Telomere maintenance and length regulation in *Trypanosoma brucei*

David Horn¹, Cheryl Spence and Alexandra K.Ingram

London School of Hygiene & Tropical Medicine, Keppel Street, London WC1E 7HT, UK

¹Corresponding author
e-mail: david.horn@lshtm.ac.uk

Transcription of telomere proximal variant surface glycoprotein genes is mono-allelic in bloodstream-form *Trypanosoma brucei*. The terminal DNA sequence at these telomeres consists of tandem T₂AG₃ repeats, which increase in length by ~8 bp per cell division balanced by occasional loss of large numbers of repeats. Here we have used targeted chromosome fragmentation to investigate the sequence requirements for telomere formation in *T.brucei*. Telomere formation is most efficient on tandem T₂AG₃ repeats, but can also occur on specific templates found within 'random' sequence substrates and on G-rich motifs proximal to a double-strand break. Newly formed telomeres are extended faster than other native telomeres, but as the telomere becomes longer the rate of extension declines. Telomere length regulation in *T.brucei* is discussed in the context of recent results from other cell types.

Keywords: telomerase/telomere/*Trypanosoma brucei*

Introduction

In eukaryotes chromosome ends form nucleoprotein complexes known as telomeres. These structures are vital to protect chromosomes from degradation or rearrangements (Sandell and Zakian, 1993) and to ensure complete chromosome replication. In most eukaryotes, including trypanosomes, telomeric DNA is composed of short G-rich tandem repeats with single-stranded 3' terminal extensions (see Chiurillo *et al.*, 1999). Despite their evolutionary divergence (Sogin *et al.*, 1989) telomeric 5'-TTAGGG-3' (T₂AG₃) repeats are identical in *Trypanosoma brucei* (Blackburn and Challoner, 1984; Van der Ploeg *et al.*, 1984) and their mammalian hosts. The 3' extensions are thought to provide a substrate for telomere extension (see Lingner and Cech, 1996) and to aid the formation of terminal DNA loops (Griffith *et al.*, 1999; Murti and Prescott, 1999). In trypanosomes, as in other organisms, telomeric chromatin has an unusual structure (see Freitas-Junior *et al.*, 1999) and some telomeres appear to be clustered at the nuclear periphery (Chung *et al.*, 1990). *Trypanosoma brucei* are tsetse-fly-transmitted protozoan parasites that cause disease in a variety of mammals, including humans and cattle. Telomeres are of special significance in *T.brucei* because the variant surface glycoprotein (VSG) genes, the central

players in antigenic variation, are expressed at these loci. VSG gene expression is mono-allelic, so although many VSGs are located at telomeres, only one is expressed by each cell (reviewed by Rudenko *et al.*, 1998).

During replication, DNA polymerases require priming to initiate unidirectional DNA synthesis. Therefore, an additional activity such as telomerase is required for complete replication of the ends of linear DNA. Telomerase contains a reverse transcriptase subunit and an RNA subunit (reviewed by Lingner and Cech, 1998), which serves as the template for the synthesis of additional G-rich repeats at the chromosome end. The C-rich strand can subsequently be synthesized by conventional DNA-dependent DNA polymerases. Thus, telomerase prevents progressive shortening of lagging strands during linear DNA replication in most eukaryotes. Alternative telomere maintenance pathways do exist, however (reviewed by Biessmann and Mason, 1997). In humans somatic cell proliferative potential is limited, in part, by loss of telomerase expression, while telomerase is constitutively expressed in germ line cells, many tumour cells and in unicellular eukaryotes such as *T.brucei* (Cano *et al.*, 1999).

Telomere length varies from <50 bp in the hypotrichous ciliates to >10 kbp in mammalian cells. While most human telomeres shorten during an individual's lifetime, human germ line cells and tumour cells and most unicellular eukaryotes maintain an average telomere length within relatively narrow boundaries (reviewed by Shore, 1997). In contrast, telomeres in *T.brucei*, including those downstream of a transcriptionally active VSG gene, increase in length by ~8 bp per generation, balanced by occasional loss of large numbers of repeats (Bernards *et al.*, 1983; Pays *et al.*, 1983), thus generating a broad length distribution in individual cells. More than 25 genes can influence telomere length in *Saccharomyces cerevisiae* (reviewed by Shore, 1997; Diede and Gottschling, 1999). To maintain telomere length homeostasis, telomerase-dependent telomere extension can be regulated by single-stranded telomere DNA-binding proteins from yeast and ciliates (Froelich Ammon *et al.*, 1998; Evans and Lundblad, 1999) or double-stranded telomere-repeat-binding proteins from yeast and human cells (Cooper *et al.*, 1997; Marcand *et al.*, 1997, 1999; van Steensel and de Lange, 1997; Smogorzewska *et al.*, 2000; reviewed by Shore, 1997). In the latter case, negative feedback operates, so that longer telomeres recruit more protein, which negatively regulates telomerase. Studies on human telomeres demonstrate that protein binding mediates t (telomere) loop formation (Griffith *et al.*, 1999). The formation of a t loop, in which the terminal overhang is sequestered in the duplex telomeric repeat array, is proposed to block telomere extension (Smogorzewska *et al.*, 2000). In addition, production of both strands of the telomere is tightly co-regulated in yeast, preventing the

formation of long single-stranded tails (Diede and Gottschling, 1999). Yeast, ciliate and human telomere-binding proteins have been studied in detail (reviewed by Brun *et al.*, 1997; Bourns *et al.*, 1998). Telomere-binding activities have been reported in *T.brucei* (Eid and Sollner Webb, 1995, 1997; Field and Field, 1996; Cross *et al.*, 1999; Berberof *et al.*, 2000), and trypanosome proteins that bind a telomere-enriched glucosylated base (J) that replaces thymine (see van Leeuwen *et al.*, 1998b) have been characterized in some detail (Cross *et al.*, 1999).

Based on the properties of *T.brucei* telomeres, in particular the possible link to *VSG* transcription regulation and the kinetics of length regulation, it is of interest to investigate telomere biology in more detail in this organism. We have characterized telomere proximal sequences and subsequently used targeted chromosome fragmentation to investigate *in vivo* telomere healing at the transcriptionally active chromosome end. Our results reveal the substrate requirements for telomere formation in *T.brucei*. We then measured telomere extension at these new telomeres. Newly formed telomeres are initially extended rapidly, but the extension rate declines as the telomere increases in length. We discuss telomere length regulation in *T.brucei* in the context of recent results from other cell types.

Results

Sub-telomere sequence organization

Sub-telomeric sequences in *T.brucei* are T rich on the strand extending 5' to 3' towards the chromosome end. In addition, they contain GC-rich elements with a highly conserved core, 11 bp of which is complementary but inverted relative to the T₂AG₃ repeats (Weiden *et al.*, 1991). Some chromosomes also have 29 bp repeats adjacent to the T₂AG₃ repeats (see Eid and Sollner-Webb, 1995). Before designing 'telomere healing' constructs, we determined whether these features were conserved among chromosomes. Using a polymerase chain reaction (PCR) strategy (see Materials and methods) we amplified and sequenced sub-telomeric fragments, including the T₂AG₃-non-T₂AG₃ junctions, from three distinct loci including the one we aimed to modify. The distance from the core of the T₂AG₃ repeat proximal GC-rich element to the T₂AG₃ repeats ranged from 155 to 481 bp. All three sequences were T rich, but lacked 29 bp repeats. The *VSG221* polyadenylation site was predicted by aligning the genomic sequence with a *VSG221* cDNA sequence. The 900 bp segment between the *VSG221* polyadenylation site and the T₂AG₃ repeats contains a T₂AG₃-like segment and a single GC-rich element (Figure 1A). This sequence has been submitted to the DDBJ/EMBL/GenBank databases under accession No. AJ271641.

Telomere formation is more efficient on substrates with terminal T₂AG₃ repeats

De novo telomere formation or chromosome healing (reviewed by Melek and Shippen, 1996) occurs on a large scale in organisms that undergo chromosome fragmentation, such as the hypotrichous ciliated protozoa. It can also occur in other cells following either spontaneous or artificially induced breakage. We chose to carry out

telomere manipulations in bloodstream-form *T.brucei* because telomeric *VSG* transcription occurs in this life cycle stage, thus facilitating selectable marker expression. In *T.brucei* a single *VSG*-associated telomere per nucleus is thought to be transcriptionally active. In MiTat 1.2 cells the active *VSG221*-associated telomere is stable and is inactivated or deleted in <1% of the cells each cell cycle (see Horn and Cross, 1997b). Since DNA predominantly integrates into the *T.brucei* genome by homologous recombination (Blundell *et al.*, 1996) and expression of the selectable marker is dependent upon insertion at a transcriptionally active locus, targeting *VSG221* allowed us to manipulate a single well characterized chromosome end reproducibly.

To determine whether telomere formation is influenced by the presence of terminal T₂AG₃ repeats we made a construct (p221:Neo:Hex; Figure 1A) designed to target the active *VSG221* locus, leaving the sub-telomeric GC-rich element intact, but eliminating the T₂AG₃ repeats and 79 bp of non-T₂AG₃ sequence. Southern blotting suggested that ~2000 T₂AG₃ repeats (12 kbp) were present at the *VSG221* locus in the majority of our cells (data not shown). The construct was designed to replace this DNA with a *NEO* cassette, a plasmid vector and a telomere seed (T₂AG₃)₃₃ 22 bp from the end following *SmaI* digestion. Transcription is polycistronic in *T.brucei*, so the same endogenous promoter should drive expression of the *VSG* and *NEO* genes (see Figure 1A). Digestion of the same construct with *EcoRI* allowed the T₂AG₃ repeats to be removed (Figure 1A), allowing a direct comparison between telomere healing in transformed cells in the presence and absence of T₂AG₃ repeats.

p221:Neo:Hex was digested with *SmaI* or *EcoRI* and introduced into *T.brucei* cells by electroporation in parallel experiments. Overall, we obtained 58 drug-resistant cell lines using DNA with T₂AG₃ repeats and one cell line using DNA lacking these repeats (Figure 1B). These results demonstrate that transfection efficiency is significantly greater when one end of the construct contains telomeric repeats. Sequencing the T₂AG₃-non-T₂AG₃ junction from the single clone generated using 'DNA lacking repeats' revealed that this clone was derived from a residual undigested or religated construct. From experiments using p221:Neo:Hex, therefore, we obtained no cell lines in which telomere healing occurred on sequence lacking repeats.

An increase in length during cell growth is a typical feature of *T.brucei* chromosome ends (Bernards *et al.*, 1983; Pays *et al.*, 1983), so to check that *NEO* genes had inserted proximal to chromosome ends, DNA was extracted for Southern blotting at different times following electroporation. Our results indicated that *NEO* was present on a terminal restriction fragment (TRF) that increased in length over time in four independent clones. Results for one of these clones is shown in Figure 1C. Identical results were obtained when this blot was rehybridized with a *VSG221* probe (data not shown), indicating that *NEO* had integrated at the *VSG221* locus as expected. All TRF fragments were longer than expected in this experiment, however, consistent with rapid telomere extension at newly formed telomeres (see below).

To determine the location of the telomere healing events we selected two clones and sequenced the *NEO*

Fig. 1. Telomere healing using DNA with or without T₂AG₃ repeats. (A) Maps of the *VSG221* locus and the targeting construct p221:Neo:Hex. The upper map represents the *VSG221* locus in MiTat 1.2 cells. Salient features downstream of the *VSG221* pA site are indicated. Homologous recombination within sequences between the vertical lines targeted p221:Neo:Hex to the *VSG221* locus. pA, predicted mRNA polyadenylation sites; E, *EcoRI*; M, *MseI* (only two relevant sites indicated); Sm, *SmaI*; X, *XhoI*. Probes used for Southern blotting (NEO and GEM) are labelled below the map. (B) Efficiency of DNA transformation. p221:Neo:Hex was digested with *SmaI* (+ T₂AG₃) or *EcoRI* (– T₂AG₃). Following electroporation each culture was distributed in 12-well plates in medium containing 2 µg/ml G418. Each well containing live cells after 1 week was scored as positive. The results represent the average of 11 experiments. Error bars represent standard deviation. The asterisk indicates that the single clone generated using DNA digested with *EcoRI* may have retained T₂AG₃ repeats (see the text). (C) Southern blot. DNA was extracted from a cell line generated using DNA with T₂AG₃ repeats 8 and 20 days following DNA transformation. The DNA was digested with *XhoI* and the blot was hybridized with the NEO probe. (D) Southern blot. One microgram of DNA from three cell lines (1, 2 and 3) generated using DNA with T₂AG₃ repeats and 1 ng of p221:Neo:Hex (P) were digested with *MseI* and the blot was hybridized with the GEM probe. Rehybridization with another *MseI* fragment confirmed complete digestion of the genomic DNA samples (data not shown).

proximal T₂AG₃–non-T₂AG₃ junctions (see Materials and methods). Neither junction was altered relative to our plasmid construct, indicating that healing had occurred within the T₂AG₃ repeats or the distal 22 bp sequence. To distinguish between these possibilities we digested genomic DNA from these two clones and an additional clone with *MseI*. This restriction enzyme cuts immediately at the *NEO* distal T₂AG₃–non-T₂AG₃ junction in p221:Neo:Hex (see Figure 1A). The last A in the TTAA site represents the first base to deviate from the T₂AG₃ sequence. If healing was within the 22 bp segment then genomic DNA will generate a 339 bp band when hybridized with the GEM probe. If healing was within the T₂AG₃ repeats then the 22 bp segment including a *MseI* site will be lost, thus generating a large TRF. Healing was within the T₂AG₃ repeats in all three clones tested (Figure 1D).

'Random' sequence can seed telomere formation

We next tested whether DNA lacking T₂AG₃ repeats could act as a template for telomere formation in *T.brucei*. We expected healing events on 'random' sequence to be rare relative to healing on T₂AG₃ repeats (Figure 1B). To ensure that detection of telomere healing on 'random' sequence was not dependent upon complete removal of T₂AG₃ repeats from the construct, as with p221:Neo:Hex, we used a second construct lacking T₂AG₃ repeats for these assays (p221:Neo; Figure 2A).

First, we digested p221:Neo with three different restriction enzymes, each of which generated different termini. *EcoRI* digestion generated a single-stranded TTAA-5' extension, *SacII*, a single-stranded GC-3' extension, and *SmaI*, a blunt end. Twenty-five independent DNA transformations were carried out with each digest, yielding three, six and seven cell lines, respectively, in which the construct was capped by the addition of T₂AG₃ repeats, as determined by PCR (see Materials and methods). This represents ~4% of the number of transformed cell lines we would expect in similar experiments using DNA with repeats (see Figure 1B). Sequencing the PCR products revealed the precise location of T₂AG₃ repeat addition. Digestion with *EcoRI*, *SacII* or *SmaI* appeared to make no significant difference to the results, so the entire dataset is presented in one figure (Figure 2B).

The locations of the healing events are indicated on a map extending downstream of the predicted *NEO* polyadenylation site. Thirteen out of 16 healing events occurred at an 8 bp TTAGGGTT motif (Figure 2B). This is the longest stretch of bases in pBluescript that conforms to the telomere repeat sequence. Although three 7 bp sequences that conform to the repeat sequence are present in the region indicated, only the 7 bp TTAGGGT motif close to the 8 bp motif primes telomere formation in a single cell line (Figure 2B). Given that two or three consecutive G bases occur >100 and >40 times, respectively, in pBluescript, the location of the other two healing

Fig. 2. Telomere healing within 'random' sequence. (A) Maps of the *VSG221* locus and the targeting construct p221:Neo. The maps are adapted from Figure 1A. Ac, *Acc65I*; Ap, *Apal*; B, *Bsp120I*; E, *EcoRI*; K, *KpnI*; S, *Sall*; Sc, *SacII*; Sm, *SmaI*. The BS probe is labelled below the map. (B) Map of the p221:Neo sequence downstream of the *NEO* polyadenylation site [~300 bp followed by pBluescript; see Figure 2A] indicating the sites and frequency of T₂AG₃ repeat addition. p221:Neo was digested with *SmaI*, *SacII* or *EcoRI* and 16 telomere healing events were obtained from a total of 75 DNA transformation experiments. The sequence that conforms to the telomeric sequence at the healing site is indicated in upper case above each bar. (C) Southern blots. DNA extracted 20 days following DNA transformation from one 'TTAGGGTT' and the single 'GG' cell line was digested with *Sall*. The blot was sequentially hybridized with the probes indicated.

events appears to be significant. One (GG) is only 8 bp from a double-strand break generated by *SmaI* digestion and the other (GGG) is close to the 7 and 8 bp motifs mentioned above. In this experiment, however, the TTAGGGTT motif effectively competes with the double-strand break as a substrate for telomere healing (but see below).

As an additional test of our methodology, two samples were selected for Southern blotting (Figure 2C). Based on the location of the healing events, the TTAGGGTT-*NEO* TRF was expected to be 2.5 kbp shorter than the GG-*NEO* TRF (see Figure 2B). In addition, the BS DNA fragment should not be present on the TTAGGGTT-*NEO* TRF. Southern blotting confirmed both of these predictions (Figure 2C). Once again TRF length suggested rapid telomere extension at newly formed telomeres (see below).

G-rich motifs proximal to double-strand breaks can seed telomere formation

Tetrahymena telomerase can add telomeric repeats to primers lacking such repeats providing a tract of G bases is present somewhere in the primer (Harrington and Greider, 1991). To determine whether a double-strand break with a proximal tract of G bases can serve as an effective substrate for telomere formation in *T.brucei* we digested p221:Neo with *EcoRI* plus *KpnI* and *EcoRI* plus *Apal*. Each digest generates a fragment containing the *VSG221* target and the *NEO* gene, but lacking pBluescript sequence (see Figure 2A), thus eliminating the sequences that served as templates for telomere healing in the previous experiment. These digests generate double-strand breaks with a proximal stretch of seven G bases (Figure 3A and B). Twenty-five independent DNA transformations with each

Fig. 3. Telomere healing proximal to double-strand breaks. p221:Neo (Figure 2A) was digested with *EcoRI* plus (A) *KpnI*, (B) *Apal*, (C) *Acc65I* or (D) *Bsp120I* to generate double-strand breaks. The numbers in parentheses represent the number of cell lines derived from 25 DNA transformations. The location (arrows) and number of healing events observed at each location are indicated above each sequence.

digest yielded 10 and 14 cell lines, respectively. This represents ~6% of the number of transformed cell lines we would expect in similar experiments using DNA with repeats (see Figure 1B). The location of T₂AG₃ repeat addition was mapped in a selection of these cell lines, indicating that double-strand breaks proximal to a G-rich element can seed telomere formation (Figure 3A and B).

To determine whether healing efficiency differs on recessed or overhanging templates we digested p221:Neo with *EcoRI* and an isoschizomer of *KpnI* (*Acc65I*). Twenty-five independent DNA transformations yielded 10 cell lines for which the location of T₂AG₃ repeat addition is shown (Figure 3C). These results suggest that healing efficiency is not significantly different on recessed or overhanging templates (Figure 3A and C). We next determined whether a G-rich element immediately at the end of the DNA (Figure 3D) could serve as a more efficient

template for telomere healing. p221:Neo was digested with *EcoRI* and an isoschizomer of *Apal* (*Bsp120I*). Twenty-five independent DNA transformations yielded 20 cell lines, suggesting that healing is not dramatically (no more than 2-fold) more efficient if the G-rich element is immediately at the end of the DNA (Figure 3B and D). The location of T₂AG₃ repeat addition was mapped in a selection of these cell lines (Figure 3D).

The kinetics of telomere extension are altered at newly formed telomeres

Telomeres seeded on T₂AG₃ repeats or on random sequences appeared to be extended faster than expected (see above). To study this phenomenon in more detail we extracted DNA and measured TRF length in four clones, 8, 20 and 50 days after electroporation. All four clones generated similar results, so only the results for one (*EcoRI*-*KpnI* digest; see Figure 3A) are shown (Figure 4A). Rehybridization of the same blot with a NEO probe gave identical results to the VSG221 probe, indicating that the construct integrated at the *VSG221* locus in all four clones, as expected. Bands that do not change in size in the VSGB and VSGbR2 panels are thought to represent chromosome internal 'basic copy' VSG genes. Bands that increase in size are TRFs, so *VSGB* has one TRF and *VSGbR2* has two TRFs. TRF length increases significantly faster at newly formed telomeres than at other telomeres in the same cells (Figure 4A).

We next measured TRF lengths after 200 days in three independent clones to see whether extension continued at an increased rate. All three clones generated similar results, so only the results for one are shown (Figure 4B). Telomere length appears to reach a state of equilibrium after extending by ~15 kbp. Long telomeres downstream of a transcriptionally active VSG appear to be prone to loss of large numbers of repeats (Bernards *et al.*, 1983; Pays *et al.*, 1983), so as expected the *VSG221*-associated telomere generated significant heterogeneity in TRF length after 200 days (Figure 4B). PhosphorImager analysis revealed *VSG221*-associated TRFs ranging from ~10 to ~20 kbp (data not shown). Subclones of this population show that cells with discreet TRF lengths exist within this population (Figure 4B, lanes 1 and 2). Rehybridization with a *VSGB* probe shows that the DNA extracted from the 200 day population is intact and that the smear does reflect many different lengths of *VSG221*-associated TRFs in the 200 day culture. This panel also indicates that the *VSGB* TRF has extended to the same length as another *VSGB* fragment (Figure 4B). This interpretation was confirmed by quantitating the bands by PhosphorImager analysis and by the reappearance of a shortened *VSGB* TRF in a 200 day subclone (Figure 4B, lane 2). The *VSGbR2* TRFs also increased in length in the 200 day culture and spontaneous length reduction can be detected in a subcloned population (Figure 4B).

Telomeres in *T.brucei* increase in length by ~8 bp per generation (Bernards *et al.*, 1983; Pays *et al.*, 1983). Based on a cell doubling time of 7 h, which was constant throughout the 200 day experiment, 686 cell divisions should result in an increase in TRF length of nearly 5.5 kbp. This is consistent with results for *VSGB* and *VSGbR2* TRFs (Figure 4). From the Southern blots we estimated maximum TRF length at the *VSG221*-associated

Fig. 4. Telomere extension in growing cells. DNA was digested with *EcoRI* and the Southern blots were sequentially hybridized with probes for different telomere proximal VSGs. An asterisk indicates a non-terminal restriction fragment (non-TRF) band. These bands stay constant over time. (A) Southern blots. DNA extracted from the same clone 8, 20 and 50 days following DNA transformation. (B) Southern blots. High-molecular-weight DNA was prepared from cells grown for 200 days following DNA transformation and from two subclones of the 200 day population (1 and 2). M, Mono Cut Mix molecular weight markers (New England Biolabs). (C) TRF extension was predicted for 'normal' telomeres (dashed line) and compared with that observed at the new *VSG221*-associated telomeres (solid line). Data were derived from four independent cell lines. (D) Extension rate according to telomere length. The extension rate for 'normal' telomeres (dashed line) was compared with that observed at the new *VSG221*-associated telomeres (solid line). We estimated the elongation rate from (C) by dividing the difference in telomere length between two experimental points by the corresponding number of generations.

telomere. For clones in which healing had occurred at the *KpnI* site we deducted 4.5 kbp from the *VSG221*-TRF generated by *EcoRI* digestion, thus giving an estimate of TRF extension (see Figure 2A). This was compared with TRF extension at other telomeres (Figure 4C). The rate of extension at the *VSG221*-associated telomere declined and appeared to reach equilibrium after TRFs were extended by ~15 kbp (Figure 4C). From Figure 4C we predicted the rate of telomere extension at different time points (Figure 4D). The new *VSG221*-associated telomeres appeared to be extended at an initial rate >40-fold greater than that seen at other telomeres and extension rate appears to be inversely related to telomere length according to a non-linear function (Figure 4D).

Discussion

Using targeted chromosome fragmentation, we have investigated telomere maintenance and length regulation in *T.brucei*. We demonstrate that sequences conforming more closely to the wild-type telomere sequence serve more efficiently as substrates for telomere formation in *T.brucei*. Free DNA ends were stabilized exclusively by the addition of T₂AG₃ repeats and we detected no evidence of alternative mechanisms. Newly formed telomeres are extended rapidly, but the extension rate declines as telomere length increases.

In vivo telomere healing has been reported in ciliated protozoa (see Wang and Blackburn, 1997), *Plasmodium falciparum* (Bottius *et al.*, 1998), yeast (see Teng and Zakian, 1999) and mammalian cells (see Hanish *et al.*, 1994). Although *Tetrahymena* telomerase can add telomeric repeats to primers lacking such repeats (Harrington and Greider, 1991) and telomere healing can occur in mouse embryonic stem cells at a non-telomeric double-strand break (Sprung *et al.*, 1999), quantitative, *in vivo* assays in human HeLa cells reveal that telomere healing is more efficient when seeded on sequences conforming to the natural telomeric sequence (Hanish *et al.*, 1994). Since human telomerase appears to have relatively relaxed DNA sequence requirements *in vitro*, telomere-binding proteins are thought to increase the specificity *in vivo* (Hanish *et al.*, 1994).

In *T.brucei*, 33 T₂AG₃ repeats represent an effective substrate for the establishment of a functional telomere. Our results indicate that healing occurs within the repeats, so we are unable to distinguish between recombination with native repeats or recruitment of telomerase in these experiments. DNA lacking terminal T₂AG₃ repeats can also seed telomere formation, consistent with the action of telomerase, but we cannot rule out recombination even in these experiments. In *T.brucei*, telomeres can be formed on a variety of substrates incorporated downstream of a transcriptionally active *VSG* gene. Within >3 kbp a single TTAGGGTT motif acts as the predominant template, while G-rich elements serve as effective substrates if they are located proximal to a double-strand break. The DNA substrate appears to be rate limiting for telomere formation and in every case examined it directed addition of T₂AG₃ repeats initiated in phase with the DNA sequence in such a way that the substrate became incorporated into the first T₂AG₃ repeat. In yeast (Kramer and Haber, 1993) and human cells (Barnett *et al.*, 1993) telomeres can be formed on non-terminal telomere seeds with retention of a distal segment. In another study in human cells, distal non-telomeric DNA was consistently lost (Hanish *et al.*, 1994), similar to our results. In summary, *T.brucei* telomere formation has relatively stringent sequence requirements similar to the situation in human cells (see Hanish *et al.*, 1994), where the telomere sequence is identical.

Various components of telomeric chromatin influence telomerase activity to maintain telomere length equilibrium. For example, changing the telomeric sequence (McEachern and Blackburn, 1995) or tethering telomerase to the telomere (Evans and Lundblad, 1999) can overcome negative regulation of telomere length in yeast cells. Single-stranded telomere DNA-binding proteins may regulate access of telomerase to chromosome termini on

the relatively short telomeres found in yeast and ciliates (Froelich Ammon *et al.*, 1998; Evans and Lundblad, 1999). Double-stranded telomere-repeat-binding proteins can also negatively regulate telomere extension in yeast and human cells (Marcand *et al.*, 1997; Smogorzewska *et al.*, 2000). In these cells negative feedback appears to regulate telomere length (Cooper *et al.*, 1997; Marcand *et al.*, 1997, 1999; van Steensel and de Lange, 1997; reviewed by Shore, 1997) and in the case of human cells inhibition of telomere extension is consistent with the t loop model of telomere length homeostasis (Smogorzewska *et al.*, 2000). In this model, elongation of telomeres by telomerase is blocked by sequestration of the telomere terminus in t loops induced by TRF1 and TRF2. The t loops from human cells and the polytene chromosomes of ciliated protozoa range in size from <1 to >25 kbp (Griffith *et al.*, 1999; Murti and Prescott, 1999).

Telomeres in *T.brucei* increase in length by ~8 bp per generation (Bernards *et al.*, 1983; Pays *et al.*, 1983). Here we demonstrate that newly formed telomeres in *T.brucei* are extended faster than normal telomeres. When short, these telomeres appear to grow >40-fold faster than other telomeres, but the extension rate declines and eventually reaches equilibrium when a length of ~15 kbp is attained, consistent with a negative feedback mechanism based on counting T₂AG₃ repeats. Following similar experiments in yeast, Marcand *et al.* (1999) hypothesized that the telomere extension rate decreased with increasing telomere length according to a linear function. In *T.brucei* the relationship between telomere length and telomere extension does not appear to be linear, but is more consistent with the t loop model. Structural constraints or the lack of sufficient bound protein may reduce the chance of forming a t loop when the telomere is short, allowing rapid telomere extension to occur. A progressive block to telomere extension may be achieved when these telomeres reach a length sufficient to allow a t loop to form. By analogy to the situation in human cells we propose a double-stranded telomere DNA-binding protein in *T.brucei* analogous to TRF1, TRF2 or both (van Steensel and de Lange, 1997; Smogorzewska *et al.*, 2000). The J-binding protein (Cross *et al.*, 1999) that binds the modified DNA base known as J (van Leeuwen *et al.*, 1998b) in trypanosomes could have such a role in bloodstream-form *T.brucei*. We considered the possibility that lack of the J base at new, rapidly extended telomeres may explain the rapid extension phenotype. However, telomeres on artificial mini-chromosomes are extended at the usual slow rate when introduced into insect-stage *T.brucei* cells (Patnaik *et al.*, 1996), which lack the J base (van Leeuwen *et al.*, 1998a).

At native telomeres, including those downstream of a transcriptionally active *VSG* and those <4 kbp in length, telomere extension is limited in such a way that only one or two repeats are added per generation (Bernards *et al.*, 1983; Pays *et al.*, 1983; this paper). This limited rate of telomere extension may be mediated via stable t loops. It is interesting in this regard that the conserved core of the sub-telomeric GC-rich elements (Weiden *et al.*, 1991) is complementary to telomere repeats. It may be that GC-rich elements interact with the chromosome end to form t loops in *T.brucei*. A similar interaction between the chromosome terminus and the telomere-non-telomere junction

has been proposed in yeast (Ray and Runge, 1999). As outlined above, there is a clear difference between extension at native and at newly formed telomeres. To explain this difference we propose that negative regulation, mediated by stable t loops, is lost or disrupted at newly formed telomeres. The introduction of a 'spacer' sequence (>2 kbp) between the sub-telomeric GC-rich elements and the telomere repeats, as is the case with our constructs, could disrupt the t loops we propose above. Transcription of a telomere in yeast caused a slight reduction in telomere length (Sandell *et al.*, 1994), so transcription proceeding to the chromosome end at the *VSG221* locus could also influence telomere extension.

Expression of telomeric *VSG* genes is mono-allelic in *T.brucei*. The ability to manipulate *T.brucei* telomeres provides us with an opportunity to test the contribution of various aspects of telomere structure to *VSG* gene expression. Indeed, several yeast proteins influence telomere length regulation and telomeric gene silencing (see Runge and Zakian, 1996). Telomere extension is remarkably well co-ordinated within a population of *T.brucei* cells. This should facilitate the determination of any effect that different length telomeres may have upon transcription or the formation of t loops, for example.

Materials and methods

Cells

Growth and DNA transformation of bloodstream-form *T.brucei* MiTat 1.2 clone 221a were carried out as described previously (Horn and Cross, 1997b), but with the following modifications. Cells were mixed with 5 µg of DNA and pulsed once at 1.4 kV and 25 µF using a Gene pulser II (Bio-Rad). G418 at a concentration of 2 µg/ml (MBI-Fermentas) was added to the medium 6 h after electroporation. If the efficiency of DNA transformation was found to be high, then electroporated cultures were distributed in multi-well plates in future experiments. Cell lines were not characterized if >40% of the flasks/wells contained drug-resistant cells. Therefore, most of the data reported here are derived from clonal cell lines. G418-resistant cultures were subsequently maintained in medium containing 2 µg/ml G418.

Plasmid constructions

To generate p221:Neo an *EcoRI* fragment containing a *NEO* cassette from pbRn (Horn and Cross, 1995) was inserted at an *XhoI* site in pNeg, a genomic clone containing the *VSG221* gene (Horn and Cross, 1997b). In order to synthesize a 'telomere seed' we used two primers in a PCR protocol. The primers were Tel1, (AGGGTT)₅ and Tel2, (ACCCTA)₅. Twenty PCR cycles at 94, 45 and 72°C, each for 15 s, were carried out in the presence of *Taq* DNA polymerase (MBI-Fermentas) and primers (at 4 µM) according to the manufacturer's instructions. This procedure generated an abundant ~200 bp product that was ligated to a T-A vector (Promega). Following sequencing we chose a clone designated pHex1 that contained 33 repeats, (T₂AG₃)₃₃. The *VSG221* targeting sequence and *NEO* cassette were removed from p221:Neo by digestion with *KpnI* and *SpeI*, and inserted into pHex1 digested with *SalI* and *SpeI* to generate p221:Neo:Hex.

PCR and sequencing

Sub-telomeric sequences from wild-type cells were amplified by PCR in the presence of *Taq* DNA polymerase. Primers were complementary to the telomere repeats (ACCCTA)₅ and either *VSG221* (ggcgttaccagcttggga) or the core of the GC-rich element (GCGGACCCTAACCCCTCCTC). Temperature was cycled 30 times through 94°C for 30 s, 58°C for 30 s and 72°C for 90 s or 30 s for the *VSG221* and GC primers, respectively. One hundred nanograms of *T.brucei* genomic DNA were included in each reaction. PCR products were cloned into T-A vectors prior to sequencing.

DNA was extracted from recombinant cells as soon as possible, usually 7 days (~25 cell generations) after electroporation. Sub-telomeric sequences were amplified from these samples essentially as described above, and sequenced using the Tel2 primer and a variety of primers

complementary to the plasmid constructs. To minimize rearrangements or errors that accumulate during PCR we sequenced PCR products directly. No PCR products were generated when the p221:Neo construct or DNA from non-transformed cells served as template, so the generation of a PCR product in our assay indicated that the construct had been modified by T₂AG₃ repeats. Six out of 46 G418-resistant cell lines generated using p221:Neo did not produce a PCR product in this assay, however. All six appeared to have the *NEO* gene integrated by homologous recombination with flanking mRNA processing signals at the *ALD* locus as determined in another PCR assay (data not shown). The *NEO* gene was clearly not linked to *VSG221* in at least two of these cell lines, as determined by Southern blotting (data not shown), so these six cell lines were not studied further.

DNA isolation and Southern blotting

DNA was isolated using DNA-Stat-60 (Tel Test Incorporated) or as described previously (Horn and Cross, 1997a) for high-molecular-weight DNA. Southern blotting was carried out as described previously (Horn and Cross, 1995). The *NEO* probe was the entire protein coding region, digesting pBluescript with *PvuII*-*SacI* generated the BS probe and digesting pGem5 with *SacII*-*MseI* generated the GEM probe. The *VSG* probes were as described previously (Horn and Cross, 1997a). Signals on Southern blots were quantitated by PhosphorImager (Molecular Dynamics) analysis.

Acknowledgements

We would like to thank John Kelly, Martin Taylor, Angeles Mondragon and Carolina Mailhos for valuable advice and for critical reading of the manuscript, and Ian Yeung for technical assistance. This work was supported by a Research Career Development Fellowship (052323) to D.H. from The Wellcome Trust.

References

- Barnett, M.A., Buckle, V.J., Evans, E.P., Porter, A.C., Rout, D., Smith, A.G. and Brown, W.R. (1993) Telomere directed fragmentation of mammalian chromosomes. *Nucleic Acids Res.*, **21**, 27–36.
- Berberof, M., Vanhamme, L., Alexandre, S., Lips, S., Tebabi, P. and Pays, E. (2000) A single-stranded DNA-binding protein shared by telomeric repeats, the variant surface glycoprotein transcription promoter and the procyclin transcription terminator of *Trypanosoma brucei*. *Nucleic Acids Res.*, **28**, 597–604.
- Bernards, A., Michels, P.A.M., Lincke, C.R. and Borst, P. (1983) Growth of chromosome ends in multiplying trypanosomes. *Nature*, **303**, 592–597.
- Biessmann, H. and Mason, J.M. (1997) Telomere maintenance without telomerase. *Chromosoma*, **106**, 63–69.
- Blackburn, E.H. and Challoner, P.B. (1984) Identification of a telomeric DNA sequence in *Trypanosoma brucei*. *Cell*, **36**, 447–457.
- Blundell, P.A., Rudenko, G. and Borst, P. (1996) Targeting of exogenous DNA into *Trypanosoma brucei* requires a high degree of homology between donor and target DNA. *Mol. Biochem. Parasitol.*, **76**, 215–229.
- Bottius, E., Bakhsis, N. and Scherf, A. (1998) *Plasmodium falciparum* telomerase: *de novo* telomere addition to telomeric and nontelomeric sequences and role in chromosome healing. *Mol. Cell. Biol.*, **18**, 919–925.
- Bourne, B.D., Alexander, M.K., Smith, A.M. and Zakian, V.A. (1998) Sir proteins, Rif proteins, and Cdc13p bind *Saccharomyces* telomeres *in vivo*. *Mol. Cell. Biol.*, **18**, 5600–5608.
- Brun, C., Marcand, S. and Gilson, E. (1997) Proteins that bind to double-stranded regions of telomeric DNA. *Trends Cell Biol.*, **7**, 317–324.
- Cano, M.I., Dungan, J.M., Agabian, N. and Blackburn, E.H. (1999) Telomerase in kinetoplastid parasitic protozoa. *Proc. Natl Acad. Sci. USA*, **96**, 3616–3621.
- Chiurillo, M.A., Cano, I., Da Silveira, J.F. and Ramirez, J.L. (1999) Organization of telomeric and sub-telomeric regions of chromosomes from the protozoan parasite *Trypanosoma cruzi*. *Mol. Biochem. Parasitol.*, **100**, 173–183.
- Chung, H.-M.M., Shea, C., Fields, S., Taub, R.N., Van der Ploeg, L.H.T. and Tse, D.B. (1990) Architectural organization in the interphase nucleus of the protozoan *Trypanosoma brucei*: Location of telomeres and mini-chromosomes. *EMBO J.*, **9**, 2611–2619.
- Cooper, J.P., Nimmo, E.R., Allshire, R.C. and Cech, T.R. (1997)

- Regulation of telomere length and function by a Myb-domain protein in fission yeast. *Nature*, **385**, 744–747.
- Cross,M., Kieft,R., Sabatini,R., Wilm,M., de Kort,M., van der Marel,G., van Boom,J., van Leeuwen,F. and Borst,P. (1999) The modified base J is the target for a novel DNA-binding protein in kinetoplastid protozoans. *EMBO J.*, **18**, 6573–6581.
- Diede,S.J. and Gottschling,E. (1999) Telomerase-mediated telomere addition *in vivo* requires DNA primase and DNA polymerases α and δ . *Cell*, **99**, 723–733.
- Eid,J.E. and Sollner-Webb,B. (1995) ST-1, a 39-kilodalton protein in *Trypanosoma brucei*, exhibits a dual affinity for the duplex form of the 29-base-pair subtelomeric repeat and its C-rich strand. *Mol. Cell. Biol.*, **15**, 389–397.
- Eid,J.E. and Sollner Webb,B. (1997) ST-2, a telomere and subtelomere duplex and G-strand binding protein activity in *Trypanosoma brucei*. *J. Biol. Chem.*, **272**, 14927–14936.
- Evans,S. and Lundblad,V. (1999) Est1 and cdc13 as comediators of telomerase access. *Science*, **286**, 117–120.
- Field,H. and Field,M.C. (1996) *Leptomonas seymouri*, *Trypanosoma brucei*: A method for isolating trypanosomatid nuclear factors which bind *T. brucei* single-stranded g-rich telomere sequence. *Exp. Parasitol.*, **83**, 155–158.
- Freitas-Junior,L.H.G., Porto,R.M., Pirrit,L.A., Schenkman,S. and Scherf,A. (1999) Identification of the telomere in *Trypanosoma cruzi* reveals highly heterogeneous telomere lengths in different parasite strains. *Nucleic Acids Res.*, **27**, 2451–2456.
- Froelich Ammon,S.J., Dickinson,B.A., Bevilacqua,J.M., Schultz,S.C. and Cech,T.R. (1998) Modulation of telomerase activity by telomere DNA-binding proteins in *Oxytricha*. *Genes Dev.*, **12**, 1504–1514.
- Griffith,J.D., Comeau,L., Rosenfield,S., Stansel,R.M., Bianchi,A., Moss,H. and de Lange,T. (1999) Mammalian telomeres end in a large duplex loop. *Cell*, **97**, 503–514.
- Hanish,J.P., Yanowitz,J.L. and de Lange,T. (1994) Stringent sequence requirements for the formation of human telomeres. *Proc. Natl Acad. Sci. USA*, **91**, 8861–8865.
- Harrington,L.A. and Greider,C.W. (1991) Telomerase primer specificity and chromosome healing. *Nature*, **353**, 451–454.
- Horn,D. and Cross,G.A.M. (1995) A developmentally regulated position effect at a telomeric locus in *Trypanosoma brucei*. *Cell*, **83**, 555–561.
- Horn,D. and Cross,G.A.M. (1997a) Analysis of *Trypanosoma brucei* VSG expression site switching *in vitro*. *Mol. Biochem. Parasitol.*, **84**, 189–201.
- Horn,D. and Cross,G.A.M. (1997b) Position-dependent and promoter-specific regulation of gene expression in *Trypanosoma brucei*. *EMBO J.*, **16**, 7422–7431.
- Kramer,K.M. and Haber,J.E. (1993) New telomeres in yeast are initiated with a highly selected subset of TG1-3 repeats. *Genes Dev.*, **7**, 2345–2356.
- Lingner,J. and Cech,T.R. (1996) Purification of telomerase from *Euplotes aediculatus*: requirement of a primer 3' overhang. *Proc. Natl Acad. Sci. USA*, **93**, 10712–10717.
- Lingner,J. and Cech,T.R. (1998) Telomerase and chromosome end maintenance. *Curr. Opin. Genet. Dev.*, **8**, 226–232.
- Marcand,S., Brevet,V. and Gilson,E. (1999) Progressive *cis*-inhibition of telomerase upon telomere elongation. *EMBO J.*, **18**, 3509–3519.
- McEachern,M.J. and Blackburn,E.H. (1995) Runaway telomere elongation caused by telomerase RNA gene mutations. *Nature*, **376**, 403–409.
- Marcand,S., Gilson,E. and Shore,D. (1997) A protein-counting mechanism for telomere length regulation in yeast. *Science*, **275**, 986–990.
- Melek,M. and Shippen,D.E. (1996) Chromosome healing: spontaneous and programmed *de novo* telomere formation by telomerase. *BioEssays*, **18**, 301–308.
- Murti,K.G. and Prescott,D.M. (1999) Telomeres of polytene chromosomes in a ciliated protozoan terminate in duplex DNA loops. *Proc. Natl Acad. Sci. USA*, **96**, 14436–14439.
- Patnaik,P.K., Axelrod,N., Van der Ploeg,L.H.T. and Cross,G.A.M. (1996) Artificial linear mini-chromosomes for *Trypanosoma brucei*. *Nucleic Acids Res.*, **24**, 668–675.
- Pays,E., Laurent,M., Delinte,K., van Meirvenne,N. and Steinert,M. (1983) Differential size variations between transcriptionally active and inactive telomeres of *Trypanosoma brucei*. *Nucleic Acids Res.*, **11**, 8137–8147.
- Ray,A. and Runge,K.W. (1999) The yeast telomere length counting machinery is sensitive to sequences at the telomere–nontelomere junction. *Mol. Cell. Biol.*, **19**, 31–45.
- Rudenko,G., Cross,M. and Borst,P. (1998) Changing the end: antigenic variation orchestrated at the telomeres of African trypanosomes. *Trends Microbiol.*, **6**, 113–116.
- Runge,K.W. and Zakian,V.A. (1996) TEL2, an essential gene required for telomere length regulation and telomere position effect in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **16**, 3094–3105.
- Sandell,L.L. and Zakian,V.A. (1993) Loss of a yeast telomere: Arrest, recovery and chromosome loss. *Cell*, **75**, 729–739.
- Sandell,L.L., Gottschling,D.E. and Zakian,V.A. (1994) Transcription of a yeast telomere alleviates telomere position effect without affecting chromosome stability. *Proc. Natl Acad. Sci. USA*, **91**, 12061–12065.
- Shore,D. (1997) Telomere length regulation: getting the measure of chromosome ends. *Biol. Chem.*, **378**, 591–597.
- Smogorzewska,A., van Steensel,B., Bianchi,A., Oelmann,S., Schaefer,M.R., Schnapp,G. and de Lange,T. (2000) Control of human telomere length by TRF1 and TRF2. *Mol. Cell. Biol.*, **20**, 1659–1668.
- Sogin,M.L., Gunderson,J.H., Elwood,H.J., Alonso,R.A. and Peattie,D.A. (1989) Phylogenetic meaning of the kingdom concept: an unusual ribosomal RNA from *Giardia lamblia*. *Science*, **243**, 75–77.
- Sprung,C.N., Reynolds,G.E., Jasin,M. and Murmane,J.P. (1999) Chromosome healing in mouse embryonic stem cells. *Proc. Natl Acad. Sci. USA*, **96**, 6781–6786.
- Teng,S. and Zakian,V.A. (1999) Telomere–telomere recombination is an efficient bypass pathway for telomere maintenance in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **19**, 8083–8093.
- Van der Ploeg,L.H.T., Liu,A.Y.C. and Borst,P. (1984) Structure of the growing telomeres of trypanosomes. *Cell*, **36**, 459–468.
- van Leeuwen,F., Dirks Mulder,A., Dirks,R.W., Borst,P. and Gibson,W. (1998a) The modified DNA base β -D-glucosyl-hydroxymethyluracil is not found in the tsetse fly stages of *Trypanosoma brucei*. *Mol. Biochem. Parasitol.*, **94**, 127–130.
- van Leeuwen,F., Taylor,M.C., Mondragon,A., Moreau,H., Gibson,W., Kieft,R. and Borst,P. (1998b) β -D-glucosyl-hydroxymethyluracil is a conserved DNA modification in kinetoplastid protozoans and is abundant in their telomeres. *Proc. Natl Acad. Sci. USA*, **95**, 2366–2371.
- van Steensel,B. and de Lange,T. (1997) Control of telomere length by the human telomeric protein TRF1. *Nature*, **385**, 740–743.
- Wang,H. and Blackburn,E.H. (1997) *De novo* telomere addition by *Tetrahymena* telomerase *in vitro*. *EMBO J.*, **16**, 866–879.
- Weiden,M., Osheim,Y.N., Beyer,A.L. and Van der Ploeg,L.H.T. (1991) Chromosome structure: DNA nucleotide sequence elements of a subset of the minichromosomes of the protozoan *Trypanosoma brucei*. *Mol. Cell. Biol.*, **11**, 3823–3834.

Received February 9, 2000; revised March 23, 2000;
accepted March 31, 2000