Primary structure of a proteinase inhibitor II gene from potato (*Solanum tuberosum*)

Michael Keil, Jose Sanchez-Serrano, Jeff Schell and Lothar Willmitzer

Max-Planck-Institut fuer Zuechtungsforschung, 5000 Koeln 30, FRG

Received 21 May 1986; Revised and Accepted 1 July 1986

ABSTRACT

The isolation and characterization of a genomic clone encoding proteinase inhibitor II of potato (Solanum tuberosum) is described. The structure of this gene was determined by sequencing a genomic fragment of about 2 kb containing the entire RNA coding as well as about 900 nucleotides of the 5′-upstream and 250 nucleotides of the 3′-downstream region. The transcription start site was determined by RNase protection experiments.

The comparison of the genomic sequence with cDNA sequences reveals the presence of one intron with a length of 117 nucleotides. The genomic clone contains an open reading frame of 462 nucleotides allowing for a protein of 154 amino acids.

The proteinase inhibitor II gene displays typical features of eucaryotic genes. The sequence TATAAA is found 26 nucleotides upstream of the transcription initiation site and the sequence CAAAT at position - 103. In the 3′-region the sequence AATAA is found 33 nucleotides in front of the poly-A addition site.

INTRODUCTION

Proteinaceous inhibitors of proteases represent ubiquitous components found in a number of tissues of plants, animals and microorganisms. Their inhibitory function is in most cases directed against a single class of proteases only. In plants many proteinase inhibitors are thought to serve as a defense reaction of the plant against insect attack. This assumption is supported by the observation that proteinase inhibitors of plant storage organs such as seeds or tubers do inhibit proteolytic activity in a wide range of microorganisms and insects but only rarely inhibit proteinases of plant origin (1).

It has been shown in the case of proteinase inhibitor I and II of tomato (2) as well as for proteinase inhibitor II of potato (3) that RNA homologous to the corresponding genes

© IRL Press Limited, Oxford, England.

accumulates in leaves after mechanical injury. The actual inducing factor was shown to be an oligosaccharide fraction (4). In addition to this environmental control proteinase inhibitor II of potato is also under developmental control; in nonwounded potato plants proteinase inhibitor II derived RNA can only be detected in potato tubers (3,5). Thus the expression of this gene is under a complex control involving both developmental and environmental factors, making this gene an interesting subject for the study of differential gene expression in higher plants.

Here we describe the primary structure of a proteinase inhibitor II gene from *Solanum tuberosum* isolated from a haploid potato plant.

MATERIALS AND METHODS

Plant Materials

The haploid potato genotype AM 80/5793 originating from the collection of the Max-Planck-Institut fuer Zuechtungsforschung Koeln was used as a source for nuclear DNA.

Isolation And Analysis of Nucleic Acids

Isolation and analysis of nuclear DNA as well as of total and poly-A^+ RNA was performed as described previously (6).

Isolation of Proteinase Inhibitor II Encoding Genomic Clones

A genomic library of the haploid potato line AM 80/5793 established in the lambda replacement vector EMBL 4 (7) was screened by plaque hybridization using the radiqactively labelled cDNA insert pcT 800 (5). Positive plaques were purified using standard conditions (7).

DNA Sequence Analysis

DNA sequence analysis was performed using the dideoxy method (8). To this end overlapping deletions obtained by exonuclease III treatment of asymmetrically cut M13mp19 and M13mp18 clones containing parts of the gene were constructed and analyzed (9).

Mapping of the Transcription Initiation Site by RNA Protection Experiments

For RNA protection experiments, radioactively labelled RNA homologous to the fragment (cf. fig. 3) starting in the intron and extending up to the HindIII site 5′-upstream of the gene was

synthezised using appropriate SP6 vectors and SP6 RNA polymerase (10). Hybridization to total RNA (50 μg) isolated from either unwounded leaves or tubers was performed for 16 hours at either 37°C or 50°C in 30 μl of a buffer containing 80% formamide, 40mM PIPES pH 6,7, 0.4 M NaCl, 1mM EDTA. After hybridization 300 μl of a RNAse digestion buffer containing 10 mM Tris pH 7.5, 5 mM EDTA, 300 mM NaCl, RNAse A (20 μg/ml) and RNAse T1 (1 μg/ml) was added and after incubation for 30 minutes at 30°C, nucleic acids were precipitated and separated on a 6% polyacrylamide sequencing gel.

RESULTS

Isolation and Characterization of a Genomic Clone

A genomic library in the lambda phage replacement vector EMBL 4 (11) of the nuclear genome of the haploid potato line AM 80/5793 (7) was screened by plaque hybridization using the cDNA clone pcT 800 (3,5) as a probe. Four independent recombinant clones were obtained by screening $8x10^5$ plaques.

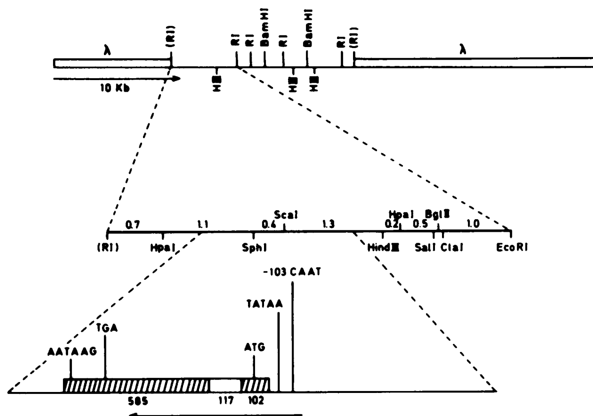

Figure 1
Upper part: Restriction map of the genomic lambda clone J32. Cleavage sites for restriction enzymes are indicated.
Middle part: Restriction map of the Eco RI fragment hybridizing to the proteinase inhibitor encoding cDNA clone pcT 800.
Lower part: Structure of the proteinase inhibitor II gene. The 2.0 kb region sequenced is shown in an expanded scale. Black boxes indicate the exons, the white box indicates the intron. The direction of transcription is indicated by the arrow.

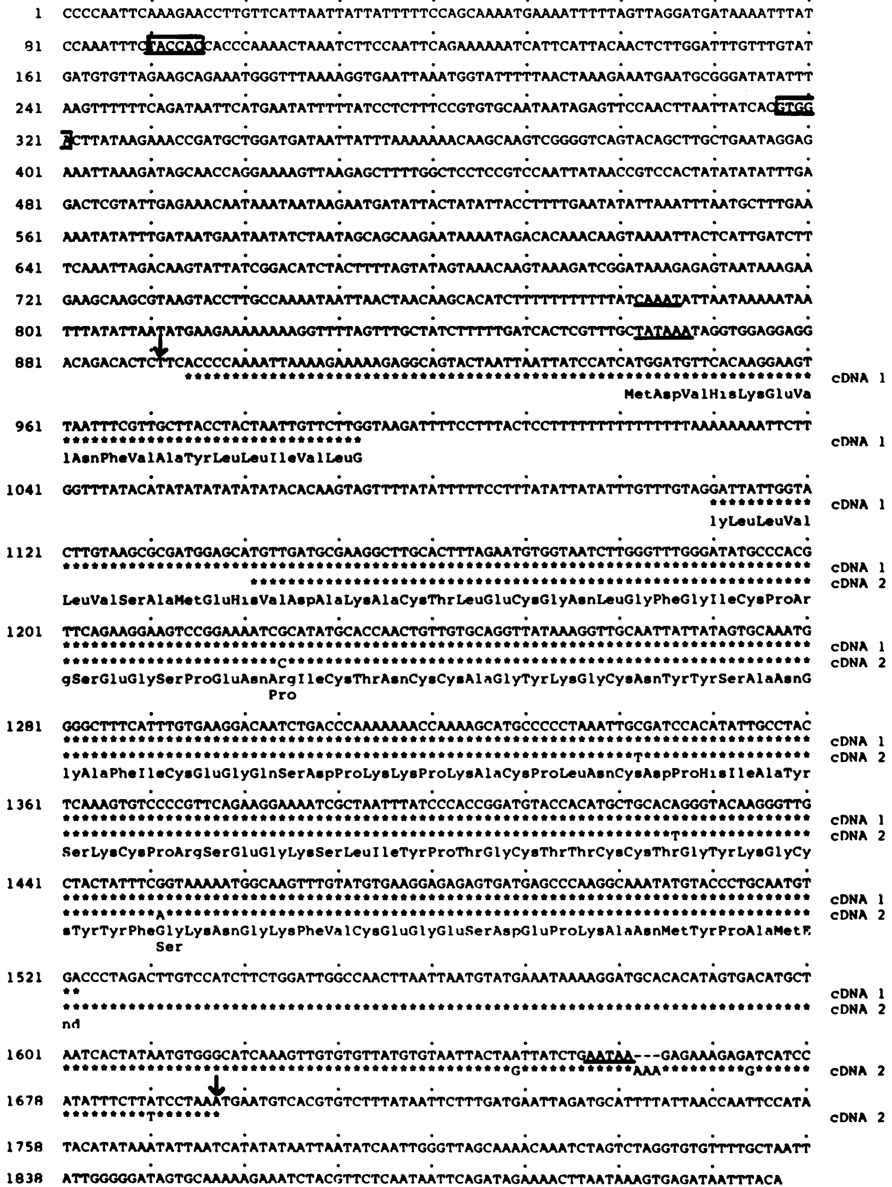

Figure 2. Nucleotide sequence and deduced amino acid sequence of the proteinase inhibitor II gene and 5′- and 3′ flanking regions. The nucleotide sequence of the genomic clone is compared to that of previously published cDNA′s (3). Homologies are indicated by a star.

Consensus sequences (CAAT box, TATA box and polyadenylation signals) are underlined. The transcription start site as

determined by RNA protection experiments as well as the presumptive poly A addition site as deduced from a comparison with the cDNA are indicated by vertical arrows. Motives in the 5′-upstream region homologous to the enhancer core structure are boxed.

Analysis of these clones by physical mapping of restriction endonuclease sites and by molecular hybridization experiments showed that they were derived from two different genomic locations. One of these clones, no.J32 was chosen for further characterization because it hybridized most strongly to the cDNA clone pcT 800 as well as to cDNA clones 1 and 2, encoding proteinase inhibitor II (3), under high-stringency washing conditions (data not shown).

A map for some restriction enzymes of the genomic lambda clone J32 is shown in figure 1. Using pcT 800 as a probe the hybridization was confined mainly to the 5.2 kb long Eco RI fragment located at the extreme left of the insert, which was therefore subcloned for further characterization. A more refined restriction map of this fragment is shown in the middle part of figure 1.

Structure of the Proteinase Inhibitor II Gene

The primary structure of the proteinase inhibitor II gene and of its flanking regions was determined by sequencing a series of progressively shortened clones obtained by treating the fragments EcoRI/SphI (1.8kb) and SphI/HindIII (1.7 kb) cloned in M13 mp18 and M13 mp 19 with exonuclease III and S1 nuclease and subsequent religation and transformation of the deleted clones. The Sph I site was taken as a start point for sequencing as this site is conserved in both the cDNA and the genomic clone. A total of 2 kb to the left and to the right of this site were sequenced on both strands. The result is shown in figure 2.

When the genomic sequence is compared to the previously determined sequence of two proteinase inhibitor II cDNA′s derived from tuber-A^+-RNA (cDNA′s 1 and 2) (3), the presence of one intron is apparent (cf. fig. 1 for a more schematic view of the gene). The size of the intron (117 nucleotides) is rather small. The exon-intron-junction obeys the GT-AG rule (12). The intron is very AT-rich (83%); in comparison with

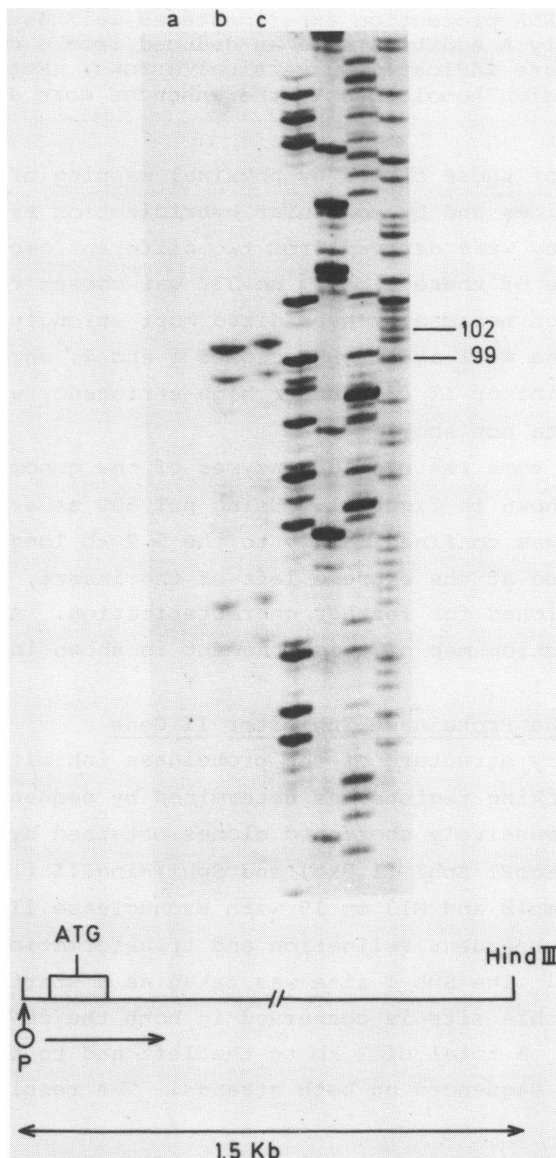

Figure 3. Determination of the transcription start site by RNA protection experiments.

The lower part of the figure shows the region of a deletion containing part of the intron, the first exon as well as the complete 5′-upstream region up to the Hind III site.Radioactively labelled RNA homologous to the noncoding strand was synthezised from this fragment using appropriate SP6 expression vectors and SP6 RNA polymerase. P indicates the

position of the promoter for the SP6 RNA polymerase, the direction of transcription of the SP6 RNA polymerase is indicated by the arrow. After hybridization to potato RNA the hybrids were treated with RNAse A and RNAse T1 and separated on 6% polyacrylamide gels (cf. Materials and Methods). The following RNA's were offered for hybridization (upper part): lane a: total RNA from unwounded potato leaf; hybridization temperature: 37°C lane b: total RNA from potato tuber; hybridization temperature: 37°C lane c: total RNA from potato tuber; hybridization temperature: 50°C The right part of the autoradiogram shows part of a sequencing gel used for determining the length of the protected RNA, the size of the two longest fragments protected is given in nucleotides.

introns in most other plant genes which have an AT-content of 60-75% (13).

The genomic sequence contains an open reading frame of 462 nucleotides which is the same length as the cDNA's (3). There is 100% homology with the coding region as well as with the 5'-untranslated leader of one of the cDNA's (cDNA 1) (3) and the genomic clone described (cf. fig. 2). However, there are several differences between the other cDNA (cDNA 2) (3) and the genomic clone. A deletion of three base pairs in the genomic clone as well as seven single nucleotide exchanges were detected.

<u>5'-Region of the Proteinase Inhibitor II Gene</u>

RNA protection experiments were performed using the SP6 system (10) in order to determine the transcription start site. The results are shown in figure 3. A deletion of the genomic clone still containing part of the intron as well as the first exon and the complete 5'-upstream region up to the HindIII site was cloned into an SP6 expression vector and a radioactive RNA synthezised from this template using SP6-RNA polymerase was offered for hybridization against total RNA from tuber (lane b,c) and unwounded leaves (lane a) of potato. Protected RNA is only visible in those cases where tuber RNA was offered for hybridization; no signal is visible in case of total leaf RNA. Northern analysis had previously demonstrated the absence of any RNA homologous to the proteinase inhibitor II gene in unwounded leaves of potato (3,5). The SP6 protection experiments shown in figure 3 therefore support this earlier finding and in addition demonstrate the specificity of the assay. The longest protected

RNA found in figure 3 has a size of 102 nucleotides which therefore identifies the transcription initiation start site at a distance of 50 nucleotides in front of the ATG serving as the translation initiation site.

The nucleotide sequence around this ATG, CATC ATG GAATG agrees well with the consensus sequence proposed for translation initiation sites in other eucaryotic genes (14).

3′-End of the Proteinase Inhibitor II Gene

At the 3′-end the sequence AATAA is present 33 nucleotides in front of the presumptive poly A addition site determined by a comparison of the genomic clone with cDNA 2 encoding proteinase inhibitor II (3). This sequence is in good agreement with the consensus sequence AATAAA serving as a polyadenylation signal in most eucaryotic genes (15) and would give the proteinase inhibitor II mRNA a 172 nucleotide long untranslated 3′ end which is in good agreement with the length of the 3′-untranslated regions of other plant genes (16). In addition to this sequence at nucleotide +767, another AATAAA sequence is found at position +681. Whether this sequence is of any functional significance is unknown.

DISCUSSION

The isolation and physical characterization of a genomic clone encoding proteinase inhibitor II of potato is described.

With regard to its primary structure the proteinase inhibitor II gene shows several characteristics which are common to other eucaryotic genes.

In the 5′-upstream region of the gene the motif TATAAA homologous to the TATA box is found 26 nucleotides in front of the transcription initiation site. No typical CAAT box is found at position -70 to -80, however the sequence CAAAT is present 103 nucleotides in front of the transcription start site.

In the 3′-untranslated region several sequences homologous to the poly-adenylation signal consensus sequence are present, which is also known from other plant genes (17,18). It is very likely that in case of the proteinase inhibitor II gene the last polyadenylation signal is actually the one used as deduced from a comparison of the cDNA and the genomic clone (cf. fig. 2).

The proteinase inhibitor is a split gene, the RNA coding part being interrupted by one intron of 117 nucleotides length. This intron is located in the N-terminal part of an open reading frame which probably codes for a signal peptide used for a co-translational import into the endoplasmic reticulum from where the proteinase inhibitor is finally distributed into the vacuole (3).

The expression of the proteinase inhibitor II in potato is under both developmental (i.e. tuber-specific) as well as under environmental (i.e. wound-induced) control (cf. introduction). Since proteinase inhibitor II genes form a small multigene family (Ref. 5 and unpublished results) their pattern of expression can reflect the function of different proteinase inhibitor II genes as a response to different controls. Alternatively the same gene(s) expressed constitutively in tubers could also be induced in leaves upon wounding by the effect of trans-acting factors. We have previously described the isolation and physical characterization of another gene of potato encoding the major storage protein of potato tuber, patatin (7). The expression of patatin is under strict developmental control being tuber-specific. Inspection of the 5′-upstream region of both the patatin and the proteinase inhibitor II genes however does not reveal any homology. This result may be understandable due to the fact that very likely the tuber-specific expression of the patatin gene and the proteinase inhibitor II gene is under different control. Thus run-off transcription experiments performed in isolated nuclei indicate that the tuber-specific expression of patatin is mainly controlled at the level of transcription whereas in the case of the proteinase inhibitor its tuber-specific expression appears to be controlled also at a post-transcriptional level (5).
To the best of our knowledge this contribution describes the first isolation and physical characterization of a proteinase inhibitor gene from plants originating from a non-seed organ. Due to its complex type of control of expression (both developmentally and by environmental factors), the proteinase inhibitor II genes offer an exciting system for the study of regulation of gene expression in higher plants.

ACKNOWLEDGEMENTS

We thank Wolfgang Schmalenbach for help during the sequencing work, Jutta Freitag for editing the manuscript and Maret Kalda for photographic work.

Jose Sanchez-Serrano is a recipient of a fellowship from the Fundacion Juan March (Spain).

This work was supported by a grant from the Bundesministerium für Forschung und Technologie (BCT 0365/2; "Molekulare Gen- und Zelltechnologie").

REFERENCES

1. Richardson, M. (1977) Phytochemistry 16, 159-169.
2. Ryan, C.A., Bishop, P.D., Walker-Simmons, M., Brown, W.E., Graham, J.S. (1985) In: Key, J., Kosuge, T. (eds) UCLA Symposia on Molecular and Cellular Biology, New Series, Vol.22, pp 319-334.
3. Sanchez-Serrano, J., Schmidt, R., Schell, J., Willmitzer, L. (1986) Mol. Gen. Genetics 203, 15-20.
4. Bishop, P.D., Makus, D.J., Pearce, G., Ryan, C.A., (1981) Proc. Natl. Acad. Sci. USA 78, 3526-3540.
5. Rosahl, S., Eckes, P., Schell, J., Willmitzer, L. (1986) Mol. Gen. Genetics 202, 368-373.
6. Eckes, P., Schell, J., Willmitzer, L. (1985) Mol. Gen. Genetics 199, 216-224.
7. Rosahl, S., Schmidt, R., Schell, J., Willmitzer, L. (1986) Mol. Gen. Genetics 203, 214-220.
8. Sanger, F., Nicklen, S., Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
9. Yanisch-Perron, A., Vieira, J., Messing, J. (1985) Gene 33, 103-119.
10. Melton, D.A.S., Krieg, P.A., Rebagliatti, M.R., Maniatis, T., Zinn, K., Green, M.R. (1984) Nucleic Acids Res. 12, 7035-7056.
11. Frischauf, A.M., Lehrach, H., Poustka, A., Murray, N. (1983) J. Mol. Biol. 170, 827-842.
12. Breathnach, R., Chambon, P. (1981) Ann . Rev. Biochemistry 50, 344-383.
13. Slightom, J.L., Sun, S.M., Hall, T.C. (1983) Proc. Natl. Acad. Sci. USA 80, 1897-1901.
14. Kozak, M. (1981) Nucleic Acids Res. 9, 5233-5252.
15. Proudfoot, M.J., Brownlee, G.G. (1976) Nature 263, 211-214.
16. Messing, J., Geraghty, D., Heidecker, G., Hu, N.T., Kridl, J., Rubenstein, J. (1983) In: Kosuge, T., Meredith, C.P., Hollaender, A. (eds) Genetic engineering of plants - an agricultural perspective. Plenum Press New York, pp. 211-227.
17. Lycett, G.W., Delauney, A.J., Croy, R.R.D. (1983) FEBS Letters 153, 43-46.
18. Dean, C., Tamaki, S., Dunsmuir, P., Farreau, M., Katayama, C., Dooner, H., Bedbrook, J. (1986) Nucleic Acids Res. 14, 2229-2240.