Supporting Information

Steczkiewicz et al. 10.1073/pnas.1015399108

SI Methods

We have utilized the human telomerase protein sequence (GenBank accession no. NM_198253.2). In order to study sequence conservation within the telomerase protein family, PSI-BLAST (1) searches were carried out against the NCBI nonredundant protein sequence database using this human telomerase sequence as query (three iterations, E-value threshold 0.005). Multiple sequence alignments of collected sequences were prepared with PCMA, a progressive alignment method (2). Additionally, to obtain secondary structure conservation patterns useful for fold assignment validation and for manual curation of the sequence-to-structure alignments, for every aligned sequence, the secondary structure was predicted with PSI-PRED (3).

Templates for comparative modeling of human telomerase domains were identified using the Gene Relational Database (GRDB) system, which stores precalculated Meta-BASIC mappings (4) between Pfam families, conserved domains, and PDB structures. Meta-BASIC is a distant homology detection method that exploits comparisons of sequence profiles enriched with predicted secondary structures (meta-profiles) (4). Meta-BASIC predictions were further validated with the consensus fold recognition server, 3D-Jury (5), followed by manual inspection of the hits obtained. All searches were carried out by using the complete telomerase sequence and the sequences of its three separate protein domains: TEN, TRBD, and RT. Sequence-to-structure alignments between human telomerase (and its closest homologs) and selected proteins of known structure were prepared using the consensus alignment approach and 3D assessment (6).

Three-dimensional models of human telomerase protein domains were generated with Modeller (7) based on manually curated, high confidence sequence-to-structure alignments. These models were built separately for (*i*) the TEN domain, using *T. thermophila* TEN (PDB ID code 2B2A) (8); (*ii*) the RT:TRBD subcomplex, using the corresponding structure of *T. thermophila* TEN T. castaneum RT–TRBD (PDB ID code 3KYL) (9) and the superimposed TRBD domain from *T. thermophila* (PDB ID code 2R4G) (10) as templates, because the *T. castaneum* TRBD domain lacks two critical α-helices that are present both in human and *T. thermophila* TRBD. The resulting 3D models of TEN and the RT:TRBD subcomplex were then assembled manually after careful consideration of the CABS (11) results for protein domain docking (described below) and published experimental data regarding specific residues proposed to mediate interdomain interactions (Table S1). Additionally, conservation of surface residues was investigated with ConSurf (12), in order to detect patterns of highly conserved amino acids that might suggest plausible interactions and the location of interfaces between domains (Fig. S6).

Automatic CABS assembly of the modeled human telomerase domains was performed using a three-stage docking procedure. First, rigid docking via an exhaustive global search in a six-dimensional space of "ligand" rotations and translations against the frozen structure of the "receptor" was carried out using FTDOCK (13). Rotational space was scanned in 12-degree increments. For translations, sampling was performed on a cubic lattice with 0.875-Å spacings. For each rotation, the three top-scoring translations were saved for subsequent analysis. The resulting 10,000 FTDOCK top-scoring structures were rescored with the CABS force field and grouped using hierarchical clustering. From each cluster, a representative with the lowest energy was selected. The number of models was thereby reduced from 10,000 to 30.

To account for protein flexibility upon complex formation, each resultant structural model was subjected to a short stochastic dynamics simulation with the CABS algorithm. From each simulation 1,000 models located in the vicinity of the initial structure were collected. Hierarchical clustering was again used to select the most populated states. Structures from the centers of the clusters were extracted to represent variants of the final telomerase tertiary structure. These representatives were used as starting models for further manual adjustments based on the consistency of the model with available experimental data (Table S1) and according to surface conservation in individual domains.

Positions of the intrinsic RNA template and single-stranded telomeric DNA substrate in the human telomerase model were copied from the *T. castaneum* telomerase structure (PDB ID code 3KYL) after superposition of their RT and TRBD domains. The sequence of the RNA:DNA hairpin containing RNA template was modified: 5′-CUGACCUGA-3′ and the complementary telomeric DNA: 5′-TCAGGTCAG-3′ were replaced with 5′-UAACCCUAA-3′ and 5′-TTAGGGTTA-3′ sequences suitable for human telomerase. The hairpin loop (5′-CTTCGG-3′) was removed and, to allow for interaction with the TEN domain, the double helix was extended by 7 base pairs (resulting in DNA: 5′-GTTAGGGTTAGGGTTA-3′, RNA: 5′-UAACCCUAACUGAGAA-3′). The telomeric DNA is less complementary to the RNA at the 5′ end; however, a nonclassical (e.g., wobble) pairing might still be sufficient for double helix formation. Finally, minor adjustments of the TEN domain were made to improve its fit between the nucleic acids and the protein domains in the complex.

To relieve steric clashes and improve internal packing, the 3D partial model of the human telomerase complex, comprising all three protein domains and the RNA:DNA heteroduplex, was energy-minimized with Tripos SYBYL using an AMBER force field (14), followed by a short molecular dynamics run (simulation time of 0.2 ns, a NTV ensemble, 100 fs coupling, Boltzmann initial velocities, Amber7 FF99 force field, Gasteiger–Hückel charges and dielectric constant = 1).

The final model was used to investigate the motions of the structure using simple coarse-grained elastic network models in conjunction with normal mode analysis. This approach has been widely used to investigate important functional motions of biomolecular structures (15). It is based on a highly cohesive model of structure and investigates particularly the larger motions that are available within the constraints of the geometry of the structure. Residues close to each another in the structure are connected with identical springs and the vibrational motions of the set of springs are analyzed with a normal mode decomposition (16). The approach has proven useful for extracting the functional motions of large domains in many structures (15). Notably the computed motions are quite insensitive to details of the structure, which means the computed motions reported here are robust and unlikely to be changed by any minor errors in the model. In the case of the human telomerase complex, our modeled structure, together with these dynamics simulations, allowed us to investigate the functional role of the TEN domain in telomere elongation.

1. Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402.

2. Pei J, Sadreyev R, Grishin NV (2003) PCMA: Fast and accurate multiple sequence alignment based on profile consistency. *Bioinformatics* 19:427–428.

3. Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* 292:195–202.
4. Ginalski K, von Grotthuss M, Grishin NV, Rychlewski L (2004) Detecting distant homology with Meta-BASIC. *Nucleic Acids Res* 32:W576–581.
5. Ginalski K, Elofsson A, Fischer D, Rychlewski L (2003) 3D-Jury: A simple approach to improve protein structure predictions. *Bioinformatics* 19:1015–1018.
6. Ginalski K, Rychlewski L (2003) Protein structure prediction of CASP5 comparative modeling and fold recognition targets using consensus alignment approach and 3D assessment. *Proteins* 53(Suppl 6):410–417.
7. Eswar N, et al. (2006) Comparative protein structure modeling using Modeller. *Curr Protoc Bioinformatics* (John Wiley & Sons, Inc., New York), Chapter 5:Unit 5.6.
8. Jacobs SA, Podell ER, Cech TR (2006) Crystal structure of the essential N-terminal domain of telomerase reverse transcriptase. *Nat Struct Mol Biol* 13:218–225.
9. Mitchell M, Gillis A, Futahashi M, Fujiwara H, Skordalakes E (2010) Structural basis for telomerase catalytic subunit TERT binding to RNA template and telomeric DNA. *Nat Struct Mol Biol* 17:513–518.
10. Rouda S, Skordalakes E (2007) Structure of the RNA-binding domain of telomerase: Implications for RNA recognition and binding. *Structure* 15:1403–1412.
11. Kolinski A (2004) Protein modeling and structure prediction with a reduced representation. *Acta Biochim Pol* 51:349–371.
12. Landau M, et al. (2005) ConSurf 2005: The projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res* 33:W299–302.
13. Jackson RM, Gabb HA, Sternberg MJ (1998) Rapid refinement of protein interfaces incorporating solvation: Application to the docking problem. *J Mol Biol* 276:265–285.
14. Cornell WD, et al. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc* 117:5179–5197.
15. Bahar I, Lezon TR, Yang LW, Eyal E (2010) Global dynamics of proteins: Bridging between structure and function. *Annu Rev Biophys* 39:23–42.
16. Atilgan AR, et al. (2001) Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys* J 80:505–515.

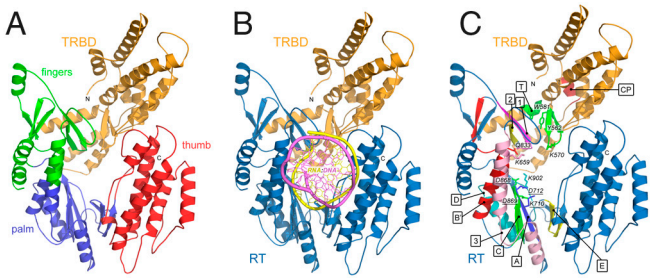


Fig. S1. The central ring-shaped part of the human telomerase model, formed by TRBD and RT domains, accommodating the RNA:DNA heteroduplex and providing catalytic residues. (*A*) Ring-shaped structure formed by RT (palm, fingers, thumb) and TRBD domains. (*B*) The channel formed by RT and TRBD domains is responsible for binding the RNA:DNA heteroduplex. (*C*) Locations of sequence motifs (in boxes) and residues shown experimentally to be important for nucleic acid binding and catalysis of template-based telomere elongation reaction.

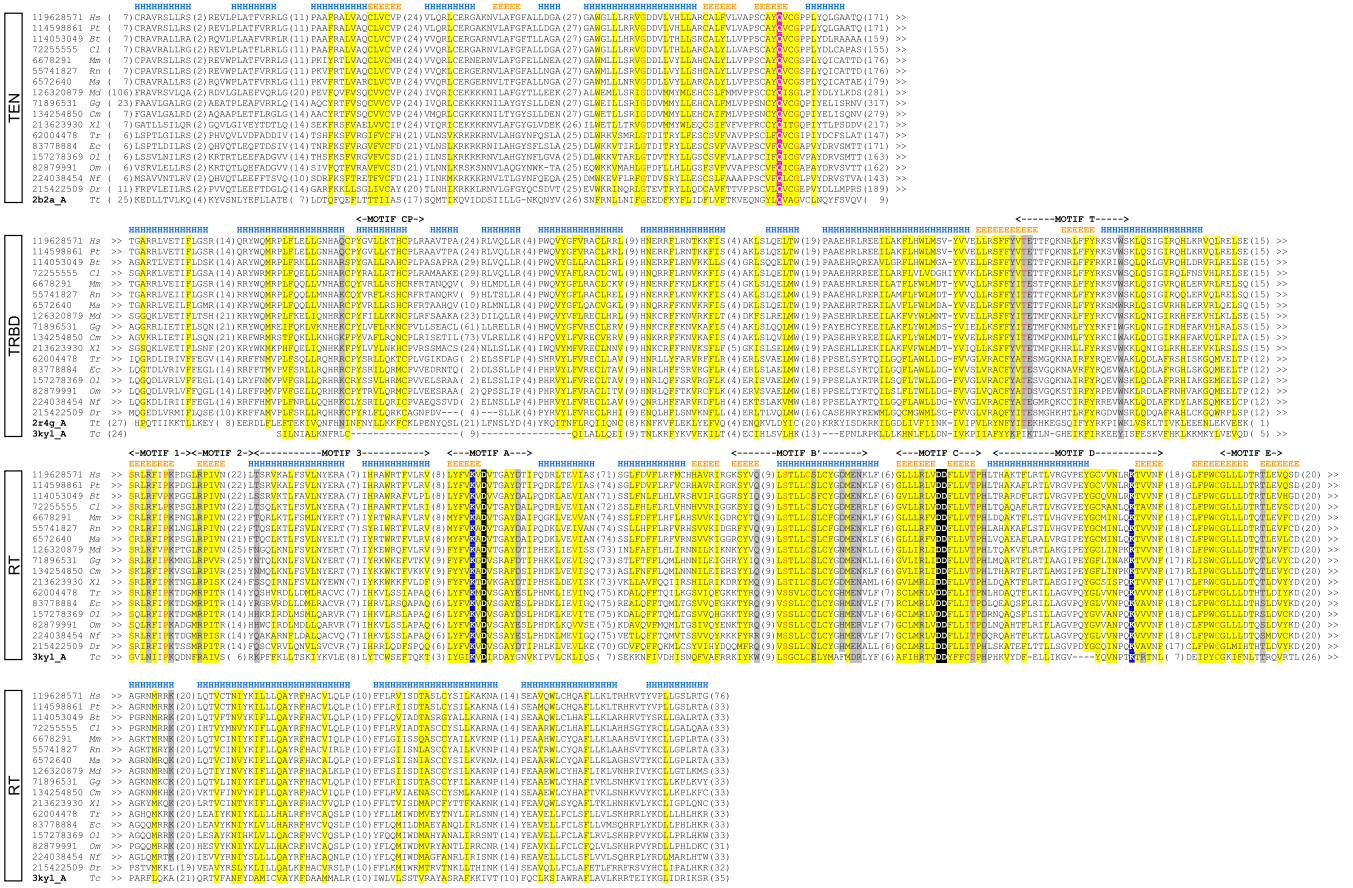


Fig. S2. Multiple sequence alignment of human telomerase protein domains with close homologs and selected proteins of known structure. Sequences are labeled with NCBI gene identification numbers or PDB ID codes. Abbreviations of the species names are: Hs, *Homo sapiens*; Pt, *Pan troglodytes*; Bt, *Bos taurus*; Cl, *Canis lupus*; Mm, *Mus musculus*; Rn, *Rattus norvegicus*; Ma, *Mesocricetus auratus*; Md, *Monodelphis domestica*; Gg, *Gallus gallus*; Cm, *Cairina moschata*; Xl, *Xenopus laevis*; Tr, *Takifugu rubripres*; Ec, *Epinephelus coioides*; Ol, *Oryzias latipes*; Om, *Oryzias melastigma*; Nf, *Nothobranchius furzeri*; Dr, *Danio rerio*; Tt, *Tetrahymena thermophila*; Tc, *Tribolium castaneum*. The number of residues not shown in the alignment is designated in parentheses. Residue conservation is denoted with the following scheme: uncharged, highlighted in yellow; charged or polar, highlighted in gray; small, letters in red. Critical acidic and basic active site residues are highlighted in black and blue, respectively. The invariant glutamine Q169 essential for telomerase processivity is highlighted in pink. Locations of predicted secondary structure elements (H, α-helix; E, β-strand) are labeled above corresponding residue columns. Positions of critical sequence motifs (CP, T, 1, 2, 3, A, B′, C, D, E) are marked above respective blocks of the alignment.

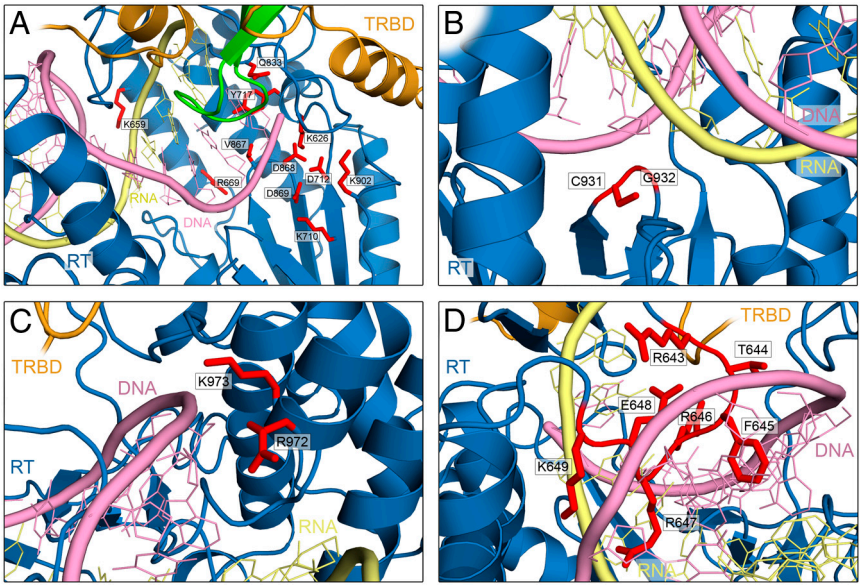


Fig. S3. Interaction between RT:TRBD domains and RNA:DNA heteroduplex. The TEN domain is not shown for clarity. (*A*) Telomerase active site region. Residues important for catalysis and nucleic acid binding are highlighted in red. TRBD motif T shown to be directly involved in RNA recognition is colored green. (*B*) Cys931 and G932 from RT palm constitute the "primer grip" crucial for proper maintenance of telomeric DNA within active site. (*C*) RT thumb residues R972 and K973 located on an α-helix that packs into the RNA:DNA minor groove, interact with the DNA backbone and may contribute to repetitive addition processivity. (*D*) The loop preceding the motif 3 α-helix in RT fingers (residues 643–649, colored in red) binds to double helix major groove.

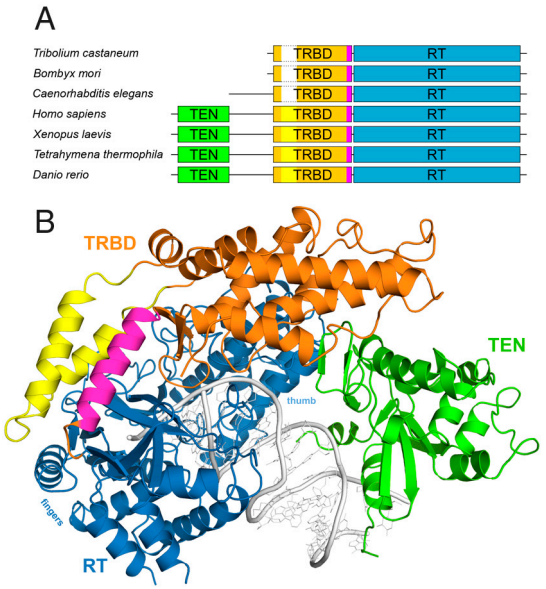


Fig. S4. Domain architecture of several telomerase proteins from different species. (*A*) Telomerases having TEN domains also possess two additional α-helices (colored in yellow) in the N-terminal region of the TRBD domain (residues 333–371 in *T. thermophila*, PDB ID code 2R4G). (*B*) The two additonal α-helices form, together with the last TRBD α-helix (violet), a three-helix bundle that packs tightly against the RT fingers. Consequently, telomerases lacking the TEN domain may have a more elastic interface between TRBD and RT, possibly allowing for some domain movements that might contribute to the enzyme processivity.

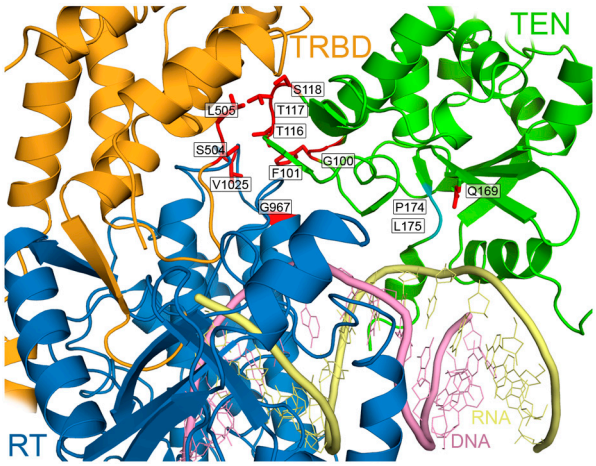


Fig. S5. Interaction of TEN with remaining components of telomerase complex. Q169 (side chain in red), whose mutation was shown to compromise telomerase processivity, is important for stabilization of the TEN local structure including the N- and C-terminal α-helices that interact with the major groove of RNA:DNA heteroduplex. The remaining red residues belong to the proposed interface between TEN and RT:TRBD.

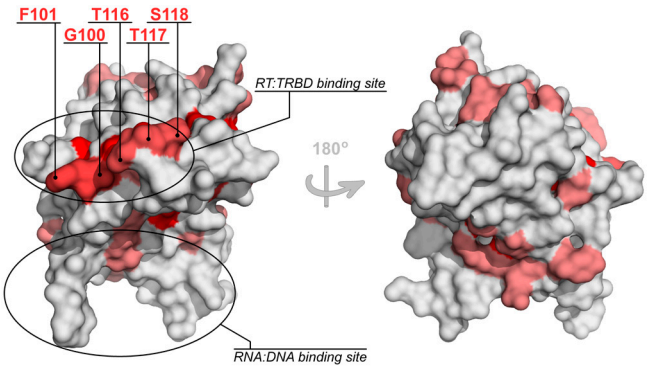


Fig. S6. Surface conservation of the TEN domain. Surface conservation generated using ConSurf is denoted in a color gradient from white (variable residues) to red (more conserved residues). Proposed RT:TRBD and RNA:DNA heteroduplex binding sites are marked with ellipses. T116, T117, and S118 were previously shown experimentally to contribute to telomerase activity in vivo but are not essential for activity in vitro.

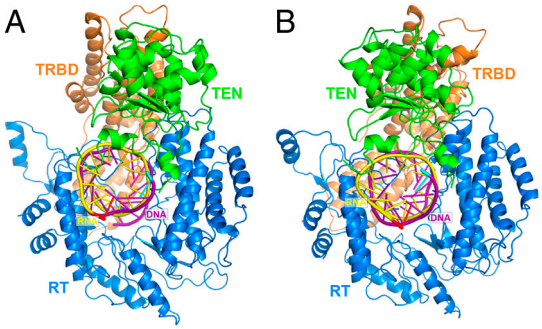


Fig. S7. Structural model for the processivity of human telomerase, shown here for the entire structure (details are shown in Fig. 2 A and B). The ANM model is constructed from one point per amino acid and two points per nucleotide with an interaction cutoff of 13 Å. The rotation of the DNA:RNA heteroduplex is evident. Procession of the heteroduplex is a critical aspect of telomerase function. The telomerase structure from Fig. 1 (same colors) and the effects of following the global mode in the (*A*) negative (−1) and (*B*) positive (+1) directions are shown. Termini closest to the viewer are highlighted: the 3′ end of RNA in red and the 5′ end of DNA in cyan. Dynamics of this slowest global motion can be seen in two views in the Movies S1 and S2.

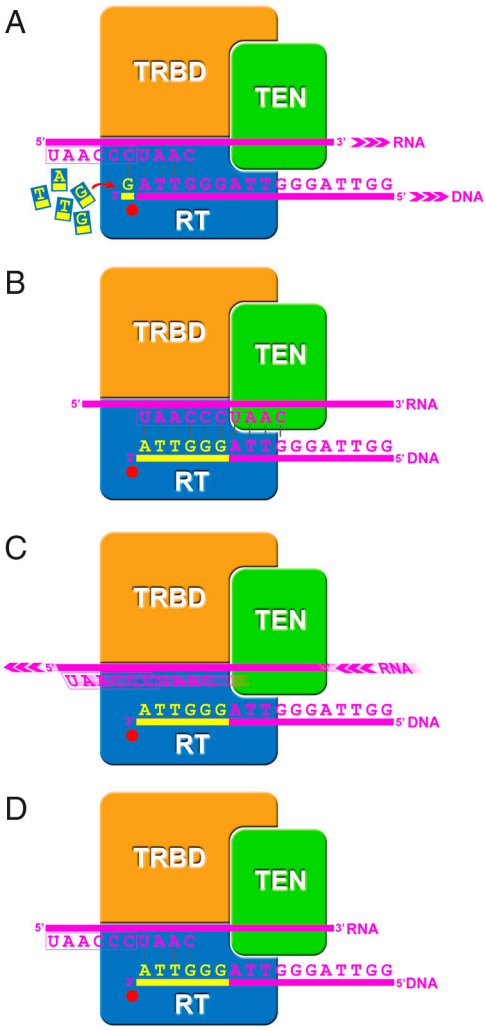
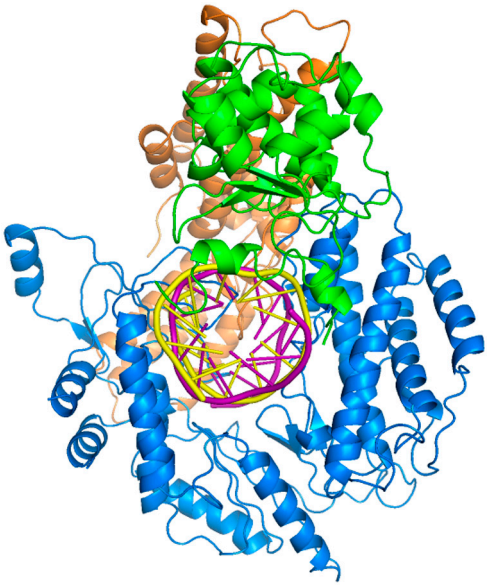


Fig. S8. Schematic representation of the mechanism of telomerase action. The RT active site is represented by a red dot. (*A* and *B*) During the RNA template-based reverse transcription, new deoxynucleotides are added until the telomeric repeat template (5′-UAACCC-3′) in TER is fully complemented by a newly synthesized DNA repeat (5′-GGGTTA-3′). (*C*) The template-encoding region of TER is realigned, while the 3′ end of the telomeric DNA substrate is stabilized by the telomerase protein domains, especially TEN. Note that this schematic representation of the RNA:DNA heteroduplex does not reflect the actual shape of the double-helix, and translocation of the RT active site relative to the template repeat in TER requires both a linear shift and a rotation of the double helix. (*D*) Once the newly synthesized telomeric DNA repeat is properly aligned with the complementary 5′-UAAC-3′ sequence immediately adjacent to the repeat template in TER, the reaction cycle is set to begin again.

Movie S1. Movie showing the motion of the heteroduplex rotating and translating through the partial telomerase structure, assisted by the motion of the TEN domain, and viewed along the double helix axis. The motion is the first normal mode, and it is shown oscillating between the positive and negative directions. Colors indicate the various domains with TEN in green, TRBD in orange, and RT in blue, and with the RNA in yellow and the DNA in purple. For an orthogonal view see Movie S2.

Movie S1 (MOV)



Movie S2. Movie showing the motion of the heteroduplex rotating and translating through the partial telomerase structure, assisted by the motion of the TEN domain, and viewed facing into the double helix axis in a view orthogonal to that in Movie S1. The motion is the first normal mode, and it is shown oscillating between the positive and negative directions. Colors indicate the various domains with TEN in green, TRBD in orange, and RT in blue, and with the RNA in yellow and the DNA in purple. For an orthogonal view see Movie S1.

Movie S2 (MOV)

Table S1. Summary of experimental evidence for telomerase interdomain interactions and telomerase-RNA:DNA heteroduplex binding interactions used for human telomerase complex assembly

Interaction	Residue	Domain	Evidence
RT–RNA:DNA			Crystallographic structure of *T. castaneum* RT and TRBD bound to RNA:DNA hairpin (PDB ID code 3KYL) (1)*
	K659	RT	Interacts with DNA primer (K210 in *T. castaneum*) (2)
	Y717, V867	RT	Forms hydrophobic pocket adjacent to catalytic aspartic acid residues and could accommodate the base of deoxynucleotide substrates (Y256, V342 in *T. castaneum*) (2)
	Q833	RT	Involved in substrate specificity in proximity of active site (Q308 in *T. castaneum*) (2)
	K902	RT	Essential for telomerase activity; K902 patients display autosomal dominant dyskeratosis congenita, anticipation, and telomere shortening) (3)
RT–TRBD			Crystallographic structure of *T. castaneum* RT and TRBD bound to RNA:DNA hairpin (PDB ID code 3KYL) (1)*
TRBD–RNA:DNA	K570	TRBD	Interacts with RNA:DNA heteroduplex and is essential for telomerase activity (K570N patients suffer from aplastic anemia) (4)
	Y562, W581	TRBD	Stabilizes RNA template by providing stacking interactions (Y477, W496 in *T. thermophila* TRBD) (5)
TEN–RNA:DNA	Q169	TEN	Essential for binding of RNA:DNA heteroduplex and repetitive addition processivity (6–9)
	N and C termini	TEN	Disordered/flexible N- and C-terminal tails of *T. thermophila* TEN domain are essential for interaction with RNA (10)
	L55	TEN	Important for RNA binding: L55Q causes 50% reduction in RNA binding and catalytic activation; L55Q patients suffer from idiopathic pulmonary fibrosis (11)
TEN–RT			TEN domain in *T. castaneum* is predicted to be 40–60 Å from reverse transcriptase active site (6)†
			Distance between W203 in TEN (W187 in *T. thermophila* TEN) and reverse transcriptase active site is between 17 and 27 Å (12)†
TEN–TRBD	T116, T117, S118	TEN	T116A, T117A, and S118A mutations compromise telomere length maintenance in vivo but do not alter DNA binding affinity or activity in vitro; may be involved in TEN–TRBD interaction (9)

*Relative orientation of respective components of human telomerase complex available for homologs with solved structure.
†Predicted distance between domains or residues.

1 Mitchell M, Gillis A, Futahashi M, Fujiwara H, Skordalakes E (2010) Structural basis for telomerase catalytic subunit TERT binding to RNA template and telomeric DNA. *Nat Struct Mol Biol* 17:513–518.
2 Gillis AJ, Schuller AP, Skordalakes E (2008) Structure of the *Tribolium castaneum* telomerase catalytic subunit TERT. *Nature* 455:633–637.
3 Armanios M, et al. (2005) Haploinsufficiency of telomerase reverse transcriptase leads to anticipation in autosomal dominant dyskeratosis congenita. *Proc Natl Acad Sci USA* 102:15960–15964.
4 Xin ZT, et al. (2007) Functional characterization of natural telomerase mutations found in patients with hematologic disorders. *Blood* 109:524–532.
5 Rouda S, Skordalakes E (2007) Structure of the RNA-binding domain of telomerase: implications for RNA recognition and binding. *Structure* 15:1403–1412.
6 Sekaran VG, Soares J, Jarstfer MB (2009) Structures of telomerase subunits provide functional insights. *Biochim Biophys Acta.* 1804:1190–1201.
7 Wyatt HD, Tsang AR, Lobb DA, Beattie TL (2009) Human telomerase reverse transcriptase (hTERT) Q169 is essential for telomerase function in vitro and in vivo. *PLoS One* 4:e7176.
8 Zaug AJ, Podell ER, Cech TR (2008) Mutation in TERT separates processivity from anchor-site function. *Nat Struct Mol Biol* 15:870–872.
9 Sealey DC, et al. (2010) The N-terminus of hTERT contains a DNA-binding domain and is required for telomerase activity and cellular immortalization. *Nucleic Acids Res* 38:2019–2035.
10 Jacobs SA, Podell ER, Cech TR (2006) Crystal structure of the essential N-terminal domain of telomerase reverse transcriptase. *Nat Struct Mol Biol* 13:218–225.
11 Armanios MY, et al. (2007) Telomerase mutations in families with idiopathic pulmonary fibrosis. *N Engl J Med* 356:1317–1326.
12 Romi E, et al. (2007) High-resolution physical and functional mapping of the template adjacent DNA binding site in catalytically active telomerase. *Proc Natl Acad Sci USA* 104:8791–8796.