The structure and the location of *chlI* genes in algae

Akira Hamada

Doctor of Philosophy

Department of Genetics
School of Life Science
The Graduate University for Advanced Studies

1997

CONTENTS

Abstract

The *chlI* gene encoding a subunit of magnesium chelatase locates on the nuclear genome of several land plants, although it locates on the chloroplast genome of several algae. During the study of the chloroplast genome structure of the green alga *Chlorella vulgaris*, I found *chlI* on the chloroplast. The amino acid sequence deduced from *C. vulgaris chlI* consists of 354 residues. It showed greater similarity to those of land plants than to those of non-green algae, suggesting that the green alga *C. vulgaris* is more closely related to land plants than non-green algae. Northern hybridization revealed a single band at 1.85 kb which was much longer than that expected from the nucleotide sequence, suggesting that *chlI* would be co-transcribe d with the neighboring *trnR* gene. Primer extension revealed a single extension product at position -294 upstream of the translation start site. This matches perfectly with the -35 and -10 regions of the *E. coli* consensus promoter sequence. RNase protection assay revealed the 3' end of the transcripts which showed the existence of the co-transcripts of *chlI* and *trnR*. Although another riboprobe revealed the existence of the tRNA-Arg, the processed *chlI* mRNA could not be detected, suggesting that *trnR* contributed to the stabilization of the transcript, and that the *chlI* transcripts were immediately degraded after the processing of tRNA.

The chloroplast genome of *C. vulgaris* has ten genes not found in land plants. Many of these genes could have been transferred to the nucleus within the green algal lineage giving

rise to land plants. Considering the possibility of the transfer of *chlI* in *C. vulgaris*, the chloroplast and nuclear DNA were analyzed by Southern hybridization. However no *chlI*-like sequence was detected in the nuclear DNA.

Several cases of gene transfer from chloroplast to the nucleus have been known during the evolution of land plants from green algae. *chlI* will be one of such genes, because it was detected in the chloroplast genomes of algae but not in those of any land plants. Therefore, I tried to detect *chlI* from Charophyceae *Nitella* and *Chara*, which were believed to be direct ancestors of land plants. Amplification of *Nitella* chloroplast DNA revealed a *chlI* which showed the highest similarity with that of *Chlorella*. Southern hybridization using the chloroplast gene as a probe revealed the existence of a *chlI*-like structure in the nuclear DNA from *Nitella*. The hybridization pattern was completely different from that observed in the chloroplast DNA. Northern hybridization revealed a single band at 1.4 kb at the poly(A)- RNA, though the signal at the poly(A)+ RNA is under detectable in this experimental condition. It is suggesting that *Nitella chlI* is still transcribed from the chloroplast DNA. To confirm the existence of a *chlI*-like sequence, amplification of the sequence of nuclear DNA was attempted using a primer pair specific to the chloroplast *chlI*. However, the *chlI*-like sequences could not be amplified, suggesting that the sequence of nuclear *chlI*-like was quite different from that of chloroplast *chlI*. Acquisition of function by a relocated gene

requires the gain of compartment-specific regulatory sequences, upstream and downstream, and an amino-terminal transit peptide sequence. The *chlI*-like sequence of *Nitella* will be in this stage which is changing the structure to adapt in the nucleus. The chloroplast encoded *chlI* is well conserved and will retain its function until the nuclear *chlI* will acquire the function. Moreover, the peculiar nucleotide sequences of *chlI* were found from *Chara*. The amino acid sequences deduced from them were the same, though the nucleotide sequences were different at 11 sites, which were found at the third letter of the codons. In connection with these *chlI* structures, gene transfer from the AT-biased chloroplast genome to the GC-biased nuclear genome is discussed.

Introduction

Chloroplasts are intracellular organelles in plants which contain the entire machinery necessary for the process of photosynthesis. They contain their own genetic systems, and a number of chloroplast components are encoded in their genomes (Palmer, 1985; Sugiura, 1992). Most chloroplast genomes in land plants consist of homogeneous circular DNA molecules which range in size from 120 to 160 kbp and appear conserved in size, gene content and organization (Ohyama *et al.*, 1986; Shinozaki *et al.*, 1986; Hiratsuka *et al.*, 1989; Wakasugi *et al.*, 1994; Maier *et al.*, 1995). On the other hand, chloroplast (plastid) genomes of algae reveal remarkable differences in gene order and structure, compared with land plants (Hallick *et al.*, 1993; Kowallik *et al.*, 1995; Reith and Munholland, 1995; Stirewalt *et al.*, 1995; Wakasugi *et al.*, 1997). Some genes not found in chloroplast genomes of land plants are located in algal plastid genomes, even though the overall size resembles that of land plants.

The genetic system of the chloroplast has many features in common with prokaryotic organisms and is distinct from the (eukaryotic) nuclear-cytoplasmic system (Sugiura, 1992). For example, (1) typical bacterial-promoter sequences (-35 and -10 regions) exist upstream of the transcriptional start site, and (2) the potential stem and loop structure as the transcriptional terminator of prokaryotic RNA polymerase is located downstream of the coding region. (3) Genes are polycistronically transcribed and processed into shorter RNA species. (4) Transcription and translation are performed using the

prokaryotic-like components. These features have been invoked in support of the hypothesis that green-plant cells evolved from a symbiosis between a eukaryotic host cell and a photosynthetic prokaryote. Especially, the chloroplasts of green plants are thought to have evolved from cyanobacterial endosymbionts (Palmer, 1991; Gray, 1992).

The size of chloroplast genomes is very small compared with general prokaryotes. The reduction reflects that genes required for free-living existence were lost, most genes useful to the symbiosis were transferred to the nucleus of the host (Palmer, 1991; Gray, 1992). Nuclear-derived mRNAs are processed and translated on 80S ribosomes in the cytosol, and the precursor polypeptides are targeted to the chloroplast, imported, processed, and assembled into their functional forms. The remaining 10% or so of genes encoding chloroplast products have been retained within endosymbionts and now comprise plastid genomes. The conservation of gene content among chloroplasts of distantly related taxa suggests that most gene transfer occurred early in organelle evolution (Palmer, 1985). However, evidence of modern gene transfer has been accumulating (Ohyama *et al.*, 1988; Baldauf and Palmer, 1990; Gantt *et al.*, 1991), suggesting that the process continues, albeit at a greatly reduced rate.

The *chlI* gene encodes a subunit of magnesium chelatase, which catalyzes the insertion of magnesium into protoporphyrin IX. Three genes (*bchD*, *bchH* and *bchI*) have been shown to be involved in the synthesis of magnesium chelatase by genetic

analyses of pigment-deficient mutants of photosynthetic bacteria (Bollivar *et al.*, 1994; Gibson *et al.*, 1995; Willows *et al.*, 1996). However, only the *chlI* gene corresponding to *bchI* has been detected in chloroplast genomes of algae; therefore, it is thought that the others are encoded in the nuclear genome (Hudson *et al.*, 1993; Gibson *et al.*, 1996; Jensen *et al.*, 1996). The *chlI* was first discovered as the *cs* gene in the nuclear genome of *Arabidopsis thaliana* (Koncz *et al.*, 1990) and then in the chloroplast genome of *Euglena gracilis* as the *ccsA* gene (Orsat *et al.*, 1992). The *chlI* gene is now known to be encoded in the plastid genomes of *Cyanophora paradoxa* (Stirewalt *et al.*, 1995), *Odontella sinensis* (Kowallik *et al.*, 1995), *Porphyra purpurea* (Reith and Munholland, 1995) and *E. gracilis* (Orsat *et al.*, 1992; Hallick *et al.*, 1993). However, it has not been found in the chloroplast genomes of land plants, though it was found in the nuclear genomes of *A. thaliana* (Koncz *et al.*, 1990) and *Glycine max* (Nakayama *et al.*, 1995).

During the study of the structure and function of the chloroplast genome of the unicellular green alga *Chlorella vulgaris C-27* (Wakasugi *et al.*, 1997), I found the *chlI* gene in a recombinant plasmid, called pKH51, which was prepared from a chloroplast-rich fraction. This is the first finding of *chlI* in the chloroplast genome of a "real" green alga, though it has been found in that of the "nominally" green alga *E. gracilis* (Orsat *et al.*, 1992; Hallick *et al.*, 1993). Concerning the *chlI* of *C. vulgaris*, the following results are described in this

paper: (1) *chlI* transcripts were started from a typical bacterial promoter sequence; (2) the *chlI* transcript was extended beyond the stem and loop structure located downstream of the *chlI* coding region into the *rpl12* coding region which was encoded in the opposite strand; (3) the *chlI* transcripts contained tRNA-Arg, which contributed to the stabilization of the transcript; and (4) no *chlI*-like sequence was detected in the nuclear genome.

Concerning the gene transfer from chloroplast to nucleus, the *chlI* of green algae *Nitella* and *Chara* which belonged to Charophyceae has been studied and described as follows: (5) *Nitella chlI* located and functioned in the chloroplast genome; (6) the GC content of *Nitella chlI* in the third letter of the codons was the highest among those known algal *chlI* sequences; (7) a *chlI*-like sequence was found in the nuclear genome of *Nitella*; and (8) two types of *chlI* sequences were found in the total DNA from *Chara*.

Materials and methods

Plant materials

Chlorella vulgaris C-27 (formerly designated *Chlorella ellipsoidea C-27* Tamiya's strain), given by Prof. T. Iwamura, Nagoya University, was grown in M-4NA medium (Tamiya *et al.*, 1953) as described in Yoshinaga *et al.*, 1988.

Nitella sp. and *Chara sp.*, given by Prof. Y. Kakuno, Kobe University, were grown in water containing unsterilized soil at room temperature.

DNA Sequencing

A recombinant plasmid pKH51, which contains a part of the chloroplast genome from *Chlorella vulgaris C-27* (Wakasugi *et al.*, 1997), was digested with restriction enzymes. The fragments were subcloned using pUC18 vector. PCR products from *Nitella* and *Chara* DNA were subcloned using a pCR-Script™ Amp Cloning Kit (STRATAGENE) according to the manufacturer's protocol. Plasmid was prepared by boiling lysis (Sambrook *et al.*, 1989), and sequencing was carried out by the dideoxy chain-termination method (Sanger *et al.*, 1977) using 7-deaza Sequenase Ver. 2.0 (US Biochemical) or using a dRhodamine Terminator Cycle Sequencing Kit (PERKIN ELMER) according to the manufacturer's protocols. The resulting sequences were treated with GENETYX software (SDC Ver. 7.06, Tokyo).

Nucleic acid extraction from *C. vulgaris*

Nucleic acid was isolated using CTAB (cetyltrimethylammonium

bromide) (Gawel and Jarret, 1991). Cells of *C. vulgaris* growing logarithmically at 25 °C under continuous fluorescent light were suspended in CTAB buffer containing 100 mM Tris-HCl (pH 8.0), 20 mM EDTA, 1.4 M NaCl and 1% CTAB. The cell suspension was frozen in liquid nitrogen drop by drop and ground into powder using a Woring blender. After thawing, it was incubated at 65 °C for 30 min and treated with chloroform/isoamylalcohol (24/1).

RNA was precipitated from the aqueous phase by adding an equal volume of 4 M LiCl. The RNA was dissolved in a buffer containing 0.1 M CH_3COONa (pH 5.0) and 5 mM $MgSO_4$ and treated with RNase-free DNase I (Boehringer Mannheim) to digest contaminated DNA. After successive treatment with phenol /chloroform/isoamylalcohol (25/24/1) and chloroform /isoamylalcohol (24/1), RNA was precipitated with ethanol and dissolved in DEPC (diethylpyrocarbonate)-treated water.

Total DNA was precipitated from the aqueous phase by adding an equal volume of isopropanol and dissolved in TE buffer. Chloroplast and nuclear DNAs were separated by CsCl equilibrium centrifugation. Nucleic acid solution containing 1.2 g/ml CsCl and 0.16 mg/ml Hoechst 33258 was centrifuged for 16 h at 208000g in a TL-100 rotor (BECKMAN). Two bands containing chloroplast and nuclear DNAs were separated in the gradient. Each band was withdrawn and further purified by re-centrifugation. Hoechst was extracted with isopropanol and CsCl was removed by dialysis against TE. Each of the DNA fractions was precipitated with ethanol and dissolved in TE.

Nucleic acid extraction from *Nitella and Chara*

Nucleic acids of *Nitella* and *Chara* were prepared from frozen plants by disruption in a mortar and pestle and extracted in a CTAB buffer supplemented with 2% sarcosyl and 1% 2-mercapto ethanol. Total DNA, chloroplast DNA and total RNA of *Nitella* were prepared as described above.

In the case of *Nitella*, the nuclear DNA could not be separated clearly in the CsCl gradient. Therefore, the DNA was prepared from the nuclei. Frozen plants were disrupted and suspended in a buffer containing 50 mM Tris-HCl (pH 9.5), 0.4 M sucrose, 10 mM $MgCl_2$, 1 mM $CaCl_2$, 5 mM 2-mercaptoethanol and 2% Triton X-100. Chloroplasts and mitochondria were lysed in the solution containing Triton. After passing through four sheets of cheese-cloth, the nucleus-rich fraction was precipitated from the suspension by centrifugation at 2000g for 20 min at 4 °C, washed twice with the same buffer and suspended in CTAB buffer. DNA was extracted and purified as described above.

Poly(A)$^+$ and poly(A)$^-$ RNAs were separated from total RNA of *Nitella* by Oligotex-dT30<super> according to the manufacturer's protocol (Takara Shuzo, Kyoto, Japan).

Hybridization

DNA treated with restriction enzyme was separated by electrophoresis on an agarose gel and transferred to a nylon membrane (Hybond-N+, Amersham) by VacuGene (LKB). Southern hybridization was carried out with a *Gene Images* (Amersham). DNA

on the membrane was hybridized with appropriate probe overnight at 60 °C. The membrane was washed twice with 1x SSC containing 0.1% SDS for 15 min at 60 °C. The signals of hybridization were detected according to the manufacturer's protocol.

RNA was separated by electrophoresis on an agarose gel containing formaldehyde (Sambrook *et al.*, 1989) and transferred to a nylon membrane (Hybond-N+, Amersham) by VacuGene (LKB). Northern hybridization against *C. vulgaris* RNA was carried out by an ECL random prime labeling and detection system (Amersham). RNA on the membrane was hybridized with appropriate probe overnight at 60 °C. The membrane was successively washed with 1x SSC containing 0.1% SDS for 15 min at 60 °C and with 0.5x SSC containing 0.1% SDS for 15 min at 60 °C. The signals of hybridization were detected according to the manufacturer's protocol. Hybridization of poly(A)$^+$ and poly(A)$^-$ RNAs from *Nitella* was carried out by an AlkPhos DIRECT (Amersham). RNA on the membrane was hybridized overnight at 55 °C and the signals were detected according to the manufacturer's protocol.

Primer extension analysis

A primer of P-CHLI d(CTGGGCGTGCTTGTTCTAATGAGTT), complementary to the *C. vulgaris chlI* transcript at the nucleotide position of 19 to 43 from the translation start site (Figure 2), was end-labeled with γ-^{32}P-ATP (148 TBq/mmol, Bresatec) using T4 polynucleotide kinase (Nippon Gene). The

labeled primer of 1.5 pmol was annealed with 100 μg RNA by heating to 65 °C for 5 min and chilling on ice. The primer was extended using 200 U of M-MLV reverse transcriptase and 1.25 mM each of four dNTPs in a buffer containing 50 mM Tris-HCl (pH 8.3), 75 mM KCl and 3 mM $MgCl_2$, in the presence of RNase inhibitor and Actinomycin D at 42 °C for 90 min. The reaction was terminated by adding 1 μl of 0.5 M EDTA, and the material was then treated with 1 μl of RNase (1 mg/ml) at 37 °C for 30 min. After the treatment with phenol/chloroform (1/1), the product was precipitated with ethanol, dissolved in TE buffer containing 10 mM Tris-HCl (pH 8.0) and 1 mM EDTA, and analyzed on a standard sequencing gel with DNA sequence ladders generated by the same primer.

RNase protection assay

A 1262 bp *Kpn*I-*Hin*dIII fragment from the insert of the recombinant plasmid pKH51 was subcloned using pBluescript (STRATAGENE). The resultant plasmid digested with restriction enzyme *Sac*I was used as a template to synthesize a riboprobe. The reaction containing 40 mM Tris-HCl (pH 8.0), 20 mM $MgCl_2$, 5 mM DTT, 0.5 mM each of ATP, CTP, GTP, 25 μM UTP, 50 μg/ml BSA, 40 U of RNase inhibitor, an adequate amount of α-^{32}P-UTP (110 TBq/mmol, Bresatec), and 65 U of T7 RNA polymerase was incubated at 37 °C for 2 h. The labeled riboprobe was 370 bp in size and contained the 3'-end of *trnR(UCU)* (Figure 6B). After digestion

of template DNA with RNase-free DNase I (Boehringer Mannheim), the riboprobe was purified with phenol/chloroform/isoamylalcohol (25/24/1), precipitated with ethanol, denatured by heating to 85 °C for 5 min, and immediately hybridized with 100 μg of total RNA by incubating at 43 °C for 12 h. Single-stranded (non-hybridized) RNA was digested with RNase A and T1 at 30 °C for 2 h, then treated with a protanase (Actinase E, Kaken Seiyaku, Tokyo). RNAs protected from RNase were purified with phenol /chloroform/isoamylalcohol (25/24/1) and analyzed on a standard sequencing gel. The length of the RNAs was determined by comparison with the radiolabeled product of the dideoxy sequencing reaction.

Amplification of *Nitella* and *Chara chlI*

Amplification reaction was performed on a PROGRAM TEMP CONTROL SYSTEM PC-700 (ASTEC) using Taq DNA polymerase (Pharmacia). The oligonucleotide primers were designed from the consensus sequences of various plant *chlI*. The locations and orientations of the primers are shown in Figure 2. A primer pair of 5-1, d(CTTTTACTGCCATCGTTGGTCAAGA) and 3-1, d (TTTCTCGGACTTTATCACCAGAATC) locating on the nucleotide positions 50-74 and 1012-1036 in the *chlI* of *C. vulgaris*, respectively, was used to amplify the inside of *chlI*. Amplification reaction of *Nitella chlI* was carried out in a total volume of 50 μl containing 10 mM Tris-HCl (pH 9.0), 50 mM KCl, 1.5 mM $MgCl_2$, 400 nM each of the primers, 0.25 mM each of the dNTPs, 2.5 U Taq DNA

polymerase, and 0.2 μg of *Nitella* total DNA. Amplification of *Chara chlI* was performed as described above using 100 nM each of the primers and 0.2 μg of *Chara* total DNA. After denaturation at 94 °C for 1 min, 10 cycles of amplification consisting of 94 °C for 0.5 min, 40 °C for 2 min and 72 °C for 3 min and then 25 cycles consisting of 94 °C for 0.5 min, 60 °C for 1.5 min and 72 °C for 3 min were performed. The final extension was for 8 min at 72 °C.

Inverted PCR was carried out to amplify toward the outside of *chlI*. *Nitella* total DNA digested with *Eco*RI was self-ligated and used as a template. Amplification was performed as described above using 400 nM each of primers: 5-2, d (CTGTTCCTCGATCCCCCATAATCAT) and 3-2, d (GAGGAGATATTGTAACAAATCGAGC) located at 130-154 and 872-896 in the *chlI* of *C. vulgaris*, respectively. After denaturation at 94 °C for 1 min, 35 cycles of amplification consisting of 94 °C for 0.5 min, 55 °C for 1.5 min, and 72 °C for 5 min were performed. The final extension was for 10 min at 72 °C.

The *chlI* of *Chlorella vulgaris* is co-transcribed with the *trnR*

Results

Location of *Chlorella chlI* and amino acid sequence

The complete nucleotide sequence of pKH51 shows that *chlI* is located about 10 kbp downstream of *rbcL* on the same strand (Figure 1). Upstream of *chlI*, *petA* coding for cytochrome f apoprotein, *atpB* coding for the β subunit of ATPase, and *atpE* coding for the ε subunit of ATPase were found in this order on the opposite strand. Downstream, *trnR(UCU)* coding for an arginine tRNA and *rpl12* coding for ribosomal protein L12 of the 50S ribosomal subunit were found on the same and on the opposite strand, respectively. Among these genes, neither *chlI* nor *rpl12* has been found in plastid genomes of land plants. The chloroplast *chlI* of *C. vulgaris* consists of 1062 bp (Figure 2). In the flanking regions, four stem and loop structures can be seen. One located on the upstream forms an 18-bp stem and a 9-base loop (positions from -82 to -38). The others located downstream form a 34-bp stem and a 4-base loop (1116 to 1187), a 19-bp stem and a 4-base loop (1328 to 1369), and a 21-bp stem and a 4-base loop (1371 to 1416). These flanking regions are quite rich in AT (21 %) showing a typical chloroplast DNA.

chlI homologues have been found in the chloroplast genomes of *C. paradoxa* (Stirewalt *et al.*, 1995), *O. sinensis* (Kowallik *et al.*, 1995) and *P. purpurea* (Reith and Munholland, 1995) (Figure 3). The *chlI* of *C. vulgaris* has no intron though the *cs* gene of *A. thaliana*, and the *ccsA* gene of *E. gracilis* has two and one intron, respectively (Koncz *et al.*, 1990; Orsat *et al.*,

1992). The intron of the *cs* gene locates in the region coding the transit peptide targeting the chloroplast. The amino acid sequence predicted from the *chlI* of *C. vulgaris* was aligned with those of homologues (Figure 3). The amino acid sequence deduced from *C. vulgaris chlI* consists of 354 residues which is almost the same length as those of the other algae, but shorter than those of land plants, because the gene products of *A. thaliana* and *G. max* have an additional 72 and 70 amino acids at their N-terminal, respectively. This reflects that the *chlI* genes of land plants moved from the chloroplast to the nuclear genome and acquired the transit peptide targeting the chloroplast. However, the *chlI* of *C. vulgaris* showed greater similarity to those of land plants than to those of non-green algae, suggesting that the green alga is more closely related to land plants than non-green algae.

Four highly conserved regions of I to IV in *chlI* (Douglas and Reith, 1993) are also found in the gene of *C. vulgaris* (Figure 3). In region I, ten consecutive amino acids (118-127) showed a significant similarity to the consensus sequence (AGXXXXGKST) of the ATP/GTP-binding site motif A (P-loop) of ATP-binding proteins (Douglas and Reith, 1993). This would be the ATP-binding site because magnesium chelatase requires ATP (Gibson *et al.*, 1995; Willows *et al.*, 1996). The tertiary structure of the *chlI* gene product is not known, but other important functions could locate in these conserved domains.

The *chlI* transcript of *C. vulgaris* is stabilized by the downstream tRNA-Arg

To investigate the expression of the *chlI*, Northern hybridization was carried out using a *chlI*-specific probe. A single band at 1.85 kb was detected in the total RNA (Figure 4). The size is larger than that predicted from the coding region of *chlI*, suggesting that *chlI* is co-transcribed with the neighboring *trnR*.

To determine the 5'-end of the transcripts, primer extension analysis was carried out using a primer P-CHLI, and a single extension product was found at position -294 upstream of the translation start site (Figure 5). This matches perfectly with the -35 and -10 regions of the *E. coli* consensus promoter sequence (Figure 2).

It was expected that the 1.85 kb transcript starting from -294 would terminate at the coding region of *rpl12* (Figure 2). To determine the precise 3'-end of the transcript, RNase protection assay was carried out. The downstream region of the *trnR* containing *rpl12* was used to prepare the antisense riboprobe (Figure 6B). The riboprobe protected from RNase gave three signals (Figure 6A). These signals, of 220, 250 and 290 bases, correspond to the transcripts of 1820, 1850 and 1890 bases, respectively. This result indicates that *chlI* and *trnR* are co-transcribed and that the transcript extends into the *rpl12* coding region. To confirm the existence of processing between *chlI* and *trnR*, another riboprobe of the spacer region was used

in the protection assay, but no signal was detected. This is consistent with the result of Northern hybridization which showed no signal around 1.5 kb corresponding to the processed *chlI* mRNA. However, the *trnR* probe clearly showed the existence of the tRNA (data not shown), suggesting that the *chlI* mRNA, once spliced, was immediately broken down.

Discussion

Most transcripts of chloroplast contain the possible stem and loop structure at the 3' end, which was originally thought to function as the terminator of transcription. However, it was reported that the structure was a stabilizer but not a terminator (Stern and Gruissem, 1987; Adams and Stern, 1990; Stern *et al.*, 1991; Stern and Kindle, 1993). Chloroplast *chlI* of *C.vulgaris* contained three possible stem and loop structures downstream of the gene which were expected to stabilize the transcript (Figure 2). However, no such transcript could be detected; instead a co-transcript of *chlI-trnR* and mature tRNA-Arg were detected. This result shows that the *chlI* transcript is immediately broken down after the processing of the tRNA. It is also reported that a tRNA functions as a stabilizer of chloroplast mRNA (Stern and Gruissem, 1987). Actually, several chloroplast mRNAs are polycistronically transcribed and tRNA locates in the 3' end (Shinozaki *et al.*, 1986). It was reported that some RNA-binding proteins stabilized mRNA by interacting with the stem and loop structure (Chen and Stern, 1991; Schuster and Gruissem, 1991; Chen *et al.*, 1995; Hayes *et al.*, 1996). In the case of the chloroplast *chlI* transcript of *C. vulgaris*, such proteins interacting with the stem and loop structure would be dissociated when the tRNA was processed.

The location of *chlI* in algae

Results

chlI-like sequence is not detected from *Chlorella* nuclear DNA

To detect *chlI*-like sequences in the nuclear genome, the chloroplast and nuclear DNA of *C. vulgaris* were partially purified from total DNA by CsCl-Hoechst centrifugation and analyzed by Southern hybridization. A *chlI* probe hybridized with restriction fragments whose size was expected from the chloroplast genome structure. Faint hybridization bands were also detected in nuclear DNA preparation at the same positions corresponding to chloroplast DNA (Figure 7). DNA was amplified from the nuclear and chloroplast DNA preparations using a primer pair specific for *chlI*, which gave the same sized fragments, and their nucleotide sequences were revealed to be the same (data not shown). Therefore, the amplified DNA from nuclear preparation would be derived from contaminated chloroplast DNA, and a *chlI*-like sequence will not exist in the nucleus of *C. vulgaris*.

chlI gene in the chloroplast of *Nitella*

Nitella is a green algae Charaphytes which is believed to be the most closely related taxon to land plants. For example, it is shown in the phylogenetic tree constructed from 18S rRNA gene sequences of various plants (Kranz *et al.*, 1995). In addition, Baldauf *et al.* (1990) have shown the gene transfer of *tufA* coding for elongation factor EF-Tu from the chloroplast to the

nucleus in Charaphytes. Therefore, it was expected that *chlI* would also be transferred to the nucleus within Charaphytes, because *chlI* as well as *tufA* has been found in the chloroplast of Chlorophytes *Chlorella* but not in that of land plants.

Nitella chloroplast DNA was purified by CsCl-Hoechst centrifugation from the total DNA and analyzed by Southern hybridization using a *chlI* probe which was amplified from total DNA of *Nitella*. Chloroplast DNA gave signals of 4.1 kb in a *Bgl*II fragment, 1.7 kb in an *Eco*RI fragment and 4.3 and 1.7 kb in *Hin*dIII fragments (Figure 8). Other signals found in the *Bgl*II-digest (lane 1) will be artifacts because no *Bgl*II site found in the region determined the nucleotide sequence. The nucleotide sequence showed that the 1683 bp of the *Eco*RI fragment which contained the putative *chlI* consisted of 1071 bp and a part of the putative *rpoC2* (Figure 9). The *rpoC2* coding for the β' subunit of DNA-dependent RNA polymerase is a typical chloroplast gene, and the flanking regions of the *chlI* are quite rich in AT (28 %).

The predicted amino acid sequence from *Nitella chlI* consisted of 357 amino acids. The sequence alignment with those from other homologous genes shows that the sequence of *Nitella* is almost the same length as those of the other algae and shows the greatest similarity to that of *C. vulgaris* (Figure 3). Four highly conserved regions of I to IV in *chlI* (Douglas and Reith, 1993) are also found in the *Nitella* homologue.

Northern hybridization was attempted using *chlI* probe. A

single band at 1.4 kb was detected in the poly(A)- RNA fraction, suggesting that *Nitella chlI* is transcribed from the chloroplast DNA (Figure 10).

chlI-like sequence in the nucleus of *Nitella*

Nuclear DNA was prepared from the isolated nuclei and Southern hybridization was carried out using *chlI* probe from *Nitella* chloroplast. Signals were detected in the *Bgl*II fragment of 4.6 kb, in the *Eco*RI fragment of 3.5 kb and in the *Hin*dIII fragment of 4.3 kb (Figure 8), suggesting the existence of *chlI*-like sequence in the nuclear DNA.

To confirm the existence of a *chlI*-like sequence in the nuclear genome, amplification of the *chlI*-like sequence of nuclear DNA was attempted using a specific primer pair for chloroplast *chlI*. However, no product was detected, suggesting that the sequence of the nuclear *chlI*-like was different from that of the chloroplast *chlI*, in at least one site of primer recognition. Northern hybridization revealed no detectable signals in poly(A)$^+$ RNA, suggesting that nuclear *chlI*-like sequence will not acquire the function (Figure 10).

Discussion

The gene content and the sequence of many genes in chloroplast DNA are relatively conserved among land plants and the Euglenophyta *E. gracilis* (Palmer, 1991; Sugiura, 1992, 1996). However, recent analyses of the entire plastid genome from non-green algae (Kowallik *et al.*, 1995; Reith and Munholland, 1995; Stirewalt *et al.*, 1995) have revealed that this is not always the case. For example, the plastid genome of the red alga *P. purpurea* contains over 70 new genes not found in the chloroplast of land plant and *Euglena* (Reith and Munholland, 1993, 1995). The chloroplast genome of *C. vulgaris* has ten genes not found in land plants (Wakasugi *et al.*, 1997). Many of these genes could have been transferred to the nucleus during the evolution of land plants from green algae. *chlI* is one of such genes, because it was detected in the chloroplast genomes of algae but not in those of any land plants. It is generally believed that land plants evolved from green algae (Graham, 1996) and that during this evolution, extensive rearrangements occurred within the chloroplast genomes. At the same time, gene transfer to the nucleus would occur.

Several cases of gene transfer from chloroplast to the nucleus have been reported: the *tufA* was transferred within the green algal lineage giving rise to land plants (Baldauf and Palmer, 1990; Baldauf *et al.*, 1990) and rather recently the *rpl22* coding ribosomal protein L22 of the 50S ribosomal subunit within the legumes (Gantt *et al.*, 1991). Baldauf *et al.* (1990)

have shown several hybridization signals for *tufA* in the nuclear genome of Charaphytes and proposed that a copy of the *tufA* gene was transferred from the chloroplast to the nucleus early in the evolution of the Charophyceae. Several mitochondrial genes were also transferred into the nucleus (Nugent and Palmer, 1991; Grohmann *et al.*, 1992; Wischmann and Schuster, 1995; Sánchez *et al.*, 1996). Especially, soybean *coxII* is encoded in the both genomes, in which both genes have the perfect coding sequences, but is transcribed from the only nuclear genome (Covello and Gray, 1992). On the other hand, rice *rps11* is transcribed from the nuclear and mitochondrial genome, but mitochondrial gene is pseudogene which has termination codon within the coding sequence (Kadowaki *et al.*, 1996). All of these results show that the real gene locates only one organella of mitochondria or nucleus.

chlI probe hybridized with the both chloroplast and nuclear genomes of the Charaphytes *Nitella*. However, *chlI* gene is transcribed only from the chloroplast genome (Figure 10). Acquisition of function by a relocated gene into nucleus requires the gain of compartment-specific regulatory sequences, upstream and downstream, and an amino-terminal transit peptide sequence. The *chlI* gene in *Nitella* will be in this stage which is changing the structure to adapt in the nucleus. The chloroplast encoded *chlI* is well conserved and transcribed, which will retain its function until the nuclear *chlI* will acquire the function.

In general, the GC content of the chloroplast genome is lower than that of the nuclear genome, especially in algae. The mutation pressure operating to alter the GC content is well reflected not only on the non-coding region but also on the third letter of the codons. The change at the third letter does not affect the amino acid sequence, except for a few substitutions between purine and pyrimidine. Genes relocated in the nucleus drastically change the codon usage. For example, rye *rpl12*, which was originally in the chloroplast genome of algae, migrated into the nuclear genome and changed the third letter of the codons to an extreme high GC content (over 90%) (Schmidt *et al.*, 1993). Furthermore, the transferred genes adjusted their base composition and codon usage, becoming more similar to the nuclear genes than to the chloroplast ones (Oliver *et al.*, 1990). It is reasonable to imagine that the *chlI*-like in the *Nitella* nucleus will be changed toward G or C from A or T. This would cause failure of the amplification of the *chlI*-like using a primer pair specific for chloroplast *chlI*. Table 1 shows the GC content of *tufA* and *chlI* and that in the third letter of the codons. *chlI* and *tufA* in *Nitella* show the highest GC content among the same genes of algae, though the nuclear genes of land plants show much higher GC content than those of *Nitella*. The high GC content of *Nitella* would be easier to adapt to the nuclear environment than the other algae examined.

Peculiar sequences (CC1 and CC4) of *chlI* were amplified from the total DNA of Charaphytes *Chara* using a primer pair specific

to the *chlI* of *C. vulgaris* (Figure 11). Amino acid sequences deduced from both nucleotide sequences were completely the same. However, the nucleotide sequences were different at 11 sites, all of which were located at the third letter of the codons. The GC contents of their third positions were 24.0% and 24.9%, suggesting that the former would locate in the chloroplast and the latter in the nucleus. Alternatively, both of them were in the chloroplast and the latter was waiting to move to the nucleus. Further studies will be required to clarify this fact.

Figures and table

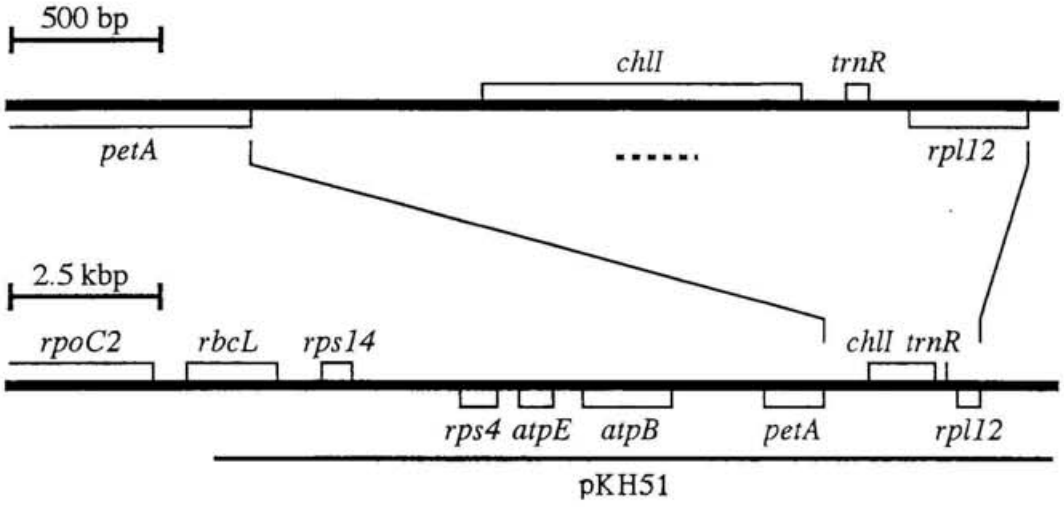


Figure 1. Chloroplast gene arrangement of *C. vulgaris*. Genes represented by open boxes; those above the bold line are transcribed from left to right and those below the bold line from right to left. The probe used for hybridization is indicated by the dotted line.

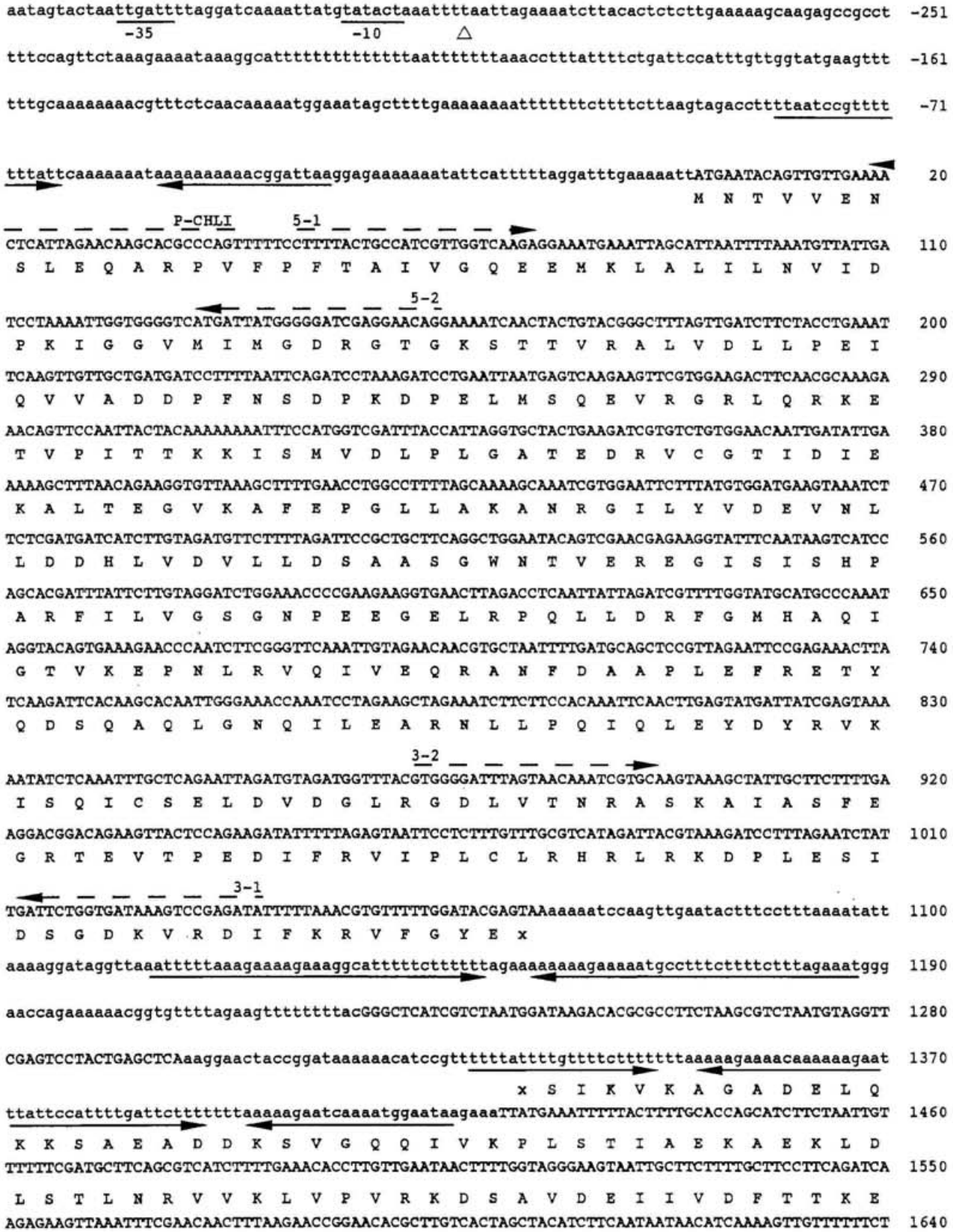

-35
-10
△
P-CHLI
5-1
5-2
3-2
3-1

Figure 2. Nucleotide sequence containing *chlI* and *trnR* of *C. vulgaris C-27*. The numbering of nucleotides shown on the right is from the translation start site of *chlI*. Genes corresponding to *chlI* (1-1065), *trnR* (1227-1298) and part of *rpl12* (1421-1640) are indicated by capital letters. The amino acid sequence deduced from this sequence is shown below by single letters, and those from the complementary sequence are shown above. Inverted repeats are indicated by horizontal arrows. The primers used for primer extension (P-CHLI), PCR (5-1 and 3-1) and inverted PCR (5-2 and 3-2) are indicated by dotted horizontal arrows. The transcription start site of *chlI* is indicated by a triangle.

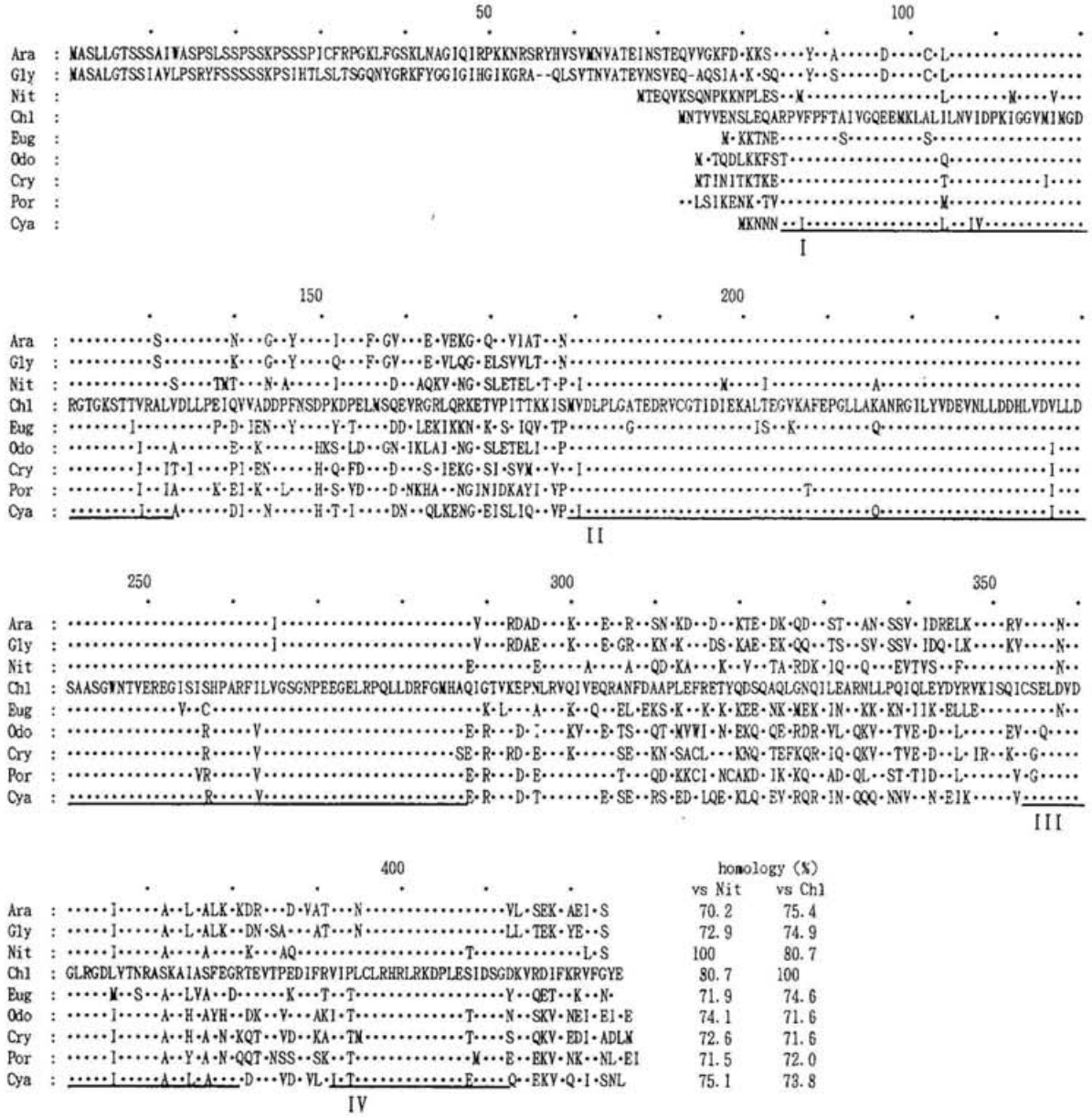

50
100
150
200
250
300
350
400
I
II
III
IV
homology (%)
vs Nit vs Chl
Ara 70.2 75.4
Gly 72.9 74.9
Nit 100 80.7
Chl 80.7 100
Eug 71.9 74.6
Odo 74.1 71.6
Cry 72.6 71.6
Por 71.5 72.0
Cya 75.1 73.8

Figure 3. Amino acid sequence alignments of *chlI* homologues. Amino acid sequences of *Arabidopsis thaliana* (Ara) (Koncz *et al.*, 1990), *Glycine max* (Gly) (Nakayama *et al.*, 1995), *Nitella sp.* (Nit), *Euglena gracilis* (Eug) (Hallick *et al.*, 1993), *Odontella sinensis* (Odo) (Kowallik *et al.*, 1995), *Cryptomonas phi* (Cry) (Douglas and Reith, 1993), *Porphyra purpurea* (Por) (Reith and Munholland, 1995), and *Cyanophora paradoxa* (Cya) (Stirewalt *et al.*, 1995) are compared with that of *C. vulgaris* (Chl). The homology (%) of the amino acid sequence with those of *Nitella* and *C. vulgaris* is shown at the end of each sequence. Four highly conserved regions, designated as I, II, III and IV, are underlined. Amino acids differing from *C. vulgaris* are shown, while those identical to *C. vulgaris* are represented by dots. Dashes were introduced to improve homology. The numbering of amino acids is denoted by the sequence of *A. thaliana*.

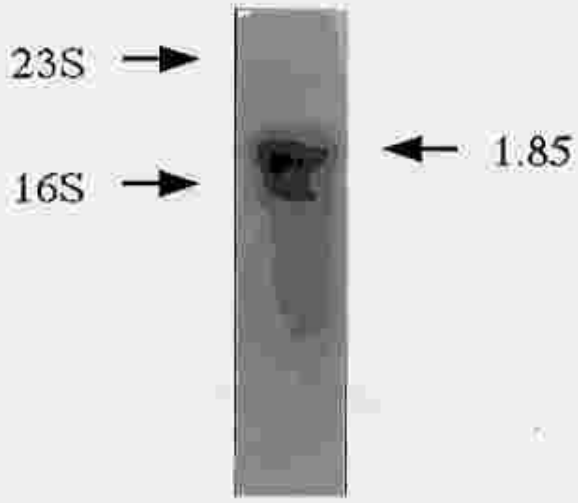


Figure 4. Northern blot analysis of the *chlI* transcripts of *C. vulgaris*. Total RNA of *C. vulgaris* was hybridized with a 0.3 kb fragment of the *chlI*. The hybridized band is indicated by a horizontal arrow and the size by kb calculated from *E. coli* 23S and 16S ribosomal RNA.

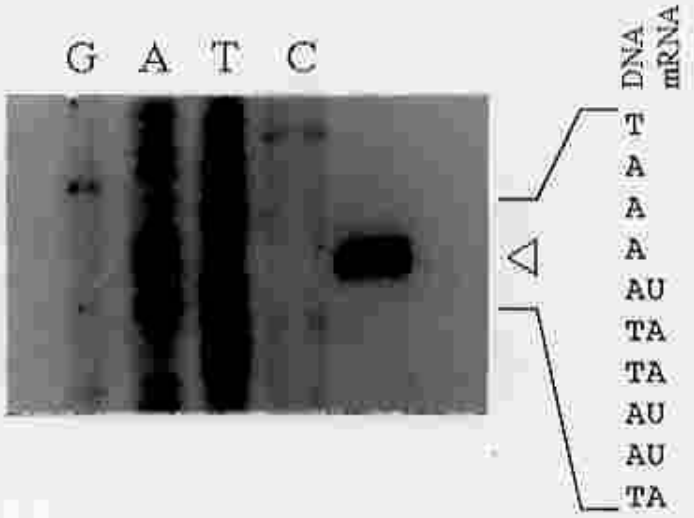
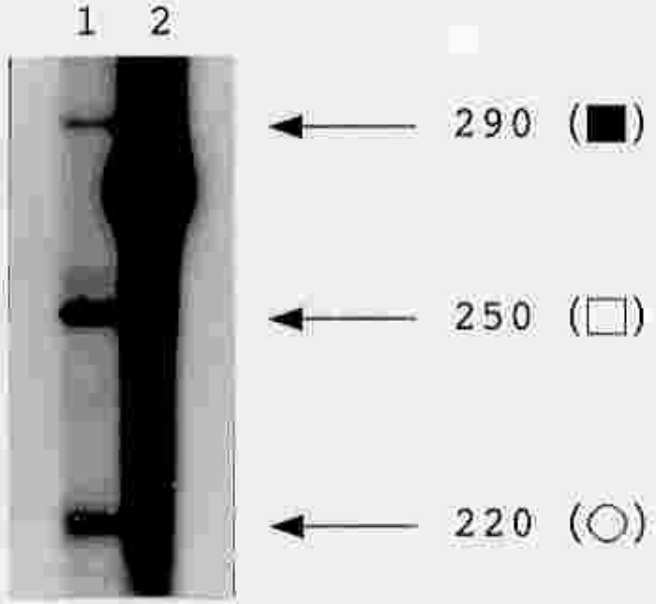


Figure 5. Primer extension analysis of the *chlI* transcripts of *C. vulgaris*. The extension product was analyzed on a sequencing gel using a dideoxy sequencing product from the same primer. The extension product is indicated by a triangle. Nucleotide sequences of template DNA and 5' part of mRNA are shown on the right side.

A


1 2
290 (■)
250 (□)
220 (○)

B

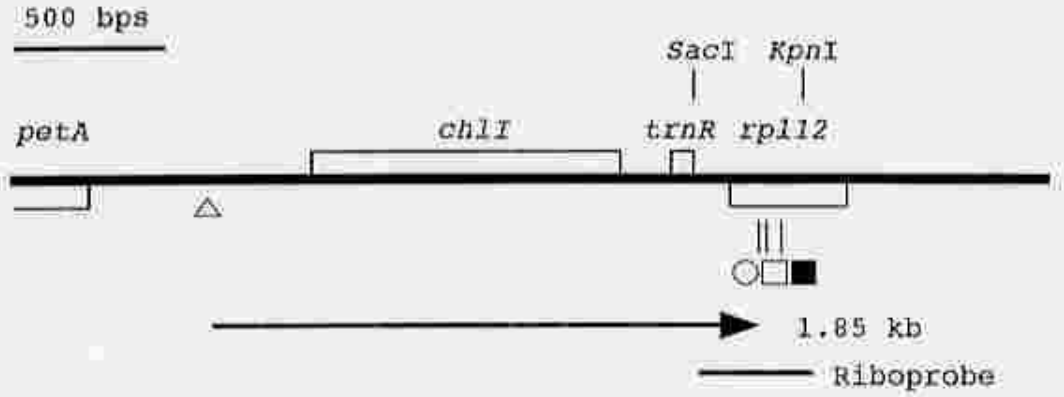

500 bps
SacI KpnI
petA
chlI
trnR rpl12
1.85 kb
Riboprobe

Figure 6. RNase protection assay to determine the 3'-termini of the *chlI* transcripts of *C. vulgaris*. (A) Total RNA from *C. vulgaris* was annealed with a labeled riboprobe, digested with RNase, and analyzed on a sequencing gel. The protected bands are indicated by horizontal arrows and the sizes by nt (lane 1). The riboprobe without RNase treatment is in lane 2. (B) Gene arrangement and *chlI* transcripts. The position of a riboprobe is indicated by a horizontal line, and the transcripts from the 5'-end (△) to 3'-end (○, □, ■) are indicated by a horizontal arrow.

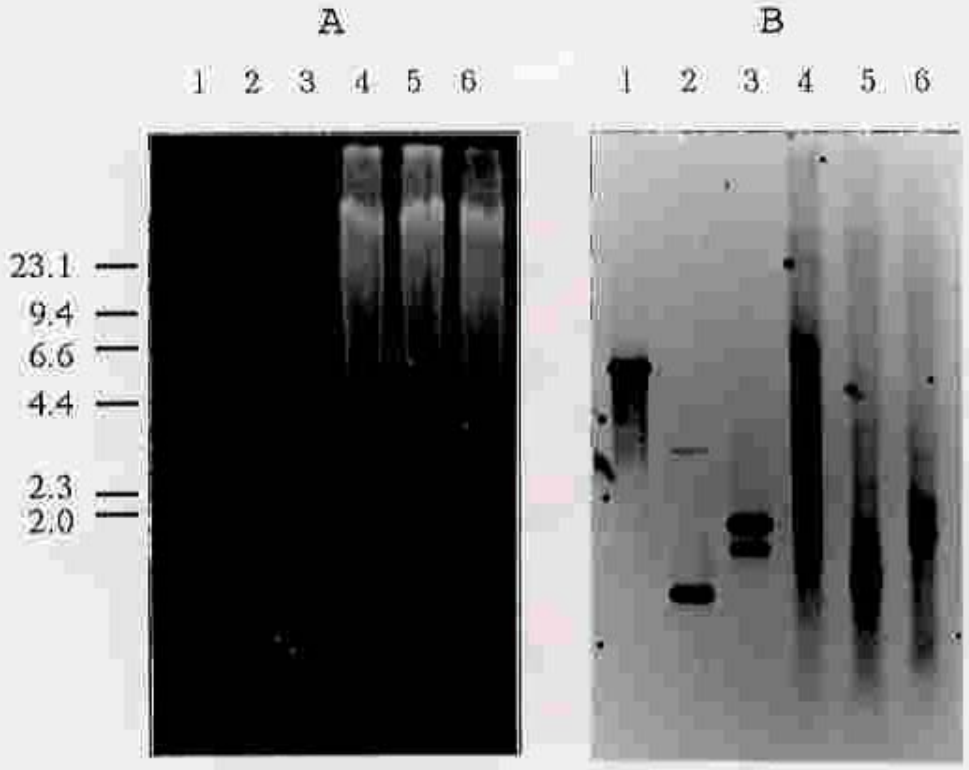


Figure 7. Southern blot analysis of chloroplast and nuclear DNA of *C. vulgaris*. (A) Electrophoresis on a 0.7% agarose gel of chloroplast (lanes 1-3) and nuclear (lanes 4-6) DNA digested with *Bgl*II (lanes 1 and 4), *Eco*RI (lanes 2 and 5) and *Hin*dIII (lanes 3 and 6). Chloroplast DNA of 0.004 µg and nuclear DNA of 5 µg were analyzed. Size markers in kb are indicated on the left. (B) A *chlI* probe amplified from *Nitella* chloroplast DNA was hybridized with a filter blot from the agarose gel.

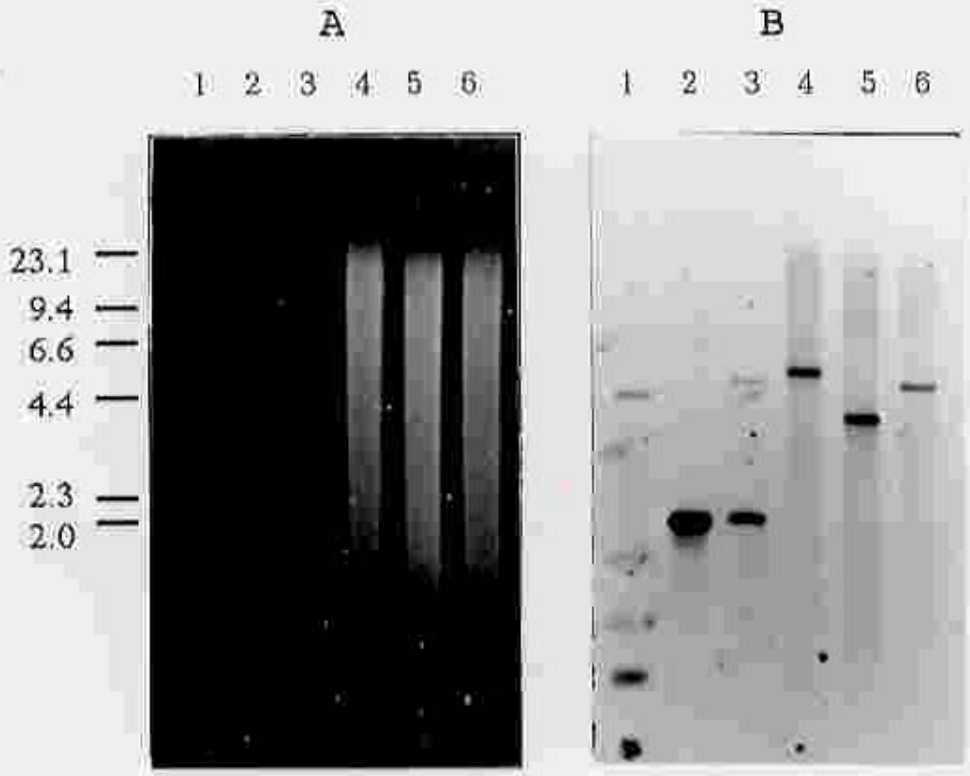


Figure 8. Southern blot analysis of chloroplast and nuclear DNA from *Nitella*. (A) Electrophoresis on a 0.7% agarose gel of chloroplast (lanes 1-3) and nuclear (lanes 4-6) DNA digested with *Bgl*II (lanes 1 and 4), *Eco*RI (lanes 2 and 5) and *Hin*dIII (lanes 3 and 6). Chloroplast DNA of 0.015 μg and nuclear DNA of 0.2 μg were analyzed. Size markers in kb are indicated on the left. (B) A *chlI* probe amplified from *Nitella* chloroplast DNA was hybridized with a filter blot from the agarose gel.

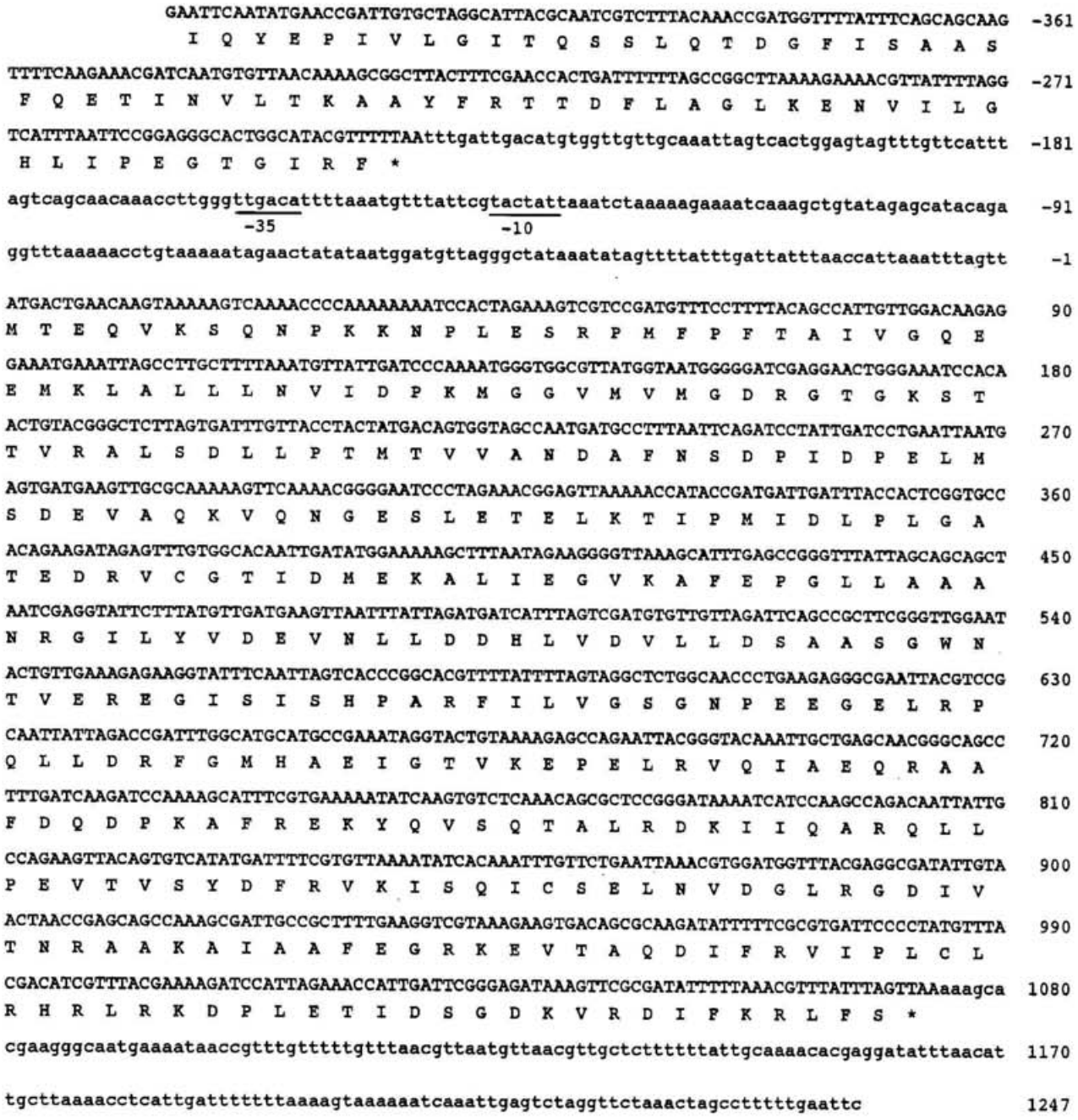

GAATTCAATATGAACCGATTGTGCTAGGCATTACGCAATCGTCTTTACAAACCGATGGTTTTATTTCAGCAGCAAG -361
I Q Y E P I V L G I T Q S S L Q T D G F I S A A S
TTTTCAAGAAACGATCAATGTGTTAACAAAAGCGGCTTACTTTCGAACCACTGATTTTTTAGCCGGCTTAAAAGAAAACGTTATTTTAGG -271
F Q E T I N V L T K A A Y F R T T D F L A G L K E N V I L G
TCATTTAATTCCGGAGGGCACTGGCATACGTTTTTAAtttgattgacatgtggttgttgcaaattagtcactggagtagtttgttcattt -181
H L I P E G T G I R F *
agtcagcaacaaaccttgggttgacattttaaatgtttattcgtactattaaatctaaaaagaaaatcaaagctgtatagagcatacaga -91
-35
-10
ggtttaaaaacctgtaaaaatagaactatataatggatgttagggctataaatatagttttatttgattatttaaccattaaatttagtt -1
ATGACTGAACAAGTAAAAAGTCAAAACCCCAAAAAAAATCCACTAGAAAGTCGTCCGATGTTTCCTTTTACAGCCATTGTTGGACAAGAG 90
M T E Q V K S Q N P K K N P L E S R P M F P F T A I V G Q E
GAAATGAAATTAGCCTTGCTTTTAAATGTTATTGATCCCAAAATGGGTGGCGTTATGGTAATGGGGGATCGAGGAACTGGGAAATCCACA 180
E M K L A L L L N V I D P K M G G V M V M G D R G T G K S T
ACTGTACGGGCTCTTAGTGATTTGTTACCTACTATGACAGTGGTAGCCAATGATGCCTTTAATTCAGATCCTATTGATCCTGAATTAATG 270
T V R A L S D L L P T M T V V A N D A F N S D P I D P E L M
AGTGATGAAGTTGCGCAAAAAGTTCAAAACGGGGAATCCCTAGAAACGGAGTTAAAAACCATACCGATGATTGATTTACCACTCGGTGCC 360
S D E V A Q K V Q N G E S L E T E L K T I P M I D L P L G A
ACAGAAGATAGAGTTTGTGGCACAATTGATATGGAAAAAGCTTTAATAGAAGGGGTTAAAGCATTTGAGCCGGGTTTATTAGCAGCAGCT 450
T E D R V C G T I D M E K A L I E G V K A F E P G L L A A A
AATCGAGGTATTCTTTATGTTGATGAAGTTAATTTATTAGATGATCATTTAGTCGATGTGTTGTTAGATTCAGCCGCTTCGGGTTGGAAT 540
N R G I L Y V D E V N L L D D H L V D V L L D S A A S G W N
ACTGTTGAAAGAGAAGGTATTTCAATTAGTCACCCGGCACGTTTTATTTTAGTAGGCTCTGGCAACCCTGAAGAGGGCGAATTACGTCCG 630
T V E R E G I S I S H P A R F I L V G S G N P E E G E L R P
CAATTATTAGACCGATTTGGCATGCATGCCGAAATAGGTACTGTAAAAGAGCCAGAATTACGGGTACAAATTGCTGAGCAACGGGCAGCC 720
Q L L D R F G M H A E I G T V K E P E L R V Q I A E Q R A A
TTTGATCAAGATCCAAAAGCATTTCGTGAAAAATATCAAGTGTCTCAAACAGCGCTCCGGGATAAAATCATCCAAGCCAGACAATTATTG 810
F D Q D P K A F R E K Y Q V S Q T A L R D K I I Q A R Q L L
CCAGAAGTTACAGTGTCATATGATTTTCGTGTTAAAATATCACAAATTTGTTCTGAATTAAACGTGGATGGTTTACGAGGCGATATTGTA 900
P E V T V S Y D F R V K I S Q I C S E L N V D G L R G D I V
ACTAACCGAGCAGCCAAAGCGATTGCCGCTTTTGAAGGTCGTAAAGAAGTGACAGCGCAAGATATTTTTCGCGTGATTCCCCTATGTTTA 990
T N R A A K A I A A F E G R K E V T A Q D I F R V I P L C L
CGACATCGTTTACGAAAAGATCCATTAGAAACCATTGATTCGGGAGATAAAGTTCGCGATATTTTTAAACGTTTATTTAGTTAAaaagca 1080
R H R L R K D P L E T I D S G D K V R D I F K R L F S *
cgaagggcaatgaaaataaccgtttgtttttgtttaacgttaatgttaacgttgctctttttattgcaaaacacgaggatatttaacat 1170
tgcttaaaacctcattgatttttttaaaagtaaaaaatcaaattgagtctaggttctaaactagccttttgaattc 1247

Figure 9. Nucleotide sequence containing *chlI* and *rpoC2* of *Nitella*. The numbering of nucleotides shown on the right is from the translation start site of *chlI*. Genes corresponding to *chlI* (1-1074) and a part of *rpoC2* (-436 - -235) are indicated by capital letters. Deduced amino acids are indicated by single letters below the nucleotide sequences. Putative promoter sequences are underlined and indicated by -35 and -10.

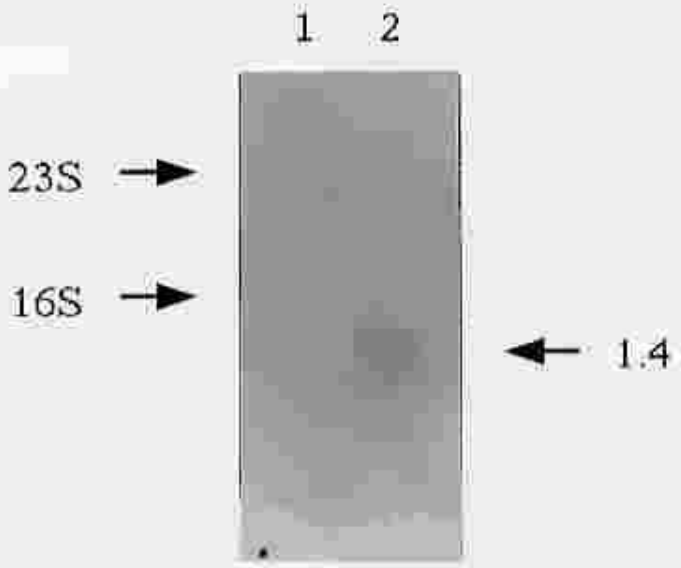


Figure 10. Northern blot analysis of the *chlI* transcripts of *Nitella sp.*. Poly(A)$^+$ (lane 1) and poly(A)$^-$ (lane 2) RNA of *Nitella* was hybridized with a part of the *chlI*. The hybridized band is indicated by a horizontal arrow and the size by kb calculated from *E. coli* 23S and 16S ribosomal RNA.

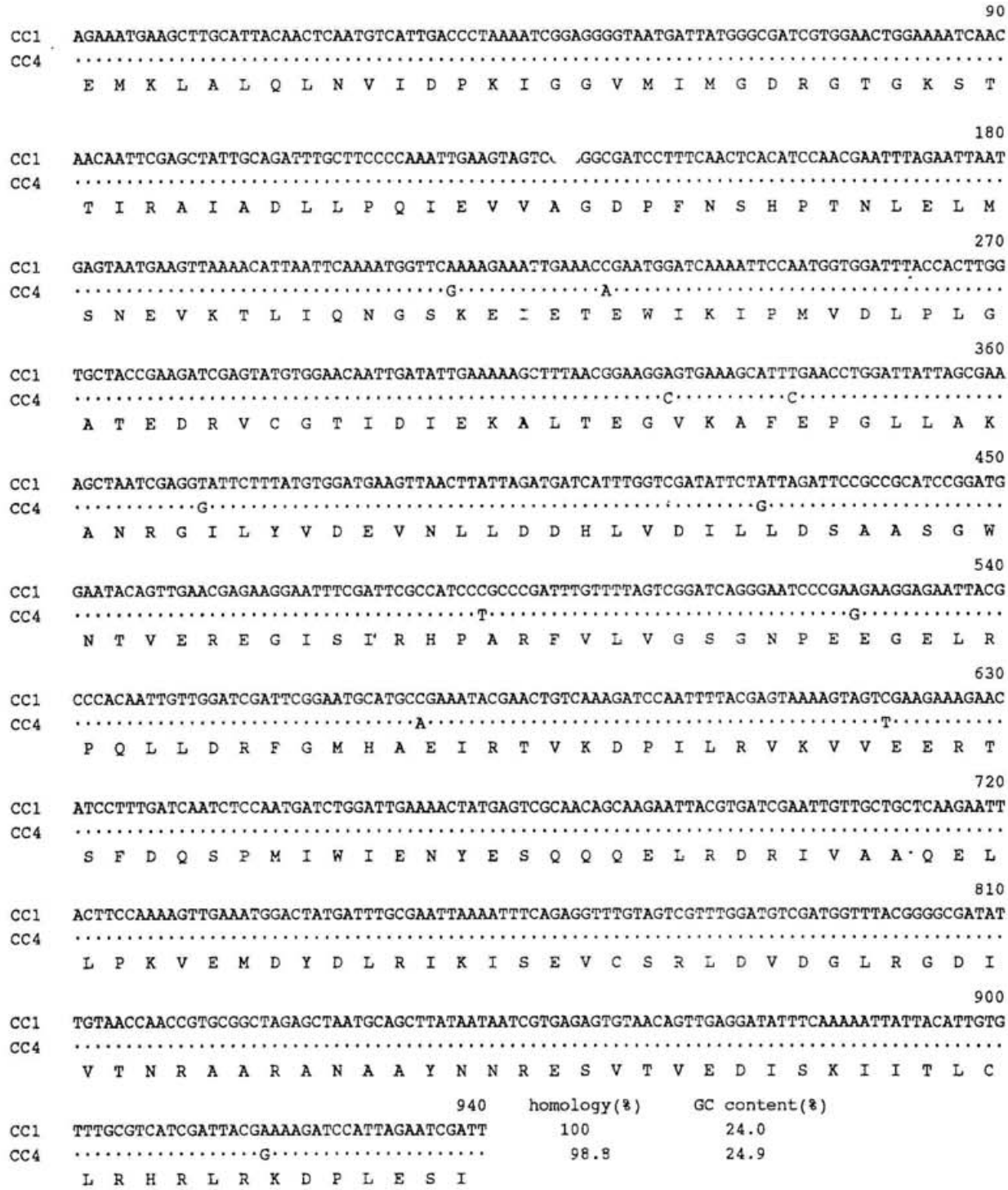

90
CC1 AGAAATGAAGCTTGCATTACAACTCAATGTCATTGACCCTAAAATCGGAGGGGTAATGATTATGGGCGATCGTGGAACTGGAAAATCAAC
CC4 .....
E M K L A L Q L N V I D P K I G G V M I M G D R G T G K S T
180
CC1 AACAATTCGAGCTATTGCAGATTTGCTTCCCCAAATTGAAGTAGTC GGCGATCCTTTCAACTCACATCCAACGAATTTAGAATTAAT
CC4 .....
T I R A I A D L L P Q I E V V A G D P F N S H P T N L E L M
270
CC1 GAGTAATGAAGTTAAAACATTAATTCAAAATGGTTCAAAAGAAATTGAAACCGAATGGATCAAAATTCCAATGGTGGATTTACCACTTGG
CC4 .....G.....A.....
S N E V K T L I Q N G S K E I E T E W I K I P M V D L P L G
360
CC1 TGCTACCGAAGATCGAGTATGTGGAACAATTGATATTGAAAAAGCTTTAACGGAAGGAGTGAAAGCATTTGAACCTGGATTATTAGCGAA
CC4 .....C.....C.....
A T E D R V C G T I D I E K A L T E G V K A F E P G L L A K
450
CC1 AGCTAATCGAGGTATTCTTTATGTGGATGAAGTTAACTTATTAGATGATCATTTGGTCGATATTCTATTAGATTCCGCCGCATCCGGATG
CC4 .....G.....G.....
A N R G I L Y V D E V N L L D D H L V D I L L D S A A S G W
540
CC1 GAATACAGTTGAACGAGAAGGAATTTCGATTCGCCATCCCGCCCGATTTGTTTTAGTCGGATCAGGGAATCCCGAAGAAGGAGAATTACG
CC4 .....T.....G.....
N T V E R E G I S I R H P A R F V L V G S G N P E E G E L R
630
CC1 CCCACAATTGTTGGATCGATTCGGAATGCATGCCGAAATACGAACTGTCAAAGATCCAATTTTACGAGTAAAAGTAGTCGAAGAAAGAAC
CC4 .....A.....T.....
P Q L L D R F G M H A E I R T V K D P I L R V K V V E E R T
720
CC1 ATCCTTTGATCAATCTCCAATGATCTGGATTGAAAACTATGAGTCGCAACAGCAAGAATTACGTGATCGAATTGTTGCTGCTCAAGAATT
CC4 .....
S F D Q S P M I W I E N Y E S Q Q Q E L R D R I V A A Q E L
810
CC1 ACTTCCAAAAGTTGAAATGGACTATGATTTGCGAATTAAAATTTCAGAGGTTTGTAGTCGTTTGGATGTCGATGGTTTACGGGGCGATAT
CC4 .....
L P K V E M D Y D L R I K I S E V C S R L D V D G L R G D I
900
CC1 TGTAACCAACCGTGCGGCTAGAGCTAATGCAGCTTATAATAATCGTGAGAGTGTAACAGTTGAGGATATTTCAAAAATTATTACATTGTG
CC4 .....
V T N R A A R A N A A Y N N R E S V T V E D I S K I I T L C
940 homology(%) GC content(%)
CC1 TTTGCGTCATCGATTACGAAAAGATCCATTAGAATCGATT 100 24.0
CC4 .....G..... 98.8 24.9
L R H R L R K D P L E S I

Figure 11. Comparison of nucleotide sequences (CC1 and CC4) amplified from the total DNA of *Chara* using a *chlI*-specific primer pair. Nucleotides of CC4 differing from CC1 are shown while those identical to CC1 are represented by dots. Deduced amino acids are indicated by single letters below the nucleotide sequences. The homology (%) of the nucleotide sequence with those of CC1 and the GC content (%) in the third letter of the codons is shown at the end of each sequence.

		total		third letter	
		tufA	*chlI*	*tufA*	*chlI*
Algae	*Nitella sp.*	37.5[a]	39.7	18.0[a]	22.0
	Chlorella vulgaris	38.9[b]	36.5	16.2[b]	11.0
	Euglena gracilis	35.5[c]	32.5[c]	12.4[c]	13.3[c]
	Odontella sinensis	39.0[d]	34.8[d]	15.7[d]	11.7[d]
	Cryptomonas phi	37.4[e]	36.3[j]	13.5[e]	18.6[j]
	Porphyra purpurea	37.6[f]	35.3[f]	14.9[f]	17.0[f]
	Cyanophora paradoxa	37.0[g]	33.7[g]	11.5[g]	11.5[g]
Land plants	*Nicotiana tabacum*		44.6[k]		38.6[k]
	Arabidopsis thaliana	45.7[h]	42.5[l]	39.8[h]	32.2[l]
	Glycine max	55.6[i]	44.2[m]	66.7[i]	36.8[m]

Table 1. The GC content of *tufA* and *chlI*, and that in the third letter of the codons. The GC content of land plants was calculated from only the region coding the mature protein.

Accession number

a: U09439 b: AB001684 c: X70810

d: Z67753 e: X52912 f: U38804

g: U30821 h: X52256 i: X66062

j: Z21976 k: AF014053 l: X51799

m: D45857

References

Adams CC, Stern DB: Control of mRNA stability in chloroplasts by 3'-inverted repeats: effects of stem and loop mutations on degradation of *psbA* mRNA *in vitro*. Nucleic Acids Res. 18: 6003-6010 (1990).

Baldauf SL, Manhart JR, Palmer JD: Different fates of the chloroplast *tufA* gene following its transfer to the nucleus in green algae. Proc. Natl. Acad. Sci. USA 87: 5317-5321 (1990).

Baldauf SL, Palmer JD: Evolutionary transfer of the chloroplast *tufA* gene to the nucleus. Nature 344: 262-265 (1990).

Bollivar DW, Suzuki JY, Beatty JT, Dobrowolski JM, Bauer CE: Directed mutational analysis of bacteriochlorophyll *a* biosynthesis in *Rhodobacter capsulatus*. J. Mol. Biol. 237: 622-640 (1994).

Chen HC, Stern DB: Specific binding of chloroplast proteins *in vitro* to the 3' untranslated region of spinach chloroplast *petD* mRNA. Mol. Cell Biol. 11: 4380-4388 (1991).

Chen Q, Adams CC, Usack L, Yang J, Monde RA, Stern DB: An AU-rich element in the 3' untranslated region of the spinach chloroplast *petD* gene participates in sequence-specific RNA-protein complex formation. Mol. Cell Biol. 15: 2010-2018 (1995).

Covello PS, Gray MW: Silent mitochondrial and active nuclear genes for subunit 2 of cytochrome *c* oxidase (*cox2*) in soybean: evidence for RNA-mediated gene transfer. EMBO J. 11: 3815-3820 (1992).

Douglas SE, Reith M: A *bchI* homolog, encoding a subunit of Mg chelatase, is located on the plastid genomes of red and cryptomonad algae. J. Mar. Biotechnol. 1: 135-141 (1993).

Gantt JS, Baldauf SL, Calie PJ, Weeden NF, Palmer JD: Transfer of *rpl22* to the nucleus greatly preceded its loss from the chloroplast and involved the gain of an intron. EMBO J. 10: 3073-3078 (1991).

Gawel NJ, Jarret RL: A modified CTAB DNA extraction procedure for Musa, and Ipomea. Plant Mol. Biol. Reptr. 9: 262-266 (1991).

Gibson LCD, Marrison JL, Leech RM, Jensen PE, Bassham DC, Gibson M, Hunter CN: A putative Mg chelatase subunit from *Arabidopsis thaliana* cv C24: sequence and transcript analysis of the gene, import of the protein into chloroplast, and in situ localization

of the transcript and protein. Plant Physiol. 111: 61-71 (1996).

Gibson LCD, Willows RD, Kannangara CG, von Wettstein D, Hunter CN: Magnesium-protoporphyrin chelatase of *Rhodobacter sphaeroides*: Reconstitution of activity by combining the products of the *bchH, -I, and -D* genes expressed in *Escherichia coli*. Proc. Natl. Acad. Sci. USA 92: 1941-1944 (1995).

Graham LE: Green algae to land plant: an evolutionary transition. J. Plant Res. 109: 241-251 (1996).

Gray MW: The endosymbiotic hypothesis revisited. Int. Rev. Cytol. 141: 233-357 (1992).

Grohmann L, Brennicke A, Schuster W: The mitochondrial gene encoding ribosomal protein S12 has been translocated to the nuclear genome in *Oenothera*. Nucleic Acids Res. 20: 5641-5646 (1992).

Hallick RB, Hong L, Drager RG, Favreau MR, Monfort A, Orsat B, Spielmann A, Stutz E: Complete sequence of *Euglena gracilis* chloroplast DNA. Nucleic Acids Res. 21: 3537-3544 (1993).

Hayes R, Kudla J, Schuster G, Gabay L, Maliga P, Gruissem W: Chloroplast mRNA 3'-end processing by a high molecular weight protein complex is regulated by nuclear encoded RNA binding proteins. EMBO J. 15: 1132-1141 (1996).

Hiratsuka J, Shimada H, Whittier R, Ishibashi T, Sakamoto M, Mori M, Kondo C, Honji Y, Shun CR, Meng BY, Li YQ, Kanno A, Nishizawa Y, Hirai A, Shinozaki K, Sugiura M: The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. Mol. Gen. Genet. 217: 185-194 (1989).

Hudson A, Carpenter R, Doyle S, Coen ES: *Olive*: a key gene required for chlorophyll biosynthesis in *Antirrhinum majus*. EMBO J. 12: 3711-3719 (1993).

Jensen PE, Willows RD, Petersen BL, Vothknecht UC, Stummann BM, Kannangara CG, von Wettstein D, Henningsen KW: Structural genes for Mg-chelatase subunits in barley: *Xantha-f, -g and -h*. Mol. Gen. Genet. 250: 383-394 (1996).

Kadowaki K, Kubo N, Ozawa K and Hirai A: Targeting presequence acquisition after mitochondrial gene transfer to the nucleus

occurs by duplication of existing targeting signals. EMBO J. 15: 6652-6661 (1996).

Koncz C, Mayerhofer R, Koncz-Kalman Z, Nawrath C, Reiss B, Redei GP, Schell J: Isolation of a gene encoding a novel chloroplast protein by T-DNA tagging in *Arabidopsis thaliana*. EMBO J. 9: 1337-1346 (1990).

Kowallik KV, Stoebe B, Schaffran I, Kroth-Pancic P, Freier U: The chloroplast genome of a chlorophyll *a+c*-containing alga, *Odontella sinensis*. Plant Mol. Biol. Reptr. 13: 336-342 (1995).

Kranz HD, Miks D, Siegler ML, Capesius I, Sensen CW, Huss VAR: The origin of land plants: Phylogenetic relationships among charophytes, bryophytes, and vascular plants inferred from complete small-subunit ribosomal RNA gene sequences. J. Mol. Evol. 41: 74-84 (1995).

Maier RM, Neckermann K, Igloi GL, Kössel H: Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. J. Mol. Biol. 251: 614-628 (1995).

Nakayama M, Masuda T, Sato N, Yamagata H, Bowler C, Ohta H, Shioi Y, Takamiya K: Cloning, subcellular localization and expression of CHLI, a subunit of magnesium-chelatase in soybean. Biochem. Biophys. Res. Commun. 215: 422-428 (1995).

Nugent JM, Palmer JD: RNA-mediated transfer of the gene *coxII* from the mitochondrion to the nucleus during flowering plant evolution. Cell 66: 473-481 (1991).

Ohyama K, Fukuzawa H, Kohchi T, Shirai H, Sano S, Umesono K, Shiki Y, Takeuchi M, Chang Z, Aota ST, Inokuchi H, Ozeki H: Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. Nature 322: 572-574 (1986).

Ohyama K, Kouchi T, Sano T, Yamada Y: Newly identified groups of genes in chloroplasts. Trends Biochem. Sci. 13: 19-22 (1988).

Oliver JL, Marín A, Martínez-Zapater JM: Chloroplast genes transferred to the nuclear plant genome have adjusted to nuclear base composition and codon usage. Nucleic Acids Res. 18: 65-73 (1990).

Orsat B, Monfort A, Chatellard P, Stutz E: Mapping and

sequencing of an actively transcribed *Euglena gracilis* chloroplast gene (*ccsA*) homologous to the *Arabidopsis thaliana* nuclear gene *cs (ch-42)*. FEBS Lett. 303: 181-184 (1992).

Palmer JD: Comparative organization of chloroplast genomes. Annu. Rev. Genet. 19: 325-354 (1985).

Palmer JD: Plastid chlomosomes: structure and evolution. In: Bogorad L, Vasil LK (ed) Cell culture and somatic cell genetics of plants, vol 7A: The molecular biology of plastids. Academic Press, San Diego,5-53 (1991).

Reith M, Munholland J: A high-resolution gene map of the chloroplast genome of the red alga *Porphyra purpurea*. Plant Cell 5: 465-475 (1993).

Reith M, Munholland J: Complete nucleotide sequence of the *Porphyra purpurea* chloroplast genome. Plant Mol. Biol. Reptr. 13: 333-335 (1995).

Sambrook J, Fritsch EF, Maniatis T: Molecular cloning: A Laboratory Manual, 2nd edn. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1989).

Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74: 5463-5467 (1977).

Sánchez H, Fester T, Kloska S, Schröder W, Schuster W: Transfer of *rps19* to the nucleus involves the gain of an RNP-binding motif which may functionally replace RPS13 in *Arabidopsis* mitochondria. EMBO J. 15: 2138-2149 (1996).

Schmidt M, Pichl L, Lepper M, Feierabend J: Identification of the nuclear-encoded chloroplast ribosomal protein L12 of the monocotyledonous plant *Secale cereale* and sequencing of two different cDNAs with strong codon bias. Biochim. Biophys. Acta 1172: 349-352 (1993).

Schuster G, Gruissem W: Chloroplast mRNA 3' end processing requires a nuclear-encoded RNA-binding protein. EMBO J. 10: 1493-1502 (1991).

Shinozaki K, Ohne M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, Ohto C, Torazawa K, Meng BY, Sugita M, Deno H, Kamogashira T, Yamada K, Kusuda J, Takaiwa F, Kato A, Thodou N,

Shimada H, Sugiura M: The complete nucleotide sequence of the tobacco chloroplast genome: Its gene organization and expression. EMBO J. 5: 2043-2049 (1986).

Stern DB, Gruissem W: Control of plastid gene expression: 3' inverted repeats act as mRNA processing and stabilizing elements, but do not terminate transcription. Cell 51: 1145-1157 (1987).

Stern DB, Kindle KL: 3' end maturation of the *Chlamydomonas reinhardtii* chloroplast *atpB* mRNA is a two-step process. Mol. Cell Biol. 13: 2277-2285 (1993).

Stern DB, Radwanski ER, Kindle KL: A 3' stem/loop structure of the *Chlamydomonas* chloroplast *atpB* gene regulates mRNA accumulation *in vivo*. Plant Cell 3: 285-297 (1991).

Stirewalt VL, Michalowski CB, Löffelhardt W, Bohnert HJ, Bryant DA: Nucleotide sequence of the cyanelle genome from *Cyanophora paradoxa*. Plant Mol. Biol. Reptr. 13: 327-332 (1995).

Sugiura M: The chloroplast genome. Plant Mol. Biol. 19: 149-168 (1992).

Sugiura M: Structure and replication of chloroplast DNA. In: Anderson B, Salter AH, Barber J (eds) Frontiers in molecular biology: molecular genetics of photosynthesis. IRL Press, Oxford, pp. 58-74 (1996).

Tamiya H, Iwamura T, Shibata K, Hase E, Nihei T: Correlation between photosynthesis and light-independent metabolism in the growth of *Chlorella*. Biochim. Biophys. Acta 12: 23-40 (1953).

Wakasugi T, Nagai T, Kapoor M, Sugita M, Ito M, Ito S, Tsudzuki J, Nakashima K, Tsudzuki T, Suzuki Y, Hamada A, Ohta T, Inamura A, Yoshinaga K, Sugiura M: Complete nucleotide sequence of the chloroplast genome from the green alga *Chlorella vulgaris*: The existence of genes possibly involved in chloroplast division. Proc. Natl. Acad. Sci. USA 94: 5967-5972 (1997).

Wakasugi T, Tsudzuki J, Ito S, Nakashima K, Tsudzuki T, Sugiura M: Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. Proc. Natl. Acad. Sci. USA 91: 9794-9798 (1994).

Willows RD, Gibson LCD, Kannangara CG, Hunter CN, von Wettstein D: Three separate proteins constitute the magnesium chelatase of

Rhodobacter sphaeroides. Eur. J. Biochem. 235: 438-443 (1996).

Wischmann C, Schuster W: Transfer of *rps10* from the mitochondrion to the nucleus in *Arabidopsis thaliana*: evidence for RNA-mediated transfer and exon shuffling at the integration site. FEBS Lett. 374: 152-156 (1995).

Yoshinaga K, Ohta T, Suzuki Y, Sugiura M: *Chlorella* chloroplast DNA sequence containing a gene for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase and a part of a possible gene for the β' subunit of RNA polymerase. Plant Mol. Biol. 10: 245-250 (1988).

Acknowledgements

I thank Dr. T. Ikemura for providing the opportunity for this study and for helpful advice. I also thank Dr. H. Morishima, Dr. T. Gojobori, Dr. S. Horai, and Dr. N. Saitou for helpful advice. I thank Dr. K. Yoshinaga, Shizuoka University, for advising on various experimental technique and helpful discussions.