Identification of a Plastid-specific Ribosomal Protein in the 30 S Subunit of Chloroplast Ribosomes and Isolation of the cDNA Clone Encoding Its Cytoplasmic Precursor

(Received for publication, March 28, 1990)

Clayton H. Johnson, Volker Kruft, and Alap Raman Subramanian

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

We describe the isolation and characterization of a chloroplast ribosomal protein and a clone of its cDNA. This protein has no homology to any Escherichia coli ribosomal protein or to any known proteins. Due to this novel finding we propose it be called PSrp-1, i.e. a plastid-specific ribosomal protein. The precursor form of PSrp-1, deduced from the cDNA sequence, is 302-amino acid residues long. The mature PSrp-1, identified by amino-terminal sequencing, is a protein of 236 residues. The NH2-terminal 66 amino acids form the transit peptide that targets PSrp-1 into the chloroplast. We show that PSrp-1 is a protein of the chloroplast 30 S ribosomal subunit by Western blotting and sequencing the excised protein after two-dimensional gel electrophoresis. The possible evolutionary origin of PSrp-1 from the nucleated host cell of the endosymbiont theory is discussed.

Chloroplast ribosomes belong to the eubacterial 70 S class and resemble Escherichia coli ribosomes in functional properties (1), and in the size, nucleotide sequence, and secondary structure of the rRNA (2, 3). The rRNA is encoded in the chloroplast DNA (4). A total of 21 ribosomal proteins have also been identified, in four plant species (reviewed in Refs. 5 and 6), in the chloroplast DNA. The sequence identities and ALIGN scores between the chloroplast ribosomal proteins and that of E. coli range between 25–70% and 3–55 S.D.; those at the lower end of the similarity spectrum (e.g. L22, L23) need further data in support of the assignment (6). Thirteen chloroplast ribosomal proteins have been shown to cross-react specifically with antisera to purified E. coli ribosomal proteins (7–9); in other cases antisera prepared against chloroplast ribosomal proteins have been shown to cross-react (also specifically) with E. coli ribosomal proteins (10, 11). These results provide the support for the general view that there exists a 1:1 correspondence between chloroplast and eubacterial ribosomes in their structural components.

Fifty-four ribosomal proteins are securely identified in the E. coli ribosome (12–14). Determination of the precise protein composition of chloroplast ribosomes is still in progress in several laboratories (7, 8, 10, 11, 15–17). The proteins of spinach and maize chloroplast ribosomes have been examined in our laboratory using two-dimensional gel electrophoresis, antibody cross-reactions, HPLC purification, and amino acid sequencing (7, 8, 18–20). The results have indicated a larger number of polypeptides than in the E. coli ribosome (reviewed in Ref. 21), suggesting that the chloroplast ribosome has apparently evolved to a more complex state than its eubacterial counterpart.

Two of the spinach chloroplast ribosomal proteins purified to homogeneity and sequenced by us yielded data indicating no significant homology to any of the E. coli ribosomal proteins. Antibodies were raised to the proteins and used to screen a spinach cDNA expression library (22, 23). In this paper we describe the results for one of these "extra" proteins and discuss its significance.

MATERIALS AND METHODS

RESULTS

Characterization of a 26-kDa Protein in Spinach Chloroplast Ribosomes—Fractionation of the total ribosomal proteins from spinach chloroplast ribosomes on a large gel filtration column and subsequent ion-exchange chromatography of the pooled fractions yielded over 50 pools of varying purity (19). One of the pools (No. 43) contained two polypeptides of molecular mass 21 and 26 kDa. They were purified by reverse phase HPLC on a RPSC C3 column, as shown in Fig. 1, Miniprint.

The purified 26-kDa protein subjected to 20 cycles of the peptide sequencer gave the following NH2-ter"minal sequence:

S-W-D-G-P-L-S-S-V-K-L-I-L-Q-G-R-N-N-L-E-V

This sequence exhibited no significant identity to any of the known ribosomal protein sequences in the RIBO data base. Therefore additional sequence data was generated by sequencing internal peptides. Fig. 2 shows the HPLC purification of the peptides produced by digesting the 26-kDa protein with lysine-specific protease, Lys-C (26) and Fig. 3 shows the results obtained from automated sequencing of five of the peaks from this purification.

The abbreviations used are: HPLC, high performance liquid chromatography; PSrp, plastid-specific ribosomal protein; bp, base pair; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis.

A. R. Subramanian, unpublished results.

1 Portions of this paper (including "Materials and Methods," Figs. 1, 5, and 7, and Tables I and II) are presented in miniprint at the end of this paper. Miniprint is easily read with the aid of a standard magnifying glass. Full size photocopies are included in the microfilm edition of the Journal that is available from Waverly Press.
A Chloroplast Ribosomal Protein with No Bacterial Homologue

The positions of the residues in the cDNA-derived complete sequence of the 26-kDa mature protein (Fig. 8) are also shown. Cycles of the sequenator run per peptide are given in parentheses.

**FIG. 2.** Purification of the peptides from a Lys-C protease digest of the 26-kDa protein. HPLC purification conditions similar to that of Fig. 1, Miniprint. The peaks which are numbered were analyzed in an Applied Biosystems automated sequenator.

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Amino acid sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pep 1</td>
<td>L-L-... (20)</td>
</tr>
<tr>
<td>Pep 2</td>
<td>S-R-... (21)</td>
</tr>
<tr>
<td>Pep 3</td>
<td>R-V-... (20)</td>
</tr>
<tr>
<td>Pep 4</td>
<td>V-R-... (20)</td>
</tr>
<tr>
<td>Pep 5</td>
<td>D-R-... (20)</td>
</tr>
</tbody>
</table>

**FIG. 3.** The NH₂-terminal amino acid sequences of five internal peptides from the 26-kDa protein. The positions of the residues in the cDNA-derived complete sequence of the 26-kDa mature protein (Fig. 8) are also shown. Cycles of the sequenator run per peptide are given in parentheses.

**FIG. 4.** Immunostaining of a Western blot of chloroplast 30 S and 50 S subunit (TP₃₀, TP₅₀) and pool 43 proteins. The blot was adsorbed with a 1:2000 diluted rabbit antiserum to pool 43 and the bound antibodies were revealed with peroxidase-conjugated goat anti-rabbit IgG.

**FIG. 6.** Schematic diagram of the PSrp-1 cDNA and the nucleotide sequencing strategy. The open reading frame is shown as the two functional sequences it encodes. Sequencing was done with universal and reverse primers, and synthetic oligonucleotides designed from the previous rounds of sequencing (arrows with rectangles). (A)ₗ₀ poly(A) tail.

**Fig. 7.** Miniprint.

These experiments yielded the sequence for a total of 144 amino acid residues. Comparison of this data with the sequences in RIBO/NBRF data bases again gave no significant matches.

**Immunological Identification of the 26-kDa Protein.** The protein mixture used for isolating the 26-kDa protein was not derived from highly purified (e.g. through sucrose gradient) chloroplast 70 S ribosomes. Thus, the possibility of the 26-kDa protein being a contaminating nonribosomal protein could not be ruled out. This point was answered using gradient-purified 30 S and 50 S ribosomal subunits (which were themselves isolated from gradient-purified 70 S ribosomes) and an antiserum raised against the two proteins of pool 43.

Immunostaining of SDS-PAGE separated proteins from the two subunits of spinach chloroplast ribosomes (Fig. 4) showed the 26-kDa protein to be present in TP₃₀ but absent in TP₅₀. It is thus a 30 S subunit protein. The 21-kDa protein, on the other hand, was present in TP₅₀ but absent in TP₃₀. This protein has recently been identified as the chloroplast homologue of *E. coli* L21.

Antibodies against the 26-kDa protein showed no cross-reaction with *E. coli* ribosomal proteins (data not shown). This result provided immunoechemical evidence, in line with the sequence data, that the 26-kDa protein represents a chloroplast-specific ribosomal protein.

We should mention here that the buffers commonly used for dissociating *E. coli* ribosomes (e.g. 10 mM Tris-HCl, 50 mM KCl, 0.1 mM MgCl₂, 7 mM 2-mercaptoethanol, pH 7.6) give very incomplete dissociation of chloroplast ribosomes. We used a phosphate buffer (10 mM potassium phosphate, 100 mM KCl, 1 mM MgCl₂, 7 mM 2-mercaptoethanol, pH 7.0) which was found to be more effective. Since data in the literature on chloroplast ribosome dissociation are limited, details of the conditions we used, including gradient profiles (Fig. 5), are given in the Miniprint.

**Immunosolation of a cDNA Clone Encoding the 26-kDa Protein.** The antiserum described above was used to screen two previously described (22, 23) spinach cDNA expression libraries. Positive recombinant phage plaques were obtained from both. These were purified to homogeneity by repeated replating and phage DNA prepared. Several restriction enzyme sites were mapped in the cDNA. The results showed two classes of inserts: one of 1050 bp length with an internal EcoRI site, and the other of 1240 bp length with internal BamHI and HindII sites. The latter class produced the fusion protein which reacted with antibodies specific to the 26-kDa protein (the cDNA with internal EcoRI site was of the 21-kDa protein, see Ref. 24).

**Nucleotide Sequence of the cDNA and Identification of a Reading Frame.** The 1240-bp cDNA was subcloned into pT7/T3-19 U and the nucleotide sequence was determined (Fig. 6). It was found to be 1242 bp long including a poly(A) tail of 30 adenines (Figs. 7, Miniprint).

The nucleotide sequence was searched for possible protein coding regions and a long open reading frame was identified beginning at nucleotide 27 and terminating at nucleotide 935. It predicts a protein of 302 amino acid residues in length and 33,752 kDa in molecular mass. The initiating AUG codon, proximal to the 5' end (Ref. 27), is in the context AAAGATGGC with an A at position -3 and a G at position +4 (the A of the initiating codon is +1), as is found in the majority of plant initiation sequences (28). The terminating codon is TAG, which is followed by a 3' downstream noncoding sequence of 277 bp before the poly(A) tail begins. The 5' upstream noncoding sequence (before the AUG codon) is relatively short, i.e. 26 bp.
Fig. 8. The complete amino acid sequence of the precursor form of spinach chloroplast PSrp-1. The cleavage site where the transit peptide is removed from the mature functional ribosomal protein is shown by an arrow. The sequence stretches determined on the protein sequenator are underlined. Hydroxylated (serine/threonine) and charged amino acids concentrated in the transit peptide, and the negatively charged region of PSrp-1 are marked.

Identification of the Transit Peptide and the Mature Protein Sequences—The protein predicted from the cDNA sequence is considerably larger than the 26-kDa protein identified in the chloroplast TPo, indicating a precursor form. Chloroplast proteins synthesized in the cytosol have NH2-terminal transit sequences that are removed during organelle import. Examination of the predicted protein sequence (Fig. 7) in light of the NH2-terminal sequence from the mature protein shows an exact correspondence beginning at residue 67. Hence, the mature protein is 236 amino acid residues long with $M_s$ 26,791.

The five peptide sequences determined previously could now be precisely identified in the mature protein sequence. They are peptide 1, positions 11–30; peptide 2, positions 39–59; peptide 3, positions 74–99; peptide 4, positions 124–151; and peptide 5, positions 164–202 (Figs. 3 and 8). The $M_s$ of the mature protein (26,791) is in reasonable agreement with that determined from SDS-gel electrophoresis (29,000). Also, the amino acid composition of the purified protein (determined after acid hydrolysis) agreed reasonably well with that calculated from the predicted protein sequence (data not shown).

The transit peptide of the 26-kDa protein is a sequence of 66 residues, rich in serine and threonine (30%) but lacking aspartic/glutamic acid, tyrosine, and tryptophan. This feature in composition is characteristic of all the chloroplast ribosomal protein transit sequences we have identified so far (Table I, Miniprint). The proline content is, however, only 3%, lower than the proline in the others.

The complete amino acid sequence shown in Fig. 8 was compared against all the known protein sequences in the RIBO and NBRF data bases. No significant homology to any known protein was found. Thus the chloroplast ribosomal protein we have identified and characterized here does not appear to be a homologue of any known protein. We have therefore designated this protein PSrp-1 (plastid-specific ribosomal protein-1), i.e., a member of a small group of chloroplast ribosomal proteins that have no homologues in the eubacterial ribosome.

The cDNA encoding PSrp-1 carries features common to the mRNAs of other nuclear-coded chloroplast ribosomal proteins that have been characterized (22–24, 34). These include a relatively short (26 bp) 5' noncoding region, a long 3' noncoding region, and two neighboring AUG codons, 9 codons apart, at the beginning of the transit peptide coding region. We have recently shown that ribosomes initiate at 41.1% frequency at the 1st and 2nd of the two initiation codons in the mRNA of ribosomal protein L12. This feature distinctly enhances the efficiency of synthesis of nuclear-coded chloroplast ribosomal proteins. The latter are encoded by one/a few genes in contrast to the chloroplast-encoded ribosomal proteins, which have gene dosage of 103–104/cell (29).

The codon preference of PSrp-1 is rather unusual for a nuclear ribosomal protein gene. A preference for A and T in the third nucleotide of the codon is apparent (Table II, Miniprint). In the previous cases of nuclear-coded chloroplast

1 Chloroplasts are one of the several terminally differentiated forms (e.g., chromoplast, amyloplast, elaioplast) of the organelle of the general name plastid (29).
ribosomal proteins (22–24) the preference at the 3rd position is for G and C. So it appears that for some reason the PSrp-1 gene has maintained a codon preference (A/T) found primarily in the ribosomal protein genes that are located in the chloroplast DNA (5, 6). At this time the significance for this codon preference is difficult to determine due to lack of necessary information.

The polypeptide of PSrp-1 has some regional features of protein–chemical interest. The net charge of the first 48 residues of PSrp-1 is zero at the physiological pH, which could perhaps aid in membrane transport. The next 86 residues carry clusters of acidic and basic groups and a net charge of +11, i.e. is highly basic as most ribosomal proteins. The next 59 residues have a high concentration of aspartic/glutamic acids giving it a net charge of −16 (Fig. 8). This region (25% of the molecule) contains 43% of all the acidic residues of PSrp-1, and possibly has a specific functional role. Finally, the last 43 residues has again a net charge of zero. Thus the slight overall acidic character of PSrp-1 is concentrated in one very acidic region of the protein leaving the remainder of the molecule basic. The mature protein would have a net charge of −5 with nonprotonated and +2 with protonated histidines.

The inference that PSrp-1 is a weakly acidic/neutral protein agrees with the migration position of the protein in the two-dimensional gel pattern of chloroplast ribosomal proteins, which was determined using purified PSrp-1 and TP50, and is shown in Fig. 9. The presence of PSrp-1 within the labeled spot in Fig. 9 was confirmed by excising the spot from a Western blot of TP50 (derived from sucrose gradient-purified 30 S ribosomal subunits) and analyzing it in an automated protein sequenator. The sequence obtained (7 cycles of degradation) was S-W-D-G-P-L-S-, identical to the first 7 residues of PSrp-1. Chloroplast ribosomal protein S5, which partially overlaps with the PSrp-1 spot (Fig. 9) appears to have a blocked NH2 terminus, a fact that helped to provide a clean result in this experiment.

An endopeptidase functioning in the periplasm of the double-membraned chloroplast envelope is thought to cleave off the NH2-terminal transit peptide (36). The cleavage site of the PSrp-1 precursor has the sequence context VVC[M]SWDG (Fig. 8), which has no apparent similarity to the cleavage sites of the other chloroplast ribosomal proteins that we have precisely identified earlier (22–24). 5

While this work was in progress a paper describing the isolation and nucleotide sequence of a cDNA clone encoding a spinach chloroplast ribosomal protein designated CS-S5 has appeared (37). According to a recent correlation of the chloroplast ribosomal protein nomenclatures used by different groups (21), CS-S5 corresponds to the protein spot designated PSrp-1 here. The published sequence (1184 bp long including the poly(A) tail encodes a polypeptide of 321 residues compared to the 1242-bp sequence given here encoding 302 residues. The nucleotide sequence of Ref. 37 is entirely contained within our data (beginning at position 14, Fig. 7, Miniprint) except for a missing T at position 836 and two nucleotide differences at positions 378 and 715. The missing nucleotide, however, changes the reading frame after residue 137 (-TREEV, Fig. 8) of the protein sequence, resulting in a predicted protein with a different, and longer COOH-terminal region.

Data supporting the correctness of the sequence given here are of three kinds. 1) We have always obtained four T's in one strand and four A's in the other strand; and not three as in Ref. 37: (Fig. 10). 2) The nucleotide sequence in Ref. 37 predicts a mature protein of net positive charge +33. Such a highly basic protein will not migrate to the position of PSrp-1 shown in Fig. 9. This large positive charge is due to loss of the acidic region (beginning at position 135) pointed out earlier and its replacement by a sequence with many basic residues, due to the frameshift. 3) The sequenced segments of PSrp-1 include peptide 5, which spans this disputed region of four T's (Figs. 3, 7, and 8). The amino acid sequence of peptide 5 (Fig. 3) is exactly that predicted by our nucleotide sequence after the disputed region, i.e. LEEVE- and not WRKWN- as in Ref. 37. The peptide sequences of peptides 2 and 5 also confirm our nucleotide reading at the positions 378 and 715. Thus we believe that the amino acid sequence (Spinacia oleracea cv. Alwaro, Ref. 19) and the nucleotide sequence (S. oleracea cv. Matador, Refs. 22 and 23) described in this paper are correct in the case of the two cultivars of spinach we have used.

Acknowledgments—We thank Dr. S. H. Phua for the initial screening of one of the cDNA expression libraries and Prof. H. G. Wittmann for a critical experimental suggestion and a Max-Planck Fellowship to C. H. J.

REFERENCES

---

**Fig. 9.** The position of PSrp-1 in a two-dimensional gel pattern of spinach chloroplast 30 S subunit proteins. Shown for reference is S5 which was identified from protein sequence homology. Two-dimensional electrophoresis as described in Ref. 35.

**Fig. 10.** Parts of two sequencing gels showing the ACGT ladder at the nucleotide sequence positions 629–642 (Fig. 7).
A Chloroplast Ribosomal Protein with No Bacterial Homologue

### Table I. Characteristics of amino acid composition of known chloroplast r-protein-ligating proteins where the storage site is precisely identified. Data taken from PSPr-1, this paper, 121, ref. 22, 122, ref. 24, 125, see footnote 5.

<table>
<thead>
<tr>
<th>Chloroplast</th>
<th>r-protein</th>
<th>Length</th>
<th>Net charge</th>
<th>Tripeptide</th>
<th>Not percent of</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSPr-1</td>
<td>86</td>
<td>+5</td>
<td>0 0 36 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>112</td>
<td>85</td>
<td>+5</td>
<td>0 0 26 16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LSR</td>
<td>86</td>
<td>+5</td>
<td>0 0 30 6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table II. Codon usage (in %) of PSPr-1 and chloroplast encoded r-proteins.

<table>
<thead>
<tr>
<th>Codon PSPr Ch</th>
<th>Codon PSPr Ch</th>
<th>Codon PSPr Ch</th>
<th>Codon PSPr Ch</th>
<th>Codon PSPr Ch</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATT Phe 67 64</td>
<td>TCT Ser 35 26</td>
<td>TAG Trp 40 36</td>
<td>TGA Cys 100 41</td>
<td></td>
</tr>
<tr>
<td>TCC Phe 45 35</td>
<td>TCT Ser 10 25</td>
<td>TAC Trp 40 36</td>
<td>TCC Cys 0 9</td>
<td></td>
</tr>
<tr>
<td>TTA Leu 5 5</td>
<td>TCA Ser 20 16</td>
<td>TAA End 0 75</td>
<td>TGA End 0 5</td>
<td></td>
</tr>
<tr>
<td>ATT Leu 40 22</td>
<td>TCC Ser 0 0</td>
<td>TAG End 100 60</td>
<td>TGC Trp 100 100</td>
<td></td>
</tr>
</tbody>
</table>

| CTT Leu 20 24 | GTT Pro 30 0  | GCT Ala 74 80 | GCT Arg 2 40 |
| CTC Leu 11 5  | CCC Pro 27 25 | CAC His 29 16 | GCA Arg 5 3  |
| GTA Leu 0 1   | CCA Pro 30 25 | CCA Gin 40 20 | GCA Arg 21 25 |
| CTC Leu 5 5   | CCG Pro 0 15  | CCA Gin 40 15 | GCA Arg 5 6  |

| ATT Leu 37 42 | ATT Thr 2 2  | ATT Asp 49 47 | ATT Arg 2 2 |
| TAC Leu 15 16 | ACC Thr 22 22 | ACC Asp 63 67  | ACC Ser 15 5 |
| GTA Leu 8 16  | AGA Thr 64 04 | AGA Lys 80 70 | AGA Arg 06 21 |
| GTA Met 100 100 | AGC Thr 0 0  | AGG Lys 60 24  | AGG Arg 37 10 |

| GTT Val 32 34 | GCT Ala 44 37 | GAT Asp 88 74 | GCT Gly 21 26 |
| GTC Val 21 16 | GCC Ala 11 17 | GAC Asp 32 26  | GCC Gly 21 26 |
| GCA Val 21 36 | GCA Ala 44 41 | GAA Gin 60 70 | GCA Gin 36 48 |
| GCG Val 22 12 | GCG Ala 0 5   | GAG Gin 40 25  | GGC Gly 21 16 |

* Computed from the data for tobacco chloroplast ENL in ENL database.*