Structure and expression of the nuclear gene coding for the plastid CS1 ribosomal protein from spinach

Bruno Franzetti, Dao-Xiu Zhou and Régis Mache*

Laboratoire de Biologie Moléculaire Végétale, CNRS (URA 1178) and Université Joseph Fourier, BP 53X, F-38041 Grenoble cédex, France

Received June 11, 1992; Accepted July 22, 1992

EMBL accession no. X66135

ABSTRACT

The chloroplast ribosomal protein CS1 is an essential component of the plastids translational machinery involved in translation initiation. Southern analysis suggests that the corresponding nuclear gene is present in one copy in the spinach genome. We have isolated and sequenced the gene (*rps1*) to study its expression at the transcriptional level. The gene consists of 7 exons and 6 introns including an unusually large intron in the 5' coding region. No canonical TATA-box is found in the 5' upstream region of the gene. *rps1* transcripts are detected early during germination and a significant accumulation is observed after the protrusion of the radicle. CS1 mRNAs are present in all organs of young seedlings although there are dramatic differences in the steady state level of the mRNAs between leaves and roots tissues. Transcripts accumulate independently of the presence or absence of light. Band shift analysis shows that the +1, -400 bp region of the gene can bind different sets of proteins isolated from roots and leaves nuclei. We suggest that the expression of the housekeeping plastid-related *rps1* gene is regulated in a tissue-specific manner by transcriptional *trans*-acting factors.

INTRODUCTION

Plastids biogenesis is the result of the interaction of two separate genomes. The mechanisms governing the coordinated expression of nuclear and plastid genes are largely unknown. It has been shown that plastid genes are regulated largely at the post-transcriptional level (1, 2, 3), whereas the nuclear gene expression appears to be regulated primarily at the transcriptional level (4). Most studies of transcriptional regulation have been focused on chloroplast nuclear genes involved in photosynthesis (5, 6). In contrast, there is little information about the expression of genes coding for plastid housekeeping proteins like ribosomal proteins (r-proteins). The plastid ribosome biogenesis is a complex process which requires the coordinate expression of more than 50 r-protein genes that are encoded in two genetic compartments (7). Previous studies have shown that the components of the plastid ribosome are not synchronously accumulated during the spinach

seed germination (8). In this respect, we wanted to investigate how r-protein nuclear gene expression is regulated.

We focused our interest on the nuclear gene coding for the recently characterized chloroplast r-protein CS1 from spinach (Franzetti *et al* submitted). This protein is homologous to the S1 r-protein of *E. coli* but shows specific structural features. The CS1 protein is actively engaged in the initiation complex formation via a strong mRNA-binding activity. Specifically, it possesses a poly(A)-binding activity which might play a role as a control element in chloroplast mRNA translation (Franzetti *et al* submitted). In this paper, we present the nucleotide sequence of the *rps1* gene and the analysis of its expression. The upstream region of the gene exhibits housekeeping-like features. We report on a correlation between tissue-specific mRNA differential accumulation and protein-DNA interactions in the *rps1* gene upstream region.

MATERIALS AND METHODS

Plant material

Spinach seeds (*Spinacia oleracea* L. cv Géant d'Hiver) were imbibed, germinated and grown in soil at 25°C in the light or in the dark. Seeds were collected at different stages of germination from dry seeds to three day-old seedlings after the protrusion of the roots. Dark- or light-grown leaves and roots were collected from 15 day-old seedlings. The collected material was immediately frozen in liquid nitrogen and stored at -80°C before usage.

Construction and screening of a genomic library

Genomic DNA was prepared by the method of Rogers *et al.* (9) from young spinach cotyledons. DNA fragments (around 10 kb) from partial *EcoRI* digests were cloned into the λ NM 1149-phage vector. The genomic library was screened with the α -³²P-labeled 290-bp *PstI-SalI* cDNA fragment. Standard procedures (10) were used for hybridization, genomic DNA isolation, restriction enzymes mapping, and subcloning.

DNA sequencing

Defined restriction fragments of the genomic clone were ligated into pBluescript KSII (Stratagene). The longest fragments (up to 600 bp) were deleted by exonuclease III digestion in order to obtain

* To whom correspondence should be addressed

overlapping clones (10). The nucleotide sequence was determined by using the dideoxy chain termination method (11) using double-stranded DNA and a T7 DNA polymerase sequencing kit (Pharmacia). DNA was prepared from small bacterial cultures according to a modified alkaline lysis method (12).

Southern and Northern blotting and hybridization

For Southern blot, 15 μ g of total spinach DNA were digested to completion with different restriction enzymes, fractionated in a 1% agarose gel and capillary blotted on Hybond-N⁺ filter (Amersham). For Northern blot, total RNA was extracted from frozen tissues as described (13) and 15 μ g of the RNAs were electrophoretically separated on formaldehyde gel and blotted on Hybond-N⁺ (Amersham) according to the supplier's protocol. The filters were probed with a ³²P-labeled RNA. The riboprobe was generated from the 290-bp *Pst*I-*Sal*I fragment of the previously characterized CS1 cDNA (Franzetti *et al.*, submitted) cloned into pBluescript using the RNA transcription kit from Stratagene. High stringency hybridization conditions were performed following the manufacturer's protocol.

Gel retardation analysis

Band shift analysis was done using different *rps1* genomic fragments of the 5' upstream region. A fragment containing the 5' upstream regulatory region (−26 to −441 from the translational start codon) was obtained by Exonuclease III deletion of the *Eco*RI-*Eco*RV 5' genomic fragment (10). Two subfragments, F1A and F1B (154 and 275 bp, respectively), were obtained using an internal *Eco*RI site artificially generated by mutagenesis. F1A contains the *rps1* upstream region from position −197 to +64 from the first transcription start S1 and 14 bp of the pUC 18 polylinker. F1B contains the 5' upstream part of the gene comprised between positions −197 to −351 from the first transcription start S1. Nuclear extracts from the mature leaves and roots of spinach were prepared as described by Green *et al.* (13). For binding reactions, 10 ng of 5'-end terminal γ -³²P-labeled DNA fragment were incubated with 5 to 10 μ g of nuclear extract in the presence of 2 μ g of poly(dI-dC) as non-specific competitor. The reaction mixtures were finally electrophorized on a 4% acrylamide gel in 0.5×TBE buffer and the DNA-protein complexes were revealed by autoradiography.

RESULTS

Gene copy number

To estimate the number of *rps1* genes in the spinach nuclear genome, a Southern blot of genomic DNA digests was probed with an internal fragment of the CS1 cDNA. One band was detected in all the *Eco*RI, *Pst*I or *Hind*III digests of this DNA (Figure 1). These results suggest the presence of one gene per haploid genome. A similar result has been reported for other nuclear-encoded chloroplast r-proteins (14, 15, 16).

Isolation and organisation of the *rps1* gene and of its upstream region

A genomic library constructed in the λ NM 1149 phage vector has been screened with a fragment of the previously characterized cDNA (Franzetti *et al.*, submitted) coding for the r-protein CS1. One positive clone was purified. An *Eco*RI digest of the inserted DNA releases three fragments of 4.9 kb, 2.8 kb and 2.2 kb, respectively. Hybridization with different labelled fragments of the cDNA revealed that the 4.9-kb genomic fragment contained

Figure 1. Southern analysis of spinach genomic DNA. 15 μ g of DNA were digested with *Eco*RI (E), *Pst*I (P), and *Hind*III (H). The blot was hybridized with a labeled riboprobe generated from the *in vitro* transcription of a *Sal*I-*Pst*I cDNA fragment as described in 'Materials and Methods'. The probe corresponds to the exon IX and a part of exon V. Lane W: molecular weight markers.

the entire CS1 coding sequence. Figure 2 A and B shows a restriction map and sequence of the *rps1* gene and of the 5' upstream region. The sequence of the gene matches exactly the cDNA sequence and shows that this gene is expressed. The gene encodes a polypeptide of 40 kDa and consists of seven exons separated by six introns. The size of the introns varies considerably (from 84 to 1350 bp). These introns do not contain any significant open reading frame. Exon-intron splice junctions fit well the consensus sequence (AG/GTAAG and TGCAG/G respectively) of pre-mRNA introns in plants (17). The transcription start sites of the *rps1* gene have been mapped by primer extension and S1 nuclease analysis (Lagrange *et al.*, submitted) and are reported in Figure 2.B. The start site of the largest transcript (S1) is marked +1. It is separated by a 40-bp pyrimidine-rich stretch from the second start site. There is no canonical TATA-box at a suitable distance from the two transcription initiation sites.

Expression of the *rps1* gene

Northern analysis shows that the *rps1* gene is expressed early during the germination of spinach seeds (Figure 3A). Transcripts are detectable in dry seeds and are gradually accumulated during the four days after imbibition. Then, at day 5 or 6, after the extrusion of the radicle, an increase in the amount of accumulated transcripts is observed (lane 5 and 6). The CS1 r-protein mRNA is detected in all organs (Figure 3A and B), although at a very low level in dry seeds and roots. The *rps1* transcripts are massively accumulated in leaves (Figure 3B) and the steady state level is the same in young cotyledons as in developed leaves (not shown). This accumulation is not dependent on light as shown by the comparison of the CS1 mRNA level in dark or light grown

Figure 2. Organisation of the nuclear gene coding for the plastid r-protein CS1 from spinach. **A.** Partial restriction map and structure of the *rps1* gene. White boxes corresponds to exons, black boxes to introns. Restriction sites are: *Eco* RI (E); *Hind* III (H); *Eco* RV (X); *Xho* I (X); *Bgl* I (Bg); *Pst* I (P); *Sal* I (S); *Kpn* I (K); *Xba* I (Xb). **B.** Nucleotide and deduced amino acid sequence of the *rps1* gene and sequence of the -441 upstream region of the *rps1* gene. Transcription start sites are indicated with bent arrows.

Figure 3. Northern blot analysis of the *rps1* mRNA. 15 μ g of total spinach RNA were denatured, fractionated and transferred onto a nitrocellulose membrane. The blots were probed with a CS1 riboprobe as described in 'Materials and Methods'. **A**, Accumulation of *rps1* mRNA during the germination of spinach seeds. Number on top corresponds to the development stages in days. Exposition time: 72 h. **B**, Accumulation of mRNAs in seeds (S), Leaf (L), and roots (R). Exposition time: 48 hours. **C**, Absence of light effect on the *rps1* mRNA accumulated in light (+) or dark (–) grown spinach cotyledons. Exposition time: 12 h.

spinach cotyledons (Figure 3C). From these results we conclude that *rps1* expression is probably constitutive but modulated in a tissue specific manner. This is supported by the functional analysis of the *rps1* upstream sequence as will be shown now.

Tissue-specificity of nuclear protein binding to the *rps1* upstream region

As a first step to investigate the transcriptional regulation of the *rps1* gene, a gel retardation analysis was performed in order to investigate DNA-protein interaction in the 5' upstream region. The different promoter fragments which were used in these experiments are reported in Figure 4A. The whole –400-bp fragment was shifted by nuclear protein extracts from both leaf or root tissues (fragment F1 in Figure 4B). The retardation profiles indicate that different sets of proteins factors from leaves and roots interact with the *rps1* gene promoter region. In order to further define the DNA-binding pattern the 5' upstream region was fractionated into two sub-fragments (F1A and F1B in Figure 4B). For the proximal fragment which contains the two transcription starts, a shifted band was found with leaf extracts. Although more discrete, the same retarded DNA-protein complex is present with the root extract. We conclude on the existence of a common factor interacting with this region. In contrast, a retardation assay with the upstream F1B subfragment exhibit a shifted band which is only present in leaves in addition to a second one common in both tissues.

DISCUSSION

In this paper we present the organisation and the functional analysis of the *rps1* gene coding for the chloroplast r-protein CS1. The gene contains an unusually large intron (1350 bp) as is the case for two other spinach chloroplast-related r-protein genes, the *rps22* and *rp121* genes (16 and Lagrange *et al.*, submitted). The meaning of this observation remains obscure. In contrast, the position of the introns can be interpreted in respect of the molecular evolution of the *rps1* gene. The r-protein CS1 contains

Figure 4. Nuclear proteins binding to the –400 bp upstream region of the *rps1* gene. **A**, Schematic representation of the fragments used for gel shift assays. Genomic DNA (bold line) and pUC18 vector DNA (thin line) are represented. Fragment F1 and subfragments F1A and F1B are represented. **B**, Autoradiograms of the retarded DNA-proteins complexes are shown. Lanes C: DNA probe. Lanes L: DNA-protein complex obtained with leaf nuclear extract. Lanes R: DNA-protein complex obtained with root nuclear extract.

a central core homologous to the RNA-binding domain of the *E. coli* S1 surrounded by chloroplast-specific polypeptide extensions (Franzetti *et al.*, submitted). Such extensions has been observed also in the case of several other large size chloroplast r-proteins (7). Five of the six introns are positioned in the sequence part of the gene which code for the central homologous core. No intron is present in the sequence coding for the transit peptide and the N-terminal extension. The sixth intron is located in the C-terminal part of the coding sequence. This insertion site might correspond to the end of the *rps1* ancestor gene or else could be the consequence of an intron-slipping mechanism as observed in some genes (18). In accordance with the current hypothesis on gene transfer correlative of the endosymbiotic theory (14), the ancestor *rps1* gene would have been transferred during the evolution from an ancient organelle genome to the nuclear genome and the extensions would have been added after. The localisation of the intron in *rps1* would therefore reflect the dual origin of the present gene. In several genes, the introns separate exons coding for functional domains as it has been previously reported in the case, for example, of the genes coding for the glyceraldehyde-3-phosphate dehydrogenase (18). The CS1 r-protein contains three repetitive units which are homologous to the RNA-binding domain of the the *E. coli* r-protein S1 (Franzetti *et al.*, submitted) and could form separately or together functional domains. This hypothesis is not verified by the emplacement of the introns in the gene sequence.

The features of the promoter region of the *rps1* gene are those of housekeeping genes (19) as shown by the absence of TATA box and the presence of a polypyrimidine rich sequence close to the transcription initiator. Ribosomal protein genes in animal cells have these characteristic features also (20, 21).

The presence of transcripts in roots as in leaves shows that the *rps1* gene is constitutively expressed. mRNAs are massively accumulated in green tissues. This accumulation is not dependent on light as observed for nuclear encoded photosynthetic genes (5, 22, 23), suggesting that *rps1* gene expression is transcriptionally regulated in a tissue-specific manner. A confirmation of this assumption was provided by the study of protein-DNA interaction in the -400-bp upstream region of the *rps1* gene. Band shift analysis show that in addition to constitutive factors, at least one DNA-binding protein is specific to leaves. We suggest that the gene is transcriptionally regulated by tissue-specific factors. These factors may be associated to the proplastid-chloroplast transition.

Usually, organ-specific expression of chloroplastic genes is related to their photosynthetic role and is under light control (24). Here we report on evidence for a transcriptional tissue-specific regulation of a nuclear gene coding for a chloroplast housekeeping protein. Experiments are in progress to characterize the regulatory factors and their targets.

ACKNOWLEDGEMENT

This research was supported by the Ministère de la Recherche et de la Technologie (to B. F.).

REFERENCES

1. Deng, X.-W. and Gruissem, W. (1987) *Cell*, **49**, 379–387.
2. Mullet, J.E. (1988) *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, **39**, 475–502.
3. Gruissem, W. (1989) *Cell*, **56**, 161–170.
4. Kuhlmeier, C., Green, P.J. and Chua, N.H. (1987) *Annu. Rev. Plant Physiol.*, **38**, 221–257.
5. Fluhr, R., Kuhlmeier, C., Nagy, F. and Chua, N.H. (1986) *Science*, **232**, 1102–1112.
6. Kuhlmeier, C., Cuzzo, M., Green, P.J., Goyvaerts, E., Ward, K. and Chua, N.H. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 4662–4666.
7. Mache, R. (1991) *Plant Sci.* **72**, 1–12.
8. Bisanz-Seyer, C., Li, Y.F., Seyer, P. and Mache, R. (1989) *Plant. Mol. Biol.* **12**, 201–211.
9. Rogers, S.O. and Bendich, A. (1988) *Plant Molecular Biology Manual*. AG: 1–10. Kluwer Academic publishers-Dordrecht.
10. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor Laboratory press, Cold Spring Harbor, NY.
11. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
12. Saunders, S.E. and Burke, J.F. (1990) *Nucl. Acids. Res.*, **18**, 4948.
13. Green, P.T., Kay, S.A. and Chua, N.H. (1988) *EMBO. J.*, **6**, 2543–2549.
14. Martin, W., Lagrange, T., Li, Y.F., Bisanz-Seyer, C. and Mache, R. (1990) *Curr. Genet.* **18**, 553–556.
15. Carol, P., Li, Y.F. and Mache, R. (1991) *Gene*, **103**, 139–145.
16. Bisanz-Seyer, C. and Mache, R. (1992) *Plant Mol. Biol.*, **18**, 337–344.
17. Brown, J.W.S. (1986) *Nucleic Acids Res.*, **14**, 9549–9559.
18. Quigley, F., Martin, W.F. and Cerff, R. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 2672–2676.
19. Mager, W.H. (1988) *Biochim. Biophys. Acta*, **949**, 1–15.
20. Dudov, K.P. and Perry, R.P. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 8545–8549.
21. Harihanan, N. and Perry, R.P. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 1526–1530.
22. Langdale, J.A., Rothermel, B.A. and Nelson, T. (1988) *Genes Dev.*, **2**, 106–115.
23. Tobin, E.M. and Silverthorne, J. (1985) *Annu. Rev. Plant Physiol.*, **36**, 569–593.
24. Guilmarin, P.M., Sarokin, L., Memelink, J., and Chua, N.-H. (1990) *Plant Cell.*, **2**, 369–378.