Molecular Biology, Vol. 29, No. 6, Part 2, pp. 801–805, 1995

Structure of Minicircle Kinetoplast DNA from a River Perch Trypanosome

L. I. Baron,[1] V. Yu. Yurchenko,[1] and A. A. Kolesnikov[1]

UDC 577.113.5

Translated from *Molekulyarnaya Biologiya*, Vol. 29, No. 6, pp. 1341–1348, November-December, 1995.

Original article submitted May 5, 1995.

For the first time the complete primary structure (1599 bp) was determined for a minicircle kinetoplast DNA (kpDNA) of a lower-vertebrate trypanosome. Some features were common with all *Trypanosomatidae* minicircles, such as the presence of conserved and variable regions within one minicircle, and invariant dodecameric 5′-d(GGGGTTGGTGTA)-3′ and hexameric 5′-d(GGGCGT)-3′ sequences. On the other hand, the kpDNA under study had some features distinguishing it from the minicircles of higher-vertebrate trypanosomes. It proved to be an incomplete dimer based on two unversal dodecamers as the nuclei of symmetry, the sequences around which form a 130-bp direct repeat with a few substitutions. No bent-helix structures were found. The regions of maximal homology between the minicircle examined and those of other trypanosomatids formed discrete clusters along the molecule. PCR analysis demonstrated that such dimeric type of organization (with two conserved regions) is typical of many minicircle kpDNAs from freshwater fish trypanosomes. At the same time, the kpDNA associates of the latter appear to contain a minor portion of minicircles with a single conserved region.

Key words: trypanosomatids; minicircle kinetoplast DNA; polymerase chain reaction

The kinetoplast DNA (kpDNA) in Protozoa of the order *Kinetoplastida* accounts for 10–30% of total cell DNA, and comprises two types of molecules: maxicircles and minicircles.

The 20–50 maxicircles in the cell are transcribed and function analogously to the mitochondrial DNA in other organisms [1]. By contrast, minicircles are unique both in size (0.47–5.2 kbp) and in number ($5 \cdot 10^3$–$5 \cdot 10^4$ per cell) [2]. The kpDNA is spatially organized in a multidimensional network associate 1–2 μm in size. Until recently, no clear function could have been asigned to minicircles. Various hypotheses were advanced, from regulation of the maxicircle transcriptional activity by selective recombination therewith [3] to accumulating the genotypic variation that would compensate for the lack of a sexual process in trypanosomatids [2].

To date, primary structures have been published for some 50 full-sized minicircles of trypanosomes from genera *Crithidia, Leishmania, Trypanosoma,* and *Blastocrithidia* [4–6].

Being highly heterogeneous, the minicircles still have some structural features in common: their sequences have conserved and variable regions (CR and VR). For all objects studied, the CR nuclei are the dodecamer 5′-d(GGGGTTGGTGTA)-3′ and the hexamer 5′-d(GGGCGT)-3′ [7], acting as replication origins for the L and the H chains, respectively. The length of CR ranges 100–200 bp among different species and is virtually constant for minicircles of different classes isolated from the same associate. Next in the 5′ direction is a region of relative (species-specific) conservation, which is practically the same for all individuals of one species. The rest is represented by hypervariable DNA [8].

Also common for minicircle kpDNAs are the high copy number and the presence of regular oligo(A) stretches prone to form bent-helix structures. A region of this type has been found in minicircles of most trypanosomatids, except those of *T. cruzi* [9] that have tetrameric organization with four CR.

For a long time the attempts at finding a minicircle transcript failed to yield an unequivocally positive result. The interest to minicircle transcription has been spurred quite recently by the discovery of the phenomenon of uridyl editing of template RNA, which

[1]Chair of Molecular Biology, Biological Faculty, Lomonosov State University, Moscow, 119899.

0026-8933/95/2906-0801$12.50

involves special RNA templates coded for by minicircle kpDNA ("guide RNAs" [10]).

In the genus *Trypanosoma*, minicircles have been characterized in detail only for the parasites of warm-blooded organisms. For other groups of trypanosomes, the information of minicircle structure is fragmentary, and mainly concerns their size and sequence heterogeneity based on restriction analysis and cross-hybridization. At the same time, fish and amphibian trypanosomes were found to have substantially larger minicircle kpDNAs (1.6 and 5.2 kbp, respectively), which suggests specific features in their structure.

MATERIALS AND METHODS

The *T. carrasii* isolate from *Perca fluviatilis* (Ts-Pf) is deposited in the collection of the Institute of Parasitology, Czech Republic; its characteristics, taxonomic position, and conditions for kpDNA isolation were described elsewhere [11]. The mean yield was 1–3 μg per 10 mln cells.

Use was made of restriction endonucleases from Fermentas (Vilnius), Exo III and S1 nucleases from Pharmacia and Boehringer, Klenow fragment from Fermentas or Pharmacia.

Labeled dNTPs were from Amersham.

Molecular cloning was done using *E. coli* JM 109 and pBlueScript KS+.

The DNA was sequenced by the Sanger dideoxy technique on single-stranded templates, using the unidirectional deletion strategy by the USB protocol.

The PCR was run as follows: 5 min at 93°C; 30 cycles of 1 min at 66°C, 1.5 min at 72°C, 1 min at 93°C; completion for 6 min at 72°C. The primers were:

1) 5′-d(GCCCTGCCCGCACGTTGA)-3′;
2) 5′-d(TGGTGCTTCGATAGGGGTT)-3′;

(henceforth the prefix "d" is omitted, sequences are spelled in the 5′→3′ direction).

The DNA sequence was analyzed by the DNASIS (Hitachi Software Eng. Co. Ltd), GENEBEE (Kombi cooperative), PCGENE-6.5 (IntelliGenetics Inc. & Genofit SA 1991) packages, graphic similarity matrix building program RUSTEM (R. F. Nakipov, Institute of Protein Research, Russian Academy of Sciences), and the MULTALIN multiple sequence alignment program (Corpet 1988).

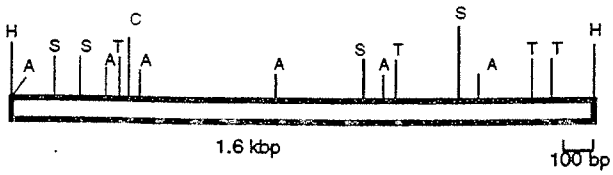

Fig. 1. Restriction map of pmiTca2: A = *Alu*I; C = *Cfr*13I; H = *Hind*III; S = *Sau*3A; T = *Taq*I.

RESULTS AND DISCUSSION

Sequencing of the Ts-Pf Minicircle

Assessment of the representativity of the minicircle under study by comparative restriction analysis and cross-hybridization (data not shown) demonstrated that this pmiTca2 minicircle belongs to a major class of sequences. Figure 1 presents its restriction map.

The sequencing strategy for the minicircle insert is displayed in Fig. 2. The overwhelming majority of nucleotides were determined for the H chain. Practically throughout the insert every region has been read from several deletion derivatives independently.

Analysis of Primary Structure

The complete H-chain sequence in the Ts-Pf minicircle is 1599 nt (Fig. 3). The minicircle is of the dimeriic type, with a 130-bp imperfect direct repeat whose nuclei are universal blocks GGGGTTGGTGTA (ori-L) and GGGCGT (ori-H) spaced by 85 nt. The H chain outside the CR has 23 direct repeats more than 10 nt long; the longest one (14 nt) is TAATTAGTTCTATA appearing twice, at positions 383 and 1176. No sequences over 10 nt were found to occur more than twice.

One of the 14-nt copies contains the most part of the TCTATAGA palindrome (position 1184). In addition, there are five 8-nt palindromes; one of these (GAAATTTC) is within the CR and therefore occurs twice (positions 82 and 882). Including these palindromes, there are 20 inverted repeats of 8 nt or more (three 9-nt ones); two of these are in the CR: AAAGCCATA at positions 123 and 924.

We found no inverted 18-nt repeats or at least similar clusters flanking the CRs; according to Jasmer and Stuart [12], the distribution of such repeats along the chain reflects the evolutionary process among African trypanosomes.

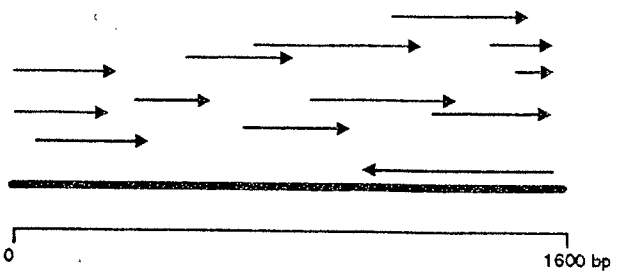

Fig. 2. Sequencing strategy for the Ts-Pf minicircle kpDNA.

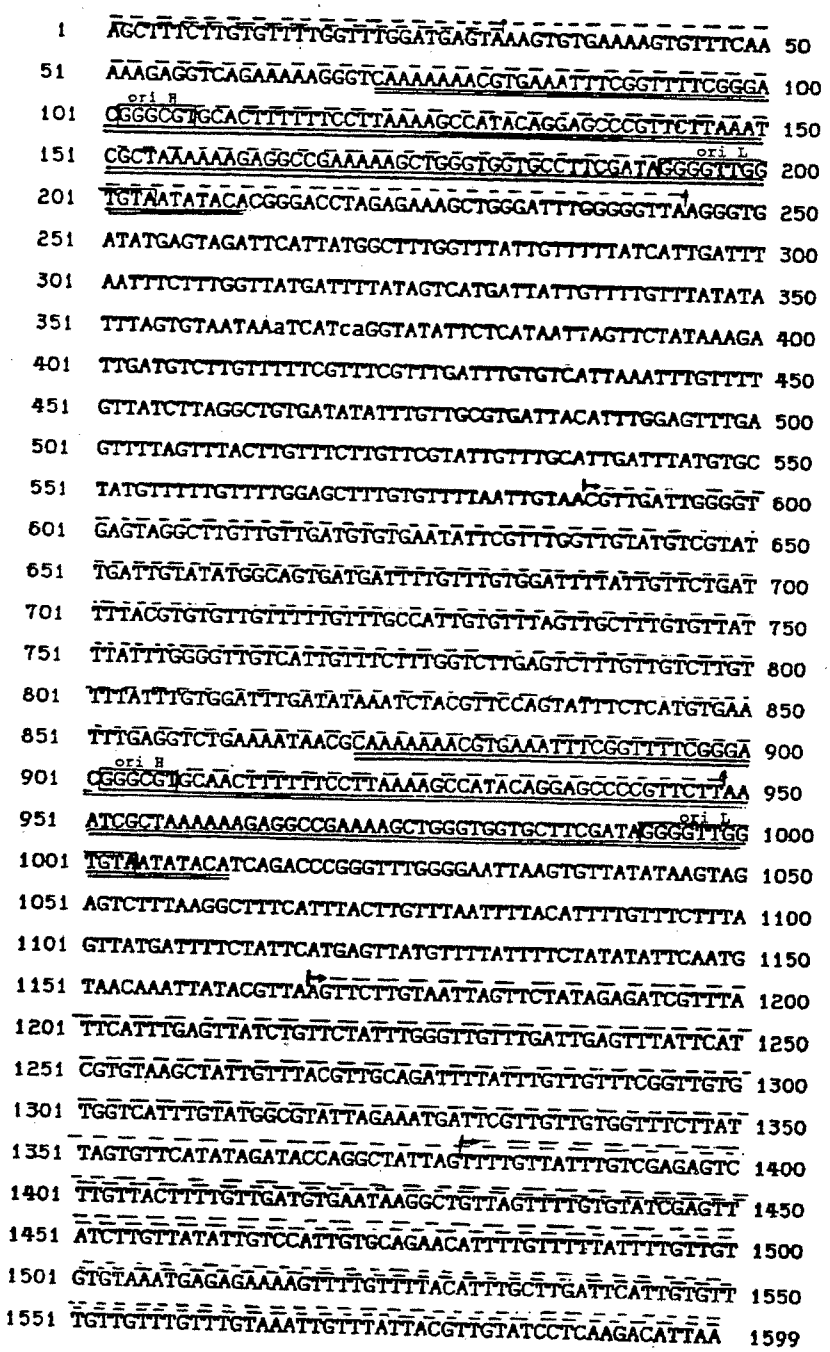

Fig. 3. Primary structure of the full-sized minicircle clone pmiTca2 of Ts-Pf kpDNA; double underlining marks 130-bp direct repeats; replication origins are framed; dashed lines above the sequence mark the largest ORFs (→, start; ↑, stop codon).

The internal homology in the minicircle undr study is mainly determined by the CR repeat. Interestingly, within the CR the homology is neither uniform nor constantly declining away from the universal dodecamer, but forms blocks spaced by rather long poorly homologous stretches.

Analyzing the admissible (in free energy) secondary structures, we failed to detect any hairpins flanking the potential open reading frames and bent helices, which, in Simpson's opinion [1], are typical of trypanosomatids. A potential bent-helix region is at the 5′ end of one CR. However, experimental checking revealed no anomalous migration of restriction fragments in 6% PAG electrophoresis. Maybe such a structure is formed upon interaction with other components of the kpDNA associate (as in *T. cruzi*).

Comparison of pmiTca2 with Other Minicircles

Comparison of the pmiTca2 H chain with the minicircles of other trypanosomatids in respect of nucleotide composition and distribution profiles demosntrates that for evry taxon the percentage of a particular nucleotide varies around some mean value. Thus minicircles of leishmaniae are much richer in C than the trypanosomal ones; the crithidial minicircles contain far less A than those of leishmaniae or trypanosomes. The H chain of our pmiTca2 has the highest T (44.65%) and the lowest A content (22.45%). Incidentally, the *T. cruzi* minicircles, whose H chain is also poor in A (24.62%), also have no bent-helix regions.

As already mentioned, the H chain of trypanosomal minicircles is usually poor in C (6–10%). However, they are nonuniformly distributed along the chain, forming C-enriched (up to 30%) regions, which in the distribution profile appear as a bidentate peak with the first tooth substantially higher. Interestingly, in all trypanosomes the universal dodecamer in the H chain is at the 3′ end of the second tooth. In the minicircles of leishmaniae and crithidiae, the C residues are almost uniformly distributed along the H chain.

Sequence alignment for full-sized minicircles demonstrated that, with a view to evolutionary assessment, it is reasonable to compare regions close to the CR 2]. In this work, the argument for subalignment and constructing the maximal similarity tree was the vicinity of the second dodecamer in pmiTca2. The evolutionary tree built on the basis of this subalignment is depicted in Fig. 4. The partners were the corresponding regions of the minicircles of various trypanosomes (*T. brucei, T. congolese, T. cruzi, T. equiperdum, T. lewisi*), with those of leishmaniae and crithidiae as representatives of more systematically remote trypanosomatids (EMBL data). The unrooted similarity tree, where the branch length is proportional to the number of noncoincidences in subalignments, distinctly shows the separateness of the Ts-Pf under study. It is noteworthy that the conservedness not simply declines farther from the universal dodecamer, but is manifest as quite discrete blocks 5′-ward from the dodecamer and spaced by unlike regions often containing extensive deletions. This corresponds to the block homology of the two CRs within one pmiTca2, and contradicts the idea of a gradient conservedness of the minicircle constant region [2].

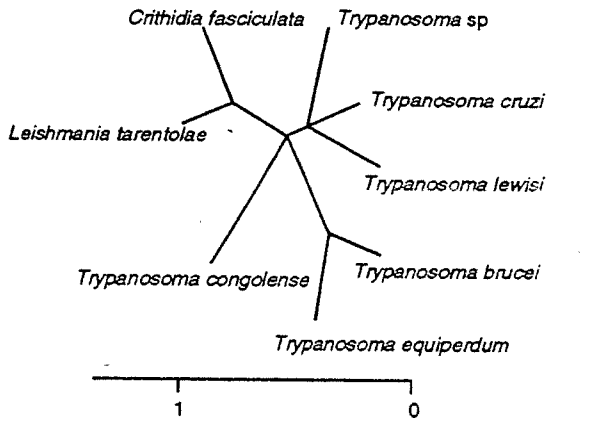

Fig. 4. Evolutionary (maximal topological similarity) tree based on aligning the conserved regions of minicircles with the GENEBEE package (module TREE).

Potential Open Reading Frames

The search through both chains (using the mitochondrial code dictionary of protozoa) revealed four large (over 100 amino acids) potential ORFs: three (ORF 1–3) in the H chain and one (ORF 4) in the L chain. The potential ORF was understood as the DNA sequence from one stop codon to the next. The hydrophobicity profiles of the putative polypeptides (built using the SOAP program) corresponded to integral proteins comprising several domains. In the domain structure and hydrophobicity profiles the products of ORFs 1–3 (especially the first) are generally similar to those of polypeptides encoded by ORFs of *C. fasciculata, L. tarentolae,* and *T. brucei* [1] although the primary structures are quite different. The hydropathy profile of the ORF 4 product corresponds to a hydrophilic matrix protein (mean index 11.76 arb.un.), which is unlikely for proteins encoded by the mitochondrial genome. This conclusion is completely in line with the established viewpoint on the preferential association of the mitochondrial genome with the H chain.

It should, however, be admitted that the coding probability estimates according to the Shepherd method with the PCGENE program proved extremely low for all four potential ORFs. Nonetheless, such a possibility cannot be ruled out.

Organization of the Minicircle *In Vivo*

To check the unified pattern of organization of the minicircle component of the kpDNA associate in freshwater fish trypanosomes, we conducted a series of PCR amplifications of kpDNA with primers corresponding to two sequences from the CR of pmiTca2 [13]. These sequences are localized at opposite ends of the CR, and primers were chosen so that the PCR would be directed outside the CR, i.e., so that the products would be kpDNA regions beyond the CR. The rationale was that the product size would be

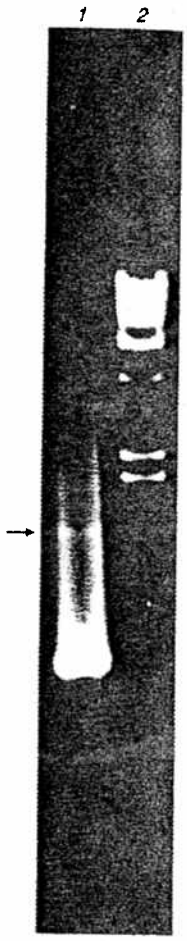

Fig. 5. PCR analysis with the Ts-Pf associate as template (*1*); the minor 1500-bp component is indicated by arrow; 2) λ/*Pst*I markers.

750–800 bp if the minicircle has two CRs, about 1500 bp if there is a single CR, and different otherwise.

Upon amplification of the plasmid studied, the PCR product proved to be 750 bp long, testifying to two CRs in pmiTca2. The same was observed with the native kpDNA associate (Fig. 5), but overloading revealed a minor component of about 1500 bp (indicated with arrow). To check whether this was a PCR artifact, the minor fraction wa isolated from the gel and amplified again; the product was the same (not shown). These data can be interpreted as that the overwhelming majority of Ts-Pf minicircles are organized according to the dimeric type, while a small portion thereof harbor a single CR. A similar pattern was observed in PCR analysis of native kpDNA of trypanosomal isolates from freshwater fishes (*Abramis bramis, Esox, Rathus*). Like results were obtained in the Institute of Parasitology, Czech Academy of Sciences (Jirku et al., 1995, in press) on 12 trypanosomal isolates.

DISCUSSION

The literature reports four trypanosomatids wherein the kpDNA minicircles are organized according to the dimeric type, i.e., have two CRs harboring universal blocks (being thus also duplicated): these are *T. lewisi* [14], *C. fasciculata, Phytomonas serpens,* and *Proteomonas brevicola* [1]. Among the trypanosomatids studied, *T. cruzi* and a systematically close *T. rangeli* were found to have tetrameric-type minicircles with four CRs [15]. Intriguingly, the conserved blocks divide the minicircle into equal parts. Thus in the dimeric-type minicircles the CRs map at 180° to each other, so that the hypervariable parts (70–80% of total length in all cases) separating them are roughly equal in length; in tetrameric-type minicircles the CRs are found at every 90°. It appears logical to regard the blocked organization of minicircles as a consequence of some evolutionary events, e.g., amplification of the region carrying vital DNA replication origins. After such hypothetical amplification, the CR and VR sequences must have evolved independently: while the CRs of different trypanosomatids have substantial and extensive homology, their VRs are totally dissimilar.

It is noteworthy that the two CRs of the Ts-Pf minicircle differ in the extent of homology to those of other trypanosomatids. Thus overall alignment, using different software and various parameters, yielded the highest homology of the **second** Ts-Pf CR (positions 890–1005) with the minicircle CRs of other trypanosomatids, except for all clones of *T. congolense* and two of the five *T. cruzi* clones available in the database, which exhibited most homology to the **first** CR of Ts-Pf. It is a further challenge to find out whether this is a happenstance or evidence for the greater functionality of the second Ts-Pf minicircle.

REFERENCES

1. L. Simpson, *Int. Rev. Cytol.*, **99**, 119–179 (1986).
2. L. Simpson, *Annu. Rev. Microbiol.*, **41**, 363–382 (1987).
3. P. Borst, F. Fase-Fowler, J. H. J. Hoeijmakers, and A. C. C. Frasch, *Biochim. Biophys. Acta*, **610**, 197–210 (1980).
4. S. S. Challberg and P. T. Englund, *J. Mol. Biol.*, **138**, 447–472 (1980).
5. P. T. Englund, *J. Biol. Chem.*, **284**, 4895–4900 (1979).
6. A. C. C. Frasch, S. G. Goijman, J. J. Cazzulo, and A. O. M. Stoppani, *Mol. Biochem. Parasitol.*, **4**, 163–170 (1981).
7. D. S. Ray, *Mol. Cell. Biol.*, **9**, 1365–1367 (1989).
8. D. P. Jasmer and K. Stuart, *Mol. Biochem. Parasitol.*, **18**, 257–269 (1986).
9. A. C. C. Frasch, D. O. Sanchez, and A. O. M. Stoppani, *Biochim. Biophys. Acta*, **782**, 26–33 (1984).
10. B. Blum, N. Bakalara, and L. Simpson, *Cell*, **60**, 189–198 (1990).
11. A. A. Kolesnikov and P. Zajichek, *Biol. Nauki*, **12**, 24–30 (1990).
12. D. P. Jasmer and K. Stuart, *Mol. Biochem. Parasitol.*, **18**, 321–331 (1986).
13. H. Avila, A. Goncalves, S. N. Nehme, C. M. Morel, and L. Simpson, *Mol. Biochem. Parasitol.*, **42**, 175–188 (1990).
14. M. Ponzi, C. Birago, and P. A. Battagila, *Mol. Biochem. Parasitol.*, **13**, 111–119 (1984).
15. R. P. Recinos, L. V. Kirchholf, and J. E. Donelson, *Mol. Biochem. Parasitol.*, **63**, 59–67 (1994).