Plant Omics Journal POJ

POJ 4(3):120-125 (2011) ISSN:1836-3644

Cathepsin B-like protease from chili pepper revealed by *in silico* approach

Sanjog T. Thul[1], Feroz Khan[2], Suman P. S. Khanuja[1,3]*

[1]Genetic Resources and Biotechnology Division, Central Institute of Medicinal & Aromatic Plants, Lucknow (CSIR), India-226015
[2]Metabolic and Structural Biology Department, Central Institute of Medicinal & Aromatic Plants, Lucknow (CSIR), India-226015
[3]Present address: NutraHelix Biotech (P) Ltd, C41-42, Double Storey, Ramesh Nagar, New Delhi 110 015, India

***Corresponding author: khanujazy@yahoo.com**

Abstract

The cathepsin B-like proteases of higher plants are mostly related to stress / damage. The expression of cathepsin B-like transcript in the plant system is regarded the response towards abiotic stimuli, wounding of tissues, organ abscission. We isolated a putative cathepsin B-like proteases partial cDNA from chili pepper (*Capsicum frutescens*). A cDNA library of wound induced placental tissue transcripts was constructed in the phage vector system. Partial sequencing and *in silico* analysis revealed high levels of sequence homology to cathepsin B-like cysteine protease from the plants of solanaceae family, but much lower levels with other plant cysteine proteinases. Sequence alignment using ClustalW revealed the consensus sequences of the family solanaceae for cathepsin B-like proteases. Further, translated amino acid sequences by BLASTx revealed the conserved domains among the unrelated families. Name assignment to this cDNA as cathepsin B-like protease was based on nucleotide and translated amino acid sequence similarity which is of 91% and 97% respectively with cathepsin B-like cysteine proteinase of *Nicotiana rustica*. Our current hypothesis towards the function of this cDNA is that it encodes cathepsin B-like proteases in response to mechanical wounding in plant tissues.

Keywords: *In silico* analysis, Solanaceae, Wound induction, ESTs, Conserved domains
Abbreviations: ABA _ Abscisic acid, cDNA _ Complementary DNA, EST _ Expressed sequence tags, GA_ Gibberellic acid

Introduction

The cathepsin B proteases were originally identified in mammalian systems as lysosomal, hydrolytic enzymes as it can degrade a wide range of peptide/protein substrates (Bond and Butler, 1987). More recently, a role for cathepsin B has been demonstrated in cellular apoptosis, where it activates caspase-11 by processing the pro-form, and can also directly induce nuclear apoptosis (Vancompernolle et al., 1998). Cathepsin B is an ancient family of eukaryotic cysteine proteases. In the process of storage protein utilization, it has been shown that many proteinases are involved in the degradation of storage proteins for nutrient mobilization. They are synthesized as pre-proteins that are processed either auto-catalytically or with the aid of a processing enzyme, and are stored in the vacuole or the lysosome, or are externally secreted. Among the cysteine proteases, the K, S, L, O, B, C and H cathepsins have been widely studied in mammals (Wex et al., 1999) and more recently H and L in plants (Ueda, et al., 2000, and references therein). By contrast, only a few cathepsin B-like proteases of plant origin have been described so far. Hansen and Hannapel (1992) reported a cathepsin D inhibitor cDNA, p749, which were identified in the genomic DNA of tomato (*Lycopersicon esculetum*) by southern hybridization and of two non-tuber-bearing potato species (*Solanum tuberosum* and *S. brevidens*). To date, there have been few reports of cathepsin B-like sequences in plants (Ward et. al., 1997). A cDNA encoding a thiol protease similar to cathepsin B from mammalian cells was isolated from aleurone layers of wheat (Cejudo et al., 1992*a*). The corresponding mRNA accumulated in the scutellum and the aleurone layers of germinating grains where it was under the regulation of gibberellin (GA) and abscisic acid (ABA). The analysis of the promoter from the corresponding gene showed that the regulation of its expression was at the level of transcription (Cejudo et al., 1992b and Gubler et al., 1999). In *Nicotiana rustica*, a wounding-responsive mRNA was isolated from roots and was shown to be expressed in most plant organs (Lidgett et al., 1995). In addition, the sequences of cathepsin B-like proteases from *Ipomoea batata* and *Arabidopsis thaliana* are available in the data banks, but there is no record of cathepsin B-like proteases from the genus *Capsicum*.

Materials and methods

Plant material

Fruits of chili pepper collected from Assam (North-east India) were used as an experimental material. Fruits were dissected to induce wounding and separate out the placental tissues from seeds and fruit wall. The dissected placental tissues were further used for RNA extraction.

RNA extraction and cDNA library construction

mRNA was isolated from placental tissues according to published method (Shukla et al., 2005). cDNA synthesis of mRNA and cloning was performed using ZAP Express cDNA synthesis and ZAP Express cDNA Gigapack III Gold cloning kit (Stratagene, USA). Amplified phage library was screened for recombinant bacterial colony by blue/white

Table 1. Details of identified protein domain, family and active sites in *C. frutescens*.

S.No.	Domain/Motif	Description	PSSM ID	E-value
1.	CD02620	Peptidase_C1A_CathepsinB [cd02620], Cathepsin B group; composed of cathepsin B and similar proteins, including tubulointerstitial nephritis antigen (TIN-Ag). Cathepsin B is a lysosomal papain-like cysteine peptidase which is expressed in all tissues and functions primarily as an exopeptidase through its carboxydipeptidyl activity.	30294	1.25e-14
2.	CD02698	Peptidase_C1A_CathepsinX [cd02698], Cathepsin X; the only papain-like lysosomal cysteine peptidase exhibiting carboxy- monopeptidase activity. It can also act as a carboxydipeptidase, like cathepsin B, but has been shown to preferentially cleave substrates through a monopeptidyl carboxypeptidase pathway.	30296	7.98e-06
3.	Pfam00112	Peptidase_C1 [pfam00112], Papain family cysteine protease.	143889	2.99e-10
4.	Smart00645	Pept_C1[smart00645], Papain family cysteine protease.	128893	1.16e-09
5.	CD02248	Peptidase_C1A [cd02248], Peptidase C1A subfamily (MEROPS database nomenclature); composed of cysteine peptidases (CPs) similar to papain, including the mammalian CPs (cathepsins B, C, F, H, L, K, O, S, V, X and W).	30292	1.18e-04
6.	CD02621	Peptidase_C1A_CathepsinC [cd02621], Cathepsin C; also known as Dipeptidyl Peptidase I (DPPI), an atypical papain-like cysteine peptidase with chloride dependency and dipeptidyl aminopeptidase activity, resulting from its tetrameric structure which limits substrate access.	30295	4.97e-03
7.	PTZ00364	Dipeptidyl-peptidase I precursor [PTZ00364]	173557	4.24e-03

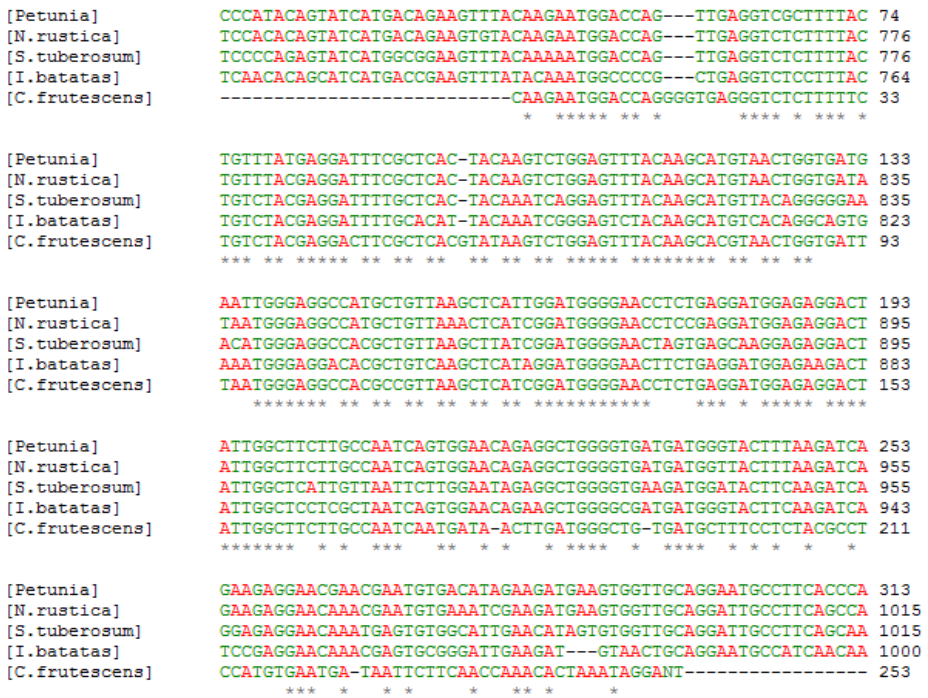

Fig 1. Multiple sequence alignment of cDNA of different solanaceae plants showing conserved residues.

colony selection as per recommended procedure of manufacturer. Recombinant white colonies were selected and grown in 5 ml LB media with selective antibiotic. The plasmids were isolated by modified polyethylene glycol precipitation method (Sambrook et al., 1989). Isolated plasmids were screened for the inserts by restriction digestion using *Eco RI* and *Hind III* restriction enzymes.

Nucleotide sequencing and in silico analysis

Single-pass sequencing was performed to obtain partial sequences by BigDye Terminator cycle sequencing (Applied Biosystems, Foster City, CA, USA). Each sequence obtained was assessed manually to determine sequence quality. Significant numbers of clones were found to hold the signature sequence of cathepsin gene when searched for homology in NCBI database. Out these, the clone having longest stretch of cDNA nucleotide of candidate gene was selected for further confirmation of sequence. The amplification of candidate gene was carried out by growing respective single recombinant cloned bacterial colony in liquid LB medium. No mixed or overlapping sequence was observed from sequencing of plasmids. Sequence homology search was performed using BLASTn program (Altshcul et al., 1997) at the DNA sequence databases of NCBI (www.ncbi.nlm.nih.gov.). The sequences then were analyzed using clustalW software package (www.ebi.edu.uk./EMBL) for consensus sequence among the homologous sequences of different plant origin [*Nicotiana rustica* (CAA57522.1), Petunia X hybrida (AAU81590.1), *Solanum tuberosum* (AAR25800.1), *Ipomia batatas* (AAK69541.1)], which were derived from datatabase. Further, the cDNA sequences were

Table 2. Similarity of *N. rustica* Cathepsin B-like cysteine proteinase with *C. frutescens*.

BLAST sequence similarity	Query seq.	*Nicotiana rustica*	*C. frutescens*
	Database	Plants	Plants
	Max. match	emb|CAA57522.1	emb|CAA57522.1
	Description	Cathepsin B-like cysteine proteinase; *N. rustica*; 356 aa	Cathepsin B-like cysteine proteinase; *N. rustica*; 356 aa
	Score	745 bits (1923)	87.4 bits (215)
	Expect	0.0	3e-17
	Identities	356/356 (100%)	38/39 (98%)
	Positives	356/356 (100%)	39/39 (100%)
	Gaps	0/356 (0%)	0/39 (0%)
Sequence length		356 AA; complete	51 AA; partial
Domain	CDD (NCBI)	Belongs to the peptidase C1 family	Peptidase_C1A_CathepsinB [cd02620]
UniProtKB/TrEMBL		Q40413	-
Protein name		Cathepsin B-like cysteine proteinase	Cathepsin B-like cysteine proteinase
Gene name		catch B	gi|70972097|, gb|DR741973.1
Gene Ontology (GO)	Biological process	Proteolysis	Proteolysis
	Molecular function	Cysteine-type endopeptidase activity; Hydrolase; Protease; Thiol protease	Cysteine-type endopeptidase activity
Protein family and domains	InterPro	IPR000169 Active site (Peptidase, cysteine peptidase active site) IPR013128 Family (Peptidase C1A, papain) IPR000668 Domain (Peptidase C1A, papain C-terminal) IPR015643 Peptidase_C1A_cathepsin-B IPR012599 Propeptide_C1A.	IPR000169 Active site (Peptidase, cysteine peptidase active site) IPR000668 Domain (Peptidase C1A, papain C-terminal)
	PANTHER	PTHR12411:SF16 CathepsinB_like PTHR12411 Peptidase_C1A	PTHR12411:SF16 CathepsinB_like
	Pfam	PF00112. Peptidase_C1 PF08127. Propeptide_C1	PF00112. Peptidase_C1
	PRINTS	PR00705. PAPAIN.	
	SMART	SM00645. Pept_C1	Pept_C1[smart00645], Papain family cysteine protease
	PROSITE	PS00139. THIOL_PROTEASE_CYS PS00639. THIOL_PROTEASE_HIS	PS00639. THIOL_PROTEASE_HIS

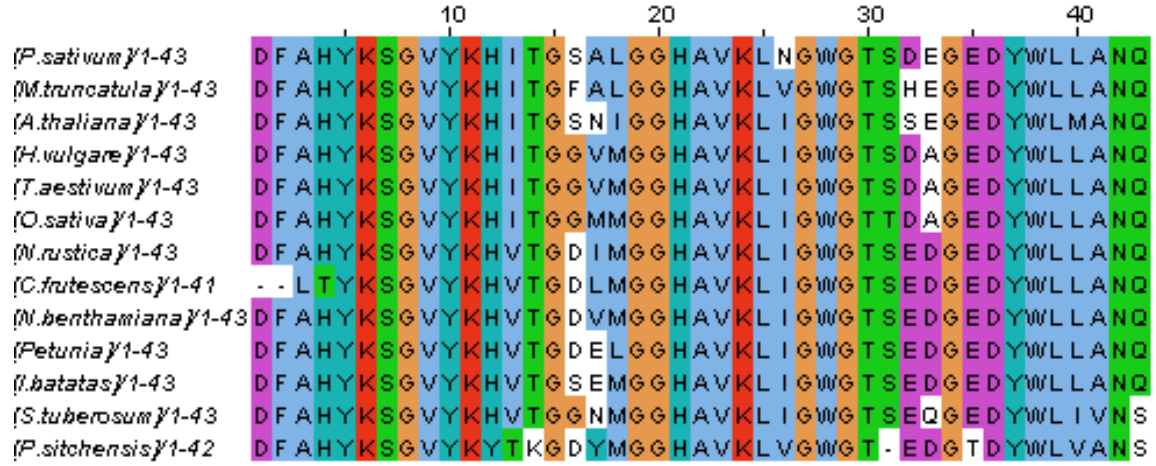

Fig 2. Sequence alignment of translated amino acid residues revealing conserved domain of cysteine peptidase active site.

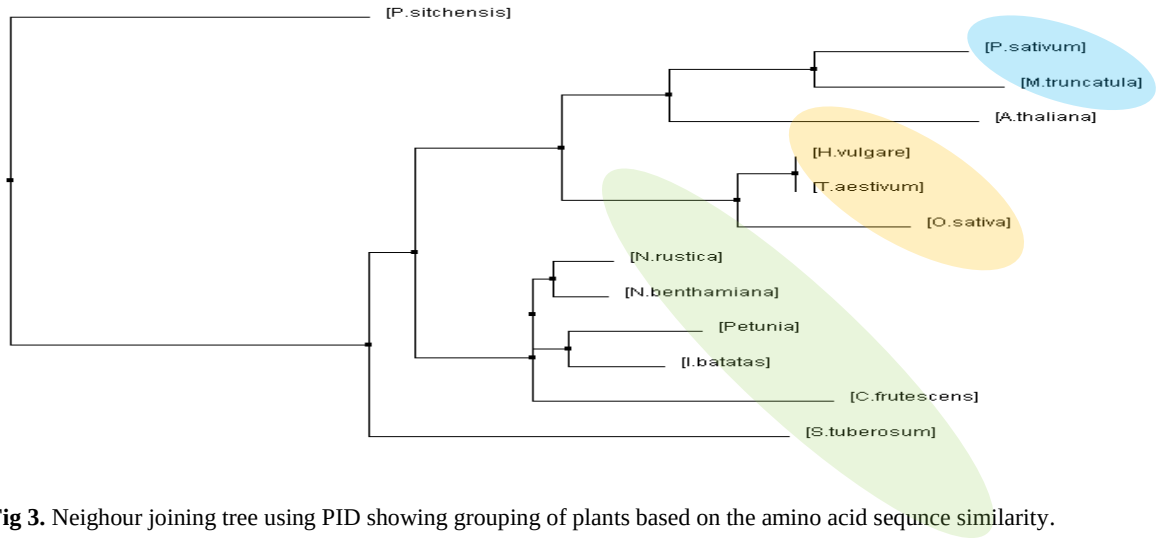

Fig 3. Neighour joining tree using PID showing grouping of plants based on the amino acid sequnce similarity.

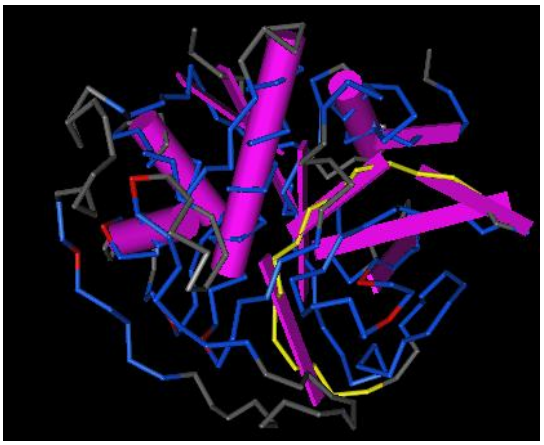

Fig 4. Three dimensional structure of peptidase C1A Cathepsin-B showing conserved domain recognized by amino acids of *C. frutescens* (highlighted by yellow stick).

translated using BLASTx and the amino acid sequences were compared with the available amino acid sequences in the database of plants of different families like poaceae [*Hordeum vulgare* (AJ310426), *Triticum aestivum* (X66013) and *Oryza sativa* (AY916493)] fabaceae [*Pisum sativum* (AJ251536), *Medicago truncatula* (AY336982)], brassicaceae [*A. thaliana* (NM_178950)] and pinaceae [*Picea sitchensis* (ABK23329)]. The obtained EST sequences were submitted to the EST database of NCBI under the accession number DR741973. The 3D structure of candidate protein was elucidated using Cn3D version-4.2 software (www.ncbi.nlm.nih.gov) to align the proposed stretch of translated amino acid sequences.

Results and discussion

The discovery of proteinase inhibitor transcripts and proteins accumulates systemically in the leaves of wounded plants has led to the investigation of possible signal pathways (Ryan, 1988). Partial sequences of 253 base of cDNA clone obtained from dissected placental tissues of *C. frutescens* were analyzed *in silico* to identify the putative homologous sequences related to wound inducible genes. In search of novel promoter region, Fei et al., (2009) used bioinformatics-based approach towards identification of short 256 bp genomic DNA fragment from *Solanum lycopersicum*. The used approach revealed the presence of several motifs for plant transcription factors such as circadian, TGA-element and motifs involved in light responsive control. In present study, sequence homology of 253 bp of cDNA clone was searched using BLASTn program and revealed the high levels of homology with cathepsin B-like cysteine protease of *N. rustica* with 91% similarity. Further, the comparative analysis of translated amino acid showed highest positive identity with cathepsin B-like cysteine protease (Table 2). The nucleotide sequence shared significant homology with plants like *N. rustica*, *Pitunia*, *S. tuberosum*, *I. batatas*, with high score value 226, 202, 99.6 and 83.8, respectively. Interestingly, all these plants were belongs to the family solanaceae that also comprises genus *Capsicum*. Besides, results showed diversity in sequence similarity within studied sequences such as mismatches and gaps. Since mismatches indicate substitution mutation and gaps indicate probability of addition or deletion mutation, therefore we hypothesized that mutation might evolve them as a separate entity of the family solanaceae. To validate this hypothesis, the homologous sequences from different plant origin were analyzed for consensus nucleotide sequences or patterns (Fig. 1). Further nucleotide sequences were translated to amino

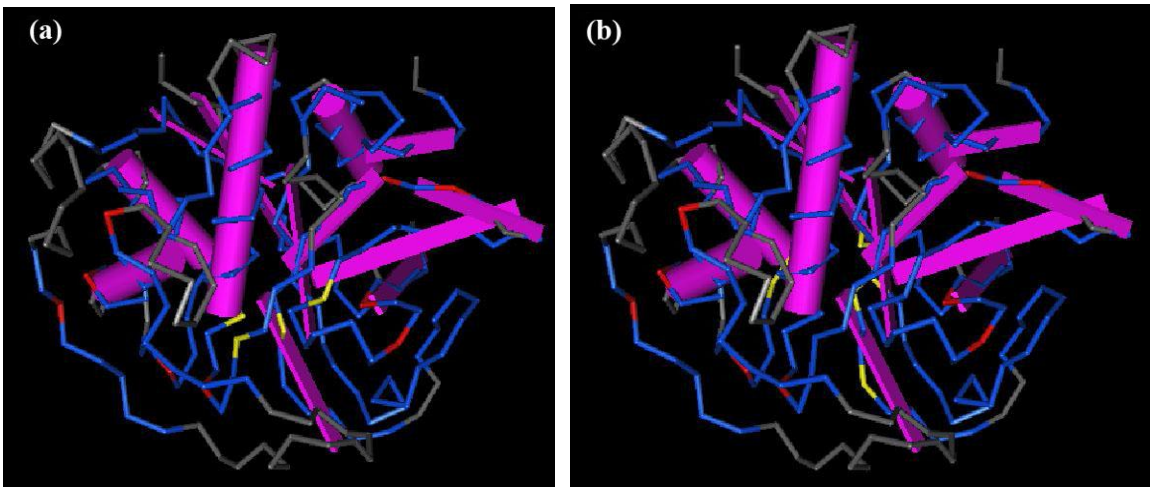

Fig 5. Three dimensional structure of peptidase C1A Cathepsin-B, the active site (a) and S2 subsite (b) shown by yellow sticks.

acid sequences using BLASTx (NCBI) and compared with the amino acid sequences obtained from database belonging to plants of different families through ClustalW (Fig. 2). The comparative analysis revealed that conserved protein domains are not only shared by these plants (*N. rustica*, *Pitunia*, *S. tuberosum*, *I. batatas* and *C. frutescens*), but also by plants of other families (poaceae, fabaceae, brassicaceae and pinaceae). Although *P. sitchensis* belongs to family pinaceae (Conifers), but still a major portion of the conserved protein domains shared the same amino acid sequences, based on which these were delineated as separate species (Fig. 3). Similarly, *cis*-acting upstream regulatory elements of polygalacturonase inhibitory protein (PGIP) encoding genes have been detected using bioinformatics based sequence analysis in seven different plant species (Kumar et al., 2009). Further, the domain/motif wide search was made (Table 1) and exhaustive comparative study was done with *N. rustica* indicating a significant similarity (Table 2). Hansen and Hannapel (1992) demonstrated the expression of p749 genes (similar to cathepsin-D) in leaves which was induced at the RNA level in response to wounding. High levels of p749 transcripts were detected in polyadenylated RNA extracted from locally wounded leaves 12 h after wounding. Interestingly, no sequences showed homology with the probe of p749 cDNA upon southern hybridization with genomic DNA from egg plant, pepper or tobacco (Hansen and Hannapel, 1992). As in present *in silico* analysis, it was quite interesting that the isolated partial cDNA sequences showed a very prominent consensus region among the nucleotide sequences and conserved domains in case of amino acid sequences, this inferred that the cathepsin-B proteases might be more conserved in family solanaceae compared to that of cathepsin-D. Also it was pointed out that the cathepsin gene showed rhythmic expression and its expression increased in response to wounding (Lidgett et al., 1995). The dissection of fruits for the separation of placental tissues from seeds and fruit wall induced the expression of mRNA and by this means the present investigation supports the study of expression of cathepsin like proteinase on wound induction (Hansen and Hannapel, 1992; Lidgett et al., 1995). It is also inferred that likewise the cathepsin variants in human (K, S, L, O, B, C and H), the variants of cathepsin also present in plant kingdom (B, D, H and L). Upon the conserved domain search in NCBI database elucidate the homology with peptidase C1 superfamily. Within the conserved domain of cathepsin-B (M*GG***H***A*VKLIGWGTS) from various origin (Fig. 4), which was elucidated using Cn3D ver. 4.2 (www.ncbi.nlm.nih.gov) showed that the His (hydrophilic amino acid) is the part of active site (Fig. 5a). Whereas the hydrophobic amino acids Gly and Ala (Italicized letters) is the part of S2 subsite (Fig. 5b), which is the dominant substrate specificity subsite of papain-like cysteine proteases. The site prefers for bulky hydrophobic or aromatic residues at the P2 side chain of the substrate to occupy the S2 subsite, indicates the perfect prediction of the function of EST isolated from the *C. frutescens*. In conclusion, despite the unfurnished work using plant host system for the identification of the role and function of the cDNA in genome of *C. frutescents*, the *in silico* analysis yields a considerable results to hypothesize the function of cDNA as a probable candidate for cathepsin B-like protease. The approach used for the achievement of aim to elucidate the identification of partial cDNA sequence was seems to be most appropriate.

Acknowledgment

The authors are thankful to Central Institute of Medicinal and Aromatic Plants, Lucknow and Council of Scientific and Industrial Research, New Delhi for providing financial support. Authors are also grateful to Dr. A. K. Shasany for extending help during cDNA library construction and sequencing.

References

Altshcul SF, Gish W, Miller W, Myers EW, Lipman DJ (1997) Basic local alignment search tool. J Mol Biol 215: 403-410

Bond JS, Butler PE (1987) Intracellular proteases. Annual Rev Biochem 56: 333-364

Cejudo FJ, Murphy G, Chinoy C, Baulcombe DC (1992*a*) A gibberellin-regulated gene from wheat with sequence homology to Cathepsin B of mammalian cells. Plant J 2: 937–948

Cejudo FJ, Ghose TK, Stabel P, Baulcombe DC (1992*b*) Analysis of the gibberellin-responsive promoter of a cathepsin B-like gene from wheat. Plant Mol Biol 20: 849–856

Fei CK, Ismail I, Ismail SI, Natorajan D, Zainal Z (2009) Identification of a short putative 5´ regulatory sequence from transgenic hairy root of tomato-regulating specific expression pattern. Plant Omics J 2(5): 206-213

Gubler F, Raventos D, Keys M, Watts R, Mundy J, Jacobsen JV (1999) Target genes and regulatory domains of the GAMYB transcriptional activator in cereal aleurone. Plant J 17: 1–9

Hansen JD, Hannapel DJ (1992) A wound-inducible potato proteinase inhibitor gene expressed in non-tuber-bearing species is not sucrose inducible. Plant Physiology 100: 164–169

Kumar GM, Mamidala P, Podile AR (2009) Regulation of Polygalacturonase-inhibitory proteins in plants is highly dependent on stress and light responsive elements. Plant Omics J 2(6): 238-249

Lidgett AJ, Moran M, Wong KA, Furze J, Rhodes MJ, Hamill JD (1995) Isolation and expression pattern of a cDNA encoding a cathepsin B-like protease from *Nicotiana rustica*. Plant Mol Biol 29: 379-384

Ryan CA (1988) Oligosaccharide signaling for proteinase inhibitor genes in plant leaves. Recent Adv Phytochem 22: 163-180

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: A laboratory manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Shukla AK, Shasany AK, Khanuja SPS (2005) Isolation of poly (A^+) mRNA for downstream reactions from some medicinal and aromatic plants. Ind J Exp Biol 43: 197-201

Ueda T, Seo S, Ohashi Y, Hashimoto J (2000) Circadian and senescence-enhanced expression of a tobacco cysteine protease gene. Plant Mol Biol 44: 649–657

Vancompernolle K, VanHerreweghe F, Pynaert G, VandeCraen M, Devos K, Totty N, Sterling A, Fiers W, Vandenabeele P, Grooten J (1998) Atractyloside-induced release of cathepsin B, a protease with caspase-processing activity. FEBS Lett 438: 150-158

Ward W, Alvarado L, Rawlings ND, Engel JC, Franklin C, McKerrow JH (1997) A primitive enzyme for a primitive cell: The protease required for excystation of *Giardia*. Cell 89: 437-444

Wex T, Levy B, Wex H, Bromme D (1999) Human cathepsins F and W: A new subgroup of cathepsins. Biochem Biophys Res Commu 259: 401-407