Volume 17 Number 12 1989 Nucleic Acids Research

Human telomeres contain at least three types of G-rich repeat distributed non-randomly

Robin C.Allshire*, Maureen Dempster and Nicholas D.Hastie

MRC Human Genetics Unit, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK

Received March 2, 1989; Revised and Accepted May 8, 1989

ABSTRACT

Using a combination of different oligonucleotides and restriction enzymes we have examined the gross organisation of repeats within the most distal region of human chromosomes. We demonstrate here that human telomeres do not contain a pure uniform 6 base pair repeat unit but that there are at least three types of repeat. These three types of repeat are present at the ends of most or all human chromosomes. The distribution of each type of repeat appears to be non-random. Each human telomere has a similar arrangement of these repeats relative to the ends of the chromosome. This could reflect differences in the functions that they perform, or might result from the mutation and correction processes occurring at human telomeres. The number of repeat units, the repeat types and arrangement differs at mouse telomeres. Analysing the change in length of the telomeric repeat region between an individuals blood and germline DNA reveals that this is due to variable amounts of the TTAGGG repeat and not the other repeat types. This organisation of repeat units at human telomeres will only be confirmed upon the isolation and sequencing of full length (10−15 kb), intact human telomeres.

INTRODUCTION

Telomeres are the specialised structures at the ends of chromosomes which allow them to be replicated completely. Telomeres also act to prevent chromosome degradation and end to end fusions and may be involved in the attachment of chromosome ends to the nuclear membrane(1,2).

To date the only reports of telomeres which have been successfully cloned have concerned those of eukaryotes with small genomes (less than 20,000kb) (3,4,5,6,7). These DNA sequences, isolated directly from the ends of chromosomes, consist of short tandemly repeated (G+C)-rich sequences of up to 14bp in length (5).

Recently, it has been demonstrated that in *Tetrahymena* cell extracts the G-rich strand acts as a primer for the addition of further copies of the *Tetrahymena* telomere repeat, TTGGGG by the enzyme telomerase (8). Subsequently it was found that synthetic oligonucleotides of the G-rich strand of many of the known telomere repeats would act as primers in these extracts for the addition of TTGGGG (9). This suggests that there is a common conserved structure formed by the telomeric repeats of these widely divergent organisms which can be recognized by the *Tetrahymena* telomerase enzyme and its associated RNA template (10).

It is likely that the telomeres of all eukaryotes are structurally similar and that they have the same mode of function. We have previously shown that the *Tetrahymena* repeat (TTGGGG) cross-hybridises to human telomeres (11). Subsequently, a human repetitive sequence was isolated which hybridised strongly with human telomeres (12). This repetitive sequence is identical to that found at *Trypanosoma* telomeres (TTAGGG). It has also been

demonstrated that the *Arabidopsis thaliana* telomere repeat (TTTAGGG) can detect human telomeric DNA (7).

Since all these different repetitive sequences (TTGGGG, TTAGGG, TTTAGGG) can hybridise with human telomeres, the exact nature of the human telomeric repeat remains unclear. All these sequences might just be cross-hybridising to a uniform tandem repeat. Alternatively there could be several types of repeat motifs which would contribute to a heterogeneous repetitive human telomeric region. If several types of repeat are present at human telomeres then these may or may not be arranged at random with respect to each other. We have utilised specific oligonucleotide probes to investigate the nature of the human telomeric repeat region under conditions which allow us to distinguish between cross-hybridisation and the presence of specific types of repeats. We conclude that on all human chromosomes the telomere is composed of at least three types of tandem G-rich repeats which can be distinguished by hybridisation with three oligonucleotides. These observations allow a more complete characterisation of human telomeric repeats which will assist in understanding their mechanism of function.

MATERIALS AND METHODS

DNA:- *Tetrahymena thermophila* macronuclear DNA was a gift from Carol Greider. *Trypanosoma brucei* DNA from David Barry. *Plasmodium falciparum* DNA from John Scaife. Cloned *Tetrahymena* telomeric repeat, pTC16Δ5, was a gift from Neal Sugawara. The plasmid pDirt3 containing the telomeric repeat of *Didymium. iridis* (identical to that of *Trypanosoma*) is as described by Forney et al (6). To simplify further discussions we refer to the repeat in this plasmid as the *Trypanosoma* telomeric repeat. The cloned *Plasmodium* telomeric repeat is as described by Ponzi et al (4). *Tetrahymena thermophila* rDNA was purified as described in (13). Extractions of human DNA from placenta, blood and sperm were carried out by the standard methods (14,15 and 16 respectively). Total spleen DNA extracted from a DBA mouse was a gift from R. Hill.

DNA was purified from BRL Ultrapure agarose gels as follows; Gel slices were melted at 68°C in an equal volume of 100 mM NaCl 10 mM Tris pH 8.0 1mM EDTA for 30 min. The molten gel slice slurry was cooled to 37°C and digested overnight with 5 u/ml of Agarase (Pharmacia). The DNA was extracted twice with phenol and chloroform and recovered by precipitation with 2.5 M NH_4OAC and two volumes of ethanol.

Purified inserts were radiolabelled with α ^{32}P TTP (Amersham, 800 Ci/mmol) to a specific activity of greater than 10^8 cpm/μg using a random priming kit (BCL).

Synthetic oligonucleotides were manufactured on an Applied Biosystems 381A DNA Synthesizer by John Inglis. Usually 25ng of a 24-mer was labelled in the presence of 100 μCi of γ ^{32}P ATP (Amersham, 5000 Ci/ml) using 10 units of Polynucleotide Kinase (BCL) in the recommended buffer.

Southern Blotting

Restriction enzyme digests were performed according to the manufacturer's recommended conditions. Restricted DNA was separated by agarose gel electrophoresis. DNA was transferred to Hybond-N (Amersham) using a Vacugene vacuum transfer apparatus (LKB). Hybridisations using oligonucleotide probes were performed in 5×SSC, 5×Denharts, 0.1% SDS and 0.1% NaPPi at appropriate temperatures. Normally the filters were washed four times for 15 minutes in 4×SSC and 0.1% SDS, stringency was varied by use of different wash temperatures. In competition experiments the filters were prehybridised in the presence of cold oligonucleotides (2μg/ml) for one hour before the addition of the labelled probe.

Hybridisations using random primed double stranded probes were performed under the conditions of Church and Gilbert (17) at 68°C. Filters were washed four times for 15 minutes in 0.1×SSC at 68°C.

Bal31 Time Course

Reaction mixes contained DNA at a concentration of 200μg/ml which was digested with 15U/ml of Bal31 (BRL Batch 12181) in the manufacturers recommended buffer at 30°C. Initially the mix, without Bal31, was preincubated at 30°C for 10 minutes when one aliquot was removed and incubated in the absence of Bal31 for the duration of the time course (this time point is referred to as T0). Subsequently, the Bal31 was added and aliquots were removed every 5 minutes (T5, T10 etc). The reaction was stopped by the addition of 0.1 volumes of 200 mM EGTA pH8.0. DNA from each aliquot was extracted twice with phenol, chloroform and ether. DNA was recovered by precipitation with 0.3 volumes of 7.5 M NH_4OAC and 2.5 volumes of ethanol. The pellet was washed twice with 70% ethanol, dried and resuspended in TE. DNA was then digested with restriction endonucleases. Generally, between 200 and 1000μg of total genomic DNA was used. Each aliquot contained 20 to 60μg of DNA.

Densitometry

Autoradiographs and photographs of ethidium bromide stained agarose gels were scanned on an automatic autoradiograph scanner (31). The signal intensity was calculated using the integrated optical density of the corresponding peak in the track density profile. The signal intensities in the tracks were corrected with respect to each other for inaccuracies in the amounts of DNA loaded to give the relative signal intensity (Table 1). The ethidium bromide intensities serve as measurements of the relative amounts of DNA loaded in each track.

RESULTS

Human telomeres cross-hybridise with synthetic oligonucleotides identical in sequence to the telomeres of three lower eukaryotes

Oligonucleotides identical in sequence to the telomeric repeats of *Tetrahymena thermopila* (TTGGGG), *Trypanosoma brucei* (TTAGGG) and *Arabadopsis thaliana* (TTTAGGG) were synthesised (*Arabidopsis* and *Plasmodium* telomere repeats are used interchangably since 35% of the *Plasmodium* telomeric repeat is identical in sequence to that of *Arabidopsis*). Radiolabelled oligonucleotides were hybridised to total genomic DNA from human, mouse, *Tetrahymena thermophila, Trypanosoma brucei*, and *Plasmodium falciparum*, which had been digested with HinfI (Figure 1.). To assess the amount of cross-hybridisation between the various probes similar amounts of DNA fragments bearing *T.brucei* (W), *T. thermophila* (X and Z) or *Plasmodium berghei* (Y) telomeric repetitive sequences were run in a single track beside the digested genomic DNA's (Track M). Identical filters were hybridised with each oligonucleotide at 50°C. These filters were then washed in 4×SSC at 55°C and after autoradiography washed again in 4×SSC at 68°C and re-autoradiographed. The higher temperature wash should reduce the levels of any cross-hybridisation seen after washing at 55°C by melting missmatched hybrid molecules. The marker track allows any differential melting of hybrids between the probe and the three types of telomeric repeats to be assessed. Comparison of melting in this marker track, with any observed in total human and mouse DNA, allows conclusions to be drawn as to which sequences are present at human and mouse telomeres.

Figure 1a shows that the oligonucleotide $(TTAGGG)_4$ hybridises very strongly to the

Table 1

Track	Tissue	Enzyme	Signal Intensity under Main Peak $(TTGGGG)_4$	Signal Intensity under Main Peak $(TTAGGG)_4$	Estimate of Main Peak size (bp)
1	Sperm	Sau 3a	759	873	15563
2	Blood	Sau 3a	932	478	9325
3	Sperm	MnlI	367	708	13098
4	Blood	MnlI	254	306	7579
5	Sperm	HphI	218	578	12703
6	Blood	HphI	236	201	8214
7	Sperm	HinfI	550	816	14291
8	Blood	HinfI	527	239	10092

high molecular weight telomeric smear in human, mouse and *Trypanosoma* DNA. No signal is seen in *Tetrahymena* DNA but a slight signal is detected in *Plasmodium* DNA. The control track (M) reveals that under these hybridisation conditions the $(TTAGGG)_4$ oligonucleotide not only hybridises to the *Trypanosoma* like telomeric repeat but also cross-hybridises weakly with the *Plasmodium* telomeric repeat. To reduce the level of this cross-hybridisation the filter was rewashed at 68°C in 4×SSC (Figure 1b). At this temperature the small amount of cross-hybridisation of $(TTAGGG)_4$ to the *Plasmodium* telomeric repeat seen in Figure 1a is reduced. It is clear that most hybridisation with the $(TTAGGG)_4$ oligonucleotide is specific for the *Trypanosoma* telomeric repeat, as indicated by the strong signal seen in the marker track (M) and in total *T. brucei* DNA. In human and mouse DNA the $(TTAGGG)_4$ oligonucleotide hybridises very strongly with the high molecular weight telomeric material. This indicates that the TTAGGG repeat hybridises with a major component of the telomeric repeat in these mammals.

In Figure 1c an autoradiograph of a similar filter hybridised with the oligonucleotide $(TTGGGG)_4$ and washed at 55°C in 4×SSC is shown. Under these conditions this probe hybridises strongly with the *Tetrahymena* (TTGGGG) repeat and can also cross-hybridise with the *Trypanosoma* (TTAGGG) telomeric repeat. This is demonstrated in the marker track (M) and the lower eukaryote genomic DNA tracks. Under these conditions, therefore, one cannot effectively distinguish between the presence in mammalian DNA of bona fide TTGGGG like repeats and cross-hybridisation with TTAGGG like repeats.

Rewashing this filter at 68°C in 4×SSC resulted in the autoradiograph shown in Figure 1d. The level of crosshybridisation of the $(TTGGGG)_4$ oligonucleotide to both *Trypanosoma* and *Plasmodium* telomeric repeats and genomic DNA's is now reduced relative to the hybridisation of $(TTGGGG)_4$ to *Tetrahymena* telomere repeats. In contrast the amount of hybridisation to human DNA remains equivalent to that seen in Figure 1c. Reduction of cross-hybridisation of this $(TTGGGG)_4$ probe to TTAGGG sequences relative to the signal in human DNA indicates that the $(TTGGGG)_4$ oligonucleotide detects

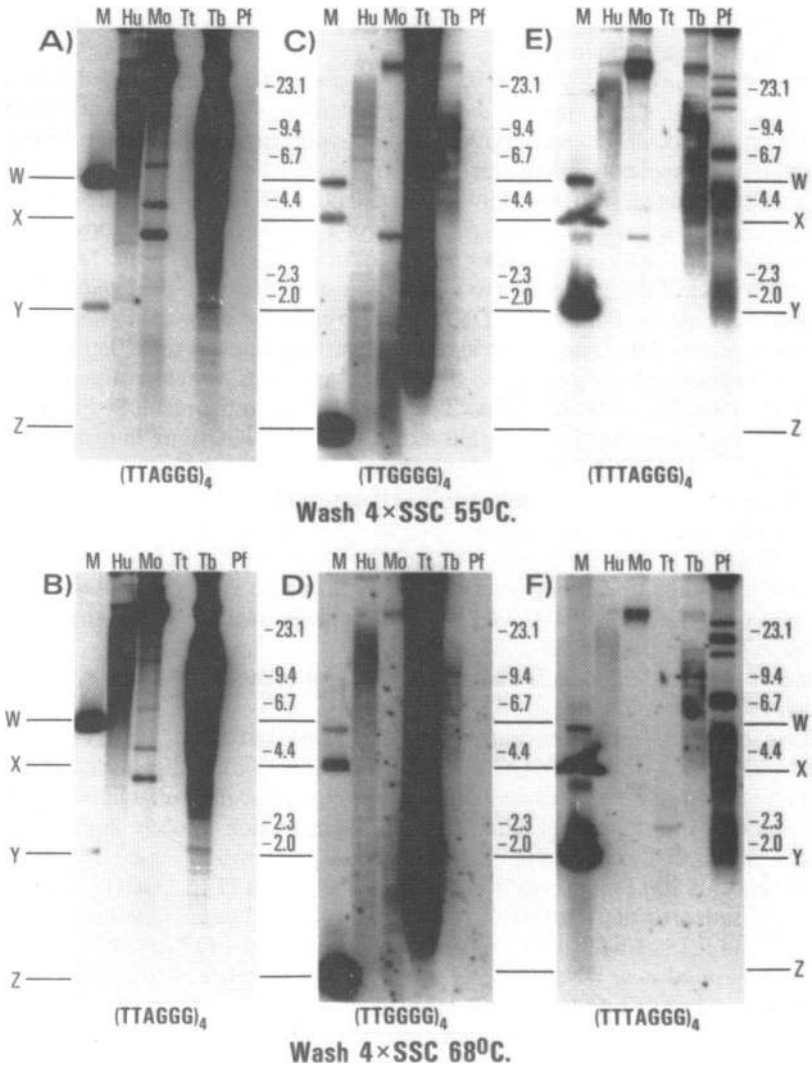


Figure 1. Effect of high stringency washes on oligonucleotide cross-hybridisation. Samples of Human (Hu 5μg), Mouse (Mo 5μg), *T. thermophila* (Tt 0.5μg), *T.brucei* (Tb 0.5μg), and *P. falciparum* (Pf 0.5μg) DNA's were digested with HinfI and electrophoresed through a 0.8% agarose gel and transferred to a nylon filter. The track M contains linearised pDirt3 *Trypanosoma* like TTAGGG repeats (W), terminal BglII fragment from *Tetrahymena* rDNA containing TTGGGG repeats (X), 1.4 kb TaqI/HpaI insert with Plasmodium TTT/CAGGG repeats (Y) and 300 bp BamHI/XhoI insert from pTC1685 with *Tetrahymena* TTGGGG repeats (Z). These three filters were hybridised with the oligonucleotides indicated at 50°C as in Materials and Methods. The filters were washed as indicated. HindIII digest of lamda DNA is indicated as size marker (kb).

sequences at human telomeres which differ from the TTAGGG repeat. This conclusion is strengthened by the fact that hybridisation of this oligonucleotide to mouse DNA is also reduced relative to human DNA at the higher stringency washes. Therefore the

$(TTGGGG)_4$ oligonucleotide detects a specific type of repeat at human telomeres which is absent from those of mouse.

Additional evidence for this differential melting comes from comparing the total lower eukaryote DNA tracks in Figures 1c and d. The intensity of hybridisation of $(TTGGGG)_4$ to total *T. brucei* DNA is reduced relative to total *Tetrahymena* DNA and the Tetrahymena telomeric repeat fragments in the track M.

The results obtained with the third oligonucleotide $(TTTAGGG)_4$ are presented in Figures 1e and f. The autoradiographs shown are of a similar filter to the above, hybridised at 50°C and washed in 4×SSC at either 55°C (1e) or 68°C (1f). In Figure 1e it is apparent that this oligonucleotide, in addition to hybridising strongly to the *Plasmodium* telomeric repeat and total genomic *Plasmodium* DNA, can also cross-hybridise to *Trypanosoma brucei* telomeric sequences. Increasing the stringency of washing causes a reduction in the amount of cross-hybridisation to *Trypanosoma* like telomeric repeats as demonstrated in both the track M and total *Trypanosoma* DNA (T.b.). In contrast to the result in Figures 1c and d, at this level of stringency, hybridisation of $(TTTAGGG)_4$ to human and mouse telomeric material was reduced by an equivalent amount. This indicates that most hybrids formed between the oligonucleotide $(TTTAGGG)_4$ and human and mouse DNA melt simultaneously with hybrids between this oligonucleotide and *Trypanosoma* like telomeric repeats (TTAGGG). Therefore we conclude that most material detected at both human and mouse telomeres by the heptameric TTTAGGG probe is due to cross-hybridisation with copies of the hexamer TTAGGG and does not represent bona fide *Arabidopsis* like telomeric repeats in the human telomeric repeat.

Competition with cold oligonucleotides confirms the presence of two types of repeat unit at human telomeres

To confirm that human telomeres contain two distinct types of repeats, which can be detected with either $(TTAGGG)_4$ or $(TTGGGG)_4$, hybridisation with each oligonucleotide was also performed in the presence of excess of the other two unlabelled oligonucleotides. Such competition experiments should eliminate cross-hybridisation of each oligonucleotide to sequences similar to the competing oligonucleotides. As shown in Figure 2a hybridisation in the presence of 100 fold excess of cold $(TTGGGG)_4$ and $(TTTAGGG)_4$ eliminates all cross-hybridisation of the labelled $(TTAGGG)_4$ probe to *Plasmodium* telomere repeats (Tracks M and P.f.). Under these conditions hybridisation of $(TTAGGG)_4$ with human, mouse and *Trypanosoma* DNA remains strong. In the presence of 100 fold excess cold $(TTAGGG)_4$ and $(TTTAGGG)_4$ the labelled $(TTGGGG)_4$ probe no longer cross-hybridised with *Trypanosoma* like telomeric repeats (Figure 2b, Track M and T.b.). Hybridisation to the human telomeric smear, however, can still be detected (track Hu). Therefore the human telomeric repeat must contain copies of a repeat similar in sequence to TTGGGG but different from TTAGGG.

In Figure 2c the labelled probe, $(TTTAGGG)_4$, was hybridised to a filter in the presence of 100 fold excess $(TTAGGG)_4$ and $(TTGGGG)_4$. This abolished all cross-hybridisation of the probe to *Trypanosoma* telomere repeats (tracks M and Tb). Although this probe still hybridises strongly with *Plasmodium* telomeric repeats (track M and P.f.), all hybridisation to both human and mouse telomeric material is lost. We conclude that neither human or mouse telomeric DNA contain substantial amounts of the seven base pair TTTAGGG repeat identical to sequences found at telomeres of *Arabidopsis* and *Plasmodium*.

From these results it is clear that the telomeric smear in human DNA is composed of

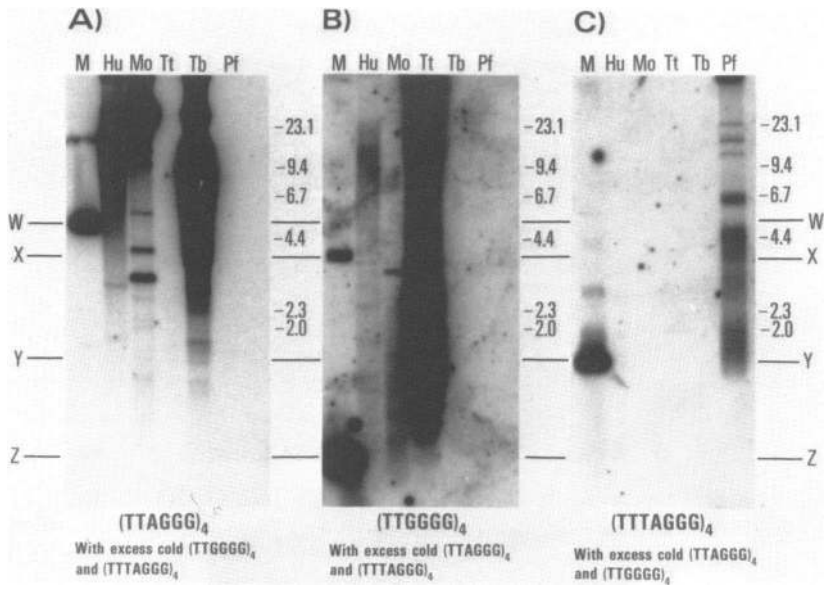
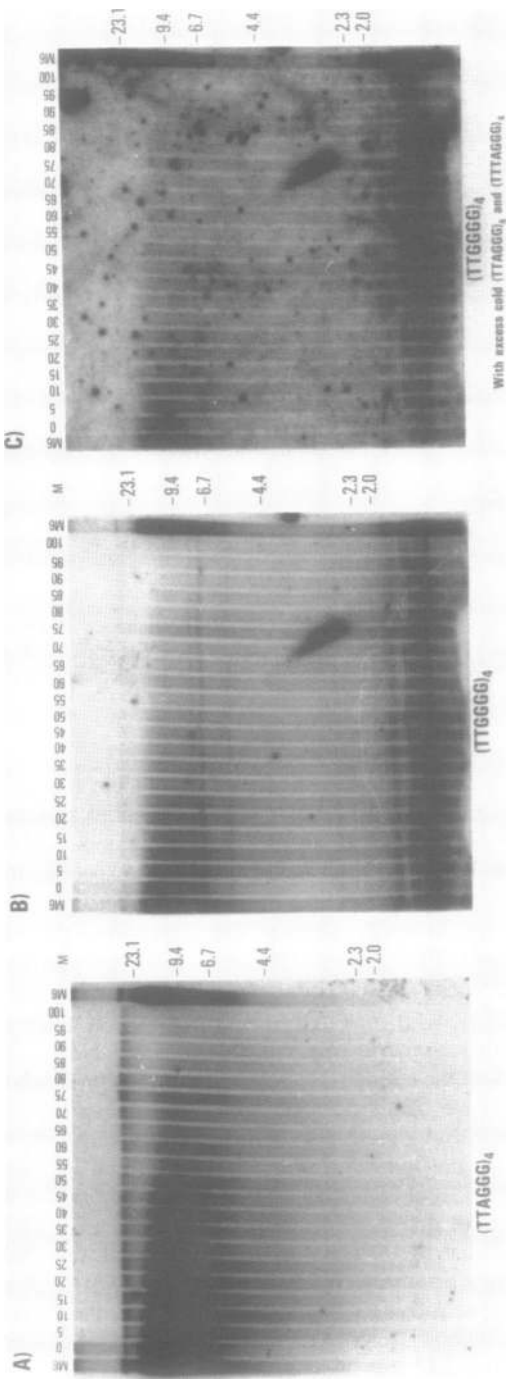


Figure 2. Excess cold oligonucleotides compete out cross-hybridisation between repeats. The three nylon filters used for Figure 1 were boiled for 10 minutes in 1% SDS. These filters were rehybridised with the labelled oligonucleotides in the presence of 100×excess cold competing oligonucleotides as indicated. After hybridisation at 50°C these filters were washed in 4×SSC at 55°C.

at least two types of G-rich six base pair repeat, similar but not necessarily identical to, TTAGGG and TTGGGG. From the relative signal intensities and the length of exposures we estimate that there are between 10 and 100 fold more copies of repeats that hybridise specifically with $(TTAGGG)_4$ than those which hybridise specifically with $(TTGGGG)_4$.

In mouse DNA we find that there is little or no material detected by the TTGGGG repeat in the high molecular weight smear, since all hybridisation of this probe is competed out by cold $(TTAGGG)_4$ and $(TTTAGGG)_4$ (Figure 2b). This presumptive mouse telomeric repeat could consist of pure TTAGGG or alternatively it may contain other variant repeats not detected by the $(TTGGGG)_4$ or $(TTTAGGG)_4$ probes. It is obvious from the data in Figures 1 and 2 that the telomeric smear in this mouse DNA is beyond the resolution of these gels and therefore the telomeric repeat in mouse spleen DNA is greater than 50 kb.

The presence of discrete bands detected with the probes $(TTAGGG)_4$ in mouse and $(TTGGGG)_4$ in human DNA should also be noted (Figure 1). The strong bands seen in mouse DNA, which do not hybridise with $(TTGGGG)_4$ at low stringency but remain hybridised with $(TTAGGG)_4$ at high stringencies, must be more like the TTAGGG repeat. The fact that they are not detected or competed with the $(TTGGGG)_4$ oligonucleotide provides further proof that the $(TTGGGG)_4$ probe recognises different sequences than $(TTAGGG)_4$ in the telomeric smear. We have previously shown that in human DNA these discrete bands remain insensitive to the exonuclease Bal31 and therefore reside proximal to the telomere. These 'internal' telomeric repeats are discussed further elsewhere (Allshire and Hastie in preparation).


A)
B)
C)
M
23.1
9.4
6.7
4.4
2.3
2.0
(TTAGGG)4
(TTGGGG)4
With excess cold (TTAGGG)4 and (TTTAGGG)4

Repeats detected specifically with the $(TTAGGG)_4$ or $(TTGGGG)_4$ oligonucleotides are sensitive to the exonuclease Bal31.

In high molecular weight DNA prepared by standard methods, the only non-random DNA 'ends' should be those located at the natural ends of chromosomes, that is the telomeres. All telomeres isolated to date have been shown to be sensitive to the exonuclease Bal31. Bal31 treatment followed by restriction endonuclease digestion of high molecular weight DNA should cause a continual decrease in the size of the most telomeric restriction fragments with increasing time of exposure to Bal31. We have previously shown that in both human blood and sperm DNA the material which cross-hybridises to the *Tetrahymena* telomere probe is sensitive to Bal31 digestion (11). We show above that the *Tetrahymena* telomeric repeat can cross-hybridise to the *Trypanosoma* telomeric repeat. Since this cross-hybridisation can be melted off at high stringencies or competed out, to reveal distinct *Tetrahymena* telomere like repeats in human DNA, the question arises as to whether or not these TTGGGG like repeats are sensitive to Bal31 and therefore at a telomeric location.

In Figure 3 total genomic DNA extracted from a male human placenta was digested with the same concentration of Bal31 for increasing lengths of time. Subsequently, the DNA from each time point was digested with the endonuclease HinfI and hybridised with either $(TTAGGG)_4$ (Figure 3a), $(TTGGGG)_4$ in the absence (Figure 3b), or presence (Figure 3c), of cold competing oligonucleotides $(TTAGGG)_4$ and $(TTTAGGG)_4$. These filters were also washed at a stringency which would reduce repeat cross-hybridisation.

Figure 3a shows that the $(TTAGGG)_4$ oligonucleotide detects a Bal31 sensitive high molecular weight smear, confirming the results of Moyzis et al (12). In Figure 3b it is clear that this high molecular weight Bal31 sensitive smear is also detected by the oligonucleotide $(TTGGGG)_4$. This hybridisation was performed at 55°C and washed at 65°C. In the presence of excess competing $(TTAGGG)_4$ and $(TTTAGGG)_4$ oligonucleotides a similar result is obtained (Figure 3c). We conclude therefore that both the TTAGGG and TTGGGG like components of the human high molecular weight smear are sensitive to Bal31 and are therefore telomeric in location.

Restriction sites cleaved by MnlI and HphI reside within the human telomeric repeat

Telomeric repeats are non-palindromic in nature. As a result most of these repeats (with the exception of the *S. pombe* repeat) lack cleavage sites for type II restriction endonucleases. We have previously shown that most restriction enzymes with 4 base recognition sites will generate a similar sized high molecular weight smear, demonstrating that the human telomeric repeat is barren for restriction enzyme sites (11). Indeed a triple digest with three such enzymes leaves the 10−20 kb telomeric repeat intact (see Figure 4).

There is a class of restriction endonucleases, sometimes referred to as type IIS enzymes (18), which cleave to one side of non palindromic recognition sequences. MnlI and HphI

Figure 3. Bal31 sensitivity of TTAGGG and TTGGGG like repeats at the human telomere. Bal31 time course was performed and DNA recovered as described in Materials and Methods. 5μg of DNA from each time point was digested with HinfI and electrophoresed through a 0.8% gel, transferred to a nylon filter which was successively hybridised with either (A) $(TTAGGG)_4$ (B) $(TTGGGG)_4$ or (C) $(TTGGGG)_4$ in the presence of 100×excess cold $(TTAGGG)_4$ and $(TTTAGGG)_4$. After hybridisation the filter was washed in 4×SSC at 65°C. The filter was boiled for 10 minutes in 1% SDS between each hybridisation. The tracks M6 contain the starting placental DNA digested with HinfI. T0, T5 etc refer to the time of exposure to Bal31 (minutes). HindIII digested lamda DNA size markers are indicated (kb).

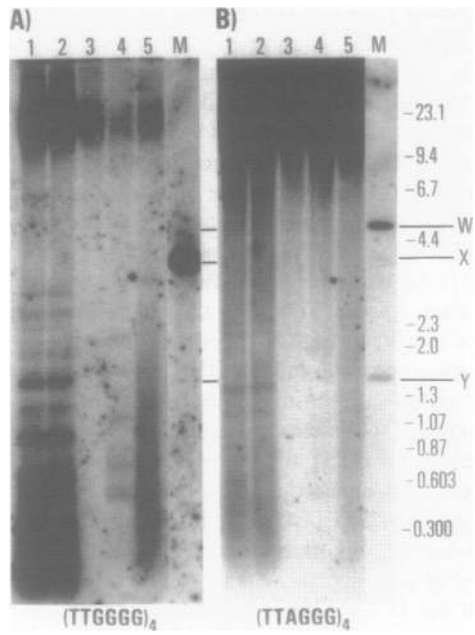


Figure 4. MnlI sites reside within the human telomeric repeat region. Track 1 and 2:- Total sperm DNA (5μg) digested simultaneously with Sau3a, HinfI and HaeIII; Track 3:- Aliquot of isolated 10−20 kb fraction of Sau3a/HinfI/HaeIII digested sperm DNA recovered from low melting point agarose gel; Track 4:- as in track 3 but digested with MnlI; Track 5:- 5μg of total sperm DNA digested with MnlI; Track M:- as in Figure 1 but without Z. The filter resulting from vacuum transfer was hybridised successively with the oligonucleotides indicated in A and B. The filter was washed in 4×SSC at 60°C in each case. Size markers are HindIII digested lamda DNA and HaeIII digested ϕX174 DNA (kb).

are two such enzymes (19). the recognition sites of which are GAGG and GGTGA respectively. Since these sites are G-rich these enzymes might cleave within degenerate versions of the human telomeric repeat. Initial experiments showed that digestion of the human DNA with either MnlI or HphI reduced the size of the human telomeric repeat relative to conventional restriction enzymes with 4 base pair palindromic recognition sites. This size reduction allowed a number of small MnlI and HphI bands to be detected. These bands were only visible when the 300 bp *Tetrahymena* telomere repeat probe or $(TTGGGG)_4$ oligonucleotide were used (data not shown—see Figure 4 and 6).

It is likely that the small MnlI and HphI bands are derived from the high molecular weight telomeric smear, but it is also possible that they are further digest products of internal copies of the TTGGGG like repeat. To distinguish between these two possibilities, total human sperm DNA was digested simultaneously with three restriction endonucleases, Sau3a, HinfI and HaeIII. This DNA was then fractionated on a 1% low melting point agarose gel from which DNA migrating between 10 and 20 kb was isolated. This fraction should contain the heterogeneous telomeric DNA whilst internal copies of the repeat are digested to fragments of a smaller size. Bal31 time course experiments demonstrate, that all material detected in the 10−20 kb size range with both telomere repeats in this sperm DNA is sensitive to the enzyme and therefore telomeric in location (data not shown). Sperm DNA was used since all of the telomeric repeat is in the 10−20 kb range.

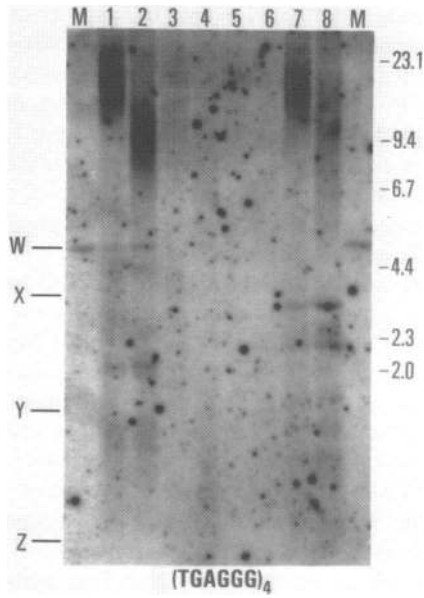


Figure 5. The sequence TGAGGG also resides within the human telomeric repeat. 5μg of DNA extracted from an individuals sperm (Tracks 1,3,5 and 7) or Blood (Tracks 2,4,6 and 8) was digested with Sau3a (Tracks 1 and 2), MnlI (Tracks 3 and 4), HphI (Tracks 5 and 6) or HinfI (Tracks 7 and 8), electrophoresed through a 0.8% agarose gel and transferred to a nylon filter. This filter was hybridised with the oligonucleotide $(TGAGGG)_4$ at 48°C as described in Materials and Methods. After hybridisation the filter was washed in 4×SSC at 60°C. Track M contains the control telomeric repeats W, X, Y and Z as in Figure 1. Size markers are lamda DNA digested with HindIII (kb).

This triple digest of sperm DNA was run on a gel along side the isolated 10−20 kb fraction and an aliquot of that fraction digested with MnlI. The resulting filter was probed with the oligonucleotide $(TTGGGG)_4$ (Figure 4a). Digestion of the 10−20 kb fraction with MnlI releases at least 5 bands of between 0.5 and 2 kb confirming that MnlI sites do indeed reside within the human telomeric repeat. These bands are obscured in total sperm DNA digested with MnlI, due to the large amount of small material hybridising with the oligonucleotide in this region of the filter. Reprobing of the filter with the oligonucleotide $(TTAGGG)_4$ barely detects the small MnlI bands (Figure 4). Therefore these MnlI fragments are composed of a sequence more likely to hybridise with the TTGGGG repeat than with the TTAGGG repeat.

Since human telomeres appear to be composed of repeats similar in sequence to the *Trypanosoma* (TTAGGG), and *Tetrahymena* (TTGGGG) repeats, MnlI sites could occur wherever divergent repeats such as TGAGGG or TTGAGG are located. HphI sites could be formed when a variant TGAGGG repeat follows a TTGGGG or TTAGGG repeat to give the sequence GGTGAGGG. In this case an MnlI site is also created adjacent to the HphI site. If the sequence TGAGGG occurs as a tandem repeat unit within the human telomeric repeat then an oligonucleotide $(TGAGGG)_4$ should detect such repeats. Digestion with MnlI or HphI would reduce such tandem arrays to unit length fragments thereby destroying any signal detected. To test this possibility we have synthesised the

oligonucleotide $(TGAGGG)_4$. Sperm and blood DNA's were digested with Sau3a, MnlI, HphI and HinfI, and hybridised with labelled $(TGAGGG)_4$ (Figure 5). It has been previously shown that human telomeres are longer in an individuals sperm DNA than in his blood DNA (11, 16). It is clear in Figure 5 that this oligonucleotide also detects this telomeric material when the DNA's are cleaved with Sau3a or HinfI. Cleavage with MnlI or HphI, however, destroys most hybridisation of this oligonucleotide to the high molecular weight telomeric smear. This oligonucleotide only cross-hybridises slightly with the *Trypanosoma* telomeric repeat and not at all to the *Tetrahymena* or *Plasmodium* repeats. Therefore the $(TGAGGG)_4$ oligonucleotide must hybridise with a specific type of repeat unit distinct from those composed of TTAGGG or TTGGGG repeats (see rehybridisation of this filter in Figure 6). The human telomeric repetitive region must contain tandem arrays of TGAGGG which are destroyed by the action of MnlI and HphI.

The three types of repeat unit appear to be arranged similarly on most human telomeres

We have presented data above which demonstrate that human telomeric repeats contain at least three types of hexameric repeat unit. Using defined conditions, these three types of repeat can be detected specifically with the oligonucleotides $(TTAGGG)_4$, $(TTGGGG)_4$ and $(TGAGGG)_4$. The TTAGGG repeat represents the major component of the human telomeric repeat. One of the minor repeats TGAGGG must occur in tandem arrays which can be destroyed by cleavage with the restriction enzymes MnlI and HphI. Cleavage with these enzymes also results in the formation of DNA fragments only detectable with the TTGGGG repeat. These three types of repeat unit could be organised in a number of different arrangements with respect to each other and the ends of human chromosomes, for example:-

1. Individual human telomeres could be composed of different types of repeats. Therefore MnlI and HphI sites might be confined to a subset of telomeres.
2. All human telomeres might contain a random mixture of each type of repeat so that MnlI and HphI cleave each telomere differently.
3. Each telomere might have a similar arrangement of all repeat types so that each type of repeat, the MnlI and HphI sites reside in a definable position with respect to all chromosome ends.

Using oligonucleotide probes in conjunction with restriction enzymes which cut within (MnlI and HphI) and proximal (Sau3a and HinfI) to the telomeric repeat, we have attempted to define the gross structural organisation of the different types of repeat units on all endogenous human telomeres. Comparisons of telomeric smear signal intensity and size were performed on an individuals sperm and blood DNA digested with the above enzymes, and hybridised at high stringency with both the $(TTGGGG)_4$ and $(TTAGGG)_4$ probes (Figure 6 and Table 1). A schematic representation of the proposed arrangement of repeats at human telomeres is presented in Figure 7.

In Table 1 the size estimates of the main telomeric smear peak demonstrate that the human telomeric smear is 5 kb smaller in this individuals blood DNA when compared with his sperm DNA, regardless of the restriction enzyme used. Since the whole telomeric smear shifts by 5 kb all telomeres must be shorter in blood DNA. We assume that blood telomeres are shortened by erosion from the ends of chromosomes. This decrease in telomere length allows us to investigate which types of repeat unit are lost from somatic telomeres. It is clear from Figures 5 and 6 and Table 1 that all three oligonucleotides detect this decrease in telomere length. However, the signal intensity with $(TTGGGG)_4$ in both sperm and blood telomeric material is equivalent with all enzymes used, therefore there is no loss

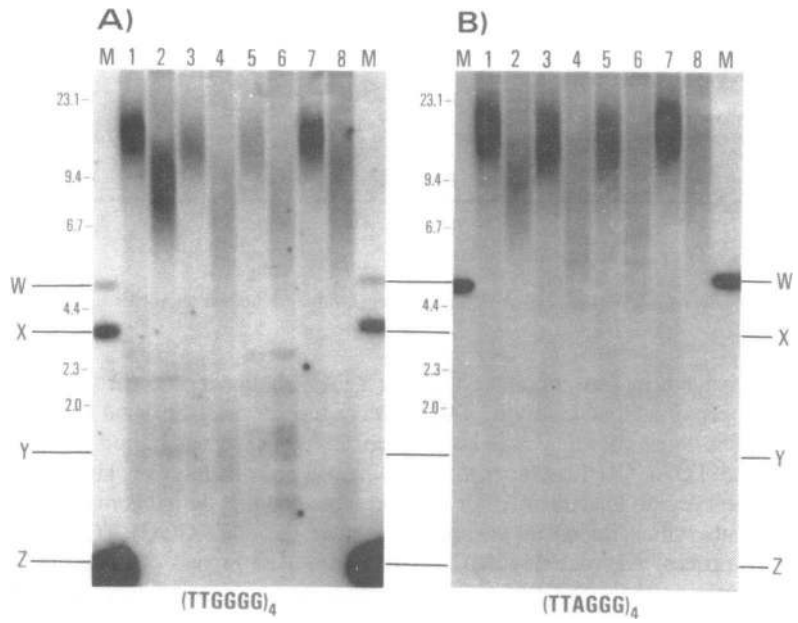


Figure 6. Hybridisation of $(TTGGGG)_4$ and $(TTAGGG)_4$ to Human sperm and blood DNA. The nylon filter used in Figure 5 was hybridised successively with the labelled oligonucleotides as indicated in A and B. The filter was boiled for 10 minutes in 1% SDS between each hybridisation. Size markers are HindIII digested lamda DNA (kb).

of TTGGGG like repeat units from blood telomeres. In contrast the signal intensity detected in blood DNA with the $(TTAGGG)_4$ probe is only 40% of that seen in sperm with all these enzymes. Therefore, 60% of TTAGGG like repeat units are lost from blood telomeres. This suggests that the repeats which are specifically detected with the TTGGGG probe lie proximal to most TTAGGG repeats. A similar argument allows us to conclude that the TGAGGG repeats also lie internal to most TTAGGG repeats (see Figure 5 tracks 1,2,7 and 8).

The intensity of hybridisation of $(TTGGGG)_4$ to the telomeric smear is decreased in MnlI and HphI digests of sperm and blood DNA relative to Sau3a and HinfI digests (Figure 6). This confirms the observations made in Figure 5. Table 1 shows that digestion of human telomeric repeats with MnlI or HphI leads to a loss of 61% of the material detected with $(TTGGGG)_4$ as compared to Sau3a and HinfI digests. This loss of TTGGGG like repeats is accompanied by, on average, a 1.9 kb decrease in size of main telomeric smear peak, and is identical in both sperm and blood DNA's. This decrease in size is also detected with $(TTAGGG)_4$. Therefore the MnlI and HphI sites and 61% of TTGGGG like repeats must reside within the most proximal 1.9 kb of the telomeric repeat, that is approximately 13 kb from the end of each chromosome in sperm. All remaining TTGGGG like repeats (1.2 kb) must reside distal to all MnlI sites.

Although our estimates show a reduction in size of the telomeric smear detected by $(TTAGGG)_4$ of approximately 5 kb or 30%, on going from sperm to blood, the signal intensity decreases by 60%, double the size estimate. This suggests that 60% of the TTAGGG like repeats reside in the last 5 kb of these sperm telomeres. Therefore there

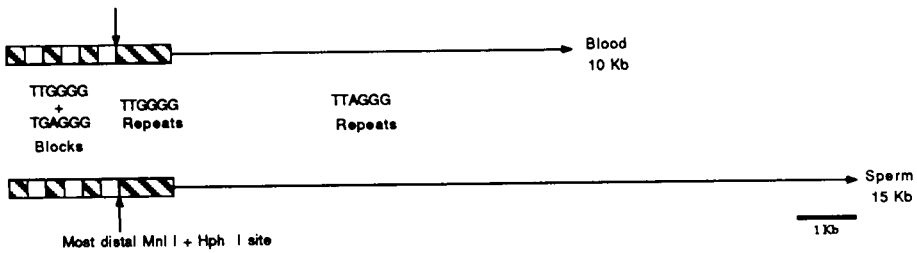


Figure 7. Schematic representation of proposed repeat organization at human telomeres.

can only be, on average, a maximum of 8.3 kb of TTAGGG like repeats per sperm telomere. In addition approximately 1.2 kb of TTGGGG like repeats must be distal to all MnlI and HphI sites. The most proximal 1.9 kb of each telomeric repeat region is made up of TTGGGG and TGAGGG like repeats, this leaves approximately 3.6 kb of the human telomeric repeat region unaccounted for. It is possible that this deficit is made up by other types of repeats which do not hybridise with either the $(TTAGGG)_4$, $(TTGGGG)_4$ or $(TGAGGG)_4$ probes. Alternatively the 1.2 kb of TTGGGG repeats distal to the MnlI and HphI sites is interspersed with TTAGGG and/or other repeat types and these are not detected efficiently with the probes used. Human telomeres can also be detected with probes containing the pentameric repeats, TTAGG, TAGGG or TTGGG (data not shown) perhaps these or other variants are also represented.

DISCUSSION.

We have shown that intact endogenous human telomeres contain a non random mixture of at least three types of G-rich hexameric repeat units. These different repeats are detected specifically with probes bearing the sequences TTAGGG, TTGGGG or TGAGGG. The majority of the human telomeric repeat units in the DNA's examined here are similar to the repeat TTAGGG, most of which forms the extreme end of all human chromosomes. We have positioned two other types of repeat relative to TTAGGG like repeats (Figure 7). Firstly, tandem repeats detected with the sequence $(TGAGGG)_4$ are present on both sperm and blood telomeres and are destroyed by the action of MnlI and HphI which reduces the size of the human telomeric repeat region by 1.9 kb. Therefore the most distal TGAGGG like repeats can only reside 1.9 kb from the proximal end of the repetitive region, that is approximately 13 and 8 kb from the chromosome end in sperm and blood DNA respectively. Secondly 61% of those repeats detected with the $(TTGGGG)_4$ oligonucleotide are cleaved from the telomeric smear by MnlI and HphI to form multiple small DNA fragments which are detectable with only this probe. The signal intensity of the telomeric smear and the band pattern are identical in both sperm and blood DNA. Therefore most of these TTGGGG like repeats must also reside in the proximal 2 kb of the telomeric repeat region (that is between 13 and 15 kb from the end of sperm chromosomes). Blocks of TTGGGG and TGAGGG like tandem repeats may or may not be interspersed with each other. Since all telomeric material detected with all three probes behaves in a similar fashion, on either comparison of sperm and blood DNA or upon cleavage with MnlI and HphI, it is likely that all human telomeres have a similar arrangement of these three repeats.

The significance of several types of non-randomly distributed repeats within the most terminal 10 or 15 kb of most or all human chromosomes is unknown. It could reflect functional organisation of human telomeres or alternatively might be due to the acquisition and accumulation of mutations in the most proximal repeats. It can be seen from the data presented in Figure 2 that the mouse telomere does not appear to contain many copies of repeat which specifically hybridise with $(TTGGGG)_4$. It is possible that the human TTGGGG and TGAGGG like repeats arose by the acquisition of mutations in copies of the TTAGGG repeat to give variant forms which subsequently became amplified and transferred to all telomeres by recombinogenic interactions (20). In this case the mouse may have lost all copies of the telomeric TTGGGG repeat by deletion. It is also possible that the mouse telomere contains a type of repeat, in place of TTGGGG, which cannot be detected by hybridisation with TTGGGG repeats. Such changes in simple repetitive sequences are well documented in relation to satellite sequences. It is of interest to note that in the guinea pig, *Cavia procellus*, and kangaroo rat, *Dipodoymus ordii*, the sequence TTAGGG forms well characterised satellite DNA sequences (21,22 and 23). The HSα satellite sequence of *D. ordii* constitutes approximately 19% of the genome. It is known that only 25% of this satellite consists of TTAGGG repeats and the remaining repeats differ by single and double base pair changes (23). If the TTGGGG and TGAGGG like and other variant units in the human telomeric repeat arise by random mutation and amplification then one might expect that most of this heterogeneity would remain at a proximal location, since this region of the telomeric repeat would be less likely to be exposed and therefore corrected by a putative human telomerase.

The different telomeric repeats may play different roles in telomere function perhaps interacting with different proteins. Preliminary band shift assays suggest that this is indeed the case in both human and mouse (Mitchell and Allshire, unpublished results). The TTGGGG and/or TGAGGG like repeats could conceivably be involved in the pairing of homologous chromosomes and recombination during meiosis. The TTAGGG like repeats would then form the functional telomere which interacts with a putative human telomerase and protect the telomere and chromosome by binding other proteins.

The presence of two types of repeat at telomeres is not unique to human telomeres. Recently it has been found that the *Paramecium tetraurelia* telomeric repeat consists of two types of repeats, TTGGGG and TTTGGG, with the occasional TTTTGGG (24,25). However, unlike human telomeres, in *Paramecium*, the two repeats are mixed at random and there does not appear to be separation of the two into distinct regions. Other organisms such as *Schizosaccharomyces pombe* contain highly variable telomeric repeats, $T_{1-2}ACA_{0-1}C_{0-1}G_{1-6}$, which allow the repeat unit to vary in length between four and twelve base pairs (5).

It remains unclear as to why the human and mouse telomeric repeat region should be so long. In many of the lower eukaryotes the length of the repeat is variable but is seldom greater than 1 kb (3,4,5 and 6), even when grown under conditions which favor prolonged telomere elongation (26,27). In human sperm the telomeric repeat is 15 times longer and in mouse 50 to 100 times the length of these lower eukaryotic telomeres. It would appear that other higher eukaryotes apart from mammals also contain these longer stretches of telomeric repeat (7,11). This could reflect additional functions performed by higher eukaryotic telomeres not required of their lower eukaryote counterparts. Longer stretches of telomeric repeats on the longer mammalian chromosomes might enable the ends of homologous chromosomes to come together more readily at synapsis. Alternatively, it might

be that the machinery for repairing the ends of chromosomes in lower eukaryotes is more efficient, due perhaps to a higher activity of telomerase. It is possible that longer stretches of telomeric repeats might minimise the chance of deleterious telomere-telomere fusion events occuring. The outcome of such events in mammalian cells could lead to aneuploidy, which in turn might initiate the formation of a neoplastic cell.

The human telomeric repeat is longer in sperm than somatic tissue. If, in mammals telomerase were only expressed at very low levels, or perhaps only in the germline, the telomeric repeat would have to be longer to withstand the continual erosion of the ends of chromosomes with every round of replication throughout development. Other data (Allshire and Hastie unpublished results) suggests that the human telomeric repeat can increase in length in some cell lines, therefore this putative human telomerase activity is not necessarily germ line specific. Recombination-dependent replication has also been proposed as a mechanism by which telomeres could expand and contract (20,28,29,30). It is possible that telomere length is greater in the germline as a result of meiotic recombination processes which maintain their length. It is interesting to note that G-rich sequences similar to the repeats at human telomeres have the ability to form Hoogsteen-bonded four-stranded structures. It has been proposed that such structures may play a role in the pairing of homologous chromosomes during meiosis and may initiate recombination (31). A combination of both a telomerase like enzyme and recombination dependent replication could also be responsible for the variation in telomere length seen on comparison of blood and sperm.

The method most favoured for cloning human telomeres requires functional complementation in *Saccharomyces cerevisiae* on a modified YAC vector. This method has recently proven successful for isolating human telomeres, however only 1−2 kb of the human telomeric repeat is retained on these YAC's (33,34). The arrangement of repeats at human telomeres described here shall only be confirmed upon the cloning and sequencing of full length (10−15 kb), intact human telomeres.

The further characterisation of the DNA and associated proteins at the telomeric region of mammalian chromosomes should allow the complete dissection of the role they perform in many processes. This should ultimately allow a better understanding of one aspect of chromosome function in mammalian cells.

ACKNOWLEDGEMENTS

We thank Sally Cross, Howard Cooke, Wendy Bickmore, Veronica van Heyningen and other colleagues for their invaluable advice and discussion; Carol Greider for same and help with Figure 7. John Inglis for synthesising oligonucleotides; D. Green for scanning autoradiographs and densitometry measurements; N. Davidson, S. Bruce and D. Stewart for photographic and art work; and Prof. H.J. Evans for his encouragement and support.

*To whom correspondence should be addressed at Cold Spring Harbor Laboratory, PO Box 100, Cold Spring Harbor, NY 11724, USA

REFERENCES

1. Blackburn, E.H. and Szostak, J.W. (1948) A. Rev. Biochem. **53**, 163−194.
2. Agard, D.A. and Sedat, J.W. (1983) Nature **302**, 676−681.
3. Shampay, J., Szostak, J.W., and Blackburn, E.H. (1984) Nature **310**, 154−157.
4. Ponzi, M., Pace, T., Dore, E., and Frontali, C. (1985) EMBO J. **4**, 2991−2995.
5. Sugawara, N., and Szostak, J.W. (1986) Yeast **2**, (supplement) p.373.

6. Forney, J., Henderson, E.R., and Blackburn, E.H. (1987) Nucl. Acids Res. **15**, 9143–9152.
7. Richards, E.J. and Ausubel, F.M. (1988) Cell **53**, 127–136.
8. Greider, C.W. and Blackburn, E.H. (1985) Cell **43**, 405–413.
9. Greider, C.W. and Blackburn, E.H. (1987) Cell **51**, 887–898.
10. Greider, C.W. and Blackburn, E.H. (1989) Nature **337**, 331–337.
11. Allshire, R.C., Gosden, J.R., Cross, S.H., Cranston, G., Rout, D., Sugawara, N., Szostak, J.W., Fantes, P.A., and Hastie, N.D. (1988) Nature **332**, 656–659.
12. Moyzis, R.K., Buckingham, J.M., Cram, L.S., Dani, M., Deaven, L.L., Jones, M.D., Meyne, J., Ratliff, R.L., and Wu, J-R. (1988) Proc. Natl. Acad. Sci. USA **85**, 6622–6626.
13. Wild, M.A. and Gall, J.G. (1979) **16**, 565–573.
14. Gross-Bellard, M., Oudet, P. and Chambon, D. (1973) Eur. J. Biochem. **36**, 32–40.
15. Gosden, J.R., Gosden, C.M., Christie, S., Cooke, H.J.,, Moresman, J.M., and Rodeck (1984) **66**, 347–351.
16. Cooke, H.J. and Smith, B.A. (1986) Cold Spring Harbor Laboratory Sym. Quan. Biol. **51**, 213–219.
17. Church, G.M., and Gilbert, W. (1984) Proc. Natl. Acad. Sci. USA **81**, 1991–1995.
18. Szybalski, W. (1985) Gene **40**, 169–173.
19. Roberts, R.J. (1987) Nucl. Acids Res. **15**, Supplement r189–r217.
20. Pluta, A.F. and Zakian, V.A. (1989) Nature **337**, 429–433.
21. Southern, E.M. (1970) Nature **227**, 794–798.
22. Mazrimas, J.A. and Hatch, F.T. (1977) Nucl. Acids Res. **4**, 3215–3227.
23. Fry, K. and Salser, W. (1977) Cell **12**, 1069–1084.
24. Forney, J.D. and Blackburn, E.H. (1988) Mol. Cell Biol. **8**, 251–258.
25. Baroin, A., Prat, A., and Caron, F. (1987) Nucl. Acids Res. **15**, 1717–1728.
26. Larson, D.D., Spangler, E.A. and Blackburn, E.H. (1987) Cell **50**, 477–483.
27. Shampay, J. and Blackburn, E.H. (1988) Proc. Natl. Acad. Sci. USA **85**, 534–536.
28. Walmsley, R.M., Szostak, J.W. and Petes, T.D. (1983) Nature **302**, 84–86.
29. Bergold, P.J., Campbell, G.R., Littau, V.C. and Johnson, E.M. (1983) Cell **32**, 1287–1299.
30. Formosa, T. and Alberts, B.M. Cell **47**, 793–806.
31. Sen, D. and Gilbert, W. (1988) Nature **334**, 364–366.
32. Elder, J.K., Green, D.K. and Southern, E.M. (1986) Nucl. Acids Res. **14**, 417–424.
33. Cross, S.H., Allshire, R.C., McKay, S., McGill, N., and Cooke, H.J. (1989) Nature **338**, 771–774.
34. Brown, W.R.A. (1989) Nature **338**, 774–776.

This article, submitted on disc, has been automatically converted into this typeset format by the publisher.