Published in final edited form as:

Photochem Photobiol. 2019 January ; 95(1): 252–266. doi:10.1111/php.12991.

Photocrosslinking of G-Quadruplex Forming Sequences found in Human Promoters†

Jillian E. Smith-Carpenter and John-Stephen Taylor*

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

Abstract

While it is well known that human telomeric DNA sequences can adopt G-quadruplex structures, some promoter sequences have also been found to form G-quadruplexes, and over 40% of promoters contain putative G-quadruplex forming sequences. Because UV light has been shown to crosslink human telomeric G-quadruplexes by cyclobutane pyrimidine dimer (CPD) formation between T's on adjacent loops, UV light might also be able to photocrosslink G-quadruplexes in promoters. To investigate this possibility, fifteen potentially UV-crosslinkable G-quadruplex forming sequences found in a search of human DNA promoters, were UVB irradiated *in vitro*, and three were confirmed to have formed non-adjacent CPDs by mass spectrometry. In addition to non-adjacent T=T CPDs found in human telomeric DNA, a non-adjacent T=U CPD was discovered that presumably arose from deamination of a non-adjacent T=C CPD. Analysis of the three sequences by circular dichroism, melting temperature analysis, and chemical footprinting confirmed the presence of G-quadruplexes that could explain the formation of the non-adjacent CPDs. The formation of non-adjacent CPDs from the sequences *in vitro*, suggests that they might be useful probes for the presence of non-B DNA structures, such as G-quadruplexes, *in vivo*, and if they were to form *in vivo*, might also have significant biological consequences.

Graphical Abstract Figure

UVB irradiation of 15 potentially photocrosslinkable G-quadruplex forming sequences from a search of human promoter sequences resulted in three sequences containing non-adjacent cyclobutane pyrimidine dimers identified by combined enzyme-coupled mass spectrometric analysis. This suggests that non-adjacent photoproducts could be used as probes for formation of non-B DNA structures, such as G-quadruplexes *in vivo*, and if formed *in vivo* may contribute to the toxicity and mutagenicity of sunlight.

†This article is part of a Special Issue celebrating Photochemistry and Photobiology's 55th Anniversary.

*Corresponding author: taylor@wustl.edu.

INTRODUCTION

Non-canonical DNA structures, such as hairpins, G-quadruplexes and cruciforms, have been intensely studied over the last several decades (1) (2). Human telomeric DNA, (TTAGGG)_n forms a variety of different G-quadruplex structures in vitro (3–6). The structure of G-quadruplexes is influenced by loop length, 5' and 3' overhangs and coordinating cations (7–9). Telomeric G-quadruplexes structures can adopt antiparallel basket (3, 10), antiparallel chair (11), hybrid (3+1) (4, 12) and parallel (13) structures under physiologically relevant conditions (Fig. 1). G-quadruplex forming sequences are not confined to telomeric DNA and have been found in G-rich sequences of promoters (14, 15). Over 40% of gene promoters contain putative G-quadruplex forming sequences (15), including c-Myc (16), bcl-2 (17), KRAS (18), VEGF (19), and c-Kit (20). The potential for G-quadruplex formation in promoters has suggested that they may serve a regulatory role during transcription. Recently, researchers have used a fluorescently tagged antibody against a variety of G-quadruplex structures to show that about 80% of the binding sites were outside of telomeres (21). They also observed a 5-fold increase in antibody binding in going from G0/G1 to S phase indicating replication dependent formation of the G-quadruplex structures.

One of the often-overlooked aspects of G-quadruplexes is their unique photochemistry which may have important biological consequences if G-quadruplexes were to form *in vivo*, and which could also be used to confirm the presence of such structures. While *cis,syn* cyclobutane pyrimidine dimers (CPDs) that form between adjacent bases in duplex DNA are the major photoproducts produced by UVB light (22), non-adjacent CPDs (Fig. 2) formed between non-contiguous nucleotides have been shown to form between adjacent loops in human telomeric G quadruplex forming DNA in the presence of potassium ion but not sodium ion (23). The major product is a *trans,anti* cyclobutane thymine dimer whose yield depends on the DNA sequence and metal cations present, and was proposed to arise from a basket-type structure (24). It is also possible, however, that a long hairpin intermediate, that has been proposed to be on the folding pathway to chair structures (25), is involved. In the latter regard, non-adjacent *trans,anti* CPDs were also found to form in the presence of lithium ion, which does not stabilize G quadruplexes, and in non-G quadruplex forming sequences, and was proposed to involve the intermediacy of a reverse Hoogsteen hairpin

(26). Nothing is known, however, about non-adjacent CPD formation in other G-quadruplex forming sequences, such as those that have, or might be found in promoters.

In this paper, non-adjacent CPD formation in three putative G-quadruplex forming sequences (PQS) in promoters is investigated. By searching six human chromosomes in a G quadruplex-forming sequence database (27, 28), nearly 300 potential G-quadruplex forming sequences were found upstream of the transcription start sites (TSS) with a three G stem, and with loops 1 and 3 containing between 3 and 7 nucleotides and at least one thymine. Of these sequences, 15 were UVB irradiated, and three of these were found to form readily identifiable non-adjacent CPDs. These three sequences were further shown to form G-quadruplexes by chemical footprinting, circular dichroism, and melting temperature analysis. Formation of non-adjacent CPDs in these G quadruplex-forming sequences by UVB light suggests that such photoproducts might also form *in vivo* and if they did, could substantiate the presence of non-B DNA structures, like G quadruplexes, *in vivo*, as well as have significant biological consequences.

MATERIALS AND METHODS

Materials.

Oligodeoxynucleotides were obtained from Integrated DNA Technologies, Inc. (IDT) (Coralville, Iowa) and Nuclease P1 (NP1) from *Penicillium citrinum* was from Sigma (St. Louis, MO). Ammonium citrate and 3-hydroxypicolinic acid (3-HPA) MALDI matrices were purchased from Fluka (Milwaukee, WI). HPLC solvents were from Fisher (Fair Lawn, NJ).

Search for potential photocrosslinkable G-quadruplex-forming promoter sequences.

G-rich motifs upstream of transcription start sites in human chromosomes 1, 2, 3, 6, 7, and 9, were searched for using the EuQuad G-quadruplex database (quadbase.igib.res.in/EuQuad) (27, 28). The search was restricted to a G-stem of three nucleotides and a loop length ranging between 1 and 7 nucleotides, with the general form $G_3 - \text{Loop } 1_{(1-7)} - G_3 - \text{Loop } 2_{(1-7)} - G_3 - \text{Loop } 3_{(1-7)} - G_3$. The query set was imported into excel and sequences were filtered for those that a) resided within 500 bases upstream from a transcription start site (TSS), b) had an overall length from 22 to 30 nucleotides, and c) had loops 1 and 3 that contained three to seven nucleotides and at least one thymine.

Preparation of the G-Quadruplexes.

The ODNs (50 μM) were heated at 95 °C for 10 min in 10 mM Tris-HCl, pH 7.5, with 150 mM NaCl or KCl, and then rapidly cooled down in ice. Samples were left on ice for at least 2 h.

UVB irradiation and analysis.

UVB irradiation (270–400 nm with peak intensity at 312 nm) was carried out with two Spectroline XX-15B UV 15-W tubes (rated UV intensity of 1.15 mW/cm² at 25 cm) filtered by a LONGLIFE filter (Spectronics Corporation) immediately after sample preparation. G-quadruplex samples were contained 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. One μL of 1 U/ μL aqueous nuclease P1 from *Penicillium citrinum* (Sigma) and 1 μL 10 mM ZnCl_2 were added to each 100 μL of 50 μM UVB irradiation sample and digested at 37 °C for >36 h. Reverse-phase HPLC was carried out with an X-Bridge column (C18, 4.6×75 mm, 2.5 μm , 135 Å; Waters Corporation) on a System Gold BioEssential HPLC with a Model 125 binary gradient pump and a Model 168 diode array detector (Beckman Coulter). NP1 digestion products were separated with a 1 mL/min gradient of 100% A (50 mM triethylammonium acetate, pH 7.5) for 3 min and 0–20% B (50% acetonitrile in 50 mM triethylammonium acetate, pH 7.5) for 3–53 min with detection at 260 nm. HPLC peaks were collected and dried *in vacuo*. The ODNs were resuspended in 5 μL of Milli-Q water. MALDI mass spectra were collected on a Voyager-DE STR tandem time-of-flight mass spectrometer (PerSeptive Biosystems). MALDI spots (1 μL were composed of 50/50 (v/v) of the ODN solution and 3-HPA matrix solution. Spectra were taken in reflector positive ion mode and are the sum of 100 accumulations. ESI-MS/MS spectra were obtained in the negative-ion mode with a Thermo Finnegan Orbitrap XL mass spectrometer. A solution of 50/50 acetonitrile/water with 0.05 % ammonium hydroxide was used as the spray solvent. The spray voltage was 3.7 V. The capillary voltage and temperature were –42 V and 225 °C, respectively. MS/MS experiments employed CAD with helium as the collision gas.

Circular Dichroism (CD).

CD spectra from 220 to 320 nm were obtained in a 1 mm path quartz cell under nitrogen at 4 °C on a J-810 spectropolarimeter (Jasco) using 100 μL of 50 μM ODN that had been kept for over a week at 4 °C in 10 mM Tris-HCl buffer, pH 7.5, with 150 mM of either NaCl or KCl. Spectra are an average of three scans taken with a band width of 1 nm, response time of 1 s, data pitch of 0.2 nm and a scan speed of 50 nm/min. A blank sample of buffer alone was used and subtracted from the G-quadruplex spectra.

Melting temperature studies.

UV melting curves were obtained on a Cary 100 Bio UV-Vis Spectrometer (Varian). Typically, 100 μL ODN samples in 10 mM Tris, 150 mM KCl, pH 7.5 buffer were preheated in a cuvette to 95 °C for 10 min and any air bubbles that formed were removed by tapping the cuvette. The absorbance was recorded at 295 nm and all samples were zeroed with buffer only. The temperature was cycled from 20 to 95 °C and back twice (heat-cool-heat-cool) with a rate of 1 °C per minute. The T_m 's were determined from the average of the minima of the first derivative curves of the 1st cooling curve and the 2nd heating curve.

Dimethyl sulfate protection assay.

The 5' termini of the oligodeoxynucleotides were labeled by a forward reaction in 50 mM Tris•HCl, pH 7.6, 10 mM MgCl_2 , 5 mM DTT, 0.1 mM spermidine, 0.1 mM EDTA with 10 U of T4 polynucleotide kinase and 20 pmol [γ - ^{32}P]-ATP at 37 °C for one h. Unreacted ATP was removed with a G-25 spin column. One μL aliquots of the labeled ODNs were added to the appropriate buffer for a final concentration of approximately 1 μM . G-quadruplexes were formed by heating to 95°C for 5 min and cooling on ice for 2 h. Dimethylsulfate (DMS) (0.75 μL) was added to each sample and allowed to react for varying times on ice. The reaction was quenched by the addition of 50 μL of stop solution (25 μL 3M NaOAc, 17 μL

dH₂O, 3.5 μ L 2-mercaptoethanol, 4.5 μ L of 10 mg/mL salmon sperm DNA), followed by recovery of the modified ODN by ethanol precipitation. To the recovered pellet, 100 μ L of 1 M piperidine was added and placed in 90 °C water bath for 30 min to induce strand cleavage. Samples were evaporated under vacuum, after which 50 μ L of formamide was added and the samples were analyzed on 20% denaturing PAGE gel.

RESULTS AND DISCUSSION

Selection of the putative G-quadruplex forming promoter sequences.

A number of G-quadruplex forming sequences in promoter regions have been characterized that form well-defined and characterized G-quadruplexes (29, 30). Some of these previously characterized G-quadruplex forming sequences have thymine or cytosine in each of loops 1 and 3 that would be necessary for CPD photocrosslinking in chair or basket-type structures. At least one of these pyrimidines, however, was either in a very short loop (<3 nucleotides), and/or were in sequences that were shown to form parallel or hybrid type structures that would not be expected to photocrosslink. To find potentially photocrosslinkable G-quadruplex forming sequences with possible relevance to transcription we therefore searched the EuQuad database for 22–30 nt sequences within 500 bases upstream of a transcription start site (TSS). The search was restricted to a program option of searching a three G long stem that had loops from 1 to 7 nucleotides. The results were further restricted to sequences of lengths 22–30 that had at least one thymine, which we knew from previous work could photocrosslink, in each of loops 1 and 3, that had lengths of 3 or more nucleotides that we expected would be necessary to get efficient crosslinking. From a combined set of over 300 sequences from well annotated genes from chromosomes 1, 2, 3, 6, 7, and 9 that met these criteria, we randomly selected 15 (2–3/chromosome) for further analysis of non-adjacent photoproduct formation (Table S1).

Photoproduct Analysis.

The 15 potentially photocrosslinkable G-quadruplex forming sequences were then analyzed for UVB-induced non-adjacent photoproducts using an NP1 enzyme digest-HPLC assay. Nuclease P1 degrades adjacent CPDs to trinucleotides of the form pPy=Py(N) (31), and non-adjacent CPDs to tetranucleotides of the form pPy(N)=pPy(N), where the nucleotide in parentheses is attached to the 3'-side of the pyrimidine involved in CPD formation (32) (Fig. 2). Photodimer containing tetranucleotides are readily distinguished from tetranucleotides resulting from partial digestion of undamaged DNA by having an 18 unit higher mass due to the presence of a second 5'-phosphate in place of a phosphodiester linkage. G-quadruplexes were formed by heating 50 μ M samples of DNA in either 150 mM NaCl or 150 mM KCl in 10 mM Tris buffer, pH 7.5 to 95°C for 5 min. Samples were then quickly cooled on ice and left to equilibrate for at least two hours before irradiation with UVB light centered at 312 nm for 2.5 h on ice.

Nuclease P1 was then incubated with the irradiated samples for at least 36 h before HPLC analysis. Using the HPLC method described in the experimental section, mononucleotides elute in the first 20 min, pPy=Py(N) trinucleotides formed from adjacent pyrimidines elute between approximately 20 and 30 min, and pPy(N)=pPy(N) tetranucleotides formed from

non-adjacent pyrimidines elute between approximately 30 and 36 minutes. Most of the 15 sequences showed some evidence of non-adjacent CPD formation by the appearance of peaks eluting between 30 and 36 minutes (Fig. 3, S1 & S2). Many of the peaks, however, were of low intensity and/or were present with equally sizable peaks above 36 min (Fig. S1A & C, S2A, B, C, & H) suggestive of incomplete digestion, and were not further pursued. In some cases peaks did not have m/z values in MALDI corresponding to the expected NP1 digestion products of either the photodamaged or undamaged DNA sequence (Fig. S1A & B, S2E). Of the fifteen sequences, CALU, STARD3NL and SUS3D (Table 1) showed evidence of non-adjacent CPD formation by MALDI mass spectrometry in both Na^+ and K^+ solution (Fig. 3 and Fig. S1A & C & S2E). The photoproducts of CALU and SUS3D in Na^+ were chosen for more detailed study by MS/MS mass spectrometry because non-adjacent photoproducts had not been previously seen to form in human telomeric DNA sequences the presence of Na^+ (23, 24, 26). STARD3NL in K^+ was also chosen for more detailed MS/MS study because one of the non-adjacent photoproducts appeared to involve C, which had also not been detected in previous studies of G-quadruplexes.

The NP1 digestion mixtures of UVB irradiated CALU, SUS3D, and STARD3NL in both Na^+ and K^+ were complex, presumably due to photoproduct formation with different stereochemistries between the different pyrimidines in putative loop 1 and putative loop 3 (Table 1). Unfortunately, we were unable to isolate and then identify all of these products in sufficient purity for analysis. We were, however able to isolate and identify some of the major products from the irradiation of CALU and SUS3D in the presence of Na^+ ion, and STARD3NL in the presence of K^+ ion by way of MALDI and ESI-MS/MS (Fig. 3). Non-adjacent photoproducts were also detected by MALDI for CALU and SUS3D in K^+ solution, and for STARD3NL in Na^+ solution, (Fig. S1) but were not further characterized by ESI MS/Ms.

Irradiation of the CALU sequence in the presence of Na^+ ion with UVB light resulted in a single major product eluting in the tetranucleotide region of the HPLC chromatogram after Nuclease P1 digestion, that formed in about 1% yield (Fig. 3A). Analysis by MALDI identified the photoproduct as pT(G)=pT(G) by an m/z of 1302.5 corresponding to the $[\text{M}+\text{H}]^+$ (Fig. S3). The structure of the photoproduct was further confirmed by MS/MS analysis of the $[\text{M}-2\text{H}]^{2-}$ ion at m/z 650.3 in electrospray ionization mass spectrometry (ESI) (Fig. 4). The presence of dG was confirmed by the product ion $[\text{pdG-H}]^-$ at m/z 346.17 resulting from loss of deoxyguanosine monophosphate, and $[\text{M-Gua}-2\text{H}]^{2-}$ at m/z 574.83 corresponding to the loss of guanine. The presence of pT(G)=pT(G) was also detected on the basis of MALDI when the CALU sequence was irradiated in K^+ solution (Fig. S1A).

To determine which two of the three possible thymines of CALU were involved in the photoreaction, two of the thymines were separately replaced with uracil. Since the photocycloaddition between two thymines and a thymine and uracil occurs equally well, thymines that participate in the photocrosslinking can be identified by a change in the retention time of the tetranucleotide product, and a change in mass. When uracil replaced the thymine in putative loop 2 of the G-quadruplex (CALU- U^2), the retention time of the major peak was the same as for the parent CALU sequence. However, when uracil was positioned in loop 3 (CALU- U^3), the major peak eluted more slowly, as verified by co-injecting the two

products (Fig. 5). MALDI analysis of the products confirmed the identity of the photoproduct from CALU-U² as pT(G)=pT(G), and from CALU-U³ as pT(G)=pU(G).

Irradiation of the STARD3NL sequence in the presence of K⁺ with UVB light resulted in three distinct peaks in the tetranucleotide region of the HPLC chromatogram after Nuclease P1 digestion (Fig. 3B). Two of the peaks eluted very close to each other in about 6% yield, and could not be resolved. The molecular ion in both MALDI and ESI-MS corresponded only to pT(G)=pT(G) (Fig. S4A & B). The ESI-MS/MS product ion spectrum had many of the same diagnostic fragment ions observed for the pT(G)=pT(G) product from the CALU sequence, but with different relative abundances which could be the result of the presence of two stereoisomers of pT(G)=pT(G) (Fig. S4C and Table S2). The second photoproduct-containing tetranucleotide peak produced in about 4.5% yield showed the presence of three different ions in MALDI with *m/z* of 1289.0, 1402.2 and 1544.0 (Fig. S5A). The assignment of the ion at *m/z* 1289.0 to the [M+H]⁺ ion of pT(G)=pU(G) was confirmed by negative ion ESI-MS/MS carried out on the *m/z* ion at 643.1 corresponding to [M-2H]²⁻ (Fig S5B) that resulted in product ions [pdG-H]⁻ at *m/z* 346.1 and [M-Gua-2H]²⁻ at *m/z* 567.8 (Fig. 6 and Table S3). The uracil-containing product pT(G)=pU(G) must have arisen from deamination of the C in pT(G)=pC(G) to uracil. The other two ions observed in the MALDI cannot be readily assigned to any NP1 products expected for this DNA sequence, irradiated or non-irradiated. In K⁺ solution, only the pT(G)=pT(G) non-adjacent dimer was detected by MALDI, along with the adjacent pU=TG product (Fig. S1B).

Irradiation of SUS3 in Na⁺ ion solution resulted in several distinct NP1 peaks in the tri- and tetranucleotide region of the HPLC chromatogram (Fig. 3C). An early eluting peak could be readily identified as the adjacent dimer pT=T(G) by an *m/z* ion at 956.2 in the MALDI spectrum corresponding to the [M+H]⁺ ion, and comes from the only dipyrimidine site in the sequence (Fig. S6). We were also able to identify two non-adjacent photoproduct-containing products as pT(G)=pT(G) (6% yield), as previously described above (Fig. S7), and pT(G)=pT(T) (4% yield) by its [M+H]⁺ ion at *m/z* 637.6 in the MALDI spectrum (Fig. S8). The structure of pT(G)=T(T) was confirmed by product ions in the ESI MS/MS spectrum for [pdT-H]⁻ and [pdG-H]⁻ at *m/z* 321 and 346.25, respectively, and by the [M-Thy-2H]²⁻ and [M-Gua-2H]²⁻ product ions at *m/z* 574.9 and 562.25, respectively (Fig. 7 and Table S4). In K⁺ solution, three non-adjacent pT(G)=pT(G) products were detected on the basis of MALDI, along with two adjacent pT=TG photoproducts (Fig. S2).

Characterization of the G-Quadruplex forming promoter sequences.

The CALU, STARD3NL and SUS3 sequences were further characterized by melting temperature analysis (Fig. S9B, F & G) and circular dichroism spectroscopy (Fig. 8) in both K⁺ and Na⁺ containing buffers. All three sequences showed reversible melting behavior at 295 nm in the presence of either Na⁺ or K⁺ ions that is characteristic of G-quadruplexes (Fig. S9). The melting temperatures were all higher in the presence of K⁺ ions, which is consistent with the general observation that K⁺ stabilizes quadruplexes more effectively than Na⁺ (9). CD spectra of the other 12 sequences were also obtained for comparison purposes (Fig. S10), along with melting temperature curves for a subset of these sequences (Fig. S9).

The CD spectra of the CALU, STARD3NL and SUS3D sequences were quite similar in K^+ buffer (Fig 8A). There was a peak maximum at 260 nm and a minimum at 240 nm, which is indicative of a parallel structure (33). The presence of a peak at 295 nm, however, is evidence for the presence of conformations that contain antiparallel strands (33). This could either mean the presence of basket or chair forms, and/or the presence of hybrid structures. In Na^+ buffers, the major peak for CALU and SUS3D was at 295 nm, characteristic of basket or chair structures, whereas for STARD3NL the 295 nm peak was of about the same intensity as the 260 nm peak suggesting the possible presence of both parallel and antiparallel quadruplexes and/or hybrid structures (Fig. 8B). For the twelve other sequences, two showed only evidence of a parallel quadruplex (Fig. S10C, H) while others showed an additional contribution from an antiparallel form (Fig. S10E, F, G, I, J, K and L) that could possibly favor non-adjacent dimer formation. In Na^+ only one sequence showed a strong contribution from an only a parallel conformation (Fig. S10B), while others showed a mix of parallel and antiparallel conformations (Fig. S10D, E, F, G, H, I, J, K, L). There is insufficient information to conclude whether or not there is any correlation between the CD spectra and ability to form non-adjacent CPDs.

Dimethyl sulfate protection assay.

The CALU, STARD3NL and SUS3D G-quadruplex forming sequences contained four runs of three or more contiguous guanines that were expected to form a three G-quartet stem, and isolated guanines in what were expected to be loops. A dimethyl sulfate (DMS) protection assay was therefore conducted to establish which of the guanines were part of G-quartets. G quartets are formed by Hoogsteen hydrogen bonding to N7 which protects it from alkylation by dimethyl sulfate (34), while the N7 of a guanine in a loop, or in an unstable G-quartet, is exposed and expected to react more rapidly with dimethyl sulfate. Upon heating the DMS-treated sequences with piperidine, positions at which guanines are methylated at N7 will result in strand cleavage. For each putative G-quadruplex-forming sequence, an aliquot of 5'-end-labeled ODN was incubated with dimethyl sulfate for various times on ice. The reaction was quenched with a solution containing 2-mercaptoethanol and salmon sperm DNA, and the DNA recovered by ethanol precipitation. After treatment at 90°C in 1M piperidine for 30 min, the solvent was evaporated *in vacuo* and the residue dissolved in formamide. The strand cleavage sites were analyzed on a 20% denaturing PAGE gel (Fig. 9). When the dimethyl sulfate reaction was conducted in water in the absence of salt, the G quadruplexes do not form, and all G's of the three putative G-quadruplex-forming sequences were observed to react almost equally. In the presence of metal ions, however, the reactivity varied with position, and was more greatly inhibited at G's predicted to be involved with G quartet formation. The percentage inhibition of the 8 minute methylation reaction in the presence of the metal ions relative that in water only is plotted in Fig. 10.

The CALU sequence produced very similar DMS cleavage in the presence of both K^+ and Na^+ ions, though the overall inhibition was greater in presence of K^+ ion (Fig. 10A). The local maxima of inhibition at G2/3, G9, and G15 correlate with the center of the G-tracts that were expected to be involved in G-quartet formation. In contrast, G7 and G13 are the least protected from methylation, which would place these G's on average in loops rather than in the G-quartets. The CALU sequence has three G-tracts that contain four guanines

and one that only has three (Table 1), which could allow for the formation of multiple equilibrating G-quadruplex structures. In this regard, G20, which is in the center of the fourth G-tract appears to be unprotected, and could indicate flexibility of the 5'-end of the G-quartet and/or the presence of a triplex in which this strand is exposed.

The methylation patterns for the STARD3NL sequence were also quite similar in K^+ and Na^+ solution with some minor differences, and showed overall greater inhibition than was observed for the CALU sequence. The periodic protection pattern is again consistent with G-quartet formation with local inhibition maxima at G3, G10&11, and G17/18 coinciding with the center of the G-tracts, although G21, which was expected to be in a loop, also showed significant protection. Local minima in inhibition were observed at G7, G14 and G20 in K^+ solution, which are consistent with the expectation that they would be in loops. While it was expected that all three guanines of the two internal G_3 -tracts would be involved in G-quartet formation, the similar level of protection in Na^+ of G12 and G16 with G14 that is assigned to loop 2 suggests that these two G quartet guanines are often exposed, while in K^+ , G12 appears to be protected. This sequence could also fold into parallel quadruplexes which would explain the presence and relative sizes of peaks characteristic of both parallel and antiparallel conformations in the CD spectra depending on the cation present.

The DMS protection assay for the SUS3D sequence was, like the other sequences, also rather similar in both Na^+ and K^+ solution, but rather different from the CALU and STARD3NL sequences. Instead of showing a highly periodic protection pattern, the G's showed more inhibition towards the 5'-end that tapered off towards the 3'-end. A weak pattern could be detected that was more pronounced in K^+ than Na^+ solution, that had local minima at G6, G10 and G17 for K^+ , suggesting that these G's resided in loops or were otherwise accessible. In Na^+ , weak local minima were observed at G13 and G17 that were more pronounced at 1 minute (data not shown). It is likely that this sequence is rather dynamic and adopts a variety of conformations.

Possible photoreactive conformations of the G-quadruplex forming promoter sequences.

The photoproducts observed to form in the three promoter sequences are consistent with the presence of a basket and/or chair-type G-quadruplex structures which place the two pyrimidines in adjacent lateral loops of antiparallel quadruplexes as shown in Fig. 1. As indicated in the introduction, only folded antiparallel quadruplexes have two adjacent lateral loops that could position T's in such a way as they can form non-adjacent photoproducts (Fig. 1). The CD, melting temperature, and methylation protection patterns are consistent with the presence of basket and/or chair G-quadruplex conformations, though the CD indicates a significant contribution of parallel quadruplexes in K^+ solution that would be predicted not to be photoreactive. It is also possible that the non-adjacent CPDs are arising in part from a reverse Hoogsteen type hairpin (25).

For the CALU sequence, the formation of a non-adjacent dimer between T6 and T18 in Na^+ solution that was verified by the T/U-replacement experiment, is consistent with both a chair or basket as the photoreactive conformation. Though not verified by a U-replacement a pT(G)=pT(G) product was also observed by MALDI in K^+ and may be coming from the same conformation. For SUS3D, there are no T's in loop 2, so the pT(G)=T(T) photoproduct

in Na⁺ can only arise from CPD formation between T5 & T19 in either the chair or basket form. The pT(G)=pT(G) photoproduct observed in Na⁺ and K⁺, however, could arise from CPD formation between T5 in loop 1 and either or both T16 & T19 in loop 3 of the chair or basket conformations. It is also possible that the pT(G)=pT(G) product could arise between T16 & T19 in loop 3 of any G-quadruplex form, which could in principle be a lateral, propeller or diagonal loop.

It was surprising to see significant amounts of non-adjacent CPD formation in the SUS3 structure in Na⁺ solution, while very little was seen with the CALU sequence. In our previous work with human telomeric sequences and mutants we were unable to observe non-adjacent CPD formation in the presence of Na⁺ and attributed this to the chair conformation that forms in Na⁺ solution and the conformation of the lateral loops (24, 23). For SUS3, it must be that the two loops are long enough (3 and 4 nucleotides), or positioned in such a way that the T's can overlap in the presence of Na⁺, whereas they can't overlap well when the loops are only 2 & 3 nucleotides, as in CALU, or 3 nucleotides, as in the human telomeric sequences.

For the Stard3NL sequence, non-adjacent CPD formation leading to the pT(G)=pC(G) photoproduct in K⁺ could arise in a number of ways. It could occur between loops 1 & 3, and in this case could only form between C19 and T9 in the chair form, or between C19 and T6 in the basket form. Alternatively, the product could arise within a loop, and in this case occur between C19 and T22 or between C15 and T13 in any type of G-quartet structure. Nucleotide replacement experiments were not carried due problems in easily distinguishing between the two CG sites. The most readily available and minimally perturbing analog would be 5-methylC, but this would deaminate to T in the dimer and become indistinguishable from other non-adjacent CPDs involving T. The pT(G)=pT(G) product observed in both K⁺ and Na⁺ could also arise either between T6 & T22 in loops 1 & 3 in the chair form, or between T9 & T22 in the basket form. It could also arise between T6 & T9 in loop 1 in any form.

It is important to note that the chair and other quadruplex forms may also be in equilibrium with the long hairpin form which has been proposed as an intermediate in folding of the chair conformation (25) (Fig. 1). Long hairpins contain a bulge loop formed from the original pair of lateral loops, which could also lead to non-adjacent CPD formation. We have obtained evidence that a reverse Hoogsteen hairpin may be involved in the formation of non-adjacent CPDs in a human telomeric DNA sequence in the presence of Li⁺ which does not support G quadruplex formation (26).

CONCLUSIONS

We have shown that it is possible to identify G-quadruplex forming sequences in human promoters that can facilitate the formation of non-adjacent CPDs by searching for sequences capable of forming at least two G-tetrads with at least one thymine in loops 1 and 3. Chemical footprinting and circular dichroism experiments indicate that these sequences can fold into G-quadruplexes, though high resolution NMR spectroscopy would be required to rigorously assign the major structures present, which may not, however, be the photoreactive

ones. Despite the apparent structural polymorphisms, we have shown that promoter region G-quadruplexes can be photocrosslinked through intramolecular pyrimidine-pyrimidine photodimerization. Unlike our previous studies with human telomeric DNA G-quadruplexes, we were able to detect non-adjacent CPDs formed in the presence of Na⁺, indicating that Na⁺ binding to G-quadruplexes is not itself inhibitory. These unique non-adjacent CPD photoproducts may be useful for establishing the presence of non-B DNA structures in G quadruplex forming sequences containing pyrimidine rich loops in human promoters *in vivo*

We have also obtained the first evidence for the formation of a non-adjacent thymine-cytosine CPD. The thymine-uracil CPD, resulting from deamination of the thymine-cytosine CPD, was one of the main photoproducts isolated after irradiation of the STARD3NL G-quadruplex with UVB light in the presence of K⁺, which is the biologically relevant intracellular cation. The ability to form non-adjacent CPDs with C would greatly expand the number of G-quadruplex forming sequences capable of being photocrosslinked. While it is known that the adjacent *trans,syn* cyclobutane dimer of dCpT deaminates approximately 10 times more slowly than the corresponding *cis,syn* dimer (35), the deamination rates for *cis,anti* and *trans,anti* CPDs are unknown. It will also be interesting to know how 5-methylcytosine affects non-adjacent CPD formation in G-quadruplexes as methylation of cytosine at CpG sites is known to enhance *cis,syn* CPD formation (36, 37).

Further studies of G-quadruplex photochemistry will be needed to better understand the sequences and conformations that facilitate the formation of non-adjacent CPDs, and whether these photoproducts also form *in vivo*, and if so, how they are repaired. The high prevalence of G-quartet forming sequences in genomic DNA suggests that if these photoproducts were to form *in vivo* they would have the potential to disrupt both transcription and replication and contribute to the toxic and mutagenic effects of sunlight.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements-

We thank Henry Rohrs for performing the ESI-MS/MS experiments, Dr. Robert Blankenship for the use of his CD instrument, and Dr. Joshua Maurer for the use of his MALDI time-of-flight mass spectrometer. Funding for this work was provided by National Institutes of Health Grant CA404063.

REFERENCES

1. Saini N, Zhang Y, Usdin K and Lobachev KS (2013) When secondary comes first - the importance of non-canonical DNA structures. *Biochimie* 95, 117–123. [PubMed: 23084930]
2. Choi J and Majima T (2011) Conformational changes of non-B DNA. *Chem Soc Rev* 40, 5893–5909. [PubMed: 21901191]
3. Wang Y and Patel DJ (1993) Solution structure of the human telomeric repeat d[AG3(T2AG3)3] G-tetraplex. *Structure* 1, 263–282. [PubMed: 8081740]
4. Luu KN, Phan AT, Kuryavy V, Lacroix L and Patel DJ (2006) Structure of the human telomere in K⁺ solution: An intramolecular (3 + 1) G-quadruplex scaffold. *J. Am. Chem. Soc* 128, 9963–9970. [PubMed: 16866556]

5. Dai J, Carver M, Punchihewa C, Jones RA 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. [PubMed: 17626043]
6. Zhang Z, Dai J, Veliath E, Jones RA 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. [PubMed: 19946019]
7. Wlodarczyk A, Grzybowski P, Patkowski A and Dobek A (2005) 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* 109, 3594–3605. [PubMed: 16851398]
8. Dai J, Carver M and Yang D (2008) Polymorphism of human telomeric quadruplex structures. *Biochimie* 90, 1172–1183. [PubMed: 18373984]
9. Tran PL, Mergny JL and Alberti P (2011) Stability of telomeric G-quadruplexes. *Nucleic Acids Res* 39, 3282–3294. [PubMed: 21177648]
10. Lim KW, Amrane S, Bouaziz S, Xu W, Mu Y, Patel DJ, Luu KN and Phan AT (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. [PubMed: 19271707]
11. Lim KW, Alberti P, Guedin A, Lacroix L, Riou JF, Royle NJ, Mergny JL and Phan AT (2009) Sequence variant (CTAGGG)_n in the human telomere favors a G-quadruplex structure containing a G.C.G.C tetrad. *Nucleic Acids Res* 37, 6239–6248. [PubMed: 19692585]
12. Phan AT, Luu KN and Patel DJ (2006) Different loop arrangements of intramolecular human telomeric (3+1) G-quadruplexes in K⁺ solution. *Nucleic Acids Res* 34, 5715–5719. [PubMed: 17040899]
13. Phan A-T and Patel DJ (2003) Two-Repeat Human Telomeric d(TAGGGTTAGGGT) Sequence Forms Interconverting Parallel and Antiparallel G-Quadruplexes in Solution: Distinct Topologies, Thermodynamic Properties, and Folding/Unfolding Kinetics. *J. Am. Chem. Soc* 125, 15021–15027. [PubMed: 14653736]
14. Eddy J and Maizels N (2006) Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Res* 34, 3887–3896. [PubMed: 16914419]
15. Huppert JL and Balasubramanian S (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res* 35, 406–413. [PubMed: 17169996]
16. Mathad RI, Hatzakis E, Dai J and Yang D (2011) c-MYC promoter G-quadruplex formed at the 5'-end of NHE III1 element: insights into biological relevance and parallel-stranded G-quadruplex stability. *Nucleic Acids Res* 39, 9023–9033. [PubMed: 21795379]
17. Dai J, Chen D, Jones RA, Hurley LH and Yang D (2006) NMR solution structure of the major G-quadruplex structure formed in the human BCL2 promoter region. *Nucleic Acids Res* 34, 5133–5144. [PubMed: 16998187]
18. Cogoi S and Xodo LE (2006) G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic Acids Res* 34, 2536–2549. [PubMed: 16687659]
19. Sun D, Guo K and Shin YJ (2011) Evidence of the formation of G-quadruplex structures in the promoter region of the human vascular endothelial growth factor gene. *Nucleic Acids Res* 39, 1256–1265. [PubMed: 20959293]
20. Hsu SD, Varnai P, Bugaut A, Peszka AP, Neidle S and Balasubramanian S (2009) A G-Rich Sequence within the c-kit Oncogene Promoter Forms a Parallel G-Quadruplex Having Asymmetric G-Tetrad Dynamics. *J. Am. Chem. Soc* 131, 13399–13409. [PubMed: 19705869]
21. 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. [PubMed: 23422559]
22. Douki T and Cadet J (2001) Individual Determination of the Yield of the Main UV-Induced Dimeric Pyrimidine Photoproducts in DNA Suggests a High Mutagenicity of CC Photolesions. *Biochemistry* 40, 2495–2501. [PubMed: 11327871]
23. Su DG, Fang H, Gross ML and Taylor JS (2009) Photocrosslinking of human telomeric G-quadruplex loops by anti cyclobutane thymine dimer formation. *Proc. Natl. Acad. Sci. U.S.A* 106, 12861–12866. [PubMed: 19628696]

24. Smith JE, Lu C and Taylor JS (2014) Effect of sequence and metal ions on UVB-induced anti cyclobutane pyrimidine dimer formation in human telomeric DNA sequences. *Nucleic Acids Res* 42, 5007–5019. [PubMed: 24598261]
25. Mashimo T, Yagi H, Sannohe Y, Rajendran A and Sugiyama H (2010) Folding pathways of human telomeric type-1 and type-2 G-quadruplex structures. *J. Am. Chem. Soc* 132, 14910–14918. [PubMed: 20882978]
26. Lu C, Smith-Carpenter JE and Taylor JA (2018) Evidence for Reverse Hoogsteen Hairpin Intermediates in the Photocrosslinking of Human Telomeric DNA Sequences. *Photochem. Photobiol.*
27. Yadav VK, Abraham JK, Mani P, Kulshrestha R and Chowdhury S (2008) QuadBase: genome-wide database of G4 DNA--occurrence and conservation in human, chimpanzee, mouse and rat promoters and 146 microbes. *Nucleic Acids Res* 36, D381–385. [PubMed: 17962308]
28. Dhapola P and Chowdhury S (2016) QuadBase2: web server for multiplexed guanine quadruplex mining and visualization. *Nucleic Acids Res* 44, W277–283. [PubMed: 27185890]
29. Brooks TA, Kendrick S and Hurley L (2010) Making sense of G-quadruplex and i-motif functions in oncogene promoters. *Febs J* 277, 3459–3469. [PubMed: 20670278]
30. Rigo R, Palumbo M and Sissi C (2017) G-quadruplexes in human promoters: A challenge for therapeutic applications. *Biochim. Biophys. Acta* 1861, 1399–1413.
31. Wang Y, Taylor JS and Gross ML (1999) Nuclease P1 digestion combined with tandem mass spectrometry for the structure determination of DNA photoproducts. *Chem. Res. Toxicol* 12, 1077–1082. [PubMed: 10563833]
32. Su DG, Kao JL, Gross ML and Taylor JS (2008) 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* 130, 11328–11337. [PubMed: 18680367]
33. Vorlickova M, Kejnovska I, Sagi J, Renciuik D, Bednarova K, Motlova J and Kypr J (2012) Circular dichroism and guanine quadruplexes. *Methods* 57, 64–75. [PubMed: 22450044]
34. Sun D and Hurley LH (2010) Biochemical techniques for the characterization of G-quadruplex structures: EMSA, DMS footprinting, and DNA polymerase stop assay. *Methods Mol Biol* 608, 65–79. [PubMed: 20012416]
35. Douki T and Cadet J (1992) Far-UV photochemistry and photosensitization of 2'-deoxycytidyl-(3'-5')-thymidine: Isolation and characterization of the main photoproducts. *Journal of Photochemistry and Photobiology B: Biology* 15, 199–213.
36. Drouin R and Therrien J-P (1997) UVB-induced cyclobutane pyrimidine dimer frequency correlates with skin cancer mutational hotspots in p53. *Photochem. Photobiol* 66, 719–726. [PubMed: 9383997]
37. Tommasi S, Denissenko MF and Pfeifer GP (1997) Sunlight induces pyrimidine dimers preferentially at 5-methylcytosine bases. *Cancer Research* 57, 4727–4730. [PubMed: 9354431]

Figure 1. Common G-quadruplex folding patterns found in human telomeric DNA sequences. A) Parallel conformation with three double-chain-reversal or propeller (p) loops. B) Hybrid-1 & C) Hybrid-2 conformations with one propeller loop and two lateral loops. D) Antiparallel basket conformation with two lateral loops and one diagonal (d) loop. E) Antiparallel chair conformation with three lateral (l) loops. F) Long G-hairpin form that is proposed as an intermediate in the folding of the chair conformation, which has a hairpin loop and an interior loop. Only the basket, chair and long G-hairpin have structures that

bring the T's of loops 1 and 3 close enough for CPD formation, as indicated by the two headed dashed arrows.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 2. Nuclease P1 digestion products of adjacent and non-adjacent CPDs.

Digestion of DNA containing adjacent photoproducts leads to trinucleotides, whereas cleavage of non-adjacent photoproducts leads to tetranucleotides. A) Adjacent CPDs form between sequential nucleotides in a chain, whereas non-adjacent dimers form between nucleotides separated by one or more nucleotides in a chain or between different chains. The wiggly line refers to the various possible stereochemistries shown in panels B & C. B) Possible *syn* CPD stereochemistries resulting from parallel alignment of the bases that could arise from both adjacent and non-adjacent nucleotides. The *cis-syn-II* stereochemistry has

never been detected in an adjacent CPD. C) Possible *anti* CPD stereochemistries resulting from antiparallel alignment of the bases that can only arise from non-adjacent nucleotides.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 3. HPLC analysis of the nuclease P1 digests of irradiated G quadruplex forming sequences.

Reverse phase HPLC traces of the NP1 digested products in the tri- and tetranucleotide region for A) CALU, B) STARD3NL, and C) SUS3D that had been irradiated on ice with 150 mM of the indicated metal ion. The approximate yields of the products were ascertained from their absorption coefficients relative to those of the mononucleotides.

A. pT(G)=pT(G)

B. MS/MS of m/z 650.3, $[M-2H]^{2-}$

Figure 4. Identification of the pT(G)=pT(G) NP1 product from irradiation of the CALU sequence in 150 mM NaCl.

A) Structure of the non-adjacent CPD NP1 digestion product in which the stereochemistry is undetermined. B) ESI-MS/MS of the $(M-2H)^{2-}$ ion showing characteristic product ions of the pT(G)=pT(G) non-adjacent photoproduct.

Figure 5. NP1-HPLC assay of CALU and the U-substituted sequences after 2.5 h of UVB irradiation at pH 7.5.

A) NP1-HPLC trace for parent CALU sequence in 150 mM NaCl and sequences containing site-specific uracil substitutions. MALDI confirmed the assignment of pT(G)=pT(G) and pT(G)=pU(G) photoproducts. B) Sequences used in mapping of CALU G-quadruplex photoproduct sites.

A. pT(G)=pU(G)

B. ESI MS/MS of m/z 643.3, $[M-2H]^{2-}$

Figure 6. Identification of the pT(G)=pU(G) NP1 product from irradiation of the STARD3NL sequence in 150 mM KCl.

A) Structure of the non-adjacent CPD NP1 digestion product in which the stereochemistry is undetermined. B) ESI-MS/MS of $(M-2H)^{2-}$ ion showing characteristic product ions of the pT(G)=pU(G) non-adjacent photoproduct. This product is the deamination product of pT(G)=pC(G).

A. pT(G)=pT(T)

B. ESI MS/MS of m/z 637.9. $[M-2H]^{2-}$

Figure 7. Identification of the pT(G)=pT(T) NP1 product from irradiation of the SUS3D sequence in 150 mM NaCl.

A) Structure of the non-adjacent CPD NP1 digestion product in which the stereochemistry of the cyclobutane ring is undetermined. B) ESI-MS/MS of $(M-2H)^{2-}$ ion showing characteristic product ions of the pT(G)=pT(T) non-adjacent photoproduct.

A. 150 mM KCl **B. 150 mM NaCl**

Figure 8. CD spectra of the G-quadruplex forming sequences as a function of metal ion. CD spectra were acquired at 4°C at pH 7.5 in A) 150 mM KCl and B) 150 mM NaCl. A peak at about 260 nm and a minimum at 240 nm is indicative of a parallel structure, whereas a peak at 295 nm is indicative of an antiparallel structure.

Figure 9. DNA methylation protection assay of G-quadruplex forming sequences as a function of metal ion.

Hoogsteen base pairing inhibits alkylation of N7 by dimethylsulfate, and hence subsequent formation and cleavage of abasic sites by piperidine, reveals G's that are involved in triplex and quadruplex formation. G quadruplex sequences were incubated with dimethyl sulfate in 150 mM KCl or NaCl on ice for the indicated time in min, where lane C is the control reaction carried out for 8 min in pure water. The numbers to the side refer to the position of the G in the sequence.

Figure 10. Quantification of the extent of methylation protection as a function of nucleotide position and metal ion.

The % inhibition of DNA methylation was calculated by comparing the percent cleavage for the 8 minute methylation reaction in the presence or absence of metal cation. Data is only shown for bands that could be quantified accurately. The bars below the G's indicate runs of G's. An L signifies an isolated G that would presumably be in a linker.

Figure 11. Photodimer sites mapped onto antiparallel G-quadruplex structures.

The sites of non-adjacent CPD formation are mapped onto three tetrad A) chair and B) basket conformations of the CALU, STARD3NL and SUS3D sequences. The possible reaction pairs are indicated with dashed lines for inter-loop pairs and dotted lines for intra-loop pairs. For CALU, involvement of T12 was ruled out based on uridine substitution experiments shown in Fig. 7. The G's in grey are those that correspond to local inhibition minima in the methylation protection plots shown in Fig. 10 for the ion under consideration.

Table 1.
Sequence and melting temperatures of the G quadruplex forming sequences.

T_m 's are for the absorbances at 295 nm in 150 mM of the indicated cation. The blocks of G's that could participate in the formation of three consecutive G-tetrads are underlined.

Name	Sequence (5'–3')	T_m K ⁺	T_m Na ⁺
CALU	<u>GGGGCTGGGGCTGGGGCTGGG</u>	81.5	64
STARD3NL	<u>GGGGCTGCTGGGTGCGGGCGGTGGG</u>	70	60
SUSD3	<u>GGGATGGGGGAAGGGTGTGGG</u>	72	51.5