Mapping and sequencing of an actively transcribed *Euglena gracilis* chloroplast gene (cccA) homologous to the *Arabidopsis thaliana* nuclear gene cs(ch-42)*

Bernard Orsat, Amparo Monfort*, Philippe Chatellard and Erhard Stutz

Laboratoire de Biochimie végétale, Université de Neuchâtel, Chantemerle 18, CH-2000 Neuchâtel, Switzerland

We mapped and sequenced a novel chloroplast gene encoding a protein (348 amino acids) which shows a high sequence identity with both the decoded nuclear cs(ch-42) gene product of *Arabidopsis thaliana*, and the C-terminal half of the decoded 'crtA' gene product of *Rhodobacter capsulatus*. The chloroplast gene (cccA) is split (two exons) and transcribed into a stable mRNA of about 1200 nucleotides. The putative protein may be involved in the biosynthesis of photosynthetic pigments.


1. INTRODUCTION

During our studies on structure and function of the *E. gracilis* chloroplast genome we mapped and sequenced a DNA segment with a split ORF coding for a protein of 348 amino acids. Using the FASTA service offered by EMBL, Heidelberg, we found that the amino acid sequence was very similar to a recently published decoded sequence of the nuclear gene cs(ch-42) of *A. thaliana* [1]. According to this study cs is a light-regulated gene encoding a chloroplast protein which is imported into chloroplasts as shown by in vitro experiments. In case of *E. gracilis* the corresponding gene is located and expressed within the chloroplast. We propose to call this gene cccA. It represents a chloroplast gene homologous to the *A. thaliana* nuclear cs gene which most likely is involved in chloroplast pigment biosynthesis.

2. MATERIALS AND METHODS

The previously described DNA fragment Bgl Z [2] was cloned into the BamHI site of the vector Bluescript KSII- (+EGKS-Z). We subcloned a HindIII-BglII fragment (2935 bp) by digestion of the construct pEGKS-Z with HindIII. Fragments were separated on agarose gels (1%) and fragments of the appropriate length (the HindIII map of Bgl Z is known) were eluted from the gel (Biotrap, Schleicher & Schuell) and religated. The clone pEGKS-2.9 with the 2935 bp insert was totally sequenced (B. Orsat, Ph.D. Thesis, Neuchâtel, 1992) following standard protocols (STRATEGENE). Overlapping smaller fragments were generated by cutting with HindIII and KpnI, followed by selective degradation with exonuclease III (BRL). Blunt ending with mung bean nuclease (Promega) and religation. The 2935 bp insert carries at one end the 5'-terminal part of a tRNA-Leu (CAA) gene which is cut by HindIII as published [3].

Chloroplast RNA was isolated and purified as published [4]. Northern hybridization was done in 5 × SSPE based solutions, 50% for formamide at 42°C. Filters (Schleicher & Schuell, BAB3) were washed twice with buffer 2 × SSC, 0.1% SDS, 42°C for 15 min.

Nucleotide and amino acid sequence data were analysed using the sequence analysis software package of Genetics Computer Group (GCG), Wisconsin.

3. RESULTS AND DISCUSSION

We show in Fig. 1(I,II,III) the position of the cccA gene on the Bgl Z fragment which was previously mapped on the chloroplast genome [2]. The cccA gene is situated between the trnL (CAA) [3] and the psbD gene-sharing transcription polarity with the trnL, but not with the psbD gene. The coding part (ORF348) is split into exon 1 and 2 with 21 and 327 codons, respectively. The intron (332 bp) has canonical 5' and 3' termini and features of a chloroplast group II intron [5].

The *Euglena* cccA gene encodes a protein (Mw = 39,307) having a high sequence identity (70%) with the cs nuclear gene of *A. thaliana* [1] and the C-terminal half of the 'crtA' gene of *R. capsulatus* (Fig. 2) which according to a personal note (G.A. Armstrong, ETHZ, Zurich) and contrary to the published data [6] represents an independent ORF (bchI). Accordingly, we used in the alignment study only the C-terminal part of the...
Fig. 1. Mapping of ORF 348. I: Bgl II (3647 bp); II: Bgl II-HindIII fragment (2935 bp); III: subfragment of II (1681 bp, arbitrary cut). B, H, Ha, X respectively, are Bgl II, HindIII, and XhoI cleavage sites. D, J are HaeII fragments as published [9]. Note, however, that the two fragments D and J are separated by a small HaeII fragment (495 bp) not noted previously. a,b represent, respectively, 5'-terminal part of trnL (CAA) and N-terminal part of exon I of pshD. Exons of ccsA interrupted by intron. □ polarity of transcription. ori, origin of DNA replication [10].

Fig. 2. Amino acid sequence alignment of decoded genes. E: E. gracilis chloroplast ccsA, A: A. thaliana cs, R: R. capsulatus C-terminal part of 'cctA' (helix, 350 codons) starting with MTT... as suggested by C.A. Armstrong (see text). (+) exact matches, (-) conservative matches across all sequences; 1 and 2 mark intron positions in nuclear cs gene; 3 marks intron position in ccsA gene. (-) gap.
published sequence starting with MTTA. We notice large domains of exact and conservative matches across all three sequences what strongly suggests that these proteins not only have a common evolutionary origin but most likely have an equivalent function.

The *A. thaliana* *cs* protein sequence has a long N-terminal part which qualifies as transit peptide [1] and therefore has no equivalent sequence in the chloroplast and bacterial counterpart. According to the result of Fig. 2, a first conservative domain (RPV.,) starts at position 85 (line A), i.e. the transit sequence most likely ends upstream of that domain and not at position 93 (line A) as tentatively assumed [1]. In that context it was of interest to compare the hydropathy plot of the *csA* N-terminal part (103 amino acids) with the equivalent sequence of the *cs* gene (Fig. 3). Certainly the two profiles are congruent from position 80 on (*A. thaliana*) strongly suggesting that the processed *cs*-protein starts in that region.

The *cssA* gene is transcribed in light grown cells, as shown in Northern hybridization experiments (Fig. 4). The stable transcript is about 1200 nucleotides long. In addition to the major band a precursor of about 1430 nucleotides interacts with the probe and a very faint band around 400 nucleotides can be detected on the radiograph. From previous studies (A. Monfort, Ph.D. thesis, Neuchâtel, 1990) we know that the *trnL* gene is co-transcribed with upstream elements and all indications are that the *cssA* gene and the downstream *trnL* gene are part of a primary transcript which undergoes several steps of processing including the splicing of exon 1 with exon 2.

The function of the *cssA* gene is presently unknown. Mutations in the *cs* gene, or, e.g. a T-DNA insertion in the 3'-end of the coding part lead to loss of chloroplast pigments (pale mutant) [1]. The same holds for mutations in the C-terminal part of the *'erlA* gene (beh) of *R. capsulatus*: such mutants show loss of bacteriochlorophyll accumulation [7,8].

Considering the close structural relationship between the Euglena chloroplast *cssA* gene with both the plant and bacterial counterparts we postulate that the chloroplast gene is also involved in chloroplast pigment biosynthesis. If such is the case then *cssA* is the first identified chloroplast gene participating in chlorophyll accumulation.

The *cssA* gene represents another example for genes transferred to the nuclear DNA of higher plants but retained in the algal chloroplast genome as was shown for the *tufA* gene [11,12]. Higher plant plastids also show differences in gene composition. It was, e.g., reported that the tobacco and rice chloroplast genome contain the *rps16* but lack the *rpl21* gene while the opposite is true for *Marchantia polymorpha* [13].
Acknowledgements: We are grateful to Drs. G.A. Armstrong, ETHZ, Zurich and C. Konez, Max-Planck-Institute, Cologne for valuable information about unpublished data. This work represents part of the Ph.D. thesis of B.O. and receives support from the Fonds national suisse de la recherche scientifique, to E.S.

REFERENCES