Plant Cytosolic Ribosomal Protein S11 and Chloroplast Ribosomal Protein CS17
THEIR PRIMARY STRUCTURES AND EVOLUTIONARY RELATIONSHIPS*

J. Stephen Gantt and Michael D. Thompson
From the Department of Plant Biology and the Plant Molecular Genetics Institute, The University of Minnesota, St. Paul, Minnesota 55108

We have isolated cDNA clones specific for *Arabidopsis thaliana* cytosolic ribosomal protein S11 and plastid ribosomal protein CS17, both of which are encoded in the nuclear genome, through the use of the corresponding soybean and pea cDNAs as probes, respectively. The nucleotide sequences of all four cDNAs were determined. The amino acid sequences derived from these cDNA sequences show that the soybean and *A. thaliana* S11 cDNAs encode proteins that are homologous to rat ribosomal protein S11 and that the pea and *A. thaliana* CS17 cDNAs encode proteins that are homologous to *Escherichia coli* ribosomal protein S17. The plant S11 cytosolic ribosomal proteins also show significant sequence similarity to both *E. coli* ribosomal protein S17 and plastid CS17 indicating that these are all related proteins. Comparison of *A. thaliana* CS17 with *A. thaliana* S11 and with *E. coli* S17 suggests that CS17 is more related to S17 than it is to S11. These results support the idea that the gene encoding CS17 was derived from a prokaryotic endosymbiont and not from a duplication of the eukaryotic S11 gene.

Plants contain three types of ribosomes: one located in the cytosol that is very similar to the cytosolic ribosomes of other eukaryotes, one located in mitochondria whose properties are not well defined, and one located in the plastid that is remarkably similar to ribosomes found in eubacteria (Boynton et al., 1979). The similarity of the plastid and eubacterial ribosomes extends to the nucleotide sequences of the rRNAs (Schwarz and Kossel, 1980; Edwards and Kossel, 1981) and to the amino acid sequences of many of the ribosomal proteins (Ohyama et al., 1986; Shinozaki et al., 1986; Gantt, 1988; Giese and Subramanian, 1989; Phua et al., 1989).

Plastid rRNAs are encoded in the plastid genome; however, only one-third of the approximately 55 plastid ribosomal proteins are encoded in the plastid genome (Ohyama et al., 1986; Shinozaki et al., 1986). The remaining two-thirds are apparently encoded in the nuclear genome. These nuclear-encoded plastid ribosomal proteins are translated in the cytosol as precursor proteins that are proteolytically processed to their mature size during or subsequent to their transport into the plastid (Schmidt et al., 1984; Gantt and Key, 1986). It is generally believed that these nuclear genes were originally encoded in the genome of the eubacterial endosymbiont that evolved into modern chloroplasts and that the genes were transferred to the nuclear genome early in plant evolution (Margulis and Bermudes, 1985). This theory, in part, is generally supported by data that show significant amino acid sequence similarities between nuclear-encoded plastid proteins and bacterial proteins.

We previously isolated cDNA clones that are complementary to cytosolic and nuclear-encoded plastid ribosomal protein mRNAs from the legumes soybean and pea, respectively (Gantt and Key, 1985, 1986). We have recently used a soybean and a pea cDNA to isolate the corresponding *Arabidopsis thaliana* cDNA clones. In this report, we describe the nucleotide sequences of the isolated soybean, pea, and *A. thaliana* cDNAs. The predicted amino acid sequences from the large open reading frames contained in each of these cDNAs show that cytosolic ribosomal protein S11 is encoded by the cDNA isolated from soybean and the corresponding *A. thaliana* cDNA while the pea cDNA and its corresponding *A. thaliana* cDNA encode plastid ribosomal protein CS17. The proteins encoded in each of these cDNAs exhibit considerable sequence similarity with *Escherichia coli* ribosomal protein S17, *Halo- bacterium marismortui* S14, and rat S11. This sequence similarity, together with the localization of these proteins to the small ribosomal subunit (Gantt and Key, 1985; 1986), suggests that plant S11 and CS17 are homologous proteins. In order to explore the possible evolutionary relationship of CS17 and S11, we have compared the amino acid sequences of *A. thaliana* CS17 and *A. thaliana* S11 with each other, with the partial sequences of soybean S11 and pea CS17, and with homologous proteins from several other taxa. These comparisons suggest that the amino acid sequences of *A. thaliana* CS17 and S11 are not as similar to one another as CS17 is similar to *E. coli* S17 and *A. thaliana* S11 is similar to rat S11. These results suggest that CS17 did not descend from S11, but rather it is related to the prokaryotic lineage of the endosymbiont.

METHODS
Isolation of *A. thaliana* cDNA Clones—The *A. thaliana* cDNA clones that are complementary to S11 and CS17 mRNA were identified by their hybridization to 32P-labeled soybean S11 and pea CS17 cDNAs isolated by Gantt and Key (1985, 1986). Soybean S11 and pea CS17 were originally named GmS16 and PsCS16, respectively. However, because of the results reported here, they have been renamed according to the standard nomenclature system which is based on sequence similarity with rat proteins, in the case of cytosolic ribosomal proteins, and *E. coli* proteins, in the case of plastid ribosomal proteins. The pea and soybean cDNAs were labeled with 32P according to Feinberg and Vogelstein (1985). These labeled cDNAs

* This work was supported in part by the University of Minnesota Graduate School and United States Public Health Service Grant GM38769. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) M31024, M31025, J09215, and J09216.
were used to probe an *A. thaliana* cDNA library constructed in λgt10
(kindly provided by Dr. Nigel Crawford, Stanford University). Filters
were hybridized in 6 x SSC, 0.05 mg/ml sheared salmon sperm DNA,
0.2% SDS, 5 x Denhardt's solution, and 10⁶ cpm/ml of ³²P-labeled
probe at 60 °C for 16 h. The filters were washed for 15 min three
times in 3 x SSC, 0.1% SDS at room temperature and once in 3 x
SSC, 0.1% SDS at 60 °C.

**Results and Nucleic Acid Sequence Determination.** The cDNA
inserts from the *A. thaliana* CDNA library were subcloned into the
plasmid vector pUC119 (Viera and Messing, 1987). Overlapping dele-
tions of the cDNAs were generated as described by Dale et al. (1995).

The chain termination technique described by Sanger et al. (1977)
and modified for use with Sequenase™ (U.S. Biochemicals) was used
to sequence the cDNAs. Soybean S11 and pea CS17 were sequenced
by the procedures described by Maxam and Gilbert (1980) as modified
by Nagao et al. (1985).

**RESULTS AND DISCUSSION**

**Analysis of cDNA and Predicted Amino Acid Sequences.**—A
pea nuclear-encoded plastid ribosomal protein cDNA clone (CS17,
originally named PsCS16), obtained by Gantt and Key (1986), was used to isolate a corresponding cDNA clone from an
*A. thaliana* cDNA library. The nucleotide and derived amino
acid sequence of this cDNA and the pea CS17 cDNA, which appears to be truncated at its 5'-end, are shown in Fig.
1. The *A. thaliana* cDNA contains a long uninterrupted reading
frame of 447 nucleotides in the putative coding region,
corresponding to a 16,280-dalton protein. The 5'- and 3'
untranslated regions consist of 30 and 142 nucleotides,
respectively, in addition to a 27-nucleotide poly(A) tail. There
are 212 nucleotides in the 3'-untranslated region plus a 15
nucleotide sequence similarity in the 3'-untranslated region.

The predicted amino acid sequence of *A. thaliana* CS17 is
consistent with its being a complete sequence for a nuclear-
encoded plastid protein. These proteins are synthesized in
the cytosol as precursors that are proteolytically processed
to their mature size through the release of a transit peptide
during or shortly after their passage through the plastid
envelope membranes (Lubben et al., 1988). Transit peptides
are generally 30 to 50 amino acids long, located at the amino-
terminal end of the protein, rich in serine and/or threonine
residues, particularly near the amino terminus of the precur-
sor protein, and typically lack aspartic or glutamic acid resi-
dues (Schmidt and Mishkind, 1986). The amino acid sequence
deduced for the amino-terminal region of *A. thaliana* CS17
fits this general structure very well. Six of the eight amino
acids immediately following our proposed amino-terminal me-
thionine are either serine or threonine. Additionally, the first
acidic amino acid occurs at amino acid position 63. Both of
these observations are consistent with the amino terminus of
*A. thaliana* CS17 being encoded by the AUG codon at nucleo-
tide 1 (Fig. 1).

Although the amino-terminal region of *A. thaliana* CS17
strongly resembles a transit peptide, the site of proteolytic
cleavage to produce the mature protein is not clear. The amino
acid sequence Gly-Arg-Val has been implicated in the recog-
nition of the transit peptide by the transit peptidase (Schmidt
The codons for this sequence start at nucleotide 163 in *A. thaliana*
CS17 cDNA (Fig. 1). However, an amino acid sequence com-
parison between CS17 and *E. coli* S17 (Fig. 4, discussed below)
shows that this sequence is aligned with an identical sequence
in what we believe to be the homologous prokaryotic protein.
This alignment suggests that the peptide Gly-Arg-Val forms

![Fig. 1. Pea and *A. thaliana* plastid ribosomal protein CS17
CDNA clones. A, the strategy employed for sequencing the full length
CDNA encoding *A. thaliana* CS17. The heavy line represents the
cDNA insert, and the dashed line represents vector sequence. Arrows
represent the extent of sequence data from either the full length
insert or from deletions derived from it. E represents the EcoRI
restriction endonuclease cloning sites bordering the cDNA. B, nu-
cleotide and the predicted amino acid sequences of *A. thaliana* (AtCS17) and pea (PsCS17) CDNA clones. Asterisks (*) indicate
identical amino acids in both protein sequences. Underlined amino
acids are substitutions that occur frequently in related proteins. Gaps
in the sequences have been inserted to maintain optimal sequence
alignment.

![Fig. 1. Pea and *A. thaliana* plastid ribosomal protein CS17
CDNA clones. A, the strategy employed for sequencing the full length
CDNA encoding *A. thaliana* CS17. The heavy line represents the
cDNA insert, and the dashed line represents vector sequence. Arrows
represent the extent of sequence data from either the full length
insert or from deletions derived from it. E represents the EcoRI
restriction endonuclease cloning sites bordering the cDNA. B, nu-
cleotide and the predicted amino acid sequences of *A. thaliana* (AtCS17) and pea (PsCS17) CDNA clones. Asterisks (*) indicate
identical amino acids in both protein sequences. Underlined amino
acids are substitutions that occur frequently in related proteins. Gaps
in the sequences have been inserted to maintain optimal sequence
alignment.

![Fig. 1. Pea and *A. thaliana* plastid ribosomal protein CS17
CDNA clones. A, the strategy employed for sequencing the full length
CDNA encoding *A. thaliana* CS17. The heavy line represents the
cDNA insert, and the dashed line represents vector sequence. Arrows
represent the extent of sequence data from either the full length
insert or from deletions derived from it. E represents the EcoRI
restriction endonuclease cloning sites bordering the cDNA. B, nu-
cleotide and the predicted amino acid sequences of *A. thaliana* (AtCS17) and pea (PsCS17) CDNA clones. Asterisks (*) indicate
identical amino acids in both protein sequences. Underlined amino
acids are substitutions that occur frequently in related proteins. Gaps
in the sequences have been inserted to maintain optimal sequence
alignment.

![Fig. 1. Pea and *A. thaliana* plastid ribosomal protein CS17
CDNA clones. A, the strategy employed for sequencing the full length
CDNA encoding *A. thaliana* CS17. The heavy line represents the
cDNA insert, and the dashed line represents vector sequence. Arrows
represent the extent of sequence data from either the full length
insert or from deletions derived from it. E represents the EcoRI
restriction endonuclease cloning sites bordering the cDNA. B, nu-
cleotide and the predicted amino acid sequences of *A. thaliana* (AtCS17) and pea (PsCS17) CDNA clones. Asterisks (*) indicate
identical amino acids in both protein sequences. Underlined amino
acids are substitutions that occur frequently in related proteins. Gaps
in the sequences have been inserted to maintain optimal sequence
alignment.
Plant mRNAs (Heidecker and Messing, 1986). This sequence of the pea CS17 cDNA sequence contains possible poly(A) addition signals. A related sequence, AAUGA, consequently is probably not the polyadenylation signal. The nucleotide and the predicted amino acid sequence does not contain the entire Sll coding region, is 80% identical with the A. thaliana nucleotide sequence in the overlapping protein coding region. The three adenosines at nucleotide 133, shows that 66% are identical, leading to an average number of substitutions per site was calculated as described by Nei (1987). This type of analysis for determining the relatedness of sequences was used because it at least partially accounts for multiple substitutions at a single site which have undoubtedly occurred due to the great diversity of the organisms whose protein sequences are being compared. Insertions or deletions, which occur infrequently in the sequence alignments, are not included in the computation of the number of substitutions per site.

Poly(A) Addition Signal—The sequence AAUAA is found in most nuclear-encoded plastid mRNAs. It is possible that there are different signals or different mechanisms for poly(A) addition of the nuclear-encoded plastid ribosomal protein mRNAs. It should be noted that, with the exception of A. thaliana CS17, we do not yet have proof that these poly(A) sequences are added post-transcriptionally.

Protein Sequence Comparisons of A. thaliana CS17 and S11 with Evolutionarily Related Proteins—We have used the DFASTP program (Lipman and Pearson, 1985) to compare and align the amino acid sequences of A. thaliana CS17 and S11 with several other ribosomal protein sequences in an effort to increase our understanding of their function and evolution. Results of these analyses show that A. thaliana CS17 and S11 have amino acid sequence similarities with each other as well as to small ribosomal subunit proteins from a very diverse group of organisms including rat, E. coli, and H. marismortui representing the animal, eubacterial, and archaebacterial kingdoms. Since both A. thaliana CS17 and S11 have significant amino acid sequence similarity with a common group of proteins and both of these proteins are located on the small ribosomal subunit (Gantt and Key, 1985, 1986), it is likely that they are evolutionarily related proteins with an analogous function. Thus, by comparing their sequences, we may be able to learn about their evolution and predict which regions of the protein are crucial for its function. To perform these comparisons, the DFASTP program was used to align the amino acid sequences, and the average number of substitutions per site was calculated as described by Nei (1987). This type of analysis for determining the relatedness of sequences was used because it at least partially accounts for multiple substitutions at a single site which have undoubtedly occurred due to the great diversity of the organisms whose protein sequences are being compared. Insertions or deletions, which occur infrequently in the sequence alignments, are not included in the computation of the number of substitutions per site.

A comparison of the predicted amino acid sequences of A. thaliana and pea CS17 shows that there is poor conservation among the amino acids near the amino-terminal ends of the proteins (Fig. 1). This low degree of sequence similarity is consistent with this region functioning as the transit peptide whose sequences typically diverge more rapidly than do the sequences of mature proteins. Comparison of the remaining amino acids, starting with the proline whose codon starts at nucleotide 133, shows that 66% are identical, leading to an

2 J. S. Gantt, unpublished observations.

3 M. D. Thompson, unpublished data.
that it plays in ribosome function. There are, however, several pressure in eukaryotes possibly reflecting some critical role the cDNA sequence (Tanaka et al., 1985).

Despite this, the amino acid substitutions of the chloroplast protein reflects a greater and rat Sll, which most likely diverged early in eukaryotic evolution. Although both CS17 and Sll are nuclear-encoded, plastid CS17 (0.41 substitution per site) is evolving at a significantly higher rate than is the homologous cytosolic ribosomal protein, Sll (0.12 substitution per site). Whether the higher rate of amino acid substitutions of the chloroplast protein reflects a greater rate of evolution of the chloroplast ribosome, as these data suggest, is not known.

Comparison of the amino acid sequences of A. thaliana Sll and rat Sll, which most likely diverged early in eukaryotic evolution and more than 0.8 × 10^9 years ago (Knoll, 1985, a and b), is shown in Fig. 3. The DFASTP program yields two single-amino acid insertions/deletions to maintain alignment of the sequences. The two sequences are identical at 59% of the residues (0.53 ± 0.07 substitution per site), and 28% of the residues are known to frequently replace one another in homologous proteins. There are several regions in the sequences that are particularly well conserved such as the region from amino acid 38 to 73 which is 86% identical. This would suggest that this region is under a high degree of selective pressure in eukaryotes possibly reflecting some critical role that it plays in ribosome function. There are, however, several regions where the sequence similarity is not pronounced, for example, the amino acids 22 through 34 in which the rat Sll and soybean Sll protein sequences. Colons (:), identical amino acids; periods (.), amino acids that frequently replace one another in evolutionarily related proteins; dashes (−), insertions or deletions. Numbers refer to the amino acid positions in the rat protein which was derived from the cDNA sequence (Tanaka et al., 1985).

A. thaliana Sll is more similar to E. coli S17, and probably the original endosymbiont, than it is to A. thaliana Sll. Thus, the amino acid sequence comparison of these proteins includes members of four kingdoms and chloroplasts, the majority of plastid proteins are encoded in the nuclear genome. The genes for most of these proteins are thought to have originally been located in the genome of the prokaryotic endosymbiont, to have been physically transferred to the nucleus, and then modified to become functional. Alternatively, plastid proteins could have evolved from duplicated nuclear genes. It may be possible to differentiate between these two models of plastid gene evolution by comparing the amino acid sequences of homologous cytosolic, plastid, and prokaryotic proteins. Amino acid sequence comparison of A. thaliana CS17 with E. coli S17 shows that these sequences are similar throughout their entire overlapping region with only two single-amino acid insertions or deletions (Fig. 4). Of the 75 overlapping amino acids in the two proteins, 39% are identical (0.95 ± 0.12 substitution per site) and 45% are known to replace one another frequently in homologous proteins. The amino acid sequences of A. thaliana CS17 and Sll were compared and aligned by the DFASTP program. In order to maximize sequence alignment, there is one single position spacing added to the CS17 sequence. The amino acid sequences of the two proteins are 30% identical (1.21 ± 0.17 substitution per site) in the 34-amino acid overlap region between amino acids 60 and 144 of S11. In this same region, 38% of the amino acid substitutions are known to occur frequently in homologous proteins. These data suggest that A. thaliana Sll is more similar to E. coli S17, and probably the original endosymbiont, than it is to A. thaliana Sll. However, the standard error in calculating the amino acid substitutions per site is too large to allow definitive support of the gene transfer hypothesis and the endosymbiotic theory of chloroplast evolution.

The amino acid sequence alignments of the S17 homologues found in a eubacterium (E. coli S17), a chloroplast (A. thaliana CS17), an archaebacterium (H. marismortui S14), an animal (rat Sll), and a plant (A. thaliana Sll) are shown in Fig. 5. Thus, the amino acid sequence comparison of these proteins includes members of four kingdoms and chloroplasts, the amino acid substitutions of the chloroplast ribosome suggests that two or more insertion/deletion events have occurred in this region that result in shifts of the reading frame (data not shown). Interestingly, the amino acid sequences of soybean and A. thaliana Sll in this same region are identical except for a single serine/threonine mismatch.
was essentially identical to that shown for \([3H]\)cholesterol.

Identical studies in which 1.4 mM DTNB (final concentration) was included in the incubations showed a similar time course of cholesteryl ester transfer from HDL to LDL. In control incubations at 4 °C, less than 4% of HDL \([3H]\)cholesteryl oleate transferred to LDL in 24 h.

The LDL and HDL fractions were separately pooled from standard incubations and were extracted, and the various lipid classes were separated for the determination of cholesterol and esterified cholesterol mass and radioactivity. To assess LCAT activity in these incubations, LDL and HDL cholesterol and esterified cholesterol masses were measured in 5-h incubations and compared with that in identical samples incubated in the presence of 1.4 mM DTNB. Fig. 2 shows a cholesteryl ester mass increase of 0.36 and 0.19 pmol/ml of the incubation mixture for LDL and HDL respectively, whereas cholesterol mass decreased 0.30 and 0.21 pmol/ml for HDL and LDL, respectively. Determination of the initial (0–60 min) linear accumulation rate of cholesteryl ester (data not shown) estimated plasma LCAT activity to be 129 pmol/h/liter of plasma in these experiments, a rate similar to that reported by others (9).

After incubation of whole plasma with \([3H]\)cholesteryl oleate and cholesteryl \([14C]\)oleate-labeled HDL, a time-dependent accumulation of lipoprotein \([3H]\)cholesterol was observed following separation of cholesterol and cholesteryl esters by TLC. The data presented in Fig. 3 indicate that the formation of \([3H]\)cholesterol was linear for several hours, with an observed maximum of 4% of the initial tritiated moiety derived from \([3H]\)cholesteryl oleate having accumulated by 24 h. It is interesting to note that the absolute increase of \([3H]\)cholesterol was greater in the HDL than in the LDL fraction by 24 h of incubation. The data in Fig. 3 indicate that the formation of \([3H]\)cholesterol was greater in the HDL than in the LDL fraction by 24 h of incubation. The data in Fig. 3 indicate that the formation of \([3H]\)cholesterol was greater in the HDL than in the LDL fraction by 24 h of incubation.
Plant cytosolic ribosomal protein S11 and chloroplast ribosomal protein CS17. Their primary structures and evolutionary relationships.
J S Gantt and M D Thompson


Access the most updated version of this article at http://www.jbc.org/content/265/5/2763

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/265/5/2763.full.html#ref-list-1