*Pak. J. Bot.*, 43(4): 2059-2065, 2011.

# REQUIREMENT OF PRO-PEPTIDE IN PROPER FOLDING OF SUBTILISIN-LIKE SERINE PROTEASE TK0076

**NOUMAN RASOOL, NAEEM RASHID*, MUHAMMAD ARSHAD JAVED AND MUHAMMAD SALEEM HAIDER**

*School of Biological Sciences, University of the Punjab, Quaid-e-Azam Campus, Lahore 54590, Pakistan*
**Corresponding author E-mail: naeemrashid37@hotmail.com; naeem.ff@sbs.pu.edu.pk*
*Tel: +92-42-99231534; Fax: +92-42-99230980*

**Abstract**

Subtilisin-like proteases are characterized by a catalytic triad of the three amino acids Asp His and Ser. Tk0076 is a subtilisin-like serine protease originated from *Thermococcus kodakaraensis*. Regions corresponding to signal-peptide, pro-peptide and the mature protein were predicted by homology modeling. Homology comparison revealed that $Asp^{215}$, $His^{247}$ and $Ser^{424}$ constitute the catalytic triad of the protein. Gene encoding Tk0076 was cloned and expressed in *Escherichia coli*. The protein was produced in the soluble form when the gene contained the sequence corresponding to the pro-peptide whereas it was produced in the insoluble form without the sequence corresponding to pro-peptide under the same expression system. Attempts to refold the protein properly in the absence of pro-peptide were unsuccessful indicating that pro-peptide is essential for proper folding of Tk0076.

## Introduction

Proteases are indispensable components of any cell and play important roles not only inside but outside the cell. They perform a variety of functions in metabolism and convert zymogens into active forms (Neurath, 1984; Van de ven *et al*., 1993; Huber & Bode, 1978; Davie *et al.*, 1991). These are necessary for growth of the cell. The proteases, secreted to extracellular space, play important role in scavenging the protein based nutrition (Rao *et al*., 1998). Serine proteases are large in number and ubiquitously present in all domains of life. It is believed that serine proteases are one third of all the proteases (Hedstrom, 2002). These are among the first enzymes which have been extensively characterized and studied (Neurath, 1985).

Archaea, the third domain of life, have been extensively studied and the knowledge about their biochemical adaptation to extreme environments has been explored (Halio *et al*., 1997; Huber & Stetter, 2001). Archaea can survive at extreme conditions including high and low temperatures, high saline conditions, high and low pH, and high pressure (López-García *et al*., 2001). Microorganisms that thrive at temperature near boiling water are called hyperthermophiles. The enzymes produced by these microorganisms are usually thermostable.

*Thermococcus kodakaraensis* was isolated from a solfatara on the shore of Kodakara Island, Japan (Morikawa *et al*., 1994; Atomi *et al.*, 2004). It is an obligatory heterotroph, grows on complex organic medium containing starch, maltooligosaccharides and pyruvate, supplemented with elemental sulfur. Cells are motile having polar flagella, irregular cocci with a 1 µm diameter and multiply by constriction. Complete genome of the organism has been determined (Fukui *et al*., 2005). Analysis of the genome sequence revealed that there are total 38 open reading frames (ORFs) which encode 38 different types of proteases and peptidases. Among them 5 are ATP dependent proteases, 15 are ATP independent proteases and 18 are peptidases. Proteolysis in the microorganism is carried out by two types of proteases, energy dependent proteases and energy independent proteases.

Subtilisin-like serine proteases are well studied enzymes and have been isolated from a variety of sources. There are three ORFs (Tk1689, Tk1675 and Tk0076) in the genome of *T. kodakaraensis* encoding three subtilisin-like serine protease precursors. It is speculated that each subtilisin-like serine protease precursor is synthesized inside the cell as preproprotein. Prepeptide serves as signal recognized by the cellular machinery to export the protein and N-terminal propeptide keeps the enzyme in zymogenic form in the same fashion as in other subtilisins (Jacobs *et al*., 1985; Stahl & Ferrari, 1984). In this study Tk0076 was cloned and expressed in *E. coli*. Computational analysis was also performed including phylogenetic analysis, sequence alignment and 3D modeling. The truncated version of this enzyme was also produced and effect of deleted polypeptide on the solubility and folding of the enzyme was studied.

## Materials and Methods

**Phylogenetic analysis:** The alignment of amino acid sequence of the protease with other proteases was done by using Bioedit software. Amino acid composition, molecular mass calculation and theoretical pI of the proteins were carried out using Expasy program available at http://www.expasy.org. Phylogenetic analyses of the proteases were also performed with the help of Clustal X, Genious and Dendroscope softwares (Jeanmougin *et al*., 1998).

**Prediction of signal- and pro-peptide:** Signal peptide was predicted using a web based program SignalP provided at http://www.cbs.dtu.dk/services/SignalP (Bendtsen *et al*., 2004). Prediction of the size of the pro-peptide was done by aligning the protease sequences by using ProDom server (http://prodom.prabi.fr/prodom/current/html/home.php).

**Prediction of secondary structure and 3D modeling:** The 3D model of Tk0076 was constructed on the basis of standard homology modeling protocols. Homologue structures were found from ExPdb (modified PDB database) by submitting the query sequences to Swiss Model Server (http://swissmodel.expasy.org//SWISS-MODEL.html). Among the subtilisin like proteases whose crystal structures have been resolved subtilisin from

*Bacillus amyloliquefaciens* (lyjb(A)) exhibited highest homology with Tk0076, therefore it was used as template. The amino acid sequence of Tk0076 was aligned with the template and the alignment was submitted to the Swiss Model Server. The result model was evaluated in the Swiss Model Viewer V. 3.5 (SP5). Amino acid constraints and clashes were removed. The protein model was examined and evaluated using verify 3D server (Arnold *et al*., 2006; Kopp & Schwede, 2004). Putative $Ca^{2+}$ binding sites of the protease were determined by using web based server (http://bioinf.cs.ucl.ac.uk/MetSite/MetSite.html).

**Chemicals, plasmids and strains:** All chemicals used in this study were of high quality and purchased from Fisher Scientific (Leicestershire, UK), Fluka (Buchs, Switzerland), Merck (Darmstadt, Germany) and Sigma (Taukkirchen, Germany). Restriction enzymes, DNA markers and protein size standards, T4 DNA Ligase, Taq DNA Polymerase, dNTPs, TA cloning vector pTZ57R/T, InsTAclone $^{TM}$PCR Cloning Kit and DNA Extraction Kit were purchased from Fermentas Life Sciences (Maryland, USA). RNase was obtained from Sigma. Expression vectors pET-21a and pET-22b were from Novagen (Merck, Germany). All designed primers were synthesized by e-oligos (New York, USA). The ingredients of cultivation media were purchased from Difco Laboratories, or US Biological (CA, USA), unless stated otherwise. *E. coli* strains DH5α and BL21 CodonPlus (DE3)-RIL (Novagen) were used for the expression of protease gene.

**Primer designing:** Three sets of forward and reverse primers were designed to amplify the full-length gene, the gene without pre-region (signal-peptide) and the gene without both pre- and pro-region. The primers were designed and checked by using a program AmplifX (http://ifrjr.nord.univ-mrs.fr/AmplifX-Home-page?lang=en) and possibilities of self priming loop formation, primer dimer formation were minimized. Quality of Primers was also analyzed on Windows based software Oligoanalyzer for melting temperature (Tm), parenthesis GC contents, 3′ and 5′ end stability. Primers used in this study are given in Table 1.

**Table 1. Nucleotide sequence of primers used to amplify Tk0076 gene.**

| Name | Sequence | Length |
|---|---|---|
| 0076-R | 5′-GGAACGCTGTAAAGAGGAGGAT | 22 |
| 0076-F | 5′-GCACATATGGACAGACAGCAGCTCCAACG AGACCATCAGC | 40 |
| 0076MP-F | 5′-ACATATGAACATCAAGCTCGTTAAGGC | 28 |

**Amplification of gene:** Polymerase Chain Reaction (PCR) was performed using GeneAmp® PCR System 2700 (ABI, USA). PCR mixture was prepared in sterile 200 µL tubes. Each reaction mixture contained 100 ng of *T. kodakaraensis* genomic DNA, 1X PCR buffer (75 mM Tris-HCl pH 8.8 at 25°C, 20 mM $(NH_4)_2SO_4$ and 0.01 Tween 20), 50 mM KCl, 2.5 mM $MgCl_2$, 200 µM deoxynucleoside triphosphates (dATP, dTTP, dGTP, dCTP) 100 pmol of each forward and reverse primers. A negative control contained all the reaction components except the template DNA. The thermocycler was programmed as a single denaturation step of 5 min at 94°C, followed by a step cycle program for 30 cycles of denaturation at 94°C for 30 seconds, annealing at 55°C for 30 seconds and extension at 72°C for 1 min, and a final extension at 72°C for 10 min.

**Cloning and expression of Tk0076:** The gene encoding for Tk0076 was amplified by PCR and cloned in pTZ57R/T. After confirming the cloning and orientation of the gene in pTZ57R/T, the inserted fragment was cut with *Nde*I and *Hin*dIII. Expression vector (pET-21a) was also digested with same pair of restriction enzymes and Tk0076 gene was ligated in the expression vector. The resulting recombinant plasmid was named as pET-0076. *E. coli* DH5α competent cells were transformed using pET-0076 plasmid. Colonies on the selection plate were screened for the presence of pET-0076 plasmid. For expression of the gene *E. coli* BL21 CodonPlus (DE3)-RIL cells were transformed using pET-0076 plasmid. Heterologous expression of the gene was induced with 0.2 mM IPTG when optical density of the culture was 0.45 at 660 nm. Cells were harvested 4 h after induction and lysed by sonication. Production of recombinant Tk0076 was analyzed by sodium dodecylsulphate polyacrylamide gel electrophoresis.

**Protein estimation:** Protein concentration was estimated by dye-binding method of Bradford with the Bio-Rad protein assay reagent (Bio-Rad Laboratories, Hercules, CA, USA) using BSA as a standard (Bradford, 1976). The standard curve was obtained using 2 to 18 µg of BSA in 1 mL of protein assay reagent. The incubations were performed at 37°C for 5 min and the absorbance was measured at 595 nm.

**Refolding of inclusion bodies:** Cells were disrupted by sonication and centrifuged at 10,000 X g at 4°C in Avanti J-251, Beckman Coulter$^{TM}$ centrifuge. The insoluble inclusion bodies were taken out, resuspended in 50 mM Tris-HCl pH 8.0 containing 0.2% Triton X-100 and sonicated again for 5 min. The sonicated sample was kept on ice for 1 h. The insoluble inclusion bodies were separated from soluble proteins by centrifugation and two procedures were used to refold the recombinant protein:

1. Inclusion bodies were dissolved in 8 M urea and heated in boiling water for 5 min. It was placed on ice for 20 min and then centrifuged for 30 min at 10,000 x g. The soluble fraction was taken out and fractional dialysis from 8 M to 6 M to 4 M to 2 M to 0 M urea was performed in 50 mM Tris-HCl (pH 8.0). After dialysis, the sample was centrifuged for 15 min at 10,000 x g to separate soluble folded protein from insoluble protein.
2. In second protocol, refolding of the recombinnat protein was performed using decreasing linear gradient of urea from 8 to 1 M in the presence of 5 mM cysteine and 1 mM cystine. To refold MP-TK0076, renaturation buffer (20 mM Tris-HCl, pH 8.5, 1 M urea, 500 mM NaCl, 5 mM cysteine and 1 mM cystine) containing 250 mM imadazole was used

and dialysed against TNG buffer [20 mM Tris-HCl, pH 8.5, 50 mM NaCl and 10% (v/v) glycerol].

**Protease assay:** For protease assay azocasein was used as substrate. The reaction mixture consisted of 100 µL (2 mg) azocasein solution, 50 to 100 µL of enzyme solution, and 50 mM Tris-HCl pH 8.0 in a final volume of 1 mL. Assay was performed in a water bath set at 80°C for 10 min and stopped by adding 1 mL of 30% Trichloroacetic acid. The mixture was placed on ice for 10 min and centrifuged at 10,000 X g. Supernatant was taken out and absorbance was measured at 335 nm. A control experiment contained all the above except for the enzyme. One unit of enzyme was defined as "the amount of enzyme which produces an increase in absorbance of 0.05 at 335 nm".

## Results

**Nucleotide and protein sequence analysis:** *Tk0076* gene, consisting of 1, 572 nucleotides, encodes a subtilisin-like serine protease precursor comprising 524 amino acids. The calculated molecular weight of *Tk0076* gene product was 56.6 kDa. Signal peptide prediction using SignalP program pointed out that the putative signal-peptide consists of 23 N-terminal amino acids. The calculated molecular weight of the protein without signal-peptide was 54.1 kDa and a pI of 5.0. A prediction of the pro-peptide region was made by aligning Tk0076 amino acid sequence with various protease sequences available in database by using ProDom server which indicated a putative pro-peptide consisting of 183 amino acids after the signal-peptide. After removal of pre- and pro-peptide, the remaining mature domain or mature protein consisted of 318 amino acids. Analysis of the amino acid sequence of Tk0076 revealed that this protein was rich in hydrophobic amino acids. Ala, Gly, Ile, Leu and Val were present in high concentrations at a percentage of 9.7%, 8.2%, 8.8, 8.0% and 9.0%, respectively. Charged residues were low in number in Tk0076. Two cysteine residues were also found in the sequence indicating that the protein may form a disulphide bridge (Table 2).

**Table 2. Amino acid composition of Tk0076.**

| Amino acid | Total number | % Age | Amino acid | Total number | % Age |
|---|---|---|---|---|---|
| Ala | 51 | 9.7 | Leu | 42 | 8.0 |
| Arg | 14 | 2.7 | Lys | 31 | 5.9 |
| Asn | 18 | 3.4 | Met | 10 | 1.9 |
| Asp | 32 | 6.1 | Phe | 17 | 3.2 |
| Cys | 2 | 0.4 | Pro | 28 | 5.3 |
| Gln | 14 | 2.7 | Ser | 33 | 6.3 |
| Glu | 30 | 5.7 | Thr | 34 | 6.5 |
| Gly | 43 | 8.2 | Trp | 12 | 2.3 |
| His | 8 | 1.5 | Tyr | 12 | 2.3 |
| Ile | 46 | 8.8 | Val | 47 | 9.0 |

**Sequence comparison and phylogenetic analysis:** Amino acid sequence of Tk0076 was aligned with those of other proteases. Tk0076 displayed a highest homology of 66% with preprosubtilisine from *Thermococcus* sp. AMT11. It showed a low level of homology with other proteases even from other hyperthermophilic archaea. There was only a 25% identity between subtilisin-like serine protease from *Thermococcus onnurineus* and Tk0076. Interestingly it displayed a 34% and 33% sequence identity with intracellular protease from *Bacillus cereus* and *Bacillus thuringiensis*, respectively. The catalytically active residues $Asp^{215}$, $His^{247}$ and $Ser^{424}$ (Tk0076 numbering) were conserved in all the protease. $Asn^{354}$ which is believed to play a role in the formation of oxyanion hole at active site (Ménard & Storer, 1992) was also conserved in all the sequences (Fig. 1).

**Theoretical model of Tk0076:** Secondary and tertiary structures of Tk0076 were predicted using crystal structure of Subtilisin BPN′ (1yjb(A)) from *Bacillus amyloliquefaciens* as a template. The predicted model showed that the mature protein consisted of 7 α helices and 8 β sheets (Fig. 2). $D^{215}$, $H^{247}$ and $S^{424}$ constituted the active site where the breakage of peptide bonds of other proteins might occur. $N^{354}$ was also in close proximity of this triad (Fig. 3). The theoretical model of pro-peptide region of Tk0076 could not be constructed due to lack considerable similarity with any of the proteases whose crystal structure had been determined.

$Ca^{2+}$ ion are necessary for optimum proteolytic activity of subtilisin-like proteases and they have specific sites which are involved in binding of these metal ions. According to predicted model of Tk0076, there are two $Ca^{2+}$ binding sites Ca1 (from $I^{204}$ to $G^{208}$) and Ca2 (from $H^{222}$ to $Q^{226}$) in the predicted model.

**Amplification of Tk0076:** Polymerase chain reaction using 0076-F and 0076-R primers and *T. kodakaraensis* genomic DNA as a template resulted in the amplification of 1.5 kbp DNA fragment (Fig. 4). The amplified gene was inserted in pET-21a utilizing *Nde*I and *Hind*III sites and the resulting recombinant plasmid was named as pET-0076. *E. coli* DH5α cells were transformed using pET-0076 plasmid and selection of transformants was done using ampicillin as selection marker. Plasmid DNA was isolated from the transformant and presence of Tk0076 gene was confirmed by digestion the plasmid with *Nde*I and *Hind*III which resulted in the liberation of 1.5 kbp DNA fragment from the vector (Fig. 5). Proper insertion of the gene was confirmed by DNA sequencing.

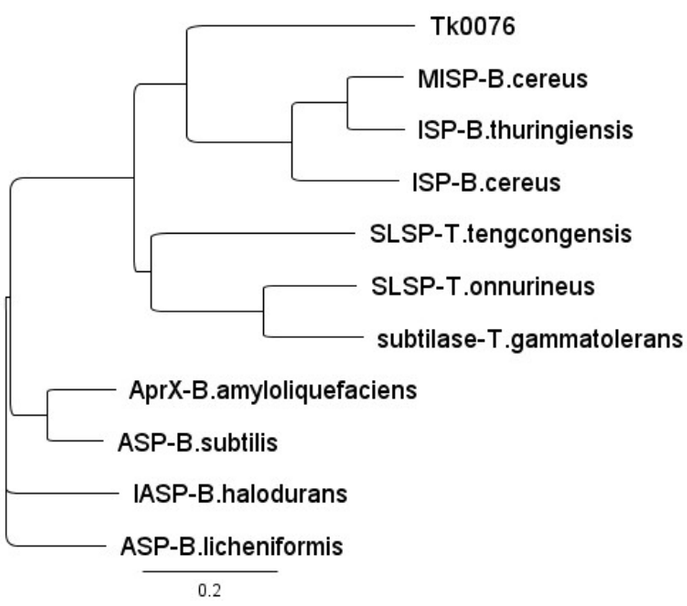


Fig. 1. Phylogenetic tree constructed by neighbour joining method. Accession numbers of the sequences used for phylogentic tree are: Subtilisin-like serine protease from *Thermoanaerobacter tengcongensis* (Accession No. NP_622477); Subtilisin-like serine protease precursor from *Thermococcus onnurineus* NA1 (Accession No. YP_002307742); AprX from *Bacillus amyloliquefaciens* (Accession No. YP_001421301); Alkaline serine protease from *Bacillus subtilis* subsp. subtilis str. 168 (NP_389608); Subtilisin-like serine protease (Subtilase) from *Thermococcus gammatolerans* EJ3 (Accession No. YP_002958734); Intracellular alkaline serine protease from *Bacillus halodurans* (Accession No. NP_242796); Intracellular serine protease from *Bacillus cereus* (Accession No. ZP_03233960); Alkaline serine protease from *Bacillus licheniformis* (Accession No. YP_079346); Intracellular serine protease from *Bacillus thuringiensis* serovar berliner (Accession No. ZP_04105404); Intracellular serine protease from *Bacillus cereus* (Accession No. NP_978391).

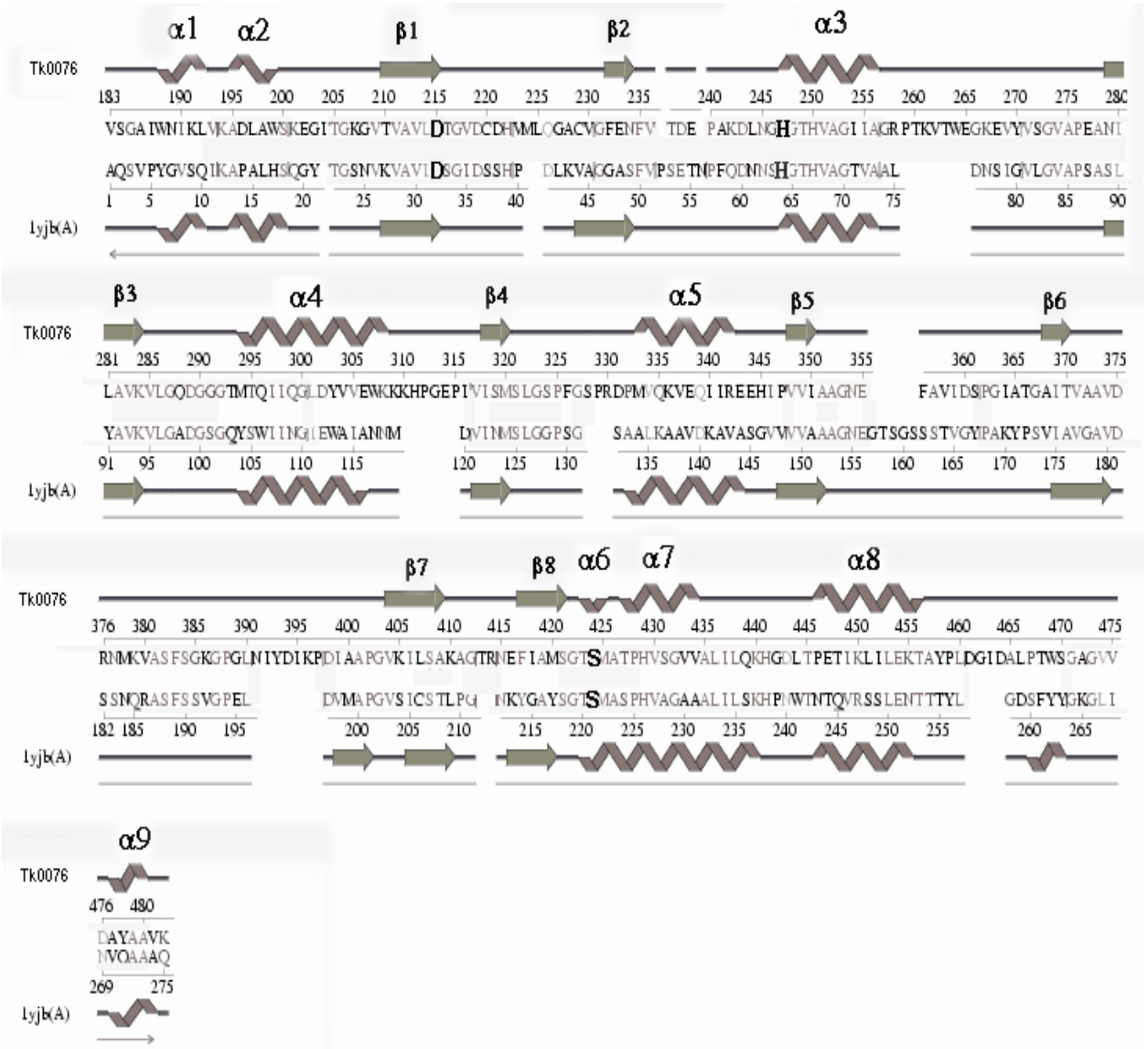


Fig. 2. Alignment of Tk0076 amino acid sequence with template ((1yjb (A)) showing secondary structure: represents the conserved residues, active site D, H and S are **shaded**. Conserved domains are shown in rectangles.

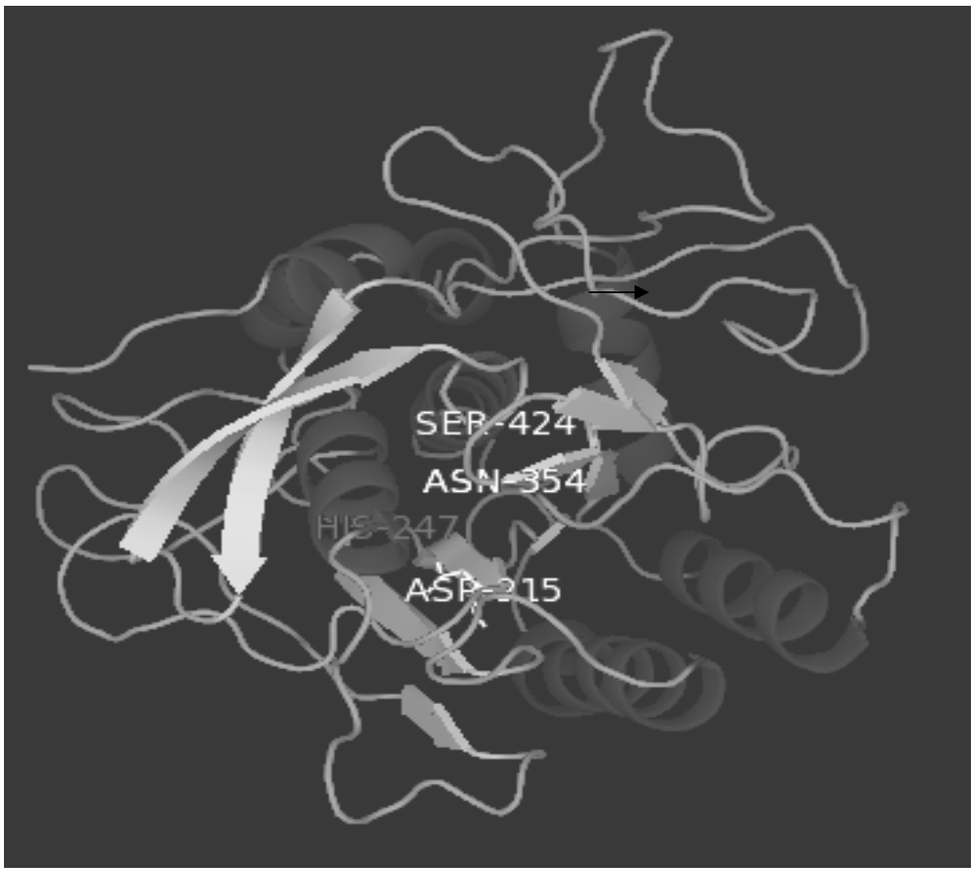


Fig. 3. Theoretical 3D structure of mature Tk0076 showing helices and sheets. Active site residues are highlighted.

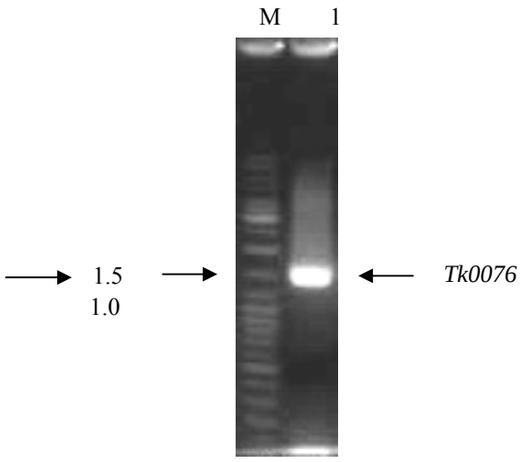


Fig. 4. Ethidium bromide stained 1% agarose gel demonstrating PCR amplified Tk0076
Lanes: M, standard marker; Lane 1, PCR amplified Tk0076

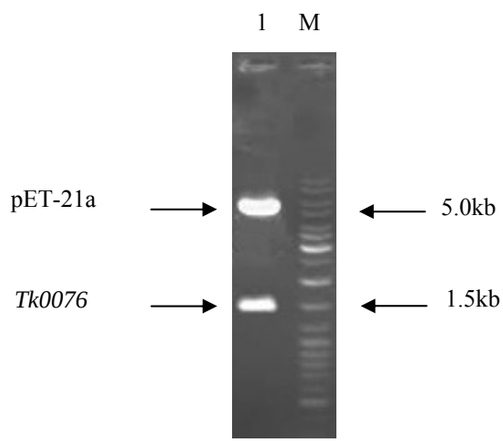


Fig. 5. Ethidium bromide stained 1% agarose gel demonstrating restriction analysis of pET-0076.
Lanes: M, standard marker; Lane 1, pET-0076 digested with *Nde*I and *Hind*III.

**Expression of Tk0076 in *E. coli*:** When *E. coli* BL21 CodonPlus (DE3)-RIL cells carrying pET-0076 were induced with 0.2 mM IPTG for 4 h, a 55 kDa protein band of recombinant Tk0076 was observed on SDS-PAGE which was absent in the cells carrying the pET-21a vector only. Analysis of the soluble and insoluble fractions, after lysis of the cells, showed that recombinant Tk0076 was produced in the soluble fraction (Fig. 6). As the recombinant Pro-Tk0076 was produced in soluble form, therefore, whole cell lysate was subjected to protease assay using azocasein as substrate in presence or absence of 2 mM $Ca^{2+}$ but no proteolytic activity could be detected in either case. When supernatant containing recombinant Tk0076 was heated at 80°C for 10 min in presence or absence of 2 mM $Ca^{2+}$, Tk0076 was precipitated indicating that although recombinant Tk0076 was soluble on production in *E. coli* but it might not folded properly and the zymogen form was unable to cleave itself to produce a catalytically active form.

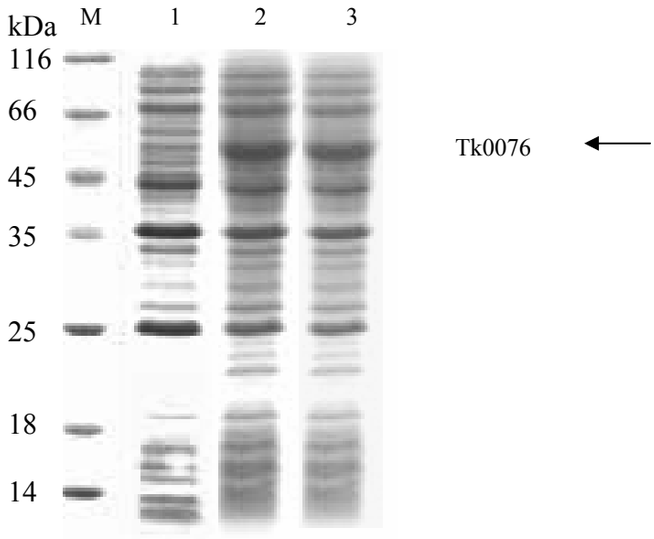


Fig. 6. CBB-stained 0.1% SDS-12% PAGE demonstrating the expression of recombinant Pro-Tk0076
Lanes: M, molecular weight standards; lane 1, total cell lysate of the host cells carrying the pET-21a(+); lane 2, total cell lysate of the host cells carrying the pPro-Tk0076 plasmid; lane 3, soluble fraction of the host cells carrying pPro-Tk0076 plasmid.

**Amplification and cloning of MP-Tk0076 in pET-21a:** As Tk0076 was enzymetically inactive we, therefore, decided to remove the predicted pro-petide region (313 amino acids at the N-terminal). For this purpose, an internal forward primer (0076MP-F) was designed to amplify the region encoding only mature domain. PCR with 0076MP-F and 0076-R primers and *T. kodakaraensis* genomic DNA as a template resulted in the amplification of 1.0 kbp DNA fragment (data not shown). The amplified DNA fragment was inserted in pET-21a utilizing the *Nde*I and *Hind*III restriction enzymes. The resulting plasmid was named as pET-MP0076. *E. coli* DH5α cells were transformed using this recombinant plasmid. The insertion of the required DNA fragment was confirmed by digesting pET-MP0076 which resulted in the liberation of 1 kbp DNA fragment from the vector (Fig. 7).

**Production and refolding of MP-Tk0076:** When *E. coli* BL21 CodonPlus (DE3)-RIL cells containing pET-MP0076 plasmid were induced with 0.2 mM IPTG for 5 h and total cell lysate was analyzed on SDS-PAGE, and a very thick band was observed which was not present in the lysate of cells carrying pET-21a vector (Fig. 8). In order to investigate whether MP-Tk0076 was produced in soluble or insoluble form, cells were lyzed by sonication and soluble fraction was separated from insoluble fraction by centrifugation and analyzed on SDS-PAGE. The recombinant MP-Tk0076 was produced as insoluble inclusion bodies (Fig. 8, lane, 4). Inclusion bodies of MP-Tk0076 were washed with 0.2% Triton X-100 and refolded as described in Materials and Methods section. The recombinant protein became precipitated after removal of urea indicating that it was not folded properly. These results suggested that pro-peptide has an important role in proper folding of the protein.

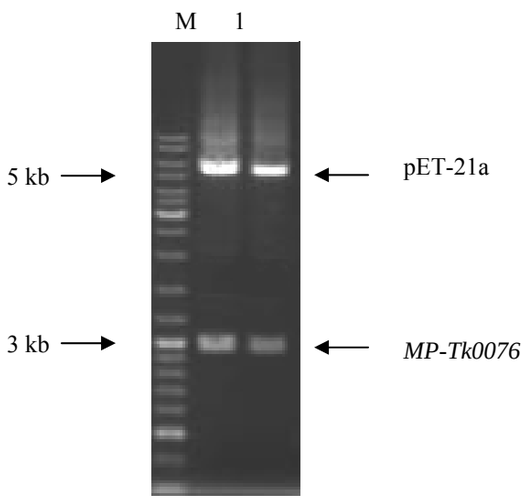


Fig. 7. Ethidium bromide stained 1% agarose gel demonstrating restriction analysis of *MP-Tk0076*/ pET-21a. Lanes: M, standard marker; Lane 1 & 2, double restricted *MP-Tk0076*/pET-21a.

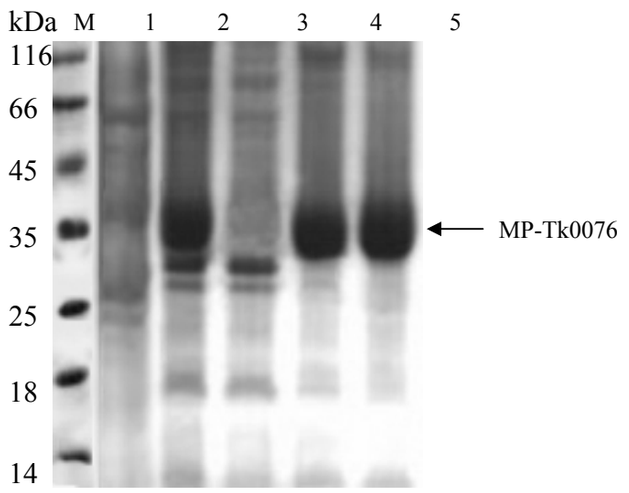


Fig. 8. CBB-stained 0.1% SDS-12% PAGE demonstrating the expression and refolding of recombinant MP-Tk0076. Lanes: M, molecular weight standards; lane 1, total cell lysate of the host cells carrying the pET-21a(+); lane 2, total cell lysate of the host cells carrying the pMP-Tk0076 plasmid; lane 3, soluble fraction of the host cells carrying pMP-Tk0076 plasmid; lane 4, insoluble fraction of the host cells carrying pMP-Tk0076 plasmid showing recombinant MP-Tk0076 protein in the form of inclusion bodies; lane 5, insoluble fraction after refolding MP-Tk10076 protein.

## Discussion

Tk0076 encoded a subtilisin-like serine protease precursor having pre-peptide, pro-peptide and mature domain. A phylogenetic tree constructed for Tk0076 revealed that it exhibited similarity with proteases characterized from genus *Bacillus* rather than closely related genus *Pyrococcus*. Variations were quite high in the pre- and pro- regions of Tk0076 compared to mature enzyme. It is speculated that subtilisin-like serine proteases are diverged from a common ancestor. Tk0076 is a good example of such proteases which seems to be a close relative of subtilisin-like serine proteases of bacterial origin.

Tk0076 was produced in *E. coli* cells in the soluble form whereas the mature protein (without pro-peptide) was produced in the insoluble form. Therefore, it can be speculated that folding of the protein might be facilitated due to the presence of this pro-peptide. It is reported that propeptide play important role in proper folding of subtilisins as intramolecular chaprons (Shinde & Inouye, 2000; Ohta *et al.*, 1991, Ohta & Inouye, 1990; Zhu *et al.*, 1989) which help them in proper folding and protect from denaturing. Tk0076 without pro-peptide could not fold properly in *E. coli* and produced in the form of inclusion bodies. The attempts to refold the protease *in vitro* went in vain and it could not attain proper conformation. It may be possible that urea could not completely denature Tk0076 as there are reports describing the enzyme activity of certain subtilisin-like serine proteases even in the presence of 8M urea (Koma *et al*., 2007). We speculate that pro-peptide of Tk0076 play an important role in proper folding of the mature protein similar to subtilisin E from *B. subtilis* (Li *et al*., 1995).

## References


Arnold, K., L. Bordoli, J. Kopp and T. Schwede. 2006. The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. *Bioinformatics*, 22: 195-201.

Atomi, H., T. Fukui, T. Kanai, M. Morikawa and T. Imanaka. 2004. Description of *Thermococcus kodakaraensis* sp. nov., a well studied hyperthermophilic archaeon previously reported as *Pyrococcus* sp. KOD1. *Archaea*, 1: 263-267.

Bendtsen, J.D., H. Nielsen, G. von Heijne and S. Brunak. 2004. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol*., 340: 783-795.

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

Davie, E.W., K. Fujikawa and W. Kisiel. 1991. The coagulation cascade: Initiation, maintenance and regulation. *Biochemistry*, 30: 10363-10370.

Fukui, T., H. Atomi, T. Kanai, R. Matsumi, S. Fujiwara and T. Imanaka. 2005. Complete genome sequence of the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1 and comparison with *Pyrococcus* genomes. *Genome Res*., 15: 352-363.

Halio, S.B., M.W. Bauer, S. Mukund, M. Adams and R.M. Kelly. 1997. Purification and characterization of two functional forms of intracellular protease PfpI from the hyperthermophilic archaeon *Pyrococcus furiosus*. *Appl Environ Microbiol*., 63: 289-295.

Hedstrom, L. 2002. Serine protease: mechanism and specificity. *Chem Rev*., 102: 4501-4524.

Huber, R. and W. Bode. 1978. Structural basis of the activation and action of trypsin. *Acc Chem Res*., 11: 114-122.

Huber, R. and K.O. Stetter. 2001. Discovery of hyperthermophilic microorganisms. *Methods Enzymol*., 330: 11-24.

Jacobs, I., M. Eliasson, M. Uhlen and J.I. Flock. 1985. Cloning, sequencing and expression of subtilisin Carlsberg from *Bacillus licheniformis*. *Nucleic Acids Res*., 13: 8913-8926.

Jeanmougin, F., J.D. Thompson, M. Gouy, D.G. Higgins and T.J. Gibson. 1998. Multiple sequence alignment with Clustal X. *Trends Biochem Sci*., 23: 403-405.

Koma, D., H. Yamanaka, K. Moriyoshi, T. Ohmoto and K. Sakai. 2007. Overexpression and characterization of thermostable serine protease in *Escherichia coli* encoded by the ORF TTE0824 from *Thermoanaerobacter tengcongensis*. *Extremophiles*, 11: 769-779.

Kopp, J. and T. Schwede. 2004. The SWISS-MODEL Repository of annotated three-dimensional protein structure homology models. *Nucleic Acids Res*., 32: 230-234.

López-García, P., A. López-López, D. Moreira and F. Rodríguez-Valera. 2001. Diversity of free-living prokaryotes from a deep-sea site at the Antarctic Polar Front. *FEMS MicrobiolEcol*., 36: 193-202.

Li, Y., Z. Hu, F. Jordan, and M. Inouye. 1995. Functional analysis of the propeptide of subtilisin E as an intramolecular chaperone for protein folding. *J. Biol. Chem.* 270, 25127-25132.

Ménard, R. and A.C. Storer. 1992. Oxyanion hole interactions in serine and cysteine proteases. *Biol Chem Hoppe Seyler*. 373:393-400.

Morikawa, M., Y. Izawa, N. Rashid, T. Hoaki and T. Imanaka. 1994. Purification and characterization of a thermostable thiol protease from a newly isolated hyperthermophilic *Pyrococcus* sp. *Appl. Environ Microbial*, 60: 4559-4566.

Neurath, H. 1984. Evolution of proteolytic enzymes. *Science*, 224: 350-357.

Neurath, H. 1985. Proteolytic enzymes, past and present. *Fed Proc*., 44: 2907-2913.

Ohta, Y. and M. Inouye. 1990. Pro-subtilisin E: purification and characterization of its autoprocessing to active subtilisin E *in vitro*. *Mol. Microbiol.* 4: 295-304.

Ohta, Y., H. Hojo, S. Aimoto, T. Kobayashi, X. Zhu, F. Jordan and M. Inouye. 1991 Pro-peptide as an intermolecular chaperone: renaturation of denatured subtilisin E with a synthetic pro-peptide. *Mol. Microbiol.* 5: 1507-1510.

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

Shinde, U. and M. Inouye. 2000. Intramolecular chaperones: polypeptide extensions that modulate protein folding. Seminars in Cell & Developmental Biology. 11: 35-44

Stahl, M. L. and E. Ferrari. 1984. Replacement of the subtilisin structural gene with an *in vitro*-derived deletion mutation. *J Bacteriol*., 158: 411-418.

Van de Ven, W.J., A.J. Roebroek and H.L. Van Duijnhoven. 1993. Structure and function of eukaryotic proprotein processing enzymes of the subtilisin family of serine proteases. *Crit Rev Oncogen* 4: 115-136.

Zhu, X., Y. Ohta, F. Jordan and M. Inouye. 1989. Pro-sequence of subtilisin can guide the folding of denatured subtilisin in an intermolecular process. *Nature* 339: 483-484.


(Received for publication 25 November 2009)