A Ribosomal Protein Is Encoded in the Chloroplast DNA in a Lower Plant but in the Nucleus in Angiosperms

ISOLATION OF THE SPINACH L21 PROTEIN AND cDNA CLONE WITH TRANSIT AND AN UNUSUAL REPEAT SEQUENCE

(Received for publication, February 27, 1990)

Peter M. Smucker, Volker Kruft, and Alap R. Subramanian

From the Max-Planck-Institut für Molekulare Genetik, Abteilung Wittmann, Ihnestrasse 73, D-1000 Berlin 33, Federal Republic of Germany.

The distribution of chloroplast ribosomal protein genes between the organelle DNA and the nuclear DNA is highly conserved in land plants, but a notable exception is rp121. This gene has been found in the completely sequenced chloroplast genome of a lower plant but not in that of two higher plants. We describe the purification and characterization of the spinach chloroplast ribosomal protein L21 and the isolation and nucleotide sequence of a cDNA clone that encodes its cytoplasmic precursor. The mature protein, identified by N-terminal sequencing, has 201 residues (M, 22,768) and is thus substantially larger than either its Escherichia coli (103 residues) or the lower plant homologue (116 residues). The extra length is in peptide extensions at both amino and carboxyl termini. The COOH-terminal extension is unusual in that it comprises seven Ala-Glu repeats, a feature not found in any other ribosomal proteins described so far. The cDNA clone also encodes a 55-residue long transit peptide (with a high proportion of the polar residues, threonine and serine), to target the L21 protein into chloroplasts. The identification of rp121 as a nuclear gene in a higher plant (spinach) and chloroplast gene in a lower plant (liverwort) suggests an organelle-to-nucleus gene relocation during the evolution of the former.

Chloroplast ribosomes of land plants contain between 56 and 65 r-proteins (1-5). Genes encoding these proteins are distributed in two cellular compartments. Cloning and characterization of the complete set of the r-protein genes located in the chloroplast DNA has been described for four plant species, tobacco (6), a liverwort (7), rice (8), and maize (37). The analysis of these genes and their transcripts has shown that ribosome-Proteins extracted from spinach chloroplast ribosomes were subjected to purification steps which are described in Results. An antiserum against pool 43 immunoreacted with the 21-kDa protein and the localization/significance of the gene that encodes it.

RESULTS

Characterization of a 21-kDa Protein of Spinach Chloroplast Ribosome—Proteins extracted from spinach chloroplast ribosomes were subjected to purification steps which are described in Results. One of the pools so obtained (i.e. pool 43) contained mainly two polypeptides of molecular mass 21 and 26 kDa (Fig. 1, lane 1). An antiserum raised against pool 43 strongly reacted with both the polypeptides (Fig. 1, lane 2). This paper is concerned with the identification of the 21-kDa protein and the localization/significance of the gene that encodes it.

In a Western blot of the total proteins from sucrose-gradient purified 30 and 50 S subunits of spinach chloroplast ribosomes, the antisem against pool 43 immunoreacted with...
a 21-kDa band in the 50 S subunit. In contrast, an immunoreaction at the 21-kDa position was not detected with 30 S subunit proteins (Fig. 1, lanes 3 and 4). This result demonstrated that the 21-kDa protein is specifically associated with the 50 S ribosomal subunits.

Immunoscreening of a cDNA Library—Having established that the antiserum reacts strongly with chloroplast r-proteins, we used it to screen a spinach cDNA expression library in Agt11 (13). Of a total of 150,000 plaques screened, 10 gave apparent positive signals and were purified. The restriction maps of their cDNA inserts showed two classes of inserts. One class comprised of an insert having an internal EcoRI restriction site and (as will be apparent) produced the fusion protein which reacted with the 21-kDa protein antibodies. The other class contained cDNA inserts which encoded the 26-kDa protein (38).

Sequencing of the cDNA Insert and Identification of a Reading Frame—The cDNA insert corresponding to the 50 S subunit protein was subcloned into pT7/T3-19U, and the nucleotide sequence of the entire insert was determined (Fig. 2) by subcloning restriction fragments or by using oligonucleotides designed from previous rounds of sequencing. The cDNA insert was found to be 1069 nucleotides in length (Fig. 3, Miniprint) and included a poly(A) tail of nine bases.

The nucleotide sequence was searched for possible protein coding regions, and a long open reading frame predicting a protein of 256 amino acid residues and $M_r$ 28,412 was found. The initiating methionine codon of this reading frame is in the context AAAAATGGC, with an A at position −3 and a G at +4, as is found in the majority of plant mRNA initiation sequences (30).

NH$_2$-terminal Sequencing of the 21-kDa Protein Identifies the Mature Protein and the Precursor Cleavage Site—The two polypeptides of pool 43 were purified to homogeneity by reverse-phase HPLC as shown in Fig. 4, and the NH$_2$-terminal sequence of the 21-kDa protein was determined. Twenty cycles of the sequenator run gave the following result:

$$1 \text{A-K-R-R-R-F-Q-E-I-P-E-E-L-K-A-E-F-E-F}$$

This sequence matches exactly with that predicted from the nucleotide sequence from residue 56 to 75. Hence the cDNA encodes a precursor form, and the NH$_2$-terminal 55 residues would comprise the targeting sequence that directs the precursor to chloroplasts. The mature protein would be 201 amino acids in length ($M_r$ 22,766). The experimentally determined amino acid composition of the purified protein agreed closely with that predicted from the nucleotide sequence (Table I, Miniprint). The cleavage site for the targeting sequence is PVAJAKR.

Identity of the 21-kDa Chloroplast R-protein—Searching the NBRF and RIBO data bases with the mature protein sequence gave two matches: the *E. coli* r-protein L21 (31), with 32% identical residues, and liverwort (*Marchantia polymorpha*) L21 (7), with 28% identical residues. This was an unusual match, as the *E. coli* and liverwort L21 proteins are 103 and 116 amino acids residues in length, respectively, while the protein predicted here is over 200 amino acid residues long. This difference in length between a chloroplast and *E. coli* r-protein is unusually large (for review, see Ref. 37).

We used the ALIGN program (32) to further investigate
this apparent homology. By using this program again two highly significant matches were found. The first, with *E. coli* L21 (align score > 17 S.D.), the second with liverwort L21 (align score > 14 S.D.). Both scores are well above the range at which a match is highly significant. The alignment of the three sequences is shown in Fig. 5.

The spinach protein is by far the largest of the three known L21 proteins, being 98 and 85 amino acids, respectively, longer than the *E. coli* and the liverwort homologues. The larger size of the mature spinach L21 protein is the result of peptide extensions at both the NH₂ terminus (6 residues) and the COOH terminus (25 residues). The liverwort protein has a pentapeptide in the central region which is absent in both the spinach and *E. coli* proteins (Fig. 5).

The net charge of the spinach L21 protein is +1, compared to +10 and +20, respectively, for the *E. coli* and liverwort proteins. Thus, unlike its *E. coli* and liverwort homologues, it is not a highly basic protein. This is reflected by the migration position of L21 in two-dimensional gels of spinach chloroplast 50 S subunit proteins, shown in Fig. 6, (determined using purified L21 protein) where it migrates to the region of acidic proteins.

**DISCUSSION**

We have described the protein-chemical characterization of a chloroplast r-protein from a higher plant (spinach) and the nucleotide sequence of a cDNA clone that encodes its cytoplasmic precursor. From the degree of amino acid sequence identity, the chloroplast protein is inferred to be the homologue of the *E. coli* ribosomal protein L21.

The nuclear-coded protein described here is much larger than its two known homologues, the L21 of *E. coli* (31) and of the lower plant *M. polymorpha* (7). However, over the region where the proteins show sequence identity the charge distribution and other similar features are well conserved. In the COOH-terminal region there is a conserved GHRQ motif which is preceded by a highly positively charged region (e.g. K-Y-K-K-K-Y-R-R in spinach). This observation implies a common function for the three proteins, although the angiosperm protein has additional sequence features of significance.

The liverwort L21 protein is 13 amino acids longer than the *E. coli* homologue, an increase not uncommon among chloroplast-encoded r-proteins (37). The difference of 98 amino acid residues between spinach L21 and *E. coli* L21 is over twice the next highest known, *i.e.* spinach chloroplast L13 and *E. coli* L13 (14). Both L13 and L21 are encoded in the nucleus, and it appears as a general rule that chloroplast r-proteins encoded in the nucleus are larger than their eubacterial homologues, predominantly due to an extension at the NH₂ terminus.

For chloroplast proteins encoded in the nucleus there is the added requirement for a transit peptide, which must also contain a recognition site for the endopeptidase that removes this peptide during or after the import of the precursor into the chloroplast. The fusion of such a transit peptide coding region to the presumably eubacterial coding region of the mature r-protein could result in the peptide extension at the NH₂ terminus.

Compared to the liverwort homologue, spinach L21 carries also a peptide extension of 25 residues at the COOH terminus. This has the remarkable feature of containing seven repeats of the dipeptide Glu-Ala, in a peptide stretch E-A-E-V-E-A-K-E-E-A-E-A-E-A-E-A (Fig. 5). Much of the negative charge in the spinach protein lies in this unusual COOH-terminal extension. Such a feature is not found in other known ribosomal proteins, including *E. coli* and liverwort L21, and *E. coli* S6 and L12 which are rich in Ala/Glu content. Probably this added structural feature signifies an added functional role for the chloroplast L21 protein of angiosperms. Since the COOH-terminal feature is absent in the organelle-encoded chloroplast L21 of liverwort, this putative function is likely to be connected with the gene location in the nuclear compartment.

The transit peptide of L21 appears similar to the others we have identified (13, 14), and of chloroplast transit peptides in general (33). It is composed of a high proportion of serine/threonine residues (22%), proline (15%), and no tryptophan or tyrosine. The amino acid sequence around the cleavage point (PVAJKR) is not conserved, although it occurs between 2 alanine residues, as for L12 (13) and L35.5

The codon usage of spinach L21 is similar to that for other nucleus-encoded r-proteins (13, 14) and distinctly different from that for the r-proteins that are encoded in the chloroplast (Table II, Miniprint). Therefore, the gene which encodes spinach chloroplast L21 is typical of nuclear genes, as noted earlier in the case of r-protein L12 (13).

The distribution of chloroplast r-protein genes between the two genomes (chloroplast and nucleus) is highly conserved in land plants. A notable exception is the gene for L21 which...
was not found in the total chloroplast DNA sequence of either tobacco (6) or rice (8), but was identified in that of M. polymorpha (7). M. polymorpha (a liverwort) belongs to the division Bryophyta comprising nonvascular plants having no true roots, stems, or leaves (34). The Bryophyta are believed to have first appeared 350-400 x 10^6 years ago (35). The origin of flowering plants (angiosperms) is placed at 150-200 x 10^6 years ago (35), or somewhat earlier from recent nucleotide sequence data (36). The localization of an r-protein gene in the organelle genome in a lower plant and in the nuclear genome in a higher plant suggests, in light of the endosymbiotic theory (15, 16), a migration of the L21 gene from the organelle to the nucleus, about 200-400 x 10^6 years ago, after the first appearance of land plants. Localization of the L21 gene in other major plant taxa should enable the time period of its evolutionary occurrence to be narrowed.

Acknowledgments—We thank Panagiotis Padas for excellent technical assistance, Michael Hearne for oligonucleotide synthesis, and Prof. H. G. Wittmann for a Max-Planck-Institute fellowship (to P. M. S.).

Note Added in Proof—The cDNA library (13) was constructed with EcoRI-methylated cDNA. Nevertheless, purified L21 protein has since been sequenced to residue number 32 (Asp), to run through the internal EcoRI site (which is at position 19-20). The experimental result has exactly corresponded to the predicted sequence.

REFERENCES

Relocation of a Chloroplast Gene in Flowering Plant Evolution

Table I. Amino Acid Composition of Spinach LSU Protein (MgS)

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Determined from 16703</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>6.0</td>
</tr>
<tr>
<td>Arg</td>
<td>6.0</td>
</tr>
<tr>
<td>Asp/Asn</td>
<td>8.0</td>
</tr>
<tr>
<td>Cys</td>
<td>0.0</td>
</tr>
<tr>
<td>Thr</td>
<td>15.0</td>
</tr>
<tr>
<td>Val</td>
<td>8.0</td>
</tr>
<tr>
<td>Phe</td>
<td>15.0</td>
</tr>
<tr>
<td>Trp</td>
<td>0.0</td>
</tr>
<tr>
<td>Met</td>
<td>0.0</td>
</tr>
<tr>
<td>Lys</td>
<td>5.0</td>
</tr>
<tr>
<td>Gin</td>
<td>5.0</td>
</tr>
<tr>
<td>His</td>
<td>5.0</td>
</tr>
<tr>
<td>Leu</td>
<td>8.0</td>
</tr>
<tr>
<td>Pro</td>
<td>0.0</td>
</tr>
<tr>
<td>Ser</td>
<td>5.0</td>
</tr>
<tr>
<td>Tyr</td>
<td>5.0</td>
</tr>
<tr>
<td>Glu</td>
<td>5.0</td>
</tr>
<tr>
<td>Gly</td>
<td>5.0</td>
</tr>
<tr>
<td>Thr</td>
<td>5.0</td>
</tr>
<tr>
<td>Glu</td>
<td>5.0</td>
</tr>
</tbody>
</table>

S.D. not determined. Values are corrected for amino acids not analyzed.

Table II. Usage of synonymous codon (percent) in nuclear (guanine) and chloroplast (leucine)-encoded LSU genes

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>TTT</td>
<td>57</td>
<td>TCT</td>
<td>57</td>
<td>TTC</td>
<td>57</td>
<td>TCA</td>
<td>57</td>
</tr>
<tr>
<td>TTA</td>
<td>57</td>
<td>TCC</td>
<td>57</td>
<td>TCG</td>
<td>57</td>
<td>TCT</td>
<td>57</td>
</tr>
<tr>
<td>TAC</td>
<td>57</td>
<td>TCA</td>
<td>57</td>
<td>TGG</td>
<td>57</td>
<td>TCA</td>
<td>57</td>
</tr>
<tr>
<td>TAC</td>
<td>57</td>
<td>TCC</td>
<td>57</td>
<td>TAC</td>
<td>57</td>
<td>TCA</td>
<td>57</td>
</tr>
<tr>
<td>TAC</td>
<td>57</td>
<td>TCC</td>
<td>57</td>
<td>TAC</td>
<td>57</td>
<td>TCA</td>
<td>57</td>
</tr>
<tr>
<td>TAC</td>
<td>57</td>
<td>TCC</td>
<td>57</td>
<td>TAC</td>
<td>57</td>
<td>TCA</td>
<td>57</td>
</tr>
<tr>
<td>TAC</td>
<td>57</td>
<td>TCC</td>
<td>57</td>
<td>TAC</td>
<td>57</td>
<td>TCA</td>
<td>57</td>
</tr>
<tr>
<td>TAC</td>
<td>57</td>
<td>TCC</td>
<td>57</td>
<td>TAC</td>
<td>57</td>
<td>TCA</td>
<td>57</td>
</tr>
<tr>
<td>TAC</td>
<td>57</td>
<td>TCC</td>
<td>57</td>
<td>TAC</td>
<td>57</td>
<td>TCA</td>
<td>57</td>
</tr>
<tr>
<td>TAC</td>
<td>57</td>
<td>TCC</td>
<td>57</td>
<td>TAC</td>
<td>57</td>
<td>TCA</td>
<td>57</td>
</tr>
<tr>
<td>TAC</td>
<td>57</td>
<td>TCC</td>
<td>57</td>
<td>TAC</td>
<td>57</td>
<td>TCA</td>
<td>57</td>
</tr>
</tbody>
</table>

Computer analyses were performed on a FACOM VME computer, using the DNASU suite of programs. Alignments were determined using the NBRF database and the BLAST data base of this institute.
A ribosomal protein is encoded in the chloroplast DNA in a lower plant but in the nucleus in angiosperms. Isolation of the spinach L21 protein and cDNA clone with transit and an unusual repeat sequence.

P M Smooker, V Kruft and A R Subramanian


Access the most updated version of this article at http://www.jbc.org/content/265/27/16699

Alerts:

- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/265/27/16699.full.html#ref-list-1