PARTS OF THE SEQUENCE BETWEEN THE COMPLETE rRNA OPERONS ARE REPEATED ON EITHER SIDE OF THE EXTRA 16 S rRNA GENE IN CHLOROPLAST DNA OF *EUGLENA GRACILIS* STRAIN Z

Barbara KOLLER and Hajo DELIUS
European Molecular Biology Laboratory, Postfach 10.2209, 69-Heidelberg, FRG

Received 10 February 1982

1. Introduction

The rRNA genes in the chloroplast DNA (cpDNA) of *Euglena gracilis* strain Z are arranged in 3 tandemly repeated operons [1–3]. With respect to the direction of rRNA transcription they will be referred to as the first, second and third complete operon. An extra 16 S rRNA gene was found upstream of the first complete operon [4]. It was described that the extra 16 S rRNA gene has the same size and is coded on the same strand as the other 16 S rRNA genes. The position of a small inverted repeat sequence between this gene and the first complete operon was determined by heteroduplex analysis and it was established that no 16 S–23 S spacer was associated with this 16 S rRNA sequence [5].

To determine whether sequences from the region between the 3′-end of the 23 S rRNA gene and the 5′-end of the 16 S rRNA gene of the complete operons may be repeated in connection with the extra 16 S rRNA gene, we analysed heteroduplexes formed between restriction enzyme fragments *Bam*HI E, or D, and *Eco*RI B or *Bam*HI B. It was found that a sequence of ~330 nucleotides which is located downstream of the 3′-end of the 23 S rRNA gene and which includes part of the 5 S rRNA gene in the complete operons [6] is repeated downstream of the 3′-end of the extra 16 S rRNA gene. In addition, ~150 nucleotides near the 5′-ends of the extra and the normal 16 S rRNA genes are also homologous. It could be demonstrated that the entire sequence of 1200 nucleotides between the 3′-end of the 23 S rRNA genes of the first and second operon and the 5′-end of the 16 S rRNA genes of the second and third operon is repeated in the *Bam*HI D fragment, upstream of the 5′-end of the 16 S rRNA gene of the first operon. This sequence includes a modified 5 S rRNA gene sequence.

2. Materials and methods

2.1. *Preparation of cpDNA from* Euglena *chloroplasts*

Euglena gracilis strain Z was grown in Difco *Euglena* Broth. The cells of a 5-day-old culture (200 ml) were pelleted at 2000 rev./min in a Beckman J6 centrifuge, suspended in 100 ml medium A (300 mM sorbitol, 50 mM Tris–HCl (pH 8.0), 3 mM EDTA) and centrifuged again. The pellet was resuspended in 35 ml medium A and homogenized with a French pressure cell at 1200 lb/in^2. The homogenate was layered onto a Percoll gradient and centrifuged at 3000 rev./min. The chloroplast layer was resuspended and lysed with 4% Triton X-100. After pelleting the cell debris the supernatant was made 1% with respect to SDS, passed over a column of Sepharose C1-2B (4 cm × 8 cm), which was equilibrated with 10 mM Tris–HCl (pH 7.4), 1 mM EDTA. The cpDNA was purified on CsCl–ethidium bromide density gradients, as was described for the isolation of *Vicia faba* cpDNA [7].

2.2. *Preparation of heteroduplexes*

Isolation of DNA fragments from *Euglena* chloroplasts was done by electroelution from agarose gels [8]. The preparation of heteroduplexes, the electron microscopy and the length measurements were done as described before [7,9,10].

3. Results

Maps of restriction enzyme cleavage sites of *Euglena gracilis* strain Z cpDNA indicate that the sequence separating the complete rRNA operons between the 3′-end of the 23 S rRNA genes and the 5′-end of the 16 S rRNA genes is repeated upstream of the 16 S rRNA gene of the first operon, up to the *Eco*RI site

Published by Elsevier Biomedical Press
00145793/82/0000–0000/$02.75

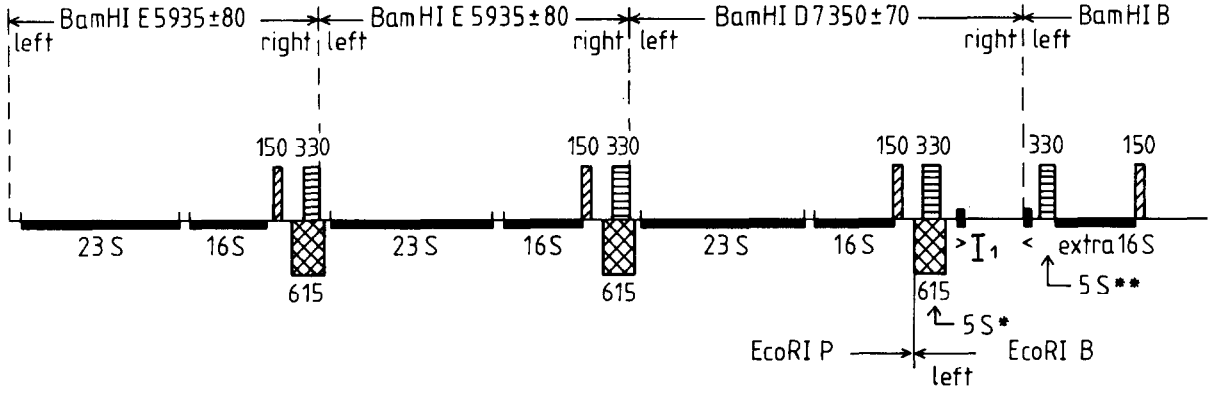


Fig.1. Schematic drawing and length measurements of the fragments *Bam*HI B, D, and E, and *Eco*RI B from *Euglena* cpDNA. The black bars indicate the positions of the rRNA genes, and of the inverted repeat I1. The areas marked ▨ (615 nucleotides) represent homologies between the sequences separating the first, second and third operon and the sequence upstream of the first operon, as described in fig.2. The areas marked ▤ (330 nucleotides) and ▨ (150 nucleotides) indicate homologies between the inter-operon sequence and sequences on either side of the extra 16 S rRNA gene as in fig.3. 5 S* indicates the position of a modified 5 S rRNA sequence. 5 S** the position of a possible partial 5 S rRNA sequence. All measurements are given in numbers of nucleotides. Double-stranded T7 DNA (39 400 basepairs [15]) and single-standed M13 DNA (6400 nucleotides [16]) were used as internal length standards.

which generates the P and the B fragments [11]. The extension of this homology into the *Eco*RI B fragment was not determined. It was also not clear whether part of this sequence may be repeated upstream of the extra 16 S rRNA gene. To search for additional repeated sequences apart from the 16 S rRNA the restriction enzyme fragments *Bam*HI B, D, E and *Eco*RI B were isolated, denatured, and reannealed in various combinations. Fig.1 shows the positions of the fragments in relation to the rRNA genes and to inverted repeat I1 [5]. The left and right ends of the fragments were defined as indicated in this drawing. Fig.2a shows a heteroduplex between *Eco*RI B and *Bam*HI E, and fig.2b its interpretation in a tracing. Both, the left and the right end of the *Bam*HI E fragment (BE1 and BEr) reannealed to the left end of the *Eco*RI B fragment (EB1), thereby forming a circle. This implies that the sequence between the operons (located at both ends of *Bam*HI E) is almost completely homologous with the sequence upstream of the *Eco*RI site which separates the P and the B fragments. The second duplex region is formed by the two 16 S rRNA genes. The 23 S rRNA gene in *Bam*HI E and the inverted repeat I1 in *Eco*RI B remain single-stranded. Length measurements on this heteroduplex are shown in fig.2c. The double-stranded region which is formed by the left end of the *Eco*RI B fragment and both, the left and right end of the *Bam*HI E fragment has a length of 615 nucleotides. Since the *Bam*HI site, which generates the E fragments is situated within the sequence of the 5 S rRNA gene [6], the structure of this heteroduplex shows the presence of another 5 S rRNA gene ~600 nucleotides upstream of the *Eco*RI site. In the B fragment this gene must have a slightly modified sequence, since no *Bam*HI site is found at this position. The presence of this sequence explains the faint hybridization of 5 S rRNA to the *Eco*RI B fragment in [12]. The *Eco*RI site is ~400–500 nucleotides upstream of the 5′-end of the 16 S rRNA sequence in the first operon. This region, within the *Eco*RI P fragment, is identical for all 3 operons [11]. Therefore, the entire sequence of 1100–1200 nucleotides between the complete operons is repeated upstream of the 16 S rRNA gene of the first operon, with some slight modifications.

The duplex formed between the 16 S rRNA genes has a length of 1655 nucleotides which is ~150 nucleotides longer than *Escherichia coli* 16 S rRNA [12]. From measurements on heteroduplexes with *E. coli* rDNA fragments [5] it was known that the 3′-end of the extra 16 S rRNA gene is ~450 nucleotides away from I1. In this heteroduplex the distance from the 3′-end of the 16 S duplex to I1 is also 435 nucleotides. Therefore, the additional homology of 150 nucleotides must be positioned at the 5′-end.

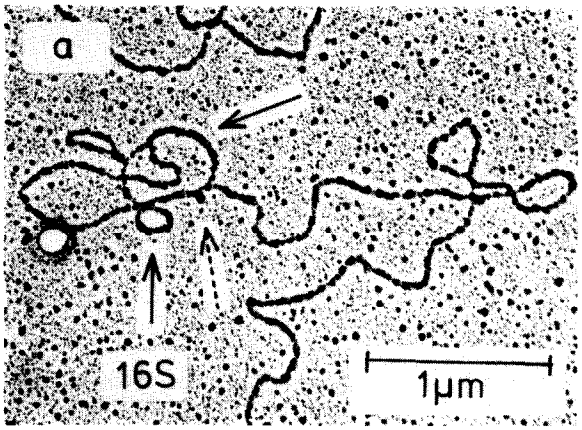
Both ends of the *Bam*HI E fragment reanneal to

the left end of the *Eco*RI B fragment, and are, therefore, not free to anneal with any possible shorter homologies upstream of the extra 16 S rRNA gene on the *Eco*RI B fragment. To avoid this formation of a heteroduplex with the left end of *Eco*RI B, *Bam*HI E was reannealed with *Bam*HI B instead. The *Bam*HI site which separates the B and D fragments is located ~500 nucleotides downstream of the 3′-end of the extra 16 S rRNA gene, very close to or within the repeated sequence I1. Single strands of the D fragment do not form a fold-back. One of the inverted sequences, therefore, must be at least partially included within *Bam*HI B.

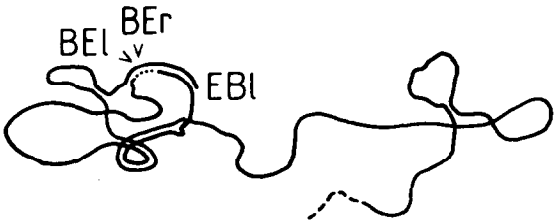
Fig.3a shows a heteroduplex between *Bam*HI E and *Bam*HI B and fig.3b a tracing of the structure. Again a circular molecule is formed, because the right end of *Bam*HI E (BEr) annealed to a sequence which is directly adjacent to the 3′-end of the extra 16 S rRNA gene. A single-stranded bridge is formed by *Bam*HI E from this short duplex to the duplex of the 16 S rRNA genes. The rest of *Bam*HI E is visible as a single-stranded tail. The left end of fragment *Bam*HI B (BB1) is visible as a short single-stranded end.

The heteroduplex between *Bam*HI D and *Bam*HI B (fig.3c) has essentially the same structure as in fig.3a, but the right end of the fragment (BDr) is ~1545 nucleotides longer than *Bam*HI E. This DNA is seen connected to the circular heteroduplex structure as a second single-stranded tail. No gap is visible at this point between the 2 heteroduplex regions formed by the 16 S rRNA genes and by the additional homology.

Length measurements of these heteroduplexes are summarized in fig.3f. The part of the sequence between the complete operons which was repeated downstream of the 3′-end of the extra 16 S rRNA gene has a length of 330 nucleotides. The right end of *Bam*HI E should



b



c

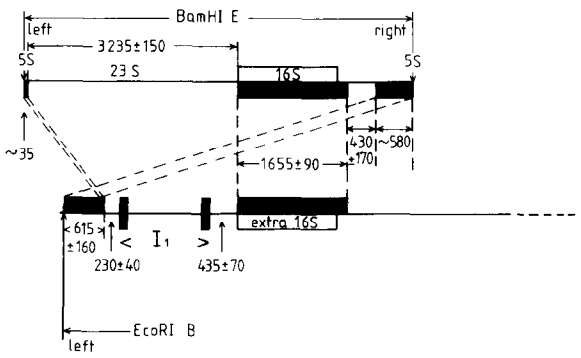
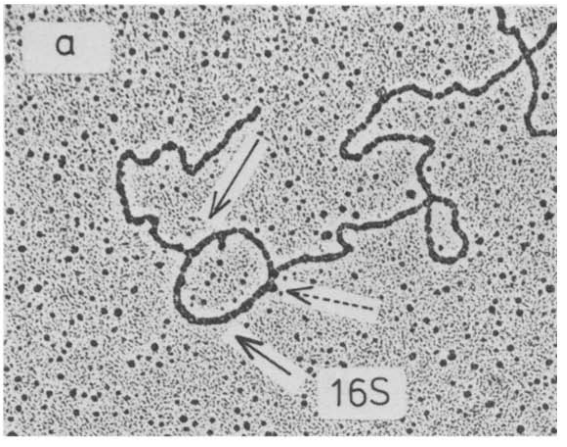
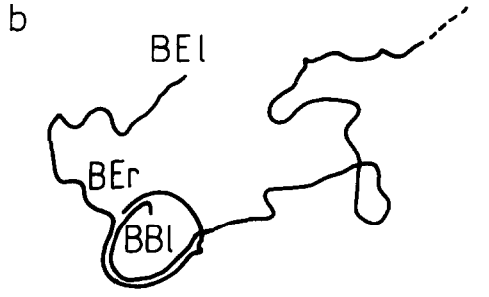


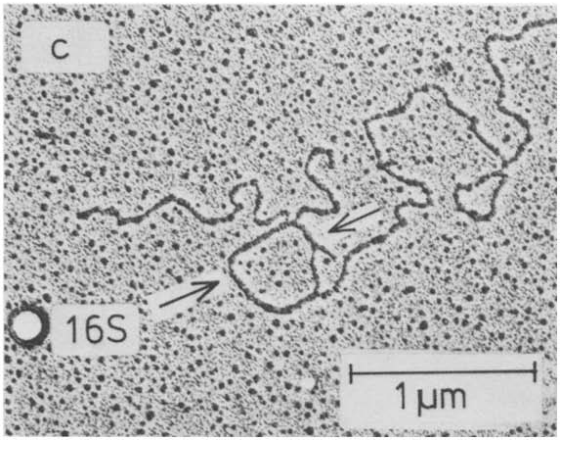
Fig.2. Analysis of heteroduplexes formed between fragments *Eco*RI B and *Bam*HI E of *Euglena* cpDNA: (a) Electron micrograph of the heteroduplex prepared by cytochrome spreading; (b) a tracing of the molecule. The left and the right ends of *Bam*HI E (marked BE1 and BEr) reannealed to the left end of *Eco*RI B (EB1, black arrow). The second duplex is formed by the 16 S rRNA genes. The dashed arrow points to the small non-homology observed in many molecules close to the 5′-end of the 16 S rRNA gene. The regions of the inverted repeat I1 in *Eco*RI B and of the 23 S rRNA gene in *Bam*HI E remain single-stranded. (c) Schematic drawing and length measurements of the heteroduplexes. The black areas indicate the positions of the double strands in each fragment. The measurements are given in number of nucleotides. The duplex of 615 nucleotides is formed partly by the left end (35 nucleotides) and partly by the right end (580 nucleotides) of *Bam*HI E. The positions of the fragment ends cannot be determined using this heteroduplex structure alone. From the heteroduplex described in fig.3e it can be calculated that the *Bam*HI site is 680 + 330 = 1010 nucleotides upstream of the 16 S rRNA gene duplex. Therefore the 615 nucleotide duplex can be divided into a sequence of 1010 − 430 = 580 nucleotides allocated to the right end and the rest of 35 nucleotides allocated to the left end of *Bam*HI E. These numbers give an estimate and may have an error of ±100 nucleotides.

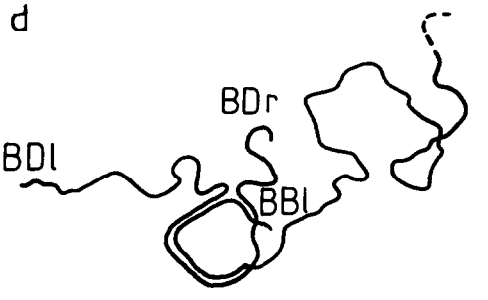
contain the 3′-end of the 5 S rRNA gene, while the left end should contain the 5′-end. Only the right end annealed to *Bam*HI B, but not the left end (fig.3b). If 5 S rRNA sequences are present near the 3′-end of the extra 16 S rRNA gene it has to be assumed that they include modifications or only part of the 5 S rRNA gene. This is also indicated by the fact that no *Bam*HI site is present and that no hybridization of 5 S rRNA to this fragment was found [6].

The lengths of the 16 S rRNA gene in both types









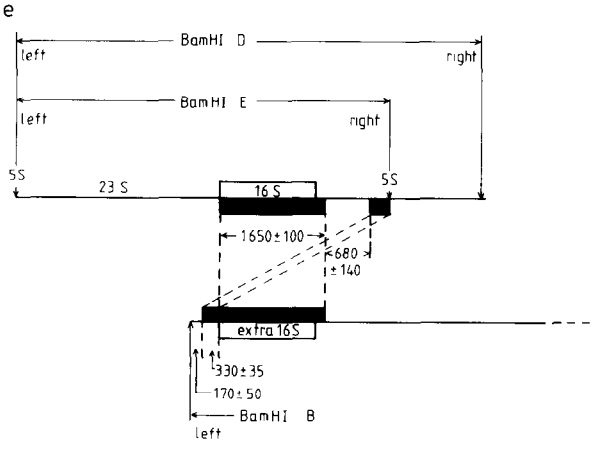


Fig.3. Analysis of heteroduplexes formed between fragments *Bam*HI E, or D and *Bam*HI B of *Euglena* cpDNA. BE1, BD1, and BEr, BDr indicate the left and right ends of *Bam*HI E and D respectively, and BB1 the left end of *Bam*HI B. (a) Electron micrograph of a heteroduplex between *Bam*HI E and *Bam*HI B; (b) a tracing of the molecule. Apart from the double strands formed by the two 16 S rRNA genes, a second homology can be seen, because the right end of *Bam*HI E (BEr) reannealed adjacent to the 3′-end of the 16 S rRNA duplex (black arrow) leaving only a short tail of the left end of *Bam*HI B (BB1) unpaired. The dashed arrow points to the small non-homology near the 5′-end of the 16 S rRNA duplex. (c) Electron micrograph of a heteroduplex between *Bam*HI D and *Bam*HI B; (d) a tracing of the molecule. The structure is almost the same as in (a,b), but the right end of *Bam*HI D (BDr) can be seen as an additional single-stranded tail, emerging from the junction of the two duplex regions, leaving no visible single-strand gap at this position on the fragment *Bam*HI B. (e) Schematic drawing and length measurements of these heteroduplexes. (▬) Positions of the double strands in each fragment.

of heteroduplexes is 1650 nucleotides. The 150 nucleotides exceeding the 16 S sequence are localised at the 5′-end of the gene. A small knob was often visible near this end of the duplex (fig.2a,3a, ——→). The length of the double strand from this knob to the 3′-end of the 16 S region was 1465 ± 100 nucleotides. This knob can be interpreted as a small insertion or deletion loop between the 5′-end of the 16 S rRNA genes and the additional homology. But it cannot be decided whether the non-complementary sequence is part of the complete operon or an addition to the extra 16 S sequence. It is clear, however, that most of the sequence of ~1200 nucleotides [14] which is located between or in front of the other 3 operons is not repeated upstream of this gene.

The positions of the homologies between the sequences on either side of the extra 16 S rRNA gene and the sequences separating the complete operons are shown in fig.1. Two possible positions of modified 5 S rRNA sequences were found:

(1) That marked 5 S* within the 615 nucleotide stretch is located ~1100 nucleotides upstream of the 5′-end of the 16 S rRNA gene of the first operon;

(2) That position in the 330 nucleotide stretch marked 5 S** which might contain a partial sequence, is located close to the 3′-end of the extra 16 S rRNA gene.

The homology upstream of the extra 16 S rRNA gene is limited to ~150 nucleotides separated by an insertion/deletion loop. It will be interesting to study whether this modification of the sequence has an affect on the promoter of the extra 16 S rRNA gene possibly interfering with its expression.

References

[1] Jenni, B. and Stutz, E. (1978) Eur. J. Biochem. 88, 127–134.
[2] Rawson, J. R. Y., Kushner, S. R., Vapnek, D., Kirby, N., Boerma, A. and Boerma, C. L. (1978) Gene 3, 191–209.
[3] Gray, P. W. and Hallick, R. B. (1978) Biochemistry 17, 284–289.
[4] Jenni, B. and Stutz, E. (1979) FEBS Lett. 102, 95–99.
[5] Koller, B. and Delius, H. (1982) FEBS Lett. 39, 86–92.
[6] Gray, P. W. and Hallick, R. B. (1979) Biochemistry 18, 1820–1825.
[7] Koller, B. and Delius, H. (1980) Mol. Gen. Genet. 178, 261–269.
[8] Koller, B., Delius, H., Buenemann, H. and Mueller, W. (1978) Gene 4, 227–239.
[9] Davis, R. W., Simon, M. and Davidson, N. (1971) Methods Enzymol. 21, 413–428.
[10] Delius, H. and Koller, B. J. (1980) Mol. Biol. 142, 247–261.
[11] Orozco, E. M., Rushlow, K. E., Dodd, J. R. and Hallick, R. B. (1980a) J. Biol. Chem. 255, 10997–11003.
[12] Hallick, R. B. (1981) in: The Biology of Euglena, (Buetow, D. E. ed) vol. 4, Academic Press, New York, in press.
[13] Brosius, J., Dull, T. J. and Noller, H. F. (1980) Proc. Natl. Acad. Sci. USA 77, 201–204.
[14] Orozco, E. M., Gray, P. W. and Hallick, R. B. (1980b) J. Biol. Chem. 255, 10991–10996.
[15] Stueber, D. and Bujard, H. (1977) Mol. Gen. Genet. 154, 299–303.
[16] Van Wezenbeck, P. M. G. F., Hulsebos, T. J. M. and Schoenmakers, J. G. G. (1980) Gene 11, 129–148.