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.
Telomeres protect the ends of linear eukaryotic chromosomes from degradation and from fusion with other chromosomes. 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. In humans, telomeres are usually protected by a six-protein complex, known as shelterin. Of the six shelterin proteins, only Pot1 binds single-stranded telomeric DNA. Although robust technologies have now emerged that enable the manipulation of ssDNA and RNA in living cells, 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, 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. They regulate RNA processing, translation, stability and editing, and possess well-characterised, modular RNA recognition properties. PPR proteins 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. 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. 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, and the plant THA8 protein was reported to bind ssRNA with >100-fold higher affinity than ssDNA. 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, 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 α helices (Fig. 1a and Supplementary Fig. 1). In addition, a C-terminal solvating helix is added after the final consensus repeat to prevent unfolding, according to the successful consensus tetra-tri-cotipeptide repeat domain design of Main et al.

Synthetic PPR proteins fold into solenoid structures like natural PPR proteins and a recent study from Shen et al. 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, we previously identified particular amino acids at positions 5 and 35 of our cPPR that enable the specific recognition of individual RNA bases. 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.
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 Pot130, 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 not31. 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 assays32 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...
extension and internal binding stimulating the processivity of telomerase. 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 (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.
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\(^3^4\). 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\(^{21}\) and others\(^{24}\) 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\(^{24}\) 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\(^{34}\). 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\(^{35}\). 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 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 the 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
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\(^\text{36}\). 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 orientated 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\(^\text{34, 37}\). 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 THAB 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\(^\text{39}\). 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\(^\text{34}\). 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\(^\text{1}\). 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 RNA-binding proteins, such as replication protein A (RPA)\(^\text{2}\), 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\(^\text{1}\). 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
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. 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. This may result from the complex competition between telomerase, POT1, the general ssDNA-binding protein RPA, and other shelterin components. 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. 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).
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 mM NaCl and 5 mM imidazole). Lysates were clarified by centrifugation and incubated with His resin (Sigma) for 60 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 Diadez Dialyzer (Bio-Vision) 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, cPPR-polyUT and cPPR-NRE were purified as described by Coquelle et al.21. 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 positions 4, 5, 6 and 10. 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 (TTAGGGTTAGGGTTAGGGTTAG-3′) 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 Diadez Dialyzer (Bio-Vision) 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-polyUT and cPPR-NRE were purified as described by Coquelle et al.21. 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 positions 4, 5, 6 and 10.

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 (TTAGGGTTAGGGTTAGGGTTAG-3′) 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 Diadez Dialyzer (Bio-Vision) 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, cPPR-polyUT and cPPR-NRE were purified as described by Coquelle et al.21. 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 positions 4, 5, 6 and 10.

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 (TTAGGGTTAGGGTTAGGGTTAG-3′) 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 Diadez Dialyzer (Bio-Vision) 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-polyUT and cPPR-NRE were purified as described by Coquelle et al.21. 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 positions 4, 5, 6 and 10.

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 (TTAGGGTTAGGGTTAGGGTTAG-3′) 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 Diadez Dialyzer (Bio-Vision) 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, cPPR-polyUT and cPPR-NRE were purified as described by Coquelle et al.21. 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 positions 4, 5, 6 and 10.

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 (TTAGGGTTAGGGTTAGGGTTAG-3′) 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 Diadez Dialyzer (Bio-Vision) 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-polyUT and cPPR-NRE were purified as described by Coquelle et al.21. 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 positions 4, 5, 6 and 10.

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 (TTAGGGTTAGGGTTAGGGTTAG-3′) 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 Diadez Dialyzer (Bio-Vision) 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-polyUT and cPPR-NRE were purified as described by Coquelle et al.21. 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 positions 4, 5, 6 and 10.
80 °C, exposed to a PhosphoImager 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 [44]. An aliquot of 100 nM of purified cPPR-Telo1 was combined with 50 μM ssDNA library oligo-nucleotide with the sequence: 5’-CTTTATCCAGCCCTACGNNNNNNNNNTATAGTGTCACCTAAATC-3’ for 1 h in EMSA buffer (10 mM HEPES, pH 8.0, 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’-tcgtcgcggcagcgtcagatgtataagagacagCTTTATCCAGCCCTAC-3’ and 5’-gtctcgtgggctcggagatgtataagagacagGATTTAGGTGACACTATAG-3’. PCR products were purified using Exo and 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 mininsert = 0) from the BBTools suite v37.02 (https://www.gi.io/doi.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 [449] (-g ^CTTTATC-) and 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 DREMEl [50] (-k 10 -dna -normc).

**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**

1. Dickey, T. H., Altschuler, S. E. & Wuttke, D. S. Single-stranded DNA-binding proteins: multiple domains for multiple functions. *Structure* 21, 1074–1084 (2013).
2. Ashton, N. W. et al. Human single-stranded DNA binding proteins are essential for maintaining genomic stability. *BMC Mol. Biol.* 14, 9 (2013).
3. Zeman, M. K. & Cinpichr, K. A. Causes and consequences of replication stress. *Nat. Cell Biol.* 16, 2–9 (2014).
4. Sanjuan, R. Mutational dynamics: automated model building. 1. Tracing error in protein–DNA interactions. *Biochemistry (Mosc.*) 42, 504–506 (2003).
5. Chen, I., Christie, P. J. & Dubnau, D. The ins and outs of DNA transfer in yeast and humans. *Cell Rep.* 22, 127–142 (2018).
6. Blackburn, E. H. Telomeres and telomerase: the means to the end (Nobel Lecture). *Proc. Natl Acad. Sci. USA* 95, 1382 (1998).
7. Blackburn, E. H. Telomeres and telomerase: the means to the end (Nobel Lecture). *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365, 1975–1982 (2010).
8. Armstrong, C. A. & Tomita, K. Fundamental mechanisms of telomerase stabilization and mitochondrial ribosome assembly. *Cell Rep.* 33, 127–142 (2018).
9. Siria, S. J. et al. LRPRRC-mediated folding of the mitochondrial transcriptosome. *Nat. Commun.* 8, 1532 (2017).
10. Rackham, O. et al. Hierarchical RNA processing is required for mitochondrial ribosome assembly. *Cell Rep.* 16, 1874–1890 (2016).
11. Coquille, S. et al. An artificial PPR scaffold for programmable RNA recognition. *Nat. Commun.* 5, 5729 (2014).
12. 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).
13. Barkan, A. et al. A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. *PLoS Genet.* 8, e1002910 (2012).
14. Shen, C. et al. Structural basis for specific single-stranded RNA recognition by designer pentatricopeptide repeat proteins. *Nat. Commun.* 7, 11285 (2016).
15. 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).
16. Wang, F. et al. The design and structural characterization of a synthetic pentatricopeptide repeat protein. *Acta Crystallogr. D Biol. Crystallogr.* 71, 196–208 (2015).
17. 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).
18. 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).
19. Cohen, S. B. & Reddel, R. R. A sensitive direct human telomerase activity assay. *Nat. Methods* 5, 355–360 (2008).
20. Wang, F. et al. The POT1-TTP1 telomere complex is a telomerase processivity factor. *Nature* 445, 506–510 (2007).
21. Yin, P. et al. Structural basis for the modular recognition of single-stranded RNA by PPR proteins. *Nature* 504, 168–171 (2013).
22. 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).
23. Miranda, R. G., McDermott, J. I. & Barkan, A. RNA-binding specificity landscapes of designer pentatricopeptide repeat proteins elucidate principles of PPR-RNA interactions. *Nucleic Acids Res.* 46, 2613–2623 (2018).
24. Gully, B. S. et al. The solution structure of the pentatricopeptide repeat protein PPR10 upon binding atp8 RNA. *Nucleic Acids Res.* 43, 1918–1926 (2015).
25. Zaug, A. J. et al. Processive and distributive extension of human telomeres by telomerase under homeostatic and nonequilibrium conditions. *Cell Rep.* 22, 297–307 (2011).
26. Rubtsova, M. P. et al. Replication protein A modulates the activity of human telomerase in vitro. *Biochemistry (Mosc.*) 74*, 92–96 (2009).
27. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* 26, 795–800 (1993).
28. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* 66, 213–221 (2010).
29. Vormbrein, A., Blanc, E., Roversi, P. & Bricogne, G. Automated structure solution with autoSHARP. *Methods Mol. Biol.* 364, 215–230 (2007).
30. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* 62, 1002–1011 (2006).
31. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* 66, 2126–2132 (2004).
32. 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).
33. Lovell, S. C. et al. Structure validation by Calpha geometry: phi, psi and Cbeta deviation. *Proteins* 50, 437–450 (2003).
34. 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