Structure
Article

Structure of the RNA-Binding Domain of Telomerase: Implications for RNA Recognition and Binding

Susan Rouda[1] and Emmanuel Skordalakes[1,*]
[1]Gene Expression and Regulation Program, The Wistar Institute, University of Pennsylvania, 3601 Spruce Street, Philadelphia, PA 19104, USA
*Correspondence: skorda@wistar.org
DOI 10.1016/j.str.2007.09.007

SUMMARY

Telomerase, a ribonucleoprotein complex, replicates the linear ends of eukaryotic chromosomes, thus taking care of the "end of replication problem." TERT contains an essential and universally conserved domain (TRBD) that makes extensive contacts with the RNA (TER) component of the holoenzyme, and this interaction is thought to facilitate TERT/TER assembly and repeat-addition processivity. Here, we present a high-resolution structure of TRBD from *Tetrahymena thermophila*. The nearly all-helical structure comprises a nucleic acid-binding fold suitable for TER binding. An extended pocket on the surface of the protein, formed by two conserved motifs (CP and T motifs) comprises TRBD's RNA-binding pocket. The width and the chemical nature of this pocket suggest that it binds both single- and double-stranded RNA, possibly stem I, and the template boundary element (TBE). Moreover, the structure provides clues into the role of this domain in TERT/TER stabilization and telomerase repeat-addition processivity.

INTRODUCTION

Any organism with linear chromosomes faces a substantial obstacle in maintaining the terminal sequence of its DNA often referred to as the "end replication problem" (Blackburn, 1984; Cavalier-Smith, 1974; Cech and Lingner, 1997; Lingner et al., 1995; Lundblad, 1997; Ohki et al., 2001). Eukaryotic cells overcome this problem through the use of a specialized DNA polymerase, called telomerase. Telomerase adds tandem, G-rich DNA repeats (telomeres) to the 3′ end of linear chromosomes that serve to protect chromosomes from loss of genetic information, chromosome end-to-end fusion, genomic instability, and senescence (Autexier and Lue, 2006; Blackburn and Gall, 1978; Chatziantoniou, 2001; Collins, 1996; Dong et al., 2005).

The core telomerase holoenzyme is an RNA-dependent DNA polymerase (TERT) paired with an RNA molecule (TER) that serves as a template for the addition of telomeric sequences (Blackburn, 2000; Lamond, 1989; Miller and Collins, 2002; Miller et al., 2000; Shippen-Lentz and Blackburn, 1990). TERT consists of four functional domains, one of which shares similarities with the HIV reverse transcriptases (RTs) in that it contains key signature motifs that are hallmarks of this family of proteins (Autexier and Lue, 2006; Bryan et al., 1998; Lee et al., 2003; Peng et al., 2001). The RT domain, which contains the active site of telomerase, is thought to be involved in loose associations with the RNA template (Collins and Gandhi, 1998; Jacobs et al., 2005). TERT, however, is unique, compared to other RTs, in that it contains two domains N-terminal to the RT domain that are essential for function. These include the far N-terminal domain (TEN), which is the least conserved among phylogenetic groups, but is required for appropriate human, yeast, and ciliated protozoa telomerase activity in vitro and telomere maintenance in vivo (Friedman and Cech, 1999; Friedman et al., 2003). The TEN domain has both DNA- and RNA-binding properties. DNA binding facilitates loading of telomerase to the chromosomes, whereas RNA binding is nonspecific. The precise role of the TEN-TER interaction is unclear at this stage (Hammond et al., 1997; Jacobs et al., 2006; Wyatt et al., 2007). A third domain, TRBD, is located between the TEN and RT domains, and, unlike the TEN domain, is highly conserved among phylogenetic groups and is essential for telomerase function both in vitro and in vivo (Lai et al., 2001). The TRBD contains key signature motifs (CP and T motifs) implicated in RNA recognition and binding and makes extensive contacts with stem I and the TBE of TER, both of which are located upstream of the template (Bryan et al., 2000; Cunningham and Collins, 2005; Lai et al., 2001, 2002; Miller et al., 2000; O'Connor et al., 2005). The TRBD-TER interaction is required for the proper assembly and enzymatic activity of the holoenzyme both in vitro and in vivo and is thought to play an important role (although indirect) in the faithful addition of multiple, identical telomeric repeats at the ends of chromosomes (Lai et al., 2001, 2002, 2003).

Unlike TERT, TER varies considerably in size between species. For example, in *Tetrahymena thermophila* TER is only 159 nucleotides long (Greider and Blackburn,

1989), whereas yeast harbors an unusually long TER of 1167 nucleotides (Zappulla and Cech, 2004). Despite the large differences in size and structure, the core structural elements of TER are conserved among phylogenetic groups, suggesting a common mechanism of telomere replication among organisms (Chen et al., 2000; Chen and Greider, 2003, 2004; Ly et al., 2003; Theimer and Feigon, 2006). These include the template, which associates loosely with the RT domain and provides the code for telomere synthesis, and the TBE, which partly regulates telomerase's repeat-addition processivity. In *Tetrahymena thermophila*, the TBE is formed by stem II and the flanking single-stranded regions, and it is located upstream and in close proximity to the template (Lai et al., 2002, 2003; Licht and Collins, 1999). Low-affinity TERT-binding sites are also found in helix IV and the template recognition element (TRE) of *Tetrahymena thermophila* TER.

TERT function is regulated by a number of proteins, some of which act by direct association with the TERT/TER complex, whereas others act by regulating access of telomerase to the chromosome end through their association with the telomeric DNA (Aisner et al., 2002; Cong et al., 2002; Dong et al., 2005; Loayza and de Lange, 2004; Smogorzewska and de Lange, 2004; Smogorzewska et al., 2000; Witkin and Collins, 2004; Witkin et al., 2007). For example, p65 in the ciliated protozoan *Tetrahymena thermophila* and its homolog p43 in *Euplotes aediculatus* are integral components of the telomerase holoenzyme (Aigner and Cech, 2004; Aigner et al., 2003; O'Connor and Collins, 2006; Prathapam et al., 2005; Witkin and Collins, 2004; Witkin et al., 2007). Both p65 and p43 are thought to bind and fold TER, a process required for the proper assembly and full activity of the holoenzyme. In yeast, recruitment and subsequent upregulation of telomerase activity requires the telomerase-associated protein Est1 (Evans and Lundblad, 2002; Hughes et al., 1997; Lundblad, 2003; Lundblad and Blackburn, 1990; Reichenbach et al., 2003; Snow et al., 2003). Est1 binds the RNA component of telomerase, an interaction that facilitates recruitment of the holoenzyme to the eukaryotic chromosome ends via its interaction with the telomere-binding protein Cdc13 (Chandra et al., 2001; Evans and Lundblad, 1999; Lustig, 2001; Pennock et al., 2001).

How telomerase and associated regulatory factors physically interact and function with each other to maintain appropriate telomere length is poorly understood. Structural and biochemical characterization of these factors, both in isolation and complexed with one another, is needed to understand a host of questions regarding telomere biology. For example, how does the interaction of the TRBD domain with stem I and the TBE of TER facilitate the proper assembly and promote the repeat-addition processivity of the holenzyme? It is also unclear if the TRBD-TER interaction allows for TRBD movement during nucleotide addition. The structure of the TRBD domain of TERT will provide clues to the physical interactions between TERT and TER, which, in turn, will help us better understand the role of this domain in telomerase assembly and telomere replication.

In order to address some of these questions, we solved and refined the structure of the RNA-binding domain of telomerase from *Tetrahymena thermophila* to 1.71 Å resolution. The structure reveals a nearly all-helical nucleic acid-binding fold. The molecule is divided into two asymmetric halves joined together by several loops forming a boomerang-like structure. The nucleic acid-binding site is formed by two conserved motifs (CP and T motifs), which are located at a crevice formed by the two lobes of the molecule, and spans two-thirds of the width of the molecule. Part of this pocket is narrow and hydrophobic in nature, and part of it is wide and positively charged. Close inspection of this cavity indicates that it could accommodate both single- and double-stranded RNA, which is consistent with previous studies that suggest that it interacts with stem I and the TBE of TER.

RESULTS AND DISCUSSION

RBD Structure

To explore the role of the essential RNA-binding domain of telomerase (TRBD), we cloned, overexpressed, and purified to homogeneity a construct identified by limited proteolysis that contained residues 254–519 from *Tetrahymena thermophila* (Figure 1A). This protein construct is monomeric in solution, as indicated by both gel filtration and dynamic light scattering (data not shown). Crystals of this construct grew readily and diffracted to 1.71 Å resolution (Table 1). The protein was phased to 2.7 Å resolution by the multiwavelength anomalous dispersion (MAD) method by using a holmium derivative, and the phases were extended with the native data set to 1.71 Å resolution (Table 1). In the refined structure, there is clear density for residues 259–265 and 277–519.

The structure contains 12 α helices linked together by several long loops and 2 short β strands (Figures 1B–1D). The helices are organized so that the molecule is divided into two asymmetric halves linked together by three extended loops. The larger half consists of nine α helices, one of which (α11) runs along the middle of the domain and spans its entire length, making contacts with every one of the other eight helices. The smaller half of the molecule consists of three helices (α4, α5, and α12), all of which are arranged at an ~120° angle to the plane of the larger half of the protein. The smaller half of the protein is somewhat more flexible than the larger half, as suggested by its high B factors, reflecting the intrinsic mobility of this region, and may result from the absence of observable contacts with the RNA substrate (Figure 4C). An interesting feature of the structure is a β hairpin formed by the 15 residues that connect helices α11 and α12 of the larger and the smaller halves, respectively (Figures 1B–1D). The β hairpin protrudes from the base of the crevice formed by the two halves of the protein and stands at a 45° angle to the plane of the smaller half of the molecule. The positioning and the fact that this hairpin is well defined in the density could be attributed to helix α7 and the loop that connects it to helix α8. Both of these elements are conveniently positioned at the back of this hairpin, holding it in place. A

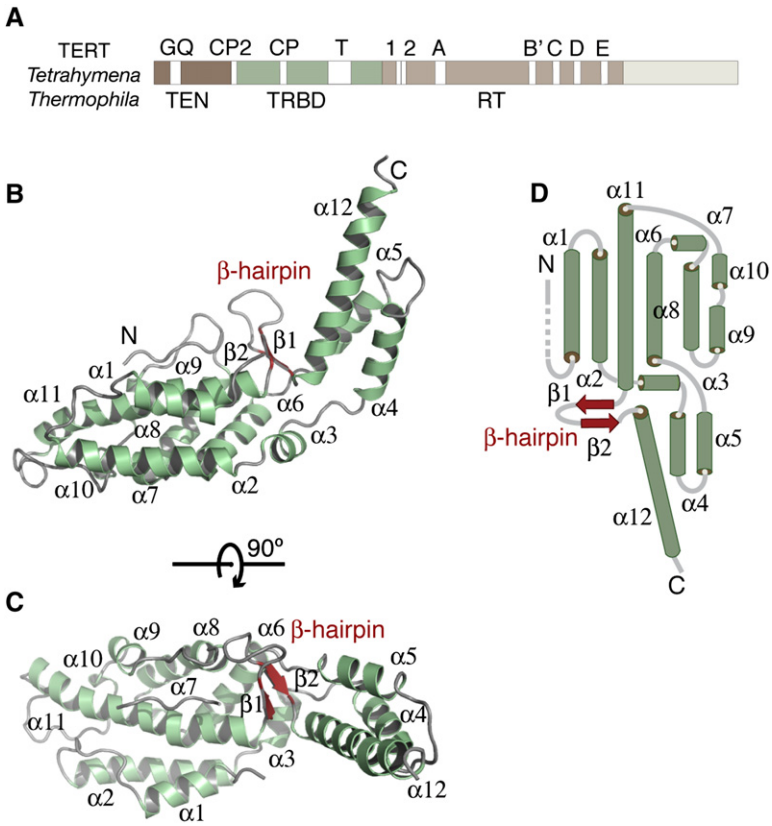

Figure 1. TRBD Structure and Fold

(A) Primary structure of *Tetrahymena thermophila* TERT showing the functional domains and conserved motifs.

(B) Structure of TRBD from *Tetrahymena thermophila*; the structure consists of two asymmetric halves (large to the left, small to the right of the plane of the paper) linked by three extended loops. The β hairpin that forms part of the T pocket is colored red. Secondary structure elements are labeled.

(C) (B) rotated 90° for clarity.

(D) Schematic of the TRBD fold.

search in the protein structure database with the Dali server (Holm and Sander, 1996) produced no structural homologs, suggesting that the TRBD domain of telomerase is a novel nucleic acid-binding fold. The overall organization of the two halves of the protein has significant implications for nucleic acid recognition and binding.

The TRBD RNA-Binding Motifs

The ability of the TRBD domain to interact with TER has been attributed to two conserved motifs known as the CP and T motifs, whereas a third motif, known as the QFP motif, is thought to be important for RNP assembly (Figure 2A) (Bosoy et al., 2003; Bryan et al., 2000; Jacobs et al., 2005; Xia et al., 2000). The TRBD structure shows that the QFP motif is formed by several mostly hydrophobic residues, which are located on the larger half of the molecule. The QFP residues are buried within the core of this domain, and they make extensive contacts with the surrounding residues, aiding in the fold of the protein (Figure 3A). The location and the contacts of the QFP residues suggest that they are not directly involved in nucleic acid binding.

The T motif is located at the center of the molecule, where the two halves of the protein meet, and it consists of residues that form both part of the β hairpin and helix α12. Together, these structural elements form a narrow (~10 Å), well-defined pocket (T pocket) that is lined by several solvent-exposed and highly conserved residues (Figures 2A, 2C, and 3B). Of particular note are the side chains of the invariant residues Y477 and W496, which are part of the β hairpin and helix α12, respectively. Together, these residues form a "hydrophobic pincer" that could sandwich the purine/pirimidine moiety of an interacting RNA nucleotide (Figure 3B). In this structure, the side chains of Y477 and W496 are only 4 Å apart, which is not sufficient to accommodate a nucleotide base. Insertion of a base between the two side chains would require structural rearrangement of the T pocket, possibly splaying the two halves of the molecules apart. In addition to its hydrophobic part, the T pocket also contains several hydrophilic residues, such as R492 and K493, both of which are solvent exposed and are located at the interface of the T and CP pockets connecting the two together.

The CP motif is formed by helix α3 and the following loop. In contrast to the T motif, which is a narrow, well-defined pocket, the CP motif comprises a shallow, wide (~20 Å), highly positively charged cavity located adjacent and beneath the entry of the T pocket (Figures 2B and 3B). Several of the conserved residues that form the CP motif are buried in the core of the larger half or the region that connects the two halves of the molecule and are contributing to the protein fold. Of particular interest are residues L327, C331, L333, and P334, all of which are buried and make direct contacts with structural elements of the T motif, thus aiding in the formation of both the CP and the T pockets. For example, L327 and C331 are within van der Waal contacts of the large hydrophobic side chain of the invariant F476 and the aliphatic part of the side chain of

Table 1. Data Collection Statistics

	Native	Holmium Derivative	
$TRBD_{(254–519)}$	λ	Ho-λ1	Ho-λ2
Wavelength (Å)	0.9795	1.5347	1.5595
Space group	$P2_1$	$P2_1$	$P2_1$
Cell dimensions (Å)	39.4, 67.2, 51.5, 90.7	39.2, 68.2, 50.1, 91.6	39.2, 68.2, 50.1, 91.6
Resolution (Å)	20–1.71 (1.77–1.71)[a]	50–2.59 (2.69–2.59)	50–2.63 (3.02–2.63)
Redundancy	3.7 (3.0)	1.7 (1.8)	1.7 (1.8)
Completeness (%)	99.3 (93.3)	92.5 (88.1)	92.9 (88.7)
R_{sym} (%)	4.7 (48.1)	7.3 (23.8)	7.0 (21.5)
I/σ (I)	27.3 (2.6)	9 (3.4)	9.4 (3.7)
Phasing Analysis			
Resolution (Å)	50–2.7		
Number of sites	2		
Mean figure of merit (FOM)	0.43		

[a] Values in parentheses correspond to the highest-resolution shell.

the conserved R492, both of which form part of the β hairpin. Interestingly, R492 is located at the base of helix α12, and its contact with L327, C331, and L333 partially helps position this helix at a 45° angle of the plane that runs parallel with the larger half of the molecule, thus further facilitating the formation of the T pocket. Moreover, the interaction of R492 with L327, C331, and L333 helps position the guanidine moiety, the only solvent-exposed part of this residue, at the interface formed by the CP and T pockets. The CP pocket also contains several surface-exposed, conserved residues that are mainly hydrophilic in nature (Figures 2C and 3B). These include K328 and K329, both of which are located beneath the T pocket and in close proximity to R492 and K493, together forming a single large, positively charged surface area that almost spans the entire side of the molecule (Figure 4A).

TRBD Structure and Existing Mutants

To date, several mutants of TERT that affect RNA binding and telomerase activity have been isolated. Several of these mutants are found in the TRBD domain and specifically within the T and CP motifs. Single- and double- as well as stretches of 4–10 amino acid alanine substitutions within these two motifs showed moderate to severe loss (20%–100%) of RNA-binding affinity and polymerase activity when compared to the wild-type enzyme (Bryan et al., 2000; Lai et al., 2002; Miller et al., 2000).

One set of mutants, F476A, Y477A, T479A, E480A, R492A, and W496A, showed severe loss (80%–100%) of RNA-binding affinity and telomerase activity, suggesting that these residues mediate direct contacts with the RNA substrate (Bryan et al., 2000; Lai et al., 2002). All 5 residues are part of the T motif, and, with the exception of F476, all of their side chains are solvent exposed (Figure 3B). In the structure, both Y477 and W496 are located at the base of the T pocket, and their side chains form a "hydrophobic pincer." Assuming that the solvent-exposed side chains of these residues are involved in stacking interactions with the single-stranded RNA, mutating them to small alanines would most likely compromise substrate binding, which would explain the dramatic loss of RNA-binding affinity and telomerase function. In contrast to Y477 and W496, F476 is buried and is not accessible for interactions with the nucleic acid substrate. Instead, F476 is located at the base of the β hairpin and contributes significantly to the formation of the T pocket. Mutating the large hydrophobic side chain of this residue to the small alanine one would most likely lead to conformational rearrangements of this pocket and loss of RNA-binding affinity and telomerase activity.

A second set of alanine mutants, L327A, K329A, C331A, and P334A, which showed moderate loss of RNA-binding affinity and telomerase activity, has also been isolated (Bryan et al., 2000; Miller et al., 2000). Both L327 and C331 make direct contacts with F476 and the aliphatic part of the side chain of R492, both of which are located at the base of the T motif. Mutation to the smaller alanine residue could result in the rearrangement of the T pocket, potentially leading to loss of interactions with the nucleic acid substrate and loss of function. Likewise, P334 is located at the back of helix α12 and makes direct contacts with residues of this structural element (Figure 3B). Helix α12 contains the invariant W496 and the conserved K493, both of which form part of the T pocket. Mutating P334 into an alanine could lead to the displacement of helix α10 and reorganization of the T pocket, leading to loss of function. K329 is also located on helix α3 and, unlike L327A, C331A, and P334A, is solvent exposed and possibly makes direct contacts with the nucleic acid substrate (Figure 3B). Mutating it to an alanine would lead to loss of RNA interactions and loss of RNA-binding affinity and telomerase activity.

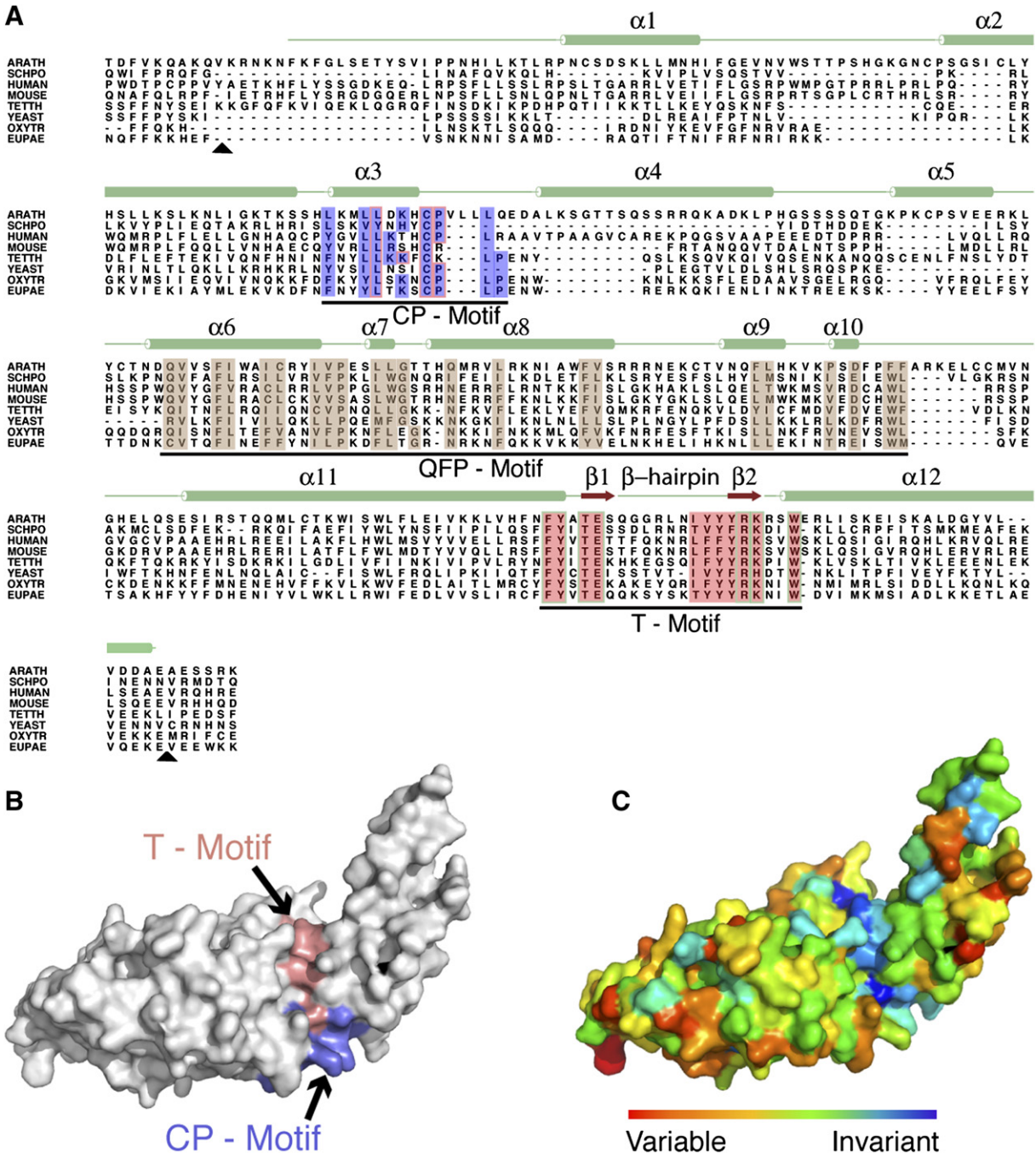

Figure 2. TRBD Sequence Alignment and Surface Conservation

(A) Sequence alignment and schematic of the secondary structure of TRBDs from a wide range of phylogenetic groups (ciliated protozoan, mammals, yeast, and plants) produced by ALSCRIPT (Barton, 1993). Conserved residues in key signature motifs are colored (CP, slate; QFP, brown; T, salmon), and mutated residues that affect RNA binding and telomerase function are also boxed. The black triangles define the boundaries of the TRBD construct that we used in our studies.

(B) Surface representation of TRBD showing solvent-exposed residues that form both the T (salmon) and CP motifs (slate).

(C) Conservation of surface-exposed residues of TRBD (red, variable; blue, invariable) produced by CONSURF (Glaser et al., 2003).

The TRBD Domain Facilitates the Formation of a Stable RNP Complex and Promotes Repeat-Addition Processivity

In vivo, telomerase exists as a stable ribonucleoprotein complex, and contacts between the protein (TERT) and the RNA components (TER) are mediated by the TEN, TRBD, and RT domains. Extensive biochemical and mutagenesis studies have shown that the TRBD is involved in extensive, specific interactions with stem I and the TBE of TER (Lai et al., 2001; O'Connor et al., 2005)

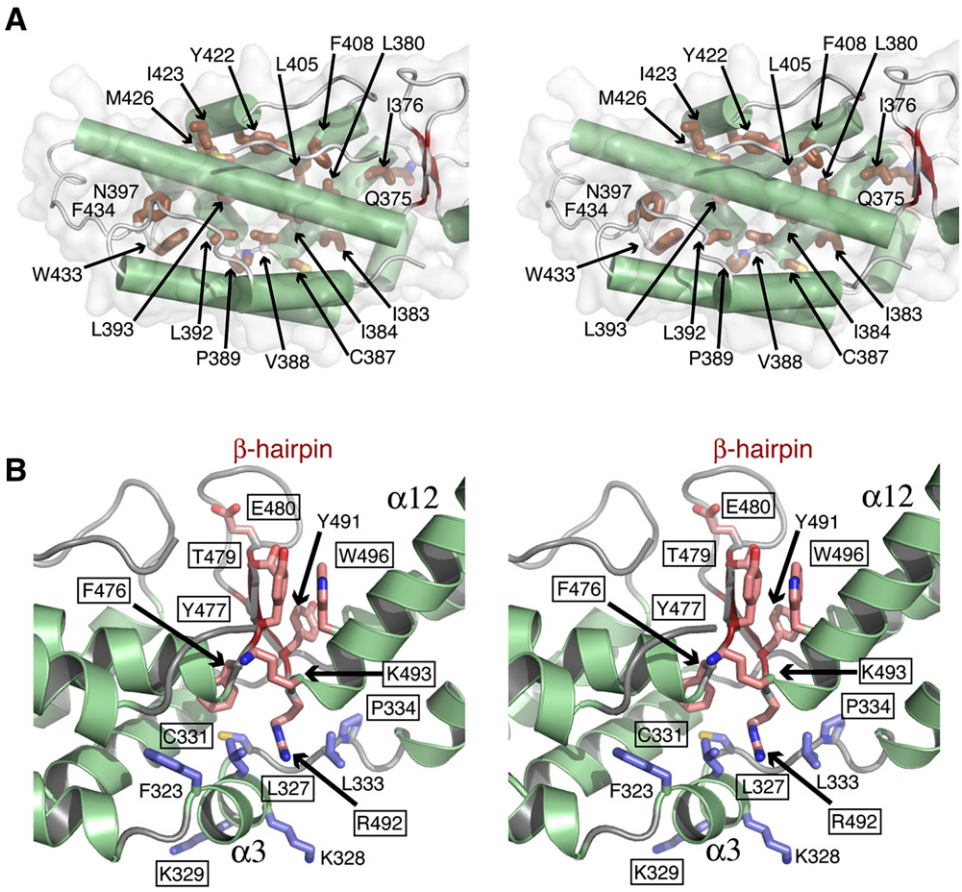

Figure 3. Conserved Residues and Existing Mutants of the CP, QFP, and T Motifs

(A) Stereo view of the larger half of the molecule showing the QFP residues (brown stick) buried in the core of the domain.

(B) Zoom in, stereo view of the CP (slate) and T motifs (salmon) showing conserved residues in stick representation. Mutated residues that affect RNA binding and telomerase activity are boxed.

(Figure 4D). Contacts between the TRBD and TER are thought to facilitate the proper assembly and stabilization of the RNP complex and promote repeat-addition processivity (Lai et al., 2003). In ciliates, in addition to the TRBD, a conserved motif (CP2) located N-terminally to the TRBD domain is thought to be required for TERT/TER assembly and template boundary definition (Lai et al., 2002; Miller et al., 2000). However, until now, it has been unclear as to how the telomerase TRBD carries out this process. The TRBD domain is divided into two asymmetric halves connected by several long loops that are shaped like a boomerang, an arrangement that has significant implications for RNA recognition and binding. The overall organization of the two lobes of the molecule results in the formation of two well-defined cavities (CP and T pockets; Figures 2B, 4A, and 4B) on the surface of the protein that consist of several solvent-exposed, invariant/conserved residues. The T pocket is a narrow, deep cavity located at the junction of the two halves of the molecule; part of the T pocket is hydrophobic in nature, whereas the part that is located the vicinity of the CP pocket is positively charged (Figure 4B). Interestingly, the hydrophobic side chains of Y477 and W496 are solvent exposed and are stacked against each other, forming a narrow "hydrophobic pincer" that, in this structure, could not accommodate a nucleotide base. It is, however, worth noting that helix α12, which contains W496, is somewhat flexible with respect to the β hairpin that contains Y477 (Figure 4C). The ability of helix α12 and therefore W496 to move could splay the two side chains apart, thus allowing for the space required for the accommodation of a nucleotide base between them. Another possibility is that the polar moiety of Y477 and W496 could act together as a nucleotide base that would allow for the formation of pseudo Watson Crick interactions with an incoming nucleotide base. Pseudo Watson Crick interactions have been previously observed for a number of protein nucleic acid complexes, including the Rho transcription termination factor (Bogden et al., 1999) and the signal-recognition particle (Wild et al., 2001). The width and the organization of the hydrophobic part of the T pocket suggest that it binds single-stranded RNA, most likely the TBE, possibly mediated by a network of stacking interactions.

In contrast to the T pocket, the CP pocket is a positively charged, shallow cavity located on the side of the molecule and forms an extension of the T pocket (Figure 4A).

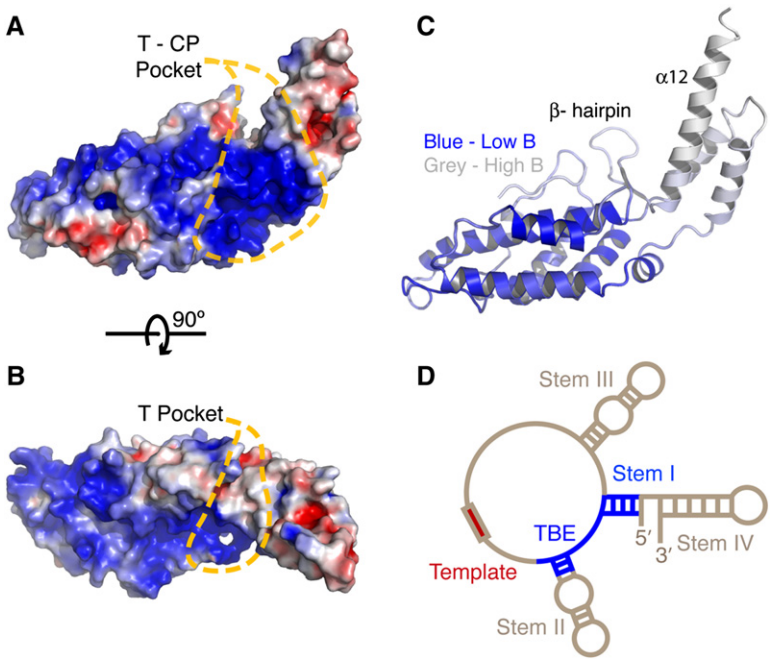

Figure 4. Surface Charge of TRBD and the TER Primary Structure

(A) Surface representation of TRBD showing positive charge in blue, negative charge in red, and hydrophobic residues in gray. The CP and T pockets and the location at which RNA is thought to bind are highlighted with an orange, dashed line.

(B) (A) rotated 90° for clarity.

(C) TRBD structure colored by B factor; blue represents regions with low mobility, and gray represents regions with high B factors.

(D) Schematic of the primary structure of the RNA component (TER) of telomerase from *Tetrahymena thermophila*. Stem I and the TBE are shown in blue, and the template is shown in red.

Together, the hydrophilic part of the T pocket and the CP pocket are lined with several lysines and arginines, the side chains of which are solvent exposed and could be involved in direct contacts with the backbone of double-stranded RNA. The width and the chemical nature of this pocket suggest that it binds double-stranded RNA, most likely stem I or stem II (Figure 4D). The nature and the extent of the protein/nucleic acid interactions mediated by the TRBD-binding pockets would provide the stability required for the proper assembly of a functional ribonucleoprotein enzyme and guide TERT to a TER-binding site (between stems I and II) that has significant implications for telomerase function.

Telomerase is unique in its ability to add multiple short oligonucleotide repeats at the end of linear chromosomes. The enzyme's ability to do so has been partly attributed to the interactions of the TRBD domain with the TBE and, in ciliates, both the TRBD and the CP2 motif (Lai et al., 2002, 2003; Miller et al., 2000). The TBE consists of stem II and the flanking single-stranded RNA regions and is located downstream of stem I and only a few nucleotides upstream of the RNA template (Figure 4D). The TRBD structure suggests that the T pocket, a narrow, hydrophobic cavity located on the surface of the protein that can only accommodate single-stranded RNA, may play an important role in this process. Assuming that the T pocket binds the single-stranded RNA that connects stem I and stem II, this interaction most likely forces stem II to act as a steric block, which, in turn, forces the TRBD domain to stay within the boundaries of stem I and stem II. The stem I- and -II-locked TRBD domains then may act as anchors that constrain the distance the RT domain can travel and prevent it from moving beyond the boundaries of the RNA template, thus promoting telomerase addition processivity. In ciliates, however, the TRBD domain alone is not sufficient for template boundary definition, and it requires the action of the CP2 motif (Lai et al., 2002; Miller et al., 2000). The precise mechanism by which the CP2 motif facilitates template boundary definition is unclear at this stage, and additional experimental evidence is required to help us better address this question. One, however, can speculate that CP2 binding to TER promotes template boundary definition either via contributing to the stabilization of the RNP complex or, like the TRBD, by acting as an anchor that prevents slippage of the active site of the RT domain beyond the RNA template.

In summary, the TRBD domain is highly conserved among phylogenetic groups and is essential for the function of telomerase. Extensive biochemical and mutagenesis studies have localized TRBD binding to stem I and the TEB, interactions that are thought to be important for the proper assembly and stabilization of the TERT/TER complex as well as the repeat-addition processivity of the holoenzyme. Here, we present the first, to our knowledge, atomic structure of the TRBD domain, which provides clues for TERT/TER binding and explains the wealth of biochemical data published to date. The RNA-binding site of TRBD is an extended groove on the surface of the protein that is partly hydrophilic and partly hydrophobic in nature and is formed by the previously identified T and CP motifs shown to be important for telomerase function. The size, organization, and chemical nature of this groove suggests that the TRBD domain interacts with both double- and single-stranded nucleic acids, possibly stems I or II, and the single-stranded RNA that connects them. Extended contacts in this region would confer stability, whereas the nature of the interactions will facilitate the enzyme's ability to add short oligonucleotide repeats at the ends of linear chromosomes. Moreover, the fact that the TRBD domain is an essential, highly conserved domain

Table 2. Model Refinement Statistics	
$TRBD_{(254–519)}$	
Refinement Statistics	
Resolution (Å)	20–1.71
R_{work}/R_{free} (%)	20.0/23.9
Rmsd bonds (Å)	0.008
Rmsd angles (°)	0.831
Number of atoms	
Protein	2145
Bromine	7
Water	213
Average B factor (Å^2)	
Protein	27.41
Bromine	42.63
Water	31.22
Ramachandran % (No Res.)	
Most favored	91.6
Allowed	8.4

among phylogenetic groups that contains a well-defined RNA-binding pocket suggests that it could serve as an ideal candidate for telomerase inhibitors.

EXPERIMENTAL PROCEDURES

Protein Expression and Purification

The *Tetrahymena thermophila* TERT residues 254–519 were identified by limited proteolysis and cloned into a modified version of the pET28b vector containing a cleavable hexahistidine tag at its N terminus. The protein was overexpressed in *E. coli* BL21 (pLysS) at 20°C for 4 hr. The cells were lysed by sonication in 50 mM Tris-HCl, 10% glycerol, 0.5 M KCl, 5 mM β-mercaproethanol, and 1 mM PMSF (pH 7.5) on ice. The protein was first purified over a Ni-NTA column, followed by TEV cleavage of the hexahistidine tag overnight at 4°C. The TRBD/TEV mix was diluted so that the concentration of imidazole was at 15 mM, and the protein mix was passed over a Ni-NTA column to remove the TEV, the cleaved tag, and any tagged protein. The Ni-NTA flowthrough was concentrated to 1 ml and diluted to a salt concentration of 0.15 M. The diluted TRBD sample was then passed over a POROS-HS column (Perseptive Biosystems). At this stage, the protein was more than 99% pure. The protein was finally passed over a sephedex-S200 sizing column preequilibrated with 50 mM Tris-HCl, 10% glycerol, 0.5 M KCl, and 2 mM DTT (pH 7.5) to remove any TRBD aggregates. The pure, monodisperse protein, as indicated by SDS-PAGE and dynamic light scattering, respectively (DynaPro, model 99-CP, Protein Solutions), was concentrated to 8 mg/ml by using an amicon 10K cutoff (Millipore), and the protein was stored at 4°C for subsequent studies. Stock protein was dialyzed in 5 mM Tris-HCl, 500 mM KCl, 1 mM TCEP (pH 7.5) prior to crystallization trials.

Protein Crystallization and Data Collection

Initial plate-like clusters of TRBD that diffracted poorly (~4 Å resolution) were grown at 4°C by using the sitting drop method by mixing one volume of dialyzed protein with one volume of reservoir solution containing 20% PEG 3350, 0.2 M $NaNO_3$. Single, well-diffracting crystals were grown in microbatch trays under paraffin oil by mixing one volume of dialyzed protein with an equivalent volume of 50 mM HEPES (pH 7.0), 44% PEG 400, 0.4 M $NaNO_3$, 0.4 M NaBr, and 1 mM TCEP at 4°C. Crystals were harvested into cryoprotectant solution that contained 25 mM HEPES (pH 7.0), 25% PEG 400, 0.2 M $NaNO_3$, 0.2 M NaBr, and 1 mM TCEP and were flash frozen in liquid nitrogen. Data were collected at the NSLS, beamline X6A, and were processed with HKL-2000 (Minor) (Table 1). The crystals belong to the monoclinic space group $P2_1$; there is one monomer in the asymmetric unit.

Structure Determination and Refinement

Initial phases were obtained from a two-wavelength MAD holmium (Ho) derivative that was prepared by cocrystallizing the protein with 5 mM $HoCl_3$. Heavy-atom sites were located by using SOLVE (Terwilliger, 2003), and the sites were refined and new phases calculated with MLPHARE (CCP4, 1994) as implemented in ELVES (Holton and Alber, 2004) (Table 1). Initial maps showed well-defined density only for the larger half of the molecule. The density for the smaller half of the molecule was weak, mostly due to its intrinsic mobility with respect to the larger half of the molecule. The problem associated with building the model into the density was exacerbated by the lack of information regarding the location of specific side chains such as selenomethionines. Key factors in building a complete model were successive rounds of PRIME and SWITCH in RESOLVE (Terwilliger, 2002), followed by manual building in COOT (Emsley and Cowtan, 2004). The model was refined by using both CNS-SOLVE (Brunger et al., 1998) and REFMAC5 (Murshudov et al., 1997). The last cycles of refinement were carried out with TLS restraints as implemented in REFMAC5 (Table 2). Figures were prepared in PYMOL (DeLano, 2002), and electrostatic surfaces were prepared in APBS (Baker et al., 2001).

ACKNOWLEDGMENTS

We thank Ronen Marmorstein and Harold C. Riethman for critical reading of this manuscript, Steve Gamblin and James Berger for useful discussions, John J. Rux for assistance with the installation of X-ray data-processing software, and Kathleen Collins for the gift of the full-length TERT. This project was funded, in part, by the Pennsylvania Department of Health.

Received: July 28, 2007
Revised: September 3, 2007
Accepted: September 6, 2007
Published: November 13, 2007

REFERENCES

Aigner, S., and Cech, T.R. (2004). The Euplotes telomerase subunit p43 stimulates enzymatic activity and processivity in vitro. RNA *10*, 1108–1118.

Aigner, S., Postberg, J., Lipps, H.J., and Cech, T.R. (2003). The Euplotes La motif protein p43 has properties of a telomerase-specific subunit. Biochemistry *42*, 5736–5747.

Aisner, D.L., Wright, W.E., and Shay, J.W. (2002). Telomerase regulation: not just flipping the switch. Curr. Opin. Genet. Dev. *12*, 80–85.

Autexier, C., and Lue, N.F. (2006). The structure and function of telomerase reverse transcriptase. Annu. Rev. Biochem. *75*, 493–517.

Baker, N.A., Sept, D., Joseph, S., Holst, M.J., and McCammon, J.A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. Proc. Natl. Acad. Sci. USA *98*, 10037–10041.

Barton, G.J. (1993). ALSCRIPT: a tool to format multiple sequence alignments. Protein Eng. *6*, 37–40.

Blackburn, E.H. (1984). The molecular structure of centromeres and telomeres. Annu. Rev. Biochem. *53*, 163–194.

Blackburn, E.H. (2000). The end of the (DNA) line. Nat. Struct. Biol. *7*, 847–850.

Blackburn, E.H., and Gall, J.G. (1978). A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in Tetrahymena. J. Mol. Biol. *120*, 33–53.

Bogden, C.E., Fass, D., Bergman, N., Nichols, M.D., and Berger, J.M. (1999). The structural basis for terminator recognition by the Rho transcription termination factor. Mol. Cell *3*, 487–493.

Bosoy, D., Peng, Y., Mian, I.S., and Lue, N.F. (2003). Conserved N-terminal motifs of telomerase reverse transcriptase required for ribonucleoprotein assembly in vivo. J. Biol. Chem. *278*, 3882–3890.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. *54*, 905–921.

Bryan, T.M., Sperger, J.M., Chapman, K.B., and Cech, T.R. (1998). Telomerase reverse transcriptase genes identified in *Tetrahymena thermophila* and *Oxytricha trifallax*. Proc. Natl. Acad. Sci. USA *95*, 8479–8484.

Bryan, T.M., Goodrich, K.J., and Cech, T.R. (2000). Telomerase RNA bound by protein motifs specific to telomerase reverse transcriptase. Mol. Cell *6*, 493–499.

Cavalier-Smith, T. (1974). Palindromic base sequences and replication of eukaryote chromosome ends. Nature *250*, 467–470.

CCP4 (Collaborative Computational Project, Number 4) (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D Biol. Crystallogr. *50*, 760–763.

Cech, T.R., and Lingner, J. (1997). Telomerase and the chromosome end replication problem. Ciba Found. Symp. *211*, 20–28.

Chandra, A., Hughes, T.R., Nugent, C.I., and Lundblad, V. (2001). Cdc13 both positively and negatively regulates telomere replication. Genes Dev. *15*, 404–414.

Chatziantoniou, V.D. (2001). Telomerase: biological function and potential role in cancer management. Pathol. Oncol. Res. *7*, 161–170.

Chen, J.L., and Greider, C.W. (2003). Template boundary definition in mammalian telomerase. Genes Dev. *17*, 2747–2752.

Chen, J.L., and Greider, C.W. (2004). Telomerase RNA structure and function: implications for dyskeratosis congenita. Trends Biochem. Sci. *29*, 183–192.

Chen, J.L., Blasco, M.A., and Greider, C.W. (2000). Secondary structure of vertebrate telomerase RNA. Cell *100*, 503–514.

Collins, K. (1996). Structure and function of telomerase. Curr. Opin. Cell Biol. *8*, 374–380.

Collins, K., and Gandhi, L. (1998). The reverse transcriptase component of the Tetrahymena telomerase ribonucleoprotein complex. Proc. Natl. Acad. Sci. USA *95*, 8485–8490.

Cong, Y.S., Wright, W.E., and Shay, J.W. (2002). Human telomerase and its regulation. Microbiol. Mol. Biol. Rev. *66*, 407–425.

Cunningham, D.D., and Collins, K. (2005). Biological and biochemical functions of RNA in the tetrahymena telomerase holoenzyme. Mol. Cell. Biol. *25*, 4442–4454.

DeLano, W.L. (2002). The PyMOL Molecular Graphics System (www.pymol.org).

Dong, C.K., Masutomi, K., and Hahn, W.C. (2005). Telomerase: regulation, function and transformation. Crit. Rev. Oncol. Hematol. *54*, 85–93.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. *60*, 2126–2132.

Evans, S.K., and Lundblad, V. (1999). Est1 and Cdc13 as comediators of telomerase access. Science *286*, 117–120.

Evans, S.K., and Lundblad, V. (2002). The Est1 subunit of *Saccharomyces cerevisiae* telomerase makes multiple contributions to telomere length maintenance. Genetics *162*, 1101–1115.

Friedman, K.L., and Cech, T.R. (1999). Essential functions of amino-terminal domains in the yeast telomerase catalytic subunit revealed by selection for viable mutants. Genes Dev. *13*, 2863–2874.

Friedman, K.L., Heit, J.J., Long, D.M., and Cech, T.R. (2003). N-terminal domain of yeast telomerase reverse transcriptase: recruitment of Est3p to the telomerase complex. Mol. Biol. Cell *14*, 1–13.

Glaser, F., Pupko, T., Paz, I., Bell, R.E., Bechor-Shental, D., Martz, E., and Ben-Tal, N. (2003). ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. Bioinformatics *19*, 163–164.

Greider, C.W., and Blackburn, E.H. (1989). A telomeric sequence in the RNA of Tetrahymena telomerase required for telomere repeat synthesis. Nature *337*, 331–337.

Hammond, P.W., Lively, T.N., and Cech, T.R. (1997). The anchor site of telomerase from *Euplotes aediculatus* revealed by photo-cross-linking to single- and double-stranded DNA primers. Mol. Cell. Biol. *17*, 296–308.

Holm, L., and Sander, C. (1996). Mapping the protein universe. Science *273*, 595–603.

Holton, J., and Alber, T. (2004). Automated protein crystal structure determination using ELVES. Proc. Natl. Acad. Sci. USA *101*, 1537–1542.

Hughes, T.R., Morris, D.K., Salinger, A., Walcott, N., Nugent, C.I., and Lundblad, V. (1997). The role of the EST genes in yeast telomere replication. Ciba Found Symp. *211*, 41–47, discussion 47–52, 71–45.

Jacobs, S.A., Podell, E.R., Wuttke, D.S., and Cech, T.R. (2005). Soluble domains of telomerase reverse transcriptase identified by high-throughput screening. Protein Sci. *14*, 2051–2058.

Jacobs, S.A., Podell, E.R., and Cech, T.R. (2006). Crystal structure of the essential N-terminal domain of telomerase reverse transcriptase. Nat. Struct. Mol. Biol. *13*, 218–225.

Lai, C.K., Mitchell, J.R., and Collins, K. (2001). RNA binding domain of telomerase reverse transcriptase. Mol. Cell. Biol. *21*, 990–1000.

Lai, C.K., Miller, M.C., and Collins, K. (2002). Template boundary definition in Tetrahymena telomerase. Genes Dev. *16*, 415–420.

Lai, C.K., Miller, M.C., and Collins, K. (2003). Roles for RNA in telomerase nucleotide and repeat addition processivity. Mol. Cell *11*, 1673–1683.

Lamond, A.I. (1989). Tetrahymena telomerase contains an internal RNA template. Trends Biochem. Sci. *14*, 202–204.

Lee, S.R., Wong, J.M., and Collins, K. (2003). Human telomerase reverse transcriptase motifs required for elongation of a telomeric substrate. J. Biol. Chem. *278*, 52531–52536.

Licht, J.D., and Collins, K. (1999). Telomerase RNA function in recombinant Tetrahymena telomerase. Genes Dev. *13*, 1116–1125.

Lingner, J., Cooper, J.P., and Cech, T.R. (1995). Telomerase and DNA end replication: no longer a lagging strand problem? Science *269*, 1533–1534.

Loayza, D., and de Lange, T. (2004). Telomerase regulation at the telomere: a binary switch. Cell *117*, 279–280.

Lundblad, V. (1997). The end replication problem: more than one solution. Nat. Med. *3*, 1198–1199.

Lundblad, V. (2003). Telomere replication: an Est fest. Curr. Biol. *13*, R439–R441.

Lundblad, V., and Blackburn, E.H. (1990). RNA-dependent polymerase motifs in EST1: tentative identification of a protein component of an essential yeast telomerase. Cell *60*, 529–530.

Lustig, A.J. (2001). Cdc13 subcomplexes regulate multiple telomere functions. Nat. Struct. Biol. *8*, 297–299.

Ly, H., Blackburn, E.H., and Parslow, T.G. (2003). Comprehensive structure-function analysis of the core domain of human telomerase RNA. Mol. Cell. Biol. *23*, 6849–6856.

Miller, M.C., and Collins, K. (2002). Telomerase recognizes its template by using an adjacent RNA motif. Proc. Natl. Acad. Sci. USA *99*, 6585–6590.

Miller, M.C., Liu, J.K., and Collins, K. (2000). Template definition by Tetrahymena telomerase reverse transcriptase. EMBO J. *19*, 4412–4422.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D Biol. Crystallogr. *53*, 240–255.

O'Connor, C.M., and Collins, K. (2006). A novel RNA binding domain in tetrahymena telomerase p65 initiates hierarchical assembly of telomerase holoenzyme. Mol. Cell. Biol. *26*, 2029–2036.

O'Connor, C.M., Lai, C.K., and Collins, K. (2005). Two purified domains of telomerase reverse transcriptase reconstitute sequence-specific interactions with RNA. J. Biol. Chem. *280*, 17533–17539.

Ohki, R., Tsurimoto, T., and Ishikawa, F. (2001). In vitro reconstitution of the end replication problem. Mol. Cell. Biol. *21*, 5753–5766.

Peng, Y., Mian, I.S., and Lue, N.F. (2001). Analysis of telomerase processivity: mechanistic similarity to HIV-1 reverse transcriptase and role in telomere maintenance. Mol. Cell *7*, 1201–1211.

Pennock, E., Buckley, K., and Lundblad, V. (2001). Cdc13 delivers separate complexes to the telomere for end protection and replication. Cell *104*, 387–396.

Prathapam, R., Witkin, K.L., O'Connor, C.M., and Collins, K. (2005). A telomerase holoenzyme protein enhances telomerase RNA assembly with telomerase reverse transcriptase. Nat. Struct. Mol. Biol. *12*, 252–257.

Reichenbach, P., Hoss, M., Azzalin, C.M., Nabholz, M., Bucher, P., and Lingner, J. (2003). A human homolog of yeast Est1 associates with telomerase and uncaps chromosome ends when overexpressed. Curr. Biol. *13*, 568–574.

Shippen-Lentz, D., and Blackburn, E.H. (1990). Functional evidence for an RNA template in telomerase. Science *247*, 546–552.

Smogorzewska, A., and de Lange, T. (2004). Regulation of telomerase by telomeric proteins. Annu. Rev. Biochem. *73*, 177–208.

Smogorzewska, A., van Steensel, B., Bianchi, A., Oelmann, S., Schaefer, M.R., Schnapp, G., and de Lange, T. (2000). Control of human telomere length by TRF1 and TRF2. Mol. Cell. Biol. *20*, 1659–1668.

Snow, B.E., Erdmann, N., Cruickshank, J., Goldman, H., Gill, R.M., Robinson, M.O., and Harrington, L. (2003). Functional conservation of the telomerase protein Est1p in humans. Curr. Biol. *13*, 698–704.

Terwilliger, T.C. (2002). Automated structure solution, density modification and model building. Acta Crystallogr. D Biol. Crystallogr. *58*, 1937–1940.

Terwilliger, T.C. (2003). SOLVE and RESOLVE: automated structure solution and density modification. Methods Enzymol. *374*, 22–37.

Theimer, C.A., and Feigon, J. (2006). Structure and function of telomerase RNA. Curr. Opin. Struct. Biol. *16*, 307–318.

Wild, K., Sinning, I., and Cusack, S. (2001). Crystal structure of an early protein-RNA assembly complex of the signal recognition particle. Science *294*, 598–601.

Witkin, K.L., and Collins, K. (2004). Holoenzyme proteins required for the physiological assembly and activity of telomerase. Genes Dev. *18*, 1107–1118.

Witkin, K.L., Prathapam, R., and Collins, K. (2007). Positive and negative regulation of tetrahymena telomerase holoenzyme. Mol. Cell. Biol. *27*, 2074–2083.

Wyatt, H.D., Lobb, D.A., and Beattie, T.L. (2007). Characterization of physical and functional anchor site interactions in human telomerase. Mol. Cell. Biol. *27*, 3226–3240.

Xia, J., Peng, Y., Mian, I.S., and Lue, N.F. (2000). Identification of functionally important domains in the N-terminal region of telomerase reverse transcriptase. Mol. Cell. Biol. *20*, 5196–5207.

Zappulla, D.C., and Cech, T.R. (2004). Yeast telomerase RNA: a flexible scaffold for protein subunits. Proc. Natl. Acad. Sci. USA *101*, 10024–10029.

Accession Numbers

Coordinates for the TRBD structure have been deposited in the RCSB PDB database with accession code 2R4G.

