Shelterin reduces the accessibility of telomeric overhangs

Sajad Shiekh [1], **Amanda Jack**[2], **Ayush Saurabh**[3], **Golam Mustafa** [1], **Sineth G. Kodikara**[1], **Prabesh Gyawali**[1], **Mohammed Enamul Hoque**[4], **Steve Pressé**[3,5], **Ahmet Yildiz**[2,6,7,*] **and Hamza Balci** [1,*]

[1]Department of Physics, Kent State University, Kent, OH 44242, USA, [2]Biophysics Graduate Group, University of California, Berkeley, CA 94720, USA, [3]Center for Biological Physics, Department of Physics, Arizona State University, Tempe, AZ 85287, USA, [4]Department of Chemistry and Biochemistry, Kent State University, Kent, OH 44242, USA, [5]School of Molecular Science, Arizona State University, Tempe, AZ 85287, USA, [6]Physics Department, University of California, Berkeley, CA 94720, USA and [7]Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA

ABSTRACT

Telomeres terminate with a 50–300 bases long single-stranded G-rich overhang, which can be misrecognized as a DNA damage repair site. Shelterin plays critical roles in maintaining and protecting telomere ends by regulating access of various physiological agents to telomeric DNA, but the underlying mechanism is not well understood. Here, we measure how shelterin affects the accessibility of long telomeric overhangs by monitoring transient binding events of a short complementary peptide nucleic acid (PNA) probe using FRET-PAINT *in vitro*. We observed that the POT1 subunit of shelterin reduces the accessibility of the PNA probe by ∼2.5-fold, indicating that POT1 effectively binds to and protects otherwise exposed telomeric sequences. In comparison, a four-component shelterin stabilizes POT1 binding to the overhang by tethering POT1 to the double-stranded telomeric DNA and reduces the accessibility of telomeric overhangs by ∼5-fold. This enhanced protection suggests shelterin restructures the junction between single and double-stranded telomere, which is otherwise the most accessible part of the telomeric overhang.

INTRODUCTION

Telomeres are 5-10 kilo basepair long repetitive DNA sequences (GGGTTA in humans) that cap the ends of eukaryotic chromosomes and serve several critical physiological roles in the protection and maintenance of chromosome ends (1,2). Telomeres terminate with a 50-300 nucleotide (nt) long 3′ single-stranded overhang (ssTEL), which plays an active role in successive rounds of cellular replication and DNA preservation (2–4). To maintain overall genomic integrity, telomeres associate with a multi-protein complex called shelterin (5,6). In humans, shelterin consists of six distinct proteins, TRF1, TRF2, TIN2, RAP1, POT1 and TPP1 (5–7). TRF1 and TRF2 recognize and bind to the double-stranded telomere (dsTEL) (8,9), while POT1 binds to ssTEL (10,11). RAP1 forms a complex with TRF2 to regulate its sequence-specific binding affinity (12). TIN2 and TPP1 orchestrate the interactions with TRF2, TRF1, and POT1, and thereby establish a direct protein-mediated link between ssTEL and dsTEL (13,14). Deletion or dysfunction of shelterin proteins activates various DNA damage repair pathways at telomeres (5,7).

The G-rich telomeric overhang is prone to fold into an array of G-quadruplex (GQ) structures. GQs have been reported to form in promoter regions and telomeres (15,16) *in vitro* (17,18) and *in vivo* (19–21). There is growing evidence that GQ formation regulates essential cellular processes such as DNA replication, transcription, and translation (22,23), as well as its pivotal role in telomere maintenance (24). These roles have made GQs potentially important therapeutic targets (25).

The multiple GQs that form in ssTEL present a major hurdle for telomeric DNA synthesis and impede the physical association of telomere-binding proteins that are dedicated to protecting and modulating the organization of telomeres (26–28). These GQs might form compact higher-order structures that further reduce the accessibility of ssTEL (29–32). They also help to distinguish the telomeres from DNA double-strand breaks and prevent activation of superfluous DNA damage response (1,2). Resolving these structures requires helicase unfolding activity (33–35) or

ssDNA binding proteins (28,36). POT1 binding destabilizes telomeric GQs (37–39), which might make the telomeric overhang more accessible for telomerase and other enzymes (36). However, POT1 binding may also serve to protect otherwise exposed segments of ssTEL that are not folded into a GQ structure and thereby reduce telomerase access (40). It also remains unclear how the length of ssTEL tracts modulates the ability of POT1 to regulate access to the overhang.

The compaction of ssTEL into higher-order structures via tandem GQs could potentially contribute to the robust protection of telomeres in coordination with telomere-associated proteins. Using single molecule FRET-PAINT (Förster Resonance Energy Transfer- Point Accumulation in Nanoscale Topography) (41), we recently investigated the accessibility of telomeric overhangs (42). In this assay, transient bindings of a complementary short peptide nucleic acid (PNA) strand to accessible regions of ssTEL are used to characterize the accessibility of the overhang. We showed that a FRET signal is detected between Cy5-PNA and Cy3-labeled telomeric DNA with up to 168 nt long overhang, presumably because ssTEL becomes highly compact upon tandem GQ formation (42). The PNA binding frequency to an overhang that contains four GGGTTA repeats (which can fold into a GQ) was an order of magnitude smaller than that of an overhang that contains a single GGGTTA repeat (cannot form a GQ), even though the 4-GGGTTA overhang contains four times as many potential binding sites (42). Therefore, the majority of ssTEL tracts remain folded into GQs while PNA primarily binds to sites not protected within a GQ (42).

In this study, we investigated the accessibility of telomeric DNA constructs with physiologically relevant overhang lengths (28–148 nt) in the absence and presence of POT1 or a 4-component shelterin that includes TRF1, TIN2, POT1 and TPP1 (shelterin hereafter). These telomeric overhangs can fold into 1–6 tandem GQs. We characterized the frequency and location of PNA binding events before and after introducing POT1 or shelterin and performed a kinetic analysis of these binding events. Shelterin provided more efficient protection and resulted in more significant changes in Cy5-PNA accessibility maps than POT1 alone. Removing the dsTEL segment of the DNA constructs also eliminated the additional protection provided by shelterin. These findings suggest that tethering of POT1 to dsTEL tracts via the rest of shelterin stabilizes its binding to ssTEL and reduces the accessibility of the telomeric overhangs (42).

MATERIALS AND METHODS

Preparation of DNA constructs

Partial duplex DNA (pdDNA) constructs were formed by annealing a long strand that contains the telomeric overhang with a short strand (30 nt) that has a biotin at 3′-end and a Cy3 at 5′-end. Long DNA oligomers (sequences in Supplementary Table S1) were purchased from Eurofins Genomics and were purified in-house using denaturing polyacrylamide gel electrophoresis (PAGE; Supplementary Figure S2). The short strand contains a TAACCCTAACCC sequence at its 5'-side which hybridizes with the complementary sequence on the long strand to create a 12 bp long

dsTEL segment. To illustrate, a pdDNA construct that contains 16 repeats of GGGTTA sequence (hereafter G-Tract) at its overhang was formed by annealing a long strand TGGCGACGGCAGCGAGGC*TTAGGGTTAGGG***TTA (GGGTTA)$_{16}$G** with the short strand Cy3-*CCCTAACCCTAA*GCCTCGCTGCCGTCGCCA-Biotin. The italic sequences are complementary to each other and form the dsTEL segment that enables the binding of TRF1 or TRF2. Similarly, the underlined sequences are complementary to each other and form an 18-bp duplex that registers the annealing of repetitive telomeric sequences. The bold sequences constitute the overhang.

For annealing, the short strand and the long strand were mixed in a molar ratio of 1:4 in a buffer that contains 150 mM KCl and 10 mM $MgCl_2$. This high $MgCl_2$ concentration was used only during annealing assay, while all FRET-PAINT measurements were performed at 2 mM $MgCl_2$. The annealing reaction was performed in a thermocycler by heating the samples at 95°C for 3 min and then gradually cooling them to 30°C in steps of 1°C/3 min.

HPLC-purified TAACCCTT-Cy5 PNA was purchased from PNA-Bio Inc. The underlined sequence is complementary to ssTEL while the last T acts as a spacer. This probe is significantly shorter than PNA strands typically used for targeting and destabilizing GQs (43). The melting temperature (T_m) of the duplex formed by Cy5-PNA and telomeric DNA (12.7 ± 0.3°C) is much lower than $T_m = 69$°C (44) for a telomeric GQ under similar ionic conditions. Therefore, the Cy5-PNA does not introduce a significant destabilization to the telomeric GQs.

Protein constructs

POT1 and 4-component shelterin (POT1, TPP1, TIN2, and either TRF1 or TRF2) were expressed in insect cells, as previously described (45). Briefly, human POT1 with an N terminal ZZ affinity tag, TEV cleavage site, and YBBR labeling site was cloned into a pOmnibac vector, and the four-component shelterin containing either TRF1 or TRF2 was cloned into a BigBac vector, where POT1 was given an N terminal YBBR tag, POT1, TRF1 and TRF2 were each given an N terminal ZZ affinity tag and a TEV cleavage site, and TIN2 and TPP1 were each given an N-terminal His-MBP affinity tag and a TEV cleavage site. Protein was purified from insect cells, as previously described (46). Briefly, plasmids containing genes of interest were transformed into DH10Bac competent cells (Berkeley MacroLab), and Bacmid DNA was purified using ZymoPURE miniprep buffers (Zymo Research, D4210) and ethanol precipitation. Insect cells were transfected using Fugene HD transfection reagent (Promega, E2311). The virus was amplified in progressively larger cultures. 1 ml of the P1 virus was used to infect 50 ml of Sf9 cells at 1 million cells/ml for 72 h. 10 ml of the P2 virus was used to infect 1 l of Sf9 cells at 1 million cells/ml and expression proceeded for 72 h. Cells expressing the protein of interest were harvested at 4000 × g for 10 min and resuspended in 50 ml lysis buffer (50 mM HEPES pH 7.4, 1 M NaCl, 10% glycerol, 1 mM PMSF, 1 mM DTT and 1 tablet of protease inhibitor (Sigma, 4693132001)). Lysis was performed using 15 loose and 15 tight plunges of a Wheaton glass dounce. The lysate was clarified using a

45 min, 360 000 × g spin in a Ti70 rotor. The supernatant was incubated with 1 ml IgG beads (IgG Sepharose 6 Fast Flow, GE Healthcare, 17096902) for POT1 or 1 ml amylose beads (New England BioLabs, E8021S) for shelterin for 1 h. Beads were washed with 40 ml of lysis buffer followed by 40 ml of storage buffer (50 mM Tris pH 7.5, 300 mM KCl, 2 mM MgCl$_2$, 10% glycerol, 1 mM DTT). Beads were then collected and incubated with TEV protease (Berkeley Macrolab, Addgene #8827) for 1 h at room temperature to elute the protein. Cleaved protein was run through a Superdex 200 Increase 10/300 GL size exclusion column (Cytiva, 28-9909-44) to remove TEV, cleaved affinity tags, and shelterin subcomplexes. POT1 and shelterin were aliquoted in a buffer that contains 50 mM Tris (pH 7.5), 300 mM KCl, 2 mM MgCl$_2$ and 10% glycerol, flash frozen and stored at –80°C.

Single-molecule FRET assay and measurements

A home-built prism-type total internal reflection fluorescence (TIRF) microscope was built around an Olympus IX-71 inverted microscope, as previously described (47). Laser-drilled quartz slides were used after they were thoroughly washed with acetone and 1 M potassium hydroxide (KOH), followed by piranha etching for 20 min. After surface functionalization with amino silane for 20–30 min, the slides and coverslips were initially passivated with a mixture of mPEG (PEG-5000, Laysan Bio, Inc.) and biotinylated PEG (biotin-PEG-5000, Laysan Bio, Inc.) in the molar ratio of 40:1 and kept overnight. The slides and coverslips were cleaned and dried with nitrogen gas and stored at –20°C. Before the experiments, the slide and coverslip were passivated with 333 Da PEG for 30–45 min to increase the density of the PEG brush. Finally, the microfluidic chamber was created between the slide and coverslip by placing double-sided tape between them followed by sealing the chamber with epoxy. In addition, the microfluidic chamber was treated with 2% (v/v) Tween-20 to reduce non-specific binding. Following the removal of excess detergent from the chamber, 0.01 mg/ml streptavidin was incubated in the chamber for 2 min. Subsequently, the freshly annealed pdDNA samples were diluted to 10–20 pM in 150 mM KCl and 2 mM MgCl$_2$ and immobilized on the surface for 2–5 min to obtain ∼300 Cy3 spots per imaging area (∼50 μm × 100 μm). Excess or unbound DNA was removed by washing the chamber with a buffer that contains 150 mM KCl and 2 mM MgCl$_2$. All FRET-PAINT measurements were carried out in an imaging buffer (50 mM Tris–HCl pH 7.5, 2 mM Trolox, 0.8 mg/ml glucose, 0.1 mg/ml glucose oxidase, 0.1 mg/ml bovine serum albumin (BSA), 2 mM MgCl$_2$, 150 mM KCl and 40 nM Cy5-PNA). In experiments with proteins, 20 nM POT1 or 20 nM shelterin were also included in the imaging buffer. The Cy5-PNA strand was preheated to 85°C for 10 minutes before it was added to the imaging buffer.

FRET-PAINT measurements with proteins

To avoid variations between surface qualities in different microfluidic chambers, the FRET-PAINT measurements for a given pdDNA were first performed in the absence of proteins in a microfluidic chamber, followed by data acquisition in the presence of POT1 and then shelterin in the same chamber. Proteins were thawed on ice for 2-5 min and diluted to desired concentration in the imaging buffer. The chamber was incubated with the proteins for 10 min before data acquisition. To ensure that bound proteins dissociate from DNA, the chamber was incubated with 150 mM KCl and 100 mM MgCl$_2$ for 15 min. The chamber was then washed with 1000 μL of 150 mM KCl and 2 mM MgCl$_2$ to remove these remnant proteins allow the overhang to refold in the original ionic conditions (Supplementary Figure S3). After waiting for 10 min for refolding of the GQs, new proteins were added to the chamber.

Data acquisition and analysis

The donor fluorophore was excited with a 532 nm laser beam (Spectra-Physics, Excelsior) and the fluorescence signal was collected by an Olympus water immersion objective (60×, 1.20 NA). An Andor Ixon EMCCD camera was used to record movies with 1500-2000 frames at 10 Hz. The recorded movies were processed and analyzed by a custom C++ script. A custom MATLAB code was used to generate and screen single-molecule trajectories of donor and acceptor intensities. The background was subtracted based on the remnant donor and acceptor intensities after donor photobleaches. The selected trajectories were used to generate the FRET histograms. FRET efficiency was calculated by $E_{\text{FRET}} = \frac{I_A}{I_A + I_D}$, where I_A is the acceptor intensity and I_D is the donor intensity. Molecules that did not show any binding event contributed to the donor only (DO) peak at $E_{\text{FRET}} = 0.06$. The DO peak was subtracted from the FRET histograms which were then rescaled such that DO peak corresponds to $E_{\text{FRET}} = 0.00$. The number of molecules used in each FRET histogram is given in Supplementary Table S2.

Binding frequency analysis

An automated and bias-free step detection program, AutoStepfinder (48), was used to determine the E_{FRET} levels and dwell times in the single molecule time traces that were used to construct the FRET histograms. The Cy5-PNA binding frequencies were determined based on the Autostepfinder fits and were calculated by dividing the number of binding events by the sum of each trajectory that shows at least one binding event or donor photobleaching. The error bars for frequency analysis were determined by Bootstrapping analysis (49) with a sample size of 2000 and 95% confidence level (32). For comparing whether the variations between frequencies of different conditions are significant, one-way ANOVA analysis was performed using OriginLab-Origin Pro.

S-parameter calculation

S-parameter was used to quantify the broadness of FRET distributions, which represents the ambiguity associated with predicting the G-Tract that Cy5-PNA binds to, or more generally the uncertainty implied by a particular probability distribution (50). FRET histograms were normalized to 100%. The broadness of FRET distributions was quantified by an S-parameter (Shannon entropy divided by the
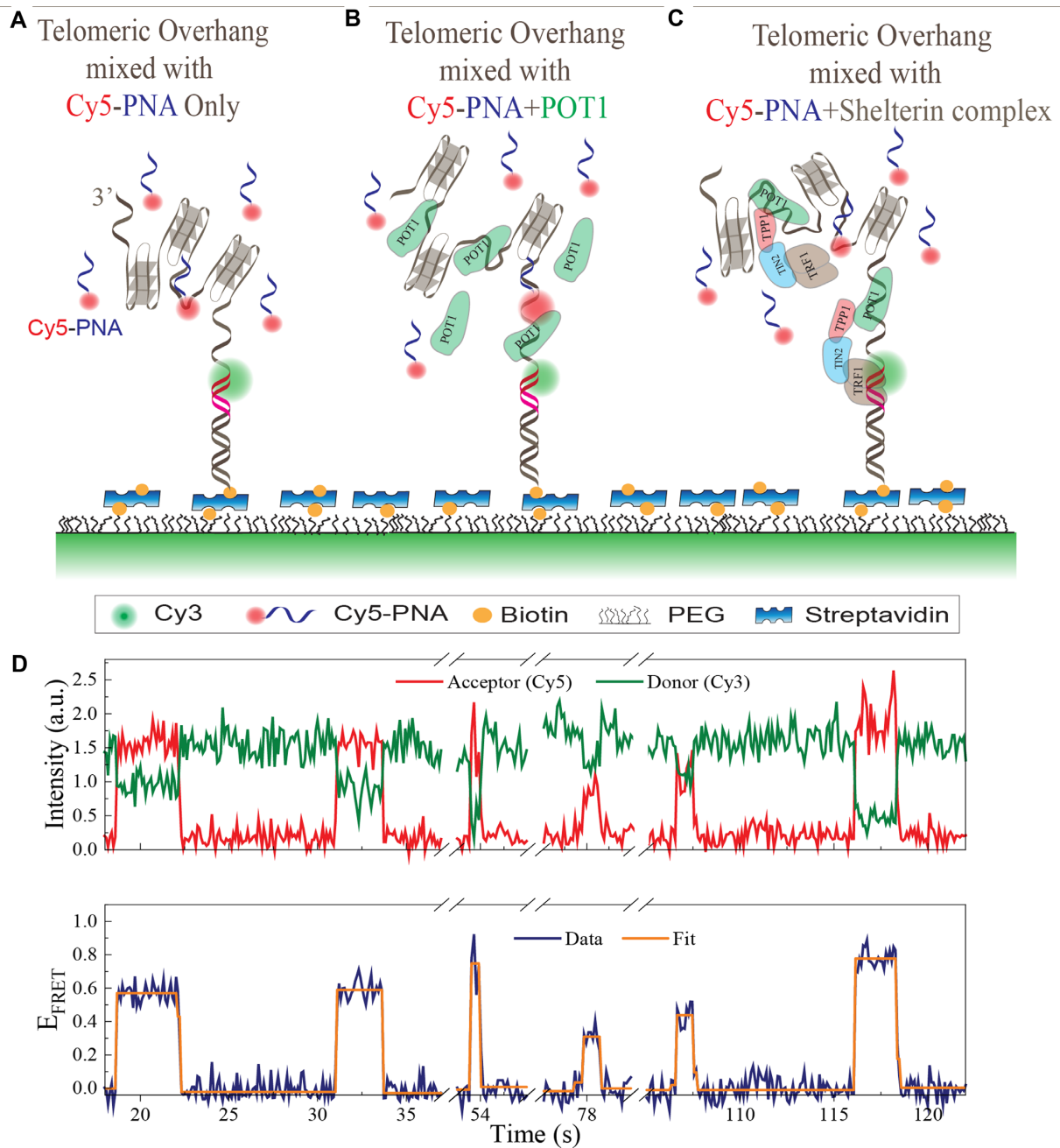
Figure 1. Schematic of FRET-PAINT measurements in the absence or presence of POT1 or shelterin. (**A**) A pdDNA construct labeled with a donor (Cy3) is immobilized on the PEGylated surface via streptavidin–biotin linkage. The pdDNA constructs consist of dsTEL (red) and ssTEL (overhang) with multiple GGGTTA (G-Tract) repeats, which can fold into tandem GQs separated by unfolded regions of varying length. The Cy5-PNA, which is complementary to a 7 nt long ssTEL, is used to access available regions within the overhang. FRET signals with different levels are observed when Cy5-PNA transiently binds to accessible G-Tracts. Schematics of these measurements in the absence of any proteins, in the presence of POT1 and in the presence of shelterin (POT1, TPP1, TIN2 and TRF1) are shown in (**A**), (**B**) and (**C**), respectively. (**D**) Representative smFRET time trace showing six Cy5-PNA binding events. The orange line is a fit to the FRET data and it is used to characterize the corresponding FRET levels.

Boltzmann constant): $S = -\sum_i p_i \ln p_i$, where p_i is the probability of a particular FRET level i. The p_i is obtained by dividing population at the corresponding FRET level by 100 (Figure 4) so that $\sum_{i=1}^{N} p_i = 1$ for a given FRET histogram. The errors in the S-parameter calculation are based on the standard error of the mean of at least four measurements at different parts of the surface for each data point.

Kinetic analysis

To estimate the binding and unbinding rates for each construct, we used Markov Chain Monte Carlo (MCMC) techniques developed in our earlier work (51–53), details of
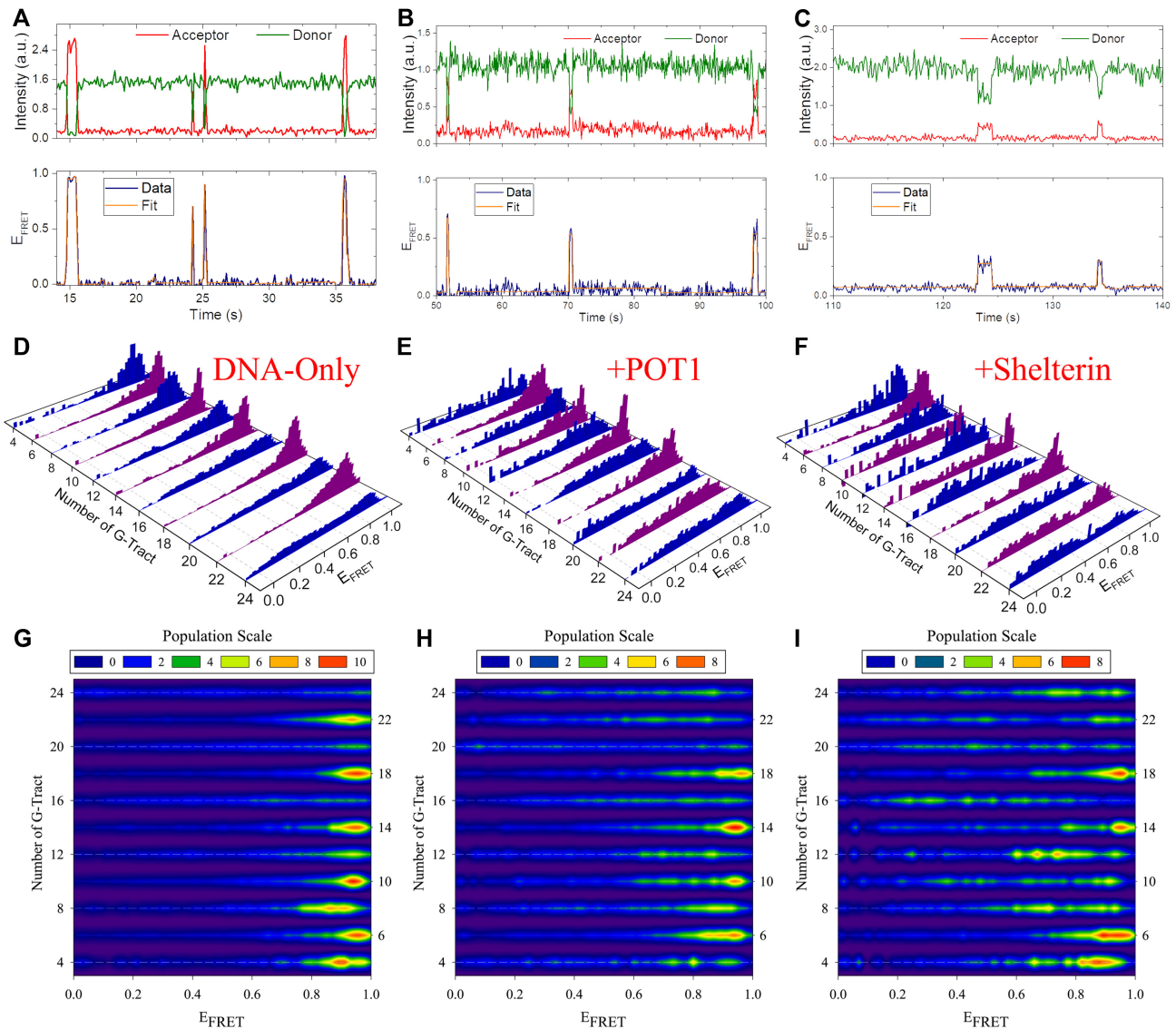
Figure 2. Normalized FRET histograms, contour plots, and sample traces for telomeric overhangs with 4-24 G-Tracts. (**A**–**C**) Sample FRET-PAINT traces for DNA-only, +POT1 and +Shelterin conditions, respectively. (**D**–**F**) FRET histograms are constructed from Cy5-PNA binding to different regions of ssTEL for DNA-only, +POT1 and + Shelterin conditions, respectively. (**G**–**I**) Contour plots of histograms in (D), (E) and (F), respectively. Regions with concentrated red color are highly accessible while broad yellow/green regions indicate lower accessibility to a broader segment of ssTEL.

which are given in the Supplementary Information (Supplementary Figures S10–S20). These techniques implement a Bayesian non-parametric (BNP) framework where the MCMC sampler treats the number of states of a DNA construct itself as an unknown along with kinetic rates and FRET efficiencies. Furthermore, the sampler also explores the space of possible state trajectories for a given smFRET trace to estimate the most probable trajectory. Specifically, for each smFRET time trace, we generated 10 000 samples for state trajectory, rates, and FRET efficiencies. The samples for each DNA construct in the presence or absence of proteins were then collected together in the form of bivariate probability distributions, as shown in Figure 6 and Supplementary Figures S10–S20. These distributions show the probability for a construct to assume specific FRET efficiency and binding/unbinding rate values. We note here that

the probability distributions shown in these figures were smoothened using kernel density estimate (KDE) for visualization purposes only. The scripts for kinetic analysis are available through a GitHub repository at https://github.com/LabPresse/BNP-FRET.

RESULTS

We studied the accessibility of telomeric overhangs with 4–24 G-tracts in the absence or presence of 20 nM POT1 or shelterin using a FRET-PAINT assay (Figure 1) (32,42). The observed FRET levels are correlated with the position of the accessible G-Tracts with respect to the Cy3 donor located at the ssDNA/dsDNA junction. In the current study, we performed these measurements in the presence of POT1, which binds and extends the overhang. Supplementary Fig-
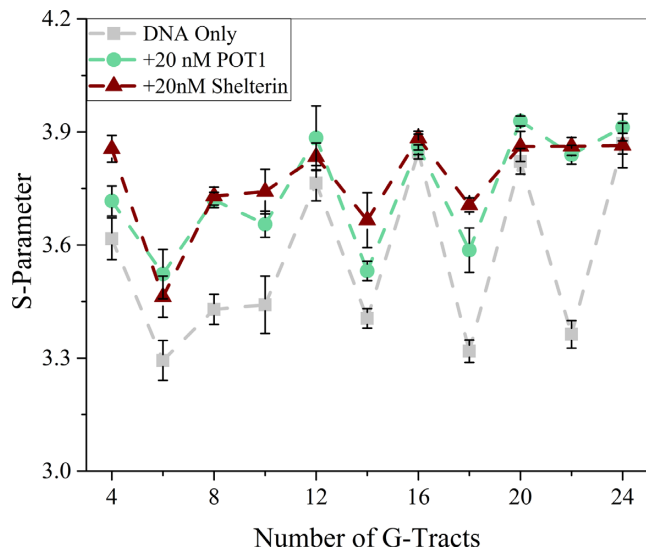
Figure 3. Scatter plot of S-parameter in the absence of proteins (DNA-only) or presence of POT1 or Shelterin. S-parameter measures the broadness of FRET distributions. Larger S-parameters represent distribution of Cy5-PNA binding sites over a greater segment of ssTEL. The variation between the S-parameters for different constructs (different number of G-Tracts) decreases in the presence of POT1 and shelterin, i.e. S-parameters flatten, suggesting the accessibilities of different constructs become more uniform. The error bars are standard error of the mean of at least four different measurements.

ure S4 shows that despite the extension of the overhang by POT1, it is still possible to detect the binding of Cy5-PNA to sites furthest from the donor molecule.

The 20 nM POT1 or shelterin concentration employed in these studies is significantly greater than the measured dissociation constant of POT1 or POT1-TPP1 (1-10 nM range, about 6-10 fold smaller for POT1-TPP1 compared to POT1) (54), and it is similar to the estimated concentration of POT1 in the nucleus of human cells (in the tens of nanomolar range) (55). Consistent with these reports, 64% of 3G-Tract constructs were bound by 5 nM POT1 and 86% are bound by 20 nM POT1 (Supplementary Figure S5A), even though this construct may form a G-Triplex secondary structure that restricts POT1 binding. Supplementary Figure S5B and C shows two example traces illustrating the stability of POT1 binding to the 3G-Tract construct.

We also measured Cy5-PNA binding frequency to 16G-Tract construct under different POT1 concentrations. While the binding frequency decreases as the POT1 concentration is increased until 7.5 nM, it remains constant at higher concentrations (Supplementary Figure S5D), indicating that 20 nM POT1 is adequate to reach a steady state for POT1 binding. However, 20 nM POT1 does not unfold GQ at a detectable range (Supplementary Figure S6).

We next monitored transient binding events of Cy5-PNA to DNA constructs with different overhang lengths in the absence or presence of 20 nM POT1 or shelterin. We analyzed FRET-PAINT trajectories that demonstrate transient binding events (Figure 2A–C) and created normalized FRET distributions (Figure 2D–F) and contour plots (Figure 2G–I) to infer how the accessible G-Tracts are distributed through the overhang. In the absence of proteins

(Figure 2D and G), the FRET histograms for [4n]G-Tract constructs (n = 3, 4, 5, 6) are broadly distributed, suggesting that accessible G-Tracts are distributed throughout the overhang. The FRET distributions for [4n + 2]G-Tract constructs are concentrated at high FRET levels, suggesting that most accessible sites are in the vicinity of ssDNA/dsDNA junction region. However, POT1 reduces the contrast between the accessibility maps of [4n]G-Tract and [4n + 2]G-Tract constructs (Figure 2E and H), and shelterin largely eliminates these differences (Figure 2F and I).

We introduced an S-parameter to quantify how broadly the FRET values are distributed in each condition (Figure 3). Broader FRET distributions indicate that the binding sites are distributed over a more extensive part of the overhang, resulting in higher S-parameter. In the absence of proteins, the S-parameter was maximum at [4n]G-Tract constructs and minimum at [4n + 2]G-Tract constructs, with large variations between these two groups of constructs. These variations significantly diminish in the presence of POT1 and are almost eliminated by shelterin (Figure 3). These changes in the S-parameter distributions can be quantified by the variance of the S-parameter data, which shows a minimum in the presence of shelterin (6.5×10^{-3}) and maximum in the absence of proteins (6.2×10^{-2}). We also calculated the weighted average, variance, skewness, and kurtosis of the FRET histograms (Supplementary Figure S1, Supplementary Table S3). Similar to the S-parameter analysis, the variance of the FRET distribution of [4n + 2]G-Tract constructs was substantially lower than that of [4n]G-Tract constructs.

To quantify telomere accessibility, we calculated the Cy5-PNA binding frequency for different overhang lengths in the presence or absence of proteins (Figure 4A) and quantified the 'relative accessibility', defined as the ratio of the binding frequencies observed in the presence of proteins to those observed in their absence (Figure 4B). Mean relative accessibilities were reduced by ∼2.5-fold in the presence of 20 nM POT1 and ∼5-fold in the presence of 20 nM shelterin (one-way ANOVA, $P < 0.001$), indicating that POT1 and to a greater extent shelterin binds to ssTEL and antagonizes PNA binding to telomeric overhang.

We also investigated whether tethering of POT1 to dsTEL via TPP1-TIN2-TRF1 impacts the accessibility of the telomeric overhang. We performed measurements in the presence of POT1 and TRF1, without TPP1 and TIN2, which are required for establishing a connection between POT1 and TRF1 (Figure 5A–C). First, TRF1 alone does not reduce the accessibility of ssTEL compared to the DNA-only case (one-way ANOVA, $P < 0.001$) (Supplementary Figure S7). Second, adding 20 nM TRF1 along with 20 nM POT1 does not reduce the accessibility beyond that provided by the POT1-only case (one-way ANOVA, $P < 0.001$). These results suggest that the connection between TRF1 and POT1, via TPP1 and TIN2, is critical in achieving the protection provided by the shelterin.

We also explored the possibility that TIN2 and TPP1 increase the telomere binding affinity of POT1 (54). To test this possibility, we created analogs of 16G-Tract and 18G-Tract constructs that do not contain dsTEL (Figure 5D), hence preventing tethering of POT1 to dsTEL by TPP1-TIN2-TRF1. The relative accessibilities for these modified
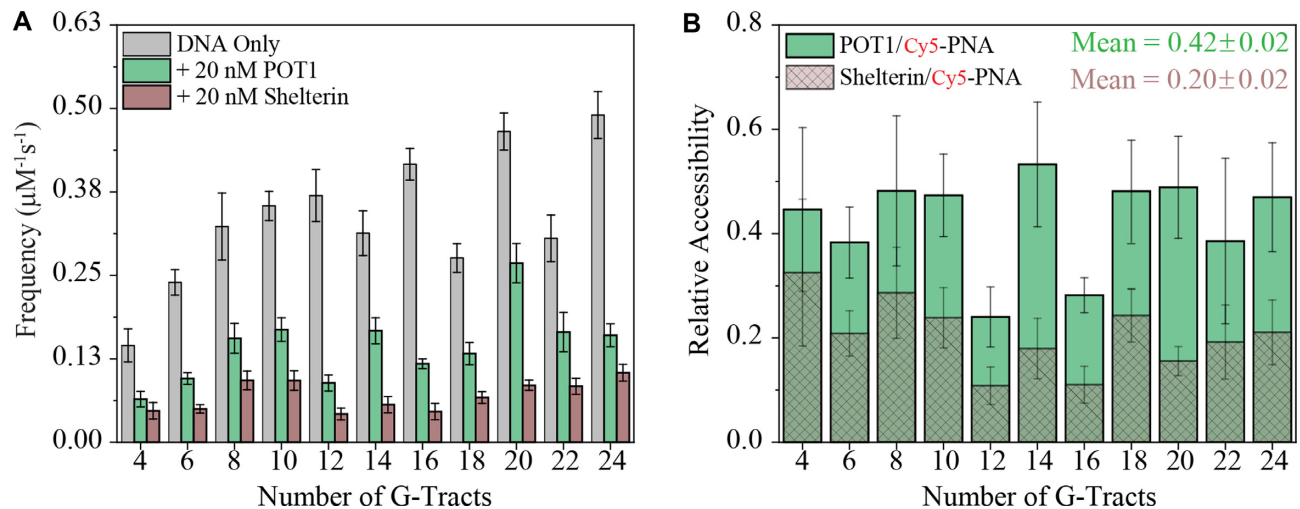
Figure 4. Cy5-PNA binding frequencies and relative accessibility of telomeric DNA overhangs. (**A**) Binding of Cy5-PNA is quantified by analyzing the binding events in the absence of proteins and in the presence of POT1 or shelterin. (**B**) Relative accessibility represents the ratio of binding frequencies in the presence and absence of proteins. Mean accessibility is reduced ∼2.5-fold in the presence of POT1 and ∼5-fold in the presence of shelterin. The error bars for frequency data in (A) are obtained from bootstrapping, while those for relative accessibility in (B) are obtained by propagation of errors based on respective frequency data.

constructs in the presence of POT1 were similar to those for unmodified constructs (Figure 5E, F). However, in the absence of dsTEL tracts, shelterin does not reduce the accessibility any more than that of the POT1-only case. These results demonstrate that the accessibility of telomere overhangs is reduced by the connection between proteins that are bound to ssTEL and dsTEL, and not by the altered ssTEL binding affinity of POT1 by other shelterin components.

We also tested if TRF2, instead of TRF1, would make a difference in binding frequencies or relative accessibilities of telomere overhangs. We repeated similar measurements with shelterin containing TRF2 (TRF2, TIN2, POT1, TPP1) using 14G-Tract and 16G-Tract constructs (Supplementary Figures S8 and S9). These measurements did not show a significant difference between TRF1 or TRF2 containing shelterin, further supporting our conclusion that the connection between ssTEL and dsTEL is sufficient to reduce the accessibility of ssTEL overhang.

Finally, we performed a kinetic analysis of Cy5-PNA binding events using MCMC sampling from a Bayesian Hidden Markov model. Figure 6 shows estimated state trajectories (Figure 6A–C) and bivariate probability distributions for escape rates (Figure 6D–F) for the 20G-Tract construct, while Supplementary Figures S10–S20 show the corresponding analysis for all constructs. The intense band around $E_{FRET} \approx 0.0$ indicates the transitions that take place from the donor-only level (before Cy5-PNA binding) to all other levels. The transitions from all other levels are almost exclusively back to the donor-only level, as transitions to any other state require overlapping of multiple Cy5-PNA binding events, which have a negligible probability. Most binding rates (escape from $E_{FRET} \approx 0.0$ state) from the donor-only level have rates <0.1 s^{-1}, indicating extended time intervals between successive Cy5-PNA binding events. Escape rates from other states are broadly distributed around 1.0 s^{-1}, suggesting ∼1.0 s dwell time in

these FRET levels before they transition back to the donor-only level.

DISCUSSION

Our comparative FRET-PAINT measurements on long telomeric overhangs provide important insights about how the distribution (location of accessible sites) and frequency (level of accessibility) of Cy5-PNA binding events are impacted by POT1 or shelterin. We previously interpreted the observed accessibility patterns in the absence of proteins in terms of elevated levels of folding frustration in [4n]G-Tract constructs, which results from combinatorial effects associated with the finite length of the overhang and the position of folded GQs on the overhang (42,56). Computational modeling suggested that the ssDNA/dsDNA junction region has the least folding stability while the 3' end has the maximum stability and that neighboring GQs interact with positive cooperativity (42).

In the current work, we primarily focus on the impact of POT1 and shelterin on the accessibility to ssTEL. Interestingly, introducing POT1 and shelterin largely eliminated the dependence of the S-parameter on the number of G-tracts in the telomeric overhang. As mentioned, variations in the S-parameter are primarily due to how the accessible sites are distributed, e.g. whether they are concentrated in the junction region (lower S-parameter) or distributed throughout the overhang (higher S-parameter). In the presence of POT1, the S-parameters of [4n + 2] G-Tract constructs increase and become similar to those of [4n]G-Tract constructs. This requires effective protection of the most accessible sites (in this case the junction region) by POT1, so that their accessibility is reduced to similar levels as other sites.

The reduction in Cy5-PNA binding frequency (Figure 4) was not readily anticipated because POT1 has a mild GQ destabilization activity (27,38,57), which would have re-
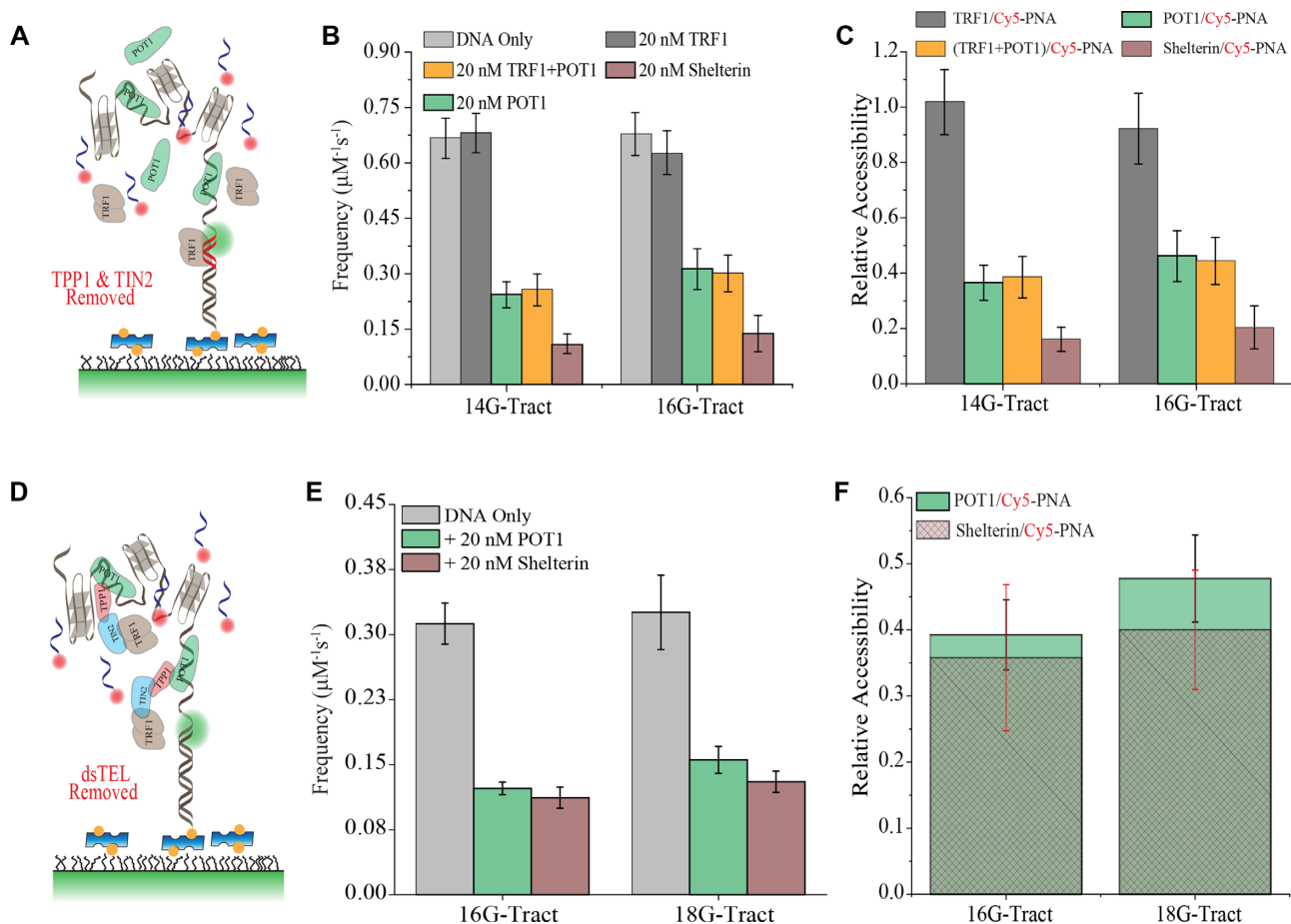
Figure 5. Shelterin-mediated tethering of POT1 to dsTEL is required to reduce overhang accessibility. (**A**) Schematics of FRET-PAINT assay where POT1 and TRF1 are used in the absence of TPP1 and TIN2. (**B**) Binding frequencies for 14G-Tract and 16G-Tract constructs. Frequencies for TRF1-only case are similar to those of DNA-only case while those of POT1 and TRF1 are similar to those of POT1-only case. (**C**) Relative accessibility for 14G-Trct and 16-G-Tract constructs. Relatively accessibility of the case when both TRF1 and POT1 are included is significantly less than that of shelterin. The errors in relative accessibility are propagated from the data and errors in the frequency of each condition presented in (B). (**D**) Schematics of FRET-PAINT assay for modified DNA constructs that do not contain dsTEL. (**E**) Binding frequencies for these modified 16G-Tract and 18G-Tract constructs (lacking dsTEL) in the absence of proteins and in the presence of POT1 or shelterin (POT1, TPP1, TIN2 and TRF1). (**F**) Relative accessibility for modified 16G-Tract and 18G-Tract constructs. Relativity accessibility of shelterin is similar to that of POT1, suggesting the dsTEL is critical for the enhanced protection provided by shelterin. The error bars for frequency data in (B) and (E) are obtained from bootstrapping, while those for relative accessibility in (C) and (F) are obtained by propagation of errors based on respective frequency data.

sulted in elevated exposure of ssTEL to Cy5-PNA unless such exposed sequences are effectively covered by POT1. Considering the 7-nt complementarity of Cy5-PNA and ssTEL and the minimum 9-nt binding site of POT1 (5′-TAGGGTTAG) (11), a tight competition between the two would have been expected. However, the ∼2.5 fold reduction in Cy5-PNA accessibility in the presence of POT1 suggests that POT1 efficiently binds and protects exposed telomeric sequences.

The further flattening of the S-parameter in the presence of shelterin compared to the POT1-only case indicates that members of shelterin also play a role in the protection of the overhang. Because they do not directly bind ssTEL, it is not clear how TRF1, TIN2 and TPP1 could impact accessibility of ssTEL. Because ssTEL/dsTEL junction is the most accessible region of ssTEL (42), a connection between TRF1 and POT1 through TPP1 and TIN2 at this junction might be an effective way to enable robust protection of ssTEL.

Interestingly, destabilization of GQs that might form in this region due to their proximity to dsTEL may be facilitating the establishment of this critical connection between shelterin components. Consistent with this possibility, disrupting the potential link between POT1 and dsTEL eliminates the enhanced protection provided by shelterin compared to the POT1-only case (Figure 5).

In this single molecule investigation, we demonstrated that a coordinated effort between members of the shelterin that bind to ssTEL and dsTEL reduces the accessibility of telomeric overhangs by ∼5-fold. We also showed that the connection between POT1 and TRF1, via TPP1 and TIN2, is required for the shelterin-induced protection of the overhang. These results suggest a consistent and compelling picture where the nature and strength of interactions between shelterin proteins and telomeric DNA complement the thermodynamic stabilities and folding characteristics of the tandem telomeric GQs. This complementarity, in return, may
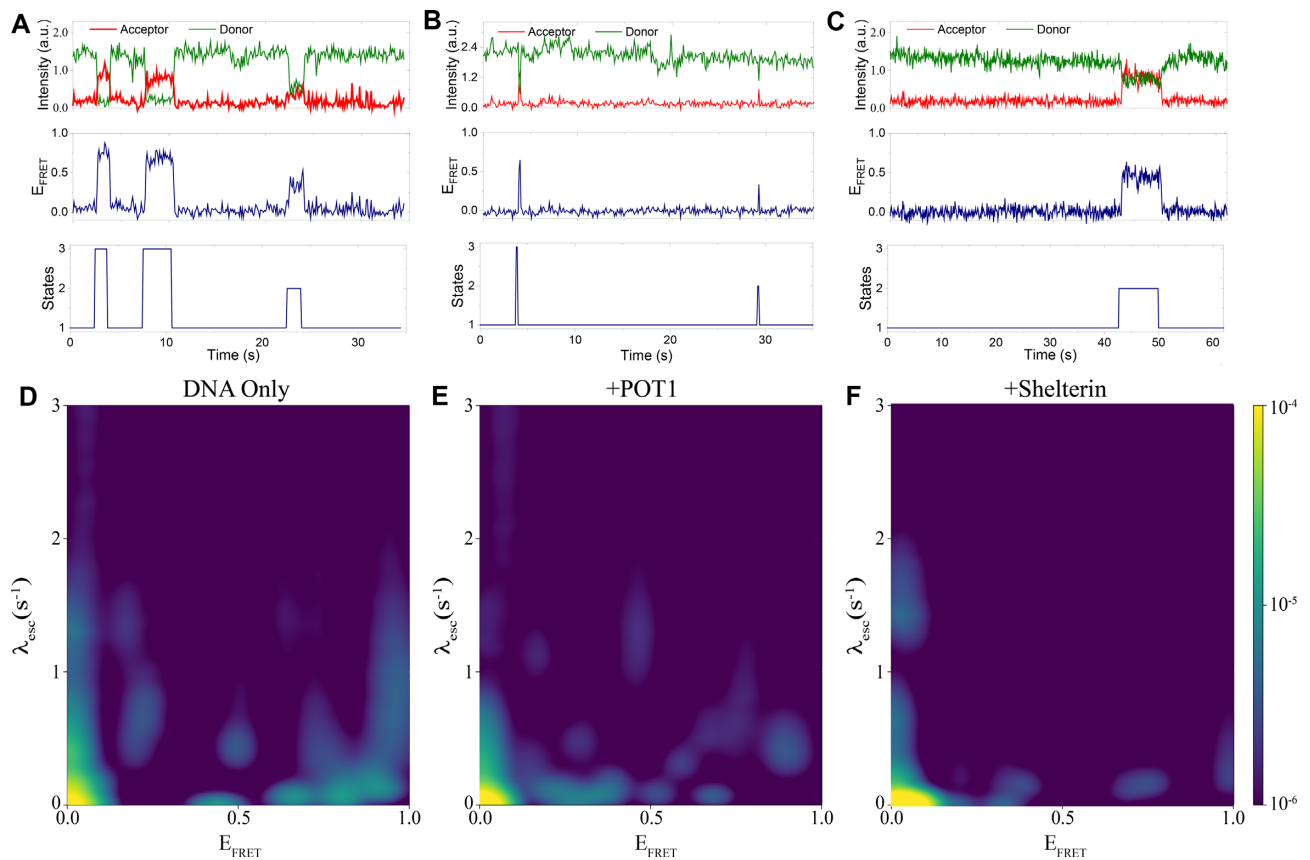
Figure 6. Estimated state trajectories and bivariate probability distributions (posterior) for escape rates (λ_{esc}) for 20G-Tract DNA construct. The trajectories and distributions were obtained using the Bayesian non-parametric FRET sampler utilizing MCMC techniques. In the top panel, we show example smFRET traces from the datasets for DNA-Only case (**A**), +POT1 case (**B**) and +shelterin case (**C**), in addition to the corresponding most probable of all the state trajectories. In these state trajectories, the molecules transition from level 1 ($E_{FRET} \approx 0.0$, donor only level) to other FRET levels (levels 2 and 3) and vice versa. In the bottom panel, we show probability distributions for the DNA only case (**D**), +POT1 case (**E**) and +shelterin case (**F**). As expected for 20G-Tract, we see a landscape of FRET efficiencies for the bound states. To clearly distinguish probability regions of different orders of magnitude we have used logarithmic scale for color map. Since binding events are rare in general, the region around $E_{FRET} \approx 0.0$ dominates the three distributions. Moreover, due to the more effective protection of binding sites by proteins, we expect fewer binding events in the case of shelterin compared to the DNA-only and POT1 cases. This is clearly evident from (F) which shows the smaller probability for the 20G-Tract construct to be in non-zero FRET efficiency regions compared to (D) and (E).

enable effective protection of these critical regions of the genome against DNA damage repair.

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Meyne,J., Ratliff,R.L. and Moyzis,R.K. (1989) Conservation of the human telomere sequence (TTAGGG)(n) among vertebrates. *Proc. Natl. Acad. Sci. U.S.A.*, **86**, 7049–7053.
2. Blackburn,E.H. (1991) Structure and function of telomeres. *Nature*, **350**, 569–573.
3. Stewart,S.A., Ben-Porath,I., Carey,V.J., O'Connor,B.F., Hahn,W.C. and Weinberg,R.A. (2003) Erosion of the telomeric single-strand overhang at replicative senescence. *Nat. Genet.*, **33**, 492–496.
4. Verdun,R.E. and Karlseder,J. (2007) Replication and protection of telomeres. *Nature*, **447**, 924–931.
5. Bandaria,J.N., Qin,P., Berk,V., Chu,S. and Yildiz,A. (2016) Shelterin protects chromosome ends by compacting telomeric chromatin. *Cell*, **164**, 735–746.
6. De Lange,T. (2005) Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev.*, **19**, 2100–2110.

7. Lim,C.J. and Cech,T.R. (2021) Shaping human telomeres: from shelterin and CST complexes to telomeric chromatin organization. *Nat. Rev. Mol. Cell Biol.*, **22**, 283–298.

8. Fairall,L., Chapman,L., Moss,H., De Lange,T. and Rhodes,D. (2001) Structure of the TRFH dimerization domain of the human telomeric proteins TRF1 and TRF2. *Mol. Cell*, **8**, 351–361.

9. Griffith,J.D., Comeau,L., Rosenfield,S., Stansel,R.M., Bianchi,A., Moss,H. and De Lange,T. (1999) Mammalian telomeres end in a large duplex loop. *Cell*, **97**, 503–514.

10. Lei,M., Baumann,P. and Cech,T.R. (2002) Cooperative binding of single-stranded telomeric DNA by the Pot1 protein of Schizosaccharomyces pombe. *Biochemistry*, **41**, 14560–14568.

11. Loayza,D., Parsons,H., Donigian,J., Hoke,K. and De Lange,T. (2004) DNA binding features of human POT1: a nanomer 5′-TAGGGTTAG-3′ minimal binding site, sequence specificity, and internal binding to multimeric sites. *J. Biol. Chem.*, **279**,13241–13248.

12. Palm,W. and De Lange,T. (2008) How shelterin protects mammalian telomeres. *Annu. Rev. Genet.*, **42**,301–334.

13. Ye,J.Z.S., Donigian,J.R., Van Overbeek,M., Loayza,D., Luo,Y., Krutchinsky,A.N., Chait,B.T. and De Lange,T. (2004) TIN2 binds TRF1 and TRF2 simultaneously and stabilizes the TRF2 complex on telomeres. *J. Biol. Chem.*, **279**,47264–47271.

14. O'Connor,M.S., Safari,A., Xin,H., Liu,D. and Songyang,Z. (2006) A critical role for TPP1 and TIN2 interaction in high-order telomeric complex assembly. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 11874–11879.

15. Huppert,J.L. and Balasubramanian,S. (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.*, **35**, 406–413.

16. Schaffitzel,C., Berger,I., Postberg,J., Hanes,J., Lipps,H.J. and Plückthun,A. (2001) In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 8572–8577.

17. Henderson,E., Hardin,C.C., Walk,S.K., Tinoco,I. and Blackburn,E.H. (1987) Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine·guanine base pairs. *Cell*, **51**, 899–908.

18. Sen,D. and Gilbert,W. (1988) Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature*, **334**, 364–366.

19. Biffi,G., Tannahill,D., McCafferty,J. and Balasubramanian,S. (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.*, **5**, 182–186.

20. Henderson,A., Wu,Y., Huang,Y.C., Chavez,E.A., Platt,J., Johnson,F.B., Brosh,R.M., Sen,D. and Lansdorp,P.M. (2014) Detection of G-quadruplex DNA in mammalian cells. *Nucleic Acids Res.*, **42**, 860–869.

21. Di Antonio,M., Ponjavic,A., Radzevičius,A., Ranasinghe,R.T., Catalano,M., Zhang,X., Shen,J., Needham,L.M., Lee,S.F., Klenerman,D. *et al.* (2020) Single-molecule visualization of DNA G-quadruplex formation in live cells. *Nat. Chem.*, **12**, 832–837.

22. Marsico,G., Chambers,V.S., Sahakyan,A.B., McCauley,P., Boutell,J.M., Antonio,M.Di and Balasubramanian,S. (2019) Whole genome experimental maps of DNA G-quadruplexes in multiple species. *Nucleic Acids Res*, **47**, 3862–3874.

23. Maizels,N. (2015) G4-associated human diseases. *EMBO Rep*, **16**, 910–922.

24. Wang,Y., Yang,J., Wild,A.T., Wu,W.H., Shah,R., Danussi,C., Riggins,G.J., Kannan,K., Sulman,E.P., Chan,T.A. *et al.* (2019) G-quadruplex DNA drives genomic instability and represents a targetable molecular abnormality in ATRX-deficient malignant glioma. *Nat. Commun.*, **10**, 943.

25. Hänsel-Hertsch,R., Di Antonio,M. and Balasubramanian,S. (2017) DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential. *Nat. Rev. Mol. Cell Biol.*, **18**, 279–284.

26. Arnoult,N. and Karlseder,J. (2015) Complex interactions between the DNA-damage response and mammalian telomeres. *Nat. Struct. Mol. Biol.*, **22**, 859–866.

27. Taylor,D.J., Podell,E.R., Taatjes,D.J. and Cech,T.R. (2011) Multiple POT1-TPP1 proteins coat and compact long telomeric single-stranded DNA. *J. Mol. Biol.*, **410**, 10–17.

28. Hwang,H., Kreig,A., Calvert,J., Lormand,J., Kwon,Y., Daley,J.M., Sung,P., Opresko,P.L. and Myong,S. (2014) Telomeric overhang length determines structural dynamics and accessibility to telomerase and ALT-associated proteins. *Structure*, **22**, 842–853.

29. Chandradoss,S.D., Haagsma,A.C., Lee,Y.K., Hwang,J.H., Nam,J.M. and Joo,C. (2014) Surface passivation for single-molecule protein studies, *J. Vis. Exp.*, 50549.

30. Neidle,S. (2016) Quadruplex nucleic acids as novel therapeutic targets. *J. Med. Chem.*, **59**, 5987–6011.

31. Abraham Punnoose,J., Cui,Y., Koirala,D., Yangyuoru,P.M., Ghimire,C., Shrestha,P. and Mao,H. (2014) Interaction of G-quadruplexes in the full-length 3' human telomeric overhang. *J Am Chem Soc*, **136**, 18062–18069.

32. Mustafa,G., Shiekh,S., Gc,K., Abeysirigunawardena,S. and Balci,H. (2021) Interrogating accessibility of telomeric sequences with FRET-PAINT: evidence for length-dependent telomere compaction. *Nucleic Acids Res.*, **49**, 3371–3380.

33. Budhathoki,J.B., Stafford,E.J., Yodh,J.G. and Balci,H. (2015) ATP-dependent G-quadruplex unfolding by Bloom helicase exhibits low processivity. *Nucleic Acids Res.*, **43**, 5961–5970.

34. Paeschke,K., Bochman,M.L., Garcia,P.D., Cejka,P., Friedman,K.L., Kowalczykowski,S.C. and Zakian,V.A. (2013) Pif1 family helicases suppress genome instability at G-quadruplex motifs. *Nature*, **497**, 458–462.

35. Huber,M.D., Lee,D.C. and Maizels,N. (2002) G4 DNA unwinding by BLM and Sgs1p: substrate specificity and substrate-specific inhibition. *Nucleic Acids Res.*, **30**, 3954–3961.

36. 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.

37. Ray,S., Bandaria,J.N., Qureshi,M.H., Yildiz,A. and Balci,H. (2014) G-quadruplex formation in telomeres enhances POT1/TPP1 protection against RPA binding. *Proc. Natl. Acad. Sci. U. S. A.*, **111**, 2990–2995.

38. Chaires,J.B., Gray,R.D., Dean,W.L., Monsen,R., Deleeuw,L.W., Stribinskis,V. and Trent,J.O. (2020) Human POT1 unfolds G-quadruplexes by conformational selection. *Nucleic Acids Res.*, **48**, 4976–4991.

39. Hwang,H., Buncher,N., Opresko,P.L. and Myong,S. (2012) POT1-TPP1 regulates telomeric overhang structural dynamics. *Structure*, **20**, 1872–1880.

40. Xu,M., Kiselar,J., Whited,T.L., Hernandez-Sanchez,W. and Taylor,D.J. (2019) POT1-TPP1 differentially regulates telomerase via POT1 His266 and as a function of single-stranded telomere DNA length. *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 23527–23533.

41. Lee,J., Park,S. and Hohng,S. (2018) Accelerated FRET-PAINT microscopy. *Mol. Brain*, **11**, 70.

42. Shiekh,S., Mustafa,G., Kodikara,S.G., Hoque,M.E., Yokie,E., Portman,J.J. and Balci,H. (2022) Emerging accessibility patterns in long telomeric overhangs. *Proc. Natl. Acad. Sci. U. S. A.*, **119**, e2202317119.

43. Kushon,S.A., Jordan,J.P., Seifert,J.L., Nielsen,H., Nielsen,P.E. and Armitage,B.A. (2001) Effect of secondary structure on the thermodynamics and kinetics of PNA hybridization to DNA hairpins. *J. Am. Chem. Soc.*, **123**, 10805–10813.

44. Qureshi,M.H., Ray,S., Sewell,A.L., Basu,S. and Balci,H. (2012) Replication protein A unfolds G-quadruplex structures with varying degrees of efficiency. *J. Phys. Chem. B*, **116**, 5588–5594.

45. Jack,A., Kim,Y., Strom,A.R., Lee,D.S.W., Williams,B., Schaub,J.M., Kellogg,E.H., Finkelstein,I.J., Ferro,L.S., Yildiz,A. *et al.* (2022) Compartmentalization of telomeres through DNA-scaffolded phase separation. *Dev. Cell*, **57**, 277–290.

46. Ferro,L.S., Can,S., Turner,M.A., Elshenawy,M.M. and Yildiz,A. (2019) Kinesin and dynein use distinct mechanisms to bypass obstacles. *Elife*, **8**, e48629.

47. Maleki,P., Budhathoki,J.B., Roy,W.A. and Balci,H. (2017) A practical guide to studying G-quadruplex structures using single-molecule FRET. *Mol. Genet. Genomics*, **292**, 483–498.

48. Loeff,L., Kerssemakers,J.W.J., Joo,C. and Dekker,C. (2021) AutoStepfinder: a fast and automated step detection method for single-molecule analysis. *Patterns*, **2**, 100256.

49. Simonoff,J.S., Efron,B., Tibshirani,R.J. and Hjorth,J.S.U. (1994) An Introduction to the Bootstrap. *J. Am. Stat. Assoc.*, **89**, 1559.

50. Phillips,R., Kondev,J., Theriot,J. and Garcia,H.G. (2014) *Physical Biology of the Cell*. The University of Chicago Press.

51. Jazani,S., Sgouralis,I., Shafraz,O.M., Levitus,M., Sivasankar,S. and Pressé,S. (2019) An alternative framework for fluorescence correlation spectroscopy. *Nat. Commun.*, **10**, 3662.

52. Tavakoli,M., Jazani,S., Sgouralis,I., Shafraz,O.M., Sivasankar,S., Donaphon,B., Levitus,M. and Pressé,S. (2020) Pitching single-Focus confocal data analysis one photon at a time with bayesian nonparametrics. *Phys. Rev. X*, **10**, 011021.

53. Saurabh,A., Safar,M., Sgouralis,I., Fazel,M. and Pressé,S. (2022) Single photon smFRET. I. Theory and conceptual basis. bioRxiv doi: https://doi.org/10.1101/2022.07.20.500887 23 November 2022, preprint: not peer reviewed.

54. Wang,F., Podell,E.R., Zaug,A.J., Yang,Y., Baciu,P., Cech,T.R. and Lei,M. (2007) The POT1-TPP1 telomere complex is a telomerase processivity factor. *Nature*, **445**, 506–510.

55. Takai,K.K., Kibe,T., Donigian,J.R., Frescas,D. and de Lange,T. (2011) Telomere protection by TPP1/POT1 requires tethering to TIN2. *Mol. Cell*, **44**, 647–659.

56. Carrino,S., Hennecker,C.D., Murrieta,A.C. and Mittermaier,A. (2021) Frustrated folding of guanine quadruplexes in telomeric DNA. *Nucleic Acids Res.*, **49**, 3063–3076.

57. Wang,H., Nora,G.J., Ghodke,H. and Opresko,P.L. (2011) Single molecule studies of physiologically relevant telomeric tails reveal POT1 mechanism for promoting G-quadruplex unfolding. *J. Biol. Chem.*, **286**, 7479–7489.