Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 405–408, January 1984
Biochemistry

Chloroplast tRNA^{Gly} gene contains a long intron in the D stem: Nucleotide sequences of tobacco chloroplast genes for tRNA^{Gly} (UCC) and tRNA^{Arg} (UCU)

(molecular cloning/precursor RNA/nuclease S1 mapping/open reading frame)

HIROSHI DENO AND MASAHIRO SUGIURA*

Department of Biology, Faculty of Science, Nagoya University, Chikusa, Nagoya 464, Japan

Communicated by Dan L. Lindsley, October 3, 1983

ABSTRACT **The nucleotide sequences of tobacco chloroplast genes for tRNA^{Gly} (UCC) and tRNA^{Arg} (UCU) have been determined. The tRNA^{Gly} gene has a 691-base-pair intron located in the D stem while the tRNA^{Arg} gene does not have any intron. The tRNA^{Gly} and tRNA^{Arg} genes are encoded on the same strand and separated by a 169-base-pair spacer. The tRNA^{Gly} gene is transcribed as a 900-base precursor RNA molecule. The tRNA^{Gly} and tRNA^{Arg} deduced from the DNA sequences show 84% and 55% sequence homologies with *Escherichia coli* tRNA^{Gly} (UCC) and phage T4 tRNA^{Arg} (UCU), respectively.**

Since the first evidence was found for the presence of introns in yeast tRNA^{Tyr} genes (1), introns have been found in many eukaryotic tRNA genes (2). Some tRNA genes from chloroplasts in plants also contain introns (3–6), although chloroplast tRNAs show high sequence homologies with prokaryotic tRNAs (7), and sequences similar to the *Escherichia coli* Pribnow box and −35 region are found in the regions upstream from chloroplast tRNA genes (6, 8–11). The introns in chloroplast tRNA genes are very long [458–949 base pairs (bp)], compared with those of nuclear tRNA genes (13–60 bp) (12). The intron sites of all tRNA genes so far analyzed are located in the anticodon loops (7).

Recently, we determined the nucleotide sequences of tobacco chloroplast genes for six tRNA species (4, 6, 8, 9). We show here the nucleotide sequences of the tRNA^{Gly} and tRNA^{Arg} genes from tobacco chloroplast DNA. The tRNA^{Gly} gene was found to have a long intron in the D stem.

MATERIALS AND METHODS

Recombinant plasmid pTB24, containing a 12.4-kilobase-pair (kbp) partial *Bam*HI fragment of *Nicotiana tabacum* (var. Bright Yellow 4) chloroplast DNA, was constructed previously (13). The plasmid DNA was digested with *Hin*dIII, and the 2.1-kbp fragment was separated from other *Hin*dIII fragments by electrophoresis in a 1% agarose gel. The DNA sequence was determined by the method of Maxam and Gilbert (14).

Total tobacco chloroplast RNA was prepared as described (15). RNA was electrophoresed and hybridized with the ^{32}P-labeled 559-bp *Hin*dIII/*Bam*HI, 391-bp *Bam*HI/*Rsa* I, and 436-bp *Rsa* I/*Rsa* I subfragments from the 2.1-kbp *Hin*dIII fragment as described (6).

The 5′ end of a 196-bp *Sau*3A subfragment from the 2.1-kbp fragment was labeled with ^{32}P. After strand separation by 5% polyacrylamide gel electrophoresis, the ^{32}P-labeled coding strand was hybridized with total tobacco chloroplast RNA (10 μg) and digested with S1 nuclease as described (6). The protected fragment was electrophoresed in a 12% polyacrylamide gel in parallel with the sequencing ladder of the coding strand.

RESULTS AND DISCUSSION

The DNA Sequence. We had previously cloned a 12.4-kbp partial *Bam*HI fragment of tobacco chloroplast DNA that codes for several tRNAs and the α subunit of proton-translocating ATPase (16). On digestion with *Hin*dIII, the recombinant plasmid pTB24 DNA yields 5.9 (containing the 4.0-kbp pBR322 sequence)-, 1.1-, 2.1-, 0.05-, 2.5-, 2.7-, and 2.4 (containing the 0.35-kbp pBR322 sequence)-kbp fragments (Fig. 1). Total tobacco chloroplast tRNA hybridized to the 5.9-, 1.1-, and 2.1-kbp fragments (data not shown). The 2.1-kbp *Hin*dIII fragment yields 0.56-, 0.17-, and 1.4-kbp subfragments on digestion with *Bam*HI, and the 1.4-kbp subfragment contains the 3′ part of the α-subunit gene (16). We isolated the 2.1-kbp *Hin*dIII fragment from pTB24 and determined the sequence of the region downstream from the α-subunit gene by the strategy shown in Fig. 1. Fig. 2 shows the nucleotide sequence of a 1,553-bp part (the left *Hin*dIII site to the second *Rsa* I site) of it.

The tRNA^{Arg} Gene. On digestion with *Fok* I, the 1.4-kbp *Bam*HI/*Hin*dIII fragment yields 1.0- and 0.4-kbp subfragments. Total tobacco chloroplast tRNA hybridized to both fragments (data not shown). A tRNA gene was found between positions 932 and 1,003 (strand A) and 126 bp apart from the α-subunit coding region (strand B) in an opposite orientation in the 1.0-kbp *Fok* I/*Hin*dIII subfragment (Fig. 2). Judging by the anticodon sequence (TCT), it is a tRNA^{Arg} gene (Fig. 3*b*). The chloroplast tRNA^{Arg} deduced from the DNA sequence shows 55%, 56%, and 56% sequence homologies with phage T4, yeast, and yeast mitochondrial tRNA^{Arg} (UCU)s, respectively (7).

The tRNA^{Gly} Gene. A sequence corresponding to a 3′ part of a tRNA gene was found between positions 715 and 762 and 169 bp apart from the tRNA^{Arg} gene in the same orientation in the 0.4-kbp *Bam*HI/*Fok* I subfragment. It seems to be a part of a gene for tRNA^{Gly}, judging by the anticodon sequence (TCC). However, the sequence immediately before position 715 was not able to form D and aminoacyl stems. We then searched for a sequence complementary to the aminoacyl stem (positions 755–761, T-A-C-C-C-G-C). We could find the complementary sequence (G-C-G-G-G-T-A) between positions 1 and 7 in the 0.56-kbp *Hin*dIII/*Bam*HI fragment (Fig. 2), and the sequence between positions 1 and 23 was able to form a cloverleaf structure with the above sequence (715–762) (Fig. 3*a*). These two sequences show 84% and 82% homologies with those of *E. coli* and *Bacillus subtilis* tRNA^{Gly} (UCC)s, respectively, as shown in Fig. 4. Based on the sequence homology, the two sequences are a gene for tRNA^{Gly} (UCC). The spacer between these two se-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: bp, base pair(s); kbp, kilobase pair(s).
*To whom reprint requests should be addressed.

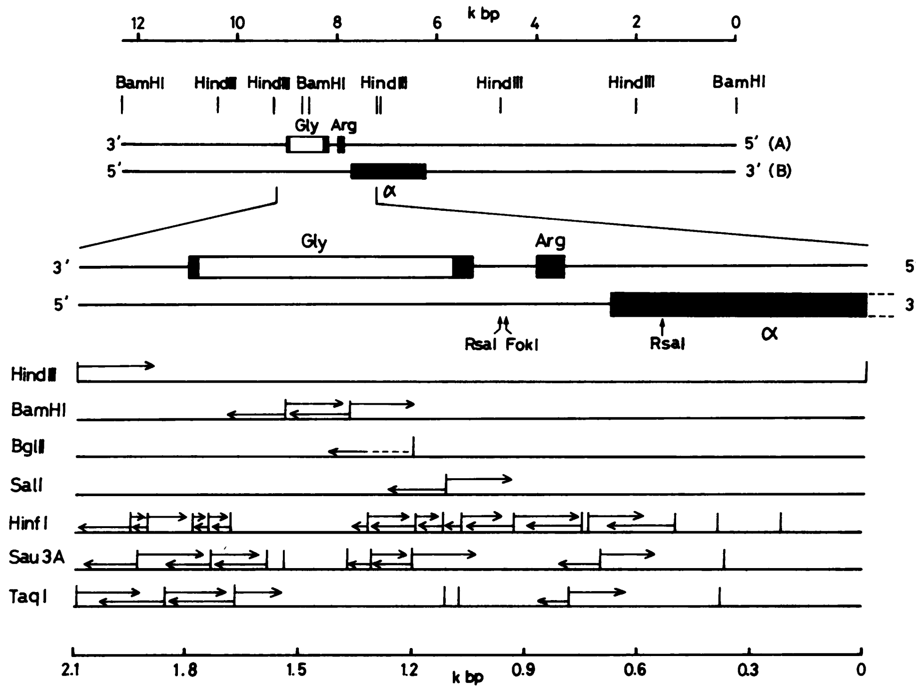

FIG. 1. Physical map of the cloned 12.4-kbp partial *Bam*HI fragment from tobacco chloroplast DNA and the strategy for determining the sequence of part of it. Strand A (*A*) codes for the tRNA^{Gly} and tRNA^{Arg} genes and strand B (*B*) codes for the α subunit of proton-translocating ATPase. Coding regions are shown by thick lines, and introns, by boxes. The lower part shows an expanded physical map of the 2.1-kbp *Hin*dIII fragment and the strategy for determining its sequence. Horizontal arrows indicate directions and extents of the DNA segments analyzed.

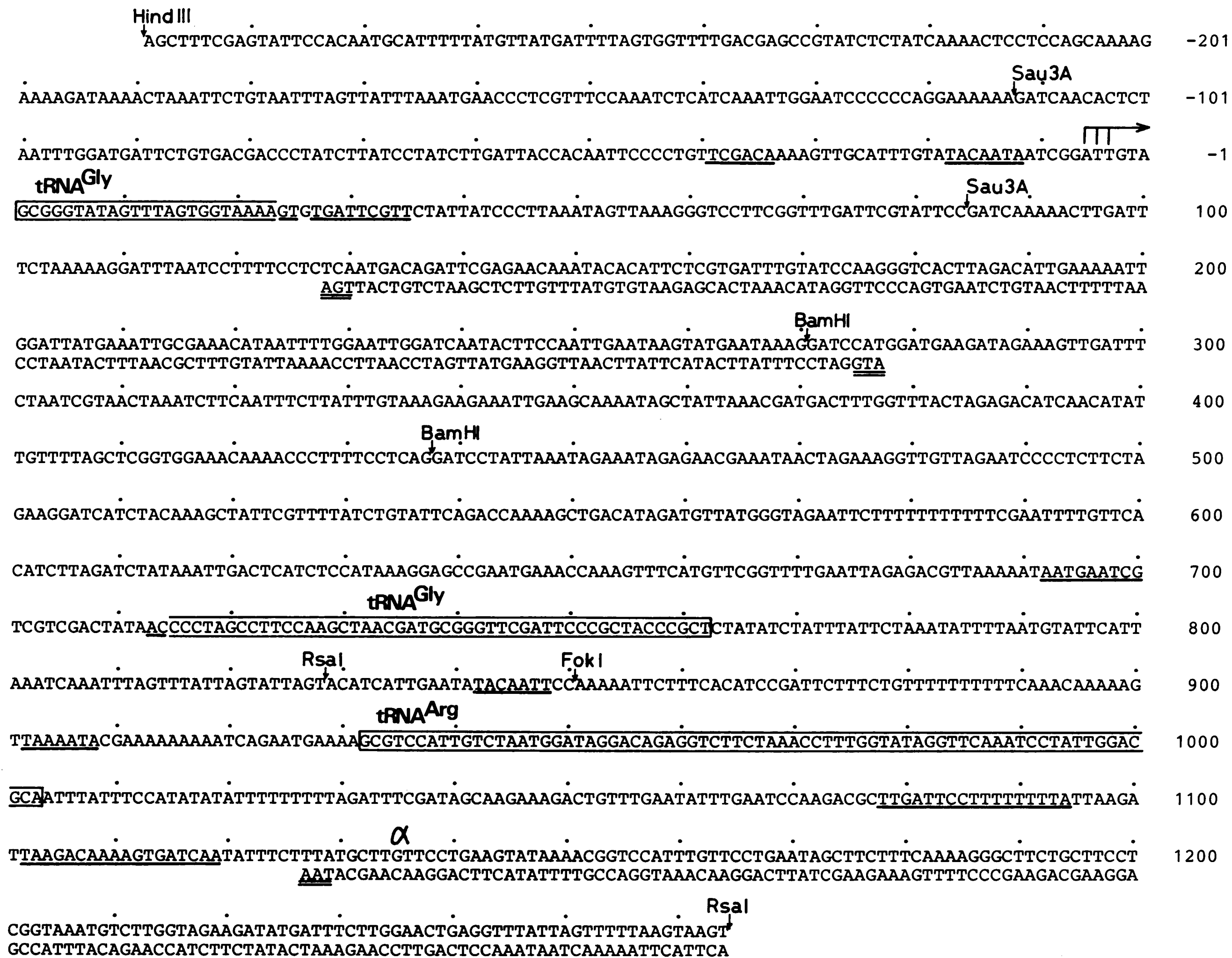

FIG. 2. DNA sequence of the 1,553-bp region containing the tRNA^{Gly} and tRNA^{Arg} genes. The tRNA-like strand (strand B) is presented. The tRNA sequences are boxed. Sequences of the RNA-like strand (strand A) for an open reading frame in the intron and the carboxyl-terminal portion of the α-subunit gene are also shown. Double-underlined sequences are start and stop codons. Sequences similar to the *E. coli* Pribnow box and −35 region and inverted repeat sequences are underlined. A horizontal arrow indicates possible transcription initiation sites.

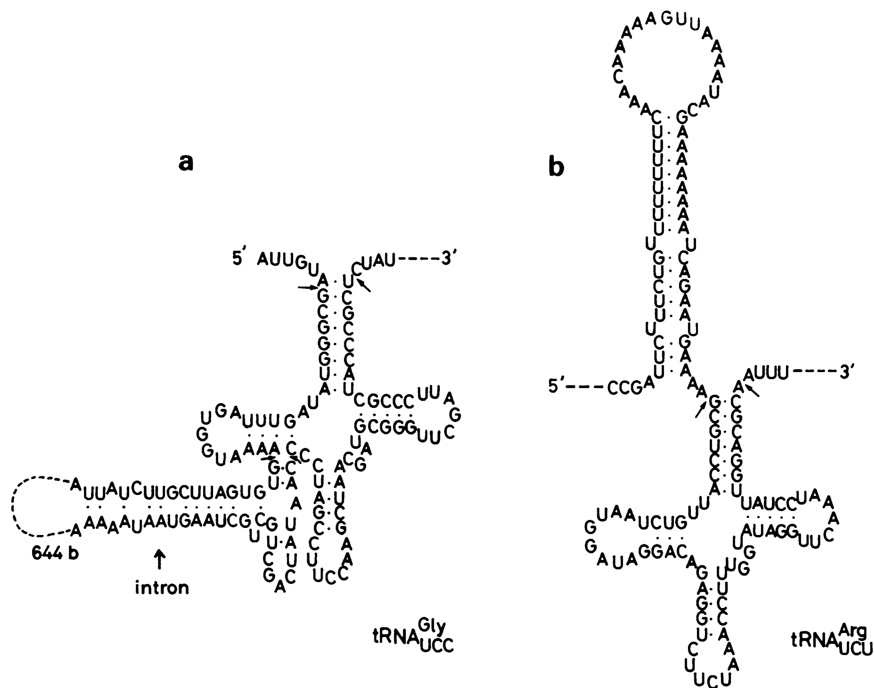

FIG. 3. Possible secondary structures of putative transcripts containing the $tRNA^{Gly}$ (*a*) and $tRNA^{Arg}$ (*b*) sequences. Arrows indicate possible processing and splicing sites.

quences should be an intron. From the sequence homologies with *E. coli* and *B. subtilis* $tRNA^{Gly}$ (UCC)s (Fig. 4), the intron is most likely located at position 24/25 in the D stem and is 691 bp long. This is the first intron found in the D stem of a tRNA gene. In both the $tRNA^{Arg}$ and $tRNA^{Gly}$ genes, the 3′ CCA sequences are not coded for by the chloroplast DNA.

Expression of the $tRNA^{Gly}$ Gene. It is important to examine whether the $tRNA^{Gly}$ gene containing a long intron in the D stem is expressed in the chloroplasts. Total tobacco chloroplast RNA extracted from young tobacco leaves was loaded on a 1.2% agarose gel. After electrophoresis, the RNA was transferred to nitrocellulose filter sheets and hybridized with the 559-bp *Hin*dIII/*Bam*HI fragment containing the 5′-half exon (positions −289 to 270) and the 391-bp *Bam*HI/*Rsa* I fragment containing the 3′-half exon (positions 438–828). These two DNA fragments hybridized to an RNA band of about 900 bases (Fig. 5, lanes a and b). Judging by its size, this RNA should be an unspliced precursor molecule for the $tRNA^{Gly}$. The DNA fragment containing the 3′-half exon hybridized to 4S RNA also, indicating that the precursor is processed to mature tRNA. The 4S RNA band is rather faint because of the low retention of small RNAs on nitrocellulose paper. No detectable hybridization was observed in a 4S RNA region when the DNA fragment containing 5′-half exon was used. This is probably because the 5′-half exon is too short (23 bases) and too A+T-rich (61%) to form a stable hybrid with the mature tRNA.

A 436-bp *Rsa* I/*Rsa* I fragment containing the $tRNA^{Arg}$ sequence (positions 829–1,264) hybridized to a 4S RNA band but not to a 900-base band (Fig. 5, lane c). This suggests that these two tRNA genes are transcribed separately. An RNA band of 23S size is mRNA for the α subunit of proton-translocating ATPase (unpublished data).

The position corresponding to the 5′ end of the precursor RNA was determined by nuclease S1 mapping. A scheme for the procedure is shown in Fig. 6*a*. The 5′-^{32}P coding strand (positions −108 to 88) of a 196-bp *Sau*3A fragment that contains the 5′-half exon was hybridized with total tobacco chloroplast RNA and treated with nuclease S1. The major S1-resistant DNA fragments were determined to be 5′ A-T-T-G-T . . . , 5′ T-T-G-T . . . , and 5′ T-G-T . . . , as shown in Fig. 6*b*. This indicates that the 5′ end of the precursor RNA is located at −6 to −4 (Fig. 2). The Pribnow box-like sequence T-A-C-A-A-T-A (positions −18 to −12) and the −35 region-like sequence T-C-G-A-C-A (positions −39 to −34) were found in the region upstream from the putative initiation sites. The 169-bp spacer between the two tRNA genes is highly A+T-rich (81%) so that transcription may terminate in this region (Fig. 2). Two Pribnow box-like sequences T-A-C-A-A-T-T (positions 842–848) and T-A-A-A-A-T-A (positions 902–908) were found in the region upstream from the $tRNA^{Arg}$ gene, and an inverted repeat sequence (positions 1,078–1,119) followed by a thymidine-rich cluster was found in the region downstream from the $tRNA^{Arg}$ gene and in the vicinity of the 3′ end of the α-subunit gene (Fig. 2). These structures may be transcriptional initiation and termination signals for the $tRNA^{Arg}$ gene.

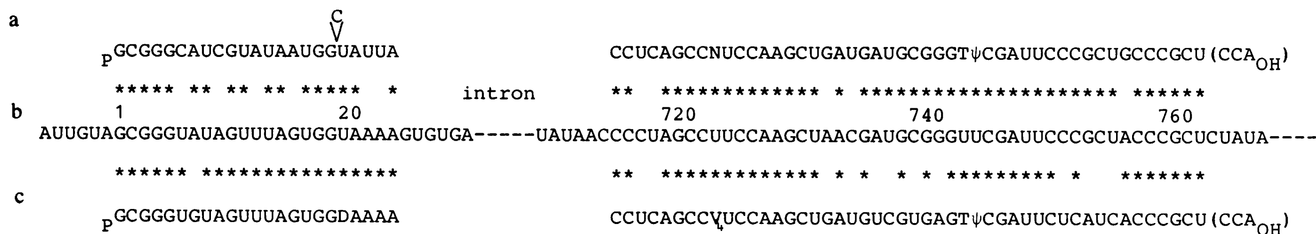

FIG. 4. Comparison of chloroplast $tRNA^{Gly}$ (UCC) (*b*) with *E. coli* (*a*) and *B. subtilis* (*c*) $tRNA^{Gly}$ (UCC)s. Sequences are aligned to give maximal homology. Asterisks indicate homologous nucleotides.

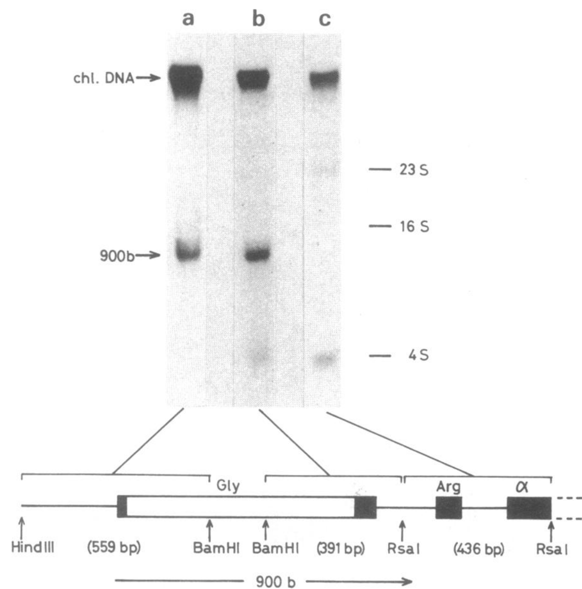

FIG. 5. Detection of precursor RNA molecules for $tRNA^{Gly}$ Autoradiographs of RNA blots of tobacco chloroplast RNA hybridized to the ^{32}P-labeled 559-bp *Hin*dIII/*Bam*HI fragment (lane a), 391-bp *Bam*HI/*Rsa* I fragment (lane b), and 436-bp *Rsa* I/*Rsa* I fragment (lane c). Size markers are *E. coli* 23S, 16S, and 4S RNAs.

Possible secondary structures of putative transcripts containing $tRNA^{Gly}$ (UCC) and $tRNA^{Arg}$ (UCU) sequences can be constructed as shown in Fig. 3. The 5′ end (G-T) of the intron in the $tRNA^{Gly}$ gene is complementary to the 3′ end (A-C) and a sequence near the 5′ end of the intron (T-G-A-T-T-C-G-T-T, positions 27–35) is nearly complementary to a sequence near the 3′ end (A-A-T-G-A-A-T-C-G, positions 692–700). Common sequences reported previously near the 5′ and 3′ ends of introns in tobacco chloroplast $tRNA^{Ile}$, $tRNA^{Ala}$, and $tRNA^{Val}$ (UAC) genes (6) were not found in the intron of the $tRNA^{Gly}$ gene. A short open reading frame (positions 277–131, 49 codons) was found in the opposite orientation in the intron. It is interesting to note that a tRNA-like structure can be constructed from a portion of phage T4 species I RNA and its apparent D loop contains a short insertion sequence (17).

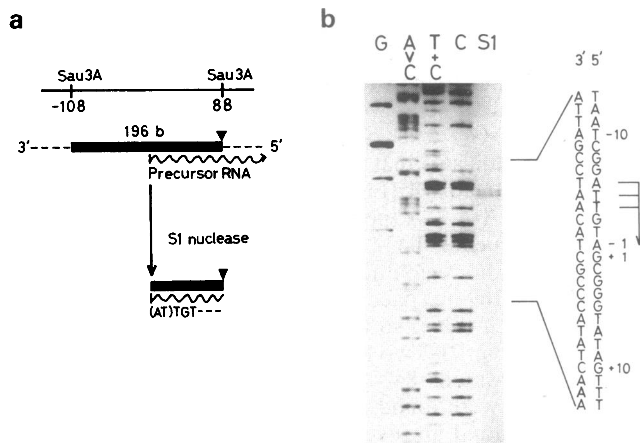

FIG. 6. Determination of the 5′ ends of precursor RNA molecules containing the $tRNA^{Gly}$ sequence by nuclease S1 mapping. (*a*) Schematic presentation of the procedure. A triangle indicates the 5′ end labeled with ^{32}P. (*b*) The nuclease S1-protected fragment (S1) was electrophoresed in parallel with sequence ladders of the coding strand (G, A>C, T+C, C).

From the codon usage in tobacco chloroplast genes for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase and the α and β subunits of proton-translocating ATPase, the $tRNA^{Gly}$ (UCC) and $tRNA^{Arg}$ (UCU) are frequently used tRNAs (16, 18, 19). The tobacco chloroplast $tRNA^{Val}$ (UAC) whose gene contains a 571-bp intron is also one of the frequently used tRNAs (6). Therefore, the presence of long introns seems to not interfere with effective expression of these tRNA genes, as indicated previously (6).

We thank Dr. K. Shinozaki for helpful suggestions and discussion. This work was supported in part by a grant-in-aid from the Ministry of Education, Science, and Culture, Japan.

1. Goodman, H. M., Olson, M. V. & Hall, B. D. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5453–5457.
2. Gauss, D. H. & Sprinzl, M. (1983) *Nucleic Acids Res.* **11,** r55–r103.
3. Koch, W., Edwards, K. & Kössel, H. (1981) *Cell* **25,** 203–213.
4. Takaiwa, F. & Sugiura, M. (1982) *Nucleic Acids Res.* **10,** 2665–2676.
5. Steinmetz, A., Gubbins, E. J. & Bogorad, L. (1982) *Nucleic Acids Res.* **10,** 3027–3037.
6. Deno, H., Kato, A., Shinozaki, K. & Sugiura, M. (1982) *Nucleic Acids Res.* **10,** 7511–7520.
7. Sprinzl, M. & Gauss, D. H. (1983) *Nucleic Acids Res.* **11,** r1–r53.
8. Tohdoh, N., Shinozaki, K. & Sugiura, M. (1981) *Nucleic Acids Res.* **9,** 5399–5406.
9. Kato, A., Shimada, H., Kusuda, M. & Sugiura, M. (1981) *Nucleic Acids Res.* **9,** 5601–5607.
10. Schwarz, Z., Kössel, H., Schwarz, E. & Bogorad, L. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 4748–4752.
11. Orozco, E. M., Jr., & Hallick, R. B. (1982) *J. Biol. Chem.* **257,** 3265–3275.
12. Ogden, R. C., Knapp, G., Peebles, C. L., Kang, H. S., Beckmann, J. S., Johnson, P. F., Fuhrman, S. A. & Abelson, J. N. (1980) in *Transfer RNA: Biological Aspects*, eds. Söll, D., Abelson, J. N. & Schimmel, P. R. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 173–190.
13. Sugiura, M. & Kusuda, J. (1979) *Mol. Gen. Genet.* **172,** 137–141.
14. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 560–564.
15. Shinozaki, K. & Sugiura, M. (1982) *Nucleic Acids Res.* **10,** 4923–4934.
16. Deno, H., Shinozaki, K. & Sugiura, M. (1983) *Nucleic Acids Res.* **11,** 2185–2191.
17. Paddock, G. & Abelson, J. (1973) *Nature (London) New Biol.* **246,** 2–6.
18. Shinozaki, K. & Sugiura, M. (1982) *Gene* **20,** 91–102.
19. Shinozaki, K., Deno, H., Kato, A. & Sugiura, M. (1983) *Gene* **24,** 147–155.