Chapter 2

REVIEW OF LITERATURE
2.1 PROTEASE

Proteases constitute a very large and complex group of hydrolytic enzymes that catalyze the complete hydrolysis of peptide bonds and degrade proteins into small peptides and amino acids, in presence of a water molecule (Rawlings et al. 2012; Sarrouh et al. 2012). They are the single class of enzymes which occupy about 65% of the global industrial enzyme market with respect to their applications in both physiological and commercial fields (Kumar et al. 2008; Ningthoujam and Kshetri 2010; Siala et al. 2012).

![Percentage distribution of various Enzymes in Global Enzyme Market](image)

**Figure 2.1: Percentage distribution of various Enzymes in Global Enzyme Market**

Proteases find major applications in food and dairy industries (Inouye 2002; Vishwanath et al. 2010), meat tenderization (Ashie et al. 2006), detergent and cleaning compositions (Ribitscha et al. 2010), pharmaceutical and medicinal industries as drugs and in therapeutics (Cudic and Fields 2009; Deu et al. 2012), leather industries, hydrolysis of gelatin layers, fibrous proteins of horn, feather and hair, in peptide synthesis, recovery of silver from used X-ray films, resolution of racemic mixture of amino acids, wool
treatment, degumming of raw silk, waste management and bioremediation processes (Soccol et al. 2003; Akcan and Uyar 2011; Stalin et al. 2012).

2.1.1 Sources of Proteases

Proteolytic enzymes are ubiquitous in occurrence, being found in a wide diversity of living organisms such as plants, animals, and microorganisms and are physiologically essential for cell growth and differentiation (Gupta et al. 2002; Deu et al. 2012).

2.1.1.1 Plant Proteases

Plant proteases are involved in many aspects of plant physiology and development (Van der Hoorn 2008; Mahajan and Badgujar 2010). Proteases like papain, bromelain and ficin have been identified from latex of several plant families such as Asteraceae, Caricaceae, Moraceae, Convulvalaceae, Asclepiadaceae, Apocynaceae, Euphorbiaceae etc. and used in various industries like food, leather, medicine, therapeutics, production of emulsifiers and essential amino acids, in waste water treatment and many more. (Domsalla and Melzig 2008, Gonzalez-Rabade et al. 2011). Production of plant proteases is governed by several factors as availability of land for cultivation, suitability of climatic conditions for growth and other environmental factors. In addition, the long cultivation periods between planting and harvesting make selection of high-yielding strains difficult, thus resulting in expensive products (Rao et al. 1998).

2.1.1.2 Animal Proteases

Animal proteases have the most ancient history of use with pancreatic, trypsin, chymotrypsin, pepsin and rennin being the most important proteases that are derived
from porcine and bovine sources (Inouye 2002; Maheshwari et al. 2010). However, their production depends on the availability of livestock for slaughter, which is controlled by government policies (Rao et al. 1998). Hence these enzymes are expensive to be used abundantly in food or baking industries. Also the presence of side effects as development of bitter taste, rancid or soapy flavors inhibit the use of these enzymes in food industry (Mahajan and Badgujar 2010). Nevertheless, they find use in the preparation of bacterial media, in specialized medical, diagnostic and analytical applications (Rao et al. 1998). Exceptionally, rennet is used extensively in the dairy industry to produce a stable curd with good flavour (Vishwanath et al. 2010).

2.1.1.3 Microbial Proteases

As the limitations of cost, availability and properties of plant and animal proteases make their applications restricted, the focus of interest has shifted to microbial proteases in view of the increased enzyme demands worldwide (Rao et al. 1998; Kumar et al. 2008). Microbial proteases are among the most important hydrolytic enzymes accounting for approximately 40% of the total worldwide enzyme sales (Mahajan and Badgujar 2010; Maheshwari et al. 2012). Microbial community provide an excellent source of enzymes owing to their rapid growth, limited space required for cell cultivation, broad biochemical diversity and simplicity for generation of new recombinant enzymes with desired properties (Soccol et al. 2003; Ningthoujam and Kshetri 2010; Wilson and Remigio 2012). A variety of microorganisms such as bacteria, fungi, yeast and actinomycetes are known to produce protease, of which bacterial and fungal enzymes are the most preferred sources (Kumar and Takagi 1999; Gupta et al. 2002).
(i) Fungal Protease

Filamentous fungi are used in many industrial processes for the production of enzymes and metabolites (Wang et al. 2005; Chutmanop et al. 2008; Ire et al. 2011). Various strains of *Penicillium, Aspergillus, Trichoderma, Rhizopus* etc. are known to produce acidic, neutral and alkaline proteases which have applications in baking, food processing, animal feed, and pharmaceutical industries (Wang et al. 2005; Xiao-Lan et al. 2005; Hajji et al. 2008; Bhavsar et al. 2012). Fungi provide a wider variety of enzymes than bacteria in terms of pH (4 to 11) and substrate specificity; however, they have a lower reaction rate and worse heat tolerance than the bacterial enzymes (Rao et al. 1998; Mahajan and Badgujar 2010).

(ii) Bacterial Protease

Proteases are produced by many bacterial strains like *Bacillus, Vibrio, Staphylococcus, Pseudomonas, Aeromonas, Alcaligenes*, (Ducros et al. 2009; Nisnevitch et al. 2010; Tang et al. 2010; Pandey et al. 2010; Yun et al. 2012) among which, *Bacillus* strains are the most important and specific producers of extracellular commercial proteases (Schallmey et al. 2004; Ghorbel-Frikha et al. 2005). *Bacillus* constitutes a diverse group of rod-shaped, endospore-forming, aerobic or facultatively anaerobic, gram-positive bacteria with broad physiological diversities, widely distributed in soil, water and air and could be isolated from almost every kind of environment including desert sands, soils from tropical to temperate zones, air, snow, cold and thermal waters, fresh and preserved foods, milk, animal and human faeces, birds, amphibians and fish (Bhunia et al. 2010; Maughan and Van der Auwera 2011; Agrahari and Wadhwa 2012). Various *Bacillus*
species as *B. cereus*, *B. anthracis*, *B. megaterium*, *B. stercorarius*, *B. licheniformis*, *B. sphaericus*, *B. mojavensis*, *B. subtilis* are known to produce protease (Chung et al. 2006; Balamaran and Prabakaran 2007; Bhunia et al. 2010; Mahajan and Badgujar 2010). *Bacillus* species have a high capacity to secrete proteases into the extracellular medium during their transition to the stationary phase; much more than that required for their physiological activities (Joo and Chang 2005; Todar 2006; Liu et al. 2010).

### 2.1.2 Physiological Properties of Microbial Proteases

Proteases play an important role in normal and abnormal biological and physiological conditions by catalyzing various metabolic reactions extending from cellular to organ and organism level including cell-cycle progression, cell signalling, proliferation and death, protein trafficking, immune response and regulating various enzymatic cascades such as haemostasis and inflammation (Barrett et al. 2004; Deu et al. 2012). Microbial proteases may be intracellular and/or extracellular (Gupta et al. 2002). The extracellular proteases are extensively exploited commercially in various industrial fermentation processes including production of enzymes, biopesticides, vitamins, antibiotics etc. (Kumar and Takagi 1999; Joo and Chang 2005; Schallmey et al. 2004).

### 2.1.3 Proteases: Nomenclature and Terminology

Proteases are also known as peptidyl-peptide hydrolases or proteinases indicating that they hydrolyze peptide bond. According to the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (1992), proteases are enzymes of class 3, hydrolases; subclass 3.4, the peptide hydrolases and represented as
EC 3.4.11-25 and 99 (Rawlings et al. 2012). A peptidase is classified into families based on sharing significant similarities in amino acid sequences and into clans based on catalytic mechanism (Rawlings and Barrett 1993). Each family of peptidases is denoted by a code letter indicating the type of catalysis, i.e. S, C, A, M or U for serine, cysteine, aspartic, metallo, or unknown type, respectively (Polgar 2005). At present over 66,000 peptidase protein sequences have been classified into 50 clans and 184 families (Rawlings et al. 2012). However, due to wide diversity of action and structure, proteases do not comply easily with the general system of enzyme nomenclature (Rao et al. 1998).

### 2.1.4 Classification of Proteases

Proteases are classified on the basis of three categories (Barrett et al. 2004; Rao et al. 1998).

i. **Site of action**: Proteases are widely subdivided into two groups- exopeptidases and endopeptidases, depending on their site of action (Wu and Chen 2011; Deu et al. 2012). Exopeptidases cleave the peptide bond proximal to the amino or carboxy termini of the substrate, whereas endopeptidases cleave peptide bonds distant from the termini of the substrate (Ire et al. 2011).

ii. **Amino acid required for the catalytic function**: There are seven distinct classes of proteases grouped according to the amino acid or ion that catalyzes peptide bond cleavage namely serine, aspartic, cysteine, metallo, glutamic acid, threonine and the newly identified asparagine peptide lyases (Rawlings et al. 2012).

iii. **Optimum pH for their activity**: Proteases are acidic, alkaline or neutral on the basis of the pH at which they exhibit their optimum activity (Lee et al. 2006).
2.1.4.1 Exopeptidases

The exopeptidases act only near the ends of polypeptide chains. Based on their site of action at the N or C terminus, they are classified as amino and carboxypeptidases, respectively (Sumantha et al. 2006). The secondary and tertiary structure of protein substrate usually prevents attack by exopeptidases (Lodish et al. 2000).

(i) Aminopeptidases

Aminopeptidases act at a free N terminus of the polypeptide chain and liberate a single amino acid residue, a dipeptide, or a tripeptide. They are known to remove the N-terminal Met that may be found in heterologously expressed proteins but not in many naturally occurring mature proteins (Rao et al. 1998; Maheshwari et al. 2010).

(ii) Carboxypeptidases

The carboxypeptidases act at C terminals of the polypeptide chain and liberate a single amino acid or a dipeptide. They are divided into four major groups - serine, metallo cysteine and omega carboxypeptidases, based on the nature of the amino acid residues at the active site of the enzymes (Rao et al. 1998; Kumar et al. 2008).

(iii) Omega Peptidases

These are exopeptidases acting on peptide as acyl aminoacyl-peptidase, cleave N-acetyl or N-formyl aminoacid from the N-terminus of a polypeptide, Y-Gly-X- carboxy peptidase 3.4.19.9) cause cleavage of a Y-glutamyl bond to release an unsubstituted terminal amino acid (Maheshwari et al. 2010).
2.1.4.2 Endopeptidases

Endopeptidases are characterized by their preferential action on the peptide bonds in the inner regions of the polypeptide chain away, from the N and C termini. The presence of the free amino or carboxyl group has a negative influence on enzyme activity (Rao et al. 1998). They are divided into seven subgroups based on their catalytic mechanism (Rawlings and Barrett 1993; Rawlings et al. 2011).

(i) Serine proteases:

Serine proteases are a family of enzymes in the exo, endo, oligo, and omega peptidase groups that utilize a serine residue as a nucleophile in the substrate-binding pocket to hydrolyze peptide bonds (Rao et al. 1998). They also have essential aspartate and histidine residues which, along with serine, forming the catalytic triad (Polgar 2005). They constitute almost one-third of all proteases and are found abundantly among viruses, bacteria, and eukaryotes in co-ordinating various physiological functions, including digestion, immune response, blood coagulation and reproduction (Rawlings and Barrett 1993; Mahajan and Badgujar 2010).

(ii) Aspartic proteases

Aspartic acid proteases also known as acidic proteases, are endopeptidases mainly of vertebrate, fungal and retroviral origin that use an aspartate residue for catalysis of their peptide substrates (Sumantha et al. 2006). More recently, bacterial and archaean aspartic endopeptidases have been studied (Taylor and La Pointe 2000; Jarrell et al. 2006). Microbial aspartic proteases can be broadly divided into two groups (i) pepsin-

(iii) **Cysteine protease**

Cysteine or thiol proteases are enzymes that degrade polypeptides by formation of a covalent intermediate and involve a nucleophilic cysteine thiol and a histidine residue in a catalytic dyad. They occur in both prokaryotes and eukaryotes. Based on their side chain specificity, they are divided into four groups: (i) papain-like (ii) trypsin-like with preference for cleavage at the arginine residue (iii) specific to glutamic acid (iv) others (Ward et al. 2009).

(iv) **Metalloproteases**

Metalloproteases are the most diverse of all catalytic protease types that include both endoproteases and exoproteases (Ward et al. 2009), characterized by requiring a divalent metal ion like zinc, cobalt, manganese or nickel for their catalytic activity and inactivated by addition of chelating agents (Keith et al. 2000) In these enzymes, a divalent cation, usually zinc, is held in place by amino acid ligands, usually three in number (Fukasawa et al. 2011). These are produced by several species of plants, animals, and microorganisms including bacteria and fungi (Mahajan and Badgujar 2010).
(v) Threonine proteases

Threonine proteases having an N-terminal threonine at the active site were first discovered from *Saccharomyces cerevisiae* in 1995 as part of proteasome catalytic subunits, the large protein-degrading apparatus complex (Ward et al. 2009).

(vi) Glutamic acid proteases

Glutamic acid proteases were first described in 2004 with an active site containing a diad glutamic acid and glutamine that play a critical role in substrate binding and catalysis. These amino acids along with their associated water molecules act as nucleophiles to exhibit an acid-base mechanism distinct from that of the aspartic proteases (Ward et al. 2009).

(vii) Asparagine peptide lyases

Asparagine peptide lyases are the new class of proteases discovered by Rawlings et al. (2011). Till recently all proteolytic enzymes have been considered to be hydrolases (EC 3.4). However, the discovery of asparagine peptide lyases that cleave themselves at asparagine residues indicates that not all peptide bond cleavage occurs by hydrolysis. These enzymes are not peptidases because breaking the peptide bonds does not involve hydrolysis. The nucleophile in the reaction is an asparagine, which represents a new proteolytic catalytic type (Rawlings et al. 2011).

Fig 2.2 demonstrates the wide classification of proteases as described above
Figure 2.2: Classification of Proteases according to the NC-IUBMB (1992)

2.2 METALLOPROTEASES

Metalloproteases are the most diverse proteases containing a divalent metal ion at the active site (Rawlings and Barrett 1993; Mansfeld 2007). In most cases the metal ion is zinc as it is an integral component of many proteins involved in virtually all aspects of metabolism of the different species of all phyla, but in some cases it is cobalt, manganese, or nickel. The catalytic metal ion is usually coordinated by three amino acid side chain ligands. The role of the metal ion is to activate a water molecule, which serves as a nucleophile in catalysis and also coordinates with the metal ion as a fourth ligand (Nagase 2001; Fukasawa et al. 2011).
2.2.1 Classification of Metalloproteases

According to the latest release of MEROPS database (9.8), metalloproteases comprise of 68 families with 43 subfamilies. Thirteen families contain the sequence HEXXH, which provides two of the three ligands for the zinc atom (Rawlings et al. 2012). Some other mono zinc proteases have different zinc-binding motifs, for example, HxxE(D)-aan-H in the carboxypeptidase family or HxD-aa12-H-aa12-H in the matrix metalloprotease family (Fukasawa et al. 2011). Nine families include (not exclusively) bacterial extracellular metalloproteases (BEMPs). As shown in Table 2.1, BEMPs are distributed among the metalloprotease families M4, M5, M9, M10, M12, M13, M23, M30, and M34. Metalloproteases in the family M4 are mostly BEMPs. In family M23, HEXXH motif is replaced by HXH (Mansfeld 2007; Wu and Chen 2011). Families of metalloproteases have been grouped into a total of 16 different clans based on the protein structure, homology and nature of the amino acid that completes the metal-binding site; e.g., clan MA has the sequence HEXXH-E and clan MB corresponds to the motif HEXXH-H (Wu and Chen 2011; Rawlings et al. 2012). Metalloproteases from clans MA, MC, MD, ME, MJ, MK, MM, MO and MP require only one catalytic metal ion, in most cases zinc ions, whereas clans MF, MG, MH, MN and MQ contain two metal ions acting co-catalytically on the substrate (M stands for metalloprotease). The majority of metalloproteases contain one catalytic metal ion, and those containing two catalytic metal ions so far described are all exoproteases (Mansfeld 2007; Wu and Chen 2011). Metalloproteases with HEXXH zinc-binding motif can be further classified into thermolysin, serralysin, and neurotoxin families, according to the location of the third zinc ligand (Fukasawa et al. 2011).
Table 2.1: Representative bacterial extracellular metalloproteases in each MEROPS family (Wu & Chen 2011)

<table>
<thead>
<tr>
<th>MEROPS Family</th>
<th>Representative</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>M4</td>
<td>Thermolysin</td>
<td>Bacillus thermoproteolyticus</td>
</tr>
<tr>
<td>M5</td>
<td>Mycolysin</td>
<td>Streptomyces cacaoi</td>
</tr>
<tr>
<td>M9</td>
<td>Bacterial collagenase V</td>
<td>Vibrio alginolyticus</td>
</tr>
<tr>
<td></td>
<td>Bacterial collagenase H</td>
<td>Clostridium histolyticum</td>
</tr>
<tr>
<td>M10 B</td>
<td>Serralysin</td>
<td>Serratia sp. E-15</td>
</tr>
<tr>
<td></td>
<td>Aeruginolysin</td>
<td>Pseudomonas aeruginosa</td>
</tr>
<tr>
<td>M10 C</td>
<td>Fragilysin</td>
<td>Bacteroides fragilis</td>
</tr>
<tr>
<td>M12</td>
<td>Flavastacin</td>
<td>Chrysebacterium meningosepticum</td>
</tr>
<tr>
<td></td>
<td>Myroilysin</td>
<td>Myroides profundiD25</td>
</tr>
<tr>
<td>M13</td>
<td>PepO</td>
<td>Lactococcus lactis</td>
</tr>
<tr>
<td>M23 A</td>
<td>β-lytic metalloendopeptidase</td>
<td>Lysobacterenzymogenes</td>
</tr>
<tr>
<td>M23B</td>
<td>Lysostaphin</td>
<td>Staphylococcus simulans</td>
</tr>
<tr>
<td>M30</td>
<td>Hyicolysin</td>
<td>Staphylococcus aureus</td>
</tr>
<tr>
<td>M34</td>
<td>Anthrax lethal factor</td>
<td>Bacillus anthracis</td>
</tr>
</tbody>
</table>

Based on the specificity of their action, metalloproteases can be divided into four groups (i) neutral (ii) alkaline (iii) *Myxobacter* I and (iv) *Myxobacter* II. The neutral proteases show specificity for hydrophobic amino acids, while the alkaline proteases possess a very broad specificity. *Myxobacter* protease I is specific for small amino acid residues on either side of the cleavage bond, whereas *Myxobacter* II is specific for lysine residue on the amino side of the peptide bond. All of them are inhibited by chelating agents such as EDTA, DTT, but not by sulfhydryl agents or DFP (Diisopropyl fluorophosphate) II (Nagase 2001; Ward et al. 2009).
2.2.2 Structure of Metalloprotease

The X-ray structures of metalloproteases show that the metal ion in catalytic sites is generally coordinated to the side chain of three amino acid residues, a combination of histidine, glutamate, aspartate and cysteine, and a solvent molecule completes the tetrahedral coordination sphere (Keith et al. 2000; Fukasawa et al. 2011). A feature common to all zinc sites is that the metal ion is surrounded by a shell of hydrophilic groups that is embedded within a larger shell of hydrophobic groups. The crystallographic studies show that most of the known metalloproteases contain a His-Glu-Xaa-Xaa-His (HEXXH) motif forming an \( \alpha \)-helix which is well conserved in the active metal-binding site where two histidine residues coordinate with the zinc ion (Fukasawa et al. 2011). The BEMPs also have the same structures with some variation (Fig 2.3) (Wu and Chen, 2011). Most BEMPs only contain one catalytic domain with molecular mass (Mm) ranging from 20 to 35 kDa, while some contain more than one catalytic domain, such as the anthrax lethal factor that has four domains with a Mm up to 90.2 kDa. Despite the difference in their sequences and size, all BEMPs contain a catalytic \( \text{Zn}^{2+} \) at their active centre. Deprivation of the catalytic \( \text{Zn}^{2+} \) by a chelator can lead to loss of the activity of a metalloprotease (Nagase 2001; Wu and Chen 2011).

2.2.3 Metal Substitutions of Monozinc Metalloproteases

Almost all metalloproteases are mono zinc enzymes; however some enzymes contain two zinc ions for catalytic domains. Few metalloproteases with dicobalt or dimanganese ions have also been reported (Nagase 2001; Fukasawa et al. 2011; Wu and Chen 2011).
Substitution studies have been conducted where zinc in numerous zinc metalloproteases for example, astacin, carboxypeptidase A and thermolysin, dipeptidyl peptidase (DPP) III etc. has been substituted by other divalent cations to probe the role of the metal for catalysis and structure. The enzymes showed complete restored or even excess activity. However, no direct relationship between the metal tolerance and the metal coordination structure of zinc metalloproteases has been established (Fukasawa et al. 2011).

### 2.2.4 Function Mechanism

The 3-D structures of many metalloproteases have suggested the mechanisms of peptide bond hydrolysis. During catalysis, the Zn$^{++}$ promotes nucleophilic attack on the carbonyl carbon by the oxygen atom of a water molecule at the active site. An active site base a glutamate residue in carboxypeptidase facilitates this reaction by extracting a proton from the attacking water molecule (Fig 2.4) (Fukasawa et al. 2011; Deu et al. 2012).
2.2.5 Properties of Metalloprotease

Metalloproteases exhibit deviant physiological and biochemical properties that account for their pathophysiological and industrial applications (Maheshwari et al. 2010).

2.2.5.1 Physiological and Biochemical properties

Metalloproteases play multifaceted role ranging from cell proliferation, differentiation and remodeling of the ECM to embryonic development, morphogenesis, processing of peptide hormones, release of cytokines and growth factors, cell-cell fusion, cell adhesion and migration, vascularization, intestinal absorption of nutrients, viral polyprotein processing, bacterial cell wall biosynthesis, and metabolism of antibiotics (Nagase 2001;
Wu and Chen 2011). The aberrant activities of metalloproteases have been implicated in diseases such as arthritis, cancer, cardiovascular diseases, nephritis, disorders in the central nervous system, fibrosis, and infection etc. (Nagase 2001).

- Chelating agents such as EDTA, 1,10-phenanthroline are commonly used to block metallopeptidase activities in vitro experiments (Van den Burg and Eijsink 2004). Synthetic inhibitors contain specific chelating moiety have also been designed to inhibit specific metalloproteases (Nagase 2001). As metalloproteases exhibit various physiological roles in therapeutics and disease manifestation, their inhibitors are considered good therapeutic targets as the enzyme activities are likely to be regulated by the endogenous inhibitors (Mansfeld 2007).

- Collagenolytic and gelatinolytic metalloproteases degrade the ECM which leads to physiological and pathological conditions such as embryonic and foetal bone development, wound repair, rheumatoid arthritis, intestinal ulceration and chronic periodontal inflammation, tissue remodelling repair and growth, proliferation metastasis and invasion of cancer cells and malignant tumours (Nagase 2001; Watanabe 2004). In contrast to vertebrate collagenases, the bacterial collagenases possess broad substrate specificity and degrade both native collagen and gelatin (Grass et al. 2004; Liu et al. 2010).

- Fibrinolytic metalloproteases play an important role in the regulation of cellular fibrinolysis thus contributing in prevention and cure of thrombotic diseases (Agrebi et al. 2010; Lee et al. 2012). These enzymes are successively discovered from different microorganisms, the most important among which is the genus Bacillus (Xiao-Lan et al. 2005).
• Microbial neutral metalloproteases are thermostable enzymes with pH optima around 7.0-8.0, while remaining active in a broad pH range. The molecular weights of the neutral BEMPs are found to be in the range 20 to 35 kDa (Fukasawa et al 2011). The alkaline metalloproteases are active in the pH range from 7 to 9 and have molecular masses in the region of 48 to 60 kDa (Rao et al 1998).

2.2.6 Sources of Metalloprotease

Metalloproteases are produced by several species of plants, animals, and microorganisms including bacteria and fungi (Rao et al. 1998). BEMPs have been identified in both gram-positive and gram-negative pathogens, but they are certainly not exclusive to pathogenic species (Ward et al. 2009). Various strains of Bacillus (aB. cereus, B. thuringiensis, B. anthracis, B. megaterium, B. thermoproteolyticus B. subtilis.), Alicyclobacillus, Staphylococcus, Pseudomonas, Streptomyces, Clostridium, Vibrio, Listeria, Serratia etc (Nagase 2001; Ghorbel-Frikha et al. 2005; Chung et al. 2006; Mansfeld 2007; Kim et al. 2007; Wu and Chen 2009; Simkhada et al. 2010; Ghorbel-Bellaaj et al. 2012) have been reported to produce metalloproteases.

2.2.7 Applications of Metalloproteases

Bacterial metalloproteases possess a wide global market base when it comes to their exploitation in various industries. A lot of commercial microbial proteases such as thermolysin (Sigma, Merck); Thermoase PC10F (Amano Enzyme Inc., Japan), Neutrase (Novo Nordisk, Denmark), Protin PC10F (Amano Enzyme Inc., Japan), and the highly stable TLP-ste variant Boilysin (Groningen, The Netherlands) are available in the enzyme market for various applications (Wu and Chen 2011). These proteases are
widely used in food, medicine, brewing, leather, film, and baking industries (Rao et al. 1998; Mansfeld 2007; Wu and Chen 2011).

2.2.7.1 Food industry

The oldest discovered metalloprotease thermolysin has been studied and used for the synthesis of N-carbobenzoxy-L-aspartyl-L-phenylalanine methyl ester (Z-Asp-Phe-OMe), the precursor to the artificial sweetener aspartame (Inouye 2002). Milk clotting metalloproteases find application in cheese making (Vishwanath et al. 2010). Some metalloproteases in combination with other proteases are used to hydrolyze food proteins to produce flavour and taste enhancing peptides in food industry (Wu and Chen 2011). Neutrase is used in a concoction of proteases to accelerate the ripening of dry fermented sausages for flavour development. Milk protein hydrolysates are used for the enhancement of cheese products, soy and wheat hydrolysates are used as the flavour enhancer of soups and sauces while meat hydrolysates find application in the enhancement of the flavour of meat products, soups, sauces and other instant products (Sumantha et al. 2006; Mansfeld, 2007). Neutral metalloproteases show specificity for hydrophobic amino acids generating less bitterness in hydrolyzed food proteins and hence are valuable in food industries, the low thermostability being an advantage for controlling their activity (Watanabe 2004; Daboor et al. 2010). Collagenolytic proteases tenderize meat by digesting collagens (Kumar and Takagi 1999; Ashie et al. 2006).

2.2.7.2 Therapeutics and Pathogenesis

Metalloproteases find important application in therapeutics and pathogenesis. Collagenolytic enzymes have been employed in clinical therapy as antimicrobial agents,
for removal of necrotic tissue from burns, wound healing, ulcers, treatment of sciatica and herniated intervertebral discs, in isolation of pancreatic islets for transplantation, treatment of dupuytren disease, pretreatment for enhancing adenovirus-mediated cancer gene (Grass et al. 2004; Liu et al. 2010; Jovanovic et al. 2012). Oral administration of collagen peptides prevents osteoporosis, gastric ulceration, relaxes hypertension, and stimulates skin metabolism and are effective as skin moisturizers (Watanbe 2004; Salamone et al. 2012). Metalloproteases with virulence have a role in pathology and can be used as targets in drug and vaccine development and have agricultural and veterinary importance (Cudic and Fields 2009). Metalloproteases from Bacillus thuringiensis degrades antibacterial proteins produced by the insect host (Saxena and Singh 2011a).

Fibrinolytic or thrombolytic agents convert plasminogen to plasmin, lyse the clot by breaking down the fibrin contained in a clot. Thrombotic diseases and disorders such as myocardial infarction, cerebrovascular thrombosis, pulmonary embolism and venous thrombolism are life-threatening for human beings and account for heavy toll in death and disability worldwide (Hassanein et al. 2011; Agrebi et al. 2010). Currently, several thrombolytic agents such as streptokinase, urokinase, prourokinase, reteplase (r-PA), alteplase (t-PA), reptilase, brinase and anisoylated purified streptokinase activator complex (APSAC) are available for clinical use which have significant shortcomings of large therapeutic dose, short plasma half-life, limited fibrin specificity, reocclusion and bleeding complications (Agrebi et al. 2010; Simkhada et al 2010). Hence cheaper and safer thrombolytic agents as fibrinolytic enzymes have been purified from many plant, animal and microbial sources for therapy of thrombotic diseases (Xiao-Lan et al. 2005; Balaraman and Prabakaran 2007; Agrebi et al. 2010).
2.2.7.3 Synthesis of Peptides

Peptide synthesis has become increasingly important for the food and pharmaceutical industries over the last few decades. The main application of peptides is their use as low-calorie sweeteners as mentioned above (Inouye 2002). Additionally, some metalloproteases as thermolysin and neutrase are often employed to hydrolyze proteins for the production of novel peptides with various bioactivities, such as antioxidant activity and angiotensin-I converting enzyme inhibitory activity owing to which they are used in health industry as drugs in the treatment of various diseases (Cudic and Fields 2009; Deu et al. 2012). Low-molecular weight hydrolysis products of protamine have been tested successfully as a delivery system for DNA in gene therapy (Mansfeld 2007).

2.2.7.4 Other Applications

Neutral metalloproteases find industrial importance as they cleave the amylase precursor into ρ and α amylases playing a role in digestion of starch and glycogen (Saxena and Singh 2011a). In baking industry, metalloproteases are used for reducing gluten–protein cohesiveness to produce the desired dough in cake baking and for the modification of proteins in bread manufacturing (Lee et al. 2006). Further commercial applications of metalloproteases can be found in the brewing industry for improved filtration of beer, reduced calorie content, in the leather industry in bating and dehauling, processing of slaughter waste, improvement in baking characteristics of flour, and in the film industry for recovery of waste silver (Mansfeld 2007; Wu and Chen 2011). An interesting application of immobilized thermolysin is the removal of protein coatings from the
surface of old documents and art work (Van den Burg and Eijsink 2004). Inhibitors of metalloproteases involved in diseases are of potential therapeutic use (Nagase 2001).

2.3 ENZYME PRODUCTION

Enzyme production by microorganisms has grown significantly during the past few decades in response to the expanding enzyme market. The overall process of enzyme production includes selection of a suitable microorganism, formulation of media, fermentation and purification process with a final quality control of the finished product (Hatti Kaul 2007; Singh and Pandey 2009).

2.3.1 Selection of Microorganism

Screening and selection of microorganism is based on the following criteria using appropriate selection procedures (Hatti-Kaul 2007; Maheshwari et al. 2010).

(i) GRAS (generally regarded as safe) status, which implies that they must be non-toxic, non-pathogenic and generally should not produce antibiotics.

(ii) Produce the desired enzyme in higher amount than other metabolites.

(iii) Rapid growth for short production and fermentation process.

(iv) Utilization of low cost culture media and minimum production requirements and other parameters for economic production of the enzyme.

2.3.2 Medium Formulation and Preparation

Enzyme production is highly influenced by media components. For mass production of enzymes, it is important to customize and formulate the fermentation media ingredients
according to the host and type of enzyme produced in maximum level (Hatti-Kaul 2007; Mahalaxmi et al. 2010). An ideal production medium includes carbon and nitrogen sources, amino acids, growth promoters, trace elements minerals and metal ions and little amount of salts (Ghorbel-Frikha et al. 2005; Bhunia et al. 2010; Singh et al. 2011).

2.3.3 Fermentation Process

Fermentation is biological conversion of complex substrates into simple compounds utilizing microorganisms such as bacteria and fungi (Mahalaxmi et al. 2010; Subramaniyam and Vimala 2012). The substrates used in fermentation vary widely utilizing any natural or synthetic material that supports microbial growth yielding varied metabolic products (Chisti 1999). These products range from industrial enzymes, bioinsecticides and antibiotics to peptides and growth factors that possess biological activity and are industrially and economically important (Raimbault 1998; Maheshwari et al. 2010; Mienda et al. 2011). Other than media components, the fermentation process is also affected by physical parameters as pH, temperature, agitation and aeration, time etc. (Walker et al. 2000; Hatti-Kaul 2007).

Two types of fermentation processes are used for production of enzymes; solid state and submerged fermentation. Both the processes are applied commercially and have their own advantages and disadvantages (Chisti 1999; Subramaniyam and Vimala 2012).

2.3.3.1 Solid State Fermentation

SSF is a well established and effective biotechnological tool, which holds tremendous potential for production of a large variety of enzymes using microorganisms (Hatti-Kaul
2007; Pandey et al. 2008). The process involves cultivation of microorganisms under controlled conditions, in absence of free-flowing water for production of enzymes (Toca-Herrera et al. 2007; Singh and Pandy 2009; Subramaniyam and Vimala 2012). Over the past decade, SSF has gained a tremendous momentum owing to certain advantages over conventional submerged fermentation like low production cost and energy requirements, higher volumetric productivity of enzymes with high thermal or pH stability, easier and simpler processing equipments, improved oxygen circulation, less waste effluent problem and minimum efforts in downstream processing (Soares et al. 2005; Chutmanop et al. 2008; Singh et al. 2010). However, direct comparison between SSF and SmF is difficult because the two processes are quite different (Toca-Herrera et al. 2007). An increase in 3 fold of product yield is always expected with SSF compared to that of submerged fermentation (Raimbault 1998; Soccol et al. 2003).

The major factors which affect a SSF process include:

(i) Selection of Microorganism

Selection of strain is one of the most important criteria in SSF and is based on the factors as the potential of the strain to grow and produce in SSF, its efficiency to produce the desired product as the major end product in the process (Walker et al. 2000). This warrants through the characterization of individual microorganism to evaluate its potential at commercial level (Prakasham et al. 2006; Mahalaxmi et al. 2010).

(ii) Substrate

The selection of a substrate for SSF process depends mainly upon several the cost and availability (Naidu and Devi 2005). In SSF process, the solid substrate not only supplies
the nutrients to the microbial culture growing in it, but also serves as an anchorage for the cells (Singh and Pandey 2009). Agro-industrial residues offer potential advantages for their application as substrates owing to their mineral content, cost and availability (Chutmanop et al. 2008; Prakasham et al. 2006; Sindhu et al. 2009). However, some essential nutrients may be present in sub-optimal concentrations, or absent in these substrates. In such cases, it would be necessary to supplement them externally (Mienda et al. 2011). Various agro-industrial residues such as wheat and rice bran, husk, oilseed cakes, sugarcane bagasse, coconut flesh, etc. have been used as single or in combination for the production of various enzymes (Soares et al. 2005; Chutmanop et al. 2008; Pandey et al. 2008).

(iii) Moisture Content

Substrate moisture and water activity play very important role in SSF. The fundamental water activities in a SSF system can be distinguished as (i) mass transfer of the water and solutes across the cell membrane to provide nutrients and scavenges wastes, or metabolites, under the dissolved form (ii) stabilization of the structure of the biopolymers, such as proteins, nucleotides and carbohydrates (iii) stabilization of the lamellar structure at cellular level. These parameters control and modify the metabolic activity of the micro-organism (Pandey et al. 2008; Gervais and Molin 2003).

(iv) Particle Size

Solid substrates should have generally large surface area per unit volume $10^3$-$10^6$ m$^2$/cm$^3$ for the ready growth on the solid/gas interface (Pandey and Ramachandran 2005).
Other factors effecting the SSF process are pre-treatment of the substrate, relative humidity, removal of metabolic heat generated during process, supplementation of carbon nitrogen sources and inducers, mineral nutrients and trace elements, maintenance of uniformity in the gaseous atmosphere and environment etc. (Wang et al. 2006; Hatti-Kaul 2007; Singh and Pandey 2009; Sindhu et al. 2009; Mienda et al. 2011)

2.3.3.2 Submerged Fermentation

Submerged fermentation is very commonly employed by enzyme manufacturers for commercial production which involves the production of enzymes through microbial growth in a rich broth of nutrients and a high concentration of oxygen (Singh et al. 2010; Bhavsar et al. 2012; Subramaniyam and Vimala 2012). Cheap carbon and nitrogen sources supplemented with protein rich sources like soybean meal, casein, fish flour are generally used in submerged fermentation, however, synthetic media is also used for enzyme production commercially (Boominadhan et al. 2009; Ninghoujam et al. 2010; Bhunia et al. 2010). Batch-fed and continuous fermentation are the most commonly employed processes (Chisti 1999).

Submerged fermentation owes its advantages to ease of sterilization, easier process control, consistent enzyme production characteristics with defined medium and process conditions and easier downstream processing in spite of the cost-intensiveness for medium components (Walker 2000; Prakasham et al. 2006). The disadvantages associated with the process include high costs due to the expensive media and equipments, catabolic repression of enzyme synthesis and risk of contamination (Soccol et al. 2003; Wang et al. 2005; Singh and Pandey 2009).
2.3.4 Bioreactors for large scale production of Enzymes

Bioreactors or fermenters are employed for industrial production of biomolecules. Industrial fermentations may be carried out either batch wise, as fed-batch or as continuous cultures (Chisti 1999). The increase in enzyme demand has led to significant improvement in understanding of designing and operating of bioreactors (Singh et al. 2011). Different types of bioreactors as tray fermenters, static bed and tunnel fermenters, rotary drum fermenters, fluidized beds, agitated tank fermenters, continuous screw fermenters have been used in various production processes (Ducros et al. 2009; Singh et al. 2011; Wilson and Remigio 2012).

2.4 OPTIMIZATION OF PRODUCTION MEDIA

All fermentation processes are greatly influenced by various media components as carbon, nitrogen concentration, their ratio, other mineral salts, metal ions and physical factors such as pH, temperature, incubation time, aeration, inoculum size etc. and biological factors such as genetic nature of the organism, its biochemical/metabolic behaviour (Gupta et al. 2002; Ducros et al. 2009; Singh et al. 2011). Designing an appropriate fermentation medium thus becomes paramount in commercial practice for higher yield and productivity as each microorganism has its own special conditions for maximum product production owing to its physiology and genetic diversity (Kumar and Takagi, 1999; Reddy et al. 2008). However, to optimize all parameters one by one and to establish the best possible conditions by interrelating all parameters involves numerous experiments which is extremely time consuming and expensive (Hajji et al. 2008; Mabrouk 2008). Experimental designs based on statistical tools are generally preferred,
due to variety of recognized advantages as minimum time and energy requirement, enhanced product yield and reduced production cost (Li et al. 2006; Saxena and Singh 2010). Fractional factorial saturated designs as Plackett Burman (Hadamard matrix) designs used in the early stages of experimentation allow initial screening of the ingredients and understanding the significance of their effect on the process response (Xiao et al. 2007; Mabrouk 2008; Gupta et al. 2010). Each factor is tested equal number of times at its low and high settings, thereby setting up a balance between each and every pair of factors. As the design is orthogonal in nature, the effects of these factors as worked out are pure in nature and not confounded with interactions among them (Plackett and Burman 1946). A number of added tools as p values, normal and half normal plots and pareto charts are helpful in accessing the significance of each factor. The p value is the probability that the magnitude of a parameter estimate is due to random process variability. A low p value indicates a significant effect and provides a base line for determining the relative significance of the variables tested with respect to the response required. Out of these, generally the most effective components with positive effects are selected for further study, while those showing large negative effects may be dropped for all further experiments.

Second level optimization is required after the 2 factorial screening to optimize the required levels of these screened factors. Response surface methodology (RSM) is a collection of mathematical and statistical techniques to design experiments, build models to search optimum conditions of factors for desirable responses, evaluate and understand the interactions between different physiological and nutritional parameters and the relative significance of several affecting factors even in the presence of complex
interactions (Hajji et al. 2008; Gupta et al. 2010). Second-order models like CCD, FFCCD, Box–Behnken and Doehlert designs are widely employed in RSM as they can take on a wide variety of functional forms and this flexibility allows them to more closely approximate the influence of key factors on the true response surface by a small number of experiments (Li et al. 2006; Xiao et al. 2007). The central composite design (CCD) is commonly used in product optimization under RSM. A complete CCD experiment design allows estimation of a full quadratic model for each response. The design matrix for a CCD involving $k$ factors is derived from a matrix, $d$, containing three different parts (i) each factor having a minimum and a maximum levels coded as $+1$ and $-1$ (ii) a matrix $C$, a set of center points (iii) a matrix $E$, a set of axial points (Fig 3.3).

![Central Composite Design](image)

**Figure 2.5: A Central Composite Design with three input parameters**

The experimental data is processed by a second order polynomial equation:

$$Y = \beta_0 + \beta_1 A + \beta_2 B + \beta_{11}A^2 + \beta_{22}B^2 + \beta_{12}AB$$
Where $A$ and $B$ are the experimental variables, $\beta_0$ is the constant coefficient, $\beta_1$ and $\beta_2$ are the linear coefficients; $\beta_{11}$ and $\beta_{22}$ are the quadratic coefficients whereas $\beta_{12}$ is second order interaction coefficient.

Validation of the model developed is then carried out by comparison between the experimental and calculated values of response observed. If the calculated values fall close to the experimental value, the model is said to be fitted well. To validate a model, the lack of fit should be less than 50% of the experiments.

RSM has eliminated the drawbacks of classical methods and has emerged as a powerful and useful modelling and optimization technique for several bioprocesses, including various fermentations processes for production of metabolites as enzymes, antibiotics etc., biomass and spore production, enzyme immobilization techniques, waste treatment and many more (Wang et al. 2008; Saxena and Singh 2010; Singh et al. 2011).

**2.5 ENZYME PURIFICATION**

Downstream processing or purification of enzyme from the raw material is final and the most significant step in the production and is governed by some general principles that produce a homogenous enzyme of interest (Carta and Jungbauer 2010; Hatti-Kaul 2007). Enzyme purification aims to isolate a specific protein from a crude mixture utilizing the physical and/or chemical properties of the individual protein (Hedhammar et al. 2006; Sarrouh et al. 2012). The fundamental strategy of enzyme purification includes separation of enzyme from the solids, concentration to reduce the volumes and separate it from other soluble contaminants while retaining as much as possible of the protein of interest (Ward and Swiatek 2009; Linke and Berger 2011). It is a multi step process
Chapter 2

Review of Literature

(Berg et al. 2002) where first step involves the removal of macromolecules, nucleic acids, lipid micelles, polysaccharides, other proteins and insoluble suspended products from the production media, which is normally done by centrifugation (Lodish et al. 2000; Carta and Jungbauer 2010; Linke and Berger 2011). The extracellular enzymes remain in the broth after fermentation, while disruption of cells or tissue is required for the recovery of the intracellular enzymes (Hedhammar et al. 2006; Sarrouh et al. 2012). The succeeding steps involve separation of the protein of interest from other proteins.

2.5.1 Precipitation

This is the basic purification step where the protein solution is brought to a desired saturation level when most of the protein in the solution precipitates with the salt/solvent. Ammonium sulphate is the most common salt used for protein precipitation. Most enzymes exist in cell fluids as soluble proteins in physiological salt conditions and neutral pH at room temperature. Salting out of proteins depends strongly on hydrophobic interactions, other features like pH and temperature affecting the solubility. An increase in pH or decrease in temperature lowers the solubility of the protein resulting in precipitation (Berg et al. 2002).

Solvent precipitation is carried out with acetone, methanol, ethanol, acidified acetone/methanol, TCA depending upon the properties and concentration of the protein and the type of impurities present (Burgess 2006; Carta and Jungbauer 2010). When large amounts of a water-miscible solvent such as ethanol or acetone are added to a protein mixture, proteins precipitate out due to decrease of the dielectric constant, which
would make interactions between charged groups on the surface of proteins stronger (Wilson and Walker 2010).

Many protein solutions after centrifugation tend to have viscosity due to the presence of DNA and viscous polysaccharides and proteoglycans that interfere in the purification steps (Wilson and Walker 2010; Carta and Jungbauer 2010). In such cases precipitation with protamine sulfate or ethanol, addition of nucleases or glycosidases is employed (Lodish et al 2000; Ward and Swiatek 2009).

2.5.2 Chromatographic methods

Liquid chromatography is one of the most effective and widely used techniques for high resolution purification of enzymes based on the principle that molecules dissolved in a solution bind and dissociate with a solid surface (Lodish et al 2000). Physicochemical properties of proteins such as size, charge, hydrophobicity and biospecific interaction are used to separate mixtures of proteins, nucleic acids and other non-protein contaminants (Hedhammar et al. 2006; Berg et al. 2002). In a typical chromatography process when a solution is allowed to flow across the surface, the product binds to the adsorbent and thus move slowly, while the impurities do not bind and hence are washed away (Ward and Swiatek 2009; Carta and Jungbauer 2010). Liquid chromatography is performed in a column packed tightly with spherical beads. The nature of these beads determines whether separation of proteins depends on differences in size (size exclusion), mass (gel filtration), charge (ion exchange), or binding affinity (bio-affinity, hydrophobicity and antibody specificity) (Berg et al 2002; Hedhammar et al. 2006; Ward and Swiatek 2009; Carta and Jungbauer 2010). Other options include preparative electrophoresis e.g. disc
gel electrophoresis or isoelectric focusing (Berg et al. 2002; Scopes and Smith 2006). Foaming offers an alternative minimizing the processing steps, preserving purification efficiency and decreasing activity losses all at the same time (Linke and Berger 2011).

2.5.3 Enzyme Recovery

Protein purification protocols involve several steps where each step has the potential to achieve a measure of purity, but narrow selection of the most highly purified fractions can greatly reduce yield. Other losses come from time-dependent, irreversible binding to chromatography media, denaturation, protein oxidation, harsh binding or harsh eluting conditions from columns, heavy metal ion exposure and separation of non-covalently bound cofactors, prosthetic groups and stabilizing agents from the protein of interest. Sometimes a protein of interest exists in cells as an active non-covalent complex with one or more other proteins. A good purification scheme takes into account both purification levels and yield (Berg et al. 2002; Ward and Swiatek 2009).

2.6 GENETIC AND MOLECULAR STUDIES

2.6.1 Phylogenetics

Phylogenetics is the study of evolutionary relationships and classification of organisms based on the evolutionary theory which states that groups of similar organisms are descended from a common ancestor. These studies are undertaken to understand evolution or conservation of a species based on their characteristics, spread of disease and evolution of traits (Alcaraz et al. 2010; Klappenbach 2012; Scott and Gras 2012).
The evolutionary history inferred from phylogenetic analysis is usually depicted as branching, tree like diagrams that represent an estimated pedigree of the inherited relationships among molecules, organisms, or both (Klappenbach 2012). Different classes of algorithm as Distance-Based Methods, Character-Based Methods, Bayesian inference are used to infer phylogeny from sequence, nucleotide or protein sequences. Distance-Based Methods measure genetic distance between two species, two populations, or even between two individuals. It includes unweighted pair group method with arithmetic mean (UPGMA), neighbour joining (NJ), Fitch-Margoliash (FM) and Minimum Evolution (ME) methods (Wieds 2005; Scott and Gras 2012). Character-based methods make use of all known evolutionary information and determine the most likely ancestral relationships through Maximum parsimony (MP) and Maximum Likelihood. Bayesian inference combines the likelihood and the prior probability distribution of evolutionary parameters (Wieds 2005).

A phylogenetic analysis includes 4 basic steps of alignment (i) building the data model and extracting a phylogenetic dataset (ii) determining the substitution model (iii) tree building and (iv) tree evaluation. Automatic alignment programs such as CLUSTAL W (Thompson et al. 1994) PileUp, ALIGN in ProPack, align sequences according to an explicitly phylogenetic criterion and are commonly used for multiple sequence alignment of DNA or protein sequences. T-Coffee (Notredame et al. 2000); MUSCLE (Edgar 2004) ABA (A-Bruijn alignment) (Raphael et al. 2004), Stemloc (Holmes 2005), StatAlign (Novak et al. 2008), MAFFT (Katoh and Toh 2010) are some other programs available for multiple sequence alignment. The guide tree data obtained after multiple
sequence alignment can be imported and formatted in various tree-building programs as Treeview, PHYLIP, TreeAlign and MALIGN.

2.6.2 16S rRNA (16S rDNA) Sequence for Phylogenetic Analysis

The small subunit 16S ribosomal RNA gene has become a recognized standard since the last decade for estimating the phylogenetic diversity in microbial communities with the development of tools and availability of DNA sequencers in terms of cost, methodologies, improved technologies and instrumentation (Miranda et al. 2008; Maughan and Van der Auwera 2011). The 16S rRNA gene is the most conserved gene sequence in all cells and has been used extensively to determine taxonomy, phylogeny, to estimate rates of species divergence. Recently it has become important as a means to identify an unknown bacterium to the genus or species level (Sacchi et al. 2002). The presence of remarkably similar sequences facilitates the design of PCR primers targeting all members of a community and more variable regions that allow for the discrimination of different microbial taxa (Rasko et al. 2005; Janda and Abbott 2007).

The classification of Bacillus sp demonstrates a typical example of phylogenetic analysis of the species (Alcaraz et al. 2010) (Fig 2.4). The family Bacillaceae in the Phylum Firmicutes, includes the large genus Bacillus that incorporates a variety of phenotypically heterogeneous, gram-positive rod-shaped bacteria with physiological and metabolic diversity and DNA base composition, exhibiting a wide range of nutritional requirements includes both free living and pathogenic species with a world-wide distribution (Yakoubou et al. 2010; Maughan and Van der Auwera 2011).
Figure 2.5: Phylogenetic Tree for *Bacillus* sp. by alignment of 16S rRNA sequences (Alcaraz et al. 2010)

The family *Bacillaceae* has been classified and reclassified innumerable, earlier on the basis of phenetic characters, which were later combined 16S rDNA gene sequence alignments. In the latest classification as explained in 2nd edition of Bergey’s Manual of Systematic Bacteriology (Ludwig et al. 2009), a new taxonomy of the class *Bacilli* is presented. It is the result of phylogenetic and principal-component analyses of comprehensive datasets of 16S rDNA sequences (Ludwig et al. 2009). The *Bacillus* genus belongs to the Order *Bacillales* containing nine families: *Alicyclobacillaceae, Bacillaceae, Paenibacillaceae, Pasteuriiaceae, Planococcaceae, Sporolactobacillaceae, Listeriaceae, Staphylococcaceae* and *Thermoactinomycetaceae*. All nine families contain a total of 51 genera (Yakoubou et al. 2010)
2.7 PROTEOMICS AND ENZYME PROFILING

Proteomics is the study of structure and function of proteins supported by the vast technological advances in computing and data processing (Webster and Oxley 2012). Proteins are made up of amino acids linked like a chain in different ways to form thousands of proteins, each with a unique, genetically defined sequence that determines the protein’s specific, shape and function (Wilson and Walker 2010). In addition, each protein can undergo a variety of post translational modifications (PTMs) that further influence its shape and function (Pelton and McLea 2000; Steen and Mann 2004).

2.7.1 Protein Identification and Characterization

In traditional protein chemistry, proteins were identified by de novo sequencing using automated Edman degradation which is based on successive removal of N-terminal amino acids by chemical methods (Chen et al. 2001; Deutzmann 2004). However since the last decade, this technique has been ousted by mass spectrometry, which has emerged as the most powerful analytical tool for protein and peptide identification in protein chemistry due to increased sensitivity (femto mole level) and 10 fold increase in speed (Steen and Mann 2004; Thiede et al. 2005; Bechara et al. 2012).

Peptide identification using mass spectrometry is based on a simple principle where a peptide is ionized and the peptide bonds are fragmented in an MS–MS spectrometer (Ma 2010; Webster and Oxley 2012). Each resulting fragment ion will form a peak in the spectrum at the corresponding mass to charge (m/z) ratio of the ions which are obtained as peptide masses spectra or the peptide mass fingerprint (PMF) that contain sequence information characteristic of its generating peptide (Xu and Ma 2006). A good quality
spectrum might consist of a ladder of peaks of y-ions (fragment ions containing the C terminus) and a ladder of peaks of b-ions (fragment ions containing the N terminus) (Chen et al. 2001). Consequently, the peptide sequence can be derived by the mass differences of adjacent peaks in each of the two ladders (Thiede et al. 2005; Xu and Ma 2006). However, in practice, many factors as contamination of sample, imperfect fragmentation, simultaneous fragmentation of two different peptides, PTMs and low signal-to-noise ratio make MS–MS peptide identification significantly harder (Ma et al. 2003; Thiede et al. 2005). To minimize these errors in particular, it is critical to assess and take proper account of the strength of evidence for each putative peptide identification (Xu and Ma 2006; Nielsen et al. 2006).

MALDI and ESI (electrospray ionization) are the two different ionization methods commonly used in combination with a variety of mass analyzers as TOF, quadrupole, ion trap to generate peptide/protein ions for MS analysis (Chen et al. 2001; Deutzmann 2004; Nielsen et al. 2006; Webster and Oxley 2012).

2.7.2 Protein Bioinformatics

PMF obtained from MS studies provide the key to identification of new proteins. PMF are compared with the theoretical peptide masses of proteins stored in databases by means of mass search programs generating a score for each comparison or using fragmentation data (raw MS/MS spectra) or sequence tags (Thiede et al. 2005; Ning et al. 2010; Webster and Oxley 2012). Numerous software programs have been developed for MS–MS peptide identification that can be categorized into four classes: database searching, de novo peptide sequencing, peptide sequence tagging and consensus of
multiple search engines (Chen et al. 2001; Xu and Ma 2006; Resing et al. 2004). A database search finds the best matching peptide from a protein sequence database. This approach requires that the protein is known and listed in the database (Deutzmann 2004; Webster and Oxley 2012). De novo sequencing computes a peptide directly from the spectrum; sequence tagging combines the two approaches by first conducting de novo sequencing to obtain a partial sequence tags, and then searches the sequence database using the sequence tags; consensus combines several different programs to increase the confidence and coverage (Xu and Ma 2006).

Mascot (Perkins 1999), MS-Fit and ProteinProspector (Clauser et al 1999), and Profound (Zhang and Chait 2000), Peptident, MultiIdent, Aldente from Expasy (Gasteiger 2005), SEQUEST (Eng et al. 1994), Tandem (Craig and Beavis 2003), OMSSA (Geer et al 2004) are some of the most frequently used internet-accessible search programs for PMF. Other similar programs include SCOPE (Bafna and Edwards 2001), ProbID (Zhang et al. 2002), OLAV (Colinge et al. 2003), PepHMM (Wan et al. 2005) Probidtree (Zhang et al. 2005). Recent de novo sequencing programs include PEAKS (Ma et al. 2003) PepNovo (Frank and Pevzner 2005), NovoHMM (Fischer et al. 2005). PEAKS was initially a de novo sequencing program, however in its latest versions, database searching was included (Zhang et al. 2012). Sequence tagging was first proposed by Mann and Wilm (1994) for the error-tolerant peptide identification. MS-Shotgun (Domon and Aebersold 2010), MS-BLAST (Shevchenko et al. 2001), FASTS (Mackey et al. 2002), DBDigger (Tabb et al. 2005), GutenTag (Tabb et al. 2003), SPIDER (Han et al. 2005) and DeNovoID (Halligan et al. 2005) are some software program employing the sequence-tagging approach. Protein identification and
sequencing by obtaining consensus of multiple search engines was first used by Resing et al. (2004) by combining the search results of MASCOT and SEQUEST. Scaffold (Searle et al. 2005) is one of the first programs dedicated to combining the results of multiple search engines. Once the peptides are correctly identified, the later step of grouping the peptides and identifying the proteins becomes much simpler.

2.7.3 Protein Secondary Structure

The ability to predict protein function from structure is becoming increasingly important, hence elucidation and determination of protein structure becomes a major step in proteomics (Deu et al. 2011). Several methods for determination of the secondary structure of proteins by spectroscopic measurements have been reported (Pelton and McLea 2003; Webster and Oxley 2012). Circular dichroism (CD), Infrared (IR) and Raman spectroscopy are the most commonly used techniques employed to analyze the secondary structure of an enzyme. CD spectroscopy provides rapid determinations of protein secondary structure with dilute solutions and a way to rapidly assess conformational changes resulting from addition of ligands. Both CD and Raman spectroscopies are particularly useful for measurements over a range of temperatures (Pelton and McLea 2003).

2.8 SCOPE AND SIGNIFICANCE OF THE STUDY

Proteases are a distinct class of enzymes, ubiquitous in origin with vast physiological and commercial importance. Microbial community is the most preferred source of these enzymes owing to their inimitable properties as rapid growth, minimum space required
for cultivation, and ready accessibility to genetic manipulation. Of various proteases, metalloproteases are the most diverse catalytic proteases with wide physiological and biochemical characteristics that make the base of their implication in various diseases and pathophysiological conditions as arthritis, cancer, cardiovascular diseases, nephritis, disorders in the central nervous system, fibrosis, and infection and also wide ranged applications in various industries. The search for new metalloprotease with novel characteristics thus becomes imperative and remains in the core of research in enzymology. Exploration of the biodiversity in search of microorganisms that produce metalloproteases with unique and inimitable properties hence becomes paramount in response to their wide applications in various commercial applications. The biochemical and molecular characterization of the enzyme also play a very important role with regard to the application of the enzyme in various industries, drug discovery, diagnostics and molecular medicine. Biochemical studies of enzymes characterize their functions and the 3D structures have provided the molecular basis for our understanding of how these multi-domain proteinases function and interact with ECM molecules and inhibitors. Structural and functional studies also provide us with clues as to how to manipulate their enzymatic activities. Advances in research and methodologies in microbiology and biotechnology have created a favourable niche for further development in characterization and application of metalloproteases in various industrial and medicinal fields, provide a sustainable and balanced environment for mankind and to improve the quality of human life.