Gene, 148 (1994) 305–308

GENE 08181

A pair of genes coding for lipid-transfer proteins in *Sorghum vulgare*

(Genomic cloning; sequence; plantlet library; maize; rice)

Florence Pelèse-Siebenbourg[a], Carme Caelles[a], Jean-Claude Kader[b], Michel Delseny[c] and Pere Puigdomènech[a]

[a]*Departamento de Genética Molecular, CID-CSIC, Jordi Girona 18, 08034 Barcelona, Spain;* [b]*Laboratoire de Physiologie Céllulaire et Moléculaire, URA CNRS 1180, Université Pierre et Marie Curie Paris VI, 75252 Paris Cedex 05, France. Tel. (33-1) 4427-5918; and* [c]*Laboratoire de Biologie Moléculaire et Physiologie Végétales, URA CNRS 565, Université de Perpignan, 66860 Perpignan Cedex, France. Tel. (33) 6866-2119*

Received by J. Messing: 11 March 1994; Revised/Accepted: 26 April/3 May 1994; Received at publishers: 16 June 1994

SUMMARY

Approximately five genes coding for lipid-transfer proteins (LTP) can be detected in *Sorghum vulgare* by DNA blots using a specific genomic probe. Two of these genes have been identified and sequenced. The two genes (*ltp1* and *ltp2*) code for very similar (91.8% identity) proteins, they are separated by approx. 4 kb of DNA and their open reading frames may be read in the same direction. The gene (*ltp1*) located upstream has an intron placed in the same position already described for other *ltp* in maize and rice. Gene *ltp2* has no intron. cDNAs corresponding to *ltp1* have been identified in a 6-day-old plantlet library, but not for *ltp2*. The results of the comparison between the two sequences indicate the presence of a gap between the two genes in their promoter region. LTP seem to be coded for in plants by a small family of genes. At least in sorghum, two of its components are tightly clustered in the same genomic region.

INTRODUCTION

A number of different functions have been attributed to lipid-transfer proteins (LTP). These include the transport of lipids between membranes (Kader et al., 1984) or antibacterial and antifungal activity (Molina et al., 1993). The LTP share homology with a large family of proteins that includes proteins that are supposed to have either storage or protease and amylase inhibitory functions (Mundy and Rogers, 1986; Henrissat et al., 1988). Genes or cDNAs encoding proteins of the LTP family have been cloned due to their properties of being induced in barley aleurone during germination (Linnestad et al., 1991) or low temperature (Hughes et al.,1992), induced by salt in tomato (Torres-Schumann et al., 1992), present in the medium of carrot somatic embryo cultures (Sterk et al., 1991) or having a tapetum-specific expression in tobacco (Koltunow et al., 1990). In fact it has been shown that genes encoding LTP have a high tissue specificity both in carrot (Sterk et al., 1991), tobacco (Fleming et al., 1992) and maize (Sossountzov et al., 1991), where an accumulation in the epidermic cell layer of the embryo has been observed.

The fact that different possible functions have been proposed for LTP has suggested the possibility that different genes code for similar proteins that have different functions or expression patterns. It has been proposed, for instance, that a mechanism of alternative splicing may exist in the RNA coding for these proteins in maize (Arondel et al., 1991). Moreover, the data on the genomic structure of LTP indicate that a number of different genes code for this protein. Therefore, different subfamilies within the LTP-encoding gene family could be responsible for the different patterns of expression or

Correspondence to: Dr. P. Puigdomènech, Departament de Genètica Molecular, CID-CSIC, Jordi Girona 18, 08034 Barcelona, Spain. Tel. (34-3) 204-0600; Fax (34-3) 204-5904; e-mail: pprgmp@cid.csic.es

Abbreviations: aa, amino acid(s); bp, base pair(s); cDNA, DNA complementary to RNA; kb, kilobase(s) or 1000 bp; LTP, lipid-transfer protein(s); *ltp*, gene(s) encoding LTP; nt, nucleotide(s); ORF, open reading frame; *Sv*, *Sorghum vulgare*.

SSDI 0378-1119(94)00392-6

functions attributed to these genes. In the present paper the evidence of the existence of a genomic structure of two genes coding for LTP placed in tandem in the sorghum genome is presented.

EXPERIMENTAL AND DISCUSSION

(a) Genomic cloning and sequencing

A genomic library of *Sorghum vulgare (Sv)* in λEMBL4 (a gift of Dr. Claude Crétin, Laboratoire de Physiologie Végétale Moléculaire, Orsay, France) was screened with a maize cDNA *ltp* probe (Tchang et al., 1988) using standard procedures. A number of positive clones were isolated. One of the clones showed the presence of two types of inserts hybridizing to the probe, indicating the possibility that the insert had two distinct nt sequences placed in tandem. The fragments hybridizing to the probe were subcloned and sequenced and the results confirmed this hypothesis. Two segments of 1499 bp (corresponding to gene *ltp1*) and 1800 bp (corresponding to gene *ltp2*) were sequenced. The two genes homologous to maize LTP cDNA are separated by 4 kb as calculated by restriction mapping (Fig. 1). It appeared that two ORFs coding for proteins very similar to maize LTP and very similar among themselves were contained in this region. The two genes were called *ltp1* and *ltp2*. The nt sequences have been submitted to EMBL database under accession Nos. X71667 and X71668 for *ltp1* and *ltp2*, respectively.

A cDNA library, constructed in λgt11 using poly(A)$^+$RNA from 6-day-old *Sv* plantlets, was screened and six positive clones were obtained and sequenced. The six clones were different in length but they had the same sequence at their 3′ ends. This sequence was identical to that of gene *ltp1*, except for the presence of an intron in the region corresponding to the C-terminal domain of the protein. The cDNA sequence has also been submitted to EMBL database under accession No. X71669.

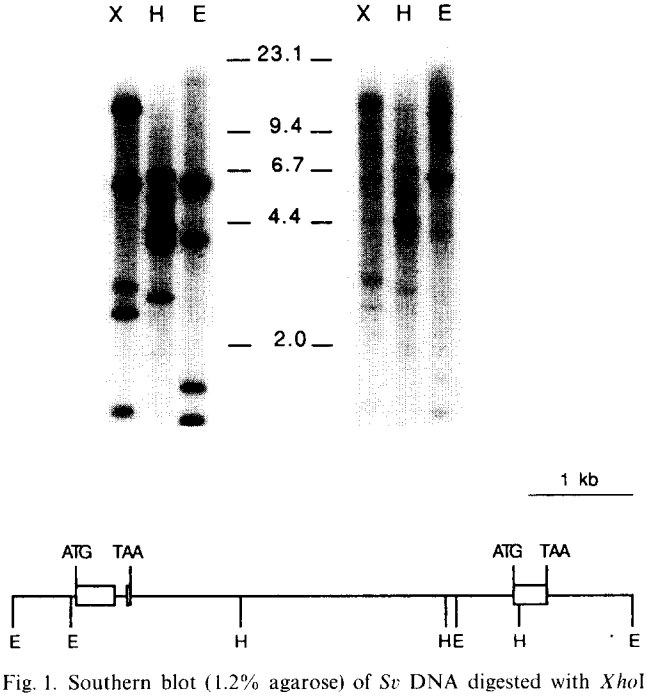

Fig. 1. Southern blot (1.2% agarose) of *Sv* DNA digested with *Xho*I (lane X), *Hin*dIII (lane H) and *Eco*RI (lane E) and hybridized with a cDNA probe (left) and a *Hin*dIII-*Acc*I probe specific for gene *ltp2*. The same filter was consecutively hybridized with the two probes. In the lower part is a map of the two genes in *Sv* genomic clones. Gene *ltp1* is located upstream (left in the figure) from *ltp2*.

(b) Analysis of the nt sequence

The nt sequence of the regions hybridizing to the maize *ltp* probe has been obtained and the coding regions can easily be recognised by comparison to the already published LTP sequences. The structure of the *ltp* genes reported so far in plants is simple. Only a single intron has been found in barley (Linnestad et al., 1991), tobacco (Fleming et al., 1992) and rice (Vignols et al., 1994), all in the same position within the region corresponding to the C terminus of the protein. The existence of an intron had already been postulated by the presence of an intervening sequence in maize cDNA clones (Arondel et al., 1990). When the two genes from *Sv* are compared between themselves and to the cDNA sequence, it appears that *ltp1* has a 114-bp intron in the same place than the other plant *ltp* genes, while *ltp2* lacks the intron. This is the first case where an intron is lacking in a genomic sequence of *ltp* genes. The comparison of sequences of genomic and cDNA clones shows that the six cDNAs sequenced correspond to *ltp1* transcripts (result not shown).

The two genes can easily be aligned in the coding region with the exception of the intron. The transcribed 3′-end is very variable and it shows stretches of identity intermixed with long fragments having no homology. In this region the overall homology is 78.5%. The 5′ non-coding region can also be aligned between the two genes (Fig. 2). In the sequence immediately upstream from the ATG it appears that the transcribed leader regions have a low degree of similarity but the identity of the sequences increases when the putative transcription initiation site is reached. Upstream from this point the similarity is very high except for a large gap between the two sequences that corresponds to the fragment between −184 and −457 of *ltp1*. This region is flanked by a number of small repeated sequences (underlined in Fig. 2) and it is highly rich in thymidine. This insertion is approximately palindromic and thus it shears some of the features of the *tourist* elements described in maize (Bureau and Wessler, 1992), although there is no significant sequence similarity between this element and the *Sv* sequence here described.

Fig. 2. Comparison of the genomic sequences of the 5′ region upstream from the ATG start codon of genes *ltp1* and *ltp2*. The small repetitive sequences located at the two ends of the gap existing between the two genes are underlined. A number palindromic sequences conserved between the two genes are boxed. Dashed lines indicate the gaps necessary to align the sequences and colons show identical nt.

Further upstream, a region with high similarity between the two genes is also observed. It is interesting to note that in this 5′ region it is possible to observe the existence of a number of conserved sequences that are formed by palindromic sequences (boxed in Fig. 2). In summary, gene *ltp2* has some of the features of *ltp1*, in particular in the promoter indicating that it is probably not a pseudogene although no cDNA has been found in the plantlet cDNA library and it lacks the intron.

(c) The aa sequence

The proteins encoded by the two *ltp* genes identified in *Sv* show a high degree of identity. Their sequences are shown on Fig. 3 aligned with the LTP polypeptides from maize (Tchang et al., 1988) barley (Molina et al., 1993), rice (Vignols et al., 1994), spinach (Bouillon et al., 1987), tomato (Torres-Schumann et al., 1992) and wheat (Dieryck et al., 1992). Only the sequences of the mature proteins have been shown. In fact, the main difference between the two sorghum sequences is that the signal peptide of the protein encoded by gene *ltp2* is 3-aa longer than the one encoded by gene *ltp1*.

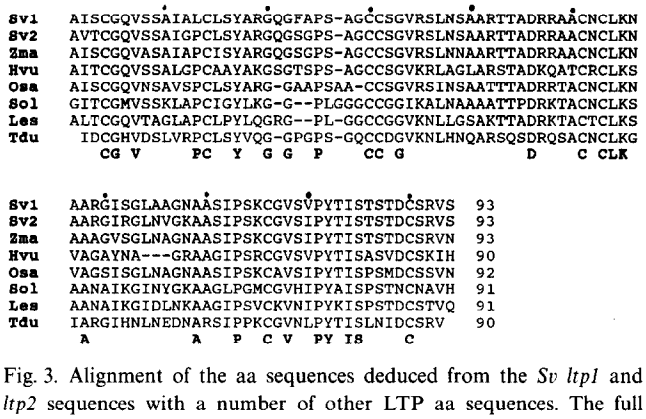

Fig. 3. Alignment of the aa sequences deduced from the *Sv ltp1* and *ltp2* sequences with a number of other LTP aa sequences. The full mature *Sv* LTP (Sv1 and Sv2) sequences (excluding the putative signal peptide) are compared to the maize (Zma), barley (Hvu), rice (Osa), spinach (Sol), tomato (Les) and wheat (Tdu) LTP. The total number of aa residues is indicated for each sequence. The conserved aa are shown in the bottom lines.

In Fig. 3 the two aa sequences are compared with the sequences of LTP from other species both cereals and dicots. At least 31 sequences from 16 different species have been published for LTP (Vignols, 1993). The sequences in the alignment presented have been chosen in order to emphasize the high similarity existing between some of the sequences of related species (such as *Sv*, maize, barley and rice), between LTP of unrelated species (such as spinach and tomato) and the relatively low similarity between different LTP sequences of related species (such as *Sv* and wheat). The sequences have been ordered in the figure in relation to the proximity to the *Sv* LTP1. Comparison between *Sv* LTP1 and tomato or spinach LTP give a percentage of matching that is in general lower than that between LTP from cereals. However, the lowest similarity of the sequences presented is between *Sv* LTP1 and the wheat LTP. This result is in agreement with the existence of distinct subfamilies of LTP in the different species.

(d) DNA blot analysis of *ltp* genes in *Sv*

The number of genes coding for a given mRNA can be estimated by DNA blot analysis. This is shown on Fig. 1 together with the structure of the *Sv ltp* genes found by genomic cloning and sequencing. The filters were hybridized with a probe corresponding to the *Sv* cDNA and with a 3′ *ltp2* probe. It appears that around five genes could be approximately counted, while some of the bands disappear when using a short gene-specific probe.

(e) Conclusions

LTP are encoded in *Sv* plants by approx. 5 genes that code for relatively similar proteins. At least two of these

genes are placed in the same genomic region. In barley, two of the distinct cDNAs described are present in chromosome 3H as results from the analysis of wheat-barley addition lines (Molina and García-Olmedo, 1993). In our case we demonstrate that in *Sv* two *ltp* genes are forming a tandem within 5 kb. Gene *ltp1* is probably the most highly expressed *ltp* gene in *Sv* plantlets because homologous cDNAs are the only ones that have been found in the cDNA library. It contains an intron as it is the case for the majority of *ltp* genes in plants, located in the same position in relation to the protein sequence. The intron in *Sv* is very short (114 bp) as it is the case in the other published cereal *ltp* genes. Instead, the tobacco *ltp* gene (Fleming et al., 1992) contains a 920-bp long intron. Gene *ltp2* does not contain any intron and it lacks a long nt stretch in the promoter region that may belong to a family of palindromic sequences similar to the *tourist* repeated element described in maize (Bureau and Wessler, 1992). However in the 5′ region a number of boxes are well conserved indicating that the essential elements needed for directing its expression may be present in the gene. It is possible that the gene is expressed in organs of the plant or in physiological situations not studied by now. *Sv* LTP protein sequences are clearly homologous with the maize, rice and barley proteins. They show a lower but still very high similarity to the tomato or spinach proteins reported. Interestingly, one of the wheat sequences published is less similar to the *Sv* sequences than the dicot sequences. It appears that the tandem of genes found in *Sv* forms a small subfamily of genes that may be as high as five or six members according to the Southern blot results. Other genes less similar to the *ltp* genes described here may still exist in the *Sv* genome. The complex structure of this family of genes may explain the different functions and patterns of expression attributed to *ltp* and their products.

ACKNOWLEDGEMENTS

The authors acknowledge the gift of *Sv* libraries from Dr. C. Crétin (Laboratoire de Physiologie Végétale Moléculaire, Orsay, France) to whose memory this article is dedicated. F.P. was a recipient of a EC fellowship (BRIDGE) and C.C. from Generalitat de Catalunya. The work has been supported by Laboratoire Européen Associé CNRS-CSIC and grant BIO91/648 to P.P. from Plan Nacional de Investigación Científica y Técnica.

REFERENCES

Arondel, V., Tchang, F., Baillet, B., Vignols, F., Grellet, F., Delseny, M., Kader, J.-C. and Puigdomènech, P.: Multiple mRNA coding for phospholipid-transfer protein from *Zea mays* arise from alternative splicing. Gene 99 (1991) 133–136.

Bouillon, P., Drischel, C., Vergnolle, C., Duranton, H. and Kader, J.-C.: The primary structure of spinach-leaf phospholipid transfer protein. Eur. J. Biochem. 166 (1987) 387–391.

Bureau, T.E. and Wessler, S.R.: *Tourist*: a large family of small inverted repeat elements frequently associated with maize genes. Plant Cell 4 (1992) 1283–1294.

Dieryck, W., Gautier, M.F., Lullien, V. and Joudrier P.: Nucleotide suquence of a cDNA encoding a lipid transfer protein from wheat (*Triticum durum* Desf.). Plant Mol. Biol. 19 (1992) 707–709.

Fleming, A.J., Mandel, T., Hofmann, S., Sterk, P. De Vries, S. and Kühlemeier, C.: Expression pattern of a tobacco lipid transfer protein gene within the shoot apex. Plant J. 2 (1992) 855–862.

Henrissat, B., Popineau, Y. and Kader, J.-C.: Hydrophobic cluster analysis and lipid transfer proteins. A domain homology between storage and lipid transfer proteins. Biochem. J. 255 (1988) 901–905.

Hughes, M.A., Dunn, M.A., Pearce, R.S., White, A.J. and Zhang, L.: An abscisic-acid-responsive, low temperature barley gene has homology with a maize phospholipid transfer protein. Plant Cell Environ. 15 (1992) 861–865.

Koltunow, A.M., Truettner, J., Cox, K.H., Wallroth, M. and Goldberg, R.D.: Different temporal and spatial gene expression patterns occur during anther development. Plant Cell 2 (1990) 1201–1224.

Linnestad, C., Lönnenborg, A., Kalla, R. and Olsen, O.A.: Promoter of lipid transfer protein gene expressed in barley aleurone cells contains similar *myb* and *myc* recognition sites as the maize Bz-McC allele. Plant Physiol. 97 (1991) 841–843.

Molina, A. and García-Olmedo, F.: Developmental and pathogen-induced expression of three barley genes encoding lipid transfer proteins. Plant J. 4 (1993) 983–991.

Molina, A., Segura, A. and García-Olmedo, F.: Lipid transfer proteins (nsLTP) from barley and maize leaves are potent inhibitors of bacterial and fungal plant pathogens. FEBS Lett. 316 (1992) 119–122.

Mundy, J. and Rogers, J.C.: Selective expression of a probable amylase/protease inhibitor in barley aleurone cells: comparison to the barley amylase/subtilisin inhibitor. Planta 169 (1986) 51–63.

Sossountzov, L., Ruiz-Avila, L., Vignols, F., Jolliot, A., Arondel, V., Tchang, F., Grosbois, M., Guerbette, F., Miginiac, E., Delseny, M., Puigdomènech, P. and Kader, J.-C.: Spatial and temporal expression of a maize lipid transfer protein gene. Plant Cell 3 (1991) 923–933.

Sterk, P., Booij, H., Schellenkens, G.A., Van Kammen, A. and De Vries, S.C.: Cell-specific expression of the carrot EP2 lipid transfer protein gene. Plant Cell 3 (1991) 907–921.

Tchang, F., This, P., Stiefel, V., Arondel, V., Morch, M.D., Pagès, M., Puigdomènech, P., Grellet, F., Delseny, M., Bouillon, P., Huet, J.C., Guerbette, F., Beauvais-Cante, F., Duranton, H., Pernollet, J.C. and Kader, J.-C.: Phospholipid transfer protein: full-length cDNA and amino acid sequence in maize. Amino acid sequence homologies between plant phospholipid transfer proteins. J. Biol. Chem. 263 (1988) 16849–16856.

Torres-Schumann, S., Godoy, J.A. and Pintor-Toro, J.A.: A probable lipid transfer protein gene is induced by NaCl in stems of tomato plants. Plant Mol. Biol. 100 (1992) 749–757.

Vignols, F.: Etude de l'Expression des Gènes Codant pour les Protéïnes de Transfert de Lipides chez le Riz (*Oryza sativa* L.). Thèse de l'Université Pierre et Marie Curie, Paris VI, 1993.

Vignols, F., Lund G., Pammi, S., Trémousaygue, D., Grellet, F., Kader, J.-C., Puigdomènech, P. and Delseny, M.: Characterization of a rice gene coding for a lipid transfer protein. Gene 142 (1994) 265–270.