Proteolytic inventory of *Thermococcus kodakaraensis*

Nouman Rasool¹, Naeem Rashid²*, Qamar Bashir² and Masood Ahmed Siddiqui³

¹Punjab Forensic Science Academy, Thokar Niaz Beig, Lahore, Pakistan.
²School of Biological Sciences, University of the Punjab, Quaid-e-Azam Campus, Lahore 54590, Pakistan.
³Department of Chemistry, University of Balochistan, Saryab Road, Quetta, Pakistan.

Accepted 21 May, 2013

Proteolysis is a very crucial process for development and survival of the cell. Genome sequence data can be employed to estimate proteolysis inventories of different organisms. In this review, we exploit genome sequence data of hyperthermophilic archaeon *Thermococcus kodakaraensis* to have an overview of the proteolysis in this microorganism. The overview is based on those peptidases that have been characterized, and on putative peptidases that have been identified but have yet to be characterized. In contrast to bacteria, the number of proteolytic enzymes in archaea is quite low. By analyzing the genome sequence data, we tried to establish how *T. kodakaraensis* maintains its life cycle and other processes by using such a small number of protein scavengers while harboring at high temperature where chances of protein denaturation are high.

Key words: *Thermococcus kodakaraensis*, proteolysis, protein denaturation, genome sequence.

INTRODUCTION

Peptidases are known to function for a variety of processes both inside and outside the cell. These not only hydrolyze the external sources for nutrition but also degrade the unwanted and abnormal proteins, especially produced due to environmental and chemical stress (Tomoyasu et al., 2001). The proteins from the microorganisms which thrive at extreme conditions have more threat to be denatured; hence, the elimination of such malfunctioned proteins is crucial for cell survival. Since cell has many useful protein workers and their loss might collapse the whole machinery of the cell, therefore, a strict check must be maintained to prevent lysis of useful proteins. *Thermococcus kodakaraensis*, a member of order Thermococcales, is an anaerobic obligate heterotrophic hyperthermophilic archaeon, isolated from a solfatara on the shore of Kodakara Island in Japan (Morikowa et al., 1994; Atomi et al., 2004). The strain displays heterotrophic growth on a variety of organic substrates. Like other sulfur dependent hyperthermophiles, *T. kodakaraensis* can feed on proteinaceous substrates to achieve carbon and essential amino acids (Hoaki et al., 1993). In order to do so, it has to produce several extracellular enzymes to hydrolyze the polypeptides outside the cell. Complete genome of the strain has been sequenced and many of the enzymes have been characterized (Akiba et al., 2005; Rashid et al., 2004, 2002, 1996, 2001; Sato et al., 2004; Watanabe et al., 2007; Imanaka et al., 2002; Fujiwara et al., 2001). The genome size of *T. kodakaraensis* is 2,088,737 nucleotides, having 92.1% coding region (Fukui et al., 2005). Here, we exploit the genome sequence data to investigate the proteolytic inventory of *T. kodakaraensis*.

INTRACELLULAR PROTEOLYSIS

Energy dependent proteolysis

Energy dependent proteolysis in cell plays important role in rapid turnover of short lived regulatory proteins, and
eliminating the damaged proteins as quality control mechanism. Presence of ATP-dependent peptidases has been shown in all the prokaryotes, including hyper-thermophiles, whose genomes have been sequenced. A number of bacterial peptidases have been identified for instance Lon, Clp AP, ClpXP, HslUV (ClpYQ) and FtsH in *Escherichia coli* (Gottesman, 1999; Gottesman and Maurizi, 1992). Genomic sequence of *T. kodakaraensis* indicated the presence of two types of ATP-dependent intracellular peptidases that is, Lon protease and proteasome (Table 1). The genes corresponding to FtsH and the Clp family proteases are missing in *T. kodakaraensis* genome.

**Lon protease**

Energy dependent peptidase, Lon protease from *E. coli* (Lo*nt*Ec) consists of N-terminal ATPase and C-terminal peptidase domain (Gottesman et al., 1997; Goldberg et al., 1994) and forms a homotetramer or homoocatamer of identical subunits (Chung and Goldberg, 1981; Goldberg et al., 1994). The proteolysis of the abnormal proteins is done in coordination with heat shock system DnaK/DnaJ/GrpE (Jubete et al., 1996; Kihara et al., 1998). Nucleotide plays important role in LonEc and there is very little or no activity in the absence of nucleotide. Interestingly, ATP binding gives conformational changes to LonEc that allows the cleavage of peptide bonds. Lon protease from *T. kodakaraensis* (Lo*ntTk*) also consists of an N-terminal ATPase and C-terminal protease domain (Fukui et al., 2002). The ATPase domain belongs to AAA+ superfamily, possessing several conserved motifs like nucleotide binding walkers A and B (Neuwald et al., 1999). There is another motif named sensor 2 motif present in Lo*ntTk* which is thought to be involved in recognizing the specific substrate, as reported in Lo*ntEc* (Smith et al., 1999).

There are two possible transmembrane regions within the ATPase domain that anchor it in membrane of *T. kodakaraensis*, a characteristic feature of archaeal Lon proteases. Unlike Lo*ntEc*, Lo*ntTk* has the open conformation without the nucleotide binding. It possesses thermostable ATPase activity and this activity is much higher than peptide cleavage activity of this enzyme (Fukui et al., 2002). So, Lo*ntTk* performs two types of functions; in the absence of ATP it recognizes and degrades the unfolded proteins into small polypeptides while the folded proteins are degraded by the expenditure of ATP (Fukui et al., 2002). It is reported that ATPase domain also works as molecular chaperons to unfold the protein structures so that peptide bond could be available to catalytic residue sitting in catalytic pocket (Fukui et al., 2002; Gottesman et al., 1997; Hoskins et al., 1998, 2000; Singh et al., 2000; van Melder et al., 1996; Wickner et al., 1999). It is believed that conformational changes in Lo*ntTk* might be induced by the nucleotide binding for degradation of folded proteins, when intracellular ATP concentration is sufficient. In case of *E. coli*, ATP hydrolysis is needed for the translocation and degradation of large number of unfolded proteins. Therefore, ATP independent proteolytic activity against the unfolded proteins is quite unique characteristic of *T. kodakaraensis* which might be needed to overcome the burden caused by the denaturation of proteins due to high temperature (Fukui et al., 2002).

It is also considered that membrane bound Lo*ntTk* protease might have replaced FtsH, another important ATP dependent membrane bound peptidase which is absent in archaea (Ward et al., 2002). We have noticed that N- and C-terminal of Lo*ntTk* are quite similar with the other members of the order Thermococcales. However, the central part is missing in Lo*ntTk*. To our knowledge, Lo*ntTk* is the only Lon protease characterized from the order Thermococcales and it has quite distinct characteristics. It would be interesting to examine the characteristics of Lon proteases from other members of the order Thermococcales in order to elucidate the effect of the additional central region.

**Proteasome**

The other type of energy dependent proteolysis involves proteasome. To avoid unwanted loss of proteins, the cells have developed a mechanism to confine the protein degradation process to special compartments. Some cells have special membrane enclosed organelles with proteolytic activity and the proteins required to be degraded and eliminated are dragged to these structures for proteolysis (Baumeister et al., 1998). While others do not have membrane bound vessels and have developed the self- or auto-compartmentalizing molecule (Lupas et al., 1997). The proteins form complex structure, enclosing an inner cavity having proteolytic activity, which is only accessible to misfolded proteins. ATPase complex facilitates the translocation of such misfolded proteins to

### Table 1. ATP dependent intracellular proteases found in the genome sequence of *T. kodakaraensis*.

<table>
<thead>
<tr>
<th>Protease</th>
<th>ORF</th>
<th>Type</th>
<th>AA</th>
<th>Size (Da)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lon</td>
<td>TK1264</td>
<td>Serine</td>
<td>635</td>
<td>70257.57</td>
</tr>
<tr>
<td>Proteasome subunit alpha</td>
<td>TK1637</td>
<td>Threonine</td>
<td>288</td>
<td>32098.03</td>
</tr>
<tr>
<td>Proteasome, beta subunit</td>
<td>TK1429</td>
<td>Threonine</td>
<td>231</td>
<td>24813.84</td>
</tr>
<tr>
<td>Proteasome, beta subunit2</td>
<td>TK2207</td>
<td>Threonine</td>
<td>227</td>
<td>24864.96</td>
</tr>
<tr>
<td>Proteasome regulatory subunit</td>
<td>TK2252</td>
<td>Threonine</td>
<td>397</td>
<td>44844.77</td>
</tr>
</tbody>
</table>
inner cavity. These have resemblance with chaperones, so sometimes named as reverse chaperones or unfold-dedases (Lupas et al., 1993).

Proteasome is an example of self-compartmentalizing energy dependent proteolytic complex. It was first reported by Dahlmann and co-workers (Dahlmann et al., 1985) as "multicatalytic proteinase" and then named "pro- teasome" (Arrigo et al., 1988). It is ubiquitously present in eukaryotes and archaea whereas, there is no report for its existence in bacteria except for the gram-positive actinomycetes (Ward et al., 2002). Eukaryotes have 26S proteasome complex while in archaea it is 20S complex. The building units of proteasome complex belong to family Ntn hydrolases (Brannigan et al., 1995). In hypthermophilic archaea, it consists of one α and two β subunits while in eukaryotes, there are seven α and β subunits (Lupas et al., 1997). The α subunits of proteasome make outer ring; whereas β subunits form inner ring enclosing a channel that runs along the length of proteasome (Lupas et al., 1993). The unfolded proteins enter in this channel and are degraded. The α ring restricts the translocation of folded protein and only allows the unfolded protein to enter the channel. 20S proteasome encoding genes have been found in all members of Crenarchaeota and Eukaryarchaeota, whose full genomes have been sequenced.

The genome sequence of *T. kodakaraensis* has revealed the presence of an open reading frame having homology with α subunit and two with β subunits of proteasome indicating the presence of a proteasome complex consisting of one α and two different types of β subunits similar to other hyperthermophilic archaea (Ward et al., 2002).

### Energy independent proteolysis

#### Zinc peptidases

Most of the enzymes acting as indigenous workers in pro- teolytic machinery of *T. kodakaraensis* are Zn\(^{2+}\) con- taining metalloproteases as depicted from the homology comparison. Zinc containing neutral peptidases, stabi- lized by Ca\(^{2+}\), are most defined and well studied metallo- proteases. A few examples are neutral peptidases from *Bacillus cereus* (Pauptit et al., 1988), elastase from *Pseudomonas aeruginosa* (Thayer et al., 1991) and thermolysin from *Bacillus thermoproteolyticus* (Colman et al., 1972; Holmes and Matthews, 1982). These are ubi- quitous and found in both gram positive and negative bacteria. The primary sequence motif HEXXH was known to be found conserved in zinc containing peptidase. These are also associated with the virulence of patho- genic species for example, lethal factor (LF) of *Anthrax* contains HEXXH motif and is considered as a member of zinc containing peptidases (Klimpel et al., 1994). Similarly *Bacillus thurengensis*, a pathogen for insects, also secretes immune inhibitor A (zinc containing peptidase), that degrades the antibacterial proteinase produced by host (Dalhammar and Steiner, 1984).

Several *Bacillus* species produce a number of zinc containing peptidases (Abakov et al., 1990; Takekawa et al., 1991; Tran et al., 1991; Yang et al., 1994; Kuhn and Fortnagel, 1993; Stoeva et al., 1990). Like these micro- bes, *T. kodakaraensis* also has several Zn\(^{2+}\) dependent metalloproteases. These are intracellular workers. Complete list of these peptidases has been summarized in Table 2.

#### Deblocking aminopeptidases

These are the enzymes that release amino acids from proteins and peptides modified with various N-terminal acyl type blocking groups in addition to their amino- peptidase activity. Deblocking aminopeptidases (DAPs) are different from D-aminopeptidases that have different substrate specificities. DAPs are widely distributed among all domains of life. There are three ORFs in *T. kodakaraensis* (DAP\(_n\)) displaying homology to DAPs. They have very close similarity with those of the genus *Pyrococcus*. The DAP from *Pyrococcus horikoshii* has been cloned and characterized (Mori and Ishikawa, 2005), it is reported that DAPs contain zinc ions and on removal become inactivated. Some retain activity on acquiring Co\(^{2+}\) ions. In archaea and eukaryotes proteins are hydrolyzed into small oligopeptides by the activity of proteasome (Akopian et al., 1997). It is believed that these DAPs help the cells to get rid of those small polypeptides.

### Methionine aminopeptidases

Proteins are translated starting from methionine (in eu- karyotes and archaea) and formyl-methionine (in bac- teria). Cells have machinery to remove this N-termini amino acid. This process has been accomplished in non- processive manner. Methionine aminopeptidases in the cell regulate many cellular processes like functional regulation, intracellular targeting and protein turn over (Prchal et al., 1986). Methionine aminopeptidases are classified into two groups, types I and II, depending upon their structure. Bacteria contain type I, archaea have type II while eukaryotes contain both type I and II (Bradshaw et al., 1998). All forms are appeared to be member of metalloproteases and activated by Mn\(^{2+}\), Co\(^{2+}\), Fe\(^{2+}\) and Zn\(^{2+}\) (Lee et al., 2006). Methionine aminopeptidases from *E. coli*, *Staphylococcus aureus*, *Pyrococcus furiosus* and *Homo sapiens* have been crystallographically characterized (Roderick and Matthews, 1993; Oefner et al., 2003; Tahiriof et al., 1998; Liu et al., 1998). They all show homologous catalytic domains that contain two metal ions coordinated by strictly conserved residues, two glutamate, two aspartate and one histidine and two water molecules (Lowther and Matthews, 2000).

*T. kodakaraensis* has one putative ORF for methionine...
aminopeptidase with a highest homology of 85% with Thermococcus sp NA1. The homology comparison shows that both have two aspartic acid residues (Asp 83, Asp 94), two glutamic acid (Glu 188 Glu 281) and a histidine (His 154) as active site residues involved in metal ion coordination (Lee et al., 2006).

**D-aminopeptidases**

D-aminopeptidases are known to form energy independent complex in cells and degrade D-amino acids containing polypeptides. These are member of zinc dependent metalloproteases. D-aminopeptidases play an important role in synthesis and remodeling of peptidoglycan in bacterial cell wall (Goffin and Ghuysen, 1998) and also give resistance against β-lactum based antibiotics. In T. kodakaraensis, there are three ORFs encoding D-aminopeptidases. T. kodakaraensis lacks D-amino acids in the cell wall. Therefore, it is hard to narrate the function of this peptidase in this organism. The predicted use is to give resistance to the strain against β-lactum based antibiotics and to degrade the protein from cell walls of other microbes and used them as carbon/energy sources.

### Carboxypeptidases

Carboxypeptidases play crucial role in the cell for protein degradation/turnover and processing of precursor proteins (Steiner, 1998). They also help aminopeptidases which play an important role in regulation of several cellular functions. Carboxypeptidases cleave single amino acids from C-terminal and nature of C-terminal amino acid decides the fate of the protein. Occurrence of non-polar amino acids on C-terminal facilitates the carboxypeptidases to degrade those proteins (Keiler et al., 1995; Parsell et al., 1990). Based upon nature of the amino acids in active sites, carboxypeptidases are classified into three major groups, cysteine-carboxypeptidases, metallo-carboxypeptidases and serine-carboxypeptidases. There are two ORFs encoding carboxypeptidases in T. kodakaraensis. One of them has close homology of 91% with carboxypeptidase I from Thermococcus sp NA1 and belongs to family metallo-carboxypeptidase.

The other ORF exhibited highest homology of 73% with that of *P. horikoshii*. Carboxypeptidase Taq from *Thermus aquaticus* (Lee et al., 1994), *Thermus thermophilus* (Nagata et al., 2004), *Thermococcus* sp NA1 (Lee et al., 2006)

<table>
<thead>
<tr>
<th>Protease</th>
<th>ORF</th>
<th>Type</th>
<th>AA</th>
<th>Size (Da)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zinc-dependent protease</td>
<td>TK0512</td>
<td>Metallo-protease</td>
<td>234</td>
<td>27498.80</td>
</tr>
<tr>
<td>TldD/PmbA family</td>
<td>TK0698</td>
<td>Metallo-protease</td>
<td>440</td>
<td>49495.99</td>
</tr>
<tr>
<td>TldD/PmbA family</td>
<td>TK0699</td>
<td>Metallo-protease</td>
<td>473</td>
<td>53589.04</td>
</tr>
<tr>
<td>TldD/PmbA family</td>
<td>TK0499</td>
<td>Metallo-protease</td>
<td>441</td>
<td>48553.25</td>
</tr>
<tr>
<td>TldD/PmbA family</td>
<td>TK0502</td>
<td>Metallo-protease</td>
<td>445</td>
<td>50902.23</td>
</tr>
<tr>
<td>TldD/PmbA family</td>
<td>TK2169</td>
<td>Metallo-protease</td>
<td>455</td>
<td>50148.83</td>
</tr>
<tr>
<td>TldD/PmbA family</td>
<td>TK2170</td>
<td>Metallo-protease</td>
<td>429</td>
<td>47502.78</td>
</tr>
<tr>
<td>DUF45 family</td>
<td>TK0002</td>
<td>Metallo-protease</td>
<td>215</td>
<td>25255.37</td>
</tr>
<tr>
<td>Metalloprotease</td>
<td>TK1033</td>
<td>Metallo-protease</td>
<td>132</td>
<td>15162.72</td>
</tr>
<tr>
<td>Intracellular protease I</td>
<td>TK1284</td>
<td>Serine</td>
<td>166</td>
<td>18415.12</td>
</tr>
<tr>
<td>Hycl</td>
<td>TK2004</td>
<td>Metallo-protease</td>
<td>186</td>
<td>19925.47</td>
</tr>
<tr>
<td>D-Aminopeptidase</td>
<td>TK0160</td>
<td>Metallo-protease</td>
<td>244</td>
<td>36822.2</td>
</tr>
<tr>
<td>D-Aminopeptidase</td>
<td>TK1022</td>
<td>Metallo-protease</td>
<td>276</td>
<td>30032.94</td>
</tr>
<tr>
<td>Deblocking aminopeptidase</td>
<td>TK0775</td>
<td>Metallo-protease</td>
<td>337</td>
<td>37201.76</td>
</tr>
<tr>
<td>Deblocking aminopeptidase</td>
<td>TK0781</td>
<td>Metallo-protease</td>
<td>348</td>
<td>38297.16</td>
</tr>
<tr>
<td>Deblocking aminopeptidase</td>
<td>TK1177</td>
<td>Metallo-protease</td>
<td>346</td>
<td>38197.95</td>
</tr>
<tr>
<td>Xaa-Pro aminopeptidase</td>
<td>TK0967</td>
<td>Metallo-protease</td>
<td>348</td>
<td>39219.87</td>
</tr>
<tr>
<td>Xaa-Pro aminopeptidase</td>
<td>TK1455</td>
<td>Metallo-protease</td>
<td>356</td>
<td>40022.91</td>
</tr>
<tr>
<td>Methionin aminopeptidase</td>
<td>TK1183</td>
<td>Metallo-protease</td>
<td>295</td>
<td>33021.26</td>
</tr>
<tr>
<td>M28 aminopeptidase</td>
<td>TK1912</td>
<td>Metallo-protease</td>
<td>561</td>
<td>62851.61</td>
</tr>
<tr>
<td>Prolyl endopeptidase</td>
<td>TK0423</td>
<td>Serine</td>
<td>616</td>
<td>70098.91</td>
</tr>
<tr>
<td>O-Sialoglycoprotein</td>
<td>TK2126</td>
<td>Metallo-protease</td>
<td>325</td>
<td>35366.06</td>
</tr>
<tr>
<td>Acylaminoacyl peptidase</td>
<td>TK0725</td>
<td>Serine</td>
<td>632</td>
<td>72086.07</td>
</tr>
<tr>
<td>Acylaminoacyl peptidase</td>
<td>TK2049</td>
<td>Serine</td>
<td>622</td>
<td>72151.98</td>
</tr>
<tr>
<td>Carboxypeptidase/aminoclylase</td>
<td>TK0494</td>
<td>Serine</td>
<td>384</td>
<td>42430.57</td>
</tr>
<tr>
<td>Pyrroldione-Carboxylate</td>
<td>TK1835</td>
<td>Metallo-protease</td>
<td>206</td>
<td>22397.89</td>
</tr>
<tr>
<td>Family M32 protease</td>
<td>TK1840</td>
<td>Metallo-protease</td>
<td>499</td>
<td>58994.57</td>
</tr>
<tr>
<td>Microsomal dipeptidase</td>
<td>TK2040</td>
<td>Serine</td>
<td>310</td>
<td>35342.07</td>
</tr>
</tbody>
</table>
and P. furiosus (Chang et al., 2001) have been characterized. Carboxypeptidase from Thermococcus sp NA1 shows broad substrate specificity which makes it a potential candidate in C-terminal sequencing.

**Intracellular peptidases**

Intracellular peptidases are ubiquitous proteases present in all the three domains of life. The intracellular serine protease in T. kodakaraensis genome is highly homologous (83% identical) to protease I (Pfpl), an internal peptidase of P. furiosus (Halio et al., 1997). Pfpl, in vitro, is found in three functional forms, a trimer, a hexamer and a dodecamer (Chang et al., 2001). It appeared that it may exist as an ATP independent protease complex and behaves like proteasome. Another intracellular peptidase from P. horikoshii (Php l) has been characterized. Its three dimensional structure revealed that it exists in dodecameric form having two identical six membered rings, each with axes of symmetry such that it consists of a dimer of trimers or trimer of dimers (Ward et al., 2002).

**Prolyl oligopeptidases**

Most of the peptidases cannot hydrolyze peptide bond at proline residue. Thereby, proline residue protects many proteins from proteolytic degradation (Wilk, 1983; Cunningham and O’Connor, 1997). Prolyl oligopeptidases cleave the polypeptide at the carboxylic end of proline residue. They were discovered in human uterus and were known as oxytocin-degrading enzymes (Walter, 1971; Koida and Walter, 1976). Prolyl oligopeptidases have been characterized from various sources including E. coli, Flavobacterium meningosepticum, Sarcophagi peregrine, Aeromonas hydrophila, P. furiosus, human lymphocytes and mouse brain (Szeltn et al., 2000; Yoshimoto et al., 1991; Diefenthal et al., 1993; Ohtsuki et al., 1997; Robinson et al., 1995; Harwood and Schreier, 2001; Vanhoff et al., 1994; Shirasawa et al., 1994). There is also found an ORF encoding prolyl oligopeptidase in the genome of T. kodakaraensis.

**Xaa-Pro aminopeptidases**

Xaa-Pro aminopeptidases are exopeptidases which hydrolyze the peptide bond between N-terminal amino acid and a penultimate proline. For enzyme activity, a free amino group must be present at the N-terminal amino acid, penultimate residue must be proline and scissile bond be in trans configuration (Lin and Brandts, 1979; Yoshimoto et al., 1994). Xaa-Pro aminopeptidases are widely distributed among mammalian and microbial sources (Yaron and Mlynar, 1968; Mars and Monnet, 1995; Orawski et al., 1987; Holtzman et al., 1987; Hooper and Turner, 1988; Harbeck and Mentlein, 1991).

There are two ORF, TK0967 and TK1455, annotated as putative Xaa-pro aminopeptidases in T. kodakaraensis. Amino acid residues important for catalytic activity (three histidine residues) and the metal binding (two aspartic acid residues, a histidine residue, and two glutamic acid residues), conserved in bacterial, nematode, insect and mammalian aminopeptidase P were also conserved in these ORF. They are expected to be responsible for the specific cleavage of N-terminal Xaa-pro peptide bond in both short and long peptides.

**Acylaminoacyl peptidases**

In post translational modifications, N-termini of many polypeptides are modified by the attachment of different acyl groups such as acetyl, chloroacetyl, formyl and carbamyl (Jones et al., 1986). Peptidases find difficulty in cleaving at these modified amino acids. Acylaminoacyl peptidases are enzymes known for hydrolysis of such N-terminally modified amino acids (Jones et al., 1994). Acylaminoacyl peptidases have been purified from different eukaryotic sources like bovine liver, rabbit muscle and ovine liver (Farries et al., 1991; Gade and Brown, 1978) and found that these are composed of four identical subunits (Polgár, 2002). Among hyperthermophilic archaia, the enzyme has been cloned and characterized from P. horikoshii (Ishikawa et al., 1998). In contrast to eukaryotic sources, it exists in a dimeric form.

There are two putative ORFs encoding acylaminoacyl peptidase in T. kodakaraensis. They are predicted to play important role in removal of N-acylated amino acid from blocked peptides similar to P. horikoshii.

**Hydrogenase maturation peptidases**

Hydrogenases are composed of two subunits; a 30 to 35 and 60 to 65 kDa. Larger subunit contains a metal binding site and once metal binds, the subunit is cleaved near the C-terminal by hydrogenase maturation peptidases. Hydrogenase maturation protease Hyc1 is a member of this family which cleaves an Arg-Met bond immediately following the metal binding site of the hydrogenase (Theodoratou et al., 2000). Hyc1 from E. coli have been characterized and found to be an aspartate peptidase (Rossman et al., 1995). T. kodakaraensis has one gene encoding hydrogenase maturation peptidase in its genome.

**Membrane associated proteolysis**

Membrane associated proteolysis in T. kodakaraensis is evident by the presence of various homologues. The membrane associated peptidase may reside between lipid bilayer of the membrane, or one part anchored in the membrane and other flanking outside containing proteolytic activity. There are the following two types of membrane associated proteins in T. kodakaraensis (Table 3).

**Intramembrane peptidases**

For a long time, peptidases were considered as water soluble enzymes, either present in aqueous environment or with a hydrophobic domain anchored in membrane. It
has been discovered that there is another class which is totally embedded within hydrophobic environment of the lipid bilayer. This unusual peptidases require unusual type of substrates, which have typically folded α-helices (Wolfe and Kopan, 2004). On the basis of amino acid sequence homology, the presence of a few putative intramembrane peptidases has been evidenced in *T. kodakaraensis*. The amino acid sequences reveal that these belong to S2P (site 2 peptidase) family metalloproteases. S2P are zinc dependent metalloproteases which require two histidine and one aspartate residue for enzyme activity (Rawson et al., 1997).

**Signal peptide peptidases**

All organisms are known to have secretory machinery to export the proteins outside the cells. These are well developed in higher eukaryotes and still at early stages in archaea (Ring and Eichler, 2004). The proteins needed to be secreted have typical signal peptide, which is cleaved at the surface of membrane by special membrane bound peptidases named as signal peptide peptidases. Eukaryotic signal peptide peptidases are intramembranous enzymes and have two aspartate residues at catalytic site (Lemberg et al., 2001; Wolfe and Kopan, 2004).


Similar candidate is present in *Bacillus* for signal sequence removal which is further degraded by cytosolic peptidase *Tep A* into small fragments (Bolhuis et al., 1999). In archaea, a few signal peptides have been studied which include type I signal peptide peptidase and Flak signal peptidase for preflagellum cleavage from *Methanococcus voltae* (Ng and Jarrell, 2003). In crenarchaeota, prepillin (PibD) from *S. solfataricus* has been studied; that is a homologue of bacterial type IV signal peptide peptidase (Albers et al., 2003). There are two ORFs for signal peptide peptidases in *T. kodakaraensis* as described in Table 3. One of them, Tk-1164 have been cloned and characterized as signal peptidase A (*SppA*). It belongs to S49 family of serine protease. It is observed that it cleaves the signal peptide on C-terminal of Ala, Leu, Val, Gly and Phe (Matsumi et al., 2005). Further investigations indicated the function of *Tk* 

Kinetic analyses indicated that Ser 

**Extracellular proteolysis**

**Archaeal serine proteases**

Serine proteases are well studied enzymes and have been isolated from a variety of sources. These are divided into two groups including subtilisin like serine protease and (chymo)trypsin like serine protease. Former is widely distributed than the later and has a conserved catalytic triad comprised of Asp, His and Ser residues. They are classified into six families including subtilisin, thermotais, lantibiotic peptidase, kexin and pyrolysin (Siezen and Leunissen, 1997). The subtilisin family is quite large and many of its members have been characterized for example, subtilisin BNP from *Bacillus amyloliquefaciens* (Wells et al., 1983), subtilisin E from *B. subtilis* (Stahl and Ferrari, 1984) and subtilisin Carlsberg from *B. licheniformis* (Jacobs et al., 1985). Subtilisins are synthesized inside the cell as preproproteins and prepeptide serves as a signal which is recognized by the cellular machinery to export the protein; whereas, N-terminal prepropeptide keeps the enzyme inzymogen form (Jacobs et al., 1985; Stahl and Ferrari, 1984).

There are three subtilisin-like serine protease precursors present in the genome of *T. kodakaraensis* (Table 4) and two of them, Tk-1675 and Tk-1689 have been characterized (Kannan et al., 2001; Rasool et al., 2010). Subtilisin like serine protease from *T. kodakaraensis* exhibit low similarity with other characterized subtilisins isolated from

<table>
<thead>
<tr>
<th>Protease</th>
<th>ORF</th>
<th>Type</th>
<th>AA</th>
<th>Size (Da)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal peptidase I</td>
<td>TK2037</td>
<td>Serine</td>
<td>164</td>
<td>18168.08</td>
</tr>
<tr>
<td>Endopeptidase IV</td>
<td>TK1164</td>
<td>Serine</td>
<td>334</td>
<td>36082.21</td>
</tr>
<tr>
<td>Family M50 protease</td>
<td>TK1247</td>
<td>Metalloprotease</td>
<td>436</td>
<td>47257.38</td>
</tr>
<tr>
<td>Family M50 protease</td>
<td>TK1467</td>
<td>Metalloprotease</td>
<td>214</td>
<td>23672.24</td>
</tr>
<tr>
<td>Family M50 protease</td>
<td>TK1820</td>
<td>Metalloprotease</td>
<td>386</td>
<td>41957.21</td>
</tr>
<tr>
<td>NfeD-like protease</td>
<td>TK0347</td>
<td>Serine</td>
<td>447</td>
<td>46856.15</td>
</tr>
<tr>
<td>Stomatrin-like protease</td>
<td>TK1316</td>
<td>Unknown</td>
<td>317</td>
<td>35887.24</td>
</tr>
<tr>
<td>Stomatrin-like protease</td>
<td>TK0348</td>
<td>Unknown</td>
<td>268</td>
<td>29631.91</td>
</tr>
<tr>
<td>CaaX family protease</td>
<td>TK0439</td>
<td>Metalloprotease</td>
<td>195</td>
<td>21863.21</td>
</tr>
</tbody>
</table>

Table 3. Membrane associated proteases found in the genome sequence of *T. kodakaraensis*.
different sources. These are Ca\(^{2+}\) dependent enzymes and have two Ca\(^{2+}\) binding sites, Ca1 and Ca2, which are far from active site. Ca\(^{2+}\) ions are required for maturation of protease, in which it provides conformational change to the enzyme to auto lyze its own prosequence. Recombinant Tk-1689 has shown cytotoxic activity indicating that the propeptide is cleaved inside the host cells converting the protease to active form and causing the death of the host (Rasool et al., 2010). There is another ORF in the genome of *T. kodakaraensis* that encodes for archaeal serine protease which exhibits a highest homology of 63% with that of *Pyrococcus abyssi* GE5.

**Thiol protease**

There is an extracellular thiol protease in *T. kodakaraensis*. The thermostable thiol-protease is the first proteolytic enzyme of *T. kodakaraensis* that was purified and characterized (Morikawa et al., 1994). It is a cysteine protease, about 44 kDa in size, showing highest activity at 110°C. The enzyme was completely inhibited by thiol protease inhibitors such as pCMB and F64, while partially inhibited by DFP and EDTA. The serine protease inhibitor PMSF has no effect on the enzyme activity (Morikawa et al., 1994).

**Serpins**

Serpins or serine protease inhibitors are proteins which inactivate serine proteases, mostly chymotrypsin and subtilisin like serine protease (Carrell and Travis, 1985; Dahlen et al., 1997). These are present in all three domains of life. Serpins usually bind to respective enzymes like standard substrate binding mechanism. They are relatively small (from 29 to 190 residues) and share an exposed, rigid binding loop with a very characteristic canonical confirmation which inserts itself into an active site cleft of the enzyme (Mottonen et al., 1992). Serpin genes appeared to be absent in several archaeal genera. Only 4 out of 13 archaeal genera, whose complete genome sequence is available, are known to have serpin genes in their genomes (Thomas et al., 2004). *T. kodakaraensis* has a putative serpin gene (TK1782) encoding 426 amino acids. It has highest homology of 41% with proteinase inhibitor 14 of *Methanoculleus marisnigri* JR1.

A putative signal peptide on N-terminal is expected to lead the inhibitor outside the cell.

**DISCUSSION**

Proteolysis not only serves to get rid of unwanted and abnormal proteins, but also to limit the concentrations and time of crucial regulatory proteins. In microbes, it also helps to get nourishment from media by degrading complex polypeptides into smaller fragments, even into amino acids so that these can easily be ingested. This is also a well defined regulatory mechanism of controlling any protein population inside the cell in addition to its regulation at transcription level. No doubt, proteolysis is very advantageous and benign mechanism for cell, but if uncontrolled and non specific, it can cause severe affects even the cell death. Eukaryotic system relatively provides an ease in term of protein degradation by making special compartments. Such compartmentalization is not present in bacteria and archaea (Baumeister et al., 1998); which certainly demands for more strict regulation to prevent any undesirable conditions. Identification of the proteins which are needed to be degraded is very critical and crucial. Intracellular proteolysis is mainly ATP dependent. Among ATP dependent peptidases, Lon protease holds vital position. *E. coli* mutant in Lon protease lost 50% of the ability to degrade abnormal proteins (Coux et al., 1996). Lon protease from *T. kodakaraensis* has an ATPase domain which is necessary for unwinding/ linearization of the misfolded proteins in order to expose the scissile peptide bond to peptidase (Fukui et al., 2002). Another energy dependent peptidase, proteasome, is also associated with the degradation of proteins damaged during any environmental or chemical stress. These are large enzymes formed by the assembly of various subunits. The recognition of proteins to be degraded is quite independent of peptide bond cutting activity of a peptidase and resides in the ATPase domain of peptidase, which has chaperon like activity (Gottesman, 1996; Gottesman et al., 1997; van Melderen et al., 1996; Wickner et al., 1999). Lon protease from *T. kodakaraensis* exhibits catalytic activity independent of ATP towards improperly folded proteins, while ATP is only required to cleave peptide bonds of properly folded proteins (Fukui et al., 2002). This is a unique property of this peptidase. *T. kodakaraensis* does not harbor homologues of Cip family proteases which are present in bacteria and absent in eukaryotes. It is assumed that Cip family proteases play same role as performed by proteasome in other organisms. Like proteasome, these also form a channel and are believed to degrade the proteins malfunctioned.

<table>
<thead>
<tr>
<th>Protease</th>
<th>ORF</th>
<th>Type</th>
<th>AA</th>
<th>Size (Da)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subtilisin-like protease</td>
<td>TK0076</td>
<td>Serine</td>
<td>524</td>
<td>56487.21</td>
</tr>
<tr>
<td>Subtilisin-like protease</td>
<td>TK1675</td>
<td>Serine</td>
<td>422</td>
<td>43655.13</td>
</tr>
<tr>
<td>Subtilisin-like protease</td>
<td>TK1689</td>
<td>Serine</td>
<td>663</td>
<td>70823.97</td>
</tr>
<tr>
<td>Archaeal serine protease</td>
<td>TK2168</td>
<td>Serine</td>
<td>643</td>
<td>70022.12</td>
</tr>
<tr>
<td>Thiol protease</td>
<td>TK1295</td>
<td>Cystein</td>
<td>1103</td>
<td>122371.24</td>
</tr>
</tbody>
</table>

Table 4. Extracellular proteases found in the genome sequence of *T. kodakaraensis*.  

Rasool et al. 3145
during heat and chemical stress (Bochtler et al., 1997). In energy independent proteolysis, Zn dependent metalloproteases and peptidases hold predominant place among other intracellular protein degrading enzymes. Exact roles of these peptidases are still to be elucidated. N- and C-terminal rule of protein degradation is also implemented by these peptidases. The presence of Arg and Lys on N-terminal reduces the half-life of a protein, while presence of last five nonpolar amino acids at C-terminal makes protein resistant to degradation (Keiler et al., 1995; Parsell et al., 1990).

C-terminus amino acid cleavage is accomplished by carboxypeptidase, while many of uncharacterized Zn-dependent aminopeptidases might be potential candidate to implement N-terminal cleavage. It is believed that some energy independent peptidases also form homo multimers enclosing a cavity inside, as observed in proteasomes. How and which type of polypeptides enter the cavity, is still unsatisfied with any answer. The genome sequence of *T. kodakaraensis* revealed the presence of four putative extra cellular peptidases. Three of them are subtilisin like serine proteases and the fourth one is a thiol protease. They help the cells to utilize the proteins in extracellular media as source of essential amino acids (Hoaki et al., 1993). Two subtilisin like serine proteases (Tk-subtilisin and Tk-SP) and a thiol protease have been characterized and they have wide substrate specificity and are highly tolerant to temperature and pH (Kannan et al., 2001; Hirata et al., 2013; Morikawa et al., 1994). Leu→Pro mutation in Tk-subtilisin accelerated the maturation of Pro-Tk-subtilisin by reducing the binding ability of Tk-propeptide to Tk-subtilisin (Uehara et al., 2013). Tk-SP is reported to have a great potential for technological applications such as thermo-stable detergent additives (Hirata et al., 2013).

A comparison of genes encoding putative peptidases found in the genome sequence of various microorganisms is shown in Table 5. Although, genes encoding proteases and peptidases can be listed with their putative roles, but sometime it does not match with the true function. For example, signal peptide peptidase characterized from *T. kodakaraensis* was considered as member of endopeptidase family based on amino acid homology (Matsumi et al., 2005). Similarly, some of the proteins were first identified as peptidase and later on found to have other activities instead of proteolytic activities (Cho and Cronan, 1994). One can deduce the mechanism of proteolysis just over viewing the list of genes present in the genome and speculating a function to the gene products as shown in Figure 1 for *T. kodakaraensis*. However, analysis and characterization of the individual gene product is needed in order to assign a function. Furthermore, questions aroused during the assignment of functions on the basis of gene structure.
Table 5. A comparison of total number of proteases in different microorganisms.

<table>
<thead>
<tr>
<th>Organism</th>
<th>ATP-dependent proteases</th>
<th>ATP-independent proteases</th>
<th>Peptidases</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>T. kodakaraensis</td>
<td>5</td>
<td>15</td>
<td>18</td>
<td>38</td>
</tr>
<tr>
<td>P. furiosus</td>
<td>5</td>
<td>13</td>
<td>22</td>
<td>40</td>
</tr>
<tr>
<td>P. Abyssi</td>
<td>5</td>
<td>9</td>
<td>20</td>
<td>34</td>
</tr>
<tr>
<td>P. Horikoshii</td>
<td>5</td>
<td>9</td>
<td>20</td>
<td>34</td>
</tr>
<tr>
<td>E. coli</td>
<td>13</td>
<td>16</td>
<td>31</td>
<td>60</td>
</tr>
<tr>
<td>P. caldifontis</td>
<td>4</td>
<td>7</td>
<td>21</td>
<td>32</td>
</tr>
<tr>
<td>M. maripaludis</td>
<td>4</td>
<td>3</td>
<td>14</td>
<td>21</td>
</tr>
<tr>
<td>S. solfataricus</td>
<td>4</td>
<td>13</td>
<td>16</td>
<td>33</td>
</tr>
<tr>
<td>A. fulgidus</td>
<td>4</td>
<td>7</td>
<td>11</td>
<td>22</td>
</tr>
<tr>
<td>T. maritima</td>
<td>9</td>
<td>14</td>
<td>16</td>
<td>39</td>
</tr>
</tbody>
</table>

can only be answered when the purified gene product is analyzed for its enzymatic activity.

REFERENCES
Gottesman S, Maurizi MR (1992). Regulation by proteolysis: energy-


