Curr Genet (1990) 18:553–556

Current Genetics

© Springer-Verlag 1990

Hypothesis for the evolutionary origin of the chloroplast ribosomal protein L21 of spinach

W. Martin *, T. Lagrange, Y. F. Li, C. Bisanz-Seyer, and R. Mache

Laboratoire de Biologie Moléculaire Végetale, Université Joseph Fourier, BP 53X, F-38041 Grenoble Cédex, France

Received October 2, 1990

Summary. A full size cDNA clone encoding the chloroplast ribosomal protein L21 from spinach is presented. The identity of the clone and the location of the transit peptide processing site were determined by comparison with the N-terminal amino acid sequence of the spinach chloroplast protein CS-L7 previously identified. L21 r-protein sequences from spinach, *Marchantia polymorpha* and *Escherichia coli* are compared. Quite surprisingly, the data do not suggest that the *rpl*21 nuclear gene from spinach was derived through intracellular gene transfer from the chloroplast genome. The possibility of a mitochondrial origin for *rpl*21 gene of spinach is discussed.

Key words: Molecular evolution – Chloroplast ribosomal protein – Gene transfer - *Spinacia oleracea*

Introduction

Chloroplast are generally believed to be of endosymbiotic origin, yet the genome of higher plant chloroplasts encodes but a minority of the proteins required by the organelle (Sugiura 1989). Many of the genes encoding proteins integral to plastid metabolism which were originally encoded in the chloroplast genome are thought to have been transferred to the nucleus during the course of plant evolution (Weeden 1981). Indeed, several clear cases of intracellular gene transfer have been reported (Martin and Cerff 1986; Shih et al. 1986; Brinkmann et al. 1987; Baldauf et al. 1990; Liaud et al. 1990).

To further investigate the nature of intracellular gene transfer events during chloroplast-nuclear DNA coevolution, we have isolated clones encoding the spinach L21 ribosomal protein (r-protein). The homologue of *rpl*21 gene from the liverwort *Marchantia polymorpha* is still encoded within cpDNA (Ohyama et al. 1986), yet is absent in higher plant cpDNAs (Sugiura 1989), suggesting that the functional gene for the L21 r-protein has been lost from the chloroplast genome subsequent to the emergence of land plants. Comparison of the spinach L21 r-protein and the *Marchantia* and *Escherichia coli* homologues suggests that the *rpl*21 gene of spinach was derived through transfer from the mitochondrial rather than the chloroplast genome.

Materials and methods

RNA preparation, cDNA cloning and sequencing. PolyA$^+$ mRNA was isolated from young spinach plants. An initial cDNA library in λgt11 was constructed as described (Zhou and Mache 1989). The library was screened with antibodies raised against spinach 50S r-proteins. A partial cDNA was identified through sequencing and comparison to the *E. coli* L21 r-protein sequence (Heiland and Wittmann-Liebold 1979).

A second cDNA library was constructed from 5 μg of polyA$^+$ mRNA with a kit from Pharmacia (Uppsala) according to the supplier's protocol but with following modifications. Subsequent to ligation of EcoRI-NotI adaptors, the cDNA was size fractionated on 1.5% agarose gel. The 0.5-3 kb DNA was electroeluted, purified on DE-52 cellulose (Whatman), phosphorylated with polynucleotide kinase (Pharmacia), purified on DE-52, and ligated into EcoRI cut λnm1149 (Murray 1983). Packaging and screening was performed as described by Schwarz-Sommer et al. (1985). 30 ng of cDNA were ligated and plated to yield 450 000 recombinants. This library was screened with the original partial L21 cDNA fragment. Approximately 300 positives were obtained. 20 of these were chosen at random and shown to contain the roughly 1.1 kb *rpl*21 inserts expected from Northern blots (data not shown). Several of these were subcloned into the Bluescript plasmids and sequenced by the dideoxy method (Sanger et al. 1977). Sequence analysis was performed with the WISGEN package (Devereux et al. 1984) with support from the computer center of the GBF Braunschweig and with the support of CITI2 in Paris.

Southern blotting. High-molecular weight *Spinacia oleracea* genomic DNA was prepared from young seedlings according to Maniatis (1982). Samples of 15 μg each were digested by restriction enzymes, subjected to electrophoresis in 1% agarose and blotted on a nylon membrane (Amersham Hybond N+). The filter was probed with the cDNA-fragment EcoRI-SphI (positions 240-1046 in Fig. 1) la-

* *Current address:* Institut für Genetik, Biozentrum, Technische Universität Braunschweig, Postfach 3329, W-3300 Braunschweig, Federal Republic of Germany

Offprint requests to: R. Mache

belled with [α-^{32}P]dATP using the nick-translation kit from Boehringer (Mannheim). Prehybridation was performed in a solution of 6 × SSPE (0.1% SDS, 0.02% PVP, 0.02% Ficoll, 50 μg/ml Salmon sperm DNA) and hybridization in a solution of 3 × SSPE at 65 °C for 18h. The filter was washed once with 3 × SSPE, 0.1% SDS for 10 min at 65 °C and once with 0.3 × SSPE, 0.1% SDS for 10 min at 65 °C.

Protein N-terminal sequencing. 50S r-proteins were isolated by HPLC and the N-terminal sequencing was performed at the Service Central d'Analyse du CNRS (Solaize, France).

Results and discussion

A full size cDNA clone coding for a spinach protein homologous to the *E. coli* L21 r-protein, was isolated from a cDNA library in λnm1149. The sequence is shown in Fig. 1. Underlined are the 12 amino acids which correspond precisely to the N-terminal sequence of the native spinach chloroplast r-protein termed CS-L7 in our previous nomenclature (Mache et al. 1980). The sequence of the cDNA is therefore unambiguously identified as that of the chloroplast r-protein L21. The transit peptide sequence (Fig. 1.) shows the characteristics of chloroplast transit peptides (Von Heijne et al. 1989). The apparent molecular weight of the in vivo protein (Mr = 28 000) is slightly higher than that derived from the cDNA (MW = 22 800), while the net charge (+1) is in good agreement with the position of the protein in the 2-D gel (Mache et al. 1980; Dorne et al. 1984).

The number of *rpl*21 genes was determined using southern blots of genomic DNA (Fig. 3). One band was detected when using EcoRV or BamHI. An HindIII site is present in one intron of the *rpl*21 gene (data not shown) thus explaining the presence of two HindIII fragments in the southern blot. These results show the presence one gene per haploid genome.

The amino acid sequence has been aligned with the sequences for L21 r-protein from *M. polymorpha* (Kochi et al. 1988) and *E. coli* (Heiland and Wittmann-Liebold 1979) as shown in Fig. 2. The spinach L21 r-protein contains roughly 100 residues common to all three sequences, a region which we term the central core. The protein also contains a 65 amino acid N-terminal and a 25 amino acid C-terminal extension found in neither the *E. coli* nor the *Marchantia* homologues. The C-terminal extension carries a net charge of −10. A series of Glu-Ala residues is present at the end of this extension. This series has been verified with several cDNA clone for the L-21 r-protein. Polypeptide extensions to a central homologous core have also been found for several nuclear encoded chloroplast r-proteins (Mache 1990) and for the chloroplast encoded L22 r-protein (Zhou et al. 1989). The N-terminal extension of the spinach L21 is the longest yet observed for a chloroplast r-protein. The function and origin of the extensions are not known. They may prove to be valuable markers in reconstruction of the phylogeny of *rpl*21 genes as more sequences become known. L21 r-protein sequences are currently available only from spinach, *M. polymorpha and E. coli.*

Since *rpl*21 is encoded in the cpDNA in *M. polymorpha* but is absent from the cpDNA of higher plants, our working hypothesis at the onset of these experiments was that the gene for *rpl*21 had been transferred from chloroplast to the nucleus after the emergence of land plants. The sequence comparison of L21 r-proteins from spinach, *M. polymorpha* and *E. coli* (Table 1) are very difficult to reconcile with this hypothesis. Within the L21 central core, amino acid identities for the three pairwise comparisons are *E. coli-Marchantia* (chloroplast) 27.4%, *E. coli*-spinach (nuclear) 31.3%, and *Marchantia* (chloroplast)-spinach (nuclear) 30.4%. This is a surprising result, since we could expect the chloroplast protein sequences to show a greater homology to each other than

```
                         M  A  S  A  T  L  A  F  S  C  S  S  L  C  A  T  L  K  L     19
TTTAGAGCTCAACTAAAAATGGCGTCTGCAACTTTAGCATTCTCTTGTTCCTCACTCTGTGCAACACTGAAACTT      75

 P  Q  N  L  N  P  L  L  L  N  V  P  P  L  S  K  P  F  S  G  V  V  S  P  P     44
CCTCAAAATCTAAACCCACTTCTTTTGAATGTTCCACCTCTCTCTAAACCCTTTTCCGGCGTAGTTTCACCGCCC     150
                                    ↓
 S  L  S  R  L  S  L  L  P  V  A  A  K  R  R  R  F  Q  E  I  P  E  E  L  K     69
TCTCTATCTCGGCTGTCTCTTCTCCCCGTCGCCGCCAAACGACGCCGATTTCAGGAAATTCCAGAAGAGTTGAAG     225

 A  E  F  E  E  F  Q  R  P  P  N  Q  K  P  Q  L  S  D  V  L  P  D  D  F  Q     94
GCGGAATTTGAAGAATTCCAAAGGCCCCCAACCAAAAGCCCCAATTATCTGATGTTCTGCCAGACGATTTTCAG     300

 A  P  E  P  G  T  P  E  Y  N  D  I  I  N  Q  F  L  P  K  K  G  P  P  P  P    119
GCTCCTGAGCCTGGTACCCCTGAATACAACGATATTATCAACCAATTCCTTCCCAAAAAAGGCCCTCCTCCTCCC     375

 R  E  E  I  F  A  V  V  V  I  G  S  R  Q  Y  I  V  I  P  G  R  W  I  Y  T    144
CGTGAAGAGATTTTTGCTGTTGTTGTGATTGGTTCTAGACAGTACATTGTCATTCCTGGAAGATGGATTTACACC     450

 Q  R  L  K  G  A  T  V  N  D  K  I  V  L  N  K  V  L  L  V  G  T  K  A  S    169
CAGAGGCTCAAAGGTGCTACTGTCAATGATAAGATTGTTCTGAACAAAGTTTTGCTAGTGGGAACTAAAGCTAGC     525

 T  Y  I  G  T  P  I  V  T  N  A  A  V  H  A  V  V  E  E  Q  L  L  D  D  K    194
ACCTACATAGGAACACCTATCGTCACAAATGCTGCAGTACATGCTGTTGTAGAAGAACAGTTGCTGGATGACAAG     600

 V  I  V  F  K  Y  K  K  K  K  N  Y  R  R  N  I  G  H  R  Q  P  I  T  R  I    219
GTTATCGTCTTCAAGTATAAGAAGAAGAAGAACTATAGACGAAATATTGGTGATAGACAGCCCATCACCCGTATA     675

 K  I  T  G  I  T  G  Y  E  D  Y  P  A  S  T  L  E  A  E  V  E  A  K  E  E    244
AAGATAACAGGAATCACTGGTTATGAAGATTACCCAGCTTCTACACTAGAAGCAGAGGTAGAGGCAAAGGAAGAG     750

 A  E  A  E  A  E  A  E  A  V  P  V  *                                         256
GCAGAGGCAGAGGCAGAGGCAGAGGCAGTTCCAGTTTAAGTTGAACAACAAATCTGCAACTTTATACCCTTTTTT     825

TTTTTCCGGTTCTGTCAGTAGGGTTTTTGATTCATACAGCTTCATCTTTTTGCCAAAGTTAGGCTGTTTGTAAGA     900
TGTAATTTGATAGGATTGATATTCAAAGAATTATGATTGACCAGAGCTTGGTGTATGTAGTATTTCTCCAACAGA     975
AGGAGCAATGTTGTAAATTAATTATCTGAAGATGCAGCTTCATGTTAATGGACAGATATAGGGTTACTGCATGCC    1050
ATCAACTTAAAAAGGACTTCAA                                                        1072
```

Fig. 1. Nucleotide and deduced amino acid sequence (single-letter code) of a cDNA coding for a chloroplast r-protein homologous to the *E. coli* L21 r-protein. The *arrow* indicates the processing site of the chloroplast ribosomal precursor deduced from the 12 amino acid terminal residues of the mature protein that are indicated by a *solid line*. The 3′ terminal non-coding sequence without polyA stretch is confirmed by the sequence of a genomic clone (not shown)

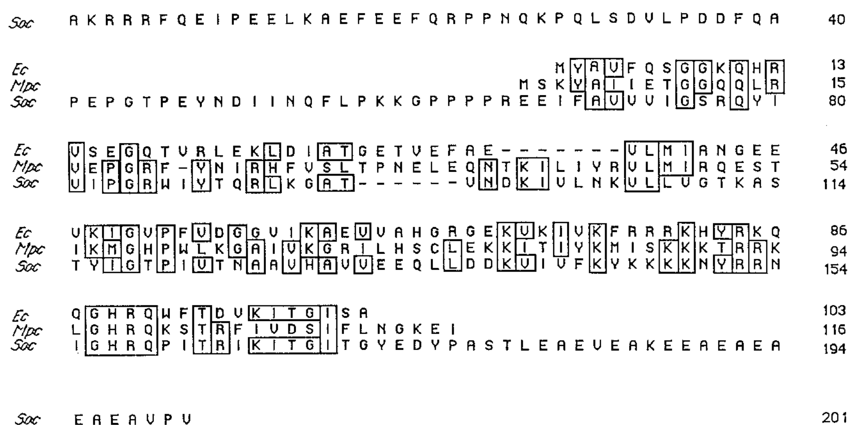

Fig. 2. Identical aminoacids in the L21 r-proteins from *E. coli* (*Ec*), *Marchantia polymorpha* chloroplasts (*Mpc*) and *Spinacia oleracea* chloroplasts (*Soc*). The alignments which favours common residues between the 3 sequences were chosen

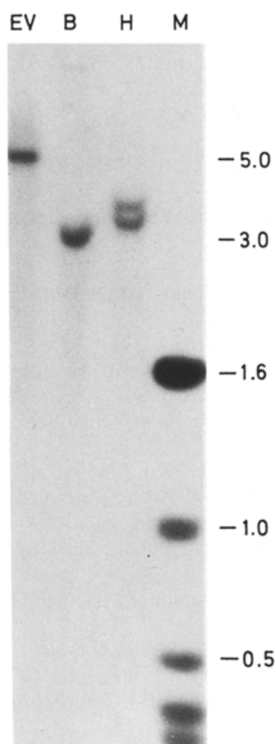

Fig. 3. Southern blot analysis of spinach genomic DNA cut with EcoRV (*EV*), BamHI (*B*) or HindIII (*H*). The sizes of the molecular weight markers (*M*; 1 kb ladder from BRL) are given in kilobase pairs. High molecular weight DNA was isolated from young spinach plants. Samples of 15 μg each were digested to completion with restriction enzymes. Filter was probed with ^{32}P-labelled cDNA fragments (see Material and methods)

Table 1. Comparison of homology data in conserved regions of different 70S r-proteins (S12, L22, L21) between *E. coli*, chloroplasts and plant mitochondria

Pair-wise comparison [a]	Percent amino acid identity	Amino acid substitutions per site (corr.) [b]
a) S12		
eco-mar(cp)	68.5	0.378
eco-tob(cp)	71.0	0.342
tob-mar(cp)	90.3	0.102
b) L22		
eco-mar(cp)	41.6	0.877
eco-spi(cp)	37.6	0.978
spi(cp)-mar(cp)	63.3	0.457
c) S12		
eco-mar(cp)	68.5	0.378
eco-mz(mt)	55.6	0.587
mar(cp)-mz(mt)	58.9	0.529
d) L21		
eco-mar(cp)	27.4	1.295
eco-spi(nu)	31.3	1.161
mar(cp)-spi(nu)	30.4	1.191

[a] Chloroplasts (cp) are represented by *Marchantia polymorpha* (*mar*), tobacco (*tob*), and spinach (*spi*). Plant mitochondria (*mt*) are represent by maize (*mz*). Eubacteria are represented by *Eschericha coli* (*eco*). The data for S12 were calculated from the alignment in Gualberto et al. (1988). The data for L22 were calculated from Zhou et al. (1989) on the basis of a 101 amino acid "central core" alignment

[b] Amino acid substitutions per site were corrected for multiple substitutions at identical sites assuming a Poisson distribution (Dickerson 1971). These values reflect the divergence which has taken place between sequences

to an enterobacterial homologue as in the case of the chloroplast encoded genes for S12 of L22 (Table 1, a and b), were the spinach *rpl*21 gene of chloroplast origin. This observation requires the consideration of alternatives which may account for the experimental data.

It is conceivable that (i) *rpl*21 may have arisen through duplication of a nuclear gene for a corresponding 80S protein. This is unlikely since 70S and 80S ribosomes are distinct and distantly related structures. Furthermore it is not known whether a true homologue for L21 exists in 80S ribosomes which could assume L21 function in 70S ribosomes. We consider this possibility improbable, but it cannot at present be completely excluded. A tempting alternative is that (ii) the spinach L21 protein is derived from a gene which originally belonged to the mitochondrial DNA, i.e. loss of *rpl*21 from the cpDNA was compensated through the product of a nuclear gene of mitochondrial origin. The substitution of a 70S r-protein by a 70S rather than 80S counterpart can be envisaged. Since the lineages leading to chloroplasts, mitochondria and gram negative bacteria are thought to have diverged roughly concomitantly (Woese 1987), we would expect in the case of mitochondrial origin for the spinach *rpl*21 gene an equidistant relationship for the three known sequences. Precisely this result is found (Table 1, d) which may be taken as evidence in favor of

alternative (ii). For comparison, we have included the data for S12 (Table 1, c), an r-protein for which sequences from mitochondria and chloroplasts are known. The pattern of homology found in Table 1 (c) is clearly much more similar to the data for L21 than are the patterns found in Table 1 (a) and (b). Judging by comparison to the analyses for 16S rRNA sequences (Woese 1987; Gray 1989), the slightly lower divergence observed between S12 of *E. coli* and *Marchantia* may be due to stochastic variation in rate; indeed, many comparisons involving *Marchantia* cpDNA-encoded r-proteins show lower values of divergence than their counterparts from other species (Christopher and Hallick 1989). However, sequences for plant mitochondrial L21, necessary for verification of a putative mitochondrial origin, are not at present available for comparison.

A further alternative is that (iii) the spinach *rpl*21 gene is indeed recently derived from the chloroplast genome. In this case, we must postulate a drastically accelerated evolutionary rate for *rpl*21 subsequent to transfer to the nucleus in order to reconcile the data given in Table 1. If the *rpl*21 gene was transferred from the chloroplast and underwent a dramatic increase in substitution rate, we would then expect divergence between the spinach and *E. coli* sequences to markedly exceed that observed between *Marchantia* and *E. coli*. Such a difference in the level of divergence was not observed (Table 1, d). This renders the chloroplast to nucleus gene transfer origin for *rpl*21 rather unlikely. Thus, the available data do not permit sound discrimination between the three alternatives presented above concerning the evolutionary origin of the spinach *rpl*21 gene. They are, however, compatible with alternative (ii). Additional reference sequences from other sources are needed to establish a firm gene genealogy. The study of mitochondrial, cyanobacterial and other higher plant genes for L21 should yield insights to this question.

The free-living antecedents of chloroplasts relinquished most of their genetic information during the course of endosymbiosis. Fixation of deletion events from cpDNA involving genes for products essential to chloroplast function could only have occurred subsequent to or concomitant with replacement of function from alternative sources (Gray 1989). The analysis of cDNA and genomic clones for *rpl*21 from spinach presented here suggests that the alternatives for origin of nuclear encoded chloroplasts proteins put forth by Weeden (1981) should be expanded to include the mitochondrial genome.

Note added in proof

When this paper was submitted, results concerning a spinach cDNA encoding the chloroplast L21 protein have been published by Smooker et al. (1990) J Biol Chem 265:16699–16703 although with a different interpretation of data.

The sequence data will appear in the EMBL library under the accession number X56691.

Acknowledgement. We are indebted to A. M. Dorne for help in the preparation of the protein CS-L7.

References

Baldauf SL, Palmer JD (1990) Nature 344:262–265
Brinkmann H, Martinez P, Quigley F, Martin W, Cerff R (1987) J Mol Evol 26:320–328
Christopher D, Hallick R (1989) Nucleic Acids Res 17:7591–7608
Devereux J, Haeberli P, Smithies O (1984) Nucleic Acids Res 12:387–395
Dickerson R (1971) J Mol Evol 1:26–45
Dorne AM, Lescure AM, Mache R (1984) Plant Mol Biol 3:83–90
Gualberto J, Wintz H, Weil J-H, Grienenberger J-M (1989) Mol Gen Genet 215:118–127
Gray M (1989) Trends Genet 5:294–299
Heiland I, Wittmann-Liebold B (1979) Biochemistry 18:4605–4612
Kochi T, Shirai H, Fukuzawa H, Sano T, Komano T, Umesono K, Inokuchi H, Ozeki H, Ohyama K (1988) J Mol Biol 203:353–372
Liaud M-F, Zhang D-X, Cerff R (1990) Proc Natl Acad Sci USA (in press)
Mache R, Dorne AM, Marti-Batlle R (1980) Mol Gen Genet 77:333–338
Mache R (1990) Plant Sci (in press)
Maniatis T, Fritsch EF, Sambrook J (1982) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
Martin W, Cerff R (1986) Eur J Biochem 159:323–331
Murray N (1983) Phage lambda and molecular cloning. In: Hendrix R, Roberts JW, Stahl FW and Weisberg RA (eds)
Lambda II. Cold Spring Harbor Laboratory Press, New York, pp 395–432
Ohyama K, Fukuzawa H, Kohchi T, Shirai H, Sano T, Sano S, Umesono K, Shiki Y, Takeuchi M, Chang Z, Aota S, Inokuchi H, Ozeki H (1986) Nature 322:572–574
Sanger F, Nickler S, Coulson A (1977) Proc Natl Acad Sci USA 74:5463–5467
Schwarz-Sommer Z, Gierl A, Cuypers H, Peterson P, Saedler H (1985) EMBO J 3:1021–1028
Shih MC, Lazar G, Goodman H (1986) Cell 47:73–80
Sugiura M (1989) Ann Rev Cell Biol 5:51–70
Von Heijne G, Stepphuhn J, Herrmann RG (1989) Eur J Biochem 180:535–545
Weeden NF (1981) J Mol Evol 17:133–139
Woese C (1987) Microbiol Rev 51:221–271
Zhou DX, Mache R (1989) Mol Gen Genet 219:204–208
Zhou DX, Quigley F, Massenet O, Mache R (1989) Mol Genet 216:439–445

Communicated by K. Wolf