AN ABSTRACT OF THE THESIS OF

Sean M. Amberg for the degree of Master of Science in Genetics Program presented on August 7, 1989

Title: Nucleotide Sequence of Two Chloroplast Genes from a *Chlorella*-like Green Alga: the Large Subunit of Ribulose-1,5-bisphosphate Carboxylase/Oxygenase and Ribosomal Protein S14

Abstract approved: Redacted for Privacy
Russel H. Meints

Chlorella is a genus of unicellular, eukaryotic green algae. *Chlorella*-like algae are found as endocellular symbionts within a number of animal species. Two chloroplast genes were sequenced from the exsymbiotic strain *Chlorella* N1a, originally isolated from *Paramecium bursaria*. The genes for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (*rbc*L) and the ribosomal protein S14 (*rps*14) are oriented in the same direction and are separated by 402 bp. A comparison of the exsymbiont *rbc*L and a free-living *Chlorella rbc*L with other reported *rbc*L sequences was made. The gene of the exsymbiont was very closely related to the gene of the free-living species. There were 80 nucleotide differences between the exsymbiont and the free-living species, mostly in the third position of the codon. These substitutions translate into twelve predicted amino acid differences. From this information, it appears as though the chloroplast genome of *Chlorella* N1a has not diverged significantly from that of free-living *Chlorella*, at least on a functional level.

Nucleotide Sequence of Two Chloroplast Genes from a *Chlorella*-like Green Alga: the Large Subunit of Ribulose-1,5-bisphosphate Carboxylase/Oxygenase and Ribosomal Protein S14

by

Sean M. Amberg

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Master of Science

Completed August 7, 1989

Commencement June 1990

APPROVED:

Redacted for Privacy

Professor of Botany and Plant Pathology in charge of major

Redacted for Privacy

Chairman of Genetics Program

Redacted for Privacy

Dean of Graduate School

Date thesis is presented August 7, 1989

Candidate Sean M. Amberg

TABLE OF CONTENTS

LIST OF FIGURES

Nucleotide Sequence of Two Chloroplast Genes from a Chlorella-like Green Alga: the Large Subunit of Ribulose-1,5-bisphosphate Carboxylase/Oxygenase and Ribosomal Protein S14

INTRODUCTION

Symbiotic Chlorella. The genus *Chlorella* is a broad taxon of unicellular, asexually reproducing, eukaryotic green algae. Various members of this genus have been recognized to exist in a symbiotic relationship with protozoa and primitive animals, including sponges, coelenterates, turbellaria, and molluscs (Reisser and Wiessner 1984). The evolutionary origin of symbiotic *Chlorella* is unknown, but due to their similarity to free-living *Chlorella*, it has always been assumed that the symbionts have a common ancestor with the free-living *Chlorella*. Of the *Chlorella* symbionts which have been assigned to a particular species, all belong to one of three closely related species: *C. vulgaris*, *C. sorokiniana*, and *C. saccharophila* (Reisser 1984). Many symbiotic strains can be isolated as exsymbionts into culture (Karakashian and Karakashian 1965). A common feature of cultured exsymbionts is a significant sugar release in the form of maltose (Mews 1980) or glucose (Wilkinson 1980). Symbiotic *Chlorella* have been shown to release up to 86% of the photosynthate to the host (Muscatine et al. 1967). The transport of fixed carbon to the host is thought to be a significant feature of the symbiosis.

The discovery of a family of infectious, plaque-forming viruses that infect exsymbiotic *Chlorella* cultured from *Paramecium bursaria* has led to an extensive investigation into the viral biology of this system (Van Etten et al. 1988). The prototype virus, PBCV-1, infects only exsymbiotic strains NC64A, N1a, and ATCC-30562; it is a large (190 nm diameter) polyhedral virus with 330 kb of double-stranded DNA. Principally two host strains have been investigated, NC64A and N1a. *Chlorella* NC64A has been designated as a *Chlorella vulgaris* based on DNA hybridization studies (Douglas and Huss 1986). *Chlorella* N1a, on which this work was performed, is believed to be

nearly identical to NC64A. The major evidence of this is derived from studies of the chloroplast DNA from both strains which demonstrate identical restriction fragment patterns (Meints, unpublished).

The interest in performing a DNA sequence analysis on a part of the *Chlorella* N1a chloroplast genome stems from the need to have molecular data in addition to what has principally been morphological data. An analysis of sequence divergence may help establish a more rigid taxonomic classification. The evolution of an organism in a symbiotic relationship may diverge significantly from free-living organisms of the same species. For example, a symbiotic alga might be able to utilize a carbon source from the host, circumventing the photosynthetic route of carbon fixation. This consideration makes the chloroplast genome a reasonable place to investigate, as one would expect the genome to reflect the selective pressure placed upon it. In addition, chloroplast genomes are generally considered good molecular rulers of evolution by nature of their small size and high level of conservation (relative to nuclear genomes).

Chloroplast Genomes. Chloroplasts contain their own DNA in the form of a single, circular molecule that can range in size from 89 kilobases for the alga *Codium fragile* (Hedberg et al. 1981) to well over 400 kb for the alga *Acetabularia mediterranea* (Tymms and Schweiger 1985). Land plants, on the other hand, have a more strictly conserved size of 120 to 160 kb (Palmer 1985). The genes of land plant chloroplast genomes tend to be arranged similarly, while the genomes of algae that have been studied have almost no conspicuous similarity (Palmer and Stein 1986).

It has been suggested that the chloroplast genome may serve as an appropriate marker for phylogeny, in terms of the size of the genome, the gene arrangement, presence of inverted repeats, and gene sequence (Palmer 1985). This idea has been widely accepted, particularly in the case of land plants. In the case of algae, the information for such an analysis has been lacking. The observation that chloroplast genomes evolve

more slowly than nuclear genomes has been borne out by studies of the rate of nucleotide substitution within 23 chloroplast and 3 nuclear genes from higher plants representing 16 species (Wolfe et al. 1987).

The chloroplast genome of *Chlorella* N1a has been shown to be distinct from free-living species on several accounts (Schuster et al. 1989). The N1a genome has no inverted repeat of the ribosomal RNA genes. Inverted repeats are found in free-living *Chlorella* chloroplast DNA, and in most other plants examined. The chloroplast genome size of the exsymbiont, 120 kb, is significantly smaller than that of free-living species (175 kb for *Chlorella saccharophila* var. ellipsoidea and *Chlorella ellipsoidea*; Yamada 1982). This size difference is too large to be accounted for by the absence of an inverted repeat. Finally, the gene arrangement on the N1a chloroplast genome is unique among reported chloroplast maps.

Ribulose-1,5-bisphosphate carboxylase/oxygenase. One of the most abundant proteins in the world is ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco), an enzyme which catalyzes two competing reactions: carboxylation of ribulose-1,5-bisphosphate to yield 3-phosphoglycerate, or oxygenation of the same molecule to yield 3-phosphoglycerate and 2-phosphoglycolate (photorespiration). The holoenzyme is composed of sixteen proteins: eight large subunits of 50-55 kilodaltons each, and eight small subunits of around 15 kDa each. In eukaryotes, the gene for the large subunit is located on the chloroplast genome, while the gene for the small subunit is located in the nucleus. In photosynthetic prokaryotes, both genes are found together and are co-transcribed (Shinozaki and Sugiura 1983). This prokaryotic context has given rise to the view that the gene for the small subunit moved to the nucleus very early in the evolution of plants. The discovery of a eukaryotic alga in which the small subunit is in fact still located in the chloroplast has strengthened this view (Reith and Cattolico 1986).

The large subunit of Rubisco contains the active sites and is conserved between

species to a much greater extent than the small subunit. The locus of the large subunit, *rbc*L, has been sequenced from at least 19 species (Wolfe 1989). *Chlorella* N1a *rbc*L was identified by low stringency hybridization with a portion of pea *rbc*L provided by Dr. Jeffrey Palmer at the University of Michigan (Schuster et al. 1989).

Chloroplast Ribosomes. The endosymbiotic theory of chloroplast origin holds that a primitive endosymbiotic, photosynthetic prokaryote evolved into the modern chloroplast. In fact, the genetic machinery of the chloroplast is very much like that of prokaryotes: transcriptional promoters are nearly identical to those of *E. coli*, transcriptional terminators appear similar to prokaryotic terminators (although they may function very differently; Stern and Gruissem 1987), messenger RNA is neither "capped" with 7-methylguanylic acid nor polyadenylated at the 3' end as eukaryotic mRNA is, and messages are translated via prokaryotic-like, 70S ribosomes (Weil 1987). All of the ribosomal RNA for these ribosomes is encoded within the chloroplast. Some of the ribosomal proteins, of which there are about 60, are encoded in the nucleus and imported to the chloroplast; the remainder are chloroplast-encoded.

The entire chloroplast genome has been sequenced from three species: the liverwort *Marchantia polymorpha* (Ohyama et al. 1986), tobacco species *Nicotiana tabacum* (Shinozaki et al. 1986), and the rice species *Oryza sativa* (Hiratsuka et al. 1989). Both the liverwort and the tobacco chloroplast genomes were found to code for 19 ribosomal proteins with extensive homology to those found in *E. coli*, but there was one important difference. Liverwort chloroplast has a gene for the large subunit protein L21 (*rpl*21), but tobacco does not. Tobacco chloroplast, on the other hand, has a gene for the small subunit protein S16 (*rps*16), which is absent in liverwort. Presumably, this reflects a difference in the way in which some genes have moved to the nucleus. The chloroplast genome of rice encodes a complement of 20 ribosomal proteins. Like tobacco, *rpl*21 is absent, but unlike either of the first two genomes, rice has a copy of the large subunit

protein L36.

The access to large databanks of both DNA sequences and protein sequences such as GenBank (National Institute of Health) and the Swiss-Protein sequence databank (University of Geneva, Switzerland) has made it possible to identify unknown sequences rapidly. Although any particular homology can by no means conclusively identify a sequence by itself, in many cases homology searches are the most rapid means of initial identification. In the case of the *Chlorella* N1a genome, a portion of the 3' untranslated region of the target gene (*rbc*L) bore a very high homology to the gene for the ribosomal small subunit protein S14 from liverwort. This information led to an extension of the sequencing project, which confirmed a complete open reading frame with homology to S14 sequences from liverwort, tobacco, *E. coli*, spinach, maize, pea, and broad bean mitochondria.

Nucleotide sequence of two chloroplast genes from a *Chlorella*-like green alga: the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase and ribosomal protein S14

Sean M. Amberg and Russel H. Meints[1]

Department of Botany and Plant Pathology, Oregon State University, Corvallis, Oregon 97331-2906

Running Title: Nucleotide sequence of *rbc*L and *rps*14 from a green alga

[1]To whom offprint requests should be addressed

All correspondence should be addressed to:
Dr. Russel Meints
Department of Botany and Plant Pathology
Oregon State University
Corvallis, OR 97331-2906
USA

SUMMARY

Two chloroplast genes were isolated and sequenced from an exsymbiotic strain of the eukaryotic, unicellular green alga *Chlorella*. The genes for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (*rbc*L) and the ribosomal protein S14 (*rps*14) are oriented in the same direction and are separated by 402 bp. A comparison of the exsymbiont *rbc*L and a free-living *Chlorella rbc*L with other reported *rbc*L sequences was made. The gene of the exsymbiont is very closely related to that of the free-living species, coding for only 12 predicted amino acid differences.

Key words: *Chlorella*, *rbc*L, chloroplast *rps*14, symbiont

INTRODUCTION

Chlorella-like green algae are found as endocellular, hereditary symbionts within a number of animal species (Trench 1979). One such alga, designated *Chlorella* N1a, was isolated as an exsymbiont from *Paramecium bursaria* and is readily cultured axenically (Meints et al. 1986). This particular alga has been of significant interest because it serves as a host for a family of large dsDNA-containing, plaque-forming viruses (Van Etten et al. 1988).

The endocellular, symbiotic origin of these algae places them in a novel situation with regard to chloroplast evolution. As obligate photoheterotrophs, these organisms rely on their host for nitrogen in the form of amino acids, so a carbon source independent of CO_2 fixation is available. The evolution of the chloroplast genome of such an organism might be expected to diverge significantly from free-living algae and land plants.

N1a chloroplast DNA is a circular genome of 120 kb with a G/C content of 38% (Schuster et al. 1989). The genome of this exsymbiont is distinct from the genome of free-living *Chlorella* in terms of its reduced size, gene organization, and the lack of an inverted repeat. The purpose of this work is to determine the extent of sequence divergence for a well-characterized chloroplast gene.

Ribulose-1,5-bisphosphate carboxylase/oxygenase is the primary enzyme responsible for the fixation of inorganic carbon in photosynthesis. In eukaryotes, the functional form of this enzyme consists of eight large subunits of about 55 kilodaltons each and eight small subunits of 15 kD each. The gene for the large subunit (*rbc*L) is encoded in the genome of the chloroplast, while the gene for the small subunit (*rbc*S) is encoded in the nuclear genome, with the exception of prokaryotes and the chromophytic alga *Olisthodiscus luteus* (Reith and Cattolico 1986). As the gene for the large subunit is highly conserved and well-characterized, it serves as an excellent evolutionary marker. The nucleotide sequence of *rbc*L has been determined for at least 19 species (Wolfe 1989), including three species of Chlorophyta: *Chlamydomonas reinhardtii* (Dron et al.

1982), *Chlamydomonas moewusii* (Yang et al. 1986), and *Chlorella ellipsoidea* (Yoshinaga et al. 1988).

The gene for S14 (*rps*14), a ribosomal protein with substantial homology to that found in *E. coli*, was discovered from sequence analysis to exist just downstream of *rbc*L in the *Chlorella* N1a chloroplast genome. It is reported here in its entirety.

MATERIALS AND METHODS

Algal cultures. Algae were grown axenically in culture flasks under constant fluorescent light (40 $\mu Ei/m^2 \cdot sec$) at 25°C with moderate shaking. Growth medium was a modified form of Bold's Basal Medium (Nichols and Bold 1965) containing 0.25% sucrose and 1% proteose peptone (KBBM); tetracycline (25 µg/ml) was used as an antibiotic. Cultures reach stationary phase at a density of about 8×10^8 cells/ml.

DNA Preparation. DNA was isolated as described (Schuster et al. 1989). DNA preparations were centrifuged on CsCl density gradients and bands were separated on an ISCO fractionator (Instrument Specialties Co., Lincoln, NE) with a UV absorbance monitor. Fractions containing chloroplast DNA were identified as described (Schuster et al. 1989).

Cloning and sequencing. Cloning vectors included pUC19, M13mp18/19, pUC118, and pUC119. Host strains used were *Escherichia coli* strains JM83, JM101, and JM109. pUC118/119 clones were used to construct directional deletions using exonuclease III (Boehringer Mannheim Biochemicals) and S1 nuclease (Bethesda Research Laboratories) as described (Henikoff 1984). Phage M13KO7 was employed as a helper phage to generate single-stranded sequencing templates of pUC118/119 clones (Vieira and Messing 1987). Dideoxy chain-termination sequencing reactions utilized either Klenow (large subunit of *E. coli* DNA polymerase) or a Sequenase kit (using a modified T7 DNA polymerase), both purchased from U. S. Biochemical Corp.; supplier's instructions were followed. [α-^{32}P]dATP (800 Ci/mmol) was purchased from New England Nuclear Research Products.

RESULTS AND DISCUSSION

A chloroplast DNA fragment hybridizing to a portion of pea *rbc*L (kindly provided by Dr. Jeffrey Palmer, University of Michigan) was previously identified as a 3.8 kb *Sal*I-*Xho*I segment of clone Kpn4 (Schuster et al. 1989). Subclone probing narrowed this region to a 2 kb *Eco*RI-*Hin*dIII piece, which was subsequently sequenced. Both coding and non-coding strands were sequenced (**Fig. 1**). A routine search of the Protein Identification Resource databank (National Biomedical Research Foundation) using IntelliGenetics software revealed a downstream open reading frame to have significant homology to several S14 proteins.

The DNA sequence presented in **Figure 2** is 64.9% A/T, which is similar to the previously calculated A/T content for the entire chloroplast genome (Schuster et al. 1989). Regions identified as untranslated are more A/T-rich than the open reading frames (72.9% vs. 59.9%). Sequences resembling a prokaryotic promoter are indicated for *rbc*L. The "TTGTGA" sequence located 111 bp upstream of the *rbc*L initiation codon resembles the "-35" sequence (Rosenberg and Court 1979) of TTGACA, while the "TAGAAT" sequence located 88 bp upstream suggests a Pribnow box (TATAAT) motif (Pribnow 1975). This potential promoter would suggest a transcription start site around -70 to -75. *rps*14 has several potential promoter sequences, but none with a conspicuous homology to a prokaryotic promoter. The best fit to a consensus promoter exists at -28 to -23 (TTGAAA) and at -7 to -2 (TATAAC); promotion from this position would result in a truncated protein with respect to *E. coli* (Yaguchi et al. 1983). It should be noted that an in-frame ATG is found at the sixth codon; translational initiation at this site would yield a protein of 95 amino acids instead of 100. No Shine-Dalgarno sequences (ribosome binding sites) were detected for either *rbc*L or *rps*14.

Two potential transcriptional termination signals are identified for *rbc*L, either or both of which may be involved in the prokaryotic-like termination found in chloroplast genes (Shinozaki et al. 1986). These signals are in the form of inverted repeats capable

of forming stem-loop structures, and are found 18 and 52 nt downstream of the termination codon; each repeat contains 8 bases. The function of these inverted repeats may not be the same as in *E. coli*, however. Evidence obtained *in vitro* suggests that 3' inverted repeats have no effect on transcriptional termination, but are instead effective transcript stabilizing elements (Stern and Gruissem 1987).

A potential "termination signal" was also identified 14 nt downstream of the *rps*14 termination codon; this is a 9 bp inverted repeat with one mismatch. A longer inverted repeat of 20 bp was identified 358 to 401 bp 5' to *rbc*L, between positions 198 and 240, which could be a terminator for an upstream gene.

The *rpo*C-like gene identified in *Chlorella ellipsoidea* (Yoshinaga et al. 1988) at a position 447 bp upstream of *rbc*L was not detected within the region sequenced (597 bp of 5' untranslated). Also, there is no similarity in the *rbc*L 3' untranslated sequences of *C. ellipsoidea* and N1a, except that both are A/T rich (data not shown). However, only 189 bp of 3' untranslated sequence was reported for *C. ellipsoidea*.

The predicted amino acid sequence of the N1a rbcL enzyme indicates a high degree of conservation with other published rbcL sequences, particularly those from *Chlorella ellipsoidea* and *Chlamydomonas reinhardtii* (**Fig. 3**). There are only 12 amino acid differences between the free-living *C. ellipsoidea* and the exsymbiont N1a. This is equivalent to a homology of 97.5%. The difference in rbcL amino acid sequences between these two species of *Chlorella* is intermediate between two species of *Chlamydomonas* (18 amino acid differences between the rbcL's of *C. reinhardtii* and *C. moewusii*; Dron et al. 1982, Yang et al. 1986), two species of wheat (3 amino acid differences between *Triticum* and *Aegilops*; Terachi et al. 1987), and three species of tobacco (2 amino acid differences between *Nicotiana tabacum*, *N. otophora*, and *N. acuminata*; Lin et al. 1986). There are only 80 individual base-pair differences between *Chlorella ellipsoidea* and *Chlorella* N1a at the nucleic acid level, for an identity value of 94.4% (data not shown). Sixty-five of the 80 base substitutions (81%) occur in the 3rd

position of the codon. Though the two *Chlorella* sequences are nearly identical, a comparison with sequences from other organisms indicates that representative sequences of a prokaryote, two algae species, and a land plant are more similar to the sequence of the exsymbiont than to the free-living form (**Fig. 4**).

The ribosomal protein S14 is less conserved (**Fig. 5**). Of the sequences published, the liverwort chloroplast gene is the most homologous to N1a, with 58% identical amino acids; no other algal sequences have been reported for this particular gene. The predicted N1a S14 protein is only slightly closer to *E. coli* than the liverwort chloroplast S14 is to *E. coli* S14 (47% vs. 46% identical residues). *rps*14 codes for a protein of 11,737 Daltons. It is a very basic protein, with 14 arginine residues, 9 lyseines, 3 histidines, and only 8 acidic residues.

The codon usage of these two chloroplast genes shows a strong bias toward codons ending in A or T (**Fig. 6**). Only 19% of the codons in both genes end in G or C; this bias is consistent with other chloroplast genes (Steinmetz and Weil 1987). The only significant difference in codon utilization between *rbc*L and *rps*14 is a very strong bias toward CGT for arginine within *rbc*L, while *rps*14 has no such bias. A comparison with *C. ellipsoidea rbc*L codon utilization disclosed no remarkable differences (data not shown).

These data suggest that *Chlorella* N1a is very similar to *C. ellipsoidea*, despite the differences in overall chloroplast genomic structure (Schuster et al. 1989). If the symbiotic relationship has altered the selective pressure upon the endosymbiont in any major way, then this particular symbiosis has been of insufficient duration to reveal that fact at the DNA sequence level. From limited sequence information, it seems clear that the chloroplast genome of this symbiotic organism has not diverged significantly, at least on a functional level.

Acknowledgements. We wish to thank Marcia Zeigelbein for her skilled technical assistance. This work was supported in part by grant DCB-8417373 from the National Science Foundation and contract XK-5-05073-2 from the Solar Energy Research Institute.

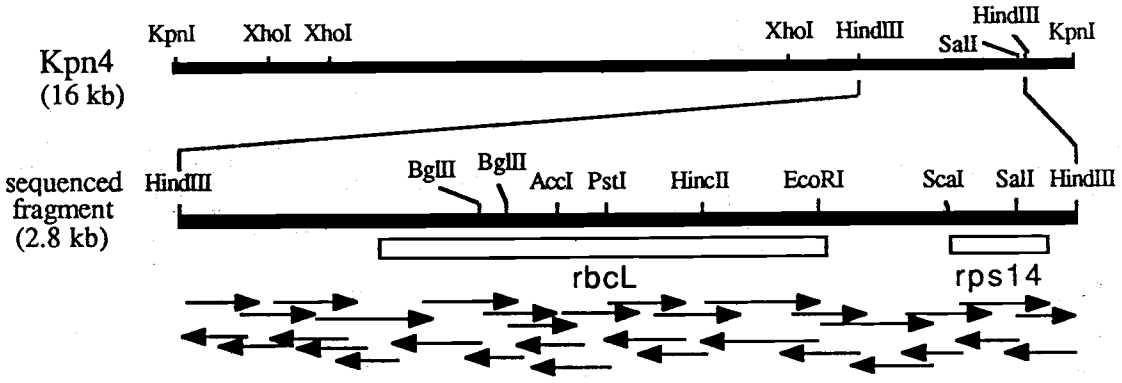

Figure 1. Sequencing strategy of a portion of *Chlorella* N1a chloroplast clone Kpn4; open blocks denote open reading frames 5' to 3', left to right (both genes are coded on the same strand); arrows represent regions sequenced from individual templates

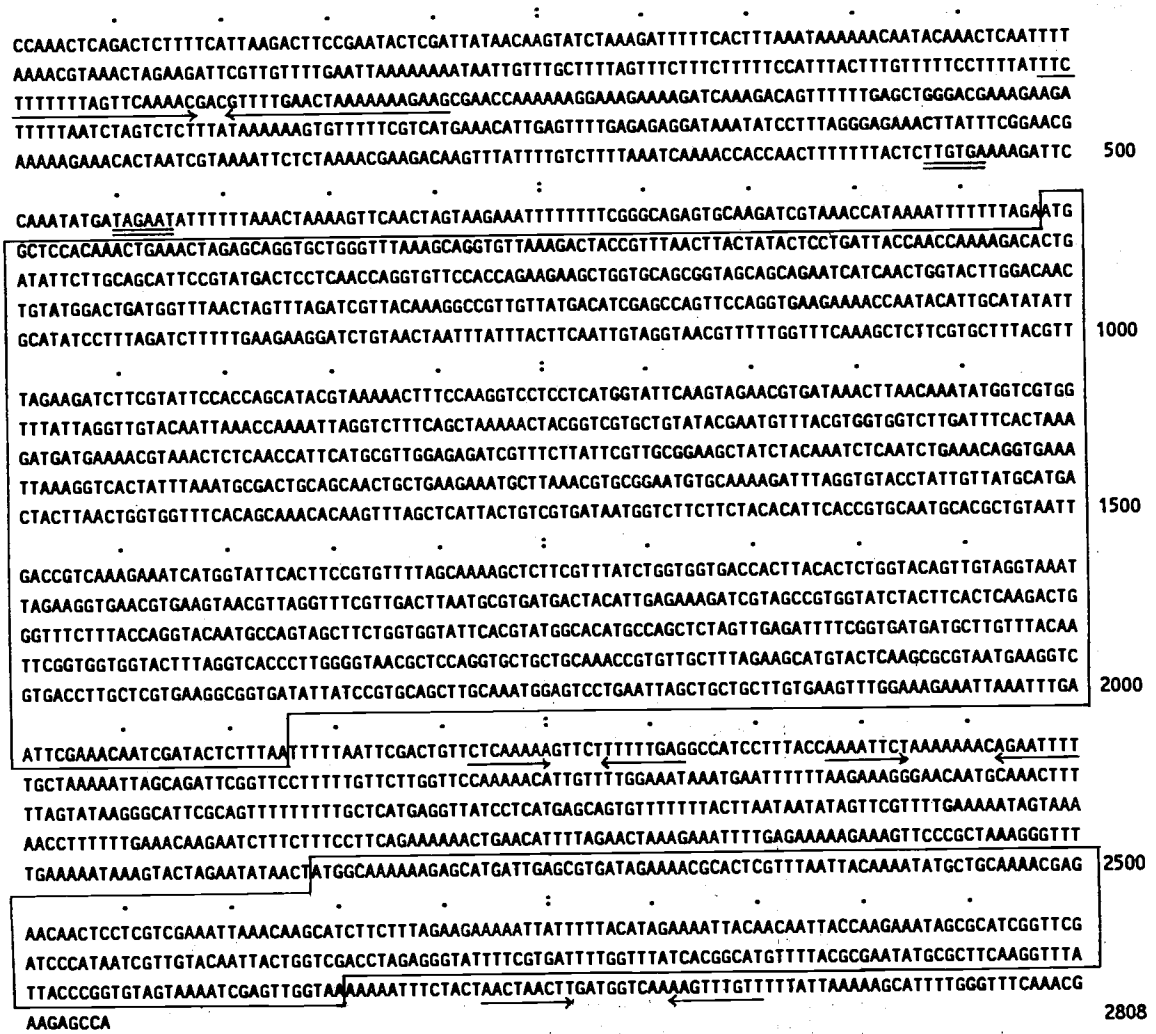

Figure 2. Nucleotide sequence 5' to 3' of a 2.8 kb portion of *Chlorella* N1a chloroplast clone Kpn4; the top box encloses the reading frame of *rbc*L, and the bottom box denotes the *rps*14 reading frame; inverted repeats (possible transcriptional terminators) are underlined; potential promoter regions of *rbc*L are underlined twice

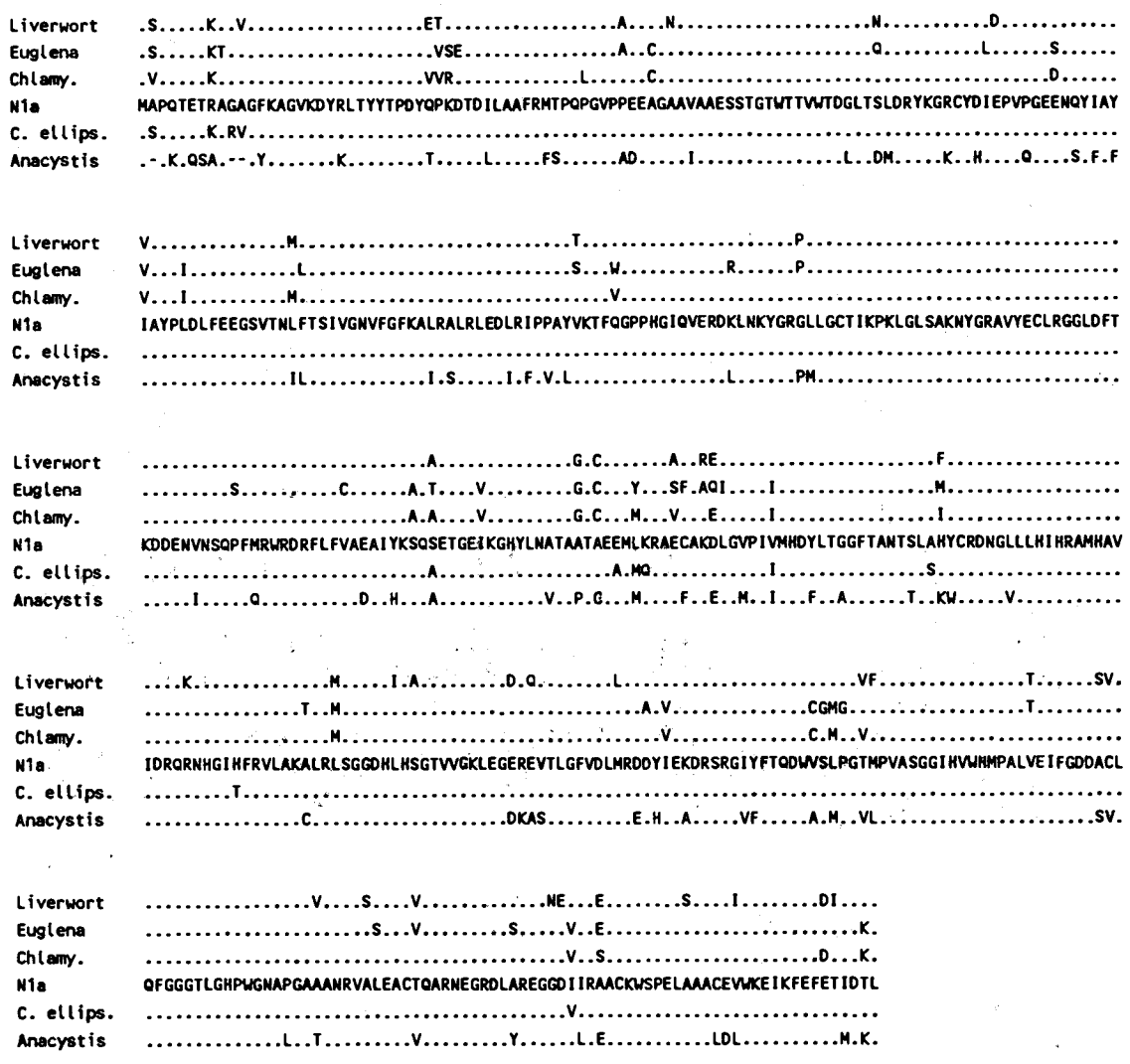

Figure 3. Predicted amino acid sequence for N1a *rbc*L; differences are indicated between N1a and the predicted amino acid sequences of the liverwort *Marchantia polymorpha* (Ohyama et al. 1986), *Euglena gracilis* (Gingrich and Hallick 1985), *Chlamydomonas reinhardtii* (Dron et al. 1982), *Chlorella ellipsoidea* (Yoshinaga et al. 1988), and the cyanobacterium *Anacystis nidulans* (Shinozaki et al. 1983); all dots indicate identical residues, while dashes denote missing residues

	N1a	*C. ellipsoidea*
Liverwort	42	46
Euglena	48	50
C. reinhardtii	31	32
Anacystis	82	84

Figure 4. Number of animo acid differences (out of 475 total) within the *rbc*L protein predicted from DNA sequence between two species of *Chlorella*, N1a and *C. ellipsoidea*, and the liverwort *Marchantia polymorpha, Euglena gracilis, Chlamydomonas reinhardtii*, and the cyanobacterium *Anacystis nidulans* (see **Figure 3** for references)

```
                    :         :         :         :
Liverwort  .....L.Q.EK..QN.EK..KIL.NS.KKK.TET...D.-.WEFQK...S.
N1a        MAKKSMIERDRKRTRLITKYAAKREQLLVEIKQASSLEE-KLFLHRKLQQL  50
E. coli    ...Q..KA.EV..VA.AD..F...AE.KAI.SDVNASD.DRWNAVL...T.

                    :         :         :         :
Liverwort  .....PT.L.R..FL....KANY........L...M.HAC.....T....
N1a        PRNSASVRSHNRCTITGRPRGYFRDFGLSRHVLREYALQGLLPGVVKSSW  100
E. coli    ..D.SPS.QR...RQ....H.FL.K.....IKV..A.MR.QI..LK..--
```

Figure 5. Chloroplast ribosomal protein S14 sequence as predicted from the DNA sequence; *Chlorella* N1a is compared to the liverwort *Marchantia polymorpha* chloroplast S14 (Umesono et al. 1984) and to the *Escherichia coli* homologue S14 (Yaguchi et al. 1983); although the mature *E. coli* protein has no N-terminal methionine, it was included for purposes of comparison; dots show identical residues, dashes represent gaps; numbering is in reference to the two chloroplast sequences

TTT-Phe	5/3	TCT-Ser	8/2	TAT-Tyr	6/3	TGT-Cys	8/1
TTC-Phe	14/0	TCC-Ser	0/1	TAC-Tyr	13/0	TGC-Cys	1/0
TTA-Leu	27/10	TCA-Ser	4/1	TAA-ter	1/1	TGA-ter	0/0
TTG-Leu	0/0	TCG-Ser	0/2	TAG-ter	0/0	TGG-Trp	8/1
CTT-Leu	13/1	CCT-Pro	8/1	CAT-His	4/3	CGT-Arg	28/4
CTC-Leu	0/2	CCC-Pro	0/1	CAC-His	10/0	CGC-Arg	0/2
CTA-Leu	2/0	CCA-Pro	15/1	CAA-Gln	12/5	CGA-Arg	0/3
CTG-Leu	0/0	CCG-Pro	0/0	CAG-Gln	0/0	CGG-Arg	0/1
ATT-Ile	17/4	ACT-Thr	22/2	AAT-Asn	5/2	AGT-Ser	3/1
ATC-Ile	5/0	ACC-Thr	0/0	AAC-Asn	9/0	AGC-Ser	1/2
ATA-Ile	0/0	ACA-Thr	8/2	AAA-Lys	22/8	AGA-Arg	3/4
ATG-Met	9/2	ACG-Thr	1/0	AAG-Lys	0/1	AGG-Arg	0/0
GTT-Val	13/2	GCT-Ala	24/1	GAT-Asp	19/2	GGT-Gly	44/4
GTC-Val	0/1	GCC-Ala	0/0	GAC-Asp	10/0	GGC-Gly	2/0
GTA-Val	14/2	GCA-Ala	19/4	GAA-Glu	29/5	GGA-Gly	1/0
GTG-Val	0/0	GCG-Ala	5/1	GAG-Glu	3/1	GGG-Gly	1/1

Figure 6. Codon utilization of two N1a chloroplast genes; the first number in each column denotes number of occurrences in *rbc*L and the second number refers to *rps*14

REFERENCES

Douglas AE, Huss VAR (1986) On the characteristics and taxonomic position of symbiotic *Chlorella*. Arch Microbiol 145:80-84

Dron M, Rahire M, Rochaix JD (1982) Sequence of the chloroplast DNA region of *Chlamydomonas reinhardii* containing the gene of the large subunit of ribulose bisphosphate carboxylase and parts of its flanking genes. J Mol Biol 162:775-793

Gingrich JC, Hallick RB (1985) The *Euglena gracilis* chloroplast ribulose-1,5-bisphosphate carboxylase gene. II. The spliced mRNA and its product. J Biol Chem 260:16162-16168

Hedberg MF, Huang YS, Hommersand MH (1981) Size of the chloroplast genome in *Codium fragile*. Science 213:445-447

Henikoff S (1984) Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene 28:351-359

Hiratsuka J, Shimada H, Whittier R, Ishibashi T, Sakamoto M, Mori M, Kondo C, Honji Y, Sun CR, Meng BY, Li YQ, Kanno A, Nishizawa Y, Hirai A, Shinozaki K, Sugiura M (1989) 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

Karakashian SJ, Karakashian MW (1965) Evolution and symbiosis in the genus *Chlorella* and related algae. Evolution 19:368-377

Lin CM, Liu ZQ, Kung SD (1986) *Nicotiana* chloroplast genome: X. Correlation between the DNA sequences and the isoelectric focusing patterns of the LS of Rubisco. Plant Mol Biol 6:81-87

Meints RH, Lee K, Van Etten JL (1986) Assembly site of the virus PBCV-1 in a *Chlorella*-like green alga: Ultrastructural studies. Virology 154:240-245

Mews LK (1980) The green hydra symbiosis. III. The biotrophic transport of carbohydrate from alga to animal. Proc R Soc Lond B 209:377-401

Muscatine L, Karakashian SJ, Karakashian MW (1967) Soluble extracellular products of algae symbiotic with a ciliate, a sponge and a mutant hydra. Comp Biochem Physiol 20:1-12

Nichols HW, Bold MC (1965) *Trichosarcina polymorpha* Gen. et Sp. Nov. J Phycol 1:34-38

Ohyama K, Fukuzawa H, Kohchi T, Shirai H, Sano T, Sano S, Umesono K, Shiki Y, Takeuchi M, Chang Z, Aota S, Inokuchi H, Ozeki H (1986) Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. Nature 322:572-574; *rbc*L protein sequence obtained from the Protein Identification Resource databank (National Biomedical Research Foundation)

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

Palmer JD, Stein DB (1986) Conservation of chloroplast genome structure among vascular plants. Curr Genet 10:823-833

Pribnow D (1975) Nucleotide sequence of an RNA polymerase binding site at an early T7 promoter. Proc Natl Acad Sci USA 72:784-788

Reisser W (1984) The taxonomy of green algae endosymbiotic in ciliates and a sponge. Br Phycol J 19:309-318

Reisser W, Wiessner W (1984) Autotrophic eukaryotic freshwater symbionts. In: Linskens HG, Heslop-Harrison J (eds) Cellular interactions, Encyclopedia of plant physiology, vol 17, Springer, Berlin Heidelberg New York, pp 59-74

Reith M, Cattolico RA (1986) Inverted repeat of *Olisthodiscus luteus* chloroplast DNA contains genes for both subunits of ribulose-1,5-bisphosphate carboxylase and the 32,000-dalton Q_B protein: Phylogenetic implications. Proc Natl Acad Sci USA 83:8599-8603

Rosenberg M, Court D (1979) Regulatory sequences involved in the promotion and termination of RNA transcription. Annu Rev Genet 13:319-353

Schuster AM, Waddle JA, Korth K, Meints RH (In press) The chloroplast genome of an exsymbiotic *Chlorella*-like green algae. Plant Mol Biol

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

Shinozaki K, Sugiura M (1983) The gene for the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase is located close to the gene for the large subunit in the cyanobacterium *Anacystis nidulans* 6301. Nucleic Acids Res 11:6957-6964

Shinozaki K, Yamada C, Takahata N, Sugiura M (1983) Molecular cloning and sequence analysis of the cyanobacterial gene for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase. Proc Natl Acad Sci USA 80:4050-4054

Steinmetz A, Weil JH (1987) Protein synthesis in chloroplasts. In: A. Marcus (ed), The biochemistry of plants, Academic Press

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

Terachi T, Ogihara Y, Tsunewaki K (1987) The molecular basis of genetic diversity among cytoplasms of *Triticum* and *Aegilops*. VI. Complete nucleotide sequences of the *rbcL* genes encoding H- and L-type Rubisco large subunits in common wheat and *Ae. crassa* 4x. Jpn J Genet 62:375-387

Trench RK (1979) The cell biology of plant-animal symbiosis. Annu Rev Plant Physiol 30:485-531

Tymms MJ, Schweiger HG (1985) Tandemly repeated nonribosomal DNA sequences in the chloroplast genome of an *Acetabularia mediterranea* strain. Proc Natl Acad Sci USA 82:1706-1710

Umesono K, Inokuchi H, Ohyama K, Ozeki H (1984) Nucleotide sequence of *Marchantia polymorpha* chloroplast DNA: a region possibly encoding three tRNAs and three proteins including a homologue of *E. coli* ribosomal protein S14. Nucleic Acids Res 12:9551-9565

Van Etten JL, Schuster AM, Meints RH (1988) Viruses of eukaryotic *Chlorella*-like algae. In: Koltin Y, Leibowitz MJ (eds) Viruses of fungi and simple eukaryotes. Marcel Dekker, New York and Basel, pp. 411-428

Vieira J, Messing J (1987) Production of single-stranded plasmid DNA. Methods Enzymol 153:3-11

Weil JH (1987) Organization and expression of the chloroplast genome. Plant Sci 49:149-157

Wilkinson CR (1980) Nutrient translocation from green algal symbionts to the freshwater sponge *Ephydatia fluviatilis*. Hydrobiol 75:241-250

Wolfe KH, Li WH, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci USA 84:9054-9058

Wolfe KH (1989) Compilation of sequences of protein-coding genes in chloroplast DNA including cyanelle and cyanobacterial homologues. Plant Mol Biol Rep 7:30-48

Yaguchi M, Roy C, Reithmeier RAF, Wittmann-Liebold B, Wittmann HG (1983) The primary structure of protein S14 from the small ribosomal subunit of *Escherichia coli*. FEBS Lett 154:21-30

Yamada T (1982) Isolation and characterization of chloroplast DNA from *Chlorella ellipsoidea*. Plant Physiol 70:92-96

Yang RCA, Dove M, Seligy VL, Lemieux C, Turmel M, Narang SA (1986) Complete nucleotide sequence and mRNA-mapping of the large subunit gene of ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) from *Chlamydomonas moewusii*. Gene 50:259-270

Yoshinaga K, Ohta T, Suzuki Y, Sugiura M (1988) *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