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

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

(Received for publication, February 27, 1990)

From the Max-Planck-Institut für Molekulare Genetik, Abteilung Wittmann, Ihnessstee 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 rpl21. 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 NH2-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 rpl21 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 the ribosome-proteins extracted from spinach chloroplast ribosomes were subjected to purification steps which are described under "Materials and Methods." One of the pools so obtained (i.e. pool 43) contained mainly two polypeptides of molecular mass 21 and 26 kDa (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.

With a complete picture of chloroplast-encoded r-proteins and their genes emerging, the current emphasis is on isolation and characterization of the chloroplast r-protein genes which are located in the nuclear DNA. These r-proteins will be synthesized on the 80 S ribosomes, concurrent with the synthesis of the organelle-encoded components, at the specific times when ribosome assembly occurs in chloroplasts. This takes place most actively in the cells of enlarging leaf primordium and is much reduced in mature leaves (11). Since ribosome assembly is a precisely ordered process, several regulatory mechanisms, coordinated at different levels, are expected to be involved in this process.

In previous reports from two laboratories cDNA clones for four nuclear coded chloroplast r-proteins, each a homologue of a corresponding E. coli r-protein (L9, L12, L13, and L24), have been described (12–14). In this article we describe the immunosolation and characterization of a spinach cDNA clone encoding the precursor of the chloroplast r-protein L21. The gene encoding r-protein L21 is localized in the chloroplast DNA sequence of a lower plant (7) but the corresponding sequence has not been found in the fully sequenced chloroplast genomes of two flowering plants (6, 8). These results are discussed here in light of the endosymbiont theory (15, 16). The spinach L21 sequence has a striking COOH-terminal extension which is not found in the chloroplast-encoded L21 of liverwort.

MATERIALS AND METHODS

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 under "Materials and Methods." 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 antiserum against pool 43 immunoreacted with...
**FIG. 1.** Molecular mass of the proteins in pool 43 and immunostaining of the 30 S and 50 S subunit proteins of spinach chloroplast ribosomes with the pool 43 antiserum. Proteins were electrophoresed in 15% sodium dodecyl sulfate-polyacrylamide gel and either stained directly (lane 1), or transferred to Immobilon membrane, incubated with antiserum, and immunostained. Lanes 1, 2, 3, and 4 contained 1, 0.1, 2.5, and 2.5 μg, respectively, of protein.

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 Pl-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 $M$, 28,412 was found. The initiating methionine codon of this reading frame is in the context AAAAAATGGC, 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 20
```

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 ($M$, 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 PVK/AKR.

**Identity of the 21-kDa Chloroplast R-protein**—Searching the NBRF and RIB0 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 \textit{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 \textit{E. coli} and the liverwort homologues. The larger size of the mature spinach L21 protein is the result of peptide extensions at both the NH$_2$ terminus (~65 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 \textit{E. coli} proteins (Fig. 5).

The net charge of the spinach L21 protein is +1, compared to +10 and +20, respectively, for the \textit{E. coli} and liverwort proteins. Thus, unlike its \textit{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.

\section*{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 \textit{E. coli} ribosomal protein L21.

The nuclear-coded protein described here is much larger than its two known homologues, the L21 of \textit{E. coli} (31) and of the lower plant \textit{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-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 \textit{E. coli} homologue, an increase not uncommon among chloroplast-encoded r-proteins (37). The difference of 98 amino acid residues between spinach L21 and \textit{E. coli} L21 is over twice the next highest known, i.e. spinach chloroplast L13 and \textit{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 NH$_2$ 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 NH$_2$ terminus.

Fig. 5. Alignment of the amino acid sequence of spinach chloroplast L21 with those of the liverwort chloroplast and \textit{E. coli}. Invariant residues are boxed. The transit sequence and the NH$_2$ and COOH-terminal extensions of the nuclear-coded protein are shown. The Glu-Ala (E-A) repeat of the COOH-terminal extension is underlined.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig5.png}
\caption{Fig. 5. Alignment of the amino acid sequence of spinach chloroplast L21 with those of the liverwort chloroplast and \textit{E. coli}. Invariant residues are boxed. The transit sequence and the NH$_2$ and COOH-terminal extensions of the nuclear-coded protein are shown. The Glu-Ala (E-A) repeat of the COOH-terminal extension is underlined.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig6.png}
\caption{Fig. 6. The position of the L21 protein in a two-dimensional gel (19) pattern of spinach chloroplast 50 S subunit proteins. Shown for reference are two other previously described nuclear-coded chloroplast r-proteins (13, 14).}
\end{figure}

\section*{Acknowledgments}

This work was supported by grants from the National Science Foundation to the University of Illinois and to Dr. P. M. Smooker. We thank Dr. M. Pollard for help in the structural analysis of the spinach chloroplast L21 protein and Dr. W. H. Sprouls for help with the nucleotide sequence analysis. We also thank Dr. C. J. Tzagoloff for providing us with the cDNA clone that encodes the spinach L21 protein.

\section*{References}

1. P. M. Smooker, T. Choli, and A. R. Subramanian, manuscript submitted for publication.
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 × 10^6 years ago (35). The origin of flowering plants (angiosperms) is placed at 150–200 × 10^6 years ago (35), or somewhat earlier from recent nucleotide sequence data (36). The localization of an r-protein 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 exactly corresponded to the predicted sequence.

REFERENCES

Relocation of a Chloroplast Gene in Flowering Plant Evolution

Supplement to

A REMOVAL PROTEIN IS ENCODED IN THE CHLOROPLAST DNA

IN A LOWEST PLANT IN THE NOVELELLA (H. NOVELELLA)

Activation of aminotransferase (AT) protein and RNA sequence with

trans- and an unusual repeat sequence

by

Peter W. Samuels, Volker Eulalio and Alpa R. Venclovasen

Walter-Schottky-Inst. For Molecular Genetic, Abteilung Mittel, Imstetten 70, D-0400 Berlin 33.
Federal Republic of Germany

METHODS AND MATERIALS

Isolation of chloroplast ribosomes, protein, and antigens preparation

Chloroplast ribosomes were isolated from freshly harvested spinach leaves (Spinacia oleracea) by the method of Schachtele and coworkers (21). Chloroplast antigens were isolated by the method of Samuels et al. (21).

Fresh spinach leaves (2 kg) were homogenized in 5 liters of 0.02 M Na-acetate, pH 6.5, 0.05 M 2-mercaptoethanol, and subjected to centrifugation. The resulting supernatant was subjected to a second centrifugation at 20,000 g for 1 hour. The supernatant was then precipitated with 0.5 M Na-acetate, 5 liters of 0.05 M 2-mercaptoethanol, and 0.1 M Na-acetate, pH 6.5, 0.05 M 2-mercaptoethanol. The precipitate was collected by centrifugation.

Ribosomes and ribosomal proteins were isolated according to the method of Schachtele and coworkers (21).

Western blotting and immunoprecipitation

After electrophoresis, proteins were blotted to immunoblot membranes (Millipore Corp.) using the method described by Samuels et al. (21).

Protein blots were probed with antibodies to tRNA synthetases. The resulting membranes were probed with antibodies to tRNA synthetase.

Additional Protein Methods

To determine the presence of protein in the chloroplast, the following procedures were performed: electrophoresis and immunoblotting. The membranes were probed with antibodies to tRNA synthetase.

Computer Analysis

Computer analysis was performed on a FACORA-60S computer using the DNACOMPUTER software package.

Table 1

<table>
<thead>
<tr>
<th>Alanine</th>
<th>Cysteine</th>
<th>Glutamate</th>
<th>Proline</th>
<th>Serine</th>
<th>Threonine</th>
<th>Tryptophan</th>
<th>Tyrosine</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>C</td>
<td>G</td>
<td>P</td>
<td>S</td>
<td>T</td>
<td>W</td>
<td>Y</td>
</tr>
</tbody>
</table>

Table 2

<table>
<thead>
<tr>
<th>Codon Sp.</th>
<th>Leu</th>
<th>Cys</th>
<th>Trp</th>
<th>Gly</th>
<th>Pro</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TGC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TGG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Codons usage for the chloroplast RNA. The usage in the transcript peptide is for comparison to that in the mature protein.

Fig. 3: Nucleotide sequence of the DNA clone for spinach chloroplast A protein (ATP). The predicted amino acid sequence is shown in parentheses. The length of the open reading frame is indicated by the boxes. The translation starts at the first AUG codon.
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