ELSEVIER

Biochimica et Biophysica Acta 1365 (1998) 541–544

BIOCHIMICA ET BIOPHYSICA ACTA
BBA

Short sequence-paper

Extraordinary features in the *Chlamydomonas reinhardtii* chloroplast genome: (1) rps2 as part of a large open reading frame; (2) a *C. reinhardtii* specific repeat sequence

Stefan Leu *

The Doris and Bertie Black Center for Bioenergetics, Ben Gurion University of the Negev, Beer Sheva 84105, Israel

Received 17 March 1998; revised 11 May 1998; accepted 13 May 1998

Abstract

We have determined the DNA sequence of the 3574-bp chloroplast DNA fragment of *Chlamydomonas reinhardtii* formed by the overlap of *Bam*HI fragment 3 and *Eco*RI fragment 5. This sequence encodes most of rps18 and orf570, an unidentified open reading frame that contains a 150 amino acid domain with high homology to the N-terminal part of 30 S ribosomal protein S2 of other chloroplast, cyanobacterial and bacterial genomes. Between these two sequences lies a highly repetitive sequence element of 500 bp, that is composed of multiple direct and inverted repeat sequences that occur in rearranged, but highly conserved form in at least 36 locations in the *C. reinhardtii* chloroplast genome. Among the conserved repeat sequences in the *C. reinhardtii* chloroplast genome we identified the borders of the inverted repeats near atpB and rps4. This might indicate that the conserved sequence elements are remainders of gene rearrangements in the chloroplast genome that occurred by relocations of the inverted repeats.

The green alga *Chlamydomonas reinhardtii* has become an ideal model organism for the study of photosynthesis. *C. reinhardtii* can be easily cultivated, the components of its photosynthetic apparatus are very similar to those of higher plants and many mutants affecting the photosynthetic apparatus have been isolated and genetically mapped. Since *C. reinhardtii* can grow heterotrophically, it is possible to inactivate or delete genes encoding components of the photosynthetic apparatus [1]. *C. reinhardtii* can be transformed in the nuclear and chloroplast genomes [2,3] and suitable selection markers and transformation techniques for both compartments are available [4–7]. The chloroplast DNA sequence of *C. reinhardtii* has not yet been completely determined. About 70 genes have been mapped on the circular 197-kbp genome [8]. Including the most recent additions to the gene bank around 90% of its sequence has now been determined. The arrangement of the genes in the plastid genome of *C. reinhardtii* deviates from that in higher plants. With the exception of tufA, orf2971 and tscA the chloroplast DNAs of *C. reinhardtii* and higher plants contain essentially the same genes.

We have determined the DNA sequence of the 3574-bp fragment formed by the overlap of *Bam*HI fragment 3 and *Eco*RI fragment 5 of the *C. reinhardtii* chloroplast genome[1]. This fragment was ex-

* Corresponding author. Fax: +972 (7) 6472 890; E-mail: stefen@netvision.net.il

[1] This sequence appears in the GeneBank under the accession nr. Y16473.

0005-2728/98/$19.00
PII: S0005-2728(98)00107-8

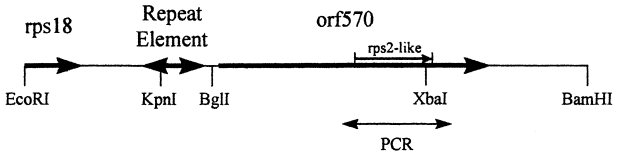

Fig. 1. Physical map of the sequenced DNA fragment, with the *Eco*RI site to the left and the *Bam*HI site to the right. The locations of the rps2 like sequence within orf570 and of the PCR product used to confirm the sequence are indicated.

cised with *Bam*HI and *Eco*RI from plasmid p-62 containing the subcloned *Eco*RI fragment 5 (obtained from E. Harris, the *Chlamydomonas* Genetics Center, Duke University) and subcloned into pBluescript. Further subclones were produced by digestion with *Hin*dIII, *Kpn*I, *Pml*I, *Ase*I and *Ssp*I. The sequences of these fragments were determined in both directions by the dideoxy chain termination method using modified T7 DNA polymerase and universal primers as well as sequence specific oligonucleotide primers. Analysis of the resulting DNA fragment using the GCG software package (version 9.1, Genetics Computer Group, Madison, WI, USA) revealed at least two open reading frames. BLAST search of the gene banks for homologies at the nucleotide or amino acid sequence level revealed three domains of interest (Fig. 1).

The C. reinhardtii rps18 gene: The first 377 bp (starting at the *Eco*RI site) form an open reading frame that is homologous to 30S ribosomal protein S18 of chloroplasts and bacteria. This sequence corresponds to the end of a recently deposited GenBank sequence containing rps9, ycf3, ycf4 and rps18 (accession nr. Y13655).

A C. reinhardtii specific highly repetitive chloroplast DNA element: Peptide fragments of 10–21 amino acid residues, deduced by translation of any of the six reading frames of bases 800–1200 showed 90–100% homology to parts of hypothetical open reading frames of the *C. reinhardtii* chloroplast genome, orf112 in the petA-petD intergenic region (accession nr. X78133), orf2 upstream of the atpB sequence (accession nr. M13704) and orf101 in the trnR-chlB intergenic region (accession nr. P37825). BLAST search for homologies of the domain from bases 750–1250 at the nucleotide level revealed about 20 highly homologous genebank entries from the *C. reinhardtii* chloroplast genome. Further analysis revealed that sequences homologous to this 500-bp fragment occur in at least 36 intergenic locations in the *C. reinhardtii* chloroplast genome (Table 1). All these homologous domains are composed of multiple identical or very highly conserved sequence fragments of 12 to more than 100 bases in length which form multiple direct or inverted repeat structures of various lengths and in various forms of rearrangement.

Several of the homologous domains were previously described as repetitive elements with extensive secondary structure [9]. Without exception these domains lie in intergenic regions, normally near promotors or transcription stops. In the case of the well mapped P-A promotor sequence [10], these repeat elements are located between and around the −35 and −10 sequences but those promotor consensus sequences are not part of the conserved repeat elements. It has been speculated that the highly structured repeat element within the 16S–23S rDNA spacer region is involved in expansion-contraction of this domain that among others is responsible for the large differences in the sizes of *Chlamydomonas* chloroplast genomes [9]. Involvement of the rDNA spacer region and the borders of the inverted repeats in rearrangement of *Chlamydomonas* chloroplast genomes could be demonstrated by comparing the *C. moewusii* and *C. pitschmannii* chloroplast genomes [11]. A conserved 200-bp sequence forming the borders of the inverted repeat at the end of atpB and rps4, respectively, was identified among these conserved *C. reinhardtii* repeat elements. This could confirm that rearrangements in the *C. reinhardtii* chloroplast genome occurred by relocation of the inverted repeats. The conserved repeat elements could thus be the remainders of such rearrangement events. On the other hand, such a repeat element is also found directly upstream of one of the two copies of the transposon-like Wendy sequences [12] and might have been dispersed across the chloroplast genome by transposition of this sequence.

A rps2-like sequence is part of a long unidentified reading frame: The sequence from base nr. 1245 to base nr. 2954 forms an open reading frame of 570 amino acids (orf570) with extraordinary features: The first 300 amino acids of this open reading frame show some homology (25–30% identity) to the α subunit of chloroplast ATP synthase, heatshock protein 60 and other protein sequences but cannot be

clearly identified. The next 150 amino acids show high homology to the N-terminal half of 30S ribosomal protein S2 of other chloroplast and bacterial genomes (Fig. 2). The C-terminal 106 amino acids have only very low homology to the C-termini of some chloroplast S2 30S ribosomal proteins.

To confirm that this extraordinary phenomenon is not the result of a mutation introduced during subcloning and maintenance of this DNA fragment, we amplified the critical sequence domains by PCR from total *C. reinhardtii* DNA and determined the DNA sequences of the PCR products (Fig. 1). The sequence of the PCR products was identical to the sequence of the plasmid DNA, which confirmed that indeed the N-terminus of rps2 is integrated into another, unidentified open reading frame. A similar phenomenon was observed with rps3, that is part of orf712 in the chloroplast genome of *C. rein-*

Table 1
Locations of conserved repeat sequences in the chloroplast genome of *C. reinhardtii*

	Coordinates	Location	orfs	Acc. Nr
1	2.1–2.5	Upstream of petA	–	D01036
2	4.2–5.2	petA-petD intergenic region	orf112	X78133
3	5.8–6.4	P-A promotor	–	M14865/X78133
4	8–8.4	petD-chlB intergenic region	orf101	L13782
5*	14.2–14.3	Downstream of tufA	–	X52257
6	20.5–20.6	Between trnL and petB	–	X62905
7	21.6–21.7	petB-frxC intergenic region	–	X62905
8	23.3–23.7	Downstream of frxC	–	X62905
9	34.8–35	Downstream of rps4, border of IR	–	U17357
10	37.6–37.8	Upstream of 16S rRNA gene	–	J01395
11	40–41.5	rDNA spacer region	–	J01395
12	45.3–45.4	Upstream of 5S rRNA gene	–	X03271
13	54.8–55	Downstream of psbA exon1	–	X01424
14	57.8–58.4	Downstream of ycf12 and trnS	orf50/orf52	U40346
15	61.2–61.5	Between atpE and rps7	–	X53977
16	65.7–66.5	Between ycf9 and ycf5	–	U81552
17	70.3–70.9	Upstream of psaA exon 3	–	X05847
18	76–76.2	Downstream of psbH	–	L13303
19	77.8–78.8	Downstream of psbN	–	L13303
20	82.6–82.9	Upstream of trnD	–	L26265
21	89.9–90.5	Between orf570 and rps18, this work	–	Y16473
22	92.6–93	Between rps9 and ycf4	–	Y13655
23	102.4–102.6	Between psbF and psbL	–	X66250
24	103.4–103.6	Between psbL and petG	–	X66250
25*	121.4–121.6	Between rpcL and atpA	–	J01399
26	131.4–131.7	Upstream of tscA	–	M61198
27	135–135.2	Downstream of psbA exon 1	–	X01424
28	143.9–144	Upstream of 5S rRNA gene	–	X03271
29	148.5–150	rDNA spacer region	–	J01395
30	151.6–151.8	Upstream of 16S rRNA gene	–	J01395
31	155.2–155.5	Downstream of atpB	–	M13704
32	157.5–157.9	Between atpB and orf1995	orf2	M13704
33	163.7–163.9	Upstream of orf1995	–	X92726
34	168.5–168.9	Between psbJ and psaA exon 2	–	AF025877
35	171.8–172.4	Between psbD and orf2971	–	U62943
36	195.6–195.8	Upstream of Wendy1	–	Z38069

The coordinates given are in kbp from the origin (defined by the *Bam*HI site between *Bam*HI fragments 1 and 7 [1]). 'orfs' are hypothetical open reading frames probably accidentally formed by these repeat elements that do not have any homologs in the gene banks. Sequences marked with an asterisk were not identified by BLAST search using the repeat element identified in this work, but were identified by BLAST search using the P-A promotor sequence.

```
      305                                                                                  384
odo     ~~~~MADITL AQLLEAGVHF GHKAYRWNPK MFPYIYTERN NI........ ......HILD LVQSAQLLKE ANSYLQSAAE
syn     ~~~~MPVVSL ADLLESGVHF GHQTRRWNPR MDQYIYTARN GV........ ......HIID LVQTAQLMED AYEYVRSSTI
spinach MTRRYWNINL EEMMEAGVHF GHGTRKWNPR MSPYISAKCK GI........ ......HIIN LTRTARFLSE ACDLVFDASS
chlor   ~~~MLQKVKL EDMIQSGMHF GHSTREWNPR MAPYIYGERN GR........ ......HILD LVQTYYLLNK VLAFLEEQAA
eugl    ~~~~~~MITV EKMLNSSVHL GHKVKQWNPR MRIYIYGERK GL........ ......HIID LLQTIVCLKK ACNFLIRSVR
orf570  QKKAFVRSSL RQMLKSGMHY GEKAIKCNAR MKNYVWTRKK GTDTKVEARP LIKKGRNLIN LLKTRRCLTK ALAQLTKYAA
                 *  *******I* I*   **I** I  I**  ** *               **** I**I   * * *   *

      385                                                                          454
odo     KGQTFLFIGT KPQASALIAQ EAK.RCNSYY VNHRWLGGML TNWVTLKSRI ARLKTLEQEE ADQVFNLLPK
syn     NGKKFLFVGT KRQAAGIISQ EAQ.RCGAHY VNQRWLGGML TNWETIRKRV DRLKELEALE ASGGIDRRGK
spinach RGKQFLIVGT KNKAADSVAR AAI.RARCHY VNKKWLGGML TNWSTTETRL HKFRDLRMEQ TAGRLARLPK
chlor   QGKTFLFVGT KQQAAPLIAK TAL.ACESFY VNQRWLGGML TNWRTIQKSL RKLQEYRRAE ERGDWNLLKK
eugl    KGKRALFVCT KRFFSILTQK IAL.KCNSFF VTKRWLGGIL TNWITIKNCI NKLKQLSKQK EKHH.NLLTK
orf570  KGKTFLFVGT KKAASGLVAR AALFSKKAFF VNTRWLGGML TNWKTILKSI SKIRPILKEK QMIIKDILEK
        *I* *I*I*I I* **    *  I     * * I***IIII*I III I*   *  *** *   *          I
```

Fig. 2. Alignment of the rps2 like part of orf570 with the N-termini of rps2 sequences from other photosynthetic organisms. Residues conserved in all species are indicated by 'I'. Residues with only conserved replacements, or that are identical with one exception, are indicated by '*'. The aligned rps2 sequences are: odo, *Odontella sinensis* (p49490); syn, *Synechocystis* sp. (p74071); spinach (p08242); chlor, *Chlorella vulgaris* (p56351); eugl, *Euglena gracilis* (p30389).

hardtii [13]. The C-terminus of rps2 has been mapped to the neighbouring chloroplast *Bam*HI fragment 6 of *C. reinhardtii* chloroplast DNA using a probe derived from *C. eugametos* [8]. This raises the question whether rps2 in its entity is encoded in a different location, and orf570 has a different function, or whether the rps2 like sequence in orf570 is one exon of the actual rps2 gene.

We expected to detect the chloroplast atpF gene in the sequence of this 3.5-kb fragment, as it had been mapped by heterologous hybridization to this location [14]. As no sequence homologous to this gene could be detected in our sequence, the atpF gene appears to be located in another not yet sequenced location of the *C. reinhardtii* chloroplast genome.

This work was supported by the Doris and Bertie Black Center for Bioenergetics in Life Sciences.

References

[1] E. Harris (1989), The Chlamydomonas Sourcebook: A Comprehensive Guide to Biology and Laboratory Use. Academic Press, San Diego, CA.
[2] K. Kindle, Proc. Natl. Acad. Sci. USA 87 (1990) 1228–1232.
[3] J.E. Boynton, N.W. Gillham, E.H. Harris, J.P. Hosler, A.M. Johnson, A.R. Jones, B.L. Randolph-Anderson, D. Robertson, T.M. Klein, K.B. Shark, J.C. Sanford, Science 240 (1988) 1534–1538.
[4] S.M. Newman, J.E. Boynton, N.W. Gillham, B.L. Randolph-Anderson, A.M. Johnson, E.H. Harris, Genetics 126 (1990) 875–888.
[5] M. Goldschmidt-Clermont, Nucleic Acids Res. 19 (1991) 4083–4089.
[6] E. Fernandez, R. Schnell, L.P.W. Ranum, S.C. Hussey, C.D. Silflow, P.A. Lefebvre, Proc. Natl. Acad. Sci. USA 86 (1989) 6449–6453.
[7] K.L. Kindle, K.L. Richards, D.B. Stern, Proc. Natl. Acad. Sci. USA 88 (1991) 1721–1725.
[8] E. Boudreau, C. Otis, M. Turmel, Plant Mol. Biol. 24 (1994) 585–602.
[9] M. Schneider, J.D. Rochaix, Plant Mol. Biol. 6 (1986) 265–270.
[10] R.J. Thompson, G. Mosig, Cell 48 (1987) 281–287.
[11] E. Boudreau, M. Turmel, Plant Mol. Biol. 27 (1995) 351–364.
[12] W.H. Fan, M.A. Woelfle, G. Mosig, Plant Mol. Biol. 29 (1995) 63–80.
[13] X.-Q. Liu, C. Huang, H. Xu, FEBS Lett. 336 (1983) 225–230.
[14] J. Woessner, N.W. Gillham, J.E. Boynton, Plant Mol. Biol. 8 (1987) 151–158.