Title
A structurally conserved human and Tetrahymena telomerase catalytic core

Journal
Proceedings of the National Academy of Sciences of the United States of America, 117(49)

ISSN
0027-8424

Authors
Wang, Yaqiang
Gallagher-Jones, Marcus
Sušac, Lukas
et al.

Publication Date
2020-12-08

DOI
10.1073/pnas.2011684117

Peer reviewed

A structurally conserved human and *Tetrahymena* telomerase catalytic core

Yaqiang Wang[a], Marcus Gallagher-Jones[a], Lukas Sušac[a], He Song[a], and Juli Feigon[a,1]

[a]Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095-1569

Edited by Michael F. Summers, University of Maryland, Baltimore, MD, and approved October 26, 2020 (received for review June 15, 2020)

Telomerase is a ribonucleoprotein complex that counteracts the shortening of chromosome ends due to incomplete replication. Telomerase contains a catalytic core of telomerase reverse transcriptase (TERT) and telomerase RNA (TER). However, what defines TERT and separates it from other reverse transcriptases remains a subject of debate. A recent cryoelectron microscopy map of *Tetrahymena* telomerase revealed the structure of a previously uncharacterized TERT domain (TRAP) with unanticipated interactions with the telomerase essential N-terminal (TEN) domain and roles in telomerase activity. Both TEN and TRAP are absent in the putative *Tribolium* TERT that has been used as a model for telomerase for over a decade. To investigate the conservation of TRAP and TEN across species, we performed multiple sequence alignments and statistical coupling analysis on all identified TERTs and find that TEN and TRAP have coevolved as telomerase-specific domains. Integrating the data from bioinformatic analysis and the structure of *Tetrahymena* telomerase, we built a pseudoatomic model of human telomerase catalytic core that accounts for almost all of the cryoelectron microscopy density in a published map, including TRAP in previously unassigned density as well as telomerase RNA domains essential for activity. This more complete model of the human telomerase catalytic core illustrates how domains of TER and TERT, including the TEN–TRAP complex, can interact in a conserved manner to regulate telomere synthesis.

TERT | reverse transcriptase | electron microscopy | telomere | TPP1

In eukaryotic cells, the ends of chromosomes are capped by repetitive telomeric DNA, which functions as a buffer for the shortening of linear chromosomes during DNA replication and recruits telomere-binding proteins that prevent activation of DNA damage responses (1). Shortened telomeres can lead to genome instability and cellular senescence (2–4). Telomerase functions to extend the 3′ ends of chromosomes by synthesizing species-specific DNA telomere repeats using a unique telomerase reverse transcriptase (TERT) with an integral RNA template contained within its larger telomerase RNA (TER). Telomerase activity is mostly undetectable in somatic cells, but a majority (~90%) of cancers require up-regulated telomerase activity to maintain immortal cell replication (2, 5). Additionally, telomerase insufficiency or dysregulation due to mutations in TERT, TER, and telomere-associated proteins are linked to a wide variety of inherited diseases in humans, such as dyskeratosis congenita and idiopathic pulmonary fibrosis (6–9). All sequences identified as TERT to date contain the palm and fingers (reverse-transcriptase domain, RT) and thumb (C-terminal extension, CTE) domains common to reverse transcriptases, and an RNA binding domain (RBD) (Fig. 1*A*) (10). The size of TER varies among species (11, 12), but it has two evolutionarily conserved domains, template/pseudoknot (t/PK) and stem-terminus element (STE), stem-loop 4 (SL4) in *Tetrahymena* and conserved region 4/5 (CR4/5) in vertebrates, that independently bind TERT and are required for catalysis. Additional TER domains bind species-specific proteins required for biogenesis (11, 13).

While the RT domain within TERT is highly conserved, the sequences and sizes of annotated TERTs vary widely, so it has not been entirely clear which domains define TERT. Crystal structures of the putative TERT from the red flour beetle *Tribolium castaneum* (hereafter referred to as *Tribolium* TERT-like) revealed that the RBD-RT-CTE domains form a ring structure (TERT ring) (14, 15). The *Tribolium* TERT-like structure has been a useful model for guiding telomerase biochemistry and structural studies (16–24). Despite the similarity of *Tribolium* TERT-like ring structure to those subsequently identified in *Tetrahymena* and human cryoelectron microscopy (cryo-EM) structures (21–25), the annotation of *Tribolium* TERT-like as an authentic TERT (26) has been controversial as it lacks the telomerase essential N-terminal domain (TEN), which is required for telomerase's unique feature of repeat addition processivity (RAP) in human and *Tetrahymena* (16, 27–29). Absence of identification of a TER and genome sequencing of *Tribolium* telomeres that showed that they are a mixture of TCAGG telomere repeats and retrotransposons (30) further call into question annotation of *Tribolium* TERT-like as a bona fide TERT. The recently reported human telomerase model based on 7- to 10-Å cryo-EM maps revealed its overall architecture, which comprises a TERT–TER catalytic core and an H/ACA small Cajal body ribonucleoprotein (scaRNP) (24). Due to resolution limitation, TERT was modeled by rigid body-fitting the *Tribolium* TERT-like and *Tetrahymena* TERT TEN domain into the cryo-EM map. The 4.8-Å cryo-EM structure of *Tetrahymena* telomerase with telomeric DNA provides a near complete atomic model of TERT (25). One of the key features is the structurally characterized motif within the insertion in fingers domain (IFD), named TRAP, which forms an unexpected complex with the TEN domain. Like the TEN domain, residues corresponding to TRAP are absent in *Tribolium* TERT-like, which only has a short linker between two IFD helices (Fig. 1*A*).

Significance

Among reverse transcriptases, telomerase reverse transcriptase (TERT) has the unique ability to replenish DNA repeats at chromosome ends by repetitively using its integral RNA template. TERT shares a variety of common features with other reverse transcriptases as well as telomerase-specific motifs/domains. Among them, TEN and TRAP are not obviously present in all annotated TERT sequences, yet they form a complex in *Tetrahymena* telomerase that is essential for telomerase' distinctive activities. Using bioinformatic and structural analysis we show that TEN and TRAP have coevolved and propose that they are defining features of TERT. We present a model for human telomerase that fits into the published cryoelectron microscopy map that will facilitate studies of mechanism, recruitment, and mutations linked to disease.

Author contributions: J.F. designed research; Y.W., M.G.-J., L.S., and H.S. performed research; Y.W., M.G.-J., H.S., and J.F. analyzed data; and Y.W., M.G.-J., L.S., H.S., and J.F. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: feigon@mbi.ucla.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2011684117/-/DCSupplemental.

First published November 23, 2020.

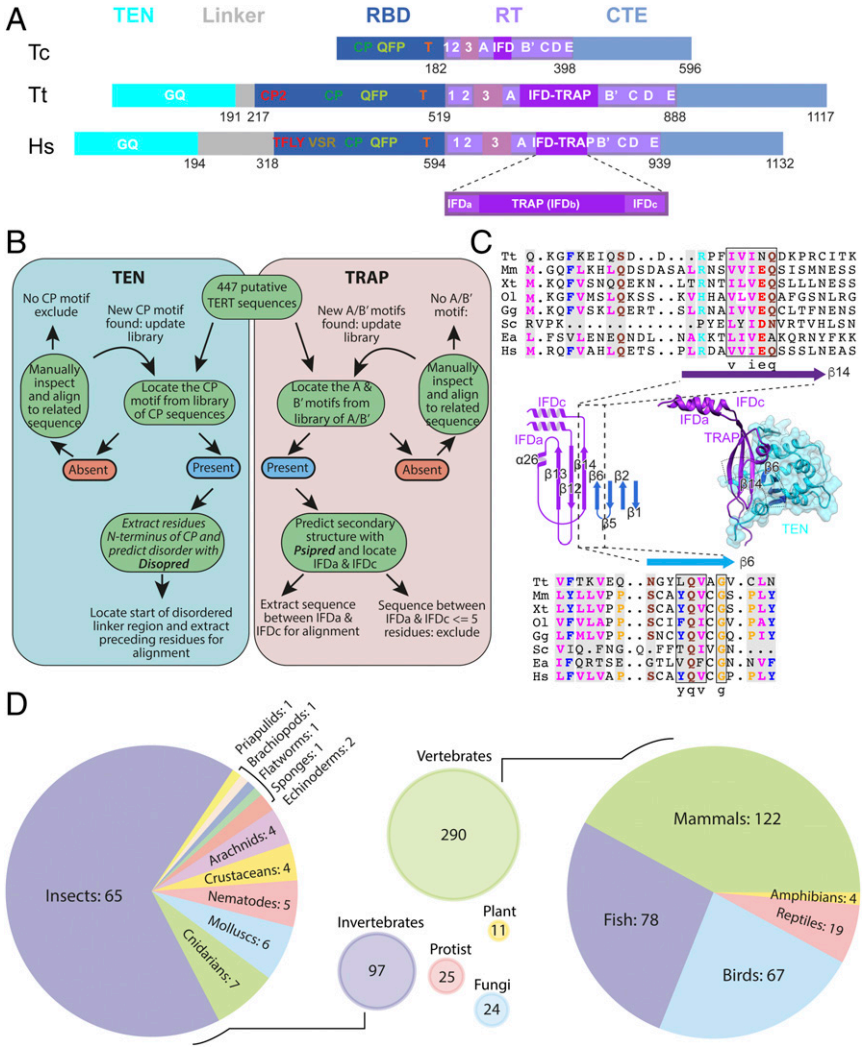

Fig. 1. TEN and TRAP have coevolved as telomerase specific domains. (*A*) Schematics comparing domains of TERT from *Tetrahymena thermophila* (*Tt*) and *Homo sapiens* (*Hs*), and TERT-like from *T. castaneum* (*Tc*), aligned at the RT domain. For the TEN domain, the conserved GQ motif is annotated. For RBD, besides the common CP, QFP, and T motifs, CP2 in ciliates (TFLY in vertebrates), and vertebrate-specific domain VSR are annotated. TERT domains are colored as follows: RT and IFD–TRAP, purple and violet; CTE, sky blue; RBD, blue; TEN, cyan. (*B*) Bioinformatic workflow for the identification of TEN and TRAP motifs from TERT sequences. (*C*) Selected sequence alignments of regions of TERT TRAP and TEN domains showing the conservation of β14 of TRAP and β6 of TEN, respectively. See *SI Appendix*, Figs. S6 and S7 for more complete alignment. The *Inset* shows the TEN–TRAP structure of *Tetrahymena* TERT (25), illustrating the interaction between these two β-strands. *Mus musculus* (*Mm*), *Xenopus tropicalis* (*Xt*), *Oryzias latipes* (*Ol*), *Gallus gallus* (*Gg*), *Saccharomyces cerevisiae* (*Sc*), *and Euplotes aediculatus* (*Ea*). (*D*) Species distribution of the 447 annotated TERT sequences spanning vertebrates, invertebrates, protists, fungi, and plants.

This led us to investigate whether TEN and TRAP are hallmarks of true TERT. To test this hypothesis, we first analyzed known TERT sequences and found that, for all organisms, TEN is always present when TRAP is present in the RT domain and that elements of TRAP secondary structure and sequences at the TEN–TRAP interface are conserved. A statistical coupling analysis across both TEN and TRAP further provided evidence that these two domains have coevolved. Using this bioinformatics information, the atomic model of *Tetrahymena* telomerase catalytic core determined from the 4.8-Å cryo-EM map (25), and NMR and crystal structures of domains of human TERT (31) and TER (32–34), we modeled the human telomerase catalytic core into the previously reported 7.7-Å cryo-EM focused refinement map of human telomerase catalytic core (24). The homology-modeled TRAP fits into previously unassigned density in the cryo-EM map. This study reveals that human and *Tetrahymena* telomerase catalytic core, especially TERT, share most features, including a conserved TEN–TRAP complex that we suggest is a defining feature of functional TERT. Our human telomerase model also provides additional structural basis for interpreting biochemistry data and mutations linked to diseases of telomerase insufficiency.

Results and Discussion

Identification of Putative TRAP and TEN Domains in Annotated TERT Sequences. The TERT RT domain has the standard RT (palm and fingers) 1, 2, A, B′, C, D, E motifs, plus two additional motifs, motif 3 and IFD (Fig. 1*A*) (12, 35–37). Motif 3 has been proposed to facilitate the realignment between DNA and RNA in telomerase (35) and to play a role in DNA primer handling (28). The IFD is located between the RT A and B′ motif, and *Tribolium* TERT-like contains two helices (IFDa and IFDc) connected by a short linker (5 aa). The linker between IFDa and IFDc is much longer for *Tetrahymena* and human telomerase (104 and 68 aa, respectively), and it was this region (IFDb) (35) that was originally identified by sequence analysis as part of IFD (36) (Fig. 1*A*). The *Tetrahymena* TERT IFDb region was recently structurally characterized as an ordered structure comprised primarily of β-strands and loops, and renamed TRAP because of its structural and

functional importance to RAP (25). Although originally thought to be unique to TERT, motif 3 and IFD (the helices IFDa and IFDc) are also found in group II intron reverse transcriptases and Penelope-like element (PLE) reverse transcriptases (38–40), as well as *Tribolium* TERT-like. In *Tetrahymena* telomerase, TRAP and TEN form a complex, in part via an extended β-sheet across the two domains (Fig. 1*C*). TEN–TRAP is located above the TERT ring, with the TER single-stranded region on the 3′-side of the template running between the CTE and TRAP (25).

To assess the general conservation of the IFD-TRAP (IFDa-TRAP-IFDc) and TEN domains within TERT, we developed a workflow (*Materials and Methods* and Fig. 1*B*) to analyze 447 putative TERT sequences spanning organisms representing vertebrates, invertebrates, protists, fungi, and plants (Fig. 1*D*). Sequences potentially corresponding to either IFD-TRAP or TEN were isolated from the full TERT sequence using the locations of the A and B′ motifs, and CP motif, respectively. To identify sequences as putative TRAP motifs, we ran a secondary structure prediction to determine the presence of the IFDa and IFDc helices that flank TRAP, as well as the potential β-strand content of each sequence. Thirty-two sequences lacking A/B′ motifs or only having a single continuous helix predicted in the IFD-TRAP region were automatically excluded from further analysis (Fig. 1*B*). Three additional sequences, corresponding to the insects *T. castaneum*, *Bombyx mori*, and *Bemisia tebaci*, had only two to five residues between IFDa and IFDc. Sequence lengths for predicted TRAP ranged from 37 to 122 residues (*SI Appendix*, Fig. S1).

To identify putative TEN domains, we analyzed the potential disorder for all residues between the CP motif and the N terminus for the same 447-sequence dataset to identify the linker region between the TEN domain and the RBD (Fig. 1*B*). We noted a greater degree of variability in this region compared to the IFD-TRAP mainly due to the large difference in the number of N-terminal residues preceding the TEN domains. Sequence lengths for predicted TEN ranged from 67 to 296 residues (*SI Appendix*, Fig. S2), after excluding 8 very short sequences that were apparent truncations due to misannotation of the start site at an internal methionine, as described below.

Of the 447 TERT sequences analyzed, we found 409 that apparently contained TRAP and 400 that contained TEN (including those anomalous sequences discussed above). All 400 that contained TEN also had TRAP, and the 9 that apparently contained TRAP but not TEN were from vertebrates, specifically *Pteropus alecto* (black flying fox), *Pelodiscus sinensis* (Chinese softshell turtle), *Chinchilla lanigera* (long-tailed chinchilla), *Corvus brachyrhynchos* (American crow), *Galeopterus variegatus* (Sunda flying lemur), *Tursiops truncatus* (common bottlenose dolphin), *Pantholops hodgsonii* (Tibetan antelope), *Oryctolagus cuniculus* (European rabbit), and *Pan paniscus* (bonobo) (*SI Appendix*, Figs. S1 and S2). Closer manual inspection of the *P. alecto* sequence revealed a lack of a conventional CP motif that was being used to identify the N-terminal linker region, and reanalysis identified a TEN domain. The *P. sinensis* sequence begins with a cysteine (C) residue (*SI Appendix*, Fig. S2), suggesting that the start site may have been incorrectly annotated. Alignment to closely related species revealed the presence of a TEN domain, with ∼40 to 80 residues of the N terminus being absent (*SI Appendix*, Fig. S3). As annotated, *C. lanigera* TERT is ∼322 residues shorter than the TERT of its close relative *Cavia porcellus*. Alignment of the two sequences showed that the first 30 residues of *C. lanigera* TERT align exactly with a sequence in *C. porcellus* TERT 300 residues downstream of the N terminus, both starting with a methionine residue (*SI Appendix*, Fig. S4). Alignment of the two sequences from this point on has an 84% sequence identity. Together, these data suggest an incorrectly identified start codon resulting in the elimination of 300 residues from the N terminus of the *C. lanigera* sequence, inclusive of the TEN domain. Similar truncations were also noted for *C. brachyrhynchos*, *G. variegatus*, *T. truncatus*, *P. hodgsonii*, *O. cuniculus*, and *P. paniscus* following alignment to related sequences.

Manual inspection of the isolated TEN domains revealed another six vertebrate sequences (*Anser cygnoides domesticus*, *Crocodylus porosus*, *Dipodomys ordii*, *Erinaceus europaeus*, *Nomascus leucogenys*, and *Panthera tigris altaica*) and two invertebrate sequences (*Aplysia californica* and *Dinoponera quadriceps*) that seemed suspicious as they were much shorter than other TEN domains, contained little secondary structure, and aligned poorly to other sequences. To determine the cause of this we aligned the full-length TERT sequence from these organisms to full-length TERT from closely related species. The start of all of the anomalous sequences aligned to a methionine residue within the disordered linker of their close relatives. The high sequence similarity past the disordered region suggest that these sequences as annotated represent truncations. Therefore, we used 394 TEN sequences, including *P. alecto* and *P. sinensis*, for the final alignment. We conclude that all of the apparently anomalous vertebrates do have TEN domains, and that all known vertebrate TERTs have TRAP and TEN sequences.

Among Insects Only *Hymenoptera* Have TERT TRAP and TEN. Multiple sequence alignments were performed on the identified TRAP and TEN sequences to determine potential conserved regions within the TEN and TRAP motifs. For all TERT sequences with a known TER (i.e., from confirmed telomerases), we find that both the TEN and TRAP have large portions of the sequence that are highly conserved irrespective of phylum (*SI Appendix*, Figs. S1 and S2). This also extends to sequences from species closely related to those with a known TER, including all vertebrate species that we analyzed. As previously noted, the TERT sequence is highly conserved across all vertebrates, with pairwise identity ranging from 20 to 99% and an average identity of 50.3% compared to the human sequence (12, 41).

The nematode TERT sequences, including *Caenorhabditis elegans*, are short compared to other TERTs (<600 residues), but still contain a significant insertion (30 to 55 residues) between the IFDa and IFDc helices, consistent with TRAP, as well as a GQ motif (42), consistent with TEN, although shortened compared to vertebrates (80 to 120 compared to 180+ in vertebrates). To date, no TERs have been identified for insects (12), but presumptive TERTs have been identified for 65 insect species, while *Drosophila melanogaster* is known to lack telomerase (43). The *Tribolium* TERT-like and similar putative TERT of *B. mori* have insertions of less than six residues between IFDa and IFDc and lack both a TEN domain and GQ motif. In fact, the majority of the sequences automatically excluded from our analysis belonged to insects of the order *Lepidoptera* and *Coleoptera* (19 sequences), for which no TER has been identified. The putative TEN domains of insects in these orders also align poorly to those of other species, with many having few residues upstream of the CP motif in the RBD or lacking a canonical CP motif entirely, as has previously been reported (44).

Among insects, members of the *Hymenoptera* family, including *Apis mellifera* (honey bee), have the largest insertion between IFDa and IFDc (40+ residues compared to an average length of 67 across all TERT sequences analyzed) between the IFDa and IFDc helices, with several secondary structure elements predicted. Alignment of the *Apis* species N-terminal regions to those of vertebrate sequences reveals a shared I/L/V, KKx motif, and asparagine (N) residue in a region associated with the GQ motif in the TEN domain that is not present in other insects (*SI Appendix*, Fig. S2). Phylogenetic clustering of IFD-TRAP from arthropods reveals that *Hymenoptera* species exist in a cluster distinct from many other insect species and have a large phylogenetic distance from both *Lepidopteran* and *Coleopteran* species (*SI Appendix*, Fig. S5). The presence of TRAP in *Hymenoptera* sequences and the dissimilarity to other insect sequences suggest

that they may have retained a functional TERT that was lost in other insects. In summary, we find that in all cases where TEN appears to be truly absent (not the result of accidental truncation due to genomic misannotation), TRAP is also absent from the sequence. Conversely, TEN and TRAP appear to always occur together, and among insects are only present among the *Hymenoptera* family.

Sequences Corresponding to the TEN–TRAP Interface in *Tetrahymena* Are Conserved and Coevolved. The IFDa and IFDc helices at the IFD N and C termini, respectively, are highly structurally conserved (*SI Appendix*, Fig. S1). Within the TRAP, the greatest variation is observed in the middle of the sequence and can be attributed to the presence of an apparently unstructured loop region that varies in length across different species. Both sequence and length are highly variable, even within a highly conserved class of TERTs, such as those found in vertebrates. Close inspection of the aligned sequences of IFD-TRAP and of TEN, however, uncovers several regions where conserved residues are clustered. Cross-referencing these regions with the cryo-EM structure of *Tetrahymena* telomerase TERT (25) reveals that they align with β6 of TEN (*SI Appendix*, Fig. S6) and β14 of TRAP (*SI Appendix*, Fig. S7), respectively. These two β-strands, far apart in TERT sequence, interact to form an extended β-sheet that is part of the TEN–TRAP interface (Fig. 1*C*). Given how well conserved both sequence and biochemical properties of the residues in these regions are, these results suggest that the interaction between TEN and TRAP is a conserved and essential element of TERT.

To investigate coevolution between TEN and TRAP more quantitively, we conducted statistical coupling analysis (SCA) to identify conserved, coevolving groups of residues (45–48). Such coevolving groups are called sectors, and typically comprise physically contiguous residues in three-dimensional (3D) space. We applied SCA to the multiple sequence alignment of 394 sequences identified as containing both TEN and TRAP and computed the positional conservation represented by the relative entropy D_i for each individual position in the alignment (Fig. 2*A*). As described above, residues comprising β6 of TEN and β14 of TRAP, located at the TEN–TRAP interface in the *Tetrahymena* structure, were found to be well conserved. To examine which regions of TEN and TRAP potentially had conserved interactions, we calculated a conservation-weighted covariance matrix $C_{i,j}$ between all aligned positions of TEN and TRAP (Fig. 2*B*). Inspection of the matrix clearly indicates there are significant correlations between many residues of TEN and TRAP. Automated sector identification based on the resulting pattern of coevolution identified a single sector composing 95 residues, ∼30% of amino acid positions of TEN and TRAP. Mapping these sector residues onto the 3D structure of *Tetrahymena* TEN and TRAP shows that they form physically contiguous units in the tertiary structure, including the β-strands at the TEN–TRAP interface (Fig. 2 *C* and *D*). The identification of coevolving residues that show structural contiguity between TEN and TRAP reveals coevolution between the two domains.

TRAP can Be Modeled into Previously Unassigned Density in the Human Telomerase Cryo-EM Map. To date, cryo-EM structures of telomerase from two species have been reported: Human at 7 to 10 Å (24) and *Tetrahymena* at 8.9 and 4.8 Å (22, 25). The human telomerase cryo-EM structure, which revealed the architecture of telomerase catalytic core and H/ACA RNP, was conservatively modeled using rigid body-fitting of then known domains (24), thus the TRAP motif was not modeled. To determine if a homologous TEN–TRAP interaction fits to the human cryo-EM map, we obtained a more complete model of the human telomerase catalytic core (Fig. 3 *A* and *B*) using information from the subsequently available 4.8-Å *Tetrahymena* telomerase structure (Fig. 3*C*), crystal and solution structures of TERT and TER domains, and sequence alignments of TERT (*SI Appendix*, Figs. S1 and S2). TERT TEN, RT including the IFD and TRAP, RBD, and CTE (*SI Appendix*, Fig. S8), were modeled as described in *Materials and Methods* (31, 49–52). Comparison of the modeled human and *Tetrahymena* TERTs show that they superimpose well, with the main differences occurring on the outside of the TERT ring, especially for the CTE (Fig. 3*D*).

In the model, the IFD helices are positioned similarly to those in *Tetrahymena* telomerase on the outer edge of the TERT ring (Fig. 4*A*). The homology model of human TRAP fits into

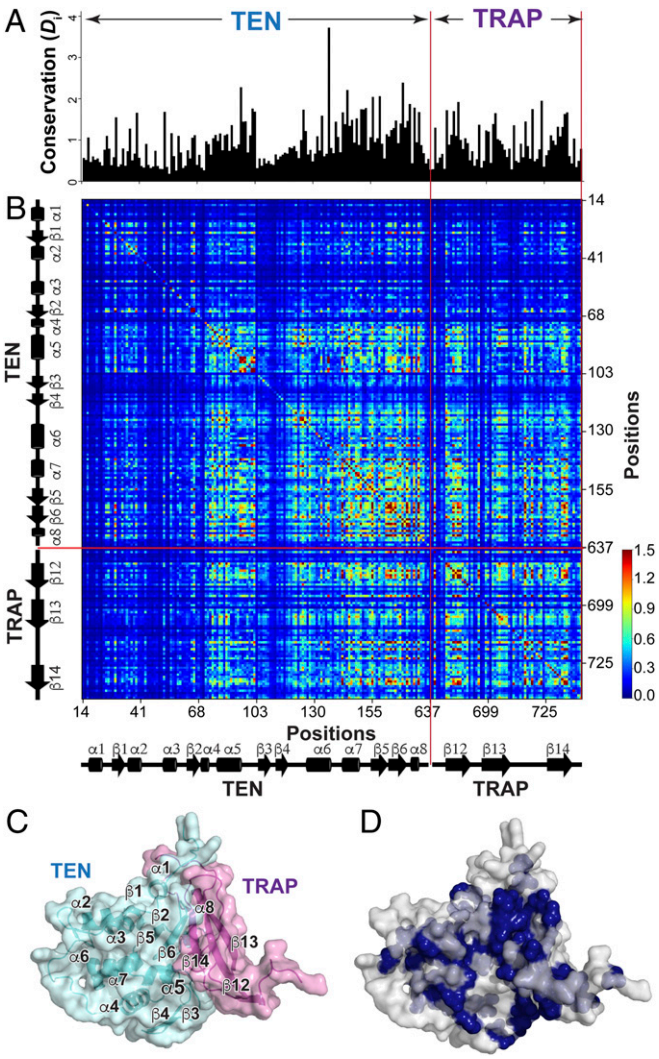

Fig. 2. Statistical coupling analysis of TEN and TRAP. (*A*) Positional conservation of TEN and TRAP. Conservation is computed as the Kullback–Leibler relative entropy (D_i) and residue positions are numbered according to *Tetrahymena* TERT. The graph is aligned with the matrix below. (*B*) Conservation-weighted positional covariance matrix $C_{i,j}$ for aligned TEN and TRAP sequences. The color of the diagonal elements reflects the conservation of individual positions, with red and blue of the spectrum corresponding to high and low conservation. Color of the off-diagonal elements reflects the correlation between two residues, with the red end of the spectrum reflecting a strong correlation, while the blue end of the spectrum reflects no correlation. (*C*) Structure of *Tetrahymena* TEN and TRAP (PDB ID code 6D6V) with secondary structure elements labeled. TEN is colored in cyan and TRAP is colored in purple. (*D*) Mapping of sector residues onto the *Tetrahymena* TEN and TRAP structure (PDB ID code 6D6V). Residues that are coevolving as a sector are shown as dark blue spheres (the light blue are interior residues). Nonsector positions are shown as a gray surface.

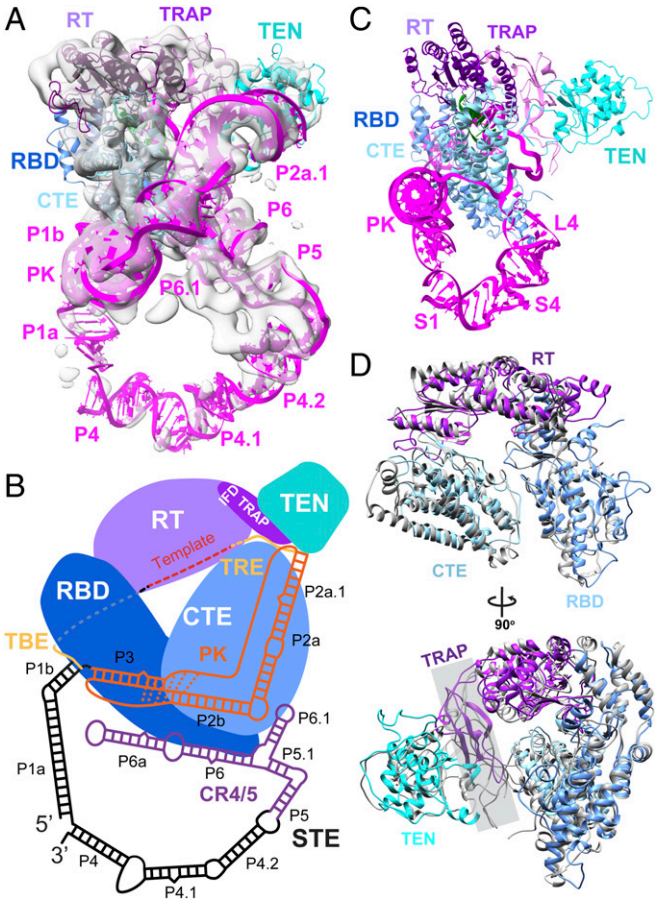

Fig. 3. Structure of human telomerase catalytic core. (*A*) Model of human TERT and TER from this work fit into the previously reported cryo-EM density map (EMDB 7518). TER is magenta and TERT subdomains are colored as in Fig. 1. (*B*) Schematic of TERT and TER subdomain arrangements of human telomerase catalytic core. (*C*) cryo-EM structure of telomerase catalytic core of *Tetrahymena* (PDB ID code 6D6V). (*D*) Model of human TERT RBD-RT-CTE (blue, purple, light blue) superimposed on *Tetrahymena* TERT (PDB ID code 6D6V) (gray). TRAP (purple in the shade) and TEN (cyan) are omitted for clarity in the *Upper* view.

previously unassigned density in the cryo-EM map and is attached at each end to the IFDa and IFDc helices, respectively (Fig. 4*A*). The overall L-shaped human IFD-TRAP is similar to the *Tetrahymena* IFD-TRAP, and it is also comprised primarily of β-strands and loops. For the homology-modeled TEN domain, due to the low resolution of the cryo-EM map especially in this region, its orientation cannot be unambiguously fixed. Based on the sequence conservation noted above, the TEN domain was positioned such that an extended β-sheet, with a shared TEN–TRAP interface (β14 in TRAP and β6 in the TEN domain using *Tetrahymena* TERT numbering), was formed as seen in the *Tetrahymena* TERT structure (Fig. 3*D*). TEN is on top of TRAP and does not directly contact the TERT ring (Fig. 4 *A*–*C*).

Tetrahymena telomerase TRAP has 26 more residues in the middle of its sequence that are absent in human TRAP. Interestingly, there is no visible cryo-EM density for those additional residues, indicating considerable flexibility (25), whereas in human the entire TRAP is apparently ordered and can be modeled (Fig. 4*A*). Several mutations in the IFD-TRAP are associated with diseases of telomerase insufficiency, including dyskeratosis congenita [M773T (53), V777L (54), R811C (55)], bone marrow failure [Q722X (56)], idiopathic pulmonary fibrosis [R742H (57), V747A (58), V777L (54)], aplastic anemia [T726M (59), Y772C (60)], and acute myeloid leukemia [P785L (61)], indicating the importance of this region for telomerase activity (Fig. 4*D*). In addition, V791, whose substitution to Y results in defects on telomere length maintenance and cell immortalization (62), is located at the TEN–TRAP interface, as previously predicted based on the *Tetrahymena* telomerase structure (25). These disease mutations highlight the importance of the TRAP for telomerase activity.

TPP1 Can Be Docked onto the TERT Model. Both TEN and TRAP have been implicated in recruitment of telomerase to telomeres and activation, via interaction with TPP1 (62, 63), a component of human shelterin (64). Human TPP1 transiently associates with telomerase TEN via a TEL patch (TPP1 glutamate [E] and leucine [L]-rich patch) and N-terminal of the oligonucleotide/oligosaccharide-binding (OB)-fold (NOB) to recruit it to telomeres (65–71). In *Tetrahymena* telomerase, the constitutively associated p50 serves a similar function as TPP1 (10, 22, 72) and is critical for telomerase recruitment and activation (73, 74). p50 interacts with both TEN and TRAP and has been proposed to stabilize TRAP folding and the TEN–TRAP complex, to help activate telomerase (25). Although TPP1 is not a component of the purified human telomerase (24), we modeled the TPP1 OB-fold onto the human telomerase catalytic core based on the orientation of the p50 OB-fold and its interaction with TEN (Fig. 4*E*). The model shows that the TEL patch and those residues (K78, N125, T128, L139) on TEN that are supposed to interact with it (71) are on facing surfaces. Additionally, the TPP1 NOB is positioned near the IFD-TRAP at the corner of the L shape, similar to the interaction between p50 and *Tetrahymena* IFD-TRAP (25). These results provide further support for the reliability of the model and essential function of the TRAP motif in telomerase recruitment, as well as activity (discussed further in *Summary and Conclusions*).

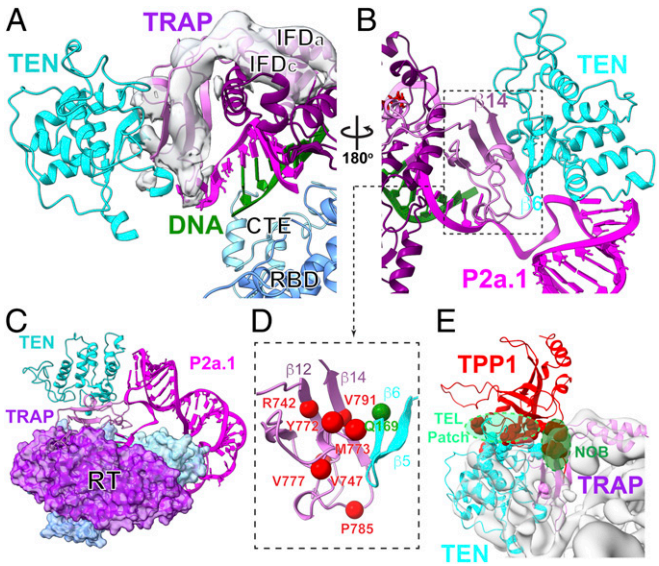

Fig. 4. TEN–TRAP interaction is conserved. (*A*) View of TRAP modeled into previously unassigned density in the cryo-EM map of human telomerase and interaction with TEN. (*B*) View of the TEN domain showing interactions with the end of P2a.1 duplex and TRAP. (*C*) Side view of the catalytic core showing TEN stacked on TRAP, which is above the TERT ring. (*D*) Ribbon depiction of the TEN–TRAP interaction, with positions of TRAP (red) and TEN (green) residues whose substitutions are associated with human disease or affect activity labeled. The numbering of human TRAP TEN secondary structure elements follows that of the *Tetrahymena* TERT structure. (*E*) Modeled interaction between TPP1, TRAP, and TEN. TEL patch and NOB are highlighted. Residues proposed to be involved in interactions between TPP1 (PDB ID code 2I46) and TEN or TRAP are shown as spheres.

Modeling of TER in the Human TERT–TER Catalytic Core. To complete the model of the human telomerase catalytic core and evaluate the TERT–TER interactions, we modeled the t/PK, STE (CR4/5), and connecting RNA into the cryo-EM map with the homology modeled TERT (Fig. 3 *A* and *B*). Previous modeling of TER used only individual subdomain structures from the human t/PK, the teleost fish medaka CR4/5, and A-form helices (24). Based on the path and the relative location of the conserved structural motifs in *Tetrahymena* telomerase (Fig. 3*C*) and biochemical and structural studies (32–34, 49, 75–77), we added previously unmodeled regions, including the human CR4/5 (Fig. 3 *A* and *B*). The general features of TER in this more complete model largely conform to those regions previously described (24), but together with TERT reveal several important TERT–TER interactions, as discussed below and in the following section. In the model, the t/PK forms a circle (Fig. 5*A*) that wraps around the equator of the TERT ring (Fig. 5*B*) with the template on RT, 5′ template-flanking single-stranded residues on the RBD, 3′ template-flanking residues between CTE and TRAP, and the double-stranded region containing the P2/P3 PK on the RBD and CTE. In the PK, there is an asymmetric internal loop (*J2a/b*) that results in an ∼90° bend between the P2a and P2b helices, consistent with the solution NMR structure of P2ab (Fig. 5*B*) (33), which facilitates wrapping of the PK around the TERT ring. P2b, the region of the PK with the loop-stem interactions that form a triple-helix, and P2a are anchored on the CTE, while the P3 region is on the RBD (Fig. 5*B*). In *Tetrahymena*, which has a smaller PK, the region corresponding to P2 is single-stranded. P2a.1 is a mammalian-specific domain and it is a linear extension on P2a (Fig. 5*B*), as predicted (33). The end of P2a.1 abuts the TEN domain (Fig. 4*B*), positioning TEN relatively further from the TERT ring than in *Tetrahymena*.

The TER P1-P4-P5 connecting t/PK and CR4/5 form a large U-shape below the TERT ring (Fig. 3 *A* and *B*), which is composed of helices interspersed with internal loops that allow it to bend. In *Tetrahymena*, the equivalent region S1-S4 forms a smaller U-shape below the TERT ring (Fig. 3*C*) that is bound by La-related protein p65. The much larger U-shape below the TERT ring in human TER (Fig. 3 *A* and *B*) provides space for the binding of one set of H/ACA RNP proteins.

In vertebrates, the CR4/5 that forms the STE has a three-way junction (3WJ) structure with P5, P6 (P6a and P6b in human TER), and P6.1 stem-loops (Fig. 5*C*), and P6a and P6.1 hairpin are required for telomerase activity (75, 76, 78, 79). Solution NMR and X-ray crystal structures of the smaller internal loop in medaka CR4/5 show that it is highly structured (Fig. 5 *D* and *F*) (34, 49). In the 3WJ, secondary structure prediction by Mfold (80), SHAPE chemical probing (81), and sequence conservation of G253-G254-C255 and G319-C320-C321 among vertebrate TERs (82) (*SI Appendix*, Fig. S9) suggest that these nucleotides form a 3-base pair helix, here named P5.1, with internal loops (*J5/6* and *J5.1*) on each side (Fig. 5*D*). This more structured internal loop along with helices P5, P5.1, P6.1, and P6 fit well into the cryo-EM map. In the modeled CR4/5 structure, P6 and P6.1 form an L shape with P6.1 as the short arm and P6 stacked on P5.1 (Fig. 5*E*). Both P6.1 and P6 interact with the RBD, which binds in the corner of L between the two helices, while only the P6.1 loop and top of the stem contact the CTE (Fig. 5*G*). Interestingly, the two conserved pseudouridines (Ψ306 and Ψ307) in the P6.1 (76) are located in the P6.1–CTE interface, suggesting they might contribute to affinity. We conclude that the human STE functions to stabilize the closed form of the TERT ring by bridging the RDB and CTE, similar to *Tetrahymena* loop 4 that inserts between RBD and CTE (Fig. 5*G*). The t/PK positions the template and template-adjacent residues on TERT for regulation of telomere repeat synthesis.

Single-Stranded Regions that Flank the Template May Function as Template Boundary Elements. The single-stranded residues between P1b that closes the t/PK circle and P2a.1 at the end of the PK traverse the RBD (5′ of the template), the RT (template), and pass between the CTE and TRAP (3′ of the template) all on one face of the TERT ring. During telomeric DNA repeat synthesis, the single-stranded telomeric DNA has to base pair with the alignment region of the template and then the template has to move through the active site as each of the 6 nt of the telomeric repeat are successively added. In *Tetrahymena* telomerase, structural studies showed that the bottom base pair and single-stranded nucleotides that flank stem 2 are anchored on the RBD to prevent nucleotides beyond the end of the template from moving into the active site, acting as a template boundary element (TBE) (19, 22), while nucleotides between the 5′ end of the template and the 3′ end of the TBE, named TBE_L (25), are proposed to allow flexibility for the template to move. Previous studies suggested that the TBE in human TER is the PK-enclosing helix P1b and flanking 3′ single-stranded

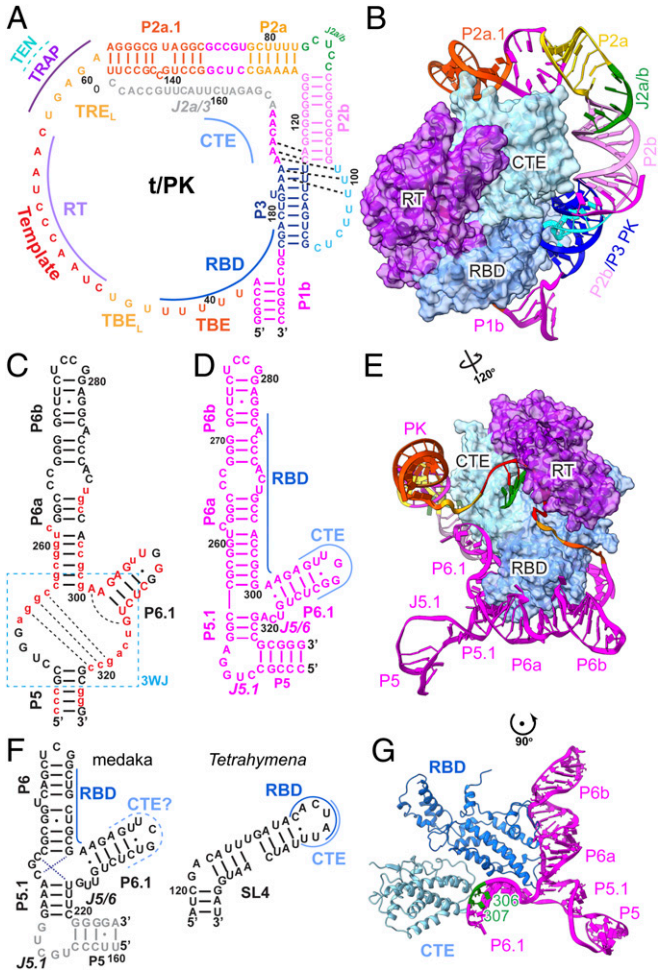

Fig. 5. Interactions between TERT and TER. (*A*) Sequence and predicted secondary structure of t/PK and site of interactions with TERT domains. (*B*) Interaction of P2-P3 PK (ribbon) with TERT ring (space-fill) showing that the P2-P3 PK wraps around the TERT ring and bends at *J2a/b*. (*C*) Sequence and predicted secondary structure of human CR4/5. Residues 100% conserved among vertebrate TER are shown as capital letters in red, and 80% conserved as lowercase letters in red. The 3WJ region is boxed by the blue dashed line. (*D*) Secondary structure of human CR4/5 model reported here. (*E*) Interactions of t/PK and CR4/5 with TERT ring. (*F*) Secondary structure of medaka CR4/5 based on crystal structure (PDB ID code 4O26) with the predicted P5 and *J5.1* loop colored in gray and *Tetrahymena* SL4 cryo-EM structure (PDB ID code 6D6V). (*G*) Model of the structure of CTE-RBD–CR4/5. The nucleotides 306 and 307 in P6.1 are in green.

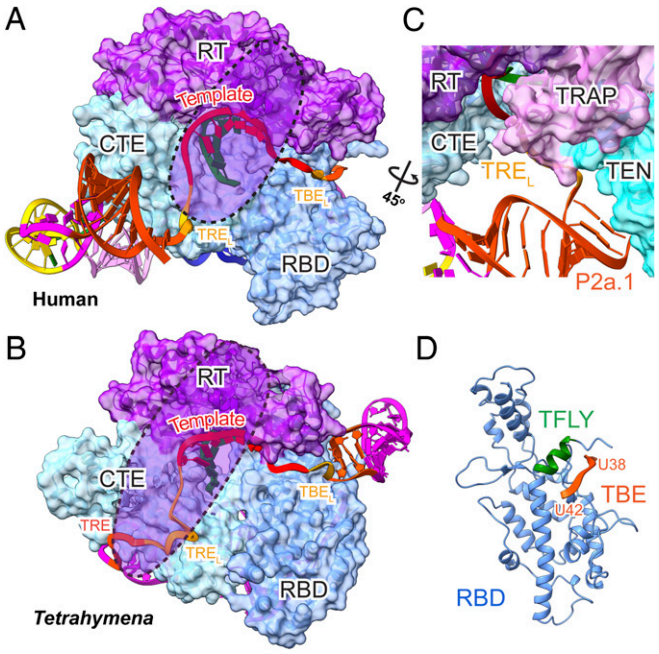

Fig. 6. TBE-template-TRE. (*A* and *B*) Comparison of human (*A*) and *Tetrahymena* (*B*) interactions of TBE-template-TRE with TERT ring. TRAP and TEN have been removed for clarity. The outline of TRAP above the TERT ring is shown as light transparent purple shadow. (*C*) Close-up view of single-stranded RNA 3′ of the template (TRE_L) passing between TRAP and CTE. P2a.1 at the end of the PK is also shown abutting the TEN domain. (*D*) Ribbon model of RBD showing the interaction of proposed human TBE $UUUUU_{38–42}$ (orange) residues 5′ of the template with the TFLY motif (green).

nucleotides (11, 83). However, P1b is located further from the template than stem 2 is in *Tetrahymena* telomerase (~8 nt vs. 4 nt) (Fig. 5*A*). The vertebrate RBD has a TFLY motif that is equivalent to CP2 in *Tetrahymena* (*SI Appendix*, Figs. S8 and S10) (49, 50). The $UUUUU_{38–42}$ located 3′ of P1b wrap around the TFLY of RBD and are positioned similarly to the TBE in *Tetrahymena* at the CP2 motif (Fig. 6*D*).

On the other side of the template, the single-stranded TER threads between the CTE and TRAP (Fig. 6 *A* and *C*), in apparent homology to the region defined as template recognition element linker (TRE_L) in *Tetrahymena* (Fig. 6*B*). In *Tetrahymena* the TRE_L is proposed to be captured by TRAP as the template moves during telomere repeat synthesis (25), while the 3′ adjacent single-stranded region (TRE, between TRE_L and PK) is anchored on the CTE (Fig. 6*B*). In human TER, with its larger PK, our proposed TRE_L is followed by the double-stranded P2. While P2a.1 sticks out from the CTE to contact TEN (Figs. 4*C* and 6*C*), P2a and P2b apparently interact with the CTE to anchor that part of the PK (Fig. 5*B*). In summary, we propose that the single-stranded TER on either side of the template in *Tetrahymena* and human telomerase that span one side of the TERT ring are structurally and functionally conserved elements, TBE_L and TRE_L, that allow the RNA to expand and compress during each round of telomere repeat synthesis, as proposed for *Tetrahymena* telomerase (Fig. 6 *A* and *B*) (19, 22, 25, 84). Furthermore, a stretch of $UUUUU_{38-42}$ 5′ of TBE_L, rather than P1b, functions as the TBE.

Summary and Conclusions

Since the first TERT was identified three decades ago (85), multiple protein domains/sequence motifs have been proposed as telomerase-specific features to separate telomerase from other reverse transcriptases. These include TEN with a GQ motif (42), motif 3 and IFD within the RT domain (12, 35–37), and the CP, CP2 or TFLY, and T motifs in the RBD (50, 86–88). The difficulties of identifying TERT arise for two reasons. First, motifs initially thought to be telomerase-specific were later discovered in other related reverse transcriptases: That is, motif 3 and IFD in group II intron RTs; and motif 3, IFD, and RBD in PLE RT. Second, identification of the corresponding TER that would help validate the presence of telomerase remains challenging, due to sequence, length, and structural diversity across species (11). Here, using an integrative approach combining bioinformatics and structural biology, we identified TEN and TRAP as hallmarks of functional TERTs. Multiple sequence alignments and statistical coupling analysis not only revealed the simultaneous presence of TEN and TRAP in TERT from all species known to have telomerase, but also showed that TEN and TRAP have coevolved. Molecular modeling of the human telomerase catalytic core into the published cryo-EM map (24) provided structural evidence for a conserved TRAP structure and TEN–TRAP interface in human telomerase, as previously observed by cryo-EM in *Tetrahymena* (25).

This is in striking contrast to *Tribolium* TERT-like (14), whose annotation as a TERT was based on the sequence alignment on RBD and RT domains and the detection of TCAGG repeats at chromosome ends (26). However, *Tribolium* TERT-like intrinsically lacks the TEN domain, and its partner telomerase RNA have not been identified to date. The TEN domain is required for telomerase's unique mechanism of RAP (16, 27–29). Indeed, only reverse-transcription activity but no RAP has been observed in *Tribolium* TERT-like (15). Analysis of the complete genome of *Tribolium* shows that its telomeres are a mix of retrotransposons and simple TCAGG repeats, suggesting that although the *Tribolium* TERT-like shares some domains with bona fide TERTs, it has either lost telomerase function or never possessed it (30). Strikingly PLE, a class of telomere-associated retroelements, possess an identical domain organization to *Tribolium* TERT-like (40). Taken together, this finding lends further support to the conclusion that the "streamlined" *Tribolium* TERT-like is not a functional TERT, but may represent an intermediate stage between a telomerase-based mechanism and a retrotransposon-based mechanism of telomere maintenance. Similarly, other putative TERTs that share *Tribolium* TERT-like's domain architecture and lack TEN and TRAP (i.e., from *B. mori*) are all expected to not have telomerase function. Indeed, to the best of our knowledge, no telomerase activity has been shown in any organisms that fall into this category.

Turning back to functional TERTs, structural data together with previous biochemical data (35, 36, 62, 67, 89) indicate that TRAP and TEN serve three essential functions in *Tetrahymena* and human telomerase that we suggest are common to all telomerases. First, TRAP and TEN are involved in what is likely the last step of assembly of the telomerase catalytic core. Formation of the TEN–TRAP complex requires that TRAP extends down from the RT, from the top of the TERT ring over the TER in the t/PK, while TEN, which is connected to the RBD by a flexible linker, extends up from the bottom of the TERT ring to bind on the outside of TRAP, thereby physically sequestering the t/PK on the TERT ring and likely stabilizing the TRAP fold (25). Second, TRAP in the TEN–TRAP complex physically interacts with the single-stranded RNA on the 3′ side of the template, thereby participating in nucleotide addition processivity and RAP by a mechanism whose structural details remain to be elucidated. TRAP may also play a role in stabilizing the initial short template-primer duplex, a function previously attributed to the TEN domain (16, 27–29). Third, both TRAP and TEN have binding epitopes with TPP1 (p50 in *Tetrahymena*), which is required for telomerase recruitment to telomeres and may also stabilize a more active form of telomerase.

Finally, comparison of the model of human telomerase catalytic core reported here to the published structure of *Tetrahymena* telomerase (25), highlights that single-stranded RNA segments (TRE-template-TBE) and their interactions with TERT largely govern telomere repeat synthesis. Thus, despite large differences in TER size and structure in both species, there are extensive mechanistic similarities. The STE holds TERT in a ring conformation by acting as a keystone; the single-stranded TER segments that flank the template are optimally positioned relative to the active site by TEN–TRAP and RBD. Given the central role of TEN–TRAP in establishing this structural framework for telomerase activity, we conclude that the presence of TEN–TRAP, as evidenced through the bioinformatics workflow presented here, can be used to identify true TERTs, and we predict conserved roles and interactions for TEN–TRAP in functional telomerases.

Materials and Methods

Alignment of IFD-TRAP Sequences. An initial list of 390 TERT sequences was retrieved from the National Center for Biotechnology Information (NCBI) gene database that were labeled as orthologous or similar to TERT. An additional 57 sequences not listed as similar to TERT but listed in the telomerase database (12) were retrieved from the NCBI protein database to bring the total number of sequences to 447. IFD-TRAP sequences were isolated from the full-length sequence using a custom Python script. The script first identified the B′ motif and worked backward through the sequence one residue at a time until the A motif was encountered, and the region in between was used in subsequent alignments. This script eliminated 19 sequences were no B′ or A motif could be located. Putative IFD-TRAP regions were prescreened for potential secondary structure by PsiPred v2.6 (90). Using the secondary structure prediction, IFDa and IFDc helices were identified and the residues in between extracted for further alignment. Sequences with only a single predicted helix or fewer than six residues between IFDa and IFDc (19) were excluded from subsequent analysis. For three of the sequences, IFDc was misidentified, causing an artificial truncation. Manual inspection was required for these sequences to extract the entire TRAP sequence. The resulting 409 sequences, ranging in length from 37 to 122 residues, were used for final alignments. Multiple sequence alignments were performed using PROMALS followed by slight manual adjustments (91, 92).

Alignment of TEN Domain Sequences. Putative TEN domains were first extracted from the 447 full-length sequences by locating the CP motif and then extracting all residues preceding it up until the N terminus. This eliminated 40 sequences from subsequent analysis as no CP motif could be identified. Closer manual inspection of the 40 eliminated sequences revealed 2 that actually contained a noncanonical CP motif and were added back into the analysis. DisoPred v3.1 (90) was used to identify the start of the linker region between TEN and RBD by identifying a stretch of disordered residues that existed for more than 10 consecutive residues. An additional seven sequences were excluded at this stage due to an inability to identify the disordered linker region. The sequence preceding this region was then used as the TEN domain for multiple sequence alignments in PROMALS. Additionally, PsiPred was used to analyze the presence of secondary structure elements within the TEN domain.

Alignment of Full-Length TERT Sequences. Alignment of the full-length *P. sinensis*, *C. lanigera*, *C. brachyrhynchos*, *G. variegatus*, *T. truncatus*, *P. hodgsonii*, *O. cuniculus*, *P. paniscus*, *A. cygnoides domesticus*, *A. californica*, *C. porosus*, *D. quadriceps*, *D. ordii*, *E. europaeus*, *N. leucogenys*, and *P. tigris altaica* TERT sequences to their respective close relative sequences was performed using Clustal Omega with default settings (93).

Statistical Coupling Analysis. For the SCA analysis of TEN and TRAP, the 394 TERT sequences containing both TEN and TRAP were used. Multiple sequence alignments of both TEN and TRAP were concatenated together using T-Coffee sequence reformatting tools for SCA analysis (48). The SCA was performed using pySCA toolbox (http://reynoldsk.github.io/pySCA/). All calculations were performed following the toolbox's instruction, and the theoretical principles underlying SCA analysis were described in detail previously (47, 94). Briefly, using the scaProcessMSA script, the final alignment was realigned with the structure of *Tetrahymena* TEN and TRAP (PDB ID code 6D6V) as a reference, during which the unaligned or highly gapped positions were truncated. After removing these positions, sequences with less than 10% identity to the reference sequence were also removed. After the preprocessing step, a total of 301 sequences composed of 216 positions were retained for subsequent calculations. Using the scaCore script, conservation values for individual amino acid positions were calculated as the Kullback–Leibler relative entropy (D_i) and a conservation-weighted covariance matrix $C_{i,j}$ between all pairs of amino acid positions was obtained (94). To identify coevolving groups of residues within the protein, spectral decomposition analysis was performed on the covariance matrix. Automated sector identification was performed using the scaSectorID script, which defined a single sector that includes 95 positions that are coevolving within TEN and TRAP. The full analysis can be downloaded from (https://github.com/marcusgj13/tert_TRAP_TEN).

Modeling of Human TERT and TER. Fitting of atomic coordinates into the cryo-EM maps was performed using Coot (95) and University of California, San Francisco (UCSF) Chimera (96). The TERT TEN, RBD, RT, and CTE domains were modeled as follows. The initial TEN homology model (residues 12 to 180) was generated by the SWISS-MODEL server using the cryo-EM structure of *Tetrahymena* TEN (PDB ID code 6D6V) and *Hansenula polymorpha* TEN (PDB ID code 5NPT). For placement of TEN into the cryo-EM map we took advantage of the published structure of *Tetrahymena* telomerase (25), and the sequence analysis described in this work, ensuring that the TEN–TRAP interface remained consistent and did not become distorted. The human RBD (residues 321 to 594) model was generated by the SWISS-MODEL server based on the *Takifugu* RBD crystal structure (PDB ID code 4LMO). The human RT domain (residues 595 to 964) is a homology model generated from *Tetrahymena* RT (PDB ID code 6D6V). The CTE domain (residues 993 to 1122) was adapted from the crystal structure of human CTE (PDB ID code 5UGW) excluding the N-terminal helix (residues 965 to 992), whose position in the crystal structure is apparently in a nonnative conformation due to crystal packing, so the N-terminal helix was manually fit to the density. The C terminus of CTE (residues 1123 to 1132) was a homology model of *Tetrahymena* CTE C terminus. All of the subdomains were connected and manually adjusted against the cryo-EM density (EMDB 7518) in Coot.

Human TER was modeled by fitting of NMR and crystal structures of subdomains into the cryo-EM map (EMDB 7518) using UCSF chimera and refining as follows. The NMR solution structures of P2ab (residues 77 to 92 and 122 to 132, P2a-J2a/b-P2b, PDB ID code 2L3E), P2a1a (residues 64 to 76 and 133 to 145, P2a-P2a.1) (33), and PK (residues 93 to 121 and 170 to 184, P2b-P3, PDB ID code 2K95). P1a, P1b, P4, P4.1, P4.2, P5, P5.1, and P6b were built by RNAComposer (97). The model of CR4/5 (residues 253 to 265, 292 to 306, and 311 to 321) was adapted from a crystal structure of medaka CR4/5 (PDB ID code 4O26) since they share a conserved 3WJ, and the P6.1 loop (residues 307 to 310) was from the NMR structure (PDB ID code 2KYE). An ideal 6-bp A-form helix for the DNA–RNA template duplex (RNA residues 50 to 55 and DNA residues 1 to 6) was fit into and refined to the density. All of the subdomains were connected by corresponding single-stranded residues (except nucleotides 148 to 168 of *J2a/3*) and manually adjusted against the cryo-EM density in Coot. The telomerase human catalytic core comprises 1,132 amino acids (TERT) and 370 nucleotides (residues 1 to 370, including t/PK, CR4/5, and connecting region), of which we modeled 981 aa and 332 nt. This model accounts for 87% of the TERT–TER catalytic core, and most of the missing residues are from the disordered region between TEN and RBD.

The entire TERT–TER complex was refined in Phenix (98) using "phenix.real_space_refine" with secondary structure, Ramachandran, and rotamer restraints. Structural models were validated using Molprobity (99).

Data Availability. All study data are included in the article and supporting information.

ACKNOWLEDGMENTS. This work was funded by NIH R01GM048123 and R35GM131901 and NSF MCB1517625 and MCB2016540 (to J.F.). M.G.-J. was supported in part by an Institute for Quantitative and Computational Biosciences Collaboratory Postdoctoral fellowship.

1. Y. Doksani, T. de Lange, The role of double-strand break repair pathways at functional and dysfunctional telomeres. *Cold Spring Harb. Perspect. Biol.* **6**, a016576 (2014).
2. M. Armanios, E. H. Blackburn, The telomere syndromes. *Nat. Rev. Genet.* **13**, 693–704 (2012).
3. J. W. Shay, Role of telomeres and telomerase in aging and cancer. *Cancer Discov.* **6**, 584–593 (2016).
4. S. E. Artandi, R. A. DePinho, Telomeres and telomerase in cancer. *Carcinogenesis* **31**, 9–18 (2010).

5. B. Bernardes de Jesus, M. A. Blasco, Telomerase at the intersection of cancer and aging. *Trends Genet.* **29**, 513–520 (2013).
6. P. F. Fogarty *et al*., Late presentation of dyskeratosis congenita as apparently acquired aplastic anaemia due to mutations in telomerase RNA. *Lancet* **362**, 1628–1630 (2003).
7. G. Sarek, P. Marzec, P. Margalef, S. J. Boulton, Molecular basis of telomere dysfunction in human genetic diseases. *Nat. Struct. Mol. Biol.* **22**, 867–874 (2015).
8. S. A. Savage, A. A. Bertuch, The genetics and clinical manifestations of telomere biology disorders. *Genet. Med.* **12**, 753–764 (2010).
9. D. M. Townsley, B. Dumitriu, N. S. Young, Bone marrow failure and the telomeropathies. *Blood* **124**, 2775–2783 (2014).
10. H. Chan, Y. Wang, J. Feigon, Progress in human and Tetrahymena telomerase structure determination. *Annu. Rev. Biophys.* **46**, 199–225 (2017).
11. J. D. Podlevsky, J. J. Chen, Evolutionary perspectives of telomerase RNA structure and function. *RNA Biol.* **13**, 720–732 (2016).
12. J. D. Podlevsky, C. J. Bley, R. V. Omana, X. Qi, J. J. Chen, The telomerase database. *Nucleic Acids Res.* **36**, D339–D343 (2008).
13. J. D. Podlevsky, Y. Li, J. J. Chen, The functional requirement of two structural domains within telomerase RNA emerged early in eukaryotes. *Nucleic Acids Res.* **44**, 9891–9901 (2016).
14. A. J. Gillis, A. P. Schuller, E. Skordalakes, Structure of the Tribolium castaneum telomerase catalytic subunit TERT. *Nature* **455**, 633–637 (2008).
15. M. Mitchell, A. Gillis, M. Futahashi, H. Fujiwara, E. Skordalakes, Structural basis for telomerase catalytic subunit TERT binding to RNA template and telomeric DNA. *Nat. Struct. Mol. Biol.* **17**, 513–518 (2010).
16. A. R. Robart, K. Collins, Human telomerase domain interactions capture DNA for TEN domain-dependent processive elongation. *Mol. Cell* **42**, 308–318 (2011).
17. C. J. Bley *et al*., RNA-protein binding interface in the telomerase ribonucleoprotein. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 20333–20338 (2011).
18. J. W. Parks, M. D. Stone, Coordinated DNA dynamics during the human telomerase catalytic cycle. *Nat. Commun.* **5**, 4146 (2014).
19. L. I. Jansson *et al*., Structural basis of template-boundary definition in Tetrahymena telomerase. *Nat. Struct. Mol. Biol.* **22**, 883–888 (2015).
20. A. J. Berman, A. R. Gooding, T. R. Cech, Tetrahymena telomerase protein p65 induces conformational changes throughout telomerase RNA (TER) and rescues telomerase reverse transcriptase and TER assembly mutants. *Mol. Cell. Biol.* **30**, 4965–4976 (2010).
21. A. Sauerwald *et al*., Structure of active dimeric human telomerase. *Nat. Struct. Mol. Biol.* **20**, 454–460 (2013).
22. J. Jiang *et al*., Structure of Tetrahymena telomerase reveals previously unknown subunits, functions, and interactions. *Science* **350**, aab4070 (2015).
23. J. Jiang *et al*., The architecture of Tetrahymena telomerase holoenzyme. *Nature* **496**, 187–192 (2013).
24. T. H. D. Nguyen *et al*., Cryo-EM structure of substrate-bound human telomerase holoenzyme. *Nature* **557**, 190–195 (2018).
25. J. Jiang *et al*., Structure of telomerase with telomeric DNA. *Cell* **173**, 1179–1190.e13 (2018).
26. M. Osanai, K. K. Kojima, R. Futahashi, S. Yaguchi, H. Fujiwara, Identification and characterization of the telomerase reverse transcriptase of Bombyx mori (silkworm) and Tribolium castaneum (flour beetle). *Gene* **376**, 281–289 (2006).
27. B. M. Akiyama, J. W. Parks, M. D. Stone, The telomerase essential N-terminal domain promotes DNA synthesis by stabilizing short RNA-DNA hybrids. *Nucleic Acids Res.* **43**, 5537–5549 (2015).
28. R. A. Wu, K. Collins, Human telomerase specialization for repeat synthesis by unique handling of primer-template duplex. *EMBO J.* **33**, 921–935 (2014).
29. S. Shastry, O. Steinberg-Neifach, N. Lue, M. D. Stone, Direct observation of nucleic acid binding dynamics by the telomerase essential N-terminal domain. *Nucleic Acids Res.* **46**, 3088–3102 (2018).
30. S. Richards *et al*.; Tribolium Genome Sequencing Consortium, The genome of the model beetle and pest Tribolium castaneum. *Nature* **452**, 949–955 (2008).
31. H. Hoffman, C. Rice, E. Skordalakes, Structural analysis reveals the deleterious effects of telomerase mutations in bone marrow failure syndromes. *J. Biol. Chem.* **292**, 4593–4601 (2017).
32. Y. Wang, J. D. Yesselman, Q. Zhang, M. Kang, J. Feigon, Structural conservation in the template/pseudoknot domain of vertebrate telomerase RNA from teleost fish to human. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E5125–E5134 (2016).
33. Q. Zhang, N. K. Kim, R. D. Peterson, Z. Wang, J. Feigon, Structurally conserved five nucleotide bulge determines the overall topology of the core domain of human telomerase RNA. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 18761–18768 (2010).
34. N. K. Kim, Q. Zhang, J. Feigon, Structure and sequence elements of the CR4/5 domain of medaka telomerase RNA important for telomerase function. *Nucleic Acids Res.* **42**, 3395–3408 (2014).
35. M. Xie, J. D. Podlevsky, X. Qi, C. J. Bley, J. J. Chen, A novel motif in telomerase reverse transcriptase regulates telomere repeat addition rate and processivity. *Nucleic Acids Res.* **38**, 1982–1996 (2010).
36. N. F. Lue, Y. C. Lin, I. S. Mian, A conserved telomerase motif within the catalytic domain of telomerase reverse transcriptase is specifically required for repeat addition processivity. *Mol. Cell. Biol.* **23**, 8440–8449 (2003).
37. J. Lingner *et al*., Reverse transcriptase motifs in the catalytic subunit of telomerase. *Science* **276**, 561–567 (1997).
38. C. Zhao, A. M. Pyle, Crystal structures of a group II intron maturase reveal a missing link in spliceosome evolution. *Nat. Struct. Mol. Biol.* **23**, 558–565 (2016).
39. G. Qu *et al*., Structure of a group II intron in complex with its reverse transcriptase. *Nat. Struct. Mol. Biol.* **23**, 549–557 (2016).
40. E. A. Gladyshev, I. R. Arkhipova, Telomere-associated endonuclease-deficient Penelope-like retroelements in diverse eukaryotes. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 9352–9357 (2007).
41. M. Mason, A. Schuller, E. Skordalakes, Telomerase structure function. *Curr. Opin. Struct. Biol.* **21**, 92–100 (2011).
42. J. Xia, Y. Peng, I. S. Mian, N. F. Lue, Identification of functionally important domains in the N-terminal region of telomerase reverse transcriptase. *Mol. Cell. Biol.* **20**, 5196–5207 (2000).
43. E. Casacuberta, Drosophila: Retrotransposons making up telomeres. *Viruses* **9**, 192 (2017).
44. A. G. Lai *et al*., The protein subunit of telomerase displays patterns of dynamic evolution and conservation across different metazoan taxa. *BMC Evol. Biol.* **17**, 107 (2017).
45. C. Rosensweig *et al*., An evolutionary hotspot defines functional differences between CRYPTOCHROMES. *Nat. Commun.* **9**, 1138 (2018).
46. D. Pincus *et al*., Engineering allosteric regulation in protein kinases. *Sci. Signal.* **11**, eaar3250 (2018).
47. N. Halabi, O. Rivoire, S. Leibler, R. Ranganathan, Protein sectors: Evolutionary units of three-dimensional structure. *Cell* **138**, 774–786 (2009).
48. C. Notredame, D. G. Higgins, J. Heringa, T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205–217 (2000).
49. J. Huang *et al*., Structural basis for protein-RNA recognition in telomerase. *Nat. Struct. Mol. Biol.* **21**, 507–512 (2014).
50. M. Harkisheimer, M. Mason, E. Shuvaeva, E. Skordalakes, A motif in the vertebrate telomerase N-terminal linker of TERT contributes to RNA binding and telomerase activity and processivity. *Structure* **21**, 1870–1878 (2013).
51. O. A. Petrova *et al*., Structure and function of the N-terminal domain of the yeast telomerase reverse transcriptase. *Nucleic Acids Res.* **46**, 1525–1540 (2018).
52. S. A. Jacobs, E. R. Podell, T. R. Cech, Crystal structure of the essential N-terminal domain of telomerase reverse transcriptase. *Nat. Struct. Mol. Biol.* **13**, 218–225 (2006).
53. L. C. Collopy *et al*., Triallelic and epigenetic-like inheritance in human disorders of telomerase. *Blood* **126**, 176–184 (2015).
54. J. Dai *et al*., Telomerase gene mutations and telomere length shortening in patients with idiopathic pulmonary fibrosis in a Chinese population. *Respirology* **20**, 122–128 (2015).
55. A. Marrone *et al*., Telomerase reverse-transcriptase homozygous mutations in autosomal recessive dyskeratosis congenita and Hoyeraal-Hreidarsson syndrome. *Blood* **110**, 4198–4205 (2007).
56. J. K. Alder *et al*., Diagnostic utility of telomere length testing in a hospital-based setting. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E2358–E2365 (2018).
57. S. Petrovski *et al*., An exome sequencing study to assess the role of rare genetic variation in pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* **196**, 82–93 (2017).
58. K. D. Tsakiri *et al*., Adult-onset pulmonary fibrosis caused by mutations in telomerase. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 7552–7557 (2007).
59. J. Liang *et al*., Mutations in telomerase catalytic protein in Japanese children with aplastic anemia. *Haematologica* **91**, 656–658 (2006).
60. H. Yamaguchi *et al*., Mutations in TERT, the gene for telomerase reverse transcriptase, in aplastic anemia. *N. Engl. J. Med.* **352**, 1413–1424 (2005).
61. H. Holme *et al*., Marked genetic heterogeneity in familial myelodysplasia/acute myeloid leukaemia. *Br. J. Haematol.* **158**, 242–248 (2012).
62. T. W. Chu, Y. D'Souza, C. Autexier, The insertion in fingers domain in human telomerase can mediate enzyme processivity and telomerase recruitment to telomeres in a TPP1-dependent manner. *Mol. Cell. Biol.* **36**, 210–222 (2015).
63. F. Wang *et al*., The POT1-TPP1 telomere complex is a telomerase processivity factor. *Nature* **445**, 506–510 (2007).
64. T. de Lange, Shelterin: The protein complex that shapes and safeguards human telomeres. *Genes Dev.* **19**, 2100–2110 (2005).
65. J. Nandakumar *et al*., The TEL patch of telomere protein TPP1 mediates telomerase recruitment and processivity. *Nature* **492**, 285–289 (2012).
66. S. Grill, V. M. Tesmer, J. Nandakumar, The N terminus of the OB domain of telomere protein TPP1 is critical for telomerase action. *Cell Rep.* **22**, 1132–1140 (2018).
67. V. M. Tesmer, E. M. Smith, O. Danciu, S. Padmanaban, J. Nandakumar, Combining conservation and species-specific differences to determine how human telomerase binds telomeres. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 26505–26515 (2019).
68. A. J. Zaug, E. R. Podell, J. Nandakumar, T. R. Cech, Functional interaction between telomere protein TPP1 and telomerase. *Genes Dev.* **24**, 613–622 (2010).
69. A. N. Sexton, D. T. Youmans, K. Collins, Specificity requirements for human telomere protein interaction with telomerase holoenzyme. *J. Biol. Chem.* **287**, 34455–34464 (2012).
70. F. L. Zhong *et al*., TPP1 OB-fold domain controls telomere maintenance by recruiting telomerase to chromosome ends. *Cell* **150**, 481–494 (2012).
71. J. C. Schmidt, A. B. Dalby, T. R. Cech, Identification of human TERT elements necessary for telomerase recruitment to telomeres. *eLife* **3**, e03563 (2014).
72. R. A. Wu, H. E. Upton, J. M. Vogan, K. Collins, Telomerase mechanism of telomere synthesis. *Annu. Rev. Biochem.* **86**, 439–460 (2017).
73. H. E. Upton, H. Chan, J. Feigon, K. Collins, Shared subunits of Tetrahymena telomerase holoenzyme and replication protein A have different functions in different cellular complexes. *J. Biol. Chem.* **292**, 217–228 (2017).
74. K. Hong *et al*., Tetrahymena telomerase holoenzyme assembly, activation, and inhibition by domains of the p50 central hub. *Mol. Cell. Biol.* **33**, 3962–3971 (2013).
75. J. R. Mitchell, K. Collins, Human telomerase activation requires two independent interactions between telomerase RNA and telomerase reverse transcriptase. *Mol. Cell* **6**, 361–371 (2000).
76. N. K. Kim, C. A. Theimer, J. R. Mitchell, K. Collins, J. Feigon, Effect of pseudouridylation on the structure and activity of the catalytically essential P6.1 hairpin in human telomerase RNA. *Nucleic Acids Res.* **38**, 6746–6756 (2010).
77. H. Ly, E. H. Blackburn, T. G. Parslow, Comprehensive structure-function analysis of the core domain of human telomerase RNA. *Mol. Cell. Biol.* **23**, 6849–6856 (2003).

78. J. L. Chen, K. K. Opperman, C. W. Greider, A critical stem-loop structure in the CR4-CR5 domain of mammalian telomerase RNA. *Nucleic Acids Res.* **30**, 592–597 (2002).
79. A. R. Robart, K. Collins, Investigation of human telomerase holoenzyme assembly, activity, and processivity using disease-linked subunit variants. *J. Biol. Chem.* **285**, 4375–4386 (2010).
80. M. Zuker, Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415 (2003).
81. C. Palka, N. M. Forino, J. Hentschel, R. Das, M. D. Stone, Folding heterogeneity in the essential human telomerase RNA three-way junction. bioRxiv:10.1101/2019.12.14.876565 (15 December 2019).
82. J. L. Chen, M. A. Blasco, C. W. Greider, Secondary structure of vertebrate telomerase RNA. *Cell* **100**, 503–514 (2000).
83. J. L. Chen, C. W. Greider, Template boundary definition in mammalian telomerase. *Genes Dev.* **17**, 2747–2752 (2003).
84. A. J. Berman, B. M. Akiyama, M. D. Stone, T. R. Cech, The RNA accordion model for template positioning by telomerase RNA during telomeric DNA synthesis. *Nat. Struct. Mol. Biol.* **18**, 1371–1375 (2011).
85. C. W. Greider, E. H. Blackburn, The telomere terminal transferase of Tetrahymena is a ribonucleoprotein enzyme with two kinds of primer specificity. *Cell* **51**, 887–898 (1987).
86. W. C. Drosopoulos, V. R. Prasad, The telomerase-specific T motif is a restrictive determinant of repetitive reverse transcription by human telomerase. *Mol. Cell. Biol.* **30**, 447–459 (2010).
87. D. Bosoy, Y. Peng, I. S. Mian, N. F. Lue, Conserved N-terminal motifs of telomerase reverse transcriptase required for ribonucleoprotein assembly in vivo. *J. Biol. Chem.* **278**, 3882–3890 (2003).
88. B. M. Akiyama, A. Gomez, M. D. Stone, A conserved motif in Tetrahymena thermophila telomerase reverse transcriptase is proximal to the RNA template and is essential for boundary definition. *J. Biol. Chem.* **288**, 22141–22149 (2013).
89. T. W. Chu, D. E. MacNeil, C. Autexier, Multiple mechanisms contribute to the cell growth defects imparted by human telomerase insertion in fingers domain mutations associated with premature aging diseases. *J. Biol. Chem.* **291**, 8374–8386 (2016).
90. D. T. Jones, Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**, 195–202 (1999).
91. J. Pei, N. V. Grishin, PROMALS: Towards accurate multiple sequence alignments of distantly related proteins. *Bioinformatics* **23**, 802–808 (2007).
92. I. M. Wallace, O. O'Sullivan, D. G. Higgins, C. Notredame, M-Coffee: Combining multiple sequence alignment methods with T-Coffee. *Nucleic Acids Res.* **34**, 1692–1699 (2006).
93. F. Madeira *et al*., The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **47**, W636–W641 (2019).
94. O. Rivoire, K. A. Reynolds, R. Ranganathan, Evolution-based functional decomposition of proteins. *PLoS Comput. Biol.* **12**, e1004817 (2016).
95. P. Emsley, K. Cowtan, Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
96. E. F. Pettersen *et al*., UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
97. M. Popenda *et al*., Automated 3D structure composition for large RNAs. *Nucleic Acids Res.* **40**, e112 (2012).
98. P. D. Adams *et al*., PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
99. C. J. Williams *et al*., MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).