

ARTICLE

DOI: 10.1038/s41467-018-04388-1

OPEN

# Modular ssDNA binding and inhibition of telomerase activity by designer PPR proteins

Henrik Spåhr<sup>1</sup>, Tionsun Chia<sup>2</sup>, James P. Lingford<sup>2</sup>, Stefan J. Siira<sup>2</sup>, Scott B. Cohen<sup>3</sup>, Aleksandra Filipovska<sup>2,4</sup> & Oliver Rackham<sup>2,4</sup>

DNA is typically found as a double helix, however it must be separated into single strands during all phases of DNA metabolism; including transcription, replication, recombination and repair. Although recent breakthroughs have enabled the design of modular RNA- and double-stranded DNA-binding proteins, there are currently no tools available to manipulate single-stranded DNA (ssDNA). Here we show that artificial pentatricopeptide repeat (PPR) proteins can be programmed for sequence-specific ssDNA binding. Interactions occur using the same code and specificity as for RNA binding. We solve the structures of DNA-bound and apo proteins revealing the basis for ssDNA binding and how hydrogen bond rearrangements enable the PPR structure to envelope its ssDNA target. Finally, we show that engineered PPRs can be designed to bind telomeric ssDNA and can block telomerase activity. The modular mode of ssDNA binding by PPR proteins provides tools to target ssDNA and to understand its importance in cells.

<sup>1</sup>Department of Mitochondrial Biology, Max Planck Institute for Biology of Ageing, D-50931 Cologne, Germany. <sup>2</sup>Harry Perkins Institute of Medical Research and Centre for Medical Research, The University of Western Australia, Nedlands, WA 6009, Australia. <sup>3</sup>Children's Medical Research Institute, University of Sydney, 214 Hawkesbury Road, Westmead, NSW 2145, Australia. <sup>4</sup>School of Molecular Sciences, The University of Western Australia, Crawley, WA 6009, Australia. These authors contributed equally: Henrik Spåhr, Tionsun Chia. Correspondence and requests for materials should be addressed to O.R. (email: [oliver.rackham@uwa.edu.au](mailto:oliver.rackham@uwa.edu.au))

DNA is predominantly found as a stable duplex in biological systems. However, accessing the genetic information stored in DNA for transcription, replication, recombination and repair requires the separation of DNA duplexes and exposure of single-stranded DNA (ssDNA). The presence of ssDNA also ensures the recruitment of the enzymatic activities required for these processes<sup>1</sup>. On the other hand, exposed ssDNA is particularly susceptible to both chemical and enzymatic damage and cells have a variety of mechanisms to protect ssDNA and ensure the rapid return of ssDNA to a duplex state<sup>2</sup>. If these mechanisms fail there are severe consequences for the cell, such as replication stalling that results in rapid increases in ssDNA. This exposed ssDNA acts as a marker of stress and activates cell signalling pathways to halt cell cycle progression and mobilise DNA repair processes and, if these fail, initiation of cell destruction via apoptosis<sup>3</sup>. Interestingly, despite its vulnerability, a large number of viruses and bacteriophages use ssDNA as the transmissible form of their genomes. The mutation rates of ssDNA viruses are extremely high and approach those of viruses with RNA genomes<sup>4</sup>. This might provide an evolutionary advantage, enabling a pool of viruses with variant genomes to be produced from each infected cell, some of which might have a selective advantage in subsequent infections. In another important biological process, bacterial conjugation, ssDNA plays an important role as it is the form by which genetic information is transferred to recipient cells<sup>5</sup>. In addition, ssDNA plays an important role in the maintenance and function of telomeres<sup>6</sup>.

Telomeres protect the ends of linear eukaryotic chromosomes from degradation and from fusion with other chromosomes<sup>6</sup>. The repetitive telomeric sequences consist of long double-stranded G-rich sequences followed by short, 50–150-nucleotide, single-stranded ends synthesised by the telomerase reverse transcriptase<sup>7–11</sup>. In humans, telomeres are usually protected by a six-protein complex, known as shelterin<sup>8, 12</sup>. Of the six shelterin proteins, only Pot1 binds single-stranded telomeric DNA<sup>13</sup>. Although robust technologies have now emerged that enable the site-specific manipulation of dsDNA and RNA in living cells<sup>14–16</sup>, the manipulation of ssDNA has not been possible to date. Because natural ssDNA-binding proteins use binding domains that recognise a combination of sequence and structure, for example, OB folds and RNA recognition motifs<sup>1</sup>, they cannot be easily repurposed to bind new ssDNA targets. Since ssDNA is very similar in structure to ssRNA we tested whether ssRNA-binding proteins that have been discovered to have modular recognition properties could be adapted for programmable ssDNA binding.

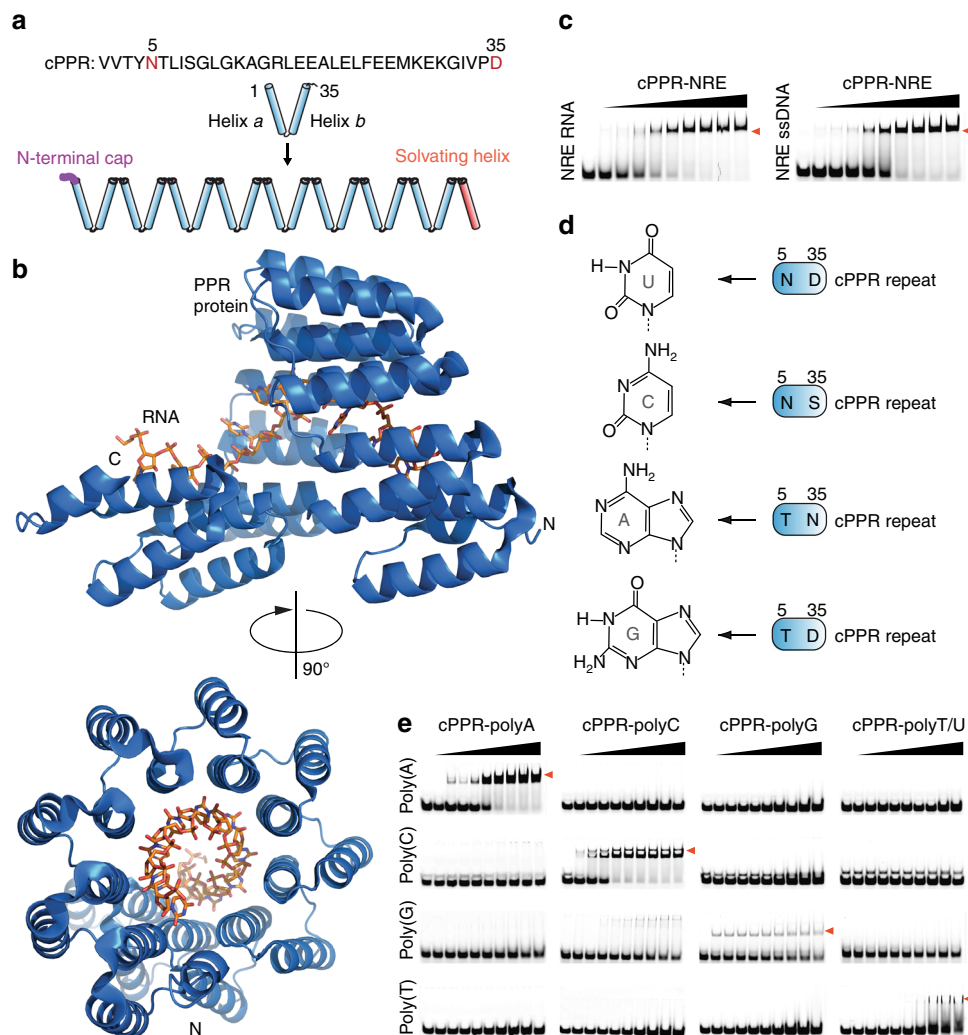
PPR proteins are RNA-binding proteins found predominantly in mitochondria and chloroplasts<sup>17</sup>. They regulate RNA processing, translation, stability and editing, and possess well-characterised, modular RNA recognition properties<sup>17–20</sup>. PPRs typically consist of tandem arrays of 2–30 degenerative repeats and each repeat usually consists of 35 amino acids. PPRs stack on each other to form an extended solenoid structure that can recognise RNA in a sequence-specific manner based on the identities of amino acids at positions 5 and 35 of each repeat<sup>21–24</sup>. Due to the highly insoluble nature of PPR proteins it has been difficult to understand their structural and RNA-binding properties. To bypass these limitations, we previously used consensus design to engineer artificial PPR domains (cPPRs) that are highly soluble and, via an appropriate choice of amino acids at position 5 and 35, can be designed to bind RNA in a predictable, sequence-specific manner<sup>21</sup>. These engineered PPR scaffolds enable the predictable binding of RNA targets and provide a starting point to use engineered PPR proteins to rationally manipulate cellular gene expression. In this study we have discovered that cPPRs bind not only RNA but also ssDNA in a modular and programmable

manner, analogous to their RNA-binding capability. Crystal structures of cPPRs with and without a ssDNA target provide an explanation for their ability to bind ssDNA and show how rearrangements of hydrogen bonds within and between protein repeats enable them to wrap around their targets. We re-design cPPRs to bind telomeric ssDNA and demonstrate that they can inhibit the activity of human telomerase *in vitro*. Unlike natural Pot1 that binds rigidly to its designated target, cPPRs can be programmed to bind any telomere sequence of interest. This opens up new opportunities to engineer artificial telomere-protective proteins and to understand fundamental aspects of telomere biology.

## Results

**Modular ssDNA binding by consensus PPR proteins.** Although PPRs are regarded as RNA-binding motifs, there is little information on their nucleic acid specificity. The plant PPR protein OTP87 was reported to bind ssRNA but not ssDNA, dsDNA or dsRNA<sup>25</sup>, and the plant THA8 protein was reported to bind ssRNA with >100-fold higher affinity than ssDNA<sup>26</sup>. However, detailed comparisons have been hampered because of PPR proteins' instability and insolubility. Because our synthetic consensus PPR (cPPR) proteins have very robust RNA-binding properties and stabilities<sup>21</sup>, we used these as a basis to examine nucleic acid-binding specificity. Typical cPPRs consist of eight identical synthetic repeats flanked by an N-terminal cap consisting of Met-Gly-Asn-Ser, because statistically, Gly, Asn and Ser have the highest propensities to occur at these N-terminal positions in a helices (Fig. 1a and Supplementary Fig. 1)<sup>27</sup>. In addition, a C-terminal solvating helix is added after the final consensus repeat to prevent unfolding, according to the successful consensus tetra-tripeptide repeat domain design of Main et al.<sup>28</sup>.

Synthetic PPR proteins fold into solenoid structures like natural PPR proteins<sup>21, 29</sup> and a recent study from Shen et al.<sup>24</sup> showed that they wrap around their RNA targets (Fig. 1b), however a molecular explanation for any potential discrimination between RNA and ssDNA is lacking. We initially tested a cPPR protein designed to bind a naturally occurring RNA sequence known as the Nanos response element (NRE). We found that this cPPR bound tightly to a ssDNA NRE oligonucleotide as well as to an RNA NRE oligonucleotide (Fig. 1c). Surprisingly, the affinity was only ~3-fold reduced between ssDNA and RNA (Fig. 1c and Supplementary Fig. 2). Based on bioinformatics predictions<sup>22, 23</sup> we previously identified particular amino acids at positions 5 and 35 of our cPPR that enable the specific recognition of individual RNA bases<sup>21</sup>. Although the NRE-targeting cPPR bound ssDNA, it was not clear if it bound the ssDNA specifically according to the PPR code. Therefore, we produced four different cPPR proteins designed to bind homopolymers of adenine, cytosine, guanine or uracil and tested their binding using ssDNA homopolymers (Fig. 1d). We found that the binding of ssDNA homopolymers conformed to the specific code for base recognition established previously for RNA (Fig. 1e). The binding of poly(A) ssDNA was slightly reduced compared to poly(A) RNA, while binding to poly(C) ssDNA was actually slightly better than for poly(C) RNA (Supplementary Fig. 2). Exact binding affinities for poly(G) ssDNA and RNA were difficult to determine, likely due to the propensity of poly(G) tracts to form G-quadruplex structures. Interestingly, binding to poly(T) ssDNA was greatly reduced compared to poly(U), possibly due to the additional methyl group on the thymine base relative to uracil. Therefore, the code for base recognition by PPRs is maintained when binding ssDNA targets, opening the way for the use of cPPRs to target ssDNAs in a programmable manner.



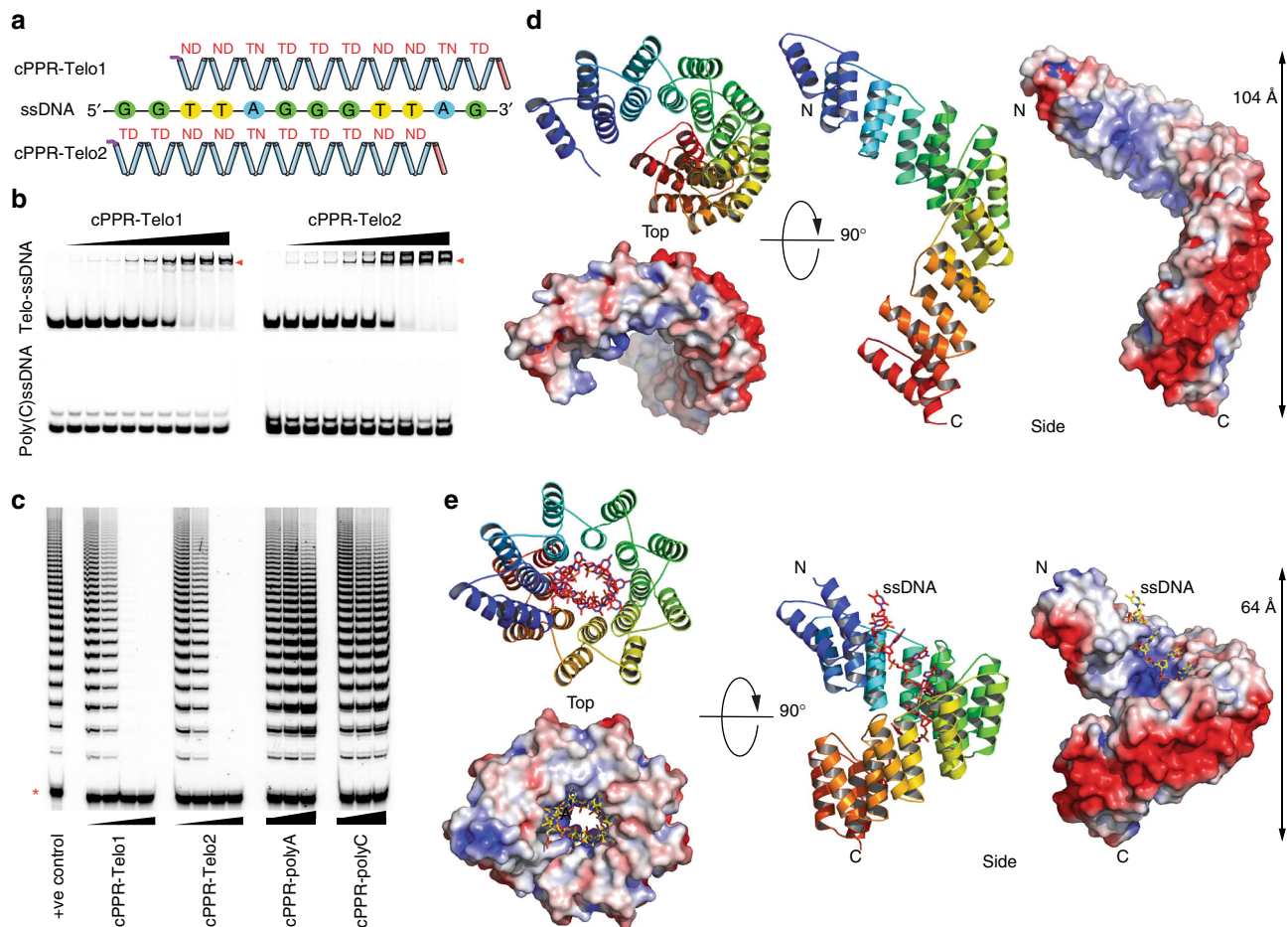
**Fig. 1** Consensus pentatricopeptide repeat (cPPR) proteins bind ssDNA. **a** The consensus PPR sequence and its assembly into a repeat protein flanked by stabilising elements. Annotation of the PPR sequence is based on the numbering scheme of Yin et al.<sup>34</sup>, where residues 5 and 35 are equivalent to residues 4 and ii according to Kobayashi et al.<sup>51</sup>, or 4 and 34 according to Fujii et al.<sup>52</sup>. **b** Crystal structure of a designer PPR protein bound to its RNA target. Structure determined by Shen et al.<sup>24</sup>. **c** Electrophoretic mobility shift assays (EMSA) of the cPPR-NRE using RNA and single-stranded DNA (ssDNA) Nanos response element (NRE) probes. Bound complexes are highlighted with red arrows. Protein concentrations used were, from left to right: 0, 0.15, 0.3, 0.5, 1, 2, 4, 6, 8 and 10  $\mu\text{M}$ . **d** The modular PPR code for RNA binding. Amino acids at positions 5 and 35 of each cPPR repeat recognise specific RNA bases. **e** Purified proteins were titrated against homopolymeric ssDNA probes in a DNA EMSA. Complexes formed between predicted cognate ssDNA-protein pairs are indicated with red arrows and demonstrate that high specificity of each cPPR protein for its cognate ssDNA target. Protein concentrations used were, from left to right: 0, 0.15, 0.3, 0.5, 1, 2, 4, 6, 8 and 10  $\mu\text{M}$

### Inhibition of human telomerase by designed cPPR proteins.

We examined if cPPRs could be designed to target biologically relevant ssDNA targets and focused on the repeating telomeric ssDNA sequences of mammalian cells. We designed two cPPRs targeting staggered sequences within the telomeric repeats (Fig. 2a and Supplementary Fig. 1b). In analogy to Pot1<sup>30</sup>, we choose to build cPPRs consisting of 10 repeats to match the length of the Pot1 recognition sequence. Both telomere-targeting cPPRs, dubbed cPPR-Telo1 and cPPR-Telo2, bound telomeric ssDNA sequences with high affinity and specificity, as determined by binding assays with off-target ssDNAs (Fig. 2b and Supplementary Fig. 3). Binding to ssDNA was robust in the presence of variable pH, salt concentrations and in the presence of non-specific competitor ssDNA (Supplementary Fig. 4). Furthermore, we compared the ability of cPPR-Telo1 and cPPR-Telo2 to bind ssDNA in G-quadruplex structures by comparing binding to two targets, where one could form a G-quadruplex structure and the other could not<sup>31</sup>. Both designed proteins could bind efficiently

independent of potential formation of a G-quadruplex (Supplementary Fig. 5), which is important since telomeric ssDNA is well known for its ability to form robust G-quadruplexes. We used the extension of telomeric ssDNA by telomerase as a model system to examine whether cPPRs could modulate ssDNA function. We performed direct telomerase extension assays<sup>32</sup> using human telomerase over-expressed and assembled in HEK-293T cells and examined whether the specific binding of the telomeric ssDNA by cPPR-Telo1 and cPPR-Telo2 could modulate telomerase activity. We found that both designed cPPRs effectively blocked telomerase activity in a dose-dependent manner, while control cPPR proteins had no effect (Fig. 2c). Furthermore, inhibition of telomerase extension was maintained even when non-specific competitor ssDNA was present in orders of magnitude excess (Supplementary Fig. 6).

The position of the Pot1-binding site relative to the 3'-end of the ssDNA template has been shown to have a dramatic effect on telomerase activity, with 3'-end binding blocking telomerase



**Fig. 2** cPPRs designed to bind to telomeric sequences. **a** Cartoon depicting the design and binding sites of cPPR-Telo1 and cPPR-Telo2. The amino acids at positions 5 and 35 of each repeat in the 10-repeat cPPRs are shown in red. The ssDNA probe is derived from the repeating telomeric DNA sequence and contains the binding sites for both cPPR-Telo1 and cPPR-Telo2. **b** ssDNA EMSA using Telo-ssDNA or poly(C) ssDNA (control). ssDNA-protein complexes are highlighted with red arrows. Protein concentrations used were, from left to right: 0, 0.15, 0.3, 0.5, 1, 2, 4, 6, 8 and 10  $\mu$ M. **c** Inhibition of telomerase extension by telomere-targeting cPPRs in telomerase activity assays with immunopurified telomerase and a primer incorporating 18 nt of telomeric sequence. Increasing concentrations of cPPR-Telo1 and cPPR-Telo2 (50, 100, 200 and 300 nM) were added to the indicated reactions and cPPR-polyA and cPPR-polyC (300 nM, 1  $\mu$ M and 3  $\mu$ M) were used as control proteins. The red asterisk indicates a 30-mer 5'-<sup>32</sup>P-labelled recovery/loading control. **d** Overall structure of cPPR-Telo1, shown in cartoon representation coloured in rainbow colours with blue at the N terminus and red at the C terminus. The surface charge distribution is coloured according to the local electrostatic potential (blue, +5 kT; red -5 kT). **e** Structure of cPPR-Telo1 in complex with DNA, which is shown as orange or yellow sticks

extension and internal binding stimulating the processivity of telomerase<sup>33</sup>. To examine whether these effects were an inherent property of Pot1 or resulted simply from the locations of protein binding we tested cPPR-Telo2's ability to modulate the extension of telomerase on a variety of different primers where the binding locations were at varied distances from the 3'-end of the ssDNA. We found that cPPR-Telo2 blocked telomerase activity regardless of its binding location, indicating that the multiple regulatory modes of Pot1 are idiosyncratic to that particular protein (Supplementary Fig. 7a). The potent inhibition of telomerase likely results from the cPPR's ability to block access of telomerase to ssDNA, since we found that cPPR-Telo2 could protect its ssDNA target from DNase I digestion (Supplementary Fig. 7b).

**Structure of a telomeric ssDNA-bound cPPR protein.** Next, we sought to obtain structural information on the two cPPR proteins specifically designed to bind to telomeric ssDNA sequences, to

understand further the cPPR scaffold's ability to bind both ssDNA and RNA. Although these proteins are highly similar in their primary sequences, only cPPR-Telo1 gave reproducible and diffracting crystals. We solved the structure of cPPR-Telo1 and cPPR-Telo1 bound to its ssDNA target using single-wavelength anomalous dispersion at 2.1 and 2.0 Å resolution, respectively (Table 1, with sample electron densities in Supplementary Fig. 8). The cPPR-Telo1 protein adopted an overall fold that closely mimicked other designed PPR structures<sup>21, 24, 29</sup> (Fig. 2d, e), where the two helices of each repeat, helix *a* and helix *b*, form a hairpin and these hairpins stack upon each other to form a right-handed superhelix. Mapping the electrostatic potential revealed a highly asymmetrical surface charge distribution with positively charged residues in the ssDNA-binding cavity and a negatively charged band along the exterior, exposed face of the superhelix (Fig. 2d, e). This charge distribution might provide the initial driving force for nucleic acids to bind the PPR proteins, based on their negatively charged phosphate backbones.

**Table 1** Data collection and refinement statistics

	apo cPPR-Telo1	cPPR-Telo1/ ssDNA
<i>Data collection</i>		
Space group	$P2_12_1$	$P4_22_2$
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	86.4, 87.1, 91.6	114.8, 114.8, 83.5
$\alpha$ , $\beta$ , $\gamma$ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	19.70–2.08 (2.15–2.08)	27.41–1.95 (2.02–1.95)
$R_{\text{merge}}$	0.111 (1.306)	0.0714 (1.823)
$CC_{1/2}$	0.998 (0.719)	0.999 (0.742)
$I/\sigma(I)$	9.92 (1.24)	30.79 (2.23)
Completeness (%)	96.62 (81.79)	99.21 (99.63)
Redundancy	10.5 (4.9)	27.1 (28.7)
<i>Refinement</i>		
Resolution (Å)	19.70–2.08 (2.13–2.08)	27.41–1.95 (2.0–1.95)
No. of reflections	40,891	40,921
$R_{\text{work}}/R_{\text{free}}$	0.240/0.280	0.189/0.228
No. of atoms		
Protein	2683	2785
DNA		210
Water	840	377
<i>B</i> -factors		
Protein	26.10	29.27
DNA		19.03
Water	44.88	41.41
R.m.s. deviations		
Bond lengths (Å)	0.014	0.010
Bond angles (°)	1.67	1.01
Ramachandran plot favoured/allowed (%) <sup>a</sup>	99.71/0.29	98.07/1.93
Clashscore <sup>a</sup>	2.93	3.17
Molprobrity score <sup>a</sup>	1.58	1.5

Statistics for the highest-resolution shell are shown in parentheses  
<sup>a</sup>According to the definition used in Molprobrity<sup>44</sup>

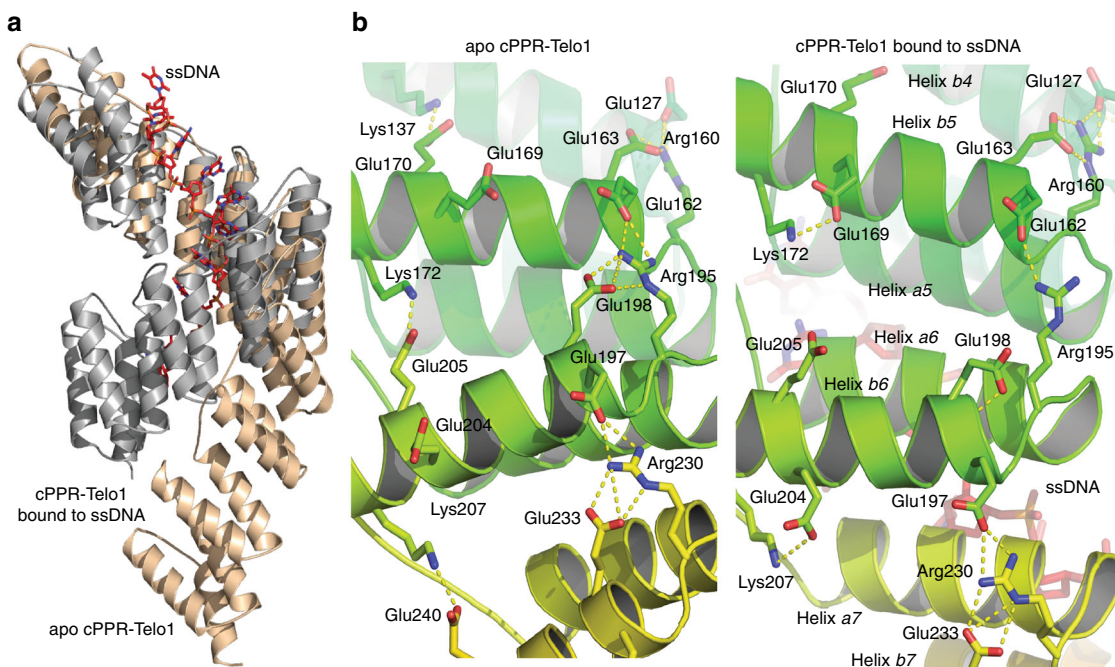
Our structures provide the first pair of structures where an identical PPR protein compacts upon binding a nucleic acid in the canonical, sequence-specific mode, providing an opportunity to understand how the conformation of the protein changes upon ligand binding. cPPR-Telo1 wraps around the ssDNA upon binding, inducing a massive compaction of the superhelical structure and decreasing the length of the protein from 104 to 64 Å (Figs. 2d, e and 3a). This conformational change more than halves the number of cPPR protein side chains exposed to water and may be the principal driving force favouring nucleic acid binding. Overall, the rearrangement is based on the *a* helices on the concave side of the protein moving into closer proximity as bonds are formed to the DNA with a concomitant separation of the *b* helices on the convex side. To allow the separation, an extensive hydrogen bond network between an arginine at position 16 and a glutamic acid at position 19 from one PPR repeat and a glutamic acid at position 18 from the preceding repeat in the apo protein rearranges at three out of nine possible positions leading to a more open arrangement when bound to DNA (Fig. 3b). Specifically, the side chains of Glu128, Glu198 and Glu303 form hydrogen bonds to their own main chain instead of to Arg125, Arg195 and Arg300, respectively. In addition, a hydrogen bond between a lysine at position 28 in each repeat and a glutamic acid at position 26 of the succeeding repeat switches to an intra-helix hydrogen bond interaction between this lysine and a nearby glutamic acid at position 25 (Fig. 3b). These and other changes in van der Waals interactions produce conformational variations

that are amplified over the entire cPPR, resulting in the notable compression of the protein superhelix.

**Interactions between cPPR-Telo1 and ssDNA.** Specific recognition of RNA by PPR proteins has been found to occur via hydrogen bonding between the side chains of residues at positions 5 and 35 of each PPR motif and the Watson–Crick faces of the nucleotide bases<sup>34</sup>. The amino acid at position 5 forms direct hydrogen bonds with the RNA base, while the amino acid at position 35 directly bonds with purine bases but makes a water-mediated interaction in the case of pyrimidines. We observed the same binding mode in our cPPR-ssDNA structure (Fig. 4) and likewise the bases were sandwiched between the valine residues at position 2 of each repeat (Fig. 4a). Furthermore, the phosphate backbone of the ssDNA was bound by lysine residues at position 13 of each repeat (Fig. 4a), as we<sup>21</sup> and others<sup>24</sup> had observed previously for RNA. The conserved modes of cPPR-base recognition between RNA and ssDNA are reflected in the cPPR's ability to bind both ssDNA and RNA (Fig. 1c, e). Interestingly, the hydrogen bond between the threonine at position 5 and adenine is rather long (3.6–3.9 Å), compared to when threonine in this position binds guanine (2.9–3.1 Å), which appears to be more ideal (Fig. 4c). This phenomenon is also true in the structures of designed PPR proteins in complex with RNA<sup>24</sup> but if another residue at position 5 might be more efficient in the binding of adenine, or whether this arrangement is advantageous for some reason remains to be determined.

In the case of uracil recognition in PPR-RNA complexes, the amide group of the asparagine side chain at position 5 donates a hydrogen bond to the O2 of the pyrimidine ring, while the carboxyl group of the aspartate makes a water-mediated hydrogen bond with the N3 of the pyrimidine<sup>34</sup>. An identical mode of thymine recognition is seen in our structure when ssDNA is bound and the methyl group at position C5 that distinguishes uracil and thymine is orientated away from the protein (Fig. 4a–c), such that it is initially hard to predict why poly(T) ssDNA is bound much less efficiently than poly(U) RNA, based on this difference. However, the C5 methyl groups likely alter the hydration pattern of the backbone and could also alter the stacking energies between bases, and thereby affect the dynamics of the RNA–protein interaction<sup>35</sup>. Support for this hypothesis comes from the observation that although the electron density is very clear for all nucleotides, it is weaker for one of the thymines (T7) in the structure (Supplementary Fig. 9). T7 sits adjacent to another thymine (T8) and the altered stacking energy between these bases might contribute to the observed flexibility at that position. Although tolerated well, or to some extent, when there is an isolated thymine or a pair of thymines in the ssDNA target, respectively, it could be that the additive disruption in stacking energies causes poly(T) tracts to bind poorly to cPPR proteins (as seen in Fig. 1e and Supplementary Fig. 2).

Comparing the recognition of bases by each of the 10 PPRs revealed that the hydrogen bonding was very similar within each type of base. This included the long hydrogen bond between the threonine at position 5 and adenine (Fig. 5a), the compact hydrogen bonding network recognising guanine (Fig. 5b) and the water-mediated hydrogen bonds recognising thymines (Fig. 5c). One interesting exception is the thymine at position 1, which does not make any specific hydrogen bonds with the protein in our structure. To further examine this observation and the specificity of base recognition at all positions in cPPR-Telo1 we performed a comprehensive Bind-n-Seq analysis using a randomised ssDNA target sequence (Fig. 5d, e). We found significant enrichment of the predicted base at each position in the selected library, further demonstrating the specificity of cPPRs for ssDNA, however the



**Fig. 3** An altered hydrogen bond network enables compaction of PPRs upon ssDNA binding. **a** Structural comparison of apo cPPR-Telo1 (wheat) and in complex with ssDNA (grey). **b** Altered hydrogen bond network between *b* helices of cPPR-Telo1 upon ssDNA binding. Amino acids participating in hydrogen bond rearrangements are shown as sticks and hydrogen bonds are shown with dashes. Arg160, Arg195 and Arg230 correspond to Arg16 in the standardised annotation of PPR motifs based on the numbering scheme of Yin et al.<sup>34</sup>; Glu162 and Glu197 to Glu18; Glu163, Glu198 and Glu233 to Glu19; Glu169 and Glu204 to Glu25; Lys137, Lys172 and Lys 207 to Lys28; and Glu170, Glu205 and Glu240 to Glu29

enrichment of thymine at position 1 was less than the other positions. Therefore, the thymine at position 1 might be in a dynamic equilibrium between bound and unbound states. A recent examination of the binding of RNA to designer PPR proteins found that the terminal positions in the target nucleic acids made the least contribution to the binding affinities of these complexes<sup>36</sup>. Our data confirm this observation in the context of ssDNA and also provide structural evidence for this effect.

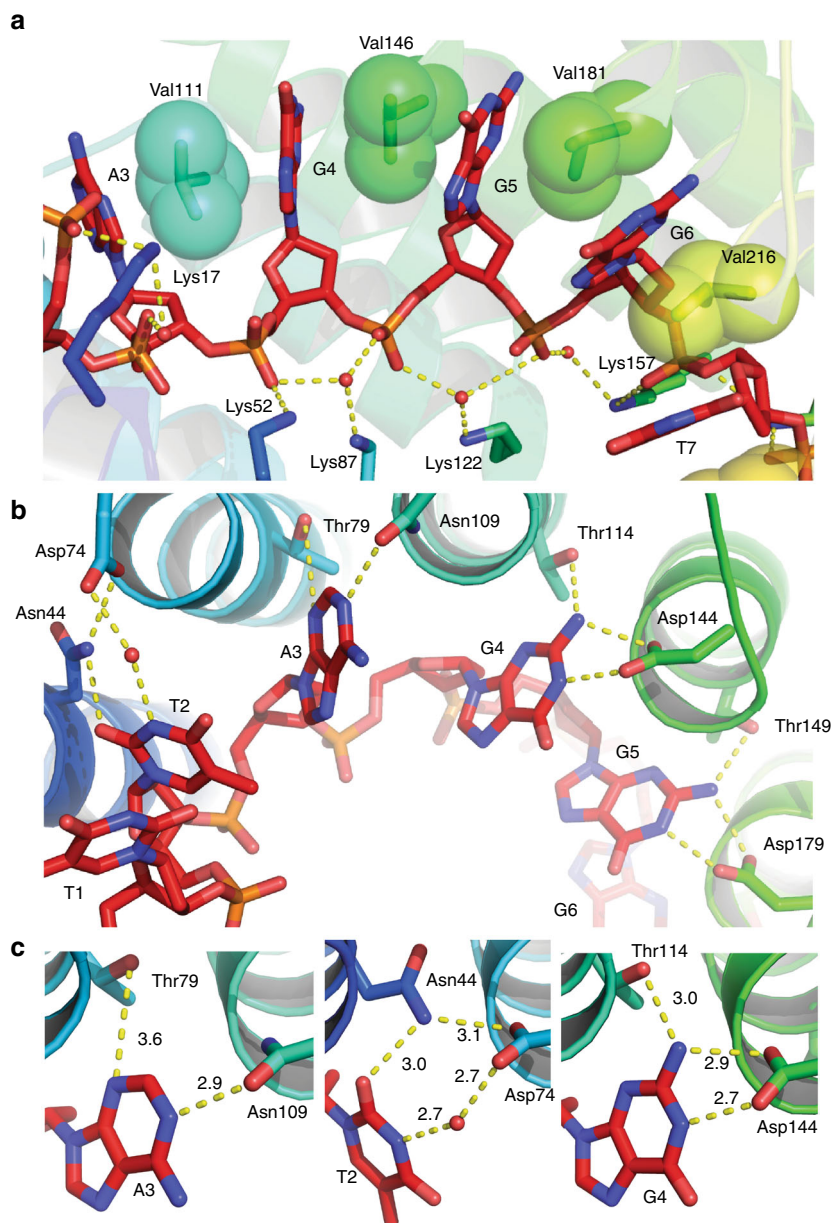
## Discussion

Here we found that designed PPR proteins can bind ssDNA in a sequence-specific manner analogous to how they bind RNA. Furthermore, the modularity of the cPPR proteins enabled the design of cPPRs that could target telomeric ssDNA and block extension by human telomerase. The ability of cPPRs to target ssDNA begs the question: do natural PPR proteins generally discriminate between RNA and DNA? The 2'-OH groups in the bound RNA are oriented towards helix *a* and discrimination between ssDNA and RNA might be achieved by interactions with the side chains of residues at positions 6 and 9 of helix *a*. However, identifying the exact residues involved is complicated by the fact that PPR arrays compress significantly upon RNA binding<sup>34, 37</sup>. This reorientates the helices so different residues might be involved in initial recognition of the backbone sugar, compared to those involved in later accommodation within the PPR solenoid. The molecular basis for potential discrimination between ssRNA and ssDNA by the wide variety of natural PPR proteins is not clear. The THA8 PPR protein binds its RNA target as a dimer and only a central G is recognised according to the PPR code by a PPR repeat from one of the THA8 subunits<sup>26</sup>. A direct hydrogen bond between an arginine residue at position 8 of that repeat and the 2'-OH group is required for high-affinity RNA binding. The ribose group of the adjacent A nucleotide is also recognised by THA8 but via a water-mediated hydrogen bond with an arginine of the other THA8 protein in the dimer.

The PPR10 protein has 19 PPRs and 3 of these interact with the 2'-OH ribose groups in its target RNA: repeats 3 and 5 make direct hydrogen bonds via arginine and serine amino acids at positions 2 and 5, respectively, while repeat 15 makes a water-mediated hydrogen bond, via a serine at position 9<sup>34</sup>. Consistent with these observations we found that recombinant PPR10 did not robustly bind ssDNA, in comparison to RNA (Supplementary Fig. 10). Therefore, the mode of recognition of ribose 2'-OH groups by different PPR proteins appears to be idiosyncratic and the specificity for RNA in natural proteins might be wiped clean in the consensus design of cPPRs, resulting in their ability to bind both RNA and ssDNA with high affinity.

In nature RNA-binding proteins do not often need to discriminate against ssDNA<sup>1</sup>. First, because RNAs are usually orders of magnitude more abundant than the genes that encode them in cells, the DNA can be simply outcompeted by RNA. Second, ssDNA is sequestered in cells by non-specific ssDNA-binding proteins, such as replication protein A (RPA)<sup>2</sup>, limiting the access of RNA-binding proteins. Third, in eukaryotic cells RNA-binding proteins are often targeted to the locations of their RNA targets, such as the cytoplasm, mitochondria and the nucleoli, which are separated from potential ssDNA targets in the nucleus<sup>1</sup>. The cPPR scaffold provides high-affinity, sequence-specific binding to aid in the goal of developing designer ssDNA-binding proteins, and our current study sets the stage for further engineering of cPPRs to discriminate against RNA, to more efficiently target ssDNA in cells. Furthermore, if combined with protein signals to efficiently target cPPRs to regions of the cell containing their ssDNA targets, these proteins could help reveal the many aspects of cell biology involving ssDNA, which are currently understudied due to a lack of appropriate tools.

In this study we provide proof of principle that cPPRs can be used to target and manipulate telomeric ssDNA *in vitro*. While further work will be required to explore the applications of these proteins in cellular systems, one attractive area in ssDNA biology



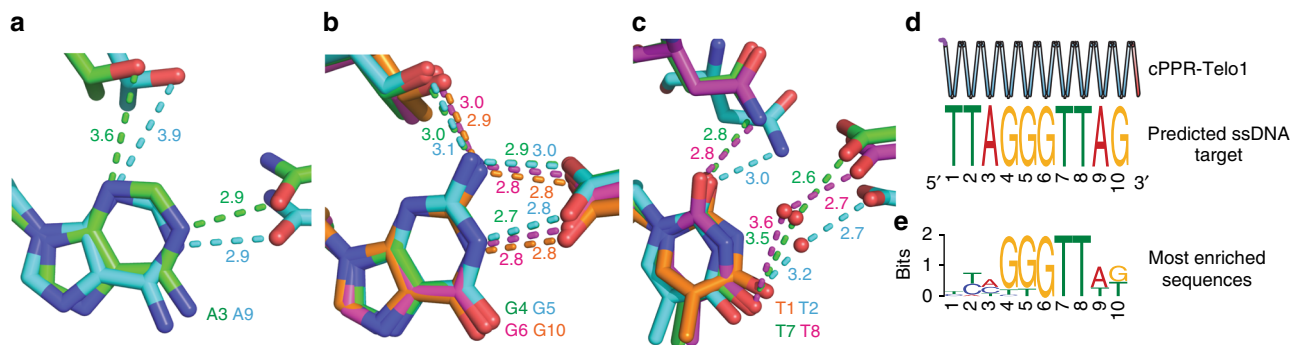
**Fig. 4** Protein-ssDNA interactions of cPPR-Telo1. **a** Non-specific protein-DNA interactions. Stacking valines are shown as sticks and spheres, and interacting polar residues and ssDNA are shown as sticks. Hydrogen bonds are shown with dashes. **b** Specific protein-ssDNA interactions. **c** Nucleotide base recognition by cPPR-Telo1 according to the PPR code. Interacting polar residues of cPPR-Telo1 and adenine (A), thymine (T) and guanine (G) are shown as sticks. Hydrogen bonds are shown with dashes and distances in Å are indicated

is to manipulate telomerase and telomeres in general. Telomerase uses a unique mode of primer extension, whereby an internal RNA template is used to specify the sequence of the added repeat<sup>10</sup>. Because this template must be re-used for each repeat, the product-template duplex must dissociate after each addition, providing ample opportunity for the engineered cPPRs to access the telomeric ssDNA. Human telomerase is proposed to extend telomeres by 5–10 repeats per cell division<sup>38</sup>. This may result from the complex competition between telomerase, POT1, the general ssDNA-binding protein RPA, and other shelterin components<sup>39</sup>. The addition of a telomeric sequence-specific cPPR in cells could be useful to skew this competition by blocking telomerase access and preventing telomere extension—to study telomere biology and as potential cancer therapeutics in the future. Overall, the manipulation of ssDNA has been neglected in biological research to date and the telomere is just one of many areas

where new tools will be very valuable. In this study we have shown that we have the ability to design proteins that recognise specific ssDNA sequences of interest, with many potential applications in biology and biotechnology.

## Methods

**Design and synthesis of cPPR coding sequences.** cPPR-polyA, cPPR-polyC, cPPR-polyG and cPPR-polyU were designed as described in Coquille et al.<sup>21</sup> and shown in Supplementary Fig. 1a. cPPR-Telo1 and cPPR-Telo2 were designed based on 10 tandem repeats with amino acids at position 5 and 35 chosen according to the recognition code of cPPRs described in this article. N-terminal cap residues (Met-Gly-Asn-Ser) and a C-terminal solvating helix (Val-Thr-Tyr-Thr-Thr-Leu-Ile-Ser-Gly-Leu-Gly-Lys-Ala-Gly) were added to the final design (see Supplementary Fig. 1b for a detailed example). Synthetic genes encoding the final cPPR design were optimised for expression in *Escherichia coli* and synthesised from overlapping oligonucleotides (GeneArt and GenScript).



**Fig. 5** Recognition of ssDNA bases across the length of the artificial PPR protein. **a** Recognition of adenines at positions 3 and 9. **b** Recognition of guanines at positions 4, 5, 6 and 10. **c** Recognition of thymines at positions 1, 2, 7 and 8. Hydrogen bonds are shown with dashes and distances in Å are indicated. **d** Bind-n-Seq analysis of cPPR-Telo1 protein ssDNA-binding specificity, compared to its designed target. **e** A logo representing the most enriched sequences (1127 in total, all with  $E$ -values  $> 8.6 \times 10^{-18}$ ) found in the cPPR-Telo1-bound ssDNA library, as determined by discriminative regular expression motif elicitation (DREME)<sup>50</sup>. No significantly enriched motifs were found in a matched unselected library

**Protein purification.** Coding sequences for cPPR-Telo1, cPPR-Telo2 and enhanced green fluorescent protein (EGFP) were subcloned into pETM30 and expressed as fusions to glutathione *S*-transferase and His tags in *E. coli* 2566 cells (New England Biolabs). Cells were lysed by sonication in lysis buffer (50 mM Tris-HCl, pH 8.0, 0.3 M NaCl and 5 mM imidazole). Lysates were clarified by centrifugation and incubated with His Select Beads (Sigma) for 30 min with gentle rocking at 4 °C. Beads were washed twice with wash buffer (same as lysis buffer but with 10 mM imidazole) and transferred into a Poly-Prep Chromatography Column (Bio-Rad). Beads were washed twice with wash buffer and proteins were eluted in elution buffer (same as lysis buffer but with 250 mM imidazole). Purified proteins were dialysed using DiaEasy Dialyzer (BioVision) in dialysis buffer (25 mM Tris, pH 7.4, 0.2 M NaCl, 0.5 mM EDTA and 2 mM dithiothreitol (DTT)) overnight at 4 °C. Protein concentration was determined by the bicinchoninic acid assay using bovine serum albumin (BSA) as a standard. cPPR-polyA, cPPR-polyC, cPPR-polyG, cPPR-polyU/T and cPPR-NRE were purified as described by Coquelle et al.<sup>21</sup>. Protein used for crystallisation was essentially purified as above with the following modifications: during dialysis, the fusion tag was removed by tobacco etch virus at a protease:protein molar ratio of 1:50, followed by purification on a Superdex 200 10/300 column in dialysis buffer and concentrated using Vivaspinn concentrators (10 000 Da molecular weight cutoff, GE).

**Crystallisation and structure determination.** Crystals of cPPR-Telo1 were grown at 23 °C by the sitting drop vapour diffusion method by mixing 1 µl protein (5 mg/ml) with an equal volume of reservoir solution (100 mM sodium acetate (pH 4.6), 20 mM calcium chloride and 40% MPD). In addition, cPPR-Telo1 was incubated with target ssDNA oligonucleotide (TTAGGGTTAG) at a protein/DNA molar ratio of 1:2 for 30 min at 4 °C and crystallised in 100 mM sodium acetate (pH 4.6), 20 mM calcium chloride and 22 % MPD. For both cPPR-Telo1 and cPPR-Telo1/ssDNA a single-wavelength anomalous dispersion data set was collected at the K-edge of selenium at beamline ID30A-3 (ESRF, Grenoble, France). The X-ray diffraction data (Table 1) were processed with X-ray diffraction spectroscopy<sup>40</sup> and the structure was solved with PHENIX<sup>41</sup> and autoSHARP<sup>42</sup>. An initial model built by Buccaneer<sup>43</sup> or Autobuild<sup>41</sup> was used as a starting model for manual building in COOT<sup>44</sup> interspersed with refinement in Buster (version 2.10.3)<sup>45</sup>, which rendered a final model (Table 1) that had no Ramachandran outliers as assessed by MOLPROBITY<sup>46</sup>. Representative portions of the electron densities of the final cPPR-Telo1 and cPPR-Telo1/ssDNA models are shown as stereo images in Supplementary Fig. 8. Figures were prepared with PyMOL (Schrodinger) and structural alignments were performed using the secondary-structure matching method.

**Electrophoretic mobility shift assays.** Purified proteins were incubated at room temperature for 30 min with 0.83 µM fluorescein-labelled ssDNA or RNA oligonucleotides (Dharmacon) in EMSA buffer (10 mM HEPES, pH 8.0, 1 mM EDTA, 50 mM KCl, 2 mM DTT, 0.1 mg/ml fatty acid-free BSA and 0.02% Tween-20). Sequences for probes are listed below:

NRE RNA: 5'-(Fl)ArUrUrGrUrArUrArUrA-3';  
 NRE: 5'-(Fl)AUUGUAUUAU-3';  
 Poly(A) RNA: 5'-(Fl)rArArArArArArArA-3';  
 Poly(C) RNA: 5'-(Fl)rCrCrCrCrCrCrCrC-3';  
 Poly(G) RNA: 5'-(Fl)rGrGrGrGrGrGrGrG-3';  
 Poly(U) RNA: 5'-(Fl)rUrUrUrUrUrUrUrU-3';  
 Poly(A): 5'-(Fl)AAAAAAAAA-3';  
 Poly(C): 5'-(Fl)CCCCCCCCC-3';  
 Poly(G): 5'-(Fl)GGGGGGGGG-3';  
 Poly(T): 5'-(Fl)TTTTTTTTT-3';

Telo-ssDNA: 5'-(Fl)GGTTAGGGTTAG-3';  
 Telo-ssDNA 1 mismatch: 5'-(Fl)GGTTAGGGTTAG-3';  
 Telo-ssDNA 3 mismatches: 5'-(Fl)GGTTACCCTTAG-3';  
 Telo-ssDNA 5 mismatches: 5'-(Fl)GGTTCCCCCTAG-3';  
 Primer GG-a: 5'-(Fl)GGTTAGGGTTAGGGTTAGGG-3';  
 Primer a: 5'-(Fl)TTAGGGTTAGGGTTAGGG-3', where *r* designates a ribonucleotide and (Fl) designates fluorescein followed by an 18-atom hexa-ethyleneglycol spacer. A non-specific competitor ssDNA derived from the 15 nt spacer DNA of the telomerase extension assay substrates with sequence 5'-CTAGACCTGTCATCA-3' was added where indicated. Reactions were analysed by 10% PAGE in TAE and fluorescence was detected using Typhoon FLA 9500 Biomolecular Imager (GE). Assays were replicated at least three times.

**Immunopurification of telomerase.** Purification of human telomerase was performed as described in Tomlinson et al.<sup>47</sup>. Briefly, an affinity purified sheep polyclonal anti-hTERT antibody (20 µg) was added per ml of cleared lysate from HEK-293T cells and rotated for 1 h at 4 °C. Protein G-agarose beads (50 µl of a 50% vol/vol slurry per ml lysate) were added and the suspension rotated for 1 h at 4 °C. The protein G suspension was collected in a 15 mm ID × 200 mm L fritted glass column (Bio-Rad), and washed with five column volumes of lysis buffer without DTT and phenylmethylsulfonyl fluoride. To elute the purified enzyme, 100 µl of 1 µM antigenic peptide (ARPAEEATSLEGALSGRH) in storage buffer (20 mM HEPES-KOH (pH 7.9), 2 mM MgCl<sub>2</sub>, 300 mM KCl, 1 mM DTT, 10% v/v glycerol and 0.1% v/v Triton X-100) was added per ml of lysate and incubated with shaking at room temperature for 1 h. The eluate was collected into a Protein Lo-Bind tube (Eppendorf), snap-frozen in liquid nitrogen and stored at -80 °C. The yield of immunopurified telomerase enzyme complex was determined by dot-blotting against hTR, using the DNA oligonucleotide probe 5'-32P-CGGTGAAGGCGGCAGGCCGAGGC-3' for detection<sup>32, 47</sup>.

**Direct primer extension activity assay.** Telomerase activity assays were performed as described in Tomlinson et al.<sup>47</sup>. Briefly, immunopurified telomerase was incubated alone or in combination with cPPR-Telo1, cPPR-Telo2 or EGFP proteins. Telomerase extension assay was performed at 37 °C for 4 h in 20 mM HEPES-KOH buffer, pH 8, 150 mM KCl, 2 mM MgCl<sub>2</sub>, 0.1% vol/vol Triton X-100, 1 mM DTT, 100 nM primer, 1 mM dTTP, 1 mM dATP, 10 µM dGTP and 20 µCi [α-32P] dGTP (PerkinElmer). The following primers were used:

Telo 14 5'-biotin-CTAGACCTGTCATCATTAGGGTTAGGGTT-3';  
 Telo 16 5'-biotin-CTAGACCTGTCATCATTAGGGTTAGGGTTAG-3';  
 Telo 18 5'-biotin-CTAGACCTGTCATCATTAGGGTTAGGGTTAGGG-3';  
 Telo 20 5'-biotin-CTAGACCTGTCATCATTAGGGTTAGGGTTAGGGTT-3';

Telomerase extension was quenched with addition of 200 µl of stop buffer (500 mM NaCl, 10 mM EDTA, 0.1% vol/vol Triton X-100) and transferred to microspin columns containing 40 µl Ultralink Neutravidin Plus beads slurry (Thermo Fisher) and rotated at room temperature for 2 h. Non-biotinylated product was removed by centrifugation at 2000 × *g* for 10 s and the beads were washed three times in 500 µl of stop buffer with rotation for 10 min at room temperature followed by centrifugation at 2000 × *g* for 10 s. Residual stop buffer was removed with centrifugation at 2000 × *g* for 10 s. Telomerase extension products were released by addition of 30 µl of denaturing formamide buffer (9 volumes of 90% vol/vol formamide, 10% TBE (90 mM Tris base, 90 mM boric acid, 1 mM EDTA), 0.01% wt/vol xylene cyanol and 0.01% wt/vol bromophenol blue to 1 volume buffer TE containing 5 mM biotin) heating at 95 °C for 10 min. A volume of 5 µl of each DNA sample was analysed by electrophoresis on a DNA sequencing gel run at a constant 75 W for 1 h. The gel was transferred to filter paper, dried for 30 min at



80 °C, exposed to a PhosphorImager screen (GE) and visualised on a Typhoon FLA 9500 scanner.

**Bind-n-Seq analysis of protein-ssDNA specificity.** Target sequences of cPPR-Telo1 were determined using a modified version of Bind-n-Seq<sup>48</sup>. An aliquot of 100 nM of purified cPPR-Telo1 was combined with 50 µM ssDNA library oligonucleotide with the sequence: 5'-CTTTATCCAGCCCTCACNNNNNNNNNNNC-TATAGTGTACCTAAATC-3' for 1 h in EMSA buffer (10 mM HEPES, pH 8.0, 1 mM EDTA, 50 mM KCl, 2 mM DTT, 0.1 mg/ml fatty acid-free BSA and 0.02% Tween-20) with gentle rocking at room temperature. An aliquot was removed to serve as the unselected library. His Select Beads (Sigma) were added and the sample was incubated for 30 min at room temperature. Complexes were washed three times with EMSA buffer for 10 min, once with wash buffer (50 mM Tris-HCl, pH 8.0, 0.3 M NaCl and 10 mM imidazole) and proteins were eluted in elution buffer (same as wash buffer but with 250 mM imidazole) for 10 min at room temperature. DNA was purified using Oligo Clean & Concentrator columns following the manufacturer's instructions (Zymo) and amplified by PCR using the following primers (with Illumina adaptor sequences shown in lower case): 5'-tcgtcggcagctcagatgtgtataagagacagCTTTATCCAGCCCTCAC-3' and 5'-gtctcgtggcctcggagatgtgtataagagacagGATTAGGTGACACTATAG-3'. PCR products were purified using Oligo Clean & Concentrator columns, following the manufacturer's instructions (Zymo), indexed and sequenced by the Australian Genomic Research Facility (Perth, Australia) on an Illumina MiSeq, according to the manufacturer's instructions. Paired-end reads were merged with bbmerge.sh (mininsert = 0 mininsert0 = 0) from the BBTools suite v37.02 (<https://jgi.doe.gov/data-and-tools/bbtools/>), removing the adapters and the merged reads were trimmed of 5' and 3' flanking sequences with cutadapt v1.1.4<sup>49</sup> (-g ^CTTTATCCAGCCCTCAC -a CTATAGTGTACCTAAATC\$ -n 2). Any trimmed reads not 10 nt in length were excluded using reformat.sh from BBTools (minlength = 10 maxlength = 10) and these sequences searched for 10-nt motifs enriched relative to the control with DREME<sup>50</sup> (-k 10 -dna -norc).

**Data availability.** Crystallography data sets are available at the RCSB Protein Data Bank with accession numbers [5ORM](#) (cPPR-Telo1) and [5ORQ](#) (cPPR-Telo1/ssDNA). The nucleotide sequence for cPPR-Telo1 is available at the NCBI GenBank database with accession code MH247127.

Received: 12 February 2018 Accepted: 25 April 2018

Published online: 07 June 2018

## References

- Dickey, T. H., Altschuler, S. E. & Wuttke, D. S. Single-stranded DNA-binding proteins: multiple domains for multiple functions. *Structure* **21**, 1074–1084 (2013).
- Ashton, N. W. et al. Human single-stranded DNA binding proteins are essential for maintaining genomic stability. *BMC Mol. Biol.* **14**, 9 (2013).
- Zeman, M. K. & Cimprich, K. A. Causes and consequences of replication stress. *Nat. Cell Biol.* **16**, 2–9 (2014).
- Sanjuan, R. Mutational fitness effects in RNA and single-stranded DNA viruses: common patterns revealed by site-directed mutagenesis studies. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **365**, 1975–1982 (2010).
- Chen, I., Christie, P. J. & Dubnau, D. The ins and outs of DNA transfer in bacteria. *Science* **310**, 1456–1460 (2005).
- Bryan, T. M. & Cech, T. R. Telomerase and the maintenance of chromosome ends. *Curr. Opin. Cell Biol.* **11**, 318–324 (1999).
- Blackburn, E. H. Telomeres and telomerase: the means to the end (Nobel lecture). *Angew. Chem. Int. Ed. Engl.* **49**, 7405–7421 (2010).
- Armstrong, C. A. & Tomita, K. Fundamental mechanisms of telomerase action in yeasts and mammals: understanding telomeres and telomerase in cancer cells. *Open Biol.* **7**, 160338 (2017).
- Moyzis, R. K. et al. A highly conserved repetitive DNA sequence, (TTAGGG)<sub>n</sub>, present at the telomeres of human chromosomes. *Proc. Natl Acad. Sci. USA* **85**, 6622–6626 (1988).
- Greider, C. W. & Blackburn, E. H. A telomeric sequence in the RNA of Tetrahymena telomerase required for telomere repeat synthesis. *Nature* **337**, 331–337 (1989).
- Greider, C. W. & Blackburn, E. H. Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. *Cell* **43**, 405–413 (1985).
- Palm, W. & de Lange, T. How shelterin protects mammalian telomeres. *Annu. Rev. Genet.* **42**, 301–334 (2008).
- Baumann, P. & Cech, T. R. Pot1, the putative telomere end-binding protein in fission yeast and humans. *Science* **292**, 1171–1175 (2001).
- Wright, A. V., Nunez, J. K. & DouDNA, J. A. Biology and applications of CRISPR systems: harnessing nature's toolbox for genome engineering. *Cell* **164**, 29–44 (2016).
- Filipovska, A. & Rackham, O. Designer RNA-binding proteins: new tools for manipulating the transcriptome. *RNA Biol.* **8**, 978–983 (2011).
- Hsu, P. D., Lander, E. S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell Rep.* **23**, 1262–1278 (2014).
- Filipovska, A. & Rackham, O. Pentatricopeptide repeats: modular blocks for building RNA-binding proteins. *RNA Biol.* **10**, 1426–1432 (2013).
- Perks, K. L. et al. PTC1 is required for 16S rRNA maturation complex stability and mitochondrial ribosome assembly. *Cell Rep.* **23**, 127–142 (2018).
- Siira, S. J. et al. LRPPRC-mediated folding of the mitochondrial transcriptome. *Nat. Commun.* **8**, 1532 (2017).
- Rackham, O. et al. Hierarchical RNA processing is required for mitochondrial ribosome assembly. *Cell Rep.* **16**, 1874–1890 (2016).
- Coquille, S. et al. An artificial PPR scaffold for programmable RNA recognition. *Nat. Commun.* **5**, 5729 (2014).
- Yagi, Y. et al. Elucidation of the RNA recognition code for pentatricopeptide repeat proteins involved in organelle RNA editing in plants. *PLoS ONE* **8**, e57286 (2013).
- Barkan, A. et al. A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. *PLoS Genet.* **8**, e1002910 (2012).
- Shen, C. et al. Structural basis for specific single-stranded RNA recognition by designer pentatricopeptide repeat proteins. *Nat. Commun.* **7**, 11285 (2016).
- Hammani, K. et al. The pentatricopeptide repeat protein OTP87 is essential for RNA editing of nad7 and atp1 transcripts in Arabidopsis mitochondria. *J. Biol. Chem.* **286**, 21361–21371 (2011).
- Ke, J. et al. Structural basis for RNA recognition by a dimeric PPR-protein complex. *Nat. Struct. Mol. Biol.* **20**, 1377–1382 (2013).
- Richardson, J. S. & Richardson, D. C. Amino acid preferences for specific locations at the ends of alpha helices. *Science* **240**, 1648–1652 (1988).
- Main, E. R. et al. Design of stable alpha-helical arrays from an idealized TPR motif. *Structure* **11**, 497–508 (2003).
- Gully, B. S. et al. The design and structural characterization of a synthetic pentatricopeptide repeat protein. *Acta Crystallogr. D Biol. Crystallogr.* **71**, 196–208 (2015).
- Lei, M., Podell, E. R. & Cech, T. R. 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).
- Zaug, A. J., Podell, E. R. & Cech, T. R. Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro. *Proc. Natl Acad. Sci. USA* **102**, 10864–10869 (2005).
- Cohen, S. B. & Reddel, R. R. A sensitive direct human telomerase activity assay. *Nat. Methods* **5**, 355–360 (2008).
- Wang, F. et al. The POT1-TPP1 telomere complex is a telomerase processivity factor. *Nature* **445**, 506–510 (2007).
- Yin, P. et al. Structural basis for the modular recognition of single-stranded RNA by PPR proteins. *Nature* **504**, 168–171 (2013).
- Warmlander, S., Sponer, J. E., Sponer, J. & Leijon, M. The influence of the thymine C5 methyl group on spontaneous base pair breathing in DNA. *J. Biol. Chem.* **277**, 28491–28497 (2002).
- Miranda, R. G., McDermott, J. J. & Barkan, A. RNA-binding specificity landscapes of designer pentatricopeptide repeat proteins elucidate principles of PPR-RNA interactions. *Nucleic Acids Res.* **46**, 2613–2623 (2018).
- Gully, B. S. et al. The solution structure of the pentatricopeptide repeat protein PPR10 upon binding atpH RNA. *Nucleic Acids Res.* **43**, 1918–1926 (2015).
- Zhao, Y. et al. Processive and distributive extension of human telomeres by telomerase under homeostatic and nonequilibrium conditions. *Mol. Cell* **42**, 297–307 (2011).
- Rubtsova, M. P. et al. Replication protein A modulates the activity of human telomerase in vitro. *Biochemistry (Mosc.)* **74**, 92–96 (2009).
- Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26**, 795–800 (1993).
- Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
- Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. Automated structure solution with autoSHARP. *Methods Mol. Biol.* **364**, 215–230 (2007).
- Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 1002–1011 (2006).
- Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
- Smart, O. S. et al. Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 368–380 (2012).
- Lovell, S. C. et al. Structure validation by Calpha geometry: phi, psi and Cbeta deviation. *Proteins* **50**, 437–450 (2003).
- Tomlinson, C. G. et al. Quantitative assays for measuring human telomerase activity and DNA binding properties. *Methods* **114**, 85–95 (2017).

48. Zykovich, A., Korf, I. & Segal, D. J. Bind-n-Seq: high-throughput analysis of in vitro protein-DNA interactions using massively parallel sequencing. *Nucleic Acids Res.* **37**, e151 (2009).
49. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12 (2011).
50. Bailey, T. L. DREME: motif discovery in transcription factor ChIP-seq data. *Bioinformatics* **27**, 1653–1659 (2011).
51. Kobayashi, K. et al. Identification and characterization of the RNA binding surface of the pentatricopeptide repeat protein. *Nucleic Acids Res.* **40**, 2712–2723 (2012).
52. Fujii, S., Bond, C. S. & Small, I. D. Selection patterns on restorer-like genes reveal a conflict between nuclear and mitochondrial genomes throughout angiosperm evolution. *Proc. Natl Acad. Sci. USA* **108**, 1723–1728 (2011).

### Acknowledgements

We thank Stéphane Thore for advice and Anima Poudyal for technical assistance. This work was supported by fellowships, scholarships and grants from the Australian Research Council (DP180101656, DP170103000 and DP140104111 to A.F. and O.R.), the National Health and Medical Research Council (APP1058442 to A.F.), Perpetual IMPACT Philanthropy (to S.B.C.) and the Cancer Council Western Australia (to O.R.). The pETM30 plasmid was a kind gift from the EMBL Protein Expression and Purification Facility. The crystallographic experiments were performed on the ID30A-3 beamline at ESRF, Grenoble, France. We thank the Protein Science Facility at Karolinska Institutet, Stockholm, Sweden for support and providing instrumentation for protein crystallisation.

### Author contributions

H.S., T.C., A.F. and O.R. designed research; H.S., T.C., J.P.L., S.J.S., S.B.C., A.F. and O.R. performed research; S.B.C., A.F. and O.R. supervised the work; H.S., T.C., S.J.S., S.B.C., A.F. and O.R. analysed data; H.S., T.C., A.F. and O.R. wrote the manuscript and all authors commented on the manuscript.

### Additional information

**Supplementary Information** accompanies this paper at <https://doi.org/10.1038/s41467-018-04388-1>.

**Competing interests:** The authors declare no competing interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018