

POT-3 preferentially binds the terminal DNA-repeat on the telomeric G-overhang

Xupeng Yu[†], Sean Gray[†] and Helder C. Ferreira^{✉*}

Biomedical Sciences Research Complex, School of Biology, University of St Andrews, North Haugh, St Andrews, UK

Received July 05, 2022; Revised December 01, 2022; Editorial Decision December 02, 2022; Accepted December 06, 2022

ABSTRACT

Eukaryotic chromosomes typically end in 3' telomeric overhangs. The safeguarding of telomeric single-stranded DNA overhangs is carried out by factors related to the protection of telomeres 1 (POT1) protein in humans. Of the three POT1-like proteins in *Caenorhabditis elegans*, POT-3 was the only member thought to not play a role at telomeres. Here, we provide evidence that POT-3 is a *bona fide* telomere-binding protein. Using a new loss-of-function mutant, we show that the absence of POT-3 causes telomere lengthening and increased levels of telomeric C-circles. We find that POT-3 directly binds the telomeric G-strand *in vitro* and map its minimal DNA binding site to the six-nucleotide motif, GCTTAG. We further show that the closely related POT-2 protein binds the same motif, but that POT-3 shows higher sequence selectivity. Crucially, in contrast to POT-2, POT-3 prefers binding sites immediately adjacent to the 3' end of DNA. These differences are significant as genetic analyses reveal that *pot-2* and *pot-3* do not function redundantly with each other *in vivo*. Our work highlights the rapid evolution and specialisation of telomere binding proteins and places POT-3 in a unique position to influence activities that control telomere length.

INTRODUCTION

Telomeres are large protein–DNA structures that protect the ends of linear chromosomes from inappropriate DNA repair and the end replication problem. Importantly, the protective functions of telomeres are mediated by proteins rather than the underlying DNA sequence *per se* (1). These protective, telomere-associated proteins tend to form a complex *in vivo*, best exemplified by the human Shelterin complex (2). Single-stranded DNA (ssDNA) binding proteins are important components of the Shelterin complex as, in most species, telomeres are processed to form a 3' ssDNA overhang (3). These overhangs help to distinguish

telomeric DNA repeats at chromosome ends (true telomeres) from internal repeats of interstitial telomere sequences (ITSS). In humans, this telomeric ssDNA binding function is carried out by the POT1 (protection of telomeres 1) protein (2). The telomere protective functions of POT1 is highly dependent on its ssDNA binding specificity through a conserved oligosaccharide/oligonucleotide binding fold (OB-fold) (4,5).

The composition of OB folds is defined by five antiparallel β strands forming a distinct β barrel (4). Variable loops connect these secondary structure elements and have a significant role in forming the binding site (6). They are also primarily responsible for variability in OB-fold lengths between 70 and 150 amino acids (7). OB folds are present across evolutionarily distant organisms and their ligands can range from RNA and ssDNA to protein (7), although they most commonly bind ssDNA. Indeed, OB fold proteins play crucial roles in manipulating ssDNA structure, which impacts telomerase activity by disrupting G-quadruplex formation and aiding telomerase translocation *in vitro* (8,9).

Within human Shelterin, POT1 is the only protein that confers ssDNA binding. It binds telomeric DNA via two OB folds and deletion of POT1 results in telomere-associated dysfunctions such as 5' end hyper-resection, and increased telomere elongation *in vivo* (10). Mice and rats have two POT1 proteins, referred to as mPOT1a and mPOT1b (11). These are homologous to the two OB folds of human POT1 (12). Despite having duplicated only recently and displaying 75% sequence similarity, the functions of mPOT1a and mPOT1b have diverged significantly (11).

The nematode worm *Caenorhabditis elegans* contains three POT1-like proteins (POT-1, POT-2 and POT-3), each containing a single OB fold. Loss of *C. elegans* POT-1 or POT-2 results in telomere elongation (13), mimicking what is seen in humans. However, *pot-1* and *pot-2* mutants do not behave identically. Loss of POT-1 results in greater initial telomere length heterogeneity (13) and reduced tethering to the nuclear periphery (14) than loss of POT-2. The closely related *pot-3* gene shares very high sequence similarity to *pot-2* (~60% amino acid identity). POT-3 initially appeared to have no telomeric function, as a *pot-3* (*ok1530*) allele was

*To whom correspondence should be addressed. Tel: +44 1334 463425; Email: hcf2@st-andrews.ac.uk

[†]The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

shown to have normal telomere length (15). Here, we characterise the behaviour of the POT-3 protein *in vitro* and the phenotypes of a novel *pot-3* mutation *in vivo* to show that POT-3 does indeed have an important telomeric function. Understanding OB-fold containing telomere binding proteins from different eukaryotes may shed light on the diversity of telomere maintenance mechanisms.

MATERIALS AND METHODS

Protein purification and expression

POT2 (HFP300) or POT3 (HFP301) encoding plasmids were transformed into BL21 (DE3) pOFX34 *Escherichia coli*. Overnight cultures were used to inoculate 2 l LB media and grown at 37°C until OD₆₀₀ reached 0.6, then placed at 4°C for 30 min to reduce the temperature. Cultures were induced with 0.1 mM IPTG and induction was carried out at 16°C for ~16 h. Cells were harvested by centrifugation at 5000 rpm for 10 min and frozen at -20°C until processed. The cell pellets were thawed on ice and resuspended in cold lysis buffer (500 mM NaCl, 50 mM Tris-HCl pH 8.0, 10 mM imidazole, 10% glycerol) + 1× cOmplete™ protease inhibitor tablet (Roche) and sonicated on ice. This was then centrifuged at 40 000 rpm for 35 min and the supernatant loaded onto a 5 ml HisTrap™ HP column (GE Healthcare). It was then wash and eluted using a gradient of Buffer A (500 mM NaCl, 50 mM Tris-HCl pH 8.0, 30 mM imidazole, 10% glycerol) to Buffer B (500 mM NaCl, 50 mM Tris-HCl pH 8.0, 500 mM imidazole, 10% glycerol). SDS-PAGE identified fractions containing protein of interest at highest purity. POT2 and POT3 fractions were concentrated using an Amicon® 10K concentrator at 4000 rpm and the concentrated samples dialysed using Snakeskin™ 10K MWCO (ThermoFisher) tubing overnight into dialysis buffer (500 mM NaCl, 50 mM Tris-HCl pH 8.0, 10% glycerol) at 4°C. This was then run on a HiLoad™ 16/60 Superdex 200 pg (GE Healthcare) gel filtration column equilibrated in gel filtration buffer (500 mM NaCl, 20 mM Tris-HCl pH 8.0, 10% glycerol) using a flow rate of 1 ml/min. Protein concentration was determined by Nanodrop 2000 and the protein frozen in small aliquots at -80°C.

Analytical gel filtration

The Superose® 12 10/300 GL column was equilibrated with 1.5 column volumes of gel filtration buffer (500 mM NaCl, 20 mM Tris-HCl pH 8.0, 10% glycerol). 250 µl binding reactions were set up in binding buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 10 mM DTT, 10% glycerol) containing POT-2 at 126 µM or POT-3 at 65 µM + no DNA or 1:1 (TTAGGC)₂ or 2:1 (TTAGGC)₄ molar ratio of ssDNA. The entire 250 µl binding reaction was loaded onto a 500 µl loop equilibrated in gel filtration buffer and run at 1 ml/min.

Generation of telomeric DNA substrates

All oligonucleotides were purchased from IDT® and resuspended in autoclaved H₂O to make 100 µM stocks. Oligos were annealed at 1 µM in ST buffer (100 mM NaCl, 10 mM Tris-HCl pH 8.0) in a PCR machine by heating to 95°C for

3 min and then cooling to room temperature at a rate of 1°C/min. Annealed oligos were stored at 4°C.

EMSA

Binding reactions were set up as follows. Appropriate concentration of protein and DNA were mixed together in binding buffer (50 µg/ml BSA, 1 mM MgCl₂, 5 mM DTT, 20 mM Tris-HCl pH8.0, 50 mM NaCl, 4% Ficoll 400). Binding reactions were incubated on ice for 10 min before loading the sample onto a 7% native polyacrylamide gel, which had been pre-run for an hour. Electrophoresis was carried out in 0.5× TBE buffer for 1 h at 100 V with an ice pack to keep the temperature low.

Competition binding

0.5 nM Cy5-labelled 12mer DNA was incubated with 50 nM of either POT-2 or POT3 in binding buffer (50 mM NaCl, 20 mM Tris-HCl pH 8.0, 50 µg/ml BSA, 1 mM MgCl₂, 5 mM DTT, 0.001% Tween-20), for 1 h at room temperature. After this, increasing concentrations of unlabelled 10mer or 12mer DNA were added and left for a further 1 h at room temperature. Ficoll 400 was added to a final concentration of 4% and the reaction cooled on ice briefly before running on 7% native polyacrylamide gels in cold 0.5× TBE buffer.

Strains

Unless noted otherwise, all strains were cultured at 20°C on nematode growth medium plates seeded with *E. coli* OP50. A full strain list is given in Supplementary Table S2.

Terminal restriction fragment (TRF) analysis

Worms were digested in 1× NTE buffer (100 mM NaCl, 50 mM Tris pH 7.4, 20 mM EDTA), 1% SDS and 500 µg/ml Proteinase K overnight at 65°C. Two consecutive phenol-chloroform extractions, followed by chloroform back-extraction and ethanol precipitation were carried out. DNA was eluted in 10 mM Tris-EDTA (pH7.5). 5 µg purified DNA was digested overnight with HinfI and HaeIII (NEB) at 37°C and resolved on a 1% agarose gel. Following a 20-min depurination in 250 mM HCl, the gel was washed 2× in denaturing buffer (1.5 M NaCl, 0.5 M NaOH) and 2× in neutralising buffer (1.5 M NaCl, 0.5 M Tris-HCl, pH 8) at room temperature. DNA was transferred onto neutral nylon membrane (Hybond-NX, GE Healthcare) by capillary transfer in 10× SSC buffer (1.5 M NaCl, 150 mM sodium citrate, pH 7). After briefly rinsing in 2× SSC buffer DNA was UV crosslinked at 1200 J/m² and hybridised with a digoxigenin-labelled telomere probe (GCCTAA)₄

C-circle assay

DNA was extracted following mechanical lysis using 0.5 mm glass beads in 73 µg/ml RNase A, 9 mM EDTA and 270 mM NaCl in a cell homogeniser for 3 × 20 s at 6 m/s. Proteins were denatured by adding 1% SDS and heating to

65°C for 10 min, followed by precipitation with a final concentration of 1.3 M potassium acetate pH 5.2. After additional purification using phenol:chloroform:isoamyl alcohol (25:24:1, pH 6.7) and chloroform back-extraction, DNA was ethanol precipitated and eluted in Tris-EDTA buffer (10 mM Tris-HCl, 100 μ M EDTA, pH 7.5). This was then used in a C-circle assay as described (16). Briefly, 0.5 μ l phi29 polymerase (NEB) was added to 1 μ g of genomic DNA and incubated at 30°C for 8 h. This was spotted onto a neutral Hybond-N membrane, UV cross-linked (1200 J/m²) and hybridized with a DIG-labelled (GCCTAA)₄ probe at 37°C using DIG Easy Hyb (Roche) according to the manufacturer's instructions.

RESULTS

Mutation of *pot-3* increases telomere length and recombination

Given the homology to POT-2, we suspected that POT-3 might also play a role at telomeres. To test whether mutation of *pot-3* caused a telomeric phenotype, we isolated a new null allele, *pot-3(syb2415)* which contains a 500 bp deletion spanning the entire OB fold (Figure 1A and Supplementary Figure S1). *pot-3(syb2415)* worms are viable and fertile. Interestingly, we observed that *pot-3(syb2415)* mutants have markedly longer telomeres than wildtype worms 5.5–23.9 kbp versus 2.2–8.6 kbp respectively (Figure 1B and Supplementary Table S1). We observed that telomere length in *pot-3* mutants is not stable but rather increases over successive generations (data not shown). This behaviour has previously been seen with *pot-2* mutants (15). Therefore, to accurately compare telomere lengths, all experiments were carried out with early generations of the relevant genotypes.

Besides telomere lengthening, *pot-2* mutants also display higher levels of telomeric C-circles (14,15). These extra-chromosomal circles of telomeric DNA are hallmarks of the alternative lengthening of telomeres (ALT) pathway (16). Strikingly, we also observe that *pot-3* mutants have significantly higher levels of C-circles than wildtype (Figure 1C). Thus, we see that *pot-3* mutants show both of *pot-2*'s telomere phenotypes, namely longer telomeres and increased levels of C-circles. These data strongly suggest that POT-3 plays a similar role to that of POT-2 at worm telomeres.

POT-3 does not act redundantly with POT-2 *in vivo*

We noticed that, although similar to *pot-2*, the telomeric phenotypes of *pot-3* mutants were always weaker. This raised the possibility that perhaps POT-2 and POT-3 were carrying out the same role (i.e. they were redundant) but that POT-2 was somehow more important or abundant. If this were the case, we would observe stronger telomeric defects in a *pot-2; pot-3* double mutant compared to either single mutant. However, we found that a *pot-2; pot-3* double mutant had weaker telomeric phenotypes than a *pot-2* single mutant. The loss of *pot-3* did not exacerbate but rather suppressed both the telomere length and C-circle phenotypes of *pot-2* mutants (Figure 1B, C, Supplementary Table S1 and Supplementary Figure S2). Thus, rather than acting redundantly, *pot-2* and *pot-3* mutants are epistatic. This indicates that they do not perform the same function but they

work together within the same genetic pathway. Indeed, this epistatic relationship is not restricted to telomeric phenotypes. We find that *pot-2; pot-3* double mutants also have a significantly higher brood size than *pot-2* single mutants (Figure 1D). This indicates that *pot-2* and *pot-3* also affect general fitness in an epistatic manner.

Loss of POT homologs in yeasts leads to telomere uncapping, resulting in loss of telomeric DNA and chromosome fusion (17,18). Although *pot-2* single mutant worms did not show signs of chromosome fusion (14), we wondered whether this phenotype might only be revealed in a *pot-2; pot-3* double mutant. The number of chromosomes in *C. elegans* can readily be counted in meiotic cells in diakinesis. However, we did not find evidence of large-scale chromosome fusions in either the *pot-3* single or the *pot-2; pot-3* double mutant (data not shown).

POT-3 specifically binds the G-strand of telomeric DNA

To understand how POT-3 might be acting on telomeres *in vivo*, we decided to test its DNA-binding properties *in vitro*. His-tagged POT-3 was recombinantly expressed in *E. coli* and purified to homogeneity using a combination of affinity chromatography and gel filtration (Figure 2A, B). Most POT homologs from other species contain multiple OB-folds. In contrast, all *C. elegans* POT homologs contain only a single OB fold. We therefore wondered whether they might multimerise. However, interestingly, POT-3 is monomeric in solution, although it is prone to forming cysteine-mediated dimers (Figure 2A, B).

To assay POT-3's specificity for telomeric DNA, we hybridised DNA oligonucleotides to generate structures that mimicked telomeres. These double-stranded molecules had single-stranded DNA overhangs containing the *C. elegans* telomere sequence. Using electrophoretic mobility shift assays (EMSA), we find that POT-3 efficiently binds DNA templates containing a G-overhang (Figure 2C). This interaction is DNA sequence-specific because if we scramble the G-overhang sequence then POT-3 no longer recognised it. Moreover, POT-3 binding is strand-specific. If we use templates containing the complementary telomeric C-overhang, POT-3 fails to bind. Recognition of the telomeric G-strand has to take place within the context of ssDNA because telomeric dsDNA is not bound by POT-3 (Figure 2C). These data are consistent with the observed telomeric phenotypes of *pot-3* mutant worms and strongly suggest that POT-3 is a *bona fide* telomere binding protein.

POT-3 and POT-2 bind a minimal six nucleotide GCTTAG motif

We initially thought that POT-3 needed more than six nucleotides of telomeric G-strand to bind. Using templates containing either one, two or three repeats of the telomeric sequence (TTAGGC) as ssDNA, we saw that POT-3 would readily bind overhangs containing two or more copies of TTAGGC but could not bind an overhang with a single repeat (Figure 2D). However, it remained possible that that we were missing the minimal binding motif of POT-3 because we were not using the correct register of telomeric repeat sequence.

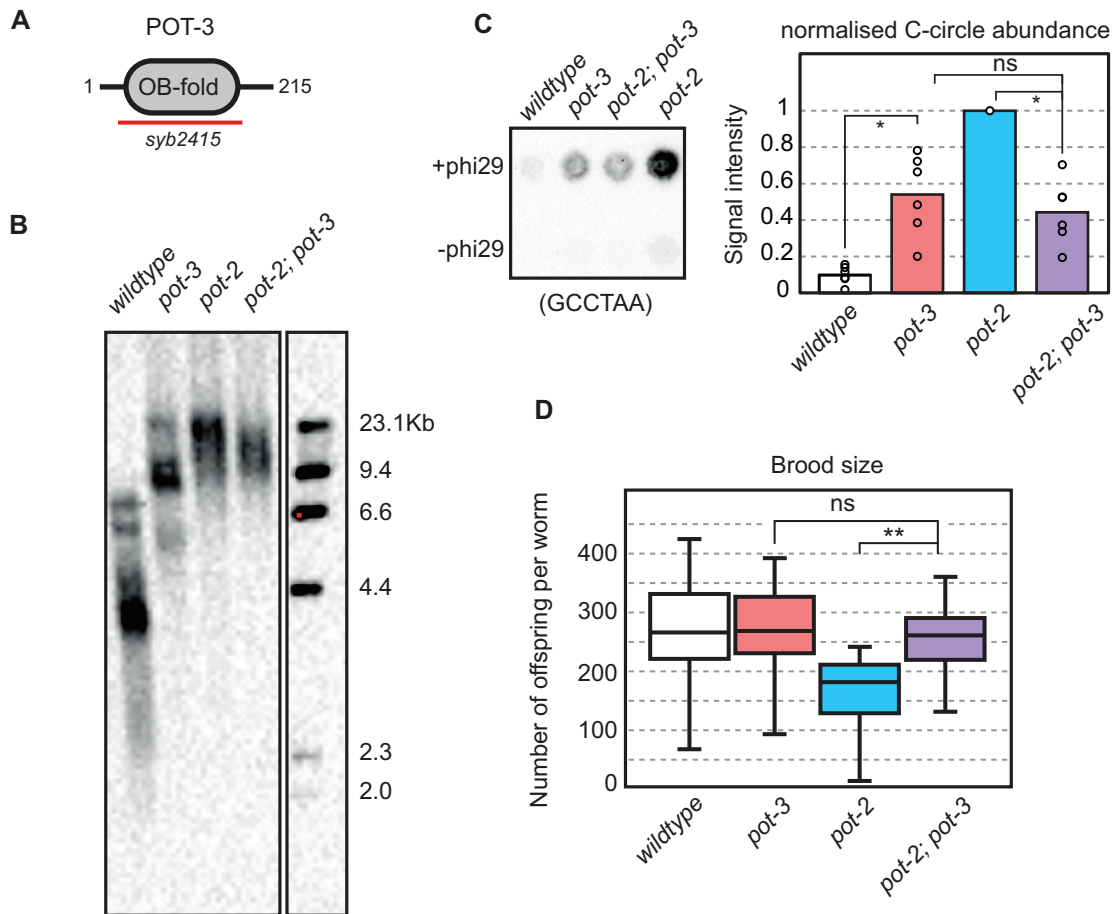


Figure 1. Mutation of *pot-3* increases telomere length and recombination in a manner that is epistatic with *pot-2*. (A) POT-3 is 215aa long, the *syb2415* allele contains a 500 bp deletion spanning the entire OB-fold region. (B) Southern blot of terminal restriction fragments from genomic DNA show that mutation of *pot-3* results in an increase in telomere length almost to that of *pot-2*. Interestingly *pot-2; pot-3* double mutants do not have longer telomeres than either single mutant. Membrane was probed with a DIG-labelled (TTAGGC)₄ oligo. (C) Telomeric C-circle assays were carried out using phi29 polymerase, spotted onto a nitrocellulose membrane and probed with a DIG-labelled (TTAGGC)₄ oligo. A representative dot blot is shown in the left panel. Signal intensity was quantified using ImageJ and results plotted relative to *pot-2* (set to 100%). The bar graph in the right panel shows the average of six independent experiments with individual results displayed as open circles. Mutation of *pot-3* increases C-circle levels above that of wildtype but suppressed the levels of C-circles in a *pot-2* background, ns = not significant, * $P < 0.05$ (Student's *t*-test, two tailed distribution, unequal variance). (D) The total number of viable offspring per adult worm (brood size) was measured for the indicated genotypes at 20°C. The box plot displays the mean, 25th and 75th percentile from at least 20 independent adults. Mutation of *pot-3* has no significant effect on brood size on its own but suppresses the lower brood size of a *pot-2* mutant. ns = not significant, ** $P < 0.005$.

To address this, we tested POT-3 binding to a series of telomeric substrates. These started with a single six nucleotide TTAGGC overhang (which we knew POT-3 did not bind) that were extended one nucleotide at a time at the 3' end until it reached a twelve nucleotide (TTAGGC)₂ overhang (which we knew POT-3 did bind). We reasoned that if we identified the point at which extending the 3' end resulted in POT-3 binding, we could then shorten the telomeric repeat from the 5' end to identify the minimal binding motif. In contrast to what we previously thought, POT-3 does indeed only require six nucleotides for complete binding. However, this has to be in the register GCTTAG and not TTAGGC (Figure 3A). This is comparable to the binding behaviour of human POT1. Although, the two OB folds of human POT1 bind a ten nucleotide motif, this also ends in TTAG. Moreover, human POT1 has a higher affinity for (GGTTAG)₂ than (TTAGGG)₂ (19).

POT-2 has previously been shown to also bind the telomeric G-strand (13). We wondered whether POT-2 and POT-3 might differ in the precise sequence they bound within the G-overhang. Therefore, we carried out the same experiment as previously described to map its minimal binding motif. The purification of POT-2 is shown in Supplementary Figure S3. Surprisingly, we find that POT-2 binds to exactly the same minimal GCTTAG motif as POT-3 (Figure 3A, B). We tried estimating the binding affinities of POT-2 and POT-3 to telomeric substrates using EMSA. When binding was carried out on ice, both proteins had remarkably tight binding in the picomolar range. Due to the Cy5 detection limits of our scanner, we struggled to detect signal at the low DNA concentrations necessary to make accurate binding measurements. We estimate that the apparent K_{ds} of POT-2 and POT-3 are 100 and 160 pM, respectively (Figure 3C, D). When binding was carried out at room temperature, the apparent K_{ds} of POT-2 and POT-3 for (TTAGGC)₂ are

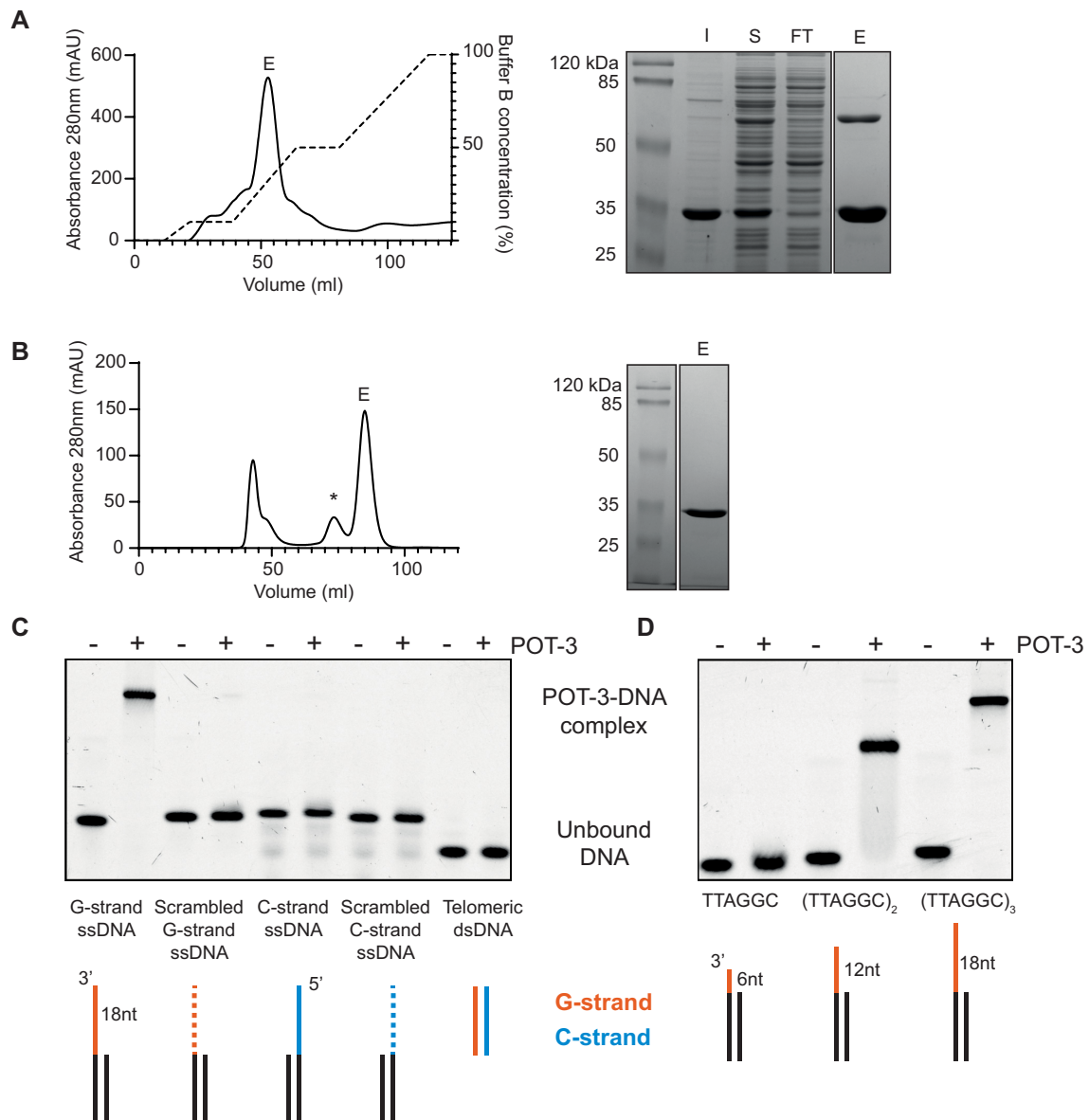


Figure 2. POT-3 is a monomer and binds the telomeric G-strand as ssDNA. (A) Nickel affinity purification of 6xHis-tagged POT-3 expressed from *E. coli*. The panel on the right is a Coomassie stained SDS-PAGE gel of different fractions collected during the purification: I – insoluble, S – soluble, FT – flowthrough, E – elution. (B) Affinity purified POT-3 was subsequently run over a HiLoad 16/60 Superdex 200 size exclusion column. The majority of POT-3 migrates as a monomer although it can form disulphide-mediated protein dimers, indicated by an asterisk. (C) 500nM Purified POT-3 was incubated with 50 nM Cy5-labelled DNA and run on a native acrylamide gel. DNA substrates contained a non-telomeric dsDNA region (black) with either a 3' G-strand telomeric ssDNA overhang (orange) or a 5' C-strand telomeric ssDNA overhang (blue). Telomeric ssDNA overhangs were 18 nucleotides long and consisted of three telomeric repeats, (TTAGGC)₃. The scrambled G- and C-strand overhangs retained the same GC content as the telomeric sequence but the nucleotide order was randomised. (D) POT-3 only binds to G-overhangs containing more than one copy of the telomeric sequence TTAGGC.

7.2nM and 5.0nM respectively (Supplementary Figure S4). Such temperature-dependent effects are expected for single-stranded nucleic acid binding proteins (20). These data allow us to conclude that POT-2 and POT-3 show similar affinities for ssDNA (TTAGGC)₂ which is tighter than that of human POT1 for (TTAGGG)₂, namely 59 nM (19). This is surprising given that the *C. elegans* POT proteins only have a single OB fold compared to the two OB-folds that human POT1 uses to bind telomeric DNA.

As POT-2 and POT-3 only have a single OB fold, we wanted to confirm that they bind their cognate ssDNA

recognition sequence as monomers. We therefore carried out analytical gel filtration of POT-2 and POT-2 with DNA fragments that contained differing numbers of GCTTAG binding sites. The elution profiles of either POT-2 or POT-3 bound to (TTAGGC)₂, which contains a single GCTTAG binding site, was consistent with only a single protein molecule bound to DNA (Figure 3E, F). When we used (TTAGGC)₄, which contains three GCTTAG binding sites, the protein complexes migrated as single species with an apparent molecular weight consistent with three protein molecules bound. These data indicate that multiple POT

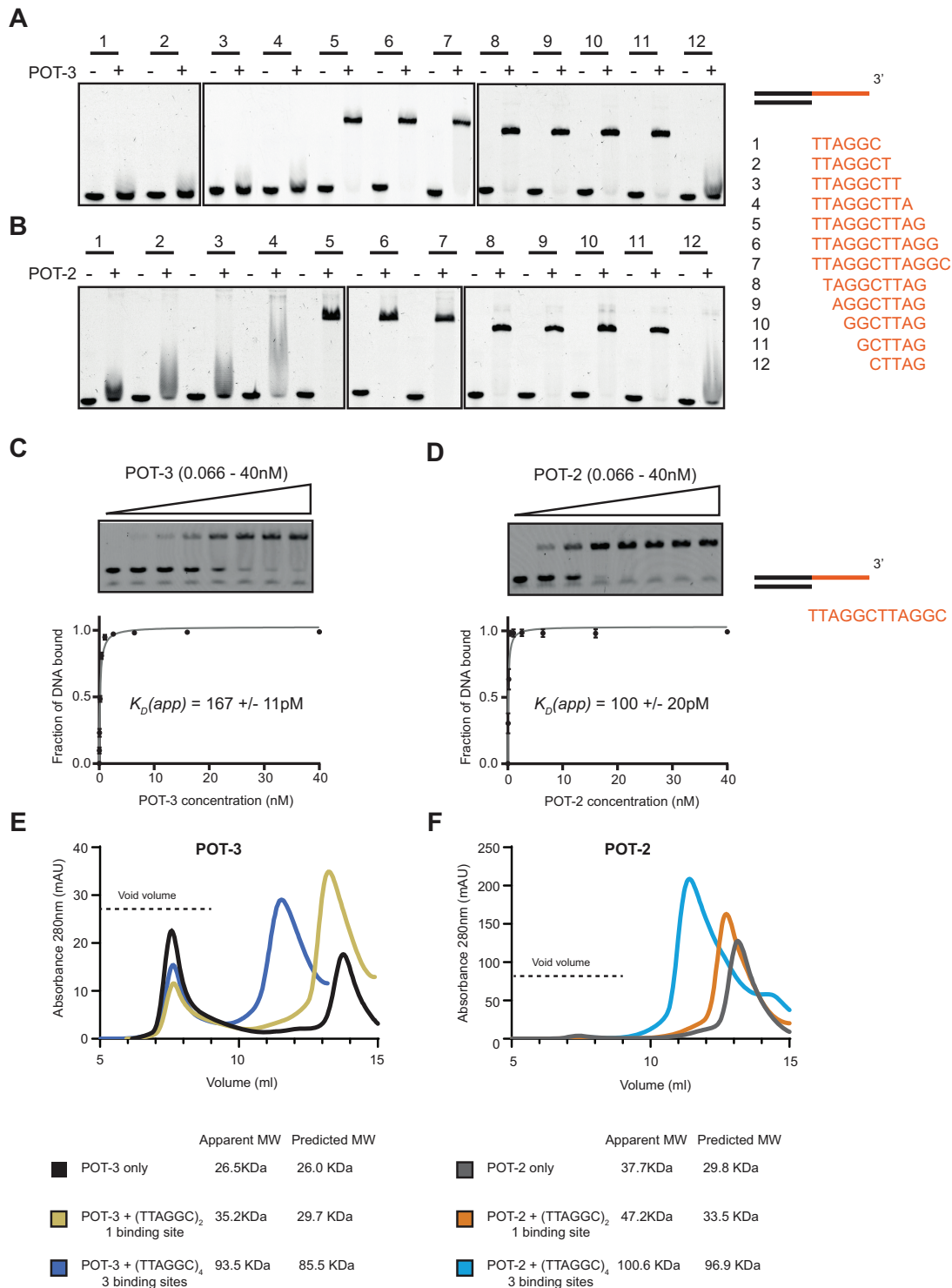


Figure 3. POT-3 and POT-2 bind the same minimal six-nucleotide recognition sequence, GCTTAG. (A) 500 nM POT-3 was bound to a series of DNA telomere fragments (50 nM) that had 3' overhangs increased by one nucleotide at a time from TTAGGC (fragment 1) to (TTAGGC)₂ (fragment 7). The first fragment in this series to show binding (fragment 5) was then shortened one nucleotide at a time from the 5' end until it was no longer able to be bound by POT-3 (fragment 12). (B) As with part A, except POT-2 was used. POT-2 binds the same minimal sequence as POT-3 but is less selective, (compare fragment 4 in parts A and B). (C) 0.2 nM Cy5 labelled DNA containing a (TTAGGC)₂ overhang was incubated with increasing amounts of purified POT3 (0.066–40 nM). The apparent dissociation constant $K_D(app)$ was calculated in Prism using a one site - specific binding equation and is the mean of three replicates \pm the standard deviation. (D) Same as in part C except using purified POT-2. (E) Analytical gel filtration of purified POT-3 incubated with either no DNA, (TTAGGC)₂ or (TTAGGC)₄ ssDNA and run on a Superose® 12 10/300 GL column. (TTAGGC)₂ and (TTAGGC)₄ contain either one or three minimal GCTTAG binding sites respectively. Comparison with molecular weight standards reveal that POT-3 binds its recognition sequence as a monomer and that three adjacent binding sites can be fully bound by three POT-3s concurrently. The chromatogram does not show peaks of DNA absorbance alone, which were removed for simplicity. (F) Same as in part C except using purified POT-2

proteins can fully occupy adjacent DNA binding sites. Due to the six nucleotide repeating register of telomeric DNA, there would be no linker DNA between adjacent POT-2/3 proteins. This suggests that POT-2/3 likely form a rigid nucleoprotein filament when bound to DNA.

POT-3 has a preference over POT-2 for binding at the 3' end of DNA

The fact that POT-2 and POT-3 bind the same minimal motif, made us wonder how they differed. These proteins are clearly carrying out distinct functions *in vivo*. Mutation of either single gene causes a telomeric phenotype, which is not additive when both genes are deleted. This argues against both proteins performing the same role or their individual phenotype being caused by a dosage effect. We surmised that although they bound the same motif, perhaps it mattered where the motif was along the telomeric repeat.

To test this, we carried out a competition experiment where we assayed the ability of different DNA templates to outcompete pre-bound POT-2 or POT-3 complexes. One of these DNA templates, which we term the 10mer, has a TTAGGCTTAG overhang. Therefore, the underlined minimal binding site is at the extreme terminus of the DNA, adjacent to the 3' hydroxyl. The second template, which we term the 12mer, has a TTAGCTTAGGC overhang. Here, the underlined minimal binding site is internal, two nucleotides away from the end of the DNA.

We pre-bound POT-2 to Cy5-labelled 12mer and titrated in either unlabelled 10mer or 12mer DNA. We found that there was no difference between these two competitors (Figure 4A, C). In contrast, when we repeated this with POT-3, we found that it was outcompeted more efficiently by the 10mer than by the 12mer (Figure 4B, D). This difference is likely driven by the relative kinetics of POT-3 binding to these DNA fragments as both POT-2 and POT-3 show a higher intrinsic affinity for 10mer DNA compared to 12mer (Supplementary Figure S4).

These data show that POT-3 binds more stably to sites at the ends of DNA. This end-binding preference requires proximity to the 3' end because if the minimal binding site is switched to the 5' end, then POT-3 is incapable of binding (Supplementary Figure S5). Interestingly, POT-2 shows reduced binding but it is still capable of binding when presented with a substrate with a binding site at the 5' end of DNA.

This increased DNA-end selectivity of POT-3 over POT-2 is also consistent with the data that POT-2 shows more promiscuous binding than POT-3 to DNA sequences that approximate its minimal binding site (substrate 4 in Figure 3B). Altogether, this shows that, although they bind the same minimal sequence, POT-3 is more selective than POT-2 and it prefers to bind the terminal telomeric repeat of the 3' G-overhang (Figure 4E). This region is the site of action for telomerase and exonucleases such as MRT-1, meaning that POT-3 is in a privileged location to influence telomere maintenance.

DISCUSSION

C. elegans contains three POT1-like proteins (POT-1, POT-2 and POT-3) which contain the characteris-

tic oligonucleotide/oligosaccharide-binding (OB) fold. A fourth *C. elegans* protein (MRT-1) also contains a similar OB fold. However, in contrast to human POT1, MRT-1 additionally contains an active nuclease domain (21), suggesting a distinct function. This is supported by *mrt-1* mutants having short telomeres (21), the opposite phenotype of *pot-1* or *pot-2* mutants (13). We show here that *pot-3* (*syb2415*) mutants also have long telomeres. This phenotype is different to that previously observed with *pot-3* (*ok1530*). When we sequenced the *pot-3* locus in the *ok1530* allele, we observed no mutations compared to wildtype strains (Supplementary figure S1 and data not shown). Thus, this strain is likely to have normal POT-3 function. In contrast, *syb2415* is a 500 bp deletion that completely removes the OB fold, the only known functional domain within POT-3.

Our work and others show that all three nematode POT proteins prevent telomere elongation; however, they do so in different ways. Telomeres in *C. elegans* are unusual in that they contain 5' C-strand overhangs (bound by POT-1) as well as 3' G-strand overhangs (bound by POT-2 and POT-3). Thus, POT-2 and POT-3 might antagonise telomerase via direct competition. However, POT-1 likely acts via a different mechanism as it binds the opposite strand of DNA to telomerase.

We also show that *pot-2* acts epistatically with *pot-3* to prevent C-circle formation. These extra-chromosomal circles of telomeric DNA occur in a subset of cancers (16) and are thought to be generated via a combination of long telomeres and replication stress (22) as well as through inappropriate DNA repair mechanisms at telomeres (23,24). It has been proposed that C-circles are formed via the cleavage of a T-loop intermediate (25). Longer telomeres may be more likely to form T-loops, which could be why *pot-2* and *pot-3* mutants have higher levels of C-circles than wildtype.

The epistasis between *pot-2* and *pot-3* in C-circle formation is interesting. We speculate that part of the function of POT-2 is to coat the bulk of the G-overhang and restrict POT-3 to the terminal repeat. Indeed, there is ~100× more POT-2 protein in worm embryos than POT-3 (26). This disparity in abundance might also explain why POT-2 but not POT-3 was detected via mass spectrometry in pulldowns using the telomere binding proteins TEBP-1 and TEBP-2 (27). Thus, in a *pot-2* mutant, POT-3 is no longer restricted to the terminal telomeric repeat and may now mislocalise to other regions such as the displaced G-strand formed next to a T-loop. We speculate that this mislocalised POT-3 makes T-loops more likely to be aberrantly processed into C-circles (Supplementary Figure S6). Such a model would explain why a double *pot-2; pot-3* mutant has lower C-circle levels than a *pot-2* single mutant.

We show that POT-2 and POT-3 bind ssDNA as monomers using a single OB-fold. This is different to most organisms where telomeric ssDNA-binding proteins often contain multiple OB folds which either contribute to DNA-binding or to protein dimerization (28). Additionally, multiple POT-2/3 proteins can fully occupy adjacent DNA binding sites without intervening linker DNA between adjacent POT-2/3 proteins. The lack of a DNA linker suggests that POT-2/3-bound G-overhangs are likely to be quite rigid and this in turn might antagonise T-loop and subsequent C-circle formation. The telomeric DNA sequence of nema-

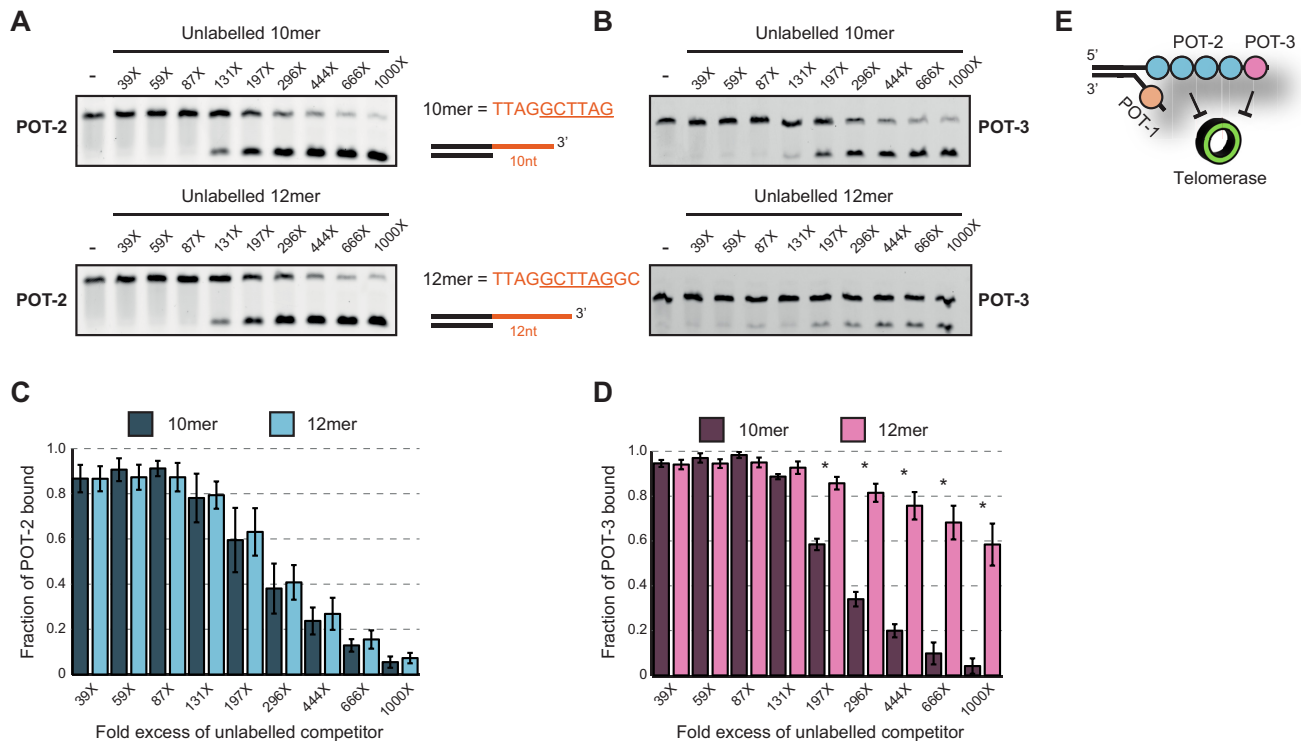


Figure 4. POT-3 preferentially binds its recognition sequence when it is immediately adjacent to the terminal 3' hydroxyl. (A) A representative EMSA of 50 nM POT-2 pre-bound to 0.5 nM Cy5-labelled 12-mer DNA and then incubated with increasing amounts (19.5–500 nM) of unlabelled DNA. (B) Same as part A except that POT-3 is used. (C, D) Quantification of parts A and B displaying amount protein still bound to labelled DNA. The graph shows the mean and standard deviation from at least three replicates. The asterisks indicate conditions at which the difference between 10mer and 12mer DNA is statistically significant (Student's *t*-test, $P < 0.01$). POT-2 can be outcompeted equally by either the 10mer or 12mer fragments, whereas POT-3 is outcompeted more efficiently by the 10mer. This indicates that it prefers its six nucleotide binding site (underlined) at the extreme 3' end. (E) Working model illustrating that POT-2 and POT-3 coat the telomeric G-overhang. They both repress telomerase activity to prevent the inappropriate lengthening of telomeres. However, in contrast to POT-2, POT-3 prefers to bind to the terminal repeat at the 3' end of the G-overhang.

todes (TTAGGC) was initially thought to be incapable of forming a G-quadruplex (29). However, it has recently been shown to fold into a non-canonical G-quadruplex structure that is less stable than the metazoan telomeric DNA sequence (TTAGGG) (30). G-quadruplexes are often associated with genome instability but they may also play protective roles at telomeres (31). Therefore, it is possible that part of POT-2/3 function is to modulate the levels of G-quadruplexes at chromosome ends.

Mutation of human POT1 causes telomere elongation (32) and is associated with cancers such as glioma (33). The complete loss of POT1 results in DNA damage activation and telomere lengthening (but not telomere fusions) in human (34) and mouse (11) cells. Interestingly, loss of POT1 homologs in simpler eukaryotes such as moss (35) and yeasts (17,18) have a distinct phenotype, causing increased chromosome fusions and telomere shortening instead of lengthening. Loss of either POT-1, POT-2 or POT-3 results in telomere elongation but not increased chromosome fusions (13). This indicates that *C. elegans* POT proteins behave more like their human, rather than yeast or plant, homologs.

The high level of sequence identity between POT-2 and POT-3 suggest that they arose from a gene duplication event that underwent rapid diversification of function. This is consistent with the observation that telomeric proteins undergo particularly rapid evolution (36). Indeed, a cursory

examination of closely related *Caenorhabditis* species reveals large variability in the number of POT-like genes (data not shown).

We show here that POT-2 and POT-3 control telomere homeostasis by binding the canonical G-overhang. *C. elegans* are unique in also having robust levels of the more unusual C-overhang and encode a dedicated protein (POT-1) to bind this ssDNA (13). The precise function of the C-overhang is unclear. However, it is likely to be important as POT-1 is required to maintain the physical interaction between the telomeric dsDNA binders (TEBP-1/2) and the telomeric ssDNA binder (POT-2) *in vivo* (27). POT-1 also promotes telomeric association with the nuclear envelope during embryogenesis (14). It will be important to understand how POT-1, POT-2 and POT-3 work together to bind both the telomeric C-strand and G-strand in *C. elegans* and how this organisation influences telomere maintenance.

DATA AVAILABILITY

Strains and data are available upon request to the corresponding author.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We would like to thank Dr Shirley Graham for help with protein purification, Dr Clarissa Melo Czekster for help with the analysis of protein binding and Janie Olver for critical reading of the manuscript.

FUNDING

The China Scholarship Scheme supported Xupeng Yu [201904910787]; some strains were provided by the CGC, which is funded by NIH Office of Research Infrastructure Programs [P40 OD010440]. Funding for open access charge: Scottish Higher Education Digital Library Consortium (SHEDL) Read and Publish agreement. *Conflict of interest statement.* None declared.

REFERENCES

- van Steensel, B., Smogorzewska, A. and de Lange, T. (1998) TRF2 protects human telomeres from end-to-end fusions. *Cell*, **92**, 401–413.
- de Lange, T. (2009) How telomeres solve the end-protection problem. *Science*, **326**, 948–952.
- Baumann, P. and Price, C. (2010) Pot1 and telomere maintenance. *FEBS Lett.*, **584**, 3779–3784.
- Murzin, A.G. (1993) OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. *EMBO J.*, **12**, 861–867.
- Horvath, M.P. (2011) Structural anatomy of telomere OB proteins. *Crit. Rev. Biochem. Mol. Biol.*, **46**, 409–435.
- Lei, M., Podell, E.R., Baumann, P. and Cech, T.R. (2003) DNA self-recognition in the structure of Pot1 bound to telomeric single-stranded DNA. *Nature*, **426**, 198–203.
- Theobald, D.L., Mitton-Fry, R.M. and Wuttke, D.S. (2003) Nucleic acid recognition by OB-fold proteins. *Annu. Rev. Biophys. Biomol. Struct.*, **32**, 115–133.
- Latrick, C.M. and Cech, T.R. (2010) POT1-TPP1 enhances telomerase processivity by slowing primer dissociation and aiding translocation. *EMBO J.*, **29**, 924–933.
- Zaug, A.J., Podell, E.R. and Cech, T.R. (2005) Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 10864–10869.
- Croy, J.E. and Wuttke, D.S. (2006) Themes in ssDNA recognition by telomere-end protection proteins. *Trends Biochem. Sci.*, **31**, 516–525.
- Hockemeyer, D., Daniels, J.P., Takai, H. and de Lange, T. (2006) Recent expansion of the telomeric complex in rodents: two distinct POT1 proteins protect mouse telomeres. *Cell*, **126**, 63–77.
- Wu, L., Multani, A.S., He, H., Cosme-Blanco, W., Deng, Y., Deng, J.M., Bachilo, O., Pathak, S., Tahara, H., Bailey, S.M. *et al.* (2006) Pot1 deficiency initiates DNA damage checkpoint activation and aberrant homologous recombination at telomeres. *Cell*, **126**, 49–62.
- Raices, M., Verdun, R.E., Compton, S.A., Haggblom, C.I., Griffith, J.D., Dillin, A. and Karlseder, J. (2008) *C. elegans* telomeres contain G-strand and C-strand overhangs that are bound by distinct proteins. *Cell*, **132**, 745–757.
- Ferreira, H.C., Towbin, B.D., Jegou, T. and Gasser, S.M. (2013) The shelterin protein POT-1 anchors caenorhabditis elegans telomeres through SUN-1 at the nuclear periphery. *J. Cell Biol.*, **203**, 727–735.
- Shtessel, L., Lowden, M.R., Cheng, C., Simon, M., Wang, K. and Ahmed, S. (2013) *Caenorhabditis elegans* POT-1 and POT-2 repress telomere maintenance pathways. *G3*, **3**, 305–313.
- Henson, J.D., Cao, Y., Huschtscha, L.I., Chang, A.C., Au, A.Y., Pickett, H.A. and Reddel, R.R. (2009) DNA C-circles are specific and quantifiable markers of alternative-lengthening-of-telomeres activity. *Nat. Biotechnol.*, **27**, 1181–1185.
- Baumann, P. and Cech, T.R. (2001) Pot1, the putative telomere end-binding protein in fission yeast and humans. *Science*, **292**, 1171–1175.
- Garvik, B., Carson, M. and Hartwell, L. (1995) Single-stranded DNA arising at telomeres in cdc13 mutants may constitute a specific signal for the RAD9 checkpoint. *Mol. Cell Biol.*, **15**, 6128–6138.
- Lei, M., Podell, E.R. and Cech, T.R. (2004) Structure of human POT1 bound to telomeric single-stranded DNA provides a model for chromosome end-protection. *Nat. Struct. Mol. Biol.*, **11**, 1223–1229.
- Jarmoskaite, I., AlSadhan, I., Vaidyanathan, P.P. and Herschlag, D. (2020) How to measure and evaluate binding affinities. *Elife*, **9**, e57264.
- Meier, B., Barber, L.J., Liu, Y., Shtessel, L., Boulton, S.J., Gartner, A. and Ahmed, S. (2009) The MRT-1 nuclease is required for DNA crosslink repair and telomerase activity in vivo in *Caenorhabditis elegans*. *EMBO J.*, **28**, 3549–3563.
- Rivera, T., Haggblom, C., Cosconati, S. and Karlseder, J. (2017) A balance between elongation and trimming regulates telomere stability in stem cells. *Nat. Struct. Mol. Biol.*, **24**, 30–39.
- Mazzucco, G., Huda, A., Galli, M., Piccini, D., Giannattasio, M., Pessina, F. and Doksan, Y. (2020) Telomere damage induces internal loops that generate telomeric circles. *Nat. Commun.*, **11**, 5297.
- Lovejoy, C.A., Takai, K., Huh, M.S., Picketts, D.J. and de Lange, T. (2020) ATRX affects the repair of telomeric dsbs by promoting cohesion and a DAXX-dependent activity. *PLoS Biol.*, **18**, e3000594.
- Oganesian, L. and Karlseder, J. (2011) Mammalian 5' C-rich telomeric overhangs are a mark of recombination-dependent telomere maintenance. *Mol. Cell*, **42**, 224–236.
- Brenes, A., Afzal, V., Kent, R. and Lamond, A.I. (2018) The Encyclopedia of Proteome Dynamics: a big data ecosystem for (prote)omics. *Nucleic Acids Res.*, **46**, D1202–D1209.
- Dietz, S., Almeida, M.V., Nischwitz, E., Schreier, J., Viceconte, N., Fradera-Sola, A., Renz, C., Ceron-Noriega, A., Ulrich, H.D., Kappei, D. *et al.* (2021) The double-stranded DNA-binding proteins TEBP-1 and TEBP-2 form a telomeric complex with POT-1. *Nat. Commun.*, **12**, 2668.
- Lloyd, N.R., Dickey, T.H., Hom, R.A. and Wuttke, D.S. (2016) Tying up the ends: plasticity in the recognition of single-stranded DNA at Telomeres. *Biochemistry*, **55**, 5326–5340.
- Skolakova, P., Foldynova-Trantirkova, S., Bednarova, K., Fiala, R., Vorlickova, M. and Trantirek, L. (2015) Unique *C. elegans* telomeric overhang structures reveal the evolutionarily conserved properties of telomeric DNA. *Nucleic Acids Res.*, **43**, 4733–4745.
- Marqueville, J., De Rache, A., Violet, B., Morvan, E., Mergny, J.L. and Amrane, S. (2022) G-quadruplex structure of the *C. elegans* telomeric repeat: a two tetrads basket type conformation stabilized by a non-canonical C-T base-pair. *Nucleic Acids Res.*, **50**, 7134–7146.
- Bryan, T.M. (2020) G-Quadruplexes at telomeres: friend or foe? *Molecules*, **25**, 3686.
- Kim, W.T., Hennick, K., Johnson, J., Finnerty, B., Choo, S., Short, S.B., Drubin, C., Forster, R., McMaster, M.L. and Hockemeyer, D. (2021) Cancer-associated POT1 mutations lead to telomere elongation without induction of a DNA damage response. *EMBO J.*, **40**, e107346.
- Amberger, J.S., Bocchini, C.A., Schiettecatte, F., Scott, A.F. and Hamosh, A. (2015) OMIM.org: online Mendelian Inheritance in Man (OMIM(R)), an online catalog of human genes and genetic disorders. *Nucleic Acids Res.*, **43**, D789–D798.
- Kim, H., Li, F., He, Q., Deng, T., Xu, J., Jin, F., Coarfa, C., Putluri, N., Liu, D. and Songyang, Z. (2017) Systematic analysis of human telomeric dysfunction using inducible telomerase/shelterin CRISPR/Cas9 knockout cells. *Cell Discov.*, **3**, 17034.
- Shakirov, E.V., Perroud, P.F., Nelson, A.D., Cannell, M.E., Quatrano, R.S. and Shippen, D.E. (2010) Protection of Telomeres 1 is required for telomere integrity in the moss *Physcomitrella patens*. *Plant Cell*, **22**, 1838–1848.
- Saint-Leandre, B. and Levine, M.T. (2020) The telomere paradox: stable genome preservation with rapidly evolving proteins. *Trends Genet.*, **36**, 232–242.