

The solution structures of higher-order human telomere G-quadruplex multimers

Robert C. Monsen¹, Srinivas Chakravarthy², William L. Dean³, Jonathan B. Chaires^{1,3,4,*} and John O. Trent^{1,3,4,*}

¹Department of Biochemistry & Molecular Genetics, University of Louisville Medical School, Louisville, KY 40202, USA, ²The Biophysics Collaborative Access Team (BioCAT), Department of Biological Chemical and Physical Sciences, Illinois Institute of Technology, Chicago, IL 60616, USA, ³James Graham Brown Cancer Center, University of Louisville Medical School, Louisville, KY 40202, USA and ⁴Department of Medicine, University of Louisville Medical School, Louisville, KY 40202, USA

Received November 14, 2020; Revised December 21, 2020; Editorial Decision December 22, 2020; Accepted January 13, 2021

ABSTRACT

Human telomeres contain the repeat DNA sequence 5'-d(TTAGGG), with duplex regions that are several kilobases long terminating in a 3' single-stranded overhang. The structure of the single-stranded overhang is not known with certainty, with disparate models proposed in the literature. We report here the results of an integrated structural biology approach that combines small-angle X-ray scattering, circular dichroism (CD), analytical ultracentrifugation, size-exclusion column chromatography and molecular dynamics simulations that provide the most detailed characterization to date of the structure of the telomeric overhang. We find that the single-stranded sequences 5'-d(TTAGGG)_n, with $n = 8, 12$ and 16 , fold into multimeric structures containing the maximal number (2, 3 and 4, respectively) of contiguous G4 units with no long gaps between units. The G4 units are a mixture of hybrid-1 and hybrid-2 conformers. In the multimeric structures, G4 units interact, at least transiently, at the interfaces between units to produce distinctive CD signatures. Global fitting of our hydrodynamic and scattering data to a worm-like chain (WLC) model indicates that these multimeric G4 structures are semi-flexible, with a persistence length of ~ 34 Å. Investigations of its flexibility using MD simulations reveal stacking, unstacking, and coiling movements, which yield unique sites for drug targeting.

INTRODUCTION

Telomeres are structures found at the end of eukaryotic chromosomes which protect genomic DNA from degradation, end-to-end fusion, and homologous recombination (1,2). The human telomere consists of the repeat d(TTAGGG)_n, and ranges from 5 to 25 kb in length with an extended single-stranded 3' overhang of a few hundred bases in non-germ cells (3). This locus has long been associated with human diseases, such as cancer (4) and telomeroopathies (5), as well as aging (6) and general genome homeostasis (7). In normal somatic cells, each round of cellular division results in a shortening of the telomere due to the so-called end replication problem—a mechanism believed to be protective against uncontrolled replication (8). Once the telomere has become critically short in normal (non-stem) cells, a DNA damage response is triggered, resulting in uncapping of the telomere-bound shelterin proteins and, eventually, apoptosis (2,8). Cancer cells avoid this fate by utilizing mechanisms that restore telomere length. In >85% of cancers, this is accomplished by reactivating human telomerase reverse transcriptase (hTERT), a ribonucleoprotein that extends the telomere 3' overhang (9–11). G-quadruplex (G4) formation in the telomere overhang can inhibit hTERT binding and extension function (12). Treating cells with telomere G4-specific small molecules leads to uncapping of the shelterin proteins and a sequestering of the free single-stranded telomere overhang, ultimately resulting in a telomere-specific DNA damage response (13–15). These findings have made telomere G4 an attractive cancer target (15).

G-quadruplexes form in guanine-rich sequences, in which guanine tracts interact to form square planar tetrads (G-tetrads) that stack atop one another and are stabilized by coordinating cations, pi-stacking interactions, and a Hoogsteen hydrogen bonding network (16). Many telomere G4

*To whom correspondence should be addressed. Tel: +1 502 852 2194; Fax: +1 502 852 7979; Email: john.trent@louisville.edu
Correspondence may also be addressed to Jonathan B. Chaires. Tel: +1 502 852 1172; Fax: +1 502 852 7979; Email: j.chaires@louisville.edu
Present address: Clinical Translational Research Building, 505 S. Hancock St, University of Louisville Medical School, Louisville, KY 40202, USA.

topologies have been characterized at the atomic level by X-ray crystallography and NMR studies. These studies have demonstrated that the monomeric form of the human telomere can exist as parallel (17), hybrid 3+1 (18,19), antiparallel (20), and two-tetrad antiparallel (21) structures under various ionic and crowding conditions. The Yang lab (18,19,22), Patel lab (23–25) and we (26) have since shown that the wild-type telomere adopts primarily the hybrid-1 and hybrid-2 topologies in physiologically relevant solution conditions. The Yang lab has shown by NMR that *in vitro* the wild-type monomeric telomere sequence of the form d(TTAGGG)₄T exists in a dynamic equilibrium of hybrid-2 (~75%) and hybrid-1 (~25%) (27).

Telomere G-quadruplexes have also been observed directly in cells. *In vivo*, G4-specific antibodies and fluorescent ligands have confirmed the formation of telomere G4s (28–30). Using the sequence d(AGGG(TTAGGG)₃), Hong-Liang and colleagues used ¹⁹F NMR cell studies to show that the hybrid-1, -2 and a two-tetrad antiparallel (hybrid-3), but not the parallel or antiparallel basket topologies, spontaneously form when injected into live HeLa cells (31). Altogether, these studies demonstrate that the most physiologically and thermodynamically relevant monomeric telomere conformations are of the hybrid type.

The minimal telomere G-quadruplex repeats in K⁺ buffers show only small energy differences between the hybrid forms (19). The formation of each topology appears to be driven by the presence or absence of flanking nucleotides on the core sequence d(GGG(TTAGGG)₃) (19,21,22). The hybrid-1 and hybrid-2 topologies have three parallel and one antiparallel strand that differ in their loop arrangements (32). Hybrid-1 has G-tracts connected consecutively with a propeller and two lateral loops (22) while the hybrid-2 is connected consecutively with two lateral loops followed by a propeller loop (19,25). In the WT sequence, these two forms are in dynamic equilibrium with each other, with hybrid-2 being the major form in the extended sequence (19). The differential stabilization of these two structures appears to be due to G-tetrad capping structures formed by the loop and terminal nucleotides (27). In the case of hybrid-1, which was solved using an artificial adenine flanked sequence d(AAAGGG(TTAGGG)₃AA) (22), it was revealed that the 5' G-tetrad was capped by a naturally occurring adenine triad, in which the third adenine is involved. Conversely, in the case of hybrid-2 (solved using the extended WT sequence d((TTAGGG)₄TT) a 3' G-tetrad capping structure was observed that consists of a T:A:T triad in which dT25 is involved. The third natural topology is a two-tetrad anti-parallel basket G-quadruplex (21,33), referred to here as hybrid-3 (otherwise known as 'natural form 3'). The hybrid-3 quadruplex exhibits two lateral loops with one diagonal loop and extensive capping networks at both 5' and 3' ends (21,33). Using the natural telomere sequence d(GGG(TTAGGG)₃T), Zhang *et al.* have shown that the hybrid-3 form accounts for as much as 65% of the G-quadruplex population in solution, and that the 3' dT22 residue is critical for its stabilization (21). However, the authors also point out that any addition of residues to the 5' end abolish this structure. For instance, adding the naturally preceding dA to this sequence reduces its formation to around 10% of the total G4 in solution.

In all other sequences with flanking nucleotides the hybrid-3 was undetectable. Thermodynamic characterization of various telomere monomer G-quadruplexes with a singular value decomposition analysis of CD melting data revealed that the hybrid-1 [PDB ID 2GKU, 5'-d(TTGGG(TTAGGG)₃A) and hybrid-3 (PDB IDs 2KKA and 2KF8, 5'-d(AGGG(TTAGGG)₃T) and d(GGG(TTAGGG)₃T), respectively] are of the highest stability, in that order (34). Although, since the 2GKU sequence is non-natural this might suggest that the hybrid-3 is overall of highest stability, however, physiologically the hybrid-3 may be irrelevant, as all endogenous telomeres are flanked at their 5' ends. The native telomere monomer sequences evaluated with extended 5' regions (5'-TA or 5'-TTA) revealed similar stabilities of hybrid-1 and hybrid-2 conformations (34). Altogether, these studies indicate that the hybrid-1 and hybrid-2 conformations are favored in the context of the extended 3' telomere overhang *in vitro*.

Conservative estimates of the length of the single-stranded overhang of the human telomere in fibroblasts indicate that the sequence exceeds the ~30 nucleotides necessary for formation of a single telomere G-quadruplex. Estimates of 'normal' single-stranded overhangs range from ~50 to >600 nucleotides (3,35), supporting the possibility of multiple G4s forming in tandem. There have been few attempts to characterize these systems at the atomic level because of the difficulties involving guanine imino overlap and structural polymorphism, which hamper NMR studies (27), and the difficulty of obtaining quality crystals for X-ray diffraction (26). Elucidating this higher-order structure is important, as its role in mediating interactions with shelterin proteins, single-stranded binding proteins, and telomerase is critical in maintaining genomic integrity (1,36,37).

To date, varieties of techniques have been applied to the study of the extended human telomere sequences. In 2006, using a combination of gel electrophoresis, CD, and UV-melting, Yu *et al.* proposed that the telomere multimer of the form d(TTAGGG)_n, where *n* is 4, 8 or 12, maximizes its usage of G-tracts by forming a 'beads-on-a-string' assembly of a variety of telomere topologies (parallel, antiparallel, and hybrid) (38). The 'beads-on-a-string' model implies that G4 units are independent and noninteracting. In 2009, Renčuk *et al.*, using CD and PAGE experiments, came to the same general conclusion that the higher-order telomere is capable of folding into multiple conformations in K⁺ buffers (parallel, antiparallel, or hybrid) but also that G4 units have the potential to interact and stack, depending on the amount of macromolecular crowding (39). The same year Xu *et al.* demonstrated the formation of higher-order G-quadruplex formation in 96 nucleotide (nt) long telomere sequences by atomic force microscopy (AFM), which provided the first direct low-resolution evidence of the 'beads-on-a-string' arrangement (40). That study did not appear to have sufficient resolution to provide the topologies of the G4 subunits. In 2013, Hansel *et al.* investigated the monomeric and higher-order human telomere sequences using NMR with site-specific ¹⁵N-labeled DNA in living *X. laevis* oocytes and *in vitro* using steady state fluorescence measurements of incorporated 2-aminopurine (41). In this study, the authors show that the parallel form does not exist in cells, consistent with results from ¹⁹F NMR of monomer

sequences (31), and also give evidence for the simultaneous existence of both hybrid-2 and hybrid-3 forms in both internal and terminal telomeric G4s. A single molecule force extension study by Punnoose and colleagues showed that telomere sequences of ~144 nt in length primarily exist as a ‘beads-on-a-string’ assembly with only a minor (~5%) population of G4 units showing any stacking interactions (42). The preference for non-interacting ‘beads-on-a-string’ conformations was also suggested by recent thermodynamic studies (43). Collectively, the above studies support the maximization of G-quadruplex formation in extended human telomere G-rich sequences. Further, these studies have demonstrated that the existence of the hybrid type G4s, but not parallel or 3-tetrad antiparallel, are the major telomere conformations under physiological buffer conditions.

In contrast to the ‘beads-on-a-string’ arrangement, some attempts at directly visualizing the higher order telomere G4 indicate that G-tract utilization is not maximized (i.e. gaps exist between G4 units). In 2010, Wang *et al.* investigated the extended telomere sequences d(TTAGGG)_n (where *n* = 4, 8 and 16) by AFM (44). While the particle sizes of the single G4 units overall agreed with prior AFM measurements of telomere G-quadruplexes (40), the authors report that the extended sequences ‘rarely’ form the maximum number of G4s. Other low-resolution studies using electron microscopy (EM) and single-molecule force ramp assays (45,46) have reached a similar conclusion. Collectively, these three studies indicate that large gap regions exist between subsequent G4 moieties. However, the techniques utilized in the latter studies often suffer from inherently harsh or non-solution experimental conditions. For instance, preparation of samples during negative stain EM requires washing away of buffer and cations with water, followed by ethanol washes which can dehydrate G4 structures and alter their conformations (47). All three techniques require the sample to physically interact with either grids, labels, or probes, bringing into question how such interactions might impinge on proper folding and stability (47,48).

Our prior biophysical studies investigating the secondary and tertiary structure of the higher-order telomere sequences d(TTAGGG)_n and d(TTAGGG)_nTT, where *n* = 4, 8, 12, 16 and 32, gave evidence that these sequences maximize G-tract usage, and preferentially form a mixture of the hybrid-1 and hybrid-2 conformations (49–51), in line with prior high resolution EPR analyses (52). Subsequent investigations by molecular dynamics (MD) simulations (53), analytical ultracentrifugation (AUC) (51), and differential scanning calorimetry (DSC) (50) studies indicated that, overall, the extended telomere G4s adopt compact, somewhat rod-like structures via stacking interactions between G4 subunits and intervening TTA linkers (50,53). The best-fit models from these analyses were alternating (5′) hybrid-1 (3′) hybrid-2, referred to as hybrid-12 and hybrid-121, for *n* = 8 and *n* = 12 runs, respectively. Interestingly, thermodynamic studies of these two higher-order systems revealed that ‘each quadruplex in the higher-order structures is not independent and identical but is thermodynamically unique and is influenced by its neighbors’ (50), which is inconsistent with a purely ‘beads-on-a-string’ arrangement. Clearly, there is disagreement about the higher-order telomere’s behavior in solution. Low-resolution imag-

ing and single-molecule studies would suggest a highly flexible beads-on-a-string arrangement with large gaps occurring between G-quadruplexes, whereas other hydrodynamic and spectroscopic investigations suggest a more rigid structure, with maximal G-quadruplex formation. Additionally, none of the above studies, apart from MD simulations (53), have directly assessed the dynamics and interactions of the putative inter-G4 stacking junctions.

Using an integrative structural biology approach (54,55), which combines CD, hydrodynamics, molecular dynamics, and small-angle X-ray scattering (SAXS), we show that the telomeric sequences form the maximal number of G4 units without any long gaps. Modeling the hydrodynamic and scattering-derived properties of sequences from 24 nt to 96 nt to a worm-like chain (WLC) model reveals a persistence length of ~34 Å, which is in between that of single-stranded DNA (ssDNA) (~22 Å) (56) and double-stranded DNA (dsDNA) (~550 Å) (57), indicating that the extended telomere G4 is semi-flexible. This flexibility is consistent with MD simulations, which show transient stacking interfaces that create potentially unique binding grooves useful in drug targeting. We follow this with an extensive sequence analysis of the sequence d(TTAGGG)₈ to determine the major constituent G4 topologies. Using CD and mutational analyses we show that the higher-order human telomere is composed of a ratio of hybrid-1 (~25%) and hybrid-2 (~75%) topologies. Our results are in excellent agreement with prior hydrodynamic, AFM, laser tweezer, EPR, and NMR analyses of the human telomere sequences (19,24,40–42,51,52). The resulting structural ensembles provide the first ‘medium-resolution’ atomistic modeling of the conformational heterogeneity and dynamics of the higher-order telomere G-quadruplex.

MATERIALS AND METHODS

Oligonucleotides

Oligonucleotide sequences were purchased from IDT (Integrated DNA Technologies, Coralville, IA, USA) with standard desalting. Upon receipt, stock oligos were dissolved in MilliQ ultrapure water (18.2 MΩ × cm at 25°C) at 1 mM and stored at –20.0°C until use. All experiments were carried out in a potassium phosphate buffer (8 mM HPO₄²⁻, 185 mM K⁺, 15 mM Na⁺, 1 mM EDTA²⁻, pH 7.2). Folding was achieved by diluting stock oligos into buffer and boiling in a water bath for 20 min, followed by slow cooling overnight. Purification was achieved using size exclusion chromatography (SEC) as detailed previously (58). Briefly, oligos were annealed at concentrations of 40–60 μM, filtered through 0.2 μm filters, and injected onto an equilibrated Superdex 75 16/600 SEC column (GE Healthcare 28-9893-33) using a Waters 600 HPLC system. The flow rate was maintained at 0.5 ml/min and sample fractions were collected every 2 min from 100 to 180 min run time. The molecular weights of fractionated species were estimated based on a regression analysis of elution time versus log(MW) of protein standards (Sigma #69385), the major folded species were visually evident as symmetric peaks when monitored at 260 nm (or 280 nm for protein standards). Fractionated samples were pooled and stored at

4°C prior to concentration. Where applicable, pooled fractions were concentrated using Pierce protein concentration devices with 3k MWCO (Thermo #88512, #88515 and #88525) which were rinsed free of glycerol. For AUC and SEC-SAXS experiments, samples were dialyzed after concentration using Spectra/Por Float-A-Lyzers G2 3.5 kDa (Sigma #Z726060) in order to buffer match. Concentrations were determined using molar extinction coefficient given in Table 1.

Size exclusion chromatography (SEC) determination of Stokes radii

Elution times from re-injections of SEC purified fractions were used in the method of Irvine (59) to determine Stokes radii, which were converted to translational diffusion (D_t) coefficients for use in Multi-HYDFIT hydrodynamic modeling (see below). Stokes radii were determined from a regression analysis of $\log(R_s)$ vs. K_d (the distribution coefficient) of protein standards (Sigma #69385).

Analytical ultracentrifugation (AUC)

Sedimentation velocity (SV) experiments were performed in a Beckman Coulter ProteomeLab XL-A analytical ultracentrifuge (Beckman Coulter Inc., Brea, CA) at 20.0°C and 40 000 rpm in standard 2-sector cells using either an An60Ti or An50Ti rotor. Samples were equilibrated in the rotor at 20.0°C for at least 1 h prior to the collection of 100 scans over an 8-h period. Initial analyses were performed in SEDFIT (60) using the continuous C(s) model with resolution 100 and S range from 0 to 10. A partial specific volume of 0.55 ml/g for DNA G-quadruplexes was used as previously determined (51). The Tel72 and Tel96 sequence sedimentation coefficients were additionally corrected for any concentration-dependence using three separate concentrations (Supplementary Figure S2).

Circular dichroism

CD spectra were collected on a Jasco-710 spectropolarimeter (Jasco Inc. Eason, MD, USA) equipped with a Peltier thermostat regulated cell holder equilibrated to 20.0°C. Spectra were collected using the following instrument parameters: 1 cm path length quartz cuvettes, 1.0 nm step size, 200 nm/min scan rate, 1.0 nm bandwidth, 2 s integration time and 4 scan accumulation. Spectra were corrected by subtracting a buffer blank and normalized to molar circular dichroism ($\Delta\epsilon$, M⁻¹ cm⁻¹) based on DNA strand concentration using the following equation:

$$\Delta\epsilon = \theta / (32982cl)$$

where θ is ellipticity in millidegrees, c is molar DNA concentration in mol/l and l is the path length of the cell in cm. Comparison or fitting of CD spectra with their monomer theoretical spectra was done manually in Microsoft Excel using spectra from a previously reported database (61). Residual sum of squares (RSS) analysis of the CD $\Delta\epsilon$ 'residuals' was carried out and plotted in Origin 2020.

Size exclusion chromatography resolved small angle X-ray scattering (SEC-SAXS)

SAXS was performed at BioCAT (beamline 18ID at the Advanced Photon Source, Chicago, USA) with in-line size exclusion chromatography. Samples in BPEK buffer (8 mM HPO₄²⁻, 185 mM K⁺, 15 mM Na⁺, 1 mM EDTA²⁻, pH 7.2) were loaded onto an equilibrated Superdex 75 10/300 GL column, which was maintained at a constant flow rate of 0.7 ml/min using an AKTA Pure FPLC (GE Healthcare Life Sciences) and the eluate after it passed through the UV monitor was directed through the SAXS flow cell, which consists of a 1 mm ID quartz capillary with 50 μm walls. A co-flowing buffer sheath was used to separate the sample from the capillary walls, helping to prevent radiation damage (62). Scattering intensity was recorded using a Pilatus3 1M (Dectris) detector which was placed 3.5 m from the sample giving access to a q -range of 0.004–0.4 Å⁻¹. A series of 0.5 s exposures was acquired every 2 s during elution and data was reduced using BioXTAS RAW 1.6.3 (63). Buffer blanks were created by averaging regions flanking the elution peak and subtracted from exposures selected from the elution peak to create the $I(q)$ versus q curves used for subsequent analyses. More information on SAXS data collection, reduction and interpretation can be found in Supplementary Table S1. SAXS sample preparation, analysis, data reduction, and data presentation has been done in close accordance with recent guidelines (64).

Molecular dynamics simulations and hydrodynamic calculations

Molecular dynamics simulations were carried out on Tel48, Tel72 and Tel96 constructs created previously (49), or modeled based on their solution NMR structures from the Protein Data Bank using the following IDs: 2GKU (hybrid-1) (23), 2JSL (hybrid-2) (24). Base modifications and optimization of starting configurations were performed in UCSF Chimera v1.12 (65) or Maestro v11.8 (66). The partial negative charges of carbonyls at the center of tetrads were neutralized with coordinated potassium counter-ions added manually in Maestro with subsequent minimization prior to simulation. The PDB structures created were then imported into the xleap module of AMBER 2018 (67), neutralized with K⁺ ions, and solvated in a rectangular box of TIP3P water molecules with a 12 Å buffer distance. All simulations were equilibrated using sander at 300 K and 1 atm using the following steps: (i) minimization of water and ions with weak restraints of 10.0 kcal/mol/Å on all nucleic acid residues (2000 cycles of minimization, 500 steepest decent before switching to conjugate gradient) and 10.0 Å cutoff, (ii) heating from 0 to 100 K over 20 ps with 50 kcal/mol/Å restraints on all nucleic acid residues, (iii) minimization of the entire system without restraints (2500 cycles, 1000 steepest decent before switching to conjugate gradient) with 10 Å cutoff, (iv) heating from 100 to 300 K over 20 ps with weak restraints of 10.0 kcal/mol/Å on all nucleic acid residues and (v) equilibration at 1 atm for 100 ps with weak restraints of 10.0 kcal/mol/Å on nucleic acids. The resulting coordinate files from equilibration were then used as input for 100 ns of unrestrained, solvated MD simulations using pmemd with GPU acceleration in the isother-

(2JSL, Tel48) or *DAMMIF* (Tel72, Tel96) bead model generation. Relevant parameters, anisotropy assumptions, normalized spatial discrepancy values (NSDs), χ^2 values, and resolutions are given in Supplementary Table S1. Single best fit models for each telomere construct were determined using the initial pool of conformers derived from MD simulations (or NMR structures for 2JSL) and calculated using *CRY SOL* (75). The best fit structure was determined by minimization of a χ^2 function:

$$\chi^2(r_o, \delta_\rho) = \frac{1}{N_p} \sum_{i=1}^{N_p} \left(\frac{I_{\text{exp}}(q_i) - cI(q_i, r_o, \delta_\rho)}{\sigma(q_i)} \right)^2$$

where $I_{\text{exp}}(q_i)$ and $I(q_i)$ are the experimental and computed profiles, respectively, $\sigma(q_i)$ is the experimental error of the measured profile, N_p is the number of points in the profile, and c is the scaling factor. Two other parameters, r_o and δ_ρ , are fitted and represent the effective atomic radius and the hydration layer density, respectively.

Flexibility analyses by swollen Gaussian chain and WLC models

Fitting of the radii of gyration, as measured by SEC-SAXS for 2JSL, Tel48, Tel72 and Tel96 was performed as outlined recently by Capp *et al.* (76) using the following relationship describing the stiffness and conformational space of a swollen Gaussian coil:

$$R_g = l_p \sqrt{\frac{N^{2v}}{(2v+1)(2v+2)}}$$

where l_p is the persistence length and v is the Flory coefficient. The R_g values with their respective errors were plotted against their G4 number and fit using a non-linear least squares fitting procedure in Origin 2020 (OriginLab Corporation, Northampton, MA, USA).

For the worm-like chain (WLC) modeling, a global analysis of three properties was used in the program Multi-HYDFIT (77,78). Measurements of two other properties, diffusion coefficient, D_t (calculated from measured Stokes radii, R_s , via the Stokes-Einstein equation), and corrected sedimentation coefficient, $S_{20,w}$, were obtained for each sequence using SEC (average \pm S.D. of four measurements each) or AUC (concentration series extrapolated to infinite dilution \pm standard error from regression analysis), respectively. Each value, with respective weighting and molecular weight (MW), was used as the input for the Multi-HYDFIT program. Multi-HYDFIT uses comparisons of the so-called equivalent radii and ratios of radii to calculate theoretical values of R_g , D_t and $S_{20,w}$, which are then compared to that of the measured values. The ratios of radii are directly related to the ratios of length to diameter (L/d) and length to persistence length (L/l_p). With starting estimates of l_p , d , and mass per unit length (M_L), the multi-HYDFIT procedure seeks to minimize a target function (78):

$$\Delta^2(l_p, M_L, d) = \frac{1}{N_s} \sum_{i=1}^{N_s} \left[\left(\frac{\sum_Y w_Y}{\sum_Y w_Y} \right)^{-1} \sum_Y w_Y \left(\frac{a_{Y(\text{cal})} - a_{Y(\text{exp})}}{a_{Y(\text{exp})}} \right)^2 \right]$$

where N_s is the number of samples of different MW, w_Y is the weighting, and a_Y is the ratio of radii for each property.

Table 2. Table of properties derived from Multi-HYDFIT fitting of the higher-order telomere experimental properties to a worm-like chain model

HYDFIT worm-like chain results	
Diameter (Å)	40 (± 5)
Persistence length (Å)	33 (± 3)
Mass per unit length (Da/Å)	163 (± 15)
Deviation from exp. equiv. radii (%)	3.7

In this equation, the outermost sum runs over the N_s samples and the innermost sum runs over the available properties of each sample. The Δ^2 is a mean-square relative deviation for the data, and 100Δ is the percent difference between experimental and theoretical values over the entire set. Additional information is required for the calculation, such as temperature (here 20.0°C was used), solvent viscosity (0.00995 poise), starting guesses for diameter, d (10–100 Å), mass per unit length, M_L (10–300 Da/Å), and persistence length, l_p (20–100 Å). Intrinsic viscosities calculated from best-fit models using HYDROPRO10 were also included with modest weighting, as they were not empirically determined but rather derived from SAXS best-fit models. The goal of the procedure is to determine the best-fit values of the latter properties, which are given in Table 2.

Force of bending curves generated in Figure 8 were calculated using the literature persistence length values for single- and double-stranded DNA at cationic conditions similar to used here (56,57), based on the relationship (79):

$$F_{\text{bend}} = \frac{1}{2} k_B T L_p R^{-2}$$

where F_{bend} is the bending force in picoNewtons, k_B is the Boltzmann constant, T is temperature in Kelvin, L_p is the persistence length in meters, and R is the radius of the arc of a curve in meters. The data was plotted such that the values on the X-axis correspond to the end-to-end length of the polymer curved 180° around the arc of a semi-circle.

Molecular visualizations

All molecular visualizations of MD trajectories and models and RMSD calculations were performed in UCSF Chimera v1.12 (65).

RESULTS

Small-angle X-ray scattering reveals G4 maximization and indicates that the higher-order telomere G4s are semi-flexible

To verify that the extended telomere sequences are in fact maximizing their G-tract usage we employed size-resolved small-angle X-ray scattering (SEC-SAXS) to assess each sequence for size, shape, and compactness (80,81). The results of the SEC-SAXS analysis for sequences 2JSL (hybrid-2), Tel48, Tel72 and Tel96 (Table 1) are shown in Figure 1, Supplementary Figure S1, and Supplementary Table S1. Figure 1A shows the scattering intensity as a function of momentum transfer (q) on a log–log scale for each sequence. Each scattering profile proceeds horizontally to the Y-axis at low values of q , indicating the absence of inter-particle interactions or repulsions (80). Scattering from 2JSL shows a distinct smooth curvature at higher q values which is indicative of a globular particle, whereas the extended telomere

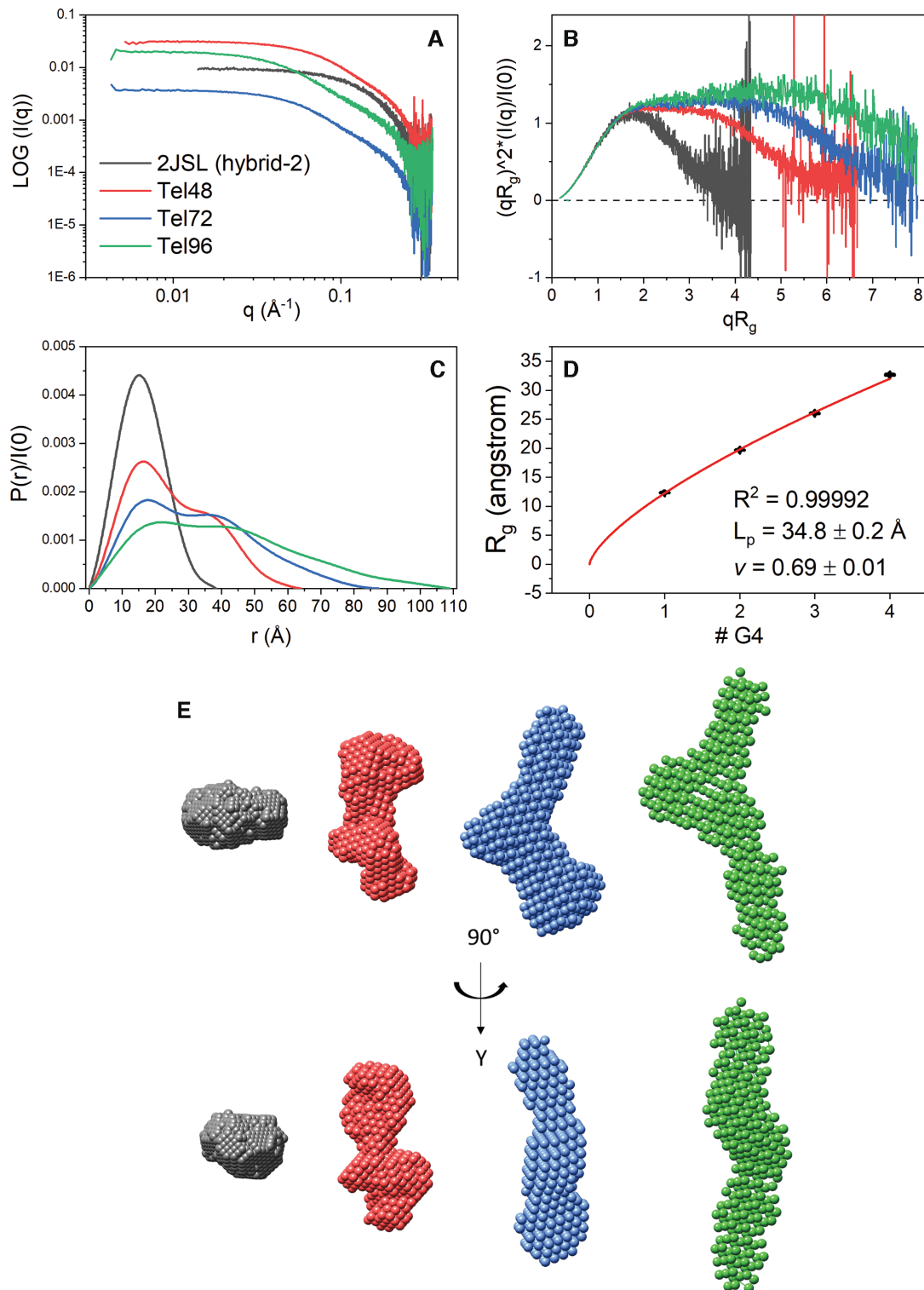


Figure 1. SEC-SAXS analysis of 2JSL (gray), Tel48 (red), Tel72 (blue) and Tel96 (green). **(A)** Log-log plot of the scattering intensity versus scattering vector, q . **(B)** Dimensionless Kratky plots of data in A. **(C)** Pair distribution function plots of data in A normalized to $I(0)$. **(D)** Scatter plot of the radii of gyration from each sequence as a function of G-quadruplex motif fit to a swollen Gaussian chain polymer model (see Materials and Methods) with (inset) derived persistence length (L_p) and Flory coefficient (ν). **(E)** DAMMIN and DAMMIF *ab initio* space-filling models from the data in C.

sequences deviate from this curvature between about 0.05 and 0.2 q , suggesting a non-globular structure (81).

Two useful transformations of the scattering data are the Kratky plot and distance distribution, $P(r)$, plots (Figure 1B and C), which allow for qualitative appraisal of compactness and overall structure, respectively (81). In Figure 1B the dimensionless Kratky plot shows that 2JSL (gray) exhibits a nearly perfect Gaussian distribution that returns to baseline at high qR_g , confirming that it is globular and folded (81). The Tel48 and Tel72 sequences also approach baseline at high qR_g , indicating that they are folded and do not contain significant amounts of flexibility (81). The higher-order sequences also exhibit distinct plateau regions above 2 qR_g , indicating that they have non-globular shapes and are likely multi-domain, consistent with tandem G4 domains. However, Tel96 (green) exhibits a slight rise in its plateau towards higher qR_g , indicating that it is flexible relative to 2JSL, Tel48 and Tel72. Figure 1C shows the corresponding $P(r)$ distributions (normalized to scattering intensity, $I[0]$), which are probability distributions of the inner-atomic distances within each macromolecule (80). 2JSL (gray) again exhibits a symmetric distribution, indicative of a globular molecule (81). Conversely, the extended sequences are all multi-modal. Tel48 exhibits a biphasic distribution (red) indicating a characteristic dumbbell-like tertiary arrangement (81), consistent with two G4 domains separated by a small linker region. Tel72 and Tel96 have tri- and tetra-phasic curves, respectively, which we take as indicating three and four contiguous globular domains in tandem, respectively.

$P(r)$ distributions also allow for quantitative characterization of macromolecules. The point on the X-axis at which each sequence converges to zero is the maximum diameter, D_{\max} , which is the diameter of the particle's longest axis (81). The D_{\max} of each sequence increases approximately linearly with a ~ 24 Å increase with each additional G4 motif. Any substantial amount of telomere species with gaps, or non-maximization of G4s, would likely result in a non-linearity (as well as large upticks in the Kratky curves at high qR_g). The radius of gyration, R_g , is the root mean square distance of the macromolecule's parts from its center of mass and reflects the particle's size (80). The R_g can be calculated by either the Guinier approximation (from plots shown in Supplementary Figure S1) or directly from its $P(r)$ distribution, the latter of which is thought to be more representative in cases where flexibility is assumed (although both values should be in general agreement) (81). D_{\max} and R_g values for each sequence are reported in Supplementary Table S1. Shown in Figure 1D is a plot of each sequence's radii of gyration plotted against G4 number. Each additional G4 motif leads to an approximate R_g increase of 6.7 Å.

The extended tails of the $P(r)$ distributions and upward trend in plateau regions of the Kratky plots (Figure 1B) signify flexibility. Although plotting the radii of gyration versus the putative number of G4 subunits appears entirely linear ($R^2 = 0.9985$), a better fit is obtained when fitting to a swollen Gaussian polymer model ($R^2 = 0.9999$, Figure 1D). The non-linear least-squares fit to this model allows for the estimation of two parameters: persistence length, L_p , and the Flory exponent, ν . The persistence length rep-

resents the distance along the telomere G4 polymer which behaves as a rigid rod. At lengths much greater than this, the polymer behaves as a flexible Gaussian chain. The Flory coefficient (also known as the excluded volume parameter) varies between 0.5 and 1.0 and describes the degree of flexibility of the system. A Flory coefficient for a theoretical freely jointed flexible chain is 0.5 (maximum flexibility), whereas that of a rigid rod is 1.0. For reference, the empirical value of chemically denatured proteins is $\nu = \sim 0.588$ (82). Fitting to this model we find that the telomere G4 has a persistence length of 34.8 ± 0.2 Å and Flory coefficient of 0.69 ± 0.01 . The persistence length is approximately 1.5x the size of a single telomere G4 (~ 20 – 25 Å (53)), suggesting that the TTA linkers allow for flexibility when the system is ~ 3 or more contiguous G4s in length. This persistence length is about 50% greater than ssDNA ($L_p = \sim 22$ Å under similar ionic conditions (83)). As an independent method of estimating the persistence length, we used the hydrodynamic modeling program Multi-HYDFIT (77). This program integrates multiple independently measured properties, such as sedimentation coefficients ($S_{20,w}$) from AUC (Supplementary Figure S2) (51), translational diffusion coefficients (D_t) from SEC, and radii of gyration (R_g) from SEC-SAXS, for a series of macromolecules of given molecular weight (MW), and uses these values to find the optimum values of the model parameters for a worm-like chain (WLC) model (78). In total, we fit 12 independent properties from three independent techniques with their respective weights (estimated from standard deviations of multiple measurements), yielding a persistence length of 33 ± 3 Å (Table 2), in excellent agreement with the L_p estimated from the swollen Gaussian chain model. Altogether, these results, along with the qualitative information from Kratky and $P(r)$ distributions, suggest that the extended telomere maximizes G4 formation, is closely packed, and is moderately flexible. The flexibility is consistent with rigid G4 units linked by flexible, hinged, interfaces.

Ab initio and atomistic modeling reveals an ensemble of conformations ranging from entirely stacked and condensed to a coiled 'beads-on-a-string' configuration

The above analyses suggest a flexible system which would render the higher-order SAXS data unsuitable for use in *ab initio* bead reconstruction methods (as the bead reconstructions are 'averages' of all conformational states observed (64)). However, upon seeing the resulting space-filling models we were compelled to include them. Figure 1E shows the resulting DAMMIN and DAMMIF space-filling models of 2JSL, Tel48, Tel72 and Tel96 created based on the $P(r)$ data in Figure 1C (with corresponding fit results tabulated in Supplementary Table S1). Consistent with predictions from the Kratky and $P(r)$ distribution plots Tel48 looks like a dumbbell with two domains roughly the size of the 2JSL reconstruction with a small linker region in the middle. Similarly, Tel72 and Tel96 have what appear to be three and four G-quadruplex domains (indicated by their distinct 'bends'), although their resolution is not quite as high as the Tel48 reconstruction (Supplementary Table S1). The similar overall shape and curvature coupled with the

flexibility assessment above indicates a non-rod-like structure for telomere sequences with more than two G4 motif repeats. These shapes are generally in accord with previous hydrodynamic investigations based on rigid structures (51), but offer a more detailed and nuanced characterization because the flexibility of the structures can be taken into account.

We next employed an ensemble modeling approach that combined explicit solvent MD-derived models with the ensemble optimization tool GAJOE (of the EOM 2.0 suite) (73). A CD analysis that will follow indicated that the telomere sequences are best represented by a combination of hybrid-1 and hybrid-2 topologies. However, the order in which they occur is not evident, and it may be that the extended sequences are dynamic and interconvert on timescales much longer than is accessible by standard MD simulations (>1 ms). Therefore, we modeled every combination of the simplest multimer system, Tel48. Using the PDB atomic structures for hybrid-1 (PDB ID: 2GKU) and hybrid-2 (PDB ID: 2JSL) we generated each of the four possible combinations: hybrid-11, hybrid-12, hybrid-21, and hybrid-22. Each structure was subjected to 100 ns of both standard MD and accelerated MD (aMD) simulations to produce a pool of 8000 conformations for use in minimal ensemble and single structure modeling efforts. In the GAJOE ensemble optimization method, a pool of PDB atomic coordinate files are generated that cover as much conformational space as possible and utilized in calculating theoretical scattering profiles. Next, a genetic algorithm acts on these scattering profiles to minimize a fitness function by weighting each scattering profile and comparing combined profiles to the experimental (see Materials and Methods). The output is an ensemble of conformers which best recapitulate the experimental scattering profile based the minimized χ^2 value. An ensemble is considered a better fit than a single conformer when its χ^2 value is reduced relative to the single best-fit conformation.

Figure 2 and Supplementary Figure S3 shows the results of modeling efforts with the Tel48 constructs. Figure 2A and B are scatter plots which show the calculated radii of gyration (Y-axis) and corrected sedimentation coefficients (X-axis) (Supplementary Figure S2B), for each of 2,000 frames across both MD (light gray) and aMD (dark gray) trajectories for the hybrid-12, -21, -11 and -22 constructs. These plots indicate that both hybrid-12 and -21 sample conformations which agree with either the experimental R_g , $S_{20,w}$, or intersect both values. The hybrid-11 and hybrid-22 constructs rarely sampled conformations that corresponded with the experimental values (see Supplementary Figure S3). Interestingly, although hybrid-12 extensively samples conformations which agree with both hydrodynamic and scattering-derived measurements, the best fit model by CRY SOL analysis was found to be a highly extended hybrid-21 conformation (cyan dot and curve in Figure 2B, C). Because this conformation appeared unnatural (e.g. maximally extended) and did not agree very well with the $P(r)$ -derived R_g and D_{max} values, we speculated that this configuration may be biased simply by our initial start configurations. The hybrid-21 clearly tended toward an overall more compact structure as indicated by the histograms.

Therefore, we next asked what the maximum number of curves could be which could reconstruct the experimental scattering without worsening the χ^2 value. We found that an ensemble of six conformations gave approximately the same χ^2 value (magenta dots in Figure 2A and B, magenta curve Figure 2C) and agreed much better with the experimental R_g and D_{max} values from the $P(r)$ analysis ($R_{g,calc} = 19.58 \text{ \AA}$ versus $R_{g,exp} = 19.69 \text{ \AA}$ and $D_{max,calc} = 66 \text{ \AA}$ versus $D_{max,exp} = 65 \text{ \AA}$, Supplementary Table S1). The resulting topologies were a 50/50 mix of hybrid-12 and -21 (Figure 2D), which sampled conformations ranging from extended to fully stacked. The flexibility of the ensemble was only marginally lower than the pool, as judged by EOM's Rflex flexibility analysis, supporting semi-flexibility. Interestingly, we found that one of the hybrid-12 conformers (bottom right of Figure 2D) was nearly identical in conformation to our previously reported hybrid-12 model (49,50), with an RMSD of just 1.6 \AA over all residue pairs (Supplementary Figure S4).

Next, we investigated the Tel72 and Tel96 constructs in the same manner as above but only using standard MD. Models were created to reflect the ratio of hybrid-1/-2 (25/72) as determined by our later CD analyses of telomere mutants (e.g. hybrid-121, -122, -212 and -221 for Tel72 and hybrid-1222, -2122, -2212 and -2221 for Tel96). The hybrid-121 was included because it was proposed previously (49). Each model was then subjected to explicit solvent MD and simulated for a total of 100 ns. From these trajectories, 1000 equally spaced frames were used as the pool for EOM's GAJOE program. For the Tel72 the best fit was obtained with an ensemble of four conformers ($\chi^2_{ensemble} = 1.09$ versus $\chi^2_{single-model} = 1.81$) which was composed of the hybrid-212 (89.7%) and hybrid-221 (10.3%) (Figure 3). Surprisingly, this configuration agrees well with Tel48 having a 5' preference for hybrid-2 followed by hybrid-1. The R_g and D_{max} values of the Tel72 ensemble (Figure 3C) agree with the experimental values ($R_{g,calc} = 25.8 \text{ \AA}$ versus $R_{g,exp} = 26.0 \text{ \AA}$ and $D_{max,calc} = 83 \text{ \AA}$ versus $D_{max,exp} = 87 \text{ \AA}$), indicating that the ensemble is an excellent solution. Similarly, Tel96 scattering was best recapitulated by an ensemble of four conformers ($\chi^2_{ensemble} = 1.15$ versus $\chi^2_{single-model} = 2.08$) but was composed entirely of different conformations of the hybrid-2122 (Supplementary Figure S5). Again, there is an agreement with a 5' hybrid-2 followed by hybrid-1. The R_g and D_{max} values of this conformer ensemble are also in agreement with the experimental values ($R_{g,calc} = 32.1 \text{ \AA}$ versus $R_{g,exp} = 32.7 \text{ \AA}$ and $D_{max,calc} = 103 \text{ \AA}$ versus $D_{max,exp} = 109 \text{ \AA}$), indicating that this ensemble is reasonable. In both cases, the EOM Rflex quantification of flexibility indicates that the ensembles are only marginally less flexible than the pool (Supplementary Table S1), consistent with the system semi-flexibility. This flexibility is also illustrated by the conformer ensembles themselves (Figure 3C and Supplementary Figure S5). Further, docking of each ensemble into their respective *ab initio* space-filling models from Figure 1E reveals excellent fits for the models of topological sequence 5'-hybrid-2,-1,-2,-2 (Figure 4). Collectively, these analyses indicate that in physiological buffer conditions the extended telomeres maximize their formation of G4 subunits, prioritize hybrid-2 at the 5' end, and are semi-flexible.

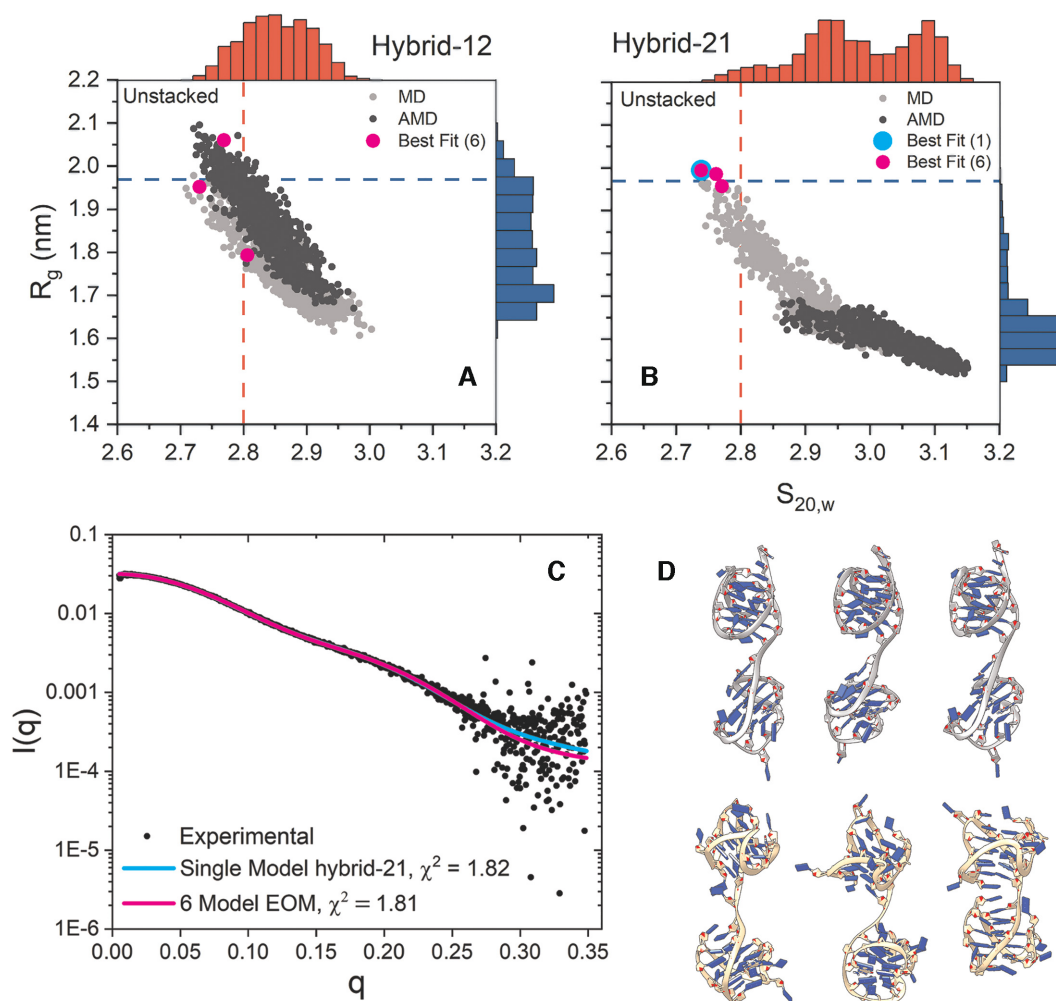


Figure 2. Results of Tel48 SAXS atomistic modeling efforts. (A, B) scatter plots of calculated radii of gyration and sedimentation coefficients for hybrid-12 (A) and hybrid-21 (B) with MD-derived values shown in light gray and aMD-derived values in dark gray. The inset dashed red and blue lines represent the experimentally measured values for sedimentation coefficient and radius of gyration, respectively. The outer histograms represent the distributions of values from both MD and aMD snapshots combined. The cyan dot represents the single best-fit model (hybrid-21) as determined by CRYSOLOG (top left model in D). Magenta dots represent the six conformers in the best fit ensemble (all six models in D). (C) Experimental SAXS scattering data with fits from single (cyan) or ensemble (magenta) calculated scattering overlaid with χ^2 values inset. (D) Single best fit model (hybrid-21, top left model) and best fit ensemble of six conformers (top row hybrid-21, bottom row hybrid-12). Models are oriented with their 5' ends at the top.

The Tel72 hybrid-212 preferentially samples a stacked conformation, forming unique electronegative binding pockets useful for drug targeting

The Tel72 ensembles reflect well the SAXS-derived properties, R_g and D_{max} . However, they do not agree as well with their measured sedimentation coefficients. SAXS scattering is exquisitely sensitive to changes in particle volume (conformation in this case). Systems which exist in an equilibrium of stacked and unstacked, or coiled and beads-on-a-string, will have a scattering profile which is composed of a continuous distribution of conformations (as the scattering intensity, $I[0]$, is directly related to the volume of the scattering particle) (84). Therefore, we next endeavored to find the most frequently sampled conformation from the MD trajectory of the Tel72 hybrid-212 construct. Figure 5A shows the top three most frequently sampled conformations across the 100 ns simulations with their respective weight-

ing (% of MD frames). This analysis suggests that approximately 47% of the frames from simulation sampled a configuration which was partially (middle) or entirely stacked (left and right models). The major stacked conformation sampled by hybrid-212 has a calculated sedimentation coefficient which is in excellent agreement with the experimental ($S_{20,w(calc)} = 3.45$ versus $S_{20,w(exp)} = 3.46$, Supplementary Figure S2) although the calculated radius of gyration is slightly lower ($R_{g(calc)} = 2.45$ nm versus $R_{g(exp)} = 2.60$). Electrostatic calculations of the most prominent form reveal highly electronegative grooves, which are appropriately sized for small molecules (Figure 5B). Overall, these analyses show that the hybrid-212 model of the Tel72 is consistent with all available spectroscopic, hydrodynamic, X-ray scattering, and MD-based analyses, and forms unique junctional grooves for selective small molecule targeting.

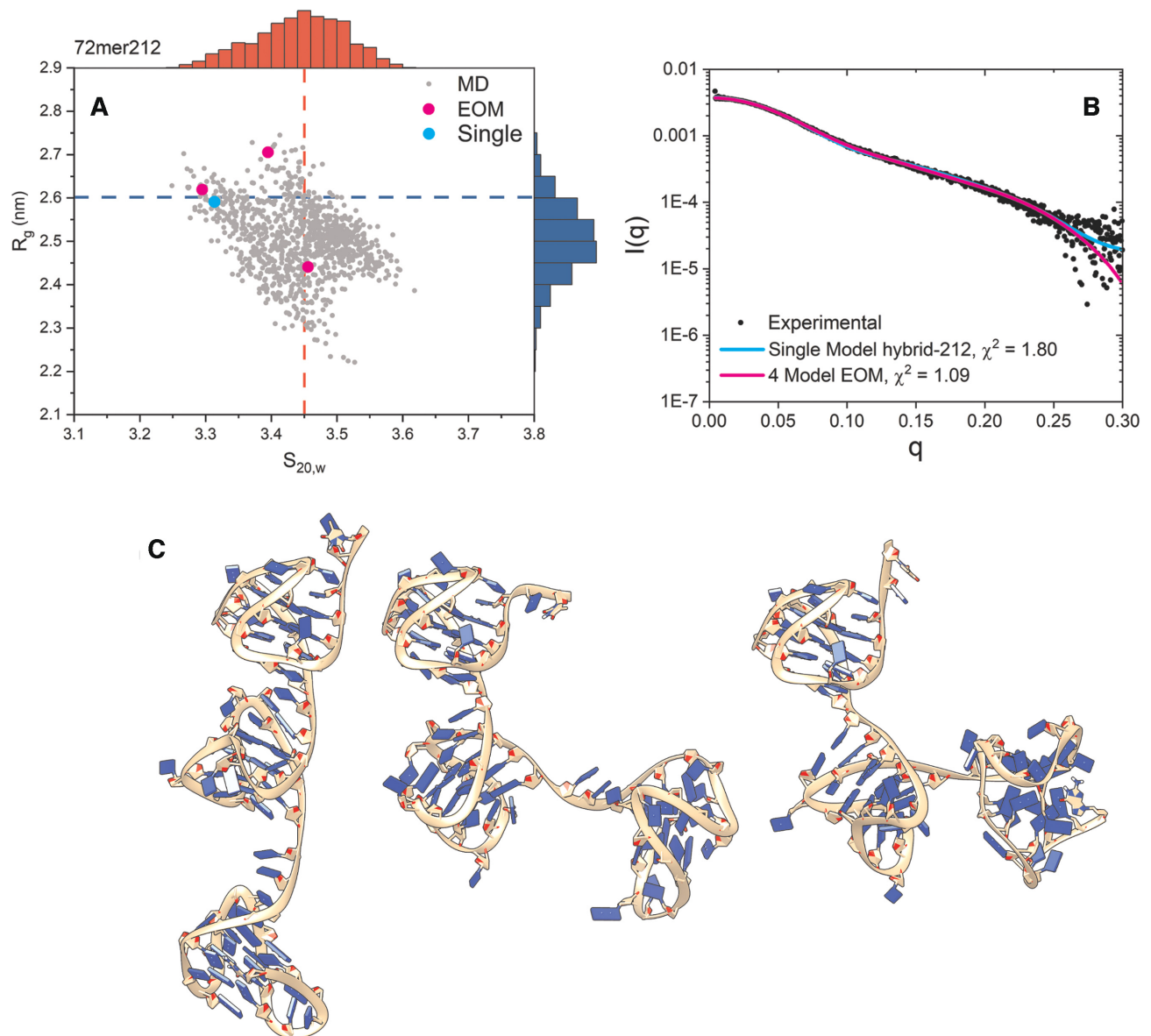


Figure 3. Results of Tel72 SAXS atomistic modeling efforts. (A) scatter plot of calculated radii of gyration and sedimentation coefficients for the hybrid-212 model from 100 ns of standard MD simulation. The inset dashed red and blue lines represent the experimentally measured values for sedimentation coefficient and radius of gyration, respectively. The outer histograms represent the distributions of values. The cyan dot represents the single best-fit model as determined by CRY SOL. Magenta dots represent the four conformers in the best fit ensemble. (B) Experimental SAXS scattering data with fits from single (cyan) or ensemble (magenta) calculated scattering overlaid with χ^2 values inset. (C) Conformations of the three hybrid-212 configurations (not showing the hybrid-221) from the best fit ensemble as determined by EOM. Models are oriented with their 5' ends at the top.

The major topologies of the extended telomere G4 are hybrid-1 and hybrid-2

Simultaneous with our structural investigations above, we investigated the conformations of G4 units within higher-order structures by using systematic sequence variations of the wild type (WT) Tel48 sequence and observing changes in circular dichroism spectra. These sequence 'mutants' were created with variation in terminal nucleotides or by changes in internal sequences that favor the various hybrid topologies (e.g. hybrid-1 and hybrid-3) (Table 1). The Tel48 spectrum (black line in Figure 6A) has a main peak at ~ 290 nm, pronounced shoulder from 265–275 nm, and a trough

at 235 nm, indicating that it is primarily composed of hybrid type folds (27). Comparing sequence variants of the form $d(T_nAGGG(TTAGGG)_7T_m)$, where $n = 1$ or 2, and $m = 0, 1$ or 2, we found no major spectral differences (Supplementary Figure S6A), indicating that changes in these flanking nucleotides have no effect on the overall topology. Removal of the 5'-end thymine residues led to a modest reduction in the shoulder at ~ 270 nm and peak at 290 nm when compared to the WT sequences of similar length (Supplementary Figure S6B, red and blue lines). We speculated that this could be due to the formation of hybrid-3 in the 5'-most G4 unit, which has been reported in shorter

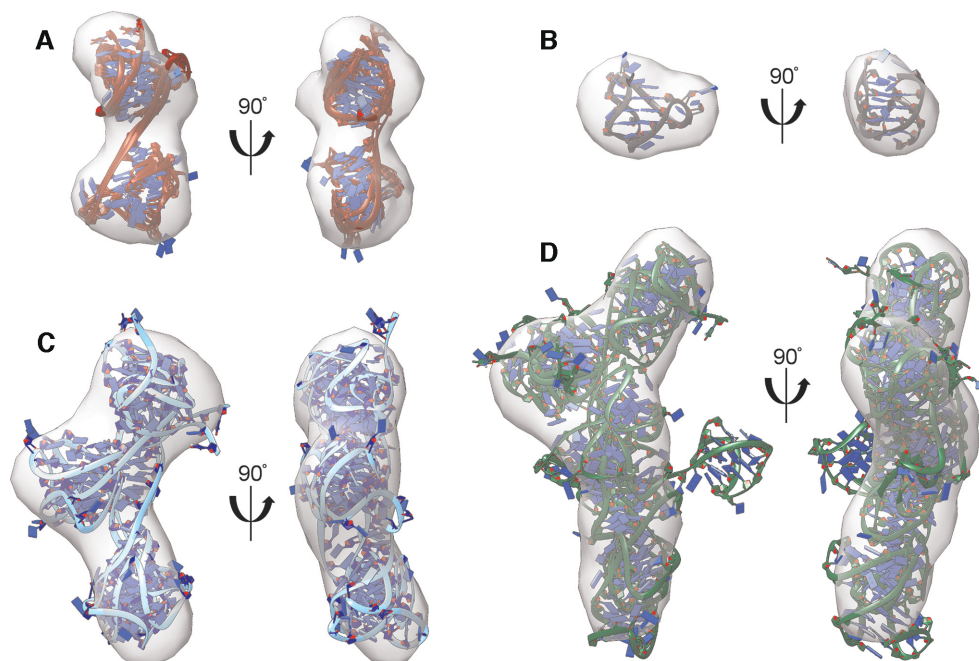


Figure 4. Telomere G4 ensembles from EOM GAJOE analysis docked into *ab initio* space-filling reconstructions from DAMMIN/DAMMIF. (A) Tel48 hybrid-21 conformers, (B) 2JSL with single best-fit NMR-derived model, (C) Tel72 hybrid-212 conformers (the same as in Figure 3C), (D) Tel96 hybrid-2122 models (the same as in Supplementary Figure S5).

sequences lacking the 5' thymine residues (21,31). Indeed, when an inosine is introduced to favor the hybrid-3 topology in the 5'-most putative G4 the CD changes observed at 270 and 290 nm become more pronounced (Supplementary Figure S6C, blue solid line). Importantly, this suggests that the hybrid-3 topology is not a major topology, as the extended telomere (in the cell) will always include 5' thymine residues. The spectral change due to hybrid-3 incorporation is made more evident when stabilized in both 5' and 3' G4 units (Supplementary Figure S6C, purple), which is of the same shape but $\sim 2\times$ the magnitude of hybrid-3. Thus, the hybrid-3 is not likely to exist in the context of the extended telomere, aside from as a potential folding intermediate (21).

Comparison of the hybrid-21, -11, and -12 sequences to Tel48 revealed subtle differences in each case (where hybrid-2 is assumed as the major conformation in unadulterated telomere sequence flanked by thymine at both ends) (Figure 6A). Overall, each spectrum was consistent in shape, but varied in magnitude at various wavelengths. We suspected that this may indicate a preference for the hybrid-22 form. A theoretical hybrid-22 spectrum overlaid nicely with Tel48 (Figure 6B, red dashed line). In contrast, a theoretical hybrid-11 (Figure 6B, purple dashed line) had a greatly increased 290 nm peak and slight reduction at ~ 250 nm and looked similar to the mutant hybrid-11 spectrum from Figure 6A. Based on the reported 25/75 ratio of hybrid-1 and hybrid-2 for the monomer telomere sequence flanked at both ends by thymine (32), we next tested a variety of computed weighted combinations of hybrid-1 and -2 spectra and found that a 30/70 ratio best reflects the Tel48 spectrum (compare blue dashed line with black in Figure 6B). Collectively, these results support a preference for both hybrid-1

and -2 topologies in the extended telomere sequence, consistent with our EOM analysis of Tel48.

CD indicates that the higher-order telomere sequences converge on a 25/75 ratio of hybrid-1 and hybrid-2 with maximization of G4 formation

Strand-normalized circular dichroism spectra are the sum of constituent secondary and tertiary structure (85). Thus, just as above we expect that the spectra of higher-order telomere G4s could be recreated by addition of their measured 'monomer' spectra. However, comparisons of the various monomer spectra (hybrid-1,-2,-3 and basket forms) to our higher-order telomere spectra resulted in non-negligible 'difference' spectra of roughly the magnitude expected for di- or tri-nucleotides. As prior studies suggest, and we have shown here, the extended telomere sequences maximize their G4 potential by leaving no G-tract gaps. The resulting stacking junctions, or other inter-G4 interactions that constrain the loop regions, could potentially give rise to a 'junctional' CD signal (85). Thus, to generate the theoretical 'junctional' spectrum, we utilized the Tel48 mutant spectra from above and subtracted from them theoretical constituent monomer spectra as appropriate. The resulting spectrum is shown in Figure 7A. The junction spectrum has a peak at 240 nm and troughs at ~ 260 and ~ 285 nm, which is consistent with the known CD spectra of adenine and thymine polynucleotides (86). This spectrum was then used as a correction factor, and was subtracted from the Tel48, Tel72, and Tel96 spectra (Figure 7B). A plot of $\Delta\epsilon_{290}$ versus the putative number of G4s yields a linear regression with a near zero Y-intercept, which is more physically relevant than the regression without the correction (Figure 7C). The

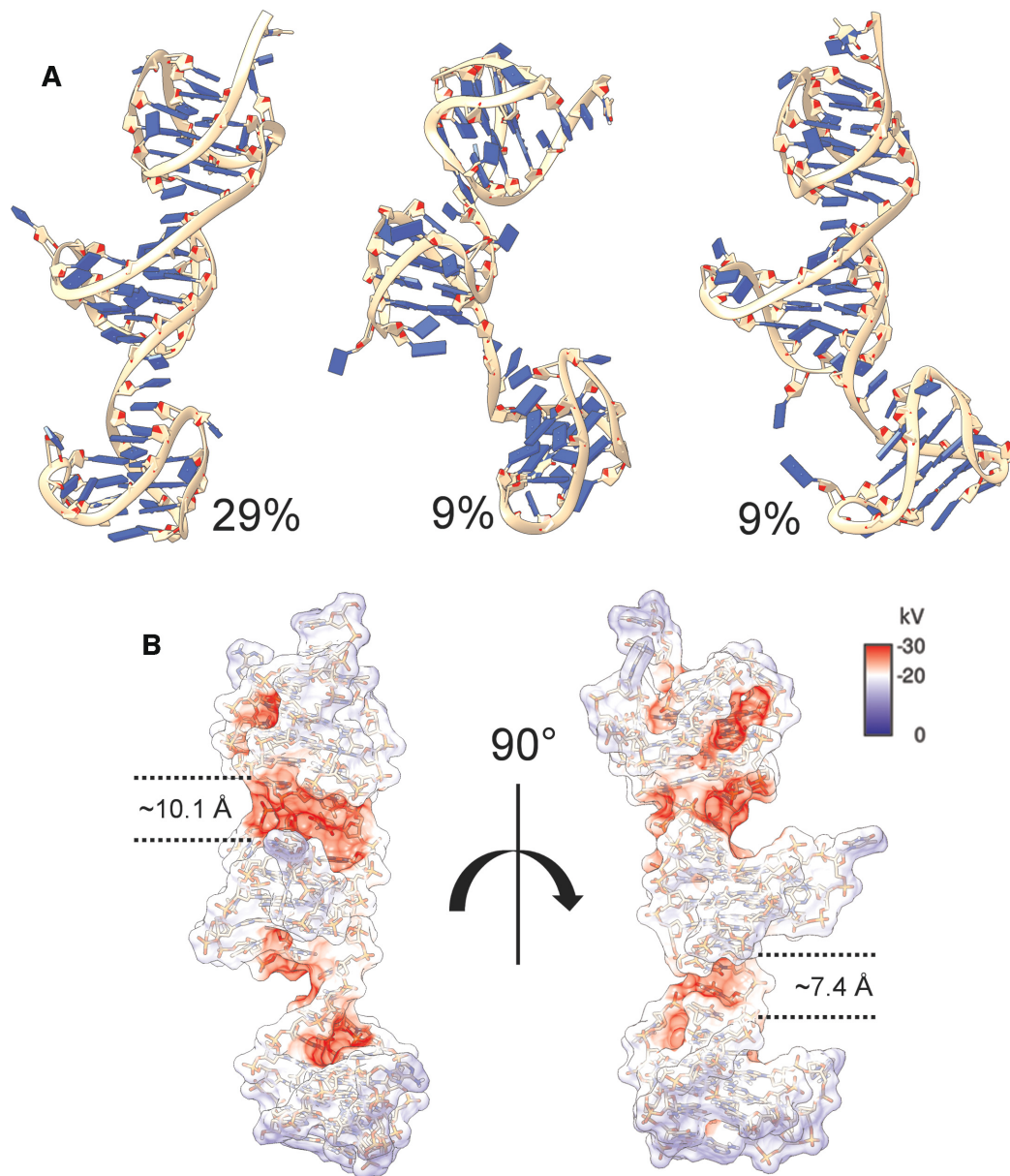


Figure 5. Results of MD clustering analysis of the Tel72 hybrid-212. (A) Top three representative centroids of DBSCAN clusters accounting for $\sim 47\%$ of frames across the entire 100 ns trajectory. (B) space-fill electrostatic APBS map of the first model from A with dashed lines indicating the approximate sizes of each groove created at the two stacking interfaces.

slope of the corrected regression data indicates that each additional putative G4 increases $\Delta\epsilon_{290}$ by $\sim 128 \text{ M}^{-1} \text{ cm}^{-1}$, in excellent agreement with the average $\Delta\epsilon_{290}$ obtained from hybrid (3+1) monomers. These corrected spectra should now be a composite spectrum of monomer G-quadruplex components. A novel finding here is that the CD spectra of higher-order G4 structures contain discernable contributions from G4–G4 interactions.

The hybrid-1, -2 and -3, as well as the antiparallel basket monomer spectra were systematically compared with the Tel48, Tel72 and Tel96 corrected spectra. Figure 7D–F show the corrected spectra in black with ‘_Corr’ indicating the corrected spectrum. Shown below are residuals from the best fit combinations of monomer spectra. In each plot

the red spectrum is the best fit, followed by blue, etc. based on residual sum of squares (RSS) analysis (Supplementary Figure S7). The optimal fit to Tel48_Corr is hybrid-22 (in agreement with Figure 6A), followed by various hybrid-1/-2 combinations; Tel72_Corr is best fit by a combination of hybrid-1, -2 and -2; Tel96 is best fit by a hybrid-1, -2, -2 -2 (not necessarily in that order as shown above). See Supplementary Figure S7 for the exhaustive residual sum of squares (RSS) ordering, PDB IDs, and distributions of CD residuals for each fit. We note here that since the monomer hybrid-1/-2 spectra are dependent on the particular sequence chosen that all fits shown in Figure 7D–F are within the experimental uncertainty, indicating that there is a range of hybrid-1/-2 ratios which could be consistent.

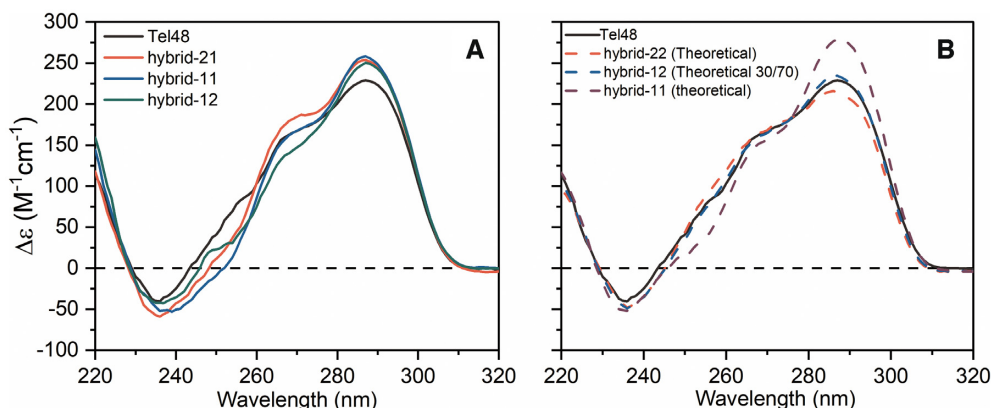


Figure 6. Normalized circular dichroism spectra of Tel48 mutants and theoretical monomer G4 spectra. (A) CD spectra comparison of the WT Tel48 G-quadruplex (black) with constructs created to favor the hybrid-1 form in the second (red), first and second (blue), or first position (green). (B) Comparison of the Tel48 CD spectrum with theoretical monomer CD combinations of hybrid-22 (red dashed), hybrid-12 with a 30/70 weighting (blue dashed), and a hybrid-11 (purple dashed). Component monomer spectra used for theoretical spectra in B were derived from sequences of the monomer telomere G-quadruplexes PDB IDs 2GKU (hybrid-1) and 2JSL (hybrid-2) annealed in potassium buffer.

These results indicate an overall preference for hybrid-2 in the longer sequences. Altogether, the above analyses confirm that the primary two topologies making up the WT telomere higher-order G4s are hybrid-1 and hybrid-2, with proportions approaching a 25% hybrid-1, 75% hybrid-2. Further, the linearity and slope of the regression analysis and excellent agreement with theoretical fits indicates that no long gaps exist in the higher-order human telomere.

DISCUSSION

Our results provide a highly detailed characterization of the inter-G4 dynamics of the extended human telomere G-quadruplex. We combine circular dichroism, hydrodynamics, and small-angle X-ray scattering experiments integrated with available high-resolution NMR studies on monomeric G4 structures (19,22) to build medium resolution higher-order structural ensembles. For dimeric structures containing two contiguous G4 units (Tel48), the best model is one featuring a mix of hybrid-1 and hybrid-2 topologies. For longer sequences with three and four G4 units, a mixture of hybrid-2 and hybrid-1 conformations seems to be present in an approximate 3:1 ratio. Our results show unequivocally that for all sequences up to 96 nt in length, the human telomere sequence maximizes its G4 formation, leaving no G-tract gaps between G4 subunits—consistent with an early AFM study (40) and in contrast to recent EM, magnetic tweezer, and AFM analyses (45,46). Our results provide the first quantitative estimates of the rigidity of folded telomeric DNA through determination of its persistence length ($L_p = \sim 34$ Å). We find that the semi-flexibility of the telomere G-quadruplex is best modeled by an ensemble of configurations which fluctuate between an entirely stacked multimer and unstacked monomeric G4s, as observed by MD simulations, providing potentially unique sites for small molecule targeting in the junctional regions. This model suggests that rigid G4 units are connected by a short, dynamic, interfacial hinge. That interface constitutes a unique structural element to target in drug design efforts. Importantly, this is the first study that integrates medium-resolution structural

data with high-resolution atomic structures of the higher-order human telomere G4.

The WT human telomere monomer sequences exhibit a high degree of polymorphism *in vitro* (27). Under physiologically relevant K^+ solution conditions the WT telomere G4 adopts a hybrid type conformation, favoring hybrid-2 over hybrid-1 (19,22). This conformational bias is seemingly dictated by the presence of 5' or 3' flanking nucleotides. Addition of 5'-TTA to the core sequence, $d(GGG(TTAGGG))_3$, leads to the favoring of hybrid-1 by a 5'-end capping adenine triplet, whereas addition of one or two thymine to the 3'-end results in a favoring of the hybrid-2 form via a T:A:T triplet cap on the 3'-end (27). The extended, end-flanked sequence, $d(TTAGGG)_4T$, forms a major configuration of hybrid-2 ($\sim 75\%$), with minor amounts of hybrid-1 ($\sim 25\%$) (32). This implies that the energy barrier between the two forms may be small. Consistent with this, our mutational analysis by CD and modeling studies agree with a 75/25 ratio of hybrid-2/-1 for the higher-order WT species, although by CD other ratios are within our experimental uncertainty. A significant result from our higher-order CD analyses is the unique junctional spectrum (Figure 7B). This spectrum was useful in providing a rationale for why the higher-order species exhibited CD signatures that were lower than expected for maximized G4 formation. Moreover, both SAXS and MD modeling studies revealed favorable, but dynamic, stacking interfaces between G4 moieties, in agreement with thermodynamic analyses (50).

Prior NMR investigations of the WT telomere sequence, $d((TTAGGG)_4T)$, indicate a dynamic equilibrium of conformations (27). If a similar equilibrium exists in the higher-order telomere, then an ensemble of both tertiary conformation and secondary structure would be required to explain both CD and SAXS results. Consistent with this, the Tel48 scattering is modeled well by an ensemble with a 50/50 ratio of hybrid-12 and hybrid-21 conformers. The single model hybrid-21 fit is comparable to the ensemble, and so this solution is, overall, somewhat ambiguous; although, the lack of thymine residues at the 3'-end would, in theory, favor hybrid-1 in the second position, giving us confidence in a

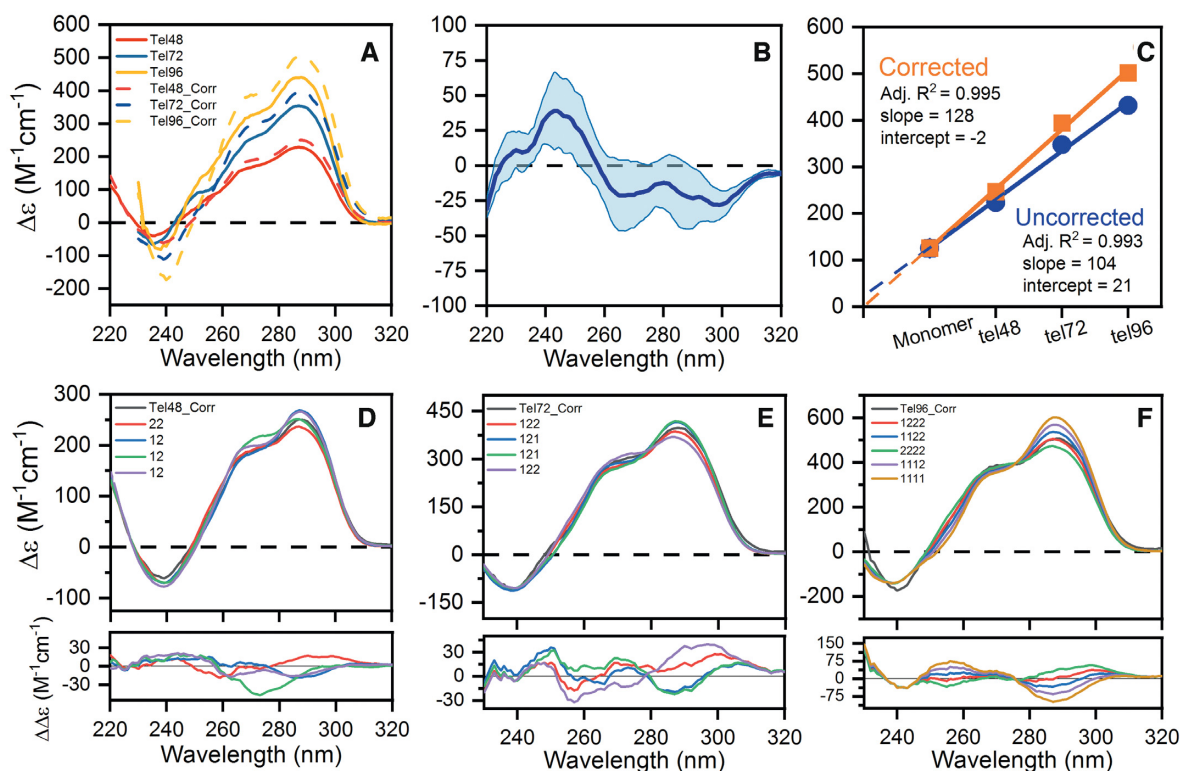


Figure 7. Circular dichroism analysis of the higher-order telomere G-quadruplexes. (A) Pre- and post-corrected ('Corr') CD spectra of the Tel48, Tel72 and Tel96 sequences by subtraction of the 'junctional' spectrum in B. (B) The average (dark blue line) and range (light blue space fill) of 'junctional' CD spectra derived from deconstruction of the Tel48 sequences in Figure 6 and Supplementary Figure S6 using constituent monomer spectra. (C) Regression analysis of the uncorrected and corrected $\Delta\epsilon_{290nm}$ values as a function of the number of G4 motifs. (D–F) Corrected CD spectra of the Tel48, Tel72 and Tel96 sequences with overlaid theoretical spectra derived from the linear addition of monomer spectra. Each '12' or '121' correspond to different combinations of telomere sequences which form either hybrid-1 or hybrid-2 (see Supplementary Figure S7). The red spectrum in each plot is the best fit as judged by RSS analysis. Residuals are shown below each figure. See Supplementary Figure S7 for the full RSS analysis along with the PDB identifiers of sequences used to compute theoretical spectra.

preferential hybrid-21 model. Our previous investigation of the WT Tel50 sequence (49) (which differs in sequence by two additional thymine residues at the 3' end) proposed that the major form is hybrid-12. We used steady-state fluorescence measurements of 2-aminopurine-substitutions to assess the solvent accessibility for each adenine site. From this it was found that residue A15 is the least solvent exposed, which agreed with SASA calculations of the hybrid-12 model (in this conformation A15 is buried in the stacking junction between the two G-quadruplex units). Coincidentally, by having the EOM algorithm increase the number of conformers in the Tel48 ensemble, we find that part of the new solution is a hybrid-12 conformer that is almost identical to the previously proposed Tel50 hybrid-12 model (Supplementary Figure S4). Thus, the collective experimental observables support an equilibrium of hybrid-1 and hybrid-2 in either position.

We have also investigated the possibility of a hybrid-3 form, which is a two-tetrad antiparallel G-quadruplex that has been characterized *in vitro* in potassium containing solution (21), and confirmed as existing intracellularly by in-cell NMR techniques (31,41). In one of the in-cell investigations, the authors analyzed a minimal monomer sequence $d(\text{AG}_3(\text{TAGGG})_3)$ which was transfected into HeLa cells (31). Using a ^{19}F -NMR technique, the authors

showed that hybrid-1, -2 and -3 type G4s formed. In the second study, the authors used both the minimal monomer sequence $d(\text{AG}_3(\text{TAGGG})_3)$ as well as extended multimer telomere sequences (8 and 12 G-tracts) to investigate the internal and terminal (5' flanked) G4 conformations both *in vivo* and *ex vivo* in *X. laevis* oocytes and cell extracts using NMR and site-specific residue labeling (41). The authors showed that both the internal and terminal telomere G4s independently form hybrid-2 and hybrid-3 conformations in a 'beads-on-a-string' assembly. The authors concluded that telomere G4 folding is 'independent of nucleotide flanking and insensitive to cellular molecular crowding' and that 'physiologically relevant conformations of G4 units in the context of the telomeric G-overhang are formed in dilute potassium-based solution'. In contrast to these studies, we show here that the extended WT telomere sequences do not favor the hybrid-3 in potassium-based solution by mutational analysis, showing that it may only occur in the 5'-most position when thymine is removed (Supplementary Figure S6). Furthermore, our combined SAXS and molecular modeling results indicate that hybrid-3 could not exist internally for two reasons, (i) the 5' and 3' termini of the hybrid-3 face the same direction, which we would expect would cause pronounced bends that weren't observed in our final ensembles and (ii) the hybrid-3 moiety would

forbid any contiguous end-to-end stacking. Altogether, our results, in combination with these studies, provide a strong case for targeting the hybrid-2 topology junctions in future drug discovery efforts.

The single-stranded telomere overhang is a critical regulator of genomic integrity. Spanning the junction of the duplex and single-stranded telomere region is a protective protein complex known as shelterin (1,87). Shelterin is composed of the proteins TRF1, TRF2, RAP1, TIN2, TPP1 and POT1. POT1 (protection of telomeres 1) is essential in sequestering the free 3' overhang, shielding it from eliciting aberrant single-stranded DNA damage responses, preventing homologous recombination, and regulating the activity of telomerase (37). POT1 binds directly to the 3' single-stranded overhang with high affinity and in a highly sequence specific manner (1,36,88). EM micrographs have revealed that POT1-TPP1 complexes coat the entirety of the extended telomere 3' overhang, forming compact, ordered complexes of ssDNA-POT1-TPP1 without gaps (89). Importantly, disruption of POT1's shielding of the single-stranded overhang elicits an ATR-dependent DNA damage response through the promiscuous ssDNA binding protein RPA (37,90). A recent AFM investigation of the Tel96 sequence with POT1 by the Opresko lab (44) found that maximization of G4 formation 'rarely' occurs, and that POT1 associates by simply recognizing the resulting gaps. We, and others (38,39,51,91), find this conclusion at odds with solution-based results. Accessible ssDNA gaps between G4s would allow RPA to compete unimpeded with POT1 binding. Indeed, RPA outcompetes POT1-TPP1 binding to single-stranded telomere DNA *in vitro* (92). Further, G-quadruplex secondary structure enhances POT1-TPP1's protection against RPA in physiologically relevant levels of K⁺ (150 mM) (93). We recently showed that POT1 unfolds and binds to telomere G4s using a conformational selection mechanism (91) and demonstrated that the kinetics of unfolding the Tel48 sequence is essentially the same as the Tel24 monomeric G4. Importantly, this suggests that a maximization of G4 formation does not impede POT1 binding. Taken together, the physiological significance of telomeric G4 maximization is 2-fold: (i) G-quadruplex secondary structure prevents promiscuous RPA binding and (ii) the G4 secondary structure promotes exclusive interaction with shelterin through specific POT1 unfolding and binding, tilting the scale in favor of POT1 over RPA.

The mechanistic details of how the shelterin complex orchestrates the sequestration of the single-stranded 3' end are still not entirely understood, but are of great importance in drug discovery (87,94). Recently, the Cech laboratory conducted a thorough investigation of co-expressed and isolated complexes of the shelterin proteins *in vitro* (87). Based on their results a shelterin 'load & search' model was proposed, whereby TRF2 and POT1 localize the shelterin complex to the telomere by specifically recognizing and binding to the single-stranded/double-stranded (ss/ds) telomere junction. The authors propose that POT1 searches for its optimal binding sequence, d(TTAG), at the 3' end by a scanning search mechanism, eventually looping the 3' end back forming a loop bridged by shelterin proteins that is 'unlike a T-loop'. An earlier report from the Cech laboratory found

that POT1 and POT1-TPP1 completely coat the long ssDNA telomere repeats *in vitro* (89). Thus, the 'normal' sequestration mechanism of the human telomere 3' overhang is seemingly a POT1-coated loop structure anchored to the ss/ds telomere junction.

DNA looping is a common theme in the cell (79). From a physical stand-point, DNA looping has been extensively studied for its relationship with genetic packaging into nucleosomes and effects on transcription (79,95). A commonly reported measure of the structural rigidity of a biological polymer is the persistence length, L_p , that defines the length over which a polymer remains unbent in solution. In this work, we have applied both SAXS and hydrodynamic modeling methods to derive an estimate of the telomere G4 persistence length. Using our telomere G4 L_p estimate, along with values reported for single- and double-stranded DNA, we can compare the relative forces required to bend each 180° around the arc of a semi-circle of a given radius (Figure 8) (79). From this plot we find that, in the case of single-stranded telomere DNA of length >200 angstroms (~63 nt) the force required to bend the polymer 180° (in the shape of a semi-circle) is negligible—energy requirements on the order of thermal fluctuations. However, if that same stretch of 63 nucleotides were to form maximal G-quadruplexes (decreasing polymer length to <100 angstrom), the increase in energy to bend increases by an order of magnitude—now requiring energy input equivalent to ATP hydrolysis. Although somewhat intuitive, this implies that in the absence of significant energy input, short telomere G4s must be made single-stranded in order to bridge the terminal 3' d(TTAG) repeat (capped by POT1) with the shelterin complex. More importantly, this figure indicates that small molecules which bind the inter-G4 grooves, thereby increasing its effective persistence length, could shift the force curve to the right (towards the dsDNA curve, red) and subsequently drive up the energy cost for associating the POT1-bound 3' end with the shelterin loop. Indeed, during the drafting of this manuscript Gao *et al.* have demonstrated that a small molecule targeting the wild-type Tel48 can shift the distribution of conformations to favor a more compact, likely stacked, conformation (96)—a transition that would directly affect persistence length. It is well established that telomere G-quadruplex interacting small molecules are able to displace shelterin components, uncap the telomeres, and ultimately, inhibit telomerase *in vitro* and *in vivo* (14,97). Thus, there is abundant rationale for future work targeting these unique G4 junctional sites with stabilizing small molecules.

Utilizing a robust integrative approach, we have presented here the highest-resolution view of the higher-order telomere G4 to date. SAXS refinement of MD-derived models constructed from high-resolution techniques is now a mainstay in structural biology. Specific combinations of hybrid forms are preferentially fitted to the SAXS data which means the resolution is less than the size of those monomers, 20–25 Å. As the major difference in the hybrid forms is the location of the d(TTA) parallel loop, and SAXS can discern this difference in fitting, the effective resolution could be considered as better than the dimensions of three nucleotides. However, SAXS refinement of MD generated atomistic models, while excellent for discarding unrealistic

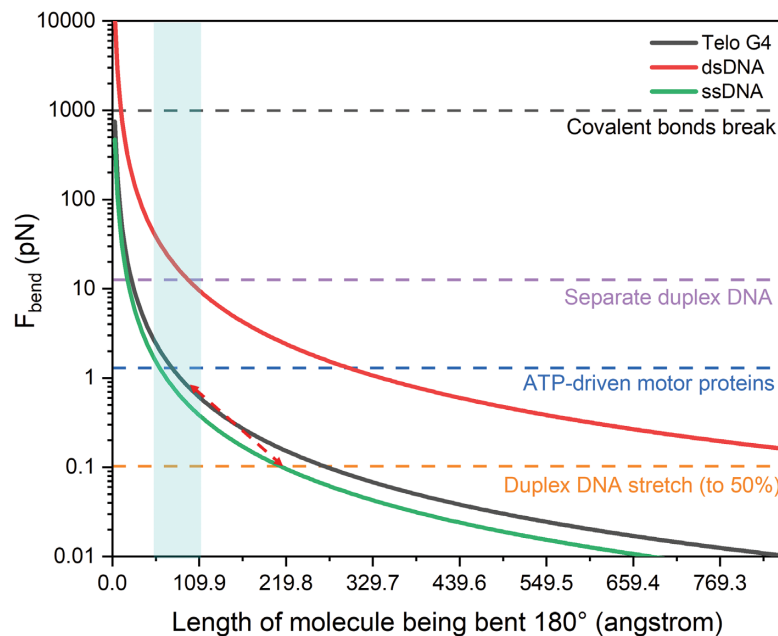


Figure 8. DNA force of bending plot for single-stranded (green), double-stranded (red) and G4 telomere DNA (black). Force curve calculations were performed similar to reference (79) using literature values of persistence length for ssDNA ($L_p = 22 \text{ \AA}$), dsDNA ($L_p = 550 \text{ \AA}$), and Telomere G4 ($L_p = 34.8 \text{ \AA}$) as measured here. The Y-axis is the estimated force (in pN) to bend a length of DNA (X-axis) 180° about the arc of a semi-circle (i.e. if you have a 330 \AA long single-stranded DNA it will require a force of $\sim 0.05 \text{ pN}$ to bend it into a semi-circle). Dashed horizontal lines are visual references to common biological forces found in the cell (orange indicates the approximate range of force from thermal fluctuations). The light blue region highlights the range in which short telomere G4s would be found, indicating that a large force would be required to bend short telomeres ($\leq 96 \text{ nt}$). The dashed red arrow illustrates that if a $\sim 200 \text{ \AA}$ long ssDNA telomere (approximately 63 nt) were to spontaneously fold into a contiguous G4 structure, the resulting bending force required for a 180° turn increases by an order of magnitude. The increase in bending force is comparable to the same length of DNA in duplex form (330 \AA long duplex requires external forces equivalent to ATP hydrolysis to bend 180°). In the case of duplex DNA, the energy requirement of 'tight' bending is usually compensated for by the highly positive charge on histones.

tic topologies and conformations, is not necessarily definitive when conformational and topological polymorphism presents itself. Thus, we await higher-resolution techniques that can inform on the distributions of topologies in the higher-order telomere G-quadruplex.

DATA AVAILABILITY

Small-angle X-ray scattering data has been deposited in the publicly accessible Small Angle Scattering Biological Data Bank (<https://www.sasbdb.org/>) under the IDs: SASDKF3 (2JSL), SASDKG3 (Tel48), SASDKH3 (Tel72) and SASDKJ3 (Tel96).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. This project was supported by grant 9 P41 GM103622 from the National Institute of General Medical Sciences of the National Institutes of Health. Use of the Pilatus 3 1M detector was provided by grant 1S10OD018090-01 from NIGMS.

The content is solely the responsibility of the authors and does not necessarily reflect the official views of the National Institute of general Medical Sciences or the National Institutes of Health.

Molecular graphics and analyses performed with UCSF Chimera, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311.

FUNDING

National Institutes of Health (NIH) [GM077422]. Funding for open access charge: NIH [GM077422].

Conflict of interest statement. None declared.

REFERENCES

- de Lange, T. (2005) Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev.*, **19**, 2100–2110.
- Stewart, J.A., Chaiken, M.F., Wang, F. and Price, C.M. (2012) Maintaining the end: roles of telomere proteins in end-protection, telomere replication and length regulation. *Mutat. Res.*, **730**, 12–19.
- Wright, W.E., Tesmer, V.M., Huffman, K.E., Levene, S.D. and Shay, J.W. (1997) Normal human chromosomes have long G-rich telomeric overhangs at one end. *Genes Dev.*, **11**, 2801–2809.
- Neidle, S. and Parkinson, G. (2002) Telomere maintenance as a target for anticancer drug discovery. *Nat. Rev. Drug Discov.*, **1**, 383–393.
- Townsley, D.M., Dumitriu, B. and Young, N.S. (2014) Bone marrow failure and the telomeropathies. *Blood*, **124**, 2775–2783.
- Aubert, G. and Lansdorp, P.M. (2008) Telomeres and aging. *Physiol. Rev.*, **88**, 557–579.

7. O'Sullivan, R.J. and Karlseder, J. (2010) Telomeres: protecting chromosomes against genome instability. *Nat. Rev. Mol. Cell Biol.*, **11**, 171–181.
8. Levy, M.Z., Allsopp, R.C., Futcher, A.B., Greider, C.W. and Harley, C.B. (1992) Telomere end-replication problem and cell aging. *J. Mol. Biol.*, **225**, 951–960.
9. Kim, N.W., Piatyszek, M.A., Prowse, K.R., Harley, C.B., West, M.D., Ho, P.L., Coviello, G.M., Wright, W.E., Weinrich, S.L. and Shay, J.W. (1994) Specific association of human telomerase activity with immortal cells and cancer. *Science*, **266**, 2011–2015.
10. Shay, J.W. and Bacchetti, S. (1997) A survey of telomerase activity in human cancer. *Eur. J. Cancer*, **33**, 787–791.
11. Hiyama, E. and Hiyama, K. (2007) Telomere and telomerase in stem cells. *Br. J. Cancer*, **96**, 1020–1024.
12. Zahler, A.M., Williamson, J.R., Cech, T.R. and Prescott, D.M. (1991) Inhibition of telomerase by G-quartet DNA structures. *Nature*, **350**, 718–720.
13. Tahara, H., Shin-Ya, K., Scimiya, H., Yamada, H., Tsuruo, T. and Ide, T. (2006) G-Quadruplex stabilization by telomestatin induces TRF2 protein dissociation from telomeres and anaphase bridge formation accompanied by loss of the 3' telomeric overhang in cancer cells. *Oncogene*, **25**, 1955–1966.
14. Zhou, G., Liu, X., Li, Y., Xu, S., Ma, C., Wu, X., Cheng, Y., Yu, Z., Zhao, G. and Chen, Y. (2016) Telomere targeting with a novel G-quadruplex-interactive ligand BRACO-19 induces T-loop disassembly and telomerase displacement in human glioblastoma cells. *Oncotarget*, **7**, 14925–14939.
15. Neidle, S. (2010) Human telomeric G-quadruplex: the current status of telomeric G-quadruplexes as therapeutic targets in human cancer. *FEBS J.*, **277**, 1118–1125.
16. Lane, A.N., Chaires, J.B., Gray, R.D. and Trent, J.O. (2008) Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res.*, **36**, 5482–5515.
17. Parkinson, G.N., Lee, M.P. and Neidle, S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*, **417**, 876–880.
18. Dai, J., Punchihewa, C., Ambrus, A., Chen, D., Jones, R.A. and Yang, D. (2007) Structure of the intramolecular human telomeric G-quadruplex in potassium solution: a novel adenine triple formation. *Nucleic Acids Res.*, **35**, 2440–2450.
19. Dai, J., Carver, M., Punchihewa, C., Jones, R.A. and Yang, D. (2007) Structure of the Hybrid-2 type intramolecular human telomeric G-quadruplex in K⁺ solution: insights into structure polymorphism of the human telomeric sequence. *Nucleic Acids Res.*, **35**, 4927–4940.
20. Wang, Y. and Patel, D.J. (1993) Solution structure of the human telomeric repeat d[AG3(T2AG3)3] G-tetraplex. *Structure*, **1**, 263–282.
21. Zhang, Z., Dai, J., Veliath, E., Jones, R.A. and Yang, D. (2010) Structure of a two-G-tetrad intramolecular G-quadruplex formed by a variant human telomeric sequence in K⁺ solution: insights into the interconversion of human telomeric G-quadruplex structures. *Nucleic Acids Res.*, **38**, 1009–1021.
22. Ambrus, A., Chen, D., Dai, J., Bialis, T., Jones, R.A. and Yang, D. (2006) Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Res.*, **34**, 2723–2735.
23. Luu, K.N., Phan, A.T., Kuryavyi, V., Lacroix, L. and Patel, D.J. (2006) Structure of the human telomere in K⁺ solution: an intramolecular (3 + 1) G-quadruplex scaffold. *J. Am. Chem. Soc.*, **128**, 9963–9970.
24. Phan, A.T., Kuryavyi, V., Luu, K.N. and Patel, D.J. (2007) Structure of two intramolecular G-quadruplexes formed by natural human telomere sequences in K⁺ solution. *Nucleic Acids Res.*, **35**, 6517–6525.
25. Phan, A.T., Luu, K.N. and Patel, D.J. (2006) Different loop arrangements of intramolecular human telomeric (3+1) G-quadruplexes in K⁺ solution. *Nucleic Acids Res.*, **34**, 5715–5719.
26. Li, J., Correia, J.J., Wang, L., Trent, J.O. and Chaires, J.B. (2005) Not so crystal clear: the structure of the human telomere G-quadruplex in solution differs from that present in a crystal. *Nucleic Acids Res.*, **33**, 4649–4659.
27. Dai, J., Carver, M. and Yang, D. (2008) Polymorphism of human telomeric quadruplex structures. *Biochimie*, **90**, 1172–1183.
28. Schaffitzel, C., Berger, I., Postberg, J., Hanes, J., Lipps, H.J. and Pluckthun, A. (2001) In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with *Styloynchia lemnae* macronuclei. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 8572–8577.
29. 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.
30. Henderson, A., Wu, Y., Huang, Y.C., Chavez, E.A., Platt, J., Johnson, F.B., Brosh, R.M. Jr, Sen, D. and Lansdorp, P.M. (2014) Detection of G-quadruplex DNA in mammalian cells. *Nucleic Acids Res.*, **42**, 860–869.
31. Bao, H.L., Liu, H.S. and Xu, Y. (2019) Hybrid-type and two-tetrad antiparallel telomere DNA G-quadruplex structures in living human cells. *Nucleic Acids Res.*, **47**, 4940–4947.
32. Lin, C. and Yang, D. (2017) Human telomeric G-quadruplex structures and G-quadruplex-interactive compounds. *Methods Mol. Biol.*, **1587**, 171–196.
33. Lim, K.W., Amrane, S., Bouaziz, S., Xu, W., Mu, Y., Patel, D.J., Luu, K.N. and Phan, A.T. (2009) Structure of the human telomere in K⁺ solution: a stable basket-type G-quadruplex with only two G-tetrad layers. *J. Am. Chem. Soc.*, **131**, 4301–4309.
34. Buscaglia, R., Gray, R.D. and Chaires, J.B. (2013) Thermodynamic characterization of human telomere quadruplex unfolding. *Biopolymers*, **99**, 1006–1018.
35. Moyzis, R.K., Buckingham, J.M., Cram, L.S., Dani, M., Deaven, L.L., Jones, M.D., Meyne, J., Ratliff, R.L. and Wu, J.R. (1988) A highly conserved repetitive DNA sequence, (TTAGGG)_n, present at the telomeres of human chromosomes. *Proc. Natl. Acad. Sci. U.S.A.*, **85**, 6622–6626.
36. Palm, W. and de Lange, T. (2008) How shelterin protects mammalian telomeres. *Annu. Rev. Genet.*, **42**, 301–334.
37. Palm, W., Hockemeyer, D., Kibe, T. and de Lange, T. (2009) Functional dissection of human and mouse POT1 proteins. *Mol. Cell. Biol.*, **29**, 471–482.
38. Yu, H.Q., Miyoshi, D. and Sugimoto, N. (2006) Characterization of structure and stability of long telomeric DNA G-quadruplexes. *J. Am. Chem. Soc.*, **128**, 15461–15468.
39. Renciu, D., Kejnovska, I., Skolakova, P., Bednarova, K., Motlova, J. and Vorlickova, M. (2009) Arrangements of human telomere DNA quadruplex in physiologically relevant K⁺ solutions. *Nucleic Acids Res.*, **37**, 6625–6634.
40. Xu, Y., Ishizuka, T., Kurabayashi, K. and Komiyama, M. (2009) Consecutive formation of G-quadruplexes in human telomeric-overhang DNA: a protective capping structure for telomere ends. *Angew. Chem. Int. Ed. Engl.*, **48**, 7833–7836.
41. Hänsel, R., Löhr, F., Trantirek, L. and Dötsch, V. (2013) High-resolution insight into G-overhang architecture. *J. Am. Chem. Soc.*, **135**, 2816–2824.
42. Abraham Punnoose, J., Cui, Y., Koirala, D., Yangyuru, 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.
43. Bugaut, A. and Alberti, P. (2015) Understanding the stability of DNA G-quadruplex units in long human telomeric strands. *Biochimie*, **113**, 125–133.
44. 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.
45. Kar, A., Jones, N., Arat, N.O., Fishel, R. and Griffith, J.D. (2018) Long repeating (TTAGGG)_n single-stranded DNA self-condenses into compact beaded filaments stabilized by G-quadruplex formation. *J. Biol. Chem.*, **293**, 9473–9485.
46. Abraham Punnoose, J., Ma, Y., Hoque, M.E., Cui, Y., Sasaki, S., Guo, A.H., Nagasawa, K. and Mao, H. (2018) Random formation of G-quadruplexes in the full-length human telomere overhangs leads to a kinetic folding pattern with targetable vacant G-tracts. *Biochemistry*, **57**, 6946–6955.
47. De Carlo, S. and Harris, J.R. (2011) Negative staining and cryo-negative staining of macromolecules and viruses for TEM. *Micron*, **42**, 117–131.
48. Neuman, K.C. and Nagy, A. (2008) Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nat. Methods*, **5**, 491–505.
49. Petraccone, L., Trent, J.O. and Chaires, J.B. (2008) The tail of the telomere. *J. Am. Chem. Soc.*, **130**, 16530–16532.
50. Petraccone, L., Spink, C., Trent, J.O., Garbett, N.C., Mekmaysy, C.S., Giancola, C. and Chaires, J.B. (2011) Structure and stability of

- higher-order human telomeric quadruplexes. *J. Am. Chem. Soc.*, **133**, 20951–20961.
51. Chaires, J.B., Dean, W.L., Le, H.T. and Trent, J.O. (2015) Hydrodynamic models of G-quadruplex structures. *Methods Enzymol.*, **562**, 287–304.
 52. Singh, V., Azarkh, M., Drescher, M. and Hartig, J.S. (2012) Conformations of individual quadruplex units studied in the context of extended human telomeric DNA. *Chem. Commun. (Camb.)*, **48**, 8258–8260.
 53. Petraccone, L., Garbett, N.C., Chaires, J.B. and Trent, J.O. (2010) An integrated molecular dynamics (MD) and experimental study of higher order human telomeric quadruplexes. *Biopolymers*, **93**, 533–548.
 54. Schneidman-Duhovny, D., Kim, S.J. and Sali, A. (2012) Integrative structural modeling with small angle X-ray scattering profiles. *BMC Struct. Biol.*, **12**, 17.
 55. Rout, M.P. and Sali, A. (2019) Principles for integrative structural biology studies. *Cell*, **177**, 1384–1403.
 56. Chi, Q., Wang, G. and Jiang, J. (2013) The persistence length and length per base of single-stranded DNA obtained from fluorescence correlation spectroscopy measurements using mean field theory. *Physica A*, **392**, 1072–1079.
 57. Bloomfield, V.A., Crothers, D.M. and Tinoco, I. Jr (2000) In: *Nucleic Acids: Structures, Properties, and Functions*. University Science Books, Sausalito.
 58. Miller, M.C., Le, H.T., Dean, W.L., Holt, P.A., Chaires, J.B. and Trent, J.O. (2011) Polymorphism and resolution of oncogene promoter quadruplex-forming sequences. *Org. Biomol. Chem.*, **9**, 7633–7637.
 59. Irvine, G.B. (2001) Determination of molecular size by size-exclusion chromatography (gel filtration). *Curr. Protoc. Cell Biol.*, doi:10.1002/0471143030.cb050506.
 60. Schuck, P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.*, **78**, 1606–1619.
 61. Del Villar-Guerra, R., Gray, R.D. and Chaires, J.B. (2017) Characterization of quadruplex DNA structure by circular dichroism. *Curr. Protoc. Nucleic Acid Chem.*, **68**, 17.18.11–17.18.16.
 62. Kirby, N., Cowieson, N., Hawley, A.M., Mudie, S.T., McGilivray, D.J., Kusel, M., Samardzic-Boban, V. and Ryan, T.M. (2016) Improved radiation dose efficiency in solution SAXS using a sheath flow sample environment. *Acta Crystallogr. D Struct. Biol.*, **72**, 1254–1266.
 63. Hopkins, J.B., Gillilan, R.E. and Skou, S. (2017) BioXTAS RAW: improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis. *J. Appl. Crystallogr.*, **50**, 1545–1553.
 64. Trewella, J., Duff, A.P., Durand, D., Gabel, F., Guss, J.M., Hendrickson, W.A., Hura, G.L., Jacques, D.A., Kirby, N.M., Kwan, A.H. et al. (2017) 2017 publication guidelines for structural modelling of small-angle scattering data from biomolecules in solution: an update. *Acta Crystallogr. D Struct. Biol.*, **73**, 710–728.
 65. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.
 66. Schrodinger (2018) Schrodinger, LLC, NY.
 67. Case, D.A., Ben-Shalom, I.Y., Brozell, S.R., Cerutti, D.S., Cheatham, T.E. III, Cruzeiro, V.W.D., Darden, T.A., Duke, R.E., Ghoreishi, D., Gilson, M.K. et al. (2018) Amber 2018.
 68. Ortega, A., Amorós, D. and García de la Torre, J. (2011) Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. *Biophys. J.*, **101**, 892–898.
 69. Le, H.T., Buscaglia, R., Dean, W.L., Chaires, J.B. and Trent, J.O. (2013) Calculation of hydrodynamic properties for G-quadruplex nucleic acid structures from in silico bead models. *Top. Curr. Chem.*, **330**, 179–210.
 70. Dolinsky, T.J., Nielsen, J.E., McCammon, J.A. and Baker, N.A. (2004) PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.*, **32**, W665–W667.
 71. Baker, N.A., Sept, D., Joseph, S., Holst, M.J. and McCammon, J.A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA*, **98**, 10037–10041.
 72. Tria, G., Mertens, H.D., Kachala, M. and Svergun, D.I. (2015) Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCr*, **2**, 207–217.
 73. Bernado, P., Mylonas, E., Petoukhov, M.V., Blackledge, M. and Svergun, D.I. (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. *J. Am. Chem. Soc.*, **129**, 5656–5664.
 74. Svergun, D. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.*, **25**, 495–503.
 75. Svergun, D., Barberato, C. and Koch, M.H.J. (1995) CRY SOL – a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.*, **28**, 768–773.
 76. Capp, J.A., Hagarman, A., Richardson, D.C. and Oas, T.G. (2014) The statistical conformation of a highly flexible protein: small-angle X-ray scattering of *S. aureus* protein A. *Structure*, **22**, 1184–1195.
 77. Ortega, A. and García de la Torre, J. (2007) Equivalent radii and ratios of radii from solution properties as indicators of macromolecular conformation, shape, and flexibility. *Biomacromolecules*, **8**, 2464–2475.
 78. Amorós, D., Ortega, A. and García de la Torre, J. (2011) Hydrodynamic properties of wormlike Macromolecules: Monte Carlo simulation and global analysis of experimental data. *Macromolecules*, **44**, 5788–5797.
 79. Garcia, H.G., Grayson, P., Han, L., Inamdar, M., Kondev, J., Nelson, P.C., Phillips, R., Widom, J. and Wiggins, P.A. (2007) Biological consequences of tightly bent DNA: the other life of a macromolecular celebrity. *Biopolymers*, **85**, 115–130.
 80. Fink, H.-P. (1989) Structure analysis by small-angle X-ray and neutron scattering. *Acta Polym.*, **40**, 224–224.
 81. Kikhney, A.G. and Svergun, D.I. (2015) A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Lett.*, **589**, 2570–2577.
 82. Bhattacharjee, S.M., Giacometti, A. and Maritan, A. (2013) Flory theory for polymers. *J. Phys. Condens. Matter*, **25**, 503101.
 83. Murphy, M.C., Rasnik, I., Cheng, W., Lohman, T.M. and Ha, T. (2004) Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy. *Biophys. J.*, **86**, 2530–2537.
 84. Rambo, R.P. and Tainer, J.A. (2011) Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers*, **95**, 559–571.
 85. Del Villar-Guerra, R., Trent, J.O. and Chaires, J.B. (2018) G-Quadruplex secondary structure obtained from circular dichroism spectroscopy. *Angew. Chem. Int. Ed. Engl.*, **57**, 7171–7175.
 86. Greve, J., Maestre, M.F. and Levin, A. (1977) Circular dichroism of adenine and thymine containing synthetic polynucleotides. *Biopolymers*, **16**, 1489–1504.
 87. Lim, C.J., Zaug, A.J., Kim, H.J. and Cech, T.R. (2017) Reconstitution of human shelterin complexes reveals unexpected stoichiometry and dual pathways to enhance telomerase processivity. *Nat. Commun.*, **8**, 1075.
 88. 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.
 89. 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.
 90. Flynn, R.L., Chang, S. and Zou, L. (2012) RPA and POT1: friends or foes at telomeres? *Cell Cycle*, **11**, 652–657.
 91. 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.
 92. Flynn, R.L., Centore, R.C., O'Sullivan, R.J., Rai, R., Tse, A., Songyang, Z., Chang, S., Karlseder, J. and Zou, L. (2011) TERRA and hnRNPA1 orchestrate an RPA-to-POT1 switch on telomeric single-stranded DNA. *Nature*, **471**, 532–536.
 93. 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.
 94. Chen, Y. (2019) The structural biology of the shelterin complex. *Biol. Chem.*, **400**, 457–466.

95. McCauley, M.J. and Williams, M.C. (2007) Mechanisms of DNA binding determined in optical tweezers experiments. *Biopolymers*, **85**, 154–168.
96. Gao, C., Liu, Z., Hou, H., Ding, J., Chen, X., Xie, C., Song, Z., Hu, Z., Feng, M., Mohamed, H.I. *et al.* (2020) BMPQ-1 binds selectively to (3+1) hybrid topologies in human telomeric G-quadruplex multimers. *Nucleic Acids Res.*, **48**, 11259–11269.
97. Phatak, P., Cookson, J.C., Dai, F., Smith, V., Gartenhaus, R.B., Stevens, M.F. and Burger, A.M. (2007) Telomere uncapping by the G-quadruplex ligand RHPS4 inhibits clonogenic tumour cell growth in vitro and in vivo consistent with a cancer stem cell targeting mechanism. *Br. J. Cancer*, **96**, 1223–1233.