PNAS RESEARCH ARTICLE | BIOCHEMISTRY OPEN ACCESS

Shelterin is a dimeric complex with extensive structural heterogeneity

John C. Zinder[a], Paul Dominic B. Olinares[b], Vladimir Svetlov[c], Martin W. Bush[d], Evgeny Nudler[c,e], Brian T. Chait[b], Thomas Walz[d,1], and Titia de Lange[a,1]

Edited by Cynthia Wolberger, Johns Hopkins University School of Medicine, Baltimore, MD; received January 28, 2022; accepted June 24, 2022

Human shelterin is a six-subunit complex—composed of TRF1, TRF2, Rap1, TIN2, TPP1, and POT1—that binds telomeres, protects them from the DNA-damage response, and regulates the maintenance of telomeric DNA. Although high-resolution structures have been generated of the individual structured domains within shelterin, the architecture and stoichiometry of the full complex are currently unknown. Here, we report the purification of shelterin subcomplexes and reconstitution of the entire complex using full-length, recombinant subunits. By combining negative-stain electron microscopy (EM), cross-linking mass spectrometry (XLMS), AlphaFold modeling, mass photometry, and native mass spectrometry (MS), we obtain stoichiometries as well as domain-scale architectures of shelterin subcomplexes and determine that they feature extensive conformational heterogeneity. For POT1/TPP1 and POT1/TPP1/TIN2, we observe high variability in the positioning of the POT1 DNA-binding domain, the TPP1 oligonucleotide/oligosaccharide–binding (OB) fold, and the TIN2 TRFH domain with respect to the C-terminal domains of POT1. Truncation of unstructured linker regions in TIN2, TPP1, and POT1 did not reduce the conformational variability of the heterotrimer. Shelterin and TRF1-containing subcomplexes form fully dimeric stoichiometries, even in the absence of DNA substrates. Shelterin and its subcomplexes showed extensive conformational variability, regardless of the presence of DNA substrates. We conclude that shelterin adopts a multitude of conformations and argue that its unusual architectural variability is beneficial for its many functions at telomeres.

telomeres | shelterin | electron microscopy | conformational heterogeneity | AlphaFold

Significance

Shelterin is a multiprotein complex that plays central roles in telomere biology. Mutations in shelterin result in premature aging diseases and familial cancer predisposition. Mechanistic understanding of these so-called telomereopathies is hampered by our lack of knowledge regarding the structure and stoichiometry of shelterin. Here, we use multiple methods to probe the stoichiometry and conformational states of shelterin and reveal that it forms a fully dimeric complex with extensive conformational heterogeneity. Our results highlight the dynamic nature of this essential complex and explain why its high-resolution structure determination has yet to be achieved.

Telomeres are protective assemblies of DNA and proteins at the ends of linear chromosomes. Human telomeres are composed of the hexanucleotide 5′-TTAGGG-3′ sequence repeated for 3 to 10 kb and terminate in an ∼100-nt 3′ single-stranded (ss) overhang (1). Telomeres need to solve two problems: 1) The DNA-replication machinery is unable to replicate all the way to the end, resulting in terminal attrition as cells divide ("the end-replication problem"), and 2) telomere ends resemble sites of DNA damage that must be protected against erroneous repair and inappropriate activation of DNA-damage checkpoints ("the end-protection problem"). The telomerase ribonucleoprotein extends telomeres, using its RNA component as a template for its reverse-transcriptase activity, to solve the end-replication problem (2, 3). Shelterin, a multisubunit DNA-binding complex (4), solves the end-protection problem and additionally recruits telomerase to solve the end-replication problem (5–7). Defects in shelterin components are associated with diverse pathologies including familial cancer predisposition and premature aging, underscoring its importance in cellular physiology and development (8–13).

Shelterin specifically binds to double-stranded (ds) telomeric repeat sequences via the C-terminal Myb domains of its TRF1 and TRF2 subunits (14–17). TRF1 and TRF2 homodimerize via their TRF homology (TRFH) domains (18, 19) and are bridged by TIN2 (20–22), which contains a variant of a TRFH domain that does not dimerize (23). In somatic cells, TRF2 is critical for the formation of t-loops (24–26), protective structures in which the telomeric 3′ overhang invades into the dsDNA repeat region, hiding the telomere end from recognition by the ATM kinase DNA-damage response (DDR) pathway and double-strand break repair by classical nonhomologous end joining (1). TIN2 is recruited to telomeres primarily via its interaction with TRF1, which is stronger than its interaction with TRF2 (20, 23, 27–29). TIN2 additionally interacts with TPP1 (30–32), which in turn binds POT1, an ssDNA-binding protein that prevents replication protein A from accumulating at telomeric ssDNA and activating the ATR kinase DDR pathway (33–38). Finally, Rap1 binds tightly to TRF2, but is not required for TRF2's primary functions at telomeres (26, 39, 40).

Shelterin acts in conjunction with several accessory factors such as the BLM and RTEL1 helicases, which facilitate the replication of telomeric DNA (41–44). After replication,

Author affiliations: [a]Laboratory of Cell Biology and Genetics, The Rockefeller University, New York, NY 10065; [b]Laboratory of Mass Spectrometry and Gaseous Ion Chemistry, The Rockefeller University, New York, NY 10065; [c]Department of Biochemistry and Molecular Pharmacology, New York University Grossman School of Medicine, New York, NY 10016; [d]Laboratory of Molecular Electron Microscopy, The Rockefeller University, New York, NY 10065; and [e]HHMI, New York University Grossman School of MedicineNew York, NY 10016

Author contributions: J.C.Z., P.D.B.O., V.S., M.W.B., E.N., B.T.C., T.W., and T.d.L. designed research; J.C.Z., P.D.B.O., and V.S. performed research; J.C.Z., P.D.B.O., V.S., E.N., B.T.C., T.W., and T.d.L. analyzed data; and J.C.Z., T.W., and T.d.L. wrote the paper.

Competing interest statement: T.d.L. is a member of the scientific advisory board of Calico Life Sciences.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: twalz@rockefeller.edu or delange@rockefeller.edu.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2201662119/-/DCSupplemental.

Published July 26, 2022.

shelterin ensures the reconstitution of the correct structure of the telomere end by recruiting the Apollo nuclease (via TRF2) (45–48) and the CST fill-in machinery (via TPP1/POT1) to telomeres (49, 50). Finally, shelterin regulates telomerase-mediated telomere extension by recruiting the enzyme via TPP1 (5, 6) and limiting its activity at the longest telomeres (via TPP1/POT1–dependent recruitment of CST) (51).

All six shelterin components contain structured domains separated by linker regions that are predicted to be largely disordered based on their low sequence conservation and their predicted low propensity to form secondary structure (1, 23, 52). These regions range in length from 25 to 150 amino acids (aa) in humans and, while their sequences and lengths are highly variable in other organisms, the trend of alternating structured domains and linker regions is maintained in shelterin components throughout metazoan evolution (53). X-ray crystallography and NMR spectroscopy have generated molecular models of all known structured domains and minimal interacting motifs within human shelterin (19, 23, 29, 54–60), but there are currently no high-resolution structural models for full-length shelterin proteins. It has been speculated that shelterin is too flexible to be structurally characterized at high resolution (52), although this has yet to be confirmed experimentally.

Purification of mouse shelterin overexpressed in HEK293 cells has been reported, but that study did not yield sufficient quantities to probe its structure or stoichiometry (61). In another study, a human shelterin subcomplex lacking TRF1 was purified from insect cells and a POT1:TPP1:TIN2:TRF2:Rap1 stoichiometry of 1:1:1:2:2 was proposed based on in-gel densitometry measurements of labeled components and size-exclusion chromatography coupled to multiangle light scattering, but no structural characterization of this complex was reported (62). Recently, POT1/TPP1/TIN2 was characterized by cryo-electron microscopy (cryo-EM) at low resolution in isolation (63) and at intermediate resolution in a cross-linked complex with human telomerase (64); however, neither of these studies was able to visualize all structured elements within this shelterin subcomplex.

Here, we report single-particle negative-stain EM characterization of purified, recombinantly produced human shelterin subcomplexes containing full-length components and the reconstituted full, six-subunit complex. We employ streptavidin-DNA labeling and cross-linking mass spectrometry (XLMS) to identify elements within reference-free two-dimensional (2D) class averages of shelterin subcomplexes and arrive at domain-scale models of their structures in different conformations. Modeling of these subcomplexes using AlphaFold further supported that they form highly flexible assemblies. Finally, we use mass photometry and native mass spectrometry (MS) to reveal that either one or two TIN2 molecules can bind the TRF1 homodimer in reconstituted complexes, arguing that shelterin can exist as a fully dimeric complex. Mass photometry of reconstituted six-subunit shelterin showed that it could form a fully dimeric stoichiometry as well. The data reveal an extraordinary level of structural variability for shelterin and several of its subcomplexes. We discuss the implications of this variability in the context of the many different functions of shelterin.

Results

High Variability in the Relative Positioning of the DNA-Binding and C-Terminal Domains of POT1. We first sought to structurally characterize a small, well-defined subcomplex of shelterin containing full-length POT1 and the N-terminal domains of TPP1 (POT1/TPP1^{N}; Fig. 1*A*). Recombinantly expressed POT1/TPP1^{N} could be purified to homogeneity as assessed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) (Fig. 1*B*) and bound to telomeric ssDNA with high specificity (Fig. 1*C*).

Negative-stain EM images of POT1/TPP1^{N} bound to ssDNA showed particles of similar size but with highly variable shapes (*SI Appendix*, Fig. 1*A*). Approximately 60,000 particles were autopicked from 100 micrographs and, after centering and curation to 40,000 particles, subjected to 2D classification using the "iterative stable alignment and clustering" (ISAC) algorithm, an approach optimized for analyzing heterogeneous samples (65). Details for this and other negative-stain EM datasets are provided in *SI Appendix*, Table 1.

Many of the reference-free 2D class averages of POT1/TPP1^{N} showed an oblong lobe connected on one side to another oblong, often bipartite, lobe of similar size, and on the other side to a third, smaller and more circular lobe (Fig. 1*D* and *SI Appendix*, Fig. 1*A*). The conformations of the two oblong lobes relative to each other ranged from almost fully extended to entirely closed. This finding suggests that the 33-aa linker connecting the N- and C-terminal domains of POT1 (Fig. 1*A*) allows the two domains to freely move with respect to each other. The smaller, more circular lobe is likely the OB fold of TPP1, which is present in many positions relative to the POT1 C-terminal domain (comprising its split OB3 and HJRL domains; Fig. 1*D*, cartoons and *SI Appendix*, Fig. 2*A*). Notably, this lobe is absent in many class averages, likely because it is connected to the RD domain by a flexible linker of 25 aa (Fig. 1*A*) and may adopt many different positions relative to POT1, causing it to become averaged out in these classes (*SI Appendix*, Fig. 1*A*).

To further resolve elements within 2D class averages of POT1/TPP1^{N}, we incubated it with a synthetic 5′-biotin ssDNA and streptavidin to add a feature of distinct size and shape to the complex which should lie close to OB1 of POT1 (58). In size-exclusion chromatography, the streptavidin-ssDNA-POT1/TPP1^{N} complex eluted as three main peaks (*SI Appendix*, Fig. 2*B*): a small first peak consistent with a complex containing two POT1/TPP1^{N} complexes bound to a single streptavidin tetramer, followed by a second, larger peak consistent with a complex containing one POT1/TPP1^{N} and one streptavidin tetramer, and finally another large peak containing excess streptavidin-ssDNA (*SI Appendix*, Fig. 2*C* and *D*). Negative-stain EM analysis of a middle-peak fraction revealed class averages containing clear signal for a single streptavidin tetramer adjacent to an oblong bipartite lobe, from which another oblong lobe extends in many different directions (Fig. 1*E* and *SI Appendix*, Fig. 1*B*). Based on the position of the biotin and known crystal structures, we interpret the bipartite oblong lobe to represent OB1 and OB2 of POT1 and the larger lobe to be its C-terminal domain bound to TPP1 (Fig. 1*E*, cartoons and *SI Appendix*, Fig. 2*A*). The OB fold of TPP1 is not visible in these averages, likely because the streptavidin-POT1 OB1/OB2 features dominate the alignment and the TPP1 OB fold is thus averaged out. The 2D class averages show complexes in a near continuum of conformations from completely closed to fully extended as well as a conformation in which the C-terminal domain is bent backward (Fig. 1*E* and *SI Appendix*, Fig. 1*B*). These data are consistent with a high degree of flexibility of the hinge region between the N- and C-terminal domains of POT1, resulting in high conformational variability.

Variable Association of the TIN2 TRFH Domain with POT1/TPP1. We performed similar analyses on POT1/TPP1/TIN2 (Fig. 2*A*), hoping that we could exploit our domain assignments of POT1/TPP1^{N} to localize elements within the larger complex. Because removal of the green fluorescent protein

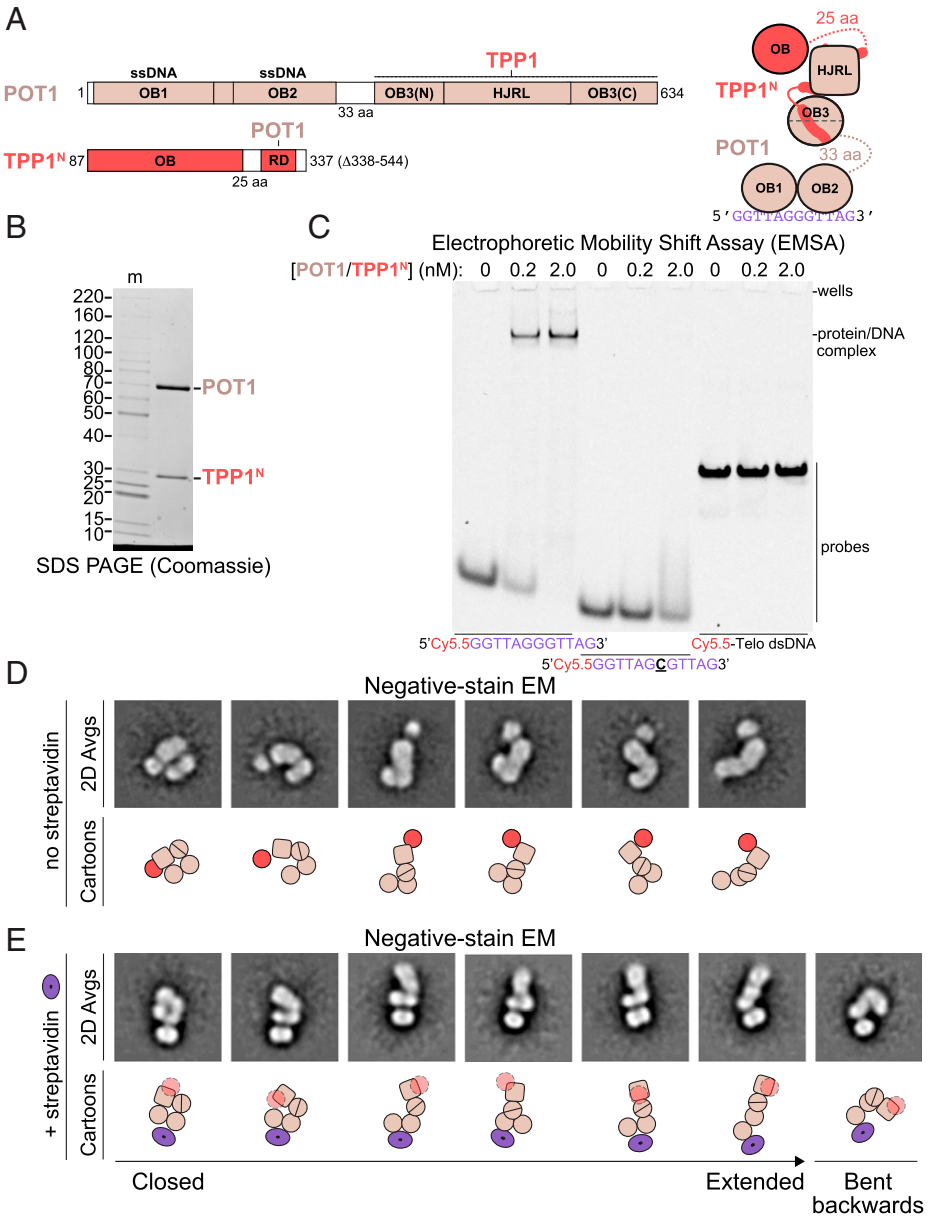

Fig. 1. Structural variability and DNA binding of POT1/TPP1^N. (*A*) Domain and cartoon schematics for POT1 and TPP1^N with starting and ending amino acids and flexible linker lengths indicated. Regions that associate with other proteins in shelterin are indicated. HJRL, Holliday junction resolvase-like domain; OB, oligonucleotide/oligosaccharide–binding fold; RD, recruitment domain. Regions of known structure are depicted in color in the domain schematics and as shapes in the cartoon. Regions of unknown structure are depicted in white in the domain schematic and as dashed lines with amino acid lengths indicated in the cartoon. (*B*) SDS-PAGE of purified POT1/TPP1^N. The image shows an 8 to 16% polyacrylamide Tris-glycine gel stained with Coomassie blue. Molecular mass markers (m) are given in kDa. (*C*) DNA-binding assay for POT1/TPP1^N with fluorescent DNA probes. Protein concentrations and DNAs are indicated. Probe concentration is 0.25 nM. The gel was imaged for Cy5.5 fluorescence. (*D*) Negative-stain EM analysis of POT1/TPP1^N bound to 5′-biotin-$(GGTTAG)_2$ DNA. Reference-free 2D class averages showing three lobes in a range of conformations were selected for display with cartoon interpretations (*Bottom*). Domains are as in *A*. (*E*) Negative-stain EM analysis of POT1/TPP1^N bound to streptavidin-5′-biotin-$(GGTTAG)_2$ DNA. Reference-free 2D class averages showing a bipartite lobe associated with streptavidin and a second oblong lobe in a range of conformations were selected for display with cartoon interpretations (*Bottom*). Domains in the cartoons are as in *A* with TPP1's OB fold depicted as transparent because its position is not discernable. The streptavidin tetramer is shown as a purple torus.

(GFP) tag on TIN2 resulted in a complex with poor solubility, we purified the POT1/TPP1/GFP-TIN2 complex (Fig. 2*B* and *C*), which specifically bound telomeric ssDNA (*SI Appendix*, Fig. 3*A*). As with POT1/TPP1^N, particles from this complex in negative-stain EM images appeared homogeneous in size and heterogeneous in shape (*SI Appendix*, Fig. 3*B*). Since the ordered domains within the constituent proteins and the GFP tag are all similar in size and shape, we were unable to unambiguously assign identities to the features in the class averages from this complex (*SI Appendix*, Fig. 3*B*). To overcome this problem, we used the streptavidin-DNA labeling strategy as with POT1/TPP1^N (Fig. 2*D*, *Top* and *SI Appendix*, Fig. 4*A*). As before, 2D class averages showed streptavidin positioned near an oblong bipartite lobe, which we interpret as OB1 and OB2 of POT1. Certain classes were clearly relatable to 2D class averages of POT1/TPP1^N, with several resembling closed states of POT1/TPP1^N and others resembling more opened conformations (compare Fig. 2*D*, *Top* with Fig. 1*E*). In all cases, we observe additional signal, which could either represent GFP or the TRFH domain of TIN2.

We asked whether the conformational heterogeneity of the POT1/TPP1/TIN2 complex was due to the flexible linkers acting

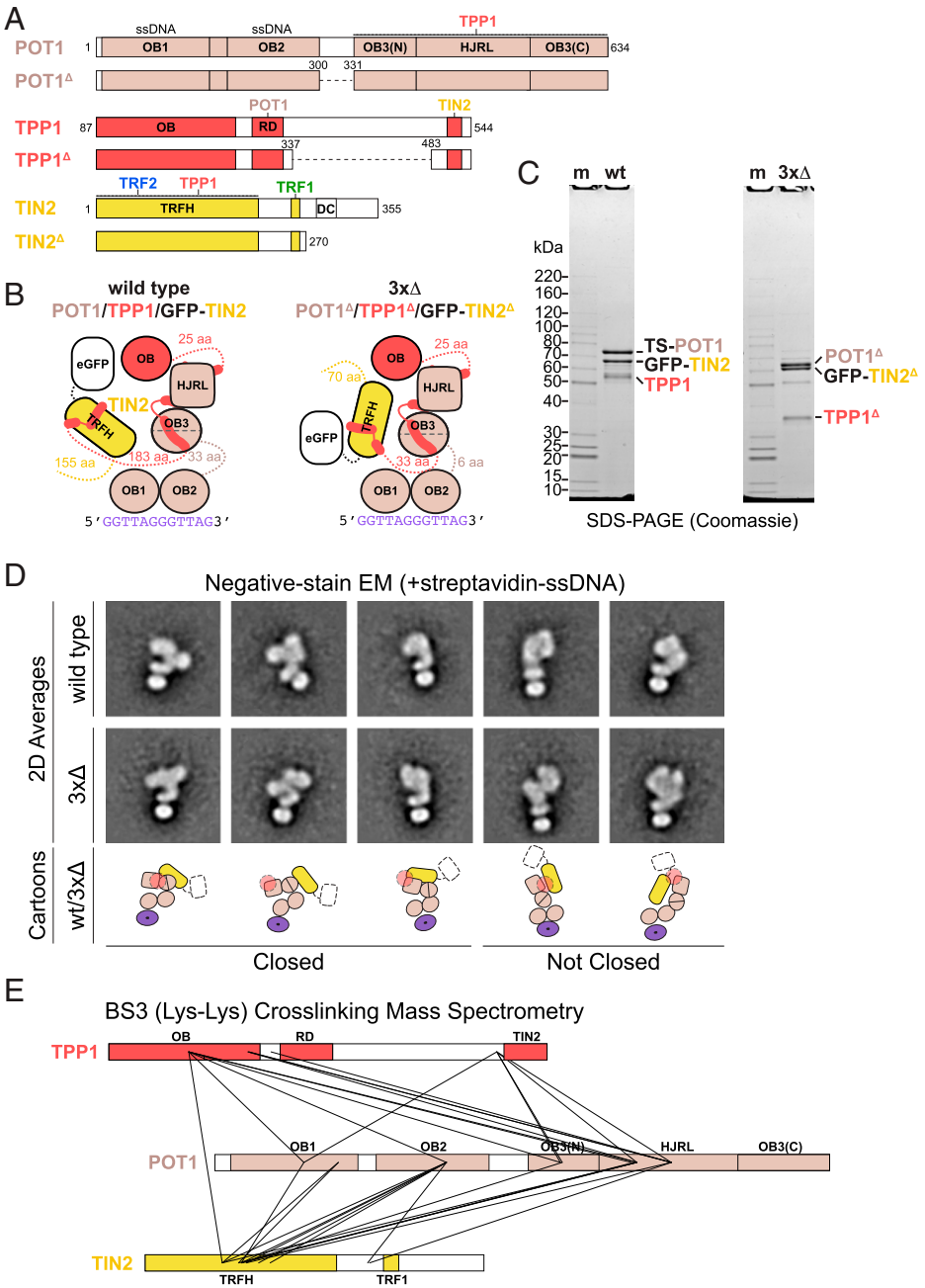

Fig. 2. Structural variability of POT1/TPP1/GFP-TIN2. (*A*) Domain schematics for POT1, TPP1, TIN2, and deletion constructs with amino acids are indicated. Coloring is as in Fig. 1*A*. In POT1 and TPP1 deletion constructs, dashed lines represent regions that have been replaced with $(Gly\text{-}Gly\text{-}Ser)_2$. (*B*) Cartoon schematics for POT1/TPP1/GFP-TIN2 wild type (wt) and deletion construct (3×Δ) bound to telomeric ssDNA. (*C*) SDS-PAGE of purified wt and 3×Δ POT1/TPP1/GFP-TIN2. TS, Twin-Strep. Protein gel as in Fig. 1*B*. (*D*) Negative-stain EM analysis of wt and 3×Δ POT1/TPP1/GFP-TIN2 bound to streptavidin-5′-biotin-$(GGTTAG)_2$ DNA. Reference-free 2D class averages that were directly comparable to those for POT1/TPP1^N were selected for display with cartoon interpretations (*Bottom*). (*E*) Cross-linking MS of wt POT1/TPP1/GFP-TIN2 using BS3. Domains are shown and colored as in *A* and cross-links are shown as solid lines. Intramolecular cross-links are not shown.

as hinges in the three proteins. We identified regions within POT1/TPP1/TIN2 (POT1 residues 301 to 329, TPP1 residues 338 to 482, and TIN2 residues 271 to 355) that were poorly conserved, predicted to have low propensity to form secondary structure, susceptible to limited proteolysis (as determined by other groups), and/or were modeled with low confidence by AlphaFold2 (53, 55, 62, 66). A complex with these regions removed (3×Δ; Fig. 2 *A*–*C*) was purified to apparent homogeneity (Fig. 2*C*) and appeared monodisperse in negative-stain EM images (*SI Appendix*, Fig. 4*B*). Single-particle analysis using streptavidin-DNA labeling revealed that this complex also exhibits considerable structural variability (*SI Appendix*, Fig. 4*B*), and that many reference-free 2D class averages closely resemble those of the wild-type complex (Fig. 2*D*, compare *Top* and *Middle*). We conclude that POT1/TPP1/TIN2 remains conformationally variable in the absence of much of the linker/hinge sequences.

We then subjected the POT1/TPP1/GFP-TIN2 complex to XLMS using the amine-reactive molecule bis(sulfosuccinimidyl)suberate (BS3). We observed many intra- and intermolecular cross-links in the complex, including between TPP1 OB and POT1 HJRL as well as between TIN2 TRFH and the

other components (Fig. 2*E* and *SI Appendix*, Tables 2 and 3), indicating that these domains are in close proximity to one another. Importantly, residues in TPP1 or TIN2 that are far apart in crystal structures cross-linked to the same or similar positions on POT1, indicating multiple conformations are present within the sample. For example, lysines 170 and 232 of TPP1 both cross-link to lysine 433 of POT1 (Fig. 2*E* and *SI Appendix*, Table 3) despite being on opposite sides of the TPP1 OB fold (57). As expected, we do not observe any cross-links between solvent-accessible lysines on the GFP or twin-Strep tags to the complex. The cross-linking data confirm the conclusion derived from the negative-stain EM analysis that the addition of TIN2 to POT1/TPP1 yields a complex with a high degree of conformational heterogeneity. These data also support the conclusion that the extra signal we see in class averages of POT1/TPP1/GFP-TIN2 relative to POT1/TPP1^{N} is the TRFH domain of TIN2 rather than GFP (Fig. 2*D*, cartoons).

We used AlphaFold-Multimer (67, 68) to generate models for wild-type POT1/TPP1/TIN2 and the 3×Δ mutant (*SI Appendix*, Fig. 5). In all cases, the predicted models were consistent with experimental results presented here and elsewhere. Regions of known structure were predicted with high confidence (pLDDT) and resembled their corresponding entries in the Protein Data Bank, while regions of unknown structure were predicted with low confidence. The predicted structures of individual polypeptides within the complex resembled their predicted structures in isolation (66), indicating that the presence of binding partners did not significantly alter the predicted models. Known interacting regions between components (e.g., TPP1 RD and POT1 OB3/HJRL) were captured as off-diagonal boxes of low predicted aligned error (*SI Appendix*, Fig. 5, dark green color in predicted aligned error charts). Both POT1's N-terminal OB folds and TPP1's OB fold have high predicted aligned errors with respect to all other features, consistent with these domains flexing about their hinge regions. Interestingly, the TRFH domain of TIN2 is placed at the tip of the POT1 HJRL in all models with intermediate predicted aligned error with respect to POT1 OB3/HJRL (*SI Appendix*, Fig. 5, red arrows), potentially consistent with a weak and/or dynamic interaction between these two domains. Finally, we observe no major differences between predictions of wild-type and 3×Δ POT1/TPP1/TIN2, consistent with our observation that truncations of the linker regions failed to significantly alter the complex's structural variability.

POT1/TPP1/TIN2/TRF1 Bound to Telomeric DNA Is a Dimeric Complex with High Conformational Heterogeneity. A complex containing POT1, TPP1, TIN2, and TRF1 (henceforth "the TRF1 complex") was coexpressed in insect cells and purified to apparent homogeneity at high stringency (350 mM NaCl) (Fig. 3 *A*–*C* and *SI Appendix*, Fig. 6). To assess the compositional heterogeneity and stoichiometry of our preparation, we employed mass photometry (69, 70). Purified POT1/TPP1/GFP-TIN2/TRF1 showed four peaks at measured masses of 80, 170, 270, and 470 kDa, which correspond to $(TRF1)_2$ (97 kDa), POT1/TPP1/GFP-TIN2 (191 kDa), POT1/TPP1/GFP-TIN2/$(TRF1)_2$ (288 kDa), and (POT1/TPP1/GFP-TIN2/$TRF1)_2$ (480 kDa), respectively (Fig. 3*D* and *SI Appendix*, Table 4). We performed native MS analysis and detected assemblies with these stoichiometries at high mass accuracy and resolution (Fig. 3*E* and *SI Appendix*, Fig. 6*C* and Table 4). Overall, we observed that the TRF1 dimer can associate with one or two POT1/TPP1/GFP-TIN2 assemblies and that the TRF1 complex partially dissociates during purification and/or dilution to low concentration.

We then sought to stabilize the dimeric complex by adding a DNA containing both ds- and ssDNA of telomeric sequence. We found that DNA containing an optimal TRF1-binding site (71) adjacent to a $(GGTTAG)_4$ ssDNA 3′ overhang gave rise to a single species when bound to the TRF1 complex in a native gel (hereafter "TeloDNA1"; *SI Appendix*, Fig. 6*B*). When TeloDNA1 was added prior to ultracentrifugation, the purified TRF1 complex behaved as a single 500-kDa species in mass photometry (Fig. 3*D*), which is close to the predicted molecular mass of (POT1/TPP1/GFP-TIN2/$TRF1)_2$ bound to TeloDNA1 (505 kDa). Similarly, addition of TeloDNA1 greatly increased the formation of the 500-kDa complex when the TRF1 complex was reconstituted by mixing purified TRF1 with POT1/TPP1/GFP-TIN2 (*SI Appendix*, Fig. 6*D*).

The TRF1 complex bound to TeloDNA1 was purified by glycerol gradient/ultracentrifugation for analysis by negative-stain EM (*SI Appendix*, Figs. 6*E* and 7*A*). Peak fractions contained all four protein components and a native gel stained for nucleic acid demonstrated the presence of a single species bound to DNA (*SI Appendix*, Fig. 6*E*). Negative-stain EM images revealed that the particles were homogeneous in size and did not aggregate, but 2D class averages showed extensive heterogeneity (Fig. 3*F* and *SI Appendix*, Fig. 7*A*). At 200 mM KCl in the absence of DNA, the TRF1 complex tended to aggregate (*SI Appendix*, Figs. 6*E* and 7*B*), and 2D class averages of the nonaggregated particles displayed less defined features compared with the DNA-bound complex (*SI Appendix*, Fig. 7*B*).

For streptavidin-DNA labeling of the TRF1 complex, two POT1-binding sites were included in the DNA, allowing two POT1 proteins in the dimeric complex to bind. Negative-stain EM of POT1/TPP1^{N} bound to this substrate showed streptavidin close to two oblong bipartite lobes, as expected from the binding of pairs of POT1 OB1 and OB2 (*SI Appendix*, Fig. 8*A*). In some cases, the two sets of OB folds adopted a curved, "U" shape consistent with the DNA path observed in the crystal structure of the OB folds bound to DNA (58). Using this feature as a guide, we analyzed POT1/TPP1/GFP-TIN2 bound to this substrate (*SI Appendix*, Fig. 8*B*). Class averages of this sample showed larger complexes than those for POT1/TPP1^{N} and, while many classes contained streptavidin associated with a U feature, other elements were poorly resolved. Adding TRF1 to the reconstitution further decreased the quality of alignment within 2D classes, indicating that streptavidin-DNA labeling using this substrate is not a suitable strategy to identify elements within 2D class averages of the TRF1 complex (*SI Appendix*, Fig. 8*C*).

Alternative Shelterin Subcomplexes Are Highly Heterogeneous. Estimates of shelterin component copy numbers suggest a subcomplex containing Rap1, TRF2, TIN2, and potentially POT1/TPP1 could be present at the telomere (72). To assess the stoichiometry and architecture of this complex, we mixed purified TRF2/Rap1 and POT1/TPP1/TIN2 in the presence or absence of TeloDNA1 and isolated resulting complexes by glycerol gradient/ultracentrifugation (*SI Appendix*, Fig. 9*A*). Mass photometry of a peak fraction in the absence of DNA displayed two peaks: one consistent with free POT1/TPP1/TIN2 (191 kDa) or free TRF2/Rap1 (200 kDa) and the other consistent with a complex of 1:1:1:2:2 POT1:TPP1:TIN2:TRF2:Rap1 stoichiometry (391 kDa), as has been previously reported (62) (*SI Appendix*, Fig. 9*B* and Table 4). Addition of TeloDNA1 resulted in a fully dimeric form (583-kDa protein + 21-kDa DNA), although minor other species were present (*SI Appendix*, Fig. 9*B*). Negative-stain EM images indicated that these complexes had largely dissociated and were thus unsuitable for single-particle

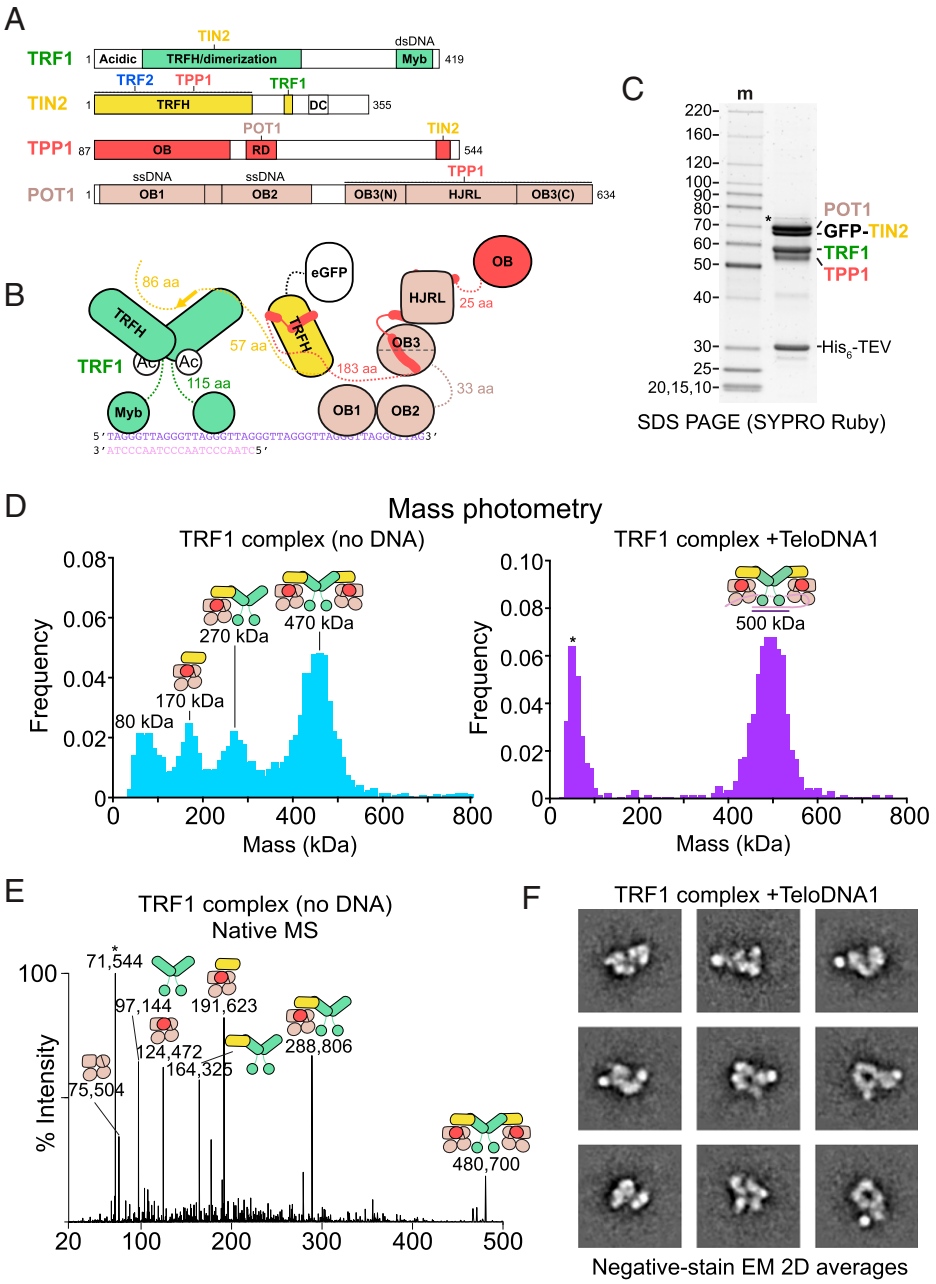

Fig. 3. Structural heterogeneity within the TRF1 complex. (*A*) Domain schematics for POT1, TPP1, TIN2, and TRF1 with amino acids indicated. (*B*) Cartoon schematic for POT1/TPP1/GFP-TIN2/TRF1 bound to a ds–ssDNA junction. (*C*) Purified POT1/TPP1/GFP-TIN2/TRF1 prior to glycerol-gradient purification fractionated on SDS-PAGE. Protein gel as in Fig. 1*B*. (*D*) Mass photometry of TRF1 complex purified in the presence or absence of ds–ss junction telomeric DNA (TeloDNA1). Peak maxima are indicated along with cartoons that correspond to complexes of that molecular mass to within 10%. (*E*) Native MS analysis of reconstituted TS-POT1/TPP1/GFP-TIN2/TRF1. The measured masses (in Da) are indicated along with cartoons that correspond to complexes of that approximate molecular mass to within 0.08%. The asterisk indicates a species that likely corresponds to HSP70 contamination. (*F*) Negative-stain EM analysis of POT1/TPP1/GFP-TIN2/TRF1 bound to TeloDNA1. Reference-free 2D class averages showing a high signal-to-noise ratio were selected for display.

analysis (*SI Appendix*, Fig. 9*C*), consistent with recent findings from a cryo-EM analysis of this complex (63).

Because POT1 and TPP1 are expressed at ∼10-fold lower levels than the other components, another subcomplex of shelterin that is likely present in cells is TRF1/TIN2/TRF2/Rap1 (shelterin core) (1, 72). Shelterin core was reconstituted by mixing TRF1/GFP-TIN2 with TRF2/Rap1 in the presence or absence of dsDNA containing four Myb-binding sites (dsTeloDNA; *SI Appendix*, Fig. 10*A*), and isolated by glycerol gradient/ultracentrifugation (*SI Appendix*, Fig. 10*B*). Mass photometry of a peak fraction in the absence of DNA displayed two peaks: one consistent with either TRF1/TIN2 (230 kDa) or free TRF2/Rap1 (200 kDa) and the other consistent with a complex of 2:2:2:2 TRF1:GFP-TIN2:TRF2:Rap1 stoichiometry (430 kDa) (*SI Appendix*, Fig. 10*C* and Table 4). Addition of dsTeloDNA resulted in a similar profile with a higher molecular mass for the second peak, but the exact identity of the complex responsible for this peak could not be determined due to overlap of the predicted sizes of possible complexes (*SI Appendix*, Fig. 10*C*). Negative-stain EM images from peak fractions revealed extreme heterogeneity, indicating that this complex was also unsuitable for single-particle analysis (*SI Appendix*, Fig. 10*D*).

These results demonstrate that TRF1 can stabilize dimeric forms of shelterin in the absence of DNA by interacting tightly

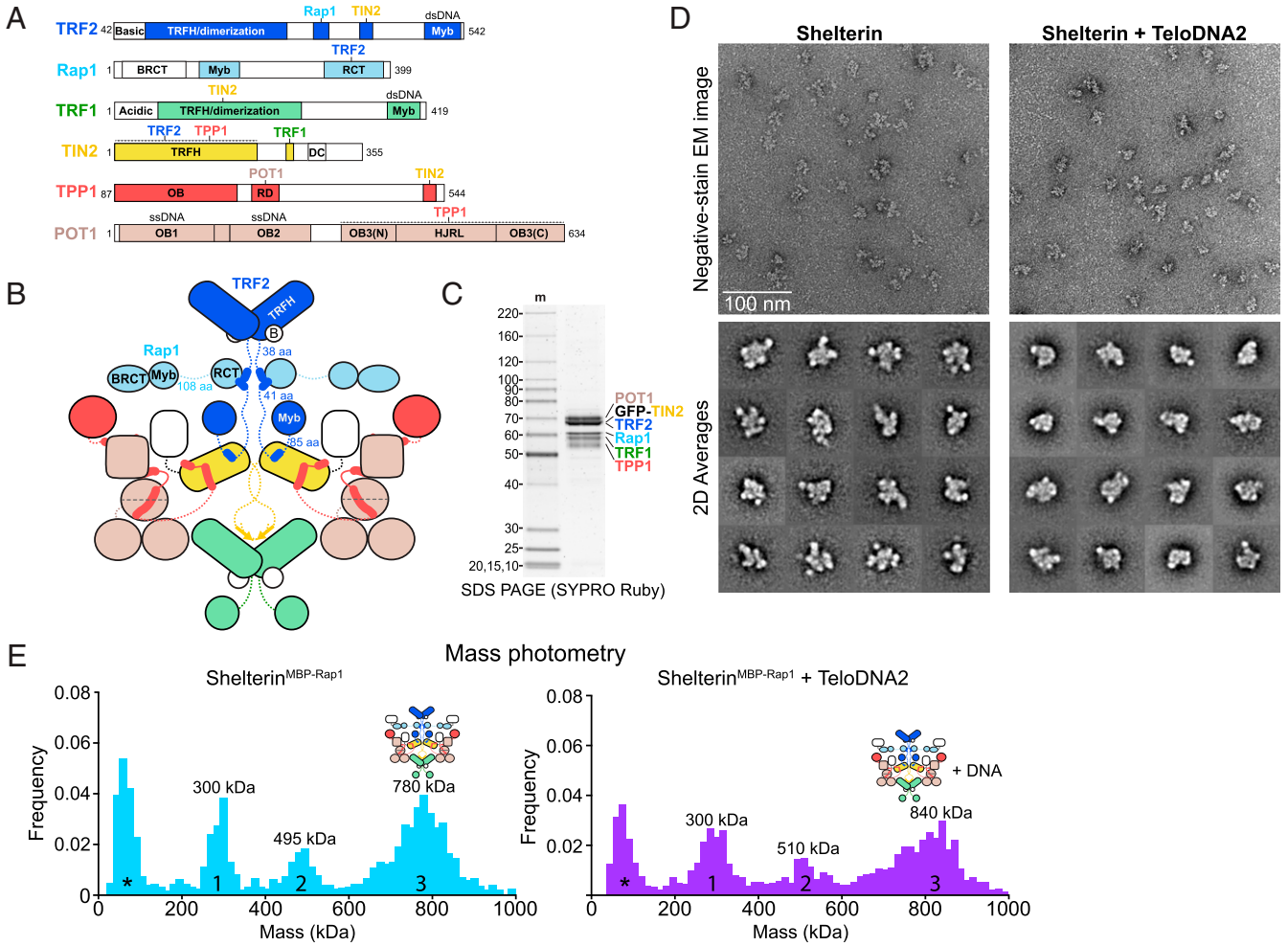

Fig. 4. Structural heterogeneity within shelterin. (*A*) Domain schematics for TRF2, Rap1, TRF1, TIN2, TPP1, and POT1 with amino acids indicated. BRCT, BRCA1 C-terminal domain; RCT, Rap1 C-terminal domain. (*B*) Cartoon schematic for dimeric shelterin. Coloring is as in Fig. 1*A*. (*C*) SDS-PAGE of purified reconstituted shelterin. Protein gel as in Fig. 1*B*. (*D*) Negative-stain EM analysis of shelterin purified in the presence or absence of ds–ss junction telomeric DNA (TeloDNA2). EM images displaying raw particles are shown above reference-free 2D class averages. Class averages showing a high signal-to-noise ratio were selected for display. (*E*) Mass photometry of reconstituted shelterin containing MBP-Rap1 and GFP-TIN2 after an amylose pull down. Data were taken in the presence or absence of exogenously added TeloDNA2 as indicated. Peaks are numbered and approximate maxima are indicated. Asterisks indicate a peak arising from minor buffer contaminants.

with two TIN2 molecules, and that TRF2/Rap1 binds TIN2 in both subcomplexes with relatively low affinity, consistent with previous measurements (23, 29). Additionally, we find that DNA-free and DNA-bound POT1/TPP1/TIN2/TRF2/Rap1 and shelterin core subcomplexes are too heterogeneous to be analyzed by single-particle EM.

Conformational Heterogeneity and Stoichiometry of Shelterin. The full shelterin complex (Fig. 4 *A*–*C*) was reconstituted by mixing purified TRF2/Rap1 (*SI Appendix*, Fig. 11 *A* and *B*) with the TRF1 complex in the presence and absence of a telomeric DNA containing four Myb-binding sites and two POT1-binding sites (TeloDNA2) (*SI Appendix*, Fig. 11 *A* and *B*). In the absence of DNA, fractions containing TRF2 and Rap1 were observed comigrating with components of the TRF1 complex at much higher glycerol concentrations than for TRF2/Rap1 alone (*SI Appendix*, Fig. 11 *A* and *B*). Unlike the TRF1 complex, DNA-free shelterin was both soluble and monodisperse at low salt. However, negative-stain EM analysis of a peak fraction from the glycerol gradient showed that shelterin behaved as a heterogeneous complex in 2D class averages (Fig. 4*D* and *SI Appendix*, Fig. 11*C*).

TeloDNA2-bound shelterin was more compact than the apo complex and migrated to higher density, but it remained extensively heterogeneous in 2D class averages (Fig. 4*D* and *SI Appendix*, Fig. 12). Thus, the addition of the TRF2/Rap1 components and/or DNA to the TRF1 complex prevented aggregation at physiological salt concentration but did not significantly limit conformational variability.

We then sought to probe the stoichiometry of the full complex, but the complex quantitatively bound to membranes during concentration/buffer exchange by spin filtration. To overcome this issue, we reconstituted shelterin by mixing the TRF1 complex with TRF2/His$_6$MBP-Rap1, separating the full complex from subcomplexes by glycerol gradient/ultracentrifugation, and performing an amylose pull down on the peak fractions (*SI Appendix*, Fig. 13 *A* and *B*). The isolated complex was analyzed by mass photometry, which showed three prominent peaks at 300, 495, and 780 kDa (not counting an additional low molecular mass peak from a buffer contamination) (Fig. 4*E* and *SI Appendix*, Fig. 13*C*). Peak 1 would be consistent with the molecular masses of either (TRF2/His$_6$MBP-Rap1)$_2$ (286 kDa) or POT1/TPP1/GFP-TIN2/(TRF1)$_2$ (285 kDa), peak 2 with either POT1/TPP1/GFP-TIN2/(TRF2/His$_6$MBP-Rap1)$_2$ (474 kDa) or (POT1/TPP1/GFP-TIN2/TRF1)$_2$ (473 kDa), and peak 3 with (POT1/TPP1/GFP-TIN2/TRF1/TRF2/His$_6$MBP-Rap1)$_2$ (759 kDa) (*SI Appendix*, Table 4). Addition of TeloDNA2 to this complex resulted in shifts of peaks 2 and 3 of approximately the

molecular mass of the DNA (41 kDa), further supporting our peak assignments (Fig. 4*E* and *SI Appendix*, Fig. 13*C*). From these observations, we conclude that shelterin can assemble into a fully dimeric complex.

Discussion

This study reports on the structure of shelterin and its subcomplexes reconstituted in vitro. No DNA was required to form the six-subunit shelterin complex, further confirming that shelterin can self-assemble in the absence of its telomeric DNA substrate (61). Single-particle negative-stain EM revealed that shelterin and its subcomplexes can adopt a wide array of conformations in both apo and DNA-bound states. Due to this extreme structural heterogeneity (Fig. 5), high-resolution structure determination of shelterin is currently intractable. Nonetheless, the data reported here provide important insights into the architecture and functions of shelterin.

The heterodimer formed by full-length POT1 and the N-terminal half of TPP1 (comprising its OB fold and RD) showed a large array of distinct conformations. The 2D class averages suggest a continuum of different positions of the C-terminal half of POT1 (composed of the split OB3/HJRL domain to which the TPP1 RD binds) and the two N-terminal OB folds that bind to ss telomeric DNA (Fig. 5*A*). Clearly, the association of TPP1 with the split OB3/HJRL domain of POT1 does not lead to a rigid POT1 structure. Therefore, POT1 may be best viewed as a DNA-binding module (OB1 and OB2) that is flexibly tethered to shelterin via the interaction of its split OB3/HJRL domain with TPP1. This architecture could enable POT1 to bind to ssDNA and fulfill its functions with greater versatility than if it were rigidly attached and need to be specially positioned by the dsDNA-binding components. Furthermore, TPP1 OB can continuously flex about its tether to POT1 OB3/HJRL, potentially aiding in its ability to recruit telomerase.

Reconstitution of the POT1/TPP1/TIN2 complex with full-length proteins resulted in a trimer with 1:1:1 stoichiometry, as expected. The structure of the POT1/TPP1 heterodimer was readily discerned within the heterotrimer, indicating that TIN2 does not significantly affect the positional variation of the POT1 halves. Interestingly, negative-stain EM and XLMS data revealed that the TRFH domain of TIN2 (which binds to the C-terminal region of TPP1) is found in close proximity to POT1 (Fig. 5*B*). The TIN2-binding site in TPP1 is separated from the POT1-binding RD domain by an unstructured and poorly conserved region of 150 aa. However, TIN2 is not freely diffusing around POT1/TPP1 on this tether but has several preferred positions yielding well-defined negative-stain EM class averages, high-confidence cross-links between the TIN2 TRFH domains of known structure in POT1/TPP1, and low predicted aligned errors between these domains in models generated with AlphaFold-Multimer. It is important to note that these data only indicate that these two domains are near one another and not that they directly interact, as it has previously been shown that TIN2 and POT1 cannot productively interact with each other in the absence of TPP1 (30). Consistent with the view that TIN2 is not randomly moving around POT1/TPP1, deletion of the 150-aa region of TPP1 did not have a significant effect on the negative-stain EM class averages, and this was further supported by the AlphaFold-Multimer models. These data suggest that POT1/TPP1/TIN2 may contain a core (composed of POT1 OB3/HJRL, TPP1 RD, and TIN2 TRFH) which, while variable in structure, may have a limited range of conformations. Contrasting with a "beads-on-a-string model," the behavior of the POT1/TPP1/TIN2 core resembles that of a toy wooden snake (Fig. 5*B*), in which the segments would represent structured domains that associate closely but can adopt many different

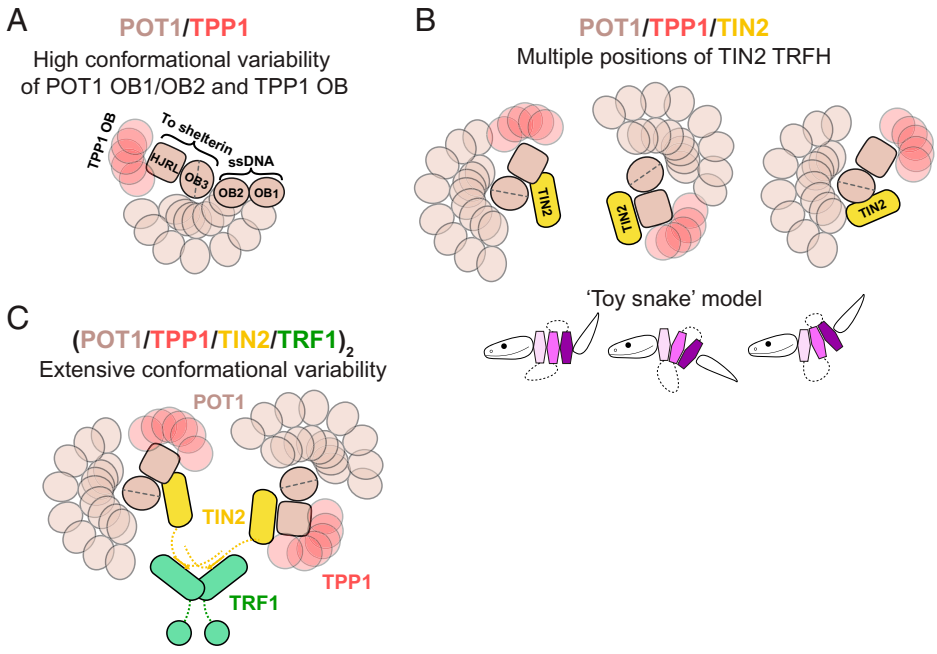

Fig. 5. Models of structural heterogeneity in human shelterin. (*A*) Visual depiction of conformational variability within POT1/TPP1^{N}. In these models, the C-terminal domains of POT1 are arbitrarily used as an anchor. Conformational variability of the DNA-binding OB folds and the TPP1 OB fold is depicted as an ensemble of possible conformations that these features can adopt relative to that anchor. (*B*) Models for flexibility within POT1/TPP1/TIN2. The TRFH domain of TIN2 is shown occupying different discrete positions on the POT1 anchor. For the "toy snake" model, structured domains are depicted as hexagonal snake segments with flexible linkers shown as dotted lines connecting them. These domains can associate with each other in a variety of conformations, and truncation of flexible linkers would result in only minor changes in structure and conformational variability. (*C*) Model for structural heterogeneity of the TRF1 complex. The POT1/TPP1/TIN2 subcomplex remains flexible and TIN2 is presumably flexibly tethered to TRF1.

conformations relative to each other. In this analogy, the flexible regions tethering the structured modules would be looping out from them such that their removal would leave the structure and flexibility of the complex unchanged. Future studies using time-resolved measurements of conformational states will be needed to determine if individual complexes can toggle freely between conformational states.

A recent cryo-EM study of a chemically cross-linked POT1/TPP1/TIN2-telomerase complex from humans (64) confirmed the interaction surface between TPP1's OB fold and the TEN-TRAP domains of telomerase (5, 7). A subset of these particles was found to contain density for OB1/OB2 of POT1 bound to DNA. However, TIN2 and the POT1 OB3/HJRL were absent from that reconstruction, consistent with our data supporting continuous flexibility between these features and the POT1 OB1/OB2. Another group performed cryo-EM on POT1/TPP1/TIN2 alone and was able to find a potentially preferred conformation of POT1, in which its OB1/OB2 and OB3/HJRL domains are oriented perpendicular to each other (63). This study failed to observe TIN2's TRFH domain or TPP1's OB fold, but these findings are complicated by the possibility that the particles may have been partially denatured due to interactions with the air–water interface.

Addition of TRF1 to the POT1/TPP1/TIN2 complex yielded a shelterin subcomplex that could have a fully dimeric stoichiometry (Fig. 5*C*). We also observed complexes containing only one POT1/TPP1/TIN2 trimer bound to the TRF1 homodimer, similar to observations of the TRF2-containing subcomplex (62). However, in both the apo state and when bound to a DNA substrate containing a TRF1- and two POT1-binding sites, the dimeric complex dominated, which was also the case for the full shelterin complex. Nonetheless, it is likely that a variety of shelterin components can form a variety of distinct complexes, including the shelterin core, TRF1/TIN2, and POT1/TPP1/TIN2/TRF2/Rap1. The idea that shelterin subcomplexes can be functional is supported by the observation that breaking of the TIN2 bridge at either the TRF1 or TRF2 site results in minor or no problems with telomere protection, as long as POT1/TPP1/TIN2 can efficiently be recruited to the telomere (27, 28). Structurally, addition of TRF1 did not significantly decrease the variability of the POT1/TPP1/TIN2 complex, resulting in many different negative-stain EM class averages with features that we were unable to unambiguously assign. Similarly, reconstitution of the full shelterin complex by addition of TRF2/Rap1 resulted in a great degree of structural variability regardless of whether it was bound to a DNA substrate.

A limitation of this study is that it assumes that there are no missing components of shelterin, which could, theoretically, lend conformational homogeneity to the complex. From numerous proteomic studies it is clear that there are no telomere-specific factors missing. However, it is not excluded that proteins that are not specific to telomeres play a role in the structure of the complex. Recent examples of this phenomenon are the telomerase and γ-TURC complexes, each of which was purified from human cells and copurified with abundant endogenous components that enhanced their stability (H2A/H2B for telomerase and actin for γ-TURC) (73–76). Additionally, it is possible that elements of shelterin are stabilized in discrete conformations when bound to one or more of its many accessory factors or when in association with the telomeric DNA maintenance factors (e.g., CST or Apollo).

All known interactions within shelterin occur between small domains or peptides flanked by disordered regions and structured domains, and not between two structured domains (1, 23, 52). This study uncovered an association between the TRFH domain of TIN2 and a structured region of POT1/TPP1, but this interaction was found to be highly variable (Fig. 5*B*, toy snake model). These features of shelterin likely underlie the observation that increasingly larger shelterin complexes resulted in increasing structural variability, which decreased the quality of alignment in single-particle analysis and precluded localization of structured domains within the larger complexes.

Why is shelterin so flexible? We speculate that structural heterogeneity is a beneficial feature of the complex that permits shelterin to fulfill its many functions. For example, flexibility may enable shelterin to accommodate the many conformations adopted by telomeric DNA. Shelterin binds to telomeres in their linear state, associating with ds repeats that are largely chromatinized and with the ds–ss transition and 3′ overhang at the terminus. Shelterin-bound telomeres also can adopt a t-loop configuration, featuring a D loop as well as branched DNA at the t-loop base. Furthermore, the telomeric long noncoding RNA, TERRA, can form R loops in telomeres, which may present additional challenges to shelterin (77). Finally, shelterin needs to accommodate numerous interacting partners involved in the maintenance of telomeric DNA (e.g., telomerase, CST, Apollo, BLM, RTEL1), which may require different conformations to position each on their respective DNA substrates.

Methods

Cloning and Protein Expression. All constructs were cloned for baculovirus-based expression of multicomponent complexes in insect cells using the biGBac system (78). Briefly, DNA fragments encoding proteins, tags, and/or protease-cleavage sites were generated by PCR and cloned between a polyhedrin promoter and SV40 terminator in a subcloning vector (pLIB) using Gibson assembly (79). Gene-expression cassettes containing the promoter, expression construct, and terminator were amplified from these vectors by PCR using a predefined primer set to allow Gibson assembly into another vector (pBIG1). In the case of POT1/TPP1/GFP-TIN2/TRF1, these vectors were digested with *Pme*I and further assembled into a final expression vector (pBIG2). pLIB subcloning vectors were verified by sequencing of the entire insert and pBIG vectors were verified by analytical restriction digestion with *Swa*I.

All insect cell incubation steps were performed at 27 °C and high humidity (68 to 74%), and all transfections/infections were performed on cells in logarithmic growth. Vectors were converted into purified bacmids using DH10Bac (Thermo Fisher) cells per the manufacturer's instructions. Bacmids were transfected into adherent Sf9 cells grown in Sf-900 II serum-free medium (Thermo Fisher) using CellFectin II (Thermo Fisher). Viruses were harvested and amplified, and then used to infect Tni cells grown in suspension in ESF900 medium. Tni cells were harvested 72 h post infection, flash-frozen in liquid nitrogen, and stored at −80 °C until use. For expression of TRF1/GFP-TIN2 used in the shelterin core reconstitution, Tni cells were coinfected with equal volumes of P3s for mCherry-ENLYFQGGS-TRF1 and eGFP-GGLEVPFQGPGS-TIN2(1–352).

Protein Purification. Prior to purification, frozen cells were disrupted by cryogenic milling (Retsch) and approximately half of the resulting powder of a 400-mL culture was used for purification while the other half was stored at −80 °C. All protein-purification steps were performed on ice or at 8 °C and using chilled buffers. All protein concentrations were determined by A_{280} and a calculated extinction coefficient from https://web.expasy.org/protparam/ for complexes lacking GFP or by A_{488} for GFP-containing complexes using an extinction coefficient for enhanced GFP (eGFP) at 488 nm of 53,000 $M^{-1}{\cdot}cm^{-1}$. Ultraviolet-visible absorbance measurements were made using a NanoDrop ND-1000 (Thermo Fisher). All buffers used in fast protein liquid chromatography (FPLC) were degassed using a benchtop vacuum and cooled to 8 °C prior to use. Glycerol gradients were mixed using a Gradient Master (BioComp) at room temperature and cooled to 4 °C prior to use. All polyacrylamide gels were purchased from Thermo Fisher, and BenchMark unstained protein ladder (Thermo Fisher) was used as a molecular mass marker for SDS-PAGE. Protein complexes were purified by a

combination of one or more affinity purification steps followed by either heparin affinity, strong-anion exchange, or strong-cation-exchange chromatography and gel filtration or glycerol gradient/ultracentrifugation. Detailed descriptions of protein purification protocols for each complex can be found in *SI Appendix*.

Electrophoretic Mobility-Shift Assays. For POT1/TPP1^{N} and POT1/TPP1/GFP-TIN2, variable concentrations of protein were mixed with 0.25 nM Cy5.5-labeled DNA in binding buffer (20 mM Hepes·KOH, pH 7.5, 100 mM KCl, 0.5 mM $MgCl_2$, 0.5 mM Tris[2-carboxyethyl]phosphine [TCEP]·HCl, 0.05% volume vol/vol IGEPAL Co-630, 8% glycerol, 50 μg/mL bovine serum albumin [New England BioLabs]) and incubated for 30 min at room temperature (~22 °C). For POT1/TPP1^{N}, 2.5 μL was separated on a 10% polyacrylamide-Tris-borate EDTA (TBE) gel in cold 100 mM Tris-borate (pH 8.3). For POT1/TPP1/GFP-TIN2, 6 μL was separated on 0.6% agarose-Tris-borate in cold 100 mM Tris-borate (pH 8.3). Gels were imaged using the near-infrared setting of a Typhoon scanner (GE). For the qualitative electrophoretic mobility-shift assay (EMSA) of POT1/TPP1/GFP-TIN2/TRF1, variable concentrations of protein were mixed with the indicated DNA (50 nM) in binding buffer containing 200 mM KCl and incubated on ice for 30 min. Each protein/DNA complex (5 μL) was then separated on 4 to 20% polyacrylamide TBE gels run in cold Tris-borate buffer, and the gel was stained with SYBR Gold per the manufacturer's instructions and imaged using the ethidium bromide setting of an Alpha imager (Alpha Innotech). For native gels of glycerol-gradient fractions, 5-μL fractions were run and the gel was stained and imaged as with the qualitative TRF1 complex EMSA.

Negative-Stain EM Data Collection and Image Processing. Protein samples (3.5 μL) were adsorbed to glow-discharged carbon-coated copper/collodion grids, washed three times with water and once with freshly prepared 0.7% weight/vol uranyl formate, and then stained for 30 s with 0.7% uranyl formate. Samples were imaged at room temperature using a Phillips CM10 electron microscope equipped with a tungsten filament and operated at an acceleration voltage of 80 kV. Images were collected with an AMT ActiveVu charge-coupled device camera at a calibrated pixel size of 2.8 Å. Particles were autopicked using the swarm (for POT1/TPP1^{N} and POT1/TPP1/TIN2 wt/3×Δ) or Gauss (for other complexes) picker in EMAN2.1 (80). Particles were then extracted and subjected to two or three rounds of 2D classification in RELION-3.0 (81) to remove junk particles. After the final round of classification, particles were reextracted using the updated coordinates from 2D classification in RELION, and that stack was used as input for 2D classification in ISAC 2.3.2 (65) using a pixel error threshold of 0.7 and variable minimum and maximum particle numbers per class (*SI Appendix*, Table 1). Classes containing readily interpretable features from the ISAC output were selected for display in the main figures while all classes are displayed in *SI Appendix*.

Cross-Linking Mass Spectrometry. BS3 cross-linker (ProteoChem, c1103) was dissolved in LC-MS–grade H_2O (ProteoChem, LC6330) at 50 mM and added to POT1/TPP1/GFP prepared at 1 mg/mL in *N*-hydroxysuccinimide–ester nonreactive buffer to a final concentration of 0.35 to 0.75 mM. Reactions were performed at 25 °C in disposable inert cuvettes (UVette, Eppendorf), and monitored by continuous looped dynamic light-scattering measurements of polydispersity (Pd < 10%; DynaPro NanoStar, Wyatt) (82). Cross-linking was quenched after 30-min incubation by addition of Tris·HCl (pH 8.0) to a final concentration of 5 mM.

Samples were dialyzed against 100 mM ammonium bicarbonate, reduced with 50 mM TCEP at 60 °C for 10 min, and alkylated with 50 mM iodoacetamide in the dark for 15 min at 37 °C. Digestion was carried out at 37 °C overnight with 125 ng/mL sequencing-grade modified trypsin (Thermo Fisher) in 25 mM ammonium bicarbonate supplemented with ProteaseMax (Thermo Fisher). The reaction mixture was supplemented with trifluoroacetic acid (TFA; Thermo Fisher) to a final concentration of 0.1%. The resulting peptides were passed through C18 Spin Tips (Thermo Fisher) before elution with 40 μL 80% acetonitrile (ACN; Thermo Fisher) in 0.1% TFA. Eluted peptides were dried and resuspended in 20 μL 0.1% formic acid (Thermo Fisher) for MS analysis. Peptides were analyzed on an Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher) coupled to an EASY-nLC (Thermo Fisher) LC system, with a 2-μm, 500-mm EASY-Spray column. The peptides were eluted over a 120-min linear gradient from 96% buffer A (water) to 40% buffer B (ACN), and then continued to 98% buffer B over 20 min with a flow rate of 250 nL/min. Each full MS scan ($R = 60{,}000$) was followed by 20 data-dependent MS2 scans ($R = 15{,}000$) with high-energy collisional dissociation (HCD) and an isolation window of 2.0 *m/z*. Normalized collision energy was set to 35. Precursors of charge state ≤3 were collected for MS2 scans in enumerative mode, and precursors of charge state 4 to 6 were collected for MS2 scans in cross-link discovery mode (both were performed for each sample); monoisotopic precursor selection was enabled and a dynamic exclusion window was set to 30.0 s. Raw files obtained in enumerative mode were analyzed by pFind3 software (83) in open search mode, and protein modifications inferred by pFind3 and comprising >0.5% of total protein were included as the variable modifications in pLink 2 (84) search parameters. pLink 2 results were filtered for false discovery rate (<5%), *e* value (<10 to 3), score (<10 to 2), and abundance (peptide spectrum matches ≥ 5). Cross-links were visualized using xiNET (85).

Mass Photometry. All data were collected using a OneMP mass photometer (Refeyn) calibrated with bovine serum albumin (66 kDa), beta amylase (224 kDa), and thyroglobulin (660 kDa). Movies were acquired for 6,000 frames (60 s) using AcquireMP software (version 2.4.0) and default settings. Final protein concentrations were empirically determined to achieve ~50 binding events per second. Raw data were converted to frequency distributions using Prism 9 (GraphPad) and a bin size of 10 to 30 Da. Details for sample preparation can be found in *SI Appendix*.

Native Mass Spectrometry. Prior to analysis, aliquots of purified POT1/TPP1/GFP-TIN2 and TRF1 were thawed and treated overnight at 4 °C with 50 μg/mL purified lambda protein phosphatase (made in-house) and 1 mM $MnCl_2$. For the TRF1 complex, the sample was prepared identically as for mass photometry except using phosphatase-treated components. The sample was immediately processed for native MS after glycerol exchange.

The samples were buffer-exchanged into 300 mM ammonium acetate (POT1/TPP1/GFP-TIN2) or 400 mM ammonium acetate (TRF1-only and TRF1 complex) solutions at pH 7.5 and with 0.01% Tween-20 using Zeba desalting microspin columns with 40-kDa molecular mass cutoff (Thermo Fisher). Sample concentrations were at least 5 μM for analysis. A 3-μL aliquot of the buffer-exchanged sample was loaded into a gold-coated quartz capillary tip that was prepared in-house (86). The sample was then electrosprayed into an Exactive Plus extended mass range instrument (Thermo Fisher) using a modified static nanospray source. The MS parameters used included spray voltage, 1.2 kV; capillary temperature, 150 °C; S lens radio frequency level, 200; resolving power, 8,750 or 17,500 at an *m/z* of 200; AGC target, 1×10^6; number of microscans, 5; maximum injection time, 200 ms; in-source dissociation, 0 V; injection flatapole, 8 V; interflatapole, 4 V (TRF1 trimer complex) or 6 V (for other samples); bent flatapole, 4 V (TRF1 complex) or 6 V (other samples); HCD, 150 to 200 V; ultrahigh vacuum pressure, 5.2 to 6.0×10^{-10} mbar; total number of scans, at least 100. The instrument mass calibration in positive EMR mode was performed using cesium iodide. The acquired MS spectra were visualized using the Thermo Xcalibur Qual Browser (version 4.2.47). Data processing and spectra deconvolution were performed using UniDec version 4.2.0 (87, 88). The general UniDec parameters used included sample mass every 1.0 Da; smooth charge state distribution, on; and peak shape function, Gaussian. No background subtraction was applied for the TRF1-only spectrum, and curved background subtraction was set to 10 for spectra from the rest of the samples. The expected masses for each component protein based on the primary sequence are the following: TRF1 (48,538 Da), Twin-Strep-POT1 (75,438 Da), TPP1 (48,939 Da), and GFP-TIN2 (67,237 Da). Mass accuracies (calculated as the % mass difference between the measured and expected mass relative to the expected mass) ranged from 0.005 to 0.08%.

AlphaFold-Multimer Modeling. For the full-length POT1/TPP1/TIN2 complex, sequences in FASTA format for human POT1, TPP1 (short isoform), and TIN2 (short isoform) were used as inputs for AlphaFold-Multimer 2.2.0 with two random seeds and default settings. For the 3×Δ mutant, FASTA files were modified to reflect the appropriate truncations. Of the 10 models generated by AlphaFold-Multimer, the top three ranked (based on average pLDDT score) were chosen for display.

Data Availability. All study data are included in the article and/or *SI Appendix*.

ACKNOWLEDGMENTS. We thank members of the T.W. and T.d.L. laboratories for their critical reading of the manuscript and discussions throughout data

collection and analysis. We also thank Yixiao Zhang for his assistance in implementing ISAC with graphics processing unit acceleration, and Lauren Vostal and Tarun Kapoor for training on and use of the mass photometry instrument. Additionally, we thank John Jumper for helpful suggestions, and Sarah Cai and Logan Sweezy for their technical assistance in running AlphaFold-Multimer. Native MS was performed by P.D.B.O. in the laboratory of B.T.C. and cross-linking MS was performed by V.S. in the laboratory of E.N. This research is supported by grants from the National Institutes of Health (5 R35 CA210036 to T.d.L., P41 GM109824 and P41 GM103314 to B.T.C., and T32 CA009673 to J.C.Z.) and a grant from the Damon Runyon Cancer Research Foundation (DRG 2337-19 to J.C.Z.). J.C.Z. is the Lorraine Egan Fellow of the Damon Runyon Cancer Research Foundation. This content is solely the responsibility of the authors and does not represent the official views of the NIH. E.N. is an Investigator of the Howard Hughes Medical Institute.

1. T. de Lange, Shelterin-mediated telomere protection. *Annu. Rev. Genet.* **52**, 223–247 (2018).
2. C. W. Greider, E. H. Blackburn, Identification of a specific telomere terminal transferase activity in *Tetrahymena* extracts. *Cell* **43**, 405–413 (1985).
3. J. Lingner *et al.*, Reverse transcriptase motifs in the catalytic subunit of telomerase. *Science* **276**, 561–567 (1997).
4. T. de Lange, Shelterin: The protein complex that shapes and safeguards human telomeres. *Genes Dev.* **19**, 2100–2110 (2005).
5. J. Nandakumar *et al.*, The TEL patch of telomere protein TPP1 mediates telomerase recruitment and processivity. *Nature* **492**, 285–289 (2012).
6. F. L. Zhong *et al.*, TPP1 OB-fold domain controls telomere maintenance by recruiting telomerase to chromosome ends. *Cell* **150**, 481–494 (2012).
7. 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).
8. J. Maciejowski, T. de Lange, Telomeres in cancer: Tumour suppression and genome instability. *Nat. Rev. Mol. Cell Biol.* **18**, 175–186 (2017).
9. S. E. Artandi, R. A. DePinho, Telomeres and telomerase in cancer. *Carcinogenesis* **31**, 9–18 (2010).
10. P. C. Haycock *et al.*; Telomeres Mendelian Randomization Collaboration, Association between telomere length and risk of cancer and non-neoplastic diseases: A Mendelian randomization study. *JAMA Oncol.* **3**, 636–651 (2017).
11. N. D. Nelson, A. A. Bertuch, Dyskeratosis congenita as a disorder of telomere maintenance. *Mutat. Res.* **730**, 43–51 (2012).
12. S. A. Savage *et al.*, Sequence analysis of the shelterin telomere protection complex genes in dyskeratosis congenita. *J. Med. Genet.* **48**, 285–288 (2011).
13. S. A. Savage, A. A. Bertuch, The genetics and clinical manifestations of telomere biology disorders. *Genet. Med.* **12**, 753–764 (2010).
14. L. Chong *et al.*, A human telomeric protein. *Science* **270**, 1663–1667 (1995).
15. D. Broccoli, A. Smogorzewska, L. Chong, T. de Lange, Human telomeres contain two distinct Myb-related proteins, TRF1 and TRF2. *Nat. Genet.* **17**, 231–235 (1997).
16. P. König, L. Fairall, D. Rhodes, Sequence-specific DNA recognition by the Myb-like domain of the human telomere binding protein TRF1: A model for the protein-DNA complex. *Nucleic Acids Res.* **26**, 1731–1740 (1998).
17. T. Bilaud *et al.*, Telomeric localization of TRF2, a novel human telobox protein. *Nat. Genet.* **17**, 236–239 (1997).
18. A. Bianchi, S. Smith, L. Chong, P. Elias, T. de Lange, TRF1 is a dimer and bends telomeric DNA. *EMBO J.* **16**, 1785–1794 (1997).
19. L. Fairall, L. Chapman, H. Moss, T. de Lange, D. Rhodes, Structure of the TRFH dimerization domain of the human telomeric proteins TRF1 and TRF2. *Mol. Cell* **8**, 351–361 (2001).
20. J. Z. Ye *et al.*, TIN2 binds TRF1 and TRF2 simultaneously and stabilizes the TRF2 complex on telomeres. *J. Biol. Chem.* **279**, 47264–47271 (2004).
21. D. Liu, M. S. O'Connor, J. Qin, Z. Songyang, Telosome, a mammalian telomere-associated complex formed by multiple telomeric proteins. *J. Biol. Chem.* **279**, 51338–51342 (2004).
22. S. H. Kim, P. Kaminker, J. Campisi, TIN2, a new regulator of telomere length in human cells. *Nat. Genet.* **23**, 405–412 (1999).
23. C. Hu *et al.*, Structural and functional analyses of the mammalian TIN2-TPP1-TRF2 telomeric complex. *Cell Res.* **27**, 1485–1502 (2017).
24. J. D. Griffith *et al.*, Mammalian telomeres end in a large duplex loop. *Cell* **97**, 503–514 (1999).
25. Y. Doksani, J. Y. Wu, T. de Lange, X. Zhuang, Super-resolution fluorescence imaging of telomeres reveals TRF2-dependent T-loop formation. *Cell* **155**, 345–356 (2013).
26. L. A. Timashev, T. de Lange, Characterization of t-loop formation by TRF2. *Nucleus* **11**, 164–177 (2020).
27. D. Frescas, T. de Lange, TRF2-tethered TIN2 can mediate telomere protection by TPP1/POT1. *Mol. Cell. Biol.* **34**, 1349–1362 (2014).
28. K. K. Takai, T. Kibe, J. R. Donigian, D. Frescas, T. de Lange, Telomere protection by TPP1/POT1 requires tethering to TIN2. *Mol. Cell* **44**, 647–659 (2011).
29. Y. Chen *et al.*, A shared docking motif in TRF1 and TRF2 used for differential recruitment of telomeric proteins. *Science* **319**, 1092–1096 (2008).
30. J. Z. Ye *et al.*, POT1-interacting protein PIP1: A telomere length regulator that recruits POT1 to the TIN2/TRF1 complex. *Genes Dev.* **18**, 1649–1654 (2004).
31. D. Liu *et al.*, PTOP interacts with POT1 and regulates its localization to telomeres. *Nat. Cell Biol.* **6**, 673–680 (2004).
32. B. R. Houghtaling, L. Cuttonaro, W. Chang, S. Smith, A dynamic molecular link between the telomere length regulator TRF1 and the chromosome end protector TRF2. *Curr. Biol.* **14**, 1621–1631 (2004).
33. P. Baumann, T. R. Cech, Pot1, the putative telomere end-binding protein in fission yeast and humans. *Science* **292**, 1171–1175 (2001).
34. D. Hockemeyer, J. P. Daniels, H. Takai, T. de Lange, Recent expansion of the telomeric complex in rodents: Two distinct POT1 proteins protect mouse telomeres. *Cell* **126**, 63–77 (2006).
35. E. L. Denchi, T. de Lange, Protection of telomeres through independent control of ATM and ATR by TRF2 and POT1. *Nature* **448**, 1068–1071 (2007).
36. Y. Gong, T. de Lange, A Shld1-controlled POT1a provides support for repression of ATR signaling at telomeres through RPA exclusion. *Mol. Cell* **40**, 377–387 (2010).
37. K. Kratz, T. de Lange, Protection of telomeres 1 proteins POT1a and POT1b can repress ATR signaling by RPA exclusion, but binding to CST limits ATR repression by POT1b. *J. Biol. Chem.* **293**, 14384–14392 (2018).
38. L. Wu *et al.*, Pot1 deficiency initiates DNA damage checkpoint activation and aberrant homologous recombination at telomeres. *Cell* **126**, 49–62 (2006).
39. A. Sfeir, S. Kabir, M. van Overbeek, G. B. Celli, T. de Lange, Loss of Rap1 induces telomere recombination in the absence of NHEJ or a DNA damage signal. *Science* **327**, 1657–1661 (2010).
40. P. Martínez *et al.*, RAP1 protects from obesity through its extratelomeric role regulating gene expression. *Cell Rep.* **3**, 2059–2074 (2013).
41. A. Sfeir *et al.*, Mammalian telomeres resemble fragile sites and require TRF1 for efficient replication. *Cell* **138**, 90–103 (2009).
42. G. Sarek, J. B. Vannier, S. Panier, J. H. J. Petrini, S. J. Boulton, TRF2 recruits RTEL1 to telomeres in S phase to promote t-loop unwinding. *Mol. Cell* **57**, 622–635 (2015).
43. M. Zimmermann, T. Kibe, S. Kabir, T. de Lange, TRF1 negotiates TTAGGG repeat-associated replication problems by recruiting the BLM helicase and the TPP1/POT1 repressor of ATR signaling. *Genes Dev.* **28**, 2477–2491 (2014).
44. Z. Yang, K. K. Takai, C. A. Lovejoy, T. de Lange, Break-induced replication promotes fragile telomere formation. *Genes Dev.* **34**, 1392–1405 (2020).
45. M. van Overbeek, T. de Lange, Apollo, an Artemis-related nuclease, interacts with TRF2 and protects human telomeres in S phase. *Curr. Biol.* **16**, 1295–1302 (2006).
46. C. Lenain *et al.*, The Apollo 5′ exonuclease functions together with TRF2 to protect telomeres from DNA repair. *Curr. Biol.* **16**, 1303–1310 (2006).
47. P. Wu, M. van Overbeek, S. Rooney, T. de Lange, Apollo contributes to G overhang maintenance and protects leading-end telomeres. *Mol. Cell* **39**, 606–617 (2010).
48. Y. C. Lam *et al.*, SNMIB/Apollo protects leading-strand telomeres against NHEJ-mediated repair. *EMBO J.* **29**, 2230–2241 (2010).
49. P. Wu, H. Takai, T. de Lange, Telomeric 3′ overhangs derive from resection by Exo1 and Apollo and fill-in by POT1b-associated CST. *Cell* **150**, 39–52 (2012).
50. M. Wan, J. Qin, Z. Songyang, D. Liu, OB fold-containing protein 1 (OBFC1), a human homolog of yeast Stn1, associates with TPP1 and is implicated in telomere length regulation. *J. Biol. Chem.* **284**, 26725–26731 (2009).
51. L. Y. Chen, S. Redon, J. Lingner, The human CST complex is a terminator of telomerase activity. *Nature* **488**, 540–544 (2012).
52. C. J. Lim, T. R. Cech, Shaping human telomeres: From shelterin and CST complexes to telomeric chromatin organization. *Nat. Rev. Mol. Cell Biol.* **22**, 283–298 (2021).
53. L. R. Myler *et al.*, The evolution of metazoan shelterin. *Genes Dev.* **35**, 1625–1641 (2021).
54. R. Court, L. Chapman, L. Fairall, D. Rhodes, How the human telomeric proteins TRF1 and TRF2 recognize telomeric DNA: A view from high-resolution crystal structures. *EMBO Rep.* **6**, 39–45 (2005).
55. C. Rice *et al.*, Structural and functional analysis of the human POT1-TPP1 telomeric complex. *Nat. Commun.* **8**, 14928 (2017).
56. C. Chen *et al.*, Structural insights into POT1-TPP1 interaction and POT1 C-terminal mutations in human cancer. *Nat. Commun.* **8**, 14929 (2017).
57. F. Wang *et al.*, The POT1-TPP1 telomere complex is a telomerase processivity factor. *Nature* **445**, 506–510 (2007).
58. M. Lei, E. R. Podell, T. R. Cech, Structure of human POT1 bound to telomeric single-stranded DNA provides a model for chromosome end-protection. *Nat. Struct. Mol. Biol.* **11**, 1223–1229 (2004).
59. Y. Chen *et al.*, A conserved motif within RAP1 has diversified roles in telomere protection and regulation in different organisms. *Nat. Struct. Mol. Biol.* **18**, 213–221 (2011).
60. S. Hanaoka *et al.*, NMR structure of the hRap1 Myb motif reveals a canonical three-helix bundle lacking the positive surface charge typical of Myb DNA-binding domains. *J. Mol. Biol.* **312**, 167–175 (2001).
61. F. Erdel *et al.*, Telomere recognition and assembly mechanism of mammalian shelterin. *Cell Rep.* **18**, 41–53 (2017).
62. C. J. Lim, A. J. Zaug, H. J. Kim, T. R. Cech, Reconstitution of human shelterin complexes reveals unexpected stoichiometry and dual pathways to enhance telomerase processivity. *Nat. Commun.* **8**, 1075 (2017).
63. E. W. Smith, S. Lattmann, Z. B. Liu, B. Ahsan, D. Rhodes, Insights into POT1 structural dynamics revealed by cryo-EM. *PLoS One* **17**, e0264073 (2022).
64. Z. Sekne, G. E. Ghanim, A. M. van Roon, T. H. D. Nguyen, Structural basis of human telomerase recruitment by TPP1-POT1. *Science* **375**, 1173–1176 (2022).
65. Z. Yang, J. Fang, J. Chittuluru, F. J. Asturias, P. A. Penczek, Iterative stable alignment and clustering of 2D transmission electron microscope images. *Structure* **20**, 237–247 (2012).
66. K. Tunyasuvunakool *et al.*, Highly accurate protein structure prediction for the human proteome. *Nature* **596**, 590–596 (2021).
67. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
68. R. Evans *et al.*, Protein complex prediction with AlphaFold-Multimer. *bioRxiv* [Preprint] (2022). https://doi.org/10.1101/2021.10.04.463034 (Accessed 28 April 2022).
69. G. Young *et al.*, Quantitative mass imaging of single biological macromolecules. *Science* **360**, 423–427 (2018).
70. A. Sonn-Segev *et al.*, Quantifying the heterogeneity of macromolecular machines by mass photometry. *Nat. Commun.* **11**, 1772 (2020).
71. A. Bianchi *et al.*, TRF1 binds a bipartite telomeric site with extreme spatial flexibility. *EMBO J.* **18**, 5735–5744 (1999).
72. K. K. Takai, S. Hooper, S. Blackwood, R. Gandhi, T. de Lange, In vivo stoichiometry of shelterin components. *J. Biol. Chem.* **285**, 1457–1467 (2010).
73. G. E. Ghanim *et al.*, Structure of human telomerase holoenzyme with bound telomeric DNA. *Nature* **593**, 449–453 (2021).
74. F. Wan *et al.*, Zipper head mechanism of telomere synthesis by human telomerase. *Cell Res.* **31**, 1275–1290 (2021).

75. M. Wieczorek *et al.*, Asymmetric molecular architecture of the human γ-tubulin ring complex. *Cell* **180**, 165–175.e16 (2020).
76. P. Liu *et al.*, Insights into the assembly and activation of the microtubule nucleator γ-TuRC. *Nature* **578**, 467–471 (2020).
77. C. M. Azzalin, J. Lingner, Telomere functions grounding on TERRA firma. *Trends Cell Biol.* **25**, 29–36 (2015).
78. F. Weissmann *et al.*, biGBac enables rapid gene assembly for the expression of large multisubunit protein complexes. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E2564–E2569 (2016).
79. D. G. Gibson *et al.*, Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
80. G. Tang *et al.*, EMAN2: An extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
81. J. Zivanov *et al.*, New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).
82. A. Meyer *et al.*, Systematic analysis of protein-detergent complexes applying dynamic light scattering to optimize solutions for crystallization trials. *Acta Crystallogr. F Struct. Biol. Commun.* **71**, 75–81 (2015).
83. H. Chi *et al.*, Comprehensive identification of peptides in tandem mass spectra using an efficient open search engine. *Nat. Biotechnol.* **36**, 1059–1061 (2018).
84. Z. L. Chen *et al.*, A high-speed search engine pLink 2 with systematic evaluation for proteome-scale identification of cross-linked peptides. *Nat. Commun.* **10**, 3404 (2019).
85. C. W. Combe, L. Fischer, J. Rappsilber, xiNET: Cross-link network maps with residue resolution. *Mol. Cell. Proteomics* **14**, 1137–1147 (2015).
86. P. D. B. Olinares, B. T. Chait, Native mass spectrometry analysis of affinity-captured endogenous yeast RNA exosome complexes. *Methods Mol. Biol.* **2062**, 357–382 (2020).
87. M. T. Marty, Eliminating artifacts in electrospray deconvolution with a softmax function. *J. Am. Soc. Mass Spectrom.* **30**, 2174–2177 (2019).
88. D. J. Reid *et al.*, MetaUniDec: High-throughput deconvolution of native mass spectra. *J. Am. Soc. Mass Spectrom.* **30**, 118–127 (2019).