Plant Molecular Biology **30**: 625–636, 1996.

Molecular characterization of the gene coding for GPRP, a class of proteins rich in glycine and proline interacting with membranes in *Arabidopsis thaliana*

Isabelle Marty[1,3], Amparo Monfort[1], Virginia Stiefel[1], Dolors Ludevid[1], Michel Delseny[2] and Pere Puigdomènech[1,*]

Laboratoire Européen Associé de Biologie Moléculaire Végétale CNRS-CSIC; [1] *Department de Genètica Molecular, CID-CSIC, Jordi Girona, 18, 08034 Barcelona, Spain (*author for correspondence);* [2] *Laboratoire de Biologie Moléculaire et Physiologie Végétale, Unité associée au CNRS 565, Avenue de Villeneuve, 66860 Perpignan, France;* [3] *Present address: INRA, Station des Technologies des Produits Végétaux, Domaine Saint-Paul, BP 91, 84143 Montfavet Cedex, France*

Received 9 May 1995; accepted in revised form 24 November 1995

Key words: *Arabidopsis thaliana*, genomic cloning, glycine-rich, proline-rich, protein-membrane interaction

Abstract

The gene coding for a new class of proteins rich in glycine and proline (GPRP) was cloned in *Arabidopsis thaliana*. In the protein sequence, five amino acids – glycine, proline, alanine, tyrosine and histidine – account for 79.4% of the total composition. The protein has two different glycine-rich domains interrupted by a hydrophobic segment having a high probability of helix formation. The protein synthesized *in vitro* interacts with microsomes possibly through the hydrophobic domain. The gene in *Arabidopsis* has two introns, one in the coding region and the other one in the 5′ non-coding region. The later one is 778 bp long. Homologous sequences are found in carrot, tomato and tobacco. GPRP mRNA is found in the different organs of the plant analyzed except in mature seeds and anthers, and mostly in epidermal and vascular tissues. Possible hypotheses about the function of GPRP are discussed.

Introduction

A growing number of proline- and glycine-rich proteins has been described in plants (for a review, see reference [16]). They are formed by highly repetitive sequences and they are mainly located in the cell wall where they are considered to be important structural components. This is the case for extensins or hydroxyproline-rich glycoproteins (HRGP) and glycine-rich proteins (GRP). Other classes of such proteins contain repetitive proline-rich segments in most of their sequences like the auxin-induced proline-rich proteins (PRP) from soybean [14] or they contain

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number X84315.

proline-rich stretches only in part of their sequence like hybrid PRPs [15]. Genes coding for proteins having proline- or glycine-rich domains have been shown to have a preferential expression in the embryo [15] or the stylar tissue [4] or induced during nodule formation in legumes [8].

Proline and glycine-rich domains have in general been supposed to have a structural function and this is probably the case for cell wall proteins such as HRGP, PRP and GRP. However, it has recently been proposed that in at least a number of cases they may have a very different function. The glycine-rich domains in RNA-binding proteins described in plants [10] seem to take part in the binding process. A proline-rich domain in gamma-zein has been shown to be essential for the protein targeting [9]. It has recently been proposed that proline-rich segments take part in protein-protein recognition in important cellular processes [24].

In particular, it has been shown that the SH3 domains, described in a number of signal transduction or cytoskeletal proteins, interact with other proteins through proline-rich segments [24]. In some cases, proteins presenting repeated sequences rich in proline or glycine identified in animals have also transmembrane domains and a cytosolic domain rich in these two amino acids. This is the case for proteins such as rhodopsin [23], synexin [7] and synaptophysin [35]. Other proteins having N-terminal domains containing proline or glycine stretches and a transmembrane domain include VAMPs or synaptobrevins [1, 38], proteins that take part in the process of vesicle fusion in transport phenomena [34].

The systematic *Arabidopsis thaliana* cDNA sequencing project [13] led to the identification of some cDNAs encoding proteins with repeated sequences rich in proline and glycine. One of these cDNAs (clone p9, EMBL accession number Z17584; Est:Atts0158) was identified in a developing silique cDNA library. The putative protein presents a large domain rich in glycine, proline and tyrosine which constitute 27.1%, 18.0% and 11.2%, respectively, of total protein composition. By comparison of sequences to EMBL database, the nucleotide and protein sequences do not present significative homology with known GRP or HRGP proteins. Only the repeated domain shows similarity to the repeated region of mammal transmembrane proteins.

In this paper, we present the characterization of this cDNA and the corresponding gene which encodes a novel class of proteins rich in glycine and proline in plants. Analysis by *in vitro* transcription and translation shows that the corresponding protein is able to interact with membranes. The expression of the cDNA was studied in different organs of *Arabidopsis thaliana*. Analysis by Southern blotting show that this protein is encoded by one or two genes and that it is conserved in different dicotyledonous species.

Materials and methods

Plant material and culture

Arabidopsis thaliana ecotype Columbia plants were grown to maturity under a continuous fluorescent illumination at 22 °C, on a mixture of vermiculite, perlite and sphagnum (1:1:1) and irrigated with mineral nutrients [12].

Mature tissues as rosette leaves, roots and stems, immature siliques, pollinated or unpollinated flowers were collected on 2-month-old plants. Seeds were collected on 3-month-old plants. Wounding of plants was performed by cutting rosette leaves of 5-week-old plants into a humidified chamber that was then covered and incubated at room temperature (method described by Niyogi and Fink [22]).

Library screening

The initial cDNA clone (clone p9) was obtained from an immature silique library (obtained by Dr Jerome Giraudat, CNRS, Gif-sur-Ivette) and identified from the systematic sequencing programme [13]. The insert was completely sequenced and it appears in the EMBL database under the accession number Z17584. This cDNA

was subcloned in pBS in *Eco*RI site and named p9. The insert was separated from the pBS vector by *Eco*RI digestion and fractionated on agarose gel run. Recovery and elution of the 790 bp fragment was made by the electro-separation technique (Biotrap; Schleicher and Schuell).

For the cloning of the genomic sequence, an *Arabidopsis thaliana* Columbia genomic library in the vector EMBL3 was purchased from Clontech. The library was plated on LE392 host cells in LB top agarose with 0.2% maltose, and incubated overnight at 37 °C. Plaques were transferred to nylon filters (Nytran; Schleicher and Schuell) in two replicas where the DNA was denaturated and immobilized as described by the manufacturer. The filters were prehybridized 4 hours at 42 °C to the following conditions: 40% formamide, 6 × SSC, 1% SDS, 50 μg/ml salmon sperm DNA. The filters were hybridized for 20 h in the same solution with denatured p9 cDNA probe, then washed twice at room temperature in 2 × SSC, three-fold at 55 °C in 2 × SSC, 1% SDS for 20 min each time. The positive clones were selected and replated until a single, well-isolated positive plaque could be picked and stocked for analysis.

Positive clones were then analyzed by DNA blot. Their DNA was digested by restriction enzymes of the phage multicloning region, *Sac*I, *Xho*I and *Sfi*I in order to separate phage arms from gene insert. Digested DNAs were size fractionated on agarose gel run in 1 × TAE (40 mM Tris acetate, 1 mM EDTA, pH 8). Gel was transferred onto nylon membrane (Nytran; Schleicher and Schuell) using 0.4 M NaOH as transfer buffer. The DNA was bound to the membranes by baking at 80 °C for 2 h, prehybridized and hybridized with p9 cDNA probe in the same conditions as described for library screening.

Southern analysis

DNAs were extracted from different plant species using CsCl-ethidium bromide centrifugation [28]. Ten μg of *Arabidopsis thaliana* digested by *Spe*I, *Sac*I and *Eco*RI, 10 μg of tobacco (*Nicotiana tabacum*), and tomato (*Lycopersicon esculentum*) digested by *Eco*RI were fractionated on agarose gel electrophoresis. DNAs were transferred to nylon membrane (Nytran; Schleicher and Schuell) using 0.2 M NaOH as transfer buffer. Membranes baked at 80 °C for 2 h were prehybridized, hybridized with p9 cDNA probe and washed as described for library screening. In the case of homologous DNA hybridization (*Arabidopsis* genomic DNA), stringent washes were carried out twice at 65 °C with 0.5 × SSC, 1% SDS for 20 min.

DNA sequencing

The nucleotide sequence of p9 cDNA was determined by using the dideoxynucleotide chain termination method [29]. The cDNA was sequenced by its two extremities with double strand DNA templates using reverse and universal primers (Amersham Multiwell system). The genomic clones were sequenced using *Exo*III nested deletions and primer walking strategies in the automated fluorescence sequencing (ALF, Pharmacia).

Northern blot analysis

Total RNAs were extracted from different tissues by previously described methods [11]. Ten μg of total RNA was denatured and fractionated on denaturing gel as described by Meinkoth and Wahl [21]. The corresponding gel was transferred to nylon membrane (Nytran; Schleicher and Schuell) using 20 × SSC as transfer buffer. The filter baked at 80 °C, prehybridized for 4 h at 42 °C with the following solution: 40% formamide, 6 × SSC, 1% SDS, 2 × Denhardt's solution, 50 μg/ml of salmon sperm DNA. The filter was hybridized in the same solution with denatured p9 cDNA probe for 20 h, then washed twice with 2 × SSC at room temperature for 10 min, twice at 65 °C with 1 × SSC, 1% SDS for 20 min.

In situ *hybridization*

Tissues were fixed in 3:1 ethanol/acetic acid for 1 h at room temperature. Once the fixative had been removed, the samples could be stored in 70% ethanol at 4 °C. Treatment of tissues prior to hybridization was performed as previously described [20]. Probes, hybridization, washes and detection of the transcripts were carried out according to the suppliers' protocols (RNA colour kit for non-radioactive *in situ* hybridization; Amersham). Both sense and antisense transcripts were synthesized starting with 1 μg of linearized DNA template. The labelling reaction was checked on a 1.5% agarose gel.

An overnight incubation with the detection reagent was required to get the strongest signals. Samples were then dehydrated, incubated in xylene for 5 min and mounted in Permount. Digoxygenin-labelled hybrids were viewed using bright-field microscopy and photographs were taken using Ektachrome 160 films.

In vitro *protein analysis*

The p9 DNA was digested by KpnI in order to transcribe GPRP sense RNA. The *in vitro* transcription of the cDNA was carried out following standard procedures [28].

Cell-free protein synthesis was carried out using an RNase-treated rabbit reticulocyte lysate (Rabbit Reticulocyte Lysate amino acid-depleted; Amersham). To 7 μl of lysate was added 0.8 μl of 35S methionine (1000 Ci/mmol, Amersham), 1 μl of amino acids minus methionine, 0.3 μl of 2 M potassium acetate, 0.5 μl of p9 RNA (200 ng). In order to analyze the relation of proteins to membranes, to this solution 0.8 μl of microsomes (canine microsomes, Promega) were added. The translation solution was incubated to 30 °C for 1 h. When indicated, the translational products were diluted in 40 μl of triethanolamine 50 mM pH 7.5, potassium acetate 150 mM, magnesium diacetate 2.5 mM (TEA buffer) or sodium carbonate buffer (100 mM sodium carbonate pH 11.5). Supernatants were separated from pellets by centrifugation. The pellets were resuspended in 15 μl of loading buffer for polyacrylamide gel electrophoresis [19]. Proteins of the supernatants were precipitated in 20% trichloroacetic acid, washed twice with iced acetone (−20 °C) and resuspended in 15 μl of loading buffer. Sucrose cushions were carried out in order to purify microsomes. The translational products were diluted until 1 ml with the corresponding buffer and loaded on 1 ml 50% sucrose cushions (in TEA or carbonate buffer). Centrifugation was carried out in a Beckman SW60 rotor at 45000 rpm, during 45 min at 4 °C. Pellets were directly resuspended in 15 μl of electrophoresis loading buffer. Interfaces and supernatants were collected and precipitated in 20% trichloroacetic acid, washed twice with iced acetone and resuspended in 15 μl of loading buffer. The microsomal fractions were recovered from the interface. The *in vitro* translation products were analyzed on polyacrylamide gel electrophoresis as described by Laemmli [19].

Results

Characterization of GPRP cDNA and features of the deduced protein sequence

The systematic *Arabidopsis thaliana* cDNA sequencing project led to the identification of a small number of cDNAs encoding proline- and/or glycine-rich proteins. One of these cDNAs (p9, accession number Z17584; Est:Atts0158) coming from the developing silique library was fully sequenced. It contains a 790 bp long insert and an open reading frame corresponding to a protein of 177 amino acids rich in proline and glycine. This cDNA presents a 5′-flanking region of 40 nucleotides and a 3′-untranslated region of 203 nucleotides followed by a 17 nucleotide long polyA tract. The predicted protein (Figs. 1 and 2) has 177 amino acids and it corresponds to a polypeptide having a molecular mass of 17829 Da and an isoelectric point of 10.33. It contains a large proportion of glycine (48 residues, 27.1%), proline (32 residues, 18.0%), alanine (22 residues,

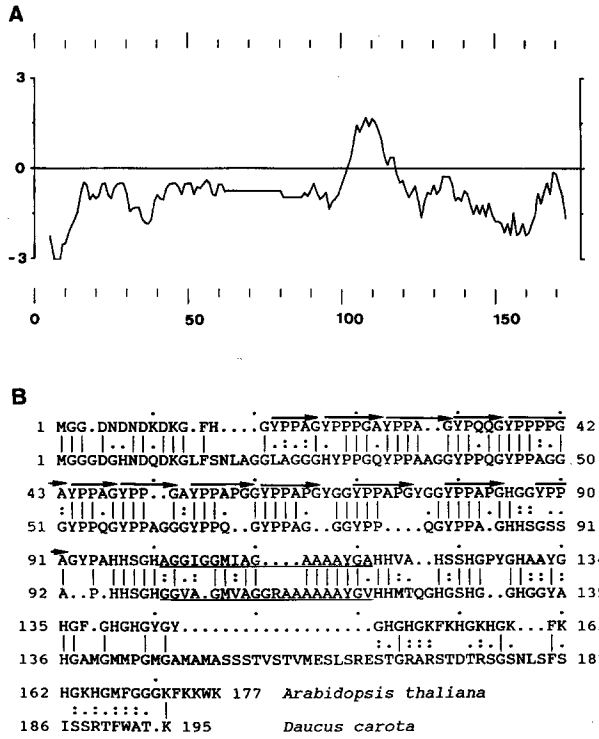

Fig. 1. A. Hydrophobicity profile of GPRP. The hydrophobicity of the GPRP sequence is plotted against the sequence according to the algorithm of Kyte and Doolittle. B. Alignment of GPRP with a similar carrot protein sequence. The GPRP sequence is compared with a protein reported from carrot having a similar distribution of amino acids. The residues forming the central hydrophobic domain are underlined and the repetitive segments in the N-terminal domain are shown with arrows.

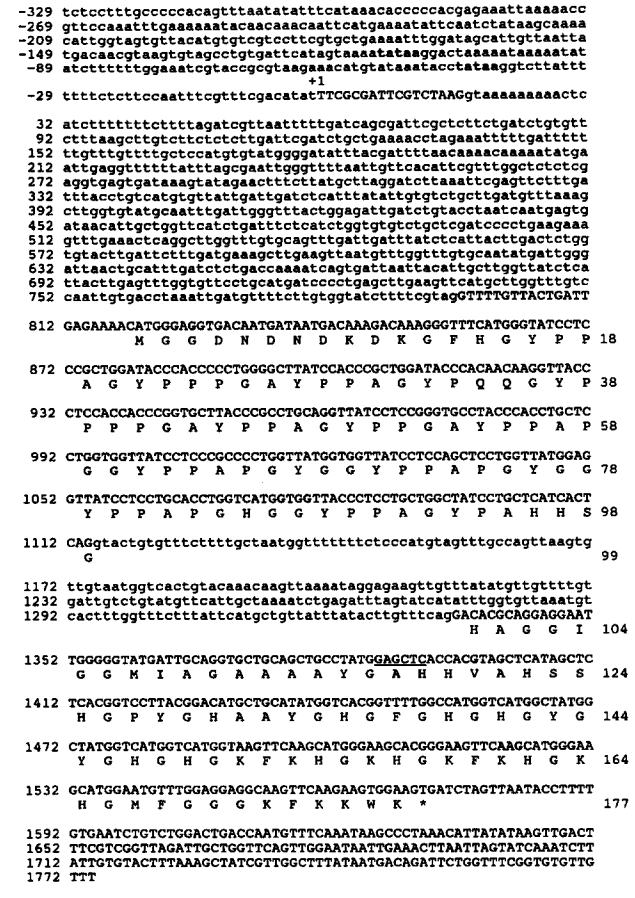

Fig. 2. Nucleotide and deduced amino acid sequences of the GPRP gene of *Arabidopsis thaliana*. The cDNA sequence is shown in upper case letter and the genomic sequences not present in cDNA in lower-case letters. Putative TATA boxes are shown in boldface letters. Numbers on the left refer to the nucleotide sequence. The deduced amino acid sequence is presented in the single-letter code below the corresponding nucleotide sequence. Numbers on the right refer to the amino acid sequence.

12.4%), tyrosine (20 residues, 11.2%) and histidine (19 residues, 10.7%) residues. These five amino acids account for 79.4% of the total protein composition. Due to the composition of the protein it has been called GPRP (glycine- and proline-rich protein).

The protein has a clear three-domain structure that can be seen in the hydrophobicity profile shown on Fig. 1A. The protein begins with a short hydrophilic sequence rich in aspartic acid and asparagine. Therefore it does not have the features expected for a signal peptide. This is followed by a repetitive glycine- and proline-rich region, the sequence YPPAG being the most frequent repeat. This region forms a particular domain of 86 amino acids (amino acids 15 to 100) and can be aligned into motifs of a pentapeptide repeated 12 times punctuated by tyrosines (Fig. 2B). The comparison of the whole GPRP protein to the EMBL and PIR databases produced an unpublished carrot sequence (accession numbers: EMBL, X72383 and PIR, S32123) that can be easily aligned to GPRP in at least part of the sequence (Fig. 2B). The similarity between these two proteins is specially significant in the

first half of the sequence including the first hydrophobic stretch and the glycine- and proline-rich and the hydrophobic domains. The general distribution of amino acids indicates that these two proteins may belong to the same family of sequences. The comparison with the databases also produced a number of proteins having proline-rich domains with similarity to the N-terminal repetitive region of GPRP. All these proteins have the characteristic to be associated to membranes with the repeated region located in the cytoplasm.

The GPRP sequence contains a central hydrophobic domain and it finishes with a hydrophilic domain rich in glycine and histidine. The C-terminal domain is also repetitive, it has the XG sequence repeated nine times, X being either Tyr, Phe or His. The motive KFKHGKHG is also repeated twice.

The hydrophobic domain is 17 amino acids long and is flanked by histidine residues. It can be well observed by applying the Kyte and Doolittle algorithm [18] as shown in Fig. 1A. It is the only region of the protein that shows high probability of helical formation according to Chou and Fasman [5]. This distribution of amino acids is also found in proteins associated with membranes described in animal systems and taking part in vesicle trafficking such as synaptobrevin or vesicle-associated membrane proteins [1]. In plants a protein showing a hydrophobic domain that could be a single transmembrane region has been found in wheat induced by pathogen attack [3]. This protein, PWIR1a, has a very short basic N-terminus and a C-terminal glycine-rich region and it was proposed that it may be an integral membrane protein that may increase adhesion of the membranes to the cell wall in the case of pathogen attack.

Cloning and sequence of the homologous gene

An *Arabidopsis thaliana* genomic library was screened with the p9 cDNA as labelled probe and phages showing the strongest hybridization intensity were selected. In order to separate the gene insert from phage arms, the DNA phage was digested by restriction enzymes from phage multicloning region (*Sac*I, *Xho*I and *Sfi*I), and hybridized to p9 cDNA. According to the resulting restriction map, two 7 and 8 kb fragments were subcloned in pBluescript corresponding respectively to the 5′ and 3′ regions up to the SacI site of the cDNA (Fig. 3). A 1.8 kb region containing the zone of homology to the cDNA was sequenced. The sequence has been submitted to the EMBL database and its has the accession number X84315.

At nucleotide and protein levels, the coding region of the gene shows 100% of homology to

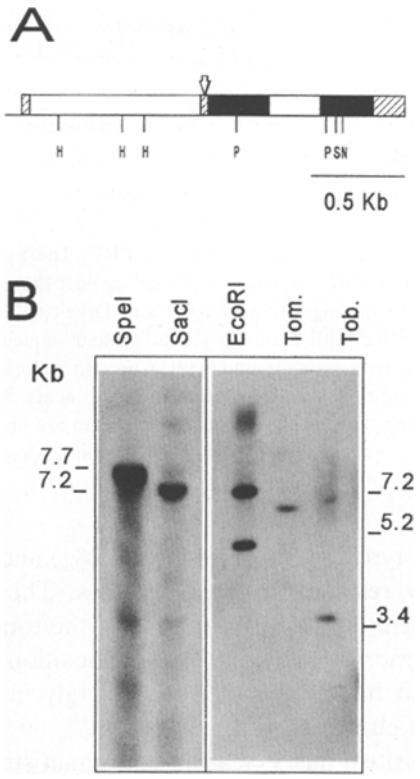

Fig. 3. Genomic structure of the GPRP gene in *Arabidopsis thaliana*. A. Schematic representation of GPRP gene structure. The open boxes indicate intron regions, hatched boxes represent 5′ and 3′ non-coding regions and black boxes indicate the coding domain. Restriction sites are: H, *Hin*dIII; P, *Pst*I; S, *Sac*I; N, *Nco*I. B. Southern blot analysis was carried out on 10 μg of *Arabidopsis thaliana* DNA, digested with *Spe*I, *Eco*RI and *Sac*I. Tobacco and tomato DNA was digested with *Eco*RI and hybridized with the GPRP cDNA probe.

GPRP cDNA (see Fig. 2). This region is interrupted by an intron which has a length of 221 bp. The intron starts with the three nucleotides GTA and ends by CAG which are consensus sequences for introns according to Brown [2] and Rogers [26]. In the 5′ non-coding region the sequence is interrupted by a long intron (778 bp) also flanked by consensus sequences. A schematic drawing of the structure of the GPRP gene is shown on Fig. 3A. Upstream the transcriptional start point the sequence changes its character and it contains putative TATA boxes (see Fig. 2). In general the sequence of the gene is rich in A/T (61.6%), however the base composition is very different in the two exons (57% and 52% GC) when compared to the rest of the sequences. In particular the intron located in the coding region is 31% rich in GC and the promoter region is 30% rich in GC. It has been described for other plant introns that they tend to be rich in AT [30] while the fact that the protein is rich in glycine and proline (whose codons are GGN and CCN respectively) may increase the GC content of the exons.

Genomic organization of the GPRP gene in Arabidopsis

Genomic DNA from *Arabidopsis* was digested with *Eco*RI and *Spe*I which have no recognition site in the cDNA, and with *Sac*I (underlined in Fig. 2) which has a recognition site at position 386 in the cDNA. The DNA was electrophoresed on an agarose gel, transferred onto a nylon membrane and hybridized to the p9 cDNA probe (Fig. 3B). The size of the hybridizing fragments is 7.2 and 5.2 kb for *Eco*RI and 7.8 kb for *Spe*I. The digestion pattern for *Sac*I is a single 7 kb band, this could appear in contrast with the existence of a *Sac*I site in the gene deduced from the sequence. However this band appears as a double band in well resolved gels (result not shown). The Southern blot results indicate that GPRP is probably encoded by either one or two genes in *Arabidopsis*.

The presence of sequences homologous to the GPRP gene in other plant species was analyzed by Southern hybridization. Genomic DNA from two different dicots, tomato (*Lycopersicum esculentum*) and tobacco (*Nicotiana tabacum*), and were digested by *Eco*RI. After electrophoresis and transfer, the corresponding Southern blot was hybridized with *Arabidopsis* GPRP cDNA probe. All species show hybridizable bands that would correlate with the presence of a small number of genes, probably a single gene (Fig. 3B). Similar experiments done on DNA extracted from monocot species such as maize (*Zea mays*) and barley (*Hordeum vulgare*) gave no clear hybridization band (result not shown) that could be attributed to the low sequence similarity due to the large distance existing between the species.

Expression in Arabidopsis *tissues and organs and response to wounding*

The accumulation of mRNA corresponding to the GPRP gene in some *Arabidopsis* tissues as adult roots, rosette leaves, stems, immature siliques, pollinated and unpollinated flowers, mature seeds was analyzed by RNA blot hybridization (Fig. 4). The mRNA band hybridizing to GPRP cDNA is ca. 900 bp long. The accumulation of this mRNA is ubiquitous in all tested tissues except in mature seeds (Fig. 4). The highest mRNA accumulation is found in stems. The result is con-

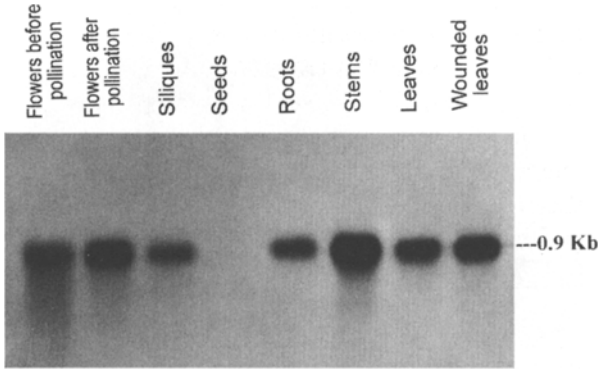

Fig. 4. Detection of *Arabidopsis thaliana* GPRP mRNA in different tissues. Ten μg of total RNAs extracted from tissues were separated on an agarose gel and transferred to membrane. The labelled GPRP cDNA probe was hybridized to the corresponding northern blot.

firmed by *in situ* hybridization (Fig. 5). In *Arabidopsis* stems the GPRP mRNA is accumulated in the vascular tissues and the epidermis. In the young flower the GPRP is highly accumulated in epidermal tissues of the siliques but not in the mature anthers (Fig. 5).

Some genes encoding proteins rich in proline, such as maize HRGP [36], bean PvPRP1 [31] and bean Hyp [6] and rich in glycine such as tomato GRP [32] are induced by wounding 15 min to 2 h after the treatment. This induction was analyzed on GPRP gene expression. In other cases GRP mRNA appear to have its level reduced by wounding [17]. After 4 h of wounding, the level of GPRP mRNA is similar to the unwounded plants (Fig. 4) and therefore GPRP differs from the behaviour of the typical GRPs described in plants.

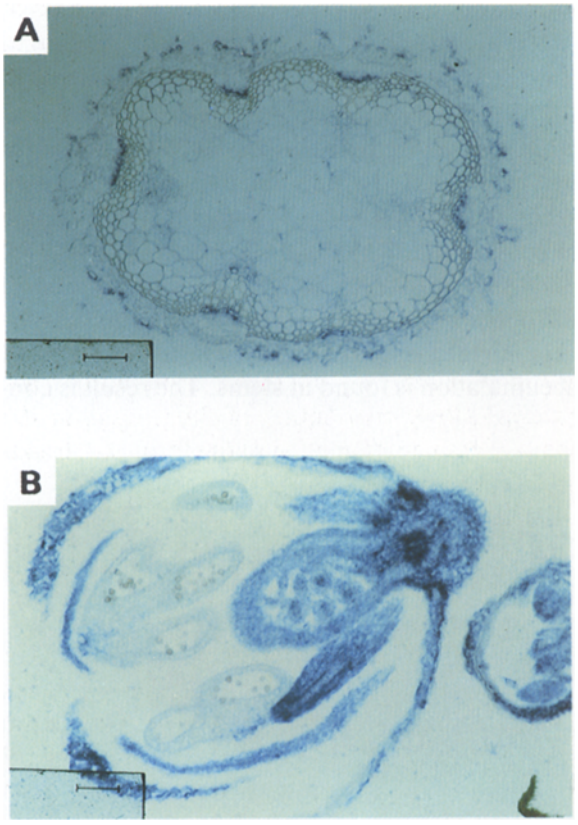

Fig. 5. In situ hybridization of two *Arabidopsis thaliana* organs using the GPRP probe. A. Localization of GPRP mRNA in the stem showing accumulation primarily in the vascular bundles and the epidermis. B. Localization of GPRP mRNA in the flower showing accumulation in the female parts of the flower and in epidermal tissues.

The GPRP protein interacts with membranes in vitro

The high similarity of the repetitive motif of GPRP protein (rich in proline, glycine and tyrosine) with regions of membrane-associated proteins led us to analyze an eventual interaction of GPRP protein with membranes. In addition, the hydrophobicity graph according to Kite and Doolittle (Fig. 1A) shows three domains: the N-terminal domain, 86 amino acids long, which corresponds to the repeated region rich in proline and glycine, and the C-terminal domain, 60 amino acids long, are both hydrophilic regions while the central domain of only 17 amino acids is highly hydrophobic. This is a feature shared by proteins interacting with membranes either forming transmembrane domains or being anchored to the membrane.

In order to characterize a possible interaction of GPRP with membranes, the GPRP cDNA was transcribed and translated *in vitro* in the presence and in the absence of canine microsomal membranes (Fig. 6). The β-lactamase protein was chosen as a control because this protein is translocated and processed but not associated with the membrane.

The translational product of β-lactamase mRNA solubilized in TEA buffer (pH 7.5) is found in the supernatant to a molecular mass of 35 kDa. As expected, the *in vitro* translated β-lactamase in the presence of microsomes was found as a 29 kDa polypeptide in the pellet after ultracentrifugation (Fig. 6A). In contrast, the translational products of GPRP mRNA diluted in TEA buffer were found in the pellet in the presence or in the absence of microsomes showing that these proteins are insoluble in this range of pH. The GPRP proteins are in fact solubilized in a sodium carbonate buffer at pH 11.5 (Fig. 6A). The apparent molecular weight of GPRP proteins is 22 kDa, which is somewhat different from the one deduced from the complete sequence

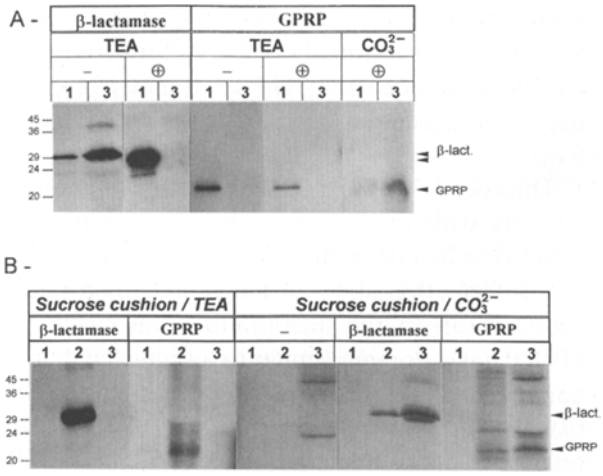

Fig. 6. Characterization *in vitro* of membrane-associated GPRP protein. A. β-lactamase or GPRP mRNAs were translated in the presence (+) or in the absence (–) of canine microsomal membranes. Translational products were diluted in TEA buffer (TEA) or carbonate buffer(CO_3^{2-}), centrifuged and proteins in the pellet (1) and in the supernatant (3) were analyzed by polyacrylamide gel electrophoresis and autoradiography. B. β-lactamase or GPRP mRNAs were translated in the presence of microsomes, diluted in TEA buffer and loaded on a sucrose cushion. After ultracentrifugation, proteins in pellet (1), interphase (2) and supernatant (3) were analyzed by polyacrylamide gel electrophoresis. A control with microsomes in the absence of RNA (–) was carried out in the CO_3^{2-} buffer experiment.

(17.7 kDa). The high content of proline could explain this retardation of migration, already noted for proline-rich proteins [27] that was attributed to the anomalous mobility of basic proline-rich regions in SDS gels.

According to these results, centrifugation through a sucrose cushion was performed in order to purify the microsomes and to separate the insoluble proteins (which are recovered in the pellet) from the proteins associated with membranes (which sediment in the interphase). When the translation products were resuspended in TEA buffer and loaded on a sucrose cushion, β-lactamase sedimented in the interphase with membranes (Fig. 6B). In the same way, the GPRP protein which normally precipitated in this buffer, sedimented in the interphase together with microsomes suggesting that they are associated with membranes (Fig. 6B). The translation product of GPRP in the absence of microsomes and loaded on a sucrose cushion was recovered in the pellet as it was previously found in TEA buffer (result not shown).

In order to identify the kind of interaction of GPRP proteins with membranes, GPRP mRNA was translated in the presence of microsomes and then dissolved in sodium carbonate buffer (pH 11.5). This treatment allows the formation of membrane sheets and consequently, to the release from the vesicles of those proteins which are soluble inside the microsomes or peripherally bounded to membranes. In these conditions, transmembrane proteins sediment in the interphase. As expected, β-lactamase was released from microsomes and found mainly in the supernatant (Fig. 6B). In general, carbonate treatment is not a quantitative process a small proportion of the β-lactamase protein (ca. 20%) not released from microsome vesicles is found in the interphase (see Fig. 6B, β-lactamase, lane 2). However GPRP protein was found in the supernatant and in the interphase. Comparison of sucrose cushions in TEA or in carbonate buffer shows that although GPRP is bound to membranes, the binding was partially removed by the sodium carbonate treatment at pH 11.5. This result suggests that GPRP proteins are anchored to membranes and that at least a percentage (around 40% according to the distribution of the protein in the sucrose gradients) may be in the form of a transmembrane protein.

Discussion

A new class of proline- and glycine-rich proteins from *Arabidopsis thaliana* is described. Its sequence shows at least three well defined domains. After a short hydrophilic stretch, the first domain is repetitive and rich in glycine and proline having a consensus sequence YPPAG. This sequence is similar to a cytosolic domain found in transmembrane proteins of the annexin VII family described in a number of animal systems ranging from invertebrates to man. This type of domains has not been reported in the already described

annexins in plants [33]. In fact, similar repeated motifs like those indicated in Fig. 1B are found in rhodopsin [23], synexin [7] and synaptophysin [35]. This domain shows 63% of similarity to octopus rhodopsin (53 amino acids organized in 7 repeats), 50% to *Dictyostelium* annexin VII (100 amino acids organized in 15 repeats) and a lower degree of homology to rat synaptophysin that contains 49 amino acids organized in 10 repeats having a consensus sequence YGP/QQG. The second domain is hydrophobic and it has the features of a transmembrane domain, in particular a highly hydrophobic core with a high probability of helix formation. Transmembrane domains are frequently flanked by amino acids of basic character such as lysine or arginine, in this case it is defined by histidine residues (see Fig. 1A and B). The protein sequence of GPRP finishes with a repetitive sequence rich in glycine and histidine. The complete protein is basic with a pI of 10.33.

The first two domains have a high similarity to a sequence recorded in carrot (see Fig. 1B) while the third domain is much more dissimilar to this protein. This observation may lead to the conclusion that the two important features of the protein are, first, the general distribution of amino acid residues and, second, the existence of a proline-rich repetitive sequence in the N-terminus of the protein and a hydrophobic domain in the center of the sequence. The existence of similar proteins in other plant species also results from Southern blots on DNA from other dicot species. GPRP appears to be a general type of plant proteins at least in dicotyledonous species.

The GPRP protein is encoded by one or two genes in *Arabidopsis*. The gene corresponding to the p9 cDNA has been cloned and it shows two introns, one in the coding region and another one in the 5′ non-coding region. This 5′ intron is very long (778 bp) for an *Arabidopsis* intron and both its location and length make it an unusual feature of the *GPRP* gene. Whether this is important for the control of its expression or not is a question that analysis of the promoter could decide in the future. The gene appears to have a general expression in the organs of the plant, in particular in vascular and epidermal tissues. The location where the gene has the lowest level of expression is in the seed and mainly in the mature anther. It can be concluded that GPRP is not a protein needed in processes such as cell division or defence.

The need for proteins linking intracellular components with the cell wall through the plasma membrane has been published repeatedly [25, 33, 37]. Information about proteins that could relate cytoskeleton to the membrane is also lacking. GPRP has a domain similar to a sequence that in annexin has been proposed to interact with cytoplasmic components [35]. GPRP is probably not an extracellular protein as it lacks a signal peptide and it has been shown to interact with membranes *in vitro*. The sequence similarity to animal proteins and the anchoring of GPRP in membranes might be taken as an indication of this protein being an element of the plant cytoskeleton-cell wall continuum [37].

Another interesting possibility is that GPRP could take part in the processes of internal cellular transport. In cytoplasm of eukaryotic cells, specific vesicles are responsible for these processes. Proteins with amino acid distribution similar to GPRP have been described in animal systems. These are VAMP (vesicle-associated membrane proteins) or synaptobrevins [1]. In some cases these are integral membrane proteins with a single transmembrane domain and they interact with the proteins responsible for vesicle fusion [1]. The distribution of GPRP mRNA could correlate with this function as it appears to be highly abundant in cells where transport is very active such as phloem and epidermis. These hypotheses are testable by using, for instance, immunological methods. In this sense, GPRP may be a useful tool to explore interesting phenomena in plant cell biology.

Acknowledgements

The authors acknowledge the help of Drs Monique Raynal and Françoise Grellet (Perpignan) for their help in providing data and material from the *Arabidopsis* cDNA sequencing programme.

I.M. was supported by a fellowship from European Community (BRIDGE Programme). The work was supported by grant 094-648 to P.P. from Plant Nacional de Investigación Científica y Técnica. The present work has been carried out within the framework of the Laboratoire Européen Associé CNRS-CSIC de Biologie Moléculaire Végétale.

References

1. Archer III BT, Özçelik T, Jahn R, Francke U, Südhoff TC: Structures and chromosomal localizations of two human genes encoding synaptobrevins 1 and 2. J Biol Chem 265: 17267–17273 (1990).
2. Brown JWS: A catalogue of splice junction and putative point sequences from plant introns. Nucl Acids Res 14: 9549–9559 (1986).
3. Bull J, Mauch F, Hertig C, Rebmann G, Dudler R: Sequence and expression of a wheat gene that encodes a novel protein associated with pathogen defense. Mol Plant-Microb Interact 5: 516–519 (1992).
4. Chen CG, Cornish EC, Clarke AE: Specific expression of an extensin-like gene in the style of *Nicotiana alata*. Plant Cell 4: 1053–1062 (1992).
5. Chou PY, Fasman GD: Prediction of conformation. Biochemistry 13: 222–245 (1974).
6. Corbin DR, Sauer N, Lamb CJ: Differential regulation of a hydroxyproline-rich glycoprotein gene family in wounded and infected plants. Mol Cell Biol 7: 4337–4344 (1987).
7. Döring V, Schleicher M, Noegel AA: *Dictyostelium* Annexin VII (Synexin). J Biol Chem 266: 17509–17515 (1991).
8. Franssen HJ, Nap J-P, Gloudemans T, Stiekema W, van Dam H, Glovers 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).
9. Geli MI, Torrent M, Ludevid D: Two structural domains mediate two sequential events in gamma-zein targeting: protein endoplasmic reticulum retention and protein body formation. Plant Cell 6: 1911–1922 (1994).
10. Gómez J, Sánchez-Martínez D, Stiefel V, Rigau J, Puigdomènech P, Pagès M: A gene induced by the plant hormone abscisic acid in response to water stress encodes a glycine-rich protein. Nature 334: 262–264 (1988).
11. Grosset J, Marty I, Chartier Y, Meyer Y: mRNAs newly synthesized by tobacco mesophyll protoplasts are wound inducible. Plant Mol Biol 25: 485–496 (1990).
12. Haughn G, Somerville C: Sulfonylurea-resistant mutants of *Arabidopsis thaliana*. Mol Gen Genet 204: 430–434 (1986).
13. Höfte H, Desprez T, Amselem H, Chiapello H, Caboche M, Moisan A, Jourjon MF, Charpenteau JL, Berthomieu P, Guerrier D, Giraudat J, Quigley F, Thomas F, Yu DY, Mache R, Raynal M, Cooke R, Grellet F, Delseny M, Parmentier Y, De Marcillac G, Gigot C, Fleck J, Philipps G, Axelos M, Bardet C, Tremousaygue D, Lescure B: An inventory of 1152 expressed sequence tags obtained by partial sequencing of cDNAs from *Arabidopsis thaliana*. Plant J 4: 1051–1061 (1993).
14. Hong JC, Nagao RT, Key JL: Characterization of a proline-rich cell wall protein fene family from soybean. A comparative analysis. J Biol Chem 265: 2470–2465 (1990).
15. José M, Ruiz-Avila L, Puigdomènech P: A maize embryo-specific gene encodes a proline-rich and hydrophobic protein. Plant Cell 4: 413–423 (1992).
16. José M, Puigdomènech P: Structure and expression of genes coding for structural proteins of the plant cell wall. New Phytol 125: 259–282 (1993).
17. Keller B, Sauer N, Lamb CJ: Glycine-rich cell wall proteins in bean: gene structure and association of the protein with the vascular system. EMBO J 7: 3625–3633 (1988).
18. Kyte J, Doolitle RF: A simple method for displaying the hydropathic character of a protein. J Mol Biol 157: 105–132 (1982).
19. Laemmli UK: Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227: 680–685 (1970).
20. Langdale JA: *In situ* hybridization. In: Freeling M, Walbot V (eds) The Maize Handbook, pp. 165–180. Springer-Verlag, New York (1993).
21. Meinkoth J, Wahl G: Hybridization of nucleic acids immobilized on solid support. Anal Biochem 138: 267–284 (1984).
22. Niyogi KK, Fink GR: Two anthranilate synthase genes in *Arabidopsis*: defense-related regulation of the tryptophane pathway. Plant Cell 4: 721–733 (1992).
23. Ovchinnikov YA, Abdulaev NG, Zolotarev AS, Artamonov ID, Bespalov IA, Dergachev AE, Tsuda M: Octopus rhodopsin. Amino acid sequence deduced from cDNA. FEBS Lett 232: 69–72 (1988).
24. Ren R, Mayer BJ, Cicchetti P, Baltimore D: Identification of a ten-amino acid proline-rich SH3 binding site. Science 259: 1157–1161 (1993).
25. Roberts K: Structures at the plant cell surface. Curr Opin Cell Biol 2: 920–928 (1990).
26. Rogers JH: The role of introns in evolution. FEBS Lett 268: 339–343 (1990).
27. Ruiz-Avila L, Ludevid MD, Puigdomènech P: Differential expression of a hydroxyproline-rich cell-wall protein gene in embryonic tissues of *Zea mays* L. Planta 184: 130–136 (1991).
28. Sambrook J, Fritsh EF, Maniatis T: Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1990).

29. Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain terminating inhibitors. Proc Natl Acad Sci USA 74: 5463–5467 (1977).
30. Sauer J, Friedländer K, Gräm-Wicke U: Primary structure, genomic organization and heterologous expression of a glucose transporter from *Arabidopsis thaliana*. EMBO J 9: 3045–305 (1990).
31. Sheng J, D'Ovidio R, Mehdy MC: Negative and positive regulation of a novel proline-rich protein mRNA by fungal elicitor and wounding. Plant J 1: 345–354 (1991).
32. Showalter AM, Zhou J, Rumeau D, Worst SG, Varner JE: Tomato extensin and extensin-like cDNAs: structure and expression in response to wounding. Plant Mol Biol 16: 547–565 (1991).
33. Smallwood M, Keen JN, Bowles DJ: Purification and partial sequence analysis of plant annexins. Biochem J 270: 157–161 (1990).
34. Söllner T, Whiteheart SW, Brunner M, Erdjument-Bromage H, Geromanos S, Tempst P, Rothman JE: SNAP receptors implicated in vesicle targeting and fusion. Nature 362: 318–323 (1993).
35. Südhof TC, Lottspeich F, Greengard P, Mehl E, Jahn R: A synaptic vesicle protein with a novel cytoplasmic domain and four transmembrane regions. Science 238: 1142–1144 (1987).
36. Tagu D, Walker N, Ruiz-Avila L, Burgess S, Martínez-Izquierdo JA, Leguay JJ, Netter P, Puigdomènech P: Regulation of the maize HRGP gene expression by ethylene and wounding. mRNA accumulation and qualitative expression analysis of the promoter by microprojectile bombardment. Plant Mol Biol 20: 529–538 (1992).
37. Wyatt SE, Carpita NC: The plant cytoskeleton-cell-wall continuum. Trends Cell Biol 3: 413–417 (1993).
38. Yamasaki S, Hu Y, Binz T, Kalkuhl A, Kurazono H, Tamura T, Jahn R, Kandel E, Niemann H: Synaptobrevin/vesicle-associated membrane protein (VAMP) of *Aplysia californica*: structure and proteolysis by tetanus toxin and botulinal neurotoxins type D and F. Proc Natl Acad Sci USA 91: 4688–4692 (1994).