The Arabidopsis Profilin Gene Family

Evidence for an Ancient Split between Constitutive and Pollen-Specific Profilin Genes

Shurong Huany, John M. McDowell, Michael J. Weise, and Richard B. Meagher*

Department of Genetics, University of Georgia, Athens, Georgia 30602 (R.B.M., M.J.W.); Lawrence Berkeley Laboratory, University of California, Berkeley, California 94720 (S.H.); and Department of Biology, University of North Carolina, Chapel Hill, North Carolina 27599 (J.M.M.)

Profilin is a ubiquitous eukaryotic protein that regulates the actin cytoskeleton and recently has been identified as a potent allergen in pollen. We examined the profilin gene family in the model plant, Arabidopsis thaliana, and found that it contained approximately 8 to 10 members. Four distinct profilin sequences, three cDNAs, PRF1, PRF2, and PRF3, and two genomic clones, PRF1 and PRF4, were isolated and characterized. These genes encoded four distinct profilin isoforms of 131 to 134 amino acids. Northern and reverse-transcriptase polymerase chain reaction analyses demonstrated that Arabidopsis PRF1 was expressed in all major plant organs, whereas PRF4 was specifically expressed in mature pollen. Gene trees constructed from amino acid sequence data revealed the presence of two ancient, distinct profilin gene classes in plants. PRF4 was in a class with previously identified pollen-specific profilins from monocot and dicot species. PRF1, PRF2, PRF3, and a distant dicot sequence formed a separate novel class, suggesting an ancient separation of plant profilins based on regulation and perhaps function. The coevolution of plant actin and profilin classes with similar patterns of expression is discussed. The similarity of plant, fungal, protist, insect, and nematode profilins and their extreme divergence from the vertebrate profilins has striking implications for the evolution of fungal-spore- and plant-pollen profilins as allergens.

The actin cytoskeleton of eukaryotic cells undergoes rapid and dynamic rearrangements in response to external stimuli and during the cell cycle. The reorganization of the cytoskeleton is highly regulated both temporally and spatially through a wide spectrum of ABPs (Stossel et al., 1985; Pollard and Cooper, 1986; Hartwig and Kwiatkowski, 1991). These ABPs cross-link actin filaments into loose networks or tight bundles, cap the ends or block the sides of the filaments, sever the filaments, or sequester actin monomers from polymerization. Specific interactions between ABP and either monomeric or polymerized forms of actin control the three-dimensional organizations and cellular locations of the actin cytoskeleton, which in turn affects the functions of the actin cytoskeleton in the cell.

Profilin is a low-molecular-mass (12-15 kD), ubiquitous actin-binding protein (Bajer and Molè-Bajer, 1971; Pollard and Cooper, 1986; Dower et al., 1988; Haarer and Brown, 1990) whose three-dimensional structure has recently been determined (Schutt et al., 1993; Vinson et al., 1993). Although its precise role in vivo is not well understood, profilin appears to be a multifunctional protein (Haarer et al., 1990). It exerts both positive and negative effects on actin polymerization (Thiriot and Mitchison, 1993). Profilin binds to actin monomers, forming a 1:1 profilin-actin complex that sequesters actin monomers and prevents their polymerization. In contrast, profilin also catalyzes the exchange of ATP for ADP on recycled monomeric G-actin. ATP-actin has a lower critical concentration, so this activity promotes actin polymerization. Profilin also reversibly binds to the membrane phospholipid PIP_2, resulting in the release of actin from the profilactin complex (Lassing and Lindberg, 1988) and in the protection of PIP_2 from the hydrolytic action of phospholipase C (Gold-schmidt-Clermont et al., 1990). This is believed to link the remodeling of the actin cytoskeleton to the phosphatidylinositol cycle-mediated signal transduction pathway, which responds to extracellular stimuli (Aderem, 1992; Machesky and Pollard, 1993).

Profilin was first isolated from calf spleen (Carlsson et al., 1977), and homologs now have been identified and characterized in a number of organisms throughout the four eukaryotic kingdoms. Yeast contains a single profilin gene (Magdolen et al., 1988), the disruption of which leads to yeasts cells displaying a temperature-sensitive phenotype with an abnormal morphology and a dramatically altered actin cytoskeleton system (Haarer et al., 1990). Some organisms have multiple profilin sequences (Binette et al., 1990; Machesky et al., 1990; Haugwitz et al., 1991; Mach-esky and Pollard, 1993; Staiger et al., 1993), but their significance is for the most part unknown. Although there are multiple copies of profilin sequences in the human genome, many of them may be pseudogenes, and only a single profilin isoform has been identified (Kwiatowski and Bruns, 1988). Differential expression and/or developmental regulation of profilin has been demonstrated in...
Dictyostelium (Haugwitz et al., 1991), Physarum (Binette et al., 1990), Drosophasis (Cooley et al., 1992), and mammals (Babcock and Rubenstein, 1993), and this will be briefly reviewed with the discussion of our data (below).

Profilin cDNAs expressed primarily in pollen in maize (Zea mays L.) and tobacco (Nicotiana tabacum) (Staiger et al., 1993; Mittermann et al., 1995) and in the leaves and root nodules of beans (Phaseolus vulgaris) (Vidali et al., 1995) have been partially characterized. Other monocot and dicot profilin cDNAs have been isolated from pollen libraries (Valenta et al., 1991), and profilin proteins have been isolated from vegetative organs (Ebner et al., 1991; Vallier et al., 1992), but the specificity of their expression patterns and the functional significance of having multiple plant profilins are not known. Plant profilins have been shown to bind plant and animal actin in vitro (Valenta et al., 1993; Giehl et al., 1994; Ruhlandt et al., 1995), and their role in transmitting signals from the phosphoinositide system to the cytoskeleton has been discussed (Drobak, 1993; Drobak et al., 1994). Much of the interest in plant profilins, including those from both pollen and vegetative organs, results from their activity as potent human allergens (Valenta et al., 1991), which elicit histamine release and a strong IgE response.

The isolation and initial characterization of the profilin multigene family in the model plant Arabidopsis thaliana are reported herein. The family contains as many as 10 gene members. We characterized four novel sequences and showed the presence of two divergent classes of profilin genes that displayed fundamentally different expression patterns. The data suggested an ancient split between constitutive and pollen-specific profilin genes and coevolution of profilins as allergens.

The isolation and initial characterization of the profilin multigene family in the model plant Arabidopsis thaliana are reported herein. The family contains as many as 10 gene members. We characterized four novel sequences and showed the presence of two divergent classes of profilin genes that displayed fundamentally different expression patterns. The data suggested an ancient split between constitutive and pollen-specific profilin genes and coevolution of profilins as allergens.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
<th>Region of Homology (codons)</th>
<th>Degeneracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRF1S</td>
<td>5'-ATGWSNTGGCARACNTAYGTNGA-3'</td>
<td>1-8</td>
<td>Figure 2</td>
</tr>
<tr>
<td>PRFp22S</td>
<td>5'-TCNGCGCNCATNGTNGGNYA-3'</td>
<td>22-28</td>
<td>8,192</td>
</tr>
<tr>
<td>PRF-71S</td>
<td>5'-AARTAYATGGTANCARGCGNCA-3'</td>
<td>71-78</td>
<td>384</td>
</tr>
<tr>
<td>PRF113S</td>
<td>5'-GGNCARTGYYAYCNGTNGTNGA-3'</td>
<td>113-120</td>
<td>2,048</td>
</tr>
<tr>
<td>PRF121S</td>
<td>5'-GNGARMGYNTNGGNGAYTA-3'</td>
<td>121-127</td>
<td>4,096</td>
</tr>
<tr>
<td>PRF1A</td>
<td>5'-TCNACRTAGNTGCGCNSAWCAT-3'</td>
<td>1-8</td>
<td>2,048</td>
</tr>
<tr>
<td>PRF22A1-4</td>
<td>5'-CRTCNTNGCRCRTRRATNGCGNC-3'</td>
<td>22-29</td>
<td>4,096</td>
</tr>
<tr>
<td>PRF29A1-4</td>
<td>5'-CTTYGNGCCCAACRTCCNCTR-3'</td>
<td>29-36</td>
<td>512</td>
</tr>
<tr>
<td>PRF71A</td>
<td>5'-YTNCYCTGYNTNYACATRTAYNT-3'</td>
<td>71-78</td>
<td>2,048</td>
</tr>
<tr>
<td>PRFp98A</td>
<td>5'-ATNCCNAYNACAGNCGGCGNC-3'</td>
<td>98-105</td>
<td>16,384</td>
</tr>
<tr>
<td>PRF113A</td>
<td>5'-YTNCACNACANRRCAYTCNGC-3'</td>
<td>113-120</td>
<td>16,384</td>
</tr>
<tr>
<td>PRF120A</td>
<td>5'-RTARTCCNCCNARCKYTCNCA-3'</td>
<td>120-127</td>
<td>32,768</td>
</tr>
<tr>
<td>PRF123A</td>
<td>5'-CCYTGNTCTCNANRARTATGCNNC-3'</td>
<td>123-130</td>
<td>16,384</td>
</tr>
<tr>
<td>PRF1-3'S15Smal</td>
<td>5'-GCCCGCGCCGACCCACGTTCCAGGTCCTTCTC-3'</td>
<td>3'UTR</td>
<td>1</td>
</tr>
<tr>
<td>PRF4-3'S1</td>
<td>5'-CTTCCTACATGGTCCACCCACCTAC-3'</td>
<td>3'UTR</td>
<td>1</td>
</tr>
<tr>
<td>PRF4-3'S1</td>
<td>5'-AATCCCAAATCCGTTATGAGGAGA-3'</td>
<td>3'UTR</td>
<td>1</td>
</tr>
<tr>
<td>PRF4-3'S1</td>
<td>5'-TCACTCGATCCCAAAGCGAAGAGAAGTAAACGCT-3'</td>
<td>3'UTR</td>
<td>1</td>
</tr>
<tr>
<td>PRF4-3's106N</td>
<td>5'-CCAGGAACTCATGACCCCCCAGACT-3'</td>
<td>3'UTR</td>
<td>1</td>
</tr>
</tbody>
</table>
GEM 11 (kindly provided by Drs. J.T. Mulligan and R.W. Davis, Stanford University, CA) was screened with probes made from PCR products of Arabidopsis genomic DNA under moderate-stringency conditions (2X SSC, 48°C) to isolate both conserved and divergent profilin sequences. Approximately eight genome equivalents of the library (40,000 λ recombinant phages) were screened based on the λ genome equivalent of approximately 5,000 phages. Positive clones were purified and further screened using PCR probes made from pollen cDNA under high-stringency conditions to separate potentially pollen-specific profilin clones from those expressed elsewhere in the plant. Representative clones showing different signal intensities were selected and subsequently subcloned into pBluescript SKII* plasmids (Stratagene). Double-stranded DNA templates of the cDNA and genomic clones were prepared by the alkaline denaturation method of Sambrook et al. (1989), and nucleotide sequence determination was carried out by a combination of double-stranded DNA sequencing using a Sequenase 2.0 kit (United States Biochemical) and PCR sequencing using an fmol DNA sequencing kit (Promega) following the manufacturers’ instructions. A set of degenerate oligonucleotides spanning the entire profilin coding region and complementary to the profilin sequence in both sense and antisense orientations (see Table I and “Results”) was used as primer to sequence the coding region, including the introns. Unique primers were made to obtain additional 5' and 3' flanking sequences. All of the clones identified by hybridization with the various probes described above were bona fide profilin-related sequences.

Computational Analysis of Profilin Sequences

The profilin sequence data were managed within the Gel program of the Intelligenetics (Mountain View, CA) suite. Alignments of the profilin sequences and building of the profilin gene tree based on amino acid sequence homology were carried out using the PILEUP program (Devereaux et al., 1984) provided by the Genetics Computer Group of the University of Wisconsin on a VAX computer. Amino acid identity of the profilin polypeptides and nucleotide identity of the profilin coding sequences were analyzed with the GAP program of the Genetics Computer Group. The neighbor-joining and protein parsimony tree-building programs were provided within version 3.5 (provided by J. Felsenstein) of PHYLIP (Felsenstein, 1989).

Southern Blotting

Genomic DNA was isolated from Arabidopsis (Columbia ecotype) plants according to the procedure of McLean et al. (1988). The DNA was restricted with different enzymes and resolved on a 0.8% agarose gel at 30 V for 18 h followed by blotting to a nylon membrane (Biotrans Plus, ICN) in a Transvac vacuum blotter (Hoefer Scientific Instruments, San Francisco, CA) according to the manufacturer’s instructions. The blot was then exposed to UV irradiation for 2 min followed by 1 h of baking at 80°C. The blot was prehybridized overnight in 6X SSC, 5X Denhardt’s solution, 0.2% SDS, and 40% formamide at 38°C followed by hybridization under the same conditions for 24 h using a profilin probe PCR amplified from Arabidopsis genomic DNA or from an Arabidopsis profilin cDNA clone and [32P]dATP labeled at approximately 5 X 10^8 cpm/μg by the random primer method of Feinberg and Vogelstein (1983). The blot was washed 4 times for 10 min each at 48°C in 3X SSC and 0.2% SDS and exposed to X-Omat x-ray film (Kodak) with two intensifying screens for 6 to 12 h at −70°C.

Northern Blotting and RT-PCR Analysis

Arabidopsis (Columbia ecotype) plants were grown in a greenhouse under 16 h of light. Seven-day-old seedlings, roots, stems, and leaves of 3- to 4-week-old plants and flowers (approximately 90% unopened) and green siliques of 5- to 6-week-old plants were collected into liquid nitrogen and stored at −70°C before use. Total RNA was prepared from these tissues essentially by the method of Logemann et al. (1987). Mature pollen was collected and RNA was prepared as described previously (S. Huang, Y.-Q. An, J. McDowell, E.C. McKinney, R.B. Meagher, personal communication). Gene-specific probes were prepared from the PCR products PCR amplified from the 3' UTR of the genes using PRF1-3'S1SmaI and PRF1-3'N1 oligonucleotides (for PRF1, see “Results” and Table I) and PRF4-3'S1 and PRF4-3'N1BamHI oligonucleotides (for PRF4, see “Results” and Table I). Northern blotting and hybridization were carried out essentially as described by S. Huang, Y.-Q. An, J. McDowell, E.C. McKinney, and R.B. Meagher (personal communication) using these gene-specific probes (approximately 2.5 X 10^6 cpm/mL) labeled to a specific activity of approximately 5 X 10^8 cpm/μg by the random primer method of Feinberg and Vogelstein (1983). Blots were washed at 48°C three times for 15 min each, 2X SSC/0.2% SDS. RNA hybridization was carried out by the method of Thompson et al. (1992) using a 26-nucleotide 185 RNA oligonucleotide probe (approximately 5 X 10^5 cpm/mL) labeled to a specific activity of approximately 1 X 10^6 cpm/pmol. RT-PCR was carried out using gene-specific primers on RNA from the above tissues following the method of An et al. (1996). cDNAs were synthesized from total RNA from these tissues, quantified, and then PCR amplified using oligonucleotides PRF-71S and PRF1-3’N1 for PRF1 and PRF-71S and PRF4-3’106N for PRF4.

RESULTS

Conserved Profilin Sequences Were Used to Amplify Diverse Plant Profilins

Profilin is a relatively small protein of approximately 130 amino acids that is moderately conserved within some groups of organisms but poorly conserved across kingdoms and among animal phyla (Fig. 1). The few plant profilin sequences that have been characterized from timothy grass (Ppr), maize (Zma1-3), and white birch (Bve) share approximately 84 to 88% of their amino acid sequence identity with the amino terminal and...
Figure 1. Amino acid sequence comparison among profilins from different eukaryotes. The profilin sequences were compared between animals: Bta, Bos taurus; Mmu, Mus musculus; Hsa, Homo sapiens; Vac, vaccinia virus; Acr, Anthoscudaris crassispina (sea urchin); Cja, Clypeaster japonicus (sand dollar); Dme, Drosophila melanogaster; and Cel, Caenorhabditis elegans (partial sequence); plants: Ath, Arabidopsis thaliana (this study, see below); Zma, Zea mays (maize); Ppr, Phleum pratense (timothy grass); Bve, Betula verrucosa (white birch); protists: Aca, Acanthamoeba castellanii; Ppo, Physarum polycephalum; and Ddi, Dictyostelium discoideum; and a fungus: Sce, Saccharomyces cerevisiae. The numbers on the right indicate the serial numbers of the last residues at the end of the profilin amino acid sequences. The two arrows indicate the positions of the introns identified in Ath PRF1 and PRF4 gene sequences. The sequence alignment was produced using the PILEUP program (Devereaux et al., 1984) from the Wisconsin Package (Genetics Computer Group). This program ranks the amino acid sequences in order of similarities. The profilin sequences were retrieved from the GenPept, PIR, and SwissProt databases.

carboxyl terminal ends of profilins from yeast, protists, and insects (Figs. 1 and 2A) but not from vertebrates (see "Discussion"). On the assumption that more diverse plant profilins are yet to be identified, highly degenerate oligonucleotides were designed based on the sequences that were conserved across portions of all four eukaryotic kingdoms (Fig. 2A; Table I). Sense and antisense oligonucleotides were used to amplify profilin sequences from Arabidopsis leaf and floral cDNA libraries and from genomic DNA by PCR. Using the amino terminal 1S and carboxyl terminal 120A primers, a DNA product of approximately 400 bp was obtained from the cDNA libraries, and 700- to 900-bp products were obtained from genomic DNA (data not shown). Partial sequencing of these PCR products revealed profilin-related coding sequences. These PCR products were used as probes to screen floral cDNA, leaf cDNA, and genomic DNA libraries.

**Arabidopsis Profilin cDNA and Genomic Sequences**

Ten profilin cDNA clones were isolated and purified from an Arabidopsis floral cDNA library. Partial sequencing of these clones revealed three unique profilin cDNA sequences, designated as PRF1, PRF2, and PRF3. The longest cDNA clone for each of the three sequences was chosen for further characterization and was completely sequenced. They contained 5' untranslated leaders of 41 to 76 bp and 3' UTR of 136 to 218 bp, followed by short poly(A+) tails.
A. Conserved amino acid sequence motifs among distant profilin proteins. Conserved regions with four or more identical amino acids are indicated by dark shaded boxes, for comparison among monocot and dicot pollen profilins, and by striped boxes for comparison among plant, insect, protist, and yeast profilins. The approximate locations of the sense (S) and antisense (A) oligonucleotides designed from the conserved regions and used for PCR, RT-PCR, and sequencing are indicated. The numbers refer to the amino-terminal codon positions covered by that oligonucleotide. B. Arabidopsis PRF1 gene structure. Coding regions are shown by shaded boxes, and the 5' and 3' UTR are indicated. The 5' and 3' flanking sequences and introns are designated by lines. Numbers above the gene structure indicate the codon positions, and the numbers below show the size of the introns. The codons split by introns I and II are labeled. Also indicated are the ATG start and TAG stop codons. The GAA repeats in the mRNA leader and the only known poly(A) addition site in the 3' UTR are indicated. C. Arabidopsis PRF4 gene structure (see B for details).

To isolate and determine the structure of genomic profilin sequences, we screened an Arabidopsis library of randomly sheared genomic DNA using profilin probes PCR amplified from genomic DNA, leaf cDNA, and floral cDNA libraries, as described above. A preliminary screen of eight genome equivalents under moderate-stringency conditions gave rise to approximately 80 positive signals (data not shown). Several of these clones were purified and sequenced to reveal genomic clones for PRF1 and a novel profilin gene sequence designated PRF4. An XhoI fragment containing the PRF1 gene (Fig. 2B) and an XhoI-EcoRI fragment encoding the PRF4 gene (Fig. 2C) were subcloned into plasmid vectors and completely sequenced. A deduced cDNA sequence for PRF4 is shown in Figure 3. PRF4 encoded a potential profilin isoform of 134 amino acids (14.4 kD) with an estimated pI of 4.93. The PRF4 protein was fairly divergent from PRF1, PRF2, and PRF3 in amino acid sequence (see below). Its greater length was due to an insertion of three amino acids (G16, D17, and G18) relative to the other three sequences. Similar to the three profilin cDNAs characterized, the potential 5' leader sequence immediately upstream of the start codon of PRF4 contained two long stretches of GAA-related repeat sequences.

The PRF1 and PRF4 genomic sequences revealed two introns in conserved positions (Fig. 2). Introns of 505 and 260 bp were located after the Gln codons Q41 and Q44, and introns of 91 and 85 bp were located after the Lys codons K87 and K90 in the PRF1 and PRF4 genes, respectively. Their positions within the two protein coding sequences are indicated in Figures 1 and 3.

Southern Blot Analysis of the Arabidopsis Profilin Multigene Family

To determine the size of the profilin multigene family in A. thaliana, a Southern analysis was carried out under relatively low-stringency conditions using a degenerate profilin probe made from profilin sequences PCR amplified from Arabidopsis genomic DNA. The Arabidopsis genomic DNA restricted with five different enzymes produced 5 to 13 DNA fragments that hybridized to the probe (Fig. 4). Similar results were obtained under low-stringency conditions using a probe made from a coding region PCR amplified from the PRF1 profilin cDNA clone (data not shown). These results, together with the detection of approximately 80 genomic clones from 8 genome equivalents during genomic library screening, and the ease with which we were able to isolate 3 unique profilin cDNAs from only 10 clones screened, suggest that Arabidopsis contains a complex profilin gene family with approximately 8 to 10 gene members, a family size similar to that of the actin gene family in this species (McDowell et al., 1995).

Differential Expression of Two Divergent Arabidopsis Profilin Genes

The presence of the multiple divergent classes of plant profilins raised the possibility that these genes might be differentially regulated. Therefore, we looked at the expression patterns of the two most divergent profilin family gene members we had characterized, PRF1 and PRF4. Northern analysis was carried out on total Arabidopsis RNA from roots, stems, leaves, flowers, pollen, and green silicles using probes PCR amplified from the 3' UTRs of these genes and is shown in Figure 5. The sequence data in Figure 3 show that the 3' UTRs were highly divergent among the four characterized Arabidopsis profilins. The size of each profilin transcript was estimated to be approximately 750 nucleotides. As shown in Figure 5A, the Arabidopsis PRF1 transcripts were detected in all of the organs examined and in pollen. The PRF1 transcript levels were highest in stems, moderate in flowers and pollen, and very low in roots, leaves, and silicles. In contrast, the tran-
scripts of PRF4 were detected only in mature pollen (Fig. 5B). No PRF4 transcripts were detected in the other tissues even after prolonged exposure (data not shown). Hybridization of these blots with an oligonucleotide probe made complementary to 18S rRNA showed that equal amounts of RNA from all samples were loaded on the gel and transferred to the blot. Therefore, expression of PRF4 was pollen specific, and PRF1 appeared to be constitutive at the organ level.

The relative transcript levels of these two profilin genes were further examined by RT-PCR using gene-specific PCR primers that amplified portions of the coding region of the 3' UTR of the transcripts from these profilin genes. cDNAs made from the above plant organs and pollen and from young seedlings were quantified and subjected to PCR amplifications using controlled amounts of cDNA input (Fig. 6). PCR products of expected sizes for cDNAs (404 bp

Figure 3. Nucleotide and predicted amino acid sequences of four Arabidopsis profilin cDNAs (PRF1, PRF2, PRF3, and PRF4) are compared. The coding sequences are shown in uppercase letters, and the 5' and 3' untranslated sequences are shown in lowercase letters. The G-T-A-rich sequences in the 5' leader are underlined. Only those amino acids that differed from the PRF1 sequence are shown for PRF2, PRF3, and PRF4 amino acid sequences. The nucleotide 236. The position of a 3 codon insertion in PRF4, relative to the other three cDNA, is indicated (codons 16–18). The positions of oligonucleotides designed from the 3' UTR of PRF1 and PRF4 are underlined and labeled, PRF1, PRF2, PRF3, and PRF4 were sequenced directly as cDNA clones, whereas the PRF4 cDNA structure was deduced from a PRF4 genomic DNA sequence and confirmed by RT-PCR and northern analysis (below). The location of the introns identified in genomic clones of PRF1 and PRF4 are indicated by a vertical line above the PRF1 sequence.

Figure 4. Genomic Southern blotting reveals a large profilin gene family. Four micrograms of Arabidopsis genomic DNA were restricted with BglII, EcoRI, EcoRV, HindII, and HindIII, respectively, resolved on a 0.8% agarose gel, blotted to nylon membrane, and hybridized with a probe made from profilin sequences PCR amplified from genomic DNA. Sequence analysis confirmed that this genomic probe contained only profilin coding sequences. Numbers at the side indicate the molecular weight standards of a HindIII DNA marker.


Downloaded on February 12, 2021. - Published by https://plantphysiol.org
Copyright (c) 2020 American Society of Plant Biologists. All rights reserved.
Evolution of Profilin Sequences

Surprising Affinities among Profilins from Distant Kingdoms

Alignment of the profilin amino acid sequences among different eukaryotic kingdoms revealed relatively low levels of sequence identity, and only the Trp residue near the amino-terminal end (W3 in most sequences) was completely conserved in all profilins examined. Typically, only 25 to 40 residues were identical in any comparison of profilin sequence pairs between two kingdoms. Thus, a quantitative approach was needed to examine the sequence relatedness among the profilins.

Gene trees (Fig. 7) were constructed from an amino acid sequence alignment similar to that shown in Figure 1 but containing some additional profilin sequences. These trees were constructed from two independent tree-building algorithms, neighbor joining in Figure 7A and protein parsimony in Figure 7B, but gave nearly identical results. The tree topologies differed significantly only in the placement of a potential profilin homolog from the protist *Eimeria acervulina* (*Eac*), which is of unknown function. Surprisingly, both trees placed the plant, fungal, protist, and lower animal (Dme, insect) profilins closer to each other than to the profilin sequences from vertebrates or echinoderms (deuterostomes). Only a partial sequence was available for the nematode *Caenorhabditis elegans* (Cel) shown in Figure 1, so it was not included in these trees. However, independently constructed trees, which included the Cel sequence using these and other methods (not shown), grouped this lower animal sequence with the plant, fungal, and protist sequences. The simplest explanation for this topology is that most organisms share a common ancient profilin sequence, whereas the deuterostomes appear to have evolved novel profilin sequences since their divergence from this ancestral type (see “Discussion”).

An Ancient Split among Plant Profilins

Examination of the sequence alignment in Figure 1 also showed that the plant profilins were essentially colinear,
Figure 7. Profilin sequence trees. Trees were constructed by the neighbor-joining (A) and protein parsimony (B) methods provided with the PHYLIP package of computer programs (Felsenstein, 1989). Both trees were based on an amino acid sequence alignment similar to that shown in Figure 1A with the addition of the following sequences from plants: Nta, Nicotiana tabacum; Tae, Triticum aestivum; and Pvu, Phaseolus vulgaris; animals: Var, Varicella major (virus) and Hsa2, Homo sapiens; and protists: Tpy, Tetrahyamina pyriformis and omission of the partial sequence from C. elegans and the sea urchin sequence, Acr. The sequence of a potential profilin homolog and protist sporozoite antigen, from Eac, was also included. The approximate scale of the trees can be determined from the Ath PRF1 versus Ath PRF2, Ath PRF1 versus Ath PRF4, Ath PRF1 versus Ppo1, and Ath PRF1 versus Hsa1 comparisons, which have approximately 87, 76, 44, and 34% amino acid identity, respectively. Based on discussions within the text, the two classes of plant profilins are labeled as constitutive or pollen-specific. The sequence alignment used in both trees was produced with the PILEUP program (Devereaux et al., 1984) from the Wisconsin package. The additional profilin-related sequences were retrieved from the GenPept, PIR, and Swiss Prot databases.

Table II. Percentage of amino acid and coding nucleotide identity of plant profilins
Comparison was done by the GAP program of the Genetics Computer Group of the University of Wisconsin. Amino acid identity is shown in the lower part and nucleotide identity in the upper part of the matrix table. See Figure 1 for gene designations.

<table>
<thead>
<tr>
<th>Gene</th>
<th>AthPRF1</th>
<th>AthPRF2</th>
<th>AthPRF3</th>
<th>AthPRF4</th>
<th>Bve</th>
<th>Ppo</th>
<th>Zma1</th>
<th>Zma2</th>
<th>Zma3</th>
</tr>
</thead>
<tbody>
<tr>
<td>AthPRF1</td>
<td>82.8</td>
<td>77.0</td>
<td>71.2</td>
<td>72.5</td>
<td>71.5</td>
<td>71.7</td>
<td>69.2</td>
<td>72.0</td>
<td></td>
</tr>
<tr>
<td>AthPRF2</td>
<td>87.0</td>
<td>75.5</td>
<td>70.2</td>
<td>72.7</td>
<td>72.5</td>
<td>72.2</td>
<td>71.0</td>
<td>72.5</td>
<td></td>
</tr>
<tr>
<td>AthPRF3</td>
<td>81.7</td>
<td>77.9</td>
<td>68.9</td>
<td>68.7</td>
<td>71.0</td>
<td>68.4</td>
<td>69.7</td>
<td>69.7</td>
<td></td>
</tr>
<tr>
<td>AthPRF4</td>
<td>75.6</td>
<td>73.3</td>
<td>74.8</td>
<td>75.1</td>
<td>76.5</td>
<td>74.5</td>
<td>73.2</td>
<td>73.2</td>
<td></td>
</tr>
<tr>
<td>Bve</td>
<td>76.3</td>
<td>76.3</td>
<td>76.2</td>
<td>82.0</td>
<td>76.3</td>
<td>75.0</td>
<td>75.3</td>
<td>75.5</td>
<td></td>
</tr>
<tr>
<td>Ppo</td>
<td>74.8</td>
<td>73.3</td>
<td>74.0</td>
<td>82.4</td>
<td>79.4</td>
<td>85.4</td>
<td>86.4</td>
<td>83.1</td>
<td></td>
</tr>
<tr>
<td>Zma1</td>
<td>71.8</td>
<td>73.3</td>
<td>69.5</td>
<td>79.4</td>
<td>77.1</td>
<td>82.4</td>
<td>90.4</td>
<td>82.8</td>
<td></td>
</tr>
<tr>
<td>Zma2</td>
<td>71.0</td>
<td>71.8</td>
<td>69.5</td>
<td>78.6</td>
<td>76.3</td>
<td>83.2</td>
<td>90.1</td>
<td>83.6</td>
<td></td>
</tr>
<tr>
<td>Zma3</td>
<td>74.0</td>
<td>73.3</td>
<td>72.5</td>
<td>78.6</td>
<td>79.4</td>
<td>84.7</td>
<td>88.6</td>
<td>89.3</td>
<td></td>
</tr>
</tbody>
</table>

ranging from 131 to 137 amino acids. A few small insertion/deletion differences were observed. For example, the Zes mays sequence Zma2 contained an additional six amino acids at the amino terminus. In addition to an insertion of 3 amino acids positioned 15 amino acids downstream of the initiating Met, as was also observed in PRF4, the white birch profilin sequence Bve had a deletion of 1 amino acid, located 25 amino acids from the initiating Met. A detailed comparison of the amino acid identity and nucleotide sequence divergence among nine of the plant profilins is shown in Table II. Approximately 70 to 90% of the amino acids were identical among any one of the 36 comparisons of plant sequence pairs. At the nucleotide sequence level, the coding sequences of these profilins were 68 to 90% identical. Although the plant profilins were more similar to each other than to sequences from other kingdoms, they
still contained a great deal of variability that may be functionally significant. For example, 14 of the 36 differences between *Ath PRF1 and PRF4* involved the charged residues H, K, R, D, or E (Fig. 3).

Two clades or classes of profilins were resolved in the two sequence trees (Fig. 7). Arabidopsis *PRF4* was more similar to the profilins from other plant species than to the other three Arabidopsis profilins. At the amino acid level, *PRF4* was 78.6 to 82.4% identical to the profilins from the monocots *Z. mays, Zma*, timothy grass, *Ppr*, and the dicot white birch, *Bee*. All of these sequences are expressed in pollen. The maize and Arabidopsis sequences have been shown to be pollen specific. These data suggest that the tobacco, *Nta*, and wheat, *Tae*, sequences also may be expressed in pollen (see “Discussion”). *PRF4* only shared 73.3 to 75.6% amino acid identity with the other three Arabidopsis profilins (*PRF1–3*). Thus, *PRF4* is significantly closer to the known profilins in other plant species than to *PRF1–3*. *PRF4* is saturated with silent nucleotide changes with respect to the other three Arabidopsis sequences. Arabidopsis *PRF1–3* showed low amino acid sequence identity to most of the plant profilins examined (69.5–76.3%), with the exception of a more closely related sequence from the common French bean, *Pou* (78–82%). These four plant profilin sequences appear to form a new clade among plant profilins. *PRF1* was constitutively expressed in most organs; *PRF1*, *PRF2*, and *PRF3* were each identified in both flower and leaf (not shown) cDNA libraries; and *Pou* was expressed in leaves and roots (Vidali et al., 1995). This second clade (Fig. 7) may represent a constitutive class of profilins (see “Discussion”). Similar sequence relationships were obtained from DNA sequence comparisons (Table II). Among the three Arabidopsis sequences (*PRF1–3*) in this novel clade, there is a significant number of silent nucleotide substitutions (>40%, Fig. 3), suggesting that they have not shared a common ancestral sequence for more than 30 to 50 million years.

**Conservation of Functional Domains among Plant Sequences**

Restricted cross-kingdom comparisons revealed conservation of amino acid domains reported to be involved in profilin functions. For example, in bovine profilin, the hydrophobic domain rich in aromatic amino acids (composed of W3, Y6, W31, H133, L134, and Y139 [Fig. 1], which have been implicated in poly-L-Pro binding [Björkergren et al., 1993; Schutt et al., 1993]) was highly conserved even in many distantly related profilins, including those from plants. It is also worth noting that the *Acanthamoeba, Aca*, profilin sequence motif VEKALDYL (amino acids 112–119, Fig. 2), which has been implicated in actin binding (Schutt et al., 1993; Cedergren-Zeppezauer et al., 1994), was partially conserved in most plant (VERLGDL), protist, fungal, and lower animal profilins but was poorly conserved in vertebrate profilins. In addition, a sequence rich in basic amino acids in the human gelsolin (KHVPIEVEVQRLFQVKGR) and villin (YNVQRLLHVGGKKNV) has also been identified as having a role in the binding of polyphospho- lipids PIP and PIP2 (Janneny et al., 1992). All of the plant profilins contained a similar sequence motif (RG-KKG(S/A/P)GG(T/V)T(I/V)KK; amino acids 84–96 in *PRF1*, Fig. 1A), with five positively charged amino acids completely conserved among the plant profilins. All protist and plant profilins shared some similarity within this region.

**DISCUSSION**

**Arabidopsis Contains a Complex Profilin Gene Family**

The four Arabidopsis profilin sequences *PRF1, PRF2, PRF3*, and *PRF4* isolated and characterized herein represent only a small portion of the genes in a complex, multigene family. Based on the data from Southern analysis and genome reconstruction experiments and the high frequency with which the profilin genomic DNA sequences were isolated, it appears that Arabidopsis contains at least 8 to 10 profilin genes. The diversity of the four Arabidopsis profilin gene family members isolated so far suggests that the several additional profilin-related sequences yet to be characterized will also be quite diverse. It seems likely that there are additional ancient classes or subclasses of profilins yet to be characterized.

Examination of the Arabidopsis profilin sequences reveals some distinct characteristics that may be important to most plant profilins. First, the 5' untranslated leader sequences of all four profilin mRNAs contain G+A-rich sequences. In nearly all plant actin genes characterized, including those from Arabidopsis, C+T-rich sequences are observed in the mRNA leader sequences (Pearson and Meagher, 1990). These G+ A- and C+T-rich sequences have been referred to as PuPy (purine- or pyrimidine-rich) sequences and may be involved in recombination, gene duplication, and/or regulation of gene expression (Pearson and Meagher, 1990). The exact function(s) of the PuPy-rich sequences associated with the profilin mRNA leader remains to be investigated. Second, like maize profilins (Staiger et al., 1993), the four characterized Arabidopsis profilins also contain sequences potentially involved in binding actin, poly-L-Pro, and PIP2, suggesting that these plant profilins have functions similar to profilins in other kingdoms. In fact, a closely related plant profilin, *Ppr* from timothy grass, binds poly-L-Pro and actin (Valenta et al., 1993). Third, characterization of genomic clones for the evolutionarily distant *PRF1* and *PRF4* genes reveals two small introns located in conserved positions that divide the coding sequences into three nearly equal parts. Fourth, the plant profilin protein sequences are far more like each other than they are like the profilins from any of the other three eukaryotic kingdoms. These data suggest that the plant profilin genes, like the actin genes (McDowell et al., 1995), have evolved from a single common ancestral plant gene.

We have shown that genes from two divergent profilin classes within the Arabidopsis profilin gene family exhibit fundamentally different expression patterns, one constitutive and the other pollen-specific. The Arabidopsis pollen-specific *PRF4* is in the gene class that includes profilin genes from the monocots maize (*Zma1–3*), wheat (*Tae*), and timothy grass (*Ppr*), and the dicots white birch (*Bee*) and...
tobacco (Nta) (Fig. 7). It is interesting that the maize profilins characterized so far are also specifically expressed in pollen (Staiger et al., 1993). The white birch and timothy grass profilins were isolated from pollen cDNA libraries as allergens reactive with IgE from pollen-allergic patients (Valenta et al., 1991), although it is not known if their expression is pollen-specific. Therefore, it is likely that PRF4 and related profilin genes from diverse plant species represent an evolutionarily conserved gene class specialized for the regulation of the cytoskeleton in pollen or other reproductive tissues (see below).

The actin cytoskeleton is essential for pollen germination, subsequent tube growth, and movement of germinial cells down the tube (Heslop-Harrison and Heslop-Harrison, 1992; Mascarenhas, 1993). The rapid F-actin-based streaming in the tube is thought to carry vesicles with the essential building materials required for tip growth as the tube grows through the gynoecium. Profilin could influence these processes by affecting the actin cytoskeleton in several ways. For example, profilin can both inhibit and promote actin polymerization because of its role in sequestering actin monomers and in catalyzing nucleotide exchange associated with monomeric actin (Theriot and Mitchison, 1993). In addition, profilin interacts with phospholipids and probably links the remodeling of the actin cytoskeleton to the signal transduction pathway in response to extracellular stimuli (Aderem, 1992; Machesky and Pollard, 1993; Drobas et al., 1994). The exact role(s) of profilin in regulating the rearrangement of the actin cytoskeleton during pollen germination and pollen tube growth, however, remains to be investigated.

The class containing the constitutively expressed Arabidopsis PRF1 and related sequences may use these same molecular tools to regulate the more general functions of the actin cytoskeleton in the majority of vegetative cell types. This would include control over cytoplasmic streaming, division plane determination, cell-wall deposition, organelle movement, and transport. Distinct profilin isoforms are expressed in the leaves and root nodules of Phaseolus vulgaris (Perez et al., 1994), and these may represent other members of this clade. We analyzed one of these bean sequences (Pev, Fig. 7), and it groups with the "constitutive" Arabidopsis sequences, as would be expected if this profilin clade represented an ancient class predating the major groups of angiosperms.

Evolution of Diverse Plant Profilin Sequences and Coevolution with Actin

The four sequences that have been isolated and characterized can be divided into at least two divergent classes. One of the Arabidopsis profilin genes, PRF4, is significantly more closely related to profilins from distant monocot and dicot species than it is to PRF1, PRF2, or PRF3, suggesting the presence of plant profilin gene lineages that predate this organismal divergence. Dicots and monocots have not had a common ancestor for about 200 million years (Wolfe et al., 1989), and thus this split in the profilin gene family is more than 200 million years old.

Based on a detailed analysis of DNA and protein sequences, RNA steady-state levels, and expression of promoter/reporter fusions in numerous transgenic plants, the 10 Arabidopsis actins were divided into two ancient classes that have not shared a common ancestral sequence for approximately 500 million years. One class is expressed constitutively in vegetative tissues and the other is expressed primarily in reproductive organs and tissues including pollen and ovules (McDowell et al., 1995). Perhaps the profilin gene classes diverged concurrently with the actin gene classes. The structure of actin-profilin co-crystals reveals intimate interaction between the two proteins over a large surface area involving 20 to 25 of profilin's 131 amino acids (Schutt et al., 1993; Sheterline and Sparrow, 1994). This suggests that individual classes of profilins would have had to coevolve intimately with the classes of actins for proper protein-protein interactions.

Considering the diversity of plant actin (Meagher and McLean, 1990) and profilin isoforms, coevolution might be required to maintain an appropriately balanced interaction between these proteins. Coexpression and coevolution of multiple plant profilins and actins may have led to a complex form of dynamic instability, as was first proposed to explain the interaction of multiple tubulin isoforms, which were coexpressed (Burns and Surridge, 1990). The dynamic instability theory suggests that tubulin polymerization can have more complex kinetic responses when multiple tubulin isoforms are involved. In other words, having multiple and diverse actins and profilins interacting in a particular cell type greatly increases the complexity of kinetic responses possible for the actin-based cytoskeleton.

Profilin Divergence and the Allergic Response

Allergies have become a serious health problem, affecting up to 15% of the population in industrialized countries. Plant profilins have been identified as potent allergens in patients with serious pollen and food allergies (Valenta et al., 1991). Plant profilin binds a significant portion of the IgE in a large percentage of these allergic patients. Comparison of the amino acid sequences shows that the vertebrate profilins are extremely divergent from plant profilins, with an amino acid identity of only approximately 30% (Figs. 1 and 7). Human allergic responses to plant profilins are likely to be directed at the amino acid domains that are conserved among plant profilins but divergent in vertebrate profilin sequences. Of particular note is that the amino acid sequences at the amino and carboxyl termini are both conserved among plant, fungal, and lower animal profilins but are highly divergent in the vertebrate profilins. We propose the following scenario for the evolution of the human IgE response to profilin. Profilin sequence comparisons suggest that the plant, protist, fungal, and protostome (e.g. insect and perhaps nematode) profilins are closer to each other than they are to the vertebrate profilins (Figs. 1 and 7). The most parsimonious interpretation of the
phylogeny is that the more common profilin sequence, shared among several kingdoms, is the more ancient one. This suggests that deuterostomes (e.g. vertebrate and echinoderm sequences) evolved a unique and highly divergent profilin sequence after they diverged from the protostomes (e.g. lower animals such as insects and worms). The IgE response is thought to have evolved during the evolution of early vertebrates approximately 500 million years ago (Margulis and Schwartz, 1982) to protect against ectoparasites (e.g. arthropods, mites, and ticks) and helminth parasites (e.g. schistosomes) (Abbas et al., 1991). The high concentration of profilin, the profilin sequence conservation among ancestors of these distant pathogens, and the divergence from vertebrate sequences made them excellent targets for the vertebrate immunological response. The fungal spore profilins and plant pollen profilins only became respiratory allergens after evolution of sporophytic fungi and wind-borne pollen 400 and 250 million years ago, respectively. The sequence homology of plant and fungal profilins to those in lower animal parasites is simply a consequence of the common evolutionary origin for profilin sequences. Future analysis of the profilin sequences and structures from diverse lower animals may shed more light on the evolution of the vertebrate IgE response to profilins from plants and fungi.

ACKNOWLEDGMENTS
We thank Rudolf Valenta for generously sharing his Bov and Ppr profilin sequences before their publication. We thank R. Valenta, M. Fechheimer, and G. Gragson for several useful comments on the material in the manuscript.

Received October 6, 1995; accepted January 30, 1996.

Copyright Clearance Center: 0032-0869/96/111/0115/12.

LITERATURE CITED


Ruhlman T, Lange U, Grolik F (1995) Profilins purified from higher plants bind to actin from cardiac muscle and to actin from a green alga. EMBO J 14: 1583–1589


