Nucleotide sequence of a *Euglena gracilis* chloroplast gene coding for the 16S rRNA: homologies to *E. coli* and *Zea mays* chloroplast 16S rRNA

L.Graf*, E.Roux and E.Stutz+
Laboratoire de Biochimie, Université de Neuchâtel, Chantemerle 18, CH-2000 Neuchâtel, Switzerland, and

H.Kössel+
Institut für Biologie III, Universität Freiburg, Schänzlestrasse 1, D-7800 Freiburg i. Br., FRG

Received 25 May 1982

ABSTRACT

The nucleotide sequence of 16S rDNA from Euglena gracilis chloroplasts has been determined representing the first complete sequence of an algal chloroplast rRNA gene. The structural part of the 16S rRNA gene has 1491 nucleotides according to a comparative analysis of our sequencing results with the published 5'- and 3'-terminal "T1-oligonucleotides" from 16S rRNA from E. gracilis. Alignment with 16S rDNA from Zea mays chloroplasts and E. coli reveals 80 and 72% sequence homology, respectively. Two deletions of 9 and 23 nucleotides are found which are identical in size and position with deletions observed in 16S rDNA of maize and tobacco chloroplasts and which seem to be characteristic for all chloroplast rRNA species. We also find insertions and deletions in E. gracilis not seen in 16S rDNA of higher plant chloroplasts. The 16S rRNA sequence of E. gracilis chloroplasts can be folded by base pairing according to the general 16S rRNA secondary structure model.

INTRODUCTION

It was recently shown that the 16S and 23S rRNA genes from Zea mays chloroplasts (1,2) have a high degree of nucleotide sequence homology with the E. coli rRNA genes, giving further support to the hypothesis that chloroplasts are descendents of cyanobacteria (3) which entered eukaryotic cells during the course of evolution. The overall organisation of the rDNA in chloroplast genomes from higher plants follows a rather uniform pattern, i.e., they have one or two rDNA regions per circular genome and the two rDNA regions are always localized in two distant inverted repeats (4). As an exception Vicia faba chloroplast DNA does not contain an inverted repeat region and its rRNA genes exist therefore as single copies (5). Contrary to that the chloroplast genome of the unicellular alga Euglena gracilis contains three rDNA regions localized in three contiguous tandem repeats (6,7,8) and it has an additional 16S rRNA gene (pseudogene ?) which is located about three kilobasepairs (kbp) away from the first rDNA repeat (9).

0305-1048/82/1020-6369$ 2.00/0

Obviously this chloroplast genome has a gene organization different from that of higher plants what is not a surprise in view of the large evolutionary distance (10). Therefore, it is of interest to analyse the rDNA region at the nucleotide level with the aim of comparing it with both a higher plant chloroplast and a bacterial rRNA gene.

In this paper we give the complete sequence of the small subunit rRNA gene of Euglena gracilis Z, including a short piece of the leader sequence. We compare our data with published oligonucleotide analysis of 16S rRNAs. We measure the degree of sequence homologies between the algal chloroplast 16S rDNA and that of E. coli and Zea mays chloroplasts, respectively.

MATERIALS AND METHODS

The BamHI-D fragment from Euglena gracilis Z, chloroplast DNA ligated into pBR322, clone pEgc 11, was used as starting material and grown as described (11). The plasmid DNA was isolated according to the unpublished protocol of M. Billeter and H. Weber (12) in order to obtain very pure supercoiled DNA. SDS-lysis and high salt precipitation (13,14) was followed by phenol extractions, RNase-treatment and precipitation by polyethyleneglycol-6000 (PEG). Further enrichment of supercoiled plasmid DNA versus nicked-circular or linear molecules was achieved by an alkali denaturation-renaturation step and phenolization (15) and final centrifugation on a CsCl-ethidiumbromide gradient.

Large subfragments (see Fig. 1) were separated by preparative electrophoresis on vertical 1-1.5% agarose slab gels and eluted electrophoretically according to (16) or (17) and subsequently purified by DEAE-cellulose chromatography (18). Smaller restriction fragments, suitable for DNA-sequencing, were prepared according to the restriction enzyme catalogue in Fig. 1. The fragment mixtures were treated with alkaline phosphatase, phenol extracted and end labeled with T4 polynucleotide kinase and [γ-^{32}P]-ATP as described in (19). Labeled fragment mixtures were separated on polyacrylamide gels (5 or 7.5%) and individual fragments were eluted by diffusion from crushed gel slices as outlined in (19), then they were either cut with a restriction enzyme or the strands were separated. Strand separation was achieved by denaturing with NaOH/glycerol (20) or dimethylsulfoxide (19) and subsequent electrophoresis on prerun polyacrylamide gels of 5 to 6% for

fragments >130 bp and 7.5-10% for fragments <130 bp. The gel dimensions were 0.2 x 20 x 40 cm, slot width 4 cm. Electrophoresis conditions were 4°C with less than 8 V/cm (19). Extracted double or single strand fragments were subjected to the hydrazine/dimethylsulphate sequencing method of Maxam and Gilbert (19). Usually we applied reactions R5 for G, R9 for A > C, R4 for C > T and R3 for T + C. The cleavage products were separated on 17% and 8% polyacrylamide-urea gels of 0.05 x 20 x 40 cm and sometimes of 0.15 x 22 x 100 cm. Autoradiography was performed either by direct exposure of gels at -20°C, or after drying the gels at room temperature, using either Agfa Osray T4 or Kodak X-AR-5 films without intensifying screen.

Restriction endonucleases were purchased from New England Biolabs or Bethesda Research Laboratories or Boehringer-Mannheim. The enzyme BglII was isolated as described in (21). Bacterial alkaline phosphatase was obtained from Worthington and dialyzed prior to use against 100 mM Tris-HCl (pH 8.0), 1 mM $MgCl_2$ and 50% glycerol. T4 polynucleotide kinase was either from New England Biolabs or New England Nuclear (NEN).

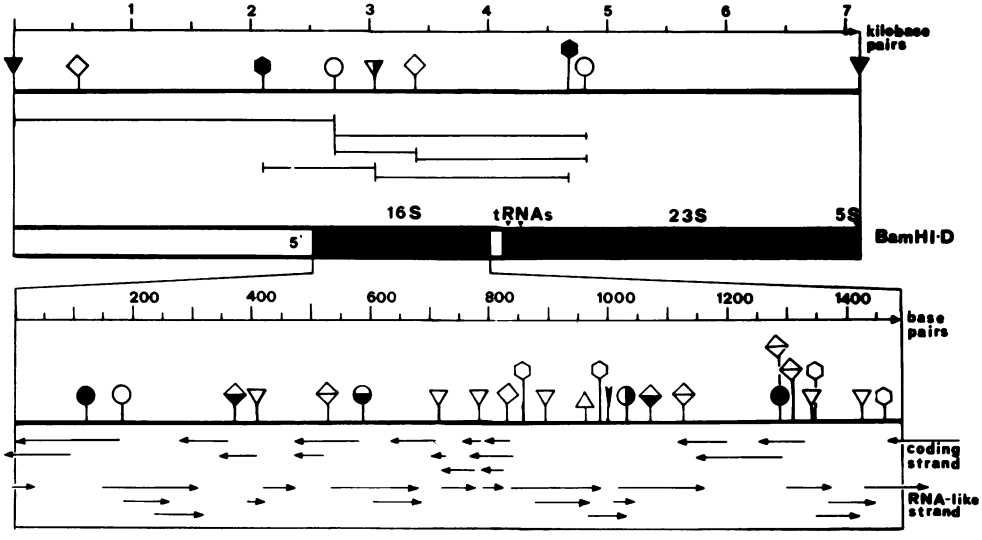
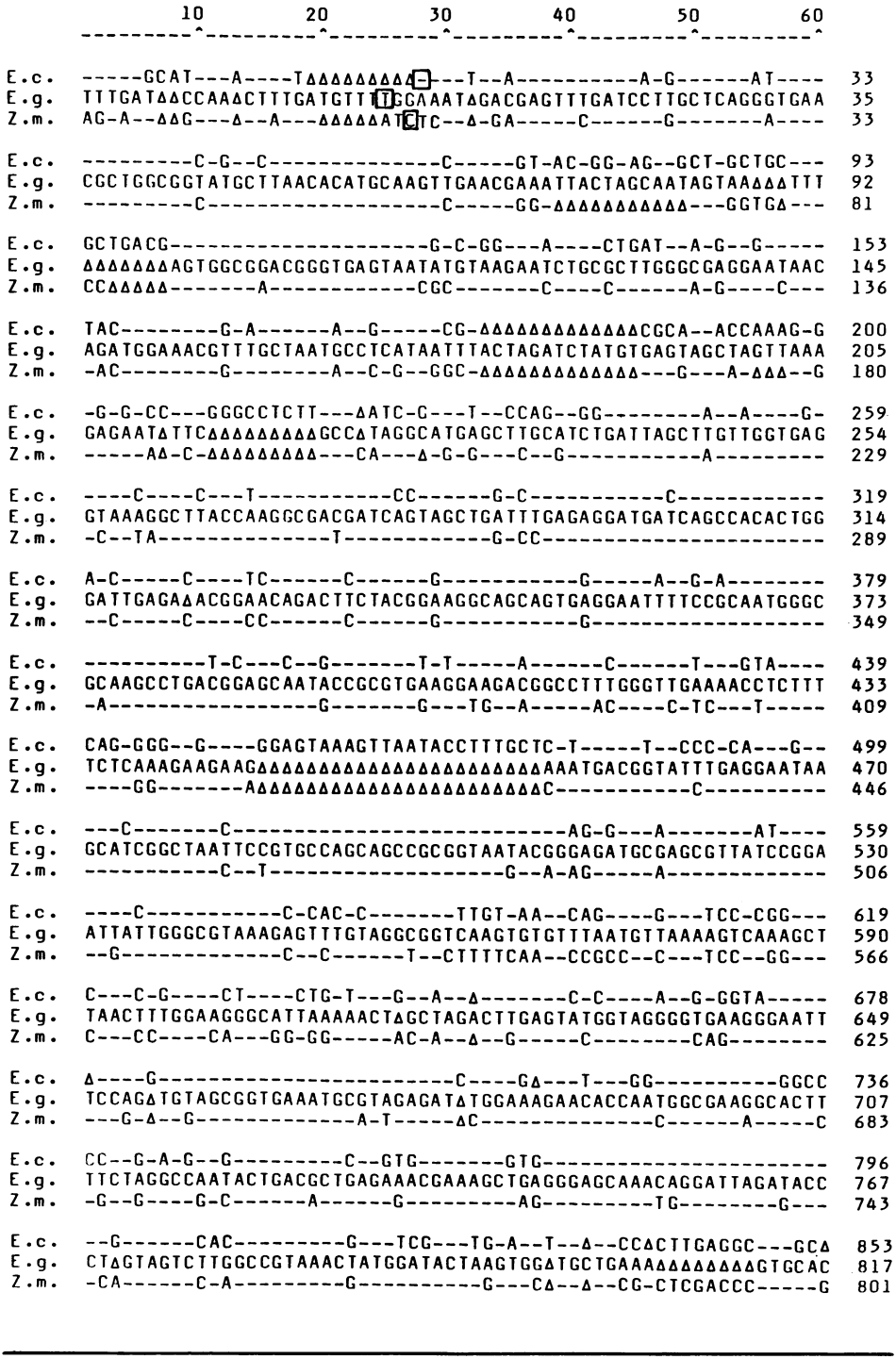


Fig. 1 : Physical mapping and use of restriction fragments for sequence analysis of 16S rDNA from E. gracilis, Z, chloroplasts. For details see text. Restriction endonuclease cleavage sites are symbolized as follows: ◑, Alu I; ▼, BamHI; ○, BglII; ⬢, EcoRI; ▽, HaeIII; ◈, HhaI; ◒, HindIII; ●, HinfI; ◇, HpaI; ◈, HpaII; ⬡, RsaI; ▼, SacII; △, Sau 3A; ▼, XbaI.


10
20
30
40
50
60
E.c. -----GCAT---A----TΔΔΔΔΔΔΔΔΔ-T--A----------A-G------AT---- 33
E.g. TTTGATΔΔCCAAΔCTTTGATGTTTTGGAAATΔGACGAGTTTGATCCTTGCTCAGGGTGAA 35
Z.m. AG-A--ΔΔG---Δ--A---ΔΔΔΔΔATCTC--Δ-GA-----C------G-------A---- 33
E.c. ---------C-G--C--------------C-----GT-AC-GG-AG--GCT-GCTGC--- 93
E.g. CGCTGGCGGTATGCTTAACACATGCAAGTTGAACGAAATTACTAGCAATAGTAΔΔΔΔTTT 92
Z.m. ---------C-------------------C-----GG-ΔΔΔΔΔΔΔΔΔΔΔ---GGTGΔ--- 81
E.c. GCTGACG--------------------G-C-GG---A----CTGAT--A-G--G----- 153
E.g. ΔΔΔΔΔΔΔAGTGGCGGACGGGTGAGTAATATGTAAGAATCTGCGCTTGGGCGAGGAATAAC 145
Z.m. CCΔΔΔΔΔ-------A----------CGC-------C----C------A-G----C--- 136
E.c. TAC--------G-A------A--G-----CG-ΔΔΔΔΔΔΔΔΔΔΔΔΔΔCGCA--ACCAAAG-G 200
E.g. AGATGGAAACGTTTGCTAATGCCTCATAATTTACTAGATCTATGTGAGTAGCTAGTTAAA 205
Z.m. -AC--------G--------A--C-G--GGC-ΔΔΔΔΔΔΔΔΔΔΔΔΔ---G---A-ΔΔA--G 180
E.c. -G-G-CC---GGGCCTCTT---ΔATC-G---T--CCAG--GG--------A--A----G- 259
E.g. GAGAATΔTTCΔΔΔΔΔΔΔΔΔGCCΔTAGGCATGAGCTTGCATCTGATTAGCTTGTTGGTGAG 254
Z.m. -----AΔ-C-ΔΔΔΔΔΔΔΔΔ---CA---Δ-G-G---C--G----------A--------- 229
E.c. ----C----C---T----------CC------G-C----------C---------- 319
E.g. GTAAAGGCTTACCAAGGCGACGATCAGTAGCTGATTTGAGAGGATGATCAGCCACACTGG 314
Z.m. -C--TA--------------T----------G-CC--------------------- 289
E.c. A-C-----C----TC------C------G----------G-----A--G-A-------- 379
E.g. GATTGAGAΔACGGAACAGACTTCTACGGAAGGCAGCAGTGAGGAATTTTCCGCAATGGGC 373
Z.m. --C-----C----CC------C------G----------G------------------- 349
E.c. ----------T-C---C--G-------T-T-----A------C------T---GTA---- 439
E.g. GCAAGCCTGACGGAGCAATACCGCGTGAAGGAAGACGGCCTTTGGGTTGAAAACCTCTTT 433
Z.m. -A----------------G-------G---TG--A-----AC----C-TC---T----- 409
E.c. CAG-GGG--G----GGAGTAAAGTTAATACCTTTGCTC-T-----T--CCC-CA---G-- 499
E.g. TCTCAAAGAAGAAGΔΔΔΔΔΔΔΔΔΔΔΔΔΔΔΔΔΔΔΔΔΔΔΔAAATGACGGTATTTGAGGAATAA 470
Z.m. ----GG-------AΔΔΔΔΔΔΔΔΔΔΔΔΔΔΔΔΔΔΔΔΔΔC----------C---------- 446
E.c. ---C-------C----------------------------AG-G---A-------AT---- 559
E.g. GCATCGGCTAATTCCGTGCCAGCAGCCGCGGTAATACGGGAGATGCGAGCGTTATCCGGA 530
Z.m. ----------C--T-------------------G--A-AG-----A------------- 506
E.c. ----C-----------C-CAC-C--------TTGT-AA--CAG----G---TCC-CGG--- 619
E.g. ATTATTGGGCGTAAAGAGTTTGTAGGCGGTCAAGTGTGTTTAATGTTAAAAGTCAAAGCT 590
Z.m. --G-------------C--C------T--CTTTTCAA--CCGCC--C---TCC--GG--- 566
E.c. C---C-G----CT----CTG-T---G--A--Δ--------C-C----A--G-GGTA----- 678
E.g. TAACTTTGGAAGGGCATTAAAAACTΔGCTAGACTTGAGTATGGTAGGGGTGAAGGAATT 649
Z.m. C---CC----CA---GG-GG-----AC-A--Δ--G-----C--------CAG-------- 625
E.c. Δ----G----------------------C----GΔ---T---GG----------GGCC 736
E.g. TCCAGΔTGTAGCGGTGAAATGCGTAGAGATΔTGGAAAGAACACCAATGGCGAAGGCACTT 707
Z.m. ---G-A--G-------------A-T-----ΔC--------------C------A-----C 683
E.c. CC--G-A-G--G---------C--GTG-------GTG---------------------- 796
E.g. TTCTAGGCCAATACTGACGCTGAGAAACGAAAGCTGAGGGAGCAAACAGGATTAGATACC 767
Z.m. -G--G----G-C------A------G---------AG---------TG--------G--- 743
E.c. --G------CAC---------G---TCG---TG-A--T--Δ--CCΔCTTGAGGC---GCΔ 853
E.g. CTΔGTAGTCTTGGCCGTAAACTATGGATACTAAGTGGΔTGCTGAAAΔΔΔΔΔΔΔΔΔGTGCAC 817
Z.m. -CA------C-A---------G----------G---CΔ--Δ--CG-CTCGACCC-----G 801

```
          10         20         30         40         50         60
          ^----------^----------^----------^----------^----------^----------^

E.c.  -T-C-G--C----G--------CGA----------------GCC---AGGT-A-------  913
E.g.  TGCTGTAGTTAACACGTTAAGTATCCCGCCTGGGGAGTACGCTTGCACAAGTGAAACTCA  877
Z.m.  --------C----G--------------------------T-C---AG-A---------  861

E.c.  --T-----------------------------------------------------G--  973
E.g.  AAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGATGCAACACG  937
Z.m.  -----------------------------------------------------GG--  921

E.c.  ------------T--TC------ΔΔΔ-Δ-C-C--------TC-G-GAT--GAATΔ--G--  1028
E.g.  AAGAACCTTACCAGGATTTGACAGGATACTAΔGGAAGTTTGΔAAAΔΔΔGAACGCAGTACC  991
Z.m.  ----------------GC------T-CCG-G-ΔATCCTC---Δ---ΔΔΔ--GA-GG--G--  976

E.c.  ------A-C-GT--G-------C----------------------T----A-------  1088
E.g.  TTCGGGTATCTAGACACAGGTGGTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGG  1051
Z.m.  C-----A-CGCG-----------------------------------C---A--G------  1036

E.c.  ---------------------------A-CC--TG--GC-A--ΔΔG---CGGCCGG---  1146
E.g.  GTTAAGTCCCGCAACGAGCGCAACCCTTTTTTTTAATTAACΔGCTATGTCAΔTTTΔAGAA  1107
Z.m.  --------T------------------CG-G----G--GC-ΔA--A--ΔΔ-G---ΔG---  1092

E.c.  CTCAAAGG-G-----CA--G--A-A-T-----------G---T-----------------  1206
E.g.  ΔΔΔΔΔΔΔΔATACTGCTGGTTATTAΔCCGGAGGAAGGTGAGGACGACGTCAAGTCATCATG  1158
Z.m.  CCCTGAAC-G--C--C---GT-A-G-----------A-----T--G-C------------  1152

E.c.  G------CGA--A----------------------ΔCGCAT----AG--AA--G-CC-C  1265
E.g.  CCCCTTATATCCTGGGCTACACACGTGCTACAATGGTTAAGΔACAATAAGTTGCAATTTT  1217
Z.m.  --------GC-------G------------------GCGG-Δ----AGG--C--G--C-C  1211

E.c.  -C--G-GCA---GG-CC--A-Δ---G-GC-T-GT---C-------G--T---C------A  1324
E.g.  GTGAAAATGAGCTAAΔΔTCTTAAAACTTAGCCTAAGTTCGGATTGTAGGCTGAAACTCGC  1275
Z.m.  -C--GGC--------CΔ--CA-----CCGT---C-----------C------C-------  1270

E.c.  --C-------T------------------TΔ--A----AΔ--G-CA--------------  1382
E.g.  CTACATGAAGCCGGAATCGCTAGTAATCGCCGGΔTCAGCTATACGGCGGTGAATACGTTC  1334
Z.m.  --G--------A----------------------Δ-----C---------C----C-----  1329

E.c.  C-----------------------------------G--ΔT--G-Δ---AAAA-----AGG-  1440
E.g.  TCGGGCCTTGTACACACCGCCCGTCACACCATGGAAGTCGGCTGTGCCCΔΔGAAGTTATT  1392
Z.m.  C----------------------------T--A-G--CT---CAG-TTTΔΔ-----C---  1387

E.c.  -GΔ---AA---ΔTCGG-----CGCT----ACTTTGΔ--AT-CA----------------  1497
E.g.  ATΔCTTGΔCCTGAAAAGAGGGΔAAATACCTAAGGCCTGGCTGGTGACTGGGGTGAAGTCG  1449
Z.m.  -CC---AA--Δ-T--G-----ΔGG--G---------TA----T-C------A---------  1445

E.c.  ----------A-----GG----CC--C--T----T--C----TTA                 1542
E.g.  TAACAAGGTAGCCGTACTGGAAGGTGTGGCTGGAACAACTCC                    1491
Z.m.  -------------------------C-------T--C----TTT                 1490
```

Fig. 2 : Nucleotide sequence of the 16S rDNA from E. gracilis, Z, chloroplasts. Only the RNA-like strand is given (E.g.) and aligned with it are the 16S rDNA from E. coli (E.c.) and Zea mays chloroplasts (Z.m.). Dashes in lines for E.c. and Z.m. indicate nucleotide identity with the E. gracilis sequence. Open triangles within sequences symbolize deleted positions as compared to the other sequences. The counting of positions starts with the 5' end of the 16S rRNA gene. For each of the three sequences the accumulated positions are given at the right margins. 21 position of the leader (E.g.) are tentatively aligned with the leader positions of E. coli (E.c.) and Zea may (Z.m.).

[γ-^{32}P]-ATP (3000 Ci/mmol) was from Radiochemical Center, Amersham or from NEN.

RESULTS AND DISCUSSION

Sequencing strategy

The circular chloroplast genome of E. gracilis is very well characterized (rev. 22) and large parts of the genome have been cloned. For the previous (23) and present sequencing studies we used the clone pEgc 11 (11) which contains the fragment BamHI·D with a complete rDNA operon as depicted in the upper part of Fig. 1. DNA fragments subsequently used for sequencing are drawn in thin lines and aligned with the respective restriction sites and the rRNA genes previously mapped in BamHI·D (24). In the lower part we show the 16S rDNA with the sites for restriction enzymes used to obtain the fragments for the 5' end labeling. The arrows represent portions of the 5' end labeled fragments (coding strand and/or RNA like strand) from which unambiguous sequences could be established.

Gene size

In Fig. 2 the RNA like strand of the entire 16S rDNA of the E. gracilis chloroplast genome including 21 positions of the leader are given. The boxed T (followed by a G) is considered to be the 5' end of the 16S rRNA gene corresponding to the 5' pUG dinucleotide listed by Zablen et al. (25). A second TG (position 7-8) is less likely to be the 5' end because the T1-oligonucleotide AAAUG which is present in three copies in the entire 16S rRNA molecule (25) would then occur only twice. The 3' end is exactly defined corresponding to the published 3'-terminal nonamer (25). With this assumption a chain length of 1491 positions results. This is 51 positions less than the 16S rRNA of E. coli and very close to the length of Zea mays chloroplast 16S rRNA for which the exact termini are not identified (1). A similar reduction of the chain length is observed for 16S rRNA from tobacco chloroplasts (26) for which the positions of the termini have been determined by the S1-technique.

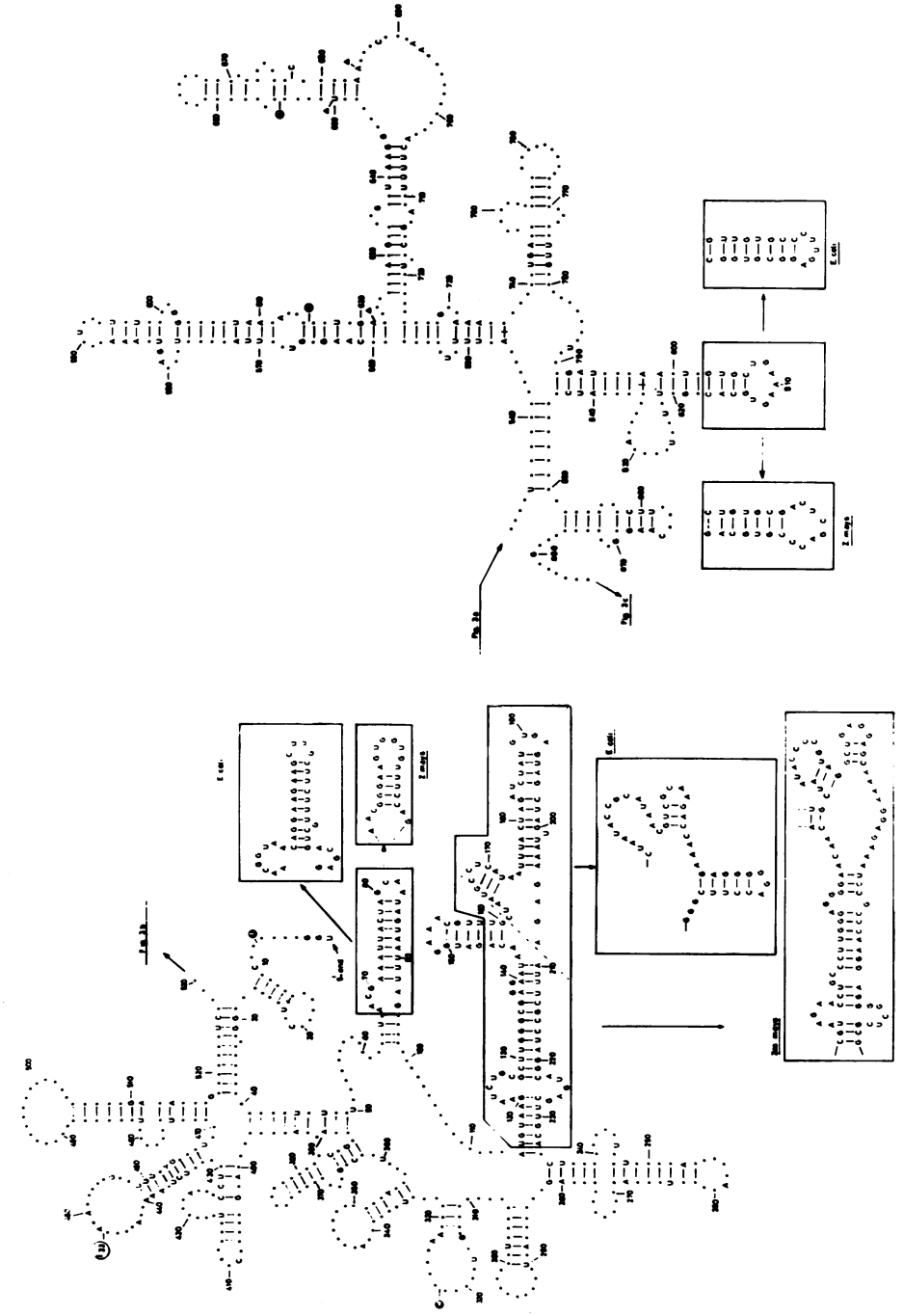
Comparison with the catalogue of "T1-oligonucleotides"

Zablen et al. (25) published a list of oligonucleotides obtained by digestion of E. gracilis 16S rRNA. Most of those oligonucleotides were found. In table 1 we complete their list with some additional information

Table 1 - Supplementary list to the published catalog from Zablen et al. (25) of T1 ribonuclease generated oligonucleotides of 16S rRNA as deduced from 16S rDNA sequencing results

	Number of oligonucleotides		
	Zablen et al.	Sequencing data	Position
Modified oligomers			
$C\overset{*}{C}CCG$	1	0	
Pentamers			
AAACG	1	2	156,736
AUCAG	1	2	281,305
UUAAG	3-2	3	838,1057,1199
ACUUG	1-2	1	625
AUUAG	2	1	762
Hexamers			
CACAAG	1	2	868,903
Heptamer			
CAACACG (new)		1	937
Octamer			
CAAUUUUG (new)		1	1218
16 mer			
CCUCAUAAUUUACUAG (new)		1	182
17 mer			
CUAAUCUUAAAACUUAG (new)		1	1245
19 mer			
AAAACCUCUUUUCUCAAAG (new)		1	441
22 mer			
CAACCCUUUUUUUUAAUUAACG (new)		1	1093

especially concerning the stoichiometry of oligonucleotides, and the sequences of four sofar uncharacterized larger oligonucleotides (16 mer, 17 mer, 19 mer, 22 mer). One of their modified oligomers ($C\overset{*}{C}CCG$) was not found, on the other hand, we found a 7 mer (CAACACG, pos. 937) which is



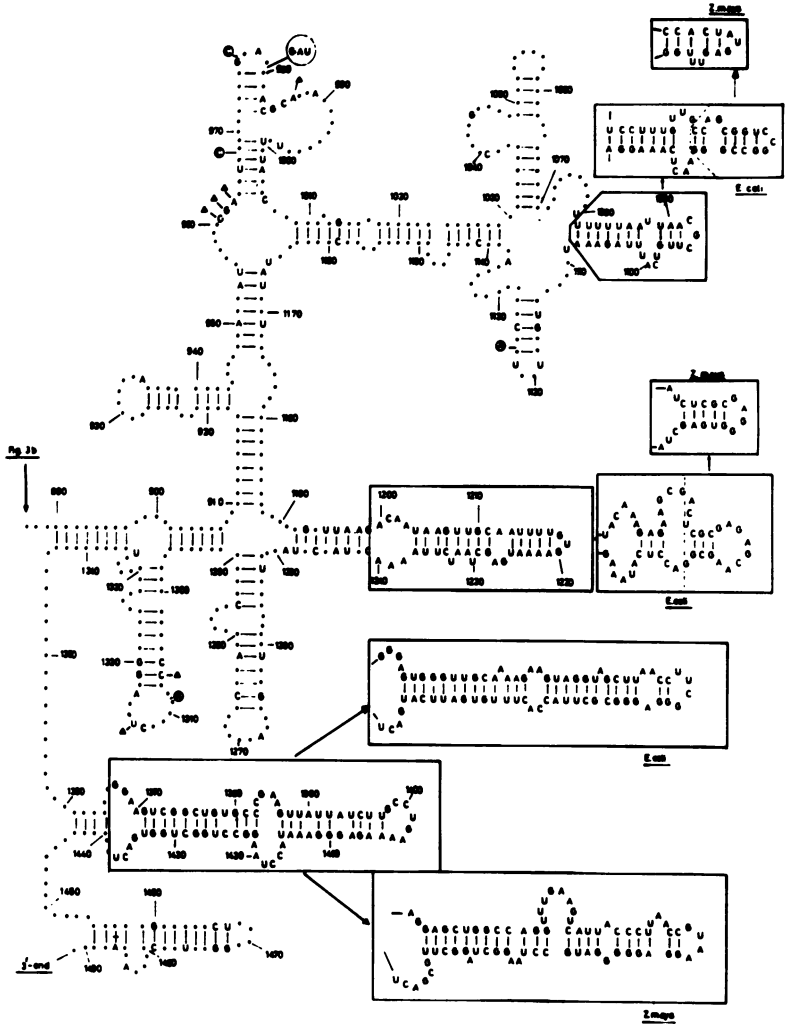


Fig. 3a-c : Secondary structure model of 16S rRNA from E. gracilis chloroplasts. The model was constructed in analogy to the E. coli / maize 16S rRNA model of Zwieb et al. (30). It is divided into the 5' terminal domain (a), a central domain (b) and a 3' terminal domain (c). The positions are numbered as given in Fig. 2 (E.g.). Single base deletions in the E. coli sequence are symbolized by △. Single base insertions in the E. coli sequence are indicated by circled bases near the points of insertion. I_{23} indicates an insertion of 23 bases in the E. coli sequence. Dots indicate sequence identity with E. coli 16S rRNA. Major differences between the three sequences (including maize) are separately given within boxes. Base pairs only possible in the E. coli but not in the Euglena 16S rRNA structure are marked by a + between the bases (e.g. at position 15/23).

related to the 7 mer CAACUCG normally found in bacterial 16S rRNA. Furthermore, Zablen et al. mention as probable sequence of one of the 8 mers AAUCUUUG; we find an 8 mer of identical base composition but of the sequence CAAUUUUG.

Sequence homology between E. gracilis chloroplast, maize chloroplast and E. coli 16S rDNA

The three 16S rDNA sequences are compared in Fig. 2. Out of the 1542 nucleotides from 16S rDNA (27) or rRNA (28) of E. coli 1117 positions or 72% and out of 1490 nucleotides of 16S rDNA of maize chloroplast 1192 positions or 80% are identical in E. gracilis chloroplast 16S rDNA. The maximum number of consecutive identical nucleotides for E. coli is 54 (positions 917-970) and for maize 65 (positions 853-917). It is noteworthy that in spite of the high degree of sequence homologies only 15 and 22 out of 40 tested restriction enzyme cleavage sites are identical in 16S rDNA of E. gracilis chloroplasts and E. coli or maize chloroplasts, respectively.

Chloroplast 16S rRNA species from maize and tobacco show 96% homology to each other (26). The decrease to 80% homology between the maize and Euglena chloroplast 16S rRNA species clearly reflects the higher divergence between higher plant and Euglena chloroplasts and substantiates an "outsider position" of Euglena within the plant kingdom as known from other criteria (10).

Deletions of 9 and 23 bases are found at positions 214 and 447, respectively, and deletions identical in size and position are observed in the maize and tobacco (26) chloroplast 16S rRNAs. The 23 bases deletion is also found in 16S rRNA from mustard chloroplasts (E. Fritzsche, G. Link and H. Kössel, unpublished). Therefore these two deletions (and partly the deletion at position 93) appear to be characteristic for all chloroplast 16S rRNA species. In contrast to this the 11 bases deletion at position 71 of the maize sequence is only found in the tobacco (26) and mustard sequences but not in the E. gracilis sequence. Therefore this deletion seems to be a characteristic of higher plant chloroplast 16S rRNA only. The 13 bases insertion at position 177 and the 8 bases deletions at position 811 and 1107 of the E. gracilis sequence are absent so far in all other chloroplast (and E. coli) 16S rRNA species. Whether they are characteristic for algal chloroplast 16S rRNA remains to be seen by

comparison with other algal 16S rRNA species.

Secondary structure model

As depicted in Fig. 3 the E. gracilis 16S rRNA sequence can be folded by base pairing according to the general 16S rRNA secondary structure model (29,30,31). This model is confirmed by many compensating base substitutions of the E. gracilis sequence as compared to the corresponding E. coli and maize sequence. The changes (substitutions, deletions and insertions) in comparison to the E. coli 16S rRNA structure are indicated in the Fig. 3. Major changes in comparison to both E. coli and maize 16S rRNA are depicted in separate boxes as these regions show more rapid divergence (on the primary and secondary structure level) than the rest of the molecule. For instance the region between position 115 and 233 differs between the three species mainly due to the insertion of 13 bases (178-190) in the algal species and to the deletion of 9 bases at position 214 (E. gracilis) in the two chloroplast species. The former gives rise to an extension of the 175-205 stem loop structure in the algal species whereas the latter amputates an extra stem loop structure present only in the E. coli species (lower part of the E. coli box in Fig. 3a). A similar amputation of an entire stem loop structure is caused by the 23 bases deletion at position 447. Other stem loop structures e.g. at positions 67-94, 803-817, 1200-1240 and 1367-1437 show only variability in lengths of the paired and unpaired regions or in the position, size and number of "bulges".

Though there is agreement in respect to the overall structure and to many details of the three 16S secondary structure models (29,30,31), the model proposed by Zwieb et al. (30) shows a considerably higher number of base paired regions such e.g., on the stem structures 11-14⌒24-27, 48-50⌒388-390, 63-66⌒95-98, 185-188⌒193-196, 334-339⌒351-356 in the 5' terminal domain and a few others in the rest of the molecule. The conservation of these base pairs in the E. gracilis structure lends support to the more tightly base paired model of Zwieb et al. (30).

ACKNOWLEDGEMENT

We thank Dr. Zs. Schwarz, who initiated sequencing in this system. We also thank Dr. K. Edwards for helpful advice. The help of B. Schlunegger, M. Eilers and E. Schiefermayr is gratefully acknowledged. This work was

supported by grants from the Fonds national suisse de la Recherche scientifique, no. 3-452.79 to E.S. and from the Deutsche Forschungsgemeinschaft (SFB 46) to H.K.

+To whom to address correspondence
*Present address: Friedrich Miescher-Institut, P.O. Box 273, CH-4002 Basel, Switzerland.

REFERENCES

1. Schwarz, Zs. and Kössel, H. (1980) Nature 283, 739-742
2. Edwards, K. and Kössel, H. (1981) Nucleic Acids Res. 9, 2853-2869
3. Bonen, L., Doolittle, W.F. and Fox, G.E. (1979) Can. J. Biochem. 57, 879-888
4. Bedbrook, J.R. and Kolodner, R. (1979) Ann. Rev. Plant Physiol. 30, 593-620
5. Koller, B. and Delius, H. (1980) Molec. Gen. Genet. 178, 261-269
6. Jenni, B. and Stutz, E. (1978) Eur. J. Biochem. 88, 127-134
7. Gray, P.W. and Hallick, R.B. (1978) Biochemistry 17, 284-290
8. Rawson, J.R.Y., Kushner, S.R., Vapnek, D., Alton, N.K. and Boerma, C.L. (1978) Gene 3, 191-209
9. Jenni, B. and Stutz, E. (1979) FEBS Lett. 102, 95-99
10. Johnson, L.P. (1968) in: The Biology of Euglena (D.E. Buetow, ed.) vol. I, pp. 1-26, Academic Press, New York
11. Knopf, U.C. and Stutz, E. (1978) Molec. Gen. Genet. 163, 1-3
12. Billeter, M. and Weber, H. (1978) Institut für Molekularbiologie I, Universität Zürich, personal communication
13. Guerry, P., LeBlanc, D.J. and Falkow, S. (1973) J. Bacteriol. 116, 1064-1066
14. Hirt, B. (1967) J. Mol. Biol. 26, 365-369
15. Currier, Th.C. and Nester, E.W. (1976) Anal. Biochem. 76, 431-441
16. Wienand, U., Schwarz, Z. and Feix, G. (1979) FEBS Lett. 98, 319-323
17. Yang, R.C.-A., Lis, J. and Wu, R. (1979) in:Methods in Enzymology (Wu, R. ed.) vol. 68, pp. 176-182, Academic Press, New York
18. Smith, H.O. and Birnstiel, M.L. (1976) Nucleic Acids Res. 3, 2387-2399
19. Maxam, A.M. and Gilbert, W. (1980) in: Methods in Enzymology (Grossman, L. and Moldave, K., eds) vol. 65, pp. 499-560, Academic Press, New York
20. Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560-564
21. Bickle, Th.A., Pirrotta, V. and Imber, R. (1977) Nucleic Acids Res. 4, 2561-2572
22. Hallick, R.B. in: The Biology of Euglena (D.E. Buetow, ed.) vol. IV, Academic Press, New York, in press
23. Graf, L., Kössel, H. and Stutz, E. (1980) Nature 286, 908-910
24. Orosco, E.M., Gray, P.W. and Hallick, R.B. (1980) J. Biol. Chem. 255, 10991-10996
25. Zablen, L.B., Kissil, M.S., Woese, C.R. and Buetow, D.E. (1975) Proc. Natl. Acad. Sci. USA 72, 2418-2422
26. Tohdoh, N. and Sugiura, M. (1982) Gene 17, 213-218

27. Brosius, J., Palmer, M.L., Kennedy, P.J. and Noller, H.F. (1978) Proc. Natl. Acad. Sci. USA 75, 4801-4805
28. Carbon, P., Ehresmann, C., Ehresmann, B. and Ebel, J.P. (1978) FEBS Lett. 94, 152-156
29. Noller, H.F. and Woese, C.R. (1981) Science 212, 403-411
30. Zwieb, Ch., Glotz, C. and Brimacombe, R. (1981) Nucleic Acids Res. 9, 3621-3640
31. Stiegler, P., Carbon, P., Ebel, J.-P. and Ehresmann, C. (1981) Eur. J. Biochem. 120, 487-495