Proc. Natl. Acad. Sci. USA
Vol. 78, No. 6, pp. 3323–3327, June 1981
Biochemistry

Complete nucleotide sequence of minicircle kinetoplast DNA from *Trypanosoma equiperdum*

(restriction enzymes/mitochondrial DNA/sequence homology)

MICHEL BARROIS*†, GUY RIOU‡, AND FRANCIS GALIBERT*

*Laboratoire d'Hématologie Expérimentale, Centre Hayem, Hôpital St. Louis, 75010, Paris, France; and ‡Laboratoire de Pharmacologie Moléculaire, Institut Gustave-Roussy, 94800, Villejuif, France

Communicated by William Trager, January 27, 1981

ABSTRACT **The kinetoplast DNA of *Trypanosoma equiperdum* is composed of about 3000 supercoiled minicircles of 1000 base pairs and about 50 supercoiled maxicircles of 23,000 base pairs topologically interlocked so as to form a compact network. Minicircles of *T. equiperdum*, which are homogeneous in base sequence, were purified by equilibrium CsCl centrifugation and used as starting material for DNA sequence analysis. One minicircle is composed of 1012 base pairs and has an adenine·thymine base pair content of 72.8%. The termination codons are uniformly distributed along the molecule and restrict the coding potentiality of the molecule to oligopeptides of about 20 amino acids. The molecule contains three dyad symmetries and a sequence of 12 nucleotides is repeated six times. We also noted the presence of a region of about 130 base pairs that is almost perfectly homologous with that of the minicircles from the closely related species *T. brucei*.**

The kinetoplast is the specialized portion of the mitochondrial apparatus of trypanosomes that contains DNA (kDNA) in high concentration. The kDNA of the flagellates so far studied is composed of thousands of minicircles and of a small number of maxicircles, linked together so as to form a large network of high molecular weight (1, 2). In *Trypanosoma equiperdum*, the kDNA is composed of about 3000 covalently closed minicircles of 1000 base pairs and of about 50 covalently closed maxicircules of 23,000 base pairs (2). The biological role of the kDNA is still unknown, although it was found that maxi- and minicircles were partly transcribed (3, 4). We have shown that the minicircle kDNA from *T. equiperdum* was homogeneous in base sequence (2) in contrast to the minicircle of other species of trypanosomes (1, 5–7), and we have mapped 10 sites of cleavage by restriction endonucleases (8). The minicircle kDNA exists as a small, double-stranded, covalently closed, circular molecule, whose physical properties allow its isolation free from DNA contaminants. Recently developed techniques of nucleotide sequence analysis provide a useful approach to an understanding of the eukaryotic gene organization and of the regulation of their expression. The purpose of the present study was to obtain the complete nucleotide sequence of the minicircle kDNA from *T. equiperdum*.

MATERIALS AND METHODS

Enzymes and Chemicals. The restriction endonucleases were purchased from New England BioLabs. Bacterial alkaline phosphatase and polynucleotide kinase were from P-L Biochemicals. The chemicals utilized were: dimethyl sulfate (Aldrich), hydrazine (Eastman), acrylamide and *N,N'*-methylenebisacrylamide (2-fold crystallized grade; Serva), and piperidine from Merck (redistilled under vacuum).

Strain and Purification of Trypanosomes. The original kinetoplastic strain of *T. equiperdum* was obtained from the Institut Pasteur (Paris) in 1961. This strain was isolated by Laveran about 60–70 yr ago and maintained at the Institut Pasteur by transfers in laboratory mammals. *T. equiperdum* induces in rabbit a chronic infection characterized by the appearance of a series of antigenic variants (9). The antigenic variant BoTat 24 (Bordeaux-Trypanozoon-antigenic type 24), which was isolated as described (10), was used in this work. Female Wistar rats, weighing about 200 g, were infected intraperitoneally with about 10^6 trypanosomes in order to obtain maximal parasitemia in 3–4 days. The trypanosomes were isolated from the blood of the rats, purified as described by Lanham and Godfrey (11), and washed three times with 0.15 M NaCl/0.015 M trisodium citrate.

Preparations and Purification of Minicircle kDNA. The total DNA was extracted from the trypanosomes as described (6). The kDNA network was fractionated in a CsCl/ethidium bromide gradient as covalently closed molecules. The kDNA minicircles form I were obtained after cleavage of the maxicircles with *Bam*HI restriction endonuclease as described (2). The minicircles were separated from linearized maxicircles in a CsCl/ethidium bromide gradient. Recovered minicircles were further purified in a second CsCl gradient in the presence of Hoechst 33258 dye, which improves the separation because of its preferential binding to A·T-rich DNA (12). The dye was removed with isopropanol.

Preparation of 5′ ^{32}P-Labeled DNA Fragments. Minicircle kDNA (10–20 pmol) was fully hydrolyzed by different restriction enzymes under the conditions recommended by the manufacturer. The DNA fragments were dephosphorylated by alkaline phosphatase, which was then inactivated by alkaline treatment (13), and the DNA was precipitated with ethanol. After redissolution in spermidine buffer, the DNA was labeled at its 5′ ends with [γ-^{32}P]ATP (New England Nuclear: 2000 Ci/mmol; 1 Ci = 3.7×10^{10} becquerels) and polynucleotide kinase (14). The DNA restriction fragments were separated by polyacrylamide gel electrophoresis and eluted from the gels. The two labeled ends were separated by polyacrylamide gel electrophoresis, after cleavage with another enzyme, or by denaturation of the DNA fragment.

DNA Nucleotide Sequence. The primary structure of the double-stranded DNA fragments was determined by the method of Maxam and Gilbert (14, 15). The sequence of the chemical reaction products were analyzed by electrophoresis in 8%, 16%, or 25% (wt/vol) acrylamide gels, 1 mm thick.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviations: kDNA, kinetoplast DNA.

† Present address: Laboratoire de Pharmacologie Moléculaire, Institut Gustave-Roussy, 94800, Villejuif, France.

RESULTS AND DISCUSSION

The restriction map of the minicircle kDNA from *T. equiperdum* had been determined (8). The 10 recognition sites allow cleavage of the minicircle into fragments sufficiently short for sequence analyses. In order to analyze independently both DNA strands and to overlap all the restriction sites used as starting points for sequence determinations, a large number of fragments were analyzed (Fig. 1). Examples of sequence analyses with gel autoradiograms from two DNA fragments are shown in Fig. 2, the DNA sequence is in Fig. 3.

Three new cleavage sites by *Rsa* I and *Dde* I were localized. The minicircle kDNA is composed of 1012 base pairs containing 72.8% adenine·thymine base pairs. One strand of the minicircle contains 39.2% adenine, 33.6% thymine, 20.6% guanine, and 6.6% cytosine. The two strands have an unequal distribution of cytosine or guanine. By using the method of Schildkraut *et al.* (16), we had found from buoyant density analysis an adenine + thymine content of 68% for the kDNA minicircles when associated in a network of compact conformation (2). This lower value could have been due to the compactness of the kDNA network. A similar phenomenon has been observed with circular DNA of high negative superhelical density, which has a very compact structure (17).

To investigate the coding potential of the minicircle kDNA, a computer program was used to analyze the relative positions of initiation and termination codons in all six reading frames. Distribution of the three nonsense codons TAA, TGA, and TAG and of the initiation codon ATG in the different reading frames is seen in Fig. 4. The number of initiation codons is particularly low in chain 2, zero in phase ϕ_1, and two in phases ϕ_2 and ϕ_3. The nonsense codons, whose number varies from 22 to 35 depending on the chain and on the phase, are more numerous in chain 1 than in chain 2 and are mostly represented by the TAA amber codon. In phase ϕ_2 of chain 2, there are 20 TAA codons and 2 TGA codons. Several stop codons are found in tandem in each phase of chain 1. Because the termination codons are uniformly distributed in each reading frame, the longest polypeptides that can be coded are restricted to 18 and 22 amino acid residues, if one takes into account the position of the initiation codons and assumes that a minicircle does not contain split genes. Fig. 4 clearly shows that the splitting frequency would have to be very high to obtain longer polypeptides.

Similar results were recently obtained by Chen and Donelson (18) in two different cloned kDNA minicircles from *T. brucei*. However, variations in the mitochondrial genetic code have already been described in man (19) and in yeast (20, 21). It was shown that TGA is used as a tryptophan codon and not as a termination codon. Moreover, ATA is a methionine and not an isoleucine codon (19). In the kDNA of trypanosomes, which is of mitochondrial origin, some of these stop codons could follow other genetic rules and code for an amino acid. However, if TGA were to code for tryptophan, as in yeast and in human mitochondria, the length of the open sequences would not be increased. In contrast, if TAA were to code for an amino acid, the open reading frame could be substantially longer; as a matter of fact, chain 2 would be one continuous reading frame, from the ATG 561 triplet located in phase ϕ_2 up to the TGA 560 triplet located in phase ϕ_3 (see Fig. 4 legend).

Another feature of the kDNA sequence is the presence of one dyad symmetry of 11 base pairs (from nucleotide 174 to nucleotide 212) and of two dyad symmetries of 9 base pairs (be-

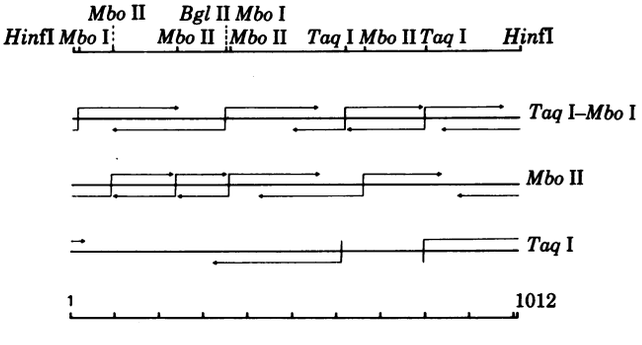

FIG. 1. Sequence determination strategy. The restriction sites used for sequence analysis are shown at the top of the figure. The horizontal arrows indicate the direction and the length of the sequences. →, Chain 1; ←, chain 2.

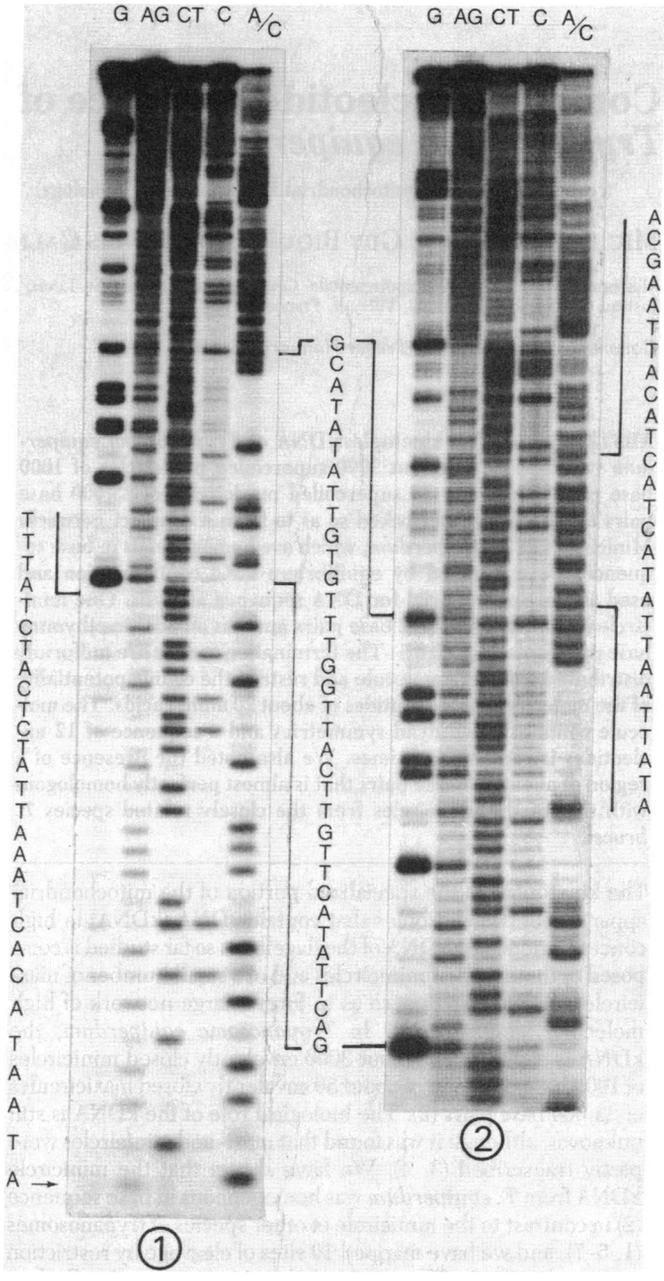

FIG. 2. Examples of autoradiograms of gels used in sequence analysis. The products, obtained by the chemical degradation method of Maxam and Gilbert, were fractionated by electrophoresis on 16% polyacrylamide gels. *Mbo* I (345) fragments contain the sequence of chain 2 from nucleotide 338 to nucleotide 262 (gel 1) and from nucleotide 310 to nucleotide 230 (gel 2).

```
                                        30                                  60                                  90
① 5' AAT CAG TGA GAG ATC AGT TAA TAT TAA TTA TAT ATT GTT ATT TAC ACC TAT TTA TTA TCT TAT TCT TTT GGT TTA GAG AAG AAA GTA ATA
     GTC ACT CTC TAG TCA ATT ATA ATT AAT ATA TAA CAA TAA ATG TGG ATA AAT AAT AGA ATA AGA AAA CCA AAT CTC TTC TTT CAT TAT
  HinfI         MboI                                                                                    MboII

                                       120                                 150                                 180
TAA TAG ATA TGA ATT GTA ATA TTT ATA ATT GAT AAT ATC ATA TAC GAA AAC AAT CAG AAA AAT GTG ATA GGA AAT ATA AAA GGT ATA ATA
ATT ATC TAT ACT TAA CAT TAT AAA TAT TAA CTA TTA TAG TAT ATG CTT TTG TTA GTC TTT TTA CAC TAT CCT TTA TAT TTT CCA TAT TAT

                                       210                                 240                                 270
TTA ATG GTT GTA TGT TTA TTT ATT ATT ATA TTT TGT TAC AAG AAG AGA GAA ATA GGA TAA TAA ATA GAT GCT TAA TGT AGT AGT ATA ATT
AAT TAC CAA CAT ACA AAT AAA TAA TAA TAT AAA ACA ATG TTC TTC TCT CTT TAT CCT ATT ATT TAT CTA CGA ATT ACA TCA TCA TAT TAA
                                                 MboII

                                       300                                 330                 MboII           360
AAT ATC GTA TAT AAC ACA AAC CAA TGA ACA AGT ATA AGT CAA ATA GAT GAG ATA ATT TAG TGT ATT ATA ATG AAG ATC TTA TTA TTT ATT
TTA TAG CAT ATA TTG TGT TTG GTT ACT TGT TCA TAT TCA GTT TAT CTA CTC TAT TAA ATC ACA TAA TAT TAC TTC TAG AAT AAT AAA TAA
                                                                                                   BglII

                                       390                                 420                                 450
TAT TAT TTT ATT GAG ATT AAT GGG AGC GAT AAT TAG GTA CAA TGA GGA TTA TAA GTT CTC TAA AGT AAA GAA AAT GAT ATA ATA GAT AGG
ATA ATA AAA TAA CTC TAA TTA CCC TCG CTA TTA ATC CAT GTT ACT CCT AAT ATT CAA GAG ATT TCA TTT CTT TTA CTA TAT TAT CTA TCC
                                           RsaI

                                       480                                 510                                 540
GAT ATA GTA ATT ATT AAT ATG ATA ACT TAT ATA TGA ATG CGT AGT AGT GTC TTG TTG GAT GAG TAT GTG TAT AAA TGA TTG TAA TTT AAA
CTA TAT CAT TAA TAA TTA TAC TAT TGA ATA TAT ACT TAC GCA TCA TCA CAG AAC AAC CTA CTC ATA CAC ATA TTT ACT AAC ATT AAA TTT

                                       570                                 600                                 630
TTA TTA TAT TAA TGA TAT CAT ATT TAT TAC TTT ATT TAA ATT AAT GGA GGA GAG ATA GGC AGG AAT AAA TAA ATC GAG AAG TAG AAG TTA
AAT AAT ATA ATT ACT ATA GTA TAA ATA ATG AAA TAA ATT TAA TTA CCT CCT CTC TAT CCG TCC TTA TTT ATT TAG CTC TTC ATC TTC AAT
                                                                                                   TaqI

                                       660                                 690                                 720
GAT TGG GCG GTA AGT GAA GAG TAT AAA TAT TGG GAA ATT GGT GGA AAT TGG TGG AAA AAT AGC AAA AAA TCA GGC TGA AAA AAC CGA AAA
CTA ACC CGC CAT TCA CTT CTC ATA TTT ATA ACC CTT TAA CCA CCT TTA ACC ACC TTT TTA TCG TTT TTT AGT CCG ACT TTT TTG GCT TTT
                  MboII

                                       750                                 780                                 810
TCT TAT GGG CGT GCA GAT TTT ACC ATA CAC AAA TCA CGT GCT ATT TTC GGG GGT TTT TTA GGT CGG AGG TAC TTC GAA AGG GGT TGG TGT
AGA ATA CCC CGA CGT CTA AAA TGG TAT GTG TTT AGT GCA CGA TAA AAC CCC CCA AAA AAT CCA GCC TCC ATG AAG CTT TCC CCA ACC ACA
                                                                                    RsaI   TaqI

                                       840                                 870                                 900
AAT ACA CAC AGG GTT TTT CTG GGA ATT TTG GAG TAT AAA ACT AGT AGT GGG GAA TTA AGT GGG GAA TTA AGT GGG GAA TTA AGT GGG GAA
TTA TGT GTG TCC CAA AAA GAC CCT TAA AAC CTC ATA TTT TGA TCA TCA CCC CTT AAT TCA CCC CTT AAT TCA CCC CTT AAT TCA CCC CTT

                                       930                                 960                                 990
TTA AGT GGG GAA TTA AGT GGG GAA TTA CAT CCT AGA GAA AAT AAA TGT TGT AAT AGA TAG AGA TAT AAA CTT AGT ATA TAT AGA TAA TTA
AAT TCA CCC CTT AAT TCA CCC CTT AAT GTA GGA TCT CTT TTA TTT ACA ACA TTA TCT ATC TCT ATA TTT GAA TCA TAT ATA TCT ATT AAT
                                                                                    DdeI

                                     1012
TAC AAA CAC TTC AAC GGA AAA G   3'
ATG TTT GTG AAG TTG CCT TTT CTT A  5' ②
```

FIG. 3. The nucleotide sequence of kDNA minicircles from *T. equiperdum* is given in reading frame 1 as defined in the legend of Fig. 4. The numbers ① and ② indicate chains 1 and 2.

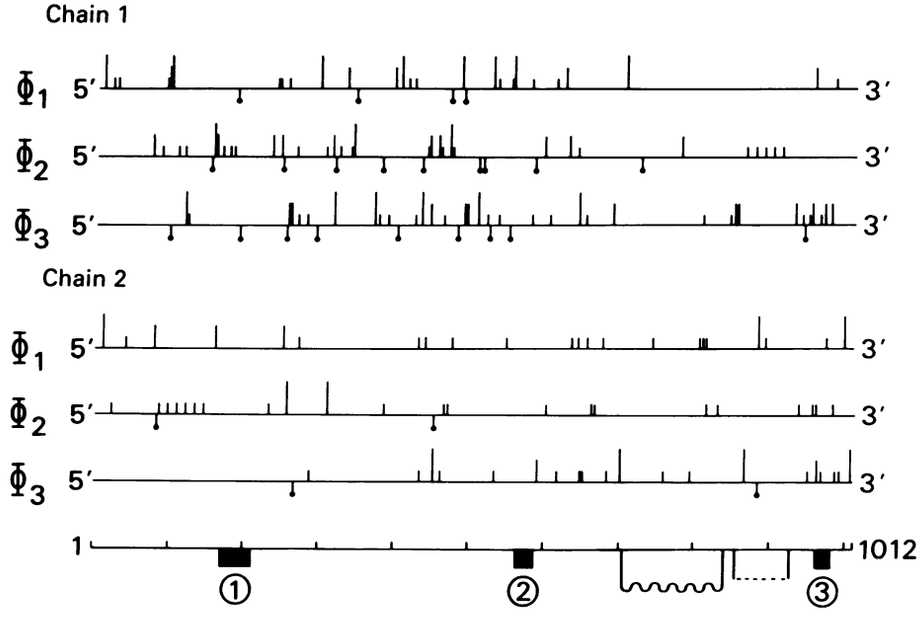

FIG. 4. Diagram showing the distribution of nonsense (⊥) and initiation (●) codons in the 6 reading frames. Three reading frames called Φ_1, Φ_2, and Φ_3 were initiated from the 5′ end of each DNA strand. On chain 1, Φ_1 is initiated by its first triplet AAT, Φ_2 by ATC, and Φ_3 by TCA. On chain 2, Φ_1 is initiated by its first triplet ATT, Φ_2 by TTC, and Φ_3 by TCT. The length of the kDNA minicircle does not correspond to a multiple of 3 nucleotides. Therefore, going through *Hinf* I site, phase Φ_1 is changed into phase Φ_3, phase Φ_2 into phase Φ_1, and phase Φ_3 into phase Φ_2. Vertical bars have different lengths, depending on the stop codon (⊥ TAA, ⊥ TAG, ⊥ TGA). (*Bottom*) Schematic representation of the minicircle, with the locations of the three dyad symmetries (①, ②, ③), the DNA region homologous to *T. brucei* minicircles (∿∿), and the six repeats of 12 base pairs each (I----I).

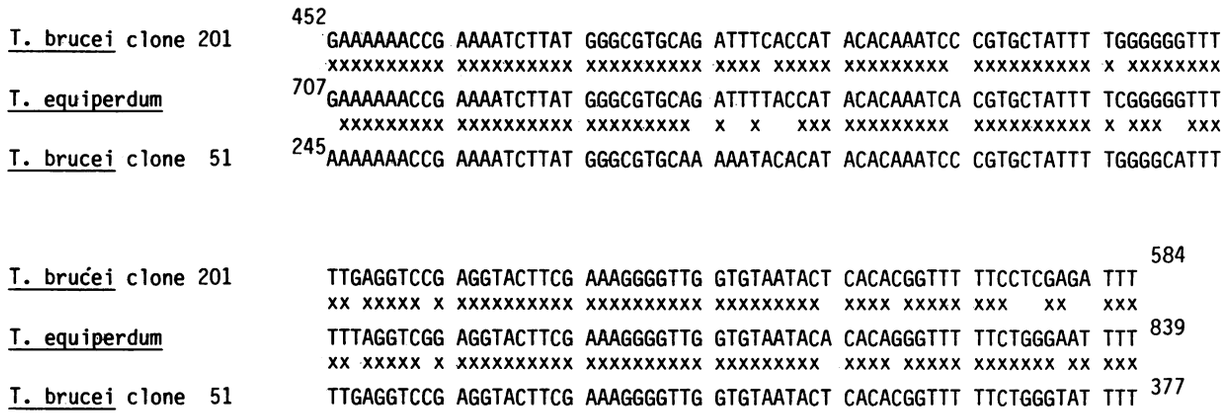

FIG. 5. Nucleotide sequence of the region of homology of *T. equiperdum* minicircles and *T. brucei* clones 51 and 201 minicircles.

tween positions 566 to 585 and 963 to 981). Dyad symmetries were found near the origin of DNA replication in simian virus 40 DNA (22) and λDNA (23). We have shown by electron microscopy that the kDNA minicircles replicated according to Cairns' model (24, 25). However, the presence of three dyad symmetries, each located on three equidistant parts of the minicircle (Fig. 4), makes it improbable that they are all three involved in the initiation of DNA replication. The short sequence composed of the 12 nucleotides A-G-T-G-G-G-G-A-A-T-T-A was repeated six times, from nucleotide 856 to nucleotide 927. One other feature of the DNA sequence shown in Fig. 3 is the uneven distribution of G·C base pairs (see below). We noticed also the presence of the sequence 5′ T-A-T-A-A-A five times in chain 1 and once in chain 2. This sequence was thought to correspond to a signal for the initiation of transcription in other DNAs (26).

As mentioned above, the sequence of minicircle DNA from *T. brucei* has been published by Chen and Donelson (18). *T. brucei* and *T. equiperdum*, which belong to the same subgenus Trypanozoon are morphologically undistinguishable. *T. brucei* is the agent of the sleeping sickness which affects cattle in Africa. This flagellate, transmitted by the tsetse fly, undergoes a complicated cell cycle. *T. equiperdum* has a quite different cell cycle: it is venereally transmitted in equines. The kDNA minicircles of *T. equiperdum* are homogeneous, as demonstrated by DNA sequence analysis; on the contrary, the kDNA minicircles of *T. brucei* are extensively heterogeneous in base sequence (27). Two individual kDNA minicircles of *T. brucei* were cloned by *in vitro* recombination in pBR322 (28), and their base sequence was determined by Chen and Donelson (18). These authors noticed the presence of a region of about 120 nucleotides presenting a strong sequence homology. The sequences of *T. brucei* and *T. equiperdum* minicircles, compared by a computer program, were found to present some interesting similarities. The most striking feature was the presence in *T. equiperdum* minicircles of a sequence of about 130 nucleotides homologous to that found in *T. brucei* minicircles (18) (Fig. 5). This region, from nucleotide 707 to 839, differs by only 12 bases from that of *T. brucei* minicircles clone 201 (nucleotides 452–584) and by 15 bases from that of *T. brucei* minicircles clone 51 (nucleotides 245–377). One of these base differences introduces a TAG 779 codon which closes in *T. equiperdum* the largest open reading frame of the two *T. brucei* minicircles. The homologous region found in *T. equiperdum* minicircle is composed of 42.2% guanine·cytosine base pairs which accounts for 24.6% of the rest of the molecule. Outside of these homologous regions, no significant homology was found between *T. equiperdum* and *T. brucei* minicircles. The presence of homologous regions accounting for about 12% of the minicircle in two closely related species of trypanosomes suggests that this region is sufficiently important to be preserved. It could contain, for example, the region of initiation of replication. It would be interesting to know whether the same homologous region exists in other species of trypanosomes. We shall not review the many speculations about the biological role of kDNA (1, 29). However, we may recall that we have induced in *T. equiperdum* treated with drugs such as ethidium bromide, acriflavine (30), or berenil (31) a complete and irreversible loss of kDNA sequences. We have also isolated a naturally occurring strain of *T. equiperdum* without kDNA (12). The problem of the biological role of the kDNA remains open.

The authors are indebted to Dr. Francis Schaeffer of the Institut Pasteur for the facilities in the computer work. This work was supported by the Centre National de la Recherche Scientifique (ATP 4253).

1. Englund, P. T. (1981) in *Biochemistry and Physiology of Protozoa*, eds. Levandowsky, M. & Hutner, S. M. (Academic, New York), 2nd Ed., Vol. 4, pp. 333–383.
2. Riou, G. & Saucier, J. M. (1979) *J. Cell Biol.* **82**, 248–263.
3. Simpson, L. & Simpson, A. M. (1978) *Cell* **14**, 169–178.
4. Fouts, D. L. & Wolstenholme, D. R. (1979) *Nucleic Acids Res.* **6**, 3785–3804.
5. Riou, G. & Yot, P. (1975) *C.R. Acad. Sci. Série D Paris* **280**, 2701–2704.
6. Riou, G. & Yot, P. (1977) *Biochemistry* **16**, 2390–2396.
7. Riou, G., Belnat, P., Barrois, M. & Gabillot, M. (1980) *J. Protozool.*, 27, 90A.
8. Riou, G. & Barrois, M. (1979) *Biochem. Biophys. Res. Commun.* **90**, 405–409.
9. Vickerman, K. (1978) *Nature (London)*, **273**, 613–617.
10. Capbern, A., Giroud, C., Baltz, T. & Mattern, P. (1977) *Exp. Parasitol.* **42**, 6–13.
11. Lanham, S. M. & Godfrey, D. G. (1970) *Exp. Parasitol.* **28**, 525–534.
12. Riou, G., Baltz, Th., Gabillot, M. & Pautrizel, R. (1980) *Mol. Biochem. Parasitol.* **1**, 97–105.
13. Kroecker, W. D. & Laskowski, M. S. R. (1977) *Anal. Biochem.* **79**, 63–73.
14. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
15. Galibert, F., Herisse, J. & Courtois, G. (1979) *Gene* **6**, 1–22.
16. Schildkraut, C. L., Marmur, J. & Doty, P. (1962) *J. Mol. Biol.* **4**, 430–443.
17. Wang, J. C. (1974) *J. Mol. Biol.* **89**, 783–801.
18. Chen, K. K. & Donelson, J. E. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 2445–2449.
19. Barrell, B. G., Bankier, A. T. & Drouin, J. (1979) *Nature (London)* **282**, 189–194.

20. Macino, G., Coruzzi, G., Nobrega, F. G., Li, M. & Tzagoloff, A. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 3784–3785.
21. Fox, T. D. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 6534–6538.
22. Subramanian, K. N., Dhar, R. & Weissman, S. M. (1977) *J. Biol. Chem.* **252**, 355–367.
23. Denniston-Thomson, K., Moore, D. D., Kruger, K. E., Furth, M. E. & Beattner, F. R. (1977) *Science* **198**, 1051–1056.
24. Brack, Ch., Delain, E. & Riou, G. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 1642–1646.
25. Riou, G. & Gutteridge, W. E. (1978) *Biochimie* **60**, 365–379.
26. Baker, C. C., Herisse, J., Courtois, G., Galibert, F. & Ziff, E. (1979) *Cell* **18**, 569–580.
27. Frasch, A. C. C., Hajduk, S. L., Hoeijmakers, J. H. J., Borst, P., Brunel, F. & Davison, J. (1980) *Biochim. Biophys. Acta* **607**, 397–410.
28. Donelson, J. E., Majiwa, P. A. O. & Williams, R. O. (1979) *Plasmid* **2**, 572–588.
29. Simpson, L. (1972) *Int. Rev. Cytol.* **32**, 139–207.
30. Riou, G., Belnat, P. & Benard, J. (1980) *J. Biol. Chem.* **255**, 5141–5144.
31. Riou, G. & Benard, J. (1980) *Biochem. Biophys. Res. Commun.* **96**, 350–354.