 Human Molecular Genetics, 1997, Vol. 6, No. 1 69–76

Comparison of the human and mouse genes encoding the telomeric protein, TRF1: chromosomal localization, expression and conserved protein domains

Dominique Broccoli, Laura Chong, Stefan Oelmann, Anthony A. Fernald[1], Nicola Marziliano[2,+], Bas van Steensel, David Kipling[2], Michelle M. Le Beau[1] and Titia de Lange*

The Rockefeller University, 1230 York Avenue, New York, NY 10021-6399, USA, [1]Department of Medicine, University of Chicago, Chicago, IL 60637, USA and [2]MRC Human Genetics Unit, Western General Hospital, Crewe Road, Edinburgh, UK

Received July 31, 1996; Revised and Accepted October 14, 1996 DDBJ/EMBL/GenBank accession nos U65586, U70992–U70994

Mammalian chromosome ends contain long arrays of TTAGGG repeats that are complexed to a telomere specific protein, the TTAGGG repeat binding factor, TRF1. Here we describe the characterization of genes encoding the human and mouse TRF1 proteins, hTRF1 and mTRF1. The mTRF1 cDNA was isolated based on sequence similarity to the hTRF1 cDNA and the mTRF1 mRNA was shown to be ubiquitously expressed as a single 1.9 kb polyadenylated transcript in mouse somatic tissues. High levels of a 2.1 kb transcript were found in testes. *In vitro* translation of the mTRF1 cDNA resulted in a 56 kDa protein that binds to TTAGGG repeat arrays. mTRF1 displayed the same sequence specificity as hTRF1, preferring arrays of TTAGGG repeats as a binding substrate over TTAGGC and TTGGGG repeats. Expression of an epitope-tagged version of mTRF1 showed that the protein is located at the ends of murine metaphase chromosomes. In agreement, conceptual translation indicated that mTRF1 and hTRF1 are similarly-sized proteins with nearly identical C-terminal Myb-related DNA binding motifs. In addition, comparison of the predicted mTRF1 and hTRF1 amino acid sequences showed that the acidic nature of the N-terminus of TRF1 is conserved and revealed a highly conserved novel domain of ~200 amino acids in the middle of the proteins. However, other regions of the proteins are poorly conserved (<35% identity) and the overall level of identity of the mTRF1 and hTRF1 amino acid sequences is only 67%. The TRF1 genes are not syntenic; the hTRF1 gene localized to human chromosome 8 band q13 while the mTRF1 gene localized to mouse chromosome 17 band E3. The data indicate that the genes for mammalian telomeric proteins evolve rapidly.

INTRODUCTION

Telomeres, the natural ends of eukaryotic chromosomes, are specialized nucleoprotein structures that ensure the complete replication and stability of linear chromosomes (reviewed in 1,2). The DNA of telomeres is usually composed of tandem repeats of sequences that are G-rich in the strand that extends to the 3′ end of the chromosome. In vertebrates, the telomeric array is comprised of TTAGGG repeats and can extend from a few kilobasepairs up to tens of kilobasepairs in length (3–5). This repeat array is maintained by telomerase, which adds telomeric repeats *de novo* to 3′ ends using an RNA template (reviewed in 6,7). Without telomeres, the ends of chromosomes are unstable possibly because they become fusogenic and susceptible to attack by nucleases. Telomere stability is believed to be accomplished by the action of telomere-specific binding proteins which sequester the end of the molecule from DNA damage checkpoints and inappropriate repair activities (8–10). In addition, telomeric DNA binding proteins have been implicated in telomere length regulation in yeast (11–14).

Evidence for the presence of a nucleoprotein complex at human chromosome ends was initially gleaned from the inspection of telomeric chromatin and the analysis of subnuclear localization of telomeres. Human telomeres are tethered to the nuclear matrix (15,16) and part of the telomeric chromatin displays an altered, diffuse nucleosomal organization (17,18). The only vertebrate telomeric protein reported to date is the human telomeric repeat binding factor, hTRF1 (19). This DNA binding protein has properties expected for a factor that binds telomeres along their length. hTRF1 specifically recognizes double-stranded repeats of the vertebrate telomeric sequence TTAGGG and does not require a DNA end to bind to its cognate sequence (20). hTRF1 was cloned through biochemical purification and found to be a novel protein containing a Myb-related DNA binding motif (19). The

*To whom correspondence should be addressed
+Present address: Dipartimento di Biologia Animale, Universita di Pavia, Piazza Botta 9–10, I 27100 Pavia, Italia

telomeric protein is a 439 amino acid nuclear protein that occupies all telomeres throughout the cell cycle (19). Since TRF1 specifically fractionates to the nuclear matrix, its binding may be responsible for the condensed structure of interphase telomeres and their association with the nuclear matrix (16). A distantly related homolog of hTRF1, called hTRF2, was recently isolated and also shown to be a component of the telomeric complex (DB, A. Smogorzewska, LC, and TdL, manuscript in preparation).

Interestingly, hTRF1 is not a homologue of Rap1p, the only other duplex telomeric DNA binding protein described to date. Rap1p is a repressor/activator protein that specifically coats the TG_{1-3} telomeric tracts of *Saccharomyces cerevisiae* (reviewed in 21,22). Rap1p recruits a series of other factors to the telomere, including Sir3p, Sir4p and Rif1p. The resulting complex forms a specific chromatin domain (telosome) suggesting that the chromosome end is sequestered in a higher order structure (23). Rap1p binding at the telomeres is required for proper telomere length regulation, telomere stability, and the telomeric complex recruited by Rap1p is responsible for telomeric silencing. Like hTRF1, Rap1p uses a Myb-related three helix bundle motif for site recognition (24). However, while hTRF1 has one Myb HTH motif, Rap1p recognizes its telomeric binding site with two such elements. Outside the DNA binding domain these proteins do not show any sequence similarities.

The lack of sequence similarity between hTRF1 and Rap1p and the absence of an hTRF1 homolog in the complete yeast genome sequence raised the possibility that the TRF1 proteins might be evolving rapidly. To address this question we have cloned and analysed the mouse homolog of hTRF1, mTRF1, and compared the human and mouse proteins and their genes in detail. The results indicate that while mTRF1 and hTRF1 are biochemically indistinguishable, their primary amino acid sequence has diverged. Comparison of the deduced reading frames of mouse and human TRF1 indicates that the most conserved features of these proteins are the Myb-homology, the acidic domain in the N-terminus, a TRF1 specific domain in the middle of the proteins, and several potential phosphorylation sites.

RESULTS

Isolation of mTRF1 cDNAs

A cDNA fragment spanning most of the hTRF1 reading frame was used to screen a mouse brain cDNA library under moderate stringency conditions and 12 cDNA clones were isolated. The clone with the largest cDNA insert, clone 12-1, was used in all subsequent analyses. This clone had an insert of 1640 bp with 65–80% sequence identity to the hTRF1 cDNA (GenBank accession number U65586). The mTRF1 cDNA showed a single open reading frame of 421 amino acids, which is comparable in length to the open reading frame encoding hTRF1 (439 amino acids). A methionine start codon flanked by a Kozak consensus sequence for translation initiation is located 10 nt from the 5′ end and 335 nt of untranslated sequence precedes an AATAAA polyadenylation signal at the 3′ end of the insert. The identified start codon is likely to represent the N-terminus of the mTRF1 polypeptide because seven out of eight amino acids that follow this methionine are identical to the N-terminus of hTRF1. *In vitro* transcription/translation of the mTRF1 cDNA resulted in the synthesis of a polypeptide with an apparent MW of ~56 kDa

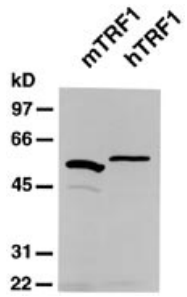

Figure 1. *In vitro* translation products of mTRF1 and hTRF1 cDNAs. mTRF1 and hTRF1 cDNAs were used in a coupled *in vitro* transcription/translation reaction and the ^{35}S-methionine labelled proteins were analysed by electrophoresis through a 10% polyacrylamide/SDS gel. Molecular weights of protein standards are indicated. Relative to these standards mTRF1 protein migrates at an apparent molecular weight of 56 kDa and hTRF1 protein migrates at ~60 kDa.

(Fig. 1). In agreement with its slightly shorter ORF, the protein encoded by the mTRF1 cDNA migrates somewhat faster than the human TRF1 protein (Fig. 1).

In order to determine whether the cDNA encoded mouse TRF1 activity, *in vitro* translation product was tested for DNA binding activity (Fig. 2, right panel). Mouse TRF1 activity, isolated from the Balb/c J558 plasmacytoma cell line, was assayed in parallel (Fig. 2, left panel). Both proteins were allowed to form a complex with a double-stranded restriction fragment containing 12 tandem TTAGGG repeats which represents an optimal site for TRF1 binding in gel-shift assays (20). As reported previously (20), mouse cells express a TRF1 activity that, like the human TRF1, forms a gel-shift complex with TTAGGG repeat probes. This activity is specific for TTAGGG repeats. While TTAGGG repeat-bearing plasmids compete with the labelled probe for complex formation, no competition occurred with plasmids containing *C.elegans* TTAGGC repeat arrays and *Tetrahymena* TTGGGG repeats had an intermediate competition effect. This relative sequence preference (TTAGGG>TTGGGG>TTAGGC) seen with mouse TRF1 is identical to that observed with human TRF1 (19,20,25). The DNA binding activity and relative sequence preference of *in vitro* translated mTRF1 protein was indistinguishable from *ex vivo* mTRF1 (Fig. 2), indicating that the isolated recombinant clone encodes the TTAGGG repeat specific DNA binding activity.

Localization of mTRF1 to the ends of mouse metaphase chromosomes

The mTRF1 reading frame was fused to an N-terminal sequence encoding the FLAG epitope and expressed transiently by transfection into mouse NIH 3T3 cells. Metaphase chromosomes from the transfected cells were examined for the location of the mTRF1 protein by indirect immunofluorescence microscopy using a monoclonal antibody directed against the FLAG epitope (Fig. 3). While metaphase chromosomes from untransfected cells did not reveal specific staining, we consistently observed signals at the ends of chromosomes from cells expressing the FLAG-tagged mTRF1 gene. In addition, a punctate pattern was observed in transfected interphase nuclei (data not shown) as expected if mTRF1 is located at telomeres throughout the cell cycle. These results show that, as anticipated from previous results (19), mTRF1 represents a mouse telomeric protein.

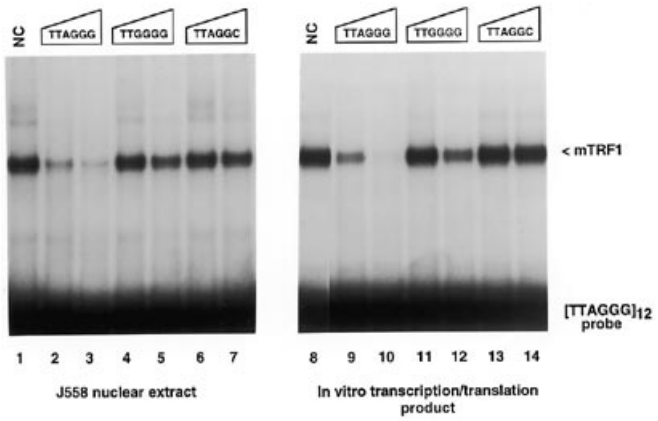

Figure 2. Identical TTAGGG repeat binding activities of *ex vivo* mTRF1 and the protein encoded by the mTRF1 cDNA. Gel-shift reactions were performed with a ^{32}P-labelled double-stranded DNA probe containing 12 tandem TTAGGG repeats in the presence of an excess (1 μg) of sheared *Escherichia coli* DNA. The reactions in lanes 1–7 were carried out with 5 μg of nuclear extract proteins isolated from the mouse J558 plasmacytoma cells. The reactions in lanes 8–14 contained 1 μl of product mix formed by *in vitro* transcription/translation of the mTRF1 plasmid. Under these conditions, no gel-shift is observed in reactions containing *in vitro* transcription/translation products formed in the absence of mTRF1 plasmid (19). Competition experiments were performed with circular plasmids containing 1.6 kb of TTAGGG repeats (lanes 2, 3, 9 and 10), 1.1 kb of TTGGGG repeats (lanes 4, 5, 11 and 12), and 2.0 kb of TTAGGC repeats (lanes 6, 7, 13 and 14). Competitor DNAs (20 and 200 ng, representing 1× and 10× excess of plasmid molecules over probe molecules) were added before the addition of protein.

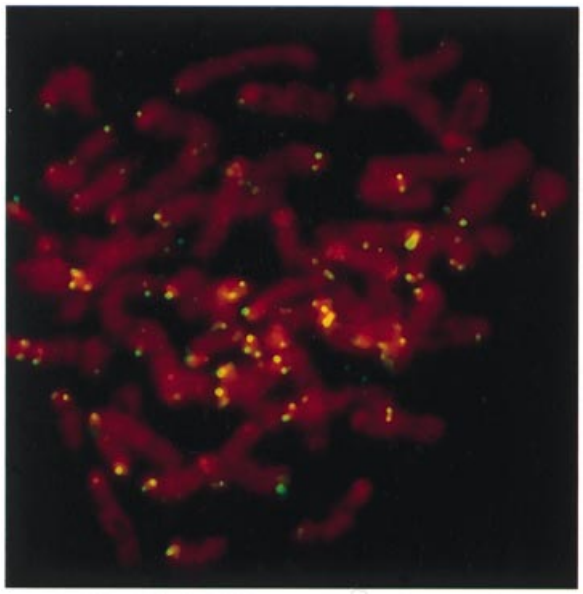

Figure 3. mTRF1 at the ends of murine metaphase chromosomes. FLAG-tagged mTRF1 was transiently expressed in NIH 3T3 cells and the metaphase chromosomes from transfected cells were stained with a monoclonal antibody directed against the FLAG epitope (green). The chromosomal DNA is stained with DAPI (converted to red).

Comparison of the mTRF1 and hTRF1 ORFs

Conceptual translation of the mTRF1 ORF (GenBank accession number U65586) and the alignment of this sequence with the hTRF1 ORF are shown in Figure 4. Although the two proteins are clearly related, they show only 67% overall amino acid identity. This low level of sequence conservation allowed identification of the most conserved domains of TRF1. A high level of sequence identity (84%) was seen in the C-terminus of the proteins where mTRF1 and hTRF1 contain nearly identical Myb-related three helix bundles. This region of the proteins is likely to be involved in the recognition of the TTAGGG repeat binding site. The sequence SVLMLKDRWRT (positions 404–413 in mTRF1), presumed to be involved in site recognition based on the structure of the Myb HTH fold (26), was identical in the two proteins.

In addition, the alignment of mouse and human TRF1 revealed that TRF1 has two other conserved domains (Fig. 4A and B). While the most N-terminal domain was not conserved in its actual amino acid sequence, it is highly conserved in its acidic nature. For both proteins the local pI of N-terminal acidic domains (see Fig. 4B for coordinates) was <3.5. The acidic domain was followed by a more highly conserved region from position 66 to position 264 with 84% amino acid identity between mTRF1 and hTRF1. This conserved middle domain did not reveal previously recognized functional or structural motifs. Both mTRF1 and hTRF1 contained putative nuclear localization signals (NLS) within the least conserved part of the protein (Fig. 4B). Sequence comparison also revealed several putative phosphorylation sites that are noteworthy because they are conserved between the mouse and human proteins (indicated in Fig. 4B).

mTRF1 is ubiquitously expressed in somatic and germline tissues

Northern analysis was carried out to determine the expression pattern of mTRF1. mTRF1 detects a single polyadenylated transcript of 1.9 kb (Fig. 5). This size suggested that the mTRF1 cDNA12-1, which contains an insert of 1.65 kb, was nearly complete. Similar to the expression pattern of hTRF1 mRNA, the mTRF1 transcript was detected in all somatic tissues tested. In addition, mTRF1 was expressed in the testes where the transcript appears slightly larger than in somatic cells. Relative to β-actin and GAPDH mRNA levels, the expression of mTRF1 varied in the different tissues with the highest expression level in testes.

Chromosomal localization of the human and mouse TRF1 genes

Genomic clones containing sequence similarities to hTRF1 and mTRF1 were isolated using the respective cDNAs as probes. To distinguish between any pseudogenes and the expressed locus, sequencing was carried out and intron–exon boundaries were identified in the N-terminal half of the two TRF1 genes (data not shown; GenBank accession numbers U70992, U70993, U70994). Sequence analysis of the mouse genomic clone indicates that, as expected, the intron–exon organization of the two genes has been conserved (data not shown).

To localize the human TRF1 gene, we performed fluorescence *in situ* hybridization (FISH) of a biotin-labeled hTRF1 probe to normal human metaphase chromosomes. Hybridization of this probe resulted in specific labelling only of chromosome 8 (Fig. 6A). Specific labelling of 8q13–21.1 was observed on four (8 cells), three (11 cells) or two (6 cells) chromatids of the chromosome 8 homologues in 25 cells examined. Of 77 signals

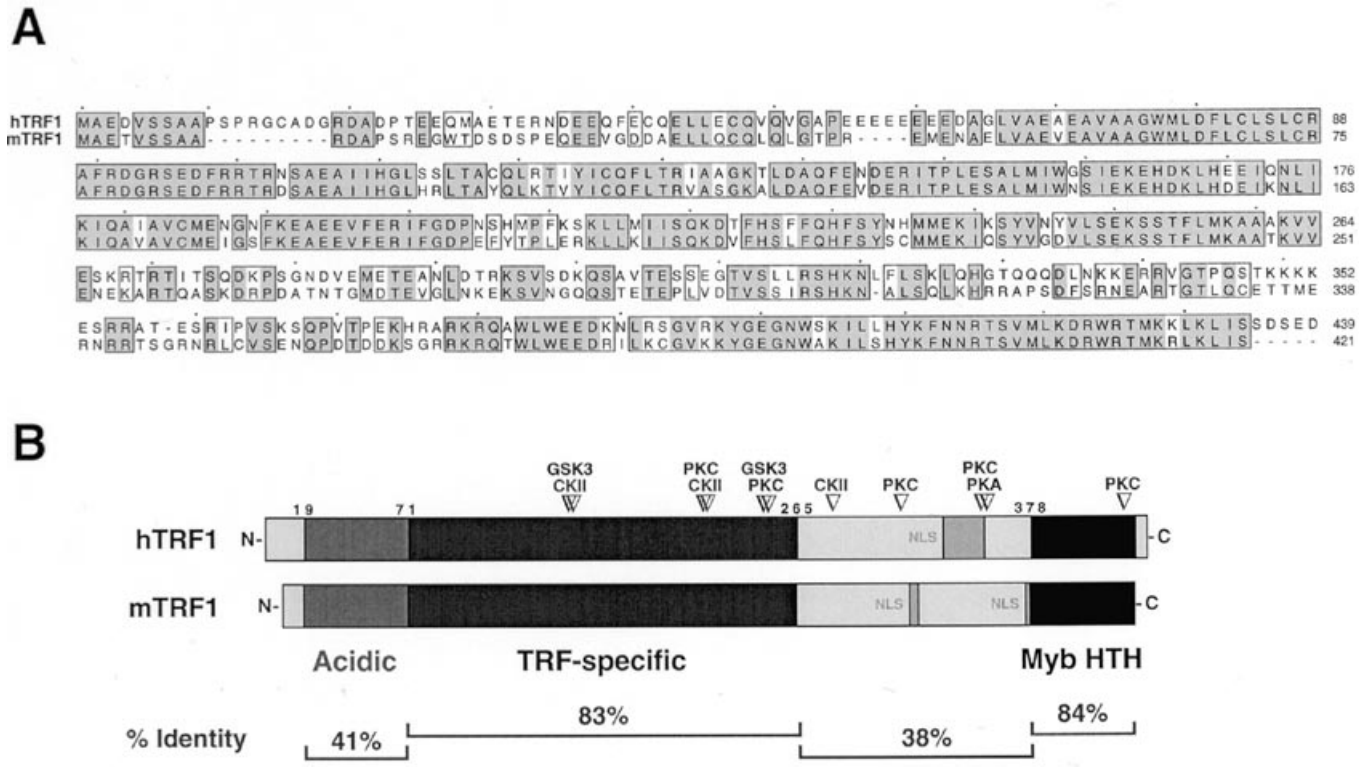

Figure 4. Comparison of the amino acid sequences of mTRF1 and hTRF1. (**A**) The mTRF1 amino acid sequence derived by conceptual translation of the mTRF1 cDNA was aligned to the hTRF1 sequence using Clustal W 1.5 with default settings and the alignment was imported into SeqVu for visual display. Identical amino acids are shaded and homologies (Rao rules) are boxed. (**B**) Schematic of the mTRF1 and hTRF1 protein domains. Three domains are identified: an acidic N-terminal domain (Acidic), a TRF1-specific unique domain (TRF1-specific), and the Myb-type helix–turn–helix region in the C-terminus (Myb HTH). The positions of putative nuclear localization signals are indicated (NLS). The hTRF1 open reading frame contains two overlapping nucleoplasmin type NLS sequences that are indicated with one block. The mTRF1 reading frame contains the putative NLS sequences KHRR (position 311) and RRKR (postition 364). Putative phosphorylation sites that are conserved between mTRF1 and hTRF1 are shown by arrowheads. (GSK3, glycogen synthetase kinase 3; CKII, casein kinase II; PKC, protein kinase C; PKA, protein kinase A). The level of conservation of the domains of TRF1 is indicated as the percentage of the hTRF1 amino acids that is identical to mTRF1.

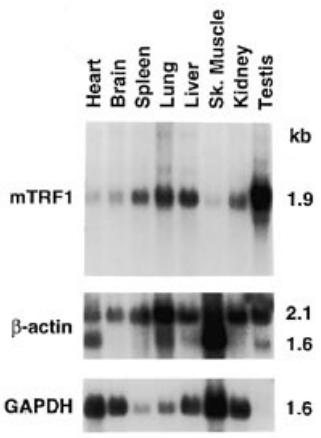

Figure 5. mTRF1 mRNA is expressed in all somatic tissues and in testes. Northern blot of polyadenylated RNA from the indicated tissues was probed for mTRF1, β-actin and GAPDH. The size of the mTRF1 mRNA is ~1.9 kb.

observed, 1 (1%) signal was located at 8q12, 64 (83%) signals were located at 8q13, 6 (8%) signals were located at the junction of 8q13–q21.1, and 6 (8%) signals were located at 8q21.1. These results suggested that the human TRF1 gene is localized to chromosome 8, band q13. This localization was verified by analyzing the segregation of the gene in DNA samples from a panel of human rodent somatic cell hybrids containing different human chromosomes using a PCR strategy (data not shown). The hTRF1 gene locus has not been implicated in cancer or chromosome instability syndromes.

FISH was also used to map the mouse TRF1 gene. Biotinylated DNA from a genomic clone containing a 5′ segment of the mouse TRF1 gene was used in FISH to normal mouse metaphase chromosomes. A single autosomal site of hybridization was observed (Fig. 6B). Preliminary chromosome identification was achieved by chromosome banding (Fig. 6B) and was confirmed by dual *in situ* hybridization of the mouse TRF1 genomic clone and a mouse chromosome 17 specific paint (data not shown). In 33 of 35 metaphases scored using this dual labeling approach, the signal was observed on a single pair of chromosomes and these chromosomes also hybridized to the chromosome 17 paint. No other sites of hybridization were observed in these cases. We conclude that the mouse TRF1 gene probe hybridizes to a single site (band E3) on mouse chromosome 17. No part of mouse chromosome 17 has been reported to be syntenic with human chromosome 8. Markers along the long arm of human chromosome 8 have been mapped to mouse chromosomes 16, 1, 4, 3 and 15 and the 8q13 region appears syntenic with parts of mouse chromosome 3 and 4. (see http://www3.ncbi.nlm.nih/Homology/ and ref. 27).

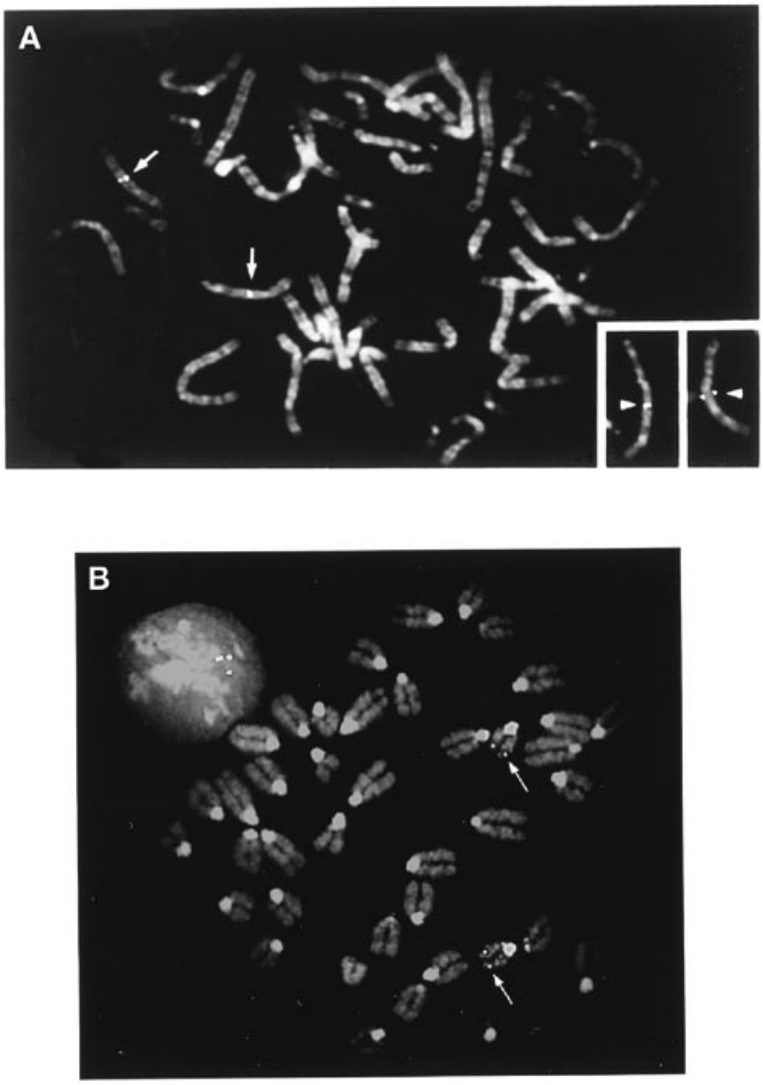

Figure 6. Mapping of the human and mouse TRF1 genes by *in situ* hybridization to metaphase chromosomes. (**A**) Mapping of the hTRF1 gene to human chromosome 8, band q13. The chromosome 8 homologues are identified with arrows. The inset shows partial karyotypes of two chromosome 8 homologues illustrating specific labeling at q13 (arrowheads). (**B**) *In situ* hybridization of biotin-labeled mTRF1 genomic probe to mouse metaphase chromosomes. Specific labelling is observed at band E3 of chromosome 17 (arrows). Images were obtained using a Zeiss Axiophot microscope coupled to a cooled charge coupled device (CCD) camera. Separate images of DAPI stained chromosomes and the hybridization signal were merged.

DISCUSSION

Telomeres function to protect the ends of linear chromosomes via the formation of a nucleoprotein complex composed of telomeric repeat DNA and specific DNA binding protein(s). The telomeric sequence TTAGGG has been conserved throughout vertebrate evolution suggesting that the protein components of this fundamental chromosomal structural element would be conserved as well. Proteins that recognize and bind to double stranded telomeric repeats have been isolated and characterized in the yeast *Saccharomyces cerevisiae* (Rap1p) and in humans (hTRF1). Here we describe mTRF1, the mouse homolog of hTRF1, and provide evidence that while the amino acid sequence of these telomeric proteins is evolving rapidly, their DNA binding domains are conserved.

The mTRF1 and hTRF1 genes show considerable sequence divergence. The overall level of sequence identity between mouse and human TRF1 is 67% and two large domains have >40% identity (coordinates 1–71 and 264–278 in the hTRF1 amino acid sequence). Human and mouse nuclear proteins typically show >80% sequence identity. For example, the amino acid identity levels of human and mouse c-Myb, CENP-B, PCNA, topoisomerase I, histones, TBP and HMG17 ranges from 91–100%, while p53, lamin B2 and HMG14 show 80–85% identity (based on data available in GenBank). Furthermore, none of these proteins show a long (>100 aa) region of low amino acid identity (<40%) as is found in TRF1. Similar to the divergence of human and mouse TRF1, telomeric proteins in yeasts and ciliates show considerable sequence divergence. For instance, the α and β subunits of the telomere terminus proteins from *Oxytricha nova* and *Stylonychia mytilis* share only 77 and 70% identity, respectively (28,29), and the Rap1 protein of *Kluyveromyces lactis* has only 47% of its amino acid sequence in common with its *S.cerevisiae* homolog (30,31). In a testimony to the rapid evolution of yeast Rap1p, the

K.lactis gene cannot complement a Rap1p deficiency in *S.cerevisiae* (31). Thus, it appears that telomeric proteins evolve relatively fast.

Rapid divergence of telomeric proteins would explain why Rap1p, the likely functional analog of TRF1, has very limited sequence similarity to the mammalian telomeric proteins. Even within the Myb homology region mTRF1 and hTRF1 show only 13% sequence identity and 30% similarity to one of the Myb repeats of Rap1p (data not shown). No sequence similarities were noted in the other parts of the protein. For comparison, Mif2, a candidate yeast homolog of the mammalian CENP-C proteins, shows much stronger levels of conservation, having two blocks in common with these proteins in which the sequence identity is 28 and 40% (32,33). Given the high rate of divergence of TRF1 proteins, it will be of interest to examine the TRF1 homologs from other vertebrates.

A high rate of sequence divergence has also been noted for another chromosomal protein, the kinetochore protein CENP-C, which shows only 53% identity between the human and mouse proteins (34,35). However, not all chromosomal proteins evolve rapidly. For example, the centromere-satellite DNA binding protein CENP-B, is highly conserved with 93% of the human and mouse amino acid sequence being identical (36,37).

Comparison with the sequence databases does not reveal convincing homologues of human and mouse TRF1 in other eukaryotes and analysis of the complete yeast genome indicates that *S.cerevisiae* does not have a TRF1-like gene. However, as first noted by Bilaud *et al*. (38), the Myb homology region in TRF1 is closely related to the single Myb domains present in the plant transcription factors IBF1 and BPF1 and to the yeast protein Tbf1p. It is interesting to note that IBF1 and BPF1 carry the Myb homology region in the C-terminus of the protein, as does TRF1 (39,40). Tbf1p, by contrast, has a centrally located Myb domain (41). With knowledge of the conserved features of TRF1, we can now address the question of whether the similarity to these proteins extends outside the Myb homology blocks. Alignment of the conserved residues of mTRF1 and hTRF1 to these three factors did not reveal striking additional homology. For example, the most conserved middle domain of the TRF1 proteins (positions 64–266) aligned to BPF1, IBP1 and Tbf1p revealed identity levels of 11–17%. Whether such low identity levels are significant is not clear at this stage.

The sequence of mTRF1 and hTRF1 revealed several putative phosphorylation sites that are conserved (Fig. 4B). Interestingly, one of these sites (at position 354/355) is placed at the same distance N-terminal of the Myb domain as a regulatory site in the c-Myb oncoproteins (42,43) and a similarly positioned conserved PKC site is found upstream of the second Myb motif in Rap1p (24,31). A second phosphorylation site that is shared between Myb oncoproteins, Rap1 proteins and TRF1 proteins is the PKC site just beyond the third (recognition) helix. While there is evidence that the DNA binding activity of Rap1p is modulated by the phosphorylation state of its DNA binding domain (44), the possible role of phosphorylation in TRF1 function has not been examined so far.

The hTRF1 cDNA sequence contains a $(GAG)_8$ triplet repeat array that codes for a stretch of glutamic acid residues in the acidic domain. Since mTRF1 carries only a single glutamic acid residue at this position, the length of the repeat region is probably not important for TRF1 function. A PCR approach was used to identify possible instability in the hTRF1 triplet repeat array. No change in array size was noted in a survey that included normal peripheral blood and germline DNA from four healthy donors, five breast carcinomas, 11 ovarian carcinomas, and 18 hereditary non-polyposis colon carcinoma samples with known microsatellite instability (S.O., D.B., R. Darnell, M. Dunlop, L. van't Veer, and T.d.L., unpublished observations). Thus, the polyglutamic acid triplet repeat array in the hTRF1 gene does not appear to be a target for microsatellite instability in human cancer.

Although the human and mouse TRF1 proteins show a high degree of amino acid divergence, biochemically they appear nearly indistinguishable. Both proteins are of approximately the same size (55–60 kDa), bind to TTAGGG repeat DNA with the same sequence preference, and form gel-shift complexes that migrate similarly (data not shown). The proteins share three conserved domains: they have nearly identical regions of homology to the Myb oncoproteins in an area of TRF1 that is implicated in DNA binding. In addition, they share a highly acidic N-terminus and a unique middle domain. Most likely these two regions are involved in protein–protein interactions. Thus, at least one component of the mammalian telomeric complex, TRF1, is functionally conserved.

Both hTRF1 and mTRF1 mRNAs are expressed in all normal somatic tissues, in cell lines, and in mouse mammary tumors (19; this work; D.B. and T.d.L., unpublished observations). This ubiquitous expression pattern is consistent with a role in chromosome maintenance. The highest level of expression of mTRF1 mRNA was found in the testes. This is intriguing given the high level of expression of telomerase in mammalian testes (45,46; D.B. and T.d.L., unpublished observations) and speculations on the potential role of telomeres during prophase of meiosis I in mediating chromosome movements and the onset of synapsis (reviewed in 47). The availability of the mouse TRF1 gene will now allow us to assess the somatic and germline phenotypes of TRF1-deficient mice.

MATERIALS AND METHODS

Isolation and sequence analysis of mouse cDNAs

A mouse brain cDNA library in the λZAPII vector (Stratagene) was screened with a fragment that spans the N-terminal 800 bp of the human TRF1 open reading frame. This fragment was excised from a low melting point agarose gel, and labelled by random priming using both [α-^{32}P]dCTP and [α-^{32}P]dGTP. Hybridizations were carried out at 55°C essentially as described by Church and Gilbert (48) except that the hybridization solution contained 100 μg of sonicated, denatured HeLa DNA. The filters were washed in 1× SSC/0.1% SDS at 55°C. Phage were converted into plasmid clones following protocols provided by the manufacturer. Plasmid DNA was isolated and sequenced following standard procedures using Sequenase v.2.0. Comparison of the mTRF1 sequence to hTRF1 was executed using Clustal W 1.5 with default settings (gap opening penalty of 10.0 and gap extension penalty of 0.05). The same settings were used to compare the mTRF1 and hTRF1 sequences to IBP1, BPF1 and Tbf1p. Database searches (GenBank) were executed using the NCBI tblastn program on June 13, 1996. Outside the Myb homology region, mTRF1 matched hTRF1 but no other sequences present in the databases (probability score <0.05). The mTRF1 GenBank accession number is U65586.

Electrophoretic mobility shift assays

The mouse myeloma cell line J558 (ATCC TIB6) was maintained in DMEM supplemented with 10% bovine calf serum, antibiotics

and glutamine. Nuclear extract was prepared as described by Zhong *et al*. (20). The *Eco*RI–*Xho*I fragment of mTRF1, representing the entire ORF, was subcloned into pET28 vector and used to produce protein in a coupled *in vitro* transcription-translation reaction (Promega). Preparation of the $[TTAGGG]_{12}$ gel-shift probe, binding reactions, detection of protein–DNA complexes, and competition experiments were carried out as described (19,20,25).

Immunofluorescence

The mTRF1 reading frame was fused to the Met-Asp-Tyr-Lys-Asp-Asp-Asp-Asp-Lys (FLAG) epitope sequences at the N-terminus and inserted in pRc/CMV (InVitrogen). Murine NIH 3T3 fibroblasts were transfected by electroporation and processed for metaphase preparation 24 h post-transfection as described (19). Immunofluorescence using Mab M2 (Eastman Kodak) directed against the FLAG epitope was as previously described by Chong *et al*. (19). Images were obtained using a Zeiss Axioplan microscope and a Kodak DCS-200 digital camera. Images were filtered with a 3×3 medium filter and merged in Adobe Photoshop. Chromosomes were stained with DAPI and converted to false-red to aid photographic reproduction.

Northern analysis

A membrane containing size-fractionated polyadenylated RNA from a variety of mouse tissues (Clontech) was hybridized sequentially with the mTRF1, β-actin, and GAPDH DNA probes. All probes were labelled by random priming using both [α-^{32}P]dCTP and [α-^{32}P]dGTP. Hybridizations were carried out at 65°C as described (19) with washing in 0.3× SSC/0.1% SDS at 65°C for 30 min. Probes were removed between successive hybridizations by incubation for 15 min in 2 mM Na_2PO_4, pH 7.2/1.5 mM NaCl/0.5% SDS at 95°C.

Isolation of genomic clones

DNA probes spanning the entire ORFs of mTRF1 and hTRF1 were used to screen mouse (Sv/129, Stratagene) or human (HT1080, Stratagene) genomic libraries, respectively. The probes were labelled by random priming using [α-^{32}P]dCTP and [α-^{32}P]dGTP. Hybridizations were carried out at 65°C as described above for the isolation of mTRF1 cDNA. The library filters were washed at 0.3× SSC/0.1% SDS at 65°C. DNA was prepared from liquid lysates of purified lambda phages. Restriction enzyme mapping was used to identify candidate phage (phage 14-2 for the mouse and phage 11-1 for the human gene) containing the expressed genes for mTRF1 and hTRF1, respectively. The *Not*I inserts from purified lambda phages were subcloned into pBluescript SK$^-$ for sequencing (see above) and FISH analysis. The GenBank accession numbers for the genomic sequences are U70992, U70993, U70994.

FISH to human and mouse chromosomes

Human metaphase cells were prepared from phytohemagglutinin-stimulated peripheral blood lymphocytes. The hTRF1 probe was a plasmid subclone of the 17.5 kb insert in a genomic lambda clone (clone p11-1), isolated from a Stratagene HT1080 library. FISH was performed as previously described (49). Biotin-labelled probes were prepared by nick-translation using biotin-11-dUTP (Enzo Diagnostics). Hybridization was detected with fluorescein-conjugated avidin (Vector Laboratories), and chromosomes were identified by staining with 4,6-diamidino-2-phenylindole-dihydrochloride (DAPI).

Metaphase chromosomes from the mouse CGR-ES cell line were prepared by methanol:acetic acid fixation (50). A plasmid carrying the *Not*I fragment from a phage containing part of the mouse TRF1 gene (plasmid m14-19) was nick translated in the presence of biotin-16-dUTP (Boehringer Mannheim) and hybridized to mouse metaphase chromosomes in the presence of salmon sperm DNA and mouse C_0t-1 DNA (Gibco-BRL) as described (50). Hybridization was detected as described for FISH on human metaphase chromosomes (above). Initial chromosome identification was made by enhancement of the DAPI counterstain and confirmed by dual *in situ* hybridization of probe m14-19 with a chromosome 17 specific paint (Cambio) following the manufacturer's instructions.

Chromosomal mapping using a somatic cell hybrid panel

The chromosomal localization of hTRF1 was determined using the NIGMS human/rodent somatic cell hybrid panel #2 (Coriell Institute for Medical Research, Camden, NJ) and PCR with oligonucleotide primers T1-F (5′-CGAGCCATTTAACATGGCGGA-3′) and T3-R (5′-TCGGCCACCAGGCCCGCG-3′). The PCR was performed in a 25 µl reaction mixture containing 50–100 ng of genomic hybrid DNA in 50 mM KCl, 10 mM Tris-HCl pH 8.3, 1.5 mM $MgCl_2$, 10 pmol of each primer, 100 µM of each dNTP (dATP, dCTP, dGTP, dTTP), and 1.5 U of *Taq* polymerase (AmpliTaq, Perkin Elmer Cetus). DNA was denatured at 94°C for 5 min, followed by 35 cycles of denaturation at 94°C for 30 s, annealing at 60°C for 1 min, extension at 72°C for 1 min and a final extension step of 10 min. The PCR was carried out in a GeneAmp 9600 PCR system and PCR products were analyzed on a 2% 1× TBE agarose gel.

ACKNOWLEDGEMENTS

We are grateful to Markus Stoffel (Rockefeller University) for initiating the genomic mapping studies on the hTRF1 gene. Robert Darnell (Rockefeller University) is thanked for providing a mouse brain cDNA library. Art Lustig (Memorial Sloan Kettering Cancer Center) and Richard Lang (New York University Medical Center) are thanked for comments on this manuscript. D.B. is the recipient of a Merck fellowship. B.v.S. is the recipient of a Human Frontiers Science Program fellowship. D.K. is a Fellow of the Lister Institute of Preventive Medicine. N.M. was supported by the United Kingdom Medical Research Council. This work was supported by NIH PHS grants to T.d.L. (GM49046) and M.M.L-B. (CA67021) and by a grant from the Lucille Markey Trust to T.d.L.

ABBREVIATIONS

FISH, fluorescent *in situ* hybridization; ORF, open reading frame; PCR, polymerase chain reaction; SSC, citrate buffered saline.

REFERENCES

1. Blackburn, E.H. (1991) Telomeres. *Trends Biochem. Sci.*, **16**, 378–381.
2. Zakian, V.A. (1996) Telomere functions: lessons from yeast. *Trends Cell Biol.*, **6**, 29–33.

3. 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) A highly conserved repetitive DNA sequence, (TTAGGG)n, present at the telomeres of human chromosomes. *Proc. Natl. Acad. Sci. USA*, **85**, 6622–6626.
4. de Lange, T., Shiue, L., Myers, R.M., Cox, D.R., Naylor, S.L., Killery, A.M. and Varmus, H.E. (1990) Structure and variability of human chromosome ends. *Mol. Cell. Biol*, **10**, 518–527.
5. Kipling, D. and Cooke, H.J. (1990) Hypervariable ultra-long telomeres in mice. *Nature*, **347**, 347–402.
6. Blackburn, E.H. (1993) In Gesteland and Atkins (eds), *The RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. pp 557–576.
7. Greider, C.W. (1995) In Blackburn and Greider (eds), *Telomeres*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. pp 35–69.
8. Sandell, L. and Zakian, V. (1993) Loss of a yeast telomere: arrest, recovery, and chromosome loss. *Cell*, **75**, 729–741.
9. Fang, G. and Cech, T.R. (1995) In Blackburn and Greider (eds), *Telomeres*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. pp 69–106.
10. de Lange, T. (1996) In search of vertebrate telomeric proteins. *Seminars Cell Biol.*, **7**, 23–29.
11. Lustig, A.J., Kurtz, S. and Shore, D. (1990) Involvement of the silencer and UAS binding protein RAP1 in regulation of telomere length. *Science*, **250**, 549–553.
12. Conrad, M.N., Wright, J.H., Wolf, A.J. and Zakian, V.A. (1990) RAP1 protein interacts with yeast telomeres in vivo: Overproduction alters telomere structure and decreases chromosome stability. *Cell*, **63**, 739–750.
13. Kyrion, G., Boakye, K.A. and Lustig, A.J. (1992) C-terminal truncation of RAP1 results in the deregulation of telomere size, stability, and function in *Saccharomyces cerevisiae. Mol. Cell. Biol.*, **12**, 5159–5173.
14. McEachern, M.J. and Blackburn, E.H. (1995) Runaway telomere elongation caused by telomerase RNA gene mutations. *Nature*, **376**, 403–409.
15. de Lange, T. (1992) Human telomeres are attached to the nuclear matrix. *EMBO J.*, **11**, 717–724.
16. Ludérus, M.E.E., van Steensel, B., Chong, L., Sibon, O.C.M., Cremers, F.F.M. and de Lange, T. (1996) Structure, subnuclear distribution, and nuclear matrix association of the mammalian telomeric complex. *J. Cell. Biol.*, in press.
17. Tommerup, H., Dousmanis, A. and de Lange, T. (1994) Unusual chromatin in human telomeres. *Mol. Cell. Biol.*, **14**, 5777–5785.
18. Lejnine, S., Makarov, V.L. and Langmore, J.P. (1995) Conserved nucleoprotein structure at the ends of vertebrate and invertebrate chromosomes. *Proc. Natl. Acad. Sci. USA*, **92**, 2393–2397.
19. Chong, L., van Steensel, B., Broccoli, D., Erdjument-Bromage, H., Hanish, J., Tempst, P. and de Lange, T. (1995) A human telomeric protein. *Science*, **270**, 1663–1667.
20. Zhong, Z., Shiue, L., Kaplan, S. and de Lange, T. (1992) A mammalian factor that binds telomeric TTAGGG repeats in vitro. *Mol. Cell. Biol.*, **13**, 4834–4843.
21. Shore, D. (1994) RAP1: a protean regulator in yeast. *Trends Genet.*, **10**, 408–412.
22. Zakian, V.A. (1995) In Blackburn and Greider (eds), *Telomeres*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. pp 107–138.
23. Wright, J.H., Gottschling, D.E. and Zakian, V.A. (1992) Saccharomyces telomeres assume a non-nucleosomal chromatin structure. *Genes Dev.*, **6**, 197–210.
24. Konig, P., Giraldo, R., Chapman, L. and Rhodes, D. (1996) The crystal structure of the DNA-binding domain of yeast RAP1 in complex with telomeric DNA. *Cell*, **85**, 125–136.
25. Hanish, J.P., Yanowitz, J. and de Lange, T. (1994) Stringent sequence requirements for telomere formation in human cells. *Proc. Natl. Acad. Sci. USA*, **91**, 8861–8865.
26. Ogata, K., Morikawa, S., Nakamura, H., Sekikawa, A., Inoue, T., Kanai, H., Sarai, A., Ishii, S. and Nishimura, Y. (1994) Solution structure of a specific DNA complex of the myb DNA-binding domain with cooperative recognition helixes. *Cell*, **79**, 639–648.
27. DeBry, R.W. and Seldin, M.F. (1996) Human/mouse homology relationships. *Genomics*, **33**, 337–344.
28. Gray, J.T., Celander, D.W., Price, C.M. and Cech, T.R. (1991) Cloning and expression of genes for the Oxytricha telomere-binding protein: Specific subunit interactions in the telomeric complex. *Cell*, **67**, 807–814.
29. Fang, G. and Cech, T.R. (1991) Molecular cloning of telomere-binding protein genes from Stylonychia mytilis. *Nucleic Acids Res.*, **19**, 5515–5518.
30. Shore, D. and Nasmyth, K. (1987) Purification and cloning of a DNA binding protein from yeast that binds to both silencer and activator elements. *Cell*, **51**, 721–732.
31. Larson, G.P., Castanotto, D., Rossi, J.J. and Malafa, M.P. (1994) Isolation and functional analysis of a Kluyveromyces lactis RAP1 homologue. *Gene*, **150**, 35–41.
32. Brown, M.T. (1995) Sequence similarities between the yeast chromosome segregation protein Mif2 and the mammalian centromere protein CENP-C. *Gene*, **160**, 111–116.
33. Meluh, P.B. and Koshland, D. (1995) Evidence that the MIF2 gene of Saccharomyces cerevisiae encodes a contromere protein with homology to the mammalian centromere protein CENP-C. *Mol. Biol. Cell*, **6**, 793–807.
34. McKay, S., Thomson, E. and Cooke, H. (1994) Sequence homologies and linkage group conservation of the human and mouse Cenpc genes. *Genomics*, **22**, 36–40.
35. Saitoh, H., Tomkiel, J., Cooke, C.A., III, H.R., Maurer, M., Rothfield, N.F. and Earnshaw, W.C. (1992) CENP-C, an autoantigen in scleroderma, is a component of the human inner kinetochore plate. *Cell*, **70**, 115–125.
36. Earnshaw, W.C., Sullivan, K.F., Machlin, P.S., Cooke, C.A., Kaiser, D.A., Pollard, T.D., Rothfield, N.F. and Cleveland, D.W. (1987) Molecular cloning of cDNA for CENP-B, the major human centromere autoantigen. *J. Cell Biol.*, **104**, 817–826.
37. Sullivan, K.F. and Glass, C.A. (1991) CENP-B is a highly conserved mammalian centromere protein with homology to the helix-loop-helix family of proteins. *Chromosoma*, **100**, 360–370.
38. Bilaud, T., Koering, C.E., Binet-Brasselet, E., Ancelin, K., Pollice, A., Gasser, S.M. and Gilson, E. (1996) The telobox, a myb-related telomeric DNA binding motif found in proteins from yeast, plants, and human. *Nucleic Acids Res.*, **24**, 1294–1303.
39. da Costa e Silva, O., Klein, L., Schmelzer, E., Trezzini, G.F. and Hahlbrock, K. (1993) BPF-1, a pathogen-induced DNA-binding protein involved in the plant defense response. *Plant J.*, **4**, 125–135.
40. Lugert, T. and Werr, W. (1994) A novel DNA-binding domain in the Shrunken initiator-binding protein (IBP1). *Plant Mol. Biol.*, **25**, 493–506.
41. Brigati, C., Kurtz, S., Balderes, D., Vidali, G. and Shore, D. (1993) An essential yeast gene encoding a TTAGGG repeat-binding protein. *Mol. Cell. Biol.*, **13**, 1306–14.
42. Lüscher, B. and Eisenman, R.N. (1990) New light on Myc and Myb. Part I. Myc. *Genes Dev.*, **4**, 2025–2035.
43. Oelgeschlager, M., Krieg, J., Luscher-Firzlaff, J.M. and Luscher, B. (1995) Casein kinase II phosphorylation site mutations in c-Myb affect DNA binding and transcriptional cooperativity with NF-M. *Mol. Cell. Biol.*, **15**, 5966–5974.
44. Tsang, J.S.H., Henry, Y.A.L., Chambers, A., Kingsman, A.J. and Kingsman, S.M. (1990) Phosphorylation influences the binding of the yeast RAP1 protein to the upstream activating sequence of the PGK gene. *Nucleic Acids Res.*, **18**, 7331–7337.
45. Kim, N.W., Piatyszek, M.A., Prowse, K.R., Harley, C.B., West, M.D., Ho, P.L.C., Coviello, G.M., Wright, W.E., Weinrich, S.L. and Shay, J.W. (1994) Specific association of human telomerase activity with immortal cells and cancer. *Science*, **266**, 2011–2015.
46. Prowse, K. and Greider, C.W. (1995) Developmental and tissue specific regulation of mouse telomerase and telomere length. *Proc. Natl. Acad. Sci. USA*, **92**, 4818–4822.
47. Dernburg, A.F., Sedat, J.W., Cande, W.Z. and Bass, H.W. (1995) In Blackburn and Greider (eds), *Telomeres*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. pp 295–338.
48. Church, G.M. and Gilbert, W. (1984) Genomic sequencing. *Proc. Natl. Acad. Sci. USA*, **81**, 1991–1995.
49. Rowley, J.D., Diaz, M.O., Espinosa, R., Patel, Y.D., Melle, E.v., Ziemin, S., Taillon-Miller, P., Lichter, P., Evans, G.A., Kersey, J.D., Ward, D.C., Domer, P.H. and Beau, M.M.L. (1990) Mapping chromosome band 11q23 in human acute leukemia with biotinylated probes: Identification of 11q23 translocation breakpoints with a yeast artificial chromosome. *Proc. Natl. Acad. Sci. USA*, **87**, 9358–9362.
50. Fantes, J.A., Bickmore, W.A., Fletcher, J.M., Bellestra, F., Hanson, I.M. and Heyningen, V.v. (1992) Submicroscopic deletions at the WAGR locus, revealed by nonradioactive in situ hybridization. *Am. J. Hum. Genet.*, **51**, 1286–1294.