A C-terminal Myb-extension domain defines a novel family of double-strand telomeric DNA binding proteins in Arabidopsis

Zemfira N. Karamysheva, Yulia V. Surovtseva, Laurent Vespa, Eugene V. Shakirov
and Dorothy E. Shippen

Department of Biochemistry and Biophysics, Texas A&M University, 2128 TAMU,
College Station, Texas 77843-2128

1To whom correspondence should be addressed:

Phone: 979-862-2342
Fax: 979-845-9274
Email: dshippen@tamu.edu

Running title: Double-strand telomere binding proteins from Arabidopsis
SUMMARY

Little is known about the protein composition of plant telomeres. We queried the *Arabidopsis thaliana* genome database in search of genes with similarity to the human telomere proteins hTRF1 and hTRF2. hTRF1/hTRF2 are distinguished by the presence of a single Myb-like domain in their C-terminus that is required for telomeric DNA binding *in vitro*. Twelve *Arabidopsis* genes fitting this criterion, dubbed TRF-like (TRFL), fell into two distinct gene families. Notably, TRFL family 1 possessed a highly conserved region C-terminal to the Myb domain called Myb-extension (Myb-ext) that is absent in TRFL family 2 and hTRF1/hTRF2. Immunoprecipitation experiments revealed that recombinant proteins from TRFL family 1, but not those from family 2, formed homo- and heterodimers *in vitro*. DNA binding studies with isolated C-terminal fragments from TRFL family 1 proteins, but not family 2, showed specific binding to double-stranded plant telomeric DNA *in vitro*. Removal of the Myb-ext domain from TRFL1, a family 1 member, abolished DNA binding. However, when the Myb-ext domain was introduced into the corresponding region in TRFL3, a family 2 member, telomeric DNA binding was observed. Thus, Myb-ext is required for binding plant telomeric DNA and defines a novel class of proteins in *Arabidopsis*. 
INTRODUCTION

Telomeres are the specialized nucleoprotein structures that comprise the natural ends of linear eukaryotic chromosomes and ensure their complete replication and stability (1, 2). In most eukaryotes, telomeric DNA is composed of tandem arrays of simple G-rich repeat sequences terminating in single-strand 3’ overhang, which is maintained through the action of the telomerase reverse transcriptase (1). Both the double- and single-strand region of the telomere are coated by non-histone proteins that provide protection for telomeric DNA and regulate telomerase access to the chromosome terminus. Proteins that bind double-strand telomeric DNA are typified in vertebrates by TRF1 and TRF2, and in budding and fission yeast by Rap1 and Taz1, respectively (3-8).

Human TRF1 behaves as a negative regulator of telomere length; over-expression results in telomere shortening, while a dominant negative allele induces telomere elongation (9, 10). hTRF1 mediates telomere length control through interactions with other telomere-associated factors including tankyrase (11), TIN2 (12), PinX1 (13), and Pot1 (14). Although hTRF2 contributes to telomere length regulation (10), its major function is to conceal telomere ends from detection as double-strand breaks (15, 16). Inhibition of hTRF2 in cultured human cells results in loss of the 3’ overhang and formation of covalently fused telomeres (15). In addition, compromised hTRF2 function culminates in cell cycle arrest and ATM/p53-mediated apoptosis (16).

The functional domains of vertebrate TRF1 and TRF2 have been studied in some detail (17). The two proteins have similar molecular masses (50-60kD), and resemble each other in domain structure. Although the N-terminus is highly acidic in hTRF1 and highly basic in hTRF2, both proteins harbor a centrally located flexible hinge region called the TRF homology (TRFH) domain that is required for homodimer formation and interactions with other telomere-associated proteins (18). The most strongly conserved feature is a Myb/homeodomain type helix-turn-helix DNA binding motif near the C-terminus (5, 19-22). The Myb domain is sufficient for specific binding to the telomeric DNA in vitro, but dimerization through the TRFH domain is required for TRF protein association with DNA in vivo (18). Biluad et al noted a sequence within the Myb domain of hTRF1 and hTRF2, VDLKDKWRT, that is also
conserved in yeast double strand telomere binding proteins and dubbed it the telobox consensus (19). An NMR structure of the Myb motif from hTRF1 revealed specific contacts between amino acid residues within the telobox consensus and the human telomere repeat sequence TTAGGG (22).

Although the plant telomere repeat sequence, TTTAGGG, is closely related to that of humans (23), almost nothing is known about the protein composition of plant telomeres. Several proteins have been shown to bind double-stranded telomeric DNA \textit{in vitro} (24-30). From \textit{Arabidopsis} these include several relatively small (~30kD) proteins with N-terminal Myb domains (30, 31). Two other telomeric DNA binding proteins, TRP1 and TBP1, more closely resemble vertebrate TRF1 and TRF2 in size (65kD and 70kD, respectively) and in architecture as they harbor a single Myb domain in their C-terminus. \textit{TRP1} from \textit{Arabidopsis} was identified in a yeast one-hybrid screen for proteins that bind double-strand telomeric DNA (26). Like vertebrate TRF1 and TRF2, full-length TRP1 shows a strong \textit{in vitro} preference for extended telomeric DNA tracts with a minimum binding site of five TTTAGGG repeats. Another \textit{Arabidopsis} gene that harbors a single Myb domain at its C-terminus is \textit{TBP1}. \textit{TBP1} encodes a homolog of the rice RTBP1, which has been shown to specifically bind plant telomeric DNA \textit{in vitro} (27, 28).

In this study we employed a BLAST search to identify \textit{Arabidopsis} homologs of hTRF1 and hTRF2 using their Myb domains as the query. Although \textit{Arabidopsis} harbors more than 100 genes with Myb domains (32), we found only 12 with a single Myb domain in their C-terminus that contains the telobox consensus motif. We designated this group of genes TRF-like (TRFL). Here we provide a molecular characterization of the TRFL proteins. Our data reveal that the TRFL genes encode two distinct families of proteins that differ dramatically in their amino acid sequences, DNA binding properties and protein interactions. Furthermore, we define a novel functional domain C-terminal to the Myb domain that is required for specific binding to duplex plant telomeric DNA.
EXPERIMENTAL PROCEDURES

Computer search for Myb-containing genes and evolutionary tree analysis

The Myb-domains of hTRF1, hTRF2, and *Arabidopsis* TRP1 were used in separate NCBI Blast searches to identify *Arabidopsis thaliana* genes predicted to encode proteins with a single Myb repeat at the C-terminus. The telobox consensus motif was used as an additional criterion (19). Twelve TRFL genes were identified (see GenBank accession numbers in Table 1). Multiple protein alignments were conducted using Oxford Molecular Group’s sequence analysis software MacVector 7.0 (Accelrys, San Diego, CA). A tree was constructed using the neighbor-joining method (33) with bootstrap mode (34).

Expression analysis, molecular cloning and production of recombinant proteins *in vitro*

RT-PCR was performed for 35 cycles using SuperScript III reverse transcriptase (Sigma) to determine whether TRFL genes were expressed in different *Arabidopsis* tissues. For most of the TRFL genes, primers flanked intron junctions, ruling out the possibility of genomic DNA contamination. Full-length cDNAs were obtained by RT-PCR using total RNA from flowers. Sequence analysis of the cloned cDNAs verified the NCBI annotations. For protein interaction studies, the full-length cDNAs of TRFL genes were subcloned into pET-28A and pCITE 4A. For DNA binding studies, PCR was used to amplify either full-length coding regions or the Myb domains and adjacent C-terminal region using plasmids that contained full-length cDNAs (see Table 1 for specific residues amplified). Overlapping PCR (35) was applied to create the fusion constructs shown in Figure 6A. A T7 phage promoter was introduced into all the constructs by PCR and the products were used as templates for *in vitro* transcription by T7 RNA polymerase (Stratagene). Transcripts were translated in either rabbit reticulocyte lysate (Promega) or a wheat germ translation system (Promega) in the presence or absence of \[^{35}\text{S}\]- methionine (Amersham). An aliquot of the labeled protein was used to verify the expected apparent molecular weight of translated protein products by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and autoradiography.
Electrophoresis mobility shift assay (EMSA)

Telomeric DNA probes and competitors were as follows: Arabidopsis
AGCATGCAGC(TTTAGGG)$_8$, AGCATGCAGC(TTTCAGGG)$_6$, AGCATGCAGC(TTTCAGGG)$_4$,
AGCATGCAGC(TTTCAGGG)$_2$, Paramecium AGCATGCAGC(TTTGGG)$_8$, human
AGCATGCAGC(TTTCAGGG)$_8$. Double-strand DNA probes were obtained by annealing the
_corresponding complementary oligonucleotides. For gel shift assays, peptides consisting of the Myb
domain and the remainder of the C-terminus (see above) were tested for DNA binding activity.
Preparation of the probes and binding reactions were carried out as described previously (36). Protein-
DNA complexes were loaded on a 4% PAGE in 0.5xTBE and subjected to electrophoresis for 4 h at 120
V. Dried gels were exposed for autoradiography.

Co-immunoprecipitation

For each TRFL protein analyzed, two constructs were made: one with a T7 protein tag and one
without. [$^{35}$S]- methionine labeled non-tagged proteins or T7 tagged unlabeled proteins were synthesized
in a TNT-coupled rabbit reticulocyte lysate translation system following the manufacturer’s
recommendations (Promega). Translation of T7 tagged proteins was verified in the presence of [$^{35}$S]-
methionine on a small aliquot from the same master mix. T7 tagged and untagged radiolabeled proteins
were combined and subjected to immunoprecipitation using agarose beads (Novagen) containing T7
monoclonal antibody as described (37). Precipitate and supernatant fractions were analyzed by SDS-
PAGE and autoradiography.

RESULTS

Identification and expression of TRFL genes in Arabidopsis

We performed a BLAST search to identify TRF-like (TRFL) genes in Arabidopsis thaliana.
Three query sequences were employed consisting of the Myb domains from human TRF1 and TRF2 and
from Arabidopsis TRP1. Twelve genes were uncovered that encode proteins with a single Myb domain at
their C-terminus (Figure 1, Table 1). As expected, the searches found TRP1 and TBP1 (26, 27), along with ten uncharacterized genes that were designated TRFL1-10. With the hTRF1 Myb domain as the query, the most closely related sequence was the Myb domain for TRFL6 (E-value=3e⁻⁰⁵), while for hTRF2, the Myb domain of TRFL3 displayed the most similarity (E-value=3e⁻⁰⁶).

A conceptual protein alignment of the C-terminal Myb-containing region of TRFL proteins with the corresponding regions in hTRF1 and hTRF2 is shown in Figure 1. The NMR structure of the Myb domain of hTRF1 bound to the human telomeric DNA repeat sequence reveals that the N-terminal arm of the first helix interacts with TT sequence in the minor groove, while the third helix recognizes TAGGG in the major groove (22). The corresponding regions are well conserved among the Arabidopsis TRFL proteins and display extensive sequence similarity to hTRF1 and hTRF2 (Figure 1A). The most highly conserved region within all of the TRFL proteins, VDLKDKWRT, lies within the telobox consensus.

Evolutionary analysis indicated the presence of two distinct TRFL gene families (Figure 1B). TRFL family 1 includes TBP1, TRP1, TRFL1, TRFL2, TRFL4, and TRFL9, while TRFL family 2 contains TRFL3, TRFL5-8 and TRFL10. The similarity within TRFL family 1 or family 2 members extends throughout their entire sequence, suggesting the DNA binding domains co-evolved with the remainder of the genes. Four of the TRFL genes reside in regions of the Arabidopsis genome known to be duplicated. These include TRP1 and TRFL1, which display 59% identity (70% similarity) and reside on chromosomes 5 and 3, respectively, and TRFL3 and TRFL6, which exhibit 52% identity, (62% similarity) and are found in a duplicated region on chromosome 1. Although TBP1 and TRFL9 are not located in a duplicated region, they display a remarkably high degree of similarity (54% identity; 66% similarity) and hence may have been derived from a recent gene duplication. The same may also be true for TRFL2 and TRFL4 (30% identity; 40% similarity), although they share somewhat less conservation overall.

Closer inspection of the predicted amino acid sequence for the TRFL proteins provided further evidence for distinct gene families. In all of the TRFL family 1 proteins, the region of amino acid conservation in the Myb domain extends further into the C-terminus, creating a Myb-extension (Myb-ext)
domain (Figure 1A). Interestingly, Myb-ext was absent from TRFL family 2 and from hTRF1 and hTRF2, which terminate immediately adjacent to the Myb domain. An additional region of substantial sequence conservation was detected in the central portion of TRFL family 1 members called the central domain (CD) (Figure 2A and B). This region bears no significant similarity to the TRFH domain in vertebrate TRF1/TRF2 proteins and is also absent in TRFL family 2. Aside from TRFL3 and TRFL6, which most likely represent a recent gene duplication, the remaining members of TRFL family 2 display no obvious sequence similarity outside their Myb domains.

RT-PCR analysis revealed that all of TRFL genes are expressed in Arabidopsis (Figure 3). Transcripts for each of the TRFL genes could be detected in all of the organs we examined (Figure 3; data not shown). Moreover, all of the genes were expressed at a relatively high level, with the exception of TRFL2 and TRFL4, which are both members of TRFL family 1. While TRFL2 and TRFL4 were ubiquitously expressed, their transcripts were scarce and could only be observed by nested PCR (Figure 3). Nevertheless, the constitutive expression of TRFL genes is consistent with a structural role in the telomere complex.

Proteins in TRFL family 1 can form homo- and heterodimers

Human TRF1 and TRF2 bind telomeric DNA as homodimers in vivo and in vitro (18, 38). Therefore, we tested whether recombinant TRFL proteins could form homodimers in vitro using co-immunoprecipitation experiments. For these studies, untagged [35S]-methionine labeled TRFL proteins were subjected to immunoprecipitation in the presence of the corresponding unlabeled T7-tagged protein using a T7 antibody (Figure 4A and B). Following immunoprecipitation, a homomeric interaction with tagged protein will allow a radiolabeled, but untagged protein to precipitate with the beads. Control reactions performed in the absence of tagged proteins showed no interaction between the T7 antibody and untagged proteins (Figure 4A and B, lane 2 in all panels). However, specific homomeric protein interactions were detected for TRP1, TRFL1, TRFL2, TRFL4 and TRFL9 (Figure 4A, upper panels; Table 1). All of these proteins are members of TRFL family 1. In contrast, TRFL3, TRFL5 and TRFL6,
members of TRFL family 2, did not exhibit the capacity for homodimerization \textit{in vitro} (Figure 4A, lower panels; Table 1).

Despite significant amino acid similarity in the TRFH domains of hTRF1 and hTRF2, steric hinderence prevents the formation of heterodimers (18). To determine whether TRFL proteins have the capacity to heterodimerize \textit{in vitro} additional co-immunoprecipitation experiments were performed with family 1 members TRP1, TRFL1 and TRFL9 (Figure 4B). These data showed that TRP1 and TRFL1 can heterodimerize, as can TRP1 and TRFL9. Interestingly, although TRP1 is closely related to TRFL1, it bears only limited sequence similarity to TRFL9 (38.2\% identity, 46.3\% similarity). Thus, it is conceivable that functionally distinct TRFL proteins directly interact \textit{in vivo}.

\textbf{DNA binding properties of TRFL proteins}

We used electrophoretic mobility shift assays (EMSA) to ask whether TRFL proteins bind telomeric DNA \textit{in vitro}. Initially, several full-length recombinant proteins expressed in rabbit reticulocyte lysate were tested for DNA binding. As expected from a previous study (26), full-length TRP1 bound a duplex telomeric DNA probe consisting of eight TTTAGGG repeats (AtTR8) and formed a discrete protein-DNA complex that migrated into the gel as well as complexes that stayed in the well (data not shown). Assays with other full-length TRFL proteins, including TRFL4 (Figure 5A), produced complexes that failed to exit the well. However, such complexes were sequence-specific (Figure 5A, left panel) and displayed length dependence in DNA binding (Figure 5A, right panel). To alleviate the concern of protein aggregation and to further analyze protein-DNA interactions, we tested whether the isolated Myb domains from TRFL proteins would bind telomeric DNA in EMSA. Previous studies showed that the corresponding regions of hTRF1, TRP1 and TBP1 were sufficient for specific interaction with duplex telomeric DNA \textit{in vitro} (26, 27, 39). We expressed the corresponding region of TRFL proteins (Myb domain and the adjacent C-terminal region) in wheat germ extract, which we found yielded a higher amount of recombinant TRFL protein than rabbit reticulocyte lysate. A summary of DNA binding results for the truncated TRFL proteins is shown in Table 1.
All members of TRFL family 1 bound specifically to plant telomeric DNA (Table 1; Figure 5A and B). Surprisingly, the minimal number of repeats required for binding varied among the different proteins. For example, TRFL1\textsuperscript{449-553} required a minimum of six telomere repeats for optimal binding (Figure 5B, lane 6). Although a weak interaction was detected with probes consisting of four telomere repeats (Figure 5B, lane 4), the complex appeared to be unstable as no discrete band was observed. In contrast, binding assays with TRFL9\textsuperscript{505-620} yielded a more discrete DNA-protein complex with a probe containing four telomeric repeats (Figure 5C, lane 4; data not shown). Although the relative migration position of the DNA-protein complexes appears different for TRFL1\textsuperscript{449-553} and TRFL9\textsuperscript{505-620} (Figure 5B, C), slightly different running conditions were used in the two gels. When the complexes were analyzed side-by-side on the same gel, they migrated at approximately the same position (Figure 5D). Competition experiments revealed that the association of TRFL family 1 proteins with plant telomeric DNA is specific (Figure 5B, lanes 8-12 and Figure 5C, lanes 10-13; data not shown). Competition was observed with probes consisting of double-stranded plant telomeric DNA, but not with non-specific duplex DNA or duplex DNA comprised of human or \textit{Paramecium} telomeric repeats (Figure 5A-C; data not shown). Assays with single-stranded telomeric DNA competitor also failed to compete with duplex plant telomeric DNA for binding (data not shown).

Strikingly, no DNA binding was detected in EMSA with TRFL family 2 members using either full-length proteins or isolated Myb domains (plus the remaining C-terminus) (Table 1; See Figure 6). As for TRFL family 1 members, all of the recombinant proteins generated from TRFL family 2 genes expressed well in the wheat germ translation system and were soluble (data not shown). Hence, the lack of telomeric DNA binding is likely to reflect inherent differences in the properties of the two TRFL families rather than protein misfolding. We conclude that, in contrast to the situation with human TRF1 (39), the capacity to bind to duplex plant telomeric DNA \textit{in vitro} is not conveyed solely by the presence of a C-terminal Myb domain, since all of the TRFL proteins we examined have this feature.

\textbf{The Myb-ext domain is necessary for telomeric DNA binding \textit{in vitro}.}
One intriguing difference between TRFL family 1 and 2 members is the presence of the Myb-ext domain in family 1 (Figure 1B). To examine the contribution of this region in telomeric DNA binding, we performed EMSA on several truncated versions of TRFL1 (Figure 6A). TRFL1\(^{449-540}\) (construct C-2) contains the Myb domain and Myb-ext, but lacks the remainder of the C-terminus. TRFL1\(^{449-509}\) (construct C-3) includes only the Myb domain. EMSA showed that TRFL1\(^{449-540}\) formed a complex with telomeric DNA that resembled that of the TRFL1\(^{449-553}\) control (Figure 6B, lanes 2 and 3). In contrast, no shifted complex was detected in the assay with TRFL1\(^{449-509}\) (Figure 6B, lane 4). One explanation for the inability of TRFL1\(^{449-509}\) to bind telomeric DNA is that the C-terminus is required for proper protein folding. To address this concern, we generated a construct in which the Myb domain of TRFL1 was fused directly to the C-terminus of TRFL3, a member of TRFL family 2 (Figure 6A; construct C-4). No binding to telomeric DNA was detected with this construct (Figure 6B, lane 5). Taken together, these data argue that the Myb-ext domain is required for binding to plant telomeric DNA.

We next asked whether the Myb-ext domain was sufficient to confer telomeric DNA binding to a TRFL family 2 protein. For this experiment, a chimeric protein was generated that consisted of the TRFL3 Myb domain fused to the TRFL1 Myb-ext and C-terminal region (Figure 6A, construct 6). Strikingly, a DNA-protein complex was observed with this construct (Figure 6B, lane 7). The capacity to bind telomeric DNA was not influenced by the residues C-terminal to the Myb-ext domain, as another chimeric construct (C-7) lacking this region also specifically bound telomeric DNA (Figure 6B, lanes 8-12). This result not only indicates that the Myb-ext domain is required for binding to plant telomeric DNA, but it also demonstrates that when placed next to a Myb domain of a TRFL family 2 protein it is sufficient to convey telomeric DNA binding specificity.

**DISCUSSION**

**TRF-like proteins from Arabidopsis**

The integrity of the nucleoprotein complex known as the telomere cap is crucial for genome stability and for cell proliferation in eukaryotes. Recent data indicate that plants display an extraordinary
tolerance to telomere dysfunction, as well as distinct mechanisms of telomere length regulation (40, 41). For example, mutations in the telomere-associated protein complex Ku70/80, which result in telomere shortening in yeast (42), lead to dramatic telomere elongation in Arabidopsis (43, 44). Unraveling these interesting evolutionary distinctions will require a greater understanding of telomere architecture in plants.

In this study we used a genomic approach to search for Arabidopsis homologs of the vertebrate TRF1 and TRF2 genes. The defining feature of this class of proteins is a single C-terminal domain reminiscent of the c-Myb DNA binding domain in oncoproteins (18, 22). Among Myb-containing proteins, TRF1 and TRF2 are unique in that their Myb domains harbor an additional telobox consensus motif, which contributes to the specific recognition of human telomeric DNA. Therefore, we searched for genes harboring a single C-terminal Myb domain with a telobox consensus motif and identified 12 TRF-like genes.

Since TRF1 and TRF2 are the only Myb-containing proteins known to bind directly to double-strand telomeric DNA in vertebrates, it was surprising to discover that Arabidopsis harbors so many TRFL genes. At least four TRFL genes (TRP1/TRFL1 and TRFL3/TRFL6) reside in regions of the Arabidopsis genome known to be duplicated. Two other pairs, TBP1 and TRFL9, and TRFL2 and TRFL4 may also reflect a recent duplication. Evidence for functional redundancy was obtained when we examined T-DNA insertion lines that disrupted the coding region in seven of the twelve TRFL genes (TRP1, TRFL1, 2, 3, 4, 5, 8). None of the single gene disruption mutants displayed defects in growth and development, nor showed perturbations in telomere length or genome stability (Vespa, Surovtseva, Karamysheva and Shippen, unpublished data). These data argue that at least a subset of TRFL genes may have overlapping function.

Nevertheless, the remarkable sequence divergence outside the DNA binding domain for many of the TRFL genes raises the distinct possibility of unique roles at telomeres or elsewhere in the Arabidopsis genome. The telomeric DNA binding protein Rap1 from budding yeast exhibits pleiotropic functions and is required for telomere maintenance as well as regulation of gene expression at internal sites in the
genome (45). Similarly, while the primary role of hTRF1 is associated with chromosome termini, TRF1 associates with interstitial blocks of telomeric DNA in some mammalian species (46). *Arabidopsis* also contains short stretches of interstitial telomeric DNA (typically one to two TTTAGGG repeats), predominantly located in promoter regions (47) that function in the control of gene expression (48, 49). Thus, some TRFL proteins may act in transcriptional regulation in a manner similar to Rap1. In this regard, it is noteworthy that TRFL family 1 members exhibit distinct minimal length requirements for telomeric DNA binding. For efficient *in vitro* binding, the Myb domain of TBP1 requires only two telomeric repeats while TRP1, TRFL4 and TRFL9 require a minimum of four repeats, and TRFL1 and TRFL2 need six repeats (this study; 26, 27). These apparent distinctions in telomere sequence recognition could potentially impact *in vivo* function. Proteins that bind efficiently to shorter stretches of telomeric DNA may preferentially act at promoters, while those that favor longer telomere tracts would have a higher probability of binding the long tracts of telomeric DNA associated with chromosome ends. Preliminary localization experiments indicate that all of the TRFL proteins accumulate in the *Arabidopsis* nucleus (Vespa, Kato, Lam and Shippen, unpublished data), however more in-depth analysis will be required to determine which of these proteins associate specifically with chromosome ends *in vivo*.

While all of the TRFL genes were identified on the basis of a C-terminal Myb-like domain, evolutionary tree analysis and inspection of the deduced protein sequence alignment indicates that they encode two distinct classes of proteins (Figure 1B). In addition to the Myb domain, all of the proteins in family 1 possess two other highly conserved regions, a Myb-ext domain and another centrally located CD domain. The CD domain was previously noted by Yaung and colleagues, who suggested that it might serve as a substrate for ubiquitination as it contains a region similar to the ubiquitin domain (29). For vertebrate TRF1 and TRF2, the only conserved region outside the Myb-domain is the TRFH domain, which not only facilitates formation of homodimers, but also interacts with other telomere-associated factors such human Rap1 and Tin2 (12, 50). Homodimerization of hTRF1 and hTRF2 is required for association with telomeric DNA *in vivo*, but these proteins cannot form heterodimers (18). In contrast, *Arabidopsis* TRFL family 1 proteins form both homo- and heterodimers *in vitro* (Figure 4). Whether the
CD domain in TRFL family 1 proteins serves as an interaction interface analogous to the TRFH domain in hTRF1 and hTRF2 remains to be determined. However, the capacity to form both homo-and heterodimers in vitro suggests that TRFL family 1 members could participate in complex structural and functional regulation in vivo.

A novel domain is necessary for TRFL proteins to bind plant telomeric DNA

One of the most striking differences among TRFL family 1 and family 2 proteins is the presence of a highly conserved Myb-ext domain in family 1. Interestingly, the maize initiator-binding protein (IBP1) and the parsley BoxP-binding factor (BPF1) also harbor a Myb domain (51, 52) with an adjacent region with striking similarity to Myb-ext (Karamysheva and Shippen, unpublished data). IBP1 is known to interact at the transcription start site of the Shrunken promoter containing a perfect telomeric repeat AGGGTTT, while BPF1 binds a series of GT-rich motifs. Based on their similarity to the TRFL family 1 from Arabidopsis, we predict that these proteins may have the capacity to efficiently bind longer stretches of plant telomeric DNA.

It is intriguing that the Myb domain of hTRF1 is sufficient for binding human telomeric repeat DNA in vitro, but this is not the case for TRFL proteins from Arabidopsis. Our data demonstrate that this latter class of proteins requires a more extended domain for telomeric DNA interactions. We found that Myb-ext is both necessary and sufficient to allow the Myb domains of TRFL proteins to bind double strand plant telomeric DNA. Precisely how Myb-ext contributes to telomeric DNA binding is unclear. One possibility is that it stabilizes the essential protein contacts between the Myb domain and telomeric DNA.

In eukaryotes a Myb-like domain is a defining feature of double-strand telomeric DNA binding proteins. The large number of such proteins in Arabidopsis confounds the identification of bona-fide telomere-associated factors. Although the Myb domains of TRFL family 2 proteins display even greater sequence similarity to the corresponding region of hTRF1 and hTRF2 than TRFL family 1 proteins, TRFL family 2 proteins fail to bind plant telomeric DNA in vitro. We cannot exclude the possibility that
TRFL family 2 proteins associate with telomeres *in vivo* through protein interactions in a manner similar to human Rap1 (reference). Furthermore, it is evident that TRFL family 1 proteins are not the only factors capable of specific binding to double-strand plant telomeric DNA *in vitro*. While the architecture of TRFL family 1 proteins closely resembles the known double-strand telomere binding proteins in higher eukaryotes, the SMH-like proteins from maize and *Arabidopsis* plant harbor a single Myb domain in their N-terminus, but lack an associated Myb-ext domain, yet still bind telomeric DNA *in vitro* (30, 31). Thus, TRFL family 1 and SMH-like proteins apparently represent two distinct classes of telomeric DNA binding proteins. Deciphering the functions and interactions of such proteins may provide new insight into the structure and complex regulation of the telomere cap.

**ACKNOWLEDGEMENTS**

We thank Tom McKnight and members of the Shippen lab for many helpful discussions. We are also grateful to Tom McKnight, Carolyn Price and Jeff Kapler for critically reading the manuscript. This work was supported by grants from NIH (GM65383) and NSF (MCB0349993) to D.E.S.

**FIGURE LEGENDS**

**Figure 1.** Identification of two TRFL gene families in *Arabidopsis*. A, Comparison of the C-terminal region of TRFL proteins with the corresponding region of hTRF1 and hTRF2. Amino acid sequences were obtained by conceptual translation of corresponding cDNAs. Sequences used for alignment include Myb domains and adjacent C-terminal region. Identical residues are shown in dark gray and similar residues in light gray. The likely positions of the three helices, which comprise the Myb domain of hTRF1, including the telobox consensus motif, are shown in boxes above the hTRF1 sequence. Asterisks indicate amino acid residues in hTRF1 that make specific contacts with human telomeric repeat DNA in the NMR structure (22). Brackets denote TRFL family 1 and 2 members. The Myb-ext domain in TRFL family 1 proteins is shown. B, Gene family analysis of TRFL proteins from *Arabidopsis*. The tree was
constructed by the neighbor-joining method after sequence alignment of the two TRFL protein families. The number at each node indicates the percentage of 10,000 bootstrap replicates for statistical support.

**Figure 2.** Sequence alignment and domain structure of TRFL family 1 members. A, Alignment of full-length TRFL family 1 proteins with hTRF1. The CD, Myb and Myb-ext domains are indicated by an over-line. The TRFH domain in hTRF1 is denoted by a wide gray box under the hTRF1 sequence. B, Comparison of the domain structures of hTRF1 and TRF2 with TRP1 and TRFL9.

**Figure 3.** Expression of TRFL genes in *Arabidopsis*. RT-PCR analysis was performed on RNA isolated from the sources indicated. All of the TRFL genes were analyzed found to be expressed in all of the organs examined. Results from TRP1, TRFL 1, 3, 4 and 9 are shown as examples. Transcripts from TRFL4 were detected only after nested PCR.

**Figure 4.** Dimerization of TRFL proteins *in vitro*. Co-immunoprecipitation experiments were conducted with full-length TRFL proteins expressed in rabbit reticulocyte lysate. In each experiment, one protein was labeled with $^{35}$S-Met (asterisk) and the other was unlabeled, but contained a T7-tag. Control reactions performed in the absence of T7-tagged proteins showed no non-specific binding of $^{35}$S-Met labeled protein to the beads. Examples of co-immunoprecipitation assay results for homodimeric (A) and heterodimeric (B) interactions are shown.

**Figure 5.** DNA binding properties of *Arabidopsis* TRFL proteins. EMSA was conducted with full-length TRFL4 expressed in rabbit reticulocyte lysate (A) or isolated C-terminal regions of TRFL1 and TRFL9 (B and C) expressed in wheat germ extract. $^{32}$P-labeled double-stranded DNA probes consisted of two (AtTR2), four (AtTR4), six (AtTR6) or eight (AtTR8) *A. thaliana* telomeric repeats. EMSA controls with unprogrammed translation extract are shown in lane 1 of panel A, lanes 1, 3, 5 and 7 of panel B, and lanes 1, 3, 5 and 7 of panel C. Competition experiments were carried out with a 50X or 100X excess of
unlabeled competitor DNA containing eight telomeric DNA repeats of *A. thaliana* (At), human (Hu), *Paramecium* (Pa) DNA, or a non-specific (NS) DNA sequence. Arrows denote position of protein-DNA complexes. Arrows and brackets denote the position of protein-DNA complexes.

**Figure 6.** Analysis of the Myb-ext domain in telomeric DNA binding. A, Schematic representation of truncation and chimeric fusion constructs. The Myb domain is highlighted in gray, the Myb-ext domain in black and the adjacent C-terminal region in white. A summary of results from EMSA is shown on the right. B, EMSA reactions were performed with the constructs shown in panel A as described in Figure 5. The probe was AtTR8. Arrows denote the position of protein-DNA complexes.
**Table 1.** DNA binding properties and protein interactions of TRFL proteins

| Locus | Accession number | Protein Regions expressed for DNA binding | Min. telo repeats bound | Homodimer | Heterodimer |
|-------|------------------|------------------------------------------|-------------------------|-----------|-------------|
| At5g13820 | NP_196886 | TBP1 | ND | 2<sup>1</sup> | ND | ND |
| At5g59430 | NP_200751 | TRP1 | ND | 4<sup>2</sup> | Yes | TRFL1,TRFL9 |
| At3g46590 | NP_190243 | TRFL1 | 449-553 | 6 | Yes | TRP1 |
| At1g07540 | NP_172234 | TRFL2 | 516-622 | 6 | Yes | ND |
| At3g53790 | NP_190947 | TRFL4 | 311-400 | 4 | Yes | ND |
| At3g12560 | NP_187862 | TRFL9 | 505-620 | 4 | Yes | TRP1 |
| At1g17460 | NP_564025 | TRFL3 | 483-605 | No binding | No | ND |
| At1g15720 | NP_173024 | TRFL5 | 303-390; 334-390 | No binding | No | ND |
| At1g72650 | NP_565045 | TRFL6 | 530-624 | No binding | No | ND |
| At1g06910 | NP_683280 | TRFL7 | 331-390 | No binding | ND | ND |
| At2g37025 | NP_671929 | TRFL8 | 297-404 | No binding | ND | ND |
| At5g03780 | NP_195998 | TRFL10 | 344-420 | No binding | ND | ND |

<sup>1</sup>Hwang et al., 2001, <sup>2</sup>Chen et al., 2001. ND=not determined.
REFERENCES

1. Blackburn, E. H. (1991) *Nature* 350, 569-573
2. Zakian, V. A. (1996) *Annual Review of Genetics* 30, 141-172
3. Chong, L., van Steensel, B., Broccoli, D., Erdjument-Bromage, H., Hanish, J., Tempst, P., and de Lange, T. (1995) *Science* 270, 1663-1667
4. Bilaud, T., Brun, C., Ancelin, K., Koering, C. E., Laroche, T., and Gilson, E. (1997) *Nature Genetics* 17, 236-239
5. Broccoli, D., Smogorzewska, A., Chong, L., and de Lange, T. (1997) *Nature Genetics* 17, 231-235
6. Shore, D., and Nasmyth, K. (1987) *Cell* 51, 721-732
7. Cooper, J. P., Nimmo, E. R., Allshire, R. C., and Cech, T. R. (1997) *Nature* 385, 744-747
8. De Ryckenaer, M., Venkatesan, R. N., Wei, C., and Price, C. M. (2003) *Biochem J* 372, 87-96
9. van Steensel, B., and de Lange, T. (1997) *Nature* 385, 740-743
10. Smogorzewska, A., van Steensel, B., Bianchi, A., Oelmann, S., Schaefer, M. R., Schnapp, G., and de Lange, T. (2000) *Molecular & Cellular Biology* 20, 1659-1668
11. Smith, S., Giriat, I., Schmitt, A., and de Lange, T. (1998) *Science* 282, 1484-1487
12. Kim, S. H., Kaminker, P., and Campisi, J. (1999) *Nature Genetics* 23, 405-412
13. Zhou, X. Z., and Lu, K. P. (2001) *Cell* 107, 347-359
14. Fairall, L., Chapman, L., Moss, H., de Lange, T., and Rhodes, D. (2001) *Molecular Cell* 8, 351-361
15. Bialaud, T., Koering, C. E., Binet-Brasselet, E., Ancelin, K., Pollice, A., Gasser, S. M., and Gilson, E. (1996) *Nucleic Acids Research* 24, 1294-1303
16. Konig, P., Giraldo, R., Chapman, L., and Rhodes, D. (1996) *Cell* 85, 125-136
17. Nishikawa, T., Nagadoi, A., Yoshimura, S., Aimoto, S., and Nishimura, Y. (1998) *Structure* 6, 1057-1065
18. Richards, E. J., and Ausubel, F. M. (1988) *Cell* 53, 127-136
19. Zentgraf, U. (1995) *Plant Molecular Biology* 27, 467-475
20. Zentgraf, U., Hinderhofer, K., and Kolb, D. (2000) *Plant Molecular Biology* 42, 429-438
21. Chen, C. M., Wang, C. T., and Ho, C. H. (2001) *J Biol Chem* 276, 16511-16519
22. Hwang, M. G., Chung, I. K., Kang, B. G., and Cho, M. H. (2001) *FEBS Letters* 503, 35-40
23. Yu, E. Y., Kim, S. E., Kim, J. H., Ko, J. H., Cho, M. H., and Chung, I. K. (2000) *J Biol Chem* 275, 24208-24214
24. Yang, S. W., Kim, D. H., Lee, J. J., Chun, Y. J., Lee, J. H., Kim, Y. J., Chung, I. K., and Kim, W. T. (2003) *J Biol Chem* 278, 21395-21407
25. Schrumpfova, P., Kuchar, M., Mikova, G., Skrisovska, L., Kubicarova, T., and Fajkus, J. (2004) *Genome* 47, 316-324
26. Mariani, C. O., Bordoli, S. J., Goltz, M., Santarella, R. A., Jackson, L. P., Danilevskaya, O., Beckstette, M., Meeley, R., and Bass, H. W. (2003) *Plant Physiol* 133, 1336-1350
32. Kranz, H. D., Denekamp, M., Greco, R., Jin, H., Leyva, A., Meissner, R. C., Petroni, K., Urzainqui, A., Bevan, M., Martin, C., Smeekens, S., Tonelli, C., Paz-Ares, J., and Weisshaar, B. (1998) *Plant J* **16**, 263-276
33. Saitou, N., and Nei, M. (1987) *Mol Biol Evol* **4**, 406-425
34. Felsenstein, J. (1985) *Evolution* **39**, 783-791
35. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. (1989) *Gene* **77**, 51-59
36. Hall, A. R., and Milner, J. (1997) *Oncogene* **14**, 1371-1376
37. Bryan, T. M., Goodrich, K. J., and Cech, T. R. (2000) *J Biol Chem* **275**, 24199-2420
38. Konig, P., Fairall, L., and Rhodes, D. (1998) *Nucleic Acids Res* **26**, 1731-1740
39. Ilia, K., Watson, J. M., Parkey, J., and Shippen, D. E. (2002) *Plant Mol Biol* **48**, 331-337
40. Regad, F., Lebas, M., and Lescure, B. (1994) *J Mol Biol* **239**, 163-169
41. Tremousaygue, D., Manevski, A., Bardet, C., Lescure, N., and Lescure, B. (1999) *Plant J* **20**, 553-561
42. Tremousaygue, D., Garnier, L., Bardet, C., Dabos, P., Herve, C., and Lescure, B. (2003) *Plant J* **33**, 957-966
43. Li, B., and de Lange, T. (2003) *Mol Biol Cell* **14**, 5060-5068
44. Lugert, T., and Werr, W. (1994) *Plant Mol Biol* **25**, 493-506
45. da Costa e Silva, O., Klein, L., Schmelzer, E., Trezzini, G. F., and Hahlbrock, K. (1993) *Plant J* **4**, 125-135
Myb-like domain

Telobox consensus

A

B

FIGURE 1

Downloaded from http://www.jbc.org/ on March 22, 2020.
FIGURE 2
FIGURE 3
FIGURE 4

A

|       | TRP1* | T7Ab | T7Ab + T7-TRFL3 |
|-------|-------|------|-----------------|
|       |       | S    | S               |
| 1     | 2     |      | S               |
| 3     | 4     |      | P               |

B

|       | TRP1* | T7Ab | T7Ab + T7-TRFL1 |
|-------|-------|------|-----------------|
|       |       | S    | S               |
| 1     | 2     |      | S               |
| 3     | 4     |      | P               |
FIGURE 5
FIGURE 6
A C-terminal Myb-extension domain defines a novel family of double-strand telomeric DNA binding proteins in Arabidopsis

Zemfira N. Karamysheva, Yulia V. Surovtseva, Laurent Vespa, Eugene V. Shakirov and Dorothy E. Shippen

J. Biol. Chem. published online September 10, 2004

Access the most updated version of this article at doi: 10.1074/jbc.M407938200

Alerts:
  • When this article is cited
  • When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts