2476–2482 *Nucleic Acids Research, 1996, Vol. 24, No. 13*

The telomeric GGGTTA repeats of *Trypanosoma brucei* contain the hypermodified base J in both strands

Fred van Leeuwen, Eric R. Wijsman[1], Esther Kuyl-Yeheskiely[1], Gijs A. van der Marel[1], Jacques H. van Boom[1] and Piet Borst*

Division of Molecular Biology, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands and [1]Leiden Institute of Chemistry, Gorlaeus Laboratories, PO Box 9502, 2300 RA Leiden, The Netherlands

Received April 22, 1996; Accepted May 9, 1996

ABSTRACT

We have previously shown that nuclear DNA of bloodstream form *Trypanosoma brucei* contains a novel base β-glucosyl-hydroxymethyluracil, called J. Base J is enriched in minichromosome fractions but not in the minichromosome internal repeats, suggesting the association of J with telomeric DNA. To test whether J is present in the long telomeric $(GGGTTA)_n$ repeat arrays, which are 2–26 kb in *T.brucei*, we have purified these arrays both by hybrid selection and by isolating 2–26 kb fragments from DNA digested with multiple restriction enzymes. We find that in purified telomeric repeats ~13% of T is replaced by J, compared to 0.8% in total DNA, and we estimate that ~50% of the total J is in these repeats. Highly purified complementary strands of the repeats were obtained by alkaline CsCl equilibrium centrifugation. In the $(TAACCC)_n$ strand 14% of T was replaced by J. In the $(GGGTTA)_n$ strand ~36% of the second T was replaced by J; the first T was not detectably replaced. Modified bases have not been found in telomeric repeats before. How the bulky base J affects telomere function and structure in bloodstream form trypanosomes remains to be determined.

INTRODUCTION

African trypanosomes have remarkable chromosome ends and there are many of them. Most of these telomeres are in some 100 minichromosomes, that vary in size between 50 and 150 kb (1). In addition there are ~20 larger chromosomes varying in size between 200 kb and 5.7 Mb (1–3).

In most organisms telomeric DNA consists of a tandemly repeated simple sequence with a strong bias towards G residues in one of the strands (reviewed in 4). The trypanosome telomeres end with $(GGGTTA)_n$ repeats (5,6), which are also found in mammalian chromosomes (reviewed in 7) and are accompanied by fairly standard subtelomeric repeats (6,8–10). What makes these telomeres remarkable is the length of the repeat arrays compared with the chromosome size and their spectacular growth and contraction (11). The length of individual repeat arrays varies between 2 and 26 kb (12,13). Most telomeres grow at an average rate of 6–10 bp/cell division (6,12). Shortening may occur by large deletions (11), but what brings them on is not clear.

Telomeres are also associated with two interesting phenomena observed in *T.brucei*, antigenic variation and subtelomeric DNA modification. *Trypanosoma brucei* can survive in the bloodstream of its mammalian host by regularly changing expression of its major coat protein, the variant surface glycoprotein (VSG). To do so, trypanosomes have a large repertoire of VSG genes which are exclusively expressed in one of the telomeric VSG expression sites (ESs) (reviewed in 14,15). There are at least six and probably ~20 ESs in a trypanosome nucleus and each of these is located near a telomere. Usually, only one of these ESs is active at a time, with transcription of VSG genes directed towards the chromosome end. Gottschling *et al.* (16) have suggested that the inactivation of all ESs but one occurs by telomeric silencing. Recent studies are compatible with this idea (17,18).

DNA modification could be one of the mechanisms involved in silencing of ESs. Transcriptional inactivation of an ES is accompanied by DNA modifications in and around the inactivated telomeric VSG gene (13,19). This modification affects *Pst*I and *Pvu*II sites, it is found in non-transcribed VSG genes near telomeres, but not in silent chromosome-internal VSG genes and it is only found in bloodstream form trypanosomes and not in insect form trypanosomes, which do not transcribe VSG genes. Modification of any *Pst*I and *Pvu*II site is only present in a fraction of the trypanosome population and this fraction increases with the length of the associated telomeric repeat array (13), linking DNA modification to telomeres. A novel hypermodified base, β-D-glucosyl-hydroxymethyluracil (β-gluc-HOMeU) or J, recently identified in trypanosome DNA (20), has the characteristics expected for the modified base postulated to be present in *Pst*I and *Pvu*II sites. J is present in bloodstream form trypanosomes and absent from insect form trypanosomes (21). J is 6-fold enriched in minichromosomes, suggesting that it is associated with chromosome ends.

* To whom correspondence should be addressed

One discrepancy between J and the previously detected subtelomeric DNA modification remained, however: whereas the modification of subtelomeric *Pst*I and *Pvu*II sites would only require 0.0001 mol% J, we find 0.2% (14). A bulky base, such as J, may be expected to block cleavage by nearly all restriction endonucleases (22) and it is therefore unlikely that substantially modified sites would have been missed. This leaves simple repeats not cut by the available restriction enzymes as potential candidates for J localization, telomeric repeats being the most obvious ones. We have therefore purified telomeric repeats of bloodstream form *T.brucei* and we show that on average as much as 16% of T residues in these repeats are replaced by J.

MATERIALS AND METHODS

Trypanosomes

221a bloodstream trypanosomes (MiTat1.2a) of *T.brucei* strain 427 (23) were grown and isolated as described (21). Procyclic (insect form) trypanosomes were grown in the semi-defined medium as described (24).

J-containing oligonucleotides

The fully protected J derivative 3′-*O*-[5-(2,3,4,6-tetra-*O*-benzoyl-β-D-glucopyranosyloxymethyl)]-5′-*O*-dimethoxytrityl-2′-deoxyuridine 2-cyanoethyl-(*N*,*N*-diisopropyl)phosphoramidite (25) was used as a building unit in automated DNA synthesis (Pharmacia Gene Assembler) of the J-containing oligonucleotides $(GGGTJA)_4$, $(GGGJTA)_4$, $(GGGJJA)_4$ and $(ACCCJA)_4$. The synthetic oligomers were used as standards in nucleotide postlabelings to determine the labeling efficiency of J.

Isolation of DNA, blotting and quantitations

Total genomic DNA was isolated as described (26) and resuspended in 10 mM Tris–HCl, 1 mM EDTA, pH 7.4. Digested DNA was transferred to nitrocellulose or Hybond-N (Amersham) by standard procedures (27). For dot blots, DNA was denatured for 20 min on ice in 0.4 N NaOH, neutralized by adding 1 vol. ice-cold 2 M ammonium acetate and blotted onto Hybond-N. Probes were labeled with [α-^{32}P]dATP by random priming. Probes for telomeric repeats were either derived from plasmid pT6, which contains a 330 bp stretch of duplex GGGTTA repeats (6), or were made by 5′ ^{32}P-labeling of oligomers consisting of five telomeric GGGTTA or CCCTAA single-stranded repeats. Other *T.brucei* probes used were a subtelomeric sequence (6,8), the conserved 3′-half of VSG genes, 177, 70 (28) and 50 bp repeats (29), a 570 bp tubulin *Hin*dIII–*Bam*HI fragment from the β-tubulin gene (30), ribosomal DNA (31) and kinetoplast DNA (32). Dot blots were scanned and quantitated on a phosphorimager (Fujix BAS 2000, TINA 2.08b).

Selective hybridization

Genomic trypanosome DNA was sonicated to fragments of an average size of 250–500 bp, denatured by heating and incubated with biotinylated oligonucleotides $(CCCTAA)_4$ or $(GGGTTA)_4$, which had been coupled to magnetic Dynabeads as described by the manufacturer (Dynal). DNA was selected in 50 mM sodium phosphate, pH 7.4, 0.9 M NaCl, 5 mM EDTA, 0.1% SDS, 0.1% BSA and 0.2 mg/ml glycogen for 2 h at 50°C. The beads were then washed three times for 5 min at 50°C with 0.9 M NaCl, twice with 0.5 M NaCl and once with 0.1 M NaCl, all in the presence of 50 mM sodium phosphate, pH 7.4, 0.1% SDS and 5 mM EDTA, and finally incubated in 0.1 N NaOH for 10 min at room temperature. The supernatant, which contained the selected DNA, was neutralized by adding 1 vol. of a mixture of 0.1 N HCl, 0.2 M Tris–HCl, pH 7.4.

Gel fractionation of digested DNA

An aliquot of 500 µg bloodstream trypanosome DNA was digested for 16 h with 1000 U *Alu*I, *Cfo*I, *Hin*fI, *Rsa*I, *Ssp*I and 500 U *Ava*II, deproteinized, phenol extracted and run on a preparative 1% LMP agarose gel. DNA was eluted from the agarose according to Sambrook *et al.* (27).

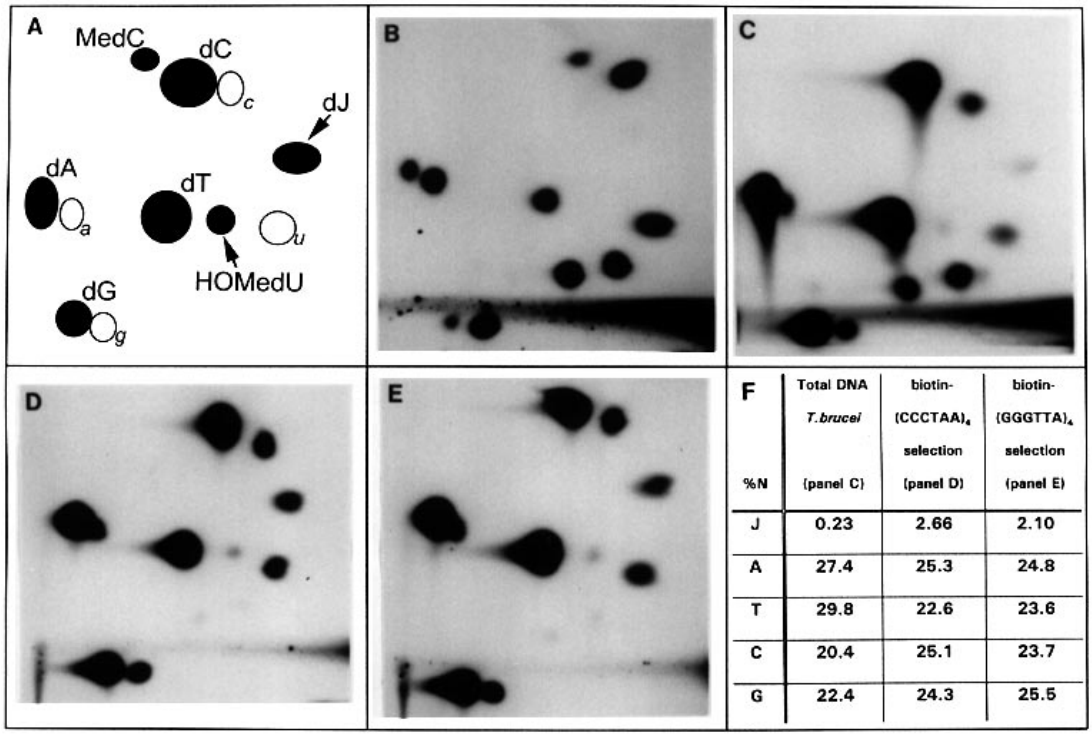
Telomeric strand separation in alkaline CsCl gradients

DNA fragments were run in alkaline cesium chloride gradients [8 ml CsCl (Suprapur; Merck) with a final density of 1.76 g/cm^3 in 0.1 N KOH, 1 mM EDTA] in polyallomer tubes (13 × 51 mm; Beckman) at 40 000 r.p.m. in a Beckman Ti50 rotor (107 000 *g*) for 72 h at 20°C. Forty fractions of 200 µl were collected from the bottom of the tube and 10 µl of each fraction was used to measure the refractive index. Samples of 2 µl were analyzed by dot blot filter hybridization as described above, but without adding NaOH. DNA from pooled fractions was desalted by ethanol precipitation and analyzed by postlabeling for nucleotide composition.

Nucleotide analysis

^{32}P-Postlabeling combined with two-dimensional thin layer chromatography was done as described (21). ^{32}P-Labeled nucleotide 5′-monophosphates were quantitated by scintillation counting of the separate spots or with a phosphorimager (Fujix). Gommers-Ampt *et al.* (21) have noted already that the labeling efficiency of J seemed to vary. This is not surprising, as nucleotides with bulky adducts are known to be labeled inefficiently (33). The chemical synthesis of J (25) made it possible to generate oligonucleotides with known amounts of J and test the labeling efficiency of J-containing nucleotides directly. We found that only 36 ± 9% of J is recovered in the postlabeling assay. The underestimation is probably due to a combination of inefficient labeling by polynucleotide kinase and inefficient dephosphorylation by nuclease P1. We have not succeeded in improving the labeling efficiency of J without severely affecting the labeling of other nucleotides. Fortunately, the labeling efficiency is not affected by excess DNA without J. This allowed us to correct all data presented in this paper to 100% recovery by including samples with known amounts of J in each series of post-labeling analyses. This correction resulted in a level of J in *T.brucei* DNA of 0.23 ± 0.06%, rather than the 0.04–0.10% of total DNA bases reported previously (21).

Depurination of DNA was performed by incubating DNA in 2% (w/v) diphenylamine and 66% (v/v) aqueous formic acid for 17 h at 30°C as described (34). Pyrimidine tracts were purified by three extractions with 6 vol. diethylether, dried by rotary evaporation and resuspended in H_2O. The 5′ pyrimidine was labeled by three consecutive reactions. First, the 5′ terminal phosphate was removed by incubation with 1 U shrimp alkaline phosphatase (USB) in the supplied reaction buffer for 30 min at



F %N	Total DNA *T.brucei* (panel C)	biotin-$(CCCTAA)_4$ selection (panel D)	biotin-$(GGGTTA)_4$ selection (panel E)
J	0.23	2.66	2.10
A	27.4	25.3	24.8
T	29.8	22.6	23.6
C	20.4	25.1	23.7
G	22.4	24.3	25.5

Figure 1. Two-dimensional thin layer chromatography of postlabeled DNA fractions enriched for telomeric repeats. (**A**) The position of the labeled nucleotides is indicated by dN; spots a, u, c and g are the corresponding ribonucleotides AMP, UMP, CMP and GMP. (**B**) The background in the assay is shown in a no-DNA control. (**C**) Bloodstream form *T.brucei* DNA and fractions enriched for telomeric repeats by selection with biotinylated oligonucleotides $(CCCTAA)_4$ (**D**) or $(GGGTTA)_4$ (**E**). (**F**) Quantitation of the percentage of nucleotides (%N) in (C)–(E).

37°C in 10 µl, after which the enzyme was inactivated by a 20 min incubation at 68°C. Second, 2 µl of the dephosphorylation reaction was used to label the first pyrimidine with [γ-^{32}P]ATP and polynucleotide kinase (PNK) at the created free 5′ hydroxyl as described (21), but in the absence of cold ATP. Third, the labeled 5′ pyrimidine was released from the dimer by nuclease P1 (NP1), as described for postlabeling (21). Labeling of the 3′ pyrimidine was achieved by the standard post-labeling reactions for small amounts of DNA. Control experiments with $(GGGTJA)_4$ and $(GGGTTA)_5$ showed that J at the 3′-T position of the dipyrimidine tract was labeled very inefficiently compared to non-depurinated DNA and was mainly recovered as HOMeU. The degree of J replacement (%J) in the standard curves and in the purified telomeric repeats (Fig. 4B) was therefore taken as the sum of HOMeU and J.

RESULTS

Purification of telomeric repeat arrays by hybrid selection

Trypanosoma brucei has a small genome (8×10^7 bp/diploid nucleus) distributed over more than 100 chromosomes, with long telomeric repeat segments of ~2–26 kb making up ~3% of the DNA. To rapidly check for DNA modifications in telomeric repeats hybrid selection was employed, using the Dynal magnetic separation system with two biotinylated oligonucleotides that consist of four telomeric repeat units of either the $(GGGTTA)_n$ or the $(TAACCC)_n$ strand. Complementary DNA strands were selected from denatured bloodstream trypanosome DNA that had been sonicated to small fragments of 250–500 bp. These small fragments can be efficiently selected and contain minimal adjacent subtelomeric sequences that can be co-selected.

To analyze DNA modifications in genomic DNA we used nucleotide ^{32}P-postlabeling combined with separation by two-dimensional thin layer chromatography, a method developed by Reddy *et al.* (35) and applied to trypanosome DNA by Gommers-Ampt *et al.* (21). This assay allows separation of the standard nucleotides (indicated by their base) A, T, C and G from hydroxymethyluracil (HOMeU) and β-D-glucosyl-HOMeU (J). While 5-methylcytosine (MeC) is also efficiently labeled (35), this base is not detectable in total genomic DNA of trypanosomes (Fig. 1C). As shown in Figure 1D and E, the telomere-enriched DNA selected with both oligonucleotides contained high levels of J. Quantitation of the nucleotide spots (see Materials and Methods for details) showed that J was 10-fold enriched compared to total DNA. However, both biotin-$(GGGTTA)_4$ and biotin-$(CCCTAA)_4$ selected DNA in which A, T, C and G were present at equal levels (Fig. 1F). As no release of oligonucleotides from the magnetic beads was detected, we conclude that during hybrid selection of the $(GGGTTA)_n$ strand the complementary $(TAACCC)_n$ strand efficiently re-annealed with its partner (and vice versa) because of the ultrashort repeats. Although it should in principle be possible to avoid re-annealing by using very dilute DNA solutions, this proved impractical. We therefore turned to a combination of classical techniques to separately purify both strands of intact telomeres.

Size fractionation of genomic DNA

Telomeric repeats lack recognition sequences for most type II restriction endonucleases. Digestion of DNA with frequent cutters

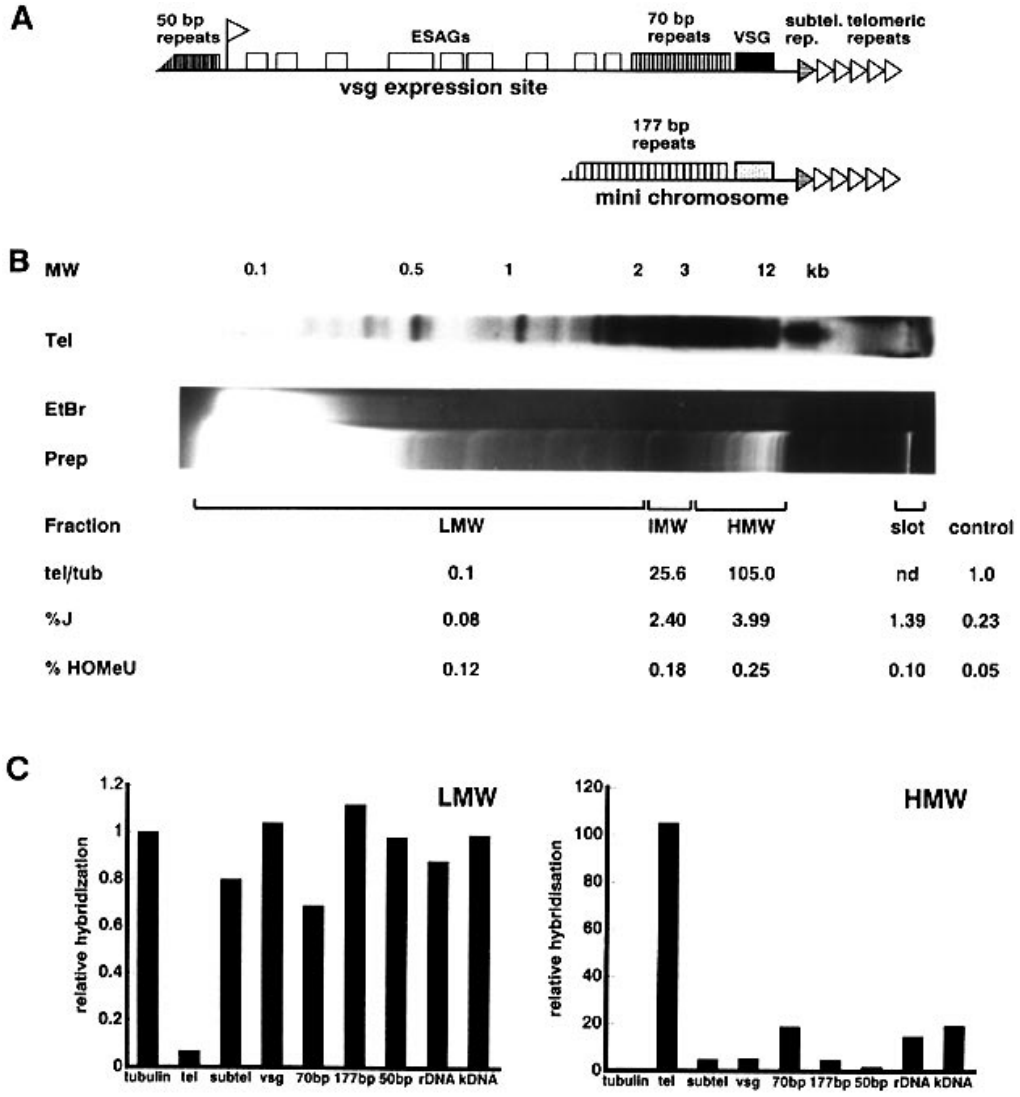


Figure 2. Characterization of DNA fractions obtained by gel fractionation of DNA digested with multiple restriction enzymes. (**A**) Schematic representation of a minichromosome and a VSG expression site, expression site-associated genes (ESAGs) and some repetitive sequences; the flag indicates the ES promoter. (**B**) Bloodstream form *T.brucei* DNA was digested with multiple restriction enzymes (see Materials and Methods), run through a 1% agarose gel and stained with ethidium bromide (EtBr). Part of the gel was blotted and probed for telomeric repeats (Tel). Five gel fractions were isolated from a preparative gel (Prep) as indicated (see text) and the DNA was analyzed by dot blot hybridization for telomeric repeats relative to tubulin genes [tel/tub; control DNA fraction is set at 1; not determined (nd) for slot DNA] and analyzed by post-labeling for %J and %HOMeU of total nucleotides. (**C**) Dot blot hybridization analysis of DNA from the LMW and HMW gel fractions with probes for repetitive sequences known to be present in *T.brucei*: subtelomeric repeats (subtel), the conserved second part of VSG genes (VSG), 70, 177 and 50 bp and rDNA repeats and kDNA minicircles; hybridization signals are quantitated relative to non-fractionated control DNA (bloodstream form DNA digest; tubulin is 1).

and subsequent isolation of the high molecular weight DNA results in DNA that contains most of the telomeric repeats, whereas the bulk of the DNA is digested to small fragments (9). We used a combination of restriction enzymes that cut frequently and/or cut in repetitive sequences known to be present in the *T.brucei* genome, such as subtelomeric repeats (Fig. 2A). Bloodstream form trypanosome DNA was digested with *Alu*I, *Ava*II, *Cfo*I, *Hin*fI, *Rsa*I and *Ssp*I and run through a preparative agarose gel. The ethidium stain of the gel confirmed that most of the DNA was in the bottom part, whereas telomeric sequences, detected by Southern blotting, ran in the top part of the gel (Fig. 2B). The gel was split into five fractions: low molecular weight DNA (LMW), intermediate molecular weight DNA (IMW), high molecular weight DNA (HMW), slot DNA and, as a control, DNA of one lane containing all four fractions (Fig. 2B). DNA eluted from the agarose was analyzed by dot blot hybridization and postlabeling.

The HMW fraction was highly enriched for telomeric sequences and the lack of subtelomeric sequences in this fraction showed that the telomeric fragments had lost most of their neighboring sequences (Fig. 2C). Since any simple repeat sequence lacking recognition sites for the enzymes used will end up in the HMW region, we analyzed the fractionation of a number of known tandemly repeated elements and satellite DNAs, including kinetoplast DNA (kDNA), ribosomal DNA (rDNA) and 70 bp repeats. Although telomeric repeats were 105-fold enriched relative to tubulin genes in the HMW fraction, none of the other repetitive sequences accumulated in this fraction (Fig. 2C). The LMW fraction contained most of the DNA; no significant change

was found for the repetitive sequences except for a clear reduction in telomeric repeats (Fig. 2C). The IMW and slot fractions contained part of the telomeric DNA (Fig. 2B) and considerable amounts of other sequences (data not shown).

Postlabeling of the fractions confirmed the result from the hybrid selection (Fig. 2B). The HMW fraction highly enriched for telomeric DNA showed a 17-fold increase in the level of J, whereas DNA deprived of most of the telomeric repeats (LMW) showed a 3-fold decrease. The intermediate fraction in which telomeres are less purified was still 8-fold enriched for J. This showed that J is not only present at high levels in the longest telomeres. Size fractionation with only four restriction enzymes (*Alu*I, *Ava*II, *Rsa*I and *Ssp*I) resulted in less enrichment for telomeric repeats and a 12-fold enrichment for J in the HMW fraction (data not shown), confirming the correlation between enrichment for telomeric repeats and J in these experiments. We also determined the level of HOMeU in the gel fractions (Fig. 2B). Very low levels of HOMeU cannot be accurately determined with our postlabeling conditions because deaminating activities result in partial conversion of deoxyCMP to deoxyUMP, which co-migrates with deoxyHOMeUMP, close to T (see Fig. 1). It is clear from Figure 2, however, that the higher levels of J in the HMW and IMW fractions correlated with increased levels of HOMeU.

Purification of single-stranded telomeric repeat arrays by alkaline CsCl centrifugation

In neutral CsCl gradients an almost linear relationship exists between the percentage of (C+G) and the buoyant density of DNA unless particular unusual bases are present (reviewed in 36). As we found substantial amounts of base J in telomeric repeats, we tested whether this affected the density of these repeats in neutral CsCl, as this would allow separation of modified from normal DNA. However, telomeric repeats of sonicated DNA from bloodstream trypanosomes and from insect form trypanosomes (which does not contain J) both banded at 1.71 g/cm^3, the density expected for DNA with 50 mol% G+C (data not shown).

In alkaline CsCl solutions mainly G and T residues are titrated and are thought to acquire a cesium ion that increases the density of the DNA (37). This proved to be useful for further purification of telomeric repeats because of their strong GT strand bias. In alkaline CsCl the $(GGGTTA)_n$ strand banded at 1.82 g/cm^3 and the $(TAACCC)_n$ strand at 1.72 g/cm^3 (Fig. 3A). The complete separation of the strands allowed us to study each strand separately, but also allowed separation of telomeric repeat DNA from sequences without a strong GT strand bias, such as 70 bp repeats, kDNA, rDNA and tubulin genes, which banded at a density of ~1.77 g/cm^3. We therefore ran the HMW DNA fraction of Figure 2A in an alkaline CsCl gradient (Fig. 3B), pooled the peak fractions and analyzed them by postlabeling. Analysis of J showed that the modified base is stable in alkaline CsCl and is present at high levels in both strands (Fig. 3C). The degree of modification is 4.9% in the $(GGGTTA)_n$ strand and 2.2% in the $(TAACCC)_n$ strand, which corresponds to 18 and 14% replacement of T respectively. The base compositions of the separated strands were in close agreement with the predicted sequences (Fig. 3C), confirming the high degree of purity of each telomeric strand.

We found equally high levels of J in telomeric repeats purified by three different methods: hybrid selection (Fig. 1), size fractionation (Fig. 2) and repetitive alkaline cesium chloride gradient centrifugation (data not shown). The highest levels, however, were found in the separated strands of the HMW fraction (Fig. 3C). This strongly suggests that we have not co-purified substantial amounts

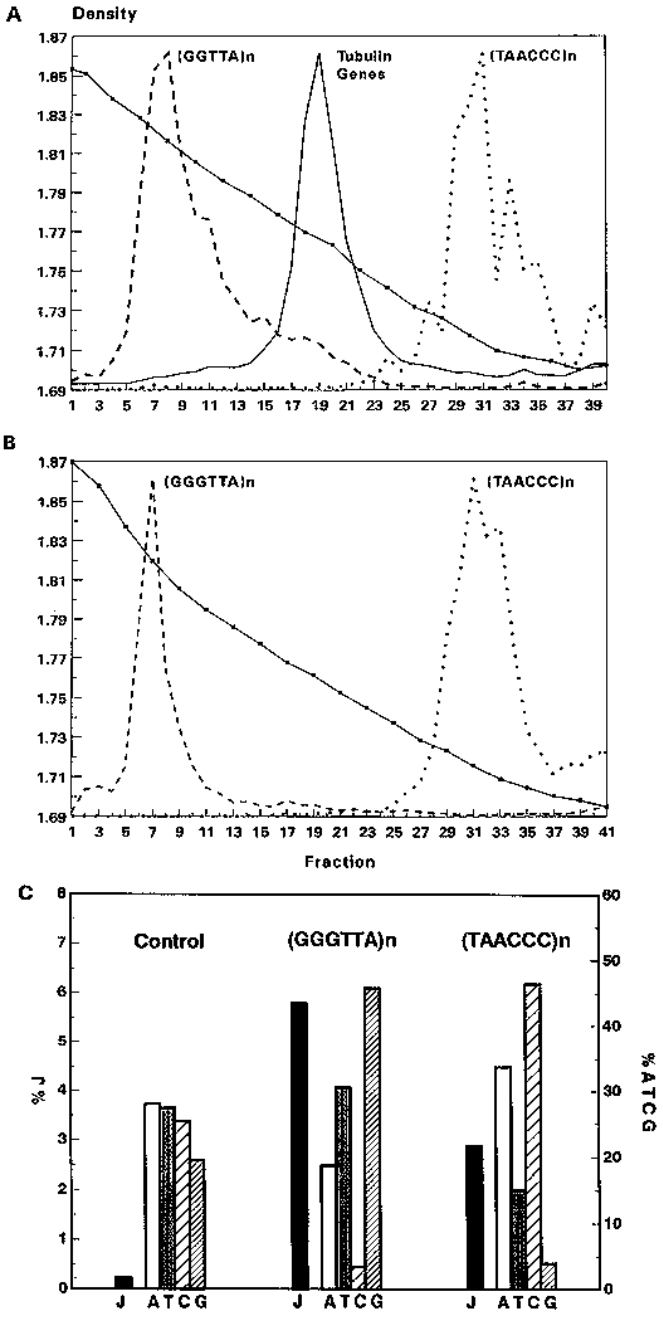


Figure 3. Strand separation of telomeric repeats in alkaline cesium chloride equilibrium gradients. Alkaline CsCl equilibrium centrifugation of (**A**) sonicated bloodstream trypanosome DNA (0.5–20 kbp) and (**B**) the HMW gel fraction (taken from an experiment as in Fig. 2A). Relative dot blot hybridization of the fractions is shown for tubulin genes (only in A) and for oligomers consisting of five telomeric repeats probing for GGGTTA and TAACCC. From gradient (B) fractions 5–11 containing the $(GGGTTA)_n$ strand and 29–35 containing the $(TAACCC)_n$ strand were pooled and further analyzed for nucleotide composition. (**C**) ^{32}P-Postlabeling analysis of the pooled fractions $(GGGTTA)_n$ and $(TAACCC)_n$ from gradient (B). Control DNA is a sample of unfractionated bloodstream form DNA. Percentage of total deoxynucleotides is shown for pdJ on the left axis and for pdA, pdT, pdC and pdG on the right axis.

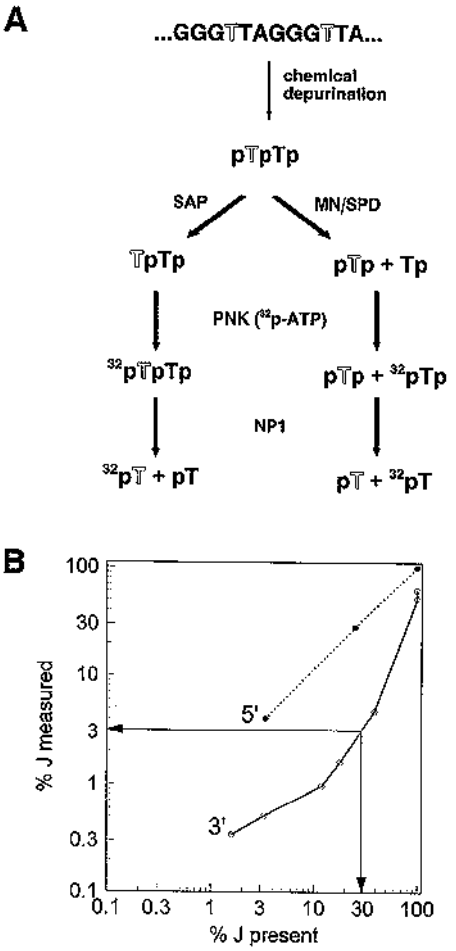


Figure 4. Differential labeling of the two pyrimidine positions in the GGGTTA repeat after chemical depurination. (**A**) Schematic representation of the two specific labeling reactions. The dinucleotide obtained by chemical depurination (see Materials and Methods) contains a phosphate at both the 5′- and the 3′-ends. Enzymes are abbreviated as SAP (shrimp alkaline phosphatase), MN (micrococcal nuclease), SPD (spleen phosphodiesterase), PNK (polynucleotide kinase) and NP1 (nuclease P1). (**B**) Plot of the labeling efficiencies of J at the two positions in the dimer obtained after chemical depurination of oligonucleotides of known composition. The 5′ position labeling was analyzed with dilutions of $(GGGJTA)_4$ in $(GGGTTA)_5$ (·●·); the labeling efficiency of J at the 3′ pyrimidine position was determined with dilutions of $(GGGTJA)_4$ in $(GGGTTA)_5$ (–○–). The arrows indicate the average value of J measured at the 3′-T position in the purified trypanosome telomeric $(GGGTTA)_n$ strand and the estimated degree of J present at that position using the standard curve (–○–).

of contaminating non-telomeric sequences. The low levels of G measured in the TAACCC fraction and C in the GGGTTA fraction (Fig. 3C) were in part caused by nucleotide contamination in the postlabeling enzyme preparations (see Fig. 1B). We have used enzymes from different sources and have dialyzed them extensively, but some background labeling remained which could not be subtracted because it varied with the amount of DNA used in the postlabeling. No background was found for J, however.

Differential labeling of the 5′-T and the 3′-T position in the $(GGGTTA)_n$ strand

To analyze the position of J in the $(GGGTTA)_n$ strand, DNA was chemically depurinated (34) and the resulting dinucleotide pyrimidine tract was purified for differential analysis of the 5′- and 3′-T positions, as outlined in Figure 4A. This method was tested on synthetic oligonucleotides consisting of four telomeric GGGTTA repeat units with J replacing T at either position. To mimic the partial replacement of T by J found in telomeric repeats of *T.brucei*, we made dilution series of the J-containing oligomers in oligomers without J (Fig. 4B). After depurination, a 5′-T, a 5′-J or a 3′-T in the dipyrimidine tract were labeled with 99% efficiency, but a 3′-J was only partially labeled (Fig. 4B). This partial labeling was only seen after chemical depurination. We used the dilution series of $(GGGTJA)_4$ in $(GGGTTA)_5$ as a standard curve to determine the approximate degree of J replacement at the 3′-T position in the purified telomeric $(GGGTTA)_n$ strand of *T.brucei* (see Materials and Methods). We found that the measured 3% J corresponds to ~30% replacement of the 3′-T (Fig. 4B). This high degree of replacement is in agreement with the 0.3% replacement measured for the 5′-T, resulting in an average replacement of ~15% of T, close to the 18% replacement of T found in the non-depurinated GGGTTA strand (Fig. 3C). The low level of J at the 5′ position is probably the 1% non-specific labeling of the other position and may be caused by breaks in the strand that have occurred during the experiment. These data clearly show that the distribution of J in telomeric DNA is not random but that J replaced only two of the three T residues, GGGTTA and TAACCC, with ~30–36 and 14% replacement of those T residues respectively.

DISCUSSION

Our results show that the telomeric repeats of *T.brucei* contain substantial amounts of J. This novel base replaces 14% of the T residues in the $(TAACCC)_n$ strand and 18% in the $(GGGTTA)_n$ strand. Replacement of T in the G-rich strand is highly asymmetric, as the 5′-T was not significantly replaced by J. Even with this high degree of replacement, not every telomeric repeat can contain J. Whether telomeres without J exist and how J is distributed within telomeres remains to be determined. We also found increased levels of HOMeU in telomeric DNA. This could either be due to loss of glucose from J *in vivo* or during DNA isolation, or due to a lag in glycosylation of HOMeU during synthesis of J. Indirect evidence suggests that J is made by modification of T in DNA and that HOMeU is an intermediate in this process (14,38).

Trypanosomes have long telomeric repeat arrays, encompassing ~3% of their genome. An average replacement of 16% of T in telomeric repeats versus 0.8% in total DNA implies that ~50% of the trypanosomal J is in telomeric repeats. The other half is mainly in subtelomeric DNA, as suggested by the distribution of blocked *Pst*I and *Pvu*II sites (13,19,21) and of restriction fragments reacting with antibodies against J (unpublished results).

Exhaustive digestion of non-repetitive DNA with restriction endonucleases has been used before to purify telomeric repeat arrays from mammals (9). Although this method resulted in highly purified trypanosomal telomeric repeat DNA, this DNA still contained significant amounts of other simple repeats, such as 70 bp repeats (Fig. 2C). We therefore tried two other purification methods, hybrid selection and CsCl equilibrium centrifugation. Hybrid selection is rapid, but we were unable to get pure telomeric repeats by this procedure, even if the procedure was successively repeated several times (results not shown). Isopycnic centrifugation in Cs_2SO_4 containing Ag^+ has been used before to obtain a limited enrichment of native human telomeric

DNA (39). Instead of this complex procedure, we have purified single-stranded DNA in alkaline CsCl. This approach is based on the extreme difference in base composition of the complementary strands of the telomeric repeats, allowing separation of these strands from each other and from bulk DNA. The GT-rich strand bands at 1.82 g/cm^3, close to the density of pure poly(dT·dG), 1.83 g/cm^3; the AC-rich strand at 1.72 g/cm^3 has an equilibrium density substantially above the 1.69 g/cm^3 reported for pure poly(dC·dA) (36,37), but still comfortably away from bulk trypanosome DNA at 1.77 g/cm^3 (Fig. 3A). This purification method may also be useful for telomeric repeats from other sources.

In neutral CsCl gradients unusual bases like hydroxymethylcytosine (HOMeC) or glucosylated HOMeC can affect the density of DNA (35,36). Despite the replacement of 16% of T by J (β-gluc-HOMeU) in bloodstream form trypanosome telomeric repeats, we found no detectable difference in density between bloodstream form and insect form telomeric repeat arrays (data not shown). This suggests that either the degree of modification is too low to give a change or that the expected increase in density due to HOMeU (0.039 g/cm^3 if HOMeU replaced every T) is counteracted by a decrease caused by the glucose moiety (the density of glucose is 1.6 g/cm^3; the presence of glucose on part of HOMeC in T-even phages results in a decrease in density of 0.005 g/cm^3; 35).

Base J is not an essential structural component of telomeres in trypanosomes because it is absent from insect stage trypanosomes (20,21). J must therefore be involved in a life cycle-specific telomeric function. The proposed role in shutting down (or tightening the shut-down) of telomeric VSG gene expression sites (13,14,18) remains a plausible possibility. Nevertheless, it remains possible that J is used for telomere-related functions in organisms other than African trypanosomes. Base modification has never been found before in telomeric repeats, but in most organisms the repeats represent such a small fraction of total DNA that the presence of unusual nucleotides may have been overlooked. Our recent isolation of J-specific antibodies (unpublished results) now provides the tools to check whether J occurs elsewhere in nature.

ACKNOWLEDGEMENTS

We thank Janet Gommers-Ampt, Martin Taylor and Wilbert Bitter for helpful discussions, Francesca Fase-Fowler for advice with the CsCl gradients, Pat Blundell, Mike Cross, Anita Dirks-Mulder, Herlinde Gerrits, Henri van Luenen, Richard McCulloch, Rudo Kieft and Gloria Rudenko for critical reading of the manuscript and Titia de Lange (Rockefeller University, New York, NY) for many helpful comments. This work was supported by grants from the Netherlands Foundation for Chemical Research, with financial support of the Netherlands Organisation for Scientific Research (NWO).

REFERENCES

1 Van der Ploeg,L.H., Schwartz,D.C., Cantor,C.R. and Borst,P. (1984) *Cell*, **37**, 77–84.
2 Johnson,P.J. and Borst,P. (1986) *Gene*, **43**, 213–220.
3 Van der Ploeg,L.H., Smith,C.L., Polvere,R.I. and Gottesdiener,K.M. (1989) *Nucleic Acids Res.*, **17**, 3217–3227.
4 Zakian,V.A. (1995) *Science*, **270**, 1601–1613.
5 Blackburn,E.H. and Challoner,P.B. (1984) *Cell*, **36**, 447–457.
6 Van der Ploeg,L.H., Liu,A.Y. and Borst,P. (1984) *Cell*, **36**, 459–468.
7 Greider,C.W. (1994) *Curr. Opin. Genet. Dev.*, **4**, 203–211.
8 Weiden,M., Osheim,Y.N., Beyer,A.L. and Van der Ploeg,L.H. (1991) *Mol. Cell. Biol.*, **11**, 3823–3834.
9 de Lange,T., Shiue,L., Myers,R.M., Cox,D.R., Naylor,S.L., Killery,A.M. and Varmus,H.E. (1990) *Mol. Cell. Biol.*, **10**, 518–527.
10 Chan,C.S. and Tye,B.K. (1983) *Cell*, **33**, 563–573.
11 Bernards,A., Michels,P.A.M., Lincke,C.R. and Borst,P. (1983) *Nature*, **303**, 592–597.
12 Pays,E., Laurent,M., Delinte,K., Van Meirvenne,N. and Steinert,M. (1983) *Nucleic Acids Res.*, **11**, 8137–8147.
13 Bernards,A., van Harten-Loosbroek,N. and Borst,P. (1984) *Nucleic Acids Res.*, **12**, 4153–4170.
14 Borst,P., Gommers-Ampt,J.H., Ligtenberg,M.J., Rudenko,G., Kieft,R., Taylor,M.C., Blundell,P.A. and van Leeuwen,F. (1993) *Cold Spring Harbor Symp. Quant. Biol.*, **58**, 105–114.
15 Pays,E., Vanhamme,L. and Berberof,M. (1994) *Annu. Rev. Microbiol.*, **48**, 25–52.
16 Gottschling,D.E., Aparicio,O.M., Billington,B.L. and Zakian,V.A. (1990) *Cell*, **63**, 751–762.
17 Horn,D. and Cross,G.A.M. (1995) *Cell*, **83**, 555–561.
18 Rudenko,G., Blundell,P.A., Dirks-Mulder,A., Kieft,R. and Borst,P. (1995) *Cell*, **83**, 547–553.
19 Pays,E., Delauw,M.F., Laurent,M. and Steinert,M. (1984) *Nucleic Acids Res.*, **12**, 5235–5247.
20 Gommers-Ampt,J.H., van Leeuwen,F., de Beer,A.L., Vliegenthart,J.F., Dizdaroglu,M., Kowalak,J.A., Crain,P.F. and Borst,P. (1993) *Cell*, **75**, 1129–1136.
21 Gommers-Ampt,J., Lutgerink,J. and Borst,P. (1991) *Nucleic Acids Res.*, **19**, 1745–1751.
22 Huang,L.H., Farnet,C.M., Ehrlich,K.C. and Ehrlich,M. (1982) *Nucleic Acids Res.*, **10**, 1579–1591.
23 Cross,G.A. and Manning,J.C. (1973) *Parasitology*, **67**, 315–331.
24 Brun,R. and Schönenberger,M. (1979) *Acta Trop.*, **36**, 289–292.
25 Wijsman,E.R., van den Berg,O., Kuyl-Yeheskiely,E., van der Marel,G.A. and van Boom,J.H. (1994) *Rec. Trav. Chim. Pays-Bas*, **113**, 337–338.
26 Bernards,A., Van der Ploeg,L.H., Frasch,A.C., Borst,P., Boothroyd,J.C., Coleman,S. and Cross,G.A. (1981) *Cell*, **27**, 497–505.
27 Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
28 Sloof,P., Menke,H.H., Caspers,M.P. and Borst,P. (1983) *Nucleic Acids Res.*, **11**, 3889–3901.
29 Zomerdijk,J.C., Ouellette,M., ten Asbroek,A.L., Kieft,R., Bommer,A.M., Clayton,C.E. and Borst,P. (1990) *EMBO J.*, **9**, 2791–2801.
30 Thomashow,L.S., Milhausen,M., Rutter,W.J. and Agabian,N. (1983) *Cell*, **32**, 35–43.
31 White,T.C., Rudenko,G. and Borst,P. (1986) *Nucleic Acids Res.*, **14**, 9471–9489.
32 Brunel,F., Davison,J., Merchez,M., Borst,P. and Weijers,P.J. (1980) In Zadrazil,S. and Sponar,J. (eds), *DNA—Recombination Interactions and Repair*. Pergamon Press, Oxford, UK, pp. 45–54.
33 Reddy,M.V. (1993) In Phillips,D.H., Castegnaro,M. and Bartsch,H. (eds), *Postlabelling Methods for Detection of DNA Adducts*. International Agency for Research on Cancer, Lyon, France, pp. 25–34.
34 Burton,K. (1967) *Methods Enzymol.*, **12**, 222–228.
35 Erikson,R.L. and Szybalski,W. (1964) *Virology*, **22**, 111–124.
36 Szybalski,W. (1968) *Methods Enzymol.*, **12**, 330–360.
37 Wells,R.D. and Larson,J.E. (1972) *J. Biol. Chem.*, **247**, 3405–3409.
38 Gommers-Ampt,J.H., Teixeira,A.J., van de Werken,G., van Dijk,W.J. and Borst,P. (1993) *Nucleic Acids Res.*, **21**, 2039–2043.
39 Brown,W.R. (1989) *Nature*, **338**, 774–776.