Gene, 147 (1994) 205–208

GENE 08174

Short Communications

Structure and characterization of the gene encoding the ferredoxin-NADP reductase-binding protein from *Zea mays L.*

(Maize; chloroplast; sequence; genomic cloning; transit peptide)

Silvina Pessino[a], Carme Caelles[b], Pere Puigdomènech[b] and Rubén H. Vallejos[a,b]

[a]*Centro de Estudios Fotosintéticos y Bioquímicos (CONICET, F.M. Lillo and Universidad Nacional de Rosario), 2000 Rosario, Argentina. Tel. (54)41371955. Fax (54)41370044; and* [b]*Departamento de Genética Molecular, Centro de Investigación y Desarrollo, CSIC, 08034 Barcelona, Spain*

Received by: J.L. Slightom: 312 January 1994; Revised/Accepted: 2 May/ 12 May 1994; Received at publishers: 13 June 1994

SUMMARY

A genomic clone encoding ferredoxin-NADP reductase binding protein (BP) from *Zea mays L.* was sequenced and characterized. The promoter region (692 bp) shows several motives resembling those involved in enhancement, tissue-specific expression and light regulation in plants, besides the typical TATA and CAAT boxes. The coding sequence is interrupted by two introns. The deduced amino acid (aa) sequence corresponds to 22.85 kDa for the precursor polypeptide, including a transit peptide of 64 aa and a mature protein of 148 aa.

INTRODUCTION

Chloroplast ferredoxin-NADP$^+$ reductase interaction with thylakoids involves a polypeptide about 16.5 kDa (Vallejos et al., 1984; Chan et al., 1987), named reductase-binding protein (BP). The N-terminal sequence and immunological studies (Soncini and Vallejos, 1989) showed that BP is identical to the 16.5-kDa polypeptide described as a component of the oxygen-evolving complex. There is both physical and strong functional evidence that this polypeptide is located on the stromal side of thylakoids (Chan et al., 1987; Soncini and Vallejos, 1989): (*i*) Antibodies against BP agglutinated thylakoids depleted of reductase and precipitated the reductase-BP complex; (*ii*) trypsin proteolyzed BP exposed on the stromal side of thylakoids; (*iii*) Fab fragments of the IgG anti-BP prevent rebinding of reductase-depleted thylakoids and (*iv*) the same Fab fragments inactivate NADP+ photoreduction activity of thylakoids depending on ferredoxin concentration. Although it is generally accepted that there is a specific reductase binding protein involved in binding there is still some controversy about how the binding takes place.

The gene coding for the 16.5-kDa polypeptide is nuclear encoded. cDNA clones from spinach, pea and *Chlamydomonas reinhardtii* have been sequenced as well as a genomic clone from the latter (Jansen et al., 1987; Mayfield et al., 1989). The *Chlamydomonas* gene is single-copy. No sequence information is available of monocot species and the genomic organization of this gene in higher plants is unknown. In this paper we report the cloning, sequencing, and characterization of the *bp* gene from *Zea mays L.*

Correspondence to: Dr. P. Puigdomènech, 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); BP, ferredoxin-NADP reductase-binding protein; *bp*, gene encoding BP; *C.*, *Chlamydomonas*; kb, kilobase(s) or 1000 bp; NADP, nicotinamide-adenine dinucleotide phosphate; nt, nucleotide(s); *Z.*, *Zea*.

EXPERIMENTAL AND DISCUSSION

(a) Cloning of the genomic sequence of the maize *bp*

A *Z. mays* (W22) genomic λEMBL4 library from Clontech was screened by in situ plaque hybridization

SSDI 0378-1119(94)00385-6

```
-691  CATTCTAATAATAATAATTTAGACATATATCAATTAAGTTAATCCAGTTTTATGCAAAATGTATTTGTATACTATTATTAGCAAGATATCGGAGATATTTATGTGCTACATTTTTACTAT

-571  AGATGAGTGAAACAGAGAGTGTCATGTAAGTTATAGAGTAGAAACAAATTCTACTAATGCATAAAATCATTTGACATCCTTCACCCCATTAATTTGATATAGGCTTATATCTGAACTTTG
                                                                                                                        P
-451  AAAAGTGGTGAAATATCAAATTCCAAACTAAATAAGTTACTTTATTGAGTGATCCAAACGCCTCATTAGGAAAAATTGGAAATCCAAGGTTATGAGGTTACCAAACTAGTACTAGTTGGC
                                                                                                  P                   P
-331  AAAATTTCGGCTAAATTAAATCTGGGAAGCATTCTGTGTGTTATTTGACAGCTGCTACCGACCAACCCAAGGGCATGCCGAGAGTGGTCTCGAACAGTATTTGACACGTGTCAGCTCTCC

-211  ATTGGCAGTGACGCAATCCTGAACGGAGGCTGCAAGTGGTTCACGCGGCCACATGGCGCTCACACACTGGCCATCAACTTATCCAGTCAGATAAGGCGGCGCCTCCAACTGCGAGCGACG

 - 91  TATATGTTTGCACGTGCCACTCGGGCAAGAGACTGCTCGACCGGACCGACGAGCGACGAAGGTGTAGAAAGAGAAGAACCTCGGAGCTGAGATGGCACAAGCCATGGCGTCGATGACCGG
   1                                                                                                                  M  A  Q  A  M  A  S  M  T  G
  30  CCTCTCGCAGGGTGTCTGCCCAGCAGCCGCCGACAGCAGAACCAGGACGGCCGTCGTCGTCGTCAGGGCGTCCGCGGAGGGCGACCGTTGCGCAGGCGGGCCGCGCTGTGATCGACTGGT
  11   L  S  Q  G  V  C  P  A  A  A  D  S  R  T  R  T  A  V  V  V  V  R  A  S  A  E  G  D  R  C  A  G  G  P  R  C  D  R  L  V
 150  GGCCACTGCATCGTCGCCGCCCTTGTCGCAGGCGGTGCACGCCGAGACCGTCAAGACCATCAAGATCGGAGCCCCGCCGCCGCCCTCCGGTGGACTCCGTGAGTGCCTGCCTGCCTGCCC
  51   A  T  A  S  S  P  P  L  S  Q  A  V  H  A  E  T  V  K  T  I  K  I  G  A  P  P  P  P  S  G  G  L
 270  GCCCTTCATATACAGTAGCCCTGTTTCTTGTATATATACGTCTCTTGAATCTTGATCGCGAGTTCAGAATCGTTTGGAGCTGATTTATAAGGGTTTAGTTTAGATTAAAAACATCGGCAA

 390  AGGATAGGTACAGATATACGAGCTCAGATCGAAAACCCCATATGTTTGCAGCTTTGCACAAAAAAAAAAATCATCTGACATTTGCTATGCACAGCCGGGACTCTCAACTCTGACCAGACGA
  83                                                                                                                  P  G  T  L  N  S  D  Q  T
 510  GGGACTTCGACCTCCCCTTGAAGGAGCGGTTCTACCTGCAGCCGCTGCCGCCGGCCGAGGCAGTGGCGAGGGTGAAGACGTCGGCGCAGGACATCATCAACCTCAAGCCGCTCATCGACA
 123  R  D  F  D  L  P  L  K  E  R  F  Y  L  Q  P  L  P  P  A  E  A  V  A  R  V  K  T  S  A  Q  D  I  I  N  L  K  P  L  I  D
 630  AGAAGGCGTGGCCGTACGTCCAGAACGACCTCCGCCTCCGGGCCTCCTACCTGCGCTACGACCTCAAGACCGTCATCGCTTCCAAGCCCAAGGAGGAGAAGAAGAGCCTCAAGGAGCTCA
 163  K  K  A  W  P  Y  V  Q  N  D  L  R  L  R  A  S  Y  L  R  Y  D  L  K  T  V  I  A  S  K  P  K  E  E  K  K  S  L  K  E  L
 750  CCGGCAAGCTCTTCAGCACCATCGACGATGTAAGTTAACCAACCCCATTGAGATCTGGGAACAGTGCAATTTTGTTTCTCCTCCAGTTTTTGACAAGAACTGTGTCTCGTTTTGGCAAGC
 203  T  G  K  L  F  S  T  I  D  D
 870  TTGACCATGCCGCGAAGATGAAGAGCACCCCTGAGGCCGAGAAATACTTCGCAGCAACCAAAGATGCTCTTGGCGATGTTCTCGCCAAGCTAGGCTAGACGCAAGGGCTAATAATGGCCA
 213  L  D  H  A  A  K  M  K  S  T  P  E  A  E  K  Y  F  A  A  T  K  D  A  L  G  D  V  L  A  K  L  G  *
 990  TGTAATTTCGGACTCGTGTTTGTTCATATGGATCCGGCAAGGCAATGTACCATTTTCTGTTGATATTGTATCAGGAAGC
```

Fig. 1. The nt sequence of the *bp* gene from maize: upper-case letters indicate coding regions. Lower-case letters indicate non-coding regions. Eukaryotic promoter elements (Cap signal, CAAT and TATA elements) are underlined, as well as directed and inverted repeated sequences within the promoter. Palindromic sequences are indicated by P. In-frame stop codon TAG is indicated by an asterisk (*). The nt sequence of this 1.9 kb fragment was determined according to the enzymatic chain-termination procedure using sequence-specific as well as commercially available primers. Sequence data were processed by computational analysis using Gene Pro or PC Gene programs. The sequence reported in the present article has been submitted to the EMBL Database under the accesion number Z26824.

with the a maize cDNA probe (Sheen et al., 1987). The identified positive phage 22B was mapped with *Bam*HI, *Sal*I, *Eco*RI, *Xba*I, *Xho*I, *Sma*I, *Hind*III and *Kpn*I. An *Eco*RI 1.9-kb fragment hybridizing to the cDNA probe and encoding the entire coding region was subcloned into pBluescript.

The coding region was identified by its similarity to the published spinach cDNA sequence (Jansen et al., 1987), but it is interrupted by a 257 bp intron at position +248 (intron 1) and a 90 bp intron (intron 2) at position +779. Fig. 1 shows a 1.76-kb sequence of this *Eco*RI fragment including the coding regions and the 5′ and 3′ untranslated regions of the *bp* gene and the deduced aa sequence. A search in the 692 nt of the promoter regions includes the typical TATA and CAAT boxes which are located at nt positions −90 and −129, respectively (Fig. 1). Different gene regulatory elements that have been described in other plant genes were identified. The genomic sequence from −400 to −690 upstream from the *bp* coding region is highly A+T rich (75%). A+T-rich regions of DNA are know to bind a variety of plant DNA binding proteins and are considered to be associated with the nuclear scaffold (Jacobsen et al., 1990).

A number of elements found within the *bp* promoter similar to those described as involved in photoregulation and tissue-specific expression and enhancement (Gilmartin et al., 1990). The hexameric G-box motif at nt position −224 has been identified in the 5′ flanking DNA region of several plant genes exhibiting regulation by a variety of environmental signals and physiological cues (Williams et al., 1992). Interestingly, the sequence is located in the middle of a 12-bp long palindromic sequence. Inverted repeats are present at several positions within the promoter sequence, as can be seen in Fig. 1. Two of these inverted repeats are large palindromic sequences. The sequence ATTTGACA is repeated directly three times at nt −501, −286 and −233. Another direct repeat is located at nt positions −426 and −349. The conserved motif ATGATAAGG, containing a single GATA element, is present in virtually all *rbcS* and *cab* genes (Gilmartin et al., 1990). Elements resembling this GATA motifs were found in four different positions in the *bp* promotor (nt positions −121, −129, −474 and −607). Sequences with more than 75% similarity to the core SV40 motif and the T-cyt box (Bichler and Herrmann, 1990) are present at nt positions −359 and −160, respectively.

(b) Nucleotide sequence and deduced aa sequence of the BP protein in *Z. mays*

The maize nt sequence that encodes the mature polypeptide shares 69.1% identity to the *bp* of spinach and 53.6% identity to that of *C. reinhardtii*. Comparison of the aa sequence that corresponds to the mature protein shows 70% overall identity between spinach and maize and 25% between *Chlamydomonas* and maize (Fig. 2) but in both cases many of the changes are conservative. The identity between the sequences of maize and pea (Ettinger and Theg, 1992) is especially high (96%). The processing site for the *Z. mays* BP was assigned by comparison with the homologous sequence in spinach. The N terminus for the mature protein was inferred to be Glu^{65}, as in spinach where the N-termianal sequence of the mature protein was determined (Soncini et al., 1989), and if so the transit peptide would be 64-aa long. The mature polypeptide is predicted to consist of 148-aa with a calculated size of 16.6 kDa. The precursor polypeptide (see below) has 212-aa residues (22.8 kDa).

The putative transit peptide and the first 18 aa of the mature protein are encoded by exon 1, the central part of the mature protein is encoded by exon 2, and the C-terminal region is encoded by exon 3 (Fig. 1). The N-terminal sequence of the transit peptide contains 8 aa that are identical to the spinach sequence (Jansen et al., 1987). This region showed no Pro, Gly, or charged aa residues. The hydropathy profile of this peptide revealed a less-pronounced hydrophobic core at the C-terminal region of the transit peptide when compared with this region in spinach, and *Chlamydomonas* transit peptides.

Introns of the *bp* gene were found to be located at +248 and +779 (Fig. 1). Intron 1 interrupts the sequence after the first nt of a codon, it belongs to the type-1 group, while intron 2 corresponds to the type-0 group. Typical GT/AG donor/acceptor sites are present at intron junctions. The sequences surrounding the splice junctions are similar to the plant consensus sequence, but are not identical (Hanley and Schuler, 1988). Recently, it was reported (Hanley and Schuler, 1988) that the purine- or pyrimidine-richness of sequences upstream from the 3′-splice site differs significantly between introns in monocotyledons and dicotyledons. It was observed that a

```
Maize     ETVKTIKIGAPPPPSGGLR-TLNSDQTRDFDLP-L-KERFYLQPLPPAEA  47
Pea       ETVKTIKIGAPPPPSGGLPGTLNSDQARDFDLP-L-KERFYLQPLPPAEA  48
Spinach   EA-RPIVVGPPPPLSGGLPGTENSDQARDGTLP-YTKDRFYLQPLPPTEA  48
Chlamydo --LTPVDLFDDRSVRD--RGFDLIYEARDLDLPQNVREGFTQARASLDET  46
             ++ +   ++ ++  +      ++** +**   ++ *  ++ + +*+
Maize     VARVKTSAQDI-INLKPLIDKKAWPYVQNDLRLRASYLRYDLKTVIASKP  96
Pea       AARVKTSAQDI-INLKPLIDKKAWPYVQNDLRLRASYLRYDLKTVIASKP  97
Spinach   AQRAKVSASEI-INVKQFIDRKAWPSLQNDLRLRASYLRYDLKTVISAKP  97
Chlamydo KKRVKESEARIDADLDVFIQKSYWTEAREQLRRQVGTLRFDLNTLASTKE  96
           *** *+  *  +++ +*+++ *+  +++** +++ *******+*+ +++
Maize     KEEKKSLKELTGKLFSTIDDLDHAAKMKSTPEAEKYFAATKDALGDVLAK  146
Pea       KEEKKSLKELTGKLFSTIDDLDHAAKIKSTPEAEKYFAATKDALGDVLAK  147
Spinach KDEKKSLQELTSKLFSSIDNLDHAAKIKSPTEAEKYYGQTVSNINEVLAK  147
Chlamydo KEAKKAALGLRKEFIQAVEDLDFALREKDQASAAKKLEITKAKLDSVLAA 146
          *+++*+  +*  +++ ++++** * + *+ +++*+  + * + +++***
Maize     LG  148
Pea       LG  149
Spinach   LG  149
Chlamydo VL  148
          +
```

Fig. 2. Comparison of the deduced mature maize BP aa sequence with those of spinach, pea and *Chlamydomonas*. Positions perfectly conserved are indicated by asterisks (*). Positions well conserved are indicated by pluses (+). Gaps within the sequence are indicated by hyphens (-).

higher proportion of dicotyledon introns is of the purine-rich type while a higher proportion of monocotyledon introns is pyrimidine-rich. Intron 1 and 2 of the *bp* gene from maize are mixed and pyrimidine-rich types, respectively, which is consistent with the previous observation.

ACKNOWLEDGEMENTS

This work was supported by grants from Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET, Argentina) and Plan Nacional de Investigación Científica y Técnica (Spain, grant BIO91–648). R.H.V. is a member of the Research Career ans SP a Fellow of CONICET. The authors are grateful to Dirección General de Investigación Científica y Técnica (Spain) for a sabbatical grant to R.H.V.

REFERENCES

Bichler, J. and Herrmann, R.G.: Analysis of the promoters of the single-copy genes for plastocyanin and subunit s of the chloroplast ATP synthase from spinach. Eur. J. Biochem. 190 (1990) 415–426.

Chan, R.L., Ceccarelli, E.A. and Vallejos, R.H.: Immunological studies of the binding protein for chloroplast ferredoxin-NADP+ reductase. Arch. Biochem. Biophys. 253 (1987) 56–61.

Ettinger, W.F. and Theg, S.M.: Sequence of the cDNA encoding the 17-kilodalton protein of the photosynthetic oxygen-evolving complex of pea. Plant Physiol. 99 (1992) 791–793.

Gilmartin, P.M., Sarokin, L., Memelink, J. and Chua N.-H.: Molecular light switches for plant genes. Plant Cell 2 (1990) 369–378.

Hanley, B.A. and Schuler, M.A.: Plant intron sequences: evidence for distinct groups of introns. Nucleic Acids Res. 16 (1988) 7159–7176.

Jacobsen, K., Laursen, N.B., Jensen, E.O., Marcker. A., Poulsen, C. and Marcker, K.A.: HMG I-like proteins from leaf and nodule nuclei interact with different AT motifs in soybean nodulin promoters. Plant Cell 2 (1990) 85–94.

Jansen, T., Rother, C., Seppuhn, J., Reinke, H., Beyreuther, K., Jansson, C., Andersson, B. and Herrmann, R.G.: Nucleotide sequence of cDNA clones encoding the complete "23 kDa" and "16 kDa" precursor proteins associated with the photosynthetic oxygen-evolving complex from spinach. FEBS Lett. 216 (1987) 234–240.

Mayfield, S.P., Schirmer-Raire, M., Frank, G., Zuber, H. and Rochaix, J.-D.: Analysis of the genes of the OEE1 and OEE3 proteins of the photosystem II complex from *Chlamydomonas reinhardtii*. Plant Mol. Biol. 12 (1989) 683–693.

Sheen, J., Sayre, R.T. and Bogorad, L.: Differential expression of oxygen-envolving polypeptide genes in maize leaf cell types. Plant Mol. Biol. 9 (1987) 217–226.

Soncini, F.C. and Vallejos, R.H.: The chloroplast reductase described as a component of the oxygen-envolving complex. J. Biol. Chem. 264 (1989) 21112–21115.

Vallejos, R.H., Ceccarelli, E.A. and Chan R.L.: Evidence for the existence of a thylakoid intrinsic protein that binds ferredoxin-NADP+ oxidoreductase. J. Biol. Chem. 259 (1984) 8048–8051.

Williams, M.E., Foster, R. and Chua, N.-H.: Sequences flanking the hexameric G-box core CACGTG affect the specificity of the protein binding. Plant Cell 4 (1992) 485–496.