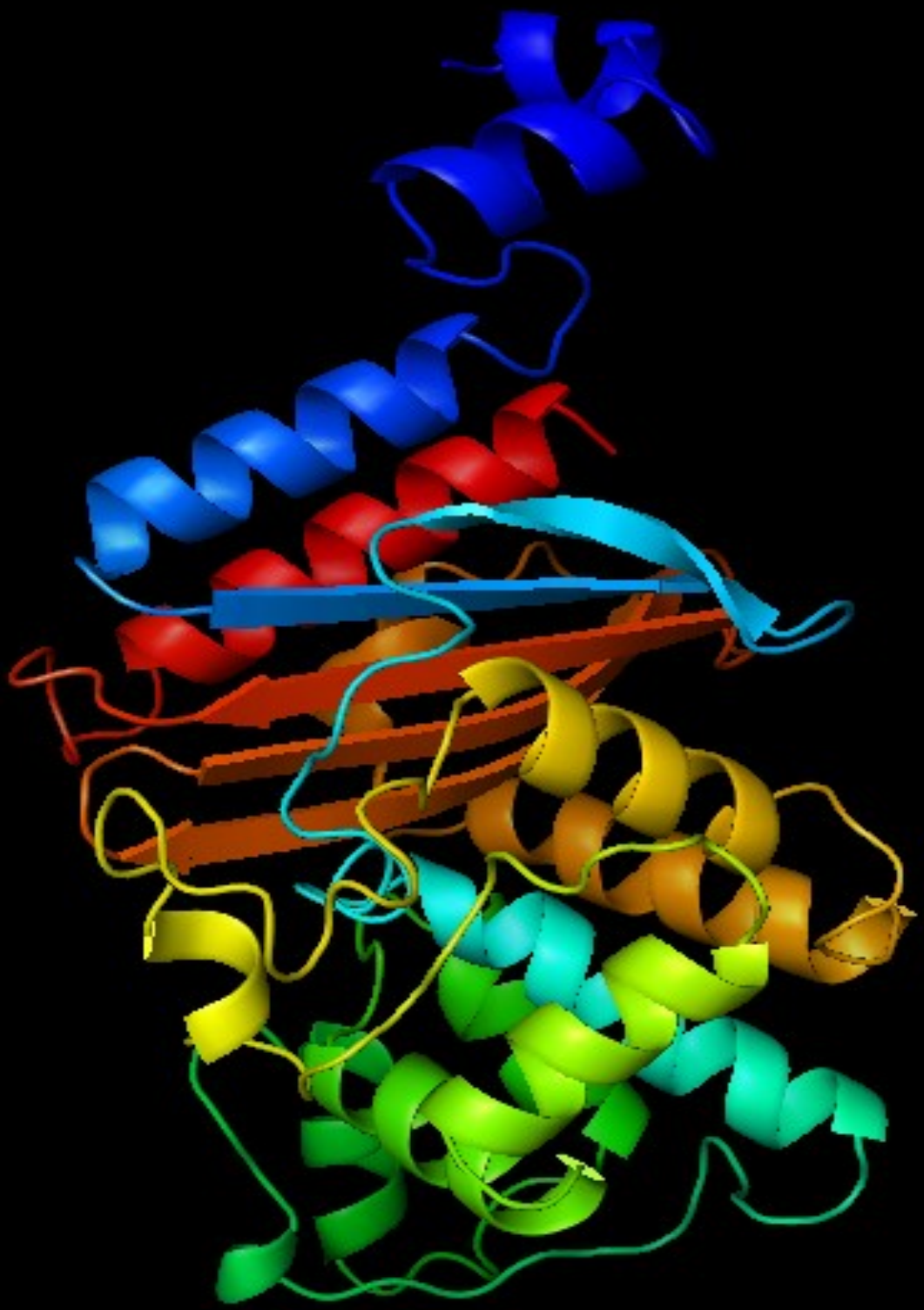
METAGENOMIC APPROACH TO SCREEN PROTEASE ENCODING GENES FROM COASTAL ENVIRONMENT OF GOA



KASHIF SHAMIM

Metagenomic approach to screen protease encoding genes from coastal environment of Goa

Thesis submitted to Goa University

For the award of the degree of

DOCTOR OF PHILOSOPHY

In

MICROBIOLOGY

By

Kashif Shamim

Under the guidance of

Professor Santosh Kumar Dubey

Department of Microbiology,

Goa University, Goa, India

2017

DEDICATED TO MY FAMILY

AND

FRIENDS

Acknowledgment

Completion of this doctoral degree thesis was possible with the support of several people. I would like to express my sincere gratitude to all of them who made this thesis possible and an unforgettable experience for me.

In the first place I would like to record my gratitude to ***Prof. Santosh Kumar Dubey****, Professor, Department of Microbiology for his valuable guidance, scholarly inputs and consistent encouragement I received throughout the research work. This feat was possible only because of the unconditional support provided by Sir. A person with an amicable and positive disposition, Sir has always made himself available to clarify my doubts despite his busy schedules. I consider it as a great opportunity to do my doctoral degree programme under his guidance and to learn from his research expertise. Thank you Sir, for all your help and support.*

I appreciate the advice of the members of the faculty research committee ***Prof. S. Ghadi****, V.C's nominee, for his meticulous appraisal and timely approvals. I thank* ***Prof. M.K, Janarthanam****, Dean Faculty of Life Sciences and Environment, for his insight, kind support and constant motivation.*

I owe my deepest gratitude to ***Prof. S. Nazareth****, Head of the Department of Microbiology, for providing necessary infrastructure as well as resources to accomplish my research work and most importantly, her valuable advice and constant support. I would also like to thank other faculty members of the Department of Microbiology,* ***Prof. S. Garg, Dr. M. Naik, Dr. L. Chari, Dr. P. D'costa, Dr. V. Khodse, Dr. T. De'souza,*** *for providing me valuable advice and assistance throughout the duration of my research work.*

Words fail to express my appreciation to my best friends, ***Jaya, Deepali Rajesh and Glen.*** *I am much indebted to them for their support, sincere encouragement, care, love, understanding, precious friendship and inspiration throughout my research work.*

I would like to express my appreciation and special thanks to my lab mates, M.Sc. students and friends for their help and cooperation throughout my research work. My heartily thanks goes to ***Dr. Pramoda, Praveen, Sajiya, Diviya, Sulochana, Imran, Mohammad, Krutika, Shruti, Vikita, Kirti, Pankaj, Alok, Preethi, Sameer, Dr. Meghnath, Krishna, Satyajit, Raghu, Salim, Ashwini, Dr. Teja, Dr. Brenda, Alisha, Sanket, Neha, Dr. Sanika, Dr. Mufeeda, Sundaram, Shashikant, Avelyno, Kiran, Samantha and Arvind.***

My sincere thanks goes to ***Milind*** *from The Centre for Cellular and Molecular Biology and* ***Amit*** *from Saurashtra University for their constant help and support in conducting some of the experiments as well as data analysis.*

I would like to thank the non-teaching staff of Department of Microbiology, ***Saraswati, Budhaji, Dominic, Tanu, Afra, Rajesh, Gajanan, Narayan*** *and* ***Deepashri*** *for providing all the necessary help. Their good sense of humor and cooperation made my working very easy in the department.*

I express my warm thanks to UGC government of India, for providing financial support in terms of fellowship as Maulana Azad National fellowship for minority student and I am also grateful for the services and facilities provided by Goa University.

My utmost appreciation to almighty God. He has made available his divine intervention in a number of ways during my challenging times.

Last but not least, I would like to pay high regards to my dad, ***Shamim Ahmad*** *who has put the fundament of my learning character, showing me the joy of intellectual pursuit ever since I was a child. My mother ,* ***Zareena Shamim*** *is the one who sincerely raised me with her care and gentle love. My brother* ***Shahzad Shamim*** *as well as other family members especially,* ***Nadeem Ahmad, Rahish Azmi, Kaleem Ahmad, Ruby Ahmad*** *and* ***Waseem Ahmad*** *for their sincere encouragement and inspiration throughout my research work and lifting me uphill this phase of life.*

Kashif Shamim

LIST OF TABLES

Chapter I

Chapter III

Chapter IV

LIST OF FIGURES

Chapter IV

LIST OF ABBREVATIONS

DNA	Deoxyribonucleic acid	**TE**	Tris-EDTA
kDa	Kilodalton	**LB**	Luria-Bertani
PCR	Polymerase chain reaction	**EDTA**	Ethylenediaminetetraacetic acid
Mb	Megabyte	**kbps**	Kilo base pairs
Gb	Gegabyte	**bps**	Base pairs
NGS	Next generation sequencing	**nm**	nanometers
ORF	Open reading frame	**DMSO**	Dimethyl sulfoxide
α	alpha	**µg/mL**	Microgram per millilitre
EC	Enzyme commission	**IPTG**	Isopropyl β-D-1-thiogalactopyranoside
HIV	Human immunodeficiency virus	**DEAE**	Diethylaminoethyl
GPS	Global Positioning System	**SDS-PAGE**	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
rpm	Revolutions per minute	**PMSF**	Phenylmethylsulfonyl fluoride
mm	millimetre	**β -ME**	Beta mercaptoethanol
µL	microlitre	**DTT**	Dithiothreitol
mL	millilitre	**ANOVA**	Analysis of variance
mg	milligram	**SRA**	Sequence read archive
mM	millimolar	**NCBI**	National center for biotechnology information

TABLE OF CONTENTS

Chapter I

Introduction and review of literature

1.1 Introduction

Microorganisms are well known source of natural bioactive compounds and biomolecules which are used extensively in various biotechnological and pharmaceutical industries (Li and Qin 2005). The enormous growth of these industries has resulted in increased focus/attention to explore microorganisms with novel biological characteristics viz. anticancer, antiviral, antibacterial and anti-helminth activities. The demand of these novel biomolecules from microorganisms has increased with time but the production is limited since only 0.01 % microorganisms are cultivable in laboratory conditions (Pace et al. 1985; Aman et al. 1995; Pace 1997; Hugenholtz 2002). In order to overcome this problem a new biological technique has emerged which is referred to as metagenomics.

The term metagenomics was first coined by a group led by Jo Handelsman (Handelsman et al. 1998). It is interesting to note that this technique eliminates the need of cultivation of microorganisms in the laboratory (Chen and Pachter 2005). Metagenomics is also known as ecogenomics, community genomics and megagenomics (Chistoserdova 2010). This new technique is currently gaining a lot of attention in the scientific community and emerging as a powerful tool to explore the vast microbial diversity present in various ecosystems which otherwise was unknown/elusive. This technique is based on isolation of metagenomic DNA directly from the environmental samples viz. air, water and soil followed by cloning of DNA fragments into suitable cloning vector in order to develop metagenomic library. The metagenomic library is analysed further using either functional or sequence based screening procedure (Handelsman 2004; Cruz et al. 2010).

Functional metagenomics is based on the expression of genes from the cloned metagenomic DNA in respective host using specific substrates. This technique is advantageous to explore novel genes encoding various industrially important enzymes and other

biomolecules including antibiotics. There are several limiting factors in functional screening viz. expression of gene (s) inserted in the foreign hosts, availability of substrates and recognition of regulatory elements such as operators and promoters (Craig et al. 2010).

Sequence based metagenomic approach relies on DNA sequencing of metagenomic library followed by identification of gene (s) encoding enzymes regulating different metabolic pathways as well as the microbial community structure that defines microbial diversity involved in specific metabolic/degradative pathways. Sequence based approach requires use of computers along with their respective softwares for genome assembly in order to find out the type of organisms as well as functions of different genes in specific samples (Schmidt et al. 1991; Stein et al. 1996; Handelsman 2004).

Microbial enzymes are biocatalysts produced by microorganisms which are in great demand for various industrial usage. Microbial enzymes are preferred nowadays since they work under extremely harsh working conditions prevalent in various industrial processes as they possess unique characteristics (Fig. 1). These conditions include extremes of temperature, pH, salinity and other factors (Nigam 2013). Although majority of industries worldwide are involved in production and sale of various microbial enzymes viz. invertase, cellulase, lipase, xylanase, chitinase, esterase, amylase, amidase, dehydratase and β-lactamase but the market share of proteases is significantly high i.e. approximately 60 % of the total sale of microbial enzymes. This high demand of proteases in industries is as a result of their usage in various industries viz. detergent, leather, food, pharmaceutical, textile and agrochemical (Rao et al. 1998; Adrio and Demain 2014).

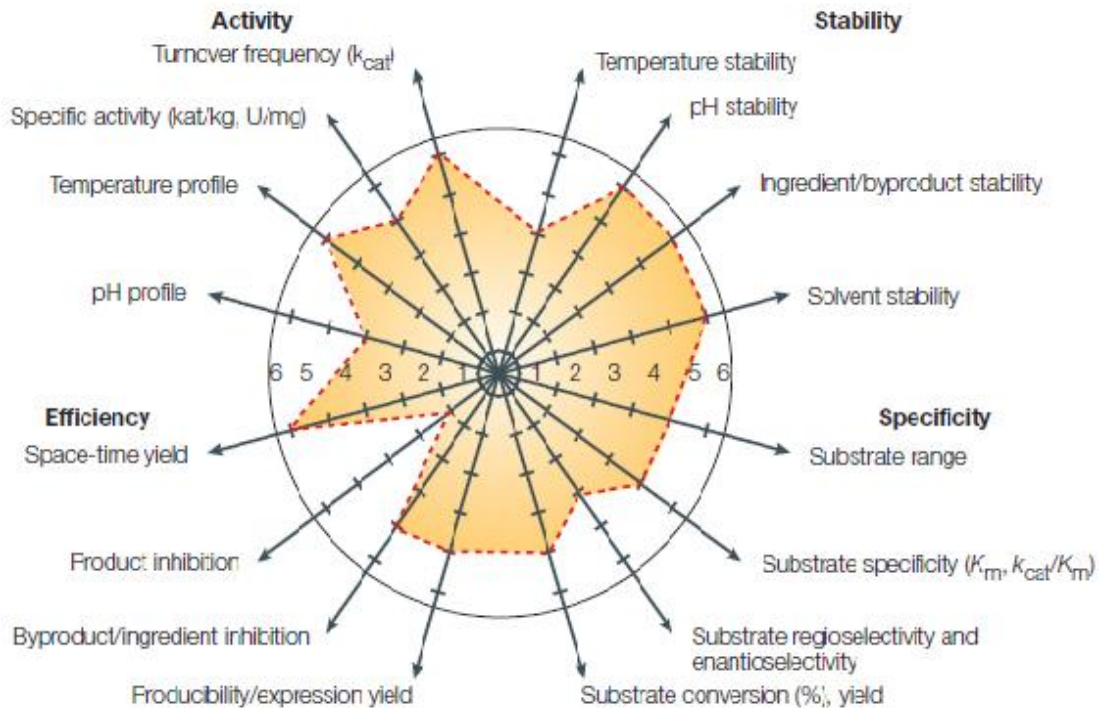


Fig. 1. Important Characteristics of an ideal biocatalyst (Lorenz and Eck 2005)

Proteases are group of enzymes that hydrolyse the peptide bonds of the proteins to form smaller peptides (Gupta et al. 2002). Proteases are highly diverse in nature which has created scientific interests among biologists to study these enzymes for applications in various pharmaceutical as well as biotechnological industries (Pohlner et al. 1986; Fox et al. 1991). Proteases have been categorized into three groups viz. neutral, acidic and alkaline on the basis of optimum pH. Similarly these enzymes have also been categorized as serine proteases (EC.3.4.21), metalloproteases (EC.3.4.24), aspartic proteases (EC.3.4.23) and sulphydryl proteases (EC.3.4.22) on the basis of functional groups present in their active sites (Kalisz 1998; Rao et al. 1998).

Alkaline proteases are one of the most important types of proteases that remain active over a range of neutral to alkaline pH conditions. These enzymes are in high demand especially in detergent industry, covering almost 30 % of the total sale of this enzyme worldwide (Gessesse 1997; Ibrahim et al. 2015). The genus "*Bacillus*" is very well known for commercial

production of microbial alkaline proteases (Varela et al. 1997). Alkaline proteases generally possess serine, aspartate and histidine residues in their active sites making a catalytic triad (Gupta et al. 2002).

Microbial ecology of terrestrial ecosystem has resulted in identification of several novel and unique bioactive compounds and biomolecules including industrially valuable proteases that has not only aided drug discovery but also enhanced the efficacy of these biomolecules. But high demand of these bioactive molecules including proteases with novel characteristics has now shifted the focus of research from terrestrial microbes to marine microbes since the marine ecosystem is poorly tapped in this regard (Li and Qin 2005). Although the marine ecosystem has got enormous microbial diversity but it has not been fully explored by the environmental microbiologists and biotechnologists to screen potential microbes possessing unique bioactive molecules including proteases (DeLong 1997).

1.2 Review of Literature

1.2.1 Metagenomics

Torsvik and Goksoyr (1980) were the pioneer researchers who gave a proof of genomic library construction after extracting and digesting genomic DNA of bacterial samples from soil. This clearly confirmed that gene library can directly be prepared from environmental DNA samples (Pace et al. 1986). Subsequent to this breakthrough, first non-PCR amplified gene library was made from marine plankton and cellulose digesters respectively (Schmidt et al. 1991; Healy et al. 1995). Among these two studies, study on cellulose digesters proved to be an important one since it showed the method for functional screening of metagenomic clones and its applications in biotechnology. Since metagenomics got popularized some aspects of this technology were patented for the first time by biotechnological companies which include Diverse (San Diego, California, USA) and TerraGen Discovery (Vancouver, Canada) in 1996

(Lorenz and Eck 2005). Cloning of PCR-generated microbial genes from soil DNA into the partial polyketide synthase gene cluster of *Streptomyces* strain has been shown for the first time by TerraGen Discovery company in order to explain the importance of cloning and expression of environmental genes (Seow et al. 1997; Wang et al. 2000).

The term metagenomics was coined by Jo Emily Handelsman and her collaborators. It is a culture independent technique to study the composition, function and diversity of microbial community in an environmental sample through direct extraction of metagenomic DNA followed by construction of metagenomic gene libraries and screening (Handelsman et al. 1998). Metagenomics became well known when two different but important research works using this technique were published in 2004. These research works clearly demonstrated application of random whole genome shotgun sequencing in understanding the microbial populations present in diverse habitats (Tyson et al. 2004; Venter et al. 2004). The research group headed by Tyson (2004) used samples from extreme environment which generated only 76 Mb DNA sequence data and resulted in complete assembly of genomes of the dominant species along with their metabolic pathway. Whereas research group of Craig Venter (2004) analysed environmental samples containing large number of species resulting into sequence database of 2 Gb. These two projects lead to flooding of information in the area of metagenomics as nearly 200 projects with over 450 different environmental samples came into existence within a very short period according to GOLD database (Liolios et al. 2010).

The development of newer technologies i.e. next generation sequencing (NGS) at the same time as that of metagenomics has proved to be a helpful tool to reduce the cost of DNA sequencing along with generation of fast and accurate DNA sequence data as compared to conventional Sanger's sequencing method (Lapidus 2009; Ansorge 2009).

Metagenomic approach to explore novel industrial enzymes from unculturable microbes has gained much attention keeping in view that besides filamentous fungi and yeast only few cultivable bacterial strains were able to withstand harsh processing and production conditions as per industrial requirements. The importance of metagenomic approach can be assessed merely on the fact that nearly one million novel open reading frames (ORFs) encoding microbial enzymes were successfully identified in a single sample of marine prokaryotic plankton obtained from Sargasso Sea (Lorenz and Eck 2005).

The development of any technique requires a demand for its continuous improvement. This demand for metagenomic technique came from its important applications in various industries especially in the field of "White biotechnology". The term White biotechnology was brought into existence in the year 2003 by the European Association for Bioindustries (EuropaBio, Belgium). They coined this term for bio-based products which are not covered under Red Biotechnology i.e. medical or Green Biotechnology (plant based industries). The White biotechnology includes a long range of products viz. vitamins, biofuels, bioplastics and enzymes used in detergent, dairy and baking products. The long term goal of the White biotechnology is to focus on replacement of commonly used fossil fuel with new renewable energy, replacing various conventional processes with bioprocesses and inventing high end bioproducts including bioactive compounds and biomolecules (Lorenz and Eck 2005). According to one of the surveys conducted by a consultancy company, Mckinsey, it was concluded that by 2010, biotechnology could be applied for production of approximately 10 to 20 % of all the chemicals sold with a value of $ 160 billion approximately.

Hence, there will be a huge demand of novel biomolecules to fulfil the urgent need of biotechnological industries that include novel biocatalysts and other bioactive compounds which has become possible through metagenomic analysis of environmental samples (Lorenz et al. 2002; Schloss and Handelsman 2003). Metagenomics along with *in vitro* evolution and

advancement in screening techniques will provide industries with an unparalleled chance to bring biological products for industrial applications (Table 1). This was achieved in case of novel α- amylase which was capable of working under hot and acidic conditions of starch liquefaction and was discovered using metagenomic approach (Richardson et al. 2002).

Table 1. Industrially important enzymes and biocatalysts from metagenomic libraries using activity based screening method (Lorenz and Eck 2005)

Function	Habitat	Library type	Average insert size (kb)	Numbers of clones Screened	Library size (Mb)	Substrate	Number of hits	Hit rate (Hit per Mb)
Esterase/lipase	Forest soil	Plasmid	8	67,000	536	Tributyrin	98	1/5.5
Esterase/lipase	Forest soil	Fosmid	40	19,968	799	Tributyrin	47	1/17
Esterase/lipase	Sandy ecosystem	Fosmid	30	29,884	903	Tributyrin	49	1/18
Esterase/Lipase	Sandy ecosystem	Fosmid	40	25,344	1,014	Tributyrin	29	1/35
Esterase/Lipase	Soil	Plasmid	6	286,000	1,716	Tributyrin	3	1/572
Esterase/Lipase	Soil	Plasmid	6	730.000	4,380	Triolein	1	1/4,380
Esterase/Lipase	Soil	BAC	27	3,648	100	Bacto Lipid	2	1/50
Oxidation of polyols	Soil	Plasmid	3	900,000	2,700	1,2-ethanediol;1,2-propanediol;2,3-butanediol	15	1/180
Alcohol oxidoreductase	Soil/enrichment	Plasmid	4	400,000	1,600	Glycerol/1,2-propanediol	10	1/160
Amidase	Soil/enrichment	Plasmid	5	193,000	965	D-phenylglycine-L-leucine	7	1/138
Amylase	Soil	Plasmid	5	80,000	400	Starch	1	1/400

Amylase	Soil/excrement enrichment	BAC	27	3,648	100	Starch	8	1/12
Biotin production	Soil/excrement enrichment	Cosmid	35	50,000	1,750	Biotin-deficient medium	7	1/250
Protease	**Soil**	**Plasmid**	**10**	**100,000**	**1,000**	**Skimmed milk**	**1**	**1/1,000**
Cellulase	Sediment/enrichment	λ phage	6	310,000	1,860	Carboxymethyl-cellulose	3	1/620
Chitinase	Seawater	λ phage	5	825,000	4,125	Methylumbelliferyl-diacetylchitobioside	11	1/375
Dehydratase	Soil/sediment enrichment	Plasmid	4	560,000	2,240	Glycerol	2	1/1,120
4-Hydroxybutyrate conversion	Soil	Plasmid	6	900,000	5,580	4-Hydroxybutyrate	5	1/1,116
B-Lactamase	Soil	Plasmid		5	400	Ampicillin	4	1/100

Marine environment has got approximately 80 % of life forms on the planet Earth and exhibits enormous biodiversity of potential microbes which can be exploited for industrial applications (Whiteman et al. 1998). The marine environment is believed to contain nearly 3.67 x 10^{30} microbes per ml including microbes at the subsurface (Kennedy et al. 2008). Therefore, marine metagenomic studies of microbial community are one of the most important data-rich areas for marine ecologists and biotechnologists. Its popularity can be assessed by the fact that it is one of the most comprehensively reviewed fields till date among the younger generation of scientific community involved in molecular microbial ecology (Schloss and Handelsman 2003, 2005; Beja 2004; Cowan et al. 2004; Falkowski and de Vargas 2004; Handelsman 2004; Riesenfeld et al. 2004; Rodriguez-Valera 2004; Delong 2005; Steele and

Streit 2005; Green and Keller 2006; Pedros-Alio 2006; Schwartz 2006; Ward 2006; Xu 2006; Edwards and Dinsdale 2007; Karl 2007; Kennedy et al. 2007; Schmeisser et al. 2007; Warnecke and Hugenholtz 2007; Marco 2008; Kennedy et al. 2008; Sleator et al. 2008; Langridge 2009; Singh et al. 2009; Steele et al. 2009; Hugenholtz and Tyson 2009; Wooley et al. 2010; Neveu et al. 2011; Pushpam et al. 2011; Biver et al. 2013; Purohit and Singh 2013). The extreme environmental conditions prevailing in the ocean with respect to pressure, temperature, salinity and nutrient availability facilitates the marine microbial community in developing novel physiological, biochemical characteristics and competence for their survival. These unique characteristics bring about the uniqueness in the type of primary or secondary metabolites they produce and might remain absent in case of microbes from terrestrial ecosystems (Li and Qin 2005). As a result more than 15,000 structurally different natural products with novel bioactivity have been discovered from marine microbes (Salomon et al. 2004). Therefore, it is logical to say that marine metagenomic approach holds a great economical potential to explore the microbial diversity for commercially valuable bioactive molecules (Morrissey et al. 2010). Moreover, with the development of newer and easier technologies the limitations of this approach will be overcome making it a promising tool for the discovery of novel bioactive molecules and compounds valuable for biotechnological and pharmaceutical industries.

1.2.2 Proteases

Proteases are type of enzymes that catalyse hydrolysis of proteins. These enzymes have gained lots of importance not only due to their vital role in cellular metabolic processes but also due to their extensive use in several industrial processes. The production of proteases around the world is more than that of any other enzyme on a commercial scale (Horikoshi et al. 2008, 2010).

Proteases are mostly present in all living organisms starting from plants to animals including microorganisms. But inability of plant as well as animal proteases to fulfil the current demand of proteases had led to an additional demand from microbial proteases. This can easily be assessed by the fact that almost two third of the commercial protease production is from microbial sources (Horikoshi 2010).

Major work in the field of microbial proteases focuses on the process of fermentation, their source and the functional role in nature. Many proteases from different microbial sources have been isolated and characterized along with their commercial applications (Dodia et al. 2008; Joshi et al. 2008; Zhang et al. 2008; Thumar and Singh 2009; Manikanda et al. 2009; Ni et al. 2009; Toyokawa et al. 2010). It has been found that many microbial proteases are capable of withstanding not only high temperature and pH but also exhibit resistance to various harsh detergents and denaturing agents (Dodia et al. 2008; Rasch et al. 2010; Manabe et al. 2010).

Proteases are of various types and according to nomenclature committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB) proteases come under the subgroup EC 3.4 of a group EC 3 (Hydrolases). On the basis of site of action, proteases are divided into two major groups i.e. exopeptidases (enzymes that break the bonds from end pieces of terminal amino acid) and endopeptidases (enzymes that break the bonds of non-terminal amino acids). They are further classified according to the groups they contain on their active site i.e. serine, aspartic, cysteine and metalloproteases (Hartley 1960). Recently proteases have also been classified into two groups based on functionality i.e. reaction they catalyse (Table 2) and molecular structure as well as sequence homology by MERPOS (Table 2).

Table 2. Classification of proteases based on functionality

Peptidase	Type	Description	NC-IUBMB
Amino-	Exo	Cleaves one aa from N- term.	EC 3.4.11
Dipeptidyl-	Exo	Cleaves two aa from N- term.	EC 3.4.14
Tripeptidyl-	Exo	Cleaves three aa from N- term.	EC 3.4.14
Carboxy-	Exo	Cleaves one aa from C- term.	EC 3. 4. 16-18
Peptidyldi-	Exo	Cleaves two aa from C- term.	EC 3.4.15
Di-	Exo	Cleaves dipeptides	EC 3.4.13
Endo-	Endo	Cleaves internal α- peptide bonds	EC 3.4.21-25
Oligo-	Endo	Endo-peptidase that only acts on peptides	EC 3.4.21-25

Table 3. Proteases are divided into 7 major protein families and 196 subfamilies based on molecular structure and sequence homology using MEROPS

Catalytic type (Major protein families)	MEROPS families
Serine	45
Cysteine	65
Metallo	59
Aspartic	14
Glutamic	2
Threonine	4
Unknown	7

(Rawlings et al. 2016).

1.2.3 Major groups of proteases

a) Metalloproteases

Metalloproteases are those proteases which require divalent metal cations for their optimum activity. They are highly diverse with respect to their catalytic activity (Manni et al. 2008; Sorror et al. 2009). According to MEROPS database metalloproteases are classified into 46 families and these families are grouped together into 10 clans depending upon the metal cation binding motifs and similarities in 3D-structures (Barrett et al. 1998).

b) Aspartic proteases

These proteases are also known as acidic proteases and contain aspartic acid residues in the active site. Some of the common pharmaceutical aspartic proteases include Renin (Chymosin), Plasmempsins (anti-malarial), Beta-secratase and HIV Protease. Their optimal activity is in the acidic pH range of 3 to 4 and they generally get inhibited by Pepstatin i.e. naturally occurring hexapeptide containing unusual amino acid statin.

c) Cysteine proteases

Cysteine proteases are also commonly known as Thiol proteases. According to MEROPS classification system cysteine proteases consist of 14 superfamilies and several others un-assigned families. These proteases have been found in a wide range of organisms viz. fungi (cathepsin B in *Aspergillus flavus*), bacteria (clostripain in *Clostridium histolyticum*), protozoa (Crizipain in *Trypanosoma cruzi*) and plants (papain and chymopapain in *Carica papaya*). There are two major groups of cysteine proteases in case of mammals i.e. lysosomal cathepsins and cytosolic calpains (Otto and Schirmeister 1997; McGrath 1999).

d) Serine proteases

Serine proteases almost share one third of all known proteolytic enzymes. Serine proteases are grouped into 13 clans and 40 families on the basis of MEROPS data base. In biological processes involvement of serine proteases are enormous (Stround 1974). These enzymes help

viruses in host infection as seen in hepatitis C and also facilitate bacteria in digesting different materials as seen in case of flesh eating *Staphylococcus* sp.. They are also responsible for regulation of development of organisms as well as degradation of flesh after snake bite (Rose and Di cera 2000).

1.2.4 Applications of alkaline proteases

The alkaline proteases attract majority of industries due to their stability in harsh industrial environment (Horikoshi 2008). Therefore, they are used by several biotechnological and pharmaceutical industries. The major alkaline protease based industries are as follows:

a) Detergent industry

The alkaline proteases got economic importance when alkaline proteases of bacterial origin i.e. *Bacillus* sp. were introduced in several detergent industries in 1960s. Interestingly it covered almost 35 % of the total sale of microbial enzymes (Godfrey and West 1985; Outtrup 1990; Kaliz 1998). At present the share of detergent enzymes is almost 89 % of the total sale of the protease enzyme in the world and among the proteases, alkaline proteases dominate the market share mainly from *Bacillus* sp. Industries such as Genencor International in the United States, Gist-Brocades in the Netherlands, Showa-Kenko in Japan and Solvay in Belgium dominate the market worldwide with a range of enzymatic products in the form of powder as well as liquid. Today, Novo Nordisk after acquiring Showa-Denko's detergent enzyme business and Genencor International after acquiring Gist-Brocades share almost 95 % of worldwide market of proteases (Gupta et al. 2002).

b) Leather industry

The major problem with any industry is the issue of environmental safety. Leather processing requires chemicals such as hydrogen sulphide along with other toxic chemicals which are of serious environmental concerns. Therefore, use of biocatalysts is a preferred

choice since it also proved to be advantageous in many ways viz. speed of leather processing, eco-friendly nature and ease of controlling the process (Andersen 1998). Currently alkaline proteases from microbial sources are extensively used in dehairing, bating and soaking processes of skin and hides. Alkaline proteases are more useful for dehairing since alkaline condition optimizes the swelling of hair roots and subsequently protease action on hair follicles leads to easy removal of hair from the skin of the animal (George et al. 1995; Hameed et al. 1996, 1999; Varela et al. 1997).

c) Photography industries

Alkaline proteases are very well known for their use in the silver recovery from X-ray photographic films. The silver is generally present 1 to 2 % by weight in the X ray films and it's recovery using alkaline proteases will be an economically viable option. The advantage of using enzyme hydrolysis process is not only limited to recovery of silver but it also makes polyester film base recyclable. Alkaline proteases in the past have been reported for several purposes. The alkaline protease from *B. subtilis* recovered silver and decomposed gelatin within 30 min (Fujiwara et al. 1989). Similarly alkaline proteases from *Bacillus* sp. B21-2 (Ishikawa et al. 1993), *Bacillus* sp. B18 (Fujiwara et al. 1991) and *B. coagulans* PB-77 (Gajju et al. 1996) have also been reported for silver recovery from X-ray films.

d) Medical Usage

The alkaline proteases from *Bacillus* sp. CK11-4 have been reported to exhibit thrombolytic activity (Kim et al. 1996) and protease from *B. subtilis* 316M was used in preparation of elastoesterase that can be applied for carbuncles, purulent wounds, burns, deep abscesses and furuncles (Kudrya and Simonenko 1994).

e) Silk degumming

This has been one of the important areas of protease applications, but unfortunately not much has been explored. This can be confirmed by the fact that only few patents have been filed in this regard (Kanehisa 2000). The smoothness of texture of silk fibres can be brought about by removal of sericin, which constitutes approximately 25 % of the total weight of the silk fibres. The conventional method of removing sericin follows shrink-proofing and twit-setting for the silk yarns with the help of starch but this process is quite expensive (Kanehisa 2000). Silk degumming using alkaline protease from *Bacillus* sp. RGR-14 has already been reported (Puri 2001).

f) Food and feed industries

Microbial proteases have also been used in food and feed industries for various purposes viz. fortification of soft drinks and fruit juices, preparation of infant dietary and specific therapeutic products of high nutritional value along with protein hydrolysates (Ward 1985; Neklyudov et al. 2000). The fish hydrolysates possessing high nutritional value were obtained using proteases from *B. subtilis* (Rebeca et al. 1991). The protein hydrolysate of sardine muscles were prepared using alkaline protease of *B. licheniformis* which showed inhibitory effects against angiotensin-I converting enzyme (ACE) consequently preventing cardiac arrest in mammals (Matsui et al. (1993). Similarly soy protein and whey protein hydrolysates were produced using alkaline proteases (Fujmaki et al. 1970; Perea et al. 1993). Alkaline protease with keratinolytic activity obtained from *B. subtilis* and *B. licheniformis* PWD-1 was used in the production of proteinaceous fodder from keratin-containing materials or waste feathers (Dalev 1994; Cheng et al. 1995).

g) Waste management

The ability of proteases to hydrolyse the proteinaceous waste helps in lowering the biological oxygen demand of aquatic ecosystem. This kind of treatment is helpful in food processing industries as well as household waste management including clearing of drainage system. This characteristic of protease has already been reported in *B. subtilis* against feather waste from poultry industries (Dalev 1994). Similarly use of keratinolytic protease against keratinous materials from poultry waste was also reported by Ichida et al. (2001) and as a depilatory agent in removal of hair from the bath room drains (Takami et al. 1992).

1.2.5 Proteases employing metagenomic approach

Proteases from various econiches using metagenomic approach have been reported with their novel characteristics. The novel fibrinolytic metalloprotease from deep-sea sediment samples was reported by Lee et al. (2007), while novel domain metalloproteases from mining shaft and compost soil sample were also isolated and characterized (Waschkowitz et al. 2009). Similarly, an alkaline serine protease was characterized from the skin of goat by Pushpam et al. (2011). Two serine proteases from Gobi desert and Death valley (Neveu et al. 2011) as well as one novel mesophilic protease from Antarctic coastal sediment sample have also been reported (Zhang et al. 2011). Biver et al. (2013) reported oxidant-stable serine protease from forest soil sample whereas alkaline protease was reported from saline habitat using metagenomic approach (Purohit and Singh 2013).

Although many proteases have been reported in the past using culture-dependent as well as metagenomic approach but very few proteases have been isolated from marine ecosystem using metagenomic approach. Therefore, keeping in view the above facts and findings, I proposed following objectives for my study:

Objectives

- Isolation of metagenomic DNA from estuarine samples of different coastal sites of Goa.
- Construction of metagenomic clone library of samples from different coastal sites of Goa using suitable cloning vector (plasmid) and *Escherichia coli* as a host following standard molecular biological techniques (Sambrook et al. 1989).
- Screening of metagenomic clone library for presence of protease using enzyme substrate (Functional screening).
- Isolation and DNA sequencing of gene encoding protease from positive metagenomic clone using standard molecular techniques.

Chapter II

Material and methods

2.1 Collection of marine sediment samples and determination of physicochemical characteristics

Marine sediment samples were collected from three different locations across Goa, India in a sterile ziplock bag and stored inside the ice box until needed for metagenomic DNA extraction. The sample codes were assigned to the samples and GPS locations were also recorded. The pH of the collected samples was recorded by resuspending one gram of sediment samples with 10 mL sterile Milli-Q water. In order to obtain the total organic content (TOC), the sediment samples were sieved through 2 mm mesh to remove the larger particulate matter followed by complete drying at 100 ^{o}C for two days and finally at 500 ^{o}C for 3 hours.

2.2 Metagenomic DNA extraction

Marine sediment samples (400 mg) were suspended in 875 µL of 0.12 M sodium phosphate buffer (pH 8.0) followed by addition of 125 µL of 20 % SDS in 2 mL eppendorf tubes which were mixed by inversion several times. Lysozyme (100 µL) from stock solution (10 mg/mL) was added to each sample, followed by vortexing of eppendorf tubes for 30 sec. Subsequently, eppendorf tubes were incubated at 37 ^{o}C for 30 min and mixed by inversion after every 5 mins. Samples were subjected to centrifugation at 13,000 rpm and 4 ^{o}C for 10 min and the supernatant was transferred to a clean eppendorf tube to which 250 µL of ammonium sulphate (2.5 M) was added. The samples were mixed by inversion several times to remove protein contamination. Samples were centrifuged at 13,000 rpm at 4 ^{o}C for 10 min and supernatant was transferred to a clean eppendorf tube again. The supernatant was mixed with equal volume of silica gel in presence of 3 M guanidine hydrochloride (pH 7.5) and incubated at 25 ^{o}C for 10 min. The mixture of supernatant and silica gel slurry was applied to the spin column (MP Biomedicals, USA) and centrifuged at 13,000 rpm at 4 ^{o}C for 10 min. Metagenomic DNA bound to the column was eluted in 1.5 mL clean eppendorf tube using 50

µL sterile Milli-Q water. Eluted metagenomic DNA was further purified by mixing with equal volume of chloroform and isoamyl alcohol (24:1). Aqueous phase was recovered after centrifugation at 13,000 rpm at 4 ^{o}C for 10 min and metagenomic DNA was precipitated by incubating aqueous phase with 0.6 volume of isopropanol at 25 ^{o}C for 1 hour. Subsequently the mixture was centrifuged at 13,000 rpm at 4 ^{o}C for 10 min and the pellet containing metagenomic DNA was washed with 70 % ethanol, vacuum dried and finally resuspended in 25 µL of TE buffer (pH 8.0) [Appendix B].

2.3 Comparative study of metagenomic DNA extraction methods

Metagenomic DNA of the environmental sediment samples were also recovered using other commonly used methods to compare the performance of our method in terms of DNA yield and purity of DNA with the help of nanodrop 2000c (Thermo Scientific, USA), duration of the protocol and type of chemicals as well as solvents used (Zhou et al. 1996; Yeates et al. 1997; Burgmann et al. 2001; Amorim et al. 2008). The extracted metagenomic DNA samples (10 µL) were electrophoresed on 0.8 % agarose gel [Appendix B].

2.4 Construction of metagenomic library

2.4.1 Bacterial strains, plasmid vectors and growth conditions

The bacterial strain *Escherichia coli* JM109 (HiMedia, India) was used as a cloning host and pUC18 (Merck life science, India) as a cloning vector, whereas *E. coli* BL-21(DE3) as an expression host and pET-22b as an expression vector. Throughout the experiment host *E. coli* cells were grown in Luria-Bertani (LB) broth at 37 ^{o}C [Appendix A].

2.4.2 Restriction digestion of metagenomic DNA and plasmid vector DNA

The Metagenomic DNAs of approximately 500 ng were subjected to partial digestion with restriction endonuclease, *Sau* 3A1 at 37 ^{o}C for 1 hour and at every 10 min sample was removed followed by deactivation of enzyme at 65 ^{o}C for 10 min [Appendix C]. After completion of one hour restriction digestion, all the samples were analysed using agarose (0.8 %) gel electrophoresis. This was done to determine the time required for partial digestion of the metagenomic DNA samples. Similarly 500 ng of plasmid DNA was also digested with *Bam* HI for one hour at 37 ^{o}C and deactivated at 65 ^{o}C for 10 min. The vector DNA was dephosphorylated [Appendix C] using 0.5 U Calf intestinal phosphatase (Merck life science, India) at 37 ^{o}C for 30 min and enzyme deactivation was achieved at 65 ^{o}C for 30 min in presence of 5 mM EDTA with pH 8.0 (Sambrook et al. 1989). The plasmid DNA was electrophoresed on 0.8 % agarose gel along with 1 kbps DNA marker (NEB, India) to check the efficiency of restriction digestion. Purification of the restriction digested metagenomic DNA from the agarose gels was done with the help of GeneiPure Gel extraction kit (Genei, India).

2.4.3 Ligation of metagenomic DNA with plasmid vector

Ligation procedure was performed using 25 ng pUC 18 plasmid DNA, 75 ng of gel purified metagenomic DNA and 1 unit of T4 DNA ligase (Genei, India) in a sterile eppendorf tube following standard protocol as per Sambrook et al. 1989 [Appendix C]. The ligation mixture was incubated overnight at 16 ^{o}C to ensure ligation and it was used for transformation of *E. coli* JM109 competent cells by heat shock method (Hanahan 1983).

2.4.4 Transformation of *E. coli* JM109 with recombinant metagenomic DNA

2.4.4.1 Preparation of Competent cells

The host *E. coli* JM109 culture was grown over night in 2 mL LB broth at 37 ^{o}C. One ml of culture suspension was inoculated next day to 100 mL fresh LB broth followed by incubation in incubator-shaker at 37 ^{o}C and 225 rpm till the absorbance at 600 nm reached to approximately 0.6 which is equivalent to approximately 6 x 10^8 cells/mL. Once the desired absorbance (i.e. 0.6) was attained the bacterial cells were incubated in ice for 15 min with swirling at regular intervals. These bacterial cells were harvested by centrifugation at 6,000 rpm for 5 min in refrigerated microcentrifuge (Eppendorf, Germany) at 4 ^{o}C. The cell pellet obtained was washed with 40 ml, 0.01 M NaCl (prechilled at 4 ^{o}C) followed by resuspension in 40 ml, 0.03 M $CaCl_2$ (prechilled at 4 ^{o}C) and incubation in ice for 20 min. The cells were again harvested at 6000 rpm and the pellets obtained were resuspended finally in 3 mL 0.03 M $CaCl_2$. The host cells after treatment with NaCl and $CaCl_2$ have become competent and may be stored at 4 ^{o}C for at least one week. These competent cells can be used within one week for better transformation results.

2.4.4.2 Transformation of E. coli JM109 with recombinant plasmids

The freshly prepared competent cells (100 µL) were taken in pre-chilled falcon 2059 polypropylene tubes and 1 µL fresh β-mercaptoethanol (1:10 dilution) was added to the cells. Addition of the β-mercaptoethanol to the mixture was keeping in view that it aids in increasing the transformation efficiency by about 2 to 3 fold. Competent cells were incubated in ice for 10 min with gentle swirling of ice bucket at an interval of 2 min. After incubation, 10 µL ligation mixture was added to the competent cells mixed gently and the falcon tubes were incubated further on ice for 30 min. The competent cells were heat pulsed in water bath at 42 ^{o}C for 45 sec followed by immediate incubation on ice for 2 min. To this The transformation

mixture was spiked with 900 µL preheated (42 ^{o}C) LB broth and competent cells were revived after heat shock by keeping the cells in incubator-shaker at 37 ^{o}C for one hour at 225 rpm. After one hour incubation 100 µL of transformation mix was plated on LB agar plate containing 50 µg/mL of antibiotic, ampicillin. Competent cells without vector DNA/recombinant DNA were also plated on LB agar plate with 50 µg/mL amplicillin and without ampicillin. As a positive control competent cells were transformed with plasmid DNA in a similar manner and plated on LB agar plate with 50 µg/mL ampicillin.

2.5 Screening of metagenomic library for protease positive clones

The ampicillin resistant clones which appeared on LB agar plate with 50 µg/mL ampicillin were subsequently screened for protease positive clones by replica plating on LB agar plate containing 50 µg/mL ampicillin and 1 % skimmed milk as a substrate for protease enzyme [Appendix A]. Once the zone of clearance around the screened colonies were obtained the colonies were picked up and grown in LB broth containing 50 µg/mL ampicillin in order to isolate recombinant plasmid DNA and diagnostic restriction of the recombinant clones with *Bam* HI. The restriction digested DNA samples were analysed on 0.8 % agarose gel electrophoresis along with 1 kbps DNA ladder to determine the size of DNA insert of recombinant clones.

2.6 Sub-cloning and IPTG - induced expression of protease gene

The protease positive clones carrying DNA inserts were used for further sub-cloning in the expression vector pET-22b to study protease expression. The metagenomic DNA insert from protease positive clone was purified from the gel using GeneiPure Gel extraction kit (Genei, India) followed by its ligation into expression vector, pET-22b and transformation

using *E. coli* BL-21 (DE3) as a host. The transformants were directly plated on LB agar plate containing 50 µg/mL ampicillin and 1 % skimmed milk. The isopropyl Beta-D-thiogalactopyranoside (IPTG) ranging from 0.2 mM to 1 mM was added separately in the growth medium after 6 hours of initial growth (i.e. mid log phase) to find out optimum concentration of IPTG required for induction of protease gene expression.

2.7 PCR amplification of *apr* gene

Amplification of alkaline protease gene i.e. *apr* was performed on mastercycler nexus gradient (Eppendorf, Germany) using gene specific primers (Bach et al. 2002) [Appendix E].

FP *apr* I : 5'-TAYGGBTTCAAYTCCAAYAC-3'

RP *apr* II : 5'-VGCGATSGAMACRTTRCC-3'

The PCR reaction comprised of an initial denaturation of the metagenomic DNA at 95 ^{o}C for 5 min followed by 30 cycles of denaturation at 95 ^{o}C for 30 sec, annealing of primers at 60 ^{o}C for 1 min followed by elongation at 72 ^{o}C for 1 min and a final extension at 72 ^{o}C for 10 min. The resulting amplicon was analysed using 1 % agarose gel electrophoresis and the image was captured using Gel documentation system (Syngene G: Box, UK).

2.8 Purification of protease enzyme

Protease enzyme from the selected positive clone was purified initially by ammonium sulphate precipitation method using 30 % to 80 % saturation level of ammonium sulphate and keeping the sample on a magnetic stirrer overnight at 4 ^{o}C. The precipitate was harvested by centrifugation at 11,000 rpm for 20 min at 4 ^{o}C (Eppendorf, Germany). It was re-suspended in 1 mL of 0.1 M sodium carbonate buffer (pH 10.0) followed by dialysis against the same buffer by keeping it at 4 ^{o}C overnight under continuous stirring condition. The enzyme obtained was further purified using a DEAE cellulose column (12 cm x 2.5 cm) pre-equilibrated with 0.1 M sodium carbonate buffer (pH 10.0). The column was washed with the same buffer to remove unbound proteins and the column bound protease was eluted by applying a gradient of

increasing concentration of NaCl (0 - 1 M). The eluted fractions were checked for presence of protein by recording the absorbance at 280 nm followed by estimation of protease activity as per Kunitz assay (Kunitz 1947) using tyrosine as standard. All the fractions showing protease activity were pooled together. One unit of enzyme is defined as the amount of enzyme required to liberate 1 µg of tyrosine per minute under standard assay conditions. Protein content of the samples was estimated using Bradford assay (Bradford 1976) taking bovine serum albumin (BSA) as a standard.

2.9 Sodium dodecyl sulphate-Polyacrylamide gel electrophoresis (SDS-PAGE)

The purified and pooled protein fraction was analysed by SDS-PAGE to determine molecular weight of protease enzyme using 12 % polyacrylamide gel following standard procedures (Laemmli 1970) and using broad range protein molecular weight marker (NEB, USA). The purified enzyme sample (equivalent to approximatly 200 µg protein) was mixed with equal volume of 2X sample solubilizing buffer [Appendix D] and the mixture was incubated in boiling water bath for 5 min. The sample was subsequently cooled down to room temperature and centrifuged at 10,000 rpm at 4 ^{o}C for 2 min. The clear supernatant was loaded on the 12 % SDS-PAGE gel [Appendix D] and electrophoresis was carried out at a constant voltage of 100 volts using Mini-PROTEAN Tetra Cell (BIO-RAD, USA) in 1X Tris-glycine electrophoresis buffer [Appendix D] at 25 ^{o}C. The SDS-PAGE gel was removed after completion of electrophoresis, stained with coomassie brilliant blue R-250 and thoroughly destained using destaining solution [Appendix D]. The destained SDS-PAGE gel was visualized using gel documentation system (Syngene G:Box, UK). The molecular weight of the protease enzyme was determined through comparision with the protein molecular weight marker (NEB, India).

2.10 Zymography

Zymogram analysis was done to check protease activity using 1 % casein co-polymerized with 12 % polyacrylamide resolving gel. Gel electrophoresis was performed at a constant voltage of 100 volts using Mini-PROTEAN Tetra Cell (BIO-RAD, USA). After completion of gel electrophoresis the gel was submerged in a glass tray containing 2.5 % (V/V) triton X-100 solution and incubated on a rocker at room temperature for 45 min with gentle shaking. Triton X-100 is used to remove SDS from the gel. Triton X-100 was removed by washing the gel several times with Milli-Q water followed by flooding the gel with renaturing buffer, Tris-Cl (pH 8.0) for 1 hour at 25 ^{0}C under gentle shaking condition. The renaturing buffer was decanted after incubation and gel was incubated overnight at room temperature in the developing buffer (pH 7.5) containing Tris base, NaCl, $ZnCl_2$, $CaCl_2$ and Brij 35 [Appendix D]. The gel was subsequently stained with Coomassie brilliant blue R-250 for 4 hours followed by destaining until clear band indicating enzyme activity appeared on the gel (D'Avila-Levy et al. 2012).

2.11 Characterization of protease activity

Protease enzyme activity was determined with reference to several environmental and other factors viz. pH, temperature, metal cations, inhibitors, surfactants and common organic solvents. In order to determine optimum pH for protease activity different buffers were used viz. glycine-HCl buffer (pH 2.0, 3.0), sodium acetate buffer (pH 4.0, 5.0, 6.0), sodium phosphate buffer (pH 7.0, 8.0) and glycine-NaOH buffer (pH 9.0 - 12.8). The optimum temperature for protease activity was determined by incubating the enzyme samples at various temperatures (i.e. 30 ^{o}C to 80 ^{o}C) under standard assay conditions. In order to determine the effect of various metal cations on protease activity several metal salts viz. $CaCl_2$, $MnCl_2$,

$ZnSO_4$, $PbSO_4$, $FeSO_4$, $MgSO_4$, $NiSO_4$ and $CoCl_2$ at 1 mM and 5 mM were used. The enzyme sample (100 µL) was mixed and incubated with different metal cations (1 and 5 mM) for 1 hour prior to assay and residual enzyme activity in terms of percentage was estimated considering the control as 100 %. Similarly effect of various enzyme inhibitors viz. β-mercaptoethanol (β-ME), dithiothreitol (DTT), disodium-ethylene diamine tetra acetic acid (EDTA-Na_2) and phenyl-methyl-sulphonyl fluoride (PMSF) at 1 mM and 5 mM on protease activity was also determined. Effect of common organic solvents on enzyme activity at 5 % and 10 % levels (V/V) was also evaluated. These solvents include butanol, isopropanol, ethanol, methanol, chloroform, benzene and toluene. Enzyme sample (100 µL) was mixed with organic solvents (1 % and 10 % v/v) separately, incubated for 1 hour at optimum temperature and residual enzyme activity (%) was estimated keeping control as 100 %.

2.12 Statistical analysis

Data analysis was carried out using the statistical package, GraphPad Prism version 7.03 (GraphPad Software, La Jolla California, USA). One way ANOVA was used to test the variation between the temperatures and pH with respect to enzyme activity as well as enzyme stability. Results are considered statistically significant if $p < 0.05$.

2.13 Sequencing and analysis of metagenomic DNA insert

The metagenomic DNA insert was sequenced employing next generation sequencing (NGS) using Illumina MiSeq (Illumina, USA) through outsourcing. The DNA sequence obtained after sequencing was submitted to Sequence Read Archive (SRA) in NCBI. It was analysed using ORF finder of NCBI to search the open reading frames present in the entire metagenomic DNA sequence.

2.14 Phylogenetic analysis and homology modelling

The ORF's were subjected to BLAST analysis (Altschul et al. 1990) to determine the identity as well as homology of the DNA sequences. The ORF corresponding to protease was used to deduce the encoded protein. The encoded peptide sequence along with closely related sequences were aligned and evolutionary relatedness of the peptide was inferred using neighbour-joining method of Mega 7.0 package (Kumar et al. 2016).

The three dimensional (3D) structure of this protein was also constructed using the Swiss Model Server (Schwede et al. 2003). The constructed model was assessed using various protein model assessment tools at the Swiss Model server comprising of various local and global quality estimation parameters (Baweja et al. 2016). This model was further assessed and verified by several programs viz. PROCHECK (Laskowski et al. 1993), VERIFY3D (Eisenberg et al. 1997) and ERRAT (Chris Colovos 1993; Singh et. al. 2016) using the verification server, SAVES. This model was also evaluated using Phyre 2.0 software which predicts the α and β strands in the 3D model (Kelly et al. 2015).

2.15 Protein-protein docking studies

Protein-protein interaction studies were carried out considering gp41 coat protein (PDB ID: 3U91) of HIV-1 and cell adhesion protein of *Listeria monocytogenes* (PDB ID: 4EZG) as targets with the identified peptide as template. Patchdock server (Schneidman-Duhovny et al. 2005) was employed as a tool to study protein-protein interaction. Structures showing minimum atomic contact energies were downloaded and viewed using PyMOL software.

Recovery of metagenomic DNA suitable for PCR and metagenomic library construction remains a challenging task since humic and fulvic acids, being the major contaminants predominantly present in soil as well as sediment interfere with downstream DNA purification process (Tsai and Olson 1992; Tebbe and Vahjen 1993; Alm et al. 2000). Recovery of high molecular weight metagenomic DNA is important for metagenomic library construction to improve the possibility of retaining the full length gene cluster involved in any biosynthetic pathway. Moreover, optimization of the lysis methods also play a crucial role in obtaining good quantity of metagenomic DNA representing all taxonomic groups (Bag et al. 2016). It is also necessary to recover pure metagenomic DNA in large quantities to ensure representation of all the genomes of a particular community (Martin-Laurent et al. 2001; Lorenz and Schleper 2002; Bertrand et al. 2005). Therefore, an efficient, cost effective and rapid method is highly desirable to recover purified metagenomic DNA from different environmental samples collected from estuarine, mangrove and salt pan ecosystems. Although molecular microbial ecologists have developed several diverse methods for recovery of metagenomic DNA from soil and sediment samples, but majority of these methods are time consuming and expensive. Besides various hazardous chemicals are involved and shearing of DNA results in an extremely poor yield of metagenomic DNA making it unsuitable for restriction digestion, gene cloning, construction of metagenomic library, PCR and DNA sequencing (More et al. 1994; Zhou et al. 1996; Clegg et al. 1997; Yeates et al. 1998; Miller et al. 1999; Niemi et al. 2001; Roose-Amsaleg et al. 2001; Santosa 2001; Lakay et al. 2006; Luna et al. 2006; Sagar et al. 2014; Solomon et al. 2016). Therefore, a simple, fast, cost effective, environment friendly, high DNA yielding protocol for extraction of metagenomic DNA from environmental samples was developed.

Chapter III

Optimization of method for extraction of metagenomic DNA from estuarine samples

(Results and Discussions)

3.1 Extraction of metagenomic DNA

The sediment samples collected were designated as GUMD1, GUMD2 and GUMD3 with their GPS location as 15º24009.800N 73º54026.000E, 15º30003.500N 73º50043.200E and 15º30021.300N 73º49031.900E. The moisture content and pH of the estuarine sediment samples ranged from 59 to 65 % and pH 7.9 to 8.2 whereas total organic carbon ranged from 9 to 15 %, respectively (Table 1). Although several metagenomic DNA extraction protocols and commercial kits are in use for isolation of metagenomic DNA from environmental samples, we have achieved substantially high yield of pure metagenomic DNA using our modified protocol which has been developed by modifying the protocol of Zhou et al. (1996). The commonly used cell lysis reagents include lysozyme, SDS, Tween-80 and DMSO (Zhou et al. 1996; Kresk and Wellington 1999; Rondon et al. 2000). Alternatively mechanical disruption of cells is also done to lyse the cells which include bead beating, freeze thawing and grinding of environmental samples, but the environmental DNA gets significantly sheared making it unsuitable for any molecular biology experiments (Amorim et al. 2008). It is interesting to note that the quantity of metagenomic DNA recovered using our method was significantly high since the amount of metagenomic DNA ranged from 1185.1 to 4579.7 µg/g of sediment for various sediment samples (Fig. 2A, B; Table 4). The recovered metagenomic DNA was considerably pure as the A260/A280 ranged from 1.88 to 1.94 and A260/A230 varied from 2.21 to 2.30 confirming that our methodology ensured successful elimination of humic acid as well as fulvic acid contamination (Table 5). In contrast, earlier methods resulted in a comparatively poor metagenomic yield ranging from 0.6 to 748.6 µg/g for soil samples. The purity of the metagenomic DNA was also very poor since the absorbance ratio (A260:A280) varied between 0.8 to 1.9 (Nair et al. 2014; Sagar et al. 2014; Sharma et al. 2014; Devi et al. 2015; Bag et al. 2016; Solomon et al. 2016). Therefore, our method is highly efficient and ideal for extraction of metagenomic DNA from sediment samples.

Table 4. Physicochemical characteristics of the estuarine sediment samples

S.N.	Sample	pH	Moisture (%)	Total Organic content (%)
1	GUMD1	8.20 ± 0.05	59	9.80
2	GUMD2	7.98 ± 0.05	65	15.25
3	GUMD3	8.12 ± 0.05	62	12.10

Table 5. Yield and purity of the metagenomic DNA from different estuarine sediment samples

S.N.	Samples	DNA yield (µg/g)	DNA Purity (A_{260}/A_{280})	DNA Purity A_{260}/A_{230}
1.	GUMD1	2255.7 ± 2.1	1.89 ± 0.03	2.21 ± 0.02
2.	GUMD2	1185.1 ± 2.2	1.94 ± 0.02	2.30 ± 0.02
3.	GUMD3	4579.7 ± 2.5	1.88 ± 0.03	2.25 ± 0.03

Ratio of A_{260} : A_{230}> 2 indicates pure DNA and <2 indicates contamination of humic acids.

Ratio of A_{260} : A_{280}>1.7 indicates pure DNA

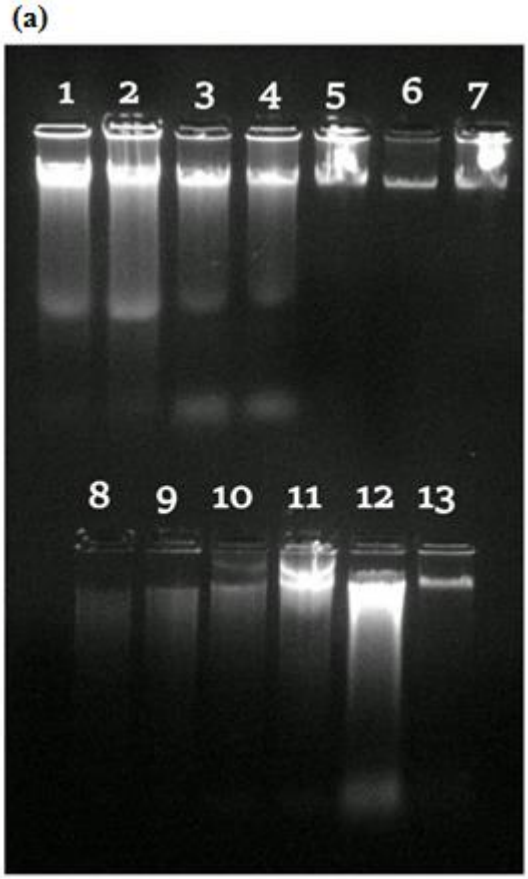


Fig. 2. (a): Agarose gel analysis of metagenomic DNA isolated from estuarine samples

Lane 1-2: DNA isolated using **our method** from sample GUMD1

Lane 3-4: DNA isolated using **our method** from sample GUMD2

Lane 5-6: DNA isolated using method of Amorim et al. (2008) from sample GUMD1 and GUMD2

Lane 7-8: DNA isolated using method of Bürgmann et al. (2001) from sample GUMD1 and GUMD2

Lane 9-10: DNA isolated using method of Yeates et al. (1997) from sample GUMD1 and GUMD2

Lane 11-12: DNA isolated using method of Zhou et al. (1996) from sample GUMD1 and GUMD2

Lane 13: Genomic DNA isolated from *E. coli* as a control

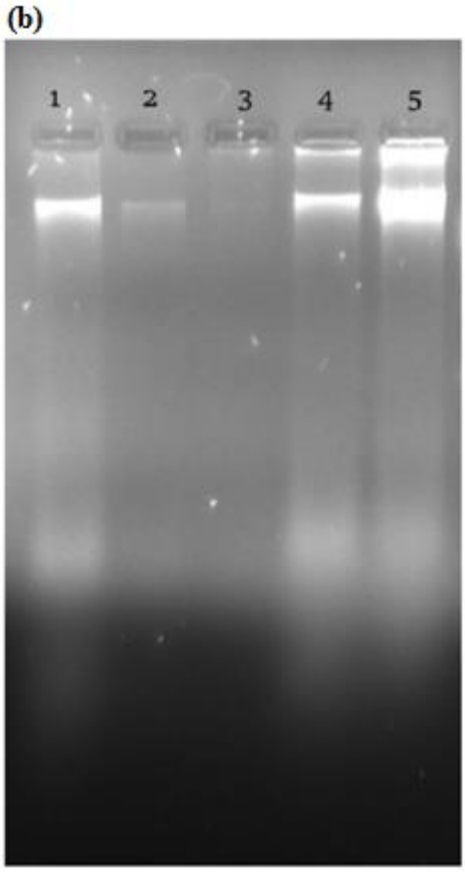


Fig. 2. (b): Agarose gel analysis of metagenomic DNA isolated from estuarine samples

Lane 1: DNA isolated using method of Amorim et al. (2008) from sample GUMD3

Lane 2: DNA isolated using the method of Bürgmann et al. (2001) from sample GUMD3

Lane 3: DNA isolated using method of Yeates et al. (1997) from sample GUMD3

Lane 4: DNA isolated using method of Zhou et al. (1996) from sample GUMD3

Lane 5: DNA isolated using **our method** from sample GUMD 3

3.2 Comparative performance of the modified method with commonly used methods

Our modified procedure for extraction of metagenomic DNA from environmental samples is comparatively simple, fast, cost effective and works well without using any toxic reagent (Table 6). We recovered metagenomic DNA from the estuarine sediments using liquid nitrogen and phenol in the range of 310 - 1000 ± 2.4 µg/g of sediment using protocol of Amorim et al. (2008). Whereas protocol of Burgmann et al. (2001) yielded metagenomic DNA in the range of 136 - 176 ± 2.3 µg/g of sediment. Similarly we could recover metagenomic DNA in the range of 15 - 23.5 ± 2.1 and 10 - 17 ± 1.8 µg/g of sediment by following the protocols of Yeates et al. (1997) and Zhou et al. (1996) respectively.

Keeping in view the comparative performance of these commonly used protocols for metagenomic DNA extraction, we may conclude that our modified method (protocol) is far more superior as the yield of metagenomic DNA was significantly high (i.e. 1185.1 - 4579.9 ± 2.2 µg/g of sediment). The absorbance ratio 260 nm : 280 nm of the recovered metagenomic DNA using our modified method was also in the range of 1.88 - 1.94 which clearly demonstrated that the metagenomic DNA was highly pure. However, this absorbance ratio ranged from 1.68 to 1.76 for the protocol of Zhou et al. (1996), 1.75 to 1.82 for the protocol of Yeates et al. (1997), 1.68 to 1.83 for the protocol of Burgmann et al. (2001) and 1.71 to 1.79 for the protocol of Amorim et al. (2008). Interestingly, our modified protocol required only lysozyme, SDS and sodium phosphate for metagenomic DNA extraction, whereas other protocols needed several expensive and few toxic chemicals viz. CTAB, proteinase K, polyethylene glycol, phenol and DTT (Zhou et al. 1996; Yeates et al. 1997; Burgmann et al. 2001; Amorim et al. 2008). The extraction of metagenomic DNA was completed in 12 hours following the protocol of Amorin et al. (2008), whereas it took 2 hours following the protocol of Burgmann et al. (2001), 5 hours using the protocol of Yeates

et al. (1997) and 3.5 hours using the protocol of Zhou et al. (1996). We could successfully obtain pure metagenomic DNA from sediment samples within 1.5 hours using our modified method (Fig. 2A, B; Table 6). Therefore, it is evident that our method for metagenomic DNA extraction is not only environment friendly and cost effective, but also less time consuming and gives high yield of metagenomic DNA with greater purity since humic and fulvic acids are eliminated.

Table 6. Comparison of various protocols used for metagenomic DNA extraction

S.N.	Method	Chemicals and reagents used	Time taken in recovery of metagenomic DNA	DNA Yield (µg/g of sediment)	Purity of DNA $A_{260} : A_{280}$
1.	Method of Amorim et al. (2008)	Liquid nitrogen, Phenol	12 hours	310 -1000 ± 2.4	1.71 – 1.79 ± 0.04
2.	Method of Burgmann et al. (2001)	CTAB, Dithiothreitol, Phenol, Polyethylene glycol	2 hours	136 - 176 ± 2.3	1.68 – 1.83 ± 0.02
3.	Method of Yeates et al. (1997)	Polyethylene glycol, Phenol	5 hours	15 - 23.5 ± 2.1	1.75 – 1.82 ± 0.05
4.	Method of Zhou et al. (1996)	CTAB, Proteinase K	3.5 hours	10 -17 ± 1.8	1.68 – 1.76 ± 0.02
3.	**Our Method/protocol (2017)**	**Lysozyme, SDS, Sodium phosphate**	**1.5 hours**	**1185.1 - 4579.7 ± 2.2**	**1.88 – 1.94 ± 0.03**

Conclusion

We have designed a new method to recover metagenomic DNA from estuarine sediment samples which is simple, fast, efficient and cost effective. Additionally, this method is environment friendly since it doesn't involve use of toxic chemicals and reagents. This protocol does not need any sophisticated equipments as required in other commonly used methods. Use of silica significantly reduced the shearing of metagenomic DNA and enhanced its purity as well as yield since proteins and humic acid contaminants commonly present in the sediments were eliminated.

It is interesting to mention that the recovered metagenomic DNA from our modified method was of good quality since it was successfully used for various molecular biology experiments viz. restriction digestion, PCR amplification and metagenomic library construction without any additional purification steps. Therefore this proves that our modified method may successfully be employed to quickly recover large quantity of pure and good quality metagenomic DNA from different environmental samples viz. sediment, sewage sludge and soil in an environment friendly manner.

Chapter IV

Construction of metagenomic library, functional screening and characterization of protease positive metagenomic clone

(Results and Discussions)

4.1 Construction and screening of metagenomic library for protease positive clone

The metagenomic DNA recovered by following the method of Shamim et al. (2017) when restriction digested (Fig. 3) and ligated with *Bam* HI cut pUC 18 cloning vector (Fig. 4) resulted in approximately 25,000 recombinant clones (Fig. 5a and b) which represented the metagenomic library in *E. coli* JM109. Since protease genes range from 0.5 to 2.8 kbps (Purohit and Singh 2013) they can be cloned and expressed easily in a plasmid vector for synthesis of a functional recombinant protein i.e. protease. Therefore, we have chosen plasmid vectors for gene cloning and expression of protease protein. Subsequently functional screening of the metagenomic library resulted in a single protease positive clone, designated as GUSK-1. This could be indicative of the difference in gene expression level in diverse taxonomical groups as per the use of different expression systems. Moreover, only 40 % of the enzymatic activity is noticeable by random shotgun cloning technique using *E. coli* as a host (Simon and Danie 2011). This protease positive clone on diagnostic restriction revealed metagenomic DNA insert of 4 kbps (Fig. 6). Sub-cloning of metagenomic insert from GUSK-1 in the expression vector pET-22b showed increased protease production for purification as well as characterization purpose. Optimum IPTG concentration was found to be 1.0 mM for maximum protease production (Fig. 7a and b). Similar studies regarding cloning and expression of protease genes have been reported from various metagenomic sources viz. oxidant-stable serine protease from forest soil (Biver et al. 2013), fibrionolytic metalloprotease from mud sample (Lee et al. 2007), two serine proteases from Gobi desert and Death valley (Neveu et al. 2011), serine protease from goat skin surface (Pushpam et al. 2011) and metalloprotease (Waschkowitz et al. 2009).Interestingly there are not many reports from marine environment in this regard.

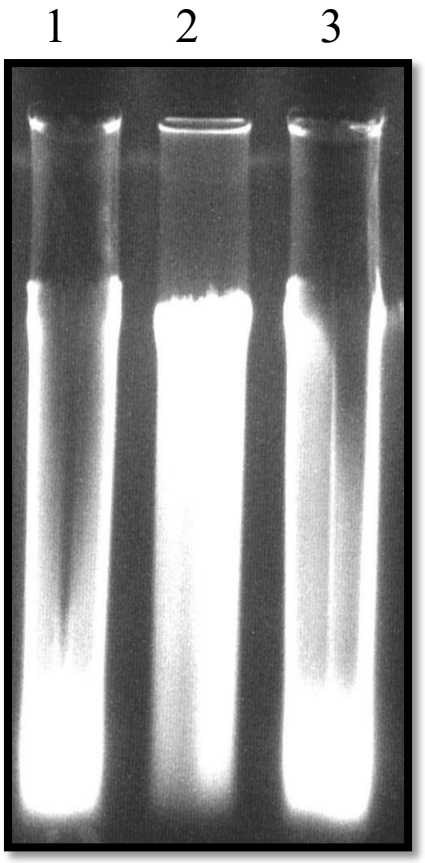
4.2 PCR amplification of alkaline protease (*apr*) gene

The *apr* gene was amplified successfully giving an amplicon of approximately 200 bps (Fig. 8). This confirms that it is alkaline protease since there was no amplification observed for neutral or acidic proteases. Further confirmation of alkalophilic nature of the protease was done using biological characterization of the enzyme.

4.3 Purification of Protease, SDS-PAGE and Zymogram analysis

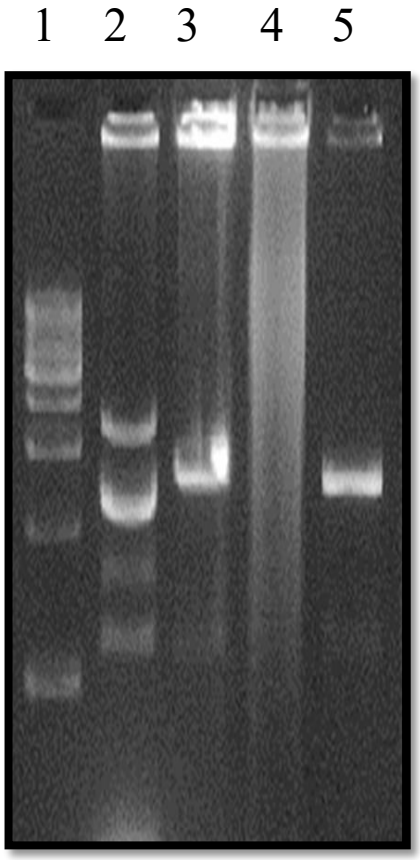
Maximum protease activity was observed at 70 % ammonium sulphate precipitation. This concentrated protein sample on further purification with DEAE cellulose column resulted in protein fractions showing high protease activity (Table 7).The fractions with high enzyme activity were pooled together to determine the molecular weight of protease on SDS- PAGE. The SDS-PAGE of purified protease enzyme showed a single band of approximately 40 kDa when compared with protein molecular weight marker (Fig. 9a). The zymogram analysis further demonstrated a single band showing zone of clearance against dark blue background confirming protease activity. This zone of clearance may be attributed to protease mediated casein hydrolysis caused due to metagenomic protease present in GUSK-1 (Fig. 9b).

The molecular weight of proteases from metagenomic sources range from 41 to 63 kDa (Lee et al. 2007; Neveu et al. 2011; Pushpam et al. 2011; Zhang et al. 2011; Biver et al. 2013) which clearly indicates the diversity of proteases as compared to proteases from culture dependent approach.



Lanes 1- 3: *Sau*3AI digested metagenomic DNA from estuarine sediments

Fig. 3. Agarose gel analysis of *Sau* 3AI digested metagenomic DNA



Lane 4 : 1kbps DNA marker
Lane 5 : Uncut pUC 18
Lane 6 & 8 : Cut pUC 18

Fig. 4. Agarose gel analysis of *Bam* HI cut pUC 18 vector

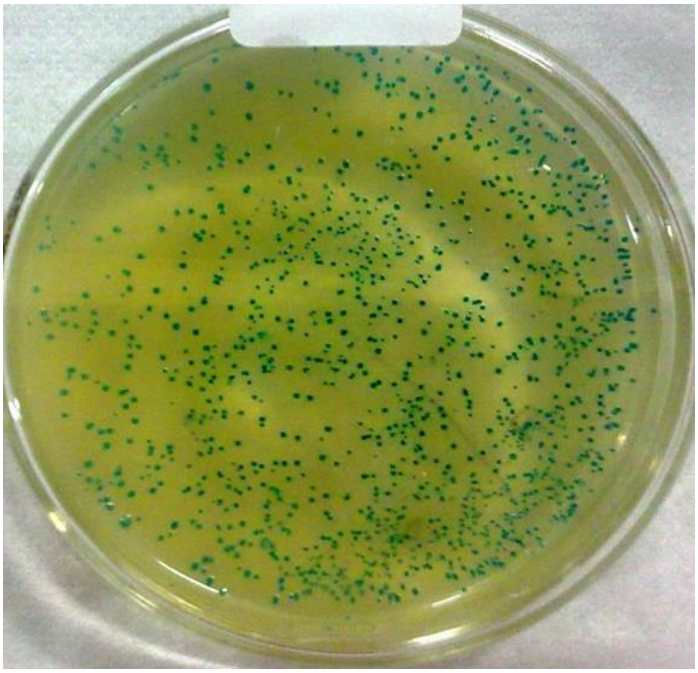


Fig. 5. (a): *E. coli* JM109 cells with pUC 18 (control) on LB agar + Amp+X-Gal+IPTG showing blue colonies

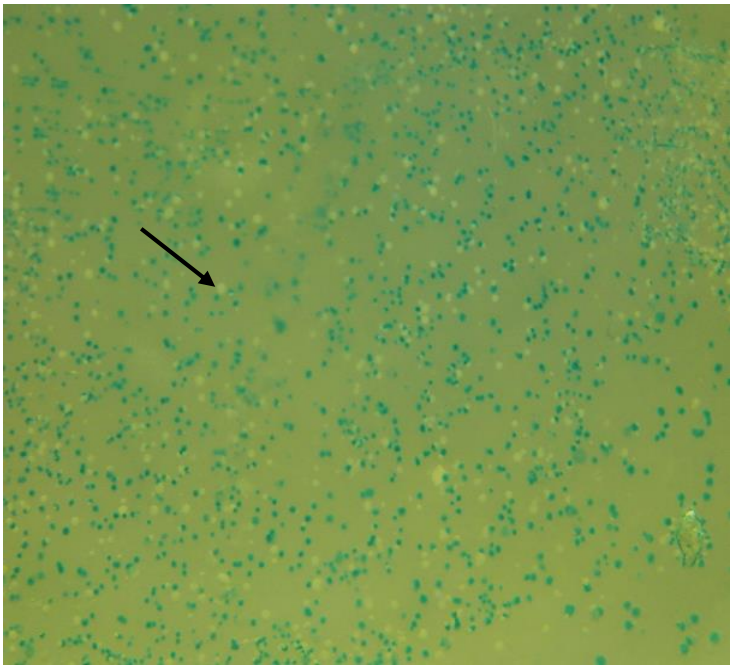


Fig. 5. (b): *E. coli* JM109 cells with recombinant plasmids on LB agar + Amp+X-Gal+IPTG showing blue & white colonies (arrow indicates recombinant colonies)

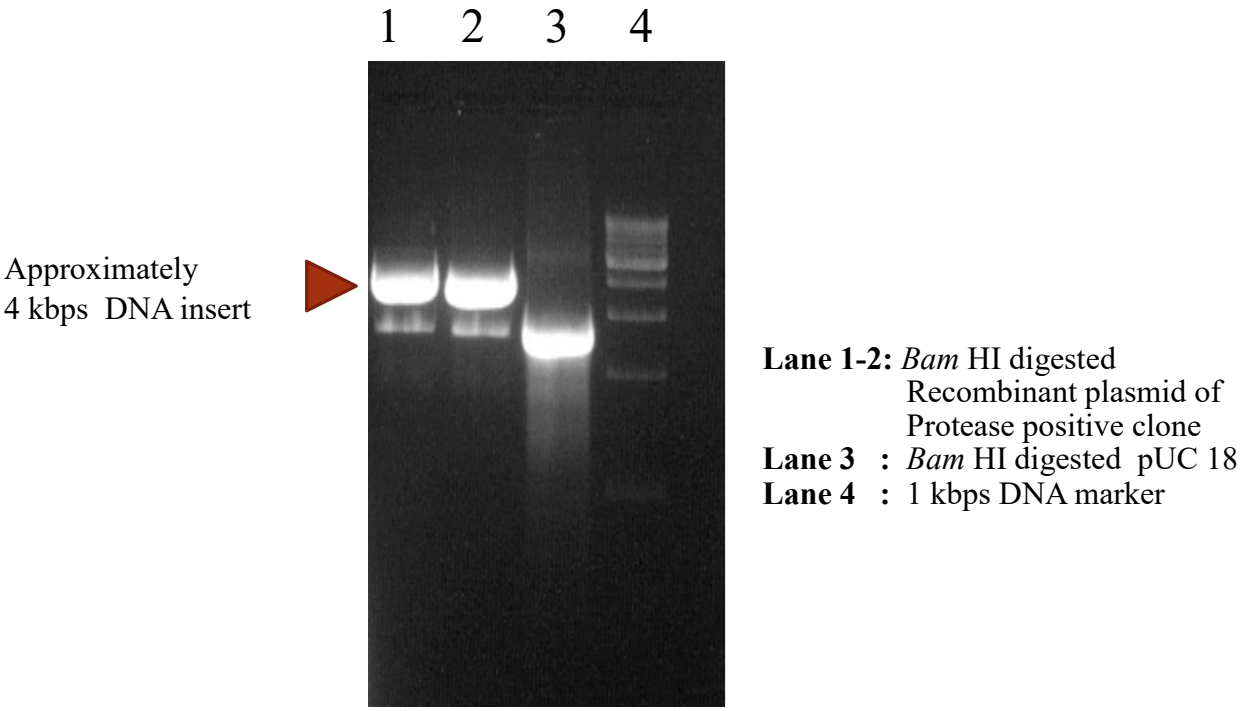


Fig. 6. Diagnostic Restriction of recombinant plasmid of protease positive clone GUSK-1

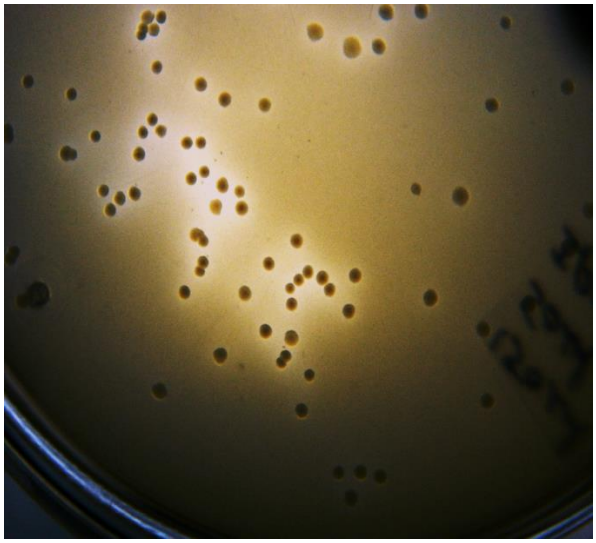


Fig. 7. (a): Protease positive clones on skimmed milk agar with ampicillin

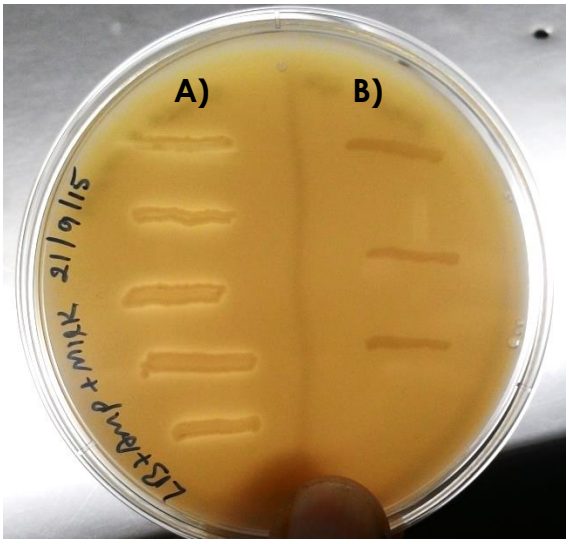


Fig. 7. (b): A) *E. coli* BL 21 with pGEX 2TK + insert

B) *E. coli* BL 21 with pGEX 2TK

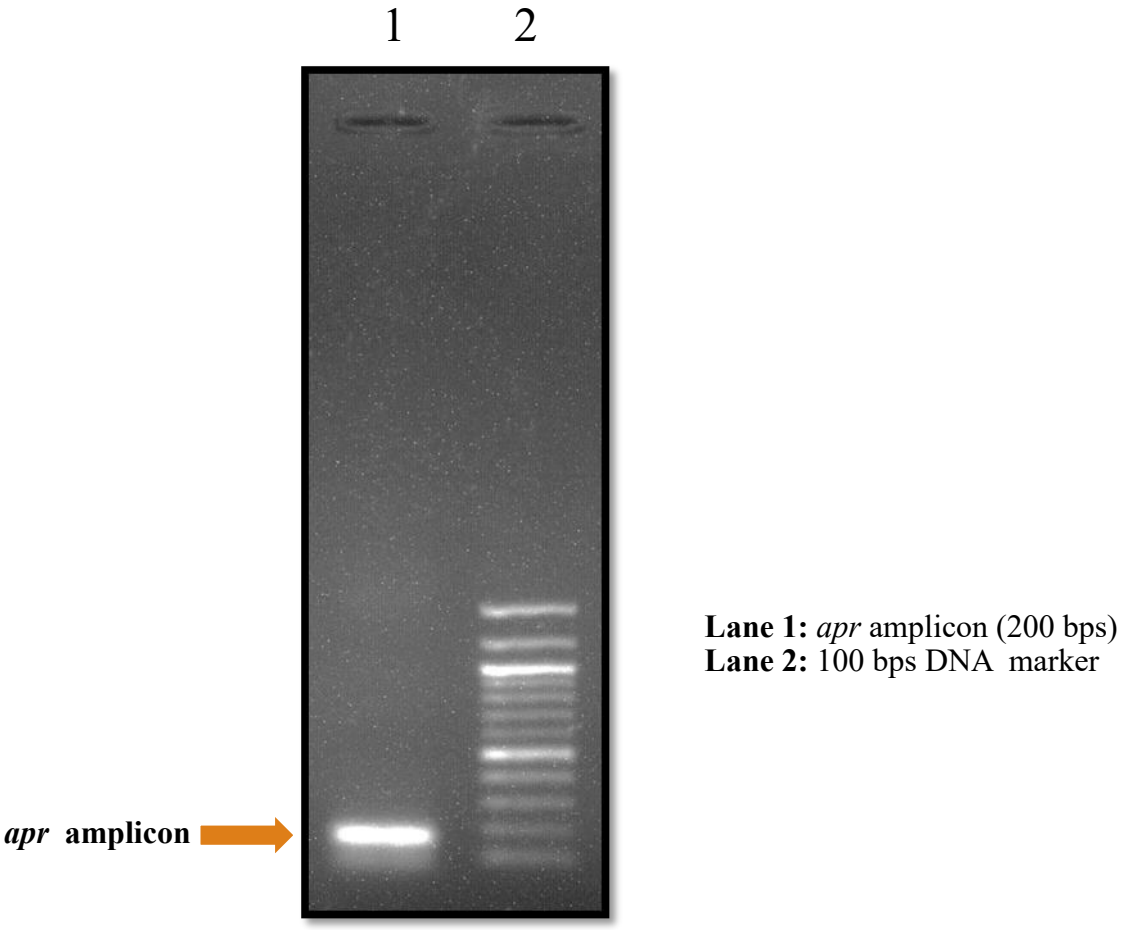


Fig. 8. PCR amplification of *apr* gene in protease positive clone GUSK-1

Table 7. Purification of Serine metalloprotease

Enzyme Sample	Total enzyme avtivity (U)	Total Protein content (mg/mL)	Specific enzyme activity (U/mg)	Purification fold
Crude enzyme	121.06	2.25	53.80	-
Purified enzyme	193.46	0.86	224.95	4.18

4.4 Biological characterization of protease enzyme from GUSK-1

Studies on the properties of any biocatalyst have been conducted over the last few decades with special focus on their industrial applications. It is important to note that major concern of industries is the stability of the enzyme under various harsh conditions. Therefore characterization of any enzyme for its maximum activity and stability are the two valuable parameters governing its commercial usage (Gupta et al. 2002).

4.4.1 Effect of temperature and pH on protease activity

The optimum temperature and pH for maximum enzyme activity of this metagenomic protease was 70 ^{o}C and 10 respectively (Fig. 10 and 11). A significant variation was observed at all temperatures and pH using one way ANOVA ($p < 0.0001$; $F= 20106$ and 32114 respectively). Interestingly, this metagenomic protease showed 100 % stability at 70 ^{o}C ($p < 0.0001$; $F= 715.9$) for one hour, however the enzyme activity declined to 78.91 % at 80 ^{o}C (Fig. 11). It also showed 100 % stability at pH 10.0 ($p < 0.0001$; $F= 2363405$) but the enzyme activity declined to 95.83 % at pH 12.0 (Fig. 13). These interesting characteristics i.e. thermostability and tolerance to high pH make it a potential candidate for industrial applications, especially in detergent and leather industries. In earlier reports, alkaline proteases from metagenomic sources have shown different temperature and pH optimas ranging from 42 ^{o}C to 55 ^{o}C with pH from 8.0 to 10.0 (Neveu et al. 2011; Pushpam et al. 2011; Zhang et al. 2011; Biver et al. 2013). Moreover, this metagenomic protease showed better characteristics than alkaline proteases reported from cultivable *Bacillus* sp. (Haddar et al. 2009; Deng et al. 2010; D'costa et al. 2013) demonstrating the metagenomic approach to be an extremely useful tool for isolation of novel and industrially valuable enzymes.

4.4.2 Effect of metal cations, organic solvents and inhibitors on enzyme activity

Metagenomic protease activity was found to be significantly enhanced in presence of most of the bivalent cations used in this study. The maximum increase in enzyme activity was observed in case of 5 mM Mg^{2+} ions i.e. by 88 % followed by Fe^{2+}, Mn^{2+} and Ca^{2+} ions i.e. 63 %, 46 % and 34 % respectively (Table 8) which suggests their possible involvement at the catalytic site of protease in addition to their role in stabilization of the enzyme (Annamalai et al. 2013). The protease activity was slightly inhibited in presence of 5 mM Pb^{2+} (3 %) and Co^{2+} (2 %). There are similar reports on enhancement of enzyme activity in presence of Ca^{2+} ions by metagenomic proteases (Lee et al. 2007; Neveu et al. 2011; Zhang et al. 2011; Biver et al. 2013) and in case of cultivable bacteria, *Bacillus* sp. JB-99 and *Bacillus cereus* SIU1 also demonstrated significant enhancement due to Ca^{2+} ions (Singh et al. 2016). However the protease activity remained unaffected in presence of Zn^{2+} and Ni^{2+} ions (5 mM) in contrary to already reported protease from metagenomic sources (Pushpam et al. 2011; Zhang et al. 2011).

The metagenomic protease also demonstrated a very high tolerance to commonly used organic solvents viz. isopropanol, chloroform, ethanol, methanol, butanol, benzene and toluene which was evident from the substantial increase in protease activity in presence of 10 % solvents (Table 9). This may be attributed to alterations in the catalytic site of protease enzyme, breakage of hydrogen bonds, hydrophobic interactions and bringing about some positive conformational changes in the protein structure (Joshi and Satyanarayana 2013). Isopropanol (1 %) has been reported to enhance the enzyme activity by 25 % for alkaline protease from metagenomic sources, whereas 2.5 % isopropanol enhanced the activity by 10 % as compared to that of control (Pushpam et al. 2011; Biver et al. 2013). Proteases from *Bacillus subtilis* strain DM-04, *Bacillus* sp. SM2014 and *Bacillus lehensis* also exhibited enhanced enzyme activity in presence of common organic solvents (Rai and Mukherjee 2009;

Jain et al. 2012; Joshi and Satyanarayana 2013). Toluene and benzene at 10 % (V/V) showed slight increase in protease enzyme activity but these solvents inhibited protease activity of *Chryseobacterium taeanense* TKU001, *Bacillus halodurans* CAS6 and *Bacillus* (Annamalai et al. 2013; Joshi and Satyanarayana 2013). This valuable characteristic of metalloprotease from GUSK-1 i.e., increase in enzyme activity in presence of solvents makes it a beneficial candidate in peptide synthesis industries.

In case of inhibitors, EDTA-Na_2 and PMSF are well known to inhibit proteases isolated from different metagenomic sources (Lee et al. 2007; Neveu et al. 2011; Zhang et al. 2011) as well as from *Bacillus lehensis* (Joshi and Satyanarayana 2013). The inhibition pattern of the isolated metagenomic protease was also in line with the above studies as significant inhibition in presence of EDTA-Na_2 (5 mM) i.e. 60 % and in presence of PMSF (5 mM) i.e. 28 % (Table 10) was observed which confirmed this enzyme to be a serine metalloprotease. Interestingly, iodoacetate (5 mM), a potent cysteine protease inhibitor also enhanced the enzyme activity proving that this metagenomic protease was not a cysteine protease (Joshi and Satyanarayana 2013). Additionally, the enzyme activity was also enhanced in presence of thiols viz. 5 mM DTT and β-mercaptoethanol indicating a possible role of these reducing agents in enzyme stability. Similar findings have also been reported in case of protease from *Bacillus subtilis* DM-04 (Rai and Mukherjee 2009).

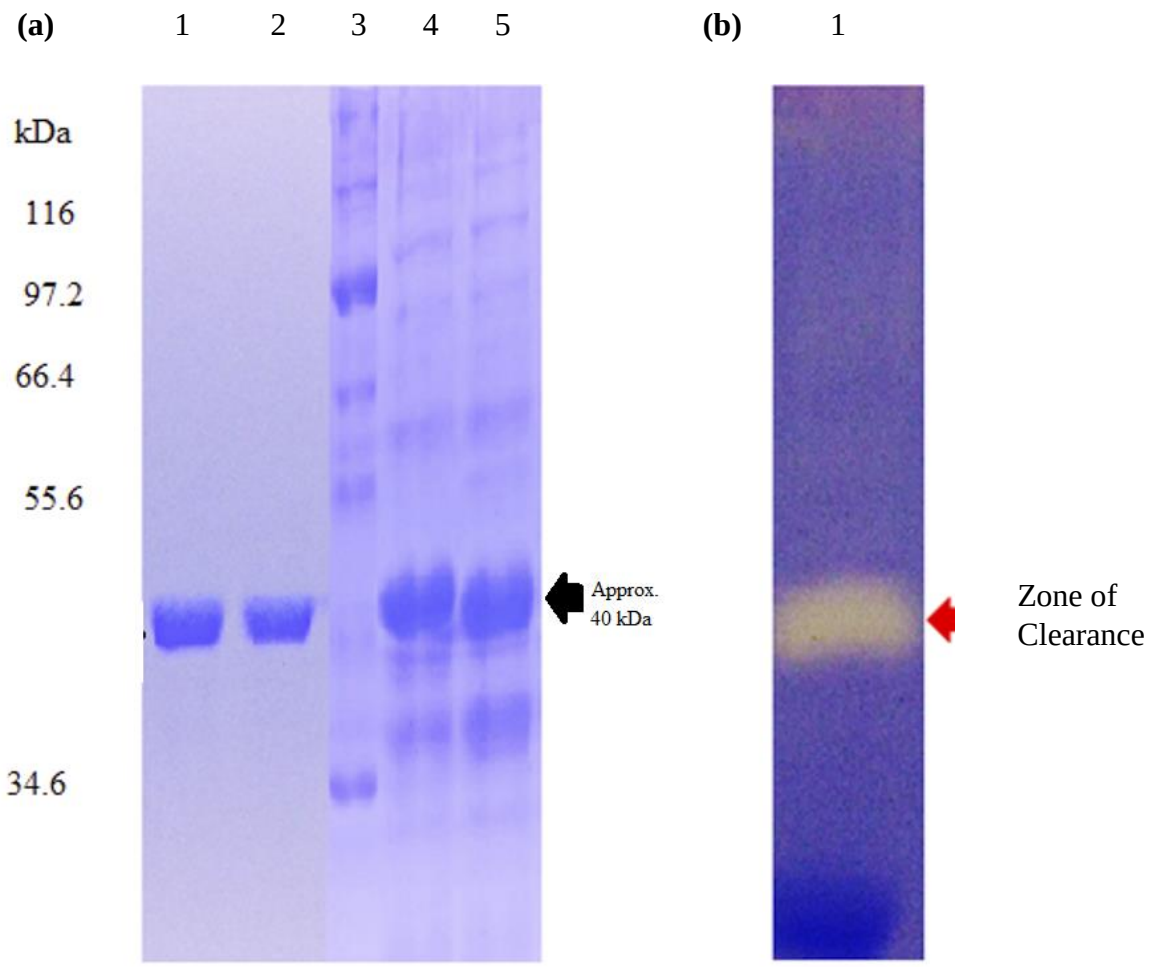


Fig. 9. (a): SDS-PAGE analysis of recombinant clone GUSK-1

Lane 1 & 2 : Column purified protease protein (40 kDa)

Lane 3 : Protein molecular weight marker

Lane 4 & 5 : IPTG induced proteins of protease positive clone

Fig. 9. (b): Zymogram analysis of clone GUSK-1 showing protease activity

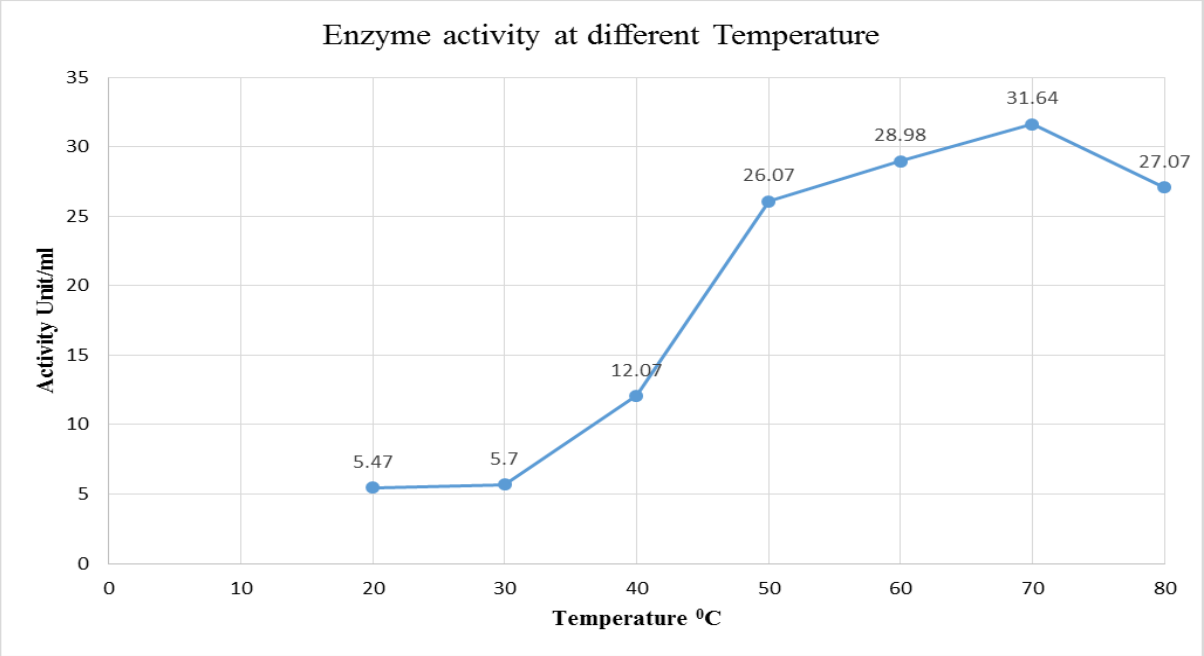


Fig. 10. Effect of temperature on protease activity ($p < 0.0001$)

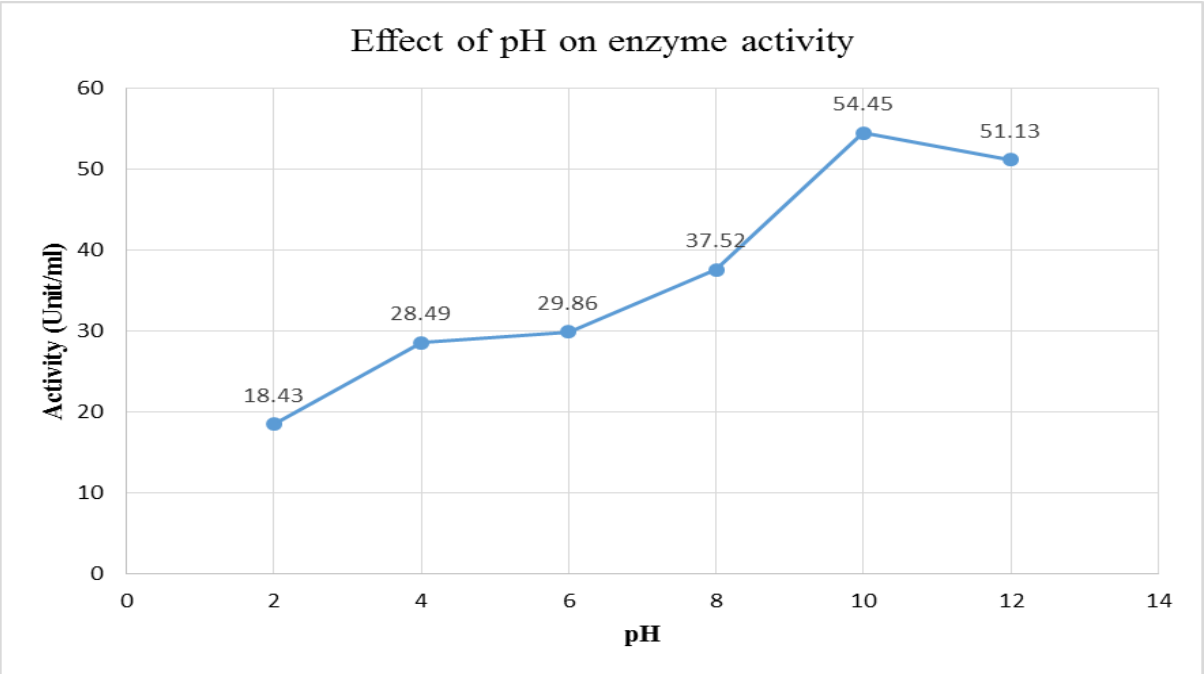


Fig. 11. Effect of pH on protease activity ($p < 0.0001$)

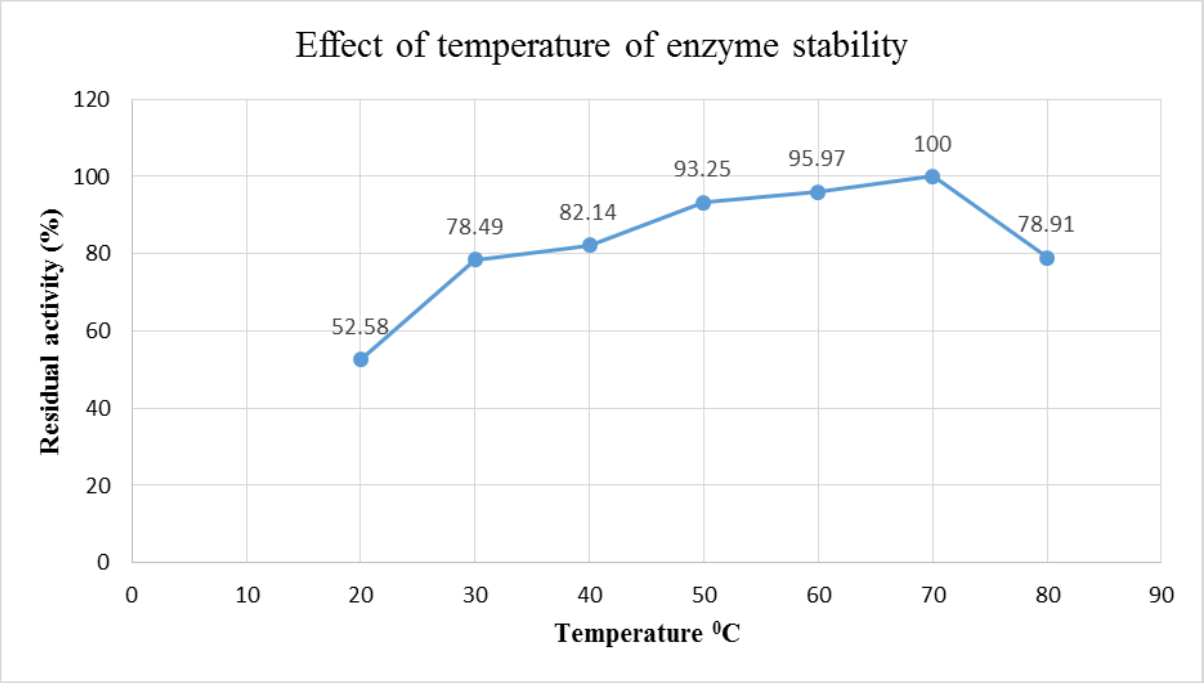


Fig. 12. Percent residual activity of protease enzyme at different temperatures ($p < 0.0001$)

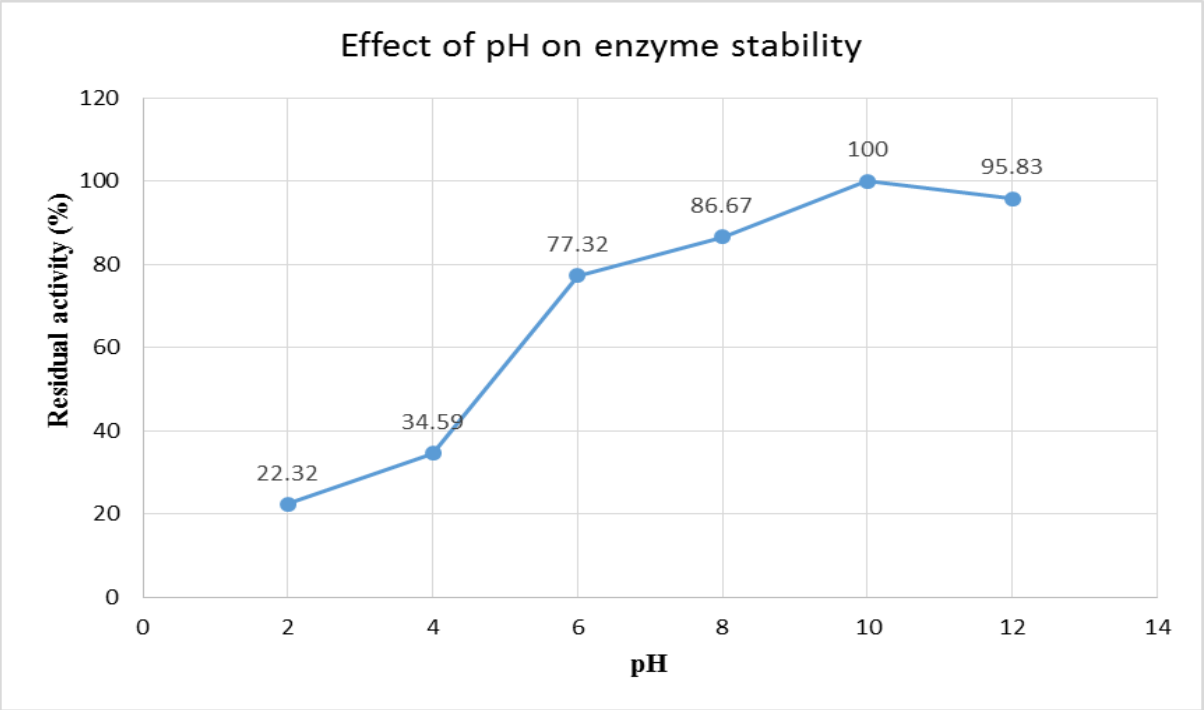


Fig. 13. Percentage residual activity of protease enzyme at different pH ($p < 0.0001$)

Table 8. Effect of various metal ions on protease activity

Metal ions	Residual enzyme activity (%)	
	1 mM	5 mM
Control	100 ± 1.33	100 ± 1.33
Ca^{2+}	126.02 ± 0.57	134.69 ± 0.33
Mg^{2+}	183.67 ± 0.88	188.88 ± 1.20
Pb^{2+}	106.89 ± 0.58	97.96 ± 0.57
Mn^{2+}	144.12 ± 1.40	146.43 ± 1.85
Zn^{2+}	130.61 ± 0.33	108.67 ± 1.33
Fe^{2+}	153.06 ± 1.92	163.78 ± 1.20
Co^{2+}	108.16 ± 1.33	98.71 ± 0.88
Ni^{2+}	110.71 ± 1.08	102.96 ± 1.66

Table 9. Effect of various solvents and inhibitors on protease activity

Solvents	Residual enzyme activity (%)	
	5 %	10 %
Control	100 ± 1.33	100 ± 1.33
Isopropanol	159.03 ± 1.15	161.6 ± 1.45
Ethanol	162.6 ± 1.33	152.4 ± 1.00
Methanol	145.4 ± 0.57	151.1 ± 0.57
Chloroform	142.7 ± 0.66	156.8 ± 0.88
Toluene	95.81 ± 0.87	103.72 ± 1.21
1-Butanol	121.86 ± 1.02	120.93 ± 1.00
Benzene	116.28 ± 1.34	114.88 ± 1.156

Table 10. Effect of various inhibitors on protease activity

Inhibitors	Residual enzyme activity (%)	
	1 mM	5 mM
Control	100 ± 1.33	100 ± 1.33
β-mercaptoethanol	104.59 ± 0.57	130.61 ± 1.18
DTT	106.12 ± 0.88	175.0 ± 0.34
PMSF	96.92 ± 1.52	78.35 ± 1.33
EDTA	60.06 ± 1.76	40.14 ± 1.20
Iodoacetate	127.55 ± 1.76	129.59 ± 1.33

Conclusion

The metagenomic library constructed using cloning vector pUC 18 and host *E. Coli* JM109 resulted in a single protease positive clone GUSK-1 after screening of approximately 25,000 metagenomic clones. This clone contained 4 kbps DNA insert encoding an alkaline protease gene which was confirmed by successful amplification of *apr* gene using gene specific primers. Protease protein of approximately 40 kDa was encoded by GUSK-1 as confirmed by SDS-PAGE analysis of the clone and its activity using zymography. This metagenomic protease demonstrated interesting biological characteristics viz. tolerance and stability at high temperature, pH, bivalent metal ions and common organic solvents. Significant inhibition of protease enzyme activity in presence of PMSF and EDTA-Na_2 confirmed it to be serine metalloprotease.

Chapter V

DNA sequencing and analysis to identify ORF encoding protease and *in silico* studies

(Results and Discussions)

5.1 DNA sequencing and sequence analysis

Next generation sequencing of the 4.0 kbps DNA insert resulted in a sequence of 4040 bps with 24 ORFs as revealed by ORF finder. The entire DNA sequence has already been submitted to NCBI (Accession No. SRP112498). Nucleotide to protein BLAST (blastx) analysis of ORFs clearly revealed that among them only ORF 4 encoded a protease which possessed 861 nucleotides and 286 amino acid residues (Fig. 14). The deduced peptide sequence showed a significant homology with a peptidase belonging to transpeptidase superfamily. Further characterization of this protease confirmed that Serine is conserved amino acid residue in the catalytic site of the protease enzyme which is one of the characteristic features of transpeptidases.

5.2 Phylogenetic analysis of serine metalloprotease

Functionally distinct proteases may show differential phylogenetic clustering therefore, phylogenetic analysis of our serine metalloprotease with functionally defined serine proteases from other microbial sources was performed in order to find out the divergence among serine proteases. Interestingly, we found that metagenomic protease showed a significant homology (i.e. 93 %) with *Shewanella baltica*, *Vibrio* sp., *Thermococcus thioreducens*, *Bacillus* sp., *Alteromonas* sp. and *Mortierella elongata* (Fig 15). The scale bar was found to be 0.2 which indicates 20 amino acid substitutions per 100 positions. Similar kind of phylogenetic analysis has also been done for serine protease from *Schistosoma mansoni* (Bos et el. 2009).

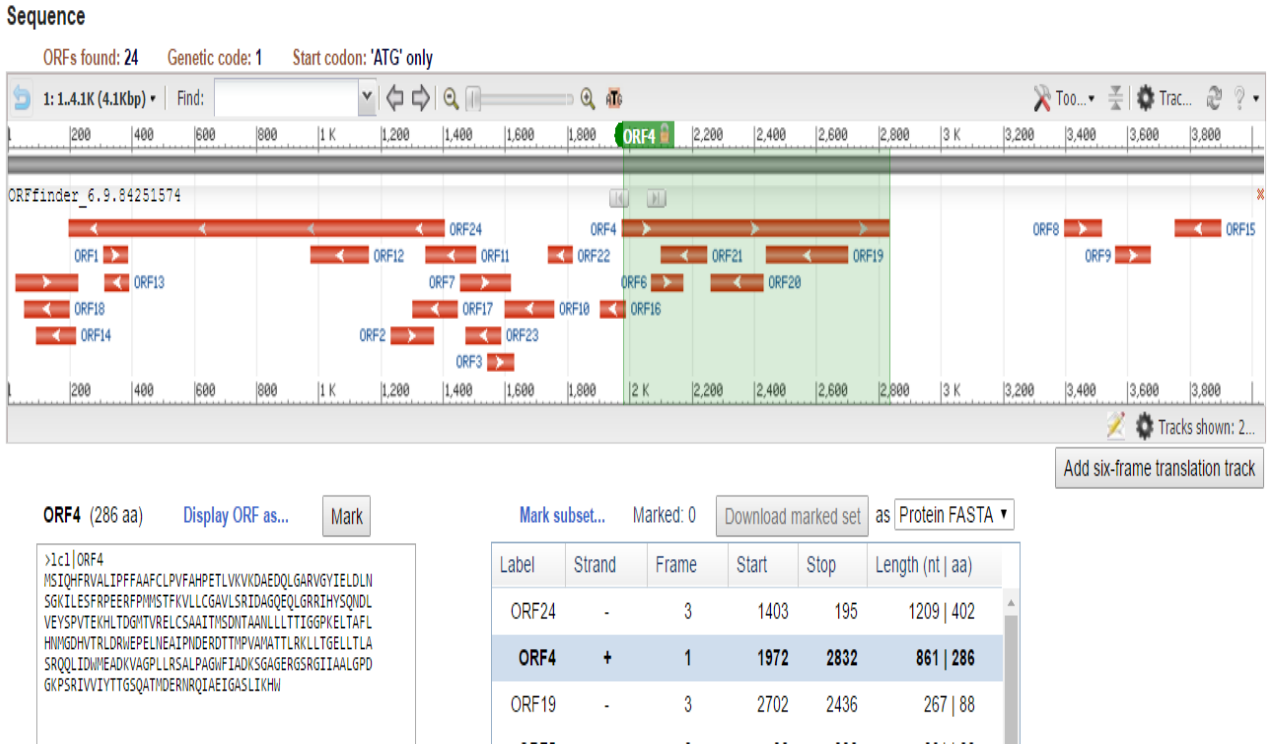


Fig. 14. Snapshot of NCBI blast showing several ORF's along with ORF4 with 861 nucleotides and 286 amino acid residues

Nucleotide sequence of ORF4 CDS

ATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCT
TCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAG
TTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTG
AGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCT
ATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCG
CATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCAT
CTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGT
GATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTA
ACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAAC
CGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAG

CAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTC

CCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCT

GCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAG

CGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTA

TCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGAC

AGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAA

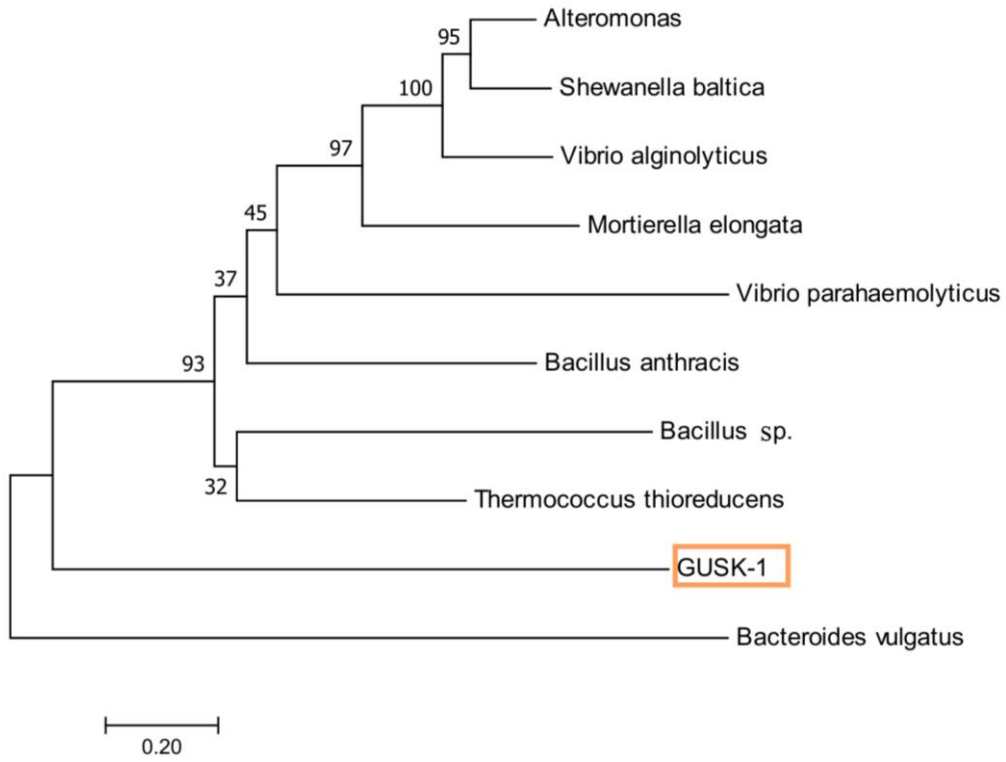


Fig. 15. The phylogenetic tree of serine metalloprotease from GUSK-1

Evolutionary relationship of the serine metalloprotease from GUSK-1 with serine proteases from other bacterial strains using Maximum Likelihood method in Mega 7. The bootstrap values are based upon 10000 replicates.

5.3 Model prediction and assessment

Three-dimensional (3D) structure determination of any protein is critically important in understanding its function. The evaluation of 3-D model of a protein provides valuable information for studies related to site specific inhibitors, disease associated mutations and site-specific mutagenesis (Fidanova and Lirkov 2009). This further helps in manipulating the cellular and biochemical functions of an enzyme/protein (Combet et al. 2002). Therefore in the present study the best model of protease enzyme (Fig. 16) was chosen using Swiss Model Server which gave Z-scores of -1.365 (Benkert et al. 2011), QMEAN6 scores of 0.653 (Benkert et al. 2009) and Dfire energy value of -414.90 kJ mol-1 (Zhou and Zhou 2002). The local model quality estimation based on Anolea, Gromos, and QMEAN6 (Fig. 17) also validated the quality of the protease model (Melo and Feytmans 1998).

Further validation of this model using PROCHECK software clearly demonstrated that the amino acid residues are located in the most favoured region i.e. up to 73.2 % which is evident from Ramachandran plot (Fig. 18). ERRAT analysis of this model resulted in overall quality factor of approximately 88.59. The VERIFY3D analysis determined the compatibility of 3-D with its own 1-D model showing 86.27 % match resulting in an average 3D-1D Score of >0.2. The SAVES software also validated this model for further studies. The evaluation of 3-D model using Phyre 2.0 revealed that α-helix is 49 %, β-strand is 14 % and only 12% is disordered confirming it to be the best possible model as the disorder percentage is < 50 % (Fig. 19). Hence the 3-D structure of serine metalloprotease from GUSK-1 has been well predicted and evaluated. There is a single earlier report describing the 3-D model of an alkaline metagenomic protease eventually identified as serine protease from a saline habitat (Purohit and Singh 2013).

Fig. 16. Best predicted 3-D model of serine metalloprotease using Swiss Model server

The best model of serine metalloprotease was chosen using Swiss Model Server which possessed Z-scores of -1.365, QMEAN6 scores of 0.653 and Dfire energy value of -414.90 kJ mol^{-1} respectively.

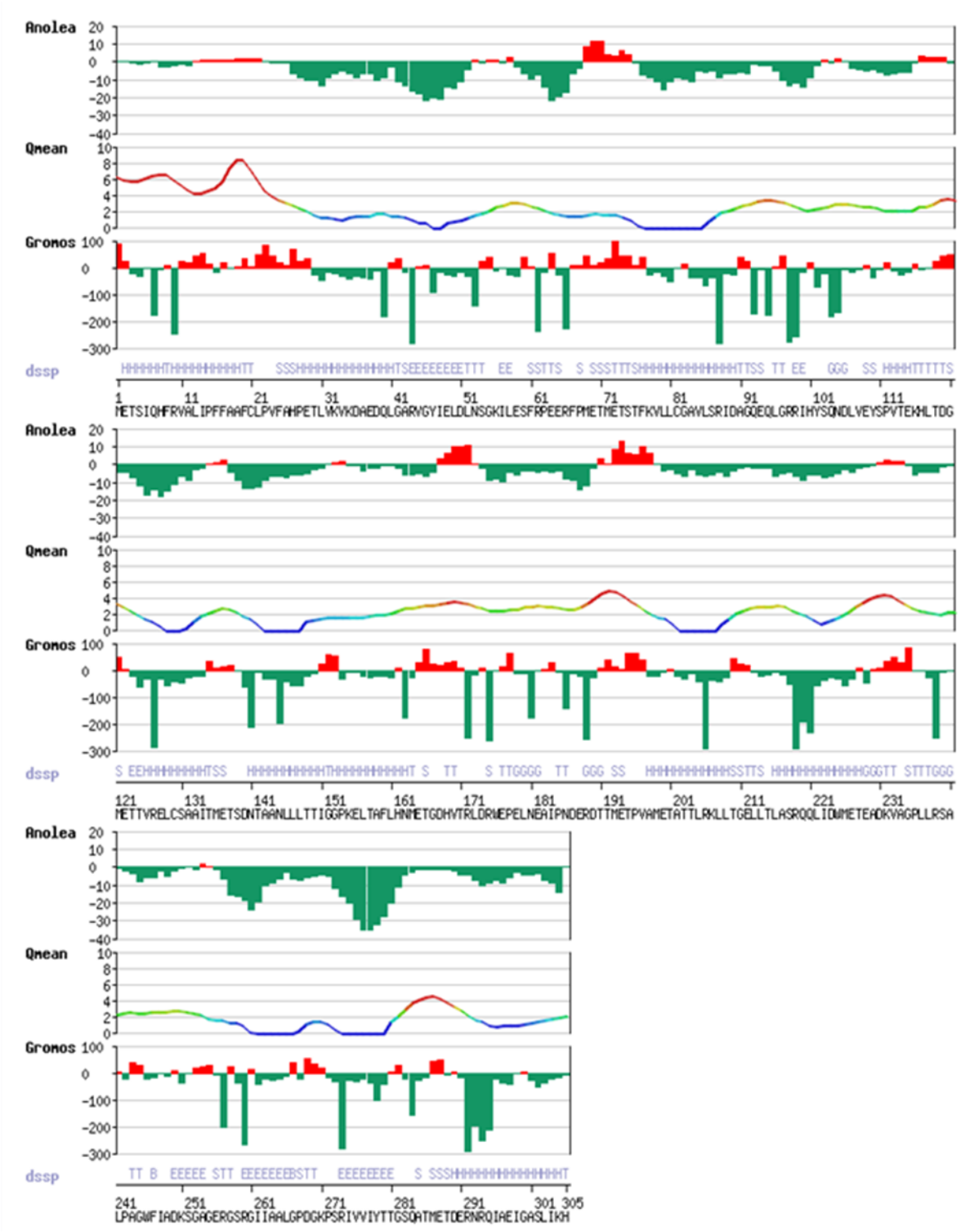


Fig. 17. Assessment of 3-D model of serine metalloprotease using Anolea, Qmean and Gromos force fields (green color indicates energy favorable environment while red indicates energy unfavorable environment)

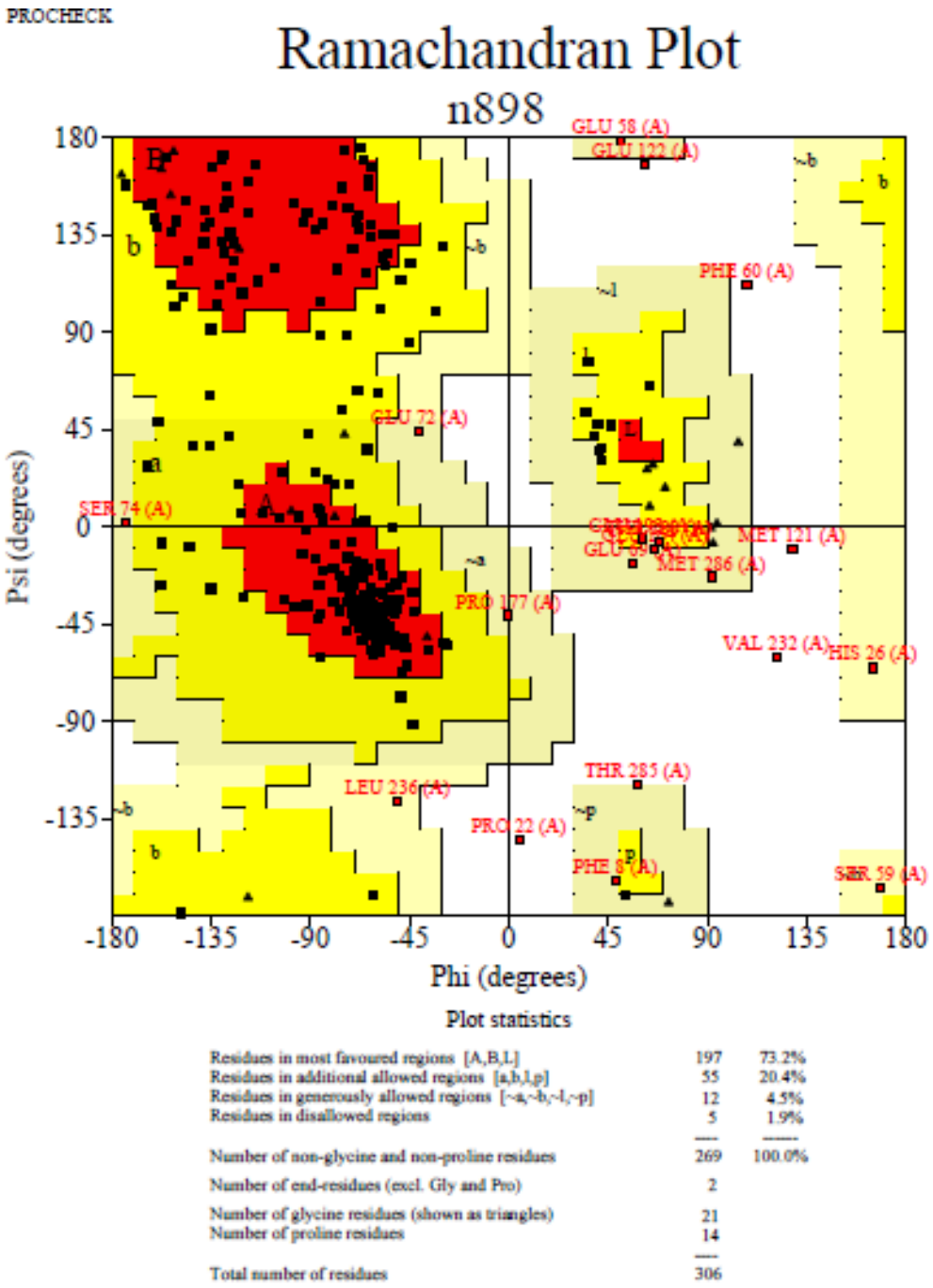


Fig. 18. Ramachandran plot for serine metalloprotease using PROCHECK software

This plot revealed that upto 73.2 % amino acid residues were located in the most favoured region confirming the validity of the 3-D model

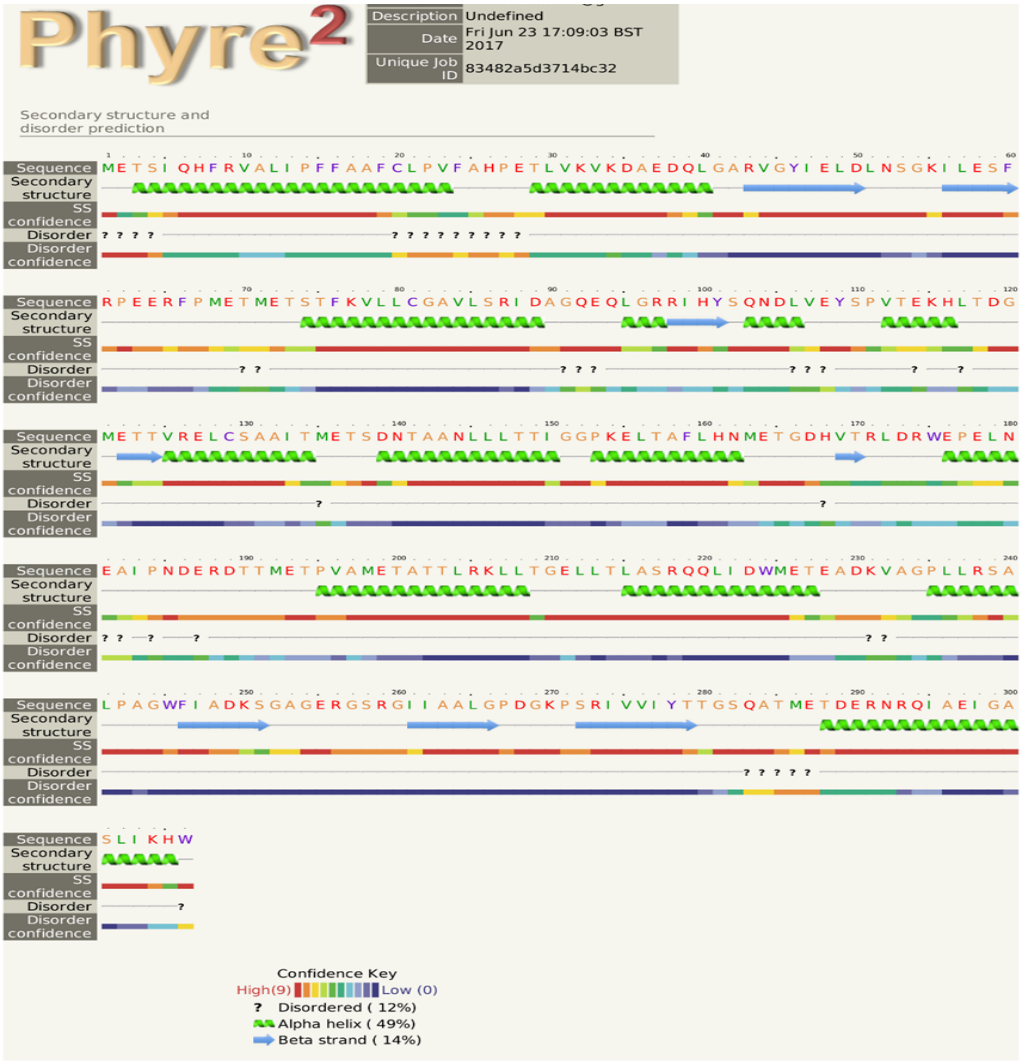


Fig. 19. Analysis of 3-D model of serine metalloprotease using Phyre 2.0 for prediction of alpha helix, beta strand and disorder. The color of the amino acids are based on their group together due to their simple properties. Amino acid residues A, S, T, G and P which are small/polar are shown in yellow, hydrophobic amino acid residue M, I, L and V are shown in green, charged amino acids K, R, E, N, D, H and Q are shown in red whereas aromatic amino acids along with cysteine W, Y, F and C are depicted in purple color. The secondary structure prediction comprises of three states: α-helix, β- strand or coil. Green helices represent α-helices, blue arrows indicate β-strands and faint lines indicate coil respectively.

5.4 *In silico* docking studies

Proteases are known for their degradative as well as synthetic characteristics. They play a crucial role in several pathological and pathophysiological processes (Rao et al. 1998). In case of HIV-1, combination therapies involve the usage of drugs targeting the HIV- protease and reverse transcriptase. However, a major drawback is that viruses may develop resistance or the drug may even cause side effects in humans. These drawbacks are mainly due to the fact that these drugs generally get involved in later stages of the infection therefore, there is a need for an alternative approach that can target the virus at an initial stage. This implies that the entry of HIV-1 needs to be blocked which is mediated primarily by two glycoproteins i.e. gp120 and gp41 (Teixeira et al. 2011). On the other hand, *Listeria monocytogenenes* being an intracellular foodborne pathogen, mostly targets an immunocompromised host. This pathogenic bacterium is transmitted via contaminated food and it traverses the epithelial barrier by intracellular or paracellular means in the gastrointestinal tract. Listeria adhesion protein (LAP) plays an important role in crossing this barrier via paracellular route (Koo et al. 2012). Therefore, blocking of LAP may serve as a important resistance mechanism against *L. monocytogenes*.

In silico studies using our serine metalloprotease interestingly revealed docking of the serine metalloprotease (shown in green) with gp41 coat protein (PDB ID: 3U91) of HIV-1 (shown in red) and cell adhesion protein of *Listeria monocytogenes* (PDB ID: 4EZG). The amino acid residues of the serine metalloprotease SER 74 and SER 138 docked with ARG 20 of gp41 coat protein of HIV-1 (Fig. 20) , whereas GLU 211 and LYS 55 of serine metalloprotease interacted with ASN 103 and SER 218 of cell adhesion protein of *L. monocytogenes* respectively (Fig. 21).

Docking studies clearly demonstrated that the protease encoded by the ORF 4 of metagenomic clone GUSK-1 may be used to design drugs against HIV-1 and *Listeria monocytogenes* which are serious human pathogens. There has also been a report of a metagenomic metalloprotease showing fibrinolytic activity that could be used to develop therapeutic agents for the treatment of thrombosis (Lee et al. 2007). It is interesting to note that there is no report so far about serine metalloprotease from marine environments using metagenomic approach targeting human pathogens viz. HIV-1 and *Listeria monocytogenes* and thus such environments should be explored more extensively for screening and development of novel therapeutics, biomolecules and protein pharmaceuticals.

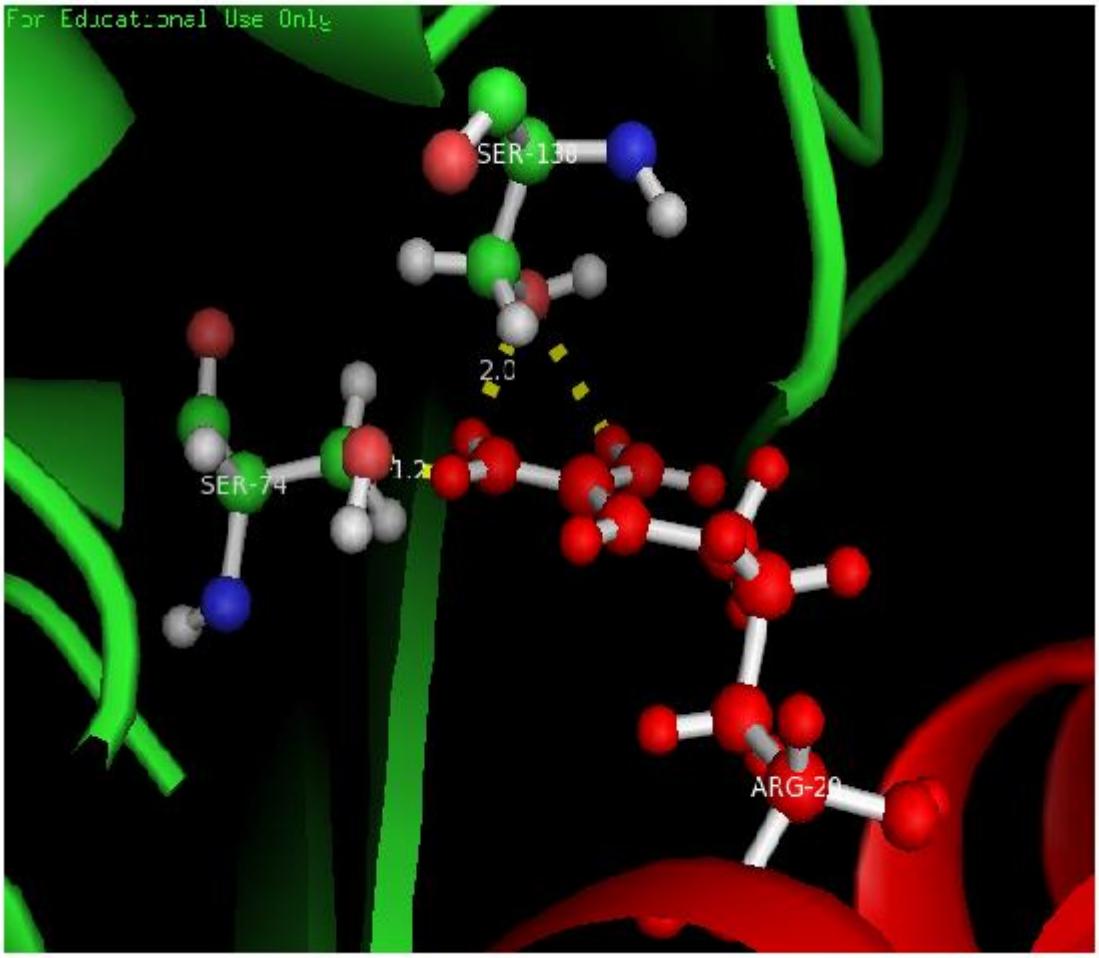


Fig. 20. Docking of serine metalloprotease (green) with gp41 coat protein of HIV-1(red)

Amino acid residue SER 74 and SER 138 of serine metalloprotease is docking with amino acid residue ARG 20 of coat protein of HIV-1.

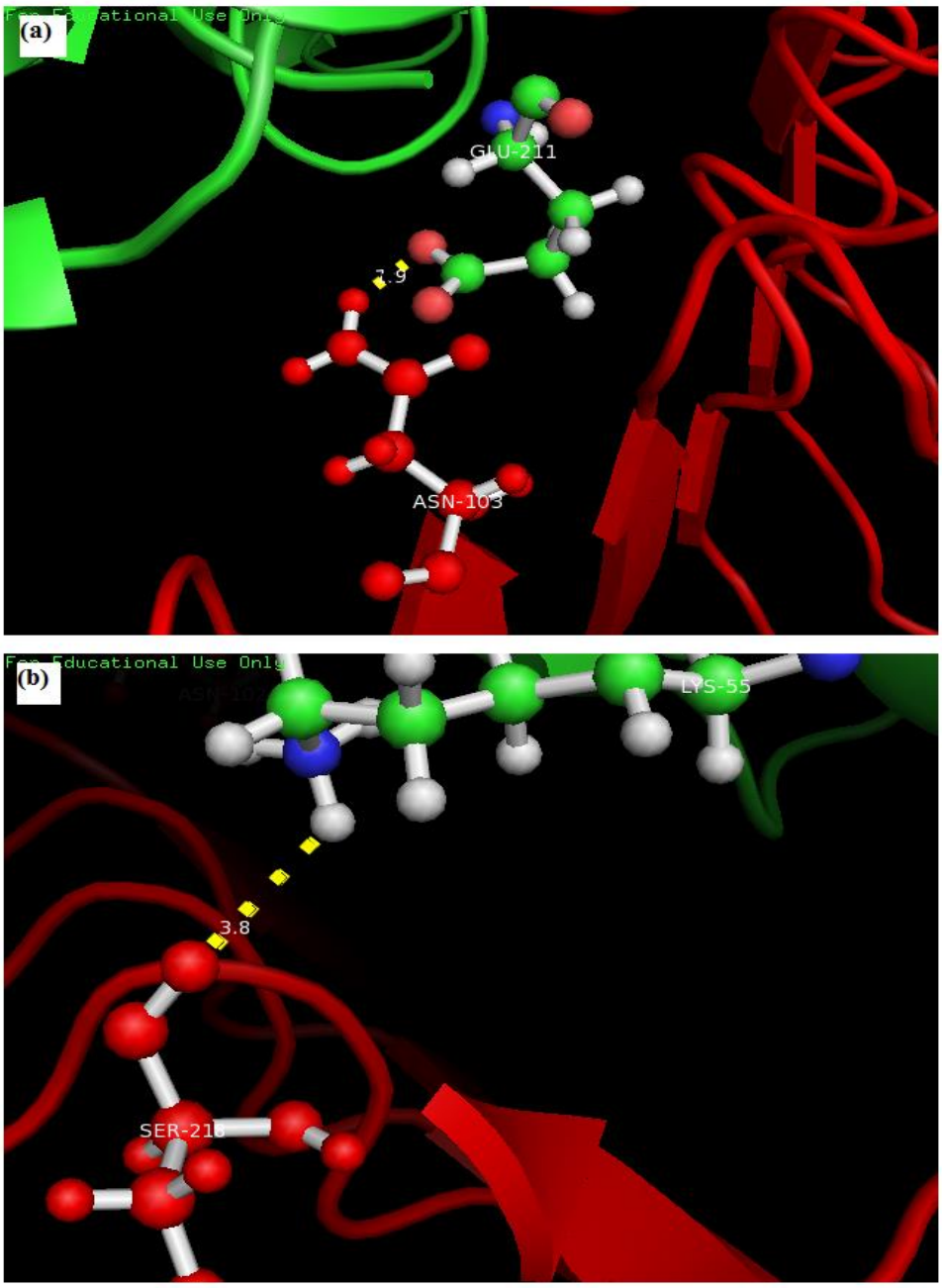


Fig. 21. Docking of serine metalloprotease (green) with cell adhesion protein of *Listeria monocytogenes* (red)

(a) Amino acid residue GLU 211 of serine metalloprotease docking with amino acid residues ASN 103 of cell adhesion protein of *Listeria monocytogenes* respectively

(b) Amino acid residue LYS 55 of serine metalloprotease is docking with amino acid residue SER 218 of cell adhesion protein of *Listeria monocytogenes* respectively

Conclusion

The 4.0 kbps metagenomic DNA insert from clone GUSK-1 possessed an ORF (i.e.ORF4) encoding a serine metalloprotease containing 286 amino acid residues belonging to transpeptidase superfamily. The 3-D structure of this serine metalloprotease was well predicted and evaluated using various bioinformatic tools. *In silico* docking studies clearly demonstrated interesting interactions of this serine metalloprotease with gp41 coat protein of HIV-1 and cell adhesion protein of *Listeria monocytogenes.* Thus, it may serve as a potential candidate for designing drugs and biopharmaceuticals against these deadly human pathogens.

The present study has been carried out keeping in view the importance of metagenomics to explore and characterize novel bioactive molecules from marine ecosystem which is not tapped fully. During this study a modified protocol was deviced for isolation of metagenomic DNA from estuarine sediment samples. After screening of approximately 25,000 metagenomic clones only one protease positive clone was found from the metagenomic library of estuarine sediment samples which was designated as GUSK-1. This DNA insert of positive clone was subcloned in the expression vector pET-22b inorder to overexpress the protease and subsequently protease was purified using ammonium sulphate and DEAE cellulose column. The purified enzyme was extensively characterized for its maximum activity and its stability with reference to temperature, pH, organic solvents, inhibitors and bivalent cations. These parameters are employed to screen industrially valued enzymes. The SDS-PAGE as well as zymogram analysis confirmed the molecular weight of the protease as 40 kDa and its activity. The metagenomic DNA insert (i.e. 4 kbps) was sequenced and aminoacid sequence was deduced.Based on amino acid sequence 3-D modelling and *in silico* docking studies of the protease were carried out with proteins of serious human pathogens viz. HIV-1 and *Listeria monocytogenes.*

The salient points of this study are as follows:

- A rapid, cost effective, high yielding and environment friendly, efficient method was designed to overcome the limitations of isolating metagenomic DNA from marine sediment samples.
- There was only one positive clone isolated after screening of approximately 25,000 clones which proved the difference in the level of gene expression of diverse taxonomical groups and also due to codon biases with reference to the use of expression systems.

- The metagenomic clone GUSK-1 encodes protease enzymes which proved to be an industrially valuable enzyme as it showed maximum enzyme activity at temperature 70 ^{o}C and pH 10.
- It is interesting to note that enzyme was 100 % stable at higher temperature and pH (i.e. 70 ^{o}C and pH 10) for 1 hour.
- Protease showed tolerance as well as enhanced in enzyme activity in presence of most of the common organic solvents which were tested.
- The enzyme activity was enhanced in presence of bivalent cations (5mM) as well as majority inhibitors and chelators.
- Inhibition of enzyme activity by (5 mM) PMSF and EDTA-Na_2 separately confirmed that the enzyme is serine metalloprotease.
- Next generation sequencing of 4 kbps DNA insert followed by ORF screening and BLAST analysis confirmed that this protease belongs to transpeptidase family.
- *In silico* docking studies further revealed a very interesting results as it showed successful docking of the protease enzyme/protein with gp41 coat protein of HIV-1 and cell adhesion protein of *Listeria monocytogenes.* These findings have proved that this protease has got importance in development of new drugs and pharmaceuticals against serious human pathogens.
- Therefore these studies have clearly proved the importance of metagenomic approach in screening and characterization of one novel protease gene from estuarine environment of Goa, India.

Appendix

Appendix-A

Media

A.1 Luria-Bertani (LB) broth

Tryptone	10g
Yeast Extract	5g
NaCl	10g
D/W	1000mL

A.2 Skimmed Milk Agar

Tryptone	10g
Yeast Extract	5g
NaCl	10g
Skim milk	10g
Agar	20g
D/W	1000mL

Appendix-B

Agarose Gel Electrophoresis

B.1 0.8% agarose

Weigh 0.8g agarose and dissolve in 100mL of 1X TAE buffer. Melt the solution in microwave oven untill clear, transparent solution is achieved. Add ethidium bromide to a final concentration of 0.5µg/mL and cast the gel.

B.2 Ethidium Bromide

Add 1g of ethidium bromide to 100mL of de-ionised water. Stir on magnetic stirrer for several hours to ensure that dye has dissolved. Transfer solution to a dark bottle and store at room temperature.

B.3 Gel Loading Buffer

0.05% (w/v) Bromophenol blue

40% (w/v) Sucrose

0.1M EDTA (pH 8.0)

0.5% (w/v) SDS

B.4 50X Tris Acetate EDTA

Tris base	24.2g
Glacial acetic acid	5.71mL
0.5M EDTA	10mL
D/W	100mL

B. 5 10X Tris EDTA (TE) Buffer (pH 8.0)

Tris Cl	100mM
EDTA	10mM

Sterilize for 20mins at 15psi.

Appendix-C

C. 1 Restriction Digestion

Component	Concentration	Quantity
DNA	50 ng/µL	10 µL
Restriction buffer	10X	2 µL
BSA	100X	2 µL
Enzyme (*Sau* 3A1)	10U/ µL	1 µL
Milli Q water	-	5 µL
Total volume		20 µL

C. 2 Dephosphorylation

Component	Concentration	Quantity
DNA	10 ng/µL	20 µL
Restriction buffer	10X	3 µL
Enzyme (Calf intestinal phosphatase)	0.5U/ µL	3 µL
Milli Q water	-	4 µL
Total volume		30 µL

C. 3 Ligation

Component	Concentration	Quantity
Dephosphorylated plasmid DNA	25 ng	1
Insert genomic DNA	25ng	3
Ligase buffer	5X	2
T4 DNA ligase	1U/ µL	1
Milli Q water	-	3 µL
Total volume		10 µL

Appendix-D

Sodium dodecyl sulphate-Polyacrylamide gel electrophoresis (SDS-PAGE) and Zymography (Sambrook et al., 1989)

D.1 Stock solutions for SDS-PAGE

D.1.1 Acrylamide-*bis*-acrylamide solution (monomer solution)

Acrylamide	29.0g
N,N' methylene *bis* acrylamide	1.0g
D/W	100mL

Acrylamide and N,N' methylene *bis* acrylamide was dissolved in 80mL of warm de-ionized water. The pH of the solution was adjusted to 7.0. The final volume of solution was made to 100mL with de-ionized water. The solution was stored in amber colour bottle at room temperature.

D. 1.2 Resolving gel buffer (Tris 1.5 M, pH 8.8)

Tris (base)	18.171 gm
D/W	100 ml

Tris was dissolved in 60mL of de-ionized water. The pH of the solution was adjusted to 8.8 with 6 N HC1 and the final volume was made up to 100mL with de-ionized water. The solution was stored at 4 °C.

D. 1.3 Stacking gel buffer (Tris 1.0 M, pH 6.8)

Tris (base)	12.114 g
D/W	100 mL

Tris was dissolved in 60mL of de-ionized water. The pH of the solution was adjusted to 6.8 with 6 N HC1 and the final volume was made up to 100mL with de-ionized water. The solution was stored at 4 °C.

D.1.4 10% ammonium per sulphate (APS)

Ammonium per sulphate	0.1 g
D/W	1 mL

D.1.5 10% Sodium dodecyl sulphate

Sodium dodecyl sulphate	10 g
D/W	100 mL

D.1.6 6N Hydrochloric acid

Concentrated HCl	51mL
D/W	100mL

D.1.7 1% Bromophenol blue

Bromophenol blue	0.1g
D/W	10mL

D.1.8 5X Tris-glycine electrophoresis buffer (pH 8.3)

25mM Tris base	3.02g
250mM Glycine	18.8g
10% (w/v) SDS	10mL
D/W	200mL

Preparation of 1X tank buffer: 100mL of 5X Tris-glycine electrophoresis buffer was made to 500mL with de-ionized water.

D.1.9 2X Sample Solubilizing buffer

1M Tris HC1 (pH 6.8)	1ml
Glycerol	2mL
1% (w/v) Bromophenol blue	2mL
10% (w/v) SDS	4 ml
200mM β-mercaptoethanol	284µL
D/W	716µL

D.1.10 Preparation of resolving and stacking gel

Solution	12% Resolving gel (10mL)	5% Stacking gel (4mL)
Monomer	4.0	0.67
1.5M Tris (pH 8.8)	2.5	-
1.0M Tris (pH 6.8)	-	0.5
10% (w/v) SDS	0.1	0.04
10% (w/v) APS	0.1	0.04
TEMED	0.004	0.004
D/W	3.3	2.7

D.2 Staining of SDS-PAGE Gels

D.2.1 Coomassie Brilliant Blue Staining Solution

Coomassie Brilliant Blue R-250	0.05g
Methanol	50mL
Glacial acetic acid	10mL
D/W	100mL

D.2.2 Destaining Solution

Methanol	30mL
Glacial acetic acid	10mL
D/W	100mL

D.3 Developing Buffer for Zymography

D.3.1 Developing Buffer (pH 7.5)

Tris base	0.6055g
NaCl	1.169g
$ZnCl_2$	0.07mg
$CaCl_2$	0.0074g
Brij35	0.2g

Appendix-E

E.1 PCR reaction mixture

Component	Concentration	Quantity
Metagenomic DNA	50 ng	4
Master mix	2X	25
Forward primer	20 mM	1
Reverse primer	20 mM	1
Milli Q water	-	19 µL
Total volume		50 µL

Standard curves

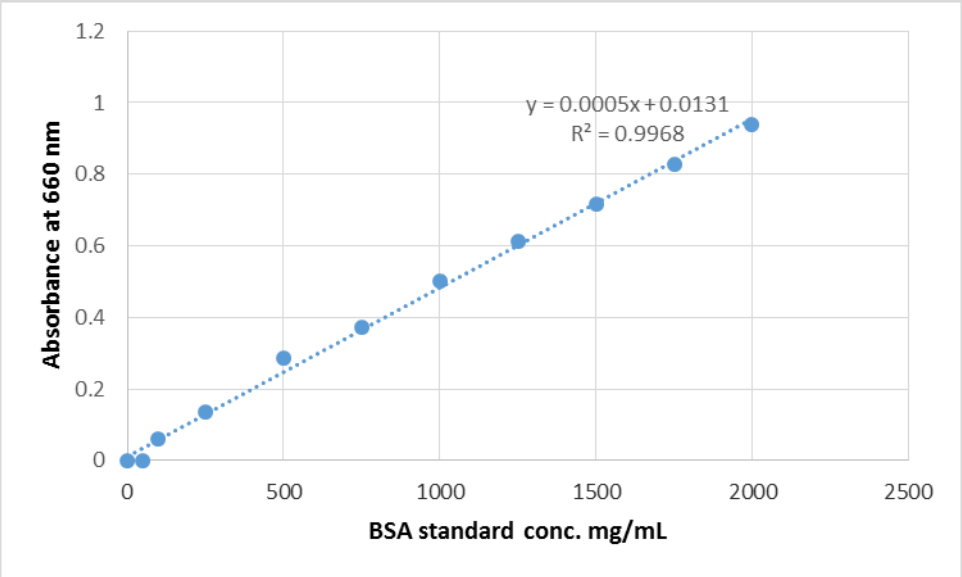


Fig. A. Standard curve for estimation of total protein content

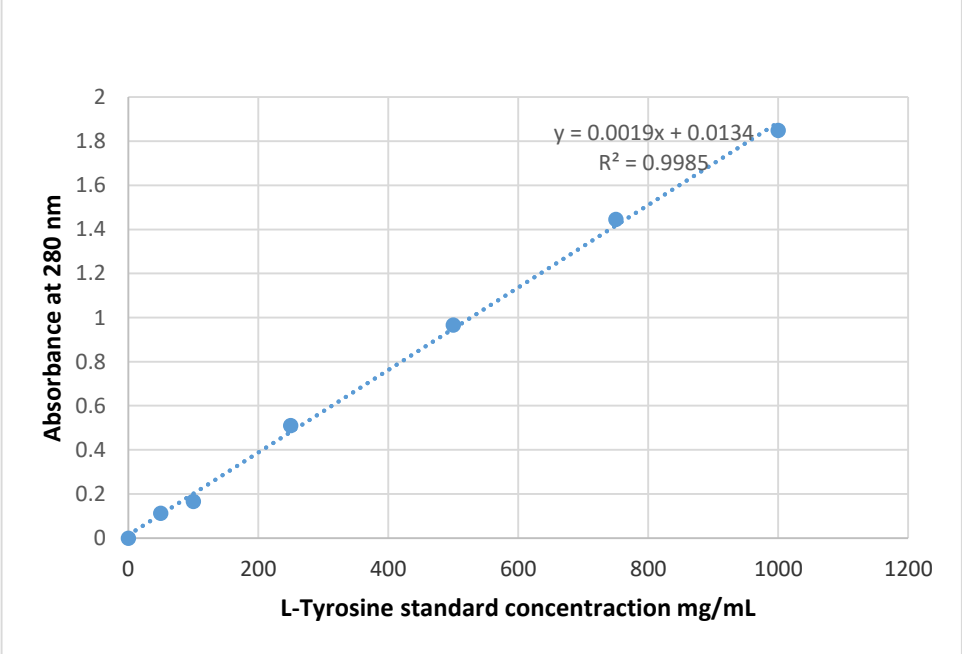


Fig. B. Standard curve for estimation of protease enzyme

Bibliography

Adrio, J.L. and Demain, A.L. (2014). Microbial enzymes: Tools for biotechnological processes. *Biomol* 4: 117-139.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990). Basic local alignment search tool. *J Mol Biol* 215: 403-410.

Alm, E.W., Zheng, D. and Raskin, L. (2000). The presence of humic substances and DNA in RNA extracts affects hybridization results. *Appl Environ Microbiol* 66: 4547- 4554.

Amann, R.I., Ludwig, W. and Schleifer, K.H. (1995). Phylogenetic identification and in-situ detection of individual microbial-cells without cultivation. *Microbiol Rev* 59: 143-169.

Amorim, J.H., Macena, T.N., Lacerda, G.V.J., Rezende, R.P., Dias, J.C., Brendel, M. and Cascardo, J.C. (2008). An improved extraction protocol for metagenomic DNA from a soil of the Brazilian Atlantic Rainforest. *Genet Mol Res* 7: 1226-1232.

Andersen, L.P. (1998). Method for dehairing of hides or skins by means of enzymes. US Patent 5,834,299.

Annamalai, N., Rajeswari, M.V., Thavasi, R., Vijayalakshmi, S. and Balasubramanian, T. (2013). Optimization, purification and characterization of novel thermostable, haloalkaline, solvent stable protease from *Bacillus halodurans* CAS6 using marine shellfish wastes: A potential additive for detergent and antioxidant synthesis. *Bioprocess Biosyst Eng* 36: 873-883.

Ansorge, W.J. (2009). Next-generation DNA sequencing techniques. *Nat Biotechnol* 25: 195-203.

Bach, H.J., Tomanova, J., Schloter, M., Munch, J.C. (2002). Enumeration of total bacteria and bacteria with genes for proteolytic activity in pure cultures and in environmental samples by quantitative PCR mediated amplification. *J Microbiol Methods* 49: 235-245.

Bag, S., Saha, B., Mehta, O., Anbumani, D., Kumar, N., Dayal, M., Pant, A., Kumar, P., Saxena, S., Allin, K.H., Hansen, T., Arumugam, M., Vestergaard, H., Pedersen, O., Pereira, V., Abraham, P., Tripathi, R., Wadhwa, N., Bhatnagar, S., Prakash, V.G., Radha, V., Anjana, R.M., Mohan, V., Takeda, T., Kurukawa, T., Nair, G.B. and Das,

B. (2016). An improved method for high quality metagenomics DNA extraction from human and environmental samples. *Sci Rep* 6: 26775.

Barrett, A.J., Rawlings, N.D. and Woessner JF. (Eds) (1998). *In*: Handbook of proteolytic enzymes. Academic Press, London, pp. xxiii-xxiv.

Baweja, M., Tiwari, R., Singh, P.K., Nain, L. and Shukla, P. (2016). An alkaline protease from *Bacillus pumilus* MP 27: Functional analysis of its binding model toward its applications as detergent additive. *Front Microbiol* 7: 1195.

Beja, O. (2004). To BAC or not to BAC: marine ecogenomics. *Curr Opin Biotechnol* 15: 187-90.

Benkert, P., Biasini, M. and Schwede, T. (2011). Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* 27: 343-350.

Benkert, P., Schwede, T. and Tosatto, S.C. (2009). QMEANclust: estimation of protein model quality by combining a composite scoring function with structural density information. *BMC Struct. Biol* 9: 35.

Bertrand, H., Poly, F., Van, V.T., Lombard, N., Nalin, R., Vogel, T.M. and Simonet, P. (2005). High molecular weight DNA recovery from soils prerequisite for biotechnological metagenomic library construction. *J Microbiol Methods* 62: 1-11.

Biver, S., Portetelle, D. and Vandenbol, M. (2013). Characterization of a new oxidant-stable serine protease isolated by functional metagenomics. *Springerplus* 2: 410.

Bos, D.H., Mayfield, C. and Minchella, D.J. (2009). Analysis of regulatory protease sequences identified through bioinformatic data mining of the *Schistosoma mansoni* genome. *BMC Genomics* 10: 488.

Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72: 248-254.

Burgmann, H., Pesaro, M., Widmer, F. and Zeyer, J. (2001). A strategy for optimizing quality and quantity of DNA extracted from soil. *J Microbiol Methods* 45: 7-20.

Chen, K. and Pachter, L. (2005). Bioinformatics for whole-genome shotgun sequencing of microbial communities. *PLoS Comput Biol* 1: 24.

Cheng, S.W., Hu, M.N., Shen, S.W., Takagi, H., Asano, M. and Tsai, Y.C. (1995). Production and characterization of keratinase of a feather-degrading *Bacillus licheniformis* PWD-1. *Biosci Biotech Biochem* 59: 2239-2243.

Chistoserdova, L. (2010). Recent progress and new challenges in metagenomics for biotechnology. *Biotechnol Lett* 32: 1351-1359.

Chris Colovos, T.O.Y. (1993). Verification Of Protein Structure: Patterns Of Nonbonded Atomic Interactions. *Protein Sci* 2: 1511-1519.

Clegg, C.D., Ritz, K. and Griffith, B.S. (1997). Direct isolation of microbial community DNA from humified upland soils. *Appl Microbiol* 25: 30-33.

Combet, C., Jambon, M., Deleage, G. and Geourjon, C. (2002). Geno3D: automatic comparative molecular modelling of protein. *Bioinformatics* 18: 213-214.

Cowan, D.A., Arslanoglu, A., Burton, S.G., Baker, G.C., Cameron, R.A., Smith, J.J. and Meyer, Q. (2004). Metagenomics, gene discovery and the ideal biocatalyst. *Biochem Soc Trans* 32: 298-302.

Craig, J.W., Chang, F.Y., Kim, J.H., Obiajulu, S.C. and Brady, S.F. (2010). Expanding small-molecule functional metagenomics through parallel screening of broad host-range cosmid environmental DNA libraries in diverse Proteobacteria. *Appl Environ Microbiol* 76: 1633-1641.

Cruz, J.M., Ortega, M.A., Cruz, J.C., Ondina, P., Santiago, R. and Rios-Velazquez, C. (2010). Unravelling activities by functional-based approaches using metagenomic libraries from dry and rain forest soils in Puerto Rico. Current research technology and education topics. *In* Mendez-Vilas (Ed.) applied microbiology and microbial biotechnology. FORMATEX, Badajoz, Spain. pp. 1471-1478.

Dalev, P.G. (1994). Utilization of waste feathers from poultry slaughter for production of a protein concentrate. *Bioresour Technol* 48: 265-267.

D'Avila-Levy, C.M., Santos, A.L.S., Cuervo, P., De Jesus, J.B. and Branquinha, M.H. (2012). Applications of Zymography (Substrate-SDS-PAGE) for Peptidase Screening in a Post-Genomic Era. *In*: Magdeldin S, editor. Gel Electrophoresis - Advanced Techniques. China: InTech; pp. 265-88.

D'Costa, B., Shamim, K. and Dubey, S.K. (2013). Characterization of thermostable serine protease from *Bacillus altitudinis* strain BR1. *J Sci Ind Res* (India) 72: 166-171.

DeLong, E.F. (1997). Marine microbial diversity: the tip of the iceberg. *Trends Biotechnol* 15: 203-207.

Delong, E.F. (2005). Microbial community genomics in the ocean. *Nat Rev Microbiol* 3: 459-69.

Deng A, Wu J, Zhang Y, Zhang G, Wen T. (2010) Purification and characterization of a surfactant-stable high-alkaline protease from *Bacillus* sp. B001. *Bioresour Technol* 101:7100-7106.

Devi, S.G., Fathima, A.A., Radha, S., Arunraj, R., Curtis, W.R. and Ramya, M. (2015). A rapid and economical method for efficient DNA extraction from diverse soils suitable for metagenomic applications. PLoS ONE 10:7.

Dodia, M.S., Rawal, C.M., Bhimani, H.G., Joshi, R.H., Khare, S.K. and Singh, S.P. (2008). Purification and stability characteristics of an alkaline serine protease from a newly isolated Haloalkaliphilic bacterium sp. AH-6. *J Ind Microbiol Biotechnol* 35: 121-131.

Edwards, R.A. and Dinsdale, E.A. (2007). Marine environmental genomics: unlocking the ocean's secrets. *Oceanography* 20: 26-61.

Eisenberg, D., Lüthy, R. and Bowie, J.U. (1997). VERIFY3D : Assessment of Protein Models with Three-Dimensional Profiles. *Methods Enzymol* 277: 397-404.

Falkowski, P.G. and Vargas, C. (2004). Shotgun sequencing in the sea: A blast from the past? *Science* 304: 58-60.

Fidanova, S. and Lirkov, I. (2009). 3D Protein Structure Prediction. Seria Matematica-Informatica 47(2): 33-46.

Fox, J.W., Shannon, J.D. and Bjarnason, J.B. (1991). Proteinases and their inhibitors in biotechnology. Enzymes in biomass conversion. *ACS Symp Ser* 460: 62-79.

Fujimaki, M., Yamashita, M., Okazawa, Y. and Arai, S. (1970). Applying proteolytic enzymes on soybean. 3. Diffusable bitter peptides and free amino acids in peptic hydrolyzate of soybean protein. *J Food Sci* 35: 215-218.

Fujiwara, N., Tsumiya, T., Katada, T., Hosobuchi, T. and Yamamoto, K. (1989). Continuous recovery of silver from used X-ray films using a proteolytic enzyme. *Process Biochem* 24:155-156.

Fujiwara, N., Yamamoto, K. and Masui, A. (1991). Utilization of a thermostable alkaline protease from an alkalophilic thermophile for the recovery of silver from used X-ray film. *J Ferment Bioeng* 72: 306-308.

Gajju, H., Bhalla, T.C. and Agarwal, H.O. (1996). Thermostable alkaline protease from thermophilic *Bacillus coagulans* PB-77. *Indian J Microbiol* 36: 153-155.

George, S., Raju, V., Krishnan, M.R.V., Subramanian, T.V. and Jayaraman, K. (1995). Production of protease by *Bacillus amyloliquefaciens* in solid-state fermentation and its application in the unhairing of hides and skins. *Process Biochem* 30: 457-462.

Gessesse, A. (1997). The use of nugmeal as low cost substrate for the production of alkaline protease by the alkalophilic *Bacillus* sp. AR 009 and same properties of the enzyme. *Biores Technol* 62: 59-61.

Godfrey, T. and West, S. (1996). Industrial enzymology. McMillan Publishers Inc. New York, USA.

Green, B.D. and Keller, M. (2006). Capturing the uncultivated majority. *Curr Opin Biotechnol* 17: 236-40.

Gupta, R., Beg, Q.K. and Lorenz, P. (2002). Bacterial alkaline proteases: Molecular approaches and Industrial Applications. *Appl Microbiol Biotechnol* 59: 15-32.

Haddar, A., Agrebi, R., Bougatef, A., Hmidet, N., Sellami-Kamoun, A. and Nasri, M. (2009). Two detergent stable alkaline serine-proteases from *Bacillus mojavensis* A21: Purification, characterization and potential application as a laundry detergent additive. *Bioresour Technol* 100: 3366-3373.

Hameed, A., Keshavarz, T. and Evans, C.S. (1999). Effect of dissolved oxygen tension and pH on the production of extracellular protease from a new isolate of *Bacillus subtilis* K2, for use in leather processing. *J Chem Technol Biotechnol* 74: 5-8.

Hameed, A., Natt, M.A. and Evans, C.S. (1996). Production of alkaline protease by a new *Bacillus subtilis* isolate for use as a bating enzyme in leather treatment. *World J Microbiol Biotechnol* 12: 289-291.

Hanahan, D. (1983). Studies on the transformation of *Escherichia coli* with plasmids. *J Mol Biol* 166: 557-580.

Handelsman, J. (2004). Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev* 68: 669-685.

Handelsman, J., Rondon, M., Brady, S., Clardy, J. and Goodman, R. (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* 5: 245-249.

Hartley, B.S. (1960). Proteolytic enzymes. *Annu Rev Biochem* 29: 45-72.

Healy, F.G., Ray, R.M., Aldrich, H.C., Wilkie, A.C., O, I.L. and Shanmugam, K.T. (1995). Direct isolation of functional genes encoding cellulases from the microbial consortia in a thermophilic, anaerobic digester maintained on lignocellulose. *Appl Microbiol Biotechnol* 43: 667-674.

Horikoshi, K. (2008). Past, present and future of extremophiles. *Extremophiles* 12: 1-2.

Horikoshi, K., Nakajima, S., Ueda, M. and Mukaiyama T. (2010). Measurement of universal thermodynamic functions for a unitary fermi gas. 327(5964): 442-445.

Hugenholtz, P. (2002) Exploring prokaryotic diversity in the genomic era. *Genome Biol* 31-38.

Hugenholtz, P. and Tyson, G.W. (2009). Metagenomics. *Nature* 455: 481-83.

Ibrahim, A.S., Al-Salamah, A.A., Elbadawi, Y.B., El-Tayeb, M.A. and Ibrahim, S.S. (2015). Production of extracellular alkaline protease by new halotolerant alkaliphilic *Bacillus* sp. NPST-AK15 isolated from hyper saline soda lakes. *Electron J Biotechnol* 18: 236-243.

Ichida, J.M., Krizova, L., LeFevre, C.A., Keener, H.M., Elwell, D.L. and Burtt, Jr. E.H. (2001). Bacterial inoculum enhances keratin degradation and biofilm formation in poultry compost. *J Microbiol Methods* 47: 199-208.

Ishikawa, H., Ishimi, K., Sugiura, M., Sowa, A. and Fujiwara, N. (1993). Kinetics and mechanism of enzymatic hydrolysis of gelatin layers of X-ray film and release of silver particles. *J Ferment Bioeng* 76: 300-305.

Jain, D., Pancha, I., Mishra, S.K., Shrivastav, A. and Mishra, S. (2012). Purification and characterization of haloalkaline thermoactive, solvent stable and SDS-induced protease from *Bacillus* sp.: A potential additive for laundry detergents. *Bioresour Technol* 115: 228-236.

Joshi, R.H., Dodia, M.S. and Singh, S.P. (2008). Production and optimization of a commercially viable alkaline protease from a haloalkaliphilic bacterium. *Biotechnol Bioprocess Eng* 13: 552-559.

Joshi, S. and Satyanarayana, T. (2013). Characteristics and applications of a recombinant alkaline serine protease from a novel bacterium *Bacillus lehensis. Bioresour Technol* 131: 76-85.

Kalisz, H.M. (1988). Microbial proteinases. *In*: Enzyme studies. Springer, Berlin, Heidelberg, pp. 1-65.

Kalisz, H.M. (1988). Microbial proteinases. *Adv Biochem Eng Biotechnol* 36: 1-65.

Kanehisa, K. (2000). Woven or knit fabrics manufactured using yarn dyed raw silk. US Patent 6,080,689.

Karl, D.M. (2007). Microbial oceanography: paradigms, processes and promise. *Nat Rev Microbiol* 5: 759-69.

Kelly, L.A., Mezulis, S., Yates, C., Wass, M. and Sternberg, M. (2015). The Phyre2 web portal for protein modelling, prediction and analysis. *Nat Protoc* 10: 845-858.

Kennedy, J., Marchesi, J.R. and Dobson, A.D. (2008). Marine metagenomics: strategies for the discovery of novel enzymes with biotechnological applications from marine environments. *Microb Cell Fact* 7: 27.

Kennedy, J., Marchesi, J.R. and Dobson, A.D.W. (2007). Metagenomic approaches to exploit the biotechnological potential of the microbial consortia of marine sponges. *Appl Microbiol Biotechnol* 75: 11-20.

Kim, W., Choi, K., Kim, Y., Park, H., Chol, J., Lee, Y., Oh, H., Kwon, I. and Lee, S. (1996). Purification and characterization of a fibrinolytic enzyme produced from *Bacillus* sp. strain CK 11-4 screened from chungkook-jangs. *Appl Environ Microbiol* 62: 2482-2488.

Krsek, M. and Wellington, E.M. (1999). Comparison of different methods for the isolation and purification of total community DNA from soil. *J Microbiol Methods* 39: 1-16.

Koo, O.K., Amalaradiou, M.A.R. and Bhunia, A.K. (2012). Recombinant probiotic expressing *Listeria* adhesion protein attenuates *Listeria monocytogenes* virulence *in vitro*. PLoS one 7: e29277.

Kudrya, V.A. and Simonenko, I.A. (1994). Alkaline serine proteinase and lectin isolation from the culture fluid of *Bacillus subtilis*. *Appl Microbiol Biotechnol* 41: 505-509.

Kumar, S., Stecher, G. and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33.

Kunitz, M. (1947). Crystalline Soybean Trypsin Inhibitor : Ii. General Properties. *J Gen Physiol* 30: 291-310.

Laemmli, U.K. (1970). Cleavage of Structural Proteins during Assembly of Head of Bacteriophage-T4. *Nature* 227.

Lakay, F.M., Botha, A. and Prior, B.A. (2006). Comparative analysis of environmental DNA extraction and purification methods from different humic acid-rich soils. *J App Microbiol* 100: 1365-1373.

Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 26: 283-291.

Langridge, G. (2009). Testing the water: marine metagenomics. *Nature* 7: 552.

Lapidus, A. (2009). Genome sequence databases (overview): sequencing and assembly. *In*: Schaechter M (ed) The encyclopedia of microbiology. New York, Elsevier, pp. 196-210.

Lee, D.G., Jeon, J.H., Jang, M.K., Kim, N.Y., Lee, J.H., Lee, J.H., Kim, S.J., Kim, G.D. and Lee, S.H. (2007). Screening and characterization of a novel fibrinolytic metalloprotease from a metagenomic library. *Biotechnol Lett* 29: 465-472.

Li, X. and Qin, L. (2005). Metagenomics-based drug discovery and marine microbial diversity. *Trends Biotechnol* 23: 539-543.

Liolios, K., Chen, I.M., Mavromatis, K., Tavernarakis, N., Hugenholtz, P., Markowitz, V.M. and Kyrpides, N.C. (2010). The Genomes On Line Database (GOLD) in 2009: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res* 38: 346-354.

Lorenz, P. and Eck, J. (2005). Metagenomics and industrial applications. *Nat Rev Microbiol* 3: 510-516.

Lorenz, P., Liebeton, K., Niehaus, F. and Eck, J. (2002). Screening for novel enzymes for biocatalytic processes: accessing the metagenome as a resource of novel functional sequence space. *Curr Opin Biotechnol* 13: 572-577.

Lorenz, P. and Schleper, C. (2002). Metagenome-a challenging source of enzyme discovery. *J Mol Catal B: Enzym* 9: 13-19.

Luna, M.G., Anno, D.V. and Danovaro, R. (2006). DNA extraction procedure: A critical issue for bacterial diversity assessment in marine sediments. *Env Microbiol* 8: 308-320.

Manabe, S., Nariya, H., Miyata, S., Tanaka, H., Minami, J., Suzuki, M., Taniguchi, Y. and Okabe, A. (2010). Purification and characterization of a clostripain-like protease from a recombinant *Clostridium perfringens* culture. *Microbiol* 156:561-569.

Manikandan, M., Kannan, V. and Pašić, L. (2009). Diversity of microorganisms in solar salterns of Tamil Nadu, India. *World J Microbiol Biotechnol* 25: 1007-1017.

Manni, L., Jellouli, K., Agrebi, R., Bayoudh, A. and Nasri, M. (2008). Biochemical and molecular characterization of a novel calcium-dependent metalloprotease from *Bacillus cereus* SV1. *Process Biochem* 43: 522-530.

Marco, D. (2008). Metagenomics and the niche concept. *Theory Biosci* 127: 241-47.

Martin-Laurent, F., Philippot, L., Hallet, S., Chaussod, R., Germon, J.C., Soulas, G. and Catroux, G. (2001). DNA extraction from soils: old bias for new microbial diversity analysis methods. *Appl Environ Microbiol* 67: 2354-2359.

Matsui, T., Matsufuji, H., Seki, E., Osajima, K., Nakashima, M. and Osajima, Y. (1993). Inhibition of angiotensin I-converting enzyme by *Bacillus licheniformis* alkaline protease hydrolyzates derived from sardine muscles. *Biosci Biotechnol Biochem* 57: 922-925.

McGrath, M.E. (1999). The lysosomal cysteine proteases. *Annu Rev Biophys Biomol Struct* 28: 181-204.

Melo, F. and Feytmans, E. (1998). Assessing protein structures with a non-local atomic interaction energy. *J Mol Biol* 277: 1141-52.

Miller, D.N., Bryant, J.E., Madsen, E.L. and Ghiorse, W.C. (1999). Evaluation and optimization of DNA extraction and purification procedure for soil and sediments sample. *Appl Environ Microbiol* 65: 4715-4724.

More, M.I., Herrick, J.B., Silva, M.C., Ghiorse, W.C. and Madsen, E.L. (1994). Quantitative cell lysis of indigenous microorganisms and rapid extraction of DNA from sediments. *Appl Environ Microbiol* 60: 1572 -1580.

Morrissey, J.P., Gara, F. and Dobson, A.D.W. (2010). Creation of effective metagenomic library for enzyme catalyst. *Mar Drugs* 8: 608-628.

Nair, H.P., Vincent, H. and Bhat, S.G. (2014). Evaluation of five in situ lysis protocols for PCR amenable metagenomic DNA from mangrove soils. *Biotechnol Rep* 4: 134-138.

Neklyudov, A.D., Ivankin, A.N. and Berdutina, A.V. (2000). Properties and uses of protein hydrolysates (review). *Appl Biochem Microbiol* 36: 452-459.

Neveu, J., Regeard, C. and Dubow, M.S. (2011). Isolation and characterization of two serine proteases from metagenomic libraries of the Gobi and Death Valley deserts. *Appl Microbiol Biotechnol* 91: 635-644.

Ni, X., Yue, L., Chi, Z., Li, Z., Wang, X. and Madzak, C. (2009). Alkaline Protease Gene Cloning from the Marine Yeast *Aureobasidium pullulans* HN2-3 and the Protease Surface Display on *Yarrowia lipolytica* for Bioactive Peptide Production. *Mar Biotechnol* 11: 81-89.

Niemi, M.R., Heiskanen, I., Wallenius, K., Lindstro¨m, K. (2001). Extraction and purification of DNA in rhizosphere soil samples for PCR-DGGE analysis of bacterial consortia. *J Microbiol Methods* 45: 155-165.

Nigam, P. S. (2013). Microbial enzymes with special characteristics for biotechnological applications. *Biomol* 3: 597-611.

Otto, H.H. and Schirmeister, T. (1997). Cysteine proteases and their inhibitors. *Chem Rev* 97: 133-171.

Outtrup, H. (1990). Boyee Col. Microbial proteases and biotechnology. *In*: Fogarty, W.M., Kelly, C.T. (eds.), Microbial Enzymes and Biotechnology. Elsevier Science, New York, pp. 227-254.

Pace, N. R., Stahl, D. A., Lane, D. J. and Olsen, G. J. (1985). Analyzing natural microbial populations by rRNA sequences. *ASM News* 51: 4-12.

Pace, N.R. (1997). A molecular view of microbial diversity and the biosphere. *Science* 276: 734-740.

Pedros-Alio, C. (2006). Genomics and marine microbial ecology. *Int Microbiol* 9: 191-97.

Perea, A., Ugalde, U., Rodriguez, I. and Serra, J.L. (1993). Preparation and characterization of whey protein hydrolysates: application in industrial whey bioconversion processes. *Enzyme Microb Technol* 15: 418-423.

Pohlner, J., Halter, R., Beyreuther, K. and T. F. Meyer, T.F. (1986). Gene structure and extracellular secretion of *Neisseria gonorrhoeae* IgA protease. *Nature* 325: 458-462.

Puri, S. (2001). An alkaline protease from a *Bacillus* sp.: Production and potential applications in detergent formulation and degumming of silk. MSc thesis, University of Delhi, New Delhi.

Purohit, M.K. and Singh, S.P. (2013). A metagenomic alkaline protease from saline habitat: Cloning, over-expression and functional attributes. *Int J Biol Macromol* 53: 138-43.

Pushpam, P.L., Rajesh, T. and Gunasekaran, P. (2011). Identification and characterization of alkaline serine protease from goat skin surface metagenome. *AMB Express* 1: 3.

Rai, S.K. and Mukherjee, A.K. (2009). Ecological significance and some biotechnological application of an organic solvent stable alkaline serine protease from *Bacillus subtilis* strain DM-04. *Bioresour Technol* 100: 2642-2645.

Rao, M.B., Tanksale, A.M., Ghatge, M.S. and Deshpande, V.V. (1998). Molecular and biotechnological aspects of microbial proteases. *Microbiol Mol Biol Rev* 62: 597-635.

Rasch, M.G., Lund, I.K., Illemann, M., Høyer-Hansen, G. and Gårdsvoll, H. (2010). Purification and characterization of recombinant full-length and protease domain of murine MMP-9 expressed in Drosophila S2 cells. *Protein Expr Purif* 72(1): 87-94.

Rawlings, N.D., Barrett, A.J. and Finn, R. (2016). Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors *Nucleic Acids Res* 44doi:10.1093/nar/gkv1118

Rebeca, B.D., Pena-Vera, M.T. and Diaz-Castaneda, M. (1991). Production of fish protein hydrolysates with bacterial proteases; yield and nutritional value. *J Food Sci* 56:309-314.

Richardson, T.H., Tan, X., Frey, G., Callen, W., Cabell, M., Lam, D., Macomber, J., Short, J.M., Robertson, D.E. and Miller, C. (2002). A novel, high performance enzyme for starch liquefaction. Discovery and optimization of a low pH, thermostable α-amylase *J Biol Chem* 277: 26501-26507.

Riesenfeld, C.S., Schloss, P.D. and Handelsman, J. (2004). Metagenomics: genomic analysis of microbial communities. *Annu Rev Genet* 38: 525-52.

Rodriguez-Valera, F. (2004). Environmental genomics, the big picture? *FEMS Microbiol Lett* 231: 153-58.

Rondon, M.R., August, P.R., Bettermann, A.D., Brady, S.F., Grossman, T.H., Liles, M.R., Loiacono, K.A., Lynch, B.A., MacNeil, I.A., Minor, C., Tiong, C.., Gilman, M., Osburne, M.S., Clardy, J., Handelsman, J. and Goodman, R.M. (2000). Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *App Environ Microbiol* 66: 2541-2547.

Roose-Amsaleg, C.L., Garnier-Sillam, E. and Harry, M. (2001). Extraction and purification of microbial DNA from soil and sediment samples. *Appl Soil Ecol* 18: 47-60.

Krem, M.M., Rose, T. and Di Cera, E. (2000). Sequence determinants of function and evolution in serine proteases. *Trends Cardiovasc. Med*. 10: 171-176.

Sagar, K., Singh, S.P., Goutam, K.K. and Konwar, B.K. (2014). Assessment of five soil DNA extraction methods and a rapid laboratory-developed method for quality soil DNA extraction for 16S rDNA-based amplification and library construction. *J Microbiol Methods* 97(1): 68-73.

Salomon, C. and Faulkner, D.J. (2002). Localization studies of bioactive cyclic peptides in the ascidian *Lissoclinum patella*. *J Nat Prod* 65: 689-92.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). Detection and analysis of protein expressed from cloned genes. 4th ed. USA: Cold Spring Harbor Laboratory Available at: http://www.molecularcloning.com/.

Santosa, D.A. (2001). Rapid extraction and purification of environmental DNA for molecular cloning application and molecular diversity studies. *Mol Biotechnol* 17: 59-64.

Schloss, P.D. and Handelsman, J. (2003). Biotechnological prospects from metagenomics. *Curr Opin Biotechnol* 14: 303-310.

Schloss, P.D. and Handelsman, J. (2005). Metagenomics for studying unculturable microorganisms: cutting the Gordian knot. *Genome Biol* 6:229.

Schmeisser, C., Steele, H. and Streit, W.R. (2007). Metagenomics, biotechnology with nonculturable microbes. *Appl Microbiol Biotechnol* 75: 955-62.

Schmidt, T.M., DeLong, E.F. and Pace, N.R. (1991). Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *J Bacteriol* 173: 4371-4378.

Schneidman-Duhovny, D., Inbar, Y., Nussinov, R. and Wolfson, H.J. (2005). PatchDock and SymmDock: Servers for rigid and symmetric docking. *Nucleic Acids Res* 33: 363-367.

Schwartz, K. (2006). Recent advances in marine metagenomics. MMG 445 *Basic Biotechnol* eJournal 2: 165-69.

Schwede, T., Kopp, J., Guex, N. and Peitsch, M.C. (2003). SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res* 31: 3381-3385.

Seow, K.T., Meurer, G., Gerlitz, M., Wendt-Pienkowski, E., Hutchinson, C.R. and Davies, J. (1997). A study of iterative type II polyketide synthases, using bacterial genes cloned from soil DNA: a means to access and use genes from uncultured microorganisms. *J Bacteriol* 179: 7360-7368.

Sharma, S., Sharma, K.K. and Kuhad, R.C. (2014). An efficient and economical method for extraction of DNA amenable to biotechnological manipulations, from diverse soils and sediments. *J Appl Microbiol* 116: 923-933.

Simon, C. and Daniel, R. (2011). Metagenomic analyses: past and future trends. *Appl Environ Microbiol* 77: 1153-1161.

Singh, J., Behal, A., Singla, N., Joshi, A., Birbian, N., Singh, S., Bali, V. and Batra N. (2009). Metagenomics: concept, methodology, ecological inference and recent advances. *Biotechnol J* 4: 480-94.

Singh, PK.., Joseph, J., Goyal, S., Grover, A. and Shukla, P. (2016). Functional analysis of the binding model of microbial inulinases using docking and molecular dynamics simulation. *J Mol Model* 22: 1-7.

Sleator, R.D., Shortall, C. and Hill, C. (2008). Under the microscope: metagenomics. *Lett Appl Microbiol* 47: 361-66.

Solomon, S., Kachiprath, B., Jayanath, G., Sajeevan, T.P., Singh, I.B. and Philip, R. (2016). High-quality metagenomic DNA from marine sediment samples for genomic studies through a preprocessing approach. *3 Biotech* 6(2): 160.

Solomon, S., Kachiprath, B., Jayanath, G., Sajeevan, T.P., Singh, I.B. and Philip, R. (2016). High-quality metagenomic DNA from marine sediment samples for genomic studies through a preprocessing approach. *3 Biotech* 6(2): 160.

Soroor, M.A.M., Hoda, H.E., Hendawy., Abd-Elhady, M., GhazyNermeen, A., Semary, E., Khalil, K.M.A. and Abd El Aziz, A.M. (2009). Characterization of an alkaline metalloprotease secreted by the Entomopathogenic bacterium *Photorhabdus* sp. strain EK1. *R J Agri Biol Sci* 5(4): 349-360.

Steele, H.L. and Streit, W.R. (2005). Metagenomics: advances in ecology and biotechnology. *FEMS Microbiol Lett* 247: 105-11.

Steele, H.L., Jaeger, K.E., Daniel, R. and Streit, W.R. (2009). Advances in recovery of novel biocatalysts from metagenomes. *J Mol Microbiol Biotechnol* 16: 25-37.

Stein, J.L., Marsh, T.L., Wu, K.Y., Shizuya, H. and DeLong, E.F. (1996). Characterization of uncultivated prokaryotes: isolation and analysis of a 40- kilobase-pair genome fragment front a planktonic marine archaeon. *J Bacteriol* 178: 591-599.

Stroud, R.M. (1974). A family of protein cutting enzymes. *Scientific American* 231: 74-88.

Takami, H., Akiba, T. and Horikoshi, K. (1992). Cloning, expression, and characterization of a minor alkaline protease from *Bacillus* sp. no. AH-101. *Biosci Biotechnol Biochem* 56: 510-511.

Tebbe, C.C. and Vahjen, W. (1993). Interference of humic acids and DNA extracted directly from soil in detection and transformation of recombinant DNA from bacteria and a yeast. *Appl Environ Microbiol* 59: 2657-2665.

Teixeira, C.J., Gomes, P., Gomes, R., Maurel, F. and Barbault, F. (2011). Viral surface glycoproteins, gp120 and gp41, as potential drug targets against HIV-1: brief overview one quarter of a century past the approval of zidovudine, the first anti-retroviral drug. *Eur J Med Chem* 46: 979-992.

Thumar, J.T. and Singh, S.P. (2009). Organic solvent tolerance of an alkaline protease from salt-tolerant alkaliphilic *Streptomyces clavuligerus* strain Mit-1. *Ind Microbiol Biotechnol* 36: 211-218.

Torsvik, V.L. and Goksoyr, J. (1980). Determination of bacterial DNA in soil. *Soil Biol Biochem* 10: 7-12.

Toyokawa, Y., Takahara, H., Reungsang, A., Fukuta, M., Hachimine, Y., Tachibana, S. and Yasuda, M. (2010). Purification and characterization of a halotolerant serine proteinase from thermotolerant *Bacillus licheniformis* RKK-04 isolated from Thai fish sauce. *Appl Microbiol Biotechnol* 86 (6): 1867-1875.

Tsai, Y.L. and Olson, B.H. (1992). Rapid method for separation of bacterial DNA from humic substances in sediments for polymerase chain reaction. *App Environ Microbiol* 58: 2292-2295.

Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., Solovyev, V.V., Rubin, E.M., Rokhsar, D.S. and Banfield, J.F. (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* 428: 37–43

Varela, H., Ferrari, M.D., Belobradjic, L., Vazquez, A. and Loperena, M.L. (1997). Skin unhairing proteases of *Bacillus subtilis*: production and partial characterization. *Biotechnol Lett* 19: 755-758.

Venter, J.C., Remington, K. and Heidelberg, J.F. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304: 66-74.

Wang, G. Y., Graziani, E., Waters, B., Pan, W., Li, X., McDermott, J., Meurer, G., Saxena, G., Andersen, R.J. and Davies, J. (2000). Novel natural products from soil DNA libraries in a streptomycete host. *Org Lett* 2: 2401-2404.

Ward, N. (2006). New directions and interactions in metagenomics research. *FEMS Microbiol Ecol* 55: 331-38.

Ward, O.P. (1985). Proteolytic enzymes. *In*: Moo-Young M (ed) Comprehensive biotechnology, the practice of biotechnology: current commodity products, vol 3. Pergamon Press, Oxford, pp. 789-818.

Warnecke, F. and Hugenholtz, P. (2007). Building on basic metagenomics with complementary technologies. *Genome Biol* 8: 231.

Waschkowitz, T., Rockstroh, S. and Daniel, R. (2009). Isolation and characterization of metalloproteases with a novel domain structure by construction and screening of metagenomic libraries. *Appl Environ Microbiol* 75: 2506-2516.

Whitman, W.B., Coleman, D.C. and Wiebe, W. J. (1998). Prokaryotes: the unseen majority. *Proc Natl Acad Sci USA* 95: 6578-6583.

Wooley, J.C., Godzik, A. and Friedberg, I. (2010). A primer on metagenomics. *PLoS Comput Biol* 6: e1000667

Xu, J. (2006). Microbial ecology in the age of genomics and metagenomics: concepts, tools, and recent advances. *Mol Ecol* 15: 1713-31.

Yeates, C., Gillings, M.R., Davison, A.D., Altavilla, N. and Veal, D.A. (1997). PCR amplification of crude microbial DNA extracted from soil. *Let App Microb* 25: 303-317.

Yeates, C., Gilling, M.R., Davison, A.D., Altaviilla, N. and Veal, D.A. (1998). Methods for microbial DNA extraction from soil for PCR amplification. *Bio Procedures Online* 1: 40-47.

Zhang, M., Zhao, C., Lianxiang, D.U., Ping, Fu and Chen, L.U. (2008). Expression, purification, and characterization of a thermophilic neutral protease from *Bacillus stearothermophilus* in *Bacillus subtilis*. *Sci China Series C: Life Sci* 51(1): 52-59.

Zhang, Y., Zhao, J. and Zeng, R. (2011). Expression and characterization of a novel mesophilic protease from metagenomic library derived from Antarctic coastal sediment. *Extremophiles* 15: 23-29.

Zhou, H. and Zhou, Y. (2002). Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. *Protein Sci* 11: 2714-26.

Zhou, J., Bruns, M.A. and Tiedje, J.M. (1996). DNA recovery from soils of diverse composition. App *Environ Microbiol* 62: 316-322.

List of publication

Published

1) D'Costa, B., **Shamim, K.,** Dubey, S.K. 2013. Characterization of thermostable serine protease from *Bacillus altitudinis* strain BR1. Journal of Scientific and Industrial Research. 72: 166-171.
2) **Shamim, K.,** Sharma, J., Dubey, S.K. 2017. Rapid and efficient method to extract metagenomic DNA from estuarine sediments. 3 Biotech. 7:182-190.

Communicated

1) **Shamim, K.,** Sharma, J., Dubey, S.K., et al. 2017. Metagenomic screening and characterization of thermo-alkalophilic M23 metalloprotease from marine environment of Goa, India. PLOS ONE (communicated).
2) **Shamim, K.,** Sharma, J., Dubey, S.K., et al. 2017. Molecular cloning, over expression, nucleotide sequencing and characterization of thermostable alkaline protease from *Chryseobacterium* sp. Process Biochemistry (communicated).

Other publications

1) Naik, M. M., **Shamim, K.**, Dubey, S.K., 2012. Biological characterization of lead resistant bacteria to explore role of bacterial metallothionein in lead resistance. Current Science 103: 426-429.
2) **Shamim, K**., Naik, M. M., Pandey, A., Dubey S.K., 2013. Isolation and identification of *Aeromonas caviae* strain KS-1 as TBT and Lead resistant estuarine bacteria. Environmental Monitoring and Assessment 185: 5243- 5249.
3) Sharma, J., **Shamim, Kashif.,** Dubey, S.K., Meena, R. M., 2016. Metallothionein assisted periplasmic lead sequestration as lead sulfite by *Providencia vermicola* strain SJ2A. Science of The Total Environment. 579:359-365.
4) Praveen Kumar, M.K., Shyama, S.K., **Shamim, K.**, Dubey, S.K., et al. 2017. Effects of gamma radiation on the early developmental stages of Zebrafish (Danio rerio. Ecotoxicology and Environmental Safety. 142: 95-101

RESEARCH PAPERS PRESENTED IN NATIONAL AND INTERNATIONAL CONFERENCES

- "Comparative studies to characterize alkaline protease using culture dependent and metagenomic approach" – presented at 56th International Annual Conference of Association of Microbiologists of India held at Jawaharlal Nehru University, New Delhi (Poster).
- "Metagenomic approach to isolate gene encoding novel alkaline proteases from coastal environment of Goa" – presented at 55th Annual National Conference of Association of Microbiologists of India held at Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu (Poster).

Workshop and Symposia attended

- Attended training program on Marine bacterial diversity at CSIR-CSMCRI institute at Bhavnagar, Gujarat, India in 2012.
- Hands-on workshop on Phylogenetics, Goa University, Goa, 2013.
- Attended UGC-Sponsored short term course in Research methodology for Science students, Goa University, Goa, 2016.
- Organised a national conference NCYR-2017, Goa University, Goa, 2017.