1294–1303 *Nucleic Acids Research, 1996, Vol. 24, No. 7*

The telobox, a Myb-related telomeric DNA binding motif found in proteins from yeast, plants and human

Thomas Bilaud, Catherine Elaine Koering, Emmanuelle Binet-Brasselet, Katia Ancelin+, Alessandra Pollice[1], Susan M. Gasser[2] and Eric Gilson*

Laboratoire de Biologie Moléculaire et Cellulaire, UMR49, Centre National de la Recherche Scientifique, Ecole Normale Supérieure de Lyon, 46 Allée d'Italie, 69364 Lyon Cedex 07, France, [1]Universita degli Studi di Napoli Federico II, Dipartimento di Genetica, Biologia Generale e Moleculare, via Mezzocannone 8, I-80134 Naples, Italy and [2]Swiss Institute for Experimental Cancer Research (ISREC), Chemin des Boveresses 155, CH-1066 Epalinges/Lausanne, Switzerland

Received December 6, 1995; Revised and Accepted February 12, 1996

EMBL accession nos X93511 and X93512

ABSTRACT

The yeast TTAGGG binding factor 1 (Tbf1) was identified and cloned through its ability to interact with vertebrate telomeric repeats *in vitro*. We show here that a sequence of 60 amino acids located in its C-terminus is critical for DNA binding. This sequence exhibits homologies with Myb repeats and is conserved among five proteins from plants, two of which are known to bind telomeric-related sequences, and two proteins from human, including the telomeric repeat binding factor (TRF) and the predicted C-terminal polypeptide, called orf2, from a yet unknown protein. We demonstrate that the 111 C-terminal residues of TRF and the 64 orf2 residues are able to bind the human telomeric repeats specifically. We propose to call the particular Myb-related motif found in these proteins the 'telobox'. Antibodies directed against the Tbf1 telobox detect two proteins in nuclear and mitotic chromosome extracts from human cell lines. Moreover, both proteins bind specifically to telomeric repeats *in vitro*. TRF is likely to correspond to one of them. Based on their high affinity for the telomeric repeat, we predict that TRF and orf2 play an important role at human telomeres.

INTRODUCTION

Telomeres are necessary to preserve the integrity of chromosomes during the cell cycle by allowing their proper segregation during cell division, by preventing their exonucleolytic degradation and end-to-end fusion, by positioning chromosomes within the nucleus and by enabling complete replication of chromosomal ends (for reviews see 1,2). In addition, yeast telomeres exert a position effect both on transcription (3,4) and on the timing of DNA replication (5). Finally, reports of nuclear movement mediated by telomeres during karyogamy and meiotic prophase in fission yeast (6) suggest that telomeres may also be involved in meiotic chromosome movement.

The DNA sequence at telomeres is generally constituted of an array of tandem repeats with clusters of G in the strand running 5′→3′ towards the chromosome extremities, ending with a 3′ overhang. The length of this repetitive DNA varies among species and cell types. For example, the irregular $(TG_{1-3})_n$ telomeric sequence of *Saccharomyces cerevisiae* spans only few hundred base pairs, while the TTAGGG repeats at vertebrate telomeres cover thousands of base pairs. A non-nucleosomal pattern of nuclease digestions is observed along the telomeric DNA of ciliates (7,8) and yeast (9). In contrast, the terminal repeats from various vertebrate and invertebrate species are arranged into regularly spaced nucleosomes smaller than bulk nucleosomes (10). An absence of nucleosomes at the very end of human telomeres was suggested from the diffuse nuclease digestion pattern observed with short human telomeres (11). This particular telomeric chromatin is associated with various subnuclear structures depending on species, cell type, stage in the cell cycle and chromosomes (for a review see 2). For example, mouse telomeres located on the long arm of chromosomes move from the interior of the nucleus to the periphery between the G1 and G2 phases of the cell cycle (12). During fission yeast meiosis telomeres are clustered near the spindle pole body and the nucleolus (13). The dynamics of these associations are largely uncharacterized and may require both stable and transient DNA–protein interactions.

In ciliate macronuclei the G-rich single-stranded tail is tightly bound to specific proteins that are believed to protect the extremities from degradation (for a review see 14). Such 'capping proteins' may also exist in vertebrates, since a similar single-stranded binding activity was detected in *Xenopus* egg extracts (15) and several G-rich strand binding proteins have been reported in *Chlamydomonas* and *S.cerevisiae* (16,17). Most of the double-stranded part of *S.cerevisiae* telomeric DNA appears to be complexed with an array of multifunctional repressor–activator protein 1 (Rap1), which distorts telomeric DNA as it binds (18). Rap1 contacts non-DNA binding proteins to initiate propagation of transcriptionally repressed chromatin into adjacent non-telomeric sequences (19; for a review see 20).

* To whom correspondence should be addressed

+Present address: Department of Pathology, McMaster University, Hamilton, Ontario L8N 3Z5, Canada

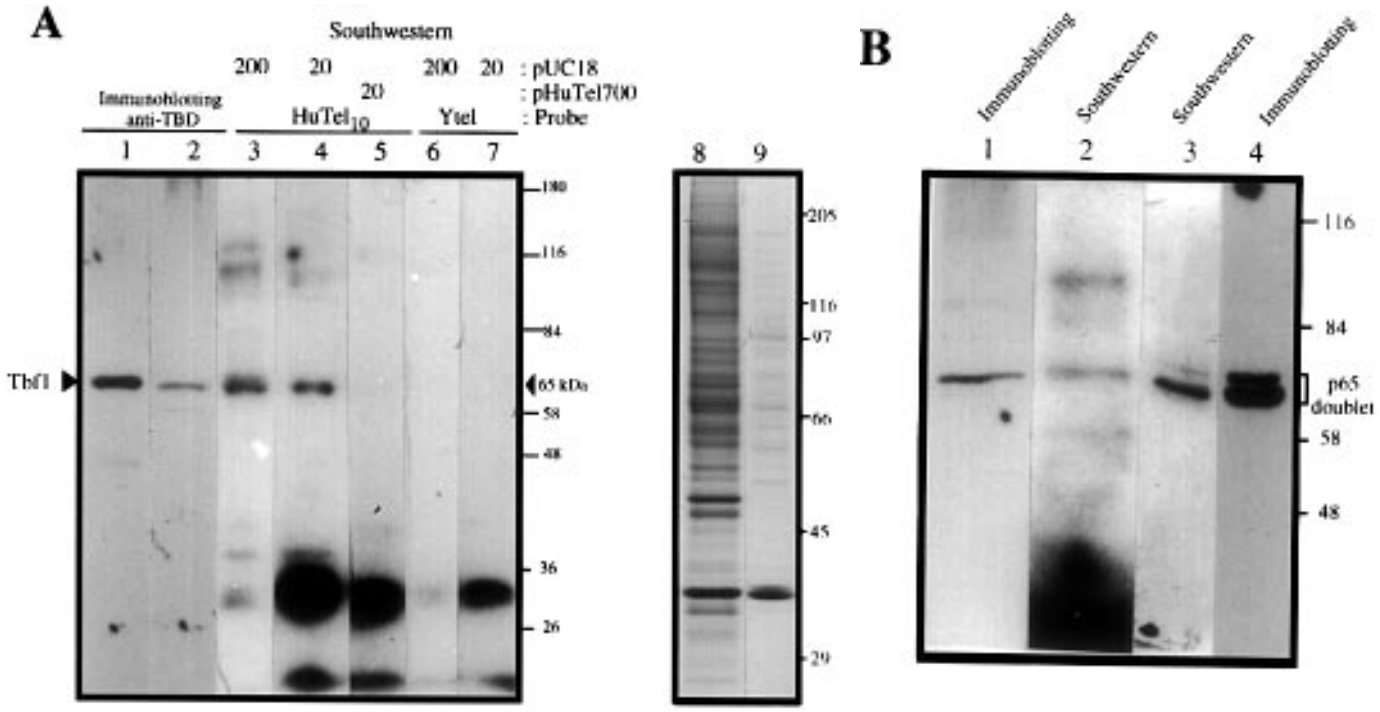

Figure 1. Human nuclear proteins analysed by immunoblotting with antibodies directed against Tbf1 and by Southwestern assay with a telomeric DNA probe. Molecular weight markers are indicated in kDa beside the gels. (**A**) Total yeast proteins (lane 1) and total HeLa nuclear proteins (lanes 2–7) separated by 8% SDS–PAGE and transferred to nitrocellulose. The filters were subjected either to an immunoblotting procedure using TBD antiserum (diluted 1/500) (lane 1) or affinity-purified antibodies against TBD (lane 2) or to a Southwestern assay (lanes 3–7) using the $HuTel_{10}$ probe (lanes 3–5) or the Ytel probe (lanes 6–7) in the presence of either linearized pUC18 DNA, a non-specific competitor or linearized $pHuTel_{700}$ DNA, a pUC18 derivative containing 700 nt of $(TTAGGG)_n$. The molar excess of plasmid DNA over the probe is indicated above each lane. The triangles indicate the position of Tbf1 and the human 65 kDa protein. The right panel shows Coomassie blue staining of an amount of total nuclear protein (lane 8) or chromosomal protein (lane 9) equivalent to that used for immunoblotting and Southwestern assays. (**B**) Total HeLa chromosomal protein or a soluble extract (S100) made from isolated HeLa nuclei were subjected either to an immunoblotting procedure with affinity-purified TBD antibodies (lanes 1 and 4 respectively) or to a Southwestern assay with the $HuTel_{10}$ probe (lanes 2 and 3 respectively). The gel was run for longer than in (A) in order to better resolve the two bands of the p65 doublet.

Several G-rich strand binding proteins have been identified in human cells. They all correspond to proteins involved in RNA metabolism (21,22). Therefore, the functional significance of DNA binding is unclear. The telomeric repeat factor (TRF) is a recently characterized protein which specifically binds $(TTAGGG)_n$ duplex DNA *in vitro* and which is localized at chromosome ends during mitosis (23). The function of TRF is unknown.

In budding yeast a protein called Tbf1 (TTAGGG repeat binding factor 1) specifically recognizes the vertebrate telomeric repeat $(TTAGGG)_n$, but not the irregular yeast telomeric repeat (24). Although the exact function of the protein is unknown, the gene encoding Tbf1 is essential for mitotic growth (25). Since TTAGGG repeats are found within the sequence of the subtelomeric elements X and Y′, Tbf1 may bind these sites *in vivo* and exert a function at the ends of yeast chromosomes (26). However, Tbf1 binding is not likely to be essential, since novel telomeres can be formed and maintained without TTAGGG sequences (3). Whatever the physiological function of Tbf1 may be in yeast, we hypothesize that this protein could be phylogenetically related to 'true' telomeric proteins in vertebrate cells, in which telomeres are exclusively constituted of TTAGGG repeats. We show here that a sequence of 60 amino acids within the Tbf1 DNA binding domain shares epitopes with two human proteins, TRF and the predicted C-terminal polypeptide from an as yet unknown protein. Both human proteins and Tbf1 contain a Myb-related motif involved in DNA binding. This so-called 'telobox' is also found in several plant proteins.

MATERIALS AND METHODS

Nuclei, chromosomes and nuclear fractionation

Nuclei from HeLa cells were isolated in polyamine buffers as previously described in Gasser *et al.* (27). Soluble nuclear extracts (S100) from HeLa and Jurkat cells were prepared according to Dignam *et al.* (28). Heparin fractions from Jurkat cells were eluted by increasing KCl concentration from a sulfopropyl 5PW column loaded with a 0.6 M KCl step elution fraction from a heparin–agarose column as previously described (29). Metaphase chromosomes were isolated from HeLa cells blocked in mitosis (30). The purity of the chromosome preparation was checked by fluorescence microscopy after DAPI staining (data not shown) and SDS–PAGE protein pattern as revealed by Coomassie staining (Fig. 1A, lane 9). The total nuclear and chromosome extracts for gel electrophoresis were obtained after digestion of the samples with micrococcal nuclease for 1 h on ice in 3.75 mM Tris–HCl, pH 7.4, 0.05 mM spermine, 0.125 mM spermidine, 20 mM KCl, 1 mM $CaCl_2$ and then adjusted to 2% SDS, 2% β-mercaptoethanol and 10% glycerol. The amount of extract loaded per well of SDS–polyacrylamide gel corresponds to ~10^6 nuclei.

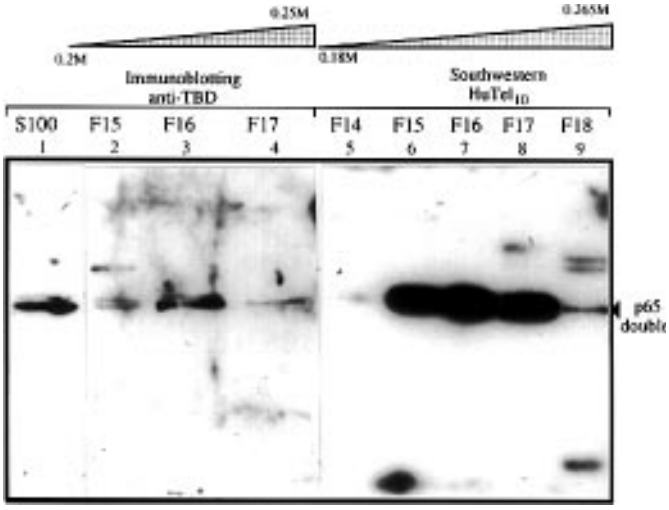

Figure 2. Partial purification of the p65 doublet. Fractionation of a Jurkat cell nuclear extract. Elution from a sulfopropyl HPLC chromatograph was with a linear KCl gradient (concentrations are given above the gels). Aliquots from fractions F15–F17 (lanes 2–4), along with the initial soluble nuclear extract (lane 1), were assayed by immunoblotting with antibodies directed against TBD and aliquots from fractions F14–F19 were assayed by Southwestern assay with the human telomeric probe $HuTel_{10}$ (lanes 5–9). The position of the p65 doublet is indicated on the right of the gel.

Expression of hybrid proteins between MalE and Tbf1 in *Escherichia coli* cells, protein purification and preparation of antibodies against Tbf1

We used the MalE expression system to produce the C-terminal part of Tbf1, hereafter called TBD, and its truncations. To construct the plasmid pMalE-TBD, pMAL™-c2 DNA (purchased from Biolabs) was cut with *Eco*RI and ligated with the 800 nt *Eco*RI fragment from pCDS47 (25). pMalE-TBD encodes a 72 kDa hybrid protein, named E–TBD, consisting of maltose binding protein (MalE) lacking its signal sequence, in order to be expressed into the cytoplasm, and the last 236 amino acids of Tbf1 separated by the recognition site of the protease factor Xa. Plasmid pΔ1 (expressing a 67 kDa hybrid protein, named E–Δ1, in which amino acids 482–562 of TBD are missing) was constructed by deleting the *Eco*NI–*Ear*I fragment from the coding region of TBD. Plasmid pΔ2 (expressing a 65 kDa hybrid protein, named E–Δ2, in which amino acids 404–468 of TBD are missing) was constructed by exchange of the *Nco*I–*Stu*I fragment of pΔ1 for a PCR-amplified fragment corresponding to residues 326–404 of Tbf1. All constructs were checked by sequencing of the cloned fragments.

In order to obtain large amount of TBD for injection into rabbits, E–TBD was purified from bacterial cells in a procedure involving two chromatographic steps. *Escherichia coli* strain DH5α transformed with plasmid pMalE-TBD was grown in 2× YT medium supplemented with 100 μg/ml ampicillin to an OD_{600} of 0.6, at which point IPTG was added to 2 mM. After a further 3 h cells were harvested and lysed as previously described (18). The supernatant was applied to a heparin HyperD™ column (Sepracor SA). Most E–TBD, as monitored by Western blot analysis with anti-MalE antibodies (kindly provided by J.M.Clément), eluted in the 0.6 M NaCl fraction. This fraction was dialyzed against 50 mM Tris–HCl, pH 7.8, 0.2 M NaCl, 1 mM EDTA, loaded onto an amylose column and E–TBD protein was eluted with maltose as described (31), producing a highly purified protein, as checked by the presence of a single band after heavily loading an SDS gel (data not shown). A similar purification procedure was applied to proteins used for DNA binding assays, except for the heparin column, which was omitted.

The TBD part from the purified E–TBD protein was separated from MalE by Xa protease cleavage, performed as indicated by Biolabs. The 28 kDa fragment corresponding to TBD was then purified from a preparative 8% SDS–polyacrylamide gel. Injection into rabbits and the bleeding schedule was performed as described (32). TBD antibody affinity purification and Western blotting were as previously described (27). As secondary antibodies we used anti-rabbit Ig peroxidase-linked $F(ab')_2$ fragments from donkey, detected by the ECL light-based system purchased from Amersham.

Band shift experiments and DNA probes

Band shift experiments were performed as described in Gilson *et al.* (18), except for the binding reaction buffer, which contained 20 mM KCl, 180 mM NaCl, 15 mM Tris–HCl, pH 7.4, 0.05 mM spermine, 0.125 mM spermidine, 1 mM dithiothreitol, 100 ng/μl poly(dI·dC), 100 ng/μl bovine serum albumin. About 2 fmol of probe was used. All the following probes were labelled at both ends with [α-^{32}P]dATP using the Klenow enzyme. The $HuTel_{10}$ probe is a 110 nt *Eco*RI–*Hin*dIII fragment from $pHuTel_{10}$ containing 10 TTAGGG repeats cloned into the polylinker region of pGEM3Zf–. The $HuTel_{2.5}$ probe is a 110 nt *Eco*RI–*Hin*dIII fragment from $pHuTel_{2.5}$ containing the sequence TAGGGTTAGGGTTAGGG inserted in between the *Sac*I and *Kpn*I sites of pUC18.

The sequences of double-stranded oligonucleotides used as competitor are, for $(TTAGGG)_{2.5}$, GTACCTAGGGTTAGGGTTAGGG annealed with TCGACCCTAACCCTAACCCTAG, and, for O.MYB, GTACAACCTAACTGACACACAT annealed with TCGAATGTGTGTCAGTTAGGTT. Note that O.Myb includes the sequence used for determination of the structure of c-Myb–DNA complexes (33).

Sequence analysis and cloning of orf1 and orf2

Computer searches in sequence databases were performed using the BLAST algorithm (34) and the e.mail servers available from NCBI. We used clustalW to produce the multiple alignment (35). The phylogenetic tree was constructed using the neighbour-joining method (36). The percentage of difference between sequences was taken as an arbitrary distance. Positions corresponding to gaps into the multiple alignment were not taken into account for calculation of the percentage of difference. The telobox consensus sequence was derived from multiple alignment of the eight members of this family, retaining amino acids found at least four times in a given position.

The sequences used for the multiple sequence alignment and tree construction were: MybSt1 (37); IBP1 (38); BPF1(39); Rap1 (40); Tbf1 (25); human c-Myb (41); human A-Myb and B-Myb (42); *Xenopus* Myb1 and Myb2 (43); a set of open reading frames derived from partial cDNA sequences (EST) with the GenBank accession nos Z26064 (orfA), D23805 (orfR1) and D22340 (orfR2); *orf1* and *orf2*, first identified from the assembly of R33191, R68526, R25990 and R70912 for *orf1* and T58911 and T11692 for *orf2*. Sets of primers containing built-in restriction sites were then designed in order to clone *orf1* and *orf2*. *Eco*RI sites were located at the 5′-end of the upstream primers (Pa and Pa′), whereas *Xba*I sites were located at the 5′-end of the downstream

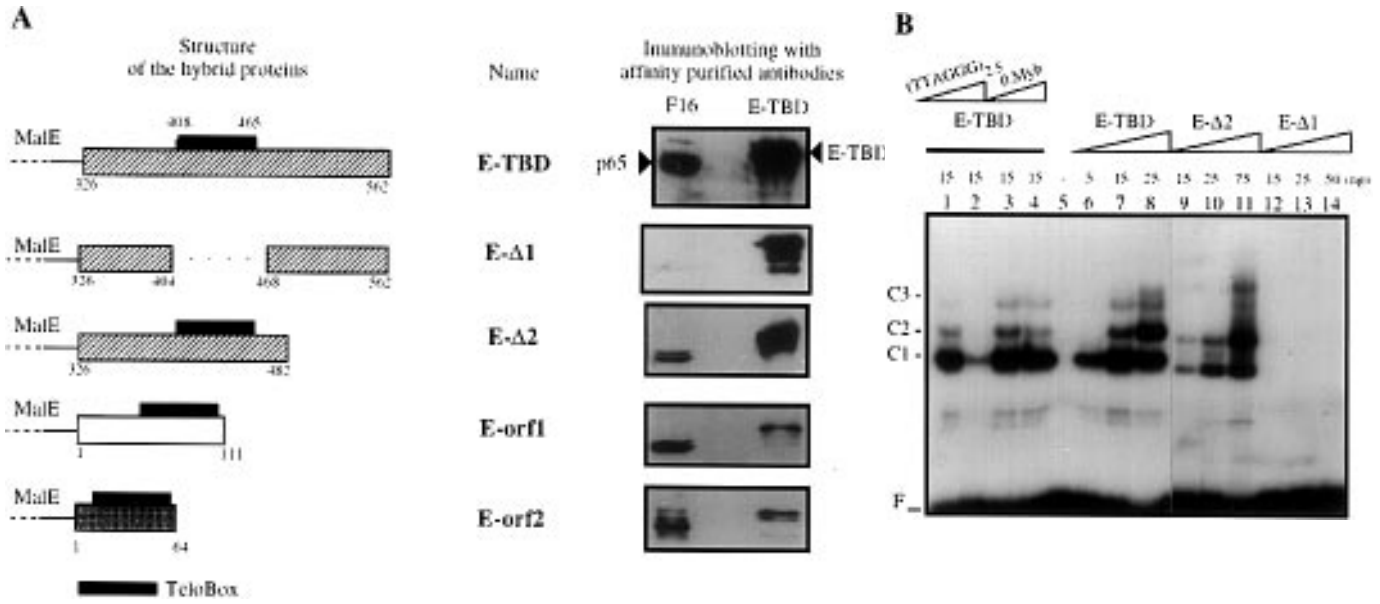

Figure 3. Immunological cross-reactivities and DNA binding capacities of the Tbf1 DNA binding domain (TBD) and truncated derivatives. (**A**) The left part shows a schematic representation of the various hybrid proteins produced in *E.coli* and used in these studies. The N-terminus part of these hybrids corresponds to bacterial maltose binding protein (MalE, thin dashed line). The C-terminus parts are constituted of the 236 C-terminal residues of Tbf1 (E–TBD, hatched rectangle) or by a truncated derivative lacking 64 internal residues (E–Δ1) or by a truncated derivative lacking the 80 terminal residues (E–Δ2) or by human orf1 (E–orf1, empty rectangle) or by human orf2 (E–orf2, grey rectangle). The black rectangles show the position of the telobox in each of these sequences. The right part shows a set of immunoblots bearing fraction F16 (see text and Fig. 2) and a bacterial extract containing E–TBD. The positions of the relevant proteins are indicated beside the gel in the upper photograph. The immunoblots were incubated in the presence of TBD antibodies, previously affinity purified against either E–TBD, E-Δ1, E–Δ2, E–orf1 or E–orf2. The bands below E–TBD are likely to be degradation products from the hybrid protein. (**B**) Band shift assays using hybrid proteins highly purified from bacterial lysates with an amylose column (see Materials and Methods) and a duplex DNA probe containing 2.5 repetitions of TTAGGG ($HuTel_{2.5}$; see Materials and Methods). The position of the free probe is indicated by F and the three DNA–protein complexes by C1, C2 and C3. The name and the quantity of the protein used in each assay are indicated above each lane, as well as the name of the double-stranded oligonucleotide ($HuTel_{2.5}$ or O.Myb) used as cold competitor. Lanes 1 and 3 correspond to a reaction with a 5-fold molar excess of the oligonucleotide over the probe and lanes 2 and 4 to a 100-fold excess.

ones (Pb, Pb′ and Pc′) (Fig. 6). After PCR amplification using the DNA of a HeLa cell cDNA library as template (44) products of expected sizes were digested with *Eco*RI and *Xba*I and inserted in between the corresponding sites of pMAL™ -c2 in order to fuse *malE* and *orf1* or *orf2* in-frame. Two independent clones emanating from each PCR experiment were sequenced twice on both strands and found to have identical sequences (Fig. 4). In each case, after IPTG induction and SDS–PAGE of a total protein extract, Coomassie staining revealed the induction of an abundant protein whose migration corresponded to the expected molecular weight for the hybrid protein (data not shown). These extracts were used either for a Southwestern analysis or to transfer the hybrid protein onto a nitrocellulose filter in order to purify TBD antibodies.

Southwestern assay

The procedures used are based on Miskimins *et al.* and von Kries *et al.* (45,46). After SDS–PAGE (8% polyacrylamide), without boiling the samples, gels were equilibrated in blotting buffer (25 mM Tris base, 192 mM glycine, 10% methanol) and electrophoretically transferred onto nitrocellulose filters (BAS85, 0.45 mm; Schleicher & Schuell). The filters were incubated overnight at 4°C in 25 mM Tris base, 192 mM glycine. They were then incubated in binding buffer (10 mM Tris–HCl, pH 7.4, 50 mM NaCl, 1 mM EDTA) supplemented with 5% non-fat milk for 90 min at room temperature with shaking, then for 90 min in binding buffer, 0.5% non-fat milk and again for 90 min at room temperature in binding buffer, 0.5% non-fat milk, end-labeled probe (~300 fmol/ml, corresponding to 10^6 c.p.m./ml), 5 µg/ml poly(dI·dC), 50 µg/ml *E.coli* DNA and competitor plasmid DNA as indicated (Figs 1 and 4). Filters were washed for 20 min in 2 l ice-cold binding buffer supplemented with 0.5% non-fat milk. Buffer was changed several times. Then filters were quickly dried on 3MM paper and exposed for autoradiography on Kodak X-Omat films and amplifying screens at –70°C.

The $HuTel_{10}$ probe is described above. The YTel probe is a 130 nt *Eco*RI–*Hin*dIII fragment from pYtCA-1X (47) that contains 80 nt of TG_{1-3} repeats. $pHuTel_{700}$ was constructed by inserting between the *Eco*RI and *Sma*I sites of pUC18 plasmid DNA a mixture of PCR products digested with *Eco*RI and *Stu*I generated in a template-free reaction (48) using primers GCGGA-ATTC$(TTAGGG)_8$ and GAAGGCCTC$(TAACCC)_8$. Sequencing of the resulting plasmid DNA revealed the presence of only TTAGGG repeats over the 700 nt of the inserted DNA in $pHuTel_{700}$.

RESULTS

The DNA binding domain of Tbf1 is immunologically related to two human nuclear proteins

Since the yeast Tbf1 protein specifically recognizes the vertebrate telomeric repeat $(TTAGGG)_n$, we hypothesized that a human telomeric protein might contain a DNA binding domain related to that of Tbf1. To explore this, a polyclonal rabbit antiserum was raised against the C-terminal 236 residues of Tbf1 (amino acids 326–562; see Materials and Methods), which were known to include the protein DNA binding domain (called TBD; 25).

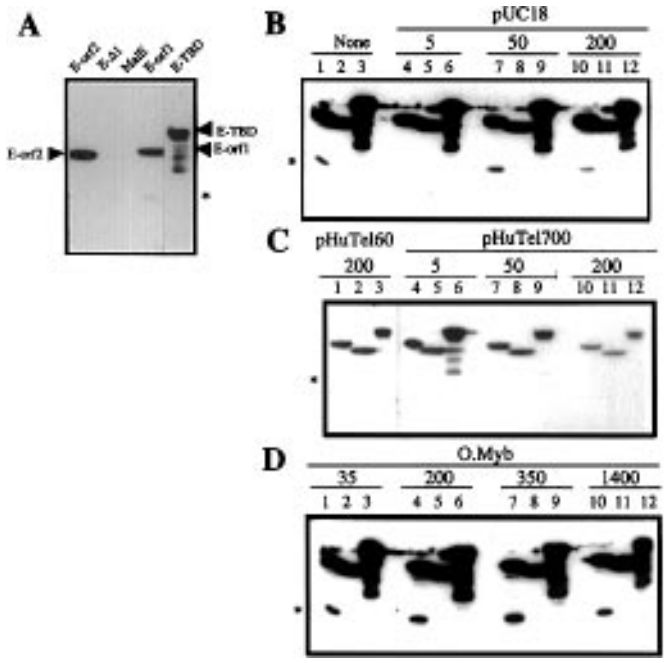

Figure 4. Southwestern assays with human orf1 and orf2. Bacterial extracts containing various MalE hybrid proteins were subjected to a Southwestern assay using a ^{32}P-labelled DNA fragment containing 10 TTAGGG ($HuTel_{10}$). The fast migrating band present in E–orf1 (marked by a star) is likely to be a breakdown product of the hybrid protein. (**A**) The name of the hybrid protein is indicated above each lane. Signals obtained are indicated by arrowheads. (**B**–**D**) Lanes 1, 4, 7 and 10 are bacterial extracts containing E–orf1; lanes 2, 5, 8 and 11 are bacterial extracts containing E–orf2; lanes 3, 6, 9 and 12 are bacterial extracts containing E–TBD. About 10 pmol hybrid protein was present in the loaded sample. The type of competitor DNA, either a linearized plasmid DNA (pUC18, $pHuTel_{60}$ or $pHuTel_{700}$) or a double-stranded oligonucleotide containing a typical Myb DNA site (O.Myb), is indicated above each lane, as well as its molar excess over the probe. Equivalent autoradiogram exposures are shown.

Western blot analysis of a total yeast protein extract revealed with the TBD antiserum (Fig. 1A, lane 1) shows a single band corresponding to the size of Tbf1 (65 kDa). When a protein extract from the same yeast strain expressing a hybrid protein between the transactivation domain of Gal4 and TBD is probed with this serum an additional band corresponding to the expected size for the Gal4–TBD hybrid protein is seen (data not shown). These results demonstrate that the serum is monospecific for Tbf1 in yeast.

The antibodies directed against TBD were affinity purified using TBD as the antigen and were used in blotting experiments with human nuclear proteins. In total protein from isolated HeLa nuclei the anti-TBD antibodies recognized a major band with an electrophoretic migration identical to that of Tbf1 and an upper band of lesser intensity (Fig. 1A, lane 2, Coomassie staining, and lane 8). The major band was also detected among total proteins from isolated metaphase chromosomes (Fig. 1B, lane 1). Using the secondary antibodies alone neither of the two bands were detected (data not shown). We ruled out that the human nuclei or chromosomes were contaminated with yeast by probing total human nuclear extracts with an antiserum directed against Rap1, an abundant yeast nuclear protein. No cross-reacting protein was detected (data not shown). A soluble extract prepared from isolated nuclei (S100) also exhibits a major signal at the level of Tbf1 closely migrating with an upper band of lesser intensity (Fig. 1B, lane 4); note that in this experiment SDS–PAGE was run for longer in order to improve the resolution of the two bands. We conclude that two human nuclear polypeptides with apparent molecular weights close to that of Tbf1 are specifically recognized by antibodies directed against the Tbf1 DNA binding domain.

The two Tbf1-related human proteins bind telomeric DNA

To determine whether the human proteins detected by TBD affinity-purified antibodies also bind human telomeric DNA a Southwestern assay was performed on the same protein samples using a duplex DNA probe containing 10 TTAGGG repeats (named $HuTel_{10}$). Remarkably, the major $(TTAGGG)_{10}$ binding activity in total nuclear extracts migrates with an apparent molecular weight of 65 kDa, exactly co-migrating with the Tbf1-related proteins (Fig. 1A, compare lane 2 with lanes 3 and 4). Since this activity is completely eliminated by the presence of a 20-fold excess of the specific competitor $pHuTel_{700}$ DNA (Fig. 1A, lane 5), but not by a 200-fold excess of non-specific pUC18 DNA (Fig. 1A, lane 3), the binding appears to be highly specific for telomeric repeats. A similar competition pattern was obtained using a DNA probe including 40 TTAGGG repeats (data not shown). Since $(TTAGGG)_n$ sequences, like most telomeric repeats, are TG-rich on one strand (1), we tested whether or not the 65 kDa band would bind the irregular $(TG_{1–3})_n$ sequence of *S.cerevisiae* telomeric DNA. With the double-stranded yeast telomeric probe YTel we were unable to detect any interaction with either 65 kDa protein (Fig. 1A, lanes 6 and 7), lending further support to the high selectivity of this human protein for $(TTAGGG)_n$ sequences. A similar band co-migrating with the major immunoblotting activity was detected in a Southwestern experiment with total mitotic chromosome proteins (Fig. 1B, lane 2).

Performing a similar Southwestern assay on human proteins from a soluble nuclear extract, two proteins exactly co-migrating with the doublet protein detected by immunoblotting bind the $HuTel_{10}$ probe (Fig. 1B, lanes 3 and 4). We reproducibly failed to detect the upper band in lysed total nuclei, both by immunoblotting and by Southwestern assays (Fig. 1A, lanes 2–4 and longer exposures of these autoradiographs; data not shown). This may be due to abundant proteins in whole nuclei, that are less abundant in soluble extracts, which interfere with either transfer to the filter or with DNA binding of the upper band. Treatment of the soluble extract with calf intestinal phosphatase did not modify the migration of either band of the doublet, as revealed by either Southwestern or immunoblotting experiments (data not shown). Thus the upper band is unlikely to correspond to a phosphorylated form of the lower band.

We next asked whether the protein bands at 65 kDa detected by immunoblotting really correspond to the proteins detected by Southwestern assay. To do this we subfractionated the soluble nuclear extract from Jurkat cells through several chromatographic steps and tested all samples by both immunoblotting and Southwestern assay. In the initial extract two closely migrating proteins ~65 kDa are recognized by anti-TBD antibodies (Fig. 2, lane 1), like those detected in a similar nuclear extract of HeLa nuclei (Fig. 1B, lane 4). After a heparin–agarose column the two proteins, as revealed by both DNA binding and immunoblotting, co-elute at 0.6 M KCl (data not shown). The 0.6 M heparin fraction was then loaded onto a sulfopropyl 5PW column, where

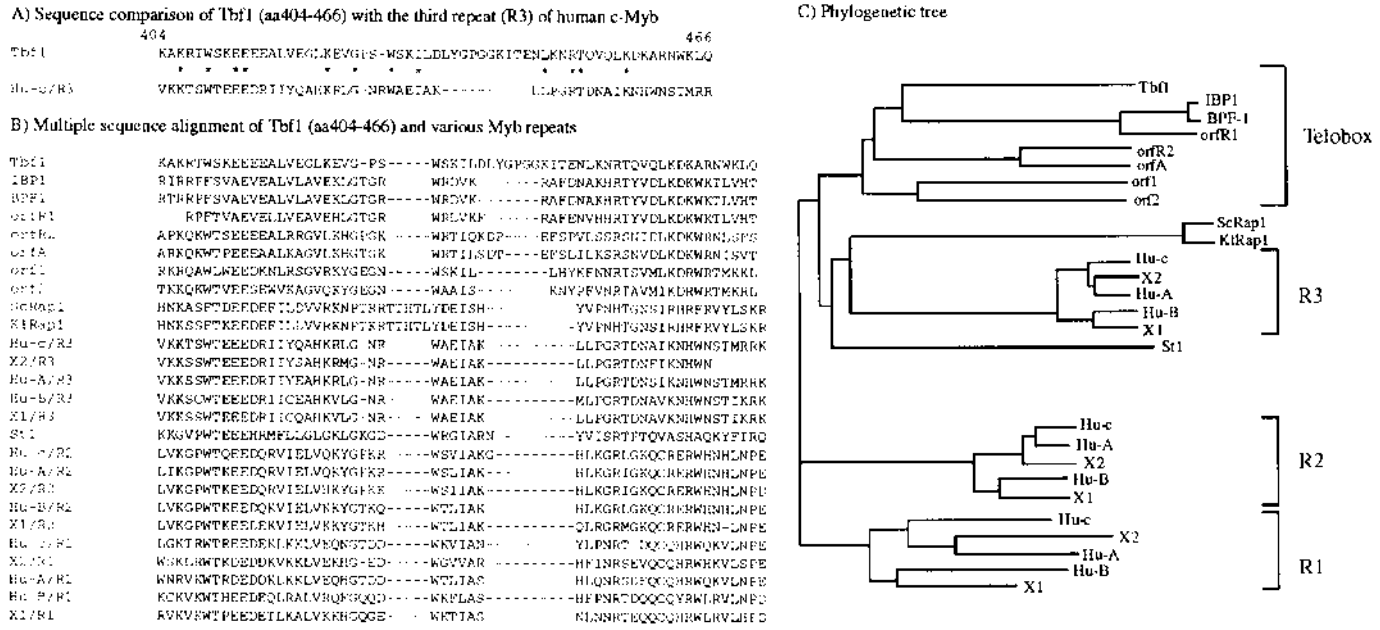

Figure 5. Sequence comparison of Myb repeats and teloboxes. Amino acids are given in the single letter code. (**A**) Alignment between amino acids 404–466 of Tbf1 and the third repeat of human c-Myb (Hu-c/R3). The amino acid identities are indicated by a star. (**B**) ClustalW (35) was used to produce a multiple alignment of the following Myb repeats: Tbf1 (amino acids 404–466 of *S.cerevisiae* Tbf1); IBP1 (amino acids 576–630 of maize IBP1); BPF1 (amino acids 581–636 of parsley BPF1); orfR1 and orfR2 (translated from rice partial cDNA sequences); orfA (translated from an *A.thaliana* partial cDNA sequence); orf1 and orf2 (translated from a human partial cDNA sequence); ScRap1 (amino acids 358–414 of *S.cerevisiae* Rap1); KlRap1 (amino acids 211–267 of *K.lactis* Rap1); Hu-c/R1, Hu-c/R2 and Hu-c/R3 (Myb repeats of human c-Myb); Hu-A/R1, Hu-A/R2 and Hu-A/R3 (Myb repeats of human A-Myb); Hu-B/R1, Hu-AB/R2 and Hu-B/R3 (Myb repeats of human B-Myb); St1 (amino acids 100–151 of potato MybSt1); X1/R1, X1/R2 and X1/R3 (Myb repeats of *Xenopus* Xmyb1); X2/R1, X2/R2 and X2/R3 (Myb repeats of *Xenopus* Xmyb2). (**C**) Phylogenetic tree showing the divergences among the sequences used for the multiple alignment presented in (B) constructed using the neighbour-joining method (36). The percentage of difference between sequences was taken as an arbitrary distance. The sequences belonging to the telobox group (also referred in the text as the 'Tbf1 family'), as well as to the R1, R2 and R3 groups, are bracketed on the right of the figure.

again immunoblotting and Southwestern assay detect two closely migrating proteins that elute between 200 and 240 mM KCl in fractions F15–F17 (Fig. 2, lanes 2–9); in the Southwestern assay the two closely migrating bands in these fractions are more easily seen in a lower exposure of the autoradiograph (data not shown). The binding specificity for $(TTAGGG)_n$ DNA of the activities recovered in fraction F16 was confirmed by competition experiments (data not shown). F16 is still a mixture of several polypeptides, as revealed by SDS–PAGE followed by Coomassie staining, and no major polypeptide at 65 kDa is visible (data not shown).

Since the two proteins revealed by Southwestern assay after two successive chromatographic steps are indistinguishable by SDS–PAGE analysis from the two revealed by immunoblotting, it is very likely that both assays identify the same proteins. Thus we conclude that there exist two human polypeptides (the p65 doublet; Fig. 1B) that bind specifically to human telomeric DNA and that are recognized by affinity-purified antibodies directed against the Tbf1 DNA binding domain.

Both polypeptides of the p65 doublet share epitopes with a single Myb repeat present at the end of Tbf1

In order to map the epitopes of Tbf1 that are shared with the p65 doublet, we tested antibodies purified against a series of truncated forms of TBD. A TBD lacking 64 central amino acids (Δ1) and a TBD lacking 80 terminal amino acids (Δ2) were expressed and purified from *E.coli* cells as a hybrid protein with the bacterial maltose binding protein MalE (E–Δ1 and E–Δ2, left part of Fig. 3A). Antibodies affinity purified against E–Δ1 detected the p65 doublet present in fraction F16 poorly, while they bind perfectly a full-sized E–TBD protein expressed in bacterial cells (right part of Fig. 3A). In contrast, when purified against E–Δ2 they efficiently recognized both p65 and E–TBD (right part of Fig. 3A). These results demonstrate that most of the epitopes shared between TBD and the p65 doublet are localized within amino acids 404–468 of Tbf1. Interestingly, this region contains an essential part of the Tbf1 DNA binding domain, since its deletion abolished DNA binding. This is demonstrated both by band shift assay with highly purified hybrid proteins (Fig. 3B, compare lanes 6–8, wild-type E–TBD, with lanes 12–14, E–Δ1) and Southwestern assay with total bacterial extracts containing either E–TBD or E–Δ1 (Fig. 4A). In contrast, a TBD with the 80 amino acid C-terminal deletion still binds telomeric DNA in a sequence-specific manner (Fig. 2B, lanes 9–11; data not shown). We conclude that the Tbf1 DNA binding domain is found between positions 326 and 482 and includes, between positions 404 and 468, most of the epitopes shared with the p65 doublet.

We attempted to identify p65 by searching human protein sequences homologous to the sequence (amino acids 404–468) of Tbf1 in various databases using the BLAST algorithm (34). We found that the sequence corresponding to amino acids 406–457 exhibits homology to the DNA binding domain of Myb proteins (Fig. 5; data not shown). These DNA binding domains are often constituted by three imperfect tandem repeats (R1, R2 and R3), with R2 and R3 making specific contacts with the cognate DNA sequence (33). An alignment between Tbf1 and the third repeat

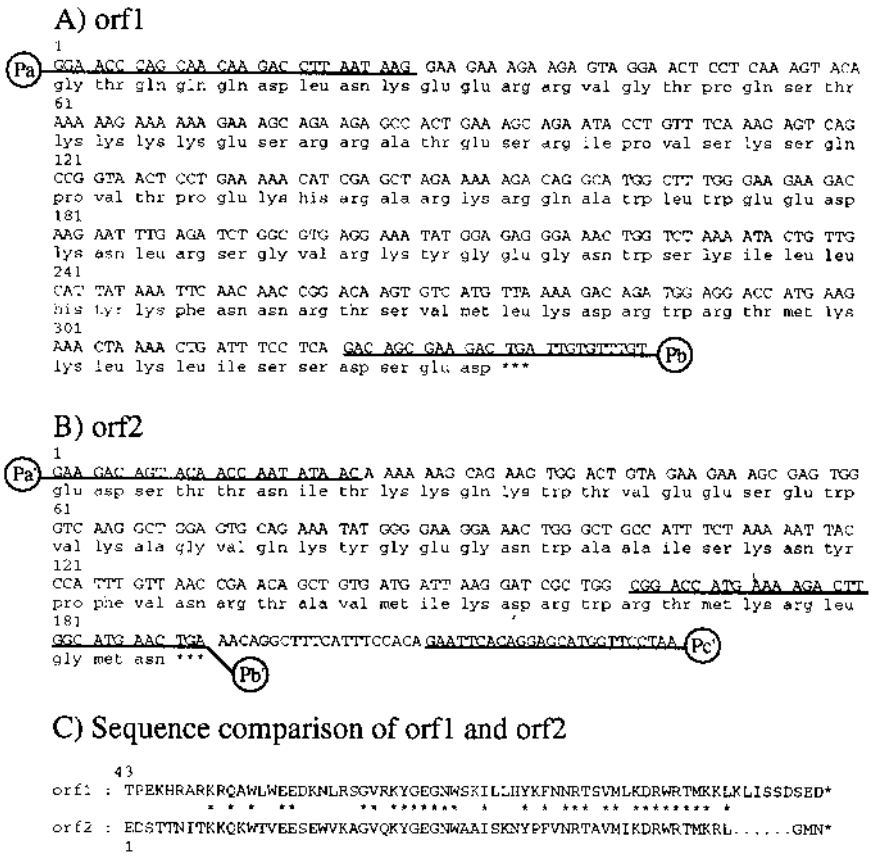

Figure 6. Sequence of the orf1 and orf2 cDNAs and their encoded peptides. (**A** and **B**) The sequences of the PCR primers (Pa, Pb, Pa′, Pb′ and Pc′) used to amplify *orf1* and *orf2* cDNA are underlined (see Materials and Methods). For the downstream primers (Pb, Pb′ and Pc′) the underlined sequences correspond to the complementary strand of the oligonucleotide. (**C**) Sequence comparison of the telobox region of orf1 and the entire orf2 amino acid sequence. Identical residues are marked with a star. Note that outside the telobox region the sequences are divergent.

of a typical member of the Myb protein family, namely human c-Myb, is shown in Figure 5A. Although the homology is partial, many of the highly conserved residues important for R3 are present within the Tbf1 sequence. In particular, two out of the three expected tryptophan residues are present. However, in contrast to most known Myb proteins, the Tbf1 sequence has only one such repeat. Intriguingly, Tbf1 does not specifically bind DNA sequences recognized by various known Myb proteins. This is demonstrated in Figure 3B, where increasing amounts of a double-stranded oligonucleotide carrying either a DNA sequence recognized by most known Myb proteins (49,33) (lanes 3–4) or 2.5 repeats of TTAGGG (lanes 1–2) were added as competitor to the band shift assay. Similarly, no competition was observed using two other DNA sequences recognized by plant MYB.Ph3 (50; data not shown).

A subfamily of single Myb repeat-containing proteins binds telomeric DNA sequences

Since Tbf1 does not bind a typical Myb DNA site and contains only a single Myb repeat (see above), Tbf1 may belong to a particular subfamily of Myb proteins. Indeed, among the Myb proteins revealed in the BLAST search for proteins homologous to the Tbf1-like Myb repeat two have a single Myb repeat and bind specific DNA sequences related to telomeric repeats, namely the maize *Shrunken* initiator binding protein IBP1 (38) and the parsley BoxP binding factor BPF1 (39). The IBP1–DNA complex at the *Shrunken* promoter covers an exact plant telomeric repeat (AGGGTTT) (38) and BPF1 binds a series of sequences rich in G_nT_m motifs (39). Both proteins share an almost identical single Myb repeat (one mismatch out of 54 amino acids) located at their C-termini (Fig. 5B) which is implicated in DNA binding (38,39). This suggests that, in addition to their function in transcriptional regulation, these proteins may play a role in plant telomere physiology. Comparison of the Myb repeat with the DNA binding domain of the well-characterized yeast telomeric factor Rap1 shows that Rap1 also contains two regions of homology with the Myb box (40).

Such observations prompted us to explore whether a subclass of Myb repeats defines a family of telomeric DNA binding factors. To this end we constructed an evolutionary tree of Myb repeat sequences. Our analysis includes amino acids 404–466 of Tbf1, the region of Rap1 which exhibits the most pronounced homology to Myb (i.e. amino acids 211–267 of *Kluyveromyces lactis* Rap1 and amino acids 358–414 of *S.cerevisiae* Rap1), IBP1, BPF-1 and another single Myb repeat plant protein (MybSt1), as well as several open reading frames (orf) derived from partial cDNA sequences that contain a single Myb repeat (orfR1 and orfR2 from rice, orfA from *Arabidopsis thaliana* and orf1 and orf2 from human). We included five representative Myb proteins of the ‘three repeats family’ by considering each repeat

A) Sequence comparison of the Telobox consensus with R2R3 of the human c-Myb.

```
            R
Telobox KQxWTVEExEALVxGVxKxGTGxWRxI  (1,12) NxKxRTxVDLKDKWRTxxxT
        *  ** **        * * *   *  *           * **     *  *
        *  ** **           *   *  *              *        *
Hu-c/R3 KTSWTEEEDRIIYQAHKRLGNR-WAEIAK-------LLPGRTDNAIKNHW-NSTMR
Hu-c/R2 KGPWTQEEDQRVIELVQKYGPKRWSVIAKG------HLKGRLGKQCRERWHNHLNP
```

B)The sequence of the telomeric repeat and of Myb binding sites are related.

```
Telomeric repeat TAACCC
                  ****
Myb binding site YAACGG
                    T
```

Figure 7. Similarities and differences between telobox and Myb repeat. (**A**) The telobox consensus sequence was derived from the multiple alignment of the eight members of this family (Fig. 5), keeping amino acids found at least four times in a given position. The underlined residues are conserved between all the family members. A central variable region of from 1 to 12 amino acids is indicated by (1,12). Identical residues between the telobox consensus and both R2 and R3 of human c-Myb are represented by a double star, while identical residues between the telobox consensus and either R2 or R3 are represented by a single star. The residues of R2 and R3 interacting with the Myb DNA site, as determined by NMR analysis, are in bold type (33). (**B**) A sequence comparison between a human telomeric repeat and the consensus sequence recognized by most of the known Myb proteins (49), but not by Tbf1, orf1 and orf2. The bases interacting with c-Myb are in bold type (33).

(R1, R2 and R3) as a separate unit, namely c-Myb, A-Myb and B-Myb from human and XMyb1 and XMyb2 from *Xenopus laevis*. The tree was constructed based on multiple sequence alignments (Fig. 5B). As expected, the members of the 'three repeats family' fall into three groups which include sequences from either R1, R2 or R3 (Fig. 5C). Remarkably, most single repeat sequences are clustered in one separated group connected to the R3 sequences. This implies that they are more closely related to one another than to any of the other repeats from the same or other species. Interestingly, this 'Tbf1 family' exhibits a pronounced similarity to the R3 sequence, which corresponds to the critical repeat for DNA binding. Rap1 and MySt1 do not fall into this family, but are, nevertheless, distantly related members of the R3 family.

Human orf1 and orf2 bind telomeric DNA sequences

To further test the possibility that all members of the 'Tbf1 family' bind telomeric DNA (see above) we investigated the DNA binding properties of the two human open reading frames derived from partial cDNA sequences, namely orf1 and orf2. The DNA coding for orf1 and orf2 was PCR-amplified from a HeLa cDNA library as described in Materials and Methods. After cloning in-frame with the *malE* gene synthesis of the E–orf1 and E–orf2 hybrid proteins was checked by Coomassie blue staining of total bacterial extracts analysed by SDS–PAGE. An abundant induced protein of the expected molecular weight was confirmed by immunoblotting using anti-MalE antibodies (data not shown). The sequence recovered by PCR largely confirmed the expected sequence from an EST assembly (Fig. 6A and B), with the presence of a stop codon in-frame with the *Myb*-containing open reading frames. Since this stop codon was present within the sequence of the PCR primers Pb and Pb′, we also amplified *orf2* with a primer located downstream of Pb′ (Pc′, Fig. 6B); sequence determination through Pb′ confirmed the presence of the stop codon for *orf2*. The amino acid sequences of orf1 and orf2 outside the Myb-related region are not related (Fig. 6C), confirming that orf1 and orf2 reflect portions of two different proteins having a Myb/Tbf1-related domain at their extreme C-terminus. Intriguingly, for all the characterized members of the 'Tbf1 family', i.e. Tbf1, IBP1, BPF1, orf1 and orf2, the single Myb repeat is located at the extreme C-terminus of the protein.

Crude extracts from *E.coli* cells expressing either E–orf1 or E–orf2 were subjected to a Southwestern analysis with $HuTel_{10}$ as probe. In both cases a telomeric DNA binding activity co-migrates with the hybrid protein (Fig. 4A). As a control extracts from bacteria expressing only MalE or the hybrid protein E–Δ1, which does not bind $(TTAGGG)_n$ in band shift assays (Fig. 3B), failed to exhibit any telomeric DNA binding activity and an extract containing E–TBD exhibits a binding activity at the level of the hybrid protein (Fig. 4A). This shows that, like TBD, orf1 and orf2 bind telomeric DNA sequences. The specific binding of E–orf1 and E–orf2 to $(TTAGGG)_n$ sequences was further analysed by specific and non-specific competition experiments. Up to a 200-fold molar excess of non-specific competitor DNA (pUC18) over the probe did not affect binding of either E–orf1 or E–orf2 or E–TBD (Fig. 4B). In contrast, a similar molar increase of the specific competitor $pHuTel_{700}$ DNA, containing 700 nt of TTAGGG repeats inserted into pUC18, greatly reduced binding of the three hybrid proteins (Fig. 4C, lanes 4–12). It is worth noting that $pHuTel_{60}$ DNA, containing 60 nt of TTAGGG repeats, was ~40 times less efficient in competition as compared with $pHuTel_{700}$ DNA, further indicating that the competition is dependent upon the number of TTAGGG repeats (Fig. 4C, compare lanes 1–3 and 7–9). Finally, the binding of either orf1, orf2 or TBD is unaffected by the presence of up to a 1400-fold molar excess of O.Myb oligonucleotide over the probe, showing that, like TBD, orf1 and orf2 do not exhibit specific binding to a typical Myb DNA site. Overall, these results demonstrate that orf1 and orf2, like TBD, specifically bind human telomeric DNA and that the minimal telomeric DNA binding domain of the proteins from which orf1 and orf2 are derived is found within the 111 C-terminal amino acids of orf1 and the 63 C-terminal amino acids of orf2.

Human orf1 and orf2 are related to the p65 doublet

The fact that orf1 and orf2 bind telomeric DNA sequences and share homologies with amino acids 404–466 of Tbf1 strongly suggests that these open reading frames may be identical to the p65 doublet. When TBD antibodies were affinity purified against either E–orf1 or E–orf2 they recognized roughly equally p65 and

E–TBD (right part of Fig. 3A). Since TBD antibodies purified against MalE alone do not react with p65 and E–TBD (data not shown), we can conclude that orf1 and orf2 are both immunologically related to Tbf1.

DISCUSSION

This study has identified a particular Myb-related protein motif that appears to be specialized for specific interaction with duplex telomeric DNA. This motif is present in proteins from yeast (Tbf1), plants (IBP1 and BPF1) and in partial sequences of human open reading frames (orf1 and orf2). We propose to name this motif the 'telobox'. Other putative members of the telobox family include open reading frames from rice (orfR1 and R2, Fig. 5B) and from *Arabidopsis* (orfA, Fig. 5B). The sequence of orf1 is identical to the C-terminal part of the human telomeric protein TRF, whose sequence was published after this work was completed (23). Since orf1/TRF is located at chromosome ends *in vivo* (23; unpublished results), this protein is expected to play an important role at telomeres. This may also be true for orf2, which binds TTAGGG *in vitro* with the same affinity as orf1/TRF (Fig. 4). Overall, these results strongly support the existence of two human telomere-associated proteins sharing a telobox at their C-terminus. The respective role of each protein in telomere physiology remains to be determined.

A telobox consensus was derived from the multiple alignment presented in Figure 5B, revealing a bipartite structure with a central region that is variable in length and sequence (Fig. 7A). Roughly 30% of residues are identical between the N-terminal 27 residues of the telobox consensus and R2/R3 of c-Myb, while only 10% are identical with the two c-Myb repeats within the C-terminal 19 residues (Fig. 5A). In particular, the C-terminal VDLKDKWRT sequence of the telobox consensus shows limited homology with the sequence of R2 and R3, while it is highly conserved among the telobox members, including orf1 and orf2 (Fig. 5B). Interestingly, the corresponding regions of R2 and R3 contain the residues that establish specific contacts with the bases of the Myb consensus site, as revealed by NMR analysis (33) (Fig. 7A). This suggests that this telobox motif might also be crucial for specific telomeric sequence recognition. In this respect it is worth noting that the sequence of a vertebrate telomeric repeat is distantly related to the Myb binding site consensus (Fig. 7B). Thus it is tempting to speculate that the various Myb motifs, through slight modifications in their DNA contacting residues, can bind different but related sequences. This is in agreement with the fact that Tbf1, orf1/TRF and orf2 recognize telomeric DNA repeats, while they are unable to specifically bind typical Myb DNA sites (Figs 3B and 4D).

The telobox constitutes the major part of the telomeric DNA binding domain, at least for orf1/TRF and orf2, which contain little more than the telobox motif. However, in IBP1 a stretch of basic residues following the telobox motif was shown to also be necessary for efficient DNA recognition (38). Thus the telobox flanking residues might also be required to stabilize or to properly fold the telobox or to contact additional DNA sites.

The level of homology between distantly related telobox sequences is high enough to allow interspecific immunological cross-reactivities. For example, antibodies directed against the yeast Tbf1 telobox specifically interact with two human telobox peptides (orf1 and orf2; Fig. 3A). Furthermore, we have shown that antibodies directed against the telobox of Tbf1 almost exclusively detect two human nuclear proteins of ~65 kDa (the p65 doublet), both of which specifically bind telomeric DNA sequences in a Southwestern assay (Figs 1 and 2). At least one of the p65 polypeptides is likely to correspond to TRF, which has a similar apparent molecular weight (23). Whether orf2 corresponds to the other p65 polypeptide or to another protein remains to be determined.

The fact that Rap1 does not fall into the telobox family based on our phylogenetic tree analysis (Fig. 5C) suggests that other types of Myb-related domains may also be used for binding telomeric repeats. In addition to the sequence divergence between Rap1 and telobox proteins, the Rap1 DNA binding domain contains two Myb-related motifs instead of one (40). It is worth noting that the sequence of the *S.cerevisiae* telomeric DNA $(TG_{1-3})_n$ is quite different from the sequences found in a wide phylogenetic range of eukaryotes, including the vertebrate TTAGGG sequence (53). This suggests that during evolution a new telomeric repeat sequence might have been added to the existing one, requiring recruitment of an ancient Rap1 precursor for efficient telomere maintenance. The presence of TTAGGG-like repeats at the junction between yeast telomeric repeats and the interior of chromosomes may thus represent 'relic' sequences from an ancient telomere (26). The conservation of Tbf1 in yeast is probably due to non-telomeric functions, which remain uncharacterized but may be linked to regulation of transcription. Like Rap1, which acts both as a structural component of yeast telomeres and as a transcriptional regulator (54), both IBP1 and BPF1 were first identified as promoter binding elements. It will be revealing to examine other members of the 'telobox' family to elucidate whether or not the presence of a Myb-related DNA binding domain and involvement in transcriptional regulation represent universal characteristics of telomere binding proteins.

ACKNOWLEDGEMENTS

We thank D. Shore and R.K. Moyzis for gifts of plasmids, R. Brent for providing the HeLa cDNA library, W. Saurin for his help in the phylogenetic tree construction, J. M. Clément for his help during the amylose–heparin chromatography, G. Lombart-Platet for generously sharing the sulfopropyl fractions and C. Brun for critical reading of the manuscript. C.E.K. thanks the Association Française de Lutte contre la Mucovisidose (AFLM) for her long-term fellowship. S.M.G. thanks the Swiss National Science Foundation and the Human Frontier Program for continued support. E.G. thanks EMBO for his long-term fellowship in the Gasser Laboratory, as well as the Ligue contre le Cancer, ARC, GREG, Region Rhône-Alpes and AFLM since moving to Lyon. The nucleotide sequence data reported in this paper will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession nos X93511 for orf1 and X93512 for orf2.

REFERENCES

1 Zakian,V.A. (1989) *Annu. Rev. Genet.*, **23**, 579–604.
2 Gilson,E., Laroche,T. and Gasser,S. (1993) *Trends Cell Biol.*, **3**, 128–134.
3 Gottschling,D.E., Aparicio,O.M., Billington,B.L. and Zakian,V.A. (1990) *Cell*, **63**, 751–762.
4 Nimmo,E.R.N., Cranston,G. and Allshire,R.C. (1994) *EMBO J.*, **13**, 3801–3811.
5 Ferguson,B.M. and Fangman,W.L. (1992) *Cell*, **68**, 333–339.
6 Chikashige,Y., Ding,D-Q, Funabiki,H., Haraguchi,T., Mashiko,S., Yanagida,M. and Hiraoka,Y. (1994) *Science*, **264**, 270–273.
7 Gottschling,D.E. and Cech,T.R. (1984) *Cell*, **38**, 501–510.

8 Price,C.M. (1990) *Mol. Cell. Biol.*, **10**, 3241–3431.
9 Wright,J.H., Gottschling,D.E. and Zakian,V.A. (1992) *Genes Dev.*, **6**, 197–210.
10 Lejnine,S., Makarov,V.L. and Langmore,J.P. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 2393–2397.
11 Tommerup,H., Dousmanis,A. and de Lange,T. (1994) *Mol. Cell. Biol.*, **14**, 5777–5785.
12 Vourc'h,C., Taruscio,D., Boyle,A.L. and Ward,D.C. (1993) *Exp. Cell Res.*, **205**, 142–151.
13 Scherthan,H., Bähler,J. and Kohli,J. (1994) *J. Cell Biol.*, **127**, 273–285.
14 Price,C. (1992) *Curr. Opin. Cell. Biol.*, **4**, 379–384.
15 Cardenas,M.E., Bianchi,A. and De Lange,T. (1993) *Genes Dev.*, **7**, 883–894.
16 Petracek,M.E., Konkel,L.M.C., Kable,M.L. and Berman,J. (1994) *EMBO J.*, **13**, 3648–3658.
17 Lin,J.J.L. and Zakian,,V.A. (1994) *Nucleic Acids Res.*, **22**, 4906–4913.
18 Gilson,E., Roberge,M., Giraldo,R., Rhodes,D. and Gasser,S.M. (1993) *J. Mol. Biol.*, **231**, 293–310.
19 Palladino,F., Laroche,T., Gilson,E., Axelrod,A., Pillus,L. and Gasser,S.M. (1993) *Cell*, **75**, 543–555.
20 Palladino,F. and Gasser,S.M. (1994) *Curr. Opin. Cell Biol.*, **6**, 373–380.
21 Ishikawa,F., Matunis,M.J., Dreyfuss,G. and Cech,T.R. (1993) *Mol. Cell. Biol.*, **13**, 4301–4310.
22 McKay,S.J. and Cooke,H. (1992) *Nucleic Acids Res.*, **20**, 6461–6464.
23 Chong,L., van Steensel,B., Broccoli,D., Erdjument-Bromage,H., Hanish,J.P., Tempst,P. and De Lange,T. (1995) *Science*, **270**, 1663–1667.
24 Liu,Z. and Tye,B.K. (1990) *Genes Dev.*, **5**, 49–59.
25 Brigati,C., Kurtz,S., Balderes,D., Vidali,G. and Shore,D. (1993) *Mol. Cell. Biol.*, **13**, 1306–1314.
26 Louis,E.J., Naumova,E.S., Lee,A., Naumov,G. and Haber,J.E. (1994) *Genetics*, **136**, 789–802.
27 Gasser,S.M., Laroche,T., Falquet,J., Boy de la Tour,E. and Laemmli,U.K. (1986) *J. Mol. Biol.*, **188**, 613–629.
28 Dignam,J.D., Lebovitz,R.M. and Roeder,R.G. (1983) *Nucleic Acids Res.*, **11**, 1475–1489.
29 Lombard-Platet,G. and Jalinot,P. (1993) *Nucleic Acids Res.*, **21**, 3935–3942.
30 Gasser,S.M. and Laemmli,U.K. (1987) *Exp. Cell Res.*, **173**, 85–98.
31 Clément,J.M. and Popescu,O. (1991) *Bull. Inst. Pasteur*, **89**, 243–253.
32 Lane,D. (1988) *Antibodies: A Laboratory Manual.* Cold spring Harbor Laboratory Press, Cold Spring Harbor, NY.
33 Ogata,K., Morikawa,S., Nakamura,A., Inoue,T., Kanai,H., Sarai,A., Ishii,S. and Nishimura,Y. (1994) *Cell*, **79**, 639–648.
34 Altschul,S.F., Gish,W., Miller,W., Myers,W.W. and Lipman,D.J. (1990) *J. Mol. Biol.*, **215**, 403–410.
35 Higgins,D.G., Bleasby,A.J and Fuchs,R. (1991) *CABIOS*, **8**, 189–191.
36 Saitou,N and Nei,M. (1987) *Mol. Biol. Evol.*, **4**, 406–425.
37 Baranowskij,N., Frohberg,C., Prat,S. and Willmitzer,L. (1994) *EMBO J.*, **13**, 5383–5392.
38 Lugert,T. and Werr,W. (1994) *Plant Mol. Biol.*, **25**, 493–506.
39 da Costa e Silva,O., Klein,L., Typermelzer,E., Trezzini,G.F. and Hahlbrock,K. (1993) *Plant J.*, **4**, 125–135.
40 Larson,G.P., Castanotto,D., Rossi,J.J. and Malafa,M. (1994) *Gene*, **150**, 35–41.
41 Slamon,D.J., Boone,T.C., Murdock,D.C., Keith,D.E., Press,M.F., Larson,R.A. and Souza,L.M. (1986) *Science*, **233**, 347–351.
42 Nomura,N., Tanahashi,M., Matsui,M., Ishii,S., Date,T., Sasamato,S. and Ishizaki,R. (1988) *Nucleic Acids Res.*, **16**, 11075–11089.
43 Bouwmeester,T., Guehmann,S., El-Baradi,T., Kalkbrenner,F., van Wijk,I., Moelling,K. and Pieler,T. (1992) *Mech. Dev.*, **37**, 57–68.
44 Zervos,A.S, Gyuris,J. and Brent,R. (1993) *Cell*, **72**, 223–232.
45 Miskimins,W.K., Roberts,M.P., Mc Clelland,A. and Ruddle,F.H. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 6741–6744.
46 von Kries,J.P., Buhrmester,H. and Strätling,W.H. (1991) *Cell*, **64**, 123–135.
47 Wellinger,R.J. and Zakian,V.A. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 973–977.
48 Ijd,J.W.I., Wells,R.A., Baldini,A. and Reeders,S.T. (1991) *Nucleic Acids Res.*, **19**, 4780.
49 Biedenkapp,H., Borgmeyer,U., Sippel,A.E. and Klempnauer,K.H. (1988) *Nature*, **335**, 835–837.
50 Solano,R., Nieto,C., Avila,J., Canas,L., Diaz,I. and Paz-Ares,J. (1995) *EMBO J.*, **14**, 1773–1784.
51 de Lange,T. (1992) *EMBO J.*, **11**, 717–724.
52 Martines-Balbas,M., Dey,A., Rabindran,S.K., Ozato,K. and Wu,C. (1995) *Cell*, **83**, 29–38.
53 McEachern,M.J. and Blackburn,E.H. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 3453–3457.
54 Gilson,E. and Gasser,S.M. (1995) *Nucleic Acids Mol. Biol.*, **9**, 308–327.