

## Nucleotide sequence of a genomic clone encoding a cowpea (*Vigna unguiculata* L.) trypsin inhibitor

**Paulraj Kirubakaran Lawrence\***

Department of Biological Sciences  
Texas Tech University  
Box No 43131  
Lubbock, TX 79409-3131 USA  
Tel: 806-742-2740  
Fax: 806-742-2963  
E-mail: p.k.lawrence@ttu.edu

**Jayaveeramuthu Nirmala**

Department of Biological Sciences  
Texas Tech University  
Lubbock, TX 79409-3131 USA  
Tel: 806-742-8084  
Fax: 806-742-2963  
E-mail: nilawren@ttacs.ttu.edu

**Kripa Ram Koundal**

National Research Center on Plant Biotechnology  
Indian Agricultural Research Institute  
PUSA, New Delhi, India-110012  
Tel: 91-011-5711554  
Fax: 91-011-5766420  
E-mail: krk\_pblo@iari.emet.in

**Financial support:** Senior Research Fellowship provided to P.K.L., by Council for Scientific and Industrial Research, New Delhi, India-110012.

**Keywords:** cowpea genomic library, nucleotide sequencing, sequence analysis.

**A protease inhibitor gene has been isolated from native cowpea (*Vigna unguiculata* L.) cv. 130 from 1-ZAP II genomic library. The nucleotide sequence of this genomic clone shared 86% homology with a cowpea trypsin inhibitor f IV mRNA, and 81% with Bowman-Birk protease inhibitor genes of soybean. The isolated gene had TATA and CAT signals in the 5' upstream region. The longest open reading frame had 504 bases, encoding 167 amino acids from the predicted coding region. The active serine centers of this protein was found between 105-158 amino acid residues, with a 69 amino acids long signal peptide in its N terminal.**

Proteinase inhibitors are potential tools of crop improvement targeting plant protection and human nutrition. This concept of plant protection is rapidly expanding and opens up diverse areas to harness the potential of these genes. The serine and its trypsin subclass of protease inhibitors exhibit antimetabolic and antinutritional effects against a wide variety of crop pests belonging to the orders of Lepidoptera, Coleoptera and Orthoptera (Shulke and Murdock, 1983). These inhibitors are lethal to the larvae of *Callosobruchus maculatus* (Gatehouse and Boulter, 1983) and *Manduca sexta* (Shulke and Murdock, 1983) at 10% of their diet. They also

promote the loss of sulphur based amino acids, leading to weak and stunted growth and ultimate death. These inhibitors are cysteine rich with a molecular mass of 8-20 kD (Peng-Wen Chen et al. 1997). Many of these are products of multigene families and exhibit different specifications towards proteases (Ryan, 1990).

The gene size and coding regions of these inhibitors are small, devoid of introns (Boulter, 1993), and comprised of readily identifiable core region covering the invariant cysteine residues and active serine centers that are bound to highly variable amino and carboxy terminal regions. The reaction sites of these inhibitors are mutating faster than amino acids in rest of the proteins, implying their role in defense against insect pests, which may exert a strong selection pressure on these proteins to conserve the reaction sites related to plant defense (Laskowski et al. 1988). The availability of these inhibitors may allow to engineer transgenic crops inherently resistant to these insect pests. These aspects therefore encourages the identification and characterization of novel protease inhibitor genes from diverse plant species to be used against insect pest. This report describes the screening of a cowpea genomic library with a tomato protease inhibitor cDNA probe to identify, characterize and clone a new protease inhibitor gene.

\* Corresponding author

## Materials and Methods

### Construction and screening of genomic library

Genomic DNA of one week old etiolated seedlings of cowpea (*Vigna unguiculata* L.) cv. V130 was isolated (Dellaporta et al. 1983), purified and partially restricted with *EcoRI*. This was fractionated on a sucrose density (10-30%) gradient and fragments of 2-7 kb were pooled together and approximately 10 µg of DNA was used for cloning in  $\lambda$ -ZAP II vector following the manufacturer's protocol (Stratagene, La Jolla, USA) and amplified in *E. coli* XL-1 Blue. The library was screened using a 700 bp cDNA clone encoding tomato protease inhibitor II (pT2, obtained from Prof. Ryan, Washington State University, Pullman, USA) according to the method of Sambrook et al. (1989). Positive clones were identified after 3 rounds of screening.

### *In vivo* excision and sequencing

The 3 recombinant lambda clones obtained after tertiary screening were excised into pBluescript II SK<sup>+</sup> plasmid using ExAssist helper phage (Short et al. 1988) and transformed into SOLR (Stratagene, La Jolla, USA) super competent cells. DNA from 10 excised positive clones (pLK1 to pLK10) was isolated using Qiagen mini-prep kit (Qiagen Inc., Valencia, USA) and digested with *EcoRI* (5 U/µg of DNA) in a volume of 20 µl. The digested plasmid DNA was analyzed on 0.7% agarose gel, transferred to Hybond N<sup>+</sup> membrane and probed using pT2 (Sambrook et al. 1989).

One of the genomic clone (pLK8) was subjected to exonuclease III deletion using MBI Exo III kit (MBI Fermentas, Hanover, USA) and transformed into *E. coli* DH5 $\alpha$ . The plasmid DNA (3-4 µg) from ten successive overlapping clones were sequenced by di-deoxy chain termination method (Sanger et al. 1977) using Sequenase ver 2.0 sequencing kit (Amersham Inc., Cleveland, USA). The clones were sequenced using both forward and reverse M13 universal primers. The reaction products were separated on 8% polyacrylamide gel containing 8 M urea using LKB Macrophor (LKB Pharmacia, Switzerland) equipment. The nucleotide sequence was analyzed by Lasergene programme (DNASTAR), Genbank BLASTN (Altschul et al. 1997) and FASTA (Pearson and Lipman, 1988).

The sequence data has been submitted to EMBL nucleotide data base (Accession No. AJ 271752).

## Results and Discussion

After 3 rounds of screening 3 positive lambda clones were obtained. *In vivo* excision resulted in 10 plasmid clones, all having an insert size of 2 kb when restricted with *EcoRI*.

Hence, one of the genomic clone designated as pLK8 was sequenced after deletion with Exonuclease III.

The deletion of genomic clone pLK8 resulted in successive overlapping clones of approximately 200-225 bp shorter, ranging from 2-1.8 kb. These clones had approximately 25 bp region overlapping with the preceding clone. These clones were sequenced using forward and reverse primers in order to confirm sequences from both DNA strands. The sequences of deleted clones were aligned (Megalign) and its total length was found to be 1997 bp.

The sequence from 1-990 bp of the original 2 kb clone did not show any significant homology with any protease inhibitor gene or its promoters, so 1007 bp are shown in [Figure 1](#). The clone had transcription control regions like TATA signals between 256-267 bases and a CAT signal at 335 bp. The longest ORF analyzed was between 355 and 858 bases in the first reading frame and the ATG for this ORF starting at 355 bp. This ORF on translation gave a protein with 167 amino acids (indicated by one letter code, Figure 1). The molecular weight of this protein was calculated to be 18.5 kD; of which 18 amino acids were strongly basic, 20 strongly acidic, 40 hydrophobic and 62 are polar amino acids with an isoelectric point of 6.5.

The protein has a N terminal signal sequence of 69 amino acids which is important in its translocation as has been reported in many proteinase inhibitors such as soybean (Hammond et al. 1984), cowpea (Hilder et al. 1989), pea (Domoney et al. 1995), maize (Rohrmeier and Lehle, 1993) and alfalfa (McGurl et al. 1995). The cowpea protease inhibitor has the signal peptide cleavage site between 69-70 amino acid residues *i.e.*, TTA-AM, which is comparable to the alfalfa trypsin inhibitor (ATI) having a signal peptide of 44 or 55 amino acid residues which may target the ATI to the ER and then eventually deposit it in the vacuole of the cell (McGurl et al. 1995). Comparison for conserved domain in the protein by NCBI CD-search showed that the regions from 105-158 have the active domain similar to serine class of Bowman-Birk inhibitors. The active centre serines (6) are located with in this above region. The total number of cysteine residues (15) are identical to that of the cowpea f IV inhibitor mRNA (Hilder et al. 1989) with most of them found in the active domain.

The total number of bases translated in the ORF is 504, with 58.28% A+T and 41.72% C+G similar to that of cowpea trypsin inhibitor mRNA f IV (Hilder et al. 1989) which has 60.5% A+T and 39.4% G+C in their expressed regions. The regions from 627-851 bp of the sequenced fragment shared 81% homology with 191-394 bp of soybean C-II protease inhibitor gene which is 645 bp long (Jourdrier et al. 1987), and with 138-341 bp of BBI mRNA (Baek et al. 1994), which is 488 bp long respectively. The maximum percentage (86%) of homology ([Figure 2](#)) was found with cowpea trypsin inhibitor mRNA coding for inhibitor f IV which is 583 bp long (Hilder et al. 1989).

Further, it was observed that sequences between 627 to 857 shared homology with the coding regions of all the above mentioned inhibitors. The protease inhibitor genes are typically devoid of any intervening sequences (Boulter et al. 1993) and none of the genes sequenced so far have any introns, similarly the isolated sequenced cowpea trypsin inhibitor gene is also devoid of introns. The amino acid sequence of this trypsin inhibitor shared a 80% homology (Figure 3a) with the protein of cowpea trypsin inhibitor f IV, 63.4% with soybean D-II inhibitor (Figure 3b) and 65% with soybean C-II inhibitor (Figure 3c). From the above analysis it could be clearly concluded that the isolated genomic clone had the complete gene encoding a trypsin inhibitor and designated as CpTi, which could be used as a potential tool against crop pests.

This gene has been cloned in Bin AR, a binary vector and used to transform pigeonpea by *Agrobacterium tumefaciens* mediated genetic transformation. The integration of this gene in transformed pigeonpea has been confirmed by Southern hybridization and also its expression by northern analysis (Lawrence and Koundal, 2001). These transformed plants are further to be tested against pod borer (*Heliothis armigera*) which is under progress.

### Acknowledgements

We are thankful to Prof. C.A. Ryan Washington State University, Pullman, for providing tomato protease inhibitor cDNA probe. One of the authors (P.K.L) also wish to thank the Council of Scientific and Industrial Research for granting fellowship during the course of this investigation.

### References

Altschul, S.F.; Madden, T.L.; Schaffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W. and Lipman D.J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research* 25:3389-3402.

Baek, J.M.; Song, J.C.; Choi, Y.D. and Kim, S.I. (1994). Nucleotide sequence homology of cDNAs encoding soybean Bowman-Birk type proteinase inhibitor and its isoforms. *Bioscience Biotechnology and Biochemistry* 58:843-846.

Boulter, D. (1993). Insect pest control by copying nature using genetically engineered crops. *Biochemistry* 34:1453-1466.

Dellaporta, S.L.; Wood, J. and Hicks, J.B. (1983). A plant DNA miniprep: Version II. *Plant Molecular Biology Reporter* 1:19.

Domoney, C.; Welham, T.; Sidebottom, C. and Firmin, J.L.

(1995). Multiple isoforms of Pisum trypsin inhibitors result from modification of two primary gene products. *FEBS Letters* 360:15-20.

Gatehouse, A.M.R. and Boulter, D. (1983). Assessment of the antimetabolic effects of trypsin inhibitors from cowpea (*Vigna unguiculata*) and other legumes on development of the bruchid beetle *Callosobruchus maculatus*. *Journal of the Science of Food and Agriculture* 34:345-350.

Hammond, R.W.; Foard, D.E. and Larkins, B.A. (1984). Molecular cloning and analysis of a gene coding for the Bowman-Birk protease inhibitor in soybean. *Journal of Biological Chemistry* 259:9883-9890.

Hilder, V.A.; Barker, R.F.; Samour, R.A.; Gatehouse, A.M.R.; Gatehouse, J.A. and Boulter, D. (1989). Protein and cDNA sequences of Bowman-Birk protease inhibitors from the cowpea (*Vigna unguiculata* Walp.). *Plant Molecular Biology* 3:701-710.

Jourdrier, P.E.; Foard, D.E.; Floener, L.A. and Larkins, B.A. (1987). Isolation and sequence of cDNA encoding the soybean protease inhibitor P I, IV and C-II. *Plant Molecular Biology* 10:35-42.

Laskowski, M. Jr.; Kato, I.; Kohr, W.J.; Park, S.J.; Tashino, M. and Whatley, H.S. (1988). Positive Darwinian selection in evolution of protein inhibitors of serine proteinases. *Cold Spring Harbor Symposia on Quantative Biology* 52:545-553.

Lawrence, P.K. and Koundal, K.R. (2001). *Agrobacterium tumefaciens* mediated transformation of pigeonpea (*Cajanus cajan* (L.) Millsp) and molecular analyses of regenerated plants. *Current Science*.

McGurl, B.; Mukherjee, S.; Kahn, M. and Ryan, C.A. (1995). Characterization of two proteinase inhibitor (ATI) cDNAs from alfalfa leaves (*Medicago sativa* var. Vernema): The expression of ATI genes in response to wounding and soil microorganisms. *Plant Molecular Biology* 27:995-1001.

Pearson, W.R. and Lipman, D.J. (1988). Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences USA* 85:2444-2448.

Peng-Wen Chen; Sue-Hui Chow and Liang-Jwu Chen. (1997). Nucleotide sequence of a cDNA (Accession No. U76004) encoding rice Bowman-Birk proteinase inhibitor (PGR97-015). *Plant Physiology* 113:663.

Rohrmeier, T. and Lehle, L. (1993). WIP1, a wound-inducible gene from maize with homology to Bowman-Birk proteinase inhibitors. *Plant Molecular Biology* 22:783-92.

Ryan, C.A. (1990). Protease inhibitors in plants: Genes for improving defenses against insects and pathogens. Annual Review of Phytopathology 28:425-449.

Sambrook, J.; Fritsch, E.F. and Maniatis, T. (1989). Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory Press, New York, USA. 8.30-10.70 pp.

Sanger, F.; Nicklen, S. and Coulson, A.R. (1977). DNA sequencing with chain terminating inhibitors. Proceedings

of the National Academy of Sciences USA 74:5463-5467.

Short, J.M.; Fernández, J.M.; Sorge, J.A. and Hose, W.D. (1988).  $\lambda$ -ZAP: A bacteriophage  $\lambda$  expression vector with *in vivo* excision properties. Nucleic Acids Research 16:7583-7600.

Shulke, R.H. and Murdock, L.L. (1983). Lipoxygenase trypsin inhibitor and lectin from soybeans: Effects on larval growth of *Manduca sexta* (Lepidoptera: Sphingidae). Environmental Entomology 12:787-791.

**FIGURES**

TGT TAC TAC ATG TTC ATC TCC TAT TCA CCA GTG TTT ACA CGA ACA AAT CTT CTT	54
TTA ATT GAA TAT ACG AAA AGT ATT CAT CTA ATT AGC AAA CAA ATT GAT TGC ATA	108
CCA TTC GAA GAA GAC TAT ATT CGG CCT AGG CAA AAT CCA CAA TCA ATT TGC TAT	162
TGT TTC ATA TAA TAA GGC CGT TTG TCT TGT ATT GTC TTC GTT ACA AAA ATC TTT	216
CAT ATC AAT TAT CAT TAA TGA TTG TGA CAT TTA TAC ATT <u>TAA TAT ATA TAT</u> ACA	270
ACT TAT CAT TAT AAT TGT ACG ACG TAT CTC ATA CAT CCC GAA CTC AGA CAT AGT	324
ACT GTT ATT <u>TCA</u> TTT TGT AAT AAG CTT ATT ATG TAC ATA AAG AGC AGT GAG ACT	378
K P I T S E N F N R S L L I Y K T S	
AAG CCA ATA ACA TCA GAA AAT TTT AAC AGA AGC CTC TTG ATT TAC AAG ACT AGC	432
Q G T L M Y S L T L Q K Q K T L K L	
CAA GGT ACA TTG ATG TAT TCA CTT ACA CTT CAA AAA CAA AAA ACT CTC AAG CTT	486
E N K M M V L K V C G V T V L V F F	
GAA AAC AAG ATG ATG GTG CTA AAG GTG TGT GGT GTT ACT GTG CTG GTA TTT TTC	540
L V G V T T A A M D C L N H L G T N	
CTT GTA GGG GTT ACT ACT GCA GCA ATG GAT TGC TTG AAC CAC CTC GGA ACT AAT	594
H H D S D S S D E P S E S S E P C C	
CAT CAT GAT TCA GAC TCA AGC GAT GAA CCT TCT GAG TCT TCA GAA CCA TGC TGC	648
D S C I C S K S I P P Q C H H T D I	
GAT TCA TGC ATC TGC TCA AAA TCA ATA CCT CCT CAA TGC CAT CAT ACA GAT ATC	702
R L N S C H S A C K S C M C T R S M	
AGG TTG AAC TCG TGT CAC TCG GCT TGC AAA TCC TGC ATG TGT ACA CGA TCA ATG	756
P G K L G A G Y F C R C L D I A D V	
CCA GGC AAG CTT GGG GCC GGT TAC TTT TGT CGT TGC CTT GAC ATT GCA GAT GTA	810
F C Y K P C K S R D E D D D E *	
TTT TGT TAC AAA CCT TGC AAG TCC AGG GAT GAA GAT GAT GAT GAG TAA GAA AAG	864
TAA AGA TGA AAC TGT CTC TAA GAT GAA GAA AAT AAG ACC TAA TAG AAG TAT GTT	918
TTA TTT TGT TGT GTA AGG AAA GAA CAG AAA TAA AAA GTA GGT AAT AAA AAG CTA	972
CGT GCT GTG CTG TTC ATG TAA TCT AAG CTA ATT TA	1007

Figure 1. Complete nucleotide and predicted amino acid sequence of cowpea trypsin inhibitor (CpTi) gene.

Nucleotide sequence of a genomic clone encoding a cowpea (*Vigna unguiculata* L.) trypsin inhibitor

```

CpTi cttacacttcaaaaacaaaaaactctcaagcttgaaaacaagatgatggtgctaaaggtg 513
      ||||| |||||||||||||||||||| |||||||||||||||||||||||||
f IV cttacactgcaaaaacaaaaaactctcaagtttgaaaacaagatgatggtgctaaaggtg 170

CpTi tgtgggtgtactgtgctggatTTTTccttgtaggggttactactgcagcaatggattgc 573
      ||||  ||||||| ||||||||||||||||||||||||||||||||||||| |||||
f IV tgtg-----tgctggacttttcttgtaggggttactactgcagccatggatt-- 219

CpTi ttgaaccacctcggaactaatcatcatgattcagactcaagcgatgaaccttctgagtct 633
      ||||| |||||||||||||||||||| |||||||||||||||||||||||||
f IV -tgaaccacctcggaagtaatcatcatgat--gactcaagcgatgaaccttctgagtct 275

CpTi tcagaaccatgctgcgattcatgcatctgctcaaaatcaatacctcctcaatgccatcat 693
      |||||||||||||||||||||||||||| | |||||||||||||||||||||||| |
f IV tcagaaccatgctgcgattcatgcatctgactaaatcaatacctcctcaatgccattgt 335

CpTi acagatatcaggttgaactcgtgtcactcggcttgcaaactcctgcatgtgtacacgatca 753
      ||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
f IV acagatatcaggttgaattcgtgtcactcggcttgcaaactcctgcatgtgtacacgatca 395

CpTi atgccaggcaag 765
      ||||||||||||
f IV atgccaggcaag 407

CpTi tgtcgttgcttgacattgcagatgtattttgttacaaaccttgcaagtcaggatgaa 843
      ||||||||||||||||||| ||| || |||||||||||||||||||||||||||||
f IV tgtcgttgcttgacattgctgat---ttctgttacaaaccttgcaagtcaggatgaa 464

CpTi gatgatga 851
      |||||||||
f IV gatgatga 472
  
```

Figure 2. Comparison of sequence homology of CpTi gene with f IV trypsin inhibitor.

	10	20	30	40	50
CpTi	MYIKSSETKPIITSENFNRSLLIYKTSQGTL	MY- SL	TLQKQKTLKLENKMMVLKVC	GV	TVL
	::: ::	: : .	::: .	: : ::: ::: ::: ::: ::: ::: ::: ::	::
f IV	MYIKSSETKPIQSENKN-----	SGTLTF	IHLTLQKQKTLKFENKMMVLKVC	---VL	
	10	20	30	40	
	60	70	80	90	100
CpTi	VFFLVGVTTAAMDCLNHLGTHHSDS	SDEP	SESESEPCCDSCIC	SKS	IPPQCHHTDIRLN
	::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::				
f IV	VLFLVGVTTAAMD-LNHLGTHHSDS	SDEP	SESESEPCCDSCICTKS	IPPQCHCTDIRLN	
	50	60	70	80	90
	120	130	140	150	160
CpTi	SCHSACKSCMCTRSMPGKLGAGYFCRCLDIADVFCYK	PCKSR	DEDDDE		
	: :				
f IV	SCHSACKSCMCTRSMPGK-----	CRCLDIAD-FCYK	PCKSRDEDDDE		
	110	120	130	140	

Figure 3a. Comparison of sequence homology of CpTi gene with f IV trypsin inhibitor.

```

          60          70          80          90          100          110
CpTi    VCGVTVLVFFLVGVTTAAMDCLNHLGTHHSDSSDEPSESSEPCCDSCICSKSIPPQCH
          :...: . .: :...: :...:
D-II    SDQSSSYDDDEYSKPCCDLCMCTRSMPQCS
          10          20          30

          120          130          140          150          160
CpTi    HTDIRLNSCHSACKSCMCTRSMPGKLGAGYFCRCLDIADVFCYKPKSRDEDDDE
          :...: :...: :...: :...: :...:
D-II    CEDIRLNSCHSDCKSCMCTRSQPGQ-----CRCLDTND-FCYKPKSRDD
          40          50          60          70
    
```

**Figure 3b. Comparison of amino acid sequence of CpTi gene with soybean D-II inhibitor.**

```

          60          70          80          90          100          110
CpTi    CGVTVLVFFLVGVTTAAMDCLNHLGTHHSDSSDEPSESSEPCCDSCICSKSIPPQCHH
          :: . . :...: :...: :...:
C-II    SDHSSSDDESSKPCCDLCMCTASMPQCHC
          10          20          30

          120          130          140          150          160
CpTi    TDIRLNSCHSACKSCMCTRSMPGKLGAGYFCRCLDIADVFCYKPKSRDEDDDE
          .: : : : : : : : : : : : : : : : : : : : : : : : : : : :
C-II    ADIRLNSCHSACDRCACTRSMPGQ-----CRCLDTTD-FCYKPKSESDDD
          40          50          60          70
    
```

**Figure 3c. Comparison of amino acid sequence of CpTi gene with soybean C-II inhibitor.**