UNIVERSITE DE NEUCHATEL
LABORATOIRE DE BIOCHIMIE

FACULTE DES SCIENCES
PROFESSEUR E. STUTZ

ANALYSE DU GENE 16S SUPPLEMENTAIRE ET DES REGIONS ADJACENTES DANS LE GENOME CHLOROPLASTIQUE D'EUGLENA GRACILIS, Z

forme réduite
de la thèse

présentée à la Faculté des Sciences
de l'Université de Neuchâtel pour
obtenir le grade de docteur ès Sciences
option Biochimie et Biologie moléculaire

par

Etienne ROUX

- 1984 -

IMPRIMATUR POUR LA THÈSE

Analyse du gène 16S supplémentaire et des régions adjacentes dans le génome chloroplastique d'Euglena gracilis, Z

de M*onsieur Etienne Roux*

UNIVERSITÉ DE NEUCHATEL

FACULTÉ DES SCIENCES

La Faculté des sciences de l'Université de Neuchâtel, sur le rapport des membres du jury,

Messieurs E. Stutz, P. Schürmann, V. Nigon (Lyon) et J.D. Rochaix (Genève)

autorise l'impression de la présente thèse.

Neuchâtel, le *le 26 octobre 1984*

Le doyen:

H. Beck

H. Beck

PUBLICATIONS

- E. Roux, L. Graf and E. Stutz (1983)
Nucleotide sequence of a 'truncated rRNA operon' of the Euglena gracilis chloroplast genome.
Nucleic Acids Research 11:1957-1968

- E. Roux and E. Stutz (1985)
The chloroplast genome of Euglena gracilis : the mosaic structure of a DNA segment linking the extra 16S rRNA gene with the rrn operon A.
Current Genetics 9:221-227

Nucleotide sequence of a 'truncated rRNA operon' of the *Euglena gracilis* chloroplast genome

Etienne Roux, Lucia Graf+ and Erhard Stutz*

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

Received 18 February 1983; Revised and Accepted 11 March 1983

ABSTRACT

An extra 16S rRNA gene (s-16S rDNA) from the Euglena gracilis chloroplast genome and several hundred positions of its flanking regions have been sequenced. The structural part has 1486 positions and is to 98% homologous in its sequence with the 16S rRNA gene in functional chloroplast rRNA operons. Sequences of about 200 positions upstream and 15 positions downstream of the structural part of the s-16S rRNA gene region are highly homologous with corresponding parts in the functional operon. Neither tRNA genes (Ala, Ile) nor parts of the 23S and 5S rRNA genes are found within 557 positions after the 3' end of the s-16S rRNA gene, i.e., the 330 bp homology, observed in electron microscopic studies of heteroduplexes (4), between the s-16S rDNA downstream region and the 6.2 kb repeated segment containing the functional rRNA operon, must be due to a DNA stretch in the interoperon spacer. A structural model of the "truncated rRNA operon" is presented. Results from S-1 endonuclease analysis suggest that the s-16S rDNA region is probably not transcribed into stable s-16S rRNA.

INTRODUCTION

We have previously shown (1) that the circular chloroplast genome of the unicellular alga Euglena gracilis, Z-strain, contains, outside and about 3.1 kbp away from three contiguous and tandemly arranged rRNA operons (2), a single region in fragment EcoRI-B which strongly hybridizes with 16S rRNA but not with 23S rRNA. Further studies with the electron microscope (3,4) revealed important sequence homology between the 16S rRNA gene of a functional operon (f-16S rDNA) and the extra 16S rRNA gene (s-16S rDNA). Some of these results also suggested that sequence homology might include regions surrounding the structural part of the s-16S rRNA gene.

One of the three functional 16S rRNA genes (5), the corresponding 16S-23S rRNA intergenic spacer (6,7) and parts of its leader (7) were sequenced. It became evident that the E. gracilis chloroplast 16S rRNA gene is homologous to 72% with the respective E. coli gene and that the intergenic

spacer contains functional genes for $tRNA^{Ile}$ and $tRNA^{Ala}$ identical to e.g. the rrnD operon of E. coli (8). Furthermore, considerable sequence homology was found to exist between the 16S-23S intergenic spacer and parts of the leader region which happens to contain a cluster of pseudo-tRNA genes (7,9). In view of these results it became interesting to compare the sequences of the s-16S rRNA gene and its flanking regions with the corresponding regions in a functional operon and also to ask the question whether this "truncated operon" could and did yield a stable 16S rRNA in a proportionate amount.

In order to answer these questions we sequenced a stretch of 2474 positions, which includes the entire s-16S rRNA gene and several hundred positions of the flanking regions. We show in the following that the s-16S rRNA gene is identical with the f-16S rRNA gene to about 98%, one major difference being a deletion of nine base pairs. Large parts of the leader region are conserved including two pseudo-tRNA genes. Sequence homology stops soon after the 3' end of the s-16S rRNA gene, i.e., the region adjacent to the s-16S rRNA gene shows neither the functional tRNA genes nor the large subunit rRNA genes as found in the functional operons.

The nine base pair deletion in the s-16S rRNA gene allowed to test, by appropriate hybridization and S1-endonuclease protection analysis, whether the s-16S rRNA gene yields stable transcripts. According to the results given in this report, this gene region is not transcribed into stable 16S rRNA in an amount proportionate to its presence in the genome.

MATERIALS AND METHODS

The fragments BglII-H and G_1 had been previously mapped on the chloroplast genome (10). They were cloned into a modified pBR322 (11). Recombinant DNA was isolated and purified as recently described (5). DNA sequencing was according to Maxam and Gilbert (12) and as specified in (5). Enzymes were purchased from Boehringer-Mannheim and New England Biolabs and used following the instructions of the supplier; details are given in (5).

The 289 bp and 280 bp XbaI-HinfI fragments which were subsequently used in the S1-endonuclease protection analysis were cut from the BamHI-D fragment (13) and the BglII-H fragment (10), respectively. The fragments were purified and 5'-end labeled with [γ-^{32}P] ATP (3000 Ci/mmol,

Amersham). Strand separation and purification was done as published (5). 5'-end labeled single strands were hybridized as specified in the legend to Fig. 5. S1-endonuclease protection analysis was according to (14).

RESULTS

1. Sequencing strategy

The circular chloroplast genome of Euglena gracilis, Z, is very well characterized (2). A survey of all BglII sites in the rDNA region is given in Fig. 1, a, and aligned with it are the three rDNA operons and the extra 16S rRNA gene region (b). The sequenced s-16S rRNA gene region is shown in greater detail on line c. Under (d) we show the various restriction sites and fragments used in the sequencing work. The arrows represent portions of the 5'-end labeled fragments from which unambiguous sequences could be established.

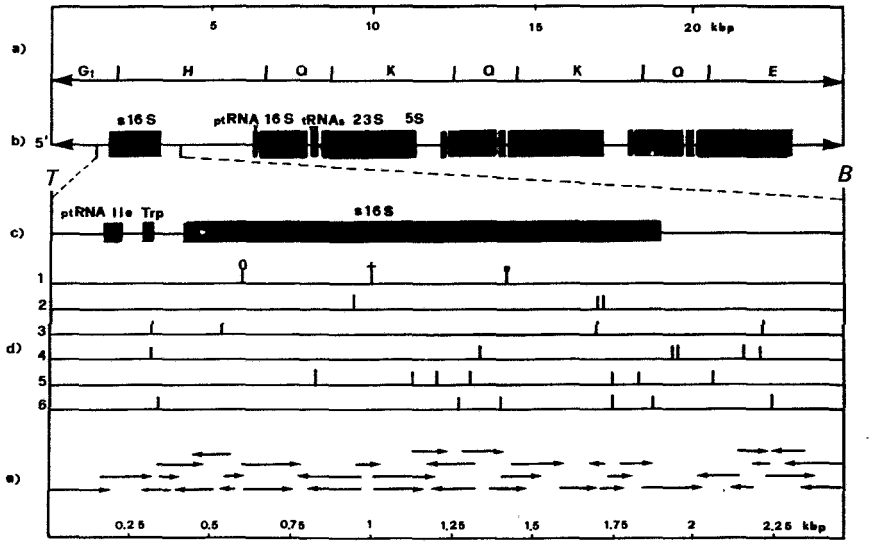

Fig. 1. Restriction endonuclease map and strategy used to sequence the s-16S rRNA gene region of the Euglena gracilis, Z, chloroplast DNA. a) BglII fragments, nomenclature see (10); b) map position of rRNA, tRNA and pseudo-tRNA genes according to published data (see text); c) s-16S rRNA gene region, T = TaqI, B = BamHI; d) restriction sites used for sequencing, 1 : BglII, HindIII, XbaI, 2 : HpaII, 3: HinfI, 4 : TaqI, 5 : HaeIII, 6 : RsaI; e) → RNA-like strand, ← coding strand.

```
TGGAAATGAC GAGTTTGATC CTTGCTCAGG GTGAACGCTG GCGGTATGCT TAACACATGC   60
AAGTTGAACG AAATTACTAG CAATAGTAAT TTAGTGGCGG ACGGGTGAGT AATATGTAAG  120
AATCTGCGCT TGGGTGAGGA ATAACAGATG GAAACGTTTG CTAATGCCTC ATAATTTACT  180
AGATCTATGT GAGTAGCTAG TTAAAGAGAA TTTCGCCTAG GCATGAGCTT GCATCTGATT  240
AGCTTGTTGG TGAGGTAAAG GCTTACCAAG GCGACGATCA GTAGCTGATT TGAGAGGATG  300
ATCAGCCACA CTGGGATTGA GAACGGAACA GACTTTTACG GAAGGCAGCA GTGAGGAATT  360
TTCCGCAATG GGCGCAAGCC TGACGGAGCA ATACCGCGTG AAGGAAGAAG GCCTTTGGGT  420
TGTAAACTTC TTTTCTCAAA GAAGAAGAAA TGACGGTATT TGAGGAATAA GCATCGGCTA  480
ATTCCGTGCC AGCAGCCGCG GTAATACGGG AGATGCGAGC GTTATCCGGA ATTATTGGGC  540
GTAAAGAGTT TGTAGGCGGT CAAGTGTGTT TAATGTTAAA AGTCAAAGCT TAACTTTGGA  600
AGGGCATTAA AAACTGCTAG ACTTGAGTAT GGTAGGGGTG AAGGGAATTT CCAGTGTAGC  660
GGTGAAATGC GTAGAGATTG GAAAGAACAC CAATGGCGAA GGCACTTTTC TAGGCCAATA  720
CTGACGCTGA GAAACGAAAG CTGAGGGAGC AAACAGGATT AGATACCCTG GTAGTCTTGG  780
CCGTAAACTA TGGATACTAA GTGGTGCTGA AAGTGCACTG CTGTAGTTAA CACGTTAAGT  840
ATCCCGCCTG GGGAGTACGC TTGCACAAGT GAAACTCAAA GGAATTGACG GGGGCCCGCA  900
CAAGCGGTGG AGCATGTGGT TTAATTCGAT GCAACACGAA GAACCTTACC AGGATTTGAC  960
AGGATCTAGG AGGAAGTTTG AAAGAACGCA GTACCTTCGG GTATCTAGAC ACAGGTGGTG 1020
CATGGCTGTC GTCAGCTCGT GTCGTGAGAT GTTGGGTTAA GTCCCGCAAC CCTTTTTTTT 1080
AATTAACGCT TGTCATTTAG AAATACTGCT CGTTATTACC AGAGGAAGGT GAGGACGACG 1140
TCAAGTCATC ATGCCCCTTA TATCCTGGGC TACACACGTG CTACAATGGT TAAGACAATA 1200
AGTTGCAATT TTGTGAAAAT GAGCTAATCT TAAAACTTAG CCTAAGTTCG GATTGTAGGC 1260
TGAAACTCGC CTACATGAAG CCGGAATCGC TAGTAATCGC CGGTCAGCTA TACGGCGGTG 1320
AATACGTTCT CGGGCCTTGT ACACACCGCC CGTCACACCA TGGAAGTTGG CTGTGCCCGA 1380
AGTTATTATC TTGCCTGAAA AGAGGGAAAT ACCTAAGGCC TGGCTGGTGA CTGGGGTGAA 1440
GTCGTAACAA GGTAGCCGTA CTGGAAGGTG TGGCTGGAAC AATTCGC
```

Fig. 2. Nucleotide sequence of the s-16S rDNA. Only the RNA-like strand is given. Nucleotides differing from that of the f-16S rRNA gene are underlined [consult Table 1]. We mark the two possible 3' ends [Zablen et al. (15) versus Steege et al. (16), see text]. A 28 bp sequence in the 5' terminal part which is invertely repeated in the leader part [see Fig. 3] is marked by a wave line. An XbaI and HinfI site which are used to generate a 280 bp fragment [see Fig. 5] are boxed.

2. Comparison between the structural parts of the s-16S rRNA gene region and the f-16S rRNA gene

In Fig. 2 we show the nucleotide sequence of the entire s-16S rRNA gene which contains 1486 or 1487 positions depending on whether we take, respectively, as terminal RNase T1 oligonucleotide the one reported by Zablen et al. (15) or Steege et al. (16). The s-16S rRNA gene is five nucleotides shorter than the f-16S rRNA gene (5). Comparing the two sequences reveals a total of 21 mismatches, i.e., 98% sequence homology. We underlined in Fig. 2 those positions which differ from that in the f-16S rRNA gene and in Table 1 we qualify each mismatch. There are eight nucleotide changes (pos. 135, 336, 409, 423, 428, 1121, 1368, 1483) two insertions of one (pos. 771) and of three nucleotides (pos. 969 to 971), respectively, and a nine nucleotide deletion (pos. 1069). All these changes in the primary structure of the gene do not interfere with the secondary structure model of a potential 16S rRNA (17,18). Some of these base changes e.g. pos. 336 and 428 allow formation of an additional base pairing within a stem region. A minor secondary structure change could occur due to the nine bp deletion in the helix 29 which is part of the variable domain E

Table 1 : Sequence mismatches between the f-16S and the s-16S RNA gene of the E. gracilis, Z, chloroplast genome

f-16S (position)	s-16S (position)	Potential RNase T1 oligonucleotides (base change underlined)
C (135)	T (135)	
C (336)	T (336)	5'-ACUU**U**UACG [9 mer]
C (409)	A (409)	
A (423)	T (423)	5'-**U**AAAC**U**UCUUUUCUCAAAG [19 mer]
C (428)	T (428)	
△ (770)*	G (771)	
△ (967)*	GGA (969-971)	
CGACGCCAA (1066-1074)	△ (1069)*	
G (1126)	A (1121)	5'-UUAUUACC**A**G [10 mer]
C (1373)	T (1368)	
C (1488)	T (1483)	

*Last identical position before mismatch.

[nomenclature according to Stiegler et al. (17)]. But as a whole we may consider the s-16S and f-16S rRNA genes to be structurally equivalent.

3. Analysis of the leader part

Electron microscopic analysis suggested that the sequence homology between the s-16S rRNA gene region and the f-16S rRNA gene includes also 150 bp preceeding the 5' end of the structural gene (4). Furthermore, Orozco et al. (7) sequenced parts of the leader of a f-16S rRNA gene and they found it to contain pseudo-tRNA genes or partial tRNA genes for isoleucin and alanine. R. Helling and collaborators have also sequenced the leader (pos. -1 to about -400) of a f-16S rRNA gene of Euglena gracilis, B-strain and they made the same observation (personal communication). It was therefore of interest to sequence the leader part of the s-16S rRNA gene. In Fig. 3 we show 415 positions upstream to the 5' end and align it with parts of the leader of a functional operon of the Z-strain (7). It contains a partial tRNA gene for alanine (codon change to tryptophane)

```
a)                              ACAAGAAA TAACTTAGCT TAATTCCTAG TGTAAATTTT
b) CTTATTTGCA AATTTAAAAG GCTTTTATCC TTTAGTAGTT AAGAAATCCA AGGATTTACT  -356

a) TCCGTATCTT AACGTGTGGA ACAATTTTTT TATTTTAATT AAAATACCGT ATAAATTTTA
b) GAAATTAAAC AATACTTATC ATTATGATGC GATATTTTTG TCAACTCAAA TATCCTTGAA  -296

a) AAAAATAAAA AAATCAAAAG GAAAATTTTT TATTTTTCAA AGATCACTAT TATCTTTTA.
b) TATCAAAATG TAAATGAGAT AAAAAACATG TTCAACTCTT TAGAAACGAC GTTAACTGCT  -236

a) ......A... ..C..... ..G..... .........G ..T....*** **********
b) TATTATGGAA TATCTTTGAT AAAGGTAGGG TCGTGGATTA AAGCCTTCGT TCAACTTGCA  -176

a) *******GTA T....C.... ...T..... T.......CC ..... .....
b) TGTGTTAAGC ATAGCTAAAT ATTGCTTTTC GTTAAAATTG TAAGGCGTAG GTCTCCAAAA  -116

a) .........A
b) CCTGATGTAG TAGGTTCGAA TCCTACAAAG CGCGCTTTTA GTGTACACAT TATAGTAAAT  - 56

a)                                        ..... T..... .....TGGAA
b) GTGCCCCTTG CTTGGTCACC AAGAGGGTGA AAGGATTTGA CCAACTTTGA TGTTTTGGAA  +  5
```

Fig. 3. Nucleotide sequence of 415 positions preceeding the s-16S rDNA. Only the RNA-like strand is given (lines b) and aligned with a partially sequenced leader (lines a) of the f-16S rRNA gene as published (5,7). Points on line a indicate nucleotide identity with the sequence given on line b. Positions -21 to -105 on line a are left open. This segment has been sequenced in case of the B-strain and is to 98% identical with the sequence given in line b (R. Helling, authorized us to use his unpublished results). Stretches of pseudo-tRNA character (7,9) are underlined. The 20 bp insert is marked by asteriks. The inverted repeat of 28 bp corresponding to the sequence in the structural part of the 16S rRNA gene [see Fig. 2] is marked by a wave line. The first five positions of the 16S rRNA gene are framed.

and a pseudo-tRNA^{Ile} gene, but homology stops at position -236. We know that the DNA segment from pos. -1 to -106 of the s-16S rRNA gene leader is to 98% homologous with the corresponding leader part of the f-16S rRNA gene from the B-strain (R. Helling, personal communication). In particular the BstEII site (5'TTGGTCACCAA, pos. -44 to -34) is preserved where transcription might start according to a report of R. Hallick (NATO-FEBS Meeting, Porto-Portese, 1982). With a leader quite similar to that found in functional operons and an almost intact structural gene, the s-16S gene might serve as template for the RNA polymerase yielding 16S rRNA.

We mark on the s-16S leader (Fig. 3) an insert of 20 nucleotides which is not present in the f-16S leader. These 20 positions are part of a 28 bp sequence which occurs as an exact inverted repeat in the 5' terminal part of the s-16S rRNA gene (pos. 44 to 72).

4. Analysis of the sequence flanking the 3' end

In Fig. 4 we show 572 positions of the DNA segment flanking the 3' end of the s-16S rRNA gene. Sequence homology with the corresponding segment in the operon continues for 15 positions but within the next 557 positions we do not find any significant homology with the DNA segment following the f-16S rRNA gene, which contains the genes for tRNA^{Ile} and tRNA^{Ala} and the 5' terminal region of 23S rRNA (6,7). However, according to an interpretation of electron microscopic data of Koller and Delius (4) this segment should

```
TTTAGTTTTT TAACTGAATT TATTTATTAA TATTAACTCG ATTTTCGAGT AATTTTTATG   +60
ACTAACCGCT ATGTTAGTAT TTATGATGAT AATATGTGCA ATTAATTAAA CCAACAAATT   +120
GTCAAAAACT TCTCCTTTGA GTTTGAATTT TTCTGGCCAA AGCAGGATGC ATGGGAAGAA   +180
ATTAAAAAAT TTTTATATAG AAATCCGTGG ATTTCCGAGG ATATCGCTTT TAAACTTTAA   +240
ATGATATACG ATATTGTCGA GTGTTGGCAA AATGATTATG TTCTGCCAGA ATAGAACTGA   +300
TTAATCTGTC GAGGGAATCT TTTGGTTCTC TTTTAAAAGT CCGAAGAGTA CAAAATCAAG   +360
GTAAGCACTT TGACTTTTTC TCCTATGTTT TTGAATGGTT TTTTGCGCCG TCTTACGCGA   +420
AAAATAAAAC AACATGGCTA TGCGTTTTTT GTCGTTTCTG TTGATGGCAT AGCTACTGGC   +480
GTAAAAGGCT CCGTAAAAGG ATCTTTTTCT GTTGCTTCAC TTGACCCGAT GGTCATGGAA   +540
AATATAGAAA AATGTGCAAT ACTGCACAAT GA
```

Fig. 4. Nucleotide sequence of 572 positions following the 3'end (16) of the s-16S rDNA. The 15 positions homologous with the f-16S-23S intergenic spacer in a functional operon (6,7), the HaeIII site [see text], and the start of an open reading frame are underlined.

contain approximately 330 bp homologous with sequences flanking the 3' end of the 23S rRNA gene, i.e., it might contain a partial 5S rRNA sequence (consult Fig. 1). A computer search for 5S rRNA-like sequences within the 572 positions gave a negative result. According to previous restriction site analysis it seemed that sequence homology might extent at least up to the HaeIII site in the $tRNA^{Ile}$ gene of the intergenic spacer (1,7). Our sequencing data now reveal that a HaeIII site incidentally exists at the same distance from the 3' end yielding also a fragment of 225 bp, as was found in the intergenic spacer of the functional operon. However, the corresponding tRNA sequence is not present. We may add at this point that an open reading frame of unknown length starts at position 386. Whether this represents the start of a functional gene remains to be shown.

5. Search for s-16S rRNA

The nine bp deletion (pos. 1069) was used as a handle to test whether

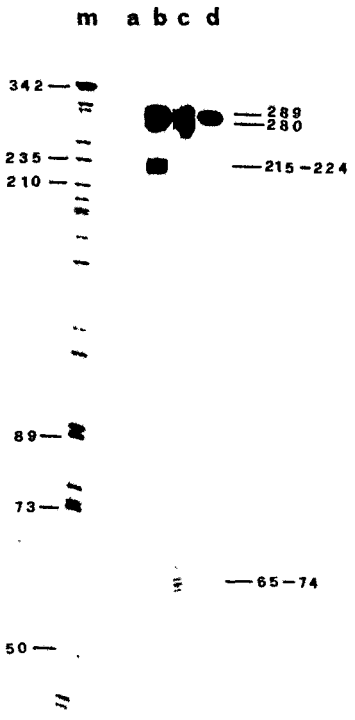

Fig. 5. Autoradiographs of the S-1 endonuclease protection analysis of DNA:DNA and DNA:RNA hybrids. 50 ng of 5' end labeled purified single strand DNA from the 289 bp fragment were hybridized with 25 ng of 5' end labeled purified single strand DNA from the 280 bp fragment in reciprocal experiments. Conditions : 3 h, 34°C, 40 mM PIPES, pH 6.4, 1 mM EDTA, 0.4 M NaCl, 80% formamide. 50 ng of the 5' end labeled purified coding strand from the 289 bp fragment were hybridized with 40 ng of total chloroplast RNA. Conditions : 3 h, 52°C, otherwise as mentioned above. S1-endonuclease digestion was in all cases for 30 min, 52°C, in 0.28 M NaCl, 0.05 M Na-acetate, pH 4.6, 4.5 mM $ZnSO_4$, 20 µg/ml carrier single strand DNA. Reaction volume in all cases 30 µl. Hybrids were analysed in sequencing gels (12). m) sizing marker, a) [289 bases, coding] alone, b) [289 bases, coding]X [280 bases, non coding], c) reciprocal version of (b), d) [289 bases, coding] X RNA. In panels b and c the top strong band is composed of the 289 and 280 bases ss DNA.

the "truncated rRNA operon" containing a leader and an intact 16S rRNA gene is transcribed yielding a proportionate amount of s-16S rRNA. To this end we cut from the f-16S rRNA and s-16S rRNA gene region a XbaI-HinfI DNA fragment of 289 and 280 bp, respectively, containing the nine bp insertion / deletion (see Table 1). These two DNA fragments were used in heterologous DNA:DNA and DNA:RNA hybridization experiments. The hybrids were analysed in S1-endonuclease protection experiments. In order to test the feasability of this approach we first constructed DNA heteroduplexes with purified 5' end labeled single strand DNA (ss DNA) from the 289 and 280 bp fragments. In Fig. 5 we show the electrophoretic analysis of the S1-endonuclease digestion products of the heteroduplexes. According to the position of the nine bp deletion in the XbaI-HinfI (280 bp) fragment (see Fig. 2) we should see in the autoradiograph ss DNA fragments of 216 and 65 bases in both cases, i.e., the combinations [289 bases, coding] X [280 bases, non-coding] and [289 bases, non-coding] X [280 bases, coding]. This is, however, under the given experimental conditions not the case. The first combination (panel b) yields a cluster of fragments in the range of 216-224 in addition to the intact fragments and the second combination yields a cluster of bands in the range of 65 to 74 (panel c) in addition to the intact fragments. The conclusion is that only the looping (longer) strand is cut while the shorter strand in the heteroduplexes remains essentially intact. In both cases the S1-endonuclease digestion of the heteroduplexes does not go to completion, yielding therefore a strong signal in the autoradiograph for the 289 and 280 bases ss DNA. On the other hand S1-endonuclease digestion is essentially complete if the 289 bases coding strand is used alone in the hybridisation experiment yielding a very faint signal (panel a) which is indicative for a pure ssDNA preparation.

Total chloroplast RNA was hybridized to the 5' end labeled 289 bases coding strand. The result of the S1-endonuclease protection analysis is shown in panel d. We see the expected strong signal in the range of 289 bases but we do not see any signal in the range of 215 to 224 bases. This strongly suggests that the s-16S rRNA is essentially absent in the chloroplast RNA preparation.

DISCUSSION

Comparison of the s-16S rDNA region with the f-16S rDNA region reveals

for the structural part of the gene essentially complete identity (98%). The most important change, namely the nine base pair deletion, would most likely not interfere with the functionality of the corresponding 16S rRNA since it does not involve any of the conserved regions (17). The only mutation of immediate functional impact might be the change in the so-called Shine-Dalgarno sequence (19) at pos. 1483 where a T(U) replaces a C. Note, however, that the chloroplast f-16S rRNA gene has already a modified Shine-Dalgarno sequence, i.e., instead of 5'-ACCUCC- as seen in E. coli DNA and, e.g., in the chloroplast DNA of maize (20) there is a 5'-AACUCC- in E. gracilis (5). Whether chloroplast mRNA is correspondingly adapted is not yet known.

Of considerable functional and evolutionary interest is a comparison of the leader parts. The first 164 positions proximal to the 5' end of the structural gene are almost identical with those from the f-16S rRNA gene of both the Z-strain and B-strain; homology continues for another 48 positions after an insert of 20 bp. Orozco et al. (7) and Myata et al. (9) have already discussed the fact that the leader region displays considerable homology with the intergenic spacer, including parts of the 3' end of the 16S rRNA gene and the $tRNA^{Ile}$ and $tRNA^{Ala}$ genes. The situation is very similar in our case, i.e., the same kind of pseudo-tRNA sequences are present in the leader part of the s-16S rRNA gene. Sequence homology between the "s-leader" and "f-leader" decreases to 40% and lower between position -235 and -415. This means that this part of the "s-leader" either has mutated during evolution more rapidly than the corresponding part in the "f-leader" or the postulated gene duplication event (9) never included that part of the s-leader.

Koller and Delius (4) have analysed DNA heteroduplexes formed between the s-16S and f-16S rDNA region. According to their electron microscopic measurements, sequence homology includes about 150 positions upstream of the structural part. Furthermore they observed a small loop (knob-like structure) upstream and in close proximity of the 5' end of the 16S rRNA gene. Our sequencing results of the leader part are in line with the electron microscopic analysis. The small loop seen in the heteroduplexes is most likely the result of a reannealing event between the 28 bases inverted repeats which are about 200 bases apart. The 20 bases insert in the leader is certainly too small to be seen in the electron microscope and therefore could not possibly be the reason for the knob-like structure (B. Koller, personal communication).

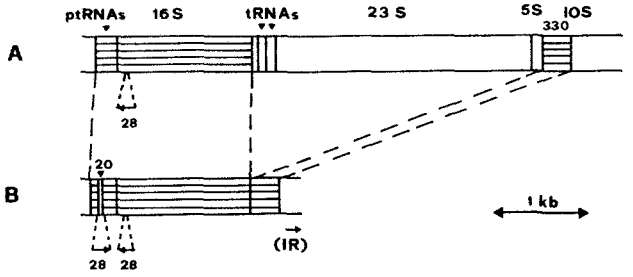

Fig. 6. Comparison of the anatomy of the functional rRNA operon [A] with the "truncated rRNA operon" [B]. Segments of high sequence homology are marked by boxes with horizontal bars; IOS : interoperon spacer; (IR) region of a small inverted repeat described in (4).

According to electron microscopic analysis a stretch of about 330 nucleotides downstream of the s-16S rRNA gene interacts with a region in the functional operon adjacent to the 3' end of the 23S rRNA gene possibly including parts of a modified 5S rRNA gene (4). We did not find a 5S rRNA-like sequence within the 330 bp adjacent and downstream of the s-16S rRNA gene. Therefore, the observed interaction must be due to a matching sequence located after the 5S rRNA gene of one operon and before the leader part of the next operon, i.e., within the interoperon spacer (2) as shown in Fig. 6. The average A+T content of this 330 bp stretch is about 70% (Fig. 2). Accordingly, the s-16S rRNA gene region would contain a leader part and a 16S rRNA gene very similar to the functional operon, including 15 positions of the intergenic spacer and about 330 bp of the interoperon spacer as schematically drawn in Fig. 6. It is noteworthy that s-16S rDNA also occurs in the chloroplast genome of the bacillaris strain of Euglena gracilis (21).

s-16S rRNA was not detectable in our hybridisation experiments and none of the potential T1-oligonucleotides (see Table 1) was listed by Zablen et al. (15). This could mean that the s-16S rDNA is not transcribed at all or that it is transcribed at normal rate like the normal rDNA regions but the steady state concentration of the s-16S rRNA remains below detectability due to rapid degradation of the transcription product after synthesis.

ACKNOWLEDGEMENTS

We are grateful to R.B. Helling (Ann Arbor) for providing unpublished sequencing data, M. Fasnacht, B. Jenni and B. Schlunegger for providing

the clones, P.E. Montandon for suggesting the experiments concerning the rDNA transcription and Ch. Bachmann for secretarial help. The computer programs were obtained from R. Staden, Cambridge. This research is supported by Fonds national suisse de la Recherche scientifique (3.183.82 to E.S.). This work is part of a Ph.D. thesis (E.R.).

+Present address: Friedrich Miescher-Institute, P.O. Box 2543, CH-4002 Basel, Switzerland

*To whom correspondence should be addressed

REFERENCES

1. Jenni, B. and Stutz, E. (1979). FEBS Lett. 102, 95-99.
2. Hallick, R.B. The Biology of Euglena, vol. IV, Buetow, D.E., ed., Academic Press, New York, in press.
3. Koller, B. and Delius, H. (1982). FEBS Lett. 139, 86-92.
4. Koller, B. and Delius, H. (1982). FEBS Lett. 140, 198-202.
5. Graf, L., Roux, E., Stutz, E. and Kössel, H. (1982). Nucleic Acids Res. 10, 6369-6381.
6. Graf, L., Kössel, H. and Stutz, E. (1980). Nature 286, 908-910.
7. Orozco, E.M. jr., Rushlow, K.E., Dodd, J.R. and Hallick, R.B. (1980). J. Biol. Chem. 255, 10997-11003.
8. Young, R.A., Maklis, R. and Steitz, J.A. (1979). J. Biol. Chem. 254, 3264-3271.
9. Miyata, T., Kikuno, R. and Ohshima, Y. (1982). Nucleic Acids Res. 10, 1771-1780.
10. Jenni, B., Fasnacht, M. and Stutz, E. (1981). FEBS Lett. 125, 175-179.
11. Schlunegger, B., Fasnacht, M., Stutz, E., Koller, B. and Delius, H. (1983). Biochim. Biophys. Acta 739, 114-121.
12. Maxam, A.M. and Gilbert, H. (1980). Methods in Enzymology, vol. 65, pp. 499-560, Grossman, L. and Moldave, K., eds, Academic Press, New York.
13. Gray, P.W. and Hallick, R.B. (1978). Biochemistry 17, 284-290.
14. Favaloro, J., Treisman, R. and Kamen, R. (1980). Methods in Enzymology 65, pp. 718-749, Grossman, L. and Moldave, K., eds, Academic Press, New York.
15. Zablen, L.B., Kissil, M.S., Woese, C.R. and Buetow, D.E. (1975). Proc. Natl. Acad. Sci. USA 72, 2418-2422.
16. Steege, D.A., Graves, M.C. and Spremulli, L.L. (1982). J. Biol. Chem. 257, 10430-10439.
17. Stiegler, P., Carbon, P., Ebel, J.P., Ehresmann, C. (1981). Eur. J. Biochem. 120, 487-495.
18. Zwieb, Ch., Glotz, C. and Brimacombe, R. (1981). Nucleic Acids Res. 9, 3621-3640.
19. Shine, J. and Dalgarno, L. (1974). Proc. Natl. Acad. Sci. USA 71, 1342-1346.
20. Schwarz, Zs. and Kössel, H. (1980). Nature 283, 739-742.
21. Koller, B. and Delius, H. (1982). Mol. Gen. Genet. 188, 305-308.

Current Genetics (1985) 9:221–227

Current Genetics

The chloroplast genome of *Euglena gracilis:* the mosaic structure of a DNA segment linking the extra 16S rRNA gene with the *rrn* operon A

Etienne Roux and Erhard Stutz

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

Summary. We have completed the analysis of a DNA segment of the chloroplast genome of *Euglena gracilis* Klebs, Z-strain, which links the 3′ end of the extra 16S rRNA gene with the 5′ end of the 16S rRNA gene of the *rrn* operon A. This region is a mosaic of several structural elements and contains an intact *rrn* interoperon spacer of 1,080 bp, an extra 5S rRNA gene, an open reading frame for 406 codons (ORF 406) which is flanked by short inverted repeats and a short direct repeat originating from the *rrn* interoperon spacer. It seems that a once complete *rrn* operon underwent in the past an insertion/deletion event leaving intact the 16S and 5S rRNA but totally excising the 16S-23S intergenic spacer and the 23S rRNA gene. Instead a protein coding gene of yet unknown function was inserted along with other structural elements.

Key words: Chloroplast DNA – *Euglena gracilis* – *rrn* operons – DNA deletion/insertion

Introduction

Several years ago, Jenni and Stutz 1979, reported for the first time that the circular chloroplast genome of *Euglena gracilis* Klebs, Z-strain, contains in addition to three tandemly arranged *rrn* operons [5′-16S-*trn*A-*trn*I-23S-5S-3′] an extra 16S rRNA gene (s16S) about 3.1 kbp upstream of the 5′ end of the next 16S rRNA gene (*rrn* operon A). We sequenced both the 16S rRNA gene of the *rrn* operon A (Graf et al. 1982) and the s16S rRNA gene (Roux et al. 1983) and found a very high sequence homology of 98%, suggesting that this s16S rRNA gene could yield functional 16S rRNA. More recently the s16S rRNA gene was found in other strains of *Euglena gracilis*, e.g., *bacillaris* and Z-S strain (Koller et al. 1984) which contain three and one *rrn* operon, respectively. It was also found in the strain American Type Culture Collection ATCC 10,616 which contains five *rrn* operons (Koller and Delius 1982a; Flamant et al. 1984). In this particular case there are two s16S genes located upstream of the first and third *rrn* operon, respectively. Also an X-ray induced mutant (Y3 BUD) of the *bacillaris* wild type which has lost the *rrn* operon A has retained the s16S rRNA gene in front of the first *rrn* operon (Ravel-Chapuis et al. 1984). From these results, we may conclude that the number of *rrn* operons per circular chloroplast genome can vary without impairing normal growth, however, a s16S rRNA gene is always present and located upstream of the first complete *rrn* operon. This strongly suggests that this DNA stretch or parts of it carries genetic information vital for the *Euglena gracilis* chloroplast.

Very recently, Koller et al. (1984) have studied in the electron microscope homo- and heteroduplexes formed between chloroplast DNAs from the *bacillaris* and Z-S strain. This study included also the DNA region between the s16S and 16S rRNA genes. From these and similar studies with the Z-strain (Koller and Delius 1982b) it became evident that this DNA segment is composed of several structural elements like, e.g., short direct and indirect repeats and that the arrangement of these elements in the various strains follows a somewhat similar pattern. Parts of this segment have recently been sequenced both in the *bacillaris* strain (El-Gewely et al. 1984) and in the Z-strain (Roux et al. 1983). Based on these sequencing data and the electron microscopic results it was suggested that the s16S rRNA gene is a relic of a once complete *rrn* operon which was "truncated" during a DNA insertion/deletion event in a distant past.

We report here the nucleotide sequence of the entire DNA segment between the s16S and 16S rRNA genes of

Offprint requests to: E. Stutz

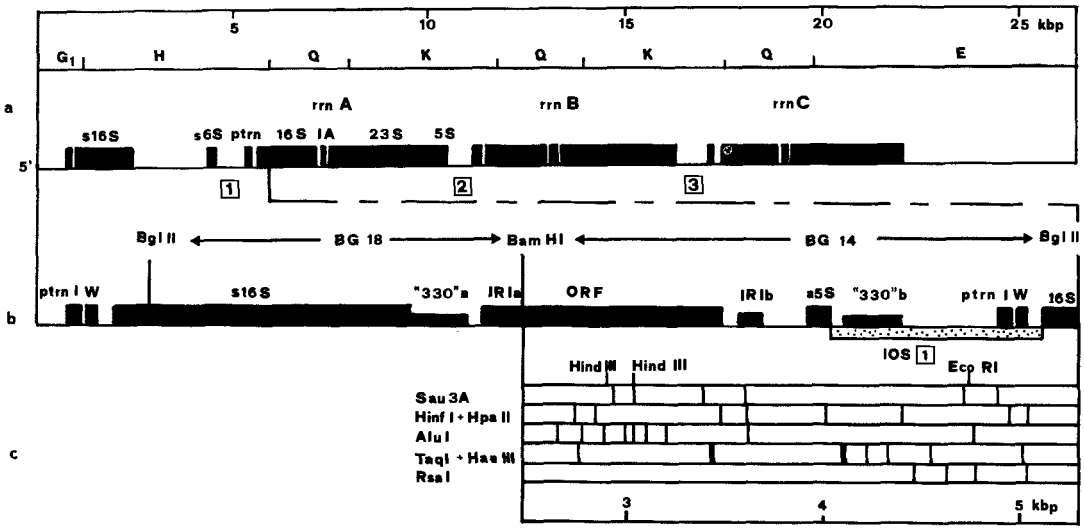

Fig. 1 a–c. Map position and strategy of sequencing. a BglII site restriction map (Jenni et al. 1981) of the rDNA region with the extra 16S rRNA gene and the *rrn* operons A, B, C (Hallick 1984). b Structural elements of the previously sequenced DNA stretch (Roux et al. 1983) and the fragment BG14 analysed in this report. c Restriction fragments used for sequencing according to Sanger et al. (1980). The interoperon spacers (IOS) are ordered 1, 2, 3 and defined in the text. The structural element "330", IRIa,b and ORF are defined in the text. s16S and s5S are the structural parts of 16S and 5S rRNA genes of the "truncated" *rrn* operon upstream of interoperon spacer 1, ptrn are the pseudo-tRNA genes for Ile and Trp previously described (Roux et al. 1983), I, W are the *trn*I and *trn*A genes of the intergenic spacer (Graf et al. 1980)

the Z-strain. This allows to precisely define the structural elements observed in the electron microscope and to compare on a nucleotide level the interoperon spacers of the Z- and *bacillaris* strain. These sequencing studies also reveal a major open reading frame which is co-transcribed with the s16S rRNA gene as will be shown elsewhere (Roux, Koller, Montandon and Stutz, to be published).

Materials and methods

Euglena gracilis, Klebs, Z-strain was purchased some time ago from the Culture Collection of Algae at Indiana University, strain number 753 (Starr 1964). The chloroplast genome of this strain and in particular the rDNA region has been described in great detail (Hallick 1984).

Enzymes were purchased from Boehringer-Mannheim and used following the instructions of the supplier. [α-^{32}P]ATP 400 Ci/mmole was from Radiochemical Center, Amersham.

The chloroplast DNA fragment BglII-H was cloned into the BglII site of a modified pBR322 (Schlunegger et al. 1983).

As starting material for DNA sequencing we used the BamHI-BglII DNA fragment BG14 (Jenni and Stutz 1979), which was obtained from the cloned fragment BglII-H (4.7 kbp). Subfragments of BG14 were routinely filled with the Klenow fragment of DNA polymerase (Wartell and Reznikoff 1980) and cloned into the HincII site of phage M13 mp9 (Messing et al. 1980; Messing and Vieira 1982) with ligation conditions as reported by Tait et al. (1980). Transformation of *E. coli* JM103 was according to Cohen et al. (1972), however, competent cells were prepared in 0.02 M Tris-HCl, pH 7.2, 0.1 M $CaCl_2$, 0,001 M NaCl.

DNA sequencing was according to Sanger et al. (1980).

Results

Mapping and strategy of sequencing

In Fig. 1 (a) we show the BglII fragments of a DNA segment of the circular chloroplast genome as mapped previously (Jenni et al. 1981). Line (b) gives the position of all the structural elements as defined and discussed in this report and under (c) we show the restriction fragments used in nucleotide sequencing experiments.

Nucleotide sequence and major structural elements

In a previous paper (Roux et al. 1983) we reported the nucleotide sequence of a DNA segment which started within the fragment BglII-G1 and ended with the BamHI site in BglII-H (see Fig. 1). In order to relate those sequencing results with the new data given here we display again in Fig. 2 the 3′ end of the s16S rRNA gene (pos. 1902) and the adjacent part up to position 2477 (see BamHI site). This stretch contains one of the "330" elements and one of the IR1 elements, seen by Koller and Delius (1982) in the electron microscope. With the complete sequence at hand we can now define both the "330" and the IR1 elements. The "330"a (293 nucleotides) and the "330"b (303 nucleotides) are homologous to

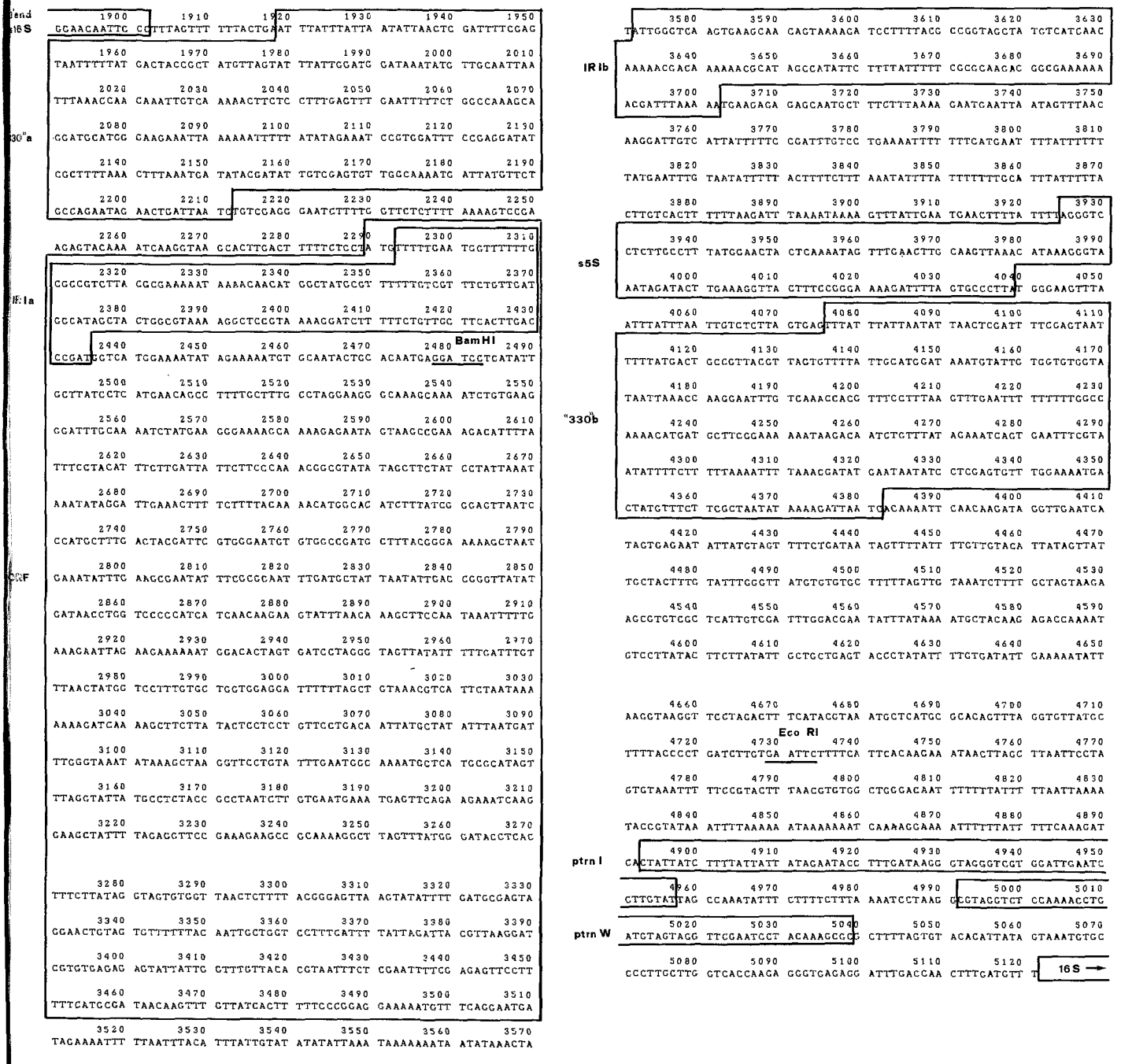

Fig. 2. Nucleotide sequence of a DNA segment linking the 3′ end of the s16S rRNA gene with the 5′ end of the 16S rRNA of *rrn* operon A. Only the DNA strand corresponding to the rRNA strand is shown. Major structural elements are boxed and important restriction sites underlined

about 75%, they are relatively rich in A+T (around 70%) and represent the only major direct repeats in the analyzed segment.

The two elements IR1a (143 nucleotides) and IR1b (131 nucleotides) match to 84%. They have an A+T content of about 60% and no other indirect repeats of this size are found in the analyzed segment. Interesting enough IR1a is part of an open reading frame sitting right at the 5′ end, while IR1b is 61 positions downstream of the protein coding region.

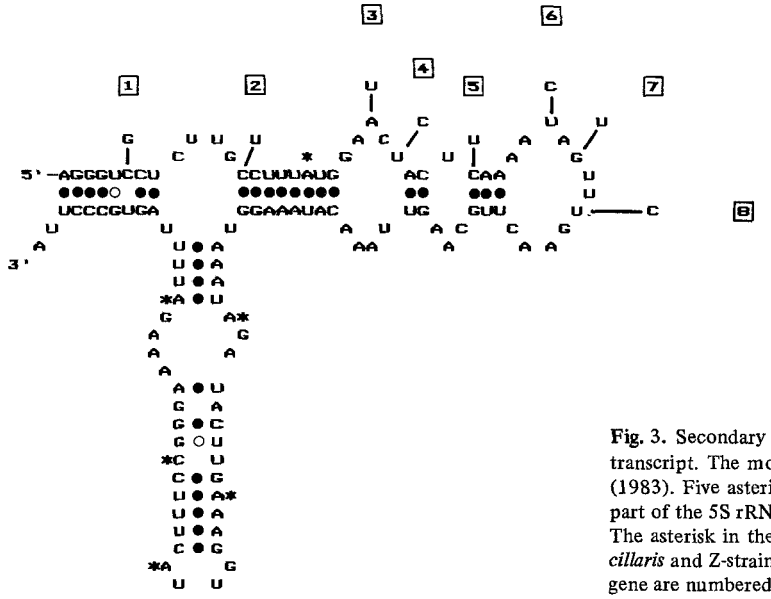

Fig. 3. Secondary structure model of a hypothetical s5S rDNA transcript. The model is constructed according to Karabin et al. (1983). Five asterisks denote the polymorphism found in the 3′ part of the 5S rRNA gene of *rrn* operon A, B and C, respectively. The asterisk in the 5′ part marks a base change between the *bacillaris* and Z-strain. Changes in the 5′ part found in the s5S RNA gene are numbered 1 to 8

```
MFLNGFLRRLTRK1KQHGYAFFVVSVDGIATGVKGSVKGSFSVASLDPMVMENIEKCAIL
HNEDPHIAYPHEQPFALPRKGKAKSVKGFAKSMKGKAKENSKPKEILFPTFLDYSSQTAY
IAEILLNKYRIETFLLQNMAHLYRELIPCFDYDSWECVADALREKANEIFEANISRNLML
LILTGLYDNLVPHHQQEVFNKASNKFLKELEEKMDTSDPRVVIFLICLTMVLCAGGGFFS
CKRHSNKKRSKASYTPPVPDNYAIFNDLGKYKAKVPVFEWQNAHAHSLGIMPLFPNLVNE
MSSEEIKEAILEVPKEAAKGLVYGIPHFLIGSVVNSFTGVKYILMRVGTVVFFTIAGPFI
LLDYVKDRVREYYWFVTRNFSNFREFLFHAITSLLSLFPGGKMFQE
```

Fig. 4. Aminoacid composition of the translation product of ORF 406 as deduced from the nucleotide sequence, the deduced molecular weight is 46,100

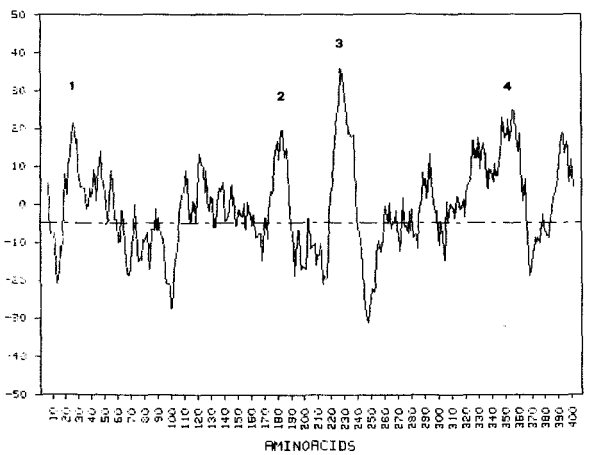

Fig. 5. Hydropathy plot of the 46,100 d protein of ORF 406. Each point on the profile corresponds to the sum of the hydropathy values of 11 adjacent aminoacids calculated according to Kyte and Doolittle (1982). Four hydrophobic region of 20 or more residues are marked 1 to 4. This computer analysis was kindly done by J. M. Erickson, University of Geneva

A complete 5S rRNA gene was retrieved, starting at position 3925. This was anticipated for the Z-strain according to results based on electron microscopic observations from Koller and Delius (1982b). An identical situation occurs also in the *bascillaris* and Z-S strain (Koller et al. 1984). We show in Fig. 3 the secondary structure model of a potential transcript and compare it with the sequence of Z-strain chloroplast 5S rRNA genes known to slightly diverge among the three *rrn* operons (Karabin et al. 1983). It is noteworthy that the s5S rRNA gene differs in eight positions all situated in the 5′ part, while sequence divergences of the 5S rRNA genes of *rrn* operons A, B, C are all located in the 3′ part. It seems that the s5S rRNA is not transcribed, since no corresponding rRNA product was identified (Karabin et al. 1983).

As structural interoperon spacer we define the DNA stretch between the 3′ end of the 5S rRNA gene and the 5′ end of the 16S rRNA gene of the next *rrn* operon. The functional *rrn* operon starts and ends somewhere within this region. According to a suggestion of EL-Gewely et al. (1984), the sequences 5′-TGGGACA (pos. 4782) and 5′-TAAAAT- (pos. 4826) may be promoter sites and the short indirect repeat 5′-ACTCGA- (pos. 4093) which is integral part of the "330" element may act as transcription terminator. These sequences are preserved in all interoperon spacers analysed so far. It is known for some time that two pseudo-tRNA genes (Ala, Trp) are proximate to the 5′ end of the 16S rRNA gene. We see them also in interoperon spacer 1 of the Z-strain (pos. 4906 and 4996).

A major open reading frame (ORF 406) starts at position 2290 (ATG) and terminates at position 3510 (TGA).

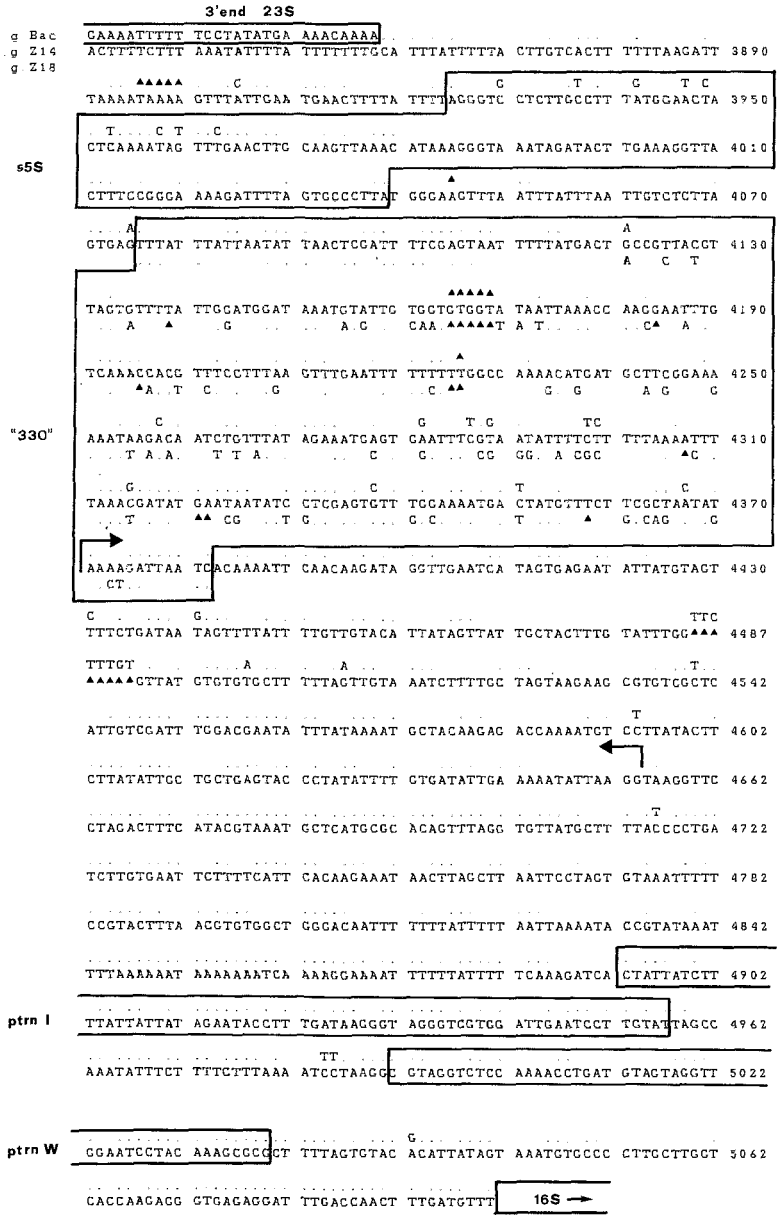

Fig. 6. Sequence comparison of interoperon spacer 1 of the Z- and *bacillaris* strain. Major structural elements are boxed. Within the "330" box we add the sequence of the "330"a element of the BG18 fragment of the Z-strain. Bent arrows mark the deletion in interoperon spacer 2 of the *bacillaris* strain (El-Gewely et al. 1984). We add the 3' part of the 23S rRNA gene of the bacillaris strain to show that in the Z-strain counterpart there is no 23S rRNA gene. Note, however, that upstream of the s5S rRNA gene of the *bacillaris* strain exists an inverted repeat of the "330" element (consult Fig. 7)

We give in Fig. 4 the deduced aminoacid sequence and in Fig. 5 the corresponding hydropathy plot. We recognize four hydrophobic domains (1 to 4) comprising at least 20 aminoacids required, e.g., to traverse a thylakoid membrane. The protein is not identical or related to any of the sequenced chloroplast proteins, but as will be shown elsewhere (Roux et al. in preparation) the gene is actively transcribed.

Discussion

Interoperon spacer 1 and the "330" element

El-Gewely et al. (1984) sequenced the *rrn* interoperon spacers 2 and 3 of the *Euglena gracilis bacillaris* strain including large parts of the interoperon spacer 1 and parts of the s5S rRNA gene. We compare in Fig. 6 the corres-

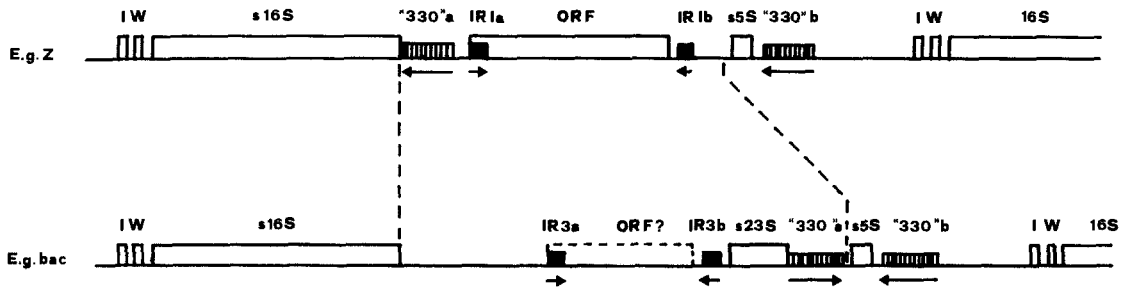

Fig. 7. Comparison of the mosaic structure of the DNA segment linking the s16S rRNA genes of the Z- and *bacillaris* strain. The structural elements are drawn in scale based on results from Koller et al. (1984) and El-Gewely et al. (1984) (*bacillaris*) and Roux et al. (1983) and this report (Z-strain). I, W = pseudogenes for Ile and Trp. Arrows indicate the relative orientation of repeats. The segment up-stream of the s5S rRNA gene including the s16S gene and its leader part of the *bacillaris* strain have not been sequenced

ponding region of the Z- and *bacillaris* strain and find a sequence homology of 97%. A most conspicuous difference is a five bases deletion (pos. 4165, Z-strain). The two strains may have separated about 15×10^6 years ago (see El-Gewely et al. 1984, for details of strain origin). Under this assumption it is surprising that the interoperon spacer 1, linking a "truncated" with a functional *rrn* operon is so well conserved, if on the other hand interoperon spacer 2 is 293 nucleotides shorter than interoperon spacer 3, which both link functional *rrn* operons (*bacillaris* strain). In the Z-strain the three interoperon spacers are of identical length and composition according to restriction site mapping (Hallick 1984), electron microscopic data (Koller and Delius 1982) and this and previous sequencing results (Graf et al. 1982).

A major DNA segment of the interoperon spacer ("330" element) is repeated proximate to the 3′ end of the s16S rRNA gene of the Z-strain. Probably the same segment is also repeated and inversely inserted upstream of the s5S of the *bacillaris* and Z-S strain according to electron microscopic observations (Koller et al. 1984). In Fig. 6 we align the two "330" elements of the Z-strain with the counterpart of the *bacillaris* strain. We calculated the degree of sequence divergence of the three "330" elements. The results are as follows: "330"a (Z-) : "330"b (Z-) = 0.23; "330"b (*bacillaris*) : "330"b (Z-) = 0.06; "330"a (Z-) : "330"b (*bacillaris*) = 0.20. According to these results the "330"a element drifted much more that the "330"b elements inside of the interoperon spacer 1 of both the Z- and *bacillaris* strain. It is noteworthy that the "330"a element of the Z-strain lacks the same five nucleotides as the "330"b element of the interoperon spacer 1 of the *bacillaris* strain. Since the "330"a element of the *bacillaris* strain has not been sequenced it is too early to establish a more elaborate pedigree of the "330" elements. We may nevertheless ask the question, whether the "330" elements qualify for some sort of mobile genetic element which may enhance DNA rearrangements like e.g. deletions and insertions (Starlinger 1977). A structural hallmark of IS elements and transposons are terminal repeats, what is not seen in our case. It will be of interest to analyse on a nucleotide level "330" elements from various Euglena strains and to search for such elements elsewhere on the chloroplast genomes.

Relics of chloroplast rrn operons and the ORF 406

Based on electron microscopic data (Koller et al. 1984) and nucleotide sequencing studies (Roux et al. 1983) it was postulated that the s16S rRNA gene is a relic of a once intact *rrn* operon. The present results give further support for this assumption in particular the findings that the interoperon spacer 1 is highly conserved and that a s5S rRNA gene is retained. It seems therefore highly plausible that due to a deletion/insertion event in the evolutionary past a DNA piece of about 3 kb was excised (the 16S-23S intergenic spacer with the *trn*I and *trn*A and the 23S rRNA gene) and replaced by an insert of about 2 kb consisting mainly of a directly repeated "330" element and a ORF flanked by short inverted repeats (IR1).

A somewhat similar situation is seen in the *bacillaris* strain. Koller et al. (1984) also observed the s16S and s5S rRNA genes separated by an insert which contains a "330" element, the short inverted repeats and a slightly shortened version of the ORF region. In addition they observed a short piece (300 to 350 nucleotides) of the 3′ part of the 23S rRNA gene. In Fig. 7 we compare the anatomical situation of the Z-strain with that of the *bacillaris* strain.

The question arises when in the evolutionary past such a major DNA rearrangement (s) occured, before or after separation of the two strains. An answer may be obtained by further sequencing the various structural elements upstream of the s5S rRNA gene of the *bacillaris* strain, notably the "330" elements, the inverted re-

peats and the segments in between. We favour right now the assumption that the major deletion/insertion event happened before strain separation what could explain the apparent identity of the ORF region and its flanking inverted repeats in both strains. The tandemly arranged *rrn* operons are prone to unequal-crossing events which easily lead to the observed variation in the number of *rrn* operons in the various Euglena chloroplast genomes. All the more it is surprising that the "truncated" *rrn* operon has survived during evolution in the various strains.

Acknowledgements. We are grateful to J. M. Erickson, University of Geneva, for the hydropathy plots and to C. Nager, Biocenter Basel, for comparing the possible ORF 406 translation product with about 2,000 known protein sequences. We thank C. Bachmann for secretarial help and the Fonds National Suisse de la Recherche Scientifique for support (E.S. 183.82). This work is part of the Ph.D thesis of Etienne Roux.

References

Cohen SW, Chang ACY, Hsu L (1972) Proc Natl Acad Sci USA 69:2110–2114
El-Gewely M, Helling RB, Dibbits JGTh (1984) Mol Gen Genet 194:432–443
Flamant F, Heizmann P, Nigon V (1984) Curr Genet 8:9–13
Graf L, Kössel H, Stutz E (1980) Nature 286:908–910
Graf L, Roux E, Stutz E, Kössel H (1982) Nucleic Acids Res 10:6369–6381
Hallick RB (1984) In: Buetow DE (ed) The Biology of Euglena, vol IV. Academic Press, New York, in press
Jenni B, Stutz E (1979) FEBS Lett 102:95–99
Jenni B, Fasnacht M, Stutz E (1981) FEBS Lett 125:175–179
Karabin GD, Narita JO, Dodd JR, Hallick RB (1983) J Biol Chem 258:14790–14796
Koller B, Delius H (1982a) Mol Gen Genet 188:305–308
Koller B, Delius H (1982b) FEBS Lett 140:198–202
Koller B, Delius H, Helling RB (1984) Plant Mol Biol 3:127–136
Kyte J, Doolittle RF (1982) J Mol Biol 157:105–132
Messing J, Vieira J (1982) Gene 19:269–276
Messing J, Crea R, Seeberg Ph (1980) Nucleic Acids Res 9:309–321
Ravel-Chapuis P, Flamant F, Nicolas P, Heizmann P, Nigon V (1984) Nucleic Acids Res 121:1039–1048
Roux E, Graf L, Stutz E (1983) Nucleic Acids Res 11:1957–1968
Sanger F, Coulson AR, Barrel BG, Smith AJH, Roe BA (1980) J Mol Biol 143:161–178
Schlunegger B, Fasnacht M, Stutz E, Koller B, Delius H (1983) Biochim Biophys Acta 739:114–121
Starlinger P (1977) Annu Rev Genet 11:103–126
Starr RC (1964) Am J Bot 51:1013–1044
Tait RC, Rodriguez RL, West RW (1980) J Biol Chem 255:813–815
Wartell RM, Reznikoff WS (1980) Gene 9:307–310

Communicated by U. Leupold

Received October 20, 1984