Volume 16 Number 6 1988 Nucleic Acids Research

Expression of the *rpl23*, *rpl2* and *rps19* genes in spinach chloroplasts

Frank Thomas, Olivier Massenet, Anne-Marie Dorne, Jean-François Briat and Regis Mache*

Laboratoire de Biologie Moléculaire Végétale, CNRS UA 1178, Université de Grenoble I, BP 68, F-38402, St Martin d'Hères, France

Received December 29, 1987; Revised and Accepted February 22, 1988

ABSTRACT

The expression of the spinach *rpl23*, *rpl2* and *rps19* chloroplast genes has been studied. The *rpl23* gene identified in tobacco and *Marchantia*, is split into two overlapping reading frames in spinach. S1 mapping has shown that initiation sites could occur upstream of each reading frames. A large transcription unit is also present covering the *rpl2* and *rps19* genes. The *rps19* and *rpl2* gene products are identified by NH2-terminal amino acid sequences. They correspond to spinach chloroplast ribosomal proteins CS-S23 and CS-L4, respectively. No product of the *rpl23* gene was detected in the chloroplast 50S ribosomal subunit. This strongly suggest that a corresponding gene has been transfered into the nucleus.

INTRODUCTION

The complete nucleotide sequence of the tobacco and of the liverwort *Marchantia polymorpha* chloroplast DNA have been recently determined (1, 2). Nineteen different sequences coding for polypeptides homologous to *Escherichia coli* ribosomal proteins have been found. Two differences appear between the two genomes: in *Marchantia* the absence of the *rps16* gene on the large single copy and the presence of an additional *rpl21* gene on the small single copy have been noticed. The number of ribosomal protein genes encoded in chloroplast DNA (cpDNA) is in good agreement with the number of ribosomal proteins synthesized in isolated chloroplasts (3,4).

In the tobacco chloroplast genome, Tanaka *et al.* (5) described a ribosomal protein gene cluster (*rpl23*, *rpl2*, *rps19*, *rpl22*, *rps3* and *rpl16*) whose arrangement is similar to the S-10 operon of *E. coli*. Northern experiments have shown that these chloroplast genes are transcribed. A precise analysis of the transcripts has to be done. A correlation has to be established between putative gene and ribosomal protein.

The spinach chloroplast DNA has been partly sequenced in the region corresponding to the S10-like operon showing the presence of the *rpl23* and *rps19* genes (6). In this paper, results about the organization and expression of the first three ribosomal protein genes of the S10-like operon in spinach are presented.

MATERIALS AND METHODS

Isolation of ribosomes and separation of ribosomal proteins

Ribosomes were prepared and ribosomal proteins isolated as previously described (7,8). Ribosomal proteins were separated by two-dimensional electrophoresis using system I of Madjar *et al* (9). The ribosomal proteins were numbered according to Mache *et al* (7) The ribosomal proteins L4 and S23 were purified by high performance liquid chromatography (10 and unpublished results). The NH_2 terminal sequences of these two proteins were determined by automatic Edman degradation (11) and were performed at the Service Central d'Analyse of CNRS.

Cloning and preparation of DNA

Chloroplast DNA was prepared as described in (12) Cloning experiments were performed using conventional methods described in Maniatis *et al* (13). Sal I-9 fragment of spinach chloroplast DNA was digested by PstI and inserted into the sites SalI and PstI of the plasmid pUC8. The recombinant plasmids containing the *rps19* gene were detected, characterized and one of them was designated pFT 1. This recombinant plasmid DNA was prepared by the method of Birnboim and Doly (14) followed by centrifugation to equilibrium in cesiumchloride ethidium bromide gradients. Purification of restriction fragments of DNA was achieved by electroelution when electrophoresis were performed in agarose gel and by the crushing-soaking method of Maxam and Gilbert (15) for DNA fragments separated by polyacrylamide gel electrophoresis. The 3' extremities of the EcoRI site and of the SalI site (Fig.1) were labeled respectively with (α^{32}P) ddATP and (α^{32}P) dTTP and the large Klenow fragment according to Maniatis *et al* (13). 5' end labeling of the BamHI-XhoI DNA fragment was done as described in Maniatis *et al* (13)

DNA-directed translation assays

The prokaryotic DNA-directed translation kit (Zubay system Amersham) was used with 5 µg of the recombinant plasmid DNA from pFT 1 and with 1,875 MBq of ^{35}S-Methionine (30 TBq/mM, Amersham). The translation

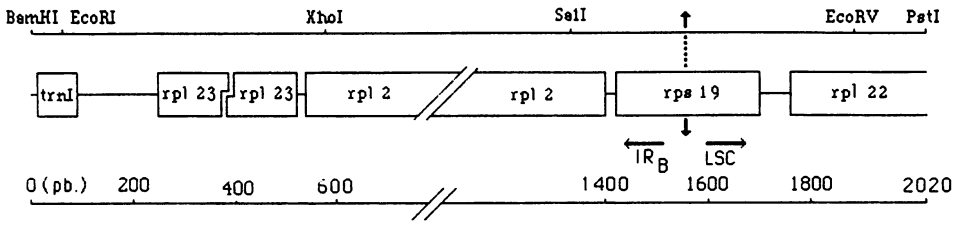


Figure 1: Localization of the different genes on a BamHI-PstI DNA fragment from spinach chloroplast located at the junction between IRB and LSC. The arrows indicate the position of the junction between the inverted repeat sequence (IRB) and the large single copy (LSC). *trnI*, *rpl23*, *rpl2*, *rps19* and *rpl22* are genes coding respectively for the tRNA-Ile(CAU), and the putative ribosomal proeins CL-23, CL-2, CS-19 and CL-22. The *rpl23* gene is organised in Spinach in two different ORFs as shown by the two boxes on the figure. The different restriction sites used in S1 mapping are indicated.

products were added to 30S ribosomal proteins and were analysed by 2D-electrophoresis as indicated above. After electrophoresis, gels were stained with Coomassie blue, dried and exposed to Kodak XAR-5 films and Dupont Cronex Quanta IIB intensifying screens at -70° C for 5 days.

S1 mapping

Chloroplastic RNAs used for S1 mapping were prepared as previously described (16) using intact chloroplasts purified by discontinuous Percoll gradients according to Douce *et al.* (17).

The two strands of end labeled DNA fragments were separated by non-denaturing electrophoresis on 4% polyacrylamide gel (40 cm x 16 cm x 0.15 cm) run at 250 volts and 4° C as described by Maniatis *et al.* (13) Single stranded probes were hybridized to RNAs and the hybrids were treated by S1 nuclease as described by Link and Langrige (18). Products resulting from this S1 treatment were analysed on 6% or 8% polyacrylamide 8M urea sequencing gels (40 cm x 16 cm x 0.04 cm) at 23 mA according to Maxam and Gilbert (13). These gels were indexed by running sequencing reactions (13) of the respective single strand end labeled DNA probes used in the different S1 mapping experiments.

RESULTS

The *rpl23*, *rpl2*, *rps19* and *rpl22* genes in spinach chloroplasts. The nucleotide sequence of part of a BamHI-PstI fragment of spinach chloroplast DNA has been previously determined by Zurawski *et al.*, (6) and

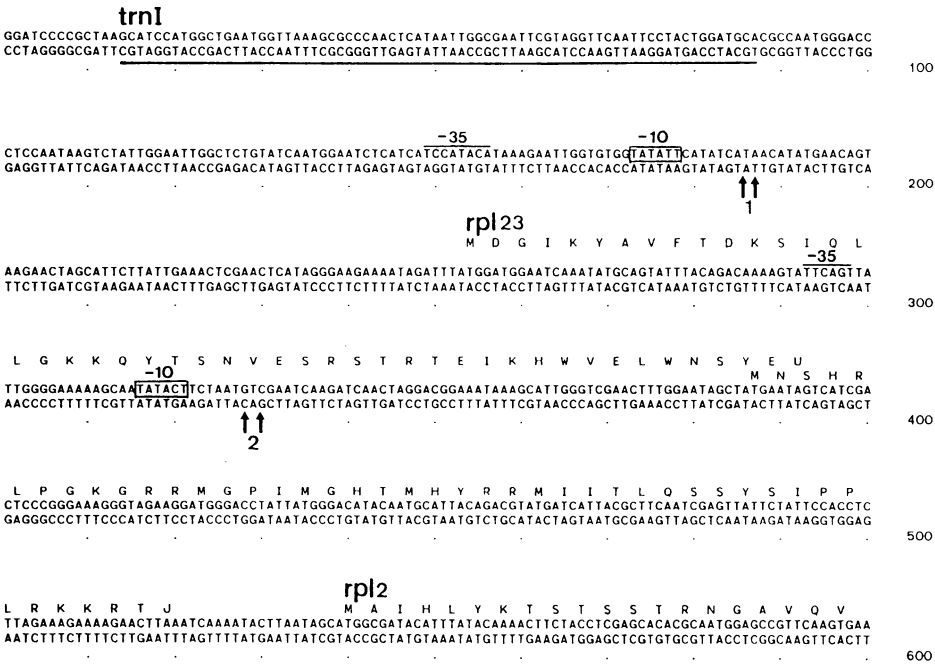


Figure 2: DNA nucleotide sequence of the *trnI*-CAU and *rpl23* genes from *Spinacia oleracea*. The corresponding amino acids from the two ORFs of the *rpl23* gene are indicated. The *trnI*-CAU gene is underlined. Arrows marks the two protections observed in figure 3 corresponding to the transcription start sites. The -10 regions are boxed and the -35 regions are doubled underlined.

the putative ribosomal protein genes *rpl2*, *rps19* have been identified on this DNA fragment by comparison with the aminoacid sequence of the corresponding ribosomal proteins L2 and S19 from *E coli*

A *rpl23* gene of 94 codons has been identified upstream of the *rpl2* gene in the chloroplast genome of tobacco (1) and of *Marchantia* (2). Two ORFs coding respectively for 47 and 45 aminoacids are present in the upstream region of spinach *rpl 2* (Fig. 1 and 2). They overlap by 8 nucleotides residues. Each one has a high degree of homology with fragments of the *rpl23* coding sequence from tobacco. This split spinach gene has resulted from a 14 bp deletion by comparison with the tobacco chloroplast genome. We have verified by sequencing (not shown) the sequence data of Zurawski *et al.* (6) in this region. The results obtained fit perfectly those reported by these authors. Similarly the *rpl22* gene

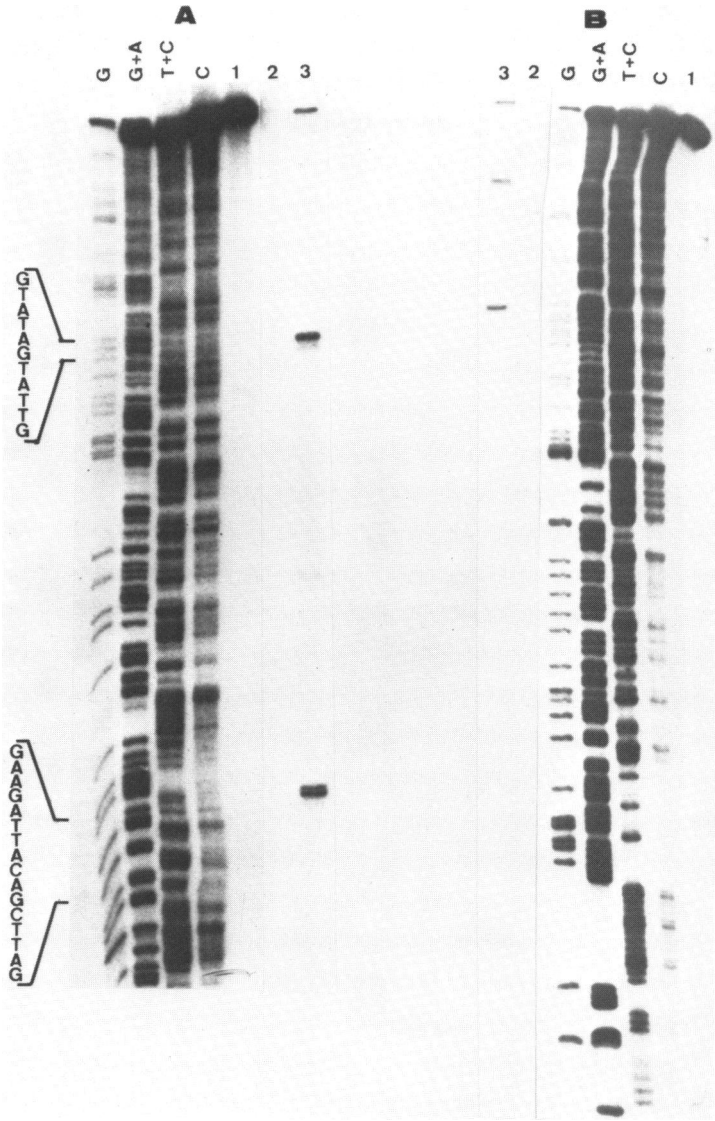


Figure 3: 5' S1 analysis of the upstream region of the *rpl2* gene of spinach chloroplast DNA using *in vivo* isolated RNAs. A 570 bp BamHI-XhoI DNA fragment was labeled at the 5' ends by using (γ^{32}P)-ATP and T4 polynucleotide kinase. **A**: long gel; **B**: short gel. For the two gels, the first four tracks correspond respectively to G, G + A, T + C and C sequence ladders of the coding strand. **Lane 1**: single stranded DNA probe labeled at the 5' end of the Xho I site. **Lane 2**: control using 30 µg of yeast tRNA and the DNA probe. **Lane 3**: S1 mapping of *in vivo* RNAs (30 µg) purified from spinach chloroplasts.

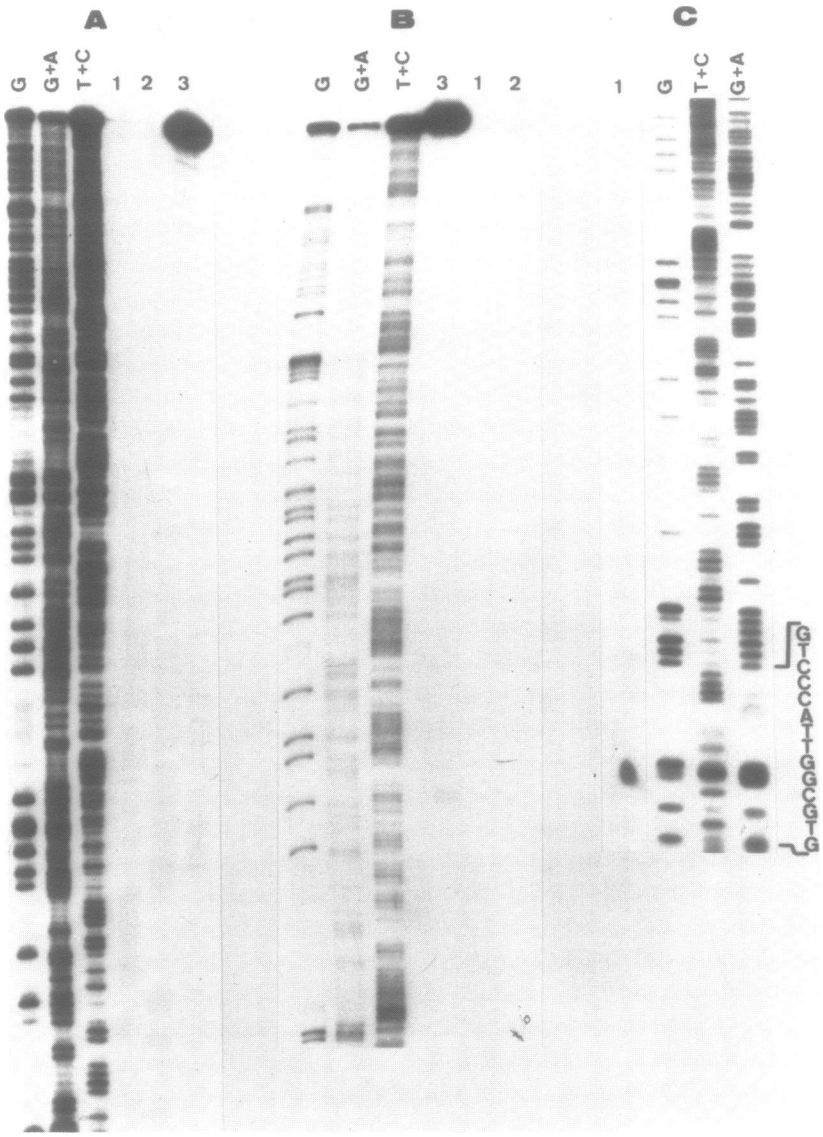


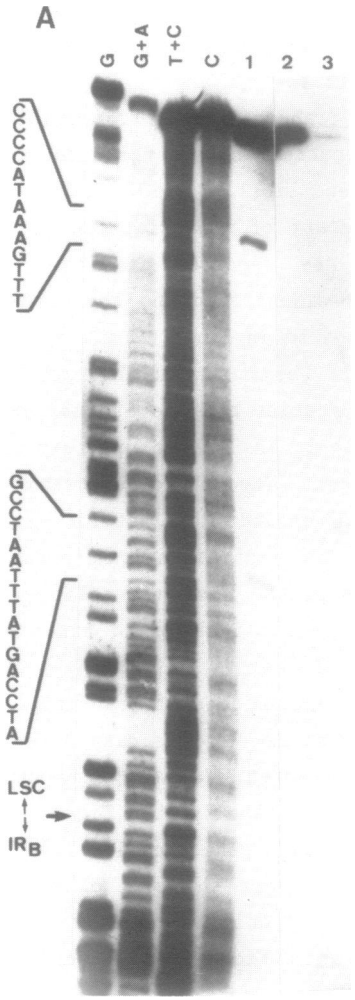
Figure 4: 3' S1 analysis of the dowstream region of the *trnI*-CAT gene of spinach chloroplast DNA using *in vivo* isolated RNAs. A 510 bp EcoRI-XhoI DNA fragment labeled at the 3' end of the EcoRI site by using (α^{32}P)ddATP and the Klenow fragment. **A**: long 6% gel; **B**: intermediate 6% gel, **C**: short 8% gel. For the two gels, the first three tracks correspond respectively to G, G + A, T + C sequence ladders. **Lane 1**: S1 mapping of *in vivo* RNAs (30 µg) purified from spinach chloroplasts. **Lane 2**: Control using 30 µg of yeast tRNA and the DNA probe. **Lane 3**: single stranded DNA probe labeled at the 3' end of the Eco RI site.

was not identified in spinach. The comparison with tobacco data shows that the part of the BamHI-11 DNA fragment which has been sequenced (6) contains the beginning of the *rpl22* gene (Fig. 1 and 5B).

S1 mapping of in vivo transcripts. A transcription study of the split *rpl23* gene and of its upstream region was undertaken by S1 mapping of *in vivo* transcripts. A 5' end labelled BamHI-XhoI probe (Fig. 1) was used in order to identify the start of transcription or processing sites within this region. Results are reported in Fig.3. The 5' ends of two transcripts were clearly identified and their positions on the sequence of nucleotides are indicated in Fig.2. One start (numbered 1) is located upstream of the first ORF (beginning of the *rpl23* gene). A Pribnow box is present at -10 from this start but the -35 box is not typical. A second start (numbered 2 in Fig.2) is located upstream of the second ORF and a Pribnow box is also present. The -35 box upstream of this second start is better than for the first one but does not fit either the consensus. Full length protected probe is significantly present (Fig.3) indicating a possible cotranscription of the *trnI*, the split *rpl23* and the *rpl2* genes. The main conclusion of this 5'S1 mapping experiment is that the split *rpl 23* gene is transcribed.

However, the two 5' extremities of transcripts described above could correspond to processing sites. In this case 3'-end of transcripts within this region should be detected. We test this hypothesis by S1 mapping using a 3'-end labelled EcoRI-XhoI probe (Fig.1) which includes the *trn* I gene and goes down to the *rpl2* gene. Results reveal the absence of 3'-end transcript between the *trnI* and the *rpl2* genes (Fig. 4). The processed product of the *trnI* gene is only observed (Fig. 4C, lane 1). The 5'-end of transcripts observed in Fig.3 could therefore correspond to true initiation sites utilized in addition to a larger co-transcript in this region.

An hairpin can be constructed immediately beyond the stop codon of the *rps19* gene, followed by an AT rich sequence. Zurawski *et al* (6) have suggested that this structure could define the end of a transcriptional unit. To test this hypothesis S1 analysis has been made by using 3'-end labelled SalI-EcoRV DNA fragment (Fig. 1). No protection was obtained which would correspond to this putative terminator but instead the end of a transcript in front of the *rpl22* start codon (Fig. 5) was observed. A very small amount of 3'-end transcripts which end within the *rps19* gene is also observed. No hairpin structures are possible immediatly upstream



B

rps 19 (3'-end)

G H K L G E F A P T L N F R G H A K N D N K S R R J

GGTCATAAATTAGGCGAATTTGCACCAACTCTCAATTTCCGGGGACATGCAAAAAACGATAATAAGTCTCGTCGTTAAGTTT
CCAGTATTTAATCCGCTTAAACGTGGTTGAGAGTTAAAGGCCCCTGTACGTTTTTTGCTATTATTCAGAGCAGCAATTCAAA

2

rpl22

SD

M G F F

AAGTTAATGTTAATTAACTTAATATAATAAATTATAGCTACTTATCATTTATTAATGGAGGTAAACTTTATGGGGTTTTTT
TTCAATTACAATTAATTGAATTATATTATTTAATATCGATGAATAGTAAATAATTACCTCCATTTGAAATACCCCAAAAAA

1

of these 3'-OH RNA extremities. These transcripts ending either within the *rps19* gene or near the start codon of the *rpl22* correspond most probably to processed RNAs. The large amount of full length protected probe indicates the presence of a long transcriptional unit extending from the *rpl2* gene to the *rpl22* gene.

rps19, rpl2 and *rpl23* gene products. Translational product of the *rps19* gene in a Zubay system were analysed by 2D-electrophoresis after the addition of spinach chloroplast 30S ribosomal proteins As shown on Fig. 6 one gene product co-migrates with the protein CS-S23. The second spot observed on the autoradiograph was also detected in the assay using the pUC8 vector alone, showing that this product did not derive from the *rps19* gene In order to demonstrate the presence of the *rps19* gene product in the 30S ribosomal subunit, the NH_2-terminal sequence of the HPLC purified CS-S23 was determined. Results reported in Fig. 7A show the identity of the sequence of 12 aminoacids with the sequence deduced from the *rps19* gene. It is also evidenced that the first methionine residue was cleaved after translation, *in vivo*.

The identification of the *E. coli* L2 protein with the CS-L4 protein was previously achieved (8) using immunological reaction. The NH_2-terminal sequence of CS-L4 was determined (Fig. 7B). Its identity with the terminal sequence deduced from the *rpl2* gene demonstrates that CS-L4 is the *rpl2* gene product, confirming the recent result of Kamp *et al.* (19).

An analysis of the putative products of the split *rpl23* gene was made by the use of a coupled transcription-translation assay in the same conditions as for the *rps* 19 gene. No specific gene products were detected after analysis by 2D-gel electrophoresis (not shown) We then searched for the presence of the two products from the split *rpl* 23 gene into the chloroplast 50S ribosomal subunits. As the putative products would have a Mr of 5500 and 5200, respectively, we have determined the

Figure 5. **A** : 3'-end S1 analysis of the downstream region of the *rpl2* gene of spinach chloroplast DNA using *in vivo* isolated RNAs A 550 bp SalI-EcoRI DNA fragment from pFT 1 was labeled at the 3' end of the SalI site by using (α^{32}P)dTTP and the Klenow fragment. The first four tracks correspond respectively to G, G + A, T + C and C sequence ladders of the coding strand. **Lane 1**: S1 mapping of *in vivo* RNAs (30 µg) purified from spinach chloroplasts. **Lane 2**: single stranded DNA probe labeled at the 3' end of the Sal I site. **Lane 3**: control using 30 µg of yeast tRNA and the DNA probe. **B** : Nucleotide sequence of the 3' part of *rps19* and of the 5' part of *rpl22* genes. Arrows 1 and 2 indicate the 2 protected regions schematized in A.

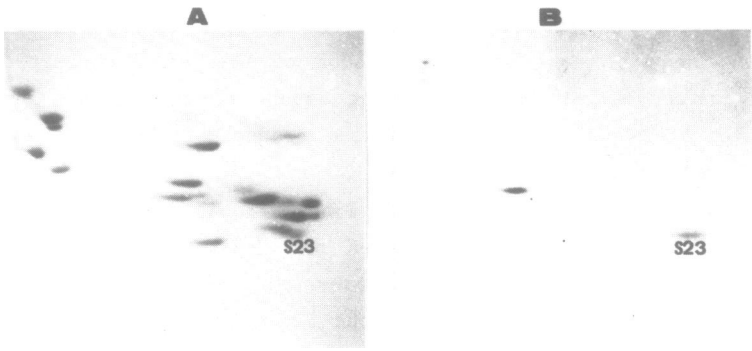


Figure 6. Analysis of the *in vitro* transcription-translation products of the pFT1 plasmid containing the *rps19* gene by two-dimensional gel electrophoresis. **A** : Coomassie staining of the 30S ribosomal proteins isolated from spinach chloroplast to which was added the products of the *in vitro* transcription-translation of pFT 1. **B** : Fluorogram of A after 5 days of exposure at -70 °C.

NH_2-terminal aminoacid sequence of the smallest chloroplast 50S ribosomal proteins which have a similar size. These are proteins CS-L31, CS-L33 and CS-L34 which have an apparent Mr ranging from 5000 to 8000 in SDS gels. None of these proteins correspond to any of the two ORF products at least for the 10 first aminoacids (not shown). We conclude that the split *rpl23* gene products are not into the ribosome. The possibility that the two ORFs could be translated into one polypeptide will be considered in the discussion.

DISCUSSION

In several plants, organisation of chloroplast ribosomal protein genes has been found in the same order as for one part of the *E. coli* S10 operon (1,2). Such an organisation also exists in the spinach chloroplast genome but the first structural gene *rpl23* is split into two ORFs which overlap by 8 nucleotides. The other structural genes *rpl2*, *rps19* and at least the beginning of the *rpl22* gene are present.

All these genes are expressed in spinach at a transcriptional level. Our results suggest that a transcriptional unit could start upstream the *trnI* gene and runs down into the *rpl22* gene. In addition, the presence of abundant transcripts has been observed with 5'-P ends located upstream of each reading frame present in the split *rpl23* gene. We suggest that the 5' end of each one of these two populations of transcripts correspond to

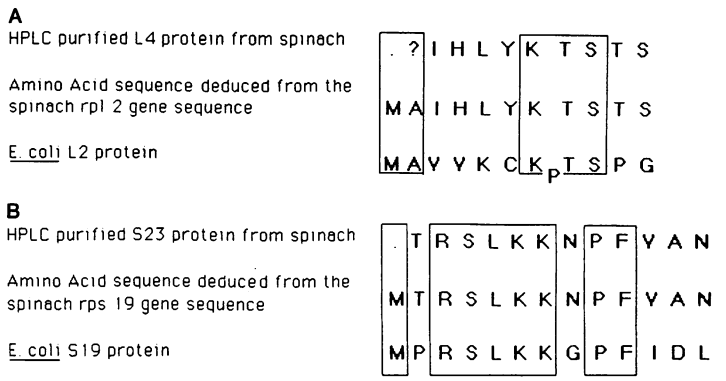


Figure 7: Comparison of amino-terminal sequences. **A** : Comparison of amino-terminal sequences of the ribosomal protein L4 purified from spinach chloroplast, the deduced protein from the chloroplast rpl2 gene (6) and of the ribosomal protein L2 from *E. coli*. **B** : Comparison of amino-terminal sequences of the ribosomal protein S23 purified from spinach chloroplast, the deduced protein from the chloroplast rps19 gene (6) and of the ribosomal protein S19 from *E. coli*.

an initiation site. However we cannot completely rule out the possibility that a *trnI-rpl23-rpl2-rps19* RNA precursor would be the only initiated transcript and therefore the transcripts 1 and 2 shown in Fig. 3 would arise from a rapid processing by more or less simultaneous cuts at multiple sites. Nevertheless, for our purpose, it is clearly demonstrated that the *rpl23*, *rpl2* and *rps19* genes are transcribed in spinach chloroplasts. The presence of these two populations of transcripts is in accordance with an independent expression of the two reading frames.

However, transcripts covering both these two ORFs exist, coming from the large transcription unit or starting at position 1 in Fig.2. Therefore, the hypothesis that a single protein could be synthesized from the split *rpl23* gene has to be considered. A shift from the first reading frame to the second one would be required during the translation in order to obtain a single product. Such a mechanism has been described in both eukaryotes and prokaryotes and has been reviewed recently (20). Characteristics (basicity and Mr) of the single putative protein deduced from the two-framed *rpl23* are very similar to the *E.coli* L23 protein. Only the CS-L30 protein migrates closely to the *E.coli* L23 in a 2D-gel (8) and is therefore the only two-framed protein candidate. We failed to find a correlation between the amino acid compositions of the CS-L30 and those from the putative *rpl23* gene product. The absence of any product

corresponding to the coding sequence of the *rpl23* gene, suggest that this gene has been transferred into the nucleus. This conclusion is in accordance with a recent observation based on nucleotide sequences comparison (21). Therefore the spinach chloroplast *rpl23* gene would be the first chloroplast pseudogene transcribed, identified so far.

In addition, results presented in this paper demonstrate that two putative ribosomal protein genes (*rps19* and *rpl2*) encode for identified ribosomal products (CS-S23 and CS-L4, respectively).

*To whom correspondence should be addressed

REFERENCES

1. Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Mastsubayachi, T., Zaita, N., Chunwongse, J., Obokata, J., Yamagushi-Shinozaki, K., Ohto, C., Torazawa, K., Meng, B.Y., Sugita, M., Deno, H., Kamogashira, T., Yamada, K., Kusuda, J., Takaiwa, F., Kato, A., Tohdoh, N., Shimada, H. and Sugiura, M. (1986) Embo J. 5, 2043-2049.
2. Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umesoro, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S., Inokushi, H. and Ozeki, H. (1986) Nature 322, 572-574.
3. Eneas-Filho, J., Hartley, M.R. and Mache R. (1981) Mol. Gen. Genet. 184, 484-488.
4. Dorne, A.M., Lescure, A.M. and Mache, R. (1984a) Plant Mol. Biol. 3, 83-90.
5. Tanaka, M., Wakasugi, T., Sugita, M., Shinozaki, K. and Sugiura, M. (1986) Proc. Natl. Acad. Sci. USA 83, 6030-6034.
6. Zurawski, G., Bottomley, W. and Whitfeld, P.R. (1984) Nucleic Acids Res. 12, 6547-6558.
7. Mache, R., Dorne, A.M. and Marti-Batlle, R. (1980) Mol. Gen. Genet. 177, 333-338.
8. Dorne, A.M., Eneas-Filho, J., Heizmann, P. and Mache, R. (1984b) Mol. Gen. Genet. 193, 129-134.
9. Madjar, J.J., Michel, S., Cozzone, A.J. and Reboud, J.P. (1979) Anal. Biochem. 12, 174-182.
10. Mache, R., Audren, H., Bisanz-Seyer, C., Dorne, A.M., Massenet, O., Rozier, C., Thomas, F. (1987) In Biggins, J. (ed), Progress in Photosynthesis Research, Martinus Nijhoff Publishers, Dordrecht, Vol. IV, pp. 547-551.
11. Edman, P. and Begg, G. (1967) Eur. J. Biochem. 1, 80-91.
12. Aguettaz, P., Seyer, P., Pesey, H. and Lescure, A.M. (1987) Plant Mol. Biol. 8, 169-177.
13. Maniatis, T., Fritsch, E.P. and Sambrook, J. (1982) *In:* Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Eds.
14. Birnboim, H.C. and Doly, J. (1979) Nucleic Acids Res. 7, 1513-1523.
15. Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci USA 74, 560-564.
16. Lescure, A.M., Bisanz-Seyer, C., Pesey, H. and Mache, R. (1985) Nucleic Acids Res. 13, 8787-8796.
17. Douce, R. and Joyard, J. (1982) *In:* Edelmal et al. (eds), Methods in chloroplat molecular biology, Elsevier Biomedical Press, pp. 239-256.
18. Link, G. and Langridge, U. (1984) Nucleic Acids Res. 12, 945-958.
19. Kamp, R.M., Bachally, R.S., Knoblauch, K.V., Subramanian, A.R. (1987) Biochemistry 26, 5866-5870.
20. Craigen, WW.J. and Caskey, C.T. (1987) Cell 50, 1-2.
21. Zurawski, G. and Clegg, M.T. (1987) Ann. Rev. Plant Physiol. 38, 391-418