

Nucleotide sequence of two anther-specific cDNAs from sunflower (*Helianthus annuus* L.)

Claire Domon¹, Jean-Luc Evrard¹, Françoise Herdenberger¹, Datta T.N. Pillay² and André Steinmetz^{1,*}

¹Institut de Biologie Moléculaire des Plantes, 12, rue du Général Zimmer, F-67084 Strasbourg-Cedex, France (*author for correspondence); ²University of Windsor, Windsor, Ontario, Canada

Received and accepted 29 March 1990

We have recently isolated and characterized several flower-specific clones from a sunflower cDNA library [1]. Two of these clones, SF2 and SF18, were shown to hybridize exclusively to RNA from anthers in the late developmental stages (none of the two RNA species is detected in anthers from a closed sunflower inflorescence). The two mRNAs, which are abundant RNAs in anthers from male-fertile plants, are barely detectable in anthers from male-sterile plants. *In situ* hybridization studies using the labeled SF2 cDNA clone as probe have shown that the SF2 mRNA is present only in the epidermal anther cells (Evrard *et al.*, submitted). The lack of cross-hybridization between the two cDNAs suggested that the SF2 and SF18 mRNAs code for different proteins.

We describe here the nucleotide sequence of the two cDNAs SF2 and SF18. The sequence analysis revealed a size of 653 bp for SF2 cDNA and 656 bp for SF18 cDNA (Fig. 1); this is a little smaller than that of the corresponding mRNAs estimated in northern experiments (800 nucleotides for SF2 and 750 nucleotides for SF18) and indicates that the two cDNAs are incomplete. Translation of the cDNA sequence of SF2 revealed an open reading frame starting with a methionine and extending over 121 amino acids. That translation must start at this specific methio-

nine residue *in vivo* has recently been confirmed by the study of the corresponding gene (C. Domon, unpublished). SF18 cDNA carries a 161 amino acid-long reading frame, which is incomplete at the amino terminus. The first 8 amino acids of this incomplete SF18 reading frame are also found in the amino terminal region of SF2 (only one valine substituting for a leucine). In SF2, this region carries a typical signal peptide sequence (21 amino acids, 16 of which are hydrophobic).

Not only are the SF2 and SF18 mRNAs similar in size, they also share short segments of significant sequence homology: one of these is found in the region coding for the signal peptide (where 25 out of 28 nucleotides are identical); another one is found in the untranslated 3' region which contains the polyadenylation site (AATAAA) and a short poly(A) stretch (21 A residues in SF2 and 18 A residues in SF18). This region is considered to be involved in mRNA stability [2].

Although no additional sequence similarity can be detected in the SF2 and SF18 polypeptides (except for the signal peptide sequence), they share some features suggesting that they are functionally related. One of these features is a pattern of 7 cysteine residues with a nearly identical spacing in the amino moiety of the two mature polypeptides. This 'cysteine domain' (Fig. 2A), is also rich in charged amino acids (35% in SF2 and

Signal peptide

M A N N S V S Y L V L L L L V F V L A I S E
TCGAAAATGGCGAACAAATTCAGTTTCTTACTTAGTATTGCTTTTGCTCGTCTTCGTTCCTGCAATATCAG⁷⁰ SF2
||||| |||||
.....TGCTCGTCTTTGTTGTTGCAATCTCAG²⁷ SF18
L V F V V A I S D

S A P V Q Y C D R V T N L Y H E K C D E
AAAGCGCACCGGT.....ACAATAT.TGCGACAGGGTAACCAACCTTTATCATGAAAAGTGTGACGAG¹³² SF2
|
ATATTGCGACTGTCAATGGAAAAATATGTGAAAAACCGAGTA.AGACTTGGTTTGGA...CTGCAAAGAC⁹⁵ SF18
I A T V N G K I C E K P S K T W F G N C K D

K Q C T E H C K T N E K A E S G Y C L V V E
A...AGCAGTGTACTGAACACTGCA.AGACAAATGAGAAGGCAGAATCAGGATACTGCCCTCGTAGTTGAA¹⁹⁸ SF2
|
ACAGATAAGTGTGACAAGCGGTGTATAGAC.TGGGAGGGTGC AAAGCATGGAGCTTGTCATCAACGTGAA¹⁶⁴ SF18
T D K C D K R C I D W E G A K H G A C H Q R E

K Q Q L S I C S F D C S K Y K P S T P A P P P
AAGCAACAACCTTAGCATATGCAGCTTCGATTGT...TCCAAGTATAAACCGAGTACTCCGGCTCCACCTC²⁶⁵ SF2
|
GCCAAACACATGTGCTTTTGCTACTTTGATTGTGATCCCCAGAAGAATCCAGGGCCACCACCAGGGGCAC²³⁴ SF18
A K H M C F C Y F D C D P Q K N P G P P P G A P

P P P K L F Y S G S W L Q A K V E N V M L P G
CACCTCCACCAAAGCTCTTCTATCTGTTTCATGGTTGCAGGCCAAAGTCGAGAATGTGATGCTTCTCTGG³³⁵ SF2
|
CGGGGACGCCAGGGACGCCTCCTGCTCCTCCAGGAAAGGGCGAAGGAGATGCTCCTCATCTCCTCTCTAC³⁰⁴ SF18
G T P G T P P A P P G K G E G D A P H P P P T

Q K N M N C T Q C P K *
GCAAAAGAAAT....ATGAATTGCACAAATGTCCCAAATAATCAACAGTCTGCATGTT.....T³⁹⁰ SF2
|
TCCAAGTCTCCTGGAGGTGATGGAGGCAGTGGACCAGCTCCTCCTGCAGGTGGTGGATCACCTCCTCTCT³⁷⁴ SF18
P S P P G G D G G S G P A P P A G G G S P P P

GTATGTTTTGAGTCTTTGCATGAAGTGCT.....TAAATTAAGGTTATACTTGTCATTTTGAGT⁴⁵⁰ SF2
|
GCAGGTGGTGACGGAGGTGGTGAGCACCTCCTCCCGCAGGTGGTGACGGAGGTGGTGAGCTCCTCCTC⁴⁴⁴ SF18
A G G D G G G G A P P P A G G D G G G G A P P P

CATATTCTAGAAAAGAGTCGTTGGGTATAACTGGATAAAG..TGGGTAGTCCATGTTGTCAAGTAGTGGG⁵¹⁸ SF2
|
CTGCAGGTGGTGATGGA.GGCGGAGCACCACCAGGAGCTTAGGC.GAGAAAGAGCCAAAGTCTTGAA⁵¹² SF18
A G G D G G G A P P P G A *

TAGATTTTCTCTGTTTGTCATGTTTGTGTTTATGTGGCCATCG.....TGTGTTTACAGTGTGTTGT.....⁵⁷⁶ SF2
|
TAGAAAA...CTTTATACATA..TGATTATGAATAAGGAGAACATATATGCTTACAGATTGAAATTAATT⁵⁷⁷ SF18

..GTTTCCCTTGTGTATTTAAG.CTCTTGTGTGATGAATAAATTATCC...AGTTCCAACA₂₁⁶⁵³ SF2
|
TAGTTTGCA.TGTGTAATAAACCTCATGAGTTCAATTATAAGTCTATAATAAATTTACATGA₁₈⁶⁵⁶ SF18

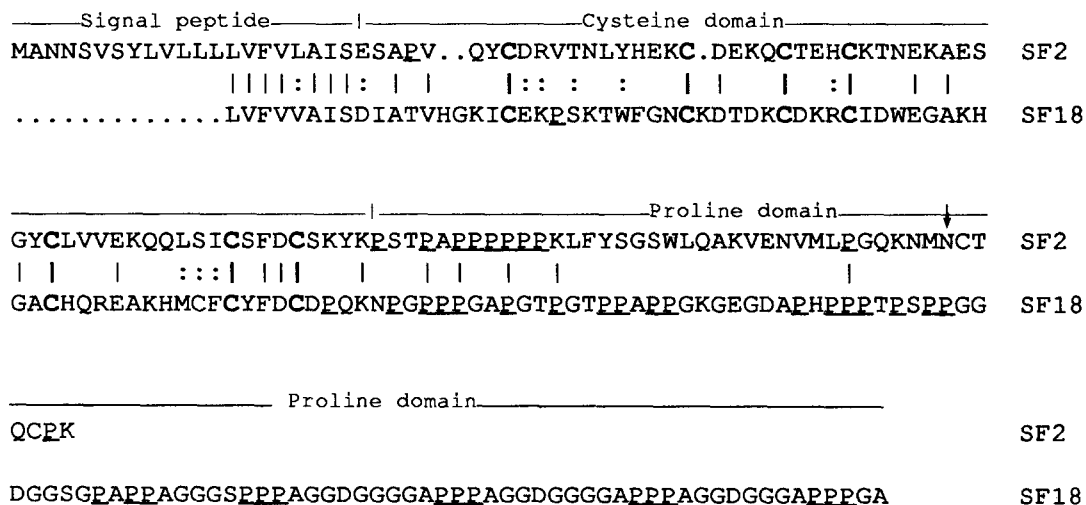
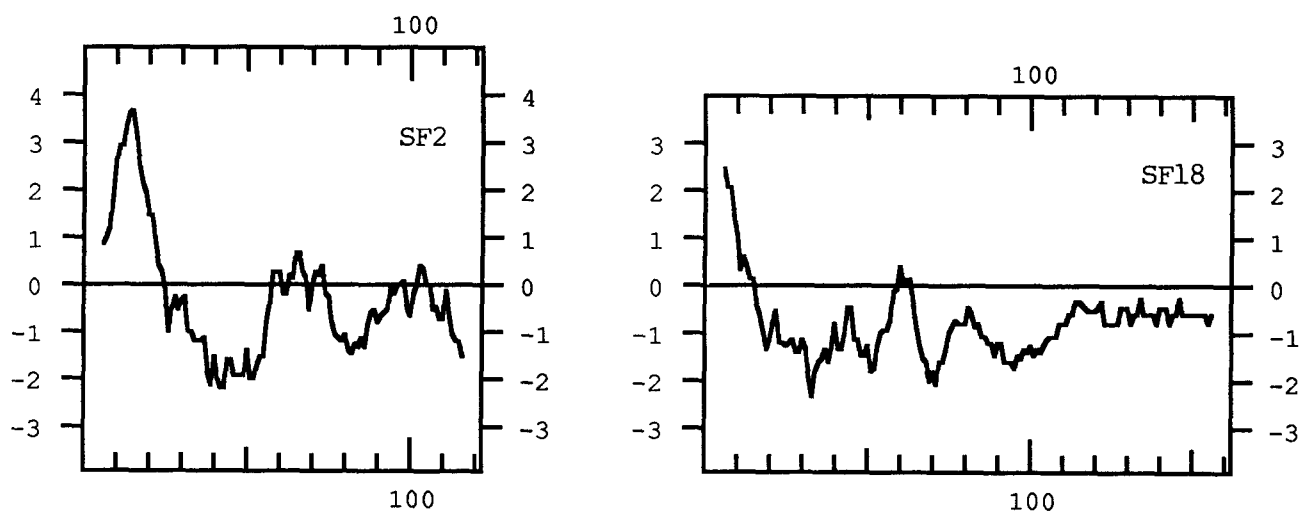
A**B**

Fig. 2. A. Comparison of the amino acid sequence of the SF2 and SF18 proteins derived from their respective cDNA sequence. Signal peptide, cysteine domain and proline domain are shown; the junctions are marked by a small vertical bar. The conserved cysteine residues in the cysteine domain are shown in bold-faced type. An N-glycosylation site in SF2 is indicated by a vertical arrow. Proline residues are underlined. B. Hydropathy plots of the SF2 and SF18 polypeptides according to Kyte and Doolittle [6].

41% in SF18). Another feature shared by the two polypeptides is the relatively high concentration of proline residues in the carboxy moiety. This 'proline domain' is poor in charged amino acids

(12% in SF2 and 9% in SF18). In SF2 the proline residues are clustered in a single stretch, whereas in SF18 they are spread over the entire second moiety. Most remarkable in SF18 is the tripli-

cation of a glycine plus proline-rich oligopeptide AGGD(G)GGGAPPP at the carboxy terminus.

The presence of a hydrophobic signal peptide at the amino terminus of the two polypeptides (Fig. 2B) suggests that they are synthesized on membrane-bound ribosomes and follow the export pathway. Since most of the hitherto known (hydroxy)proline-rich [3] or glycine-rich [4, 5] proteins have been identified in cell walls we propose that the SF2 and SF18 polypeptides are anther-specific cell wall proteins which could contribute to the cell wall architecture of epidermal anther cells via intermolecular disulfide bridges.

Acknowledgement

We thank A. Boennec (CARGILL, France) for providing the sunflower seeds (*Helianthus annuus* L. male-fertile line HA401B) used in the present study.

References

1. Herdenberger F, Evrard JL, Kuntz M, Tessier LH, Klein A, Steinmetz A, Pillay DTN: Isolation of flower-specific cDNA clones from sunflower. *Plant Sci* 69: 111–122 (1990).
2. Wilson T, Treisman R: Removal of poly(A) and consequent degradation of c-fos mRNA facilitated by 3' AU-rich sequences. *Nature* 336: 396–399 (1988).
3. Showalter AM, Varner JE: Plant hydroxyproline-rich glycoproteins. In: Marcus A (ed.) *The Biochemistry of Plants: Molecular Biology*, Vol 15, pp. 485–520. Academic Press, San Diego (1989).
4. Condit CM, Meagher RB: A gene encoding a novel glycine-rich structural protein of petunia. *Nature* 323: 178–181 (1986).
5. Keller B, Sauer N, Lamb CJ: Glycine-rich cell-wall proteins in bean: gene structure and association of the protein with the vacuolar system. *EMBO J* 7: 3625–3633 (1988).
6. Kyte J, Doolittle RF: A simple method for displaying the hydropathic character of a protein. *J Mol Biol* 157: 105–132 (1982).