The gene coding for the mustard trypsin inhibitor-2 is discontinuous and wound-inducible

L.R. Ceci[a,*], N. Spoto[a], M. de Virgilio[b], R. Gallerani[b]

[a]*Centro di Studio sui Mitocondri e Metabolismo Energetico, CNR, Sezione di Trani, Via Corato 17, 70059 Trani, Italy*
[b]*Dipartimento di Biochimica e Biologia Molecolare, Università di Bari, Via Amendola 165/A, 70126 Bari, Italy*

Received 20 March 1995

Abstract The gene coding for the mustard trypsin inhibitor-2 has been isolated from a genomic library and characterized. Comparison of genomic and cDNA sequences indicates that the gene is interrupted by an intron of 193 bp. The eukaryotic peculiar regulatory sequences have been detected in the 5′ flanking region of the gene. In addition, a decanucleotide has been detected that is highly similar to the proposed G-box and to the ABRE motifs required for the gene expression induced by methyl jasmonate and abscissic acid. Northern blot analysis demonstrates that the gene is expressed in immature seeds as well as in wounded leaves.

Key words: Serine proteinase inhibitor gene; Wound-inducible gene; White mustard; *Sinapis alba* var. Albatros

1. Introduction

In the plant kingdom, proteinase inhibitors are usually present in seeds and tubers, particularly in the families *Graminacae, Leguminosae* and *Solanaceae* [1,2].

The most studied inhibitors are those active toward serine proteinases. They are usually classified by using several chemical and biochemical parameters including primary structure, molecular weight, cysteine content and heat stability [3]. Another criterion of classification concerns the role of such inhibitors in plant defense against insect attack [3].

A trypsin inhibitor of low molecular weight (7,000 Da) has been isolated and characterized from white mustard seeds (*Sinapis alba*, cv. Albatros) [4]. This protein (mustard trypsin inhibitor-2; MTI-2), the first to have been sequenced in *Cruciferae*, shows a primary structure different from those of other serine proteinase inhibitors studied so far. A further specific feature of this protein concerns the positions of the eight cysteine residues which are more similar to those of specific toxins (such as the erabutoxin [5]) than to those of Bowman-Birk proteinase inhibitors. A similar inhibitor, specific for trypsin and chymotrypsin has recently been identified and purified from another *Crucifera* (rapeseed) [6]. It shows almost identical properties with respect to the MTI-2 protein, including a high aminoacid sequence similarity of about 70%.

The common and distinctive features of these proteins strongly suggest that *Cruciferae* contain protein inhibitors which can be classified within a new family.

This paper describes the identification and characterization from a white mustard genomic library of the discontinuous gene coding for the MTI-2 protein (*mti*-2 gene).

*Corresponding author. Fax: (39) (80) 544-3317.

The gene is expressed toward the end of seed maturation and in leaves under mechanical wounding.

2. Materials and methods

2.1. Construction and screening of a mustard genomic library

Mustard genomic DNA was isolated from 5-day-old hypocotyls as described in [7].

A mustard genomic library was constructed in the phage vector Charon40 [8]. The screening of the library was carried out by standard procedures [9].

2.2. PCR amplification of genomic DNA

The DNA was heat-denatured in the PCR reaction mixture at 92°C for 3 min. The reaction mixture contained 500 ng of mustard genomic DNA, 10 μl of 10 × PCR buffer (500 mM KCl, 100 mM Tris-HCl, pH 8.3, 0.01% gelatin), 6 μl of 25 mM $MgCl_2$, 8 μl of dNTP mix (25 mM of each dNTP), 400 pmol of suitable primers (see section 3) and 2.5 units of Taq DNA polymerase (Perkin-Elmer, 8 U/μl). The amplification was run for 30 cycles of 60 s at 94°C, 60 s at 56°C, and 60 s at 72°C.

2.3. Isolation of poly(A)$^+$ mRNA

White mustard plants (*Sinapis alba*, cv. Albatros) were grown under standard greenhouse conditions of light and temperature. Seeds were collected every 4 days starting from the 21st day after pollination (DAP), immediately frozen in liquid nitrogen and stored at −80°C until RNA extraction.

Leaves were wounded twice using a razor blade: 24 and 4 h before harvesting.

RNA was extracted by the guanidine hydrochloride method of Arrand [10] and poly(A)$^+$ mRNA was isolated by affinity chromatography on an oligo(dT)-cellulose column.

2.4. cDNA synthesis and amplification

Single-stranded cDNA was synthesized from poly(A)$^+$ mRNA by the SuperScript II reverse trascriptase enzyme, according to the manufacter's indications (Gibco-BRL).

PCR amplification was run for 30 cycles of 60 s at 94°C, 60 s at 60°C and 60 s at 72°C, in the presence of specific oligonucleotides: *P3* = 5′-CAGTTTCTTCGTTCACCCTC-3′ and *P4* = 5′-CTGTCGTTGCAGTAGTCGCA-3′.

The PCR amplification product was purified from agarose gel and directly sequenced.

2.5. Miscellaneous methods

Southern and Northern analysis, manipulation and sequencing of specific DNA fragments were carried out according to standard procedures [10].

3. Results

3.1. Detection and sequencing of the mti-2 gene

The knowledge of the primary structure of the purified mustard trypsin inhibitor-2 [4] made it possible to choose particular regions of the protein the amino acid residues of which corresponded to codons having the minimum degree of degenerancy. These regions correspond to amino acid 6–11 and 53–59. Two

SSDI 0014-5793(95)00386-X

```
                                                                    -408
GGATCATGCAGTTATTTTTCTACATAAGAGCTTCGTAATAAGTAAAAACAAGCTGTCTGA
                                                                    -348
TCCTGGTTAGGGGGTAGAAAGTAAAGTTTGCACCTGGTGGAGTAGTTCCTGGTTCGATTC
                                                                    -288
TTGATTTAACTCGGGTTTGTTCAATTCGGTTTCCATCTTTAGCATGTATTGGGCTTTTAA
                                                                    -228
ATTAATACTGGGCTGGGCCTAGAGTAATAATTTATTTTAAGAAAACCTTTCATATTACAA
                                                                    -168
ACAAATCAATTTTATCTCGTTTTTCAGGCATATGAAAAACTCAAAAAGTAGTTAAAACTC
                                                                    -108
GATATATATATATATACATGTATTTATATATATATTTAAAACAGATGGCAGCTAAATAAC
                                                                     -48
TTTAATGATTATTTTTTTCTCCAACGACAAAGATAAATGGCACACCCGGGTTCATTTTAC
                                                                1     9
GTATACGTTAACTCCAAAGTTCAGACACATAGAGAGAGAGAGAAG  ATG GCC ATG
                                                               M   G   M
                                                                      54
GCA AAA AAA TCA GTT TCT TCG TTC ACC CTC ATT TTC  ATC CTC GTT
 A   K   K   S   V   S   S   F   T   L   I   F    I   L   V
                 67                                                   109
TTG GTT ATT TTT G gtaaagtaactaactctttattttatttatttttttggtaaaa
 L   V   I   F
                                                                     169
actaactccgtatttttttcacatatttgttaatatttgtgtggttcatagtaggattct
                                                                     229
tctctccgtgactgcattgtagcacgttttgtcatatactatatgatatctgttttctct
                                  261                                280
aatactctctatgtatatgtatattgagcag   AA GTG CCG GAG ATA AAA GCG
                                  E   V   P   E   I   K   A
      *                                                              325
CAG  GAT AGC GAG TGC CTG AAA GAA TAC GGT GGT GAT GTT GGC TTC
 Q    D   S   E   C   L   K   E   Y   G   G   D   V   G   F
                                                                     370
CCT  TTC TGC GCA CCT CGG ATA TTT CCG ACG ATT TGC TAC ACA AGA
 P    F   C   A   P   R   I   F   P   T   I   C   Y   T   R
                                                                     415
TGC  CGT GAG AAC AAG GGG GCT AAA GGT GGA AGA TGC ATT TGG GGA
 C    R   E   N   K   G   A   K   G   G   R   C   I   W   G
                                                                     460
GAA  GGA ACT AAC GTT AAG TGC TTA TGC GAC TAC TGC AAC GAC AGC
 E    G   T   N   V   K   C   L   C   D   Y   C   N   D   S
                 *                          493                     507
CCT  TTT GAT CAG ATT CTA AGA GGT GGC ATT TGA  GTTCCTGCATGTGC
 P    F   D   Q   I   L   R   G   G   I
                                                                     567
CTTGGTTTGTGTAATAATAAATGTTAAAATGGTCCAAGTAATAATGTAATAACGAGTATG
                                                                     627
TGAAATAAAAAAAAAGCTCGGAAGTATCTGAGCCTCACATGGTTTGATAATAGATGTGGTT
                                                                     687
ATTGTGGTCGTGAGTCAAAGAGAGTAGTTATTAACTTTCATACTTCCCAACTTGCTAGAA
                                                                     747
GCCAAAAGATTCGTCCTAGCCCAAAGAAAAAAATAATATCTACACTTGTTTTAAAATCTT
                                772
TTATGTTAAAAAGTAACTCGAATTC
```

Fig. 1. Sequence of the mustard DNA region containing the *mti*-2 gene. The figure shows the sequence of part of the 1.3 kb *Eco*RI fragment obtained after digestion of Charon40–4F4 DNA with *Eco*RI. The codons corresponding to the first and last amino acid of the mature protein are indicated by an asterisk. Putative consensus sequences for the RNA polymerase II are overlined by a dotted line, ACGT and polyadenylation signals are underlined.

families of oligonucleotides: *P1* 5′-AAR GAR TAY GGN GGN GAY GT-3′ and *P2* 5′-NSW RTC RTT RCA RTA RTC RCA-3′ (R = G,A; Y = C,T; S = G,C; W = T,A), corresponding to these regions were synthesized and used in PCR amplification experiments of mustard genomic DNA. The amplification products were fractionated on agarose gel and among them, the one present in the greatest amount and having also the size expected for a continuous gene (not shown) was eluted and cloned in the *Sma*I site of pUC18 vector. Its deduced amino acid sequence was in full agreement with the corresponding region of MTI-2 protein.

The same insert was also used as a probe for the screening of a genomic library constructed in the Charon40 phage vector and for the Northern blot experiments described at the end of this section.

Among several positive clones, a recombinant phage, here named 4F4, was selected. Southern blot experiments made possible the identification of a specific *Eco*RI sub-fragment of about 1.3 kb which was entirely sequenced. Fig. 1 shows the sequence of the mustard DNA region present in the 1.3 kb *Eco*RI fragment. The *Eco*RI site belonging to the Charon40 polylinker was not reported.

The region coding for the MTI-2 protein shows a complex organization which implies the presence of an intron of 193 bp. The tentative gene structure reported in Fig. 1 begins with the first of two alternative AUG codons. The mature protein is entirely encoded in the second exon.

Several signals related to the expression of eukaryotic genes and most probably also to their regulation, have been detected (see section 4).

The hydropathy profile of the whole protein, deduced according to Kyte and Doolittle [12], shows in the correspondence of the presequence the typical features of a signal peptide (not shown). Indeed it posseses a hydrophobic core of 16 amino acids, flanked by charged amino acids. Similar sequences have been detected in plant proteinase inhibitors for which a vacuolar localization has been demonstrated [13].

3.2. cDNA amplification and sequence

In order to confirm the existence of the intron sequence, two oligonucleotides located in the first and second exon of the gene in position 20–39 (P3) and 420–459 (P4) were selected. The 5′ ends of P3 and P4 have a relative distance of 440 nucleotides (see also Fig. 1). They were used in a reverse transcription-PCR amplification assay in which poly(A)$^+$ mRNA purified from immature seeds (at 54 DAP) was used as template. The PCR amplification product, characterized by agarose gel electrophoresis, corresponds to a fragment of about 250 bp (not shown). Its nucleotide sequence confirms the existence of the proposed intron sequence shown in Fig. 1.

3.3. Northern blot analysis

Northern blot experiments were carried out in order to verify in which organ and/or under which physiological conditions the *mti*-2 gene is expressed.

The results reported in Fig. 2 clearly show that the gene is transcribed in the late stage of seed maturation (54–61 DAP) and in leaves, only under mechanical injury. Before 54 and after 61 DAP no hybridization signals were detected.

4. Discussion

A fragment of the *Sinapis alba* genomic DNA containing the gene coding for the MTI-2 proteinase inhibitor has been isolated and characterized.

The sequence of the gene and in particular the region corresponding to the mature protein is different from those of other serine proteinase inhibitor genes sequenced so far. This result suggests, as expected from the primary structure of the protein, that the MTI-2 identifies a new family of proteinase inhibitors.

The gene is discontinuous for the presence of an intron of

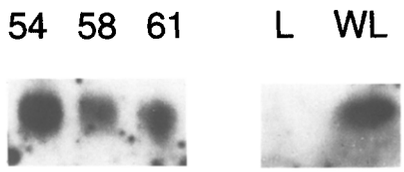

Fig. 2. Expression of *mti*-2 gene. Hybridization signals after a Northern blot analysis of mustard poly(A)$^+$ mRNAs extracted from immature seeds at 54, 58, 61 DAP and from leaves (L) and wounded leaves (WL).

193 bp. The existence of the intervening sequence is confirmed by the sequence of a cDNA fragment obtained by a reverse transcription-PCR amplification assay, carried out on poly(A)$^+$ mRNA isolated from immature seeds collected at 54 DAP (see section 3). Although most of the plant serine proteinase inhibitor genes are continuous, examples of discontinuity can be found in several wound inducible proteinase inhibitor genes [14–17]. Interestingly, in all the cases an intron is present in the region of the gene coding for the protein presequence.

In mustard the presequence contains a cluster of hydrophobic residues (11–25) flanked by charged amino acids, as also reported for other plant proteinase inhibitors for which the compartimentation into the vacuole has been demonstrated [13].

Among the presequences detected in the other inhibitors, that of the MTI-2 shows a high amino acid similarity only with the corresponding identified for the tomato matallocarboxypeptidase [18] (not shown).

The region coding for the mature protein is completely included in the second exon.

The 5′ flanking region of the gene contains regulatory motifs typical of eukaryotic genes. A CAAT box is present 221 bp upstream of the putative AUG initiation codon. The TATA box corresponding to the CAAT box is most probably included in the first of the two TATA clusters detectable from nucleotide −165 to nucleotide −134 (see Fig. 1). Other sequence elements (ACGT) which are the central part of specific motifs involved in the regulation of expression of plant genes mediated by abscissic acid (ABRE motifs) and jasmonic acid (G-boxes) [19,20], are present in tandem in position −49 and −43.

In the 3′ flanking region of the gene at a distance of 30 and 78 bp from the UGA stop codon two polyadenylation signals (AATAAA) can be detected.

The expression of the gene is apparently under both developmental and environmental control. Indeed its expression is detectable in the late stage of seed maturation and in leaves, only after mechanical wounding.

The possibility of the existence of a gene family and therefore that two different genes could be expressed in seeds and in wounded leaves can be ruled out. In fact the gene is present as a single copy as demonstrated by restriction and hybridization analysis of a 1.6 kb *Pst*I fragment detected by Southern blot analysis in a total digest of nuclear DNA (not shown). It must be mentioned that other proteinase inhibitor genes, present as single copies, are both developmentally and environmentally regulated [17,18]. Finally it has been also demonstrated that a single gene of a potato proteinase inhibitor II family is under either developmental or environmental control [21].

The complete knowledge of the organization of the *mti*-2 gene makes possible the study of its expression mainly upon wounding stress. This investigation is of great interest also because it has been shown that the MTI-2 protein is highly effective against specific pathogens (pyralis) and inhibits their proteinases (S. Palmieri, personal communication). In particular, the knowledge of the control mechanism(s) and of the DNA regions involved, will be useful to set up plant transformation experiments aimed at increasing the resistence of specific crop plants.

Acknowledgements: We are grateful to Drs. M. Kuntz and A. Steinmetz (IBMP, Strasbourg) for helpful advice in starting with this project. This work was financially supported by the Italian Ministry of Agricolture, Food and Forestry Resources (MRAFF) target oriented projects: 'Tecnologie avanzate applicate alle piante', progetto no. 9 and 'Resistenze genetiche delle piante agrarie agli stress biotici ed abiotici', progetto no. 17.

References

[1] Ryan, C.A. (1973) Annu. Rev. Plant Physiol. 24, 173–196.
[2] Ryan, C.A. (1981) in: The Biochemistry of Plants (Stumpf, P.K. and Conn, E.E. eds.) vol. 6, pp. 351–370, Academic Press, New York.
[3] Ryan, C.A. (1990) Annu. Rev. Phytopathol. 28, 425–449.
[4] Menegatti, E., Tedeschi, G., Ronchi, S., Bortolotti, F., Ascenzi, P., Thomas, R. M., Bolognesi, M. and Palmieri, S. (1992) FEBS Lett. 301, 10–14.
[5] Drenth, J., Low, B.W., Richardson, J.S. and Wright, C.S. (1980) J. Biol. Chem. 255, 2652–2655.
[6] Ceciliani, F., Bortolotti, F., Menegatti, E., Ronchi, S., Ascenzi, P. and Palmieri, S., (1994) FEBS Lett. 342, 221–224.
[7] Domon, C., Evrard, J.L., Pillay, D.T.N. and Steinmetz, A. (1991) Mol. Gen. Genet. 229, 238–244.
[8] Slightom, J.L. and Drong, R.F. (1988) in: Plant Molecular Biology Manual (Gelvin, S.B. and Schilperoot, R.A. eds.), pp. 1 42, Kluwer Academic Press, Dordrecht.
[9] Benton, W.D. and Davis, R.W. (1987) Science 196, 180–182.
[10] Arrand, J.E. (1985) Preparation of Nucleic Acid Probes; in: Nucleic Acid Hybridization: a Practical Approach (Hames, B.D. and Higgins, S.J. eds) pp. 17–45, IRL Press, Oxford/Washington.
[11] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: a Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
[12] Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105 132.
[13] Walker-Simmons, M. and Ryan, C.A. (1977) Plant Physiol. 60, 61–63.
[14] Lee, J.S., Brown, W.E., Graham, J.S., Pearce, G., Fox, E.A., Dreher, T.W., Ahern, K.G., Pearson, G.D. and Ryan, C.A. (1986) Proc. Natl. Acad. Sci. USA 83, 7277–7281.
[15] Cleveland, T.E., Thornburg, R.W. and Ryan, C.A. (1987) Plant Mol. Biol. 8, 199–207.
[16] Keil, M., Sanchez-Serrano, J., Schell, J. et al. (1986) Nucleic Acids Res. 14, 5641–5650.
[17] Rohrmeier, T. and Lehle, L. (1993) Plant Mol. Biol. 22, 783–793.
[18] Martineau, B., McBride, K.E. and Houck, C. (1981) Mol. Gen. Genet. 228, 281–286.
[19] Guiltinan, M.J., Marcotte Jr., W.R., Quatrano, R.S. (1990) Science 250, 267–271.
[20] Ishikawa, A., Yoshihara, T. and Nakamura, K. (1994) Plant Mol. Biol. 26, 403–414.
[21] Keil, M., Sanchez-Serrano, J. and Willmitzer, L. (1989) EMBO J. 8, 1323–1330.