Published in Nature 286, issue 5776, 908 - 910, 1980
which should be used for any reference to this work

Sequencing of 16S–23S spacer in a ribosomal RNA operon of *Euglena gracilis* chloroplast DNA reveals two tRNA genes

L. Graf, H. Kössel* & E. Stutz

Laboratoire de Biochimie, Université de Neuchâtel, Chantemerle 20, CH-2000 Neuchâtel, Switzerland
* Institut für Biologie III der Universität Freiburg, Schänzlestrasse 1, D-7800 Freiburg i. Br, FRG

Chloroplasts of the unicellular flagellate eukaryote *Euglena gracilis* contain several copies of a circular 135–140-kilobase pair DNA[1] which codes for chloroplast-specific stable RNAs (16S, 23S (refs 2, 3), 5S rRNAs[4] and tRNAs[5]) and for an unknown number of chloroplast-specific proteins. The rRNA genes are located within three tandemly repeated DNA regions of approximately 5.6 kilobase pairs each[6–8] and the arrangement of the structural genes within each repeat follows the prokaryotic pattern, being 5′-16S-23S-5S-3′ (ref. 9). Total chloroplast tRNA hybridizes to fragments of rDNA[9] and it was suggested that the 16S–23S spacer region contains tRNA coding sequences as is observed in *Escherichia coli*[10,11] and in spinach chloroplast[12] rDNA. We have therefore analysed *E. gracilis* strain Z 16S–23S spacer DNA at the nucleotide level, hoping this would allow identification of tRNA genes together with the processing sites of the respective primary transcripts. Maize chloroplast 16S rDNA shows strong sequence homology with *E. coli* 16S rRNA[13]. Sequence analysis of a total spacer in *E. gracilis* should demonstrate whether such similarities are also preserved in the chloroplast rDNA spacer region, or if this region has suffered a higher genetic drift rate. The latter is suggested from the 189 bases which have been sequenced from the 2.4-kilobase pair rDNA spacer from maize chloroplasts[14]. Flanking sequences, coding for the 3′-terminal region of 16S rRNA and for the 5′-terminal region of 23S rRNA have also been sequenced, to compare the drift rates between the spacer and its adjacent structural genes.

Figure 1 shows the protocol which was followed for mapping and identification of the spacer region within the restriction fragment *Bam*HI·D. This fragment, which contains one of the three rDNA repeats, has previously been cloned using the plasmid pBR322 as vector[15] and the respective clone pEgc11 (formerly pEgcKS11) was used for isolation of fragment *Bam*HI·D and its subfragments. Partial sequencing of several of the subfragments[16] revealed strong homology with regions from *E. coli*[17,18] and maize chloroplast[13] 16S rRNA (rDNA) and thereby allowed location of the 3′-terminal region of the 16S rRNA gene on the right-hand third of fragment *Hae*III$_{225}$. Since hybridization and fine-mapping data had indicated[6,7,19] that the spacer should not exceed 300 base pairs, the left-hand two-thirds of fragment *Hae*III$_{225}$ together with the neighbouring portion of fragment *Bgl*II-*Hae*III$_{650}$ should comprise the entire spacer sequence. These fragments were therefore chosen for sequence analysis according to the dimethylsulphate/hydrazine procedure[20] and the necessary overlap between the two *Hae*III fragments was obtained by sequencing the left-hand portion of fragment *Hin*fI$_{430}$.

In Fig. 2 the RNA-like strand of the *E. gracilis* spacer DNA sequence is depicted alongside the respective sequences from the *E. coli* rrnD operon. The first 63 positions comprise the 3′-terminal region of the 16S rRNA gene, the 'prokaryotic character' of which is documented by its high degree of homology with the corresponding rDNA region from *E. coli*. The reported 3′-terminal oligonucleotide AACAACUCN of *E. gracilis* 16S rRNA[21] allows exact positioning of the 3′-terminal end of the structural part of this gene. The reported T_1 oligonucleotide UAACAAG[21] which is highly conserved in prokaryotic 16 S rRNA[22] is found at position 22–28. The non-

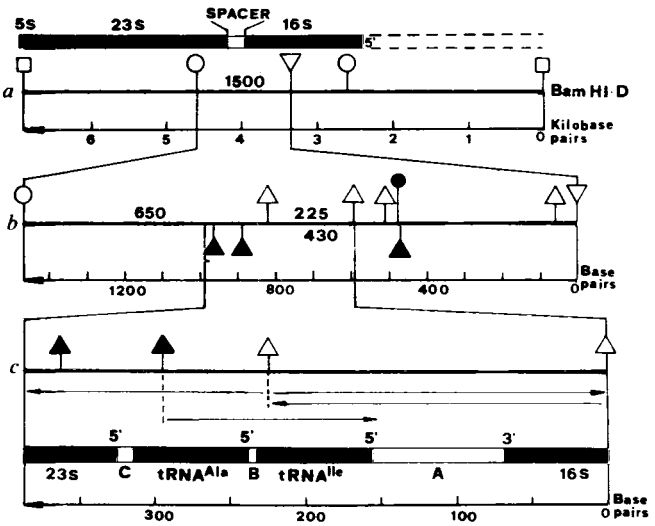

Fig. 1 *a*, Position of the *E. gracilis* rDNA region and of the spacer-containing fragment *Bgl*II–*Hpa*I$_{1,500}$ within the cloned fragment *Bam*HI·D (pEgc11)[8]. *b*, Fine mapping of fragment *Bgl*II–*Hpa*I$_{1,500}$ using the enzymes *Hae*III, *Hpa*II and *Hin*fI. *c*, Fragments used for sequence analysis of the spacer region. The arrows represent portions of 5′-terminally labelled single fragments from which unambiguous sequences could be obtained by a series of overlapping gels. Nearly all positions of fragment *Hae*III$_{225}$ have been analysed from both strands. Sequences of the right-hand portion of fragment *Bgl*III–*Hae*III$_{650}$ could be confirmed by analysing part of the complementary strand of fragment *Hin*fI$_{430}$, which also gives the necessary overlap between fragments *Bgl*II–*Hpa*I$_{650}$ and *Hae*III$_{225}$. Positions of rRNA and tRNA coding sequences and of the intergenic regions A, B and C are indicated. □, *Bam*HI; ▽, *Hpa*I; △, *Hae*III; ○, *Bgl*I; ●, *Hpa*II; ▲, *Hin*fI.

conserved position 59 is of particular interest, as it partly changes the recognition sequence for translational initiator regions[23,24]. This suggests that *E. gracilis* chloroplast ribosomes have a slightly altered pattern of recognition for mRNA in comparison with *E. coli* whereas in maize chloroplast ribosomes the corresponding sequence is identical to the *E. coli* 16S rRNA 3′ end[14].

A stretch of 66 nucleotides shown in the lower part of Fig. 2 represents the 5′-terminal portion of the 23S rRNA gene as evidenced by its extensive homology with the 5′-terminal region of *E. coli* 23S rDNA[10,11]. However, since the 5′-terminal oligonucleotides of *E. gracilis* 23S rRNA are not known, exact positioning of the 5′ terminus of the 23S rRNA gene is not possible. The 16S–23S spacer DNA is thus represented by the 259 base pairs confined within positions 63 and (probably) 322. The analogous region in the *E. coli* rrnD operon is considerably larger with 437 nucleotide positions. In *E. coli* the spacer region is known to be transcribed as a part of the single transcriptional unit of the rrnD operon[10]. A large rRNA precursor containing both the 16S and 23S rRNA sequences was also observed in spinach[25,26] and *Chlamydomonas*[27] chloroplasts. A rRNA of similar large size has been observed also in *E. gracilis* chloroplasts[28] and therefore it is reasonable to assume that a similar situation exists in this case although direct experimental evidence for transcription of the intergenic spacer is lacking.

Screening for GTTC sequences, which are indicative for the GTΨC sequence present in the Ψ loop of all tRNA species[29], and for homology with the *E. coli* rDNA spacer region reveals two tRNA genes coding for an isoleucine (AU^U_C)- and an alanine (GC^A_G)-accepting species, respectively. The tRNA primary structures deduced from the two tRNA genes allow folding to a cloverleaf model completely consistent with all other structural criteria of tRNAs[29] (see Fig. 3). The deduced cloverleaf models contain several compensating base changes in the stem regions

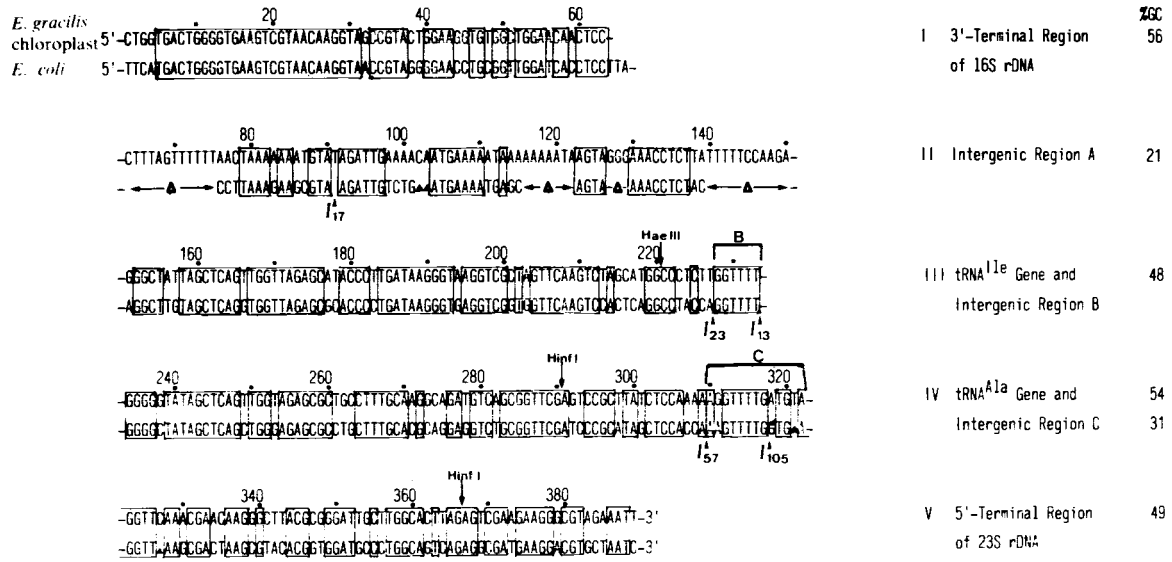

Fig. 2 Nucleotide sequence of the 16S–23S intergenic spacer region and of the flanking region of the 16S and 23S rRNA genes from *E. gracilis* chloroplast DNA (top row) and from the rrnD operon of *E. coli*[10,11]. Only the sequences of the RNA-like strands are depicted. Homologous sequences are marked by boxing. Within the *E. coli* sequence the symbols ▲, △ and *I* indicate single point deletions, multiple deletions and insertions in respect to the *E. gracilis* sequence. The number *a* of inserted bases is indicated by I_a. Note that the *E. gracilis* sequence is depicted in the opposite orientation from that of Fig. 1. Cleavage sites for the restriction enzymes used in Fig. 1 are indicated.

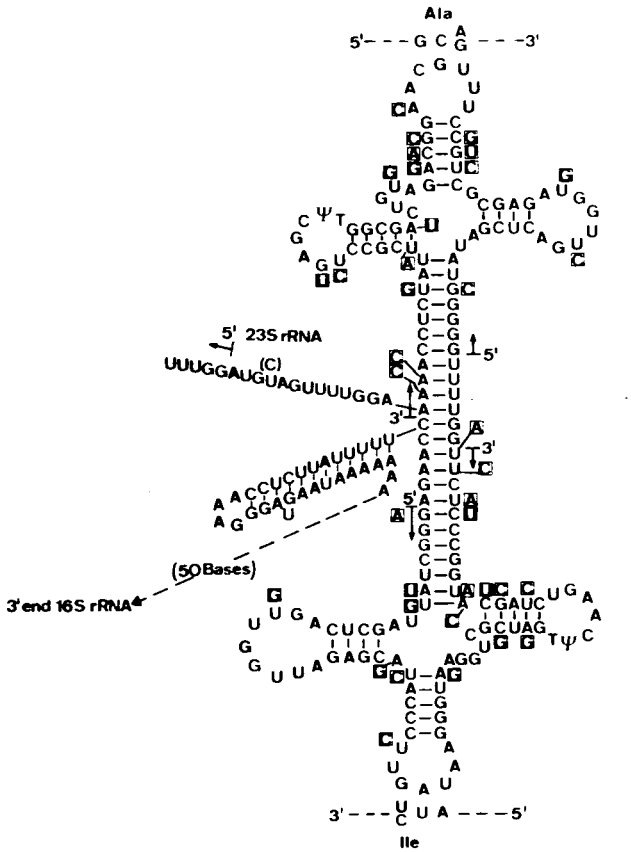

Fig. 3 Possible secondary structure of the dimeric tRNA precursor region as predicted from the corresponding DNA sequence of Fig. 2. The two tRNA cloverleaf structures are depicted without modifications of the bases except for the TΨ sequences in the T-arms. Bases different in the respective *E. coli* tRNA species[10,11] are indicated within boxes.

in addition to changes in the unpaired regions, which excludes the possibility that the two genes are derivatives of the *E. coli* genome due to a cloning or recombinational artefact. Also, the two chloroplast tRNA genes lack the CCA termini found in the respective *E. coli* tRNA genes[10,11]. The two tRNA genes are separated from each other by only six base pairs, which is reminiscent of the clustering of tRNA genes on the phage T_4 genome[30]. Within the precursor transcript, the two acceptor stems together with their neighbouring sequences can form a continuous stack of 24 base pairs (plus 2 × 5 base pairs of the two T-stems according to the tRNA tertiary model[31,32]) as drawn in Fig. 3. A possible hairpin structure directly preceding the 5' end of the tRNA^{Ile} is also depicted in Fig. 3, since its stem region could possibly also form an uninterrupted stack with the acceptor stem (and the T-stem[31,32]) of this tRNA within the precursor molecule. It is tempting to speculate that some of these possible secondary structures constitute at least part of the signal structures necessary for proper processing of the common precursor chain. Alternatively, or in addition to this, recognition of the sequence GGUUUUG, which is common to both the tRNA precursors immediately beyond the 3' ends, may act as a signalling step during the processing events.

A comparison of the rRNA and tRNA coding sequences with the intergenic regions (Fig. 2) shows that the former are highly conserved contrary to the latter which are subject to many alterations, mainly by deletions but including some base substitutions. The short homologous stretches in the intergenic regions may have some functional significance for processing steps. Furthermore it is interesting that, similar to *E. coli* rDNA[10,11], the G + C content is high in the structural and low in the intergenic parts.

We thank Drs K. Edwards and R. Hertel for helpful criticism of the manuscript. This work was supported by grants from Fonds National Suisse de la Recherche Scientifique No. 3.452.79 to E.S. and from Deutsche Forschungsgemeinschaft (SFB 46) to H.K.

Received 30 April; accepted 20 June 1980.

1. Manning, J. E. & Richards, O. C. *Biochim. biophys. Acta* **259**, 285–296 (1972).
2. Scott, N. S. *J. molec. Biol.* **81**, 327–336 (1973).
3. Kopecka, H., Crouse, J. E. & Stutz, E. *Eur. J. Biochem.* **72**, 525–535 (1977).
4. Gray, P. W. & Hallick, R. B. *Biochemistry* **18**, 1820–1825 (1979).
5. Schwartzbach, S. D., Hecker, L. I. & Barnett, W. E. *Proc. natn. Acad. Sci. U.S.A.* **73**, 1984–1988 (1976).
6. Gray, P. W. & Hallick, R. B. *Biochemistry* **17**, 284–290 (1978).
7. Rawson, J. R. Y., Kushner, S. R., Vapnek, D., Alton, N. K. & Boerma, C. L. *Gene* **3**, 191–209 (1978).
8. Jenni, B. & Stutz, E. *Eur. J. Biochem.* **88**, 127–134 (1978).
9. Hallick, R. B., Gray, P. W., Chelm, B. K., Rushlow, K. E. & Orozco, E. M. Jr *Chloroplast Development* (eds Akoyunoglou, G. *et al.*) 619–622 (Elsevier, Amsterdam, 1978).
10. Young, R. A., Maklis, R. & Steitz, J. A. *J. biol. Chem.* **254**, 3624–3271 (1979).
11. Sekya, T. & Nishimura, S. *Nucleic Acids Res.* **6**, 575–592 (1979).
12. Bohnert, H. J. *et al. FEBS Lett.* **103**, 52–56 (1979).
13. Schwarz, Z. & Kössel, H. *Nature* **283**, 739–742 (1980).

14. Schwarz, Z. & Kössel, H. *Nature* **279,** 520–522 (1979).
15. Knopf, U. C. & Stutz, E. *Molec. gen. Genet.* **163,** 1–6 (1978).
16. Graf, L., Schwarz, Z., Kössel, H. & Stutz, E. *Experientia* **86,** 34 (1980).
17. Brosius, J., Palmer, M. L., Kennedy, P. J. & Noller, H. F. *Proc. natn. Acad. Sci. U.S.A.* **75,** 4801–4805 (1979).
18. Carbon, P., Ehresmann, C., Ehresmann, B. & Ebel, J. P. *FEBS Lett.* **94,** 152–156 (1978).
19. Jenni, B. & Stutz, E. *FEBS Lett.* **102,** 95–99 (1979).
20. Maxam, A. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74,** 560–564 (1977).
21. Zablen, L. B., Kissil, M. S., Woese, C. R. & Buetow, D. E. *Proc. natn. Acad. Sci. U.S.A.* **72,** 2418–2422 (1975).
22. Woese, C. R. *et al. Nature* **254,** 83–86 (1975).
23. Shine, J. & Dalgarno, L. *Proc. natn. Acad. Sci. U.S.A.* **71,** 1342–1346 (1974).
24. Steitz, J. A. *Biological Regulation and Control* (ed. Goldberger, R.) 349–399 (Plenum, New York, 1979).
25. Hartley, M. R., Head, C. W. & Gardiner, J. *Acides Nucléiques et Synthèse des Protéines chez les Végétaux* (eds Bogorad, L. & Weil, J. H.) 419–423 (CNRS, Paris, 1977).
26. Bohnert, H.-J., Driesel, A. J. & Herrmann, R. G. *Acides Nucléiques et Synthèse des Protéines chez les Végétaux* (eds Bogorad, L. & Weil, J. H.) 213–218 (CNRS, Paris, 1977).
27. Rochaix, J. D. & Malnoe, P. *Cell* **15,** 661–670 (1978).
28. Wollgiehn, R. & Parthier, B. *Plant Sci. Lett.* **16,** 203–210 (1979).
29. Sprinzl, M., Grueter, F., Spelzhaus, A. & Gauss, D. H. *Nucleic Acids Res.* **8,** r1–r22 (1980).
30. Abelson, J. *A. Rev. Biochem.* **48,** 1035–1069 (1979).
31. Suddath, F. L. *et al. Nature* **248,** 20–24 (1974).
32. Robertus, J. D. *et al. Nature* **250,** 546–551 (1974).