Published in final edited form as:

Photochem Photobiol. 2018 July ; 94(4): 685–697. doi:10.1111/php.12898.

Evidence for Reverse Hoogsteen Hairpin Intermediates in the Photocrosslinking of Human Telomeric DNA Sequences

Chen Lu, Jillian E. Smith-Carpenter, and John-Stephen A. Taylor

Department of Chemistry, Washington University, St. Louis, MO 63130, USA

Abstract

UVB irradiation of human telomeric (GGGTTA)₃GGG sequences in potassium ion solution crosslinks the first and third TTA segments through *anti* cyclobutane pyrimidine dimer (CPD) formation. The photocrosslinking reaction was first proposed to occur through a form 3 two tetrad G-quadruplex in which the lateral 4-nucleotide GTTA loop can interact with an adjacent TTA loop. Curiously, the reaction does not occur with sodium ion, which was explained by the formation of a basket structure which only has 3-nucleotide TTA loops that cannot interact. Sequences known or expected to favor the two-tetrad basket did not show enhanced photocrosslinking, suggesting that some other structure was the reactive intermediate. Herein, we report that *anti* CPDs form in human telomeric DNA sequences with lithium ion that is known to disfavor G-quadruplex formation, as well as with potassium ion when the bases are modified to interfere with G-quartet formation. We also show that *anti*-CPDs form in sequences containing A's in place of G's that cannot form Hoogsteen hairpins, but can form reverse Hoogsteen hairpins. These results suggest that reverse Hoogsteen hairpins may play a hitherto unrecognized role in the biology and photoreactivity of DNA in telomeres, and possibly in other purine-rich sequences found in regulatory regions.

Graphical abstract

Irradiation of human telomeric DNA in the presence of potassium ion produces *anti* cyclobutane pyrimidine dimer photoproducts which we attributed to the two tetrad form 3 G quadruplex with 3 and 4 nucleotide loops. No anti photoproducts are produced in the presence of sodium ion which we attributed to the basket G quadruplex structure that only has 3 nucleotide loops and possible chelation of one of the thymines. We show that an *anti* CPD is formed in high yield in the presence of lithium ions which cannot stabilize G quadruplex structures which we attribute to a reverse Hoogsteenhairpin structure.

Correspondence to: John-Stephen A. Taylor.

Supporting Information: Additional Supporting Information may be found in the online version of this article:

Introduction

Anti cyclobutane pyrimidine dimers (CPDs) (Fig. 1A) were first discovered in the ultraviolet (UV) light irradiation products of thymidine in frozen aqueous solutions (1). Later, these products were detected in the irradiation products of dehydrated or alcoholic calf thymus DNA (2). A *cis,anti* cyclobutane thymine dimer was then accidentally discovered to form in high yield in aqueous solution during the irradiation of a 14-mer oligodeoxynucleotide at acidic pH but not neutral pH (3). The photocrosslinked T's in the 14-mer were separated by four nucleotides, suggesting the involvement of an as yet unknown folded structure. This result prompted a study of the photochemistry of human G quadruplex forming telomeric sequences (Fig. 1B) which can fold in aqueous solution at biologically relevant pH's and cations to form basket structures. Basket structures formed from the human telomere sequence have adjacent lateral loops containing T's that were expected to be able overlap and form *anti* CPDs (4). Indeed, *anti* thymine dimers were discovered to be produced in the human G-quadruplex forming telomeric sequence Tel22 (Fig. 1C) under neutral pH, though not in an entirely expected manner. What was puzzling was that irradiation of the Tel22 sequence in Na⁺ solution, which favors a basket structure (5), did not lead to significant amounts of *anti* CPDs, whereas irradiation in K⁺ solution, which favors the hybrid-1 (6, 7) and hybrid-2 (8) structures (Fig. 1B), did lead to the *anti* CPDs, in spite of their lack of adjacent lateral loops. Formation of *anti* CPDs has implications for the biological effects of sunlight and for the structure of telomeric DNA that is therefore important to understand.

We proposed that the hybrid structures favored in K⁺ solution were also in equilibrium with a more recently discovered two tetrad basket-like form 3 structure that the sequence NF3 adopts (9). The form 3 structure has one lateral loop (GTTA) that is one nucleotide longer than the corresponding TTA loop in the three tetrad basket structure adopted in Na⁺ solution, that was expected to better overlap with the adjacent TTA loop to enable formation of the *anti* CPD (4). This hypothesis was tested recently (10) by comparing the photochemistry of the Tel26 sequence (Fig. 1C), which exists primarily in a non-photoreactive hybrid-1 conformation (6), and produces almost exclusively the *trans,anti* T(A)=T(A) CPD, with that of the NF3 sequence, which exists primarily in the two G-tetrad form 3 conformation (9). While both sequences produced the *trans,anti* T(A)=T(A) CPD in about the same yield upon UVB irradiation, *anti* CPD formation was faster and more selective in Tel26 than in NF3 suggesting that the form 3 structure was not the major photoreactive conformation. In

addition, irradiation of a number of mutants specifically designed to stabilize a four-nucleotide loop-containing basket structure in both K^+ and Na^+ solution, while producing the *trans,anti* T(A)=T(A) CPD in K^+ solution, failed to produce any *anti* CPDs in Na^+ solution. The conclusion of our study was that some other conformation must also be involved, such as a chair or hairpin structure, in spite of QM/MM modeling calculations suggesting that the *trans,anti* CPD could arise from the form 3 structure (11).

To gain further insight into the photoreactive conformations involved in *anti* CPD formation, we now report the photochemistry of G-quadruplex forming sequences in the presence of various metal ions known to either stabilize or destabilize G-quadruplexes. We also report on the photochemistry of truncated sequences and sequences in which inosine and adenosine are used in place of guanosine to disrupt G-quartet formation in Na^+ and K^+ solution. From these studies we have discovered that photocrosslinking of T's can occur in structures other than quadruplexes, and provide evidence that implicates antiparallel reverse-Hoogsteen base-paired hairpins as the photoreactive conformation leading to the *trans,anti* T(A)=T(A) CPD. The results suggest that reverse Hoogsteen hairpin structures co-exist with G-quadruplexes, and possibly play a hitherto unrecognized role in DNA and RNA structure and function.

Materials and Methods

General

Oligodeoxynucleotides were purchased from Integrated DNA Technologies, Inc. (IDT) (Coralville, Iowa). Nuclease P1 (NP1) from *Penicillium citrinum*, was from Sigma (St. Louis, MO). HPLC solvents were from Fisher (Fair Lawn, NJ). HPLC separation and analysis were carried out on System Gold HPLC system with a binary gradient Model 125 pump and a Model 168 diode array detector (Beckman Coulter, Inc., Fullerton, CA). An X-Bridge column (C18, 4.6×75 mm, $2.5 \mu m$, $135 \text{ \AA}$) from Waters Corporation (Milford, MA) was used for reverse-phase HPLC. UVB (280-320 nm) irradiation was carried out with two Spectroline XX-15B UV 15W tubes (312 nm) with peak UV intensity of $1150 \mu W/cm^2$ at 25 cm filtered through a Longlife filter glass from Spectronics Corporation (Westbury, NY). CD experiments were carried out on a J-810 spectropolarimeter (Jasco). UV melting curves were obtained on a Cary 100 Bio UV-VIS Spectrometer (Varian).

Preparation and characterization of G-quadruplexes

Typically, 50 μM ODN (IDT) in 10 mM Tris-HCl, pH 7.5, with 100 mM LiCl, NaCl or KCl was heated at 95 °C for 10 min and then rapidly cooled down in ice and kept at 4 °C. CD experiments were carried out with 5 μM samples in a 1 cm pathlength cell at 4°C.

Melting temperature measurements

ODN samples (100 μL of 50 μM) in 10 mM Tris, 150 mM KCl, pH 7.4 buffer were preheated in cuvette to 95 °C for 10 min and any air bubbles that formed were removed by tapping the cuvette. The absorbance was recorded at 260 or 295 nm and the initial absorbance for all samples were zeroed against buffer only. Samples were heated and cooled at a rate of 1 °C per min from 5 to 95 °C. Samples were first denatured, then annealed, and

then denatured again. The T_m was calculated from the intersection of the annealing curve and the mean of linear fits to the upper and lower limits of the annealing curve.

UVB Irradiation and product characterization

UVB irradiation was carried out immediately at 4°C after sample preparation or after storage at 4°C for overnight to days with similar results. G-quadruplex samples (50 μ M) were irradiated in a polyethylene microcentrifuge tube and irradiated on a bed of ice for 2.5 h at approximately a distance of 1 cm from the UVB lamp. Irradiated samples (100 μ L of 50 μ M DNA) were digested with 1 μ L of 1 U/ μ L aqueous NP1 from *Penicillium citrinum* (Sigma) and 1 μ L of 10 mM $ZnCl_2$ at 37 °C for >36 h. Reverse-phase HPLC was carried out with an X-Bridge column (C18, 4.6 $\times$ 75 mm, 2.5 μ m, 135 Å; Waters Corporation) NP1 digestion products were eluted with 1 mL/min of 100% solvent A (50 mM triethylammonium acetate, pH 7.5) for 3 min followed by a linear gradient of 0-20% B (50% acetonitrile in 50 mM triethylammonium acetate, pH 7.5) in solvent A for 3-53 min and detected at 260 nm. The yield of each photoproduct-containing digestion product was calculated from the integral of the corresponding HPLC peak at 260 nm by using molar extinction coefficients for the appended undamaged nucleotide or nucleotides at 260 nm relative to that for one of the undamaged nucleotides. HPLC fractions corresponding to photoproduct-containing NP1 digestion fragments were dried *in vacuo*. Milli-Q water was added to each sample to yield a final ODN concentration of approximately 50 μ M. One microliter of sample was mixed with 1 mL of 3-HPA with 10% ammonium citrate and spotted on a stainless-steel MALDI plate. Spectra were taken in reflectron positive ion mode and were the average of 20 accumulations.

Denaturing PAGE analysis of photoproducts

The 5' termini of the ODNs were labeled by a forward reaction with 1 $\times$ reaction buffer (500 mM Tris•HCl, pH 7.6, 100 mM $MgCl_2$, 50 mM DTT, 1 mM spermidine, 1 mM EDTA; Fermentas Life Science) with 10 U T4 polynucleotide kinase and 20 pmol [α - ^{32}P]-ATP at 37 °C for 1 h. The reaction was quenched by adding to an equal volume of 2 $\times$ loading buffer (98% formamide, 10 mM EDTA), boiling for 2 min, and cooled to room temperature. The radiolabeled samples were then analyzed on a 15% denaturing polyacrylamide gel and scanned with a Personal Molecular Imager (Bio-Rad Laboratories) and analyzed with Quantity One software.

Native PAGE analysis of DNA structures

Native PAGE was performed as generally described (12). Briefly, 5'- ^{32}P -end labeled oligonucleotides were purified by 15% Denaturing PAGE. Then, each sample was prepared by mixing 5 μ L of 20 μ M radiolabeled DNA, 1 μ L 100 mM Tris (pH 7.5), 1 μ L 1 M LiCl, NaCl, or KCl, and 3 μ L dd H_2O . The samples were boiled for 5 min and cooled immediately on ice for 2 h. An equal volume of 2 $\times$ loading dye containing 46% glycerol and 5% xylene cyanol were added to the samples and 10 μ L of each sample was loaded on 15% Native PAGE containing 100 mM the respective monovalent cations. Metal ion-free 5 $\times$ TBE buffer was prepared with 0.0125 M EDTA (ethylenediaminetetraacetic acid, acid form), 0.455 M Tris and 0.445 M boric acid in dd H_2O . The 15% native PAGE gels were prepared using 29:1 acrylamide:bisacrylamide, in 1 $\times$ TBE buffer containing the indicated concentration of

cation. The gels were pre-equilibrated with ice-cold 1× TBE running buffer containing the monovalent cation and run at 400 V for at about 30 min at 4°C in a cold room prior to loading the samples. Then the gel was run at 400 V at 4°C until xylene cyanol moved 14~15 cm on the gel.

Results

Methodology

The method used for assaying CPD formation in DNA makes use of an enzyme-coupled HPLC assay in which nuclease P1 degrades adjacent photoproducts to pN=N(N) trinucleotides (13), and non-adjacent photoproducts to pT(N)=pT(N) tetranucleotides (4, 3), hereafter referred to as T(N)=T(N) (Fig. 1A). The identity of the nucleotides N can be deduced from MALDI MS, and the stereochemistry of the CPD from acid hydrolysis to release the base portion and comparison to authentic by HPLC. The TTA loops that are crosslinked can be deduced by replacement of specific T's with U's which changes the retention times of CPD-containing tetranucleotides.

Photoproduct formation in Tel26 as a function of metal ion

Tel26 was irradiated with UVB light in the presence of the monovalent alkali metal ions, Li⁺, Na⁺, and K⁺. In Li⁺ and K⁺ solutions the *trans,anti* T(A)=T(A) CPD was observed to be the major photoproduct, whereas in the presence of Na⁺, *anti* CPD formation was insignificant, (Fig. 2). Most surprisingly, *trans,anti* T(A)=T(A) CPD formation was higher with Li⁺ (23%) than with K⁺ (15%). To determine whether or not intermolecular base-paired structures might be intermediates in the formation of any of the products, the irradiated samples of 5'-³²P-end-labeled Tel26 were electrophoresed (See Supporting Information, Fig. S1A). In all cases only a band corresponding to a monomeric species was detected suggesting that the *anti* CPDs were arising from an intramolecularly folded structure.

Effect of metal ions on the conformation and stability of Tel26

The formation of the *trans,anti* CPD in higher yield in Li⁺ solution than in K⁺ solution was quite surprising as this ion is well known to inhibit G-quadruplex formation (14-16). To verify that G-quadruplexes were not forming in the presence of the Li⁺ ion and to establish the types of structures that might be present in the presence of the other ions, we obtained the CD spectra and thermal denaturation curves at 295 nm, a wavelength known to be characteristic for antiparallel G-quadruplexes (17-19) (Fig. 3). The structurally well characterized Tel26 showed large positive peaks at around 265 and 290 nm in the CD spectrum taken in K⁺ solution, characteristic of a hybrid quadruplex structure containing parallel and antiparallel strands, respectively. In Na⁺ solution, there was a single large positive peak at 300 nm characteristic of an antiparallel quadruplex. In contrast, the CD spectrum in Li⁺ had a weak positive peak at 275 nm, and the absence of the 265 or 295 nm peaks characteristic of parallel or antiparallel G-quadruplexes. Whereas the T_m values for Tel26 in Na⁺ and K⁺ were about 52°C and 60°C respectively when monitored at 295 nm, no melting behavior was observed in Li⁺. These results are consistent with previous studies showing decreasing stability for intramolecular quadruplexes of d(TTAGGG)₄ in the order K

$^+ > \text{Na}^+ \gg \text{Li}^+$ (20) and intermolecular quadruplexes of d(CGCG₃GCG), in the order $\text{K}^+ > \text{Na}^+ > \text{Li}^+$ (14).

Temperature dependence and kinetics of *anti* CPD formation in Tel26 in Li^+ and K^+ solution

In a previous study, we found that the yield of the *trans,anti* T(A)=T(A) CPD formed from the irradiation of Tel26 in K^+ solution dropped from 12% to about 1% on raising the temperature from 0° to 37°C (10). The decrease in yield with increasing temperature had been explained by either a decrease in the population of the reactive intermediate, or to a decrease in the lifetime of the photoreactive intermediate. The interpretation of this temperature dependence was complicated by the presence of the non-photoreactive hybrid-1 as the major conformation in K^+ solution (6, 8). Since quadruplex conformations are not stable in Li^+ solution, it was expected that the temperature dependence of *trans,anti* CPD formation in Tel26 in Li^+ solution might be more directly related to the stability of the conformation present under these conditions. A likely candidate is a hairpin structure composed of reverse Hoogsteen G•G base pairs (Fig. 4) that has been previously been shown by NMR to form in the *Oxytricha* telomere sequence, G₄T₄G₄, in Li^+ solution (21), though more recent studies have suggested the formation of symmetric G•G base pairs (Fig. 4, S2) (22).

Analysis of the temperature dependence of the UV absorption of Tel26 at 260 nm in Li^+ solution showed that it had a T_m of about 40°C (Fig. 5A). The decrease in the fraction of the helical form with an increase in temperature roughly paralleled the decrease in the relative yield of *trans,anti* T(A)=T(A) CPD from 0 °C to 40 °C (Fig. 5A). The temperature dependence of *trans,anti* CPD formation in Li^+ was similar to what we had previously observed in K^+ solution (10), supporting the idea that the reverse Hoogsteen hairpin might be the photoreactive conformation. The kinetics of *trans,anti* CPD formation in Li^+ and K^+ also supported this notion. The initial rate of *trans,anti* T(A)=T(A) CPD formation was 2.4 times greater in Li^+ than in K^+ solution at 0 °C (Fig. 5B) as would be expected if the reverse Hoogsteen hairpin conformation was the photoreactive conformation, since there would be no competing non-photoreactive hybrid-1 quadruplex conformation in Li^+ solution. It was also found that the initial rate of *cis,syn* T=T CPD formation was 4.8-fold faster in Li^+ than in K^+ solution (Fig. 5C). The greater rate with Li^+ could be explained by the fact that all of the TTA CPD sites in the quadruplex forms are in loops, which may inhibit *cis-syn* CPD formation, whereas only one TTA is in a loop in a hairpin structure, and the other two are in the stem where the T's can stack on each other.

Photoproduct formation in 15-mer sequences

Because the *trans,anti* photoproduct was observed to be formed in the presence of Li^+ which does not stabilize G-quadruplex structures (14-16), but is known to stabilize reverse Hoogsteen hairpins (21), further attention was focused on hairpin structures as potential photoreactive intermediates. To this end, the photochemistry of Tel15, which consists only of loop 1 through loop 3 of Tel26, was investigated, and while too short to form an intramolecular quadruplex or triplex, it could still form a reverse or normal Hoogsteen hairpin, or a bimolecular quadruplex. To assess the importance Hoogsteen-type base pairing on *anti* CPD formation, the photochemistry of HP15WC, which can only form a Watson

Crick base paired hairpin, was also investigated. UVB irradiation of Tel15 gave the *trans,anti* T(A)=T(A) product in about the same yield in Li⁺ and K⁺ solutions as observed for Tel26 in solutions of these same ions (Fig. 6). Some of the *trans,anti* CPD formed in Na⁺ and K⁺ may have come, however, from a bimolecular G-quadruplex, as a denaturing gel of the irradiated duplexes showed some slower moving bands that would be consistent with dimers in addition to the faster moving intramolecular crosslink product (Fig. S1B).

The dimer bands observed in the gel could have arisen either through photocrosslinking of a bimolecular G-quadruplex with adjacent lateral loops, or a Hoogsteen or reverse Hoogsteen duplex. The dimer bands could not, however, have arisen from dimeric quadruplex forms of the types proposed for the *Oxytricha* d(GGGGTTTTGGGG) that have loops on opposite ends (23). The CD of Tel15 in K⁺ is very similar to that of Tel26 in K⁺ solution, with maxima at 265 nm and 290 nm suggesting the formation of a bimolecular hybrid type quadruplex (Fig. S3). The CD of Tel15 in Na⁺ solution had a maximum at 290 nm indicating formation of an antiparallel quadruplex as observed for Tel26 in Na⁺ solution but was much weaker in intensity. The Li⁺ sample, however, showed almost no discernable CD from 260-300 nm, indicating the lack of possible involvement of quadruplex type structures in *trans,anti* CPD formation. Most interestingly, *trans,anti* CPD formation was now observed in Na⁺ solution, which had not been the case for the intramolecular quadruplex forming sequences, indicating that the DNA was now able to adopt a photoreactive conformation in the presence of Na⁺. In contrast to the G-quadruplex forming sequences, UVB irradiation of HP15WC, which had a Watson Crick base paired stem, gave very little in the way of *anti* CPD products in the presence of any of the ions, indicating that a hairpin with a B DNA stem does not have the proper geometry for *anti* CPD formation. The presence of a B type conformation in HP15WC was confirmed from the CD (Fig. S3) which showed a positive band at about 280 nm and a negative band at 250 nm for all the ions (24).

Photoproduct formation in inosine and Watson Crick base pair containing 21-mer sequences

To further test the hypothesis that *anti* CPDs could arise from hairpin conformations of Tel26 in K⁺, five 21-mer sequences (Fig. 1C) were prepared in which G's were strategically replaced with inosine (I) to interfere with intramolecular G-quadruplex formation, but not with reverse or normal Hoogsteen hairpin formation (25). In addition, a sequence was prepared in which the two terminal G₃-tracts were replaced by complementary Watson-Crick sequence to prevent intramolecular quadruplex formation and enforce hairpin formation. Inosine is less able to serve as a Hoogsteen or reverse Hoogsteen H-bond donor due to the absence of the N2-NH₂ group to form a second hydrogen bond, while it is still able to serve as a Hoogsteen H-bond acceptor (Fig. 4). In Tel21_3'con and Tel21_5'con, the two consecutive (con) G₃-tracts at the 3'-end or 5'-end of the 21-mer were replaced with I's so that the entire 3'-side or 5'-side of the hairpin would function as the Hoogsteen accepting strand. In Tel21_3'alt and Tel21_5'alt, alternating (alt) sets of G₃-tracts were replaced with I's so that a conformation with a two-fold axis of symmetry (disregarding the hairpin loop) would be produced in which the Hoogsteen accepting strand switches to the opposite strand following the T•T crosslink site at the center. With Tel21_mix, every other G was replaced with inosine so that hairpins formed from this sequence would have the Hoogsteen donor

alternating with the Hoogsteen acceptor. To obtain a *trans,anti* CPD stereochemistry from any of the sequences in a hairpin form, one of the T's of the T•T mispair must be in a *syn* glycosyl conformation. As shown in Fig. 4, only for the reverse Hoogsteen conformation do the two T's come close enough together and with the proper alignment for the required [2+2] cycloaddition to take place. The two TA base pairs flanking the T•T mispair site could in principle form either a reverse or normal Hoogsteen base pair as shown.

Of the five inosine-containing sequences, irradiation of only Tel21_3'con and Tel21_3'alt produced the *trans,anti*T(A)=T(A) CPD with the greatest yield (8% and 10% respectively) and selectively (Fig. 7 & Fig. S4 & S5). The common feature of these sequences is that they share the same 3'-I₃ tract. While the CD for these two sequences in K⁺ are similar as far as having a positive peak at about 290 nm, there is another positive peak at 255 nm of almost equal amplitude in the 3'alt sequence that is much less pronounced in the 3'con sequence (Fig. S3). The peak at 290 nm is indicative of an antiparallel quadruplex that may have arisen from dimerization of the two intact G₃ tracts in each sequence. The CD spectrum in Li⁺ is very weak with a weak positive peak at 250 nm for both sequences. One cannot therefore rule out on the basis of the CD that some of the *trans,anti* CPDs observed may have arisen from such a bimolecular duplexes. On the other hand, the quadruplexes that could form would have quite different structures due to the different relative positions of the G₃-tracts in the two sequences, which would have been expected to lead to different patterns of *anti* CPD products. Furthermore, while Tel21_5'con had a very similar CD to that of Tel21_3'con in K⁺ (Fig. S3) it gave a much more complex mixture of products (Fig. S6), suggesting that the positioning of the G₃ and I₃-tracts is important. Tel21_5'alt also led to a complex mixture of products (Fig. S7) and had a sizable positive peak at 260 nm in the CD spectrum, and a weaker one at 300 nm, indicative of a parallel quadruplex with some antiparallel quadruplex. The CD for these latter two sequences in Li⁺ were very similar to those of the first two sequences. The higher selectivity and yield with sequences having the 3'-I₃ tract could be explained by a hairpin structure in which *trans,anti*T(A)=T(A) CPD formation is favored with the 3'-terminal I₃-tract is serving as the Hoogsteen or reverse Hoogsteen acceptor strand. Replacing the 3'-I₃ sequence with a Watson Crick accepting GCG sequence in HP21WC abolished formation of the *trans,anti* product (Fig. 7). The CD of this sequence in both K⁺ and Li⁺ solution (Fig. S3) shows a weak positive peak at about 270 nm and a weak negative peak at 240 nm that is similar to that of HP15WC and supportive of the presence of a B DNA duplex and an absence of a quadruplex. Replacing the 3'-I₃ sequence with an alternating 3'-GIG sequence in Tel21_mix also greatly reduced the yield to 3% (Fig. 7 & Fig. S8). This sequence cannot form either an intramolecular or intermolecular quadruplex, and the CD spectrum in K⁺ and Na⁺ and similar and show a weak positive peak at 270 nm, but in Li⁺ showed a broad positive peak centered at about 260 nm (Fig. S3).

In Na⁺ solution, the highest apparent yield and selectivity for *trans,anti*T(A)=T(A) CPD formation was for Tel21_3'alt (18%) and Tel21_mix (14%) (Fig. 7 & S8), though Tel21_3'con also gave approximately the same yield (6%) and selectivity as was observed in K⁺ solution. Coinjection experiments with authentic *trans,anti* product suggested, however, that what appears to be the *trans,anti* product in Tel21_3'alt in Na⁺ may actually be a mixture of the *trans,anti* product with a slightly earlier eluting compound (Fig. S5), which

may be the *cis,anti* T(A)=T(A) based on prior work (4). The CD spectra for these three sequences in Na⁺ were quite different from each other, with Tel21_3'alt showing a weak positive peak at 255 nm like in Li⁺, whereas in Tel21_mix there was positive peak at 270 nm like observed in K⁺, while in Tel21_con there were two strong positive peaks at 250 nm and 290 nm. The high yield of the *trans,anti* T(A)=T(A) CPD in Tel21_mix in Na⁺ (Fig. 7) contrasts with the low yield observed in K⁺ indicating that Na⁺ and K⁺ can also greatly affect the conformation of non-quadruplex forming sequences. Sequences with 5'-I₃ arrangements of inosine (5'-alt and 5'-con) led to much lower yields and poor selectivity (Fig. S6-S7). The HP21WC sequence containing a Watson-Crick stem produced very little *trans,anti* CPD (Fig. 7).

The photoreactivity of the inosine substituted sequences is more consistent with the involvement of a hairpin intermediate in the formation of the *trans,anti* T(A)=T(A) CPD than with an intermolecular G-quadruplex intermediate, especially for Tel21_mix in Na⁺, which cannot form any type of G-quadruplex. Furthermore, the preference for a 3'-terminal I₃ sequence indicates that the photoreactive hairpin involves normal or reverse Hoogsteen (Fig. 4) rather than C2-symmetric base pairing (Fig. 4, S2) which would not be expected to be affected by the presence of inosine. The observation of large quantities of the *trans,anti* T(A)=T(A) CPD in the inosine-modified sequences in Na⁺ solution, especially Tel21_mix, indicates that it must be the basket quadruplex conformation that forms in Na⁺ solution, and not Na⁺ by itself, that inhibits *anti* CPD formation in Tel22 and Tel23. Also of interest is the increased yield of the *cis-syn* T=TA CPD products in the inosine-containing sequences. This may be due to the CPD enhancing effect of inosine relative to guanine (26) that was proposed to result from the higher ionization potential of I than G that prevents competitive exciplex formation and deactivation (27).

***Trans,anti* CPD formation in non-telomeric sequences**

At this point it is not possible to unambiguously conclude whether the photoreactive conformation leading to the *trans,anti* T(A)=T(A) CPD is a Hoogsteen or reverse Hoogsteen hairpin or some other type of hairpin. To determine whether or not a reverse Hoogsteen hairpin could be the photoreactive conformation, three 21-mer sequences, HP21GAG, HP21GGA, and TelAAG, were designed in which multiple G•G base pairs of the human telomere sequence were mutated to A•A base pairs. A cannot form Hoogsteen base pairs or quartet structures with itself, though both A and G can form reverse Hoogsteen base pair with themselves (28, 29), as well as symmetric base pairs (Fig. 4 & Fig. S2) (30). All three mutated sequences, led to the formation of the *trans,anti* T(A)=T(A) CPD photoproduct upon irradiation with UVB in both K⁺ and Na⁺ solution (Fig. 8). To determine what structures might be present, CD spectra were obtained for all three sequences in Na⁺ and K⁺ as well as in Li⁺ in which G-quadruplexes would be inhibited.

The CD spectra of HP21GAG in Li⁺, Na⁺ and K⁺ solution are almost identical with a minimum at about 285 nm and a positive peak at 260 nm (Fig. S3) which is the opposite of B DNA and very similar to left-handed or Z form DNA (24). A slightly longer version of this sequence, (TTAGAG)₄ has previously been shown not to adopt a G-quadruplex structure, and has a similar CD spectrum to what we observe (31). One can therefore

conclude that no G-quadruplexes are involved in the formation of the *trans,anti* CPD in this sequence, and that the sequence is adopting a similar distribution of conformations in each ion. On the other hand, The CD spectra of HP21GGA in both K^+ and Na^+ showed a positive peak at about 300 nm (Fig. S3) that could indicate the presence of an antiparallel two tetrad G-quadruplex that may have formed from the four intact G_2 -tracts. The spectrum in Li^+ , however was quite different, showing a negative peak at about 245 nm and a positive peak close to 270 nm. As a result one cannot rule out that the formation of the *trans,anti* T(A)=T(A) product in TelGGA isn't due to a G-quadruplex. The CD spectra of HP21AAG in all three ions was very weak. The CD spectra in K^+ and Li^+ closely resemble each other with a positive peak at about 270 nm and a negative peak at 250 nm, whereas the spectrum in Na^+ was weaker and showed a negative peak at about 285 nm (Fig. S3). The CD spectra establish that HP21GAG and HP21AAG are able to form the *trans,anti* T(A)=T(A) CPD through some conformation other than a G quadruplex, most likely a reverse Hoogsteen (Fig. 4) or possibly a symmetric base paired hairpin (Fig. 4, S2).

Electrophoretic behavior of hairpin and G-quadruplex forming sequences as a function of metal ion

To see if we could obtain evidence for the presence of a hairpin intermediate in the human telomeric Tel26 sequence in K^+ solution, we carried out gel electrophoresis experiments at 4°C (12) on Tel26 and two Tel26 mutants in which selected G•G pairs were replaced with A•A (HP26GAG and HP26AAG) in the presence of $Tris\cdot H^+$, Li^+ , Na^+ , and K^+ (Fig. 9). HP26GAG and HPAAG correspond to HP21GAG and HP21AAG with two additional A's at the 5'-end, and one at the 3'-end to make them the same length as Tel26. Neither sequence showed evidence of G quadruplex formation. The addition of the A's weakened the CD spectrum of HP26GAG compared to HP21GAG, but strengthened the CD spectrum of HP26AAG compared to HP21AAG which now showed a prominent positive peak at about 270 nm and a negative peak at 250 nm (Fig. S3). The CD spectrum of HP26AAG is very similar to that of d(AACC(GAA)₁₅GGATC) containing the Friedreich's ataxia repeat sequence GAA, which was shown to adopt a hairpin structure (32). As controls, we used triplex- and quadruplex-forming sequences d(TTAGGG)₃ (Tel18) and d(TTAGGG)₄ (Tel24) and scrambled versions of all the sequences denoted by S as a prefix.

In the absence of metal ions, Tel26 moved the slowest of all, and significantly slower than its scrambled version, while the other sequences had about the same mobility as their scrambled counterparts (Fig. 9A). In Li^+ , Tel26 went slightly slower than its scrambled counterpart, whereas d(TTAGGG)₄ went slightly faster. The G→A mutated sequences migrated together and more slowly than Tel26, and with the average mobility of their scrambled counterparts and not faster as might have been expected for hairpin formation (Fig. 9B). It may also be that there was significant unintended secondary structure in the scrambled sequences. In Na^+ and K^+ both Tel26 and d(TTAGGG)₄ moved much more quickly than their scrambled counterparts and with an equal or greater mobility than for the d(TTAGGG)₃ 21-mer or its scrambled counterpart that is consistent with G quadruplex formation (Fig. 9C & D) (12). The G→A mutated sequences, however, had about the same mobility as their scrambled counterparts. No additional band was observed in Na^+ or K^+ for d(TTAGGG)₄ or Tel26 that might correspond to another species suggesting that a hairpin

was either not present, or if present, was only present in a very low amount and in rapid equilibrium with the quadruplex structure.

Discussion

We have shown that the *trans-anti* T(A)=T(A) CPD can form in both Na⁺ and K⁺ solution in A-rich sequences only capable of forming reverse-Hoogsteen hairpins or possibly hairpins with symmetric base pairs. It is not possible to conclusively prove that a hairpin is the sole or contributing photoreactive conformation of human telomeric sequences in K⁺ solution because of the preferential formation of more stable G-quadruplexes that make it difficult to independently detect the presence of a hairpin conformation. Under conditions where G-quadruplex formation is inhibited by Li⁺, it is likely that a reverse Hoogsteen hairpin is the photoreactive conformation given previous NMR data showing the formation of this conformation in the *Oxytricha* d(G₄T₄G₄) sequence(21). While the *trans,anti* CPD also forms in Na⁺ or K⁺ solution when half the G's are replaced with I's to prevent intramolecular G quadruplex formation, we cannot be certain if this is due to a reverse or normal Hoogsteen hairpin, or a bimolecular quadruplex, except for Tel21_mix, which cannot form quadruplexes. The photoreactive conformation is unlikely to be due to a triplex conformation, which could also form under these conditions, as we have previously found that sequences that can only form a triplex give a low yield of *trans,anti* T(A)=T(A) (4). None-the-less there is evidence that suggests that a non-Hoogsteen hairpin structure, and in particular, the reverse Hoogsteen structure, is a major photoreactive conformation leading to the formation of the *trans-anti* T(A)=T(A) CPD in K⁺ solution.

The first line of evidence for a photoreactive reverse Hoogsteen hairpin structure is that the yield of the *trans,anti* T(A)=T(A) photoproduct is greater for Tel26 than for Tel22 in K⁺ solution. This can easily be explained by the fact that Tel26 results from the addition of two A's to both the 5' - and 3' -end of Tel22 which can form two additional reverse Hoogsteen base pairs at the end of the hairpin, which would further stabilize a reverse Hoogsteen hairpin structure and increase its relative concentration. This is assuming that the two additional reverse Hoogsteen base pairs could only form on the end of a reverse Hoogsteen hairpin. It is also possible that the A's and G's of the non-Hoogsteen hairpin form C2-symmetric base pairs (Fig. 4, S2) but these would all require one of the glycosyl bonds of each pair to be in the higher energy *syn* form. The symmetric G•G base pair involving the sugar face would seem highly unlikely due to a large steric interaction between the amino group and the sugar ring that is in a *syn* glycosyl conformation (base pair #4, Fig. S2). Also, replacement of either G of the other symmetric G•G base pair (Fig. 4, S2) with inosine should not affect base pairing since the NH₂ group is not involved, and hence should not affect the photochemical outcome, which is contrary to what was observed in some of the inosine substitution experiments where the position of the substitution makes a big difference (Fig. 6-7). The symmetric base pairs are also less likely because they would not have the same conformation as the normal or reverse Hoogsteen conformation needed to form the T•A base pairs, and are also much larger (Fig. 4)

A second line of evidence for a reverse Hoogsteen hairpin is that the major *anti* CPD photoproduct of Tel26 in both Li⁺ and K⁺ is the *trans,anti* T(A)=T(A) CPD, whereas three

additional products are produced in significant yield in Tel22 in K^+ solution. The reverse Hoogsteen structure affords a simple explanation for the selective formation of the *trans,anti* T(A)=T(A) product as shown in Fig. 4. It predicts that the second T in each TTA segment will be opposite each other and lead to the *trans,anti* T(A)=T(A), while the other T's would only be able to form adjacent T=TA CPDs. Furthermore, the *trans* stereochemistry for the T(A)=T(A) CPD can be explained by the requirement for one of the thymidines to adopt the *syn* glycosyl conformation for the 5,6-double bonds to be in close enough proximity to photodimerize. The possibility that the *cis,anti* T(A)=T(A) CPD may also form in the Tel21_3'alt sequence (Fig. S5) could be further explained by the T's being flanked by two reverse Hoogsteen TA base pairs which would pinch the DNA duplex and bring the two T's even closer together. In the normal Hoogsteen conformation, the two double bonds on the two T's are much farther apart and would be significantly less able to photodimerize (Fig. 4).

The third line of evidence is that the yield of the *trans,anti* T(A)=T(A) CPD in Tel26 linearly decreased in Li^+ with an increase in temperature and was almost zero at 37 °C. The temperature dependence closely parallels the melting behavior of the hairpin form in Li^+ (Fig. 5A) which has a T_m of 40 °C. We had observed a virtually identical linear decrease in yield in K^+ solution, which does not parallel the much sharper melting transition at a T_m of around 60°C for the G-quadruplex form (Fig. 3B). The melting temperature of the reverse Hoogsteen hairpin conformation of Tel26 is also expected to be low in K^+ solution based on some recent data with A and G containing sequences (33). The transition temperature (reported as the inflection point in a melting curve) of a 13 bp reverse-Hoogsteen hairpin containing 69% G's and 31% A's and no mismatches was 33 °C in Na^+ , but could not be measured in K^+ solution due to the preferential formation of a G-quadruplex. This mixed sequence has a similar composition to a Tel26 reverse Hoogsteen hairpin which would form an 11-mer with 60% G and 40% A.

We cannot rule out that some portion of the *trans,anti* T(A)=T(A) CPD is indeed coming from a basket type G-quadruplex structure in K^+ solution, as recent QM/MM calculations have concluded that there is a viable path to this product in the form 3 quadruplex conformation (11). There are two arguments, however, against the involvement of the form 3 quadruplex as the principal photoreactive intermediate. The first is that the rate of formation in Tel26, which has been shown by NMR to primarily adopt non-photoreactive type I and type II quadruplexes, was two times faster than for the NF3 sequence which has been shown to adopt the putative form 3 quadruplex conformation (10). The second is that all attempts to make mutations to further stabilize the form 3 structure in Tel26 by adding nucleotides to loop 1, with the exception of insertion of a G, reduced the yield of the *trans, anti* CPD photoproduct (10). The negative effects of such mutations could perhaps be better explained if the photoreactive structure was the reverse Hoogsteen hairpin, in which case adding nucleotides to loop 1 would disrupt the photocrosslinking site. The insertion of the G, however, would not have been expected to affect the structure as much, since the G-tract could slip and restore the photocrosslinking conformation.

At this point, we believe that we can explain the photoreactivity of human telomeric G-quadruplex forming sequences in Na^+ , K^+ and Li^+ solution as illustrated in Fig. 10. First, the lack of *anti* CPD formation in Na^+ solution may be explained by a greater thermodynamic

preference and kinetic stability for the basket form (Fig. 10B) than the reverse Hoogsteen hairpin form or Form 3 basket, and the inability of the T's in the two lateral loops to interact due to steric constraints. It is also possible that Na⁺ in the G-quartet facing the three nucleotide lateral loops coordinates with a thymine in one of the loops and holds it in such a way as to prevent *anti* CPD formation, as well as preventing strand slippage to make the form 3 basket structure. Though involving a different structure, coordination of thymine to the Na⁺ in a G quartet has been previously suggested to explain its orientation in the four nucleotide diagonal loops of the bimolecular d(G₄T₄G₄) Oxytricha quadruplex (34). In K⁺ solution, the DNA is more conformationally labile and the various hybrid and basket forms are in equilibrium with a small amount of weakly photoreactive Form 3 basket (Fig. 10C) and the more photoreactive reverse Hoogsteen hairpin (Fig. 10D). Indeed, a recent NMR study showed the presence of a non-quadruplex ensemble (referred to as unfolded) for d(TT(GGGTTA)₃GGGA) which equilibrates between hybrid 1 and hybrid 2 structures (35). The small relative amount of a photoreactive reverse Hoogsteen hairpin form of Tel22, however, would then be expected to increase by the addition of two additional A's to each end to form Tel26, which shows a higher yield and specificity for the *trans,anti* T(A)=T(A) CPD. The reverse Hoogsteen structure most favoring the *trans-anti* and potentially *cis, anti* T(A)=T(A) product in K⁺ appears to be the one with the same Watson to Hoogsteen orientation as in Tel21_3'alt and Tel21_3'con (Fig. 7), with the 3'-terminal G₃-tract functioning as the Hoogsteen acceptor. In Li⁺ solution, G-quadruplex formation is inhibited, thereby increasing the fraction of the photoreactive reverse Hoogsteen structure, which would explain why the rate of the *trans,anti* product of Tel26 is faster in Li⁺ than in K⁺.

Reverse Hoogsteen hairpins are well known and studied in the context of triplex DNA formed between two polypurine strands and a polypyrimidine strand and when formed in duplex DNA have been referred to as H* DNA (36, 37). One polypurine strand base pairs to the purine strand of a Watson-Crick duplex in an antiparallel orientation via reverse Hoogsteen base pairs with all nucleotides in the *anti* glycosyl conformation as established by NMR on an intramolecular triplex (38). Triplex formation by reverse Hoogsteen base pairing was first exploited for targeting polypurine tracts in duplex DNA by a third polypurine strand (39). It has also been employed for targeting poly-pyrimidine strands through the use of poly-purine reverse Hoogsteen hairpins and circular DNA (40-42). Given the prevalence of the reverse Hoogsteen base pair in triplexes, it would not be unexpected for a reverse Hoogsteen hairpin to form in human telomeric DNA, though the stability of such hairpins may be compromised by the mismatches resulting from the TTA loop sequence. It is still possible, however, that other, less well established base pairs are involved, such as C2-symmetric Pu•Pu base pairs (Fig. 4, S2) proposed for a parallel duplex of d(GA)_n (30) and an antiparallel hairpin of *Oxytricha* telomeric DNA (22).

In conclusion, it would appear that the photochemistry of G-quadruplex forming human telomeric sequences is rather complex, and may involve both Hoogsteen and reverse Hoogsteen hairpins, as well as G-quadruplex conformations that are highly dependent on the cation and DNA sequence. Surprisingly, the reverse Hoogsteen hairpin is rarely considered in experiments and simulations of G-quadruplex folding (43-45), but perhaps should be considered as a conformation that may precede, and may be present in equilibrium with the quadruplex structures. It would be the first structure expected to form when single strand

(GGGTTA)_n is released from B form duplex DNA which has all G's in *anti* glycosyl conformations, or when released from hPot1, which binds single strand telomeric DNA and holds three of four G's, in *anti* glycosyl conformations (46). Reverse Hoogsteen hairpins may thus play an otherwise unrecognized role in transactions involving human telomeric DNA, and possibly purine-rich regulatory sequences, *in vivo*, and the formation of *trans-anti* CPDs might be an indication of such intermediates. In addition, *anti* CPD formation may be a useful internal UV probe in addition to iodouracil (47) for following folding events in DNA *in vitro* or *in vivo*.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Dr. Blankenship for use of his CD instrument. Research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health under award number R01CA40463.

References

1. Ben-Hur E, Elad D, Ben-Ishai R. The photosensitized dimerization of thymidine in solution. *Biochim Biophys Acta*. 1967; 149:355–360. [PubMed: 6081513]
2. Douki T, Laporte G, Cadet J. Inter-strand photoproducts are produced in high yield within A-DNA exposed to UVC radiation. *Nucleic Acids Res*. 2003; 31:3134–3142. [PubMed: 12799441]
3. Su DG, Kao JL, Gross ML, Taylor JS. Structure determination of an interstrand-type cis-anti cyclobutane thymine dimer produced in high yield by UVB light in an oligodeoxynucleotide at acidic pH. *J Am Chem Soc*. 2008; 130:11328–11337. [PubMed: 18680367]
4. Su DG, Fang H, Gross ML, Taylor JS. Photocrosslinking of human telomeric G-quadruplex loops by anti cyclobutane thymine dimer formation. *Proc Natl Acad Sci USA*. 2009; 106:12861–12866. [PubMed: 19628696]
5. Wang Y, Patel DJ. Solution structure of the human telomeric repeat d[AG3(T2AG3)3] G-tetraplex. *Structure*. 1993; 1:263–282. [PubMed: 8081740]
6. Ambrus A, Chen D, Dai J, Bialis T, Jones RA, Yang D. Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Res*. 2006; 34:2723–2735. [PubMed: 16714449]
7. Dai J, Punchihewa C, Ambrus A, Chen D, Jones RA, Yang D. Structure of the intramolecular human telomeric G-quadruplex in potassium solution: a novel adenine triple formation. *Nucleic Acids Res*. 2007; 35:2440–2450. [PubMed: 17395643]
8. Dai J, Carver M, Punchihewa C, Jones RA, Yang D. 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*. 2007; 35:4927–4940. [PubMed: 17626043]
9. Lim KW, Amrane S, Bouaziz S, Xu W, Mu Y, Patel DJ, Luu KN, Phan AT. Structure of the human telomere in K⁺ solution: a stable basket-type G-quadruplex with only two G-tetrad layers. *J Am Chem Soc*. 2009; 131:4301–4309. [PubMed: 19271707]
10. Smith JE, Lu C, Taylor JS. Effect of sequence and metal ions on UVB-induced anti cyclobutane pyrimidine dimer formation in human telomeric DNA sequences. *Nucleic Acids Res*. 2014; 42:5007–5019. [PubMed: 24598261]
11. Lee W, Matsika S. Conformational and electronic effects on the formation of anti cyclobutane pyrimidine dimers in G-quadruplex structures. *Phys Chem Chem Phys*. 2017; 19:3325–3336. [PubMed: 28091673]
12. Fletcher TM, Sun D, Salazar M, Hurley LH. Effect of DNA secondary structure on human telomerase activity. *Biochemistry*. 1998; 37:5536–5541. [PubMed: 9548937]

13. Wang Y, Taylor JS, Gross ML. Nuclease P1 digestion combined with tandem mass spectrometry for the structure determination of DNA photoproducts. *Chem Res Toxicol*. 1999; 12:1077–1082. [PubMed: 10563833]
14. Hardin CC, Watson T, Corregan M, Bailey C. Cation-dependent transition between the quadruplex and Watson-Crick hairpin forms of d(CGCG3GCG). *Biochemistry*. 1992; 31:833–841. [PubMed: 1731941]
15. Sen D, Gilbert W. A sodium-potassium switch in the formation of four-stranded G4-DNA. *Nature*. 1990; 344:410–414. [PubMed: 2320109]
16. Williamson JR, Raghuraman MK, Cech TR. Monovalent cation-induced structure of telomeric DNA: the G-quartet model. *Cell*. 1989; 59:871–880. [PubMed: 2590943]
17. Gray DM, Wen JD, Gray CW, Repges R, Repges C, Raabe G, Fleischhauer J. Measured and calculated CD spectra of G-quartets stacked with the same or opposite polarities. *Chirality*. 2008; 20:431–440. [PubMed: 17853398]
18. Mergny JL, Lacroix L. UV melting of G-quadruplexes. *Curr Protoc Nucleic Acid Chem*. 2009; 37
19. Masiero S, Trotta R, Pieraccini S, De Tito S, Perone R, Randazzo A, Spada GP. A non-empirical chromophoric interpretation of CD spectra of DNA G-quadruplex structures. *Org Biomol Chem*. 2010; 8:2683–2692. [PubMed: 20440429]
20. Wlodarczyk A, Grzybowski P, Patkowski A, Dobek A. Effect of ions on the polymorphism, effective charge, and stability of human telomeric DNA. Photon correlation spectroscopy and circular dichroism studies. *J Phys Chem B*. 2005; 109:3594–3605. [PubMed: 16851398]
21. Choi KH, Choi BS. Formation of a hairpin structure by telomere 3' overhang. *Biochim Biophys Acta*. 1994; 1217:341–344. [PubMed: 8148384]
22. Ceru S, Sket P, Prislán I, Lah J, Plavec J. A new pathway of DNA G-quadruplex formation. *Angew Chem*. 2014; 53:4881–4884. [PubMed: 24644074]
23. Schultze P, Smith FW, Feigon J. Refined Solution Structure of the Dimeric Quadruplex Formed from the Oxytricha Telomeric Oligonucleotide d(GGGGTTTTGGGG). *Structure*. 1994; 2:221–233. [PubMed: 8069635]
24. Vorlickova M, Kejnovska I, Bednarova K, Rencuk D, Kypr J. Circular dichroism spectroscopy of DNA: from duplexes to quadruplexes. *Chirality*. 2012; 24:691–698. [PubMed: 22696273]
25. Risitano A, Fox KR. Inosine substitutions demonstrate that intramolecular DNA quadruplexes adopt different conformations in the presence of sodium and potassium. *Bioorg Med Chem Lett*. 2005; 15:2047–2050. [PubMed: 15808465]
26. Cannistraro VJ, Taylor JS. Acceleration of 5-methylcytosine deamination in cyclobutane dimers by G and its implications for UV-induced C-to-T mutation hotspots. *J Mol Biol*. 2009; 392:1145–1157. [PubMed: 19631218]
27. Pan Z, Hariharan M, Arkin JD, Jalilov AS, McCullagh M, Schatz GC, Lewis FD. Electron donor-acceptor interactions with flanking purines influence the efficiency of thymine photodimerization. *J Am Chem Soc*. 2011; 133:20793–20798. [PubMed: 22032333]
28. Coma S, Noe V, Eritja R, Ciudad CJ. Strand displacement of double-stranded DNA by triplex-forming antiparallel purine-hairpins. *Oligonucleotides*. 2005; 15:269–283. [PubMed: 16396621]
29. de Almagro MC, Coma S, Noe V, Ciudad CJ. Polypurine hairpins directed against the template strand of DNA knock down the expression of mammalian genes. *J Biol Chem*. 2009; 284:11579–11589. [PubMed: 19261618]
30. Rippe K, Fritsch V, Westhof E, Jovin TM. Alternating d(G-A) sequences form a parallel-stranded DNA homoduplex. *Embo J*. 1992; 11:3777–3786. [PubMed: 1396571]
31. Pedroso IM, Duarte LF, Yanez G, Burkewitz K, Fletcher TM. Sequence specificity of inter- and intramolecular G-quadruplex formation by human telomeric DNA. *Biopolymers*. 2007; 87:74–84. [PubMed: 17549693]
32. Suen IS, Rhodes JN, Christy M, McEwen B, Gray DM, Mitás M. Structural properties of Friedreich's ataxia d(GAA) repeats. *Biochim Biophys Acta*. 1999; 1444:14–24. [PubMed: 9931411]
33. Sole A, Delagoutte E, Ciudad CJ, Noe V, Alberti P. Polypurine reverse-Hoogsteen (PPRH) oligonucleotides can form triplexes with their target sequences even under conditions where they fold into G-quadruplexes. *Scientific Reports*. 2017; 7:39898. [PubMed: 28067256]

34. Schultze P, Hud NV, Smith FW, Feigon J. The effect of sodium, potassium and ammonium ions on the conformation of the dimeric quadruplex formed by the *Oxytricha nova* telomere repeat oligonucleotide d(G(4)T(4)G(4)). *Nucleic Acids Res.* 1999; 27:3018–3028. [PubMed: 10454595]
35. Bessi I, Jonker HR, Richter C, Schwalbe H. Involvement of Long-Lived Intermediate States in the Complex Folding Pathway of the Human Telomeric G-Quadruplex. *Angew Chem Int Ed.* 2015; 54:8444–8448.
36. Bernues J, Beltran R, Casasnovas JM, Azorin F. Structural polymorphism of homopurine--homopyrimidine sequences: the secondary DNA structure adopted by a d(GA.CT)₂₂ sequence in the presence of zinc ions. *Embo J.* 1989; 8:2087–2094. [PubMed: 2792082]
37. Kohwi Y, Kohwi-Shigematsu T. Magnesium ion-dependent triple-helix structure formed by homopurine-homopyrimidine sequences in supercoiled plasmid DNA. *Proc Natl Acad Sci U S A.* 1988; 85:3781–3785. [PubMed: 3375241]
38. Radhakrishnan I, de los Santos C, Patel DJ. Nuclear magnetic resonance structural studies of intramolecular purine.purine.pyrimidine DNA triplexes in solution. Base triple pairing alignments and strand direction. *J Mol Biol.* 1991; 221:1403–1418. [PubMed: 1942059]
39. Beal PA, Dervan PB. Second structural motif for recognition of DNA by oligonucleotide-directed triple-helix formation. *Science.* 1991; 251:1360–1363. [PubMed: 2003222]
40. Samadashwily GM, Dayn A, Mirkin SM. Suicidal nucleotide sequences for DNA polymerization. *Embo J.* 1993; 12:4975–4983. [PubMed: 8262040]
41. Wang S, Kool ET. Recognition of Single-Stranded Nucleic Acids by Triplex Formation: The Binding of Pyrimidine-Rich Sequences. *J Am Chem Soc.* 1994; 116:8857–8858. [PubMed: 20882109]
42. Ciudad CJ, Rodriguez L, Villalobos X, Felix AJ, Noe V. Polypurine Reverse Hoogsteen Hairpins as a Gene Silencing Tool for Cancer. *Curr Med Chem.* 2017
43. Mashimo T, Yagi H, Sannohe Y, Rajendran A, Sugiyama H. Folding pathways of human telomeric type-1 and type-2 G-quadruplex structures. *J Am Chem Soc.* 2010; 132:14910–14918. [PubMed: 20882978]
44. Gray RD, Trent JO, Chaires JB. Folding and unfolding pathways of the human telomeric G-quadruplex. *J Mol Biol.* 2014; 426:1629–1650. [PubMed: 24487181]
45. Stadlbauer P, Kührová P, Banáš P, Kořál J, Bussi G, Trantírek L, Otyepka M, Šponer J. Hairpins participating in folding of human telomeric sequence quadruplexes studied by standard and T-REMD simulations. *Nucleic Acids Res.* 2015; 43:9626–9644. [PubMed: 26433223]
46. Lei M, Podell ER, Cech TR. Structure of human POT1 bound to telomeric single-stranded DNA provides a model for chromosome end-protection. *Nat Struct Mol Biol.* 2004; 11:1223–1229. [PubMed: 15558049]
47. Xu Y, Tashiro R, Sugiyama H. Photochemical determination of different DNA structures. *Nat Protoc.* 2007; 2:78–87. [PubMed: 17401341]

Figure 1.

Structures and photoproducts of G-quadruplex forming sequences. A) Structures of the *cis,syn* and *trans,anti* CPDs of TT and their nuclease P1 degradation products. B) Known structures of human telomeric G-quadruplex forming sequences in Na^+ and K^+ solution. The nucleotide numbering system is for the Tel26 sequence, and the numbers 1-3 refer to the TTA loop numbers as shown below. C) Sequences referred to or used in this study written 5'→3'.

Figure 2.
HPLC traces of NP1-digested UVB irradiated Tel26 in the presence of Li⁺, Na⁺ and K⁺ ions. Tel26 was irradiated with UVB light on ice and then degraded overnight with NP1 at 37°C.

Figure 3.

CD spectra and G-quadruplex UV melting temperature curves. A) CD spectra of Tel26 in the presence of 100 mM of the indicated metal ion at 4 °C. B) Temperature dependent absorption curves at 295 nm for Tel26 in the presence of the indicated concentration of metal ion. The curves were normalized to the 5 °C value.

Figure 4.

Base pairing possibilities in hairpin structures. Top panel: Possible base pairing arrangements for hairpin forms of Tel22 and Tel26 with symbols: •, Hoogsteen, reverse Hoogsteen, or symmetric; ◆, reverse Hoogsteen or symmetric; ↑, allowed Watson face to Hoogsteen face base pairing. Tel26 is related to Tel22 by the addition of the underlined A's. Lower panels, the different base pairing possibilities for G•G, G•I, A•T and A•A showing the pairs of glycosyl conformations in bold and dashed lines required to have an antiparallel helix. Only reverse Hoogsteen pairing allows for all *anti* glycosyl conformations. Also shown is the reverse Hoogsteen conformation of the T•T mispair site that results in the closest approach of the circled bonds required for the formation of the *trans,anti* CPD stereochemistry which cannot be achieved by a Hoogsteen conformation. Distances are approximate distances between the C1' anomeric carbons of the nucleotides. Also shown is an estimated distance for forming a *trans,anti* TT CPD.

Figure 5.

Temperature dependence and kinetics of *trans,anti* CPD formation in Tel26. A) Temperature dependence of the relative fraction of *trans,anti* T(A)=T(A) CPD formed in Li⁺ solution vs fraction helix form. Kinetics of B) *trans,anti* T(A)=T(A) CPD formation and C) *cis,syn* T=T(A) formation in Li⁺ and K⁺ solution.

Figure 6.

Photoproduct formation in 15-mers with Li⁺, Na⁺, and K⁺. HPLC traces of NP1-digested UVB irradiated Tel15 and the Watson Crick HP15WC hairpin in 100 mM metal ion. Possible hairpin conformations for the 15-mers where base pairing is denoted by: •, Hoogsteen, reverse Hoogsteen, or symmetric; ↑, allowed Watson to Hoogsteen face base pairing; and ↓, Watson Crick. Because both Tel15 and HP15WC have a 5'-terminal TTA sequence lacking a 5'-phosphate, T=TA will arise from the 5'-terminal TTA, and pT=TA will arise from the internal TTA site following NP1 digestion.

Figure 7. Photoproduct formation in ionsine- and Watson Crick base pair containing 21-mers with Na⁺ and K⁺. HPLC traces of NP1-digested UVB irradiated 21-mers in 100 mM Na⁺, and 100 mM K⁺. Possible hairpin conformations for the 21-mers where base pairing is denoted by: •, Hoogsteen, reverse Hoogsteen, or symmetric; ↑, allowed Watson to Hoogsteen face base pairing, and |, Watson Crick. The asterisk indicates a possible mixture of *cis,anti* and *trans,anti* isomers, see Fig. S5.

Figure 8.

Photoproduct formation in non-telomeric 21-mers with K⁺ and Na⁺. HPLC traces of NP1-digested UVB irradiated HP21GAA, HP21GAG and HP21AAG in 100 mM K⁺ or Na⁺ solution. Possible hairpin conformations for the 21-mers where base pairing is denoted by: •, Hoogsteen, reverse Hoogsteen, or symmetric; ♦, reverse Hoogsteen or symmetric; and ↑, allowed Watson face to Hoogsteen face base pairing

Figure 9.

Gel electrophoresis of non-telomeric and telomeric sequences. The sequences shown at the bottom were electrophoresed at 400 V in a cold room at 4°C on a 15% native polyacrylamide gel with Tris-borate EDTA buffer in the absence (A) or presence (B-D) of 100 mM of the indicated metal ion. The bands in the dotted boxes correspond to Tel26 and Tel24.

Figure 10.

Proposed explanation for the variation in *anti* CPD formation as a function of sequence and metal ion. A) General scheme showing the effect of sequence and cation on the DNA conformations present for Tel22 and Tel26, in which the additional underlined A's on Tel26 serve to further stabilize the reverse Hoogsteen hairpin base pairing. Based on the inosine and adenine substrates, the preferred pairing is denoted by the symbol $\uparrow$ for Watson face to Hoogsteen face reverse Hoogsteen base pairing, and $\blacklozenge$, for reverse Hoogsteen or symmetric base pairing. B) Na^+ is proposed to stabilize the basket conformation and prevent the T's from photoreacting, because of the short lateral loops, and/or by chelation of the thymine. C) Intercalation of K^+ and the larger loop size of the form 3 quadruplex as a possible explanation for a lack of an inhibitory effect of K^+ on *anti* CPD formation. D) How a hairpin containing reverse Hoogsteen sections flanking a TT mismatch could facilitate *trans,anti* CPD formation, where the reverse Hoogsteen T•A base pairs are shorter than the G•G pairs, leading to a pinching effect that would bring the T T's close together.