

The *Euglena gracilis* chloroplast genome: structural features of a DNA region possibly carrying the single origin of DNA replication

Bibiane Schlunegger and Erhard Stutz

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

Summary. We sequenced a BglII-HindIII DNA fragment of the *Euglena gracilis* chloroplast genome which most likely carries the single origin of DNA replication according to recent electronmicroscopic mapping studies (Koller and Delius 1982a; Ravel-Chapuis et al. 1982). This DNA fragment contains a polymorphic region (Schlunegger et al. 1983) which is composed, as will be shown, of multiple tandem repeats (54 bp, 87% A+T). Furthermore we located on this DNA fragment a short inverted repeat element (96 positions) observed in the electronmicroscopic studies (Koller and Delius 1982b). Between the borders of the polymorphic region and the nearby inverted repeat (distance of 179 positions) we retrieved an exact copy of parts of the rDNA leader (105 positions) including 49 positions of the chloroplast *trnW* gene. A computer search for bacterial type Ori-regions did not reveal any significant sequence homology. However, the polymorphic region and its immediate vicinity have the capacity to form multiple stem and loop structures which may be involved in DNA replication initiation.

Key words: Chloroplast DNA – *Euglena gracilis* – Origin of DNA replication

Introduction

Prokaryotic and eukaryotic genome replication normally is under tight control, DNA replication initiation being an important controlling step. In some bacteria the initiation step is rather well understood (Stayton et al. 1983) and determining the nucleotide sequence of the origin

region and manipulating it in various ways have provided important clues. It was recognized, e.g., that the origin of DNA replication of *E. coli* extends over 245 nucleotides (Oka et al. 1980) and that it contains conserved recognition sequences and less or non conserved spacer sequences providing specific distances between recognition sequences. Bacterial origins of DNA replication contain short direct and indirect repeats allowing the formation of elaborate stem and loop structures.

Chloroplasts are in many respects of prokaryotic nature and their circular genome replicates by forming Cairns type replicative intermediates; but also displacement loops and rolling circle replicative intermediates were observed, especially in case of higher plants (Tewari 1979). In case of chloroplast DNA of the unicellular alga *Euglena gracilis* Cairns replicative intermediates were mainly observed (Manning and Richards 1972). More recently Koller and Delius (1982a) and Ravel-Chapuis et al. (1982) independently mapped the apparently single origin of replication of the *Euglena* chloroplast genome by electronmicroscopic analysis of replication forks about 5 to 6 kbp upstream of the 5' end of the s16S rRNA gene (Jenni and Stutz 1979) and in close vicinity of a previously described polymorphic region (Z-region) known to contain a variable number of short repeats (Jenni et al. 1981). This Z-region and its vicinity had already been mapped and cloned into pBR322 and further analysed by electronmicroscopy (Schlunegger et al. 1983). It was shown that the polymorphic region was located between two short inverted repeats (Koller and Delius 1982b) and from denaturing studies it was concluded that the DNA segment contains regions extremely rich in A+T. Very recently Waddel et al. (1984) mapped two origins of DNA replication on the circular chloroplast genome of *Chlamydomonas reinhardtii*. These origins are, respectively, 10 and 16.5 kbp upstream of one of the 16S rRNA genes. The same laboratory (Wang et al.

Offprint requests to: E. Stutz

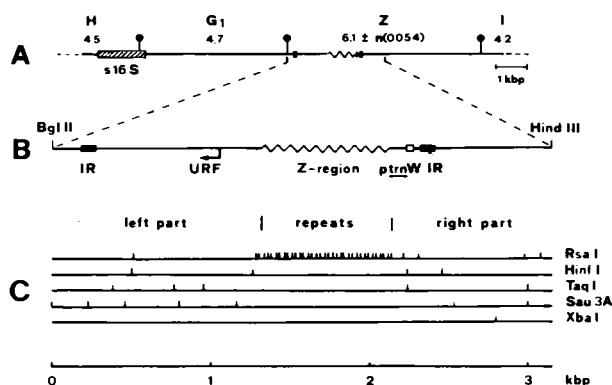


Fig. 1A-C. Map position on the *Euglena gracilis* chloroplast genome of the BglIII · Z DNA fragment carrying the size variable region (Z-region) and strategy of sequencing. A Map position of BglIII · Z relative to the extra 16S rRNA gene and previously mapped BglIII fragments, H, G1 and I (Jenni et al. 1981). Sizes are in kbp. B Structural features of the sequenced BglIII-HindIII fragment. IR, inverted repeats; URF, unidentified open reading frame; Z, region composed of a variable number of repeats of 54 bp, *PtrnW*, 49 positions of the 3' terminal part of the tRNA^{Trp} gene. C Restriction fragments which were cloned into phage M13 mp9 and sequenced according to Sanger et al. (1980). The sequenced region is subdivided into a left side part, the repeated region and the right side part (see Fig. 2A, B)

1984) identified an A+T rich 0.42 kbp fragment as most likely locus of one of two origins.

With this in mind it became interesting to analyse the Z-region and flanking parts (about 3.1 kbp) on a nucleotide level in order to see whether this segment or parts of it could structurally qualify for an origin of DNA replication. Parallel to these structural analysis went functional tests in collaboration with other laboratories, however, with no success, at least so far.

Materials and methods

pEgCH₂ DNA (Schlunegger et al. 1983) was isolated as described by Graf et al. (1982).

The BglIII-HindIII fragment containing the Z-region was separated on agarose gel, eluted by electrophoresis (Wienand et al. 1979), extracted with phenol and precipitated with ethanol.

After digestion with appropriate enzymes (shown in Fig. 1), subfragments of BglIII-HindIII fragment were 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 Vierra 1982), with the ligation conditions 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 CaCl₂ and 0.001 M NaCl. The clones to be sequenced by the method of Sanger et al. (1980) were selected after T-track analysis. As primer we used the synthetic pentadecamer nucleotide from New England Biolabs.

Results

Map position of the sequenced region and strategy of sequencing

According to Koller and Delius (1982a) and Ravel-Chapuis et al. (1982) the Ori-region is in the vicinity of the Z-region which is a single locus on the circular DNA molecule composed of a variable number of short tandem repeats (Schlunegger et al. 1983). We have previously cloned this region into pBR322 and used the recombinant pEgCH₂ for subcloning parts of this region into M13 mp9. In Fig. 1, line A we show the overall map position of the Z-region within the DNA fragment BglIII · Z and on line B we show the BglIII-HindIII fragment which was sequenced according to Sanger et al. (1980) using the restriction fragments given on lines C. The DNA fragments were sequenced in both directions. We subdivide in the following the sequenced fragment in a left side part (lp) consisting of 1,322 positions, the Z-region with n tandem repeats of 54 bases, and a right side part (rp) of 1,025 positions (see line B).

Structural features of the BglIII-HindIII fragment

In Fig. 2A we give the nucleotide sequence for both strands from the left side part and in Fig. 2B we show one repeat unit of the Z-region (54 positions) and the right side part. According to EM analysis there exists a short inverted repeat element framing the Z-region. Indeed, we identified on the left and right side part such elements of 96 positions having 82% sequence homology. This structural element must be identical with that observed in the EM analysis and according to Koller and Delius (1982a) the origin of replication should be close to the inverted repeat on the right side part. No other inverted repeat element of such length was found within the sequenced fragment.

The Z-region (Fig. 2B) is composed of tandem repeats. The 54 base element starts with 5'-GTAC (RsaI-site) and contains a second RsaI-site at position 19. The overall T+A content of the repeat unit is extremely high (87%). The primary structure of the repeats allows extensive formation of stem and loop structures as will be discussed.

On the left side part (Fig. 2A) we identified a major open reading frame (URF) potentially coding for a protein of at least Mr 41,000. Transcription would be towards the BglIII site, however, we have no clear evidence yet whether the region is transcribed. On the right side part we retrieved a pseudo-tRNA^{Trp} gene (Fig. 2B). 48 out of 49 positions of the 3' terminal part are an exact copy of the functional chloroplast *trnW* gene recently mapped as part of a tRNA gene cluster which was totally

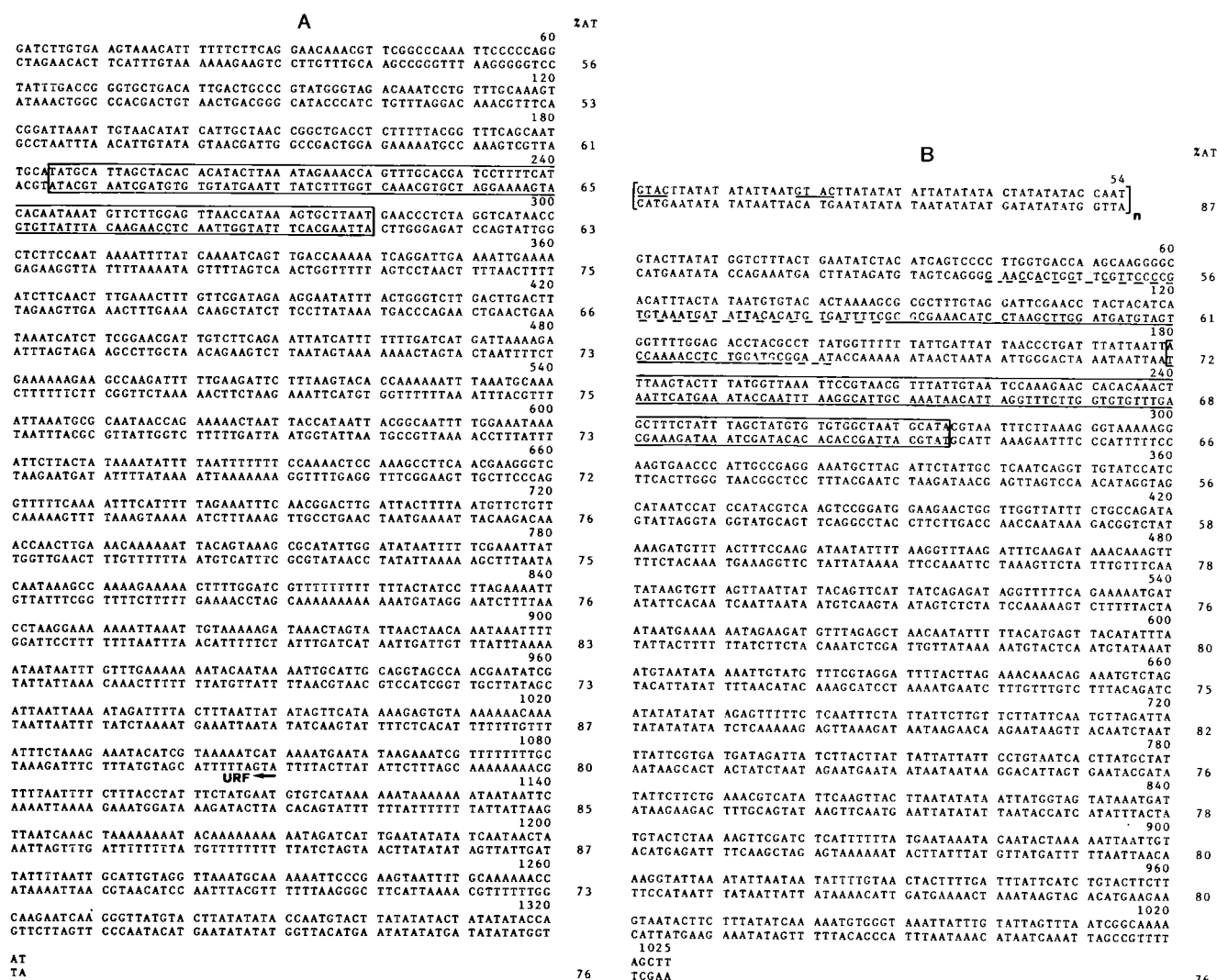


Fig. 2. A DNA sequence of the left side part. The inverted repeat is boxed. Start and reading direction of a URF are marked. The reading frame is open beyond the BglII site (top). B DNA sequence of the repeat unit and the right side part. The RsaI sites in the repeat unit are underlined. The inverted repeat homologous with that on the left side part is boxed. The *PtmW* is solidly underlined and additional sequences identical to sequences in the rDNA leader are marked by a dashed line

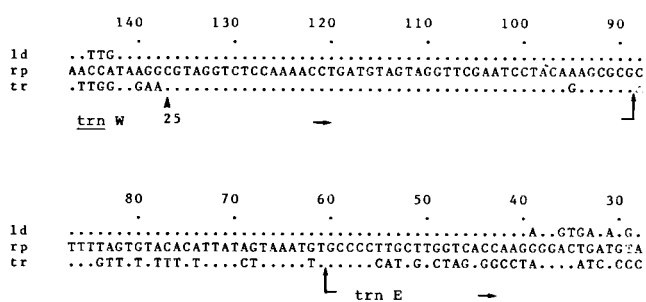


Fig. 3. Sequence comparison between the DNA segment in the vicinity of the tandem repeats (rp) a segment of the rDNA leader (1d) and parts of the tRNA gene cluster (tr). Sequence identity is marked by dots. Position 25 (D-stem) of the tRNA^{Trp} gene is marked. The 5'-3' polarity of the tRNA genes is indicated by horizontal arrows. The end of the *trnW* and the beginning of *trnE* of the gene cluster (tr) are marked by vertical bent arrows. Nucleotide positions are numbered according to Fig. 2B, note however the opposite orientation of presentation. The rDNA leader sequence is taken from Roux et al. (1983), the tRNA gene sequences from Hollingsworth and Hallick (1982)

sequenced and shown to be composed of the tRNA genes for Tyr, His, Met_E, Trp, Glu and Gly (Hollingsworth and Hallick 1982).

It is known for some time (Orozco et al. 1980; Roux

et al. 1983; El-Gewely et al. 1984) that the chloroplast rDNA leader sequence of *Euglena gracilis* contains several pseudo-tRNA gene elements. When we found the pseudo-*trnW* gene near the Z-region and compared this

Chlamydomonas reinhardtii which promote autonomous replication in yeast. These fragments contain ARS elements as defined, e.g., by Broach et al. (1983). We also see such elements in the sequenced *Euglena* chloroplast DNA fragment, e.g., at positions 887 and 897 on the right side part. However, ARS bioassays using cloned *Euglena* chloroplast DNA fragments containing the sequenced stretch did not yield positive results (J. D. Rochaix, personal communication).

Combining the electronmicroscopic results (loci cit.) with our sequencing data we come to the conclusion that within the sequenced BgIII-HindIII fragment, the Z-region and its immediate vicinity are better suited than any other part to function as an origin of DNA replication. Affirmative elements certainly are the extremely high content of A+T, the capacity to form multiple stem and loop structures and maybe the conserved ARC element II. The presence of a pseudo-tRNA^{Trp} gene in that region may be fortuitous but we cannot exclude it to have a role in DNA replication initiation as discussed in the next paragraph.

The mobility of DNA sequences within the chloroplast genome

A corollary of these sequencing studies was the observation that parts of the rDNA leader were found near the Z-region, what must be the result of a DNA transfer event of the evolutionary past. We mentioned already that the string of 105 nucleotides (see Fig. 3) contains 49 positions of the *trnW* gene which itself was transferred in an even earlier time from the *trn* gene cluster in EcoV-H to the rDNA leader region (El-Gewely et al. 1984). Most remarkable is the fact that the pseudo-*trnW* gene starts with position 25 (see Fig. 3) which precisely is the splicing site for an intron in the D-stem of the tobacco *trnG* (Deno and Sugiura 1984). The usual intron site in tRNA genes is placed in the anticodon stem (Sprinzl and Gauss (1983). We searched for the missing 5' terminal part of the *trnW* gene upstream within the right side part but without success.

The pseudo-tRNA genes in the rDNA leader as well as in the Z-region are found both in the Z-strain and B-strain (bacillaris). According to a rough estimate of El-Gewely et al. (1984) the two strains may have separated more than 150×10^6 years ago, what means that the tRNA gene transfer into the rDNA region is an old event and nevertheless 48 positions of the *trnW* region are very well conserved. El-Gewely et al. (1984) argued that the pseudo-tRNA genes in front of the 16S rRNA gene may be involved in transcription regulation. By analogy we may argue that the same DNA sequence in another genome environment (Z-region) may be involved in the DNA replication initiation step. So far RNA poly-

merases as well as DNA polymerases and other enzymatic activities involved in DNA replication initiation are very poorly understood (Bryant 1982) and appropriate enzymology is required to set up bioassays for testing whether any parts of the sequenced BgIII-HindIII fragment do indeed qualify as origins of DNA replication.

Acknowledgments. We are grateful to H. Gut for technical assistance, P. E. Montandon for helpful advice and C. Bachmann for secretarial help. This research is supported by Fonds National Suisse de la Recherche Scientifique (to E. S. 3.183.82).

References

- Broach JR, Li YY, Feldman J, Jayaram M, Abraham J, Nasmith KA, Hicks JB (1983) Cold Spring Harbor Symp. Quantitative Biol 47:1165-1173
- Bryant JA (1982) In: Parthier B, Boulder D (eds) Encyclopedia of plant physiology, new Series, vol 14B. Springer, Berlin, pp 75-110
- Calos MP, Miller JH (1980) Cell 20:579-595
- Cohen SW, Chang ACY, Hsu L (1972) Proc Natl Acad Sci USA 69:2110-2114
- Deno H, Sugiura M (1984) Proc Natl Acad Sci USA 81:405-408
- El-Gewely MR, Helling RB, Dibbib JGT (1984) Mol Gen Genet 194:432
- Graf L, Roux E, Stutz E, Kössel H (1982) Nucleic Acids Res 10:6369-6381
- Hollingsworth MJ, Hallick RB (1982) J Biol Chem 257:12795-12799
- Jenni B, Stutz E (1979) FEBS Lett 102:95-99
- Jenni B, Fasnacht M, Stutz E (1981) FEBS Lett 125:175-179
- Koller B, Delius H (1982a) EMBO J 1:995-998
- Koller B, Delius H (1982b) FEBS Lett 139:86-92
- Manning JE, Richards OC (1972) Biochemistry 11:2036-2043
- Messing J, Vieira J (1982) Gene 19:269-276
- Messing J, Crea R, Seeberg PH (1980) Nucleic Acids Res 9:309-321
- Oka A, Sugimoto K, Takanami M, Hirota Y (1980) Mol Gen Genet 178:9-20
- Orozco EM, Rushlow KE, Dodd JR, Hallick RB (1980) J Biol Chem 255:10997-11003
- Ravel-Chapuis P, Heizman P, Nigon V (1982) Nature 300:78-81
- Rochaix JD, van Dillewijn J, Rahire M (1984) Cell 36:925-931
- 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
- Sprinzl M, Gauss DH (1983) Nucleic Acids Res 11:r1-r53
- Stayton MM, Bertsch L, Biswas S, Burgers P, Dixon N, Flynn JE, Fuller R, Kaguni M, Low R, Kronberg A (1983) Cold Spring Harbor Symp Quantitative Biol 47:693-700
- Tait RC, Rodriguez RL, West RW (1980) J Biol Chem 255:813-815
- Tewari KK (1979) In: Hall TC, Davies JW (eds) Nucleic Acids in Plants, vol I. CRC Press, Florida, pp 41-108
- Vallet JM, Rahire M, Rochaix JD (1984) EMBO J 3:415-421

Waddell J, Wang XM, Wu M (1984) *Nucleic Acids Res* 12:
3843–3856
Wang XM, Chang CH, Waddell J, Wu M (1984) *Nucleic Acids Res*
12:3857–3872
Wartell RM, Reznikoff WS (1980) *Gene* 9:307–310
Wienand U, Schwarz Z, Feix G (1979) *FEBS Lett* 98:319–323

Zyskind JW, Harding NE, Takeda Y, Cleary JM, Smith DW (1982)
ICN-UCLA Symposia on molecular and cellular biology
22:13–25
Zyskind JW, Smith DW (1980) *Proc Natl Acad Sci USA* 77:
2460–2464