Plant Molecular Biology 11: 483–493
© Kluwer Academic Publishers, Dordrecht – Printed in the Netherlands

Molecular cloning of cDNAs encoding a putative cell wall protein from *Zea mays* and immunological identification of related polypeptides

Virginia Stiefel,[1] Lluis Pérez-Grau,[1,3] Fernando Albericio,[2] Ernest Giralt,[2] Luis Ruiz-Avila,[1] M. Dolors Ludevid[1] and Pere Puigdomènech[1]*
[1]*Departamento de Genética Molecular, Centro de Investigación y Desarrollo, CSIC, Jorge Girona Salgado 18, 08034 Barcelona, Spain;* [2]*Departamento de Química Orgánica, Facultad de Químicas, Universidad de Barcelona, Diagonal 647, 08028 Barcelona, Spain;* [3]*Present address: Department of Biology, University of California, Los Angeles, California, USA (*author for correspondence; telephone (34-3)2046851)*

Received 29 April 1988; accepted in revised form 14 July 1988

Key words: cDNA cloning, cell wall proteins, differential screening, extensin, immuno-detection, maize.

Abstract

Copy DNAs corresponding to a highly repetitive, proline-rich protein from maize have been cloned by differential screening of a coleoptile cDNA library. The deduced amino acid sequence contains a single repetitive element of carrot extensin (Ser-Pro-Pro-Pro-Pro). The related mRNAs have a defined distribution in tissues of the plant and are accumulated mainly in the coleoptile node and root tip. A peptide that corresponds to one of the repetitive elements of the protein has been synthesized and antisera have been obtained in rabbits. These antibodies react against crude preparations of coleoptile cell wall and against polypeptides extracted following the protocols described for the extraction of extensin. From these data it is concluded that the cDNAs correspond to a family of cell wall glycoproteins from maize.

Introduction

Several proline- or hydroxyproline-rich glycoprotein constituents of plant cell walls have been identified from dicotyledonous species. The best characterized of these proteins is extensin, a glycoprotein extracted from dicotyledonous plants. The identification of cDNAs corresponding to carrot extensin has allowed the regulation of its expression to be examined and the complete sequence of the polypeptide from carrot to be elucidated [6]. The sequence obtained through DNA cloning has confirmed that extensin is composed of highly repetitive sequences including the Ser-Pro-Pro-Pro-Pro motif [6, 7]. Extensin genes appear to form a small group of genes having interesting properties of regulation. At least some of the members of this family are induced by fungal infection [10, 38] and by wounding [6]. Other proline-rich proteins with highly repetitive sequences of which cDNAs or genes have been cloned include an auxin-induced gene [19] or nodulin-75 both from soybean [12].

It has been reported that the cell wall proteins from some graminaceous species appear to have a low proportion of hydroxyproline [24] and a low degree of glycosylation compared with dicot cell wall glycoproteins [25]. Recently, the purification of an extensin-related protein from maize cell suspensions has been reported [22]. Here we characterize cDNAs corresponding to a highly repetitive proline-rich protein from maize. The amino acid composition deduced from the cDNA sequence correlates

with that obtained for cell wall proteins extracted from this cereal [22]. Antibodies elicited againt a peptide synthesized according to the deduced protein sequence react with a small number of polypeptides present in preparations of maize cell wall proteins.

Materials and methods

Isolation of RNA, cDNA synthesis and cloning

Plantlets of maize W64A, a pure inbred line, or the double hybrid E41, were germinated by inbibition in water at 25 °C. RNA was isolated from different organs by phenol/chloroform treatment as described by Burr and Burr [4] and precipitation in 2 M LiCl, 4 M urea. Poly(A)$^+$RNA was purified by two passages through oligo(dT) cellulose. Cloning procedures followed previously published methods [28, 34, 35]. Double-stranded cDNA was prepared by the Okayama and Berg [32] procedure according to Gubler and Hoffman [17] and was fractionated by electrophoresis in 1.5% agarose and eluted from the gel according to Dretzen *et al.* [9]. Fragments between 0.5 and 2.5 kb were inserted in the *Pst* I site of pBR322 by dC/dG tailing [41]. *Escherichia coli* strain HB101 cells were transformed according to Hanahan [18]. Enzymes were purchased from Boehringer (Mannheim) except AMV reverse transcriptase (Stehelin, Basel) and DNA polymerase, restriction endonucleases AhaII and SnaBI (New England Biolabs). Labelled nucleotides were purchased from Amersham.

Nitrocellulose filter replicas of the cDNA library were incubated with ^{32}P-labelled single-stranded cDNA synthesized including a labelled precursor at a specific activity of around 15×10^7 cpm/μg from coleoptile and root poly(A)$^+$ RNA. Following hybridization in 5× SSC (1× SSC is 0.15 M NaCl, 0.015 M sodium citrate, pH 7.0), 5× Denhardt's (1× Denhardt's is 0.02% each of Ficoll 400, polyvinylpyrrolidone and bovine serum albumin), at 65 °C with $1-2 \times 10^6$ cpm/ml, filters were washed in 0.5× SSC at 65 °C, air-dried and autoradiographed.

The Northern blot analysis was carried out essentially as described [33]. 10 μg of total RNA or 0.5 μg of poly(A)$^+$ RNA was separated in 1.5% agarose gels with 2.2 M formaldehyde [26], transferred to nitrocellulose membranes (BA85, Schleicher and Schuell) and hybridized with DNA labelled by nick translation. Hybridization and washings were carried out essentially as described by Alwine *et al.* [1] to a 0.3× SSC, 0.1% SDS, 65 °C stringency, dried and autoradiographed.

Nucleotide sequencing

The nucleotide sequence was determined using both the dideoxynucleotide [37] and chemical modification procedures [29]. The A + G reaction in the Maxam and Gilbert method was modified as described by Bernard and Gough [2]. Internal *Sac*I, *Hae* III, *Hpa* II, *Sna* BI and *Aha* II sites were used for subcloning the inserts in M13mp18 and 19 and pUC18 and 19. All sequences were determined for both strands and the restriction sites confirmed by overlapping DNA fragments.

Whenever possible the sequence was determined by the M13-dideoxynucleotide method. The ends containing the C-G tails were sequenced by the chemical degradation technique. In this case subclones in pUC18 were cut at the *Hin*d III site in the polylinker, 5′ end-labelled with T4 polynucleotide kinase and restricted with a second endonuclease. Sequence alignment and treatment was done by using the computer and software of the CITI2 database (Paris) and MicroGenie (Beckman).

Peptide synthesis

Peptide synthesis was performed manually in a 50 ml polypropylene syringe fitted with a polyethylene disc. A standard coupling protocol was used on 1 g of resin. The synthesis strategy chosen was based on differential acid-labile protecting groups [16]. Two coupling cycles were performed for all residues except for lysine, which required a third coupling cycle. Final deprotection and cleavage from the resin were conducted by the high-HF procedure [39] with a 90% yield. The peptide was purified by gel filtration in Sephadex-G15 and -G25 and the purity checked

by HPLC. A cysteine residue was incorporated at the N-terminus to favour reversible coupling reactions with a carrier. The thiol group of cysteine was protected with the HF-stable group 3-nitro-2-pyridinesulphenyl (Npys). The S-Npys has the advantage of being stable in the final cleavage step and incorporates into the peptide a chromophore which allows easy monitoring of the purification. Deprotection of cysteine was performed with 1% (v/v) 2-mercaptoethanol.

Immunological methods

Antibodies against synthetic peptide (either the Npys derivative or the unprotected form) were raised in New Zealand rabbits. The animals were injected intradermally at 15-day intervals with 300 μg of synthetic peptide (without carrier) in 500 μl of 0.1 M phosphate buffer pH 7.5, 0.15 M NaCl (phosphate-buffered saline, PBS) emulsified with an equal volume of Freund's complete adjuvant (1st boost), incomplete adjuvant (2nd boost) and PBS (3rd boost). Ten days after the third injection, 5 ml sera were collected and stored in small aliquots at −20 °C. Titre and specificity of antisera were determined by the enzyme-linked immunosorbent assay (ELISA). The ELISA was used essentially according to Craig *et al.* [8]. The synthetic peptide was coupled to keyhole limpet hemocyanin (KLH) (Calbiochem) protein used as carrier to increase adhesion to microtitre plates. After incubation with antisera (1:50 final dilution), goat anti-rabbit peroxidase conjugate (Dako Patts) was used for antibody detection. Non-immunized rabbit sera were used as controls. Protein concentrations of extracts were determined by the method of Bradford [3]. The concentration of KLH-peptide is referred to as peptide concentration and was determined by amino acid analysis. The protein extracts were separated by SDS-PAGE (15% acrylamide, 0.4% bisacrylamide) according to Laemmli [23], and transferred to nitrocellulose sheets according to Towbin *et al.* [40]. The sheets were saturated with 5% (w/v) non-fat milk and incubated for 4 h at 37 °C with antisera (dilution 1:125). After subsequent washings with 0.05% Tween-20 and 1 M NaCl in PBS, goat anti-rabbit peroxidase conjugate was used for antibody detection. Protein gels were stained with silver stain [31].

Cell wall protein extraction

The upper portions (ca. 4 cm) of 6-day old W64A maize coleoptiles were cut and used for extraction of cell wall proteins. Two extraction methods (I and II) were carried out to obtain cell wall proteins extracts.
- *Method I (ethanolic/acid extraction).* 20 g tissue were ground in a chilled mortar with liquid nitrogen and the powder extracted with a mixture of ethanol: 1.25 N HCl 3:1 (v/v) overnight at 4 °C, as described by Mazeau *et al.* [30]. The homogenate was filtered through a nylon cloth and centrifuged at 8500 × *g* for 20 min. The supernatant was precipitated by 4 vol. of cold acetone, and the pellet resuspended in water. After addition of thrichloroacetic acid (TCA) to 10% (w/v) final concentration, the soluble fraction was extensively dialysed against water and freeze-dried.
-*Method II (saline extraction).* 20 g of tissue were ground as described for method I. Salt-extractable proteins were obtained by treatment with 0.2 M $CaCl_2$ as described by Cassab [5]. The final supernatant was precipitated with 10% TCA and the soluble fraction was processed as described for method I. Deglycosylation of proteins (3 mg) extracted according to method I was performed with anhydrous HF as described by Sanger and Lamport [36]. Deglycosylated material was dissolved immediately in 1 ml water and freeze-dried.

Results

Cloning of maize coleoptile-specific cDNAs

The original aim of the present study was to detect and characterize cDNAs that can be useful as markers for specific developmental stages in maize. To this end a cDNA library of maize (E41 double hybrid) coleoptile (7 days after germination) was constructed by cloning size-selected cDNA (500 to 2500 bp) in the *Pst* I site of plasmid pBR322 by dG/dC tailing. The library (800 recombinant clones)

was screened with labelled single-stranded cDNA synthesized from poly(A)$^+$ RNA extracted from coleoptiles and primary roots of plantlets collected at the same period of germination. The clones giving a positive signal with coleoptile but not with root cDNA were chosen. A further differential screening was carried out by hybridizing labelled single-stranded cDNA synthesized from coleoptile and primary root RNA to the digested DNA from mini-preparations of plasmids of several clones transferred to nitrocellulose. Twelve clones displayed differential hybridization at this stage. One of these clones was chosen the mRNA of which appears to be present at a higher abundance in coleoptile RNA than in root RNA.

A further screening of a cDNA library prepared from coleoptile poly(A)$^+$ RNA of the W64A pure inbred line was carried out with the selected clone to detect the clones with the longest insert. A clone having a 1300 bp insert (clone MC56) was chosen for further studies. This clone gave a clear signal in a Northern analysis of total coleoptile RNA and at least a 30 times higher increase in this organ than in total root RNA as measured by densitometric analysis of the autoradiographs. The size of the mRNA band hybridizing with this clone was 1500 nucleotides following Northern analysis (Fig. 1).

The level of mRNA corresponding to clone MC56 in different parts of the maize plant was measured by Northern analysis. Young maize plants were manually dissected into several portions: plumule, coleoptile node, shoot, root apex and total root. RNAs from adult leaf and whole immature (20 days after pollination) and dry embryo were also prepared. The results are presented in Fig. 1. Specific mRNA corresponding to MC56 was present mainly in coleoptile node, plumule and root apex. It was also detectable at low levels in adult leaf and in embryo two days after the beginning of germination. It was not detectable in immature (20 days after pollination) or dry embryo. Poly(A)$^+$ RNA from coleoptile gave a strong signal (Fig. 1) confirming that the RNA hybridizing to this clone is present in the polyadenylated RNA fraction.

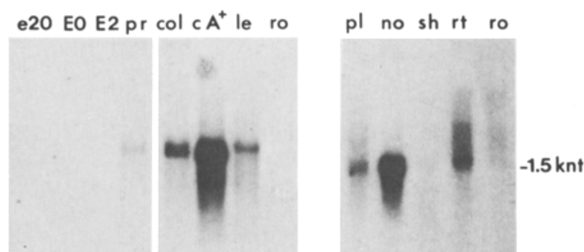

Fig. 1. Northern analysis of RNA extracted from different organs of maize and probed with the insert of MC56 clone. Total RNA (10 μg) from immature embryo 20 days after pollination (e20), dry embryo (EO), embryo 2 hours after inhibition (E2), primary root (pr), coleoptile (col), adult leaf (le), adult root (ro), plumule (pl), coleoptile node (no), shoot (sh), root tip (rt), and young root minus root tip (ro) were used. Also shown is poly(A)$^+$ RNA (0.5 μg) from coleoptile (cA$^+$).

Nucleotide and protein sequence

The nucleotide sequence of the insert of MC56 was determined by the M13-dideoxynucleotide method except when secondary structure problems required the use of the chemical degradation procedure and fragments were subcloned in pUC18 or 19. All the restriction sites used for subcloning were confirmed by overlapping sequences and all the segments were sequenced at least twice. *Aha* II and *Sna* BI sites were very useful in the subcloning as most of the usual restriction enzymes did not cut the clone. The nucleotide sequence was translated into protein sequence. Only one open reading frame was observed having the requirements of a protein sequence. The sequence has a 42-nucleotide long poly(A)$^+$ stretch at the 3′ end. Both the nucleotide and the corresponding protein sequences are shown in Fig. 2.

The protein sequence has a well defined domain structure with three zones: a 17 amino acid long N-terminal stretch that is the only fragment departing from the general composition of the protein, a highly repetitive zone in the central region and a non-repetitive 16 amino acids long carboxy-terminus. The repetitions along the sequence can be clearly observed in the hydrophilicity profile of the sequence (Fig. 3). The sequence corresponds to a 29 kDa protein. The last segment contains one element of the repetitive motif of carrot extensin Ser-Pro-Pro-Pro-Pro and the two last amino acids (Tyr-Tyr) are the same as maize 16 kDa zein-2 [34]. If the first 17 ami-

```
          tttttttag ATG TGC CCG GCT TTT AGC ATT TTT TTT AAC TCC AGA AGA TAC AGC CTC ACT    51
                    Met Cys Pro Ala Phe Ser Ile Phe Phe Asn Ser Arg Arg Tyr Ser Leu Thr    17

CCT CCG ACG TAC ACC CCT  TCC CCC AAA CCT CCG  -------------------  ACA CCT AAG CCG ACC    99
Pro Pro Thr Tyr Thr Pro  Ser Pro Lys Pro Pro  -------------------  Thr Pro Lys Pro Thr    33

CCG CCG ACG TAC ACT CCA  AGC CCC AAG CCA CCG  GCT AGC AAG CCT CCC  ACG CCC AAG CCG ACC   162
Pro Pro Thr Tyr Thr Pro  Ser Pro Lys Pro Pro  Ala Ser Lys Pro Pro  Thr Pro Lys Pro Thr    54

CCG CCG ACG TAC ACC CCT  TCT CCC AAG CCT CCG  -------------------  ACA CCT AAG CCG ACC   210
Pro Pro Thr Tyr Thr Pro  Ser Pro Lys Pro Pro  -------------------  Thr Pro Lys Pro Thr    70

CCG CCT ACG TAC ACT CCA  AGC CCC AAG CCA CCG  GCT ACC AAG CCG CCG  ACG CCC AAG CCG ACC   273
Pro Pro Thr Tyr Thr Pro  Ser Pro Lys Pro Pro  Ala Thr Lys Pro Pro  Thr Pro Lys Pro Thr    91

CCA CCG ACG TAC ACC CCT  TCC CCC AAA CCT CCG  -------------------  ACA CCT AAG CCG ACC   321
Pro Pro Thr Tyr Thr Pro  Ser Pro Lys Pro Pro  -------------------  Thr Pro Lys Pro Thr   107

CCG CCT ACG TAC ACT CCA  AGC CCC AAG CCA CCG  GCT ACC AAG CCT CCC  ACG CCC AAG CCG ACC   384
Pro Pro Thr Tyr Thr Pro  Ser Pro Lys Pro Pro  Ala Thr Lys Pro Pro  Thr Pro Lys Pro Thr   128

CCG CCG ACG TAC ACC CCT  TCT CCC AAG CCT CCG  -------------------  ACA CCC AAG CCG ACC   432
Pro Pro Thr Tyr Thr Pro  Ser Pro Lys Pro Pro  -------------------  Thr Pro Lys Pro Thr   144

CCG CCG ACG TAC ACC CCT  TCC CCC AAG CCT CCG  -------------------  ACG CCC AAG CCG ACC   480
Pro Pro Thr Tyr Thr Pro  Ser Pro Lys Pro Pro  -------------------  Thr Pro Lys Pro Thr   160

CCG CCG ACG TAC ACT CCA  AGC CCC AAG CCT CCC  --- ACA CAC CCG ---  ACG CCC AAG CCG ACC   537
Pro Pro Thr Tyr Thr Pro  Ser Pro Lys Pro Pro  --- Thr His Pro ---  Thr Pro Lys Pro Thr   179

CCA CCG ACG TAC ACC CCT  TCC CCA AAG CCT CCG  -------------------  ACA CCT AAG CCG ACC   585
Pro Pro Thr Tyr Thr Pro  Ser Pro Lys Pro Pro  -------------------  Thr Pro Lys Pro Thr   195

CCG CCG ACG TAC ACG CCT  TCC CCC AAG CCT CCG  -------------------  ACA CCC AAG CCG ACC   633
Pro Pro Thr Tyr Thr Pro  Ser Pro Lys Pro Pro  -------------------  Thr Pro Lys Pro Thr   211

CCA CCG ACG TAC ACT CCA  AGC CCC AAG CCA CCG  GCT ACC AAG CCT CCC  ACG CCC AAG CCG ACC   696
Pro Pro Thr Tyr Thr Pro  Ser Pro Lys Pro Pro  Ala Thr Lys Pro Pro  Thr Pro Lys Pro Thr   232

CCA CCG ACG TAC ACT CCC  ACA CCG AAG CCG CCG  GCC ACC AAG CCG CCC  ACC TAC ACT CCG ACG   759
Pro Pro Thr Tyr Thr Pro  Thr Pro Lys Pro Pro  Ala Thr Lys Pro Pro  Thr Tyr Thr Pro Thr   253

CCC CCG GTG TCT CAC ACC CCC AGC CCG CCG CCA CCT TAC TAC TAG aaaccgatgcctaccataccacactg   830
Pro Pro Val Ser His Thr Pro Ser Pro Pro Pro Pro Tyr Tyr end                              267

ctgtcagtctctggagcatttaggtggtctcgatcgatggaagaattgtgtcctagccagccggcaaaggtgacctgctgatgatg   916

atgatgagaggcgagtcctacgccctagtcctactaccctctttgtgtgctgccatccatccgtccccgctagacgatcgagg  1002

agagaataacgcagagctctgtgctcccggcctctgtcttctcctcccggccgtttaatttatagtctctactgtgtgttcgtccc  1088

atgtgtttagcagcagcagcagtagtgtgcgggtatgtaatggtattgcaactatattgggtgtaaaaccataataaatgtggaaa  1174

aaatgcaaggaaaaaaaaaaaaaaaaa                                                             1189
```

Fig. 2. Nucleotide and deduced protein sequence of the insert of clone MC56. The nucleotides and amino acid residues have been numbered from the first ATG codon. The sequence has been presented in order to emphasize the homology of the repetitive elements in the protein sequence. The putative polyadenylation signal is underlined.

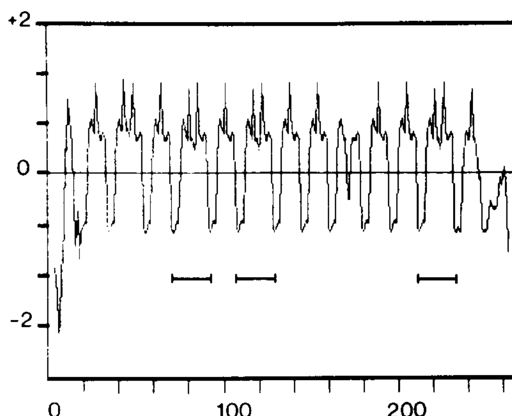

Fig. 3. Hydrophylicity profile of the amino acid sequence predicted from MC56. The mean hydrophylicity of a window of 6 consecutive residues (as described in [20]) is plotted against the amino acid number. Bars indicate the protein regions containing the sequence of the synthetic peptide used to elicit antibodies.

no acids are not considered, the composition of the predicted polypeptide is: Pro 48.2%, Thr 23.9%, Lys 12.0%, Ser 6.0%, Tyr 6.0%, Ala 2.0%, His 1.2% and Val 0.4%. It corresponds to a protein with a basic character (it has 12.0% lysine and no acidic residue) and is in agreement with a cell wall protein from maize reported by Kieliszewski and Lamport [22]: Pro+Hyp 39.3%, Thr 25.3%, Lys 13.5%, Ser 7.3%, Tyr 3.9%, Ala 1.7%, His 2.4% and Val 0.7%. The codon usage is clearly biased towards the use of a small number of codons, with C or G in the third base. This is the case for the most abundant amino acids in the sequence: Pro (66% CCC or CCG), Thr (75% ACC or ACG), Lys (93% AAG), Ser (80% AGC or TCC) and Tyr (100% TAC). That gives a high G+C content (67%) in the coding part of the cDNA, compared with the 3′ non-coding region (49%).

The central part of the protein is formed by a highly repetitive domain. It contains a sequence of six amino acids (Pro-Pro-Thr-Tyr-Thr-Pro) exactly repeated 13 times followed by two or three pentapeptides having homology among themselves. These segments have the sequences: Ser-Pro-Lys-Pro-Pro, Ala-Thr-Lys-Pro-Pro and Thr-Pro-Lys-Pro-Thr. The sequence of the repeating elements is also well conserved at the nucleotide level that being specially true in the sequences coding for the dipeptides Thr-Tyr (ACGTAC) or Lys-Pro (AAGCCG) which are perfectly conserved in 13 or 12, respectively, of the repetitive units. Although the protein is perfectly conserved in most of the repeating fragments no perfect duplication was observed at the nucleotide level.

At the 3′ end a 382 nucleotide non-coding region is present, containing a canonic polyadenylation AATAAA signal (underlined in Fig. 2). This region has a large number of small repetitive sequences. The sequence of a partial cDNA cloned from the E41 variety (a double hybrid) has also been obtained (Fig. 4). Apart from a number of single nucleotide changes, the main differences between the two sequences appears to be a number of duplications in the small repetitive sequences (shown by arrows in Fig. 4). The duplicated sequences have a high degree of homology with similar duplicated sequences observed between two related cDNAs corresponding to genes expressed in maize endosperm [35].

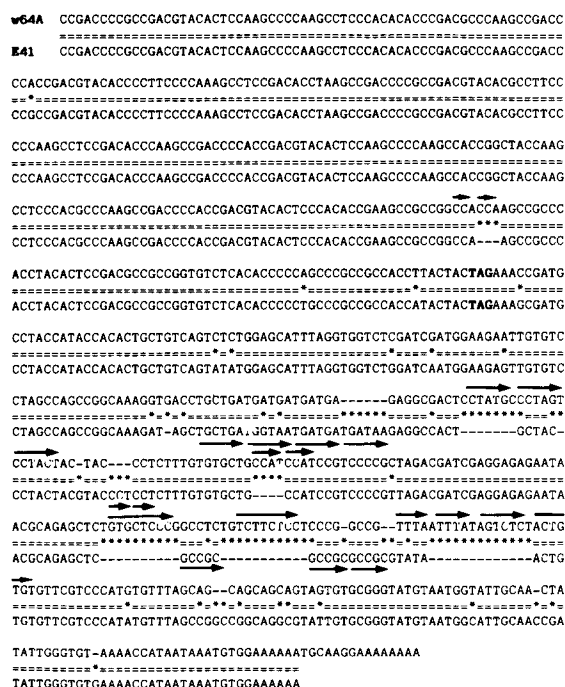

Fig. 4. Comparison of nucleotide sequences of cDNA from two maize varieties. A fragment of the cDNA sequence of clones obtained from two maize varieties (W64A and E41) is shown aligned for maximum homology. The repetitive sequences where duplications are observed are shown with arrows, and the stop codon TAG with bold characters.

Protein extraction and immunological analysis

The sequence of the protein deduced from the sequence of clone MC56 corresponds to a highly repetitive, proline-rich protein. A search in protein databanks (NBRF and Swisspro) for similarities gave a number of animal proline-rich proteins such as human basic proline-rich proteins [21] or the threonine-rich, proline-rich protein sgs-3 from *Drosophila melanogaster* [14], but in both cases the repetitive motif lacks lysine and tyrosine. Examples of proline-rich proteins are also found in plant cell wall proteins [6, 19]. In order to check whether the cDNA corresponds to a cell wall protein from maize, antibodies were raised against one of the repetitive units of the protein. A peptide (Fig. 5A) corresponding to one of the repeating units having the highest hydrophilicity was synthesized (Fig. 3) and injected into rabbits. Using ELISAs it appears that antibodies reacting with the peptide bound to a carrier protein (KLH) are present in antisera, while no reaction was observed with pre-immune antisera nor with the KLH protein (Fig. 5B).

The antisera were used to identify immunoreactive polypeptides in maize cell wall protein preparations. The analysis by ELISA shows that the antisera react with the ethanolic/acid protein extractions of maize coleoptile (Fig. 5B). The reaction was slightly increased by deglycosylation of the protein extract by treatment with anhydrous hydrogen fluoride.

The use of western blot allows the identification of some polypeptides reacting with antisera (Fig. 6). Protein patterns of total ethanolic/acid and saline extractions are presented as well as the 10% TCA-soluble fraction and the result of deglycosylation of the ethanolic/acid fraction. The blot of proteins reacting with the antisera against the synthetic peptide is also shown. In the different extracts proteins in the 50–70 kDa region react with the antisera. Deglycosylation enhances the immunoreaction and causes a slight decrease (around 6 kDa) of apparent molecular weight.

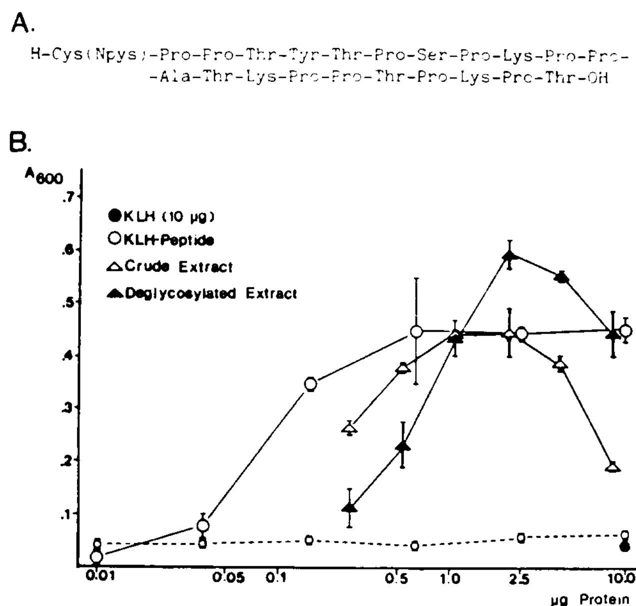

Fig. 5. Specificity of antiserum against a synthetic peptide designed from the sequence of MC56 protein as measured by ELISA. A. Sequence of the synthetic peptide used for rabbit immunization. B. Antiserum reaction against the synthetic peptide coupled to KLH and against TCA-soluble proteins from an ethanol/acid extract before (△) and after (▲) HF treatment. Non-immune serum was used as control (dotted line).

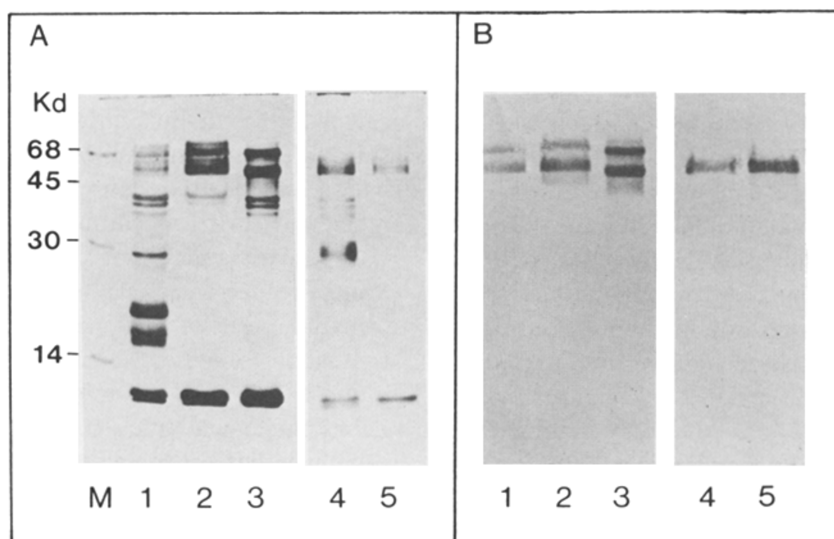

Fig. 6. SDS-polyacrylamide gel electrophoresis (A) and immunoblots (B) of two maize cell wall protein extracts. Total upper coleoptile proteins extracted with ethanol/acid (lane 1). 10% TCA-soluble proteins (lane 2) and deglycosylated proteins (lane 3). Total cell wall proteins soluble in 0.2 M $CaCl_2$ (lane 4) and 10% TCA-soluble proteins (lane 5). Nitrocellulose filters were incubated with serum against synthetic peptide (see figure 5) and developed with peroxidase goat anti-rabbit IgG. M: molecular weight markers.

Discussion

We report here the cloning by differential screening of a cDNA hybridizing with mRNAs abundant in the young coleoptile of maize. The cDNA hybridizes with mRNAs, showing a well-defined distribution in developing organs containing a high proportion of meristematic tissues. The sequence of the protein coded by the cDNA shows a highly repetitive structure and it contains a high proportion of proline, features typical of proteins having a structural function. The best characterized of the structural proteins in the cell wall of plants is extensin. The sequences of cDNA and genomic clones of extensin for carrot have been published [6]. The main repetitive motif in dicotyledon extensin is Ser-Pro-Pro-Pro-Pro [24], a sequence occurring once near the carboxy-terminus of the maize protein (Fig. 2). Amino acids which may have a function in forming cross-links between protein chains (such as tyrosine) or in post-translational modification and glycosylation (such as proline, lysine and serine) are present in extensin as well as in the sequence of MC56 protein. Tyrosine has been shown to form covalent cross-links between extensin polypeptides giving rise to isodityrosine [13]. Nevertheless the main repetitive sequence in the maize protein clearly departs from those found in dicot tissues, showing in particular a lower proportion of proline and a high proportion of threonine. Other hydroxyproline-rich proteins have been described in plants such as the arabinogalactan-proteins [11]. No information is available on the primary structure of these proteins, however they seem to be acidic in nature in contrast to extensin and the protein encoded by MC56 cDNA. The fact that polypeptides extracted as extensin react with the antisera against the synthetic peptide and the repetitive nature of the sequence would point towards a function of this protein in the cell wall of maize. Recently cDNAs encoding other putative cell wall proteins with similar sequence features have been cloned. These include an auxin-induced mRNA from soybean [19] and nodulin-75 [12]. In both cases a proline-rich repetitive sequence is observed, Pro-Pro-Val-Tyr-Lys and Pro-Pro-(Val/Lys)-Glu-Lys-Pro-Pro respectively.

In the case of the protein studied here, the sequence of the repeating units is almost perfectly con-

served at the protein level and highly conserved at the nucleotide level. The protein may be considered as the union of 13 elements each composed of a hexapeptide (Pro-Pro-Thr-Tyr-Thr-Pro) and two (Ser-Pro-Lys-Pro-Pro; Thr-Pro-Lys-Pro-Thr) or three (Ser-Pro-Lys-Pro-Pro; Ala-Thr-Lys-Pro-Pro; Thr-Pro-Lys-Pro-Thr) pentapeptides. Secondary structure prediction with the model of Garnier *et al.* [15] indicates that the only possible structure for this sequence may be beta-sheet or turns. This potential structure is centered in the conserved hexapeptide that contains the tyrosine residue. The three pentapeptide motifs have a high degree of similarity among themselves at the protein and the nucleotide level. An interesting feature is the conservation in codon usage, such as those coding for tyrosine or lysine. The only serine residue present in the sequence of the first pentapeptide is coded by either AGC or TC(C/T). The alternation between the codons correlates with the presence of three or two pentapeptides in the repeat. Such observations suggest that the sequence may derive from two modules (a hexapeptide and a pentapeptide) by successive duplication of the sequences. Another family of proteins from maize having proline-rich repetitive sequences whose cDNAs have been cloned are glutelins [34]. These proteins are found, like storage proteins from cereals, in protein bodies. However, they are located in the inner part of the protein body membrane suggesting a structural function [27]. It is interesting to note that both in the proline-rich domain and in the 3′ non-coding region, the same type of sequence duplications can be observed both in glutelins [35] and in the polypeptide sequence deduced from MC56. This suggests that these sequences share a common origin.

It has been reported that in the maize cell wall, proteins have a lower proportion of hydroxyproline and a lower carbohydrate content [25]. In a recent report on the purification of cell wall proteins from maize the amino acid composition of the proteins extracted from cell suspension cultures was presented [22]. It is very similar to the one deduced from our sequence. The use of antisera raised against one of the repetitive units of the protein has allowed us to identify a group of polypeptides sharing antigenic determinants with the deduced protein. The apparent molecular weight of these polypeptides is in the range of 50–70 kDa, a value higher than the one deduced from the sequence. However the presence of repetitive proline-rich fragments in a protein gives anomalous electrophoretical mobilities as it has been reported in other examples [30]. In any case our results suggest that the coleoptile-specific cDNA we isolated corresponds to a cell wall protein from maize. This probably belongs to the same family of proteins that the THGRP protein described by Lamport [22]. Deglycosylation of the proteins gives rise to a slight (around 6 kDa) decrease in the apparent molecular weight and a better definition of the bands. This result is in agreement with the low proportion of glycosylation observed with graminaceous cell wall proteins.

The results of Northern analysis indicate that the main points of expression are coleoptile node, plumule, and root apex and at a lower level in leaf. The mRNA begins to be accumulated in embryo two days after germination. No mRNA is detected in adult shoot and root tissues. The fact that these genes have a well-defined pattern of expression in different tissues make them an attractive system to study gene regulation in monocotyledons.

Acknowledgements

The authors are indebted to M. Mercedes Jiménez for her skilful technical assistance and to Dr J. A. Martínez-Izquierdo for critical revision of the manuscript. The present work was supported by grants from CSIC (n. 500) and from CSIC-CDTI (plan PROA). V.S. is recipient of a fellowship from Stipendienberatung, ETH, Zürich, and L.R. from PFPI.

References

1. Alwine JC, Kemp DJ, Stark JR: Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. Proc Natl Acad Sci USA 74: 5350–5354 (1977).
2. Bernard O, Gough NM: Nucleotide sequence of immunoglobulin heavy chain joining segments between translocated VH and mu constant region genes. Proc Natl Acad Sci USA 77: 3630–3634 (1980).

3. Bradford MM: A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72: 248–254 (1976).
4. Burr FA, Burr B: In vitro uptake and processing of prezein and other preproteins by maize membranes. J Cell Biol 90: 427–434 (1980).
5. Cassab GI: Arabinogalactan proteins during the development of soybean root nodules. Planta 168: 441–446 (1986).
6. Chen J, Varner JE: An extracellular matrix protein in plants: characterization of a genomic clone for carrot extensin. EMBO J 4: 2145–2151 (1985).
7. Chen J, Varner JE: Isolation and characterization of cDNA clones for carrot extensin and a proline-rich 33 kDa protein. Proc Natl Acad Sci USA 82: 4399–4403 (1985).
8. Craig S, Millerd A, Goodchild DJ: Structural aspects of protein accumulation in developing pea cotyledons. III. Immunocytochemical localization of legumin and vicilin using antibodies shown to be specific by the Enzyme-Linked Immunoabsorbent Assay (ELISA). Aust J Plant Physiol 7: 339–351 (1980).
9. Dretzen G, Bellard M, Sassone-Corsi P, Chambon, P: A reliable method for the recovery of DNA fragments from agarose and acrylamide gels. Anal Biochem 112: 295–298 (1981).
10. Esquerré-Tugaye MT, Laffite C, Mazeau D, Toppan A, Touze A: Cell surfaces in plant-microorganism interaction. II. Evidence for the accumulation of hydroxyproline-rich glycoproteins in the cell wall of diseased plants as a defense mechanism. Plant Physiol 64: 320–326 (1979).
11. Fischer GB, Store BA, Clarke AE: Arabinogalactan-proteins: structure, biosynthesis and function. Ann Rev Plant Physiol 35: 57–70 (1983).
12. Franssen HJ, Nap JP, Gloudemans T, Stiekema W, van Dam H, Govers F, Louwerse J, van Kammen A, Bisseling T: Characterization of cDNA for nodulin-75 of soybean: A gene product involved in early stages of root nodule development. Proc Natl Acad Sci USA 84: 4495–4499 (1987).
13. Fry SC: Isodityrosine: a new cross-linking amino acid from plant cell wall glycoproteins. Biochem J 203: 493–504 (1982).
14. Garfinkel MD, Pruitt RE, Meyerowitz EM: DNA sequence, gene regulation and modular protein evolution in the *Drosophila* 68C glue gene cluster. J Mol Biol 168: 765–789 (1983).
15. Garnier J, Osguthorpe DJ, Robson B: Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J Mol Biol 120: 97–117 (1978).
16. Giralt E, Andreu D, Miro P, Pedroso E: Solid phase synthesis of Tyrosine-containing histone fragments. Tetrahedron 39: 3185–3193 (1983).
17. Gubler U, Hoffman BJ: A simple and very efficient method for generating cDNA libraries. Gene 25: 263–269 (1983).
18. Hanahan D: Studies on transformation of *Escherichia coli* plasmids. J Mol Biol 166: 557–580 (1983).
19. Hong JC, Nagao RT, Key JL: Characterization and sequence analysis of a developmentally regulated putative cell wall protein gene isolated from soybean. Proc Natl Acad Sci USA 262: 8367–8376 (1987).
20. Hopp TP, Wood KR: Prediction of protein antigenic determinants from amino acid sequences. Proc Natl Acad Sci USA 78: 3824–3828 (1981).
21. Kauffman D, Hofman T, Bennick A, Keller P: Basic proline-rich proteins from human parotid saliva: complete covalent structure of proteins IB-1 and IB-6. Biochemistry 25: 2387–2392 (1986).
22. Kieliszewski M, Lamport DTA: Purification and partial characterization of a hydroxiproline-rich glycoprotein in a graminaceous monocot, *Zea mays*. Plant Physiol 85: 823–827 (1987).
23. Laemmli UK: Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227: 680–685 (1970).
24. Lamport DTA: The protein components of primary cell wall. Adv Botan Res 2: 151–218 (1965).
25. Lamport DTA, Miller DH: Hydroxyproline arabinosides in the plant kingdom. Plant Physiol 48: 454–456 (1971).
26. Lehrach H, Diamond D, Wozney J, Boedtker H: RNA molecular weight determination by gel electrophoresis under denaturing conditions, a critical reexamination. Biochemistry 21: 4743–4751 (1977).
27. Ludevid MD, Torrent M, Martínez-Izquierdo JA, Puigdomènech P, Palau J: Subcellular localization of glutelin-2 in maize (*Zea mays* L.) endosperm. Plant Mol Biol 3: 227–234 (1984).
28. Maniatis T, Fritsch EF, Sambrook J: Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1982).
29. Maxam AM, Gilbert W: Sequencing end-labelled DNA with base-specific chemical cleavages. Meth Enzymol 65: 499–560 (1980).
30. Mazeau D, Rumeau D, Esquerré-Tugaye MT: Two different families of hydroxyproline-rich glycproteins in mellon callus: biochemical and immunochemical studies. Plant Physiol 86: 540–546 (1982).
31. Morrisey JH: Silver stain for proteins in polyacrylamide gels: a modified procedure with enhanced uniform sensitivity. Anal Biochem 117: 307–310 (1981).
32. Okayama H, Berg P: High-efficiency cloning of full-length DNA. Mol Cell Biol 2: 161–170 (1982).
33. Pérez-Grau L, Cortadas J, Puigdomènech P, Palau J: Accumulation and subcellular localization of glutelin-2 transcripts during maize endosperm maturation. FEBS Lett 202: 145–148 (1986).
34. Prat S, Cortadas J, Puigdomènech P, Palau J: Nucleic acid (cDNA) and amino acid sequences of the maize endosperm protein glutelin-2. Nucl Acids Res 13: 1493–1504 (1985).
35. Prat S, Pérez-Grau L, Puigdomènech P: Multiple variability in the sequence of a family of maize endosperm proteins. Gene 52: 41–49 (1987).
36. Sanger MP, Lamport DTA: A microapparatus for liquid hydrogen fluoride solvolysis. Sugar and aminosugar composition of *Erysiphe graminis* and *Triticum aestivum* cell walls. Anal Biochem 128: 66–71 (1983).
37. Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:

5463–5467 (1977).

38. Showalther AM, Bell JN, Cramer CL, Bailey JA, Varner JE, Lamb CJ: Accumulation of hydroxyproline-rich glycoprotein mRNAs in response to fungal elicitor and infection. Proc Natl Acad Sci USA 82: 6551–6555 (1985).
39. Tam JP, Heath WF, Merrifield RB: SN2 deprotection of synthetic peptides with a low concentration of HF in dimethyl sulfide: evidence and applications in peptide synthesis. J Am Chem Soc 105: 6442–6445 (1983).
40. Towbin H, Staehelin T, Gordon J: Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc Natl Acad Sci USA 76: 4350–4354 (1979).
41. Villa-Komaroff L, Efstratiadis A, Broome S, Lomedico P, Tizard T, Naber SP, Chick WL, Gilbert W: A bacterial clone synthesizing insulin. Proc Natl Acad Sci USA 75: 3727–3731 (1978).