Plant Physiol. (1994) 106: 403–404

Plant Gene Register

Nucleotide Sequence of a Pollen-Specific cDNA from *Helianthus annuus* L. Encoding a Highly Basic Protein[1]

Natalia Dudareva[2]*, Jean-Luc Evrard, Datta T. N. Pillay, and Andre Steinmetz

Institut de Biologie Moleculaire des Plantes, 12 Rue du General Zimmer, Strasbourg, France (N.D., J.-L.E., A.S.); and Department of Biological Sciences, University of Windsor, Windsor, Ontario N9B 3P4, Canada (D.T.N.P.)

Pollen, as the male gametophyte, plays a vital role in the reproduction of flowering plants. Mature pollen grains contain mRNAs that are the products of approximately 20,000 different genes, of which up to 4000 may be pollen specific (Mascarenhas, 1990). In recent years, several pollen-specific genes have been isolated from different plants. Some of these genes encode proteins showing amino acid homology to known hydrolytic enzymes, such as polygalacturonase (Allen and Lonsdale, 1993), pectin esterase (Albani et al., 1991), and pectate lyase (Wing et al., 1989). All of these genes are expressed after microspore mitosis and their transcripts accumulate considerably in mature pollen. So far, no sequence homology to known proteins has been detected in the other pollen-specific genes studied to date (Hanson et al., 1989).

We report here the molecular characterization of another pollen-specific cDNA clone, SF16, from sunflower (*Helianthus annuus* L.) (Table I). This gene is present in a single copy in the sunflower genome, expressed late in pollen development, and has no homology to other published sequences. In the original series of northern experiments (not including RNA from pollen), this clone was found to hybridize preferentially to RNA from pistils (Herdenberger et al., 1990); more recent and detailed expression studies revealed that the hybridization was due to contaminating pollen and that the single transcript of 1400 nucleotides can be detected exclusively in mature, free (tricellular) pollen. The transcript was not detected in yellow closed flowers, indicating that it is synthesized at a late stage of pollen development.

The cDNA is 1293 nucleotides long with a 993-nucleotide-long open reading frame, starting at the first ATG codon of the cDNA sequence (positions 146–148) and terminating with a TGA stop codon (positions 1139–1141). Although we have no experimental evidence that the cDNA is full length, the initiation codon can be identified without ambiguity because of the presence of an upstream in-frame stop codon TAA (positions 92–94).

Translation of the SF16 cDNA sequence revealed an open reading frame that codes for a putative polypeptide of 331 amino acid residues with molecular mass of 37.4 kD. The polypeptide is rich in basic amino acid residues (20%; isoelectric point = 11.3), very hydrophilic, and potentially has a high proportion of α-helix structures, two of which are 18 and 43 amino acids in length.

This polypeptide has no significant hydrophobic domain

Table I. *Characteristics of sunflower pollen-specific gene SF16*

Organism:
Sunflower (*Helianthus annuus* L. male-fertile line HA401B).
Source:
cDNA library in λgt10 constructed using poly(A$^+$) RNA from a sunflower inflorescence at anthesis.
Cloning Techniques:
cDNA library was screened with a previously identified partial-length cDNA clone. The insert was subcloned into *Eco*RI site of pUC 19. Both strands were sequenced using the dideoxy chain-termination method.
Expression Characteristics:
Northern hybridization showed that the gene is expressed at a late stage of pollen development. A single transcript of 1.4 kb was detected exclusively in mature, free (tricellular) pollen.
Characteristics of the cDNA Clone:
The cDNA is 1293 nucleotides long with a 993-nucleotide-long open reading frame. It has a 145-nucleotide 5′ untranslated region and a 155-nucleotide 3′ untranslated region including a short poly(A) tail of nine A residues. A putative polyadenylation signal motif (AATAAG) was found 62 bp downstream of the termination codon TGA.
Characteristics of the Deduced Protein:
The open reading frame encodes a putative polypeptide of 331 amino acids with a mol wt of 37.4 kD. This polypeptide is very basic (20% of basic amino acids; isoelectric point = 11.3), highly hydrophilic, and potentially rich in α-helix structures. Its intracellular location suggests a potential nucleic acid-binding function.
Sequence Comparison:
Comparison of the nucleotide and deduced amino acid sequences with sequences in EMBL data bank has not shown any homology to known proteins.
Gene Copy Number:
Single copy gene as determined by Southern blot analysis.
Antibodies:
Not available.

[1] This work was supported by a grant from the French Ministry of Research and Technology, by a Competitive Research Grant from Pioneer HiBred International, and partly by a Grant A-1984 from the Natural Science and Engineering Research Council of Canada to D.T.N.P.

[2] Present address: Department of Biological Sciences, University of Windsor, Windsor, Ontario N9B 3P4, Canada.

* Corresponding author; fax 1–519–971–3609.

and is therefore probably located either in the cytoplasm or in the nucleus. The hydrophilic, basic character suggests a potential affinity for nucleic acids; hence, the SF16 polypeptide could be a nucleic acid-binding protein, acting in the cytoplasm as a translation regulator, or in the nucleus as a transcriptional regulator, or even as a pollen-specific protein involved in the stability or maturation of pollen mRNAs. None of the known nucleic acid-binding motifs (MADS box, homeodomain, Leu zipper, zinc finger, helix-turn-helix) is found in the SF16 polypeptide, suggesting that the potential interaction with nucleic acids proceeds via a new type of domain. The expression of the SF16 gene parallels that of the pollen-specific gene PLIM-1 from sunflower (formerly called SF3) (Baltz et al., 1992a, 1992b). This raises the possibility that the expression of SF16 gene is regulated by the LIM-domain-containing protein PLIM-1, or conversely, that the SF16 and PLIM-1 polypeptides act together to regulate the expression of a number of late pollen-specific genes.

Received March 21, 1994; accepted April 10, 1994.

The EMBL accession number for the sequence reported in this article is X74722.

LITERATURE CITED

Albani D, Altosaar I, Arnison PG, Fabijanski SF (1991) A gene showing sequence similarity to pectin esterase is specifically expressed in developing pollen of *Brassica napus*: sequences in its 5′ flanking region are conserved in other pollen-specific promoters. Plant Mol Biol **16:** 501–513

Allen RJ, Lonsdale DM (1993) Molecular characterization of one of the maize polygalacturonase gene family members which are expressed during late pollen development. Plant J **3:** 261–271

Baltz R, Domon C, Pillay DTN, Steinmetz A (1992a) Characterization of a pollen-specific cDNA from sunflower encoding a zinc finger protein. Plant J **2:** 713–721

Baltz R, Evrard J-L, Domon C, Steinmetz A (1992b) A LIM motif is present in a pollen-specific protein. Plant Cell **4:** 1465–1466

Hanson DD, Hamilton DA, Travis JL, Bashe DM, Mascarenhas JP (1989) Characterization of a pollen-specific cDNA clone from *Zea mays* and its expression. Plant Cell **1:** 173–179

Herdenberger F, Evrard J-L, Kuntz M, Tessier L-H, Klein A, Steinmetz A, Pillay DTN (1990) Isolation of flower-specific cDNA clones from sunflower. Plant Sci **69:** 111–122

Mascarenhas J-P (1990) Gene activity during pollen development. Annu Rev Plant Physiol Plant Mol Biol **41:** 317–338

Wing RA, Yamaguchi J, Larabell SK, Ursin VM, McCormick S (1989) Molecular and genetic characterization of two pollen-expressed genes that have sequence similarity to pectate lyases of the plant pathogen *Erwinia*. Plant Mol Biol **14:** 17–28