Peculiar sequence organization of kinetoplast DNA minicircles from *Trypanosoma cruzi*

Wim Degrave\(^1\), Stenio P. Fragoso\(^1\), Constanca Britto\(^1\), Hugo van Heuverswyn\(^1\), Getachew Z. Kidane\(^1\), Maria A.B. Cardoso\(^1\), Rita U. Mueller\(^1\), Larry Simpson\(^2\) and Carlos M. Morel\(^1\)

\(^1\)Department of Biochemistry and Molecular Biology, Instituto Oswaldo Cruz, Rio de Janeiro, Brazil, and \(^2\)Department of Biology and Molecular Biology Institute, University of California, Los Angeles, CA, U.S.A.

(Received 7 May 1987; accepted 31 July 1987)

The sequences of two minicircles from the kinetoplast DNA of the CL strain and one of the Y strain of *Trypanosoma cruzi* are reported. These 1.4 kb molecules have a peculiar sequence organization, the most distinctive feature being the occurrence of a 120 bp sequence repeated four times, located at 0, 90, 180 and 270 degrees along each circle. We have termed these conserved regions in this species ‘minirepeats’. Minirepeats have a 3-fold higher concentration of cytosine residues in comparison with the variable regions and contain the universal 12-mer motif GGGTTGGTGA present in all sequenced minicircles and which was shown to be involved in DNA replication. A consensus sequence of *T. cruzi* minirepeats was determined using the 20 minirepeats present in five known *T. cruzi* minicircle sequences. This consensus sequence contains regions which have been remarkably well preserved in strains which show great biological diversity. In addition a low level of intraminicircle sequence similarity was also observed within the variable region, but this similarity did not extend between strains. The abundance of conserved minirepeat sequences containing invariant restriction sites in *T. cruzi* cells may prove valuable for the development of new direct diagnostic methods for Chagas’ disease based on DNA probe technology.

Key words: *Trypanosoma cruzi*; Kinetoplast DNA sequence; Minicircle DNA; DNA sequence

Introduction

Restriction endonuclease digestion of kinetoplast DNA molecules has been used as the basis of schizodeme analysis, a biochemical method for the characterization of trypanosomatids at the genotypic level [1,2]. Using this approach, isolates, stocks, strains and clones of several kinetoplastid protozoa, including the human pathogens *Trypanosoma cruzi* and *Leishmania* sp., have been analyzed [3-8].

In order to have a better understanding of the structural organization and evolution of kDNA minicircles, which represent the major sequence component of the kDNA, and also to investigate the possibility of developing new approaches for the direct parasitological diagnosis of Chagas’ disease through the use of DNA probes, a sequence analysis of several cloned *T. cruzi* minicircles was carried out. Here we describe the sequence of three minicircles isolated from the Y and CL strains of this parasite. These strains represent biologically divergent organisms which are widely used as model systems in many laboratories [9].
Materials and Methods

Cells. The CL strain of *T. cruzi* and the derived clonal population CL-1416 were the kind gifts of Dr. E. Chiari, University of Minas Gerais, Belo Horizonte and of Dr. D.F. Almeida, Federal University of Rio de Janeiro, Brazil, respectively. Clonal population y2 (clone 2) from the Y strain was the generous gift of Dr. Z. Brener, Research Center R. Rachou, Belo Horizonte, Brazil.

Preparation of kDNA and cloning of minicircles. Epimastigotes were grown in LIT B medium [1] to a density of 5 × 10⁶ and kDNA was isolated as described [11]. Minicircles were liberated from the networks by EcoRI digestion and either cloned directly in pBR325 (clone pTc-21 from wild type CL strain) or first purified by electrophoresis in 0.8% agarose gels and cloned in M13mp8 (clones KM-14 and KY-13 from the cloned cultures of CL- and Y- strains, respectively). Recombinant clones containing *T. cruzi* minicircles were identified by colony or plaque hybridization using nick-translated total kDNA from the respective strains as a probe.

DNA sequence analysis. Insert DNA from pTc-21 was sequenced using the chemical degradation method [12], and insert DNAs of KM-14 and KY-13 were sequenced by the dideoxy protocol [13].

Computer analysis. Sequence data was analysed using the programs developed by R. Staden and...
by the University of Wisconsin Genetics Computer Group (UWGCG) using a VAX 11/750 computer, and by the Micro-Genie Software Package from Beckmann Instruments and the Pustell/IBI Software Package in an EGO (IBM-PC compatible) microcomputer. The previously published $T.~cruzi$ minicircle sequences [14,15] were included in this analysis.

### Results

The sequences of the minicircle inserts of pTc-21, KM-14 and KY-13 are presented in Fig. 1. The sequencing strategies used as well as the distribution of the recognition sites of several restriction endonucleases are given in Fig. 2.

The regular distribution of some restriction sites along the molecules corroborates previous findings which were interpreted as due to the occurrence of sequence repetition [2,16]. Fig. 3A shows the DIAGON dot matrix of KM-14 versus itself, and Fig. 3B the comparison of KM-14 with KY-13. In both cases one can see sets of parallel lines running at 45 degrees and regularly distributed inside the matrix, implying that there is sequence reiteration and a non-random distribution of the minirepeats in the circles. We used a relatively low stringency (17/31, proportional algorithm) to show also the low level of homology between (parts of) the variable regions.

The sequence homologies of the minirepeats shown in the dot matrix analysis of Fig. 3 can be visualized directly in the alignments shown in Fig.

---

**Fig. 1.** Nucleotide sequences of the cloned $T.~cruzi$ minicircles from KY-13 (A), pTc-21 (B) and KM-14 (C). The 120 bp minirepeat-sequences are underlined and the 12-mer sequence, present in all minirepeat sequences, is marked with a thick bar.

**Fig. 2.** Sequencing strategies and restriction endonuclease sites of $T.~cruzi$ minicircles KY-13, pTc-21, and KM-14. Minirepeats are marked with a thick bar. Restriction sites are coded as follows: E, EcoRI; H, HaeIII; R, RsaI; T, TaqI; M, MspI; Hi, HincII; B, BspI; A, AccI; DD, DdeI; HF, HinfI; C, ClaI.

A scale in base pairs is given underneath.
Fig. 3. DIAGON dot matrix of *T. cruzi* minicircles. (A) Analysis of internal repeats in the KM-14 minicircle of *T. cruzi* by DIAGON dot matrix graphics. A window of 31 nucleotides was used with a 'proportional' match of 17/31. (B) DIAGON dot matrix analysis of KM-14 (horizontal) versus KY-13 minicircle sequences (vertical). Same conditions as for A.

4, which lead to the derivation of a minirepeat consensus sequence. Note that this consensus sequence includes two *T. cruzi* minicircles sequences previously published [14,15]. We have extended our alignments beyond the 120 bp minirepeats to where multiple alignment is still significant, but it can be seen that overall homology in these regions drops to much lower levels. When aligning the minirepeats from the pTc-21 clone, we saw that this minicircle sequence most probably lacks a small part in the beginning of the first constant region, and thus is not complete. From the alignment and comparison with the consensus sequence we deduced that the missing sequence should be a small EcoRI fragment of 45 bp. We therefore cloned and sequenced small EcoRI fragments from digested total kDNA of *T. cruzi* CL strain in M13mp8. One clone was found which contained a fragment of the expected length and sequence. This sequence was therefore included in the alignment in c11 cst 1 in Fig. 4 (nucleotides 25–69). The existence of nucleotide variations specific for each parasite strain and present within all four minirepeats in a single circle can be noted in this figure. The 'universal' 12-mer sequence GGGTTCGTTGTA [17,18], which has been shown to be involved in minicircle DNA replication [19], is present near the 3' end of the minirepeats. It should be noted that in the case of *T. cruzi* there is a conserved 28-mer sequence at this position.

The sequences between the minirepeats, which we term 'variable' regions a, b, c, and d, show a low level of local sequence homology (at a stringency of 17 out of 31), as shown by the dot matrix analyses in Fig. 3. This sequence homology extends both between variable segments within a circle as well as for variable segments from different circles from the same strain. No sequence homology at this level of stringency was apparent between variable regions from different strains. These regions of sequence homology can be seen better in the dot matrix analyses in Fig. 5.

The non-random, highly organized character of *T. cruzi* minicircle sequences is also apparent in an analysis of the local variations of T, A, C and G nucleotides along each DNA strand. Fig. 6 shows this distribution for the KY-13 minicircle sequence. The local variation of C residues is particularly noteworthy, higher levels of this nucleotide along the minicircle coinciding with the minirepeats. The variable regions contain locally only 0–5% of C residues.

An examination of the five *T. cruzi* minicircle sequences for the presence of an oligo (A) tract with a periodicity of approximately 10 nucleotides was negative. The presence of this tract has been associated in minicircles from other species with the existence of a conformational ‘bend’ in
Fig. 4. Sequence alignment of T. cruzi minirepeat regions from five different kDNA minicircles. The alignments were done using the UWGCG BESTFIT, GAP and LINEUP computer programs. A consensus sequence was obtained using the UWGCG PRETTY program. A certainty level of 60% resulted in uppercase letters in the consensus sequence and a level of 50% gave lowercase letters, described in Results.
Fig. 5. DIAGON dot matrix analysis of variable regions of *T. cruzi* minicircles. (A) Comparison of two variable regions from the same minicircle (KM-14). (B) Comparison of two variable regions from two different minicircles (KM-14 and pTc-21). Same conditions as in Fig. 3.

Fig. 6. Graphic analysis of the local variation of G, C, T and A nucleotides of the *T. cruzi* KM-13 minicircle in a 50 nucleotide window with a shift of three nucleotides. The ordinate represents the percent of each nucleotide within the moving window. The programs used are UWCG WINDOW and STATPLOT.
the DNA molecule. This negative result is in agreement with the absence of electrophoretic migration abnormalities in polyacrylamide gels of restriction fragments from *T. cruzi* minicircles (Degrave, Fragoso and Britto, unpublished results), making it unlikely that the minicircle ‘bend’ has a functional importance, at least in *T. cruzi*.

**Discussion**

**Minicircles as multipurpose molecular markers of *T. cruzi***. The peculiar sequence organization of *T. cruzi* minicircles was discovered 10 years ago when the first restriction analyses were performed on the kDNA of this organism. The fact that similar digestion patterns were obtained with different enzymes, that some endonucleases with a hexanucleotide restriction site were frequent-cutters of minicircles and that the digestion products occurred as fragments having 1/4, 2/4, 3/4 and 4/4 the molecular weight of minicircles led to the proposition of the existence of sequence repetition and peculiar structural organization of these molecules [2,16]. Analysis of preliminary sequence data supported those initial observations [7].

The determination of the sequence of several full-length minicircles from *T. cruzi* (this paper) [14,15] has shown that these molecules have a highly organized, nonrandom, conserved sequence organization. The minicircles from the so-called ‘polar’ Y and CL strains are very similar in structural organization and minirepeat primary sequence, in spite of the very different biological characteristics of these organisms [9]. The sequence data indicate that minicircle sequences can be versatile indicators of the relatedness of *T. cruzi* strains, for these molecules contain sequences which evolve at different rates. Similar results have been obtained with minicircle sequences from *Leishmania mexicana* [20], in which cloned fragments from the conserved minicircle region were found to be species-specific and cloned fragments from the variable region to be strain-specific, and with *T. cruzi* in which cloned minicircle fragments were found to be strain-specific [21].

**Evolution and function of *T. cruzi* minicircles**. A model has been proposed in which *T. cruzi* minicircles obtained four minirepeat sequences through the duplication of an ancestral molecule which contained only two of these sequences (as in *T. lewisi*), which in turn originated from an even more ancestral minicircle with only one copy (as in *Leishmania*) [22]. We would in this case expect the minirepeats to be homologous in pairs. The sequence analysis we have carried out on five minicircles and the corresponding twenty minirepeats has disclosed one such sequence arrangement, which is present in the minicircle from the Y strain sequenced by Gonzalez [15]. Minirepeats 1 and 2 of this minicircle clone are similar to the four minirepeats of the KY-13 clone from the Y strain, while minirepeats 3 and 4 differ substantially (Fig. 4). Although this finding could support the proposed model, the other four molecules did not show similar sequence arrangements. Also, in the variable regions, such a model can not be sustained, possibly due to the apparent lack of selective pressure to maintain specific sequences.

The presence of highly conserved regions in the minirepeats suggests a functional role for these sequences. However it is doubtful that the sequences could encode mRNAs to be translated to polypeptides due to the absence of significant amino acid similarities and similar hydrophobic patterns of open reading frames from the different minicircles (data not shown), and due to the occurrence of insertions and deletions in nearly all minirepeats when compared to the consensus sequence. The possibility of minicircle sequences encoding a set of structural RNAs remains, as does the possibility that the conserved sequences serve a function in DNA replication.

**Minirepeats as potential DNA probes for parasite detection**. Due to the conservation of the highly abundant minirepeat sequences between different *T. cruzi* strains, it would appear that minirepeats represent ideal targets for diagnostic methods based on DNA probe technology [23]. Using synthetic oligonucleotide probes against minirepeats it was shown that these sequences are ubiquitous in *T. cruzi* (Morel, Degrave, Duarte dos Santos, Goncalves and Simpson, unpublished results). Due to the great abundance of these sequences (four minirepeats per minicircle and 30000 minicircles per cell = ∼120000 minire-
peats/parasite), they are potentially good candidates for the development of DNA probes for the detection of *T. cruzi* and the diagnosis of Chagas’ disease. The presence of invariant restriction sites (eg: TaqI, Rsal) in minirepeats could provide yet another tool for the detection and characterization of parasite sequences through DNA probe technology, as has been done in the diagnosis of human genetic diseases [24].

**Acknowledgements**

This work received financial support from the UNDP/WORLD BANK/WHO Special Program for Research and Training in Tropical Diseases, the Conselho Nacional de Desenvolvimento Cientifico e Tecnologico (CNPq), the Financiadora de Estudos e Projetos (Finep) and the Programa de Apoio ao Desenvolvimento Cientifico e Tecnologico (PADCT) to C.M., and from a research grant from the National Institutes of Health (AI09102) to L.S.

**References**


