Characterization and Expression of Four Proline-Rich Cell Wall Protein Genes in Arabidopsis Encoding Two Distinct Subsets of Multiple Domain Proteins

Thomas J. Fowler, Christine Bernhardt, and Mary L. Tierney
Department of Botany and Agricultural Biochemistry, University of Vermont, Burlington, Vermont 05405

We have characterized the molecular organization and expression of four proline-rich protein genes from Arabidopsis (AtPRPs). These genes predict two classes of cell wall proteins based on DNA sequence identity, repetitive motifs, and domain organization. AtPRP1 and AtPRP3 encode proteins containing an N-terminal PRP-like domain followed by a C-terminal domain that is biased toward P, T, Y, and K. AtPRP2 and AtPRP4 represent a second, novel group of PRP genes that encode two-domain proteins containing a non-repetitive N-terminal domain followed by a PRP-like region rich in P, V, K, and C. Northern hybridization analysis indicated that AtPRP1 and AtPRP3 are exclusively expressed in roots, while transcripts encoding AtPRP2 and AtPRP4 were most abundant in aerial organs of the plant. Histochemical analyses of promoter/β-glucuronidase fusions localized AtPRP2 expression to regions of the root containing root hairs. AtPRP2 and AtPRP4 expression was detected in expanding leaves, stems, flowers, and siliques. In addition, AtPRP4 expression was detected in stipules and during the early stages of lateral root formation. These studies support a model for involvement of PRPs in specifying cell-type-specific wall structures, and provide the basis for a genetic approach to dissect the function of PRPs during growth and development.

Plant cell walls are dynamic and complex structures that contribute to functional differences between cell types during plant growth and development. Pro-rich proteins (PRPs) represent one of five families of structural cell wall proteins that have been identified in higher plants (for review, see Carpita and Gibeaut, 1993; Showalter, 1993; Cassab, 1998). PRPs were first identified as proteins that accumulate in the cell wall in response to physical damage (Chen and Varner, 1985; Tierney et al., 1988) and have subsequently been shown to be temporally regulated during plant development. PRP gene expression is associated with early stages of legume root nodule formation (Fransen et al., 1987; van de Wiel et al., 1990; Wilson et al., 1994), soybean seedling, leaf, stem, and seed coat development (Hong et al., 1989; Kleis-San Francisco and Tierney, 1990; Lindstrom and Vodkin, 1991; Ye et al., 1991), bean seedling growth (Sheng et al., 1991), and with early stages of tomato fruit development (Santino et al., 1997). The spatial pattern of PRP expression is also tightly regulated, as shown by in situ hybridization and reporter gene expression analysis (Wyatt et al., 1992; Suzuki et al., 1993). For example, the soybean SbPRP1 and SbPRP2 transcripts have been localized to sclereids, the inner integument of the seed coat and the epidermal, cortical, and endodermal cells of young seedlings.

Protein localization studies suggest that PRPs may function both in determining cell-type-specific wall structure during plant development and by contributing to defense reactions against physical damage and pathogen infection. Immunohistochemical analyses using antibodies raised against SbPRP2 localized PRP accumulation in soybean to protoxylem cells within the root and xylem and phloem fibers within the stem, indicating that these proteins are critical for maintaining structural integrity of mature tissues (Ye et al., 1991). PRPs may play a similar role during seed development, since seed coat integrity appears to be altered in soybean lines that fail to accumulate these proteins within their cell walls (Nicholas et al., 1993). PRPs are rapidly insolubilized within the cell wall in response to physical damage, treatment with fungal elicitors, and pathogen infection (Kleis-San Francisco and Tierney, 1990; Bradley et al., 1992; Brisson et al., 1994), indicating an active role in plant defense reactions. While the mechanism for PRP insolubilization is not known, there is evidence that this process involves the formation of intermolecular isodityrosine or di-isodityrosine residues through an oxidative coupling reaction (Cooper and Varner, 1984; Fry, 1982; Bradley et al., 1992; Waffenschmidt et al., 1993; Brady et al., 1996).

DNA sequence analysis of PRP genomic and cDNA clones indicates that these proteins can be placed into more than one class based on their primary structure. The first of these classes is characterized by PRP genes isolated from carrot and soybean, which encode tandem copies of the pentapeptide PPVX(K/T), where X is often Y, H, or E (Chen and Varner, 1985; Hong et al., 1987, 1990). SbPRP1 and SbPRP2, two members of this class, have been purified from soybean (Averyhart-Fullard et al., 1988; Kleis-San Francisco and Tierney, 1990; Lindstrom and Vodkin, 1991). Neither of these proteins appears to be highly glycosylated (Datta et al., 1989), and N-terminal sequence analysis has shown that the repetitive unit for both mature proteins is
ProHypVal(Tyr/Glu)Lys. In contrast, a second group of PRP cDNAs predicts two-domain proteins containing a Pro-rich N-terminal domain and a C-terminal domain that lacks Pro-rich or repetitive sequences. This group of PRP genes includes PoPRP1 in bean (Sheng et al., 1991) and TPRP-F1 in tomato (Salts et al., 1991; Santino et al., 1997).

We present the molecular organization and expression patterns of four PRP genes from Arabidopsis. These genes encode two unique classes of PRPs based on DNA sequence identity, repetitive motifs, and domain organization. Northern hybridization and promoter/reporter gene analysis indicate that each of these AtPRP genes has a unique temporal and spatial pattern of expression, suggesting potential functions for these proteins in determining specific extracellular matrix structures throughout plant development.

**MATERIALS AND METHODS**

**Plant Material and Growth Conditions**

For RNA isolation, Arabidopsis ecotype Columbia plants were grown in Promix:vermiculite:perlite (3:1:1) at 19°C using an 8-h light/16-h dark photoperiod, followed by a 12-h light/12-h dark regime to induce flowering. Leaf, stem, and floral tissues were harvested, frozen in liquid nitrogen, and stored at −80°C. Root tissue used for RNA isolation was obtained from plants grown in liquid culture (1× Murashige and Skoog salts, 1× Gamborg’s B5 vitamins, 1% (w/v) Suc, and 10 mM 2-(N-morpholino)-ethanesulfonic acid [MES], pH 6.0) for 10 d under continuous light, frozen in liquid nitrogen, and stored at −80°C. Tissue from transgenic Arabidopsis lines expressing AtPRP/β-glucuronidase (GUS) constructs was obtained for histochemical analysis by growing plants in Magenta boxes on either Murashige and Skoog medium with 1% (w/v) Suc (for vegetative tissues) or in Promix:vermiculite:perlite (3:1:1) (for reproductive tissues) under a light/dark regime as described above.

**Isolation of AtPRP Genomic and cDNA Clones**

AtPRP genomic clones were isolated from a genomic library (Landsberg) constructed in λ-fix (Voytas and Ausubel, 1988) using carrot (pDC16; Chen and Varner, 1985) and soybean (SbPRP1; Hong et al., 1987 and SbPRP2; Datta and Marcus, 1990) PRP genes as probes. Nucleocellular membrane filter lifts of bacteriophage λ plaques (Sambrook et al., 1989) were hybridized with the heterologous probes at 55°C in 6× SSC, 5× Denhardt’s, 0.5% (w/v) SDS, and 100 μg g l−1 single-stranded DNA. The individual probes used corresponded to: AtPRP1, bp 1,475–1,665 (Fig. 2); AtPRP2, bp 176–371 (Fig. 4); AtPRP3, bp 1,265–1,455 (Fig. 3); AtPRP4, bp 1,609–1,867 (Fig. 5). The formamide concentrations were adjusted for each probe to ensure gene-specific conditions: AtPRP1 and AtPRP3, 40% (w/v) formamide; AtPRP2, 50% (w/v) formamide; AtPRP4, 43% (w/v) formamide. Each of the filters was washed at high stringency using the following conditions: AtPRP1 and AtPRP3, 30 min at 65°C in 2× SSC, 0.5% (w/v) SDS; AtPRP2, 30 min at 65°C in 1× SSC, 0.25% (w/v) SDS; AtPRP3, 30 min at 65°C in 1× SSC, 0.25% (w/v) SDS; AtPRP4, 30 min at 65°C in 0.3× SSC, 0.2% (w/v) SDS; AtPRP4, 30 min at 65°C in 2× SSC, 0.5% (w/v) SDS; AtPRP4, 30 min at 65°C in 0.7× SSC, 0.5% (w/v) SDS. Filters were then exposed to x-ray film at −80°C with two intensifying screens.

**AtPRP RNA Analysis**

RNA was extracted from various tissues using a Tris-HCl/SDS/phenol extraction method as described previously (DeVries et al., 1988). Poly(A+) RNA was isolated from total RNA preparations using the PolyATtract kit (Promega), according to the manufacturer’s protocol. Poly(A+) RNA (1.5 μg/lane) was size-fractionated by electrophoresis in 1.4% (w/v) agarose gels containing 1× 3-(N-morpholino)-propanesulfonic acid (MOPS) buffer (Sambrook et al., 1989) and 0.44 M formaldehyde. The RNAs were capillary blotted to nucleocellular membrane sheets using 20× SSC, and the membranes were incubated at 80°C under vacuum for 2 h. RNA blots were hybridized individually at 42°C overnight with 32P-labeled gene-specific probes in 5× SSC, 5× Denhardt’s, 0.5% (w/v) SDS, and 100 mg mL−1 single-stranded DNA. The individual probes used corresponded to: AtPRP1, bp 1,475–1,665 (Fig. 2); AtPRP2, bp 176–371 (Fig. 4); AtPRP3, bp 1,265–1,455 (Fig. 3); AtPRP4, bp 1,609–1,867 (Fig. 5). The formamide concentrations were adjusted for each probe to ensure gene-specific conditions: AtPRP1 and AtPRP3, 40% (w/v) formamide; AtPRP2, 50% (w/v) formamide; AtPRP4, 43% (w/v) formamide. Each of the filters was washed at high stringency using the following conditions: AtPRP1 and AtPRP3, 30 min at 65°C in 2× SSC, 0.5% (w/v) SDS; AtPRP2, 30 min at 65°C in 1× SSC, 0.25% (w/v) SDS; AtPRP3, 30 min at 65°C in 1× SSC, 0.25% (w/v) SDS; AtPRP4, 30 min at 65°C in 0.3× SSC, 0.2% (w/v) SDS; AtPRP4, 30 min at 65°C in 2× SSC, 0.5% (w/v) SDS; AtPRP4, 30 min at 65°C in 0.7× SSC, 0.5% (w/v) SDS. Filters were then exposed to x-ray film at −80°C with two intensifying screens.
**Intron Mapping**

Intron positions within the AtPRP2 and AtPRP4 genomic clones were determined by reverse transcriptase (RT)-PCR (Kawasaki et al., 1988). First-strand cDNA was synthesized using AMV RT (Boehringer Mannheim, Basel) and 50 μg of total RNA isolated from flower tissue. For the RT-PCR reactions, a common first-strand oligonucleotide primer (5′-GATA(A/G)AAACACGATCTTGG 3′) was used with both AtPRP2 and AtPRP4 transcripts. This primer has a single degeneracy that allows it to prime both transcripts at a conserved 3′ of the splice junctions (AtPRP2 [Fig. 4], bp 785–804; AtPRP4 [Fig. 5], bp 640–659). Second-strand DNA synthesis was performed using oligonucleotide primers that allowed specific amplification of either AtPRP2 or AtPRP4 sequences. Reaction conditions used for the RT-PCR were 10 mM Tris-HCl, pH 9.0; 2.5 mM MgCl2; 50 mM KCl; 200 μM each of dATP, dCTP, dGTP, and dTTP; 0.1% (v/v) Triton-X; 6.6% of the flower cDNA (2 μL of 30 μL); 0.5 μM gene-specific oligonucleotide primer (AtPRP2; Fig. 4, bp 188–207); (AtPRP4; Fig. 5, bp 179–198); 1.0 μM degenerate primer (see above); and 2.5 units of Taq DNA polymerase. Reactions were heated to 95°C for 5 min, followed by 50 cycles of 94°C for 30 s, 45°C for 60 s, 72°C for 60 s in a thermal cycler (Perkin-Elmer, Foster City, CA). The PCR fragments generated in this manner were gel-purified and ligated into pT7Blue(R) (Novagen, Madison, WI) for sequencing.

Intron positions within the AtPRP1 and AtPRP3 genomic clones were determined by comparison of the genomic sequences to the sequences of corresponding partial or full-length cDNAs isolated from the λ-PRL2 library.

**Predicted Signal Peptide Cleavage Sites**

Cleavage sites for the signal peptide were predicted using the matrix method as described by von Heijne (1986).

**Construction of AtPRP Promoter/GUS Lines**

5′-Flanking sequences for AtPRP2 (2.5 kb), AtPRP3 (1.5 kb), and AtPRP4 (1.4 kb) were fused to the bacterial uid gene encoding GUS (vector pBI101; Jefferson et al., 1987) and transformed into Arabidopsis ecotype Columbia (AtPRP2 and AtPRP3) or Landsberg erecta (AtPRP4), respectively, using an in planta transformation method (Bechtold et al., 1993). Kanamycin-resistant lines were identified, allowed to set seed, and T2 transgenic plants were grown and analyzed for GUS expression.

**Histochemical GUS Staining**

Histochemical staining of plant tissue for GUS activity was performed as described by Jefferson et al. (1987). Samples were immediately placed in substrate solution (50 mM sodium phosphate, pH 7.5, 15% [v/v] methanol, 2 mM 5-bromo-4-chloro-3-indolyl-glucuronide, and 0.05% [v/v] Tween 20), vacuum infiltrated for 2 min at 85 kPa, and incubated at 37°C for 8 to 18 h. Removal of pigments was achieved by several washes in 50% to 70% (v/v) ethanol. Samples were analyzed under a stereomicroscope (model 2000, Zeiss, Jena, Germany) and pictures were taken on Kodak 25 film.

**RESULTS**

Screening of an Arabidopsis (Landsberg erecta) genomic library with carrot and soybean PRP probes (Chen and Varner, 1985; Suzuki et al., 1993) resulted in the identification of four distinct genomic clones encoding Pro-rich proteins (AtPRPs). cDNA clones corresponding to each of these genomic clones were isolated from a λ-PRL2 library (Newman et al., 1994) obtained through the Arabidopsis Biological Resource Center (Ohio State University, Columbus). DNA sequencing of genomic and representative cDNA clones corresponding to these isolates indicated that the PRP genes in Arabidopsis can be separated into two classes (AtPRP1 and AtPRP3 versus AtPRP2 and AtPRP4) based on DNA sequence homology, domain structure, and predicted amino acid sequence.

Southern hybridization showed that AtPRP gene sequences within each class hybridized well with each other and poorly with clones encoding other PRPs. For example, at high stringency AtPRP1 hybridizes with AtPRP3 but not with AtPRP2, AtPRP4, or PRP gene sequences from other plant species (data not shown). Figure 1 illustrates the pattern of restriction fragments that are detected when...
coding region probes for AtPRP1 and AtPRP2 were used in Southern hybridizations with EcoRI-digested genomic DNA. These fragments were analyzed by DNA sequencing and were shown to correspond to AtPRP3 (7 kb), AtPRP1 (3.8 kb), AtPRP4 (3.3 kb), and AtPRP2 (1.5 kb), indicating that each of these genes is represented as single copy within the Arabidopsis genome.

Structure of AtPRP1 and AtPRP3

The DNA sequence of the AtPRP1 and AtPRP3 genomic and cDNA clones (Figs. 2 and 3) predicts Pro-rich proteins containing a signal peptide followed by two domains and having molecular masses of 36.5 and 34.4 kD, respectively. The N-terminal domain of AtPRP1 consists of 13 imperfect copies of the amino acid repeat KPTLSPPVYT. This decapeptide motif, which contains the pentapeptide motif PPVX(K/T) that is characteristic of other PRPs, is found five times within the N-terminal domain of AtPRP3 as part of a longer repeat unit, KPTIPPPVYTPPVYKPTLSPPVYT. The C-terminal domain of both of these proteins, while rich in P, Y, and K, is unique in sequence. While the amino acid sequence of AtPRP1 and AtPRP3 is highly conserved (76% amino acid identity overall), the C-terminal domain of these proteins was found to exhibit the greatest sequence identity (Table I).

Comparison of cDNA and genomic clones showed that AtPRP1 and AtPRP3 each contain an intron within their second domain. In each case, the consensus GT/AC intron splice donor and acceptor sites are present at the intron/exon border. The relatedness of these two genes is emphasized by the conserved position of the intron that interrupts a Gly codon within the second domain of the open reading frame (ORF) (Figs. 2 and 3).

Figure 3. DNA and predicted amino acid sequence of AtPRP3. The ORF for AtPRP3 and the predicted amino acid sequence is presented in uppercase, while upstream, downstream, and intron genomic sequences are presented in lowercase. The predicted cleavage site for the signal peptide is indicated with an arrowhead. A potential TATA box and polyadenylation signal are underlined.

Figure 2. DNA and predicted amino acid sequence of AtPRP1. The ORF for AtPRP1 and the predicted amino acid sequence are presented in uppercase, while upstream, downstream, and intron genomic sequences are presented in lowercase. The predicted cleavage site for the signal peptide is indicated with an arrowhead. A potential TATA box and polyadenylation signal are underlined.

copies of the amino acid repeat KPTLSPPVYT. This decapeptide motif, which contains the pentapeptide motif PPVX(K/T) that is characteristic of other PRPs, is found five times within the N-terminal domain of AtPRP3 as part of a longer repeat unit, KPTIPPPVYTPPVYKPTLSPPVYT. The C-terminal domain of both of these proteins, while rich in P, Y, and K, is unique in sequence. While the amino acid sequence of AtPRP1 and AtPRP3 is highly conserved (76% amino acid identity overall), the C-terminal domain of these proteins was found to exhibit the greatest sequence identity (Table I).

Comparison of cDNA and genomic clones showed that AtPRP1 and AtPRP3 each contain an intron within their second domain. In each case, the conserved GT/AC intron splice donor and acceptor sites are present at the intron/exon border. The relatedness of these two genes is emphasized by the conserved position of the intron that interrupts a Gly codon within the second domain of the open reading frame (ORF) (Figs. 2 and 3).
Structure of AtPRP2 and AtPRP4

AtPRP2 and AtPRP4 constitute a second, novel class of PRPs in Arabidopsis. The genomic clones encoding these PRPs predict proteins with molecular masses of 32.6 and 46 kD, respectively. Their primary structure consists of a signal peptide followed by a unique, non-repetitive domain and ending with a basic domain containing Pro-rich repeats (Figs. 4 and 5). Like AtPRP1 and AtPRP3, the non-PRP-like domain of these proteins shares the highest degree of amino acid identity (Table I). Within the C-terminal domains of AtPRP2 and AtPRP4, the PRP consensus motif PPVX(K/T) is present only degenerately as PPV and P(V/I)YK. Instead, AtPRP2 contains nine copies of the amino acid motif PIYKPPV (Fig. 4), while AtPRP4 contains eight imperfect copies of the sequence PPPKIEHPPPPVYK (Fig. 5). In addition, AtPRP2 and AtPRP4 contain four and six copies, respectively, of the Cys-containing motif KKPCPP (Figs. 4 and 5).

RT-PCR was used to identify the position of a single intron within AtPRP2 and AtPRP4. The intron was present at a conserved position within the non-Pro-rich domain of these proteins and was flanked by consensus GT/AC intron donor and acceptor sites (Figs. 4 and 5).

Table I. Comparison of the AtPRP protein sequences

For each of the AtPRP proteins, the predicted signal sequences are underlined, the N-terminal domain is represented by uppercase letters, and the C-terminal domain is represented by lowercase letters. Dashes represent introduced sequence gaps, and dots indicate the identity of the amino acid sequence between the predicted PRPs.

<table>
<thead>
<tr>
<th>AtPRP1</th>
<th>MAITRASFAICILLSLATHLIAIAVTADYYPASSPPPYSPYTPVKNKTLPPPPPYTPHPVHKTPLPPPPVYTPHPVHKTPLSLPPPVY</th>
</tr>
</thead>
<tbody>
<tr>
<td>AtPRP3</td>
<td>.....................................................</td>
</tr>
<tr>
<td>AtPRP1</td>
<td>.....................................................</td>
</tr>
<tr>
<td>AtPRP3</td>
<td>.....................................................</td>
</tr>
<tr>
<td>AtPRP1</td>
<td>.....................................................</td>
</tr>
<tr>
<td>AtPRP3</td>
<td>.....................................................</td>
</tr>
<tr>
<td>AtPRP2</td>
<td>..................................................................</td>
</tr>
<tr>
<td>AtPRP4</td>
<td>..................................................................</td>
</tr>
<tr>
<td>AtPRP2</td>
<td>..................................................................</td>
</tr>
<tr>
<td>AtPRP4</td>
<td>..................................................................</td>
</tr>
<tr>
<td>AtPRP2</td>
<td>..................................................................</td>
</tr>
<tr>
<td>AtPRP4</td>
<td>..................................................................</td>
</tr>
<tr>
<td>AtPRP2</td>
<td>..................................................................</td>
</tr>
<tr>
<td>AtPRP4</td>
<td>..................................................................</td>
</tr>
</tbody>
</table>

Characterization of Four Novel Pro-Rich Protein Genes in Arabidopsis 1085

www.plantphysiol.org on July 20, 2017 - Published by Downloaded from Copyright © 1999 American Society of Plant Biologists. All rights reserved.
Control hybridizations were used to demonstrate the gene-specific nature of the probes under these hybridization conditions (Fig. 6B). Since \textit{AtPRP2} and \textit{AtPRP4} encode transcripts of different sizes, we analyzed the same mRNA preparation used in Figure 6A for cross-reactivity. In contrast, \textit{AtPRP1} and \textit{AtPRP3} encode transcripts that cannot be distinguished by size. Therefore, we compared the cross-reactivity of each probe using in vitro-transcribed sense RNAs for each gene. In all cases, the probes were found to hybridize specifically to a single transcript.

Since developmentally regulated changes in cell wall structure may be critical for normal growth and differentiation processes, we characterized the temporal and spatial expression patterns of the \textit{AtPRP} genes using promoter/reporter gene constructs in transgenic plants. \footnote{Fowler et al., 1999}

\footnote{Fowler et al., 1999} - Flanking sequences for \textit{AtPRP2}, \textit{AtPRP3}, and \textit{AtPRP4} were fused to the bacterial \textit{uid} gene encoding GUS and these constructs were transformed into Arabidopsis using vacuum infiltration (Bechtold et al., 1993). A minimum of four independent T2 transgenic lines for each of the \textit{AtPRP} promoter/GUS constructs were analyzed for their patterns of GUS expression.

\textit{AtPRP3}/GUS expression was exclusively detected in roots during plant development, which is consistent with data obtained using northern hybridization. Shortly after

\textbf{Figure 4.} DNA and predicted amino acid sequence of \textit{AtPRP2}. The ORF for \textit{AtPRP2} and the predicted amino acid sequence are presented in uppercase, while upstream, downstream, and intron genomic sequences are presented in lowercase. The predicted cleavage site for the signal peptide is indicated with an arrowhead. A potential TATA box and polyadenylation signal are underlined.

\textbf{Figure 5.} DNA and predicted amino acid sequence of \textit{AtPRP4}. The ORF for \textit{AtPRP4} and the predicted amino acid sequence is presented in uppercase while upstream, downstream, and intron genomic sequences are presented in lowercase. The predicted cleavage site for the signal peptide is indicated with an arrowhead. A potential TATA box and polyadenylation signal are underlined.
germination, expression was found in root epidermis and root hairs localized around the transition zone marking the root/shoot junction. With further growth of the root, GUS expression could be detected in root epidermis and root hairs along the length of the root and was the most intense in the root zone forming new hairs (Fig. 7a). No AtPRP3/GUS expression was observed in the root tip. In older seedlings, AtPRP3 expression continued to be restricted to the regions of the main root active in root hair development, and this pattern of expression was reiterated in lateral roots (Fig. 7b).

Several aspects of AtPRP2 and AtPRP4 gene expression were found to be similar during plant development. In young seedlings, AtPRP4/GUS expression was detected in the hypocotyl, cotyledons (Fig. 7, c and d), and rosette leaves. Staining was most intense in expanding leaves and gradually disappeared with age (Fig. 7, e and m). After transition to the reproductive phase of growth, AtPRP4 was found to be expressed in stems, cauline leaves, and sepals (Fig. 7, h and n). Similar patterns of expression were observed for AtPRP2/GUS in these tissues (data not shown). The timing of AtPRP2 and AtPRP4 expression during anther development was temporally controlled, with AtPRP2/GUS transcription associated with anthers of closed flowers (Fig. 7o), while AtPRP4 expression was only detected in anthers of open flowers (Fig. 7i). Later in development, both genes were found to be expressed in pedicels of developing siliques, nectaries, and along the length of maturing siliques (Fig. 7, j–l).

AtPRP4/GUS was found to be uniquely expressed in stipules of both rosette and cauline leaves (Fig. 7, e and h), the stigma surface of opening flowers (Fig. 7i), emerging lateral roots, and in spaced intervals along the root that may represent initials for lateral root development (Fig. 7g).

DISCUSSION

We have isolated and characterized genomic and cDNA clones encoding four Pro-rich cell wall proteins from Arabidopsis. The expression of each of these genes is temporally and spatially regulated during plant development and targets cell types and organs where they may function to determine cell wall structure. In addition, AtPRP2 and AtPRP4 represent novel members of this gene family of extracellular matrix proteins.

Structure of the AtPRPs

The overall structure of AtPRP1 and AtPRP3 consists of a signal sequence, an N-terminal PRP-like domain, and a highly charged, non-repetitive C terminus. This structural organization is similar to that predicted for a number of other cell wall proteins, including an extensin-like protein (ISG) from Volvox (Ertl et al., 1992), several AGP-like proteins (TTS) from tobacco (Cheung et al., 1995), and three PRP-like proteins from bean, tomato, and tobacco (Salts et al., 1991; Sheng et al., 1991; Chen et al., 1993; Santino et al., 1997). In both the Volvox and tobacco systems, the interac-
tion of these proteins with other components within the extracellular matrix was found to be critical for proper development. Disruption of the interaction between ISG and other matrix components resulted in the inability of cells to complete gamete formation, while inhibition of TTS expression using antisense or sense co-suppression trans-

**Figure 7.** Histochemical localization of AtPRP expression using AtPRP promoter/GUS analysis. a and b, AtPRP3/GUS. a, 2-d-old seedling; b, 8-d-old seedling. c to l, AtPRP4/GUS. c, 1-d-old seedling; d, 2-d-old seedling; e, 23-d-old seedling; f, detail stipules; g, detail roots; h, immature inflorescence; i, flower cluster; j, young silique; k, detail nectaries; l, maturing silique. m to o, AtPRP2/GUS. m, 23-d-old seedling; n, immature inflorescence; o, flower cluster.
genic lines resulted in a reduced rate of pollen tube growth (Cheung et al., 1995). These studies support the potential importance of matrix interactions between AtPRP1/AtPRP3 and other components within the cell wall, and indicate that such interactions may be critical for root or root hair development in Arabidopsis.

AtPRP2 and AtPRP4 represent a second, newly described subset of PRPs in higher plants. These genes encode proteins containing a signal sequence, an N-terminal domain that is non-repetitive, and a PRP-like C-terminal region. The predicted amino acid sequence of these two proteins indicates that they are highly charged polypeptides. The PRP-like, repetitive motifs present within the C-terminal domain are more degenerate than those observed for AtPRP1 and AtPRP3 and are found to border a Cys-rich motif (KKPCPP). While Cys-rich motifs have been observed in other two-domain PRPs, they have previously been found within the non-repetitive domain of these proteins (Sheng et al., 1991; Chen et al., 1993; Wu et al., 1993).

Comparison of the nucleotide sequence of the four AtPRP genes presented here indicates that these genes are likely to have evolved from two gene duplication events. This is supported by the conserved position of a single intron within the unique domain of each of the AtPRP genes and the high degree of amino acid and nucleotide identity observed in the non-Pro-rich domains. Sequence gaps between either AtPRP1 and AtPRP3 or AtPRP2 and AtPRP4 are flanked by the repetitive motifs PPVX(K/T) or PTL(P/S), suggesting a possible function for these sequences in recombination (Table I). In soybean, SbPRP1 and SbPRP2 variants differing in molecular mass and containing multiple deletions or additions of the pentapeptide PPVXK have been identified (Schmidt et al., 1994). This type of variation suggests that recombination within these sequences encoding the repetitive, Pro-rich motifs characteristic of PRPs may provide a mechanism for generating new structural cell wall proteins.

Possible Functions for AtPRPs in Determining Cell Wall Structure

PRPs are thought to contribute to the cell wall structure of specific cell types based both on their patterns of gene expression during plant development and their ability to associate with and become cross-linked to components within the cell wall (for review, see Showalter, 1993). The predicted pIs of the AtPRPs range between 9.6 and 10, suggesting that they may interact with the acidic pectin network within the cell wall. In addition, the localization of Cys-rich elements with the Pro-rich domain of AtPRP2 and AtPRP4 may facilitate disulfide bond formation between these PRPs themselves and/or other proteins within the plant extracellular matrix. Further analysis of these novel PRPs may provide clues about the relationship between structural matrix protein function and cellular aspects of growth and development.

Tyr and Lys are an abundant amino acids in both PRPs and extensins (a second family of Hyp-rich structural cell wall proteins) and have been implicated as the substrate for the peroxidase-mediated insolubilization of PRPs in soybean (Kleis-San Francisco and Tierney, 1990; Bradley et al., 1992; Brisson et al., 1994) and in the cross-linking of extensins within the cell wall of suspension-cultured cells (Brady et al., 1996; Schnabelrauch et al., 1996). An extensin-specific peroxidase has been identified in tomato cell suspension cultures, and the substrate for this enzyme has tentatively been identified as Val-Tyr-Lys. Interestingly, two soybean PRPs containing this motif were not substrates for this enzyme in vitro (Schnabelrauch et al., 1996). Pectin/extensin cross-links have been identified in cotton cell walls (Qi et al., 1995) and are thought to occur through either a 3,6-linked galactan or a ferulated sugar/amino acid cross-link (Keegstra et al., 1973; Brownleader and Dey, 1993). Thus, the insolubilization of the AtPRPs may involve either protein/protein or protein/carbohydrate linkages within the cell wall, and further investigation will be needed to determine if and how the cross-linking of these proteins within the wall contributes to the structure of the extracellular matrix.

As more structural cell wall proteins are characterized, it appears that extensins and PRPs may be considered members of a superfamily of Pro/Hyp-rich cell wall proteins, as has been suggested previously (Kieliszewski and Lamport, 1996). Several structural features of the AtPRP gene family support this suggestion. Database analysis indicated that AtPRP1 and AtPRP3 share 42% identity with a predicted extensin-like protein in Nicotiana alata. In addition, AtPRP1 and AtPRP3 contain multiple Ser-Pro-Pro repeats throughout their N-terminal domain and a single Ser-Pro sequence, both of which are reminiscent of the Ser-Hyp repetitive motif characteristic of many extensin proteins.

The potential relationship between PRP and extensin protein sequences is also apparent when a repetitive unit within the AtPRP4 gene product (FPKIEHPPPVVPVYK) is compared with a known peptide sequence found within a sugar beet extensin (SOOVHEYPOOTOVYK), where O represents Hyp. However, it will be necessary to gain a better understanding of the sequences critical for extensin and PRP function within the cell wall before we can interpret whether this level of sequence conservation represents the remnants of a common evolutionary history or simply reflects conserved functional motifs required for the interaction of HRGPs and PRPs with other extracellular matrix components.

Developmental Regulation of AtPRP Gene Expression

Each of the AtPRP genes was differentially expressed. AtPRP1 and AtPRP3 transcripts were only detected in root tissue. This was supported by histochemical promoter/GUS analysis, which localized expression of AtPRP3 to the regions of the root producing root hairs. GUS expression was not observed in older parts of the root or in the root tip, indicating that AtPRP3 may play an important role during root hair formation. Two extensin genes with root-hair-specific expression patterns have recently been identified in tomato and bean (Arsenijevic-Malisimovic et al., 1997; Bucher et al., 1997), suggesting that at least two families of structural proteins may dictate aspects of cell
wall architecture necessary for the initiation and growth of root hairs in different plant species.

Northern hybridization analysis of AtPRP2 and AtPRP4 gene expression indicated that transcripts for these cell wall protein genes are most abundant in leaf, stem, and reproductive tissue. Analysis of AtPRP2/GUS and AtPRP4/GUS expression patterns supported these observations and showed that both of these genes are highly expressed in the hypocotyl and cotyledons of young seedlings, immature rosette and cauline leaves, stems, sepals, anthers, siliques, and in nectaries at the silique-pedicel junction. AtPRP4 was also found to be expressed uniquely in the stipules and stigma of opening flowers. Furthermore, AtPRP4 may play an important role in establishing a cell wall matrix necessary for the initiation and early stages of lateral root development, as its expression was observed in spaced intervals along the root and at junctions between laterals and the main root. A similar pattern of expression has been observed for a tobacco extensin gene (Keller and Lamb, 1989). However, these two genes differ in their expression pattern, as AtPRP4 is strictly associated with the early steps of lateral root initiation, while the tobacco extensin gene is also associated with lateral tip growth. Analysis of the regulation of AtPRP4 expression in association with the hormonal regulation of lateral root development will provide additional insight into the possible relationship between AtPRP4 function and lateral root growth.

In summary, we have characterized the structure and expression of four members of the PRP gene family in Arabidopsis. These genes predict cell wall proteins that fall into two classes based on domain structure, sequence identity, intron location, and patterns of gene expression during plant development. In addition, two of these proteins (AtPRP2 and AtPRP4) represent a newly described class of structural cell wall proteins whose function may involve novel interactions within the extracellular matrix and possibly with proteins within the cell membrane. Analysis of the protein products of these genes using genetic and biochemical approaches readily available in Arabidopsis will provide an opportunity to dissect the mechanism(s) by which PRPs interact with other cell wall polymers in distinct cell types during plant development and in response to environmental stimuli.

ACKNOWLEDGMENTS

We thank Keith Davis for providing the Arabidopsis genomic library, Doreen Ware for assistance with DNA sequence analysis, and Gary Ward and Eunice Froeliger for helpful discussions. Arabidopsis cDNA libraries were obtained from the Arabidopsis Biological Resource Center.

Received April 6, 1999; accepted August 6, 1999.

LITERATURE CITED

Kawasaki ES, Clark SS, Coyne MY, Smith SD, Champlin R, Witte ON, McCormick FP (1988) Diagnosis of chronic myeloid and acute lymphocytic leukemias by detection of leukemia-


