hours [R. S. Sikorski and J. D. Boeke, in *Guide to Yeast Genetics and Molecular Biology*, C. Guthrie and G. R. Fink, Eds. (Academic Press, San Diego, CA, 1991), vol. 194, pp. 302–318]. Competent cells of K^+ uptake–deficient yeast (*14*) were transformed with the mutagenized HKT1 library. The transformants were eluted and plated in arginine phosphate medium (*28*) containing 0.1 mM K^+ and 500 mM Na^+, which inhibited the growth of the strain expressing wild-type HKT1. Growing colonies were selected, and their plasmids were isolated and reintroduced into yeast to retest for growth in the presence of high [Na^+]. Ten of 60 plasmids conferred the ability to grow at high [Na^+]. The *HKT1* cDNAs of these 10 plasmids were sequenced to identify mutations. The cDNA of the HKT1 mutants was subcloned under the control of the yeast *PMA1* gene promoter, as described (*13*).

21. F. Rubio, W. Gassmann, J. I. Schroeder, data not shown.
22. Abbreviations for the amino acid residues are as follows: A, Ala; F, Phe; L, Leu; and V, Val.
23. For example, at 0.1 mM K^+ and 500 mM Na^+, intracellular K^+ + Na^+ contents were 301 ± 35 nmol mg^{-1} (HKT1), 314 ± 20 nmol mg^{-1} (A240V), and 295 ± 13 nmol mg^{-1} (L247F).
24. W. Gassmann, F. Rubio, J. I. Schroeder, data not shown.
25. Uptake experiments in yeast, in which the background buffer [Na^+] was 11 μM (see Fig. 1), showed that the maximal velocity (V_{max}) for Na^+ uptake at [Na^+] ≥ 1 mM (17 nmol mg^{-1} min^{-1}) was nine times that of the V_{max} for Rb^+ uptake at [Rb^+] ≥ 1 mM (1.8 nmol mg^{-1} min^{-1}).
26. M. R. Sussman, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **45**, 211 (1994).
27. F. A. Smith and N. A. Walker, *J. Membr. Biol.* **108**, 125 (1989); N. A. Walker and D. Sanders, *Planta (Heidelberg)* **185**, 443 (1991).
28. A. Rodríguez-Navarro and J. Ramos, *J. Bacteriol.* **159**, 940 (1984).
29. We thank R. F. Gaber and A. Rodríguez-Navarro for providing yeast strains, W. B. Frommer for providing the plasmid pDR195 and A. Rodríguez-Navarro, N. M. Crawford, A. Schinder, and J. M. Ward for comments. Supported by U.S. Department of Agriculture (95-37304-2227) and Department of Energy (DE-FG03-34-ER20148) grants and a Presidential Young Investigator Award (to J.I.S.) and a Ministerio de Educación y Ciencia (Spain) postdoctoral fellowship (to F.R.).

14 June 1995; accepted 24 October 1995

A Human Telomeric Protein

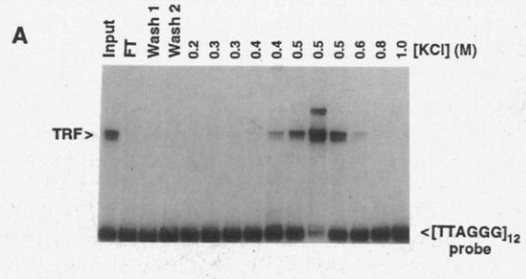
Laura Chong, Bas van Steensel, Dominique Broccoli, Hediye Erdjument-Bromage, John Hanish,* Paul Tempst, Titia de Lange†

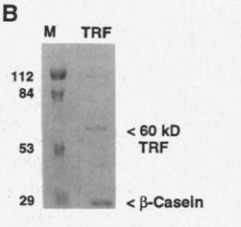
Telomeres are multifunctional elements that shield chromosome ends from degradation and end-to-end fusions, prevent activation of DNA damage checkpoints, and modulate the maintenance of telomeric DNA by telomerase. A major protein component of human telomeres has been identified and cloned. This factor, TRF, contains one Myb-type DNA-binding repeat and an amino-terminal acidic domain. Immunofluorescent labeling shows that TRF specifically colocalizes with telomeric DNA in human interphase cells and is located at chromosome ends during metaphase. The presence of TRF along the telomeric TTAGGG repeat array demonstrates that human telomeres form a specialized nucleoprotein complex.

Human chromosomes carry a long terminal array of double-stranded TTAGGG hexamers that are maintained by telomerase. Telomeric DNA is thought to form a protective nucleoprotein cap through its association with telomere-specific proteins (*1*). Because the loss of telomere function can induce cell cycle arrest and genome instability, the telomeric complex is likely to be required in all human cells. Changes in the structure and function of human telomeres are thought to play a role in malignant transformation and cellular senescence (*2*, *3*).

Protein components of the telomeric complex have been identified in ciliates and in yeast, but not in vertebrate systems (*1*). Quests for vertebrate telomeric proteins had previously yielded a single candidate factor that could potentially bind along the length of the telomeric TTAGGG repeat array (*4*–*6*). This protein, TRF (telomeric repeat binding factor), associates with double-stranded TTAGGG repeat arrays in vitro and displays strong specificity for vertebrate telomeric DNA (*4*, *5*). TRF does not bind to single-stranded telomeric sequences and does not require the proximity of a DNA terminus for its interaction (*4*). The activity is expressed in nuclei from human, monkey, rodent, and chicken cells, which all carry TTAGGG repeat arrays at their chromosome ends (*4*). Here, we show that TRF is a protein component of human telomeres.

Human TRF (hTRF) activity can be detected in HeLa cell nuclear extracts on





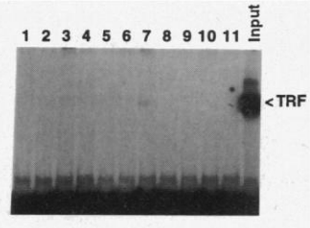


Fig. 1. Purification and identification of the 60-kD hTRF protein. (**A**) Specific DNA affinity chromatography of hTRF. Partially purified HeLa hTRF was applied to a column containing restriction fragments with the sequence $[TTAGGG]_{27}$ coupled to Streptavidin-agarose (*7*). Input, flow-through (FT), and the indicated KCl fractions were assayed for hTRF binding activity with the use of a $[TTAGGG]_{12}$ gel-shift probe. (**B**) Coomassie blue staining pattern of purified hTRF. The 60-kD TRF band is indicated. β-Casein was added to enhance hTRF activity in purified preparations (*7*). The asterisk at the right indicates a ~100-kD protein that is present in some of the hTRF preparations. Marker proteins (M) were prestained. (**C**) Recovery of hTRF activity by elution of the 60-kD protein from SDS-PAGE. Proteins from a gel similar to the one shown in (B) were eluted (*5*, *9*), and hTRF activity was assayed by gel shift with a $[TTAGGG]_{12}$ probe. Lanes 1 to 11 contain proteins isolated from successive gel slices covering the 120- to 20-kD range. Lane 7 contains proteins from the 55- to 65-kD range. (**D**) Analysis of hTRF tryptic peptides by chemical sequencing and laser-desorption mass spectrometry (*10*). Amino acids in lowercase were tentatively assigned; "x" indicates that no identification could be made (*27*). IY indicates calculated initial sequencing yields; *m/z* is the experimental mass of the peptide. [MH^+] denotes the theoretical average isotopic mass of the peptide (plus one proton), calculated from the cDNA-derived sequence (Fig. 3C). M_{ox} refers to methionine sulfoxide (singly oxidized methionine).

D

Peptide	Sequence	IY (pmol)	*m/z*	[MH^+] TRF
T7	EAEEVFEr	1.7	1009.2	1009.06
T8	TLDAqFENdEr	1.4	1337.9	1338.38
T10	TITsQDKPxxNxVxM	1.4	2672.0	2668.80 (M_{ox})
T11	ILLxYK	1.7	–	
T12	IqAIAVxm	0.65	1658.7	1659.80
T13	IFgDPNxxmpf	1.3	1408.5	1406.60 (M_{ox})
T20	xYVNYVLxEK	1.5	1201.4	1202.35
T26	QAxLxEEDK	1.1	–	
T29	TIYICQFTr	1.1	1363.0	1363.53

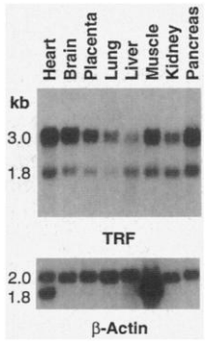


Fig. 2. Expression of hTRF mRNA in human tissues. The figure shows a Northern blot of polyadenylated RNAs (~2 μg per lane, Clontech) derived from the indicated tissues probed with a hTRF cDNA fragment encompassing most of the protein coding region. Human TRF is encoded by the 3-kb mRNA. The sequence of the 1.8-kb mRNA has not been determined. The bottom panel shows rehybridization of the same blot to a β-actin probe, which detects a 2-kb mRNA in all tissues and the muscle-specific 1.8-kb mRNA in heart and skeletal muscle.

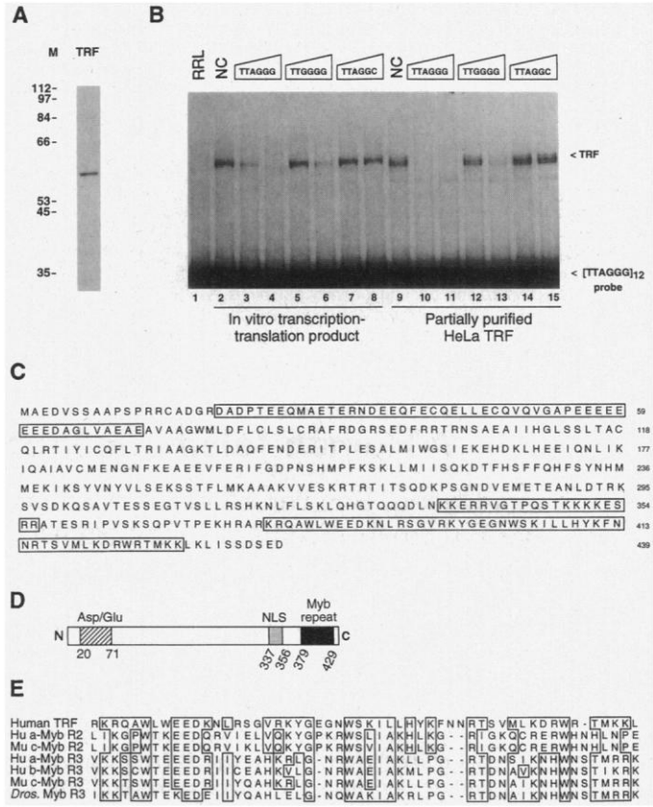


Fig. 3. Human TRF cDNA encodes a 60-kD telomeric DNA-binding protein with a Myb-type DNA-binding repeat and an acidic NH_2-terminal domain. (**A**) In vitro expression of the hTRF ORF yields a protein that migrates at ~60 kD. Shown is the [^{35}S]methionine-labeled in vitro transcription-translation product obtained with a reconstituted HeLa cDNA containing the complete TRF ORF. The position of molecular mass standards (M) is indicated. (**B**) In vitro–expressed hTRF binds telomeric DNA with the same sequence specificity as does HeLa hTRF. Unlabeled hTRF protein was produced in vitro [as in (A)] and assayed for binding to a $[TTAGGG]_{12}$ probe (lanes 2 to 8). Partially purified HeLa hTRF was assayed in parallel (lanes 9 to 15). Lane 1 contains a reaction with rabbit reticulocyte lysate (RRL) alone. Lanes 2 and 9 contain reactions without addition of competitor DNAs (NC). Lanes 3 to 8 and 10 to 15 contain binding reactions in the presence of 20 and 200 ng of circular plasmid DNAs containing stretches of 1 to 2 kb of tandemly repeated hexamers of the indicated sequences (*5*). (**C**) Conceptional translation of the ORF in hTRF cDNA (*11, 27*). The NH_2-terminal acidic region, two overlapping potential localization signals, and the region of homology to the Myb-type DNA-binding repeat are boxed. (**D**) Domain structure of hTRF (NLS, nuclear localization signal). (**E**) Alignment of hTRF to the Myb-type DNA-binding repeats of mammalian [Hu, human; Mu, mouse (murine)] and *Drosophila myb* proto-oncogenes (*11, 27*).

the basis of its ability to alter the mobility of a double-stranded DNA fragment containing the sequence $[TTAGGG]_{12}$ (*4*). Using this assay, we purified HeLa hTRF to near homogeneity by ion-exchange chromatography, by elution from columns containing nonspecific *Escherichia coli* DNA, and by fractionation over specific telomeric DNA resins (*7*) (Fig. 1A). Three independent preparations of purified hTRF contained a protein in the 60-kD apparent molecular mass range (Fig. 1B), which copurified with hTRF activity over a column containing double-stranded TTAGGG repeats (*8*). A ~100-kD protein was present in some but not all purified hTRF preparations (Fig. 1B) (*8*). Elution of the 60-kD protein from SDS–polyacrylamide gel electrophoresis (PAGE) (*9*) resulted in partial recovery of hTRF activity (Fig. 1C, lane 7), which indicates that a 60-kD polypeptide is sufficient for the formation of the hTRF complex with TTAGGG repeat probes. Amino acid sequences were obtained for tryptic peptides derived from the 60-kD band (*10*) (Fig. 1D), one of which (T29) revealed sequence identity to an anonymous partial complementary DNA (cDNA) sequence in the GenBank database. On the basis of this nucleotide sequence, cDNAs were isolated from a HeLa cell library, sequenced, and found to contain an open reading frame (ORF) encoding all sequenced peptides (*11*), as discussed below.

The hTRF cDNA hybridizes to two mRNAs of ~1.8 and ~3.0 kb that are expressed in a variety of human tissues (Fig. 2). Sequence analysis of three overlapping hTRF cDNAs derived from the larger mRNA (*11*) revealed an ORF encoding a 439–amino acid protein (Fig. 3). The predicted molecular mass of this protein is 50,341 daltons, which is 10 kD smaller than the apparent molecular mass of purified HeLa hTRF. In vitro transcription and translation of the cloned cDNA produced a protein of the same size as purified HeLa hTRF (60 kD) (Fig. 3A), which indicated anomalous migration during SDS-PAGE. To verify that the cloned gene represented hTRF, we used in vitro–expressed protein in mobility-shift assays with a $[TTAGGG]_{12}$ probe. The in vitro–expressed protein formed a complex with the telomeric DNA probe that comigrated with the largest of three closely migrating gel-shift complexes formed with HeLa hTRF (Fig. 3B). The two smaller hTRF complexes detected in HeLa extracts were never observed with in vitro–expressed hTRF. Whether the additional HeLa hTRF complexes resulted from modification in vivo or from alteration of the hTRF protein during isolation (for instance, by partial proteolytic degradation or dephosphorylation) is not known. Competition experiments with cloned telomeric DNAs in circular plasmids showed that the cloned hTRF protein and HeLa hTRF have the same sequence specificity and, as expected (*4*), do not require a DNA end for binding (Fig. 3B). For both the HeLa and cloned hTRF activities, the strongest competition was observed with a plasmid containing a TTAGGG repeat array; a plasmid with TTGGGG repeats competed to a lesser extent, and no competition occurred with a plasmid with TTAGGC repeats (Fig. 3B). These results demonstrated that the cloned cDNA encoded hTRF.

Comparison with the sequence information in the databases indicated that hTRF is a novel protein with three previously recognized sequence motifs (Fig. 3, C and D). hTRF contains two overlapping nucleoplasmin-type nuclear localization signals around position 350 and an NH_2-terminal region that is rich in aspartic and glutamic acid residues (*11*). The acidic domain of hTRF extends over a segment of 52 amino acids with 44% acidic residues and a calculated isoelectric point (pI) of 3.0. Although clus-

L. Chong, B. van Steensel, D. Broccoli, J. Hanish, T. de Lange, Laboratory for Cell Biology and Genetics, Rockefeller University, 1230 York Avenue, New York, NY 10021, USA.
H. Erdjument-Bromage and P. Tempst, Molecular Biology Program, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10021, USA.
*Present address: Laboratory of Chromosome Structure, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10021, USA.
†To whom correspondence should be addressed. E-mail: delange@rockvax.rockefeller.edu

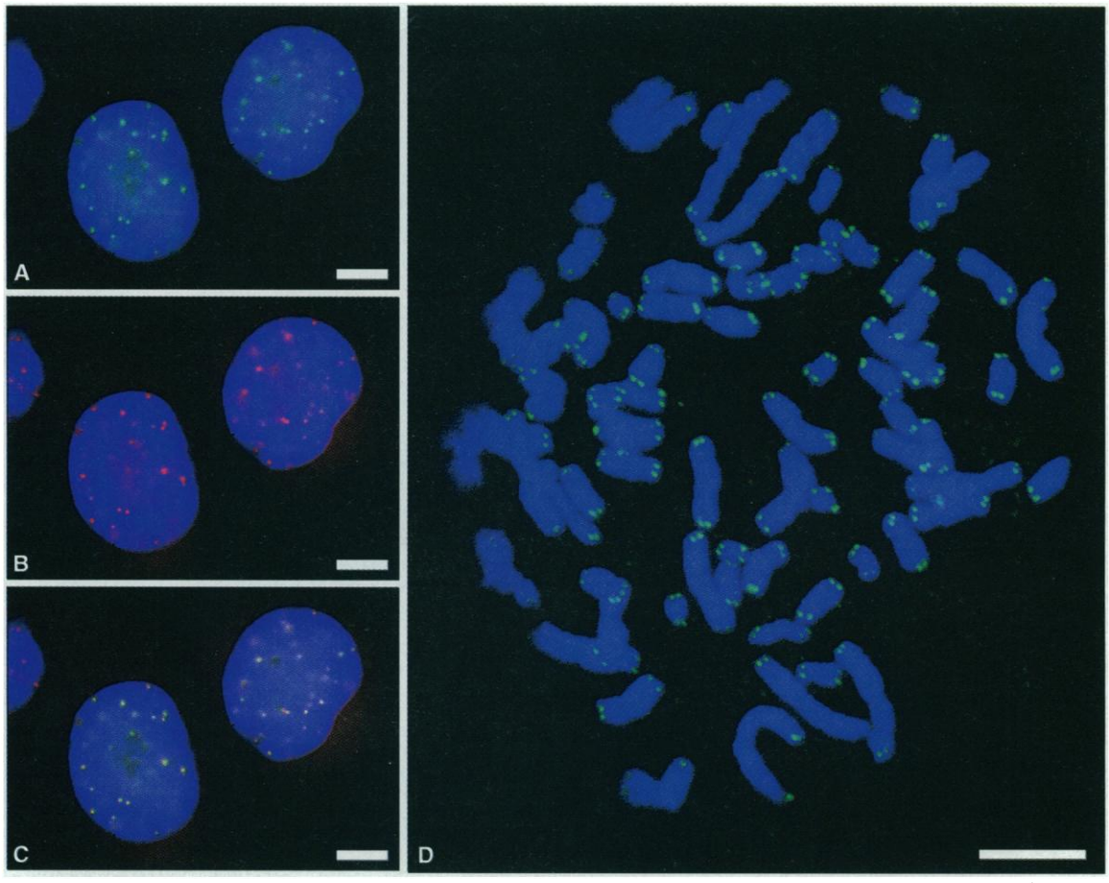


Fig. 4. Telomeric localization of TRF in human cells. (**A** through **C**) Colocalization of a FLAG epitope–tagged mTRF protein with telomeric DNA in interphase (*15, 16*). In (A), HeLa cells were transiently transfected with FLAG-mTRF expressed from the CMV promoter and labeled with an anti-FLAG mAb (M5) and FITC-conjugated donkey antibody to mouse IgG (green). In (B), telomeric DNA was visualized in the same nuclei by fluorescent in situ hybridization of a digoxigenin-labeled $[CCCUAA]_{27}$ RNA, followed by sheep anti-digoxigenin and TRITC-conjugated antibody to sheep IgG (red). In (C) is a superimposition of the images in (A) and (B). White and yellow indicate colocalization of the signals in (A) and (B). (**D**) Metaphase chromosomes of a HeLa cell expressing $[HA]_2$ epitope–tagged mTRF, labeled with an HA mAb (12CA5) and a FITC-labeled goat antibody to mouse IgG (green) (*15, 16*). In each panel, the DNA was stained with DAPI (blue). Scale bars, 5 μm.

ters of acidic residues have been primarily implicated in mediating protein-protein interactions during transcriptional regulation, they are also found in a number of proteins involved in chromosome function, including centromeric proteins, the p150 subunit of chromatin assembly factor I (CAF-I), and the largest subunit of the yeast origin recognition complex (ORC1) (*12*). The third recognizable motif in hTRF is a COOH-terminal region with strong homology to the DNA-binding repeats found in *myb* proto-oncogenes (Fig. 3D). These Myb-type DNA-binding repeats are ~50–amino acid stretches with three highly conserved tryptophan residues that form a helix-turn-helix (HTH) DNA-binding fold (*13*). Human TRF is most similar to the second and third Myb repeats in the mammalian and *Drosophila* Myb proteins. Unlike hTRF, these factors carry three tandem copies of this motif, two of which (R2 and R3) are sufficient for DNA binding (*13*). DNA-binding proteins carrying a single Myb repeat have been reported previously (*14*).

The nuclear and chromosomal location of TRF was determined by immunofluorescence microscopy of epitope-tagged mouse TRF (mTRF) protein expressed in transiently transfected HeLa cells (*15, 16*). Figure 4 shows that Met-Asp-Tyr-Lys-Asp-Asp-Asp-Asp-Lys (FLAG)–tagged mTRF displayed a punctate pattern in interphase nuclei. A similar speckled distribution was found for hemagglutinin $[HA]_2$-tagged mTRF (*17*). A minority of the transfected cells showed homogeneous nuclear staining in addition to the speckled pattern, possibly because of greater amounts of mTRF expression (*17*). No specific patterns emerged in control experiments with untransfected cells or after omission of the primary antibody (*17*). Dual labeling experiments revealed that the speckled mTRF distribution coincided with telomeric DNA detected by TTAGGG repeat–specific fluorescent in situ hybridization (Fig. 4, B and C). All telomeric loci were found to contain mTRF, and, vice versa, all mTRF speckles were associated with telomeric DNA.

The chromosomal distribution of TRF was determined by expression of a fusion protein with two tandem HA epitopes at its NH_2-terminus. In metaphase chromosomes of stably transfected HeLa cells, $[HA]_2$-tagged mTRF was predominantly detected at chromosome ends (Fig. 4D). Mouse TRF appeared to be a common feature of all chromosome ends. Occasionally, telomeres without mTRF could be found, but the absence of mTRF did not appear to be specific for any one chromosome end and is most likely a result of difficulties in detection of the protein in metaphase spreads. Discrete localization of mTRF to internal loci was not observed. In some metaphase spreads, weak staining along the axis of all chromosomes accompanied the telomeric signals (*17*). However, the strongest signals were invariably observed at chromosome ends.

These results demonstrate that TRF is a telomeric protein in vivo. TRF occupies chromosome ends both in interphase and metaphase, consistent with its role in the telomeric complex. Indirect evidence had previously suggested that mammalian chromosome ends contain one or more telomere-specific proteins (*3–6, 18*). For instance, human telomeres display an altered chromatin structure and bind to the nuclear matrix (*18*). Moreover, formation of new human telomeres upon DNA transfection occurs only with telomere seeds that contain precise TTAGGG repeat arrays (*5*). The failure of other (closely related) repetitive sequences to form new telomeres suggested that telomere healing involved a specific telomeric DNA-binding protein (*5*). However, telomeric proteins had not been isolated from human cells or other vertebrates (*1, 6*). Our

results now demonstrate directly that human telomeres form a complex with a telomere-specific protein, hTRF. In agreement with a role for hTRF at telomeres, the sequence specificity of hTRF matches the sequence requirements for telomere formation in human cells (6). On the basis of the data presented here, we suggest that TRF binding is required for telomere function at the ends of chromosomes in human and other mammalian cells.

TRF is a double-stranded telomeric DNA-binding factor. Double-stranded telomeric DNA-binding proteins had previously been implicated in telomere function in budding yeasts. In *Saccharomyces cerevisiae*, Rap1p binds along the double-stranded telomeric $TG_{1\text{-}3}$ tracts, where it regulates telomere length, chromosome stability, and telomeric position effects (*19*). Genetic alteration of the telomeric sequence of *Kluyveromyces lactis* has implicated a double-stranded telomeric DNA-binding protein in the regulation of telomerase (*20*). Although hTRF is not a Rap1p homolog, we note that the recently determined structure of a Rap1p-DNA complex reveals two Myb-type HTH motifs (*21*). Further structural and functional comparisons between these telomeric factors should be of interest. Several observations suggest that telomere maintenance in mammals, as in yeast, is subject to homeostasis (6). For instance, mammalian telomeres are maintained at constant length over the generations, and different mammals show species-specific telomere length (*22*). Moreover, telomerase expression in immortalized human cells does not lead to unlimited telomere elongation (*23*). A simple model to explain such phenomena invokes a factor (such as TRF) that binds along the length of the telomeric repeat array and modulates telomerase-mediated telomere elongation (6, *20*).

Human telomeres undergo programmed shortening in the soma (*16*, *24*). When grown in vitro, human cells enter a growth crisis at a stage when telomeres appear critically shortened and chromosome end fusions are frequent (*23*). Similarly, loss of telomeric DNA in human tumors has been suggested to contribute to genome instability in cancer (*2*, *25*). A mechanism to restore and maintain telomeres (such as activation of telomerase) may be required for tumor progression and cellular immortalization (*26*). Our findings raise the possibility that some of the deleterious consequences of telomere attrition may be caused by a failure of chromosome ends to bind protective telomeric proteins, including hTRF.

REFERENCES AND NOTES

1. G. Fang and T. R. Cech, in *Telomeres*, E. H. Blackburn and C. W. Greider, Eds. (Cold Spring Harbor Press, Cold Spring Harbor, NY, 1995), pp. 69–106.
2. T. de Lange, *ibid.*, pp. 265–293.
3. C. B. Harley, *ibid.*, pp. 247–264.
4. Z. Zhong, L. Shiue, S. Kaplan, T. de Lange, *Mol. Cell. Biol.* **12**, 4834 (1992); T. de Lange and J. Feng, unpublished data.
5. J. P. Hanish, J. L. Yanowitz, T. de Lange, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 8861 (1994).
6. T. de Lange, *Semin. Cell Biol.*, in press.
7. Human TRF was isolated from nuclear extract derived from 9.2×10^{11} HeLa cells containing 4 g of protein and ~650 pmol of hTRF (1 pmol of TRF is defined as the amount of hTRF required to bind 1 pmol of probe) over a series of ion exchange and affinity columns. Each column was loaded in dialysis buffer (buffer D; 100 mM KCl, 20 mM Hepes, 3 mM $MgCl_2$, 1 mM dithiothreitol, 0.5 mM phenylmethylsulfonyl fluoride, 20% glycerol, and 0.1% NP-40) and eluted with increasing KCl concentration in the same buffer without glycerol. Activity was determined by gel-shift assay with the use of a 142–base pair (bp) restriction fragment containing the hTRF binding site $[TTAGGG]_{12}$ (*4*). In the course of this work, it was noted that fractionation of partially purified hTRF over a column containing double-stranded $[TTAGGG]_{27}$ DNA resulted in dramatic loss of gel-shift activity of the factor. No such loss of activity was observed during fractionation of crude nuclear extract over the same resin. The lost activity of highly purified hTRF could be recovered in the 0.5 M KCl fractions by the addition of β-casein, *E. coli* lysate, guanosine 3′,5′-monophosphate phosphodiesterase, or calf thymus high mobility group–1, but not by the addition of bovine serum albumin, collagen, γ-globulin, α-transducin, or glutathione-S-transferase. By contrast, hTRF activity in crude HeLa nuclear extract was not enhanced by casein or *E. coli* lysate. Restoration of DNA-binding activity by casein and *E. coli* lysate has also been observed with highly purified fractions of the yeast centromeric binding factor CBF3 [J. Lechner and J. Carbon, *Cell* **64**, 717 (1991)]. For the final purification of hTRF, the following columns were used: DEAE-Fractogel 650 (EM Separations, Gibbstown, NJ) with hTRF eluting at 0.2 to 0.3 M KCl, P11 (Whatman, Marchstone, Kent, England) with hTRF eluting at 0.6 M KCl, and CM-Sepharose (Pharmacia) with hTRF eluting at 0.4 to 0.5 M KCl. Subsequently, hTRF was fractionated on Streptavidin-agarose beads carrying biotinylated, HinfI–digested *E. coli* chromosomal DNA from which hTRF eluted at 0.2 to 0.3 M KCl, followed by an affinity column composed of Streptavidin beads with biotinylated Asp718 restriction fragments containing $[TTAGGG]_{27}$. Human TRF eluted from this column at 0.5 M KCl (Fig. 1). The active fraction was supplied with β-casein and fractionated again by batch elution on nonspecific and specific (TTAGGG) DNA-containing resins. The final sample was precipitated with 20% trichloroacetic acid and 0.015% deoxycholate, fractionated on SDS-PAGE, and transferred to nitrocellulose. The sequenced hTRF preparation was composed of two polypeptides of ~60 kD; the larger protein was present at subpicomolar amounts and did not yield sequence information. Its relation to hTRF remains to be determined.
8. L. Chong and T. de Lange, data not shown.
9. P. A. Bäuerle and D. Baltimore, *Cell* **53**, 211 (1988).
10. Peptides were generated from the nitrocellulose-bound hTRF protein by in situ tryptic digestion [H. Erdjument-Bromage *et al.*, *Protein Sci.* **3**, 2435 (1994)] and fractionated by reversed-phase high-performance liquid chromatography [C. Elicone *et al.*, *J. Chromatogr.* **676**, 121 (1994)], and selected peak fractions were analyzed by a combination of automated Edman degradation and laser-desorption mass spectrometry [P. Tempst, S. Geromanos, C. Elicone, H. Erdjument-Bromage, *Methods* **6**, 248 (1994)]. Peptide sequences were compared to entries in various sequence databases with the use of the National Center for Biotechnology Information (NCBI) BLAST program (S. F. Altschul *et al.*, *J. Mol. Biol.* **215**, 403 (1990)]. Average isotopic masses of predicted peptides were summed from the cDNA-derived sequence with the use of ProComp software (version 1.2; P. C. Andrews, University of Michigan, Ann Arbor, MI).
11. TRF cDNAs were isolated from a HeLa cDNA library (Stratagene) by screening with a 33-nucleotide DNA probe that represents the region of homology of T29 peptide sequence to an anonymous cDNA in GenBank (accession number Z19923). Three overlapping cDNAs, together spanning 2684 bp, were sequenced (GenBank accession number U40705). The sequence contains a single long ORF of 439 amino acids starting with an AACATGG initiating codon, which is favorable for translation initiation [M. Kozak, *Nucleic Acids Res.* **15**, 8125 (1987)]. The 3′ end of the TRF cDNA contains an AATAAA polyadenylate [poly(A)] addition signal 18 bp upstream of a poly(A) stretch. The TRF ORF amino acid sequence was compared to proteins in the nonredundant database with the use of the program BLASTP (*10*). The nuclear targeting signals are of the nucleoplasmin type [J. Robbins, S. M. Dilworth, R. A. Laskey, C. Dingwall, *Cell* **64**, 615 (1991)]. Alignment to the Myb proteins was achieved with the program Clustalw 1.5; a SeqVu 1.0.1 display and Rao rules for conservative amino acid changes were used.
12. S. Hahn, *Cell* **72**, 481 (1993); W. C. Earnshaw *et al.*, *J. Cell Biol.* **104**, 817 (1987); K. F. Dohony *et al.*, *Cell* **73**, 761 (1993); J. Lechner, *EMBO J.* **13**, 5203 (1994); P. D. Kaufman, R. Kobayashi, N. Kessler, B. Stillman, *Cell* **81**, 1105 (1995); B. Stillman and S. Bell, personal communication.
13. P. Saikumar, R. Murali, E. P. Reddy, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 8452 (1990); O. S. Gabrielsen, A. Sentenac, P. Fromageot, *Science* **253**, 1140 (1991); K. Ogata *et al.*, *Cell* **79**, 639 (1994).
14. V. Delmas, D. G. Stokes, R. P. Perry, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 2414 (1993); N. Baranowskij, C. Frohberg, S. Prat, L. Willmitzer, *EMBO J.* **13**, 5383 (1994).
15. A cDNA for the mouse homolog of hTRF was isolated by cross-hybridization to the human cDNA (D. Broccoli, L. Chong, T. de Lange, unpublished data). The ORF in this cDNA was fused to a FLAG epitope. HeLa cells with long telomeres in the 15- to 30-kb size range [T. de Lange, *EMBO J.* **11**, 717 (1992)] were electroporated with a construct containing the FLAG-mTRF gene expressed from the cytomegalovirus (CMV) promoter and grown for 24 hours on Alcian Blue–coated cover slips. Cells were washed in phosphate-buffered saline, fixed in 2% formaldehyde in phosphate-buffered saline for 10 min, and permeabilized with 0.5% Nonidet P-40 in phosphate-buffered saline for 20 min. In situ hybridization was carried out essentially as described [O. C. M. Sibon *et al.*, *Histochemistry* **101**, 223 (1994)] with the use of a digoxigenin-labeled $[CCCUAA]_{27}$-repeat RNA probe made by in vitro transcription of plasmid pTH5 (*16*) using 11-digoxigenin–uridine 5′-triphosphate (Boehringer). Cells were incubated with sheep antibody to digoxigenin (anti-digoxigenin) (Boehringer) and mouse monoclonal antibody (mAb) M5 (Eastman Kodak) against the FLAG epitope, followed by tetramethyl rhodamine isothiocyanate (TRITC)–conjugated donkey antibody to sheep immunoglobulin G (IgG) and fluorescein isothiocyanate (FITC)–conjugated donkey antibody to mouse IgG (both from Jackson ImmunoResearch Labs). DNA was stained with 4,6-diamino-2-phenylindole (DAPI, 0.2 μg/ml). Cells were embedded in *p*-phenylene diamine (1 mg/ml) (Sigma) and 90% glycerol in phosphate-buffered saline (pH 7.8). After omission of the $[CCCUAA]_{27}$ probe or anti-digoxigenin, or when nuclear DNA was not denatured, only the FITC signal was found. In untransfected cells or after omission of mAb M5, only the TRITC signal was observed. Formaldehyde-fixed chromosome spreads were prepared essentially as described [D. A. Compton, T. J. Yen, D. W. Cleveland, *J. Cell Biol.* **112**,1083 (1991)]. HeLa cells were stably transformed by calcium phosphate coprecipitation of a construct containing a $[HA]_2$-tagged mTRF gene expressed from the CMV promoter and a neomycin-selectable marker (pRC/CMV, Invitrogen). A clonal G418-resistant cell line that expressed $[HA]_2$-tagged mTRF was isolated, treated with colcemid (0.1 μg/ml, 90 min), harvested by trypsinization, hypotonically swollen in 10 mM tris (pH 7.4), 10 mM NaCl, and 5 mM $MgCl_2$, and sedimented onto Alcian Blue–treated cover slips for 15 s at 3000 rpm in a Sorvall RT6000B tabletop centrifuge. Chromosome spreads were immediately fixed in 2% formaldehyde in phosphate-buffered saline,

permeabilized in 0.5% Nonidet P-40 in phosphate-buffered saline, and labeled with mouse mAb 12CA5 followed by FITC-labeled goat antibody to mouse IgG (Jackson ImmunoResearch Labs). DNA was stained with DAPI (0.2 μg/ml). Images of interphase nuclei were obtained with a Zeiss Axioplan microscope equipped with a Kodak DCS-200 digital camera. Images were noise-filtered with a 3 × 3 median filter, corrected for background with nontransfected cells as a reference, and merged to obtain triple labeling images using Adobe Photoshop. Chromosome spreads were photographed on Kodak Gold II 400 ISO film and converted to digital images with a Nikon slide scanner, after which FITC and DAPI images were corrected for background and superimposed.
16. T. de Lange *et al., Mol. Cell. Biol.* **10**, 518 (1990).
17. B. van Steensel and T. de Lange, data not shown.
18. T. de Lange, *EMBO J.* **11**, 717 (1992); H. Tommerup, A. Dousmanis, T. de Lange, *Mol. Cell. Biol.* **14**, 5777 (1994); S. Lejnine, V. L. Makarov, J. P. Langmore, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 2393 (1995).
19. M. N. Conrad, J. H. Wright, A. J. Wolf, V. A. Zakian, *Cell* **63**, 739 (1990); A. J. Lustig, S. Kurtz, D. Shore, *Science* **250**, 549 (1990); L. Sussel and D. Shore, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 7749 (1991); F. Klein *et al., J. Cell Biol.* **117**, 935 (1992); G. Kyrion, K. A. Boakye, A. J. Lustig, *Mol. Cell. Biol.* **12**, 5159 (1992); G. Kyrion, K. Liu, C. Liu, A. J. Lustig, *Genes Dev.* **7**, 1146 (1993); D. Shore, *Trends Genet.* **10**, 408 (1994); V. A. Zakian, in *Telomeres*, E. H. Blackburn and C. W. Greider, Eds. (Cold Spring Harbor Press, Cold Spring Harbor, NY, 1995), pp. 107–138.
20. M. J. McEachern and E. H. Blackburn, *Nature* **376**, 403 (1995).
21. D. Rhodes, personal communication.
22. D. Kipling and H. J. Cooke, *Nature* **347**, 400 (1990); J. A. Starling, J. Maule, N. D. Hastie, R. C. Allshire, *Nucleic Acids Res.* **18**, 6881 (1990).
23. C. M. Counter *et al., EMBO J.* **11**, 1921 (1992); A. J. Klingelhutz *et al., Mol. Cell. Biol.* **14**, 961 (1994); C. M. Counter *et al., J. Virol.* **68**, 3410 (1994).
24. H. J. Cooke and B. A. Smith, *Cold Spring Harbor Symp. Quant. Biol.* **51**, 213 (1986); R. C. Allshire *et al., Nature* **332**, 656 (1988); C. B. Harley, A. B. Futcher, C. W. Greider, *ibid.* **345**, 458 (1990); N. D. Hastie *et al., ibid.* **346**, 866 (1990).
25. N. D. Hastie and R. C. Allshire, *Trends Genet.* **5**, 326 (1989).
26. C. M. Counter, H. W. Hirte, S. Bachetti, C. Harley, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 2900 (1994); N. W. Kim *et al., Science* **266**, 2011 (1994); J. Feng *et al., ibid.* **269**, 1236 (1995).
27. Abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
28. We thank B. Stillman and S. Bell for information about ORC1 and N. Andrews, R. Benezra, A. Bianchi, R. Bose, R. Darnell, J. Feng, R. Lang, M. Lui, A. Lustig, N. Segil, O. Sibon, S. Smith, and J. Young for their contributions and advice. Supported by grants from the Lucille P. Markey Charitable Trust and NIH (GM49046) to T.d.L., grants from the National Cancer Institute (P30 CA08748-29) and NSF (BIR-9420123) to P.T., and a Irma T. Hirschl Career Scientist Award to P.T. B.v.S. is the recipient of a Human Frontier Science Program postdoctoral fellowship.

3 October 1995; accepted 1 November 1995

Conserved Initiator Proteins in Eukaryotes

Kimberley A. Gavin, Masumi Hidaka, Bruce Stillman*

The origin recognition complex (ORC), a multisubunit protein identified in *Saccharomyces cerevisiae*, binds to chromosomal replicators and is required for the initiation of cellular DNA replication. Complementary DNAs (cDNAs) encoding proteins related to the two largest subunits of ORC were cloned from various eukaryotes. The cDNAs encoding proteins related to *S. cerevisiae* Orc1p were cloned from the budding yeast *Kluyveromyces lactis*, the fission yeast *Schizosaccharomyces pombe,* and human cells. These proteins show similarity to regulators of the S and M phases of the cell cycle. Genetic analysis of *orc1*$^+$ from *S. pombe* reveals that it is essential for cell viability. The cDNAs encoding proteins related to *S. cerevisiae* Orc2p were cloned from *Arabidopsis thaliana, Caenorhabditis elegans*, and human cells. The human ORC-related proteins interact in vivo to form a complex. These studies suggest that ORC subunits are conserved and that the role of ORC is a general feature of eukaryotic DNA replication.

The replicon model for the initiation of DNA replication postulates that for DNA synthesis to occur, an initiator protein is required for recognition of a specific replicator sequence in the chromosome (*1*). In this model, recognition of the replicator by the initiator protein determines the location of an origin of DNA replication. Data from studies of prokaryotes and eukaryotic viruses support this hypothesis and suggest that the mechanism of initiation of DNA replication is conserved in eukaryotes. The initiation of DNA replication in eukaryotic cells is tightly controlled during the cell cycle and throughout development to ensure that duplication of the genome occurs only once per cell cycle. Thus, initiation is a key regulatory step in DNA replication.

In eukaryotes, the nature of initiator proteins and replicator elements remains unclear (*2*). Attempts to define precisely the origins of DNA replication have been largely unsuccessful. A notable exception is the yeast *Saccharomyces cerevisiae*, in which the origins of DNA replication have been physically mapped (*3*). Autonomously replicating sequences (ARS) have been characterized at the molecular level, revealing a modular structure (*4*, *5*). All ARS elements contain an essential ARS consensus sequence (ACS) and other elements that together are required for ARS function. In addition to characterized replication origins, an initiator protein complex has been isolated (*6*). ORC was identified by its ability to recognize and bind to the ACS in an adenosine triphosphate (ATP)–dependent manner. ORC recognizes a bipartite sequence within the replicator (*6*, *7*) and is bound to the DNA throughout the cell cycle (*8*). In cooperation with other cell cycle proteins, ORC determines the frequency of initiation in the *S. cerevisiae* genome (*9*–*12*).

ORC is a multisubunit protein complex consisting of six polypeptides with apparent molecular mass ranging from 50 to 120 kD. The genes encoding all the subunits have been cloned (*11*–*14*) and all are essential for cell viability. All six genes encode novel proteins, although *ORC1* and *ORC5* encode proteins that have regions of sequence similarity that are present in previously identified proteins. *ORC1* encodes the largest subunit of ORC, and its protein product is related to two other known yeast proteins (*13*). The protein most related to Orc1p is Sir3p, a protein involved in transcriptional silencing (*15*). Although these proteins are related over their entire length, the most striking similarity is found in the first 220 amino acids. A pairwise comparison shows that the proteins are 50% identical and 63% similar over this region. Orc1p is also related to two cell division cycle (CDC) regulators of initiation of DNA replication and control of the G_2- to M-phase transition, *S. cerevisiae* Cdc6p and *Schizosaccharomyces pombe* Cdc18p (*16*). A 270–amino acid region in Orc1p (residues 449 to 717) contains a purine nucleotide-binding motif (*17*) that is present in both Cdc6p and Cdc18p. In addition to the canonical P-loop and A-loop necessary for nucleotide metabolism (*17*), there are sequences flanking these nucleotide-binding motifs that are present only in Orc1p, Cdc6p, and Cdc18p (*13*). Orc1p is 50% identical to Cdc6p and Cdc18p across this region. We refer to this domain as the CDC–nucleoside triphosphate–binding (CDC-NTP) domain. Although Sir3p shares sequence homology with Orc1p across this domain, key residues essential for nucleotide binding are absent from Sir3p (*13*).

The nature of replicators in metazoan species is not clear, and in *Xenopus* early embryos their very existence has been questioned (*18*). To begin to address whether the replicon model applies to cell chromosome replication in higher eukaryotes, we identified genes related to *ORC1* in humans and other organisms (Fig. 1). We used a polymerase chain reaction (PCR) strategy, with primers based

K. A. Gavin, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA, and Program in Genetics, State University of New York, Stony Brook, NY 11794, USA. M. Hidaka and B. Stillman, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA.

*To whom correspondence should be addressed.