Plant Physiol. (1993) 101: 705–706

Plant Gene Register

Nucleotide Sequence of a cDNA Clone Encoding Tomato (*Lycopersicon esculentum*) Cysteine Proteinase

Huub J. M. Linthorst*, Chris van der Does, Jan A. L. van Kan, and John F. Bol

Institute of Molecular Plant Sciences, Department of Biochemistry, Leiden University, P.O. Box 9502, 2300 RA Leiden, The Netherlands (H.J.M.L., C.v.d.D., J.F.B.); and Department of Phytopathology, Wageningen Agricultural University, P.O. Box 8025, 6700 EE Wageningen, The Netherlands (J.A.L.v.K.)

When a tomato (*Lycopersicon esculentum*) leaf cDNA library was screened with a cDNA fragment from a polymerase chain reaction prepared using a degenerate primer mixture that was based on the amino acid sequence of tomato systemin (Pearce et al., 1991), a clone (LCYP-2) was obtained of 1295 bp, not including the 3′-terminal poly(A$^+$) tail (Table I). It is surprising that when sequenced the clone appeared not to correspond to systemin but to encode a protein with amino acid sequence similarity to Cys proteinases of plants and animals (H.J.M. Linthorst, C. van der Does, F.Th. Brederode, and J.F. Bol, unpublished data). The clone contains one large reading frame that is already open at the 5′ end of the coding strand but lacks a Met initiation codon. The protein encoded by clone LCYP-2 is 34% identical with papain from *Carica papaya* (Cohen et al., 1986), 38% identical with human cathepsin H (Ritonja et al., 1988; Fuchs and Gassen, 1989), and 85% identical with tobacco (*Nicotiana tabacum*) leaf Cys proteinases (H.J.M. Linthorst, C. van der Does, F.Th. Brederode, and J.F. Bol, unpublished data). The high similarity to the tobacco proteinase clones suggests that probably only the first two codons are lacking.

Cys proteinases like the mammalian cathepsins are typical housekeeping enzymes, present in lysosomes and involved in degradation of intracellular proteins. The similarity to other animal and plant Cys proteinases implies that the protein encoded by LCYP-2 is a pre-pro-protein with domains possibly involved in targeting to the vacuole (Holwerda et al., 1992). The similarity predicts that the hydrophobic N-terminal part is a signal peptide for translocation through the ER. This signal sequence is possibly cleaved off between Ala^{16} and Ile^{17}. A subsequent removal of the next N-terminal domain by cleavage between Asp^{127} and Leu^{128} would result in the mature protein, having a molecular mass of 25 kD. The active site residues (in clone LCYP-2 they are Cys^{152} and His^{294}), as well as the Cys residues involved in disulfide bridge formation in papain and other Cys proteinases, are conserved in the tomato protein. We have recently shown that the genes for the tobacco Cys proteinases are expressed in a circadian-regulated manner and are induced by incision wounding (H.J.M. Linthorst, C. van der Does, F.Th. Brederode, and J.F. Bol, unpublished data). The putative proteinase encoded by clone LCYP-2 is different from an earlier identified Cys proteinase that is induced in tomato fruit upon cold treatment (39% identity; Schaffer and Fischer, 1988).

Table I. *Characteristics of Cys proteinase cDNA from tomato*

Organism:
Lycopersicon esculentum.
Gene, Gene Product
Putative vacuolar cysteine proteinase involved in degradation of cellular proteins (EC 3.4.22).
Techniques:
A cDNA library in λZAP II (Stratagene) prepared from leaf mRNA was screened with a radiolabeled probe that resulted from a polymerase chain reaction on the library cDNA using the T3 primer and a degenerate primer mixture (5′-GCATC/TTTIGGIGGA/GTCC/TCTC/TTTA/GC/GA/TIGGIGG-3′; I, inosine). Other techniques included dideoxy sequencing on double- and single-stranded DNA and computer analysis of sequence data.
Method of Identification:
Sequence comparison revealed extended similarity to a large number of Cys proteinases such as mammalian cathepsins, papain, actinidin, etc.
Features of mRNA structure:
Poly(A$^+$) tail is present in the cDNA clone but the 5′-terminal untranslated region and adjacent coding region is not present. However, based on the high similarity to the corresponding tobacco cDNA clones, it is most probable that only the first two codons are lacking.
Codon Usage:
All codons are used. There is a slight preference for T in the wobble position.
Structural Features of Protein:
The coding region encodes 361 amino acids (probably only the N-terminal two residues are lacking). Based on similarity to other Cys proteinases, the protein is probably processed to remove an N-terminal signal peptide and subsequent vacuolar targeting domain. The mature Cys proteinase would be 234 amino acids and have a calculated M_r of 25,483.
Antibodies:
Not available.

ACKNOWLEDGMENTS

We would like to thank Jan van der Knaap, Thijs Notenboom, and Marcel Flikweert for excellent technical assistance.

* Corresponding author; fax 31-71-274537.

Received September 4, 1992; accepted October 2, 1992.
Copyright Clearance Center: 0032-0889/93/101/0705/02.
The EMBL accession number for the sequence reported in this article is Z14028.

LITERATURE CITED

Cohen LW, Coghlan VM, Dihel LC (1986) Sequencing of papain-encoding cDNA. Gene **48:** 219–227

Fuchs R, Gassen HG (1989) Nucleotide sequence of preprocathepsin H, a lysosomal cysteine proteinase. Nucleic Acids Res **17:** 9471

Holwerda BC, Padgett HS, Rogers JC (1992) Proaleurain vacuolar targeting is mediated by short contiguous peptide interactions. Plant Cell **4:** 307–318

Pearce G, Strydom D, Johnson S, Ryan CA (1991) A polypeptide from tomato leaves induces wound-inducible proteinase inhibitor proteins. Science **253:** 895–898

Ritonja A, Popovic T, Kotnik M, Machleidt W, Turk V (1988) Amino acid sequences of the human kidney cathepsins H and L. FEBS Lett **228:** 341–345

Schaffer MA, Fischer RL (1988) Analysis of mRNAs that accumulate in response to low temperature identifies a thiol protease gene in tomato. Plant Physiol **87:** 431–436