

Washington University in St. Louis

## Washington University Open Scholarship

---

Arts & Sciences Electronic Theses and  
Dissertations

Arts & Sciences

---

Spring 5-15-2023

### Non-adjacent anti Cyclobutane Pyrimidine Dimers as Intrinsic Probes of Non-B Form Secondary Structures of DNA

Natalia Eugenia Gutierrez Bayona

Follow this and additional works at: [https://openscholarship.wustl.edu/art\\_sci\\_etds](https://openscholarship.wustl.edu/art_sci_etds)

---

#### Recommended Citation

Gutierrez Bayona, Natalia Eugenia, "Non-adjacent anti Cyclobutane Pyrimidine Dimers as Intrinsic Probes of Non-B Form Secondary Structures of DNA" (2023). *Arts & Sciences Electronic Theses and Dissertations*. 2853.

[https://openscholarship.wustl.edu/art\\_sci\\_etds/2853](https://openscholarship.wustl.edu/art_sci_etds/2853)

This Dissertation is brought to you for free and open access by the Arts & Sciences at Washington University Open Scholarship. It has been accepted for inclusion in Arts & Sciences Electronic Theses and Dissertations by an authorized administrator of Washington University Open Scholarship. For more information, please contact [digital@wumail.wustl.edu](mailto:digital@wumail.wustl.edu).

WASHINGTON UNIVERSITY IN ST. LOUIS

Department of Chemistry

Dissertation Examination Committee:

Timothy Wencewicz, Chair

Kimberly Parker

Jay Ponder

John-Stephen Taylor, Advisor

Hani Zaher

Non-adjacent *anti* Cyclobutane Pyrimidine Dimers as Intrinsic Probes of Non-B Form Secondary  
Structures of DNA

by

Natalia Eugenia Gutierrez-Bayona

A dissertation presented to  
Washington University in St. Louis  
in partial fulfillment of the  
requirements for the degree  
of Doctor of Philosophy

August 2023  
St. Louis, Missouri

© 2023, Natalia E. Gutierrez-Bayona

# Table of Contents

List of Figures .....	v
List of Tables .....	xi
Acknowledgments.....	xii
Abstract of the Dissertation .....	xv
Chapter 1: Introduction.....	1
1.1    DNA G-Quadruplex Structures.....	1
1.2    Biological Relevance of G-Quadruplex Structures.....	4
1.2.1    G-Quadruplexes in Telomeres .....	4
1.2.2    G-Quadruplexes in Gene Promoters .....	7
1.2.3    G-Quadruplexes in Translation .....	10
1.2.4    G-Quadruplexes in Replication.....	11
1.3    Methods for G-Quadruplex Detection in the Human Genome.....	12
1.3.1    Sequencing of G-Quadruplex formation in the Human Genome.....	13
1.3.2    Visualization of G-Quadruplexes with Fluorescence Microscopy .....	15
1.3.3    G-Quadruplex Chromatin Immunoprecipitation Sequencing .....	19
1.4    Photoproducts as Intrinsic Probes for Non-B forms of DNA.....	20
1.4.1    UV-Induced DNA Photoproducts .....	21
1.4.3    Anti-CPD Formation in Single Stranded ODN at Low pH.....	25
1.4.4    Anti-CPD Formation in G-Quadruplex Forming Telomeric Sequences.....	26
1.4.5 <i>Trans-anti</i> CPD formation in Reverse Hoogsteen hairpins .....	28
1.4.6    Anti-CPD formation in human promoter sequences .....	30
1.5    Dissertation Overview .....	31
1.6    Figures.....	33
1.7    References.....	47
Chapter 2: Post- and Pre-radiolabeling assays for <i>anti</i> thymidine cyclobutane dimers as intrinsic photoprobes of various types of G-quadruplexes, reverse Hoogsteen hairpins, and other non-B DNA structures .....	57
2.1    Abstract.....	57

2.2	Introduction.....	58
2.3	Materials and Methods.....	60
2.3.1	Materials and Reagents .....	60
2.3.2	Large Scale Preparation of pTpT=(pT)pT .....	61
2.3.3	Large Scale Dephosphorylation of pTpT=(pT)pT with Calf Intestinal Phosphatase.....	61
2.3.4	Large Scale Phosphorylation of TpT=(T)pT with T4 PNK.....	62
2.3.5	HPLC Analysis and Purification of the Post-labeling Assay Intermediates.....	62
2.3.6	Mass Spectrometry of the Post-labeling Assay Intermediates.....	62
2.3.7	Post-labeling Assay of UVB Irradiated DNA.....	63
2.3.8	Post-labeling Assay with Competitor Plasmid DNA.....	64
2.3.9	Preparation of Internally Labelled Tel26 Sequences.....	64
2.3.10	SVP degradation of internally labeled UVB irradiated Tel26 sequences.....	65
2.3.11	SVP degradation and labeling studies with UVB irradiated 14-mer.....	66
2.4	Results and Discussion .....	66
2.5	Conclusion .....	72
2.6	Figures.....	73
2.7	Acknowledgements.....	82
2.8	References.....	82
Chapter 3: Post-labeling Assay of Non-Adjacent <i>anti</i> Cyclobutane Pyrimidine Dimers formation in known Non-B Secondary Structures of DNA.....		85
3.1	Abstract.....	85
3.2	Introduction.....	86
3.2	Materials and Methods.....	88
3.2.1	Materials and Reagents .....	88
3.2.2	Preparation of G-Quadruplex and Reverse-Hoogsteen Hairpin Structures .....	88
3.2.3	Post-labeling Assay of UVB Irradiated ODN Sequences .....	89
3.4	Results and Discussion .....	90
3.4.1	Post-labeling Assay of Tel26 and Tel22 Human Telomeric Sequences .....	90
3.4.2	Post-labeling Assay of the Model Human Telomeric Sequence hTeLo .....	92
3.4.3	Post-labeling Assay of Ciliate Telomeric Sequences Tet-4 and Oxy-4.....	94
3.4.3	Post-labeling Assay of Human Promoter Sequences .....	96

3.5	Conclusions.....	101
3.6	Acknowledgements.....	102
3.7	Figures.....	103
3.8	References.....	113
Chapter 4: Future Studies and Conclusion.....		117
4.1	Conclusions.....	117
4.2	Current and Future Studies .....	119
4.2.1	Detection and Location of <i>anti</i> CPDs in HeLa cells .....	119
4.2.1	Photochemical Signatures of Different Non-B Structures of DNA .....	120
4.3.	Concluding Remarks.....	122
4.3	Figures and Tables .....	123
4.4	References.....	126

# List of Figures

Figure 1.1. Four-stranded G-quadruplex structural polymorphism. (a) Intramolecular parallel G-quadruplex conformation. The G-tetrad building block of these structures has a Na<sup>+</sup> cation in the plane of the tetrad whereas (b) K<sup>+</sup> is found between the planes of two tetrads. (c) Intramolecular anti-parallel G-quadruplex conformations. (d) Intermolecular conformation involving two different strands of DNA. Figures created with BioRender.....33

Figure 1.2. Possible sites of G-quadruplex formation in the human genome. G-quadruplexes are believed to form in G-rich regions of DNA that become transiently single stranded, such as in (a) telomeres, (b) replication forks, (c) transcription bubbles and (d) during translation. Figure adapted from (123) and created with BioRender.....34

Figure 1.3. Potential biological role of G-quadruplexes in telomeres (a) G-quadruplex structures are believed to help regulate telomerase activity and (b) to play a capping role at the end of chromosomes. Figure adapted from (39) and created with BioRender.....35

Figure 1.4. Potential biological roles of G-quadruplexes in transcription. G-quadruplexes are believed (a) to enhance transcription by acting as binding sites for transcription factors, (b) to repress transcription by blocking transcription factor binding, (c) to enhance transcription by stabilizing the R-loop, or (d) to repress transcription by stalling RNA-polymerase. Figure adapted from (41,123) and created with BioRender.....36

Figure 1.5. G-quadruplex sequencing (G4-seq). (a) In a typical G4-seq protocol, sequencing is done twice. Once under conditions that enables accurate sequencing of DNA fragments, followed by re-sequencing the DNA under conditions that promote G-quadruplex formation. (b) Non-canonical G-quadruplex structures identified by G4-seq. Figure adapted from (77) and created with BioRender.....37

Figure 1.6. Fluorescence imaging of G-quadruplexes *in vivo*. (a) G-quadruplex detection through fluorescent antibodies. (b) Antibody imaging of G-quadruplexes in individual chromosomes (red) vs DNA (blue). (c) Representation of increased foci in cells after addition of stabilizing ligands such as PDS. (d) Two fluorescent probe molecules for G-quadruplex detection. Figures adapted from (6) and created with BioRender.....38

Figure 1.7. G-quadruplex chromatin immunoprecipitation sequencing (G4 ChIP-seq). (a) The technique involves cell fixation with formaldehyde, followed by chromatin precipitation with G-quadruplex specific antibodies and Next-Gen sequencing. (b) Relationship between G-

quadruplexes, chromatin state, and transcription. Figure adapted from (93) and created with BioRender.....39

Figure 1.8. DNA photoproducts as intrinsic probes for DNA secondary structure and conformation. (a) Different types of thymine photoproducts that form in the A and B duplex forms of DNA. (b) Formation of *anti* and (c) *syn* non-adjacent photoproducts from non-B and B conformations of DNA.....40

Figure 1.9. Structure of the different stereoisomers of T=T CPDs forming in DNA from adjacent or non-adjacent (a) head-to-head or (b) head-to-tail bases. After the Thy=Thy CPDs are released from the deoxyribose sugar by acid hydrolysis only six stereoisomers are observed, where the two *trans,syn* CPDs with the asterisk represent that they are enantiomers of each other as also is the case for the two *trans,anti* CPDs.....41

Figure 1.10. Structure determination of the *cis, anti*-CPD formed at low pH in a single stranded oligonucleotide. (a) Exonuclease-coupled MS for determination of the position where the non-adjacent photoproduct formed (b) NP1 digestion to determine flanking bases. (c) HF degradation to thymine dimer for identification of stereochemistry by comparison to authentic compounds.....42

Figure 1.11. NP1- coupled mass spectrometry assay for the detection of nonadjacent and adjacent photoproducts.....43

Figure 1.12. G-quadruplex structures and intermediates implicated in *anti* CPD formation in human telomeric DNA. (a) Basket structure that could lead to the formation of non-adjacent photoproducts from T bases found in the lateral loops (b) Hybrid-1 structure that is not expected to produce non-adjacent CPDs but that is believed to be in equilibrium with a photoreactive form of the (c) basket conformation known as Form-3. (d) A photoreactive triplex intermediate conformation between the basket, two hybrid types, and (e) chair conformation that is believed to be a minor photoreactive structure of G-quadruplexes (96,115).....44

Figure 1.13. Proposed scheme and structure that explain *anti* CPD formation in Tel22 and Tel26 as function of cation and sequence. (a) Cationic and sequence dependent formation of nonadjacent CPDs in Tel26 and Tel22 sequences (b) three- tetrad basket structure in Na<sup>+</sup> where nonadjacent photoproducts are not expected to form (c) Form 3 basket structure that is believed to be the photoreactive G-quadruplex that forms non-adjacent CPD's (d) Reverse-Hoogsteen hairpin structure that may be the major photoreactive conformation leading to *trans,anti* T2=T2 CPD (e) Hoogsteen and reverse-Hoogsteen base pairing between two guanines.....45

Figure 1.14. Sites of non-adjacent CPD formation in putative G-quadruplex forming sequences in human promoters. (a) Chair and (b) basket forms of putative G-quadruplex forming sequences from the promoters of the CALU, STARD3NM, and SUS3 genes.....46



Figure 2.1. Idea of using structure-specific photoproducts to detect and map folded DNA structures such as basket G-quadruplexes. a) Structure of adjacent *cis,syn* and non-adjacent *cis* and *trans,anti* T=T CPDs. b) Whereas UV irradiation of B DNA produces adjacent *cis,syn* T=T CPDs, UV irradiation of non-B DNA folded structures such as the basket G quadruplex form of human telomeric DNA produces unique non-adjacent *anti* T=T CPDs.....74

Figure 2.2. Post-labeling assay schemes for T=T CPDs. a) DNA containing an adjacent *cis,syn* T=T CPD is degraded to a trinucleotide with snake venom phosphodiesterase (SVP) and then dephosphorylated with calf intestinal phosphatase (CIP) followed by radiolabeling with [ $\gamma$ -<sup>32</sup>P]-ATP and kinase to give a phosphorylated trimer product. Irradiation with 254 nm light (UVC) reverts the CPD to the canonical bases. b) DNA containing a non-adjacent *anti* T=T CPD is degraded to a tetranucleotide with snake venom phosphodiesterase (SVP) and after dephosphorylation with calf intestinal phosphatase (CIP) and radiolabeling with [ $\gamma$ -<sup>32</sup>P]-ATP and kinase would give mono and dephosphorylated tetranucleotide products. Irradiation with 254 nm light would revert the tetranucleotide product to two dinucleotides. Treatment of the tetra- or dinucleotide products with NP1 would release the 5'-terminal radiolabeled nucleotide. All radiolabeled products are distinguishable by high resolution polyacrylamide gel electrophoresis.....75

Figure 2.3. HPLC analysis and purification of the intermediates in the post-labeling assay of Tel26 DNA irradiated with UVB light. a) Irradiated Tel26 digested with snake venom phosphodiesterase (SVP) for 24 h to give *trans,anti* pTpT=(pT)pT. b) Dephosphorylation of *trans,anti* pTpT=(pT)pT with calf intestinal phosphodiesterase (CIP) to give TpT=(T)pT. c) rephosphorylation of TpT=(T)pT from B with polynucleotidyl kinase (PNK) and ATP.....76

Figure 2.4. ESI-MS/MS characterization of the intermediates in the post-labeling assay of irradiated Tel26 DNA. a) MS/MS of the [M - 2H]<sup>2-</sup> ion (m/z 624.11) of *trans,anti* pTpT=(pT)pT. b) MS/MS of the [M - 2H]<sup>2-</sup> ion (m/z 545.13) of the dephosphorylated *trans,anti* TpT=(T)pT.....77

Figure 2.5. MS/MS fragmentation pathways of the *trans,anti* T=T CPD-containing products of Figure 4. Pathways for pTpT=(pT)pT are shown in normal font, whereas pathways for TpT=(T)pT are shown in bold italic font.....78

Figure 2.6. Post-labeling assay carried out on Tel26 in the absence or presence of competing plasmid DNA. Lanes 2, 3 and 5 are authentic standards prepared by radiolabeling TT and HPLC purified GT=T and TpT=(T)pT. UVC was used to photorevert T=T CPD-containing products. Lane headings for the post-labeling reactions refer to the ratio of the base pair concentrations of Tel26 to plasmid DNA. Lanes 13-18 are shown with a narrower dynamic range to make the photo-reverted products in lanes 14 and 16 more apparent.....79

Figure 2.7. Preparation and use of internally labeled Tel26 for studies of *anti* T=T CPD formation. a) ODNs (written 5'-3') used to prepare Tel26 internally labeled in loop 1 and loop 2 where the T's that undergo *trans,anti* T=T CPD formation are in bold. The 30-mer serves as a ligation scaffold. b) Degradation products the non-adjacent *trans,anti* T=T CPD formed between loop 1 and loop 3 of loop 1 internally labeled Tel26. The relative mobility of the enzymatic degradation products is compared to the mass to charge ratio assuming that all phosphate groups are fully ionized.....80

Figure 2.8. Analysis of the enzymatic degradation products of post- and pre-labeled Tel26 by denaturing polyacrylamide gel electrophoresis. Lanes 1-3: authentic products obtained by end labeling commercial ODNs. Lanes 4-7: post-labeling assay carried out on UVB irradiated Tel26 followed by dephosphorylation with CIP, with and without CPD photoreversion with UVC light as indicated. Lanes 8-13: SVP treatment of UVB irradiated Tel26 that was internally pre-labeled in loop 1 followed by CIP or CIP and NP1, with and without CPD photoreversion with UVC light, all as indicated. Lanes 14-18: SVP treatment of UVB irradiated Tel26 that was internally pre-labeled in loop 2 followed by CIP with and without CPD photoreversion with UVC light as indicated. Treatments with CIP were for either 2.5 min (a) or 30 min (b). Figure prepared from two separate gels carried out under identical conditions as indicated by the vertical line.....81

Figure 2.9. Analysis of the enzymatic degradation products of pre- and post-labeled 14-mer containing a *cis,anti* T=T CPD. a) Schematic for the enzymatic reactions carried out on the pre- and post-labeled UVB irradiated 14-mer where the T's undergoing *anti* CPD formation are in bold. b) Denaturing PAGE of the enzymatic reaction products. Lanes 5-12: the 5'-pre-labeled UVB irradiated 14-mer was digested with SVP and then dephosphorylated with CIP, where a, b and c refer to the times of 0-, 5- and 10-minutes following addition of CIP, with and without photoreversion by UVC light as indicated. Lanes 13-20: The UVB irradiated 14-mer was degraded by SVP and then labeled with PNK and [ $\gamma$ -<sup>32</sup>P]ATP, after which it was incubated for increasing time with CIP as described for the pre-labeled experiment, with and without photoreversion with UVC light as indicated.....82

Figure 3.1. The primary conformation of DNA determines the type of dipyrimidine photoproducts that form (a) In helical duplex B DNA, adjacent pyrimidine bases react to produce CPDs with the *cis,syn* stereochemistry and (6-4) photoproducts. (b) In some non-B forms of DNA, non-adjacent pyrimidine bases are brought together to produce photoproducts such as *anti* CPDs. The first thoroughly characterized *cis,anti* T=T CPD was discovered from the 14-mer sequence shown upon UVB irradiation at low pH. To this day the structure of the DNA leading to this product is unknown.....104

Figure 3.2. Enzymatic degradation products of adjacent and non-adjacent photoproducts by NP1 and SVP. The degradation products produced by these enzymes depend on whether they can cleave

the 3'-side (NP1) or 5'-side (SVP) of an undamaged nucleotide. Adjacent photoproducts degraded by these enzymes produce trimers, where NP1 is unable to cleave the nucleotide 3'-adjacent to the photoproduct, while SVP cannot cleave the nucleotide 5'-adjacent to the damage. In the presence of non-adjacent photoproducts, both enzymes produce tetramers that differ by which nucleotides are left untouched at the 3' or 5' ends of the nucleotides involved in photoproduct formation. SVP degradation for both adjacent and non-adjacent products allow for the dephosphorylation and rephosphorylation of the canonical 5'- nucleotide with radioactive <sup>32</sup>P which is the basis of the post-labeling assay.....105

Figure 3.3. Expected photochemical reactivity of G-quadruplex and hairpin structures of DNA. Only adjacent CPDs are expected to form in (a) parallel and (b & c) both types of hybrid structures whereas both adjacent and non-adjacent CPDs are expected to form in (d) basket, (e) chair, and (f) hairpin structures.....106

Figure 3.4. Post-labeling assay of UVB irradiated Tel26 and Tel22 human telomeric sequences in the presence of K<sup>+</sup>, Na<sup>+</sup>, and Li<sup>+</sup>. Lane 1 contains [γ-<sup>32</sup>P]-ATP used for radiolabeling, and lanes 2 and 3 contain radiolabeled p\*TpT=(p\*T/T)pT before and after photoreversal. Lanes 4 and 11 shows the products of SVP treatment of the 5'-end labeled Tel26 and Tel22 to test the efficiency of the digestion reaction and to produce authentic p\*dA. A 254 nm lamp was used to photoreverse T=T CPD-containing products.....107

Figure 3.5. Post-labeling assay of UVB irradiated hTeLo sequence under conditions that promote the formation of basket, hybrid, and parallel DNA G-quadruplexes. Lanes 1-4 are authentic standards of p\*GpT=pT and p\*TpT=(p\*T/T)pT. Lanes 5-8 are the products from the post-labeling assay of Tel26 and Tel22 sequences irradiated with UVB in K<sup>+</sup>. Lanes 9-10 are the products of the post-labeling assay of the 14-mer d(GTATCATGAGGTGC) irradiated with UVB under acidic conditions to produce the *cis,anti* T=T CPD. Lanes 11-20 are conditions where hTeLo was assayed with the post-labeling technique under high ionic (HIS) and low ionic conditions (LIS) in the presence of Na<sup>+</sup> (basket) or K<sup>+</sup> (hybrid) to promote G-quadruplex formation. The parallel structure was induced in the presence of K<sup>+</sup> under highly dehydrating conditions. A 254 nm lamp was used to photoreverse T=T CPD-containing products.....108

Figure 3.6. Post-labeling assay of UVB irradiated Tet-4 and Oxy-4 telomeric sequences in the presence of K<sup>+</sup>, Na<sup>+</sup>, and Li<sup>+</sup>. (a) Lanes 1 and 2 are authentic standards of p\*TpT=(p\*T/T)pT. Lanes 3 and 10 show the products of SVP digested 5'-end-labeled Oxy-4 and Tet-4 as a reference for the completeness of the digestion and to produce p\*dT. A 254 nm lamp was used to photoreverse T=T CPD-containing products. (b) Proposed chair and basket G-quadruplex structures that could explain the photocrosslinking reactions originally observed in these sequences.....109

Figure 3.7. Post-labeling assay of UVB irradiated CALU and STARD3NL human promoter sequences in the presence of  $K^+$ ,  $Na^+$ , and  $Li^+$ . (a) Lane 1 contains  $[\gamma\text{-}^{32}\text{P}]\text{-ATP}$  used in the assay, and lanes 2 and 3 are authentic standards of  $p^*\text{TpT}=(p^*\text{T}/\text{T})p\text{T}$ . Lanes 4 and 11 show the products of SVP digested 5'-end-labeled CALU and STARD3NL as a reference for the completeness of digestion and as to produce  $p^*\text{dG}$ . A 254 nm lamp was used to photoreverse T=T CPD-containing products. (b) Bar graph showing the intensity of the  $p\text{NpY}$  dinucleotide radiolabeled bands following photoreversal with 254 nm light.....110

Figure 3.8. Proposed sites of major non-adjacent CPD photoproducts formation in the CALU and STARD3NL promoter sequences with their respective SVP and NP1 tetramer digestion products. Possible chair and basket antiparallel G-quadruplex structures adopted by the (a) CALU and (b) STARD3NL sequences to produce the dipyrimidine CPDs detected by NP1-MS analysis as indicated by the dashed and dotted lines. The dashed lines refer to inter-loop CPDs whereas the dotted lines refer to intra-loop CPDs. The tables below show the products observed from the NP1-MS assay and what would be expected to be observed from a MS assay if SVP were to be used suggesting that a combination of the two should be used in future assays.....111

Figure 3.9. Post-labeling assay of UVB irradiated c-MYC and ILPR human promoter sequences in the presence of  $K^+$ ,  $Na^+$ , and  $Li^+$ . (a) Lanes 1 and 2 are authentic standards of  $p^*\text{TpT}=(p^*\text{T}/\text{T})p\text{T}$ . Lanes 3 and 10 contain the SVP digested products of 5'-end labeled c-MYC and ILPR sequences as a reference for the completeness of digestion and as a standard for  $p^*\text{dT}$  for c-MYC and  $p^*\text{dG}$  for ILPR. A 254 nm lamp was used to photoreverse T=T CPD-containing products. (b) Bar graph showing the intensity of the radiolabeled  $p^*\text{NpY}$  dinucleotide bands following photoreversal with 254 nm light.....112

Figure 3.10. Proposed G-quadruplex structures adopted by the c-MYC and ILPR promoter sequences in the presence of  $K^+$ . (a) c-MYC is proposed to adopt a set of parallel structures where the most stable conformation forms from the G bases in the II, III, IV, and V G-runs (c-MYC 2345) with a 1:2:1 loop arrangement, but it is also observed to form two more parallel structures involving the I, II, III, and IV G-runs (c-MYC 1234) with a 1:2:1 and 1:1:2 loop arrangement. These parallel structures could explain why (b) no NP1 tetramer products were detected from HPLC analysis of SVP digested UVB irradiated c-MYC in the presence of  $K^+$ . (c) The two proposed G-quadruplex structures adopted by the ILPR-a2 sequence where the antiparallel structure would be expected to produce non-adjacent photoproducts.....113

Figure 4.1. Workflow for the post-labeling assay for the detection of *anti* T=T CPDs photoproducts *in vivo*. Once HeLa cells are irradiated with UVB light, the post-labeling assay will be performed on telomeric DNA that is purified through a biotin-streptavidin pull down method.....124

# List of Tables

Table 4.1: Predicted SVP and NP1 digestion products from the possible sites of non-adjacent T=C CPD formation in the STARD3NL.....	125
Table 4.2: Predicted SVP and NP1 digestion products from the possible sites of non-adjacent T=T CPD formation in the STARD3NL sequence. ....	125
Table 4.3: Proposed pre-labeled or T/U replacement sequences to determine the bases involved in <i>anti</i> T=T and T=C CPDs that formed in the STARD3NL sequence. ....	126

# Acknowledgments

I would like to start by thanking my entire family. My mom and my dad for teaching me the strength and discipline needed to accomplish this amazing achievement. To my siblings Camila, Samuel, Juana, and Luisa for their support and belief in me. To my nieces Isabella and Luciana for showing me what perseverance, tenacity, and bravery looks like. I would like to especially thank the people that became my family away from home, Kimi, Nilab, Wagma, and Ruth. Without all these people in my life I would not have been able to be where I am today and become the person that I am today.

To Dr. Taylor I want to especially thank you for giving me the opportunity and space to grow as a scientist. For allowing me to continue to explore my fascination towards this amazing field, and for teaching and guiding me to be more curious, to be more creative, and to be more open-minded towards what research and science has to offer.

I would also like to thank Dr. Wencewicz and Dr. Ponder for offering guidance and expertise throughout my time in the program. I want to especially thank Dr. Loomis for being a mentor to me from the moment I started at WashU and for giving me countless opportunities and support in this path to discover my future in the academic world.

Last but definitely not least, I want to thank everyone that I had a chance to know and become friends with during my time at WashU. I want to thank all the faculty, staff, and fellow students that made my time at this university so pleasant and enjoyable. I want to especially thank Elena and Dan for being there for me during key and essential parts of this journey.

The research reported in this thesis was supported by the National Science Foundation under Grant No. 2003688.

Natalia E. Gutierrez-Bayona

*Washington University in St. Louis*

*August 2023*

Dedicated to my parents, my four amazing siblings, my two brave beautiful nieces, my extremely supportive strong friends, and most importantly my cats Eugene & Azula



## ABSTRACT OF THE DISSERTATION

Non-adjacent *anti* Cyclobutane Pyrimidine Dimers as Intrinsic Probes of Non-B Form Secondary Structures of DNA

by

Natalia Eugenia Gutierrez-Bayona

Doctor of Philosophy in Chemistry

Washington University in St. Louis, 2023

Professor John-Stephen Taylor

G-quadruplexes are four-stranded structures of DNA composed of G-quartets that have been proposed to play an extremely important role in replication, transcription, and translation. It has, however, been extremely difficult to unequivocally demonstrate that G-quadruplexes form in living cells due to the lack of probes that would allow for their unambiguous detection and location without disrupting the DNA or the cellular environment. To circumvent these problems, it was proposed that DNA itself could be used as an intrinsic photoprobe for certain classes of G-quadruplex structures and other non-B DNA conformations *in vivo*. Recently, we discovered that UVB irradiation of human telomeric DNA and various other G-quadruplex forming sequences found in human promoters, as well as reverse Hoogsteen hairpins, results in a unique class of non-adjacent *anti* cyclobutane pyrimidine dimers (CPDs). In this thesis, we explore the proposal that *anti* CPD formation could be used as an intrinsic photoprobe for certain classes of G-quadruplex structures and other non-B DNA conformations *in vivo*. The idea would be to use a pulse of UVB light to irreversibly trap the non-B DNA structures via *anti* CPD formation without perturbing the

dynamics of the DNA or the cell in the process, after which the *anti* CPDs could be mapped to provide unambiguous evidence for the formation and location of these structures *in vivo*

As a first step towards this goal, we report the development of radioactive post- and pre-labeling assays for the detection of nonadjacent photoproducts and show its ability to detect the *trans,anti* T=T CPD that forms in human telomeric DNA. We do so by making use of the enzyme snake venom phosphodiesterase (SVP) to degrade the *trans,anti*-CPD to a tetramer (pT)pT=(pT)pT, that after dephosphorylation with calf intestinal phosphodiesterase and rephosphorylation with kinase and [<sup>32</sup>P]-ATP produces a fragment of DNA that can be analyzed by high resolution gel electrophoresis with a fmol lower limit of detection. In the pre-labeling assay, radiolabeled phosphates are introduced into *anti* T=T CPD forming sites by ligation of 5'-end-labeled oligodeoxynucleotides and only require treatment with SVP. We also demonstrate that the assays can detect the stereoisomeric *cis,anti* T=T CPD.

To determine the scope and limitations of the post-labeling assay method to detect non-adjacent CPDs, we use it to investigate the photochemistry of a number of biologically relevant non-B secondary structures. We report the detection of non-adjacent CPDs under conditions that were previously determined by the NP1-MS assay to produce these photoproducts in insignificant amounts, as well as in structures not expected to form *anti* CPDs. We show that photoreversal of the CPDs with 254 nm light can be used distinguish DNA photocrosslinked by non-adjacent CPDs from non-photoreversible DNA photoproducts and from partially degraded DNA. We show that the post-labeling assay can be used to corroborate and complement enzyme-coupled mass spectrometry assays resulting in a more complete understanding of the photoreactivity of non-B secondary structures.

# Chapter 1: Introduction

## 1.1 DNA G-Quadruplex Structures

Guanine-rich sequences of DNA and RNA are known to fold into a right-handed four-stranded non-canonical conformation known as G-quadruplexes (G4s). G-quadruplex structures form through the  $\pi$ - $\pi$  stacking of two or more G-tetrads. G-tetrads have a square planar arrangement that forms from the Hoogsteen H-bonding between four guanine (G) bases (Figure 1a) (1,2). The stacked G-tetrads are stabilized by the presence of cations that are coordinated to the free lone pair electrons on the carbonyl groups of the H-bonded G bases, that would otherwise create unstable electrostatic repulsion forces at the center of the structure (3,4).

The stabilizing role that cations play on G-tetrads depend on the identity and size of the ions, where a preference has been observed for monovalent alkali metal cations in aqueous solutions in the order of  $K^+ > Na^+ \gg Li^+$  (4–6). Depending on the size of the cations, they can be located either in the plane of the G-tetrads where they form coordination bonds with the four guanine bases (Fig. 1a), or in between G-tetrads where the ions interact with bases from the planes above and below creating a coordination number of 8 (Fig. 1b).  $Na^+$  ions have been observed to be located in the plane of the G-tetrad as they have the proper size to fit the central cavity, while  $K^+$  ions are too large in size and can only fit in the inter-plane space between the tetrads (3,7). Ions like  $K^+$  that octa-coordinate with the G-tetrads are observed to better stabilize G-quadruplexes in aqueous solutions compared to  $Na^+$ , due to the lower energy required to dissociate the outer shell of hydration (8,9).  $Li^+$  is smaller in size compared to  $Na^+$  and is more strongly hydrated, resulting in an even greater energy penalty for binding to G-quadruplexes. Similarly,  $Li^+$  is too small in size

to properly fit in the central cavity of the G-tetrads making it incapable of creating coordination bonds with all the G bases (10).

G-quadruplex structures are characterized by having a high degree of structural polymorphism that depends on an array of factors such as: nucleotide composition and length of the DNA sequence that forms this structure, the orientation of the strands, the glycosidic bond conformation of the guanine bases, the loop connections, and the environmental factors such as the type of cations present, molecular crowding, and hydration state (7,11–14). The basic sequence motif for forming G-quadruplexes is  $G_{\geq 2}N_XG_{\geq 2}N_XG_{\geq 2}N_XG_{\geq 2}N_X$ , here N is any nucleotide that resides in the loops that connects the G-tracks, while the G-tracks (a minimum of 2 Gs) form the G-tetrads through Hoogsteen bonding of the guanine bases (Fig.1)(15).

If the four strands that form the G-quadruplex structure have the same orientation, they create a parallel type of structure that is connected through propeller loops (Fig. 1a)(16,17). If the strands have opposing orientations, an anti-parallel type of structure forms where a higher diversity of loop combinations are observed. Antiparallel strands can be connected by two types of loops, lateral (edgewise) and diagonal (central), that are found in the chair and basket types of structures (Fig, 1c) (7,18). Another antiparallel type of G-quadruplex structure that is observed is known as the hybrid type. This structure has a mix of parallel/antiparallel strands that are connected by lateral and propeller loops, where two interchangeable conformations form what is known as the hybrid-1 and hybrid-2 G-quadruplexes (Fig. 1c). These two structures differ from one another by the way the loops are ordered, where hybrid-1 structure has the propeller loop closest to the 5'-end, while

the hybrid-2 structure has the lateral loop by the 5'-end and the propeller loop by the 3'-end of the structure (19–21).

The glycosidic bond angle of the guanine bases in the G-tetrads adopt either the *anti* or *syn* glycosyl conformation depending on the relative orientation of the strands that make up the G-quadruplex. For the antiparallel and hybrid type of structures, the G bases are observed to alternate between the *syn* and *anti*-glycosidic bond angles (7,18,22), while the guanine bases in the parallel type of structure exclusively adopt the *anti*-glycosyl conformation (17,23). As the *anti*-glycosyl orientation of DNA bases are more stable than the *syn* conformation which causes steric clashes between the base and the deoxyribose sugar, the parallel conformation of G-quadruplex structures are more energetically favorable and stable compared to the antiparallel conformations (13).

Depending on the DNA sequence and the folding environment, multiple folded G-quadruplex conformations of similar energies are found to be in equilibrium with one another resulting in a highly dynamic folded structural aggregate (24). G-quadruplex structures have been observed to fold by diverse mechanisms, where some G-quadruplexes fold by a two-state pathway (25), while others proceed through many well-populated folding intermediates (26). Which folding pathway is followed depends on different factors such as: the identity of the cation present, the sequence and length of the loops, the number of guanines present in the G-tracts, and the 5'- and/ or 3'-flanking sequences that can interact and have a stabilizing stacking effect on the G-quadruplex. All these factors have an impact as to which fold a sequence will adopt, making this particular non-B form of DNA extremely polymorphic in nature (25,27,28).

In addition to folding intramolecularly, G-quadruplexes have been observed to also form intermolecularly (Figure 1d). Intramolecular G-Quadruplexes form when a strand contains all the necessary sequence motifs for the formation of the non-canonical structures. Depending on the concentration of the strands, as well as the environment and the sequence of the strands, two to four strands have been observed to associate with one another forming intermolecular G-quadruplexes (29,30). The highly diverse polymorphic nature of G-quadruplexes is believed to be used by cells for an array of cellular processes, leading to many different hypotheses for the roles that these structures could be playing *in vivo*. The problem is that to this day there is very limited data for the existence of these structures *in vivo*.

## **1.2 Biological Relevance of G-Quadruplex Structures**

The regulatory roles that G-quadruplex structures are proposed to play *in vivo* involve critical cellular processes where the DNA transiently adopts a single stranded conformation such as in transcription and replication where the helical DNA becomes partially unwound. G-quadruplexes are also proposed to arise in telomeres and messenger RNAs where the DNA and RNA are single-stranded (Fig. 2). G-quadruplexes are therefore suspected to be involved in many critical cellular roles where its exact role is still very much questioned due to the very limited set of probes that are currently available for their study (31,32).

### **1.2.1 G-Quadruplexes in Telomeres**

Telomeres are specialized DNA-protein complexes that are found at the natural ends of chromosomes. In vertebrates, telomeric DNA consists of tandem repeats of the 5'-(TTAGGG)<sub>n</sub>-3' sequence that ends in a 20-600 base long single-stranded 3'-overhang. Telomeres evolved as a

solution to the “end replication problem” that is caused by the 5' → 3' directionality of DNA polymerase synthesis and the inability of DNA polymerases to initiate DNA synthesis at the 3'-end of DNA without a primer. In DNA replication, an RNA primase produces an RNA primer that is eventually degraded, thereby leading to the consecutive shortening of the DNA with each round of DNA replication. To solve this problem the reverse transcriptase enzyme telomerase uses an RNA template to direct the elongation of the 3'-end of the DNA with a repeating sequence (telomeric DNA), that prevents loss of the complementary sequence (22,33).

The first biologically relevant sequences to be detected to form G-quadruplexes *in vitro* were the highly conserved G-rich sequences of telomeric DNA (34), suggesting that these structures might play an important role in telomere biology (Fig. 2a). *In vitro* telomerase has been shown to recognize and unwind the parallel G-quadruplex structures, and then proceed to extend the telomere end (35). On the other hand, when telomerase encounters an anti-parallel G-quadruplexes, however, they do not seem to be good binding substrates for telomerase as it is not as efficient at extending them as parallel quadruplexes (36,37). The extension of telomeres in the presence of anti-parallel G-quadruplex structures seems to depend on proteins like POT1, which are single-stranded DNA binding proteins that help unwind intramolecular G-quadruplexes and trap the linear form of DNA making it accessible for its extension (Fig. 3a) (38). The effects that parallel and antiparallel G-quadruplex structures have on telomerase *in vitro* seems to suggest that different conformations of G-quadruplexes may play different biological roles in the cell (39).

The formation of G-quadruplex structures in telomeric regions is also suspected to play a capping role where they might protect the end of chromosomes ensuring its stability and maintenance (Fig.

3b) (40). The formation of these complex structures in the single-stranded portions of telomeres are believed to prevent the degradation of these DNA regions by nucleases making sure their integrity are conserved (41). Similarly, G-quadruplexes are also believed to protect telomeres by preventing the access of incorrectly activated DNA repair machinery that may proceed to process the end of chromosomes as damaged DNA double strand breaks (DSB) (39,40,42). However, whether G-quadruplexes play these roles in telomeres, or even if they form in these regions, still remains to be established.

Various studies have shown that G-quadruplex conformations depend on the length and number of G-runs in the sequence, the counterions present, and the flanking sequences. In solutions of  $K^+$  which is the major intracellular cation, telomeric sequences have been detected to adopt parallel, basket, and the two different types of hybrid G-quadruplex structures (Fig. 1b) (20,21,43,44). In the presence of  $Na^+$ , however, basket (23) and the two hybrid (45,46) antiparallel structures have been observed to form. The sequences immediately adjacent to the 5' and 3'-ends of human telomeric sequences greatly influence the relative stability of antiparallel hybrid type-1 and type-2 which are in equilibrium with each other in  $K^+$  solution by specific capping interactions. For the hybrid type-2 structure, the 3'-flanking sequence is known to create base pairing interactions between a T base found in the 3'-end and TA bases found in the lateral loop 1, while for the hybrid type-1 structure capping interactions are known to form between bases found in both 5'- and 3'-end sequences and bases found in all three loops. The energy barrier between these two hybrid forms of G-quadruplexes seem to be low allowing them to readily shift by minor changes in their conformation (19,44,47). In a recent study, single-molecule fluorescence-force microscopy experiments have detected up to six different DNA structures made up of very stable G-quadruplex



conformations and some folding intermediates in the presence of  $K^+$ , showing the dynamic nature of human telomeric sequences (48). It has been determined that G-quadruplexes fold through the formation of folding intermediates such as G-triplex structures that form through the stacking of three G-tracks and make up a significant amount of the subpopulation of DNA structures that can form in telomeric sequences (26). Therefore, the structural biology of G-quadruplexes in telomeric DNA *in vivo* may be very complicated and difficult to elucidate.

### 1.2.2 G-Quadruplexes in Gene Promoters

Computational searches of the human genome for the putative G-quadruplex forming sequence (PQS) motif  $G_{\geq 3}N_{1-7}G_{\geq 3}N_{1-7}G_{\geq 3}N_{1-7}G_{\geq 3}$  have discovered that over 375,000 potential sites for G-quadruplex formation are distributed in a non-random way throughout the genome (12,47). They found that PQS's are highly predominant in human gene promoters, especially near transcription start sites (TSS). They found that 42.7% of all the known human genes contain at least one PQS within the one kb region upstream the transcription start site (TSS), providing evidence that these structures could potentially play a critical role during gene expression (49). The PQS's were observed to be particularly predominant in genes that are connected to cancer, as well as to genes involved in gene regulation (i.e., transcription factors). Similarly, almost 66% of the nuclease hypersensitive sites, which are regions of the genome associated with open forms of chromatin, were found to have high frequency of PQS's. Contrastingly, genes that are connected to tumor suppression and to protein biosynthesis were found to have lower frequencies of PQS's. In the regions of the genome that are particularly enriched with PQS's and are strongly associated with gene expression, G-quadruplex structures are believed to have the ability to act as a topological switch to control transcription (Fig. 4) (49,50).

The potential role that G-quadruplexes may have on transcription has been thoroughly studied *in vitro* with genes that have been linked to cancer, as well as to other diseases, that have PQS's in their promoter regions. A correlation has been found between PQS's, transcription initiation, and the development of certain diseases by studying genes such as c-MYC, BCL-2, KRAS, c-KIT, and VEGF. Depending on which DNA strand (template or non-template) the PQS is in, as well as its location with respect to the TSS, G-quadruplexes have been linked to have either a positive or negative effect in the transcription process (Fig. 4) (31,51).

In the c-MYC oncogene, which was the first gene that was described to have a PQS in its promoter region, G-quadruplexes have been linked to suppressing its expression. Because the overexpression of this gene has been determined to be the dominant carcinogenic driver for up to 70% of cancers, G-quadruplexes have become an attractive target for cancer drug development (41). c-MYC has a nuclease hypersensitive element III (NHE III) that is downstream from its promoter region that controls >80% of the transcription and contains a 46 bp PQS (PU46) that is located in the template strand (52,53). This sequence contains six G-runs that can form a basket type with three G-tetrads, a chair comprised of two G-tetrads, and parallel type with three G-tetrads in  $K^+$  depending on which G-tracts are involved (53–55). The effect that the different G-quadruplex topologies have on the transcription rate of c-MYC has been studied *in vitro*, where it is believed that the parallel conformation has the highest inhibitory effect (55). The interaction of transcription factors (TF) with the PU46 G-quadruplex is believed to be one of the many factors that contributes to the negative effect that G-quadruplexes have on the transcription efficiency of the c-MYC gene (41). For example, when TF nucleoside diphosphate kinase (NM23-H2) binds to the PU46 region it stabilizes the single-stranded form of DNA, causing the unfolding of any G-quadruplex

structures that are present in this region leading to transcription activation. On the other hand, when the transcription factor known as nucleolin binds the G-quadruplex structures it prevents transcription initiation (Fig. 4a). Nucleolin is also known to prevent the binding of another TF protein known as SP1 which promotes transcription initiation by binding either duplex DNA or other G-quadruplex conformations, where SP1 binding is known to activate the recruitment of transcription initiation factors (41,56,57).

Other modes of action that have been attributed to G-quadruplex structures during transcription, involves the presence of PQS's downstream the transcription start site. During transcription the formation of a G-quadruplex in the template strand can prevent the progression of RNA polymerase resulting in its obstruction and causing transcription termination (Fig. 4d) (58). If the PQS is in the non-template strand, however, G-quadruplex formation can have a positive effect in transcription as it can stabilize the R-loop (RNA/DNA hybrid bubble) allowing for better access of RNA polymerase to the template strand facilitating initiation and progression of transcription (Fig. 4c)(59,60).

In addition to the high abundance of G-quadruplex forming sequences in promoter regions, it is believed that these structures are prone to form during transcription due the dynamic perturbation that occurs in the DNA double helix during this stage in gene expression. Specifically, it is believed that the negative supercoiling introduced by the RNA polymerase machinery on the duplex DNA, even after the action of topoisomerases, can lead to DNA unwinding. Negative supercoiling is produced after RNA polymerase passes a region of DNA allowing for sequences upstream, such as promoter sequences, to unwind allowing PQS's to form G-quadruplexes (51,61,62).

### 1.2.3 G-Quadruplexes in Translation

Computational studies have also shown a prevalence of PQS's in the 5'- and 3'-untranslated regions (UTR) of protein encoding genes where G-quadruplexes are believed to play a regulatory role in translation (Fig. 2d). Most of the post-transcriptional controls in cells are known to happen in the 5'- and 3'-UTR regions of mRNAs, where computational searches have found that 12.6% of genes have at least one PQS at the 5'-UTR, while 15.3% of genes have these PQS's at the 3'-UTR. When looking at the distribution of PQS's in the 5'UTR of mRNAs, a high frequency of G4-forming sequences were found towards the beginning of the gene, where a decrease in their frequency was observed in the regions further into gene, suggesting that these structures could be playing a role in translational regulation, either by directly interacting with the translation machinery or by being involved with the proteins in charge of recruiting the ribosome complex (63,64).

Another region of the gene that was found to have high frequencies of PQS's was the downstream sequence of the TSS. It was found that within the first 100 bp of the first introns of genes, 48% of RNA transcripts contained a PQS that decreased in frequency within the introns found further into the gene, suggesting that these structures may be involved in the splicing of the pre-mRNA molecules (65).

The high frequency of PQS's found in the 3'-UTR of the mRNA's were found to be slightly concentrated near the transcription end sites (TES) and the 3'-downstream regions of the mRNAs. These regions of the gene have been associated with the termination of transcription, especially when another gene is located immediately adjacent to it, leading to the proposal that G-quadruplexes may act as a transcription termination element. It is speculated that when there are

back-to-back genes, G-quadruplexes may be playing a special role preventing harmful run-throughs during transcription of these genes (64). Similarly, the presence of PQS's at the 3'-UTR of mRNAs have suggested that these structures are involved in alternative polyadenylation processes, leading to the shortening of some of these mRNA molecules ultimately regulating their expression (66,67).

Up to now all the hypotheses that have been proposed for how G-quadruplex structures could be involved in modulating the transcription and translation of genes (Fig. 2), especially protooncogenes, have resulted from studies done *in silico*, *in vitro*, and only a few that have been done *in cellulo*. These proposals have made G-quadruplexes a very attractive target for drug development purposes, especially for cancer (31,41), but without methods for detecting G-quadruplex formation *in vivo*, this has become a very speculative endeavor.

#### **1.2.4 G-Quadruplexes in Replication**

During DNA replication, the antiparallel strands of helical DNA are replicated in different manners due to the unique way DNA polymerase functions. DNA polymerases are known to only read a DNA strand from the 3'→5' direction and to synthesize a new strand from the 5'→3' direction. This causes for one strand of DNA to be replicated in a continuous way (leading strand) while the other strand is replicated in a discontinuous way (lagging strand). During the discontinuous replication of the lagging strand, the DNA template becomes transiently single-stranded providing an opportunity for non-B forms of DNA to form, making the replication process another cellular event where G-quadruplex formation could be playing an influential role (Fig. 2b)(68).

In humans and other eukaryotic organisms, thousands of bidirectional replication origin sites have been identified to occur every 100 kb. PQS's containing loop sequences with up to 7 nucleotides were found in approximately 67% of these sites, while 91.4% were found to contain a PQS with up to 15 nucleotides in its loops. The formation of G-quadruplexes at these replication origin sites have been suggested to help with replication initiation by facilitating duplex DNA unwinding and by helping with the origin replication complex (ORC) (69,70). The human ORC has actually been shown to have a higher binding affinity for single-stranded DNA that contains PQS's compared to other sequence motifs of single-stranded or helical forms of DNA (71).

Similarly, other proteins involved in the replication process have also been shown to preferentially associate with PQS's (69). One class of such proteins are DNA helicases, such as the mammalian Pif1 and FANCI, that have been found to contain G-quadruplex unwinding activity which is an activity specific to many eukaryotic DNA helicases involved in replication. In humans, the loss of the unwinding activity of many of these helicases that targets G-quadruplex structures have been linked to diseases that cause genomic instability (72,73). This genomic instability has been explained by the possible failure of a DNA polymerase to continue with replication of DNA as it stalls once it encounters a G-quadruplex structure forcing replication to be terminated (73,74).

### **1.3 Methods for G-Quadruplex Detection in the Human Genome**

The majority of evidence available for the role that G-quadruplex structures play in the cell, have been obtained through *in vitro* studies that use favorable G-quadruplex forming conditions and synthetic DNA sequences. The certainty that G-quadruplexes form in human genomic DNA within cells is still very much questioned. To date, a number of enzymatic, antibody, and small molecule

probes have been developed for *in vivo* detection of G-quadruplexes (75). The problem with these techniques is that they only provide indirect evidence of the presence of G-quadruplexes *in vivo* as they either rely on enzymatic activity that could also detect other DNA structures, or antibody and small molecule binding probes that can drive the formation of the structures that they are supposed to detect. Implementation of some of these techniques involve procedures that disrupt the cellular environment and could alter the very delicate conformational dynamics that is characteristic of G-quadruplex formation.

### **1.3.1 Sequencing of G-Quadruplex formation in the Human Genome**

One of the earliest techniques that was developed to detect and locate the formation of G-quadruplex structures in human genomic DNA is the G4-dependent DNA polymerase stalling and next-generation sequencing technique (G4-seq) (Fig 5). This technique makes use of the ability of G-quadruplex structures to arrest the progression of DNA polymerase on isolated genomic DNA (76). Purified human genomic DNA is first sequenced under salt and ligand conditions that stabilize G-quadruplex structures and compared to sequencing done in the presence of  $\text{Li}^+$  that does not stabilize G-quadruplex structures (Fig. 5a) (77). Using this G4-seq method, 716,310 sites were determined to form G-quadruplex structures *in vitro*, of which 70% would not have been predicted to form based on the standard PQS motif. This result indicates that G4-seq is also detecting non-canonical DNA G-quadruplex structures that were believed to be too unstable to form in human genomic DNA. Of these non-canonical structures 21.6% contained extra-long loops with more than 7 nucleotides, while 24% contained G-runs that had single-nucleotide disruptions that would produce bulges between the stacked G-tetrads (Fig. 5B) (77).

The highest frequency of PQS's found by G4-seq were at the 5'- UTRs and splicing sites of mRNAs, supporting the hypothesis that G-quadruplex structures may be playing a role in gene expression during post-transcriptional regulation. Many of the non-canonical G-quadruplex structures detected by G4-seq were found within genes that contained little to no canonical G-quadruplex forming sequences, some of which included many important cancer genes. A significant enrichment of G-quadruplex structures were also observed in somatic copy number alterations, traits that are signatures for cancer development, adding to the interest of targeting these structures for new cancer treatments (77,78).

The frequency and location of the G-quadruplex forming sites detected in the human genome by G4-seq were highly dependent on the stabilizing conditions used to promote G-quadruplex formation in the polymerase stop assays. These stabilizing conditions consisted of using either  $K^+$  or G-quadruplex ligands such as pyridostatin (PDS) (Fig. 5a), which is a planar aromatic cationic compound that thermodynamically stabilizes G-quadruplex structures by stacking atop the G-tetrads (79). When PDS stabilizing conditions were used, almost 200,000 more sites were detected compared to  $K^+$  alone, showing that the sequences detected by G4-seq actually reflect the binding and stabilizing properties that these molecules have on G-quadruplex structures, and not that these structures form significantly in the genome (77). From all the sites that were detected by G4-seq, about half of the overall sites were found in common between  $K^+$  and PDS, suggesting that there are sequences in the human genome that are capable of forming G-quadruplex structures, independent of the stabilizing conditions, and that they may be playing biologically relevant roles in the cell environment (77).



Another method that was developed to map the locations in the human genome where G-quadruplex structures can form *in vivo* was by combining permanganate footprinting with high-throughput sequencing. Potassium permanganate (KMnO<sub>4</sub>) is known for oxidizing unpaired nucleotide bases in regions of DNA that become single stranded thereby trapping the DNA in this state and making it sensitive to the single-stranded specific S1 nuclease enzyme allowing the conformational state of DNA inside a cell to be probed (80,81). Digestion of genomic DNA with the S1 enzyme produces double-strand breaks at the site of the chemical modification, that when sequenced and compared to PQS sites predicted by computational studies, provides a high-resolution map of the potential sites in the genome where G-quadruplexes can form (80,81). This genome-wide -dependent nuclease footprinting analysis identified 20,000 sites that contained PQS supporting the hypothesis that these structures may play a role during transcription as their formation might be dependent on the state of the chromatin that is regulated by the transcription state of genes (80). But even though KMnO<sub>4</sub> is a molecule that can easily diffuse into the cell without the need to chemically fix the cells and by itself does not drive formation of G-quadruplexes, additional methods are required to demonstrate that G-quadruplexes do form in these sites in the genome.

### **1.3.2 Visualization of G-Quadruplexes with Fluorescence Microscopy**

Under intracellular conditions, a variety of DNA sequences have been shown to adopt thermodynamically stable G-quadruplex conformations *in vitro* that might be expected to form *in vivo*, yet these structures have been very challenging to detect in cells due to their very transient nature. To detect G-quadruplexes *in vivo*, fluorescent antibodies and small molecules have been

developed, but due to their high binding affinity, these probes may actually drive the formation of G-quadruplexes and actively disrupt the structural dynamics of DNA (82).

Antibodies have been engineered to have high selectivity and nanomolar affinity to both intra- and intermolecular G-quadruplexes, as well as to different G-quadruplex topologies. To image where they bind in cells, fluorescently labeled tertiary antibodies have been used (Fig. 6a). When analyzed at the level of individual chromosomes, ~75% of the fluorescent foci were found outside of telomeres while the rest were found within telomeric regions (Fig. 6b). The number of G-quadruplex structures detected was found to depend highly on the cell cycle where the highest quantity of G-quadruplexes were observed during the S phase when DNA replication occurs (83).

Addition of small molecules that bind and stabilize G-quadruplexes like pyridostatin to cells in the presence of the fluorescently labeled antibodies, causes an increase in the number of G-quadruplex structures observed (Fig. 6c) (83,84). The increase in the number of fluorescent foci clearly demonstrates the ability of G-quadruplex binders to drive the formation of G-quadruplex structures in live cells, thereby calling into question their usefulness as probes of existing G-quadruplex structures *in vivo*. Furthermore, the antibody method used requires cells to be fixed before the antibodies are introduced into the cells. As cell fixation requires the use of chemicals, such as ethanol or paraformaldehyde that actively disrupt the cellular environment, this method only shows that genomic DNA has the ability to fold into G-quadruplex structures but not that these structures are actually present *in vivo* (82,85).

Another approach to image G-quadruplex structures *in vivo*, is to use fluorescently small molecules whose fluorescence properties change upon binding G-quadruplexes. These molecules are

generally positively charged and have high selectivity for G-quadruplex DNA due to their ability to stack on top of the G-tetrads, but are bulky enough to prevent their intercalation into duplex DNA (Fig. 6d) (85,86). Because these molecules are small, they can easily diffuse into cells thereby eliminating the need to fix them which enables the dynamics of G-quadruplex structures to be monitored in live cells (85). Intrinsically fluorescent small molecules like DAOTA which are planar aromatic triarylmethyl carbocations probes (Fig. 6d), have provided additional evidence that G-quadruplex formation is occurring when the cell is in the S-phase, confirming what was initially observed with fluorescently labelled antibodies (83,86).

Imaging of live cells with fluorescent small molecule probes usually requires the use of high concentrations ( $\mu\text{M}$ ) that could perturb the cell and alter the dynamics of G-quadruplex formation. To minimize this problem, small fluorescent molecules like SiR-PyPDS were developed where the G-quadruplex ligand PDS was modified by attaching the red fluorophore Silicon-Rhodamine (SiR) (Fig. 6d), that allows for the detection of G-quadruplex structures in living cells at much lower probe concentrations ( $\text{nM}$ ) (87,88). With these single molecule probes, it was estimated that about 3000 G-quadruplex structures form in a single living cell instead of the hundreds of thousands that were estimated by the computational and sequencing studies (87). The dynamics of G-quadruplex formation was also studied using this approach by chemically trapping the unfolded conformations of G-quadruplex DNA by methylating the exposed N7 atoms of the G bases that are usually protected by Hoogsteen base pairing in G-quadruplexes with dimethyl-sulfate (DMS). By irreversibly trapping these bases preventing G-quadruplexes from re-folding it was observed that the conformational structures of DNA in a cell fluctuates between the folded and unfolded states. This was in agreement with other studies that related G-quadruplex formation with the cell-

cycle phase, as well as with biological processes that involve the active transcription and replication of DNA (87,88).

As it is still unclear if the use of these small molecules perturbs the cell and/or alters the dynamic of G-quadruplex formation, additional probes have been developed to minimize the need to use molecules with such high levels of affinity for these structures. Recently a new methodology has been established consisting of G-quadruplex binders that irreversibly trap these structures with stable covalent linkages (89–91). This technique makes use of small G-quadruplex binders that are attached to photoreactive groups that upon activation with light (UVA or visible) chemically react with G-quadruplexes. They have shown high selectivity towards G-quadruplex DNA compared to duplex DNA where these molecules photo-crosslink to the bases found in the loops, specifically with thymine bases, or with the G-quartets depending on the linking agent and on the topology of the G-quadruplex structure. So far this method shows the advantage of requiring light to trigger the detection of G-quadruplexes which can easily penetrate cells without disrupting the cell environment and has been found to be successful and reliable *in vitro*, but no studies have been done so far *in vivo* (89,90).

The development of the fluorescent antibodies and small molecules described above have allowed for the imaging of G-quadruplex structures in live cells. But due to the methodology utilized in these techniques, it still remains unclear if the G-quadruplexes that are being detected were originally present or were induced to form by the probes. While these fluorescent probes have provided evidence that the formation of G-quadruplex structures is dynamic and sensitive to the cell cycle, their biological significance still remains poorly understood. Therefore, it is imperative

to develop methods that provides additional proof that these structures form in the cell without disrupting or altering the dynamics of DNA structures being probed.

### **1.3.3 G-Quadruplex Chromatin Immunoprecipitation Sequencing**

The development of antibodies that are highly specific to G-quadruplex structures have also allowed the study of the relationship between G-quadruplexes, chromatin state, and transcription in human cells. To map the location of G-quadruplex structures in the human genome within chromatin context, a G-quadruplex chromatin immuno-precipitation followed by sequencing (G4 ChIP-seq) has been developed (Fig. 7a). Using this technique only 1% (~10,000) of the G-quadruplex structures identified by G4-seq were detected within chromatin, suggesting that the formation of these structures may be suppressed by the presence of nucleosomes (92,93), or that the G-quadruplex specific antibodies used in this technique are only detecting a subset of the G-quadruplex structures that are forming in the cell.

The study of G-quadruplex frequency in the human genome through G4 ChIP-seq experiments have shown that the conformation of DNA highly depends on the presence of nucleosomes. The highest number of G-quadruplexes were detected in regions of the genome that were depleted in nucleosomes, as well as in promoters and 5'-UTR regions of genes that show high levels of transcription (92,94). This suggests that there is a positive and dynamic relationship between G-quadruplexes found in promoter regions of genes that are associated with high transcription activity in nucleosome depleted regions of the human genome (Fig. 7b) (92–94). Similarly, the folding dynamics of G-quadruplex structures in cells are believed to aid in defining cell-specific chromatin organization in different cell lines, as well as to defining chromatin markers that are

associated with cellular differentiation (94,95). The strong relationship between G-quadruplexes and the state of chromatin in the human genome observed by G4 ChIP-seq led to the proposal that G-quadruplexes could be altering the recruitment of transcription factors and/or chromatin remodeling proteins involved in defining chromatin architecture. Also, the formation of G-quadruplex structures in DNA is believed to act as a mechanism to control transcription state in active genes influencing the nucleosome positioning and the state of chromatin (31,94).

The limitation of the G4 ChIP-seq technique (93) is that it requires cells to be fixed before the DNA fragments containing G-quadruplexes are precipitated with antibodies (Fig. 7a). Therefore, it still remains to be determined whether the G-quadruplexes being detected are actually present in the live cell or whether their formation is being driven by the high binding affinity of the antibodies and/or the chemical alterations produced by the chromatin fixation process. Consequently, there is a great need for the development of experimental methods for the detection and location of G-quadruplex structures in the cell that do not drive their formation or disrupt the cellular environment, and that can instantaneously react with G-quadruplexes when they are present.

## **1.4 Photoproducts as Intrinsic Probes for Non-B forms of DNA**

It was previously discovered by our group that G-quadruplex forming sequences produce unique, non-adjacent dipyrimidine photoproducts when irradiated with UVB light (96–99) (Fig. 8b). These non-adjacent photoproducts are different from the adjacent photoproducts formed in B form DNA, and only form when distant bases from non-B conformations of DNA are brought into close proximity (Fig. 8b)(100–103). The formation of non-adjacent photoproducts in non-B forms of

DNA such as G-quadruplexes, suggests that UV light could be used to irreversibly trap certain types of G-quadruplexes through covalent bond formation *in vivo* that could then be subsequently detected and located in the human genome. The detection of these unique photoproducts *in vivo* would provide irrefutable evidence that non-B conformations of DNA, like G-quadruplexes, are present in the cell. To understand the photochemistry of G-quadruplexes and how it could be used as an intrinsic probe of these structures, we will first review the photochemistry of B DNA.

### 1.4.1 UV-Induced DNA Photoproducts

The nucleic acid bases (A, G, T, C, U) are characterized by having main absorption peaks within the UVC (200-280 nm) range of the UV spectrum, where a slight absorption band is observed to extend to the UVB (280-315 nm) region. The absorption of UVB or UVC light is known to lead to the direct excitation of DNA that causes the photoexcitation of the pyrimidine nucleotides (T, C), that when properly oriented, lead to the formation of the vast majority of observed DNA photoproducts (104).

The major UV photoproduct of DNA is the cyclobutane pyrimidine dimer (CPD) that results from the [2+2] cycloaddition reaction between the C5=C6 double bonds of pyrimidine bases (Fig. 8a). The CPD that forms from the photodimerization of two thymine bases (T=T) is the most abundant photoproduct of duplex DNA, followed by T=C, C=T, and C=C CPDs. These photoproducts are known to have a residual absorption in the UVC range with a maximum absorption at 230 nm, where exposure to these wavelengths leads to the photoreversal of the initial photoaddition reaction with a quantum yield that approaches 1 (105,106).

UV irradiation of duplex DNA also leads to the formation of the lesion known as pyrimidine (6-4) pyrimidone (64PP) photoproduct named from the single bond linkage between the C6 atom of the 5'-base and the C4 atom of the 3'-base (Fig. 8a). The 64PP photoproduct is the second most abundant photolesion observed in duplex DNA where it is formed with a quantum yield  $\phi$  of 0.01 which is an order of magnitude lower than that for CPDs ( $\phi = 0.1$ ). The (6-4) photoproduct is characterized by having a pyrimidone moiety that has an absorption maximum at about 315 nm in the UVA range (315-400 nm). UVA irradiation of the 64PP triggers a photocyclization reaction that produces its Dewar valence isomer with a quantum yield of 0.01 (107), and so named as the Dewar photoproduct (104,108,109).

The CPD, 64PP, and Dewar photoproducts are the main photoproducts in the duplex B-form arrangement of DNA and are therefore produced in the majority of DNA life forms exposed to sunlight. The frequency of formation of these various photoproducts depends on the flanking DNA sequence, DNA bending, and the binding of transcription factors, nucleosomes, and other proteins. As a result DNA photoproducts can report on DNA conformation, as well as DNA interactions and has been used to map nucleosome and transcription factor binding by a technique known as photofootprinting (110–112).

Another type of DNA photoproduct has been observed to form in bacterial spores and is aptly named as the spore photoproduct (SP) (Fig. 8a). This photoproduct is exclusively found to form in bacterial spores due to a combination of low hydration levels that stabilizes the A-form of DNA, the presence of pyridine-1,6-dicarboxylic acid (DPA), and the binding of small, acid-soluble proteins (SASPs) to the A-form DNA. The spore photoproduct results from a two-step reaction



involving radical abstraction of an H from the C5 methyl of the 3'-thymidine by the C6 of a 5'-thymidine followed by radical recombination to give a bond between the C5 of the 5'-thymidine with the methylene radical of the 3'-thymidine (113).

#### 1.4.2 Stereochemistry of Cyclobutane Pyrimidine Dimers (CPDs)

CPD photoproducts contain a four-membered cyclobutane ring that connects two pyrimidine bases that are in close proximity to one another. These photoproducts can form several isomers that depend on the conformational constraints of DNA at the moment of photocycloaddition. There are eight possible T=T CPDs that can form in a duplex DNA but only six Thy=Thy CPDs can form between two free thymidines nucleotides (Fig. 9). If the eight T=T CPDs are released from their sugar phosphate DNA backbone by acid hydrolysis as Thy=Thy CPDs, only six of these Thy=Thy isomers would be observed, four of which exist as pairs of enantiomers (102,111).

In the B-form conformation of DNA adjacent nucleotides stack on top of each other in the same orientation as dictated by Watson-Crick base pairing. In this conformation CPDs exclusively form between adjacent nucleotides with the *cis, syn* stereochemistry in which the bases adopt a head-to-head orientation (*syn*) and the C5 substituents (methyl or H) are oriented towards the same side of the cyclobutane ring (*cis*) (Fig. 9a) (105,114). When DNA adopts any other type of secondary or tertiary structure, such as a hairpin, G-quadruplexes, cruciforms, or H-DNA, the CPD isomers that are produced will depend on the relative orientation of the two pyrimidines in the structure undergoing the cycloaddition reaction something to which has not been well studied to date (115).

Non-B conformations of DNA are known to induce the formation of both adjacent and non-adjacent CPDs with different regiochemistries and stereochemistries. CPDs can form with head-

to-head (*syn*) or head-to-tail (*anti*) regiochemistry, and with the C5 substituents pointing towards the same (*cis*) or different (*trans*) sides of the cyclobutane ring resulting in both adjacent and non-adjacent CPDs with *cis,syn*, *trans,syn*, *cis,anti*, and *trans,anti* stereochemistries (Fig. 9) (106,115)

When helical DNA is in its single stranded form, adjacent T=T CPDs with the *trans,syn* stereochemistry have been observed to form, in addition to adjacent *cis,syn* CPDs. This is due to the DNA bases no longer being constrained into the *anti* glycosyl conformation. When one of the two photoreacting pyrimidines is the *syn* glycosyl conformation and the other is in the *anti* conformation a *trans,syn* stereoisomer forms explaining why this photoproduct is not observed in B-form DNA (105,115,116).

The formation of nonadjacent CPDs with the *cis,syn* arrangement have also been reported to form within a strand from a helical B-form DNA when nucleotides between pyrimidine bases become extrahelical, forming a bulge loop. In this situation, the nonadjacent pyrimidine bases are allowed to stack on each other to form a non-adjacent *cis,syn* CPD upon UV irradiation (Fig. 8c) (115,117). The formation of non-adjacent *cis,anti* CPDs are also believed to form between strands of helical DNA between bases of opposing strands (118).

The formation of nonadjacent CPDs with the *trans,syn* stereochemistry have only been detected when DNA under dehydrating non-native conditions has been irradiated with UVC light, where the structure adopted by the DNA leading to its formation is as yet unknown (119). Non-adjacent *cis,syn*, *cis,anti*, and *trans,anti* T=T CPDs were also formed under these dehydrating conditions in significant amounts (16-21% each) compared to the adjacent *cis,syn* T=T CPD (25%), where in

comparison the *trans,syn* CPD formation seemed to be suppressed for both adjacent (4%) and non-adjacent (1%) photoproducts. (115,119)

### 1.4.3 Anti-CPD Formation in Single Stranded ODN at Low pH

The first thoroughly characterized non-adjacent *cis,anti* T=T CPD of DNA was discovered by accident following irradiation of a single strand of DNA with UVB light under acidic conditions (Fig. 10a) (118). The *cis,anti* CPD was detected to form in 40% yield between T2 and T7 in the 14mer d(GTATCATGAGGTGC) at a pH of 5. When the pH was increased to 7, the photoproduct was no longer detected suggesting that at low pH some unusual secondary structure was forming.

The structure of the T=T *cis,anti*-CPD was established by various enzyme-coupled MS assays along with 2D NMR analysis and correlation with authentic degradation products. The T2 and T7 bases were determined to form the *anti* CPD by making use of the 5'-3' exonuclease bovine spleen phosphodiesterase (BSP) that stalled at the T2 base when proceeding from the 5'-end, while the 3'-5' exonuclease snake venom phosphodiesterase (SVP) stalled once it encountered the T7 base from the 3'-end (Fig. 10a) (118). To confirm that these two bases were involved in the formation of the *anti* CPD photoproduct, they made use of the endonuclease known as nuclease P1 (NP1) that cleaves at the 3'-side of canonical nucleotides (120). When NP1 encounters a damaged nucleotide, such as a CPD containing thymidine, it is incapable of cleaving at the 3'-end of either thymidine of the CPD (Fig. 10b). As a result, NP1 produced the tetranucleotide pT(pA)=pTpG confirming that the two nonadjacent bases T2 and T7 were involved in the formation of the *cis,anti*-CPD (118).

The stereochemistry of the non-adjacent CPD was determined to be *cis,anti* by comparison of the thymine dimer released by acid catalyzed hydrolysis of the N-glycosidic bonds with HF/pyridine with an authentic sample produced by the irradiation of thymine (Fig. 10a). The stereochemistry was further confirmed by the analysis of the CPD-containing 14-mer with 2D NOE NMR. Despite site-specific mutagenesis and NMR studies of the 14-mer, it was not possible to determine the conformation of the DNA adopted at low pH leading to the *cis,anti* CPD (118).

#### 1.4.4 Anti-CPD Formation in G-Quadruplex Forming Telomeric Sequences

The discovery of the high yield formation of a nonadjacent *cis,anti*-CPD photoproduct in a single-stranded oligonucleotide, prompted the study of the formation of nonadjacent photoproducts in human telomeric DNA sequences d(GGGTTA)<sub>n</sub>. It was known that human telomeric DNA sequences could adopt various types of G-quadruplex structures, of which basket and chair types have adjacent loops that could bring nonadjacent thymine bases into close enough proximity to photodimerize (Fig. 11) (96).

To explore the photochemistry of G-quadruplex structures, the human telomeric model sequence Tel22 d[AGGG(TTAGGG)<sub>3</sub>] was studied first as it was known to form a basket type in the presence of Na<sup>+</sup> (Fig. 11a), and a hybrid type of G-quadruplex in the presence of K<sup>+</sup> (Fig. 11b)(19,20). Nonadjacent CPDs were expected to form in Na<sup>+</sup> solution as the thymines in loops 1 and 3 of basket type G-quadruplexes appeared to be close enough to photoreact (96), whereas in K<sup>+</sup>, no non-adjacent CPDs were expected as hybrid G-quadruplexes lack adjacent lateral loops (Fig. 11a,b) (19,20,96). It was quite a surprise when the opposite was found to be true! CPD formation was assayed by an NP1-coupled MS/MS in which pTp(N)=pTpN tetramers are

produced from non-adjacent CPDs upon degradation with NP1, while adjacent CPDs produce pT=pTpN trimers (Fig. 10b) (96,119). The stereochemistry of the CPDs was determined by HF degradation to thymine CPDs. In the presence of Na<sup>+</sup>, a small mixture of nonadjacent thymine dimers were detected, where the major photoproduct was the adjacent *cis,syn* T=T CPD that resulted from the photodimerization of adjacent thymines in the TTA loops (96). The irradiation of Tel22 in the presence of K<sup>+</sup> lead to the same photoproducts that were found in Na<sup>+</sup>, plus a large amount of *trans,anti* and *cis,anti* T=T CPDs. The *anti*-CPDs formed in K<sup>+</sup> were found to result from the photoreaction of either of the thymines in loop 1 and the 2<sup>nd</sup> T in loop 3 (Fig. 11). Formation of pT(pT)=pT(pA) following NP1 digestion indicated that the CPD formed between the 1<sup>st</sup> T of loop 1 and the 2<sup>nd</sup> T of loop 3 (or vice versa), while the pT(pA)=pT(pA) tetramer indicated that the CPD formed between the 2<sup>nd</sup> T's of both loop 1 and 3 (96).

To explain why non-adjacent photoproducts were detected in K<sup>+</sup>, it was proposed that the non-photoreactive hybrid G-quadruplex must be in equilibrium with a photo-reactive conformer such as the two-tetrad basket-like Form 3 G-quadruplex (Fig. 11c). Unlike the basket structure formed in Na<sup>+</sup>, the Form 3 G-quadruplex has a three-nucleotide lateral loop that is opposite to a four-nucleotide loop, that is longer by one nucleotide in the basket structure, due to a one-nucleotide slippage of the first GGG track. The increase in the length of one of the loops would allow for a better overlap of the thymines in loop 1 and loop 3 and thereby facilitate photodimerization (96,121). This would explain why non-adjacent CPDs were not observed in Na<sup>+</sup> in spite of adopting a basket structure (Fig. 11a) (96).

It was also proposed that the nonadjacent CPDs could have formed from an intermediate triplex conformation that is involved in the interconversion between various G-quadruplex structures (Fig.

11d) (19,122). Modified sequences of Tel22 that could only form the triplex structure also produced similar mixtures of nonadjacent photoproducts but in smaller yields, suggesting that this intermediate is a minor photoreactive conformation of Tel22 in  $K^+$  (96).

To further test the hypothesis that the Form 3 G-quadruplex is the major photoreactive conformation that leads to nonadjacent photoproducts in Tel22, the photochemistry of another telomeric sequence Tel26 d[AAAGGG(TTAGGG)<sub>3</sub>AA] was studied. Tel26 is known to predominantly adopt a hybrid type structure in  $K^+$  that is not expected to produce any nonadjacent photoproducts (19,96,97). Irradiation with UVB light, however, produced primarily the *trans,anti* T2=T2 CPD in contrast to Tel22 suggesting that this product was arising from a specific conformation that is preferred by Tel26. The preferential formation of the *trans,anti* CPD in Tel26 suggests that there are multiple photoreactive secondary structures that must be in equilibrium with the major non-photoreactive conformation (96,97).

#### **1.4.5 *Trans-anti* CPD formation in Reverse Hoogsteen hairpins**

To explore if there are other photoreactive secondary structures besides G-quadruplexes that can explain the preferential formation of the *trans,anti* T2=T2 CPD in both Tel26 and Tel22, photochemical studies were performed in the presence of  $Li^+$  ions (97). Because G-quadruplex structures cannot form in the presence of lithium due to its small size and strong hydration shell (10), the formation of the *trans,anti* T2=T2 CPD in the presence of this ion would suggest that human telomeric sequences can adopt a different photoreactive conformation besides G-quadruplexes (97,115).

Irradiation of Tel26 in the presence of  $\text{Li}^+$  led to the formation of the *trans,anti* T2=T2 CPD with a higher yield and selectivity compared to  $\text{Na}^+$  and  $\text{K}^+$ , indicating that other non-quadruplex forms may indeed be involved in the induction of these nonadjacent photoproducts (Fig 12.a) (97). It was proposed that this photoreactive conformation could be a hairpin structure with reverse Hoogsteen G•G base pairing, as compared to the alternative hairpin with Hoogsteen base pairs, as the G bases are in the more favorable *anti*-glycosyl conformation (Fig. 12d) (98,115). If this reverse Hoogsteen hairpin were to form in Tel26 and Tel22 it would produce an interior loop made of the loop 1 and loop 3 TTA sequences that would enable the thymine bases to photodimerize and form the *trans,anti* T2=T2 CPD (Fig. 12d). The formation of the reverse Hoogsteen hairpin would also explain why Tel26 forms this specific CPD with more selectivity than Tel22, as Tel26 contains two additional A bases at both 5'- and 3'-ends that would further stabilize the hairpin structure by reverse Hoogsteen base pairing. The *trans,anti* T=T CPD formation was also observed when the G's in the G-tracks were replaced with A's eliminating the possibility of G-quadruplex formation providing further evidence that an additional secondary structure of DNA could be involved in the formation of these photoproducts (98).

The differences in nonadjacent CPD formation for Tel26 and Tel22 in  $\text{Na}^+$ ,  $\text{K}^+$ , and  $\text{Li}^+$  (Fig. 12a) can now be explained by the DNA structures that are stabilized by these ions. In the presence of  $\text{Na}^+$ , Tel22 and Tel26 mainly adopt the basket structure conformation which has lateral loops that are made of three nucleotides that do not enable the T bases to stack on each other and dimerize (Fig 12.b). In the presence of  $\text{K}^+$ , these sequences are in equilibrium between the non-photoreactive hybrid G-quadruplex and the photoreactive Form 3 basket structure and reverse Hoogsteen hairpin (Fig. 12c-d), whereas in  $\text{Li}^+$  only the reverse Hoogsteen hairpin can form. The form 3 basket

structure produces a variety of nonadjacent photoproducts while the reverse Hoogsteen hairpin mainly produces the *trans*, *anti* T2=T2 CPD (97,98).

#### 1.4.6 Anti-CPD formation in human promoter sequences

The detection of nonadjacent photoproducts in human telomeric G-quadruplex forming sequences prompted an investigation into the formation of these photoproducts in human promoter sequences (99). To do so, a computer search was carried out on six human chromosomes for putative G-quadruplex forming sequences that had loops of three or more nucleotides with pyrimidines in loops 1 and 3. From these searches, 300 potential sequences were found upstream the TSS of genes that contained three G-runs, loops with 3 to 7 nucleotides where loops 1 and 3 had at least three pyrimidine bases, one of them being a thymine. From these 300 potential sequences, 15 sequences were randomly selected to perform irradiation studies in the presence of Na<sup>+</sup> and K<sup>+</sup>, where nonadjacent CPDs were searched with the use of the NP1 coupled MS assay (99).

In most of the 15 promoter sequences that were studied *in vitro* evidence of non-adjacent CPD formation was detected in the form of pT(pN)=pT(pN) tetramers resulting from the NP1 digestion. Of these, only three of the sequences from the promoter regions of the CALU, STARD3NL, and SUSD3 genes were analyzed in more detail. These three sequences showed significant *anti*-photoproduct formation in the presence of both Na<sup>+</sup> and K<sup>+</sup>, making it the first report of *anti*-CPD formation in Na<sup>+</sup> conditions. In addition the first nonadjacent T=C photoproduct was detected in K<sup>+</sup> conditions in the STARD3NL sequences (99).

The *anti*-CPD photoproduct pT(pG)=pT(pG) was detected in all three sequences which could be assigned, along with the C containing CPD, to putative chair (Fig. 13a) and basket (Fig. 13b) types



of G-quadruplexes (99). The photocrosslinking of these promoter sequences, as well as human telomeric sequences, suggest that nonadjacent photoproducts could be used to probe the formation of G-quadruplex structures and other non-B forms of DNA, such as reverse Hoogsteen hairpins in the human genome (99,115).

## 1.5 Dissertation Overview

The G-quadruplex structure of DNA has been shown through a variety of computational, *in vitro*, and *in vivo* studies to potentially play an extremely important role in replication, transcription, and translation. Yet it has been extremely difficult to prove if G-quadruplexes form in live cells due to the lack of techniques that would allow for their unambiguous detection and location in the human genome. Therefore, it would be ideal if a method could be developed that does not disrupt the transient nature of DNA structures, either by inducing or inhibiting G-quadruplex formation, and that could instantaneously act the moment these structures form in cells. In this thesis, we explore the idea that DNA photoproducts can be used as intrinsic probes to trap certain G-quadruplex structures *in vivo* by covalent photocrosslinking of their lateral loops via *anti* CPD formation, for which different techniques can then be developed to detect and map their location in the human genome.

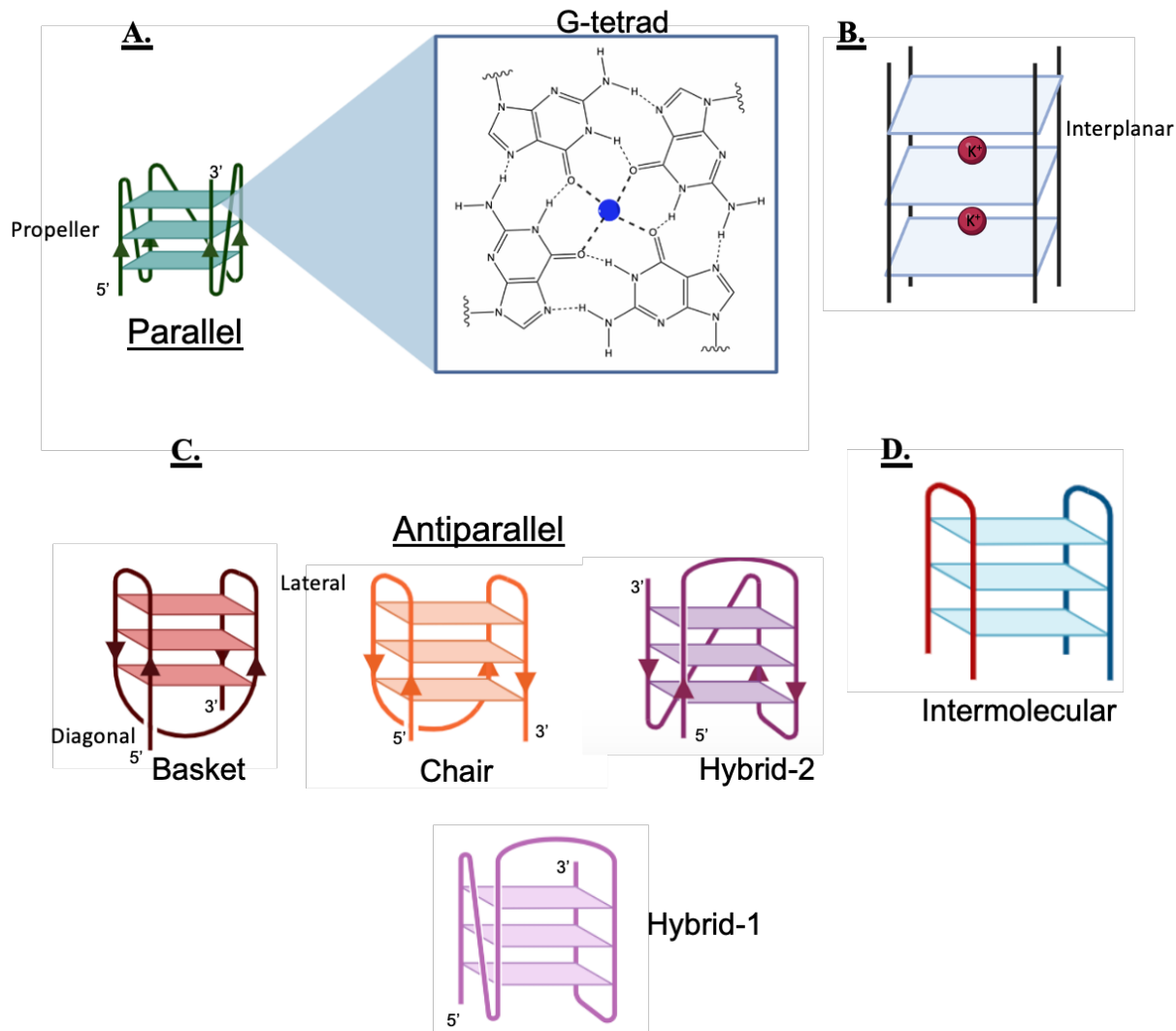
In Chapter 2, we report the development of radioactive post- and pre-labeling assays for the detection of nonadjacent photoproducts and demonstrate that it can be used to detect the *trans,anti* T=T CPD that forms in human telomeric DNA. We do so by making use of the enzyme snake venom phosphodiesterase (SVP) to degrade the *anti*-CPD containing sequence to a tetramer pTpT=(pT)pT, that after dephosphorylation with calf intestinal phosphodiesterase and

rephosphorylation with kinase and [<sup>32</sup>P]-ATP produces a fragment of DNA that can be analyzed by high resolution gel electrophoresis with an approximately ten fmol lower limit of detection.

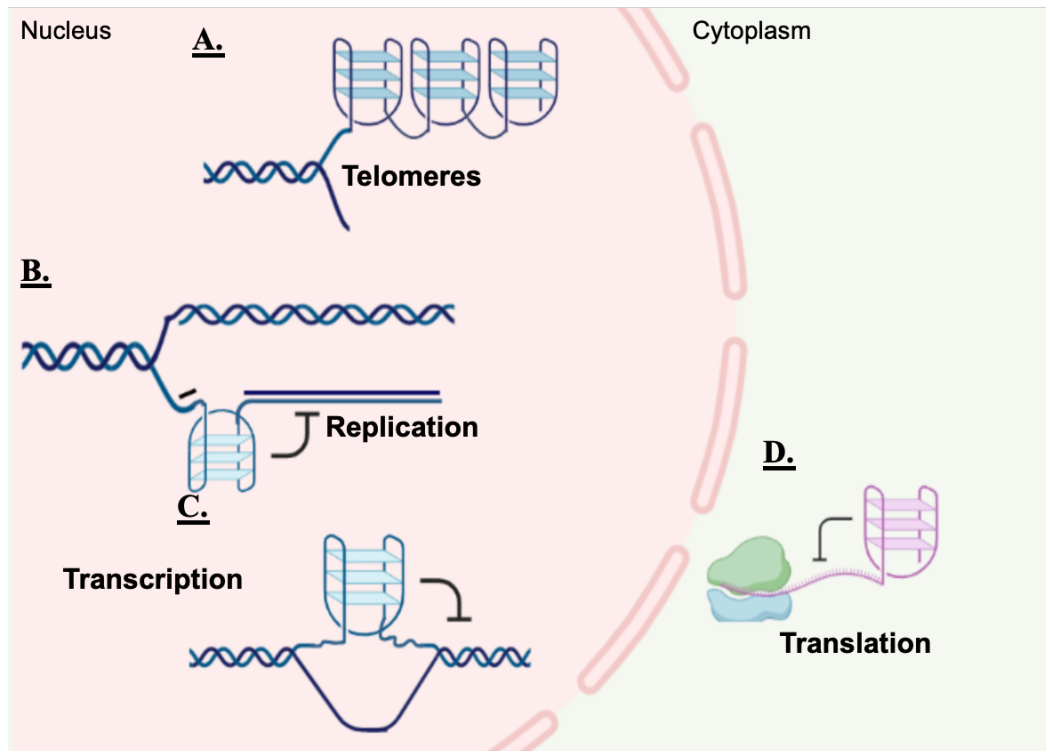
In Chapter 3, we test the radioactive post-labeling assay on different telomeric and promoter G-quadruplex forming sequences that were previously reported to form non-adjacent photoproducts and for which the type of G-quadruplex structure has been thoroughly characterized. We show that the post-labeling assay is capable of detecting non-adjacent CPDs with a variety of flanking nucleotides that can be used to provide information on the non-B conformation of the DNA that led to the CPD. We similarly show that this method can also be used to observe the effect that Na<sup>+</sup>, Li<sup>+</sup>, and K<sup>+</sup> have on *anti* CPD yields in a variety of G-quadruplex forming sequences to test if additional photoreactive conformations of DNA may be present at the moment of UV irradiation.

In Chapter 4, we propose methods that could be employed to detect *anti* CPDs in irradiated HeLa cells. We will also discuss future experimental work that can be employed to help map and detect the location of *anti*-CPDs if this method is successful at detecting non-adjacent photoproduct formation in live cells.

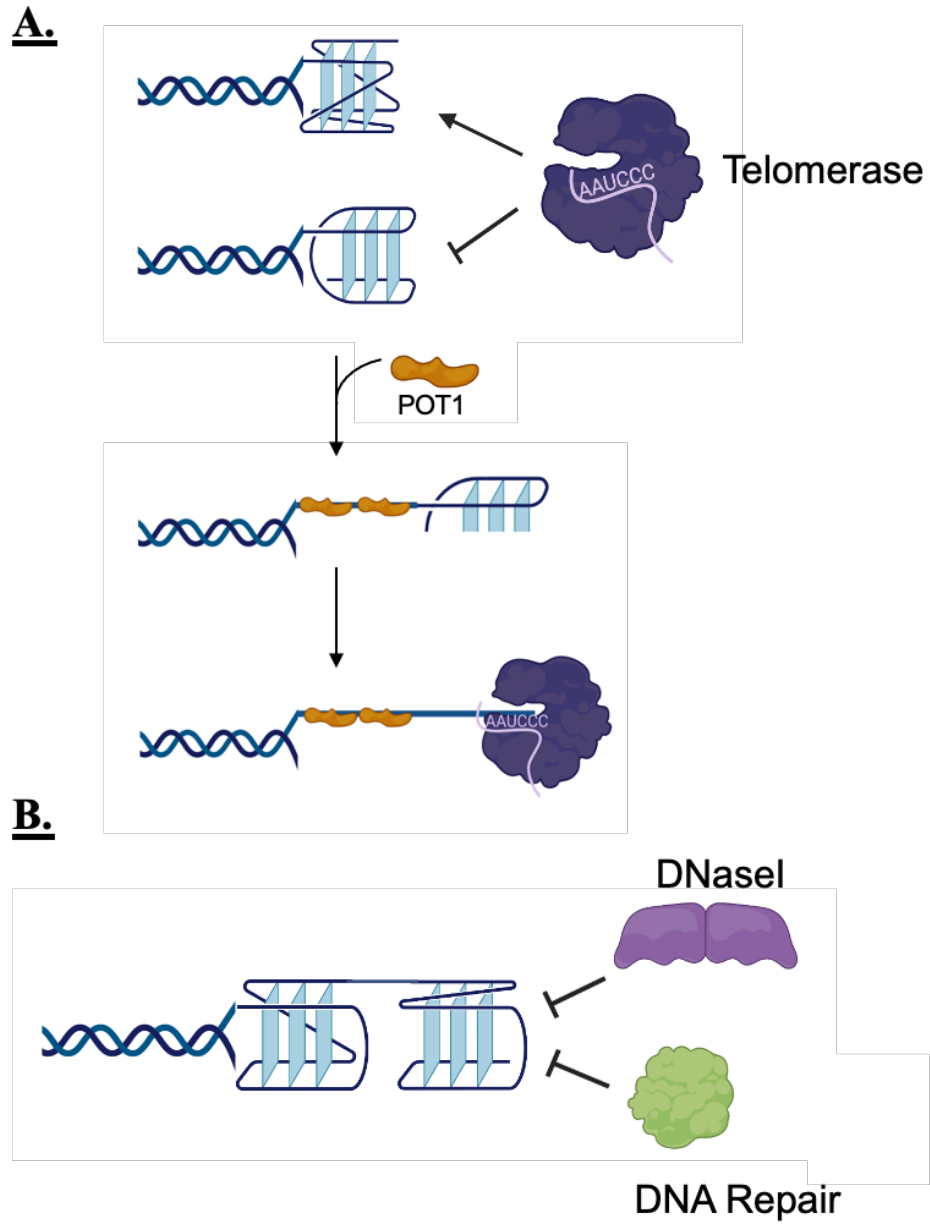
## 1.6 Figures



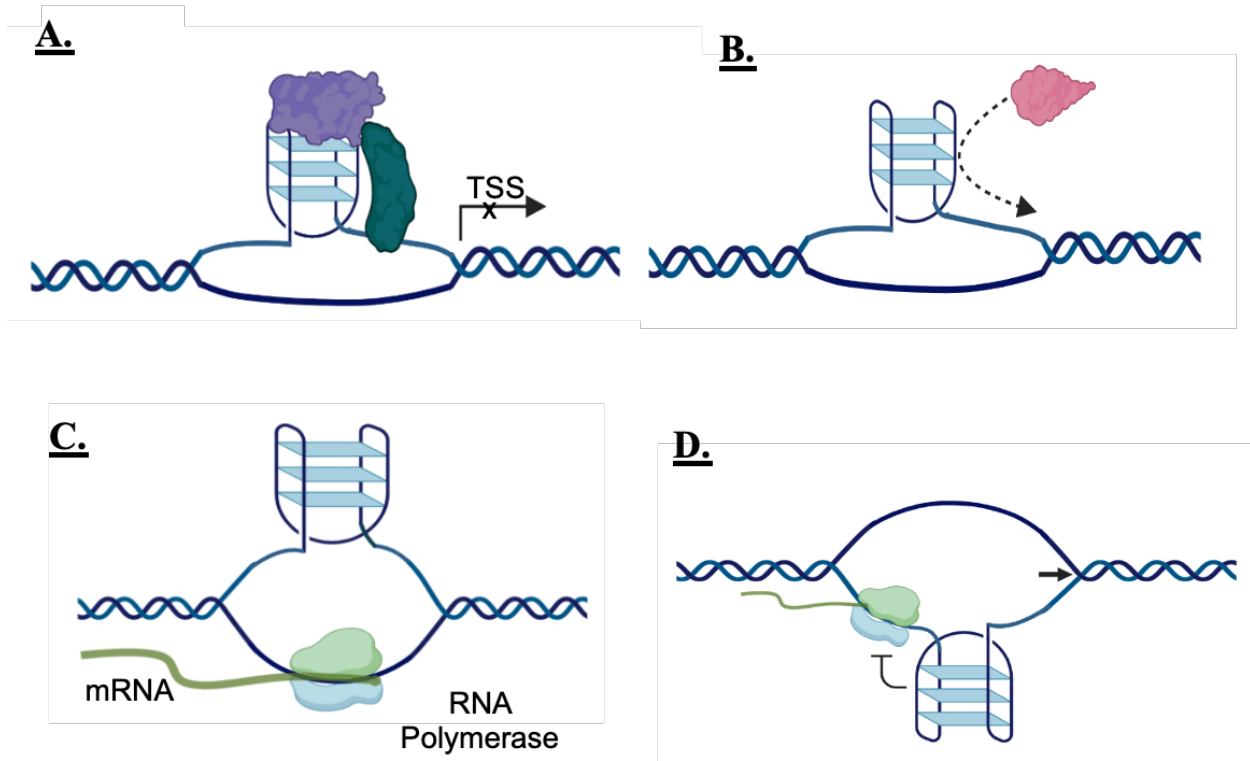
**Figure 1.1. Four-stranded G-quadruplex structural polymorphism.** (a) Intramolecular parallel G-quadruplex conformation. The G-tetrad building block of these structures has a  $\text{Na}^+$  cation in the plane of the tetrad whereas (b)  $\text{K}^+$  is found between the planes of two tetrads. (c) Intramolecular anti-parallel G-quadruplex conformations. (d) Intermolecular conformation involving two different strands of DNA. Figures created with BioRender.



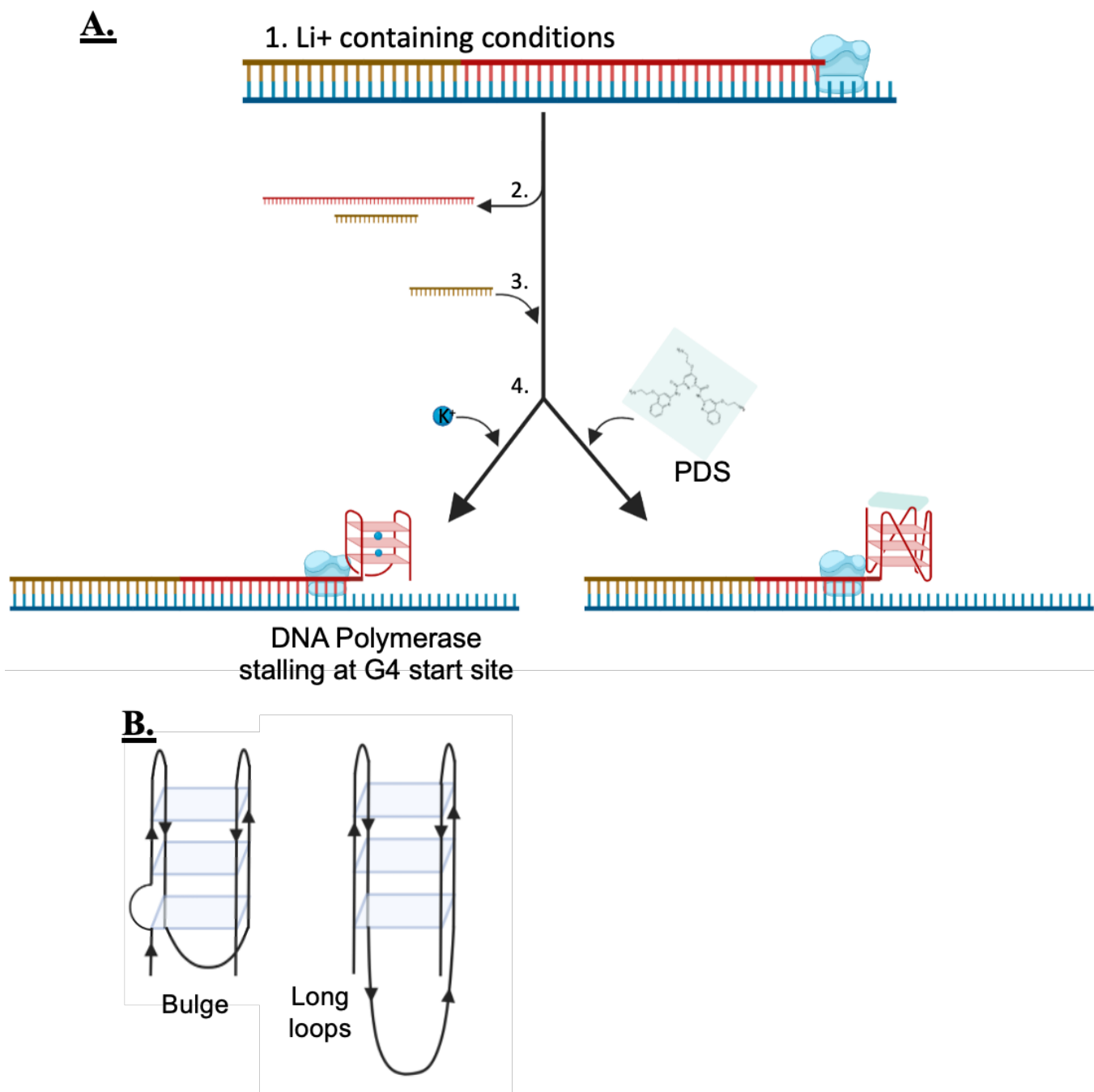
**Figure 1.2. Possible sites of G-quadruplex formation in the human genome.** G-quadruplexes are believed to form in G-rich regions of DNA that become transiently single stranded, such as in (a) telomeres, (b) replication forks, (c) transcription bubbles and (d) during translation. Figure adapted from (123) and created with BioRender.



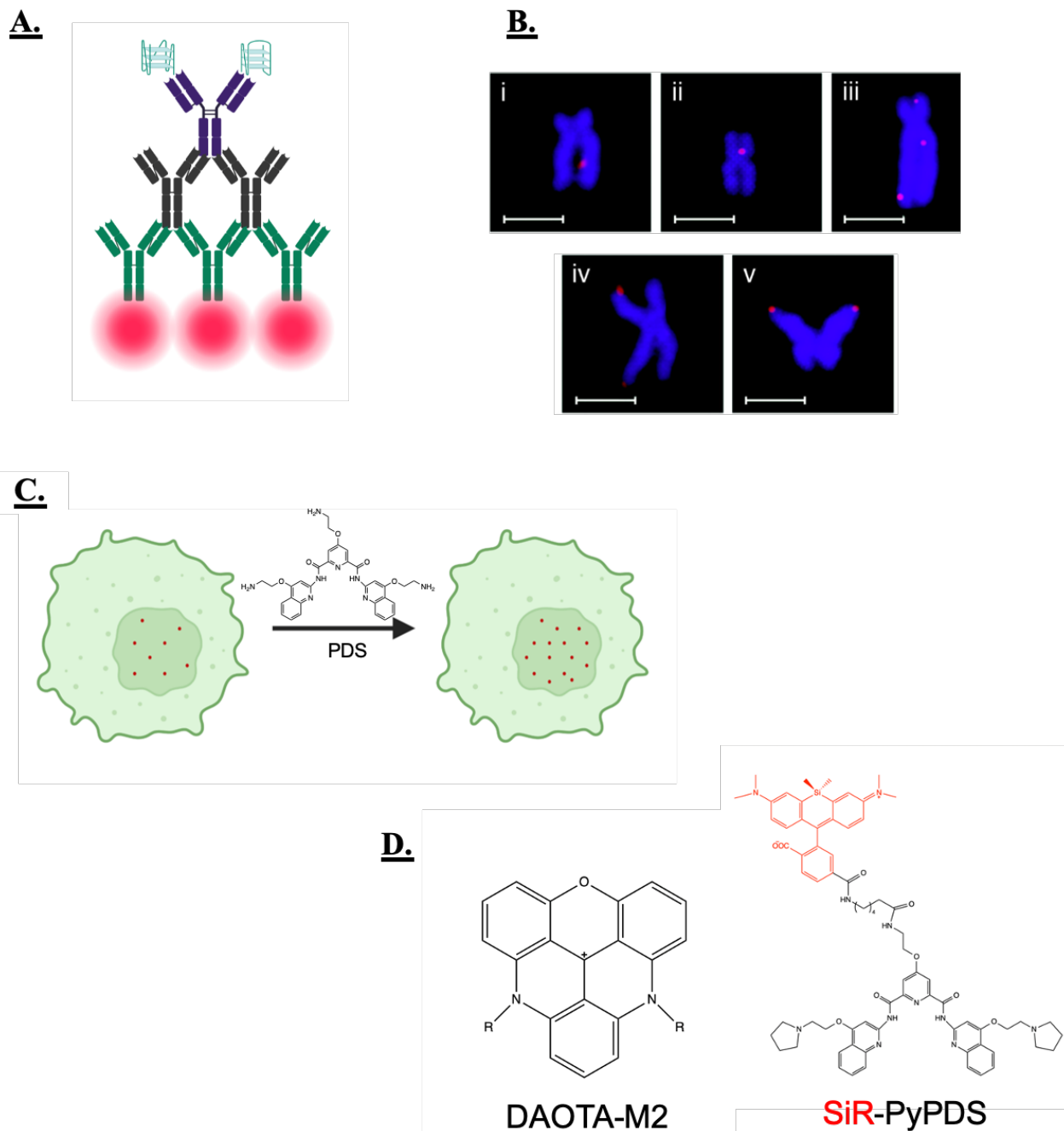
**Figure 1.3.** Potential biological role of G-quadruplexes in telomeres (a) G-quadruplex structures are believed to help regulate telomerase activity and (b) to play a capping role at the end of chromosomes. Figure adapted from (39) and created with BioRender.



**Figure 1.4. Potential biological roles of G-quadruplexes in transcription.** G-quadruplexes are believed (a) to enhance transcription by acting as binding sites for transcription factors, (b) to repress transcription by blocking transcription factor binding, (c) to enhance transcription by stabilizing the R-loop, or (d) to repress transcription by stalling RNA-polymerase. Figure adapted from (41,123) and created with BioRender.

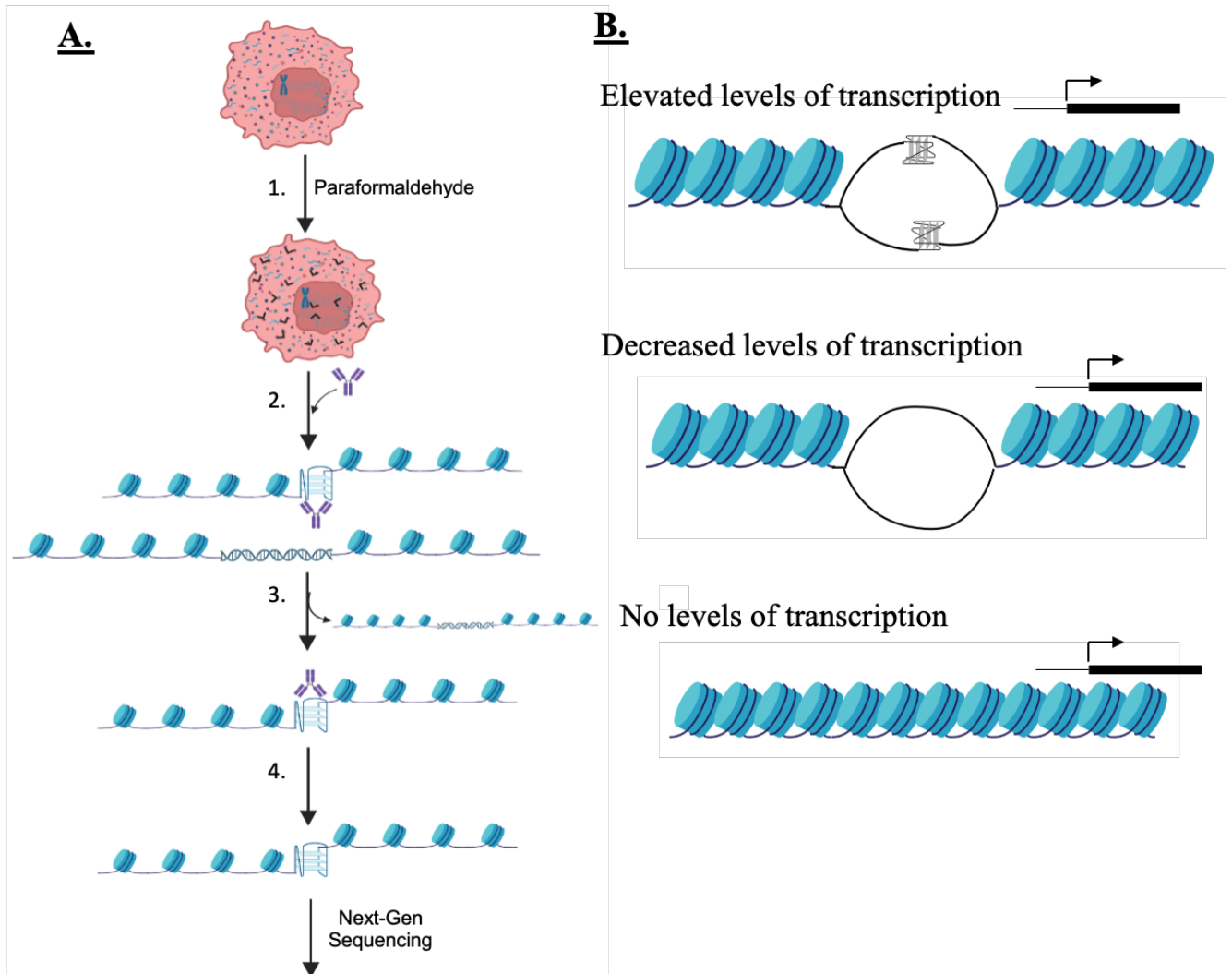


**Figure 1.5. G-quadruplex sequencing (G4-seq).** (a) In a typical G4-seq protocol, sequencing is done twice. Once under conditions that enables accurate sequencing of DNA fragments, followed by re-sequencing the DNA under conditions that promote G-quadruplex formation. (b) Non-canonical G-quadruplex structures identified by G4-seq. Figure adapted from (77) and created with BioRender.

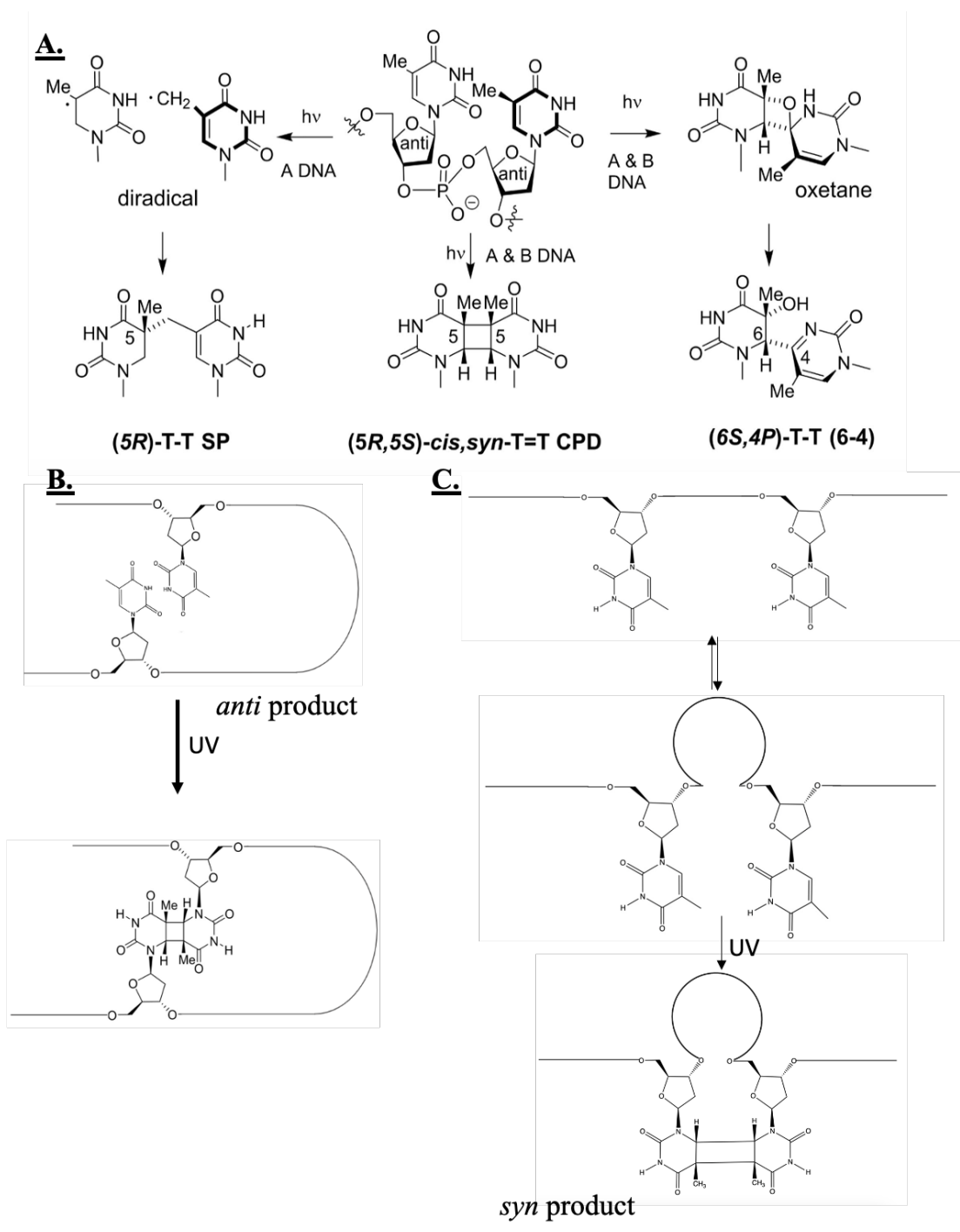


**Figure 1.6. Fluorescence imaging of G-quadruplexes *in vivo*.** (a) G-quadruplex detection through fluorescent antibodies. (b) Antibody imaging of G-quadruplexes in individual chromosomes (red) vs DNA (blue) (83). (c) Representation of increased foci in cells after addition of stabilizing ligands such as PDS. (d) Two fluorescent probe molecules for G-quadruplex detection. Figures adapted from (6) and created with BioRender.

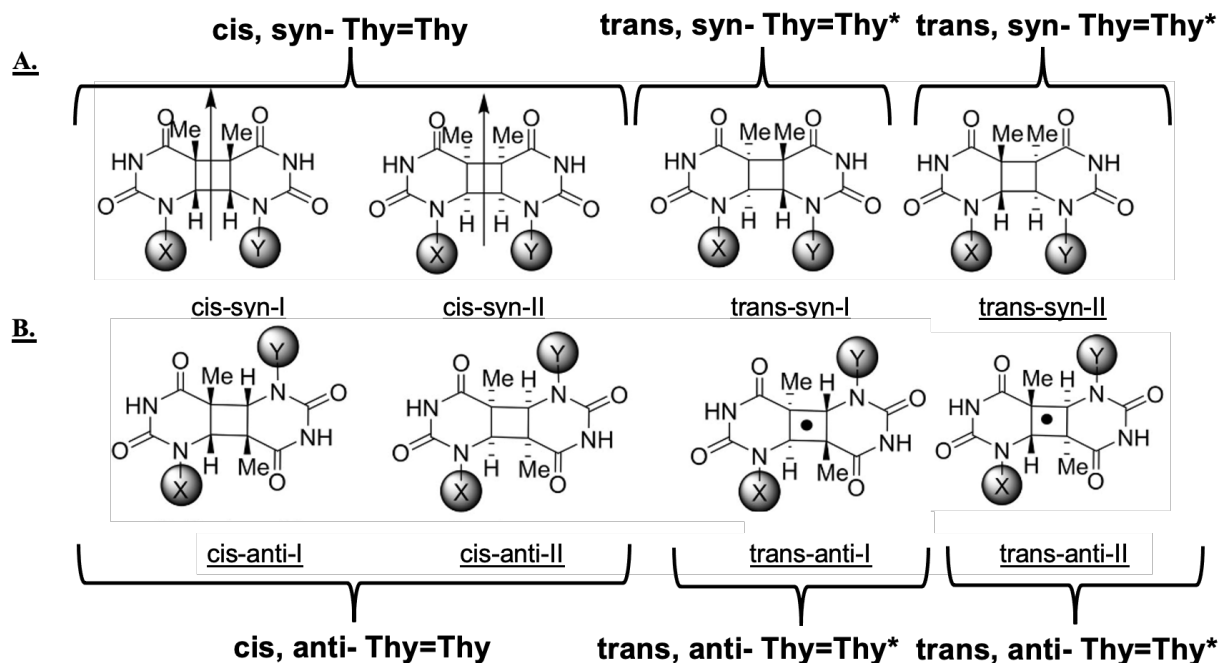




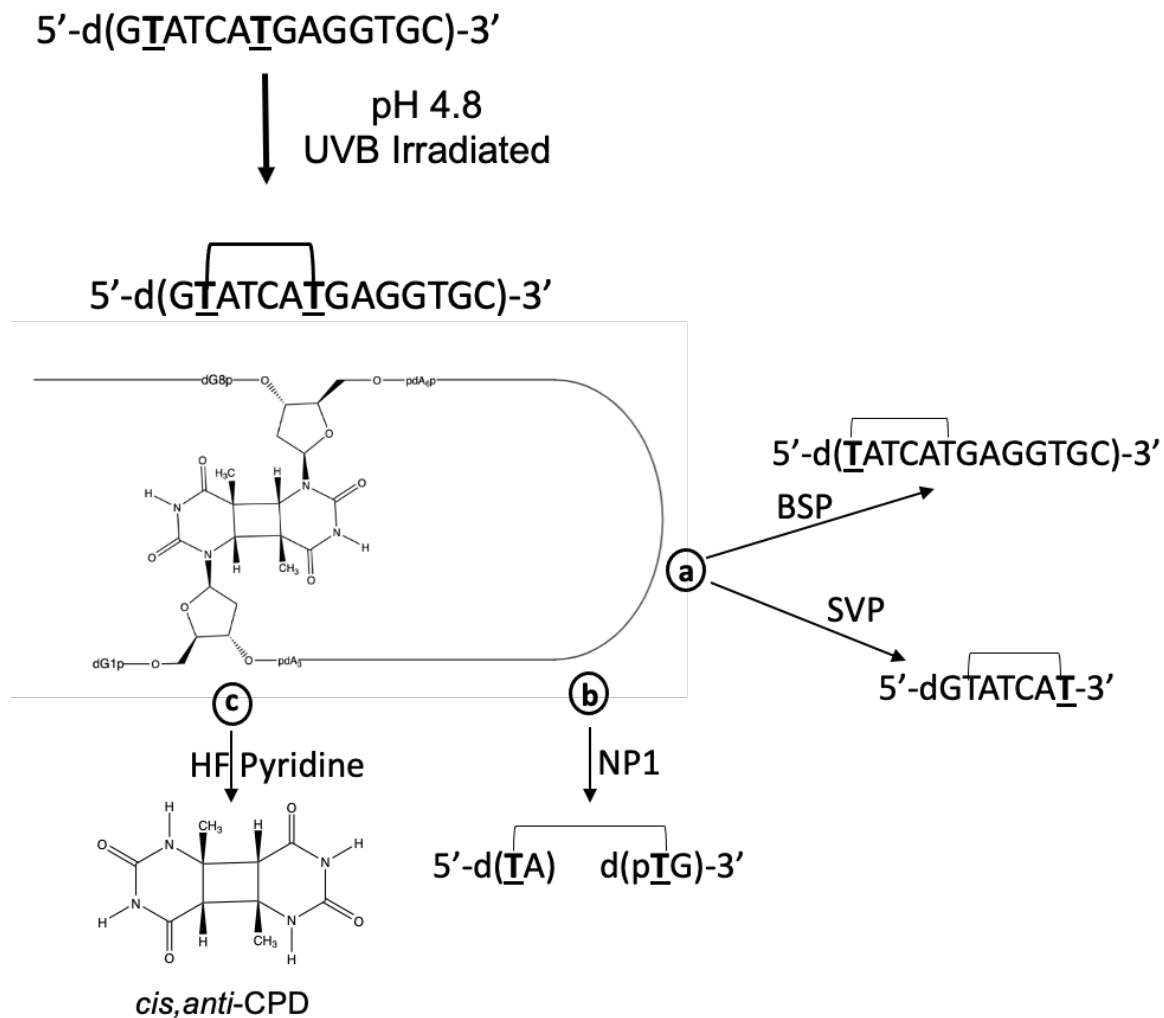
**Figure 1.7. G-quadruplex chromatin immunoprecipitation sequencing (G4 ChIP-seq).** (a) The technique involves cell fixation with formaldehyde, followed by chromatin precipitation with G-quadruplex specific antibodies and Next-Gen sequencing. (b) Relationship between G-quadruplexes, chromatin state, and transcription. Figure adapted from (93) and created with BioRender.



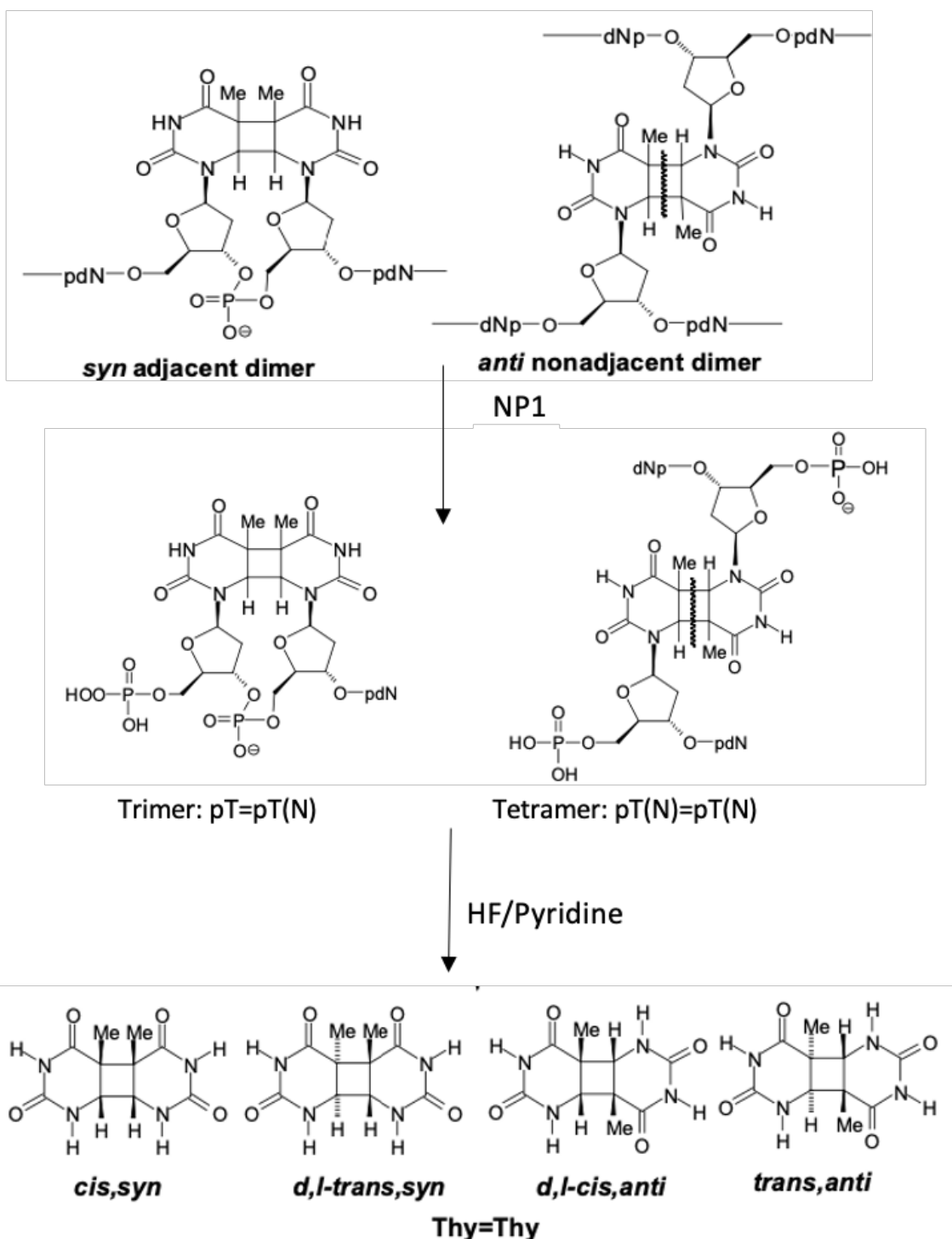
**Figure 1.8.** DNA photoproducts as intrinsic probes for DNA secondary structure and conformation. (a) Different types of thymine photoproducts that form in the A and B duplex forms of DNA. (b) Formation of *anti* and (c) *syn* non-adjacent photoproducts from non-B and B conformations of DNA.



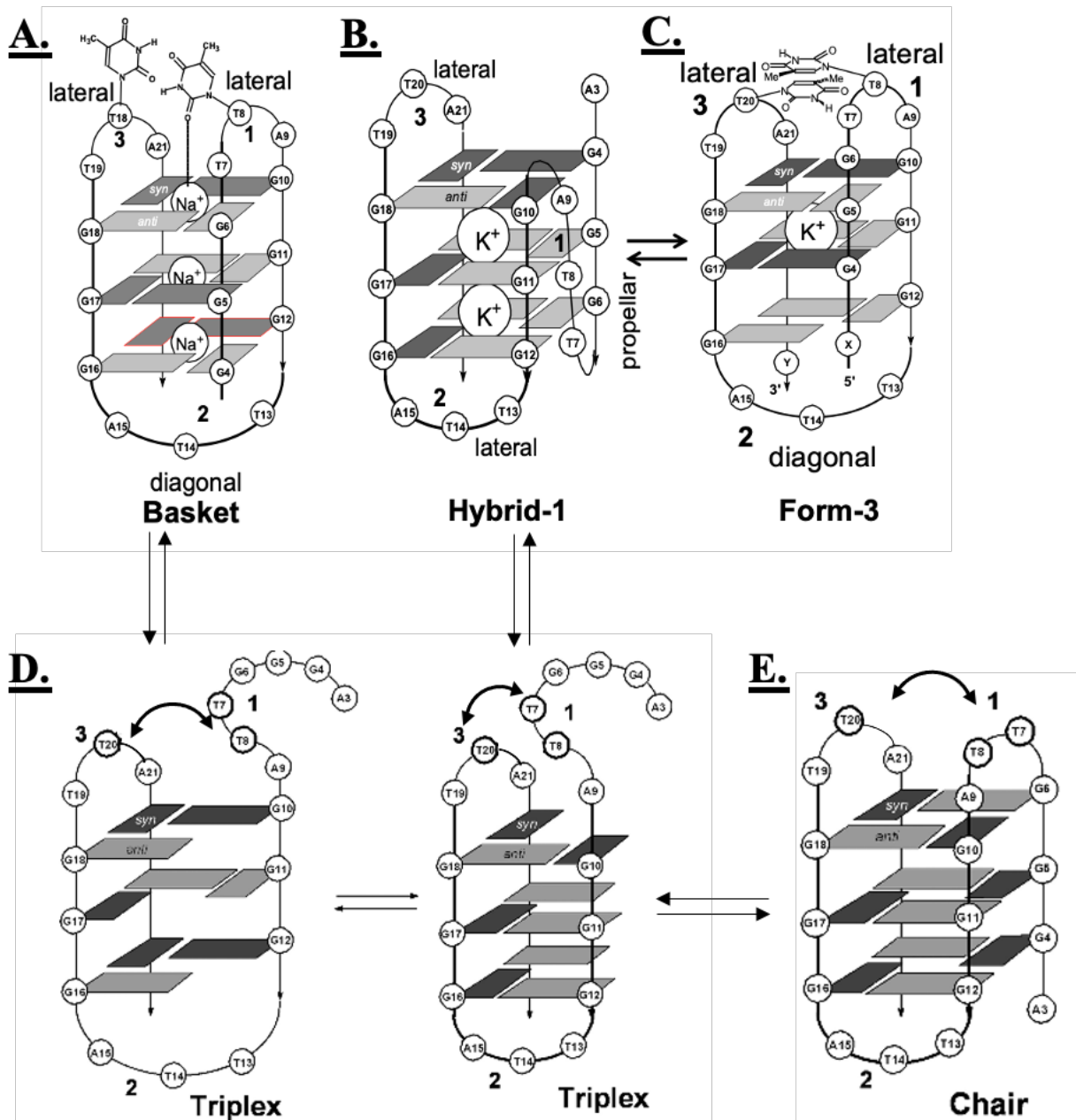
**Figure 1.9. Photochemistry of DNA pyrimidine bases.** Structure of the different stereoisomers of T=T CPDs forming in DNA from adjacent or non-adjacent (a) head-to-head or (b) head-to-tail bases. After the Thy=Thy CPDs are released from the deoxyribose sugar by acid hydrolysis only six stereoisomers are observed, where the two *trans,syn* CPDs with the asterisk represent that they are enantiomers of each other as also is the case for the two *trans,anti* CPDs.



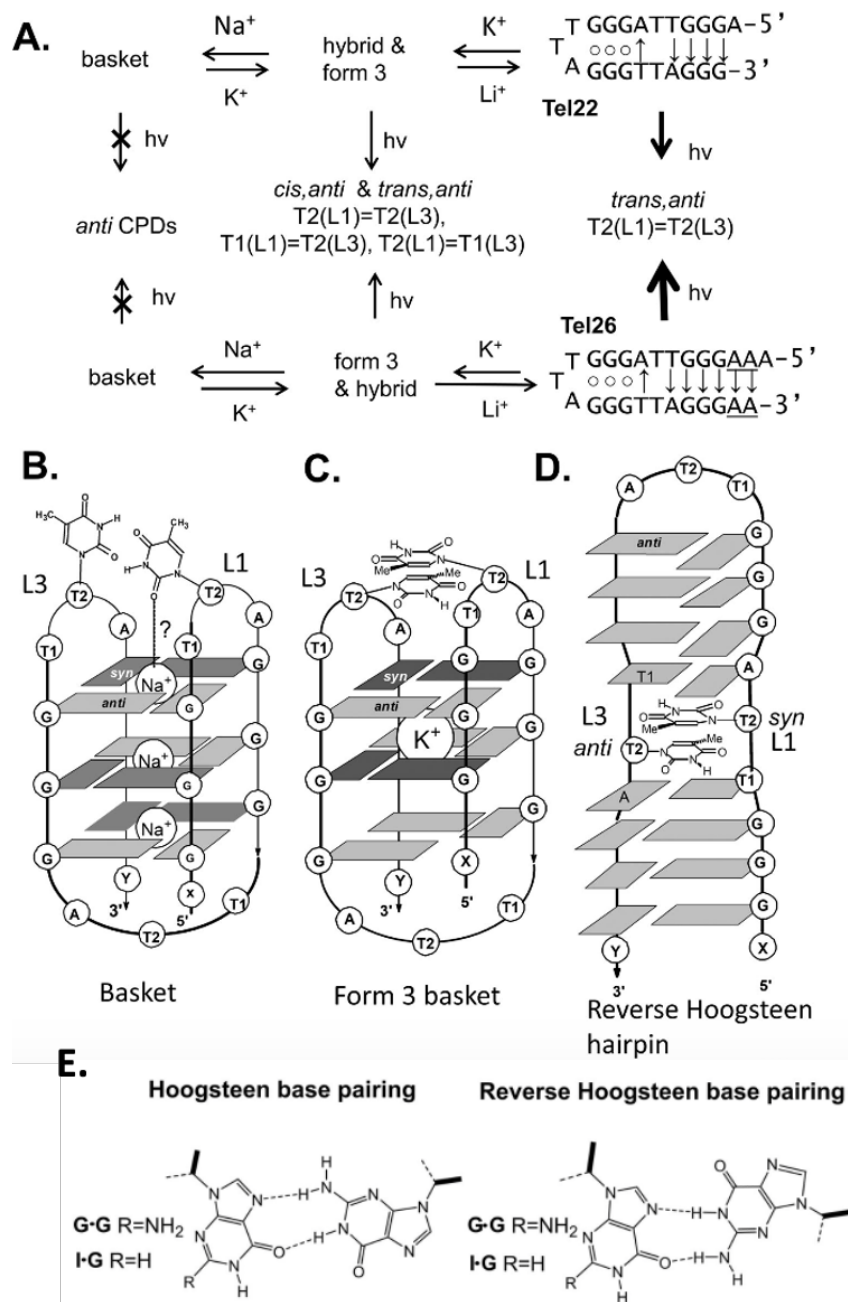
**Figure 1.10. Structure determination of the *cis,anti*-CPD formed at low pH in a single stranded oligonucleotide.** (a) Exonuclease-coupled MS for determination of the position where the non-adjacent photoproduct formed (b) NP1 digestion to determine flanking bases. (c) HF degradation to thymine dimer for identification of stereochemistry by comparison to authentic compounds (118).



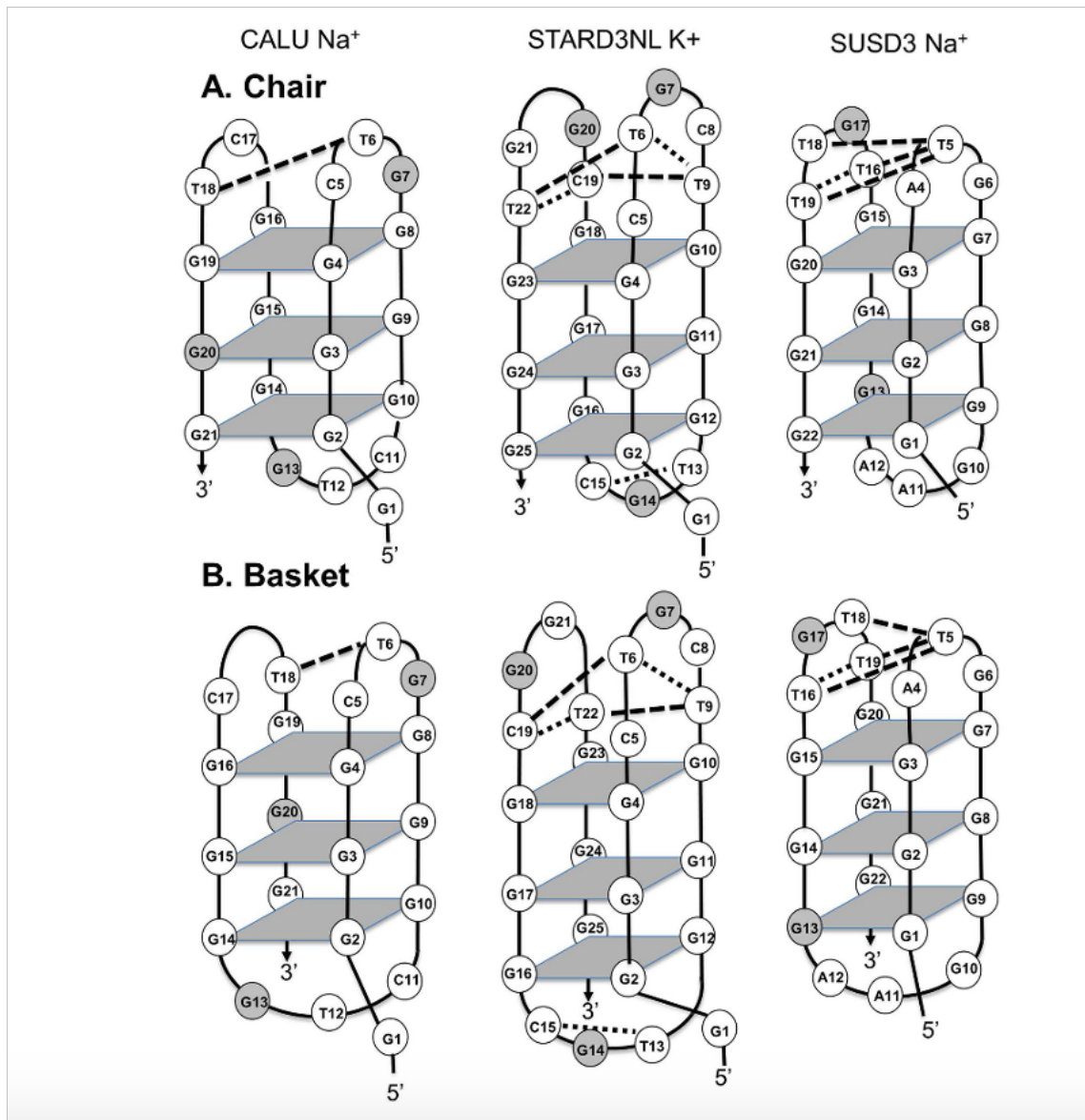
**Figure 1.11.** NP1- coupled mass spectrometry assay for the detection of nonadjacent and adjacent photoproducts (118).



**Figure 1.12.** G-quadruplex structures and intermediates implicated in *anti* CPD formation in human telomeric DNA. (a) Basket structure that could lead to the formation of non-adjacent photoproducts from T bases found in the lateral loops (b) Hybrid-1 structure that is not expected to produce non-adjacent CPDs but that is believed to be in equilibrium with a photoreactive form of the (c) basket conformation known as Form-3. (d) A photoreactive triplex intermediate conformation between the basket, two hybrid types, and (e) chair conformation that is believed to be a minor photoreactive structure of G-quadruplexes (96,115).



**Figure 1.13.** Proposed scheme and structure that explain anti CPD formation in Tel22 and Tel26 as function of cation and sequence. (a) Cationic and sequence dependent formation of nonadjacent CPDs in Tel26 and Tel22 sequences (b) three-tetrad basket structure in  $\text{Na}^+$  where nonadjacent photoproducts are not expected to form (c) Form 3 basket structure that is believed to be the photoreactive G-quadruplex that forms non-adjacent CPD's (d) Reverse-Hoogsteen hairpin structure that may be the major photoreactive conformation leading to *trans,anti* T2=T2 CPD (e) Hoogsteen and reverse-Hoogsteen base pairing between two guanines (97,115).



**Figure 1.14.** Sites of non-adjacent CPD formation in putative G-quadruplex forming sequences in human promoters. (a) Chair and (b) basket forms of putative G-quadruplex forming sequences from the promoters of the CALU, STARD3NM, and SUSD3 genes (99).



## 1.7 References

1. Zimmerman SB, Cohen GH, Davies DR. X-ray fiber diffraction and model-building study of polyguanylic acid and polyinosinic acid. *Journal of Molecular Biology*. 1975 Feb 25;92(2):181–92.
2. Arnott S, Chandrasekaran R, Marttila CM. Structures for polyinosinic acid and polyguanylic acid. *Biochem J*. 1974 Aug;141(2):537–43.
3. Pinnavaia TJ, Marshall CL, Mettler CM, Fisk CL, Miles HT, Becker ED. Alkali metal ion specificity in the solution ordering of a nucleotide, 5'-guanosine monophosphate. *Journal of the American Chemical Society*. 1978 May;100(11).
4. Sen D, Gilbert W. A sodium-potassium switch in the formation of four-stranded G4-DNA. *Nature*. 1990 Mar;344(6265):410–4.
5. Choi J, Majima T. Conformational changes of non-B DNA. *Chem Soc Rev*. 2011;40(12):5893.
6. Spiegel J, Adhikari S, Balasubramanian S. The Structure and Function of DNA G-Quadruplexes. *Trends Chem*. 2020 Feb;2(2):123–36.
7. Williamson JR, Raghuraman MK, Cech TR. Monovalent cation-induced structure of telomeric DNA: The G-quartet model. *Cell*. 1989 Dec 1;59(5):871–80.
8. Gu J, Leszczynski J. Origin of Na<sup>+</sup>/K<sup>+</sup> Selectivity of the Guanine Tetraplexes in Water: The Theoretical Rationale. *J Phys Chem A*. 2002 Jan 1;106(3):529–32.
9. Hud NV, Smith FW, Anet FAL, Feigon J. The Selectivity for K<sup>+</sup> versus Na<sup>+</sup> in DNA Quadruplexes Is Dominated by Relative Free Energies of Hydration: A Thermodynamic Analysis by <sup>1</sup>H NMR. *Biochemistry*. 1996 Jan 1;35(48):15383–90.
10. Ross WS, Hardin CC. Ion-Induced Stabilization of the G-DNA Quadruplex: Free Energy Perturbation Studies. *J Am Chem Soc*. 1994 Jul;116(14):6070–80.
11. Miyoshi D, Nakao A, Sugimoto N. Molecular Crowding Regulates the Structural Switch of the DNA G-Quadruplex. *Biochemistry*. 2002 Dec 1;41(50):15017–24.
12. Huppert JL, Balasubramanian S. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res*. 2005;33(9):2908–16.
13. Venczel EA, Sen D. Parallel and antiparallel G-DNA structures from a complex telomeric sequence. *Biochemistry*. 1993 Jun 22;32(24):6220–8.

14. Miller MC, Buscaglia R, Chaires JB, Lane AN, Trent JO. Hydration Is a Major Determinant of the G-Quadruplex Stability and Conformation of the Human Telomere 3' Sequence of d(AG<sub>3</sub>(TTAG<sub>3</sub>)<sub>3</sub>). *J Am Chem Soc.* 2010 Dec 8;132(48):17105–7.
15. Kikin O, D'Antonio L, Bagga PS. QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res.* 2006 Jul 1;34(Web Server issue):W676–82.
16. Sen D, Gilbert W. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature.* 1988 Jul;334(6180):364–6.
17. Laughlan G, Murchie AIH, Norman DG, Moore MH, Moody PCE, Lilley DMJ, et al. The High-Resolution Crystal Structure of a Parallel-Stranded Guanine Tetraplex. *Science.* 1994;265(5171):520–4.
18. Smith FW, Feigon J. Quadruplex structure of Oxytricha telomeric DNA oligonucleotides. *Nature.* 1992 Mar;356(6365):164–8.
19. 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(9):2723–35.
20. Xu Y, Noguchi Y, Sugiyama H. The new models of the human telomere d[AGGG(TTAGGG)<sub>3</sub>] in K<sup>+</sup> solution. *Bioorganic & Medicinal Chemistry.* 2006 Aug 5;14:5584–91.
21. Phan AT, Kuryavyi V, Luu KN, Patel DJ. Structure of two intramolecular G-quadruplexes formed by natural human telomere sequences in K<sup>+</sup> solution†. *Nucleic Acids Res.* 2007 Oct;35(19):6517–25.
22. Blackburn EH. Structure and Function of Telomeres. *Nature.* 1991 Apr 18;350:569–73.
23. Wang Y, Patel DJ. Solution structure of the human telomeric repeat d[AG<sub>3</sub>(T<sub>2</sub>AG<sub>3</sub>)<sub>3</sub>] G-tetraplex. *Structure.* 1993 Dec 15;1(4):263–82.
24. Ying L, Green JJ, Li H, Klenerman D, Balasubramanian S. Studies on the structure and dynamics of the human telomeric G quadruplex by single-molecule fluorescence resonance energy transfer. *Proc Natl Acad Sci U S A.* 2003 Dec 9;100(25):14629–34.
25. Harkness RW, Mittermaier AK. G-register exchange dynamics in guanine quadruplexes. *Nucleic Acids Res.* 2016 May 5;44(8):3481–94.
26. Gray RD, Trent JO, Chaires JB. Folding and Unfolding Pathways of the Human Telomeric G-Quadruplex. *J Mol Biol.* 2014 Apr 17;426(8):1629–50.

27. Harkness RW, Mittermaier AK. G-quadruplex dynamics. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*. 2017 Jun 20;1865(11):1544–54.
28. Shirude PS, Okumus B, Ying L, Ha T, Balasubramanian S. Single-Molecule Conformational Analysis of G-Quadruplex Formation in the Promoter DNA Duplex of the Proto-Oncogene C-Kit. *J Am Chem Soc*. 2007 Jun 20;129(24):7484–5.
29. Zhou J, Bourdoncle A, Rosu F, Gabelica V, Mergny JL. Tri-G-Quadruplex: Controlled Assembly of a G-Quadruplex Structure from Three G-Rich Strands. *Angew Chem Int Ed*. 2012 Oct 29;51(44):11002–5.
30. Largy E, Mergny JL. Shape matters: size-exclusion HPLC for the study of nucleic acid structural polymorphism. *Nucleic Acids Res*. 2014 Oct 29;42(19):e149.
31. Robinson J, Raguseo F, Nuccio SP, Liano D, Di Antonio M. DNA G-quadruplex structures: more than simple roadblocks to transcription? *Nucleic Acids Res*. 2021 Sep 7;49(15):8419–31.
32. Yang D. G-Quadruplex DNA and RNA. *G-Quadruplex Nucleic Acids*. 2019 Aug 23;2035:1–24.
33. Bonnell E, Pasquier E, Wellinger RJ. Telomere Replication: Solving Multiple End Replication Problems. *Front Cell Dev Biol*. 2021 Apr 1;9:668171.
34. Lipps HJ, Grussem W, Prescott DM. Higher order DNA structure in macronuclear chromatin of the hypotrichous ciliate *Oxytricha nova*. *Proc Natl Acad Sci USA*. 1982 Apr;79(8):2495–9.
35. Moye AL, Porter KC, Cohen SB, Phan T, Zyner KG, Sasaki N, et al. Telomeric G-quadruplexes are a substrate and site of localization for human telomerase. *Nat Commun*. 2015 Jul 9;6(1):7643.
36. Zahler AM, Williamson JR, Cech TR, Prescott DM. Inhibition of telomerase by G-quartet DNA structures. *Nature*. 1991 Apr;350(6320):718–20.
37. Salazar M, Thompson BD, Kerwin SM, Hurley LH. Thermally Induced DNA·RNA Hybrid to G-Quadruplex Transitions: Possible Implications for Telomere Synthesis by Telomerase. *Biochemistry*. 1996 Jan 1;35(50):16110–5.
38. Hwang H, Kreig A, Calvert J, Lormand J, Kwon Y, Daley JM, et al. Telomeric overhang length determines structural dynamics and accessibility to telomerase and ALT associated proteins. *Structure*. 2014 Jun 10;22(6):842–53.
39. Bryan TM. G-Quadruplexes at Telomeres: Friend or Foe? *Molecules*. 2020 Aug 13;25(16):3686.

40. Smith JS, Chen Q, Yatsunyk LA, Nicoludis JM, Garcia MS, Kranaster R, et al. Rudimentary G-quadruplex-based telomere capping in *Saccharomyces cerevisiae*. *Nat Struct Mol Biol*. 2011 Apr;18(4):478–85.
41. Teng FY, Jiang ZZ, Guo M, Tan XZ, Chen F, Xi XG, et al. G-quadruplex DNA: a novel target for drug design. *Cell Mol Life Sci*. 2021 Oct 1;78(19):6557–83.
42. Linke R, Limmer M, Juranek SA, Heine A, Paeschke K. The Relevance of G-Quadruplexes for DNA Repair. *Int J Mol Sci*. 2021 Nov 22;22(22):12599.
43. Parkinson GN, Lee MPH, Neidle S. Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*. 2002 Jun;417(6891):876–80.
44. Dai J, Carver M, Punchihewa C, Jones RA, Yang D. Structure of the Hybrid-2 type intramolecular human telomeric G-quadruplex in K<sup>+</sup> solution: insights into structure polymorphism of the human telomeric sequence. *Nucleic Acids Res*. 2007;35(15):4927–40.
45. Wang ZF, Li MH, Hsu STD, Chang TC. Structural basis of sodium-potassium exchange of a human telomeric DNA quadruplex without topological conversion. *Nucleic Acids Res*. 2014 Apr;42(7):4723–33.
46. Lim KW, Ng VCM, Martín-Pintado N, Heddi B, Phan AT. Structure of the human telomere in Na<sup>+</sup> solution: an antiparallel (2+2) G-quadruplex scaffold reveals additional diversity. *Nucleic Acids Research*. 2013 Dec 1;41(22):10556–62.
47. 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 Apr;35(7):2440–50.
48. Mitra J, Makurath MA, Ngo TTM, Troitskaia A, Chemla YR, Ha T. Extreme mechanical diversity of human telomeric DNA revealed by fluorescence-force spectroscopy. *Proceedings of the National Academy of Sciences*. 2019 Apr 23;116(17):8350–9.
49. Huppert JL, Balasubramanian S. G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res*. 2007;35(2):406–13.
50. Eddy J, Maizels N. Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Res*. 2006;34(14):3887–96.
51. Brooks TA, Kendrick S, Hurley L. Making sense of G-quadruplex and i-motif functions in oncogene promoters. *FEBS J*. 2010 Sep;277(17):3459–69.
52. Simonsson T, Pecinka P, Kubista M. DNA tetraplex formation in the control region of c-myc. *Nucleic Acids Res*. 1998 Mar 1;26(5):1167–72.

53. Siddiqui-Jain A, Grand CL, Bearss DJ, Hurley LH. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc Natl Acad Sci U S A*. 2002 Sep 3;99(18):11593–8.
54. Phan AT, Modi YS, Patel DJ. Propeller-Type Parallel-Stranded G-Quadruplexes in the Human c-myc Promoter. *J Am Chem Soc*. 2004 Jul 1;126(28):8710–6.
55. Seenisamy J, Rezler EM, Powell TJ, Tye D, Gokhale V, Joshi CS, et al. The Dynamic Character of the G-Quadruplex Element in the c-MYC Promoter and Modification by TMPyP4. *J Am Chem Soc*. 2004 Jul 1;126(28):8702–9.
56. Sengupta P, Bhattacharya A, Sa G, Das T, Chatterjee S. Truncated G-Quadruplex Isomers Cross-Talk with the Transcription Factors To Maintain Homeostatic Equilibria in c-MYC Transcription. *Biochemistry*. 2019 Apr 16;58(15):1975–91.
57. Raiber EA, Kranaster R, Lam E, Nikan M, Balasubramanian S. A non-canonical DNA structure is a binding motif for the transcription factor SP1 in vitro. *Nucleic Acids Res*. 2012 Feb;40(4):1499–508.
58. Broxson C, Beckett J, Tornaletti S. Transcription Arrest by a G Quadruplex Forming-Trinucleotide Repeat Sequence from the Human c-myc Gene. *Biochemistry*. 2011 May 17;50(19):4162–72.
59. Duquette ML, Handa P, Vincent JA, Taylor AF, Maizels N. Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes Dev*. 2004 Jul 1;18(13):1618–29.
60. Lee CY, McNerney C, Ma K, Zhao W, Wang A, Myong S. R-loop induced G-quadruplex in non-template promotes transcription by successive R-loop formation. *Nat Commun*. 2020 Jul 7;11(1):3392.
61. Brooks TA, Hurley LH. The role of supercoiling in transcriptional control of MYC and its importance in molecular therapeutics. *Nat Rev Cancer*. 2009 Dec;9(12):849–61.
62. Kouzine F, Sanford S, Elisha-Feil Z, Levens D. The functional response of upstream DNA to dynamic supercoiling in vivo. *Nat Struct Mol Biol*. 2008 Feb;15(2):146–54.
63. Kumari S, Bugaut A, Huppert JL, Balasubramanian S. An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. *Nat Chem Biol*. 2007 Apr;3(4):218–21.
64. Huppert JL, Bugaut A, Kumari S, Balasubramanian S. G-quadruplexes: the beginning and end of UTRs. *Nucleic Acids Res*. 2008 Nov;36(19):6260–8.

65. Eddy J, Maizels N. Conserved elements with potential to form polymorphic G-quadruplex structures in the first intron of human genes. *Nucleic Acids Res.* 2008 Mar;36(4):1321–33.
66. Beaudoin JD, Perreault JP. Exploring mRNA 3'-UTR G-quadruplexes: evidence of roles in both alternative polyadenylation and mRNA shortening. *Nucleic Acids Res.* 2013 Jun;41(11):5898–911.
67. Kostadinov R, Malhotra N, Viotti M, Shine R, D'Antonio L, Bagga P. GRSDDB: a database of quadruplex forming G-rich sequences in alternatively processed mammalian pre-mRNA sequences. *Nucleic Acids Res.* 2006 Jan 1;34(Database issue):D119–24.
68. Maizels N, Gray LT. The G4 Genome. *PLOS Genetics.* 2013 Apr 18;9(4):e1003468.
69. Besnard E, Babled A, Lapasset L, Milhavet O, Parrinello H, Dantec C, et al. Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. *Nat Struct Mol Biol.* 2012 Aug;19(8):837–44.
70. Valton AL, Hassan-Zadeh V, Lema I, Boggetto N, Alberti P, Saintomé C, et al. G4 motifs affect origin positioning and efficiency in two vertebrate replicators. *EMBO J.* 2014 Apr 1;33(7):732–46.
71. Hoshina S, Yura K, Teranishi H, Kiyasu N, Tominaga A, Kadoma H, et al. Human Origin Recognition Complex Binds Preferentially to G-quadruplex-preferable RNA and Single-stranded DNA. *J Biol Chem.* 2013 Oct 18;288(42):30161–71.
72. Bochman ML, Sabouri N, Zakian VA. Unwinding the functions of the Pif1 family helicases. *DNA Repair (Amst).* 2010 Mar 2;9(3):237–49.
73. Wu Y, Shin-ya K, Brosh RM. FANCD1 helicase defective in Fanconi anemia and breast cancer unwinds G-quadruplex DNA to defend genomic stability. *Mol Cell Biol.* 2008 Jun;28(12):4116–28.
74. Lopes J, Piazza A, Bermejo R, Kriegsmann B, Colosio A, Teulade-Fichou MP, et al. G-quadruplex-induced instability during leading-strand replication. *The EMBO Journal.* 2011 Oct 5;30(19):4033–46.
75. Kwok CK, Merrick CJ. G-Quadruplexes: Prediction, Characterization, and Biological Application. *Trends in Biotechnology.* 2017 Oct 1;35(10):997–1013.
76. Weitzmann MN, Woodford KJ, Usdin K. The development and use of a DNA polymerase arrest assay for the evaluation of parameters affecting intrastrand tetraplex formation. *J Biol Chem.* 1996 Aug 23;271(34):20958–64.

77. Chambers VS, Marsico G, Boutell JM, Di Antonio M, Smith GP, Balasubramanian S. High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat Biotechnol.* 2015 Aug;33(8):877–81.
78. Tu J, Duan M, Liu W, Lu N, Zhou Y, Sun X, et al. Direct genome-wide identification of G-quadruplex structures by whole-genome resequencing. *Nat Commun.* 2021 Oct 14;12(1):6014.
79. Bugaut A, Rodriguez R, Kumari S, Hsu STD, Balasubramanian S. Small molecule-mediated inhibition of translation by targeting a native RNA G-quadruplex. *Org Biomol Chem.* 2010 Jun 21;8(12):2771–6.
80. Kouzine F, Wojtowicz D, Baranello L, Yamane A, Nelson S, Resch W, et al. Permanganate/S1 nuclease footprinting reveals non-B DNA structures with regulatory potential across a mammalian genome. *Cell Syst.* 2017 Mar 22;4(3):344–356.e7.
81. Kouzine F, Wojtowicz D, Yamane A, Casellas R, Przytycka TM, Levens DL. In Vivo Chemical Probing for G-Quadruplex Formation. *Methods Mol Biol.* 2019;2035:369–82.
82. Hänsel-Hertsch R, Di Antonio M, Balasubramanian S. DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential. *Nature Reviews Molecular Cell Biology.* 2017 May;18(5):279–84.
83. Biffi G, Tannahill D, McCafferty J, Balasubramanian S. Quantitative Visualization of DNA G-quadruplex Structures in Human Cells. *Nat Chem.* 2013 Mar;5(3):182–6.
84. Henderson A, Wu Y, Huang YC, Chavez EA, Platt J, Johnson FB, et al. Detection of G-quadruplex DNA in mammalian cells. *Nucleic Acids Res.* 2014 Jan;42(2):860–9.
85. Shivalingam A, Izquierdo MA, Marois AL, Vyšniauskas A, Suhling K, Kuimova MK, et al. The interactions between a small molecule and G-quadruplexes are visualized by fluorescence lifetime imaging microscopy. *Nat Commun.* 2015 Sep 9;6:8178.
86. Zhang S, Sun H, Wang L, Liu Y, Chen H, Li Q, et al. Real-time monitoring of DNA G-quadruplexes in living cells with a small-molecule fluorescent probe. *Nucleic Acids Res.* 2018 Sep 6;46(15):7522–32.
87. Di Antonio M, Ponjavic A, Radzevičius A, Ranasinghe RT, Catalano M, Zhang X, et al. Single-molecule visualization of DNA G-quadruplex formation in live cells. *Nat Chem.* 2020 Sep;12(9):832–7.
88. Summers PA, Lewis BW, Gonzalez-Garcia J, Porreca RM, Lim AHM, Cadinu P, et al. Visualising G-quadruplex DNA dynamics in live cells by fluorescence lifetime imaging microscopy. *Nat Commun.* 2021 Jan 8;12(1):162.

89. Verga D, Hamon F, Poyer F, Bombard S, Teulade-Fichou MP. Photo-Cross-Linking Probes for Trapping G-Quadruplex DNA. *Angewandte Chemie International Edition*. 2014;53(4):994–8.
90. Nadai M, Doria F, Germani L, Richter SN, Freccero M. A Photoreactive G-Quadruplex Ligand Triggered by Green Light. *Chemistry – A European Journal*. 2015;21(6):2330–4.
91. Lena A, Benassi A, Stasi M, Saint-Pierre C, Freccero M, Gasparutto D, et al. Photoactivatable V-Shaped Bifunctional Quinone Methide Precursors as a New Class of Selective G-quadruplex Alkylating Agents. *Chemistry – A European Journal*. 2022;28(35):e202200734.
92. Hänsel-Hertsch R, Beraldi D, Lensing SV, Marsico G, Zyner K, Parry A, et al. G-quadruplex structures mark human regulatory chromatin. *Nat Genet*. 2016 Oct;48(10):1267–72.
93. Hänsel-Hertsch R, Spiegel J, Marsico G, Tannahill D, Balasubramanian S. Genome-wide mapping of endogenous G-quadruplex DNA structures by chromatin immunoprecipitation and high-throughput sequencing. *Nat Protoc*. 2018 Mar;13(3):551–64.
94. Lago S, Nadai M, Cernilogar FM, Kazerani M, Domíniguez Moreno H, Schotta G, et al. Promoter G-quadruplexes and transcription factors cooperate to shape the cell type-specific transcriptome. *Nat Commun*. 2021 Jun 23;12(1):3885.
95. Zyner KG, Simeone A, Flynn SM, Doyle C, Marsico G, Adhikari S, et al. G-quadruplex DNA structures in human stem cells and differentiation. *Nat Commun*. 2022 Jan 10;13(1):142.
96. Su DGT, Fang H, Gross ML, Taylor JSA. Photocrosslinking of human telomeric G-quadruplex loops by anti cyclobutane thymine dimer formation. *Proc Natl Acad Sci USA*. 2009 Aug 4;106(31):12861–6.
97. 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 Apr;42(8):5007–19.
98. Lu C, Smith-Carpenter JE, Taylor JSA. Evidence for Reverse Hoogsteen Hairpin Intermediates in the Photocrosslinking of Human Telomeric DNA Sequences. *Photochem Photobiol*. 2018 Jul;94(4):685–97.
99. Smith-Carpenter JE, Taylor JS. Photocrosslinking of G-Quadruplex-Forming Sequences found in Human Promoters. *Photochem Photobiol*. 2018 Aug 7;95(1):252–66.



100. Love JD, Nguyen HT, Or A, Attri AK, Minton KW. UV-induced interstrand cross-linking of d(GT)<sub>n</sub>.d(CA)<sub>n</sub> is facilitated by a structural transition. *Journal of Biological Chemistry*. 1986 Aug;261(22):10051–7.
101. Nguyen HT, Minton KW. Extensive photodimerization of non-adjacent pyrimidines. *Journal of Molecular Biology*. 1989 Dec 20;210(4):869–74.
102. Nguyen HT, Minton KW. Ultraviolet-induced dimerization of non-adjacent pyrimidines: A potential mechanism for the targeted –1 frameshift mutation. *Journal of Molecular Biology*. 1988 Apr 20;200(4):681–93.
103. Love JD, Minton KW. Ultraviolet-induced dimerization of non-adjacent pyrimidines in poly[d(A-T)]. *Journal of Biological Chemistry*. 1992 Dec 15;267(35):24953–9.
104. Schreier WJ, Gilch P, Zinth W. Early Events of DNA Photodamage. *Annual Review of Physical Chemistry*. 2015;66(1):497–519.
105. Cadet J, Douki T. Formation of UV-induced DNA damage contributing to skin cancer development. *Photochem Photobiol Sci*. 2018 Dec 5;17(12):1816–41.
106. Douki T. The variety of UV-induced pyrimidine dimeric photoproducts in DNA as shown by chromatographic quantification methods. *Photochem Photobiol Sci*. 2013 Aug;12(8):1286–302.
107. Taylor JS, Cohrs MP. DNA, light, and Dewar pyrimidinones: the structure and biological significance to TpT3. *J Am Chem Soc*. 1987 Apr;109(9):2834–5.
108. Douki T, Sage E. Dewar valence isomers, the third type of environmentally relevant DNA photoproducts induced by solar radiation. *Photochemical & Photobiological Sciences*. 2016;15(1):24–30.
109. Cadet J, Mouret S, Ravanat JL, Douki T. Photoinduced Damage to Cellular DNA: Direct and Photosensitized Reactions†. *Photochemistry and Photobiology*. 2012;88(5):1048–65.
110. Lu C, Gutierrez-Bayona NE, Taylor JS. The effect of flanking bases on direct and triplet sensitized cyclobutane pyrimidine dimer formation in DNA depends on the dipyrimidine, wavelength and the photosensitizer. *Nucleic Acids Res*. 2021 Apr 13;49(8):4266–80.
111. Mao P, Brown AJ, Esaki S, Lockwood S, Poon GMK, Smerdon MJ, et al. ETS transcription factors induce a unique UV damage signature that drives recurrent mutagenesis in melanoma. *Nature Communications*. 2018 Jul 6;9(1):2626.
112. Becker MM, Wang JC. Use of light for footprinting DNA in vivo. *Nature*. 1984 Jun;309(5970):682–7.

113. Setlow P, Li L. Photochemistry and Photobiology of the Spore Photoproduct: A 50-Year Journey. *Photochem Photobiol.* 2015 Nov;91(6):1263–90.
114. Weinblum D, Johns HE. Isolation and properties of isomeric thymine dimers. *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis.* 1966 Mar 21;114(3):450–9.
115. Taylor JS. Adjacent and Nonadjacent Dipyrimidine Photoproducts as Intrinsic Probes of DNA Secondary and Tertiary Structure†. *Photochemistry and Photobiology* [Internet]. [cited 2022 Nov 23];n/a(n/a). Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/php.13694>
116. Douki T. Effect of denaturation on the photochemistry of pyrimidine bases in isolated DNA. *Journal of Photochemistry and Photobiology B: Biology.* 2006 Jan 2;82(1):45–52.
117. Lingbeck JM, Taylor JS. Preparation and Characterization of DNA Containing a Site-Specific Nonadjacent Cyclobutane Thymine Dimer of the Type Implicated in UV-Induced –1 Frameshift Mutagenesis. *Biochemistry.* 1999 Oct 1;38(41):13717–24.
118. Su DGT, Kao JLF, Gross ML, Taylor JSA. 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 Aug 27;130(34):11328–37.
119. 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 Jun 15;31(12):3134–42.
120. Fujimoto M, Fujiyama K, Kuninaka A, Yoshino H. Mode of Action of Nuclease P<sub>1</sub> on Nucleic Acids and Its Specificity for Synthetic Phosphodiester. *Agricultural and Biological Chemistry.* 1974 Nov;38(11):2141–7.
121. Lim KW, Amrane S, Bouaziz S, Xu W, Mu Y, Patel DJ, et al. Structure of the Human Telomere in K<sup>+</sup> Solution: A Stable Basket-Type G-Quadruplex with Only Two G-Tetrad Layers. *J Am Chem Soc.* 2009 Apr 1;131(12):4301–9.
122. Hou XM, Fu YB, Wu WQ, Wang L, Teng FY, Xie P, et al. Involvement of G-triplex and G-hairpin in the multi-pathway folding of human telomeric G-quadruplex. *Nucleic Acids Res.* 2017 Nov 2;45(19):11401–12.
123. Rhodes D, Lipps HJ. G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.* 2015 Oct 15;43(18):8627–37.

# **Chapter 2: Post- and Pre-radiolabeling assays for *anti* thymidine cyclobutane dimers as intrinsic photoprobes of various types of G-quadruplexes, reverse Hoogsteen hairpins, and other non-B DNA structures<sup>1</sup>**

## **2.1 Abstract**

G quadruplexes are thought to play an important role in gene regulation and telomere maintenance, but compelling evidence for their existence and location is lacking owing to their highly dynamic nature and the lack of non-perturbing probes. The majority of probes for G quadruplexes to date have relied on antibody or small molecule binding agents that can also alter the dynamics and relative populations of G quadruplexes. Recently, we discovered that UVB irradiation of human telomeric DNA and various G-quadruplex forming sequences found in human promoters, as well as reverse Hoogsteen hairpins, results in a unique class of non-adjacent *anti* cyclobutane pyrimidine dimers (CPDs) which can serve as stable molecular markers. One can therefore envision using a pulse of light to irreversibly trap these non-B DNA structures via *anti* CPD formation without perturbing their dynamics, after which the CPDs could be identified and mapped to provide unambiguous evidence for the formation and location of these structures *in vivo*. As a first step towards this goal, we report the development of radioactive post- and pre-labeling assays

---

<sup>1</sup> This chapter has been submitted for publication in Gutierrez-Bayona NE, Scruggs SS, Yang HS, Gross ML, Taylor JS. Post- and Pre-radiolabeling assays for *anti* thymidine cyclobutane dimers as intrinsic photoprobes of various types of G-quadruplexes, reverse Hoogsteen hairpins, and other non-B DNA structures

for the detection of *anti* CPDs and illustrate their use in detecting *trans,anti* T=T CPD formation in a human telomeric DNA sequence. Both assays make use of snake venom phosphodiesterase (SVP) to degrade the *trans,anti* T=T CPD to pTpT=(pT)pT. In the post-labeling assay, calf intestinal phosphodiesterase is used to dephosphorylate the tetramer which is then rephosphorylated with kinase and [<sup>32</sup>P]ATP to produce a mixture of radiolabeled mono and diphosphorylated tetramers. The tetramers are confirmed to contain non-adjacent CPDs by 254 nm photoreversion to the dinucleotide p\*TpT. In the pre-labeling assay, radiolabeled phosphates are introduced into *anti* T=T CPD forming sites by ligation of 5'-end-labeled oligodeoxynucleotides and only require treatment with SVP. We also demonstrate that the assays can detect the stereoisomeric *cis,anti* T=T CPD.

## 2.2 Introduction

Non-B DNA secondary structures play many roles *in vivo*, such as Holliday junctions in recombination and slipped structures in mutagenesis (1–4). Most recently, G-quadruplexes have been proposed to play an important role in telomere maintenance and modulating transcription and translation, as well as causing genome instability (5–7). Unfortunately, most non-B DNA structures exist transiently *in vivo* and are difficult to detect and locate. Small molecule and antibody probes for detecting G-quadruplexes *in vivo* have been reported (8–15), but they are of questionable utility because these reagents drive the formation of the structures that they are seeking to probe and because of disruption to the cell caused by some assay methods.

Recently it was discovered that UVB irradiation of G-quadruplex forming sequences from human telomeres and various promoters results in the formation of a unique class of non-adjacent *anti* cyclobutane pyrimidine dimers (CPDs) with a head to tail orientation (16–21). The *anti*

regiochemistry is opposite to the head-to-head orientation of the adjacent *cis,syn* CPDs formed in B form DNA (Fig. 2.1a). The *anti* CPDs were identified by a combination of tandem mass spectrometry (MS/MS) analysis of nuclease P1 digestion products pY(pN)=pY(pN) and correlation of the thymine dimers released by acid-catalyzed glycosidic bond hydrolysis with authentic compounds. Non-adjacent dipyrimidine DNA photoproducts have also been detected in UVC-irradiated ethanolic and desiccated DNA by an enzyme-coupled LC-MS/MS assay that involved degrading the DNA to pyrimidine dimers using nuclease P1, snake venom phosphodiesterase and alkaline phosphatase (22).

The *anti* T=T CPDs formed in telomeric DNA were proposed to arise from basket type G-quadruplexes in which a pyrimidine from one lateral loop photocyclizes with a non-adjacent pyrimidine on the opposing lateral loop resulting in an interloop crosslink (19,21). It was later found that *anti* T=T CPDs can also form between T's of interior loops of reverse Hoogsteen hairpin structures (18). One could, therefore, envision irradiating cells with a pulse of UV light to trap the otherwise unstable non-B DNA secondary structures by *anti* CPD formation for subsequent analysis (Fig. 2.1b). Because of the dynamic and transient nature of the non-B DNA structures, *anti* CPDs are expected to form in small quantities requiring highly sensitive techniques to detect and map their location.

One highly sensitive method for detecting DNA damage is the radioactive post-labeling assay that has been estimated to be capable of detecting one DNA adduct per human genome (23–26). In this assay, DNA is enzymatically degraded to short oligodeoxynucleotides or mononucleotides containing the DNA adduct. The nucleotides are rephosphorylated with radioactive <sup>32</sup>P and identified chromatographically or electrophoretically by comparison to authentic adducts. It was

previously shown that degradation of UV irradiated B DNA with a combination of snake venom phosphodiesterase (SVP) and DNase I produces a trinucleotide of the form pNpY=pY (Fig. 2.2a) (27). The canonical nucleotide at the 5'-end makes it a good substrate for dephosphorylation with calf intestinal phosphatase and rephosphorylation with [<sup>32</sup>P]-ATP and kinase (27). We previously showed that SVP degrades the *cis,anti* T2=T7 CPD product of d(GTATCATGAGGTGC) to the tetranucleotide GpT=(pA)pT (corresponding to the photodimer of GpT and pApT) (20), which also has canonical nucleotides at the 5'-end.

Herein, we report a post-labeling assay for *trans,anti* T=T CPDs formed in human telomeric DNA using SVP and calf intestinal phosphatase to degrade the irradiated DNA to TpT=(Tp)T that can then be rephosphorylated with [ $\gamma$ -<sup>32</sup>P]ATP and kinase (Fig. 2.2b). Furthermore, we show that these tetramer products can be distinguished by high resolution gel electrophoresis from partial degradation products of DNA by subsequent photoreversal to pTpT dinucleotides by UVC light. We also show that a *cis,anti* T=T CPD can be similarly detected, and that *anti* CPD formation can also be easily assayed *in vitro* using pre-labeled DNA substrates.

## 2.3 Materials and Methods

### 2.3.1 Materials and Reagents

Oligodeoxynucleotides (ODNs) were purchased from Integrated DNA Technologies, Inc. (IDT, Coralville, IA, USA). Snake venom phosphodiesterase (SVP) from *Crotalus adamanteus* was obtained from Worthington Biochemical (Lakewood, NJ, USA). DNase I (RNase- free) was obtained from New England Biolabs (Ipswich, MA, USA). Calf intestinal phosphatase and T4 DNA ligase were obtained from Promega (Madison, WI, USA). T4 Polynucleotide Kinase (T4

PNK) and HPLC solvents were obtained from ThermoFisher Scientific (Waltham, MA, USA) and 6000 Ci/mmol [ $\gamma$ - $^{32}\text{P}$ ]-ATP was from Perkin Elmer (Waltham, MA, USA).

### **2.3.2 Large Scale Preparation of pTpT=(pT)pT**

Tel26 (50  $\mu\text{M}$ ) was incubated in 100  $\mu\text{L}$  of 150 mM KCl at 95  $^{\circ}\text{C}$  for 10 min and then rapidly cooled in ice to preferentially form intramolecular folded structures. Samples were left on ice for at least 1 h before irradiation with broadband UVB/UVA light (275–400 nm, centered at 330 nm). Irradiation was carried out in polyethylene microfuge tubes on a bed of ice for 2.5 h at 1 cm distance from two Spectroline XX-15B UV 15W tubes with an approximate intensity of 2.2  $\text{mW}/\text{cm}^2$  (22  $\text{J}/\text{m}^2\text{-s}$ ) after filtration through a plate of Pyrex glass. The solution was evaporated to dryness *in vacuo* and then incubated with 20 units of DNase I in 10 mM Tris-HCl (pH 7.6), 2.5 mM  $\text{MgCl}_2$  and 0.5 mM  $\text{CaCl}_2$  at 37  $^{\circ}\text{C}$  overnight followed by inactivation at 75  $^{\circ}\text{C}$  for 20 min. The digested samples were dried *in vacuo* and then dissolved in 100  $\mu\text{L}$  of 10 mM ammonium citrate, 100  $\mu\text{M}$   $\text{MgCl}_2$ , pH 9.4, to which 10 units of SVP were added and incubated at 37  $^{\circ}\text{C}$  overnight. SVP was deactivated in a boiling water bath for 20 min and removed by phenol extraction after which residual phenol were removed by multiple diethyl ether extractions. The tetramer photoproduct pTpT=(pT)pT was then isolated and purified by HPLC and analyzed by MS/MS as described below.

### **2.3.3 Large Scale Dephosphorylation of pTpT=(pT)pT with Calf Intestinal Phosphatase**

The HPLC fractions containing pTpT=(pT)pT were evaporated *in vacuo* and incubated for 3 h at 37  $^{\circ}\text{C}$  with 6 units of calf intestinal phosphatase (CIP) in 40  $\mu\text{L}$  of 50 mM Tris-HCl (pH 9.3), 1 mM  $\text{MgCl}_2$ , 100  $\mu\text{M}$   $\text{ZnCl}_2$ , and 1 mM spermidine. After incubation, the enzyme was inactivated

by heating to 80 °C for 20 min and removed by phenol extraction after which residual phenol was removed by multiple diethyl ether extractions. The dephosphorylated product TpT=(T)pT was purified by HPLC and analyzed by MS/MS as described below.

#### **2.3.4 Large Scale Phosphorylation of TpT=(T)pT with T4 PNK.**

HPLC fractions containing TpT=(TpT) were evaporated to dryness *in vacuo* and then incubated with 50 units of T4-PNK in 40 µL of 1 mM adenosine triphosphate (ATP), 50 mM Tris-HCl (pH 7.6), 10 mM MgCl<sub>2</sub>, 5 mM DTT, 100 µM spermidine for 3 h at 37°C. After incubation, the enzyme was inactivated by incubation in a boiling water bath for 20 min and removed by phenol extraction after which traces of phenol removed by multiple diethyl ether extractions. The rephosphorylated product pT(pT)=pTpT was analyzed by HPLC as described below.

#### **2.3.5 HPLC Analysis and Purification of the Post-labeling Assay Intermediates.**

Reverse-phase HPLC was carried out with an X-Bridge BEH column (C18, 4.6 x 75 mm, 3.5 µm, 130 Å) on a System Gold HPLC system with a binary gradient Model 125 pump and a Model 166 UV detector (Beckman Coulter, Inc., Fullerton, CA). Photoproducts 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 by absorbance at 260 nm. HPLC fractions corresponding to the enzymatically degraded photoproducts were evaporated *in vacuo*.

#### **2.3.6 Mass Spectrometry of the Post-labeling Assay Intermediates.**

The CPD samples were analyzed in the negative-ion mode on a Waters Synapt-G2 mass spectrometer (Waters, Milford, MA). The source and desolvation temperatures were 120 °C and



150 °C, respectively. A solution of 50/50 (vol/vol) acetonitrile/water was used as the spray solvent. The capillary voltage was 2.8 kV. The sampling cone and extraction cone voltage were 20 V and 4 V, respectively. MS/MS experiments were carried out by collisionally activated dissociation (CAD). The mass width for precursor selection was set at  $m/z$  3, and the collision gas was Argon at a pressure of  $1 \times 10^{-4}$  mbar. The trap collision energy was tuned to 20V where both the precursor ions and fragmented ions were observed on the same spectrum.

### **2.3.7 Post-labeling Assay of UVB Irradiated DNA.**

Tel26 (500 pmol) was incubated at 95 °C for 10 min in 20  $\mu$ L of 150 mM KCl and plunged into an ice water bath for 1 h after which it was irradiated with UVB light for 2.5 h on ice. The solution was then incubated with 20 units of DNaseI in 20  $\mu$ L of 10 mM Tris-HCl (pH 7.6), 2.5 mM MgCl<sub>2</sub>, and 0.5 mM CaCl<sub>2</sub> at 37 °C overnight, followed by inactivation of the DNase I at 75 °C for 20 min. After drying the sample *in vacuo*, it was incubated with 10 units of SVP in 30  $\mu$ L of 10 mM ammonium citrate, 100  $\mu$ M MgCl<sub>2</sub> pH 9.4 at 37 °C overnight. The SVP was inactivated in a boiling water bath for 20 min and then treated with 6 units of CIP in 40  $\mu$ L of 50 mM Tris-HCl (pH 9.3), 1 mM MgCl<sub>2</sub>, 100  $\mu$ M ZnCl<sub>2</sub>, and 1 mM spermidine at 37 °C for 3 h. The enzymes were removed by phenol extraction, and residual phenol was removed by multiple diethyl ether extractions. After evaporating the sample *in vacuo*, it was treated with 10 units of T4-PNK and 0.5 pmol [ $\gamma$ -<sup>32</sup>P]-ATP in 40  $\mu$ L of 50 mM Tris-HCL (pH 7.6), 10 mM MgCl<sub>2</sub>, 5 mM DTT, 0.1 mM spermidine, 0.1 mM EDTA at 37 °C for 3 h. An aliquot of the sample (4  $\mu$ L) was irradiated for 30 min with 254 nm light to reverse any CPDs. Unirradiated and irradiated aliquots (4  $\mu$ L) were then mixed with an equal volume of 2X loading buffer (98% formamide with xylene cyanol dye),

heated in a boiling water bath for 5 min, and cooled on ice. The radiolabeled samples were then analyzed on a 20% denaturing polyacrylamide gel (7 M urea, 19:1 acrylamide:bisacrylamide, 0.8 mm thick) at 1500 V until the dye reached 8 cm. The gel was then scanned with an Amersham Typhoon biomolecular imager.

### **2.3.8 Post-labeling Assay with Competitor Plasmid DNA.**

Competition experiments were carried out at a fixed base pair concentration (650  $\mu$ M bp) in 100  $\mu$ L of 150 mM KCl with varying Tel26: plasmid ratios from 1:1 to 1:1000. The solutions were heated to 95 °C and then plunged into an ice bath for 1 h to induce G-quadruplex formation after which they were irradiated with UVB light for 2.5 h on ice. After incubation with DNaseI and SVP following the conditions described above, an aliquot of 10  $\mu$ L (6.5 nmol bp) from each dilution was then treated with CIP in a total volume of 40  $\mu$ L as described above. Subsequent radiolabeling and gel electrophoresis was also performed exactly as described above except that the amount of [ $\gamma$ -<sup>32</sup>P]-ATP was increased to 5 pmol.

### **2.3.9 Preparation of Internally Labelled Tel26 Sequences.**

Tel26, d(AAA(GGGTTA)<sub>3</sub>GGGAA), was separately internally radiolabeled at the second T of the 1<sup>st</sup> and 2<sup>nd</sup> TTA loop. To introduce the radioactive phosphate in the first TTA loop, the 5'-end of the 19-mer sequence d(TA(GGGTTA)<sub>2</sub>GGGAA) (500 pmol) was first radiolabeled with [<sup>32</sup>P]-ATP and PNK and then ligated to the 3'-end of the 7-mer sequence d(AAAGGGT) (1500 pmol) in the presence of the 30-mer scaffold d(TTCCC(TAACCC)<sub>3</sub>TTTTTTT) (1500 pmol). The ligation reaction was performed with 4 units of T4 DNA ligase in 30  $\mu$ L of 30 mM Tris-HCl (pH 7.8), 10 mM MgCl<sub>2</sub>, 10 mM DTT, 1 mM ATP at 4 °C overnight. After incubation, the ligase was

deactivated at 75 °C for 20 min and dried *in vacuo*. The same general procedure was used to introduce the radioactive phosphate into the second TTA loop by using 5'-end labeled 13A-mer d(TAGGGTTAGGGAA) and 13B-mer d(AAAGGGTTAGGGT). The ligation products were purified by denaturing 15% denaturing PAGE (19:1 acrylamide: bisacrylamide) being careful to separate the 26-mer band from the 30-mer scaffold band. The radiolabeling of the 13-mer and 19-mer sequences was carried out with 10 units of T4-PNK and 2.5 pmol [ $\gamma$ - $^{32}$ P]-ATP in 30  $\mu$ L of 50 mM Tris-HCL (pH 7.6), 10 mM MgCl<sub>2</sub>, 5 mM DTT, 0.1 mM spermidine, 0.1 mM EDTA at 37°C for 1 h. This reaction was chased with 750 pmol of cold ATP and 5 units of T4-PNK for 1 h at 37 °C. After incubation, the kinase was deactivated for 20 min in a boiling water bath, and the solution dried *in vacuo*.

### **2.3.10 SVP degradation of internally labeled UVB irradiated Tel26 sequences.**

Purified ligation products of Tel26 radiolabeled in the 1<sup>st</sup> and 2<sup>nd</sup> TTA loop were irradiated with UVB light after G-quadruplex induction in 150 mM LiCl, which produces predominantly the *trans,anti* T=T CPD. The sequences were then digested with DNaseI and SVP following the conditions described above. Parts of the digested products from both sequences were treated with 0.01 units of CIP in 30  $\mu$ L of 50 mM Tris-HCl (pH 9.3), 1 mM MgCl<sub>2</sub>, 100  $\mu$ M ZnCl<sub>2</sub>, and 1 mM spermidine for 2.5 min at room temperature and immediately quenched in dry ice. CIP was deactivated in 80 °C for 20 min and dried *in vacuo*. A portion of the digested products and the dephosphorylated products was irradiated for 30 min with 254 nm light to reverse any CPDs present.

### 2.3.11 SVP degradation and labeling studies with UVB irradiated 14-mer.

The 14-mer d(GTATCATGAGGTGC) (100  $\mu$ M bp, 100  $\mu$ L) was irradiated on ice in 10 mM sodium acetate, pH 4.8 either before or after 5'-endlabeling with 0.5pmol of [ $\gamma$ - $^{32}$ P]-ATP and kinase. Degradation and post-labeling experiments were carried out as generally described above for Tel26.

## 2.4 Results and Discussion

To validate the proposed enzymatic sequence for post-labeling the *trans,anti* T=T CPD, the human telomere containing sequence Tel26, d(AAAGGGTTAGGGTTAGGGTTAGGGAA), was UVB irradiated in large scale to produce the *trans,anti* T=T CPD as the major product (18,19,21) so that the intermediates could be characterized by MS/MS. The irradiated sample was then digested with SVP to produce a major product which eluted at 34.2 min with reversed-phase HPLC (Fig. 2.3a). ESI-MS/MS of the product (Fig. 2.4a) showed an  $[M - 2H]^{2-}$  ion of  $m/z$  625, which fragmented by an electrocyclic elimination reaction to dehydrothymidinephosphate  $[pT - H_2O - H]^-$  at  $m/z$  321 and  $[M - (pT - H_2O) - H]^-$  at  $m/z$  947 (Fig. 2.5) confirming it to be  $pTpT=(pT)pT$ . A fragment ion of  $m/z$  643 was also observed that results from the loss of two dehydrothymidinephosphate groups  $[M - 2(pT - H_2O) - H]^-$  which corresponds to the monoanion of the CPD of thymidine monophosphate  $[pT=pT - H]^-$ . Treatment of  $pTpT=(pT)pT$  with calf intestinal phosphatase (CIP) resulted in a slower eluting product with a retention time of 49 min (Fig. 2.3b). ESI-MS/MS of this product confirmed the removal of the two 5'-phosphates to give  $TpT=(T)pT$  by a molecular ion  $[M - 2H]^{2-}$  at  $m/z$  545, which fragmented to  $[M - (T - H_2O)-H]^-$  at  $m/z$  867 and to  $[M - 2(T -$

H<sub>2</sub>O) - H]<sup>-</sup> corresponding to [pT=pT-H]<sup>-</sup> at *m/z* 643 (Fig. 2.4b, 2.5). Treatment of TpT=(T)pT with kinase and ATP converted it back to pTpT=(pT)pT as confirmed by HPLC (Fig. 2.3c).

Treatment of the dephosphorylated tetramer with [ $\gamma$ -<sup>32</sup>P]ATP and polynucleotide kinase, however, resulted in two radiolabeled bands on high resolution denaturing polyacrylamide gel electrophoresis, a fast-moving band and much more slowly moving band (Fig. 2.6 lane 5). Exposure of this mixture of products to UVC (254 nm) light prior to electrophoresis converted both products to a single faster moving band corresponding to the dinucleotide p\*TpT, which was expected to be produced by photoreversal of p\*TpT=(p\*T)pT CPD (Fig. 2.6 lane 6). The formation of two radiolabeled products that photoreverted to the same product suggests that the two initial products were the mono and diphosphorylated products p\*TpT=(T)pT and p\*TpT=(p\*T)pT, respectively (vide infra). Formation of the two radiolabeled products confirms that the *trans,anti* T=T CPD of the human telomere sequence can be detected by a post-labeling assay involving sequential treatment with SVP, CIP and polynucleotidyl kinase and [ $\gamma$ -<sup>32</sup>P]ATP.

To determine whether the post-labeling assay would work in the presence of competitor DNA, we irradiated decreasing amounts of Tel26 in the presence of plasmid DNA at a total concentration of 650  $\mu$ M bp. The samples were then treated with DNase I and SVP, followed by CIP and then kinase and [ $\gamma$ -<sup>32</sup>P]ATP. DNase I was added to help digest the plasmid DNA. In this case, only the slower moving radiolabeled product was observed (Fig. 2.6 lanes 7, 9, 11, 13). To confirm that the slower moving band was radiolabeled p\*TpT=(T)pT, and to distinguish it from radiolabeled partially digested ODNs, each sample was irradiated with UVC light to photorevert the tetramer products to dinucleotides. As can be seen, the slow-moving band decreases in intensity by the same amount that the fast-moving band corresponding to p\*TpT increases in intensity. The

*trans,anti* T=T CPD could be detected down to a 40 fmol level by quantification of the decrease of the band volume of the slower moving band and the increase of the band volume of the faster moving band in a 5 min exposure of the gel (1:100 lane). It was even possible to detect a 10-fold lower amount (4 fmol) of the *trans,anti* T=T CPD in the gel shown by quantifying the bands in the 1:1000 dilution lanes with a 24 h exposure. On the other hand, it was not possible to do so in another gel. We believe that this limit could be achieved or even further reduced, however, by pre-purification of the non-adjacent tetranucleotides before pre-labeling, and/or isolation of the radiolabeled tetranucleotide band prior to photoreversion to eliminate any background for the photoreverted dinucleotide bands.

The lower limit of about 4 fmol is about three times higher than 1.5 fmol of *trans,anti* T=T CPD that would be produced if one *trans,anti* T=T CPD were to be formed per telomere in 10 million human cells that have 92 telomeres/cell. There would be much more background interference from partially digested DNA and adjacent *cis,syn* T=T CPD trinucleotides if the post-labeling assay were to be carried out using human genomic DNA, but this could be circumvented by isolating the telomeres prior to the post-labeling assay (28). Interestingly, comparatively lower amounts of radiolabeled adjacent CPD products p\*NY=Y were observed in comparison to the radiolabeled p\*TpT=(T)pT product at 1:1 dilution even though GpTp=T is found to be efficiently labeled (lane 3). The lower amount of radiolabeled *cis,syn* CPDs could either be due to the inefficiency of SVP digestion of DNaseI digestion products which may be largely duplex, and/or that non-adjacent CPDs form in much higher yield than the adjacent CPDs.

To confirm the identity of the slow and fast migrating radiolabeled bands as either fully or hemiphosphorylated p\*TpT=(p\*T/T)pT internally labeled substrates were used in which the

internucleotide phosphate between the two T's in either the first TTA or the second TTA loop was radiolabeled (Fig. 2.7a). Because this substrate is internally labeled, irradiation followed by SVP would directly produce the radiolabeled diphosphorylated product  $pTp^*T=(pT)pT$ , without requiring desphosphorylation and rephosphorylation (Fig. 2.7b). The desired substrates were prepared by ligating either the 5'-end-labeled 19-mer (for loop 1) or the 5'-end-labeled 13-mer (for loop 2) to the 7-mer or 13-mer, respectively, using the 30-mer ligation scaffold (Fig. 2.7a). The radiolabeled Tel26 strand was then carefully separated by denaturing PAGE from the 30-mer strand, which would otherwise form a duplex with Tel26 and inhibit folding. As one can see from Lane 8 in Figure 2.8 UVB irradiation of Tel26 internally labeled in loop 1 followed by SVP produces the fast-moving product, which is converted to  $pTp^*T$  following photoreversal with UVC light, thereby identifying it as the doubly phosphorylated product  $pTp^*T=(pT)pT$ . A band corresponding to  $p^*T$  resulting from degradation of non-photoreacted DNA at that site was also observed, as well as a trimer corresponding to  $pGpT=p^*T$ . The same bands were also observed in Tel-26 subjected to the standard post-labeling protocol and using a chase with cold ATP to ensure complete phosphorylation (Fig. 2.8 lanes 4 & 5). Strangely, a small amount of what appears to be the mono-phosphorylated tetramer product was also observed in the pre-labeled products digested with SVP which disappears upon photoreversal, along with slower moving bands only seen following treatment with CIP (vide infra). How the mono-phosphorylated tetramer product and the higher bands would form is not obvious unless there was some phosphatase present in the commercial SVP preparation or in some other source. As expected, bands corresponding to  $pTp^*T=(pTpT)$  and  $Tp^*T=(pT)pT$  were not observable in the Tel26 loop 2 pre-labeled substrate

because *anti* CPD formation can only take place between loops 1 and 3 of a chair or basket G-quadruplex or a reverse Hoogsteen hairpin.

To confirm the identity of the slow-moving band from the post-labeling assay as the mono-phosphorylated product, samples containing the fast-moving band were incubated with CIP for various times to remove the 5'-phosphates. As one can see from Figure 2.8 lanes 6 & 7, the post-labeled products treated with CIP converts the fast-moving product to the slow-moving one, both of which disappear upon photoreversal with concomitant appearance of the p\*TpT product. The same is observed for the products of Tel26 internally labeled in loop 1 (Fig. 2.8 lane 11) except that the major products are bands that are migrating even slower that can be assigned to the completely dephosphorylated product Tp\*T=(T)pT and the dephosphorylated trimer product GpT=p\*T. There is also one additional band that cannot be assigned. Photoreversal of Tp\*T=TpT to Tp\*T with UVC results in an even slower moving product, while photoreversal of GpTp\*=T causes no observable change in mobility (Fig. 2.8 lane 12). The additional unassigned band disappears upon UVC treatment. Treatment of the Tel26 products internally labeled in loop 2 with CIP only produces GpT\*pT and essentially none of the products expected from an *anti* CPD's because they can only form between loops 1 and 3.

The large differences in mobility between the various enzymatic degradation products can be readily understood in terms of their mass to charge ratio (Figure 2.7b). Removal of a single 5'-phosphate group from pTpT=(pT)pT results in a loss of two negative charges but a minor change in molecular weight causing a substantial reduction in electrophoretic mobility. Removal of a second 5'-phosphate result in a loss of an addition two negative charges, resulting in a tetranucleotide with only two negative charges causing a further reduction in the electrophoretic



mobility. Large retardations in electrophoretic mobility upon removal of a 5'-phosphate from oligodeoxynucleotides have been observed with 3'-end-labeled substrates (29).

To determine whether the *cis* stereoisomer of the *anti* thymidine CPDs can also be detected by the pre- or post-labeling assays, we carried out the assay on the 14-mer sequence, d(GTATCATGAGGTGC), which we previously shown to form the *cis,anti* CPD between T2 and T7 (underlined) when UVB irradiated under low pH (Fig. 2.9). Because SVP degradation of the *cis,anti* T2=T7 CPD includes the 5'-terminal G, we could either pre-label the sequence with kinase and [ $\gamma$ - $^{32}$ P]ATP, or post-label the GpT=(pA)pT product following SVP degradation. When the pre-labeled product was UVB irradiated and then treated with SVP, a major fast migrating band was observed, which was converted to a faster moving band that migrated as p\*GpT upon UVC treatment (lane 6). Treatment of the SVP degraded product with CIP caused the fast-moving band to disappear and a slow-moving band to appear, both of which disappeared upon UVC irradiation with the formation of the p\*GpT band. This behavior is consistent with the fast-moving SVP degraded band being p\*GpT=(pA)pT and the slow-moving band being p\*GpT=(A)pT. The band moving like a nucleotide monophosphate in lanes 5-12 must be p\*G resulting from SVP degradation of the undamaged 14-mer.

When the post-labeling assay was carried out on the 14-mer, the fast-moving band corresponding to p\*GpT=(pA)pT was observed (Fig. 2.9b, lane 13). Upon treatment with UVC light the expected p\*GT band appeared along with a faster moving band (lane 14). The only reasonable assignment of this band is that it belongs to p\*ApT, which resulted from photoreversal of p\*GpT=(p\*A)pT that had probably become radiolabeled at A by a known phosphate exchange reaction with [ $\gamma$ - $^{32}$ P]ATP catalyzed by PNK (30). Further treatment with CIP resulted in the formation of

p\*GpT=(A)pT or GpT=(p\*A)pT and eventual loss of all labeling as in the pre-labeled case (lane 15). The apparent efficiency of the exchange labeling reaction suggests that the CIP step could be eliminated from the post-labeling sequence, thereby shortening the procedure. The formation of p\*GpT and p\*ApT as separable bands following photoreversal of p\*GpT=(p\*A)pT also demonstrates how *anti* T=T CPDs formed in different sequence contexts can be distinguished and quantified.

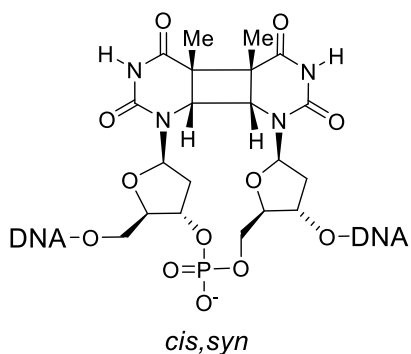
## 2.5 Conclusion

We have developed a radioactive post-labeling assay for the 10 fmol level detection of non-adjacent *trans,anti*- and *cis,anti*-T=T CPDs indicative of non-B DNA folded structures in the presence of adjacent *cis-syn* CPDs indicative of native B DNA. Furthermore, the ability of UVC light to photorevert the *anti* CPD-containing tetranucleotide products to faster moving dinucleotide products can be used to confirm their identity as non-adjacent CPDs and not partially degraded DNA, or other non-adjacent photoproducts such as (6-4) or Dewar products that cannot be photoreverted. This post-labeling method is also expected to work for non-adjacent *anti* T=U, U=T and U=U CPDs resulting from deaminated C-containing *anti* CPDs. It remains to be established whether this method will also work for regio-isomeric non-adjacent *cis,syn* and *trans,syn* CPDs. Because the assay detects *anti*-CPDs, it is specific for the presence of chair, basket, or reverse Hoogsteen hairpins containing pyrimidines in adjacent loops of sufficient length and cannot detect such structures with short loops or lacking pyrimidines, as well as parallel or hybrid G quadruplex structures. This specificity will allow for detection of non-B DNA structures not detected or identifiable by other methods. The main advantage of our method, however, is that the dynamics of G-quadruplexes is not perturbed in the probe step. Both post-labeling and pre-

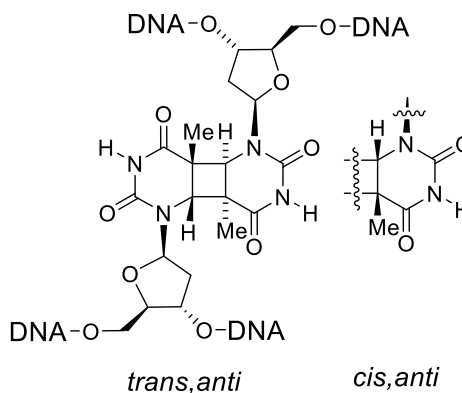
labeling assays for *anti* CPDs should also find application for *in vitro* studies of non-B structures. Once *anti* CPDs can be verified to form *in vivo*, Next-Gen sequencing assays can be developed to map the location of these products, and hence non-B.

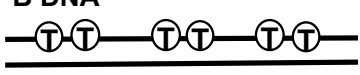
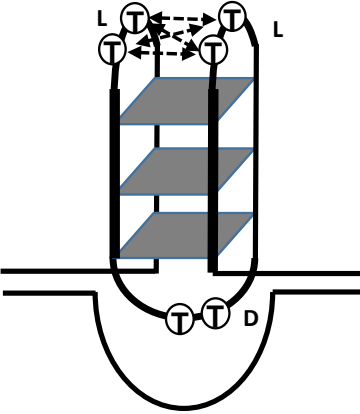
## 2.6 Figures

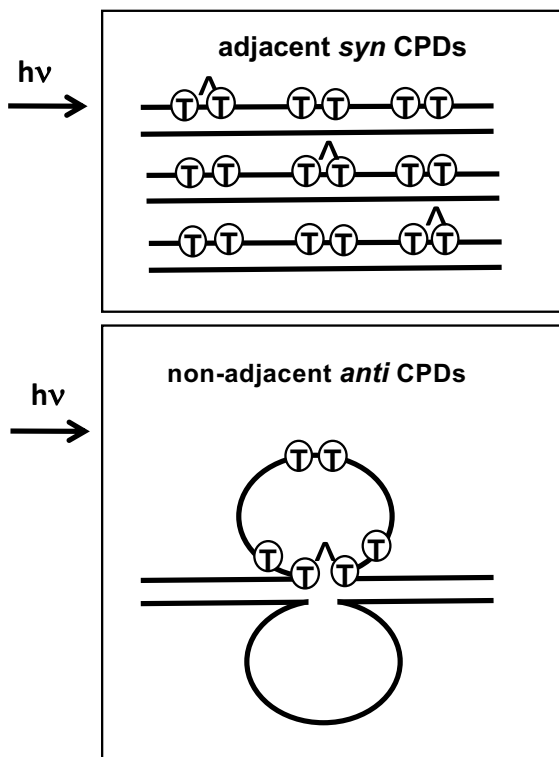
**A.** Adjacent *syn* Tp=T CPD



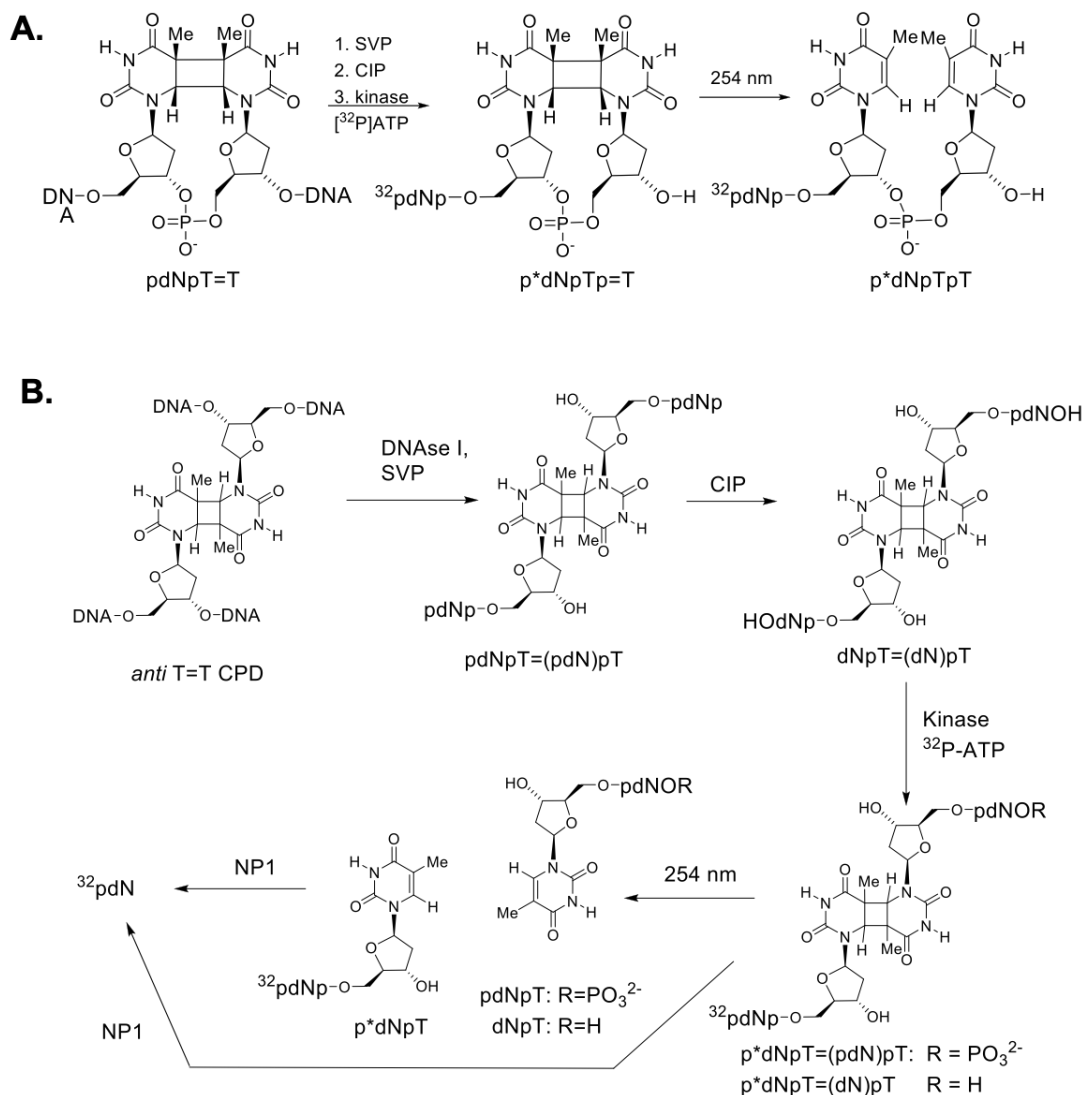
Non-Adjacent *anti* Tp=T CPD



**B.** B DNA  
  
  
 Folded non-B DNA structure

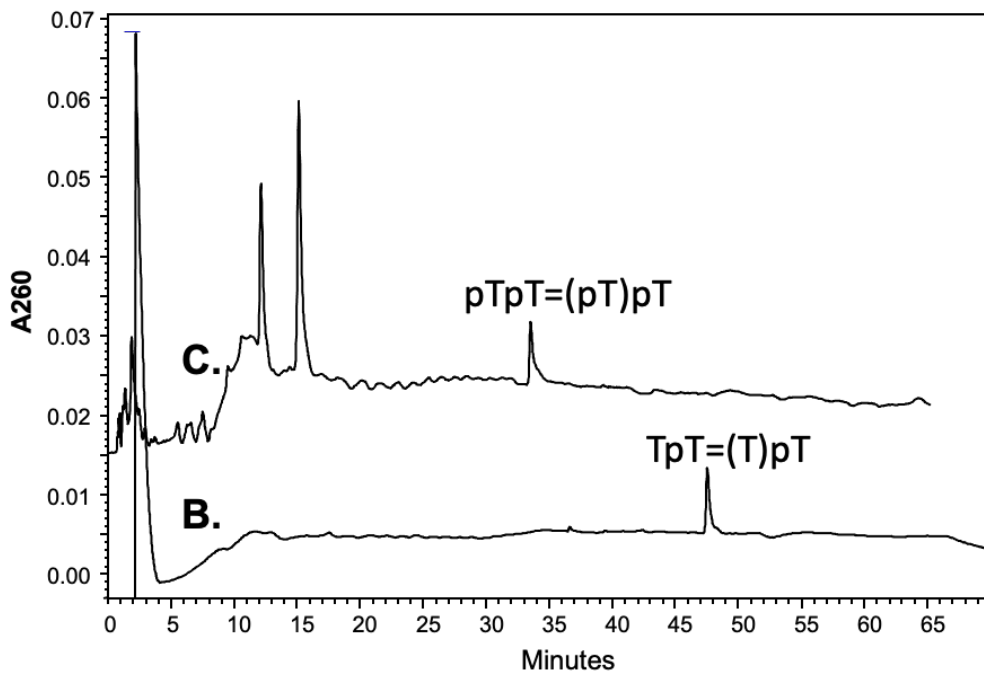
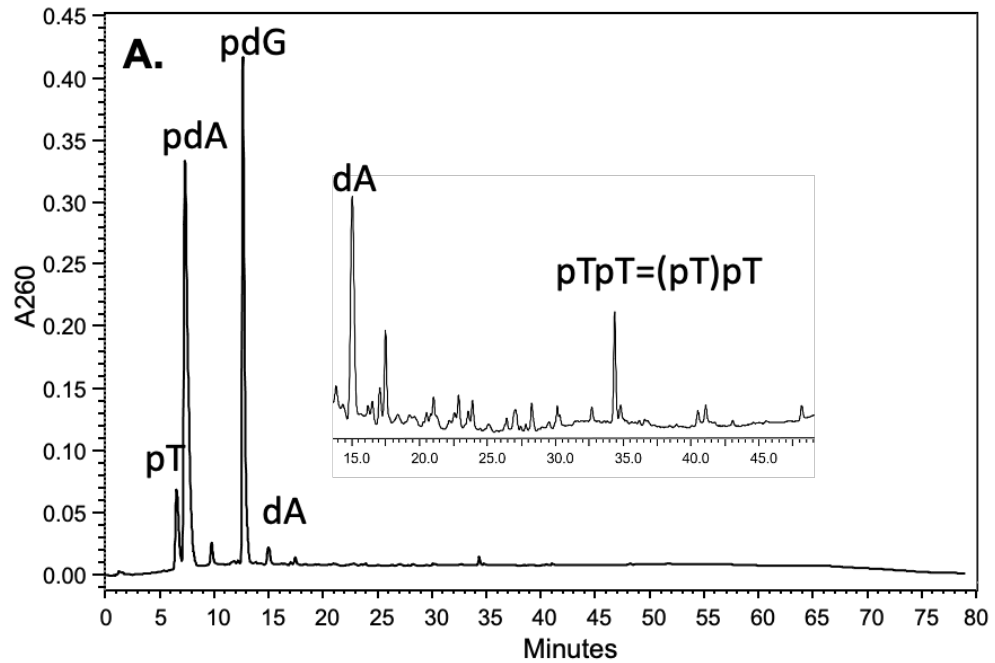


**Figure 2.1. Idea of using structure-specific photoproducts to detect and map folded DNA structures such as basket G-quadruplexes.** a) Structure of adjacent *cis,syn* and non-adjacent *trans,anti* T=T CPDs. b) Whereas UV irradiation of B DNA produces adjacent *cis,syn* T=T CPDs, UV irradiation of non-B DNA folded structures such as the basket G quadruplex form of human telomeric DNA produces unique non-adjacent *anti* T=T CPDs



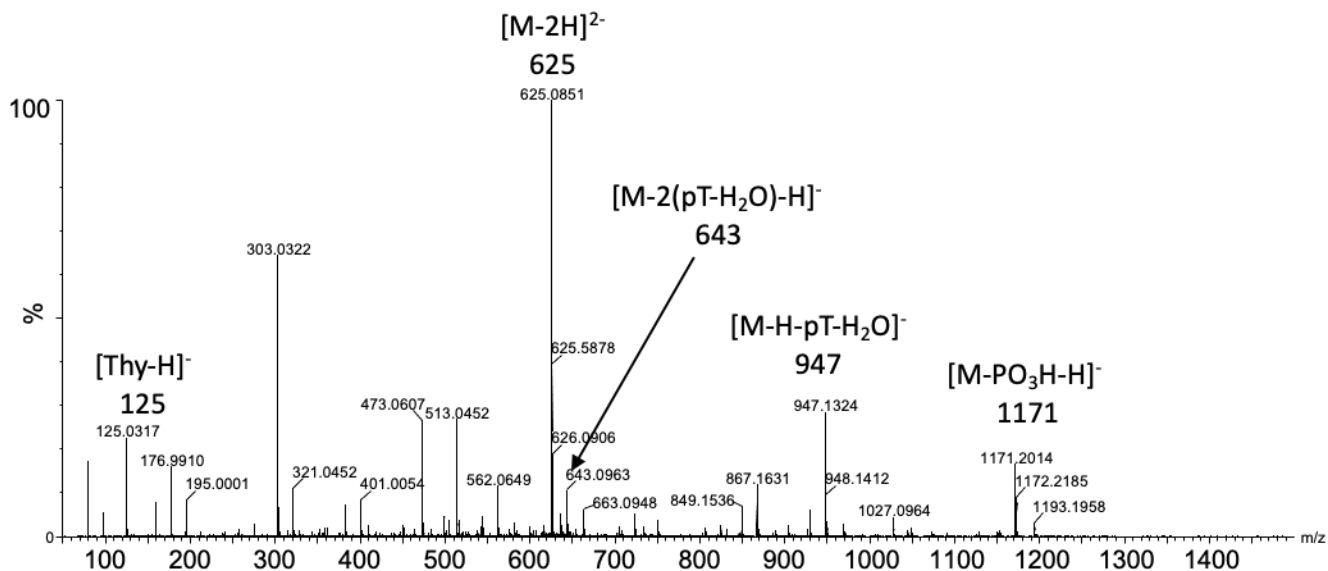
**Figure 2.2. Post-labeling assay schemes for T=T CPDs.** a) DNA containing an adjacent *cis,syn* T=T CPD is degraded to a trinucleotide with snake venom phosphodiesterase (SVP) and then dephosphorylated with calf intestinal phosphatase (CIP) followed by radiolabeling with [ $\gamma$ -<sup>32</sup>P]-ATP and kinase to give a phosphorylated trimer product. Irradiation with 254 nm light (UVC) reverts the CPD to the canonical bases. b) DNA containing a non-adjacent *anti* T=T CPD is degraded to a tetranucleotide with snake venom phosphodiesterase (SVP) and after

dephosphorylation with calf intestinal phosphatase (CIP) and radiolabeling with  $[\gamma\text{-}^{32}\text{P}]\text{-ATP}$  and kinase would give mono and diphosphorylated tetranucleotide products. Irradiation with 254 nm light would revert the tetranucleotide product to two dinucleotides. Treatment of the tetra- or dinucleotide products with NP1 would release the 5'-terminal radiolabeled nucleotide. All radiolabeled products are distinguishable by high resolution polyacrylamide gel electrophoresis.

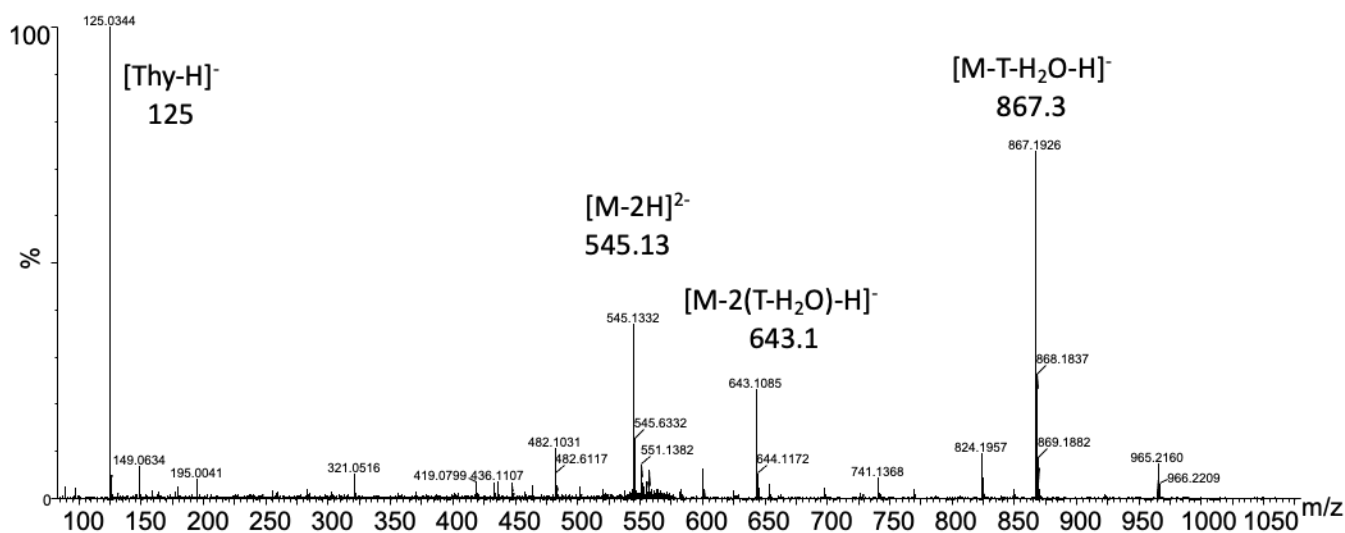


**Figure 2.3. HPLC analysis and purification of the intermediates in the post-labeling assay of Tel26 DNA irradiated with UVB light.** a) Irradiated Tel26 digested with snake venom phosphodiesterase (SVP) for 24 h to give *trans,anti* pTpT=(pT)pT. b) Dephosphorylation of *trans,anti* pTpT=(pT)pT with calf intestinal phosphodiesterase (CIP) to give TpT=(T)pT. c) rephosphorylation of TpT=(T)pT from B with polynucleotidyl kinase (PNK) and ATP.

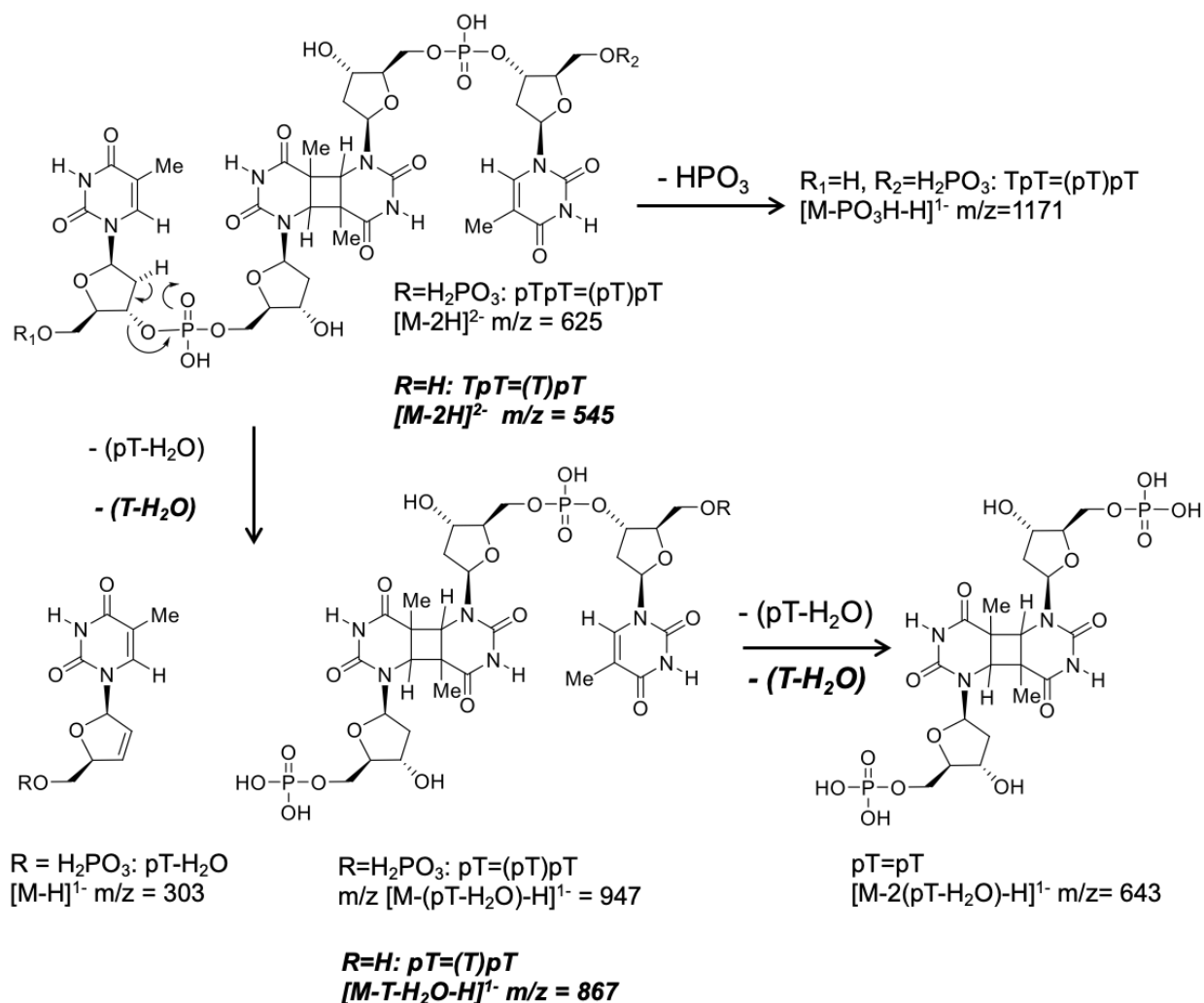
### A. Fragmentation of pTpT=(pT)pT $[M-2H]^{2-}$ at m/z 625



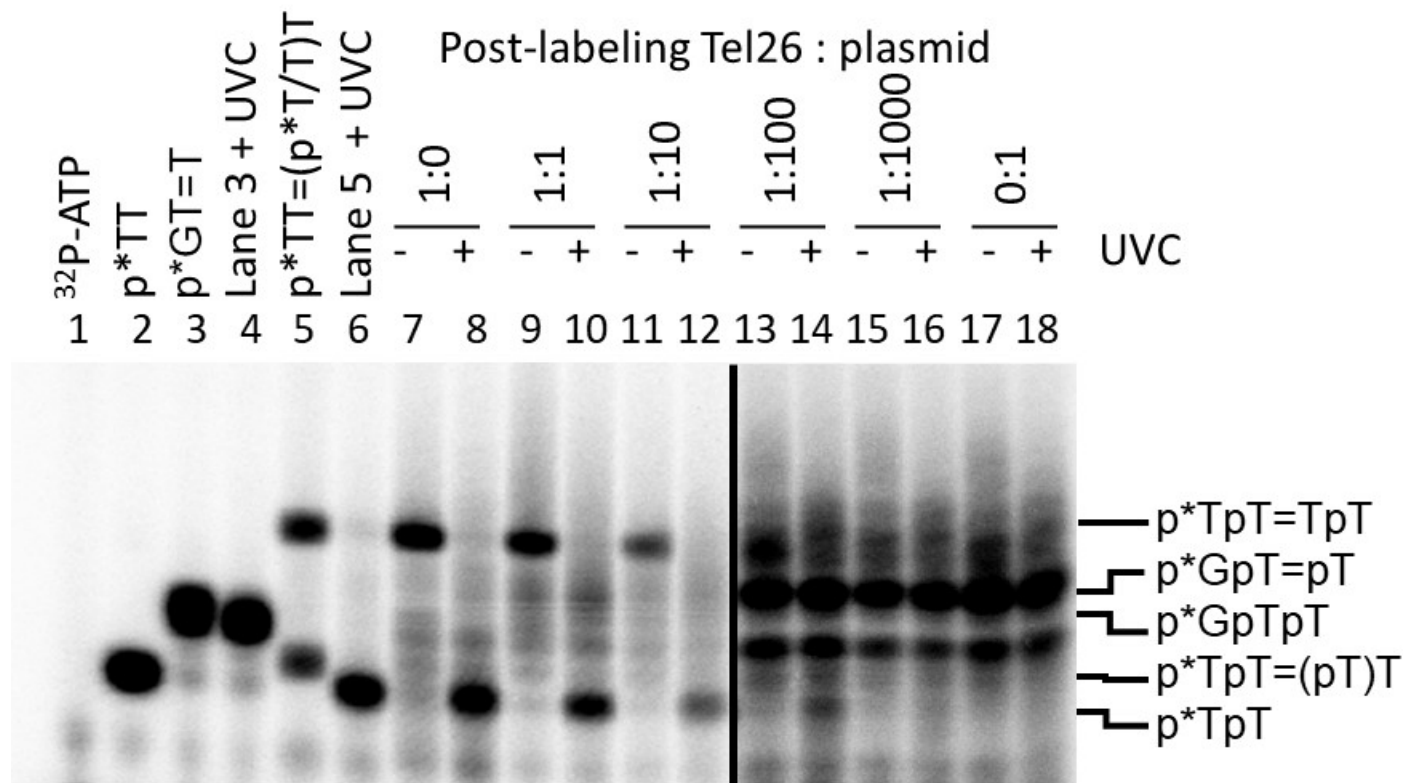
### B. Fragmentation of TpT=(T)pT $[M-2H]^{2-}$ at 545.13



**Figure 2.4. ESI-MS/MS characterization of the intermediates in the post-labeling assay of irradiated Tel26 DNA.** a) MS/MS of the  $[M - 2H]^{2-}$  ion ( $m/z$  624.11) of *trans,anti* pTpT=(pT)pT. b) MS/MS of the  $[M - 2H]^{2-}$  ion ( $m/z$  545.13) of the dephosphorylated *trans,anti* TpT=(T)pT.



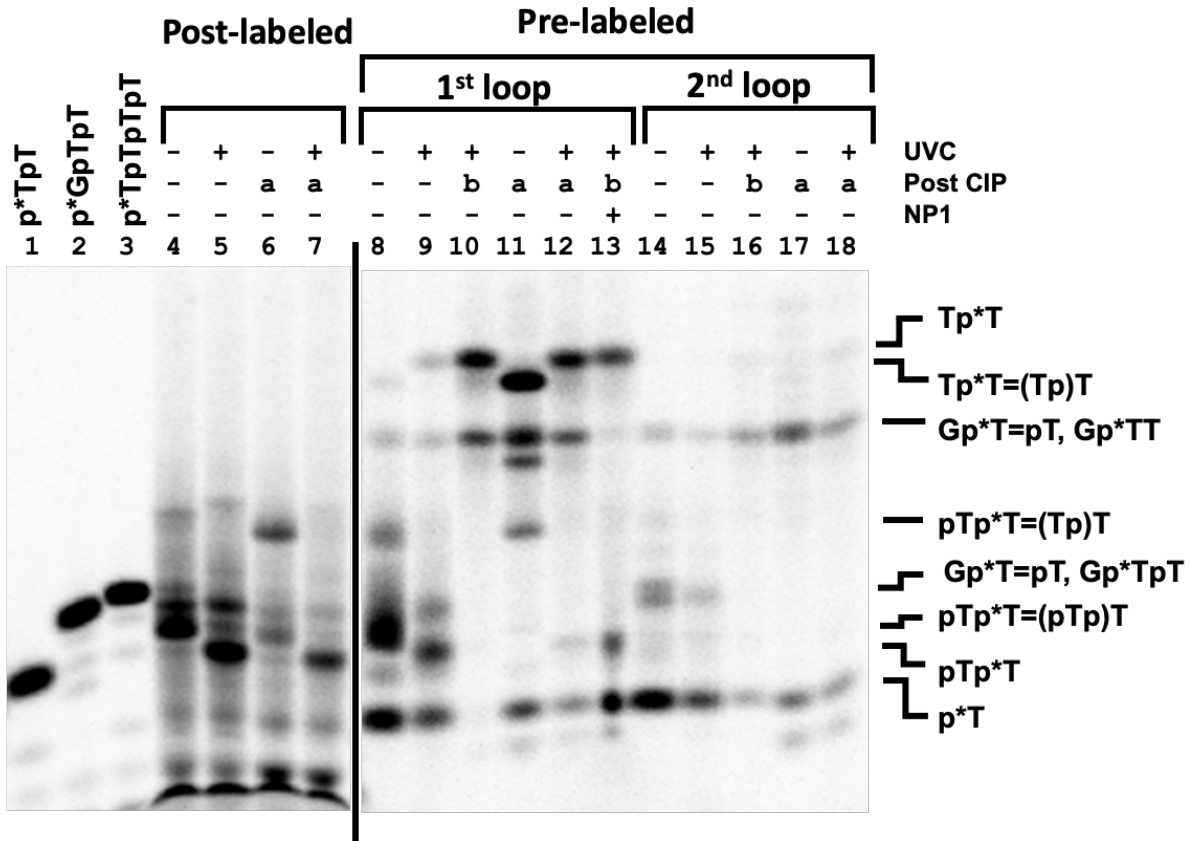
**Figure 2.5. MS/MS fragmentation pathways of the *trans,anti* T=T CPD-containing products** of Figure 4. Pathways for pTpT=(pT)pT are shown in normal font, whereas pathways for TpT=(T)pT are shown in bold italic font.



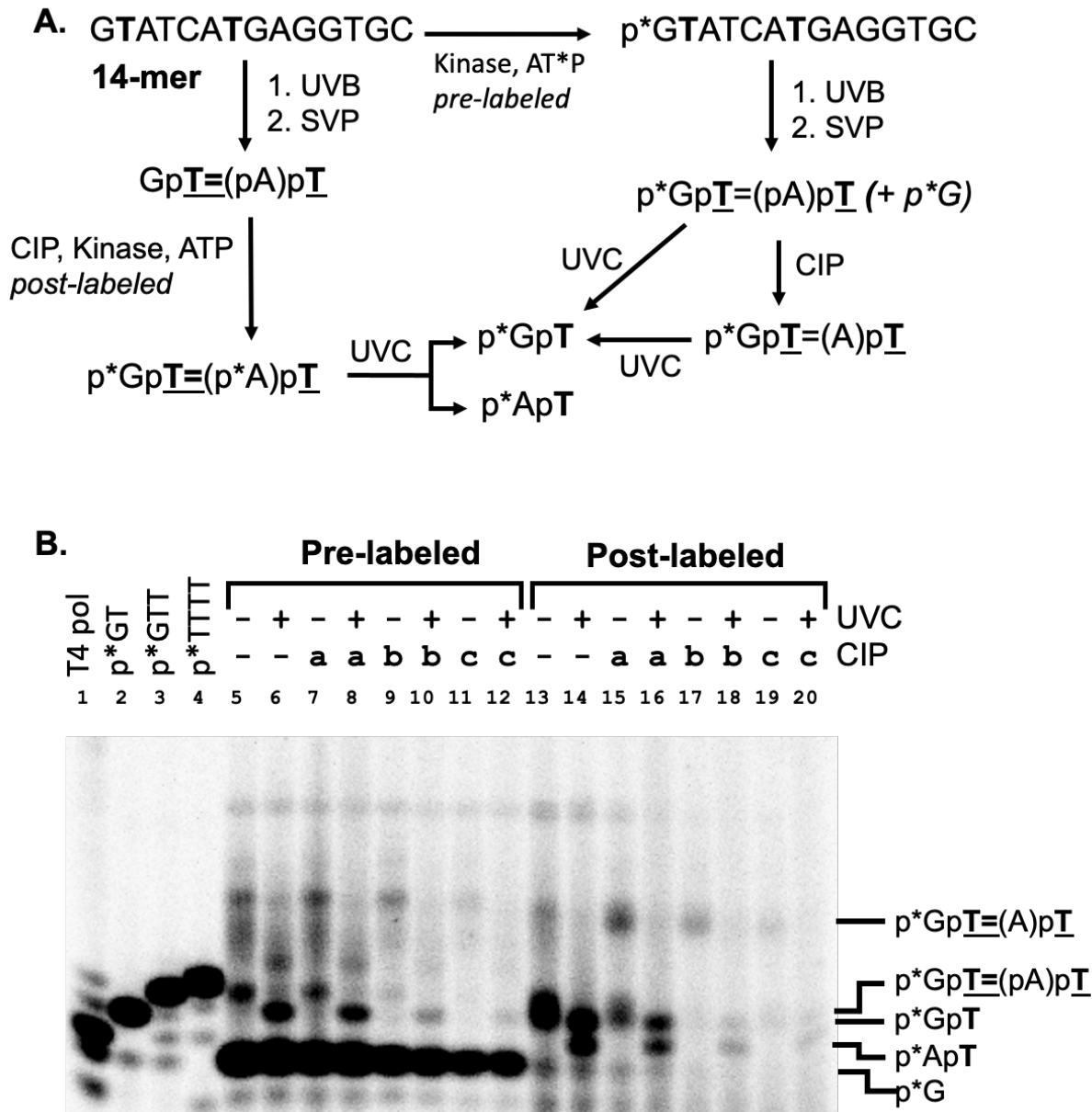
**Figure 2.6. Post-labeling assay carried out on Tel26 in the absence or presence of competing plasmid DNA.** Lanes 2, 3 and 5 are authentic standards prepared by radiolabeling TT and HPLC purified GT=T and TpT=(T)pT. UVC was used to photorevert T=T CPD-containing products. Lane headings for the post-labeling reactions refer to the ratio of the base pair concentrations of Tel26 to plasmid DNA. Lanes 13-18 are shown with a narrower dynamic range to make the photo-reverted products in lanes 14 and 16 more apparent.







**Figure 2.8. Analysis of the enzymatic degradation products of post- and pre-labeled Tel26 by denaturing polyacrylamide gel electrophoresis.** Lanes 1-3: authentic products obtained by end-labeling commercial ODNs. Lanes 4-7: post-labeling assay carried out on UVB irradiated Tel26 followed by dephosphorylation with CIP, with and without CPD photoreversion with UVC light as indicated. Lanes 8-13: SVP treatment of UVB irradiated Tel26 that was internally pre-labeled in loop 1 followed by CIP or CIP and NP1, with and without CPD photoreversion with UVC light, all as indicated. Lanes 14-18: SVP treatment of UVB irradiated Tel26 that was internally pre-labeled in loop 2 followed by CIP with and without CPD photoreversion with UVC light as indicated. Treatments with CIP were for either 2.5 min (a) or 30 min (b). Figure prepared from two separate gels carried out under identical conditions as indicated by the vertical line.



**Figure 2.9. Analysis of the enzymatic degradation products of pre- and post-labeled 14-mer containing a *cis,anti* T=T CPD.** a) Schematic for the enzymatic reactions carried out on the pre- and post-labeled UVB irradiated 14-mer where the T's undergoing *anti* CPD formation are in bold. b) Denaturing PAGE of the enzymatic reaction products. Lanes 5-12: the 5'-pre-labeled UVB irradiated 14-mer was digested with SVP and then dephosphorylated with CIP, where a, b and c refer to the times of 0-, 5- and 10-minutes following addition of CIP, with and without photoreversion by UVC light as indicated. Lanes 13-20: The UVB irradiated 14-mer was degraded by SVP and then labeled with PNK and [ $\gamma$ - $^{32}\text{P}$ ]ATP, after which it was incubated for increasing time with CIP as described for the pre-labeled experiment, with and without photoreversion with UVC light as indicated.

## 2.7 Acknowledgements

We thank Dr. Hani Zaher for allowing us to use his Amersham Typhoon biomolecular imager to image our gels. This material is based upon work supported by the National Science Foundation under Grant No. 2003688 (JST) and an NIH/NIGMS Biomedical Mass Spectrometry Resource grant 5R24GM-136766-02 (MLG).

## 2.8 References

1. Sinden RR, Pytlos-Sinden MJ, Potaman VN. Slipped strand DNA structures. *Front Biosci.* 2007 Sep 1;12:4788–99.
2. Saini N, Zhang Y, Usdin K, Lobachev KS. When secondary comes first – The importance of non-canonical DNA structures. *Biochimie.* 2013 Feb 1;95(2):117–23.
3. Thys RG, Lehman CE, Pierce LCT, Wang YH. DNA Secondary Structure at Chromosomal Fragile Sites in Human Disease. *Curr Genomics.* 2015 Feb;16(1):60–70.
4. Sullivan ED, Longley MJ, Copeland WC. Polymerase  $\gamma$  efficiently replicates through many natural template barriers but stalls at the HSP1 quadruplex. *J Biol Chem.* 2020 Dec 18;295(51):17802–15.
5. Bochman ML, Paeschke K, Zakian VA. DNA secondary structures: stability and function of G-quadruplex structures. *Nat Rev Genet.* 2012 Nov;13(11):770–80.
6. Spiegel J, Adhikari S, Balasubramanian S. The Structure and Function of DNA G-Quadruplexes. *Trends Chem.* 2020 Feb;2(2):123–36.
7. Linke R, Limmer M, Juranek SA, Heine A, Paeschke K. The Relevance of G-Quadruplexes for DNA Repair. *Int J Mol Sci.* 2021 Nov 22;22(22):12599.
8. Ma DL, Wang M, Lin S, Han QB, Leung CH. Recent Development of G-Quadruplex Probes for Cellular Imaging. *Curr Top Med Chem.* 2015;15(19):1957–63.
9. Islam MK, Jackson PJ, Rahman KM, Thurston DE. Recent advances in targeting the telomeric G-quadruplex DNA sequence with small molecules as a strategy for anticancer therapies. *Future Med Chem.* 2016;8(11):1259–90.

10. Liu HY, Zhao Q, Zhang TP, Wu Y, Xiong YX, Wang SK, et al. Conformation Selective Antibody Enables Genome Profiling and Leads to Discovery of Parallel G-Quadruplex in Human Telomeres. *Cell Chem Biol.* 2016 Oct 20;23(10):1261–70.
11. Manna S, Srivatsan SG. Fluorescence-based tools to probe G-quadruplexes in cell-free and cellular environments. *RSC Adv.* 2018 Jul 16;8(45):25673–94.
12. Raguseo F, Chowdhury S, Minard A, Di Antonio M. Chemical-biology approaches to probe DNA and RNA G-quadruplex structures in the genome. *Chem Commun (Camb).* 2020 Jan 1;56(9):1317–24.
13. Yuan JH, Shao W, Chen SB, Huang ZS, Tan JH. Recent advances in fluorescent probes for G-quadruplex nucleic acids. *Biochemical and Biophysical Research Communications.* 2020 Oct 8;531(1):18–24.
14. Han JN, Ge M, Chen P, Kuang S, Nie Z. Advances in G-quadruplexes-based fluorescent imaging. *Biopolymers.* 2022;113(12):e23528.
15. Zheng BX, Yu J, Long W, Chan KH, Leung ASL, Wong WL. Structurally diverse G-quadruplexes as the noncanonical nucleic acid drug target for live cell imaging and antibacterial study. *Chem Commun (Camb).* 2023 Feb 1;59(11):1415–33.
16. Taylor JS. Adjacent and Nonadjacent Dipyrimidine Photoproducts as Intrinsic Probes of DNA Secondary and Tertiary Structure†. *Photochemistry and Photobiology [Internet].* [cited 2022 Nov 23];n/a(n/a). Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/php.13694>
17. Smith-Carpenter JE, Taylor JS. Photocrosslinking of G-Quadruplex-Forming Sequences found in Human Promoters. *Photochem Photobiol.* 2018 Aug 7;95(1):252–66.
18. Lu C, Smith-Carpenter JE, Taylor JSA. Evidence for Reverse Hoogsteen Hairpin Intermediates in the Photocrosslinking of Human Telomeric DNA Sequences. *Photochem Photobiol.* 2018 Jul;94(4):685–97.
19. 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 Apr;42(8):5007–19.
20. Su DGT, Kao JLF, Gross ML, Taylor JSA. 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 Aug 27;130(34):11328–37.
21. Su DGT, Fang H, Gross ML, Taylor JSA. Photocrosslinking of human telomeric G-quadruplex loops by anti cyclobutane thymine dimer formation. *Proc Natl Acad Sci USA.* 2009 Aug 4;106(31):12861–6.

22. 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 Jun 15;31(12):3134–42.
23. Phillips DH. On the origins and development of the <sup>32</sup>P-postlabelling assay for carcinogen–DNA adducts. *Cancer Letters.* 2013 Jun 28;334(1):5–9.
24. Reddy MV. Methods for Testing Compounds for DNA Adduct Formation. *Regulatory Toxicology and Pharmacology.* 2000 Dec 1;32(3):256–63.
25. Phillips DH. Detection of DNA modifications by the <sup>32</sup>P-postlabelling assay. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis.* 1997 Aug 1;378(1):1–12.
26. Keith G, Dirheimer G. Post-labeling: a sensitive method for studying DNA adducts and their role in carcinogenesis. *Current Opinion in Biotechnology.* 1995 Jan 1;6(1):3–11.
27. Weinfeld M, Liuzzi M, Paterson MC. Enzymatic Analysis of Isomeric Trithymidylates Containing Ultraviolet Light-induced Cyclobutane Pyrimidine Dimers. *Journal of Biological Chemistry.* 1989 Apr;264(11):6364–70.
28. Parikh D, Fouquerel E, Murphy CT, Wang H, Opresko PL. Telomeres are partly shielded from ultraviolet-induced damage and proficient for nucleotide excision repair of photoproducts. *Nat Commun.* 2015 Sep 9;6(1):8214.
29. Hertzberg RP, Dervan PB. Cleavage of DNA with methidiumpropyl-EDTA-iron(II): reaction conditions and product analyses. *Biochemistry.* 1984 Aug 14;23(17):3934–45.
30. van de Sande JH, Kleppe K, Khorana HG. Reversal of bacteriophage T4 induced polynucleotide kinase action. *Biochemistry.* 1973 Dec 4;12(25):5050–5.

# **Chapter 3: Post-labeling Assay of Non-Adjacent *anti* Cyclobutane Pyrimidine Dimers formation in known Non-B Secondary Structures of DNA**

## **3.1 Abstract**

Irradiation of non-B secondary structures of DNA, such as G-quadruplex and reverse Hoogsteen hairpins, have been found to produce a unique class of non-adjacent *anti* cyclobutane pyrimidine dimer (CPDs) photoproducts. The formation of these *anti* CPDs has been explained by the ability of certain folded conformations of DNA to bring distant pyrimidine bases into close enough proximity for photodimerization. Because the structure and stereochemistry of the *anti* CPDs is determined by the structure of the DNA at the time of photoirradiation, it was realized that *anti* CPDs could be used as intrinsic probes for non-B DNA structures *in vivo* and *in vitro*. Before CPDs can be used as such probes, however, a better understanding of the photoreactivity of different non-B DNA structures is needed. We therefore report the use of the post-radiolabeling assay to detect and further study the formation of non-adjacent *anti* CPDs in G-quadruplex forming sequences found in the promoter and telomeric regions of human DNA, as well as in telomeric DNA from other organisms. The structures adopted by these sequences have been previously determined by an array of structural and photochemical studies. We report the detection of non-adjacent CPDs in conditions that were previously determined to produce these photoproducts at insignificant amounts, as well as in structures not expected to form *anti* CPDs, demonstrating the

use of this highly sensitive post-radiolabeling assay to investigate the photochemistry of biologically relevant non-B secondary structures of DNA.

## 3.2 Introduction

The study of the photochemistry of DNA has been mainly focused on the native double helical B-form conformation of DNA since the discovery of the thymine dimer in 1958 (1). These studies were originally motivated by trying to understand the mutagenic properties of UV light on living organisms. As a consequence, the structure of the main photoproducts of B form DNA were discovered, including the *cis,syn* cyclobutane butane pyrimidine dimers (CPDs) and the pyrimidine (6-4) pyrimidone photoproducts (Fig. 3.1a) (2,3). In contrast, very few studies have been carried out on the photochemistry of biologically significant non-B secondary structures of DNA such as hairpins, G-quadruplexes, cruciforms, H-DNA, and bulges (4–8).

Photochemical studies on non-B secondary structures of DNA was recently prompted by the discovery of an interstrand-type *cis,anti*-CPD that was accidentally found to form between non-adjacent thymine bases in a single-stranded oligonucleotide (ODN) under acidic conditions (Fig. 3.1b) (9). The structure of this non-adjacent CPD photoproduct was characterized by a nuclease P1-coupled mass spectrometry (MS) assay that is capable of distinguishing between adjacent and nonadjacent DNA photoproducts (9,10). Nuclease P1 (NP1) is an endonuclease that hydrolyzes the phosphodiester linkage at the 3'-end of a native nucleotide but cannot induce cleavage when the base is modified. NP1 digestion of DNA photoproducts that form between adjacent nucleotides in a DNA sequence, therefore results in the production of a trimer pY=pYpN, but when



photoproducts form between nonadjacent nucleotides, however, a tetramer pY(pN)=pY(pN) results (Fig. 3.2) (9–12).

The discovery and characterization of a non-adjacent CPD with a head-to-tail or *anti* orientation with the use of the NP1-coupled MS assay, led to the discovery that *anti* CPDs could also form from the UVB irradiation of human telomeric and promoter G-quadruplex forming sequences (7,13,14). G-quadruplex DNA structures are known to have a high degree of structural polymorphism and adopt many different structures characterized by the relative orientation of the strands and connections by the loops (Fig. 3.3) (5,15,16). Structures that have adjacent lateral loops, as is the case for the basket and chair structures (Fig. 3.3.d, e), can result in non-adjacent *anti* CPD formation when exposed to UVB light when the reacting pyrimidine bases are correctly positioned, and the loops are long enough. It was also discovered that non-adjacent photoproducts could also arise from the reverse Hoogsteen hairpin structures that are in equilibrium with the G-quadruplex structures and other types of DNA conformations (Fig. 3.3.f) (4,8).

The observation that non-adjacent CPDs can readily form in non-B structures of DNA led to the proposal that non-adjacent CPDs could be used as stable molecular signatures for the existence of non B- structures *in vitro* and *in vivo* (4,14). To be used effectively as probes, a fundamental understanding of the photoreactivity of non-B DNA structures is required. It will be necessary to determine the extent to which specific non-B DNA structures lead to specific non-adjacent photoproducts or if they produce a specific pattern of photoproducts (Fig. 3.3). We therefore report a first step towards a comprehensive investigation into the photochemistry of G-quadruplexes and other non-B conformations of DNA by making use of the highly sensitive post-radiolabeling assay that we have developed (Chapter 2) (Fig. 3.2). This post-labeling assay is able to detect the

formation of non-adjacent *anti* CPDs in telomeric G-quadruplex forming sequences, as well as reverse Hoogsteen hairpins (8,13,14,17–19). The assay makes use of snake venom phosphodiesterase (SVP) to degrade non-adjacent CPDs to tetramers of the form pNpY=(pN)pY (Fig. 3.2) that can be distinguished from the partial degradation of DNA by photoreversal with 254 nm light to pNpY dinucleotides. We now report using this method for the detection of non-adjacent CPD formation under conditions that previous methods detected at insignificant amounts, as well as in structures that were not expected to form *anti* CPDs. These results demonstrate that this post-radiolabeling assay complements previous mass spectrometry techniques for studying the photochemistry of biologically relevant non-B secondary structures of DNA.

## 3.2 Materials and Methods

### 3.2.1 Materials and Reagents

Oligodeoxynucleotides (ODNs) were purchased from Integrated DNA Technologies, Inc. (IDT, Coralville, IA, USA). Snake venom phosphodiesterase (SVP) from *Crotalus adamanteus* was obtained from Worthington Biochemical (Lakewood, NJ, USA). DNase I (RNase- free) was obtained from New England Biolabs (Ipswich, MA, USA). Calf intestinal phosphatase and T4 DNA ligase were obtained from Promega (Madison, WI, USA). T4 Polynucleotide Kinase (T4 PNK) was obtained from ThermoFisher Scientific (Waltham, MA, USA) and 6000 Ci/mmol [ $\gamma$ - $^{32}$ P]-ATP from Perkin Elmer (Waltham, MA, USA).

### 3.2.2 Preparation of G-Quadruplex and Reverse-Hoogsteen Hairpin Structures

The post-labeling assay was tested on nine sequences that have been shown by previous studies to fold into G-quadruplex and reverse Hoogsteen secondary structures (8,13,14,17–19). Unless

otherwise stated, the ODNs (500 pmol) were heated at 95°C for 10 min in 100 µL of 150 mM KCl, NaCl, or LiCl, and then rapidly cooled in ice to preferentially form intramolecular structures. Samples were left on ice for at least 1 h before irradiation with UVB light.

The different G-quadruplex structures adopted by hTeLo (100 µM, 100 µL) were induced with different cations under different ionic strength conditions. The basket fold was induced under high ionic strength (HIS) conditions in 1 M NaCl and in low ionic strength (LIS) conditions in 140 mM NaCl. The hybrid structure was induced under HIS conditions in 950 mM LiCl and 50 mM KCl, and for LIS conditions in 140 mM KCl. The parallel conformation was only induced under HIS conditions in 5 M LiCl and 20 mM KCl. These conditions only deviated from the original studies by omitting EDTA and Tris buffer as these compounds were found to inhibit SVP activity (19).

### **3.2.3 Post-labeling Assay of UVB Irradiated ODN Sequences**

After the respective G-quadruplex and/or reverse Hoogsteen hairpin structure was induced and kept on ice for at least 1 h, the ODN sequence was irradiated with broadband UVB/UVA light (275–400 nm, centered at 330 nm) for 2.5 h on ice. Irradiations were carried out in polyethylene microfuge tubes on a bed of ice for 2.5 h at 1 cm distance from two Spectroline XX-15B UV 15 W tubes with an approximate intensity of 2.2 mW/cm<sup>2</sup> (22 J/m<sup>2</sup>-s) after filtration through a plate of Pyrex glass. After irradiation, the solution was evaporated in vacuo and incubated with 20 units of DNaseI in 20 µL of 10 mM Tris-HCl (pH 7.6), 2.5 mM MgCl<sub>2</sub>, and 0.5 mM CaCl<sub>2</sub> at 37 °C overnight, followed by inactivation of the DNase I at 75 °C for 20 min. After the sample was dried in vacuo, it was resuspended with 10 to 15 units of SVP in 30 µL of 10 mM ammonium citrate, 100 µM MgCl<sub>2</sub> pH 9.4 at 37 °C overnight and the SVP was inactivated in boiling water for 20

min. After the sample was dried, it was then treated with 6 units of CIP in 40  $\mu$ L of 50 mM Tris-HCl (pH 9.3), 1 mM MgCl<sub>2</sub>, 100  $\mu$ M ZnCl<sub>2</sub>, and 1 mM spermidine at 37 °C for 3 h. The enzyme and other denatured proteins were removed by phenol extraction and residual phenol was removed by multiple diethyl ether extractions. After evaporating the sample in vacuo, it was treated with 10 units of T4-PNK and 0.5 pmol [ $\gamma$ -<sup>32</sup>P]-ATP in 40  $\mu$ L of 50 mM Tris-HCL (pH 7.6), 10 mM MgCl<sub>2</sub>, 5 mM DTT, 0.1 mM spermidine, 0.1 mM EDTA at 37 °C for 3 h. An aliquot of the sample (4  $\mu$ L) was irradiated for 30 min with 254 nm light to reverse any CPDs. Unirradiated and irradiated aliquots (4  $\mu$ L) were then mixed with an equal volume of 2X loading buffer (98% formamide with xylene cyanol dye), heated in a boiling water bath for 5 min, and cooled on ice. The radiolabeled samples were then analyzed on a 20% denaturing polyacrylamide gel (7 M Urea, 19:1 acrylamide:bisacrylamide, 0.8 mm thick) at 1500 V until the dye reached 8 cm. The gels were then scanned with an Amersham Typhoon biomolecular imager.

## **3.4 Results and Discussion**

### **3.4.1 Post-labeling Assay of Tel26 and Tel22 Human Telomeric Sequences**

The UVB photoproducts of the G-quadruplex structures Tel22 d[AGGG(TTAGGG)<sub>3</sub>] and Tel26 d[AAAGGG(TTAGGG)<sub>3</sub>AA] in the presence of either Na<sup>+</sup>, K<sup>+</sup> or Li<sup>+</sup> was originally studied by an NP1-coupled MS assay. A cation and DNA sequence dependent distribution of *anti* CPDs was observed and was proposed to arise from the presence of multiple types of interconverting DNA structures (7,8,13). We therefore choose these sequences first to test if the post-labeling assay would reproduce the original results. Post-labeling analysis of UVB irradiated Tel26 and Tel22 sequences in the presence of all three metal ions produced radiolabeled bands that corresponded

to the monophosphorylated products of the non-adjacent CPDs p\*NpT=(N)pT (Fig. 3.4 lanes 5,7,9,12,14,16). Exposure of the degradation mixture to UVC (254 nm) light prior to electrophoresis converted the radiolabeled tetranucleotide products to a faster moving band corresponding to the dinucleotide p\*NpT expected to occur from the photoreversal of the nonadjacent *anti* p\*NT=(N)T CPDs (Fig. 3.4 lanes 6, 8, 10, 13, 15, 17). Detection of bands corresponding to the formation of *anti* T=T CPDs in both Tel26 and Tel22 sequences, confirm that these sequences adopt photoreactive conformations in native conditions in the presence of K<sup>+</sup>, Na<sup>+</sup>, and Li<sup>+</sup>.

Previous enzyme-coupled MS studies of Tel22 and Tel26 in Na<sup>+</sup> found that they produced only low to insignificant (1%) amounts of *anti* CPDs (7,13). It was proposed that these sequences adopt a basket type structure in the presence of Na<sup>+</sup> that inhibits interloop photoreaction of the thymidines due to steric constraints imposed by the three nucleotide loops. When the post-labeling assay was carried out on both sequences irradiated in the presence of Na<sup>+</sup>, a significant, but low amount of *anti* CPDs were observed. These results support the hypothesis that the human telomere basket structure is not very photoreactive (Fig. 3.3d). An equally small amount of the radiolabeled adjacent CPD product p\*GpT=pT was observed in these samples that could be explain by the low photoreactivity of adjacent T's in loops, or that radiolabeling of *cis,syn* CPDs is less efficient than non-adjacent photoproducts.

Irradiation of Tel22 and Tel26 the presence of K<sup>+</sup> and Li<sup>+</sup>, produced a higher yield of bands corresponding to *anti* CPDs compared to Na<sup>+</sup> in accord with previous studies (7,8,13). Tel22 in the presence of K<sup>+</sup> was reported to produce a variety of nonadjacent *anti* CPDs at different yields, where the NP1 digestion products with both *trans,anti* and *cis,anti* stereochemistries were found

to form between the 1<sup>st</sup> T of loop 1 and the 2<sup>nd</sup> T of loop 3 (T1=T2 CPD) (6.5%), as well as between the 2<sup>nd</sup> T bases of loop 1 and 3 (T2=T2 CPD) (16%) (7,13). In the post-labeling assay of Tel22 in the presence of K<sup>+</sup> these photoproducts would produce the SVP digestion products pGpT1=(pT)pT2 and pTpT2=(pT)pT2, respectively. The radiolabeled products corresponding to these different *anti* CPDs are unable to be differentiated from one another by gel electrophoresis, but it has been shown previously that the post-labeling assay can detect both *trans,anti* and *cis,anti* stereoisomers of the T=T CPDs. Similarly, only one radiolabeled band corresponding to non-adjacent CPDs is observed when Tel26 was irradiated in KCl which was shown to produce similar yields of *anti* CPDs as Tel22, except that a preference for the *trans,anti* T2=T2 CPD was observed. The preferred formation for the *trans,anti* T2=T2 CPD photoproduct in Tel26 was attributed to the reverse Hoogsteen hairpin structure being furthered stabilized by this sequence compared to Tel22 (8) which is evident by the higher amount of non-adjacent CPDs detected by the post-labeling assay in Tel26 compared Tel22 in the presence of Li<sup>+</sup> (Fig. 3.4 lane 9,10,16,17) .

### **3.4.2 Post-labeling Assay of the Model Human Telomeric Sequence hTeLo**

To further explore the photochemistry of the multi-structural and dynamic human telomeric sequences, the hTeLo d(TAGGG(TTAGGG)<sub>3</sub>TT) model sequence was assayed with the post-radiolabeling technique for non-adjacent CPDs as the structures adopted by this sequence have been thoroughly characterized through a variety of structural analysis studies (20–27). Most recently this sequence was analyzed by a single-molecule nanopore technology where the fluctuations in the current levels were used to discriminate between different G-quadruplex structures. As the nanopore experiments usually require samples to have high levels of electrolytes, the structural studies were performed under two distinct ionic strength conditions classified as low

(LIS) and high (HIS). The low ionic strength conditions were consistent with the biological concentrations of ions, specifically  $K^+$ , inside the cell, while the HIS levels contained the required elevated amount of electrolyte concentrations needed for nanopore sequencing while in the presence of the respective G-quadruplex stabilizing ions. In the presence of  $K^+$  under the low (LIS, 140 mM  $K^+$ ) and high (HIS, 50 mM KCl, 950 mM LiCl) conditions, the results were consistent with a predominance of hybrid-type G-quadruplex structures along with the corresponding triplex folding intermediates (24,28). In the presence of  $Na^+$  under low (LIS, 140 mM NaCl) or high (HIS, 1 M NaCl) ionic strength, the results were consistent with a basket structure (19). When analyzed by the UVB coupled post-labeling assay, tetramer bands corresponding to the monophosphorylated products of *anti* CPDs were observed under both high (HIS) and low (LIS) salt conditions with  $Na^+$  and  $K^+$  (Fig. 3.5 lane 11, 13, 15, 17). The efficient formation of *anti* CPDs in hTeLo in the presence of  $K^+$  that favored the hybrid and triplex structures is consistent with the presence of the triplex intermediate which has been previously shown to form *anti* CPDs (7). Alternatively, the highly efficient formation of *anti* CPD photoproducts reported by the post-labeling assay may also be due to the hTeLo sequence adopting photoreactive form 3 basket G-quadruplex or reverse Hoogsteen structures that may be in equilibrium with the hybrid structures as was the case for the Tel22 and Tel26 sequences in the presence of  $K^+$  (8,13). More surprisingly, high yields of tetramer product were observed in the presence of  $Na^+$ , which is different than observed for Tel22 or Tel26. Perhaps hTeLo in  $Na^+$  solution can also adopt a form 3 basket which is photoreactive.

Unexpectedly, bands corresponding to *anti* CPDs were also observed along with what appear to be partially digested product (Fig. 3.5 lane 19) under highly dehydrating conditions (5 M LiCl) in

the presence of  $K^+$  where the parallel structure has been shown by nanopore analysis to be the predominant conformation adopted by the hTeLo sequence (Fig. 3.3a). The parallel G-quadruplex contains propeller loops that do not allow for thymine bases to interact in a way that non-adjacent CPDs could form. We speculate that these *anti* CPDs could be forming from the presence of a reverse Hoogsteen hairpin structure that may be in equilibrium with the non-photoreactive parallel G-quadruplex structure as we found for Tel26.

### 3.4.3 Post-labeling Assay of Ciliate Telomeric Sequences Tet-4 and Oxy-4

The G-quadruplex forming telomeric sequences found in the ciliate organisms *Oxytricha* d(T<sub>4</sub>G<sub>4</sub>)<sub>4</sub> (Oxy-4) and *Tetrahymena* d(T<sub>2</sub>G<sub>4</sub>)<sub>4</sub> (Tet-4) (29,30) were reported in 1989 by Cech and coworkers to form metal ion dependent intra-strand photo-crosslinks when exposed to 254 nm light (17). For the Oxy-4 sequence the photo-cross-linking efficiency was in the order  $Na^+ > K^+ > Li^+$ , while for the Tet-4 sequence the order was  $Na^+ > Li^+ > K^+$ . In the presence of  $Na^+$ , Oxy-4 was proposed to adopt a chair structure with two G-quartets, where the photo-crosslink was mapped to occur between the T11 base in loop 1 and T27 base in loop 3 (Fig. 3.6b). For the Tet-4 sequence, a chair structure with three G-quartets was proposed to account for the photoreaction between the T1 in the 5'-tail and the T13 in loop 2 (Fig. 3.6b). The involvement of these two bases in the photocrosslinking product of the Tet-4 sequence suggests a new way for nonadjacent photoproducts to form in G-quadruplex structures (4,31,32).

When the UVB-coupled post-labeling assay was used to probe the structure of the Oxy-4 and Tet-4 sequences in the presence of  $K^+$ ,  $Na^+$ , and  $Li^+$ , radiolabeled bands with similar electrophoretic mobility to the monophosphorylated tetramer products  $p^*NpT=(N)pT$  were detected (Fig. 3.6 lane



4,6,8,11,13,15). After exposure to 254 nm light the slower moving tetranucleotide bands seen in the Na<sup>+</sup> and Li<sup>+</sup> photoreactions were converted to fast-moving bands corresponding to the dinucleotide p\*NpT suggesting that non-adjacent CPDs had formed. In the presence of K<sup>+</sup>, only a small amount of the dinucleotide band was observed to form after photoreversal for both the Tet-4 and Oxy-4 sequences, but the slower moving band corresponding to the monophosphorylated tetramer was not very discernable (Fig. 3.6 lane 4,5,11,12). On the other hand, a very strong band migrating slightly faster than the band for p\*TpT=(T)pT was observed for both Oxy-4 and Tet-4 in K<sup>+</sup> solution that was unchanged by 254 nm irradiation suggesting that it may be 6-4PP and/or Dewar photoproduct. This band is very prominent in the Tet-4 photoreaction in K<sup>+</sup> (lanes 11 & 12) and should be further investigated. It is very puzzling, however, that it does not appear in Na<sup>+</sup> or Li<sup>+</sup> solution.

It was initially proposed that the photocrosslinking of Tet-4 and Oxy-4 sequences in the presence of Na<sup>+</sup> could have arisen from distinct chair structures, but it is also possible that they could have occurred from the basket type G-quadruplex conformation (Fig. 3.6b) (4,33). The detection of non-adjacent *anti* CPDs in the presence of Li<sup>+</sup> in both sequences suggest that a non-G-quadruplex secondary structure of DNA, such as a Hoogsteen or reverse Hoogsteen hairpin, may be involved. A reverse Hoogsteen hairpin has been shown to form in a fragment of the Oxy-4 sequence in the presence of Li<sup>+</sup> by NMR studies (34). In the presence of K<sup>+</sup>, the Oxy-4 sequence has been shown to adopt a mixture of parallel and antiparallel structures (35–37). The small amount of *anti* CPD formed from Oxy-4 in K<sup>+</sup> solution is consistent with the presence of an antiparallel G-quadruplex such as a chair or basket structure but may also come from a reverse Hoogsteen hairpin. Further studies of this sequence could greatly benefit from the internal labeling strategy to confirm the

crosslinking sites and the enzyme coupled- MS/MS assay to identify the photoproducts formed in  $K^+$ .

### 3.4.3 Post-labeling Assay of Human Promoter Sequences

We also used the UVB-coupled post-labeling assay to examine four putative G-quadruplex forming sequences from promoters of the CALU, STARD3NL, c-MYC, and ILPR human genes, where the CALU and STARD3NL sequences were previously found to form non-adjacent CPDs when exposed to UVB light in the presence of  $Na^+$  and  $K^+$  (14).

#### *CALU promoter sequence*

The assay was carried out in the CALU promoter sequence  $d(GGGGCT)_3GGG$  in solutions of KCl, NaCl, and LiCl (Fig. 3.7a, lanes 4-10). Photoreversible radiolabeled bands corresponding to the singly phosphorylated tetramers  $p^*NpY=(N)pY$  were observed in all cases, though it was maximal in  $Li^+$  (Fig. 3.7b). Previously it was shown by circular dichroism (CD), melting temperature analysis, and chemical foot printing experiments that this sequence adopts antiparallel G-quadruplex structures in the presence of  $K^+$  and  $Na^+$  that would explain the formation of *anti* CPDs (Fig. 3.8a) (14). The formation of non-adjacent CPDs in  $Li^+$  suggests the presence of reverse Hoogsteen hairpins or other non-G-quadruplex structures, where the greater yield of *anti* CPD formation in  $Li^+$  compared to  $Na^+$  and  $K^+$  suggests that the most photoreactive conformation is not a G-quadruplex (Fig. 3.7b).

Interestingly, the UVB-coupled post-labeling assay in the presence of  $K^+$  produced non-photoreversible tetramer bands (Fig. 3.7a lane 5,6) much like was observed for Oxy-4 and Tet-4.

This suggests that, as with the Tet-4 and Oxy-4 sequences, they may be due to non-photoreversible non-adjacent 6-4PP or Dewar photoproducts and merit further investigation.

Previous NP1-coupled MS analysis of this CALU sequence in the presence of Na<sup>+</sup> and K<sup>+</sup>, found that pT(pG)=pT(pG) was the major CPD produced from UVB irradiation. This tetramer was explained to arise from the inter-loop photoreaction between thymidines in lateral loops 1 and 3 in the chair and/or basket conformations through T/U replacement experiments (Fig. 3.8a) (14). Replacing thymidine bases with uracil allowed the reacting bases to be identified by NP1-coupled mass spectrometry, as the photocycloaddition reaction between a T and a U occurs as equally well as between two Ts.

#### *STARD3NL promoter sequence*

When the assay was carried out in the STARD3NL d(GGGGCTGCTGGGTGCGGGCGGTGGG) sequence in solutions with K<sup>+</sup>, Na<sup>+</sup>, and Li<sup>+</sup>, photoreversible radiolabeled bands corresponding to the singly phosphorylated tetramers p\*NpY=(N)pY were observed in all cases, where maximal formation was observed in Na<sup>+</sup> (Fig. 3.7a lanes 11-17). The previous photochemical studies done on this sequence showed that the formation of *anti*-CPDs in the presence of K<sup>+</sup> and Na<sup>+</sup> could be explained by the formation of antiparallel G-quadruplex structures (Fig. 3.8b) (14). The formation of non-adjacent CPDs in Li<sup>+</sup> suggests the presence of reverse Hoogsteen hairpins or other non-G-quadruplex structures, but the greater yield of *anti* CPD formation in Na<sup>+</sup> suggests that the most photoreactive conformation of this sequence may be a G-quadruplex structure. (Fig. 3.7b). In the presence of K<sup>+</sup> non-photoreversible tetramer bands were produced suggesting the formation of non-photoreversible non-adjacent 6-4PP or Dewar photoproducts.

The previous NP1-coupled MS analysis of the STARD3NL sequence in the presence of  $K^+$  detected the formation of two *anti* CPDs produced from UVB irradiation: a major pT(pG)=pT(pG) product and a novel pT(pG)=pU(pG) which would have resulted from the deamination of a nonadjacent T=dC CPD. The uracil containing NP1 product pT(pG)=pU(pG) was proposed to arise from the inter-loop photoreaction between loops one and three from the T9 and C19 bases in the chair form or between T6 and C19 in the basket form (Fig. 3.8) . Alternatively, it could have arisen from an intra-loop photoreaction between the T13 and C15 bases in loop two and/or between T22 and C19 in loop three. The exact site of formation for the T=C CPD could not be determined with nucleotide replacement experiments due to problems in distinguishing between the two possible CG sites (Fig. 3.8) (14). If, however, SVP degradation is used, the digestion products between the T=C CPD intra- and inter-loop reactions could be distinguished. SVP digestion would produce pCpT=(pG)pC which would deaminate to pUpT=(pG)pU if it formed from the inter-loop photoreaction between T6 and C19 if in the basket form or between T9 and C19 in the chair form. But if it formed from intra-loop reactions of the T and C bases in loop two or three, SVP would produce pGpT=(pG)pC which would deaminate to pGpT=(pG)pU.

The T=T *anti* CPD photoproduct detected from the NP1 analysis of the STARD3NL sequence was proposed to arise from the inter-loop photoreaction between the T9 and T22 bases in loops one and three from a basket structure, or between the T6 and T22 from a chair structure. It was also proposed that it could have arisen from a possible intra-loop photoreaction from the bases T6 and T9 in loop one (Fig. 3.8b)(14). If digestion analysis with SVP is performed, the possible intra- and inter-loop products could also be distinguished as pCpT=(pC)pT (deaminating to pUpT=(pC)pT)

would form from the loop one photoreaction between T6 and T9, whereas if it formed between loop one and loop three pCpT=(pG)pT (deaminating to pUpT=pG)pT) would form.

As we are unable to determine the sequence composition of the tetramer and photoreversed bands in the post-labeling assay using a simple denaturing gel, we cannot distinguish between non-adjacent T=T CPD and the T=U CPD arising from deamination of a T=C CPD, or where they are arising based on flanking sequence information. It may be possible to get more information from a 2D gel using low pH which differentiates nucleotides, or by 2D paper chromatography. Alternatively, the pre-labeling strategy could be used to identify the site of the CPDs. Clearly, an approach that combines the post-labeling strategy to identify sequences that form non-adjacent CPDs with an NP1 and SVP coupled MS/MS to identify the type of CPDs formed and their flanking sequence, followed by a pre-labeling assay would be best.

#### *c-MYC promoter sequence*

The G-quadruplex forming sequence found in the promoter region of the c-MYC gene was also tested for non-adjacent CPD formation with the post-labeling assay in the presence of K<sup>+</sup>, Na<sup>+</sup>, and Li<sup>+</sup> (Fig. 3.9). The G-quadruplex structures adopted by this sequence have been extensively characterized *in vitro* and have been proposed to play key roles in the expression of this oncogene (38). The c-MYC sequence has been shown to adopt an extremely dynamic G-quadruplex structure in K<sup>+</sup>, where the parallel structure is believed to predominate but is in active equilibrium with antiparallel structures (Fig. 3.10a) (39–41). Even though the exact type of antiparallel structures adopted by this sequence has not been completely characterized, non-adjacent photoproducts were not expected to form due to the short loops found in this sequence.

When the c-MYC sequence was irradiated in the presence of  $K^+$  and assayed with the NP1 enzyme, no tetramer products were detected by HPLC with the expected retention time between 30 and 36 minutes (Fig. 3.10b). Likewise, there were no photo-revertible bands leading to dinucleotides expected from *anti* CPDs (Fig. 3.9a lanes 4,5) indicating that no photoreactive conformations were present. Strangely, there was a significant amount of almost full-length products and bands with the mobility approximating tetramers that were not photoreversible (Fig. 3.9a lane 5,6) suggesting that they might contain non-photoreversible 64PP or Dewar products.

When the post-labeling assay was performed in the presence of  $Na^+$ , a small amount of a photoreversible tetramer was observed without any of the partially digested products observed in  $K^+$  or the non-photoreversible tetramer-like products (Fig. 3.9a lane 6,7). The lack of significant *anti* CPD formation would be consistent with the presence of antiparallel G-quadruplex structures that are unphotoreactive due to the short loops. In the presence of  $Li^+$  ions (Fig 3.9a ,8,9 & b) a significant amount of *anti* CPD was observed which would be consistent with the formation of reverse Hoogsteen hairpins. The photoreactive conformation could either be a two-nucleotide bulge loop formed with T2 and T3, or alternatively, a shorter hairpin in which T1 can react with T2. The involvement of T1 could be easily determined as it is the 5'-terminal nucleotide by pre-labeling the 5'-end.

#### *ILPR promoter sequence*

The G-quadruplex forming sequence found in the promoter region of the insulin (ILPR) genes was similarly tested for non-adjacent CPD formation with the post-labeling assay in the presence of  $K^+$ ,  $Na^+$ , and  $Li^+$  (Fig. 3.9). This sequence has previously been shown by *in vitro* structural studies

to primarily adopt a mixture of antiparallel and parallel G-quadruplex structures in  $K^+$ , while in  $Na^+$  it predominantly adopts antiparallel types of G-quadruplex structures (Fig. 3.8a) (42–44).

Even though the exact type of antiparallel structures adopted by this sequence has not been completely characterized, the T bases in loops one and three are present in the first and third positions, which may not be optimal for *anti*-CPD formation. The post-labeling assay showed a cation dependent formation of *anti* CPDs, where in the presence of  $K^+$  a high amount of non-photoreversible short degradation products were observed (Fig. 3.9 a lanes 11 & 12). A small amount of the photo-revertible band was observed in  $Na^+$  with a sizable amount of partially degraded products that were not photoreversible (Fig. 3.9 a lanes 13 & 14). In  $Li^+$  very little *anti* CPD was observed (Fig. 3.9 a lanes 15 & 16), which is puzzling but may be because the reverse Hoogsteen structure that is adopted by this sequence may disfavor cycloadditions between T's in positions 1 or 3 of the opposing loops and needs more investigation. Also worthy of more investigation is the nature of the non-photoreversible tetramer-like products observed in both the  $K^+$  and  $Na^+$  samples.

### 3.5 Conclusions

We have shown that the radioactive post-labeling assay developed in Chapter 2 for the detection of non-adjacent *anti* CPDs is a useful method for screening DNA sequences for photocrosslinkable non-B structures and corroborating and complementing previous enzyme-coupled mass spectrometry assays. The ability of 254 nm light to efficiently photoreverse CPD photoproducts, enables DNA that is photocrosslinked by CPDs to be distinguished from non-photorevertible DNA photoproducts such as 64PPs or Dewar products, or simply partially degraded DNA. The assay

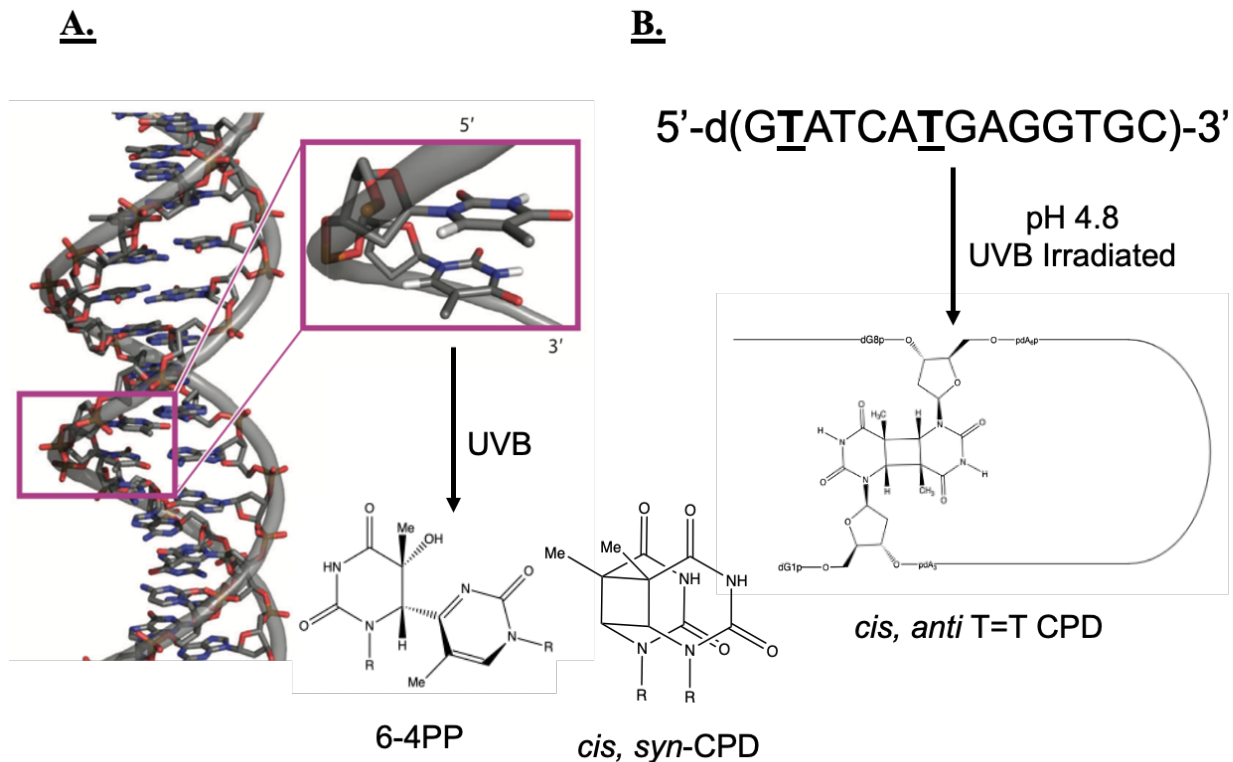
also facilitates the study of sequences for which the CPDs have been previously identified by MS under varying conditions such as metal ions and ionic strength but could also be used for kinetic studies. In its present form, however, the post-labeling assay cannot distinguish between *trans,anti*- and *cis,anti*-T=T CPDs or identify the flanking bases, but it may be possible to do so in the future by using two-dimensional electrophoresis or paper chromatography. None-the-less, the post-labeling assay together with the NP1-MS assay enables a more complete study of the photoreactivity of non-B secondary structures will be required if non-adjacent photoproducts are to be used as intrinsic probes for these type of DNA structures *in vivo*.

### **3.6 Acknowledgements**

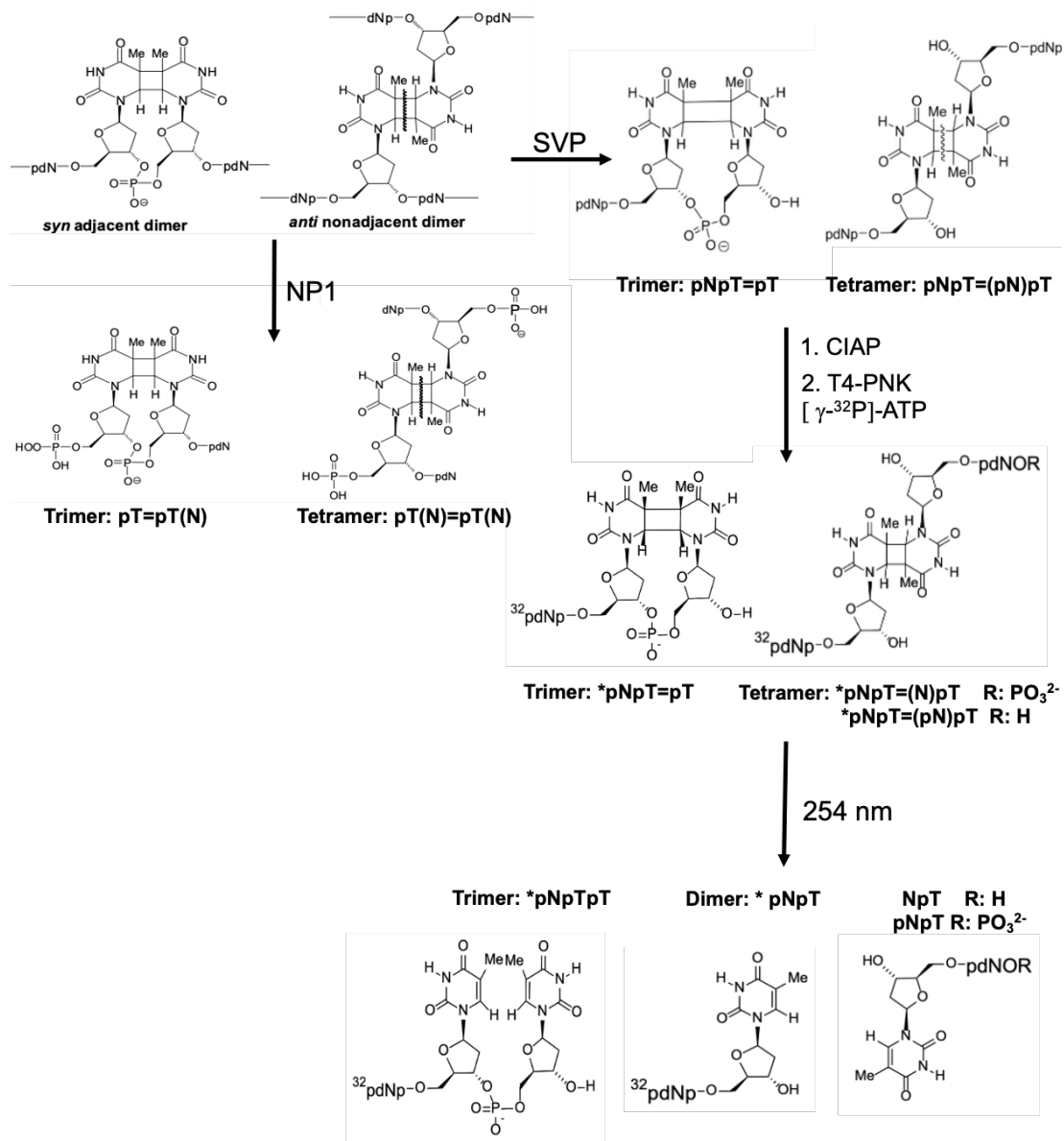
We thank Dr. Hani Zaher for allowing us to use his Amersham Typhoon biomolecular imager to image our gels. Research reported in this chapter was supported by the National Science Foundation under Grant No. 2003688.



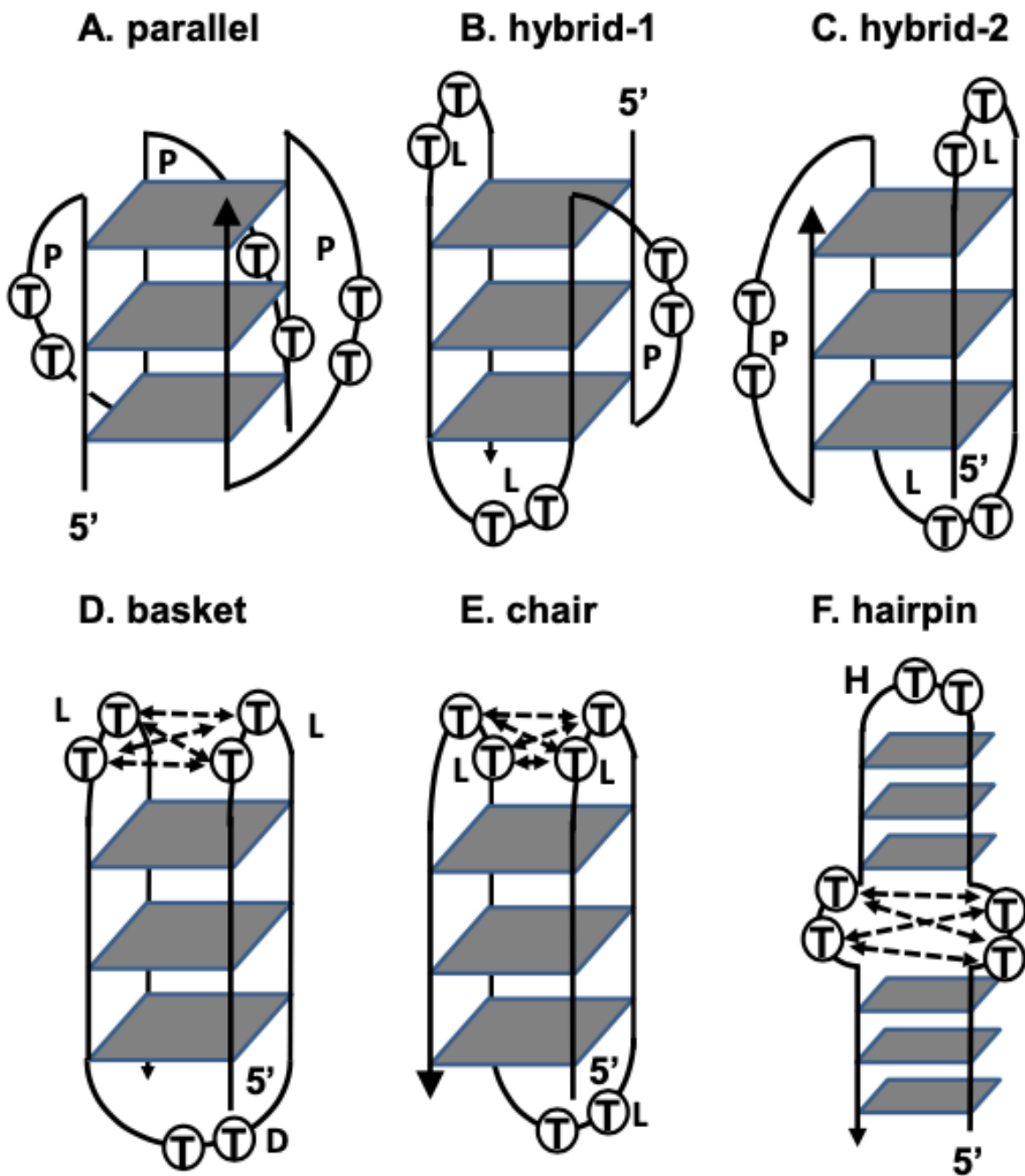
### 3.7 Figures



**Figure 3.1. The primary conformation of DNA determines the type of dipyrimidine photoproducts that form** (a) In helical duplex B DNA, adjacent pyrimidine bases react to produce CPDs with the *cis, syn* stereochemistry and (6-4) photoproducts. (b) In some non-B forms of DNA, non-adjacent pyrimidine bases are brought together to produce photoproducts such as *anti* CPDs. The first thoroughly characterized *cis, anti* T=T CPD was discovered from the 14-mer sequence shown upon UVB irradiation at low pH. To this day the structure of the DNA leading to this product is unknown.

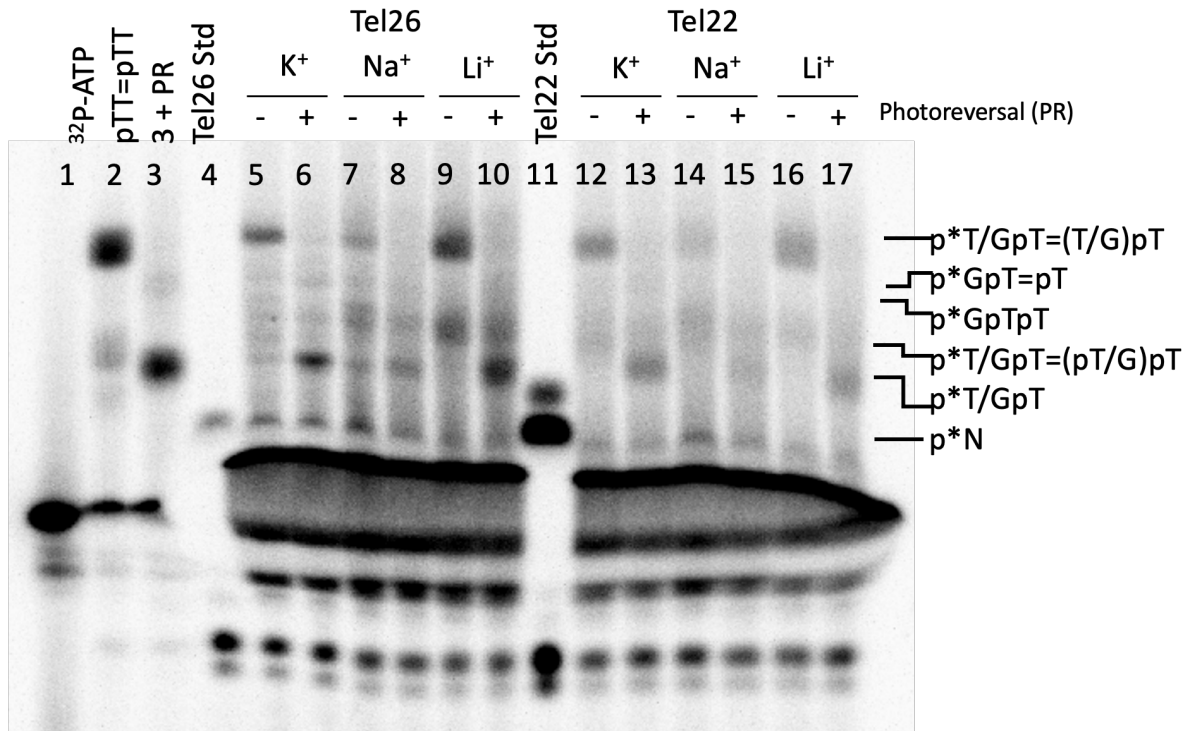


**Figure 3.2. Enzymatic degradation products of adjacent and non-adjacent photoproducts by NP1 and SVP.** The degradation products produced by these enzymes depend on whether they can cleave the 3'-side (NP1) or 5'-side (SVP) of an undamaged nucleotide. Adjacent photoproducts degraded by these enzymes produce trimers, where NP1 is unable to cleave the nucleotide 3'-adjacent to the photoproduct, while SVP cannot cleave the nucleotide 5'-adjacent to the damage. In the presence of non-adjacent photoproducts, both enzymes produce tetramers that differ by which nucleotides are left untouched at the 3' or 5' ends of the nucleotides involved in photoproduct formation. SVP degradation for both adjacent and non-adjacent products allows for the dephosphorylation and rephosphorylation of the canonical 5'- nucleotide with radioactive  $^{32}\text{P}$  which is the basis of the post-labeling assay.

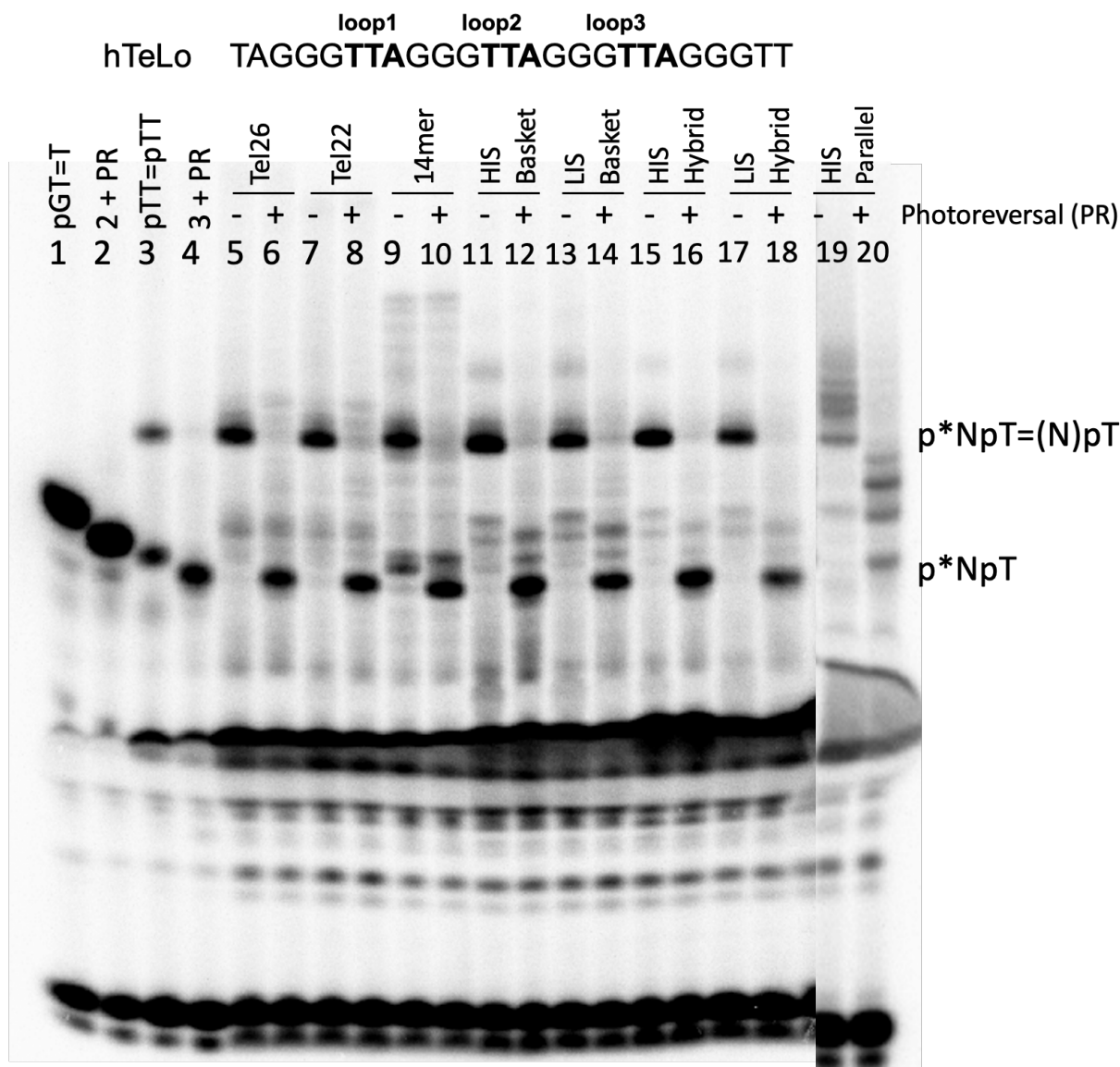


**Figure 3.3. Expected photochemical reactivity of G-quadruplex and hairpin structures of DNA.** Only adjacent CPDs are expected to form in (a) parallel and (b & c) both types of hybrid structures whereas both adjacent and non-adjacent CPDs are expected to form in (d) basket, (e) chair, and (f) hairpin structures.

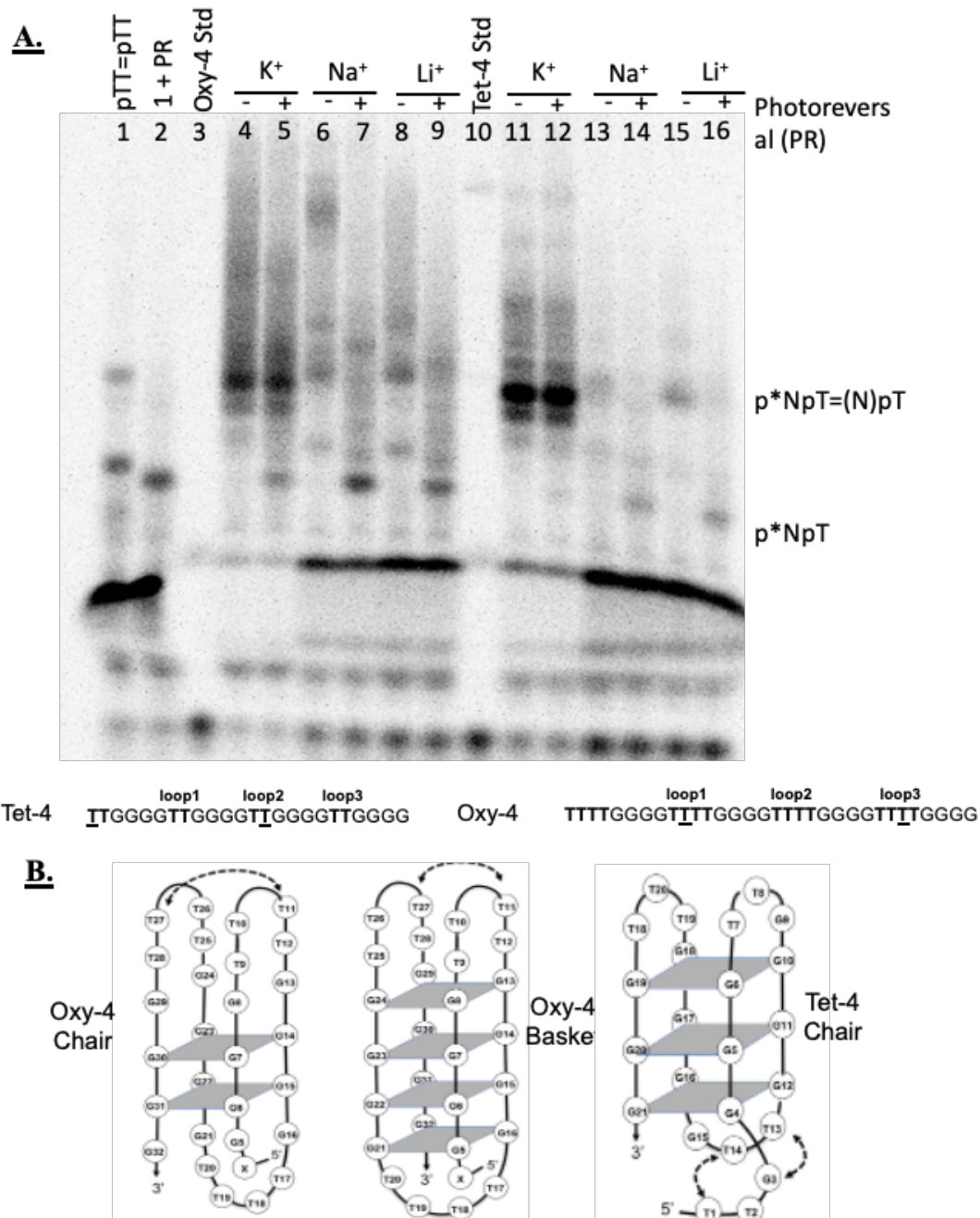
Tel26 AAAGGGT<sup>loop1</sup>TAGGGT<sup>loop2</sup>TAGGGT<sup>loop3</sup>TAGGGAA      Tel22 AGGGT<sup>loop1</sup>TAGGGT<sup>loop2</sup>TAGGGT<sup>loop3</sup>TAGGG



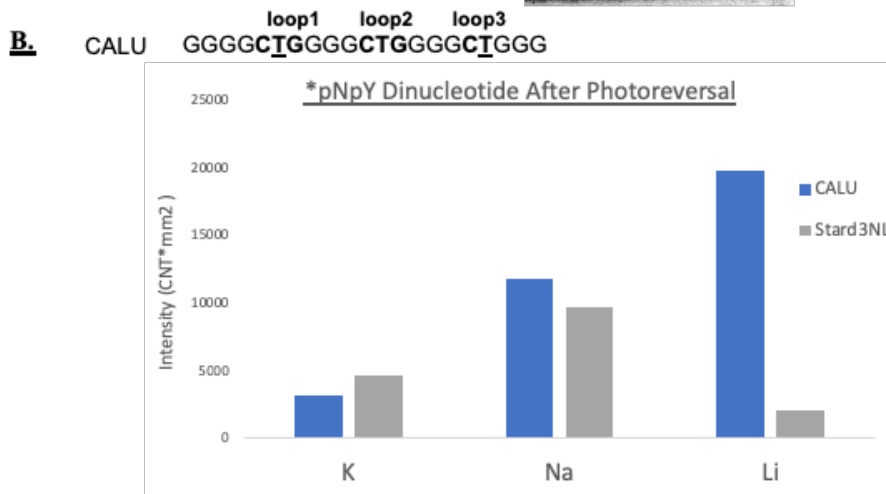
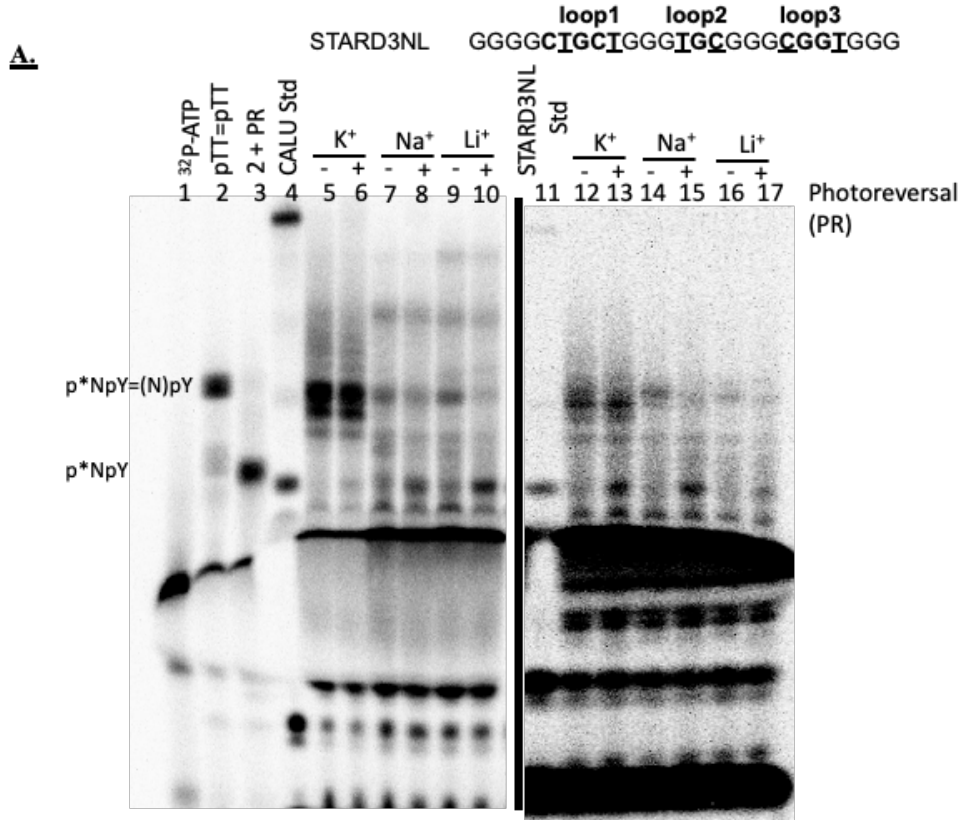
**Figure 3.4. Post-labeling assay of UVB irradiated Tel26 and Tel22 human telomeric sequences in the presence of K<sup>+</sup>, Na<sup>+</sup>, and Li<sup>+</sup>.** Lane 1 contains [γ-<sup>32</sup>P]-ATP used for radiolabeling, and lanes 2 and 3 contain radiolabeled p<sup>\*</sup>TpT=(p<sup>\*</sup>T/T)pT before and after photoreversal. Lanes 4 and 11 shows the products of SVP treatment of the 5'-end labeled Tel26 and Tel22 to test the efficiency of the digestion reaction and to produce authentic p<sup>\*</sup>dA. A 254 nm lamp was used to photoreverse T=T CPD-containing products.



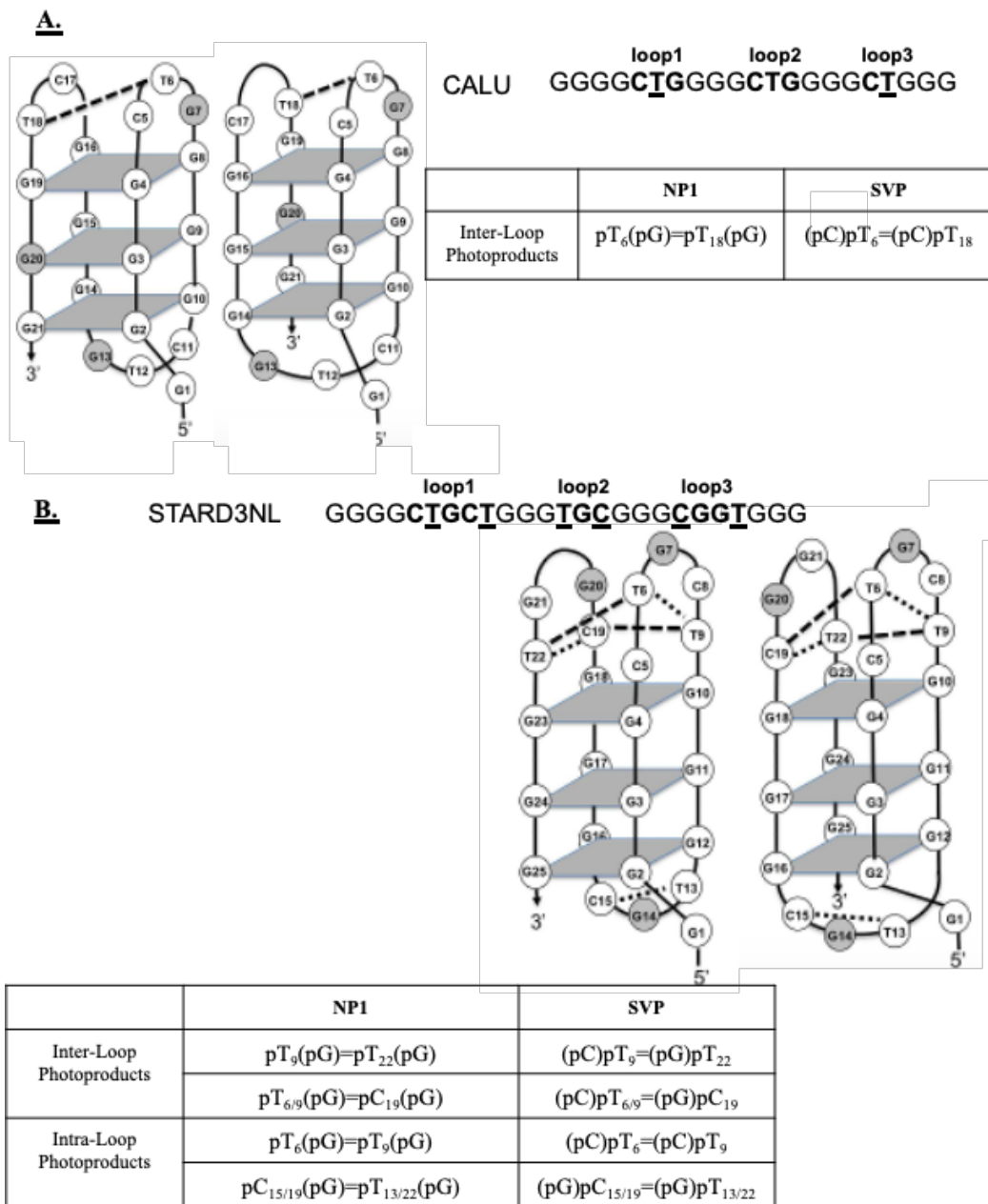
**Figure 3.5. Post-labeling assay of UVB irradiated hTeLo sequence under conditions that promote the formation of basket, hybrid, and parallel DNA G-quadruplexes.** Lanes 1-4 are authentic standards of p\*GpT=pT and p\*TpT=(p\*T/T)pT. Lanes 5-8 are the products from the post-labeling assay of Tel26 and Tel22 sequences irradiated with UVB in K<sup>+</sup>. Lanes 9-10 are the products of the post-labeling assay of the 14-mer d(GTATCATGAGGTGC) irradiated with UVB under acidic conditions to produce the *cis,anti* T=T CPD. Lanes 11-20 are conditions where hTeLo was assayed with the post-labeling technique under high ionic (HIS) and low ionic conditions (LIS) in the presence of Na<sup>+</sup> (basket) or K<sup>+</sup> (hybrid) to promote G-quadruplex formation. The parallel structure was induced in the presence of K<sup>+</sup> under highly dehydrating conditions. A 254 nm lamp was used to photoreverse T=T CPD-containing products.



**Figure 3.6. Post-labeling assay of UVB irradiated Tet-4 and Oxy-4 telomeric sequences in the presence of K<sup>+</sup>, Na<sup>+</sup>, and Li<sup>+</sup>.** (a) Lanes 1 and 2 are authentic standards of p\*TpT=(p\*T/T)pT. Lanes 3 and 10 show the products of SVP digested 5'-end-labeled Oxy-4 and Tet-4 as a reference for the completeness of the digestion and to produce p\*dT. A 254 nm lamp was used to photoreverse T=T CPD-containing products. (b) Proposed chair and basket G-quadruplex structures that could explain the photocrosslinking reactions originally observed in these sequences.

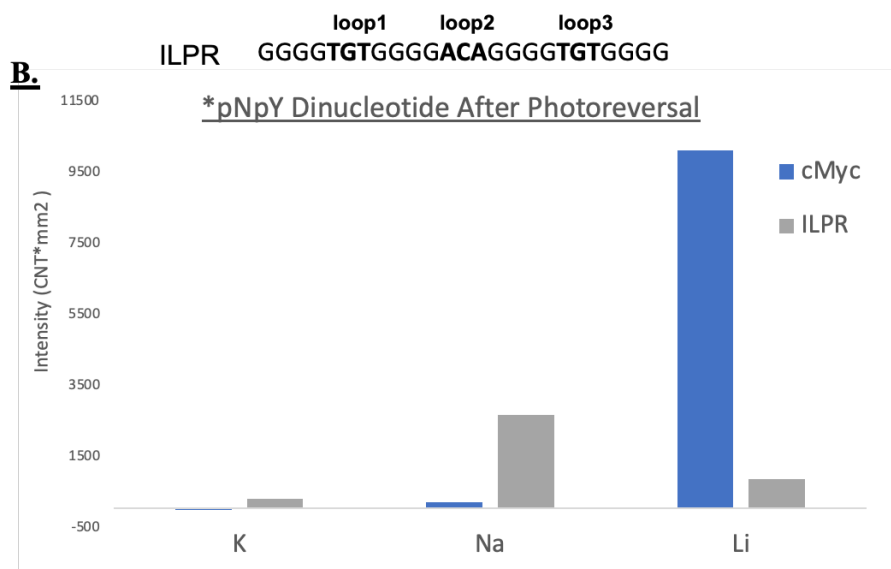
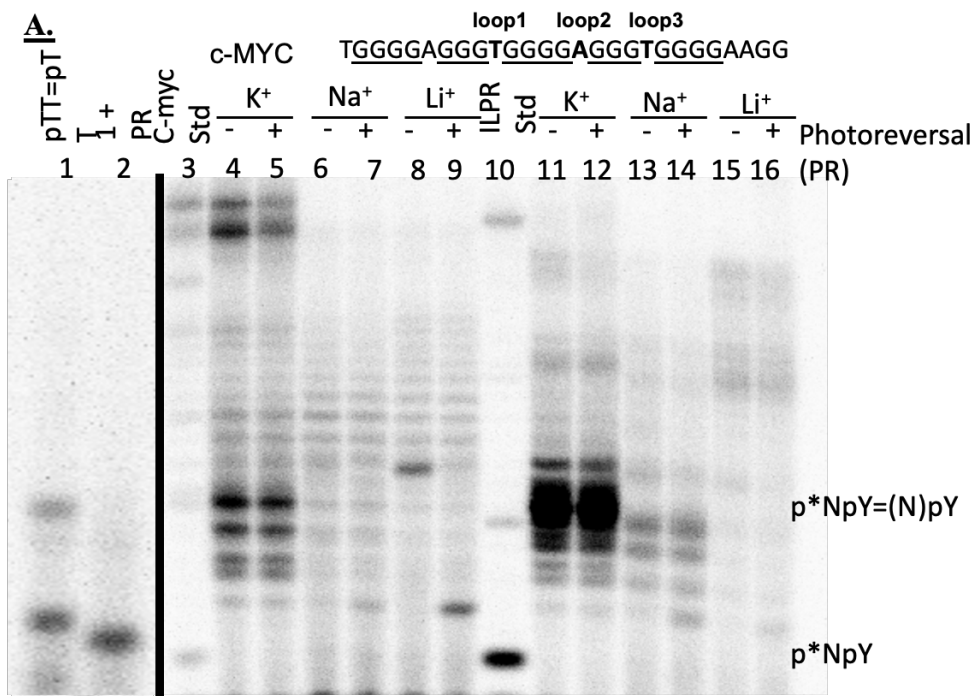


**Figure 3.7. Post-labeling assay of UVB irradiated CALU and STARD3NL human promoter sequences in the presence of K<sup>+</sup>, Na<sup>+</sup>, and Li<sup>+</sup>.** (a) Lane 1 contains [γ-<sup>32</sup>P]-ATP used in the assay, and lanes 2 and 3 are authentic standards of p\*TpT=(p\*T/T)pT. Lanes 4 and 11 show the products of SVP digested 5'-end-labeled CALU and STARD3NL as a reference for the completeness of digestion and as to produce p\*dG. A 254 nm lamp was used to photoreverse T=T CPD-containing products. (b) Bar graph showing the intensity of the pNpY dinucleotide radiolabeled bands following photoreversal with 254 nm light.

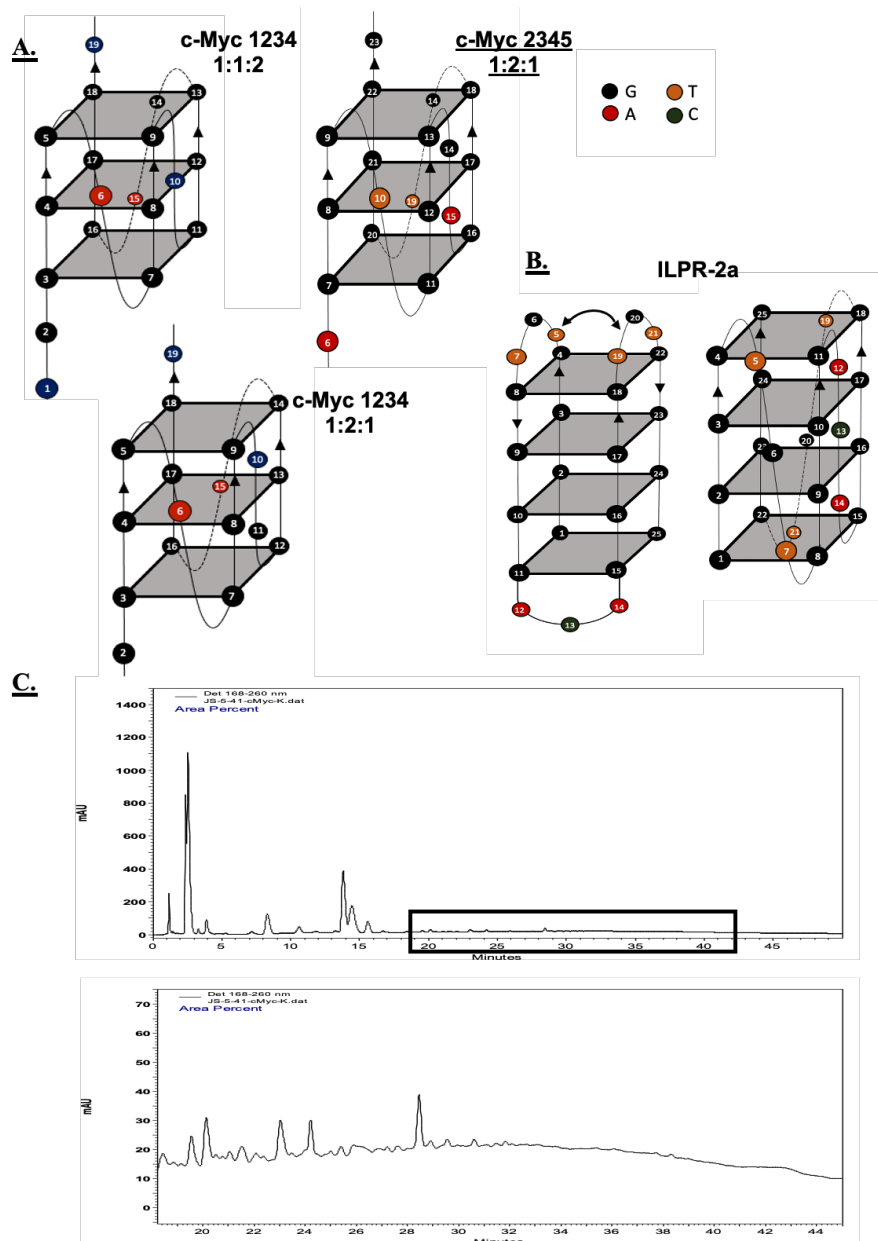


**Figure 3.8. Proposed sites of major non-adjacent CPD photoproducts formation in the CALU and STARD3NL promoter sequences with their respective SVP and NP1 tetramer digestion products.** Possible chair and basket antiparallel G-quadruplex structures adopted by the (a) CALU and (b) STARD3NL sequences to produce the dipyrimidine CPDs detected by NP1-MS analysis as indicated by the dashed and dotted lines. The dashed lines refer to inter-loop CPDs whereas the dotted lines refer to intra-loop CPDs. The tables below show the products observed from the NP1-MS assay and what would be expected to be observed from a MS assay if SVP were to be used suggesting that a combination of the two should be used in future assays.





**Figure 3.9. Post-labeling assay of UVB irradiated c-MYC and ILPR human promoter sequences in the presence of K<sup>+</sup>, Na<sup>+</sup>, and Li<sup>+</sup>.** (a) Lanes 1 and 2 are authentic standards of p\*TpT=(p\*T/T)pT. Lanes 3 and 10 contain the SVP digested products of 5'-end labeled c-MYC and ILPR sequences as a reference for the completeness of digestion and as a standard for p\*dT for c-MYC and p\*dG for ILPR. A 254 nm lamp was used to photoreverse T=T CPD-containing products. (b) Bar graph showing the intensity of the radiolabeled p\*NpY dinucleotide bands following photoreversal with 254 nm light.



**Figure 3.10. Proposed G-quadruplex structures adopted by the c-MYC and ILPR promoter sequences in the presence of  $K^+$ .** (a) c-MYC is proposed to adopt a set of parallel structures where the most stable conformation forms from the G bases in the II, III, IV, and V G-runs (c-MYC 2345) with a 1:2:1 loop arrangement, but it is also observed to form two more parallel structures involving the I, II, III, and IV G-runs (c-MYC 1234) with a 1:2:1 and 1:1:2 loop arrangement. These parallel structures could explain why (b) no NP1 tetramer products were detected from HPLC analysis of SVP digested UVB irradiated c-MYC in the presence of  $K^+$ . (c) The two proposed G-quadruplex structures adopted by the ILPR-a2 sequence where the antiparallel structure would be expected to produce non-adjacent photoproducts.

### 3.8 References

1. Beukers R, Eker APM, Lohman PHM. 50 years thymine dimer. *DNA Repair*. 2008 Mar 1;7(3):530–43.
2. Cadet J, Douki T. Formation of UV-induced DNA damage contributing to skin cancer development. *Photochem Photobiol Sci*. 2018 Dec 5;17(12):1816–41.
3. Schreier WJ, Gilch P, Zinth W. Early Events of DNA Photodamage. *Annual Review of Physical Chemistry*. 2015;66(1):497–519.
4. Taylor JS. Adjacent and Nonadjacent Dipyrimidine Photoproducts as Intrinsic Probes of DNA Secondary and Tertiary Structure†. *Photochemistry and Photobiology* [Internet]. [cited 2022 Nov 23];n/a(n/a). Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/php.13694>
5. Saini N, Zhang Y, Usdin K, Lobachev KS. When secondary comes first – The importance of non-canonical DNA structures. *Biochimie*. 2013 Feb 1;95(2):117–23.
6. Patrick MH, Gray DM. INDEPENDENCE OF PHOTOPRODUCT FORMATION ON DNA CONFORMATION. *Photochem Photobiol*. 1976 Dec;24(6):507–13.
7. Su DGT, Fang H, Gross ML, Taylor JSA. Photocrosslinking of human telomeric G-quadruplex loops by anti cyclobutane thymine dimer formation. *Proc Natl Acad Sci USA*. 2009 Aug 4;106(31):12861–6.
8. Lu C, Smith-Carpenter JE, Taylor JSA. Evidence for Reverse Hoogsteen Hairpin Intermediates in the Photocrosslinking of Human Telomeric DNA Sequences. *Photochem Photobiol*. 2018 Jul;94(4):685–97.
9. Su DGT, Kao JLF, Gross ML, Taylor JSA. 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 Aug 27;130(34):11328–37.
10. 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 Nov 1;12(11):1077–82.
11. Fujimoto M, Fujiyama K, Kuninaka A, Yoshino H. Mode of Action of Nuclease P<sub>1</sub> on Nucleic Acids and Its Specificity for Synthetic Phosphodiester. *Agricultural and Biological Chemistry*. 1974 Nov;38(11):2141–7.
12. 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 Jun 15;31(12):3134–42.

13. 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 Apr;42(8):5007–19.
14. Smith-Carpenter JE, Taylor JS. Photocrosslinking of G-Quadruplex-Forming Sequences found in Human Promoters. *Photochem Photobiol.* 2018 Aug 7;95(1):252–66.
15. Choi J, Majima T. Conformational changes of non-B DNA. *Chem Soc Rev.* 2011;40(12):5893.
16. Harkness RW, Mittermaier AK. G-quadruplex dynamics. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics.* 2017 Jun 20;1865(11):1544–54.
17. Williamson JR, Raghuraman MK, Cech TR. Monovalent cation-induced structure of telomeric DNA: The G-quartet model. *Cell.* 1989 Dec 1;59(5):871–80.
18. Siddiqui-Jain A, Grand CL, Bearss DJ, Hurley LH. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc Natl Acad Sci U S A.* 2002 Sep 3;99(18):11593–8.
19. An N, Fleming AM, Middleton EG, Burrows CJ. Single-molecule investigation of G-quadruplex folds of the human telomere sequence in a protein nanocavity. *PNAS.* 111(4):14325–31.
20. 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(9):2723–35.
21. Xu Y, Noguchi Y, Sugiyama H. The new models of the human telomere d[AGGG(TTAGGG)3] in K<sup>+</sup> solution. *Bioorganic & Medicinal Chemistry.* 2006 Aug 5;14:5584–91.
22. Parkinson GN, Lee MPH, Neidle S. Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature.* 2002 Jun;417(6891):876–80.
23. Phan AT, Kuryavyi V, Luu KN, Patel DJ. Structure of two intramolecular G-quadruplexes formed by natural human telomere sequences in K<sup>+</sup> solution†. *Nucleic Acids Res.* 2007 Oct;35(19):6517–25.
24. Dai J, Carver M, Punchihewa C, Jones RA, Yang D. Structure of the Hybrid-2 type intramolecular human telomeric G-quadruplex in K<sup>+</sup> solution: insights into structure polymorphism of the human telomeric sequence. *Nucleic Acids Res.* 2007;35(15):4927–40.
25. 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 Apr;35(7):2440–50.

26. Lim KW, Ng VCM, Martín-Pintado N, Heddi B, Phan AT. Structure of the human telomere in Na<sup>+</sup> solution: an antiparallel (2+2) G-quadruplex scaffold reveals additional diversity. *Nucleic Acids Research*. 2013 Dec 1;41(22):10556–62.
27. Mitra J, Makurath MA, Ngo TTM, Troitskaia A, Chemla YR, Ha T. Extreme mechanical diversity of human telomeric DNA revealed by fluorescence-force spectroscopy. *Proceedings of the National Academy of Sciences*. 2019 Apr 23;116(17):8350–9.
28. Rajendran A, Endo M, Hidaka K, Sugiyama H. Direct and Single-Molecule Visualization of the Solution-State Structures of G-Hairpin and G-Triplex Intermediates. *Angewandte Chemie International Edition*. 2014;53(16):4107–12.
29. Blackburn EH, Gall JG. A Tandemly Repeated Sequence at the Termini of the Extrachromosomal Ribosomal RNA Genes in Tetrahymena. *Journal of Molecular Biology* [Internet]. 1977 Aug 19 [cited 2023 Mar 14];120(1). Available from: <https://reader.elsevier.com/reader/sd/pii/0022283678902942?token=55C4E91C2F81EEEE85C5612C41534E525945A33F4791761917596E3FFCB3824070EFFC20D6A96B900D6A6BAF6E7B8105&originRegion=us-east-1&originCreation=20230314175600>
30. Klobutcher LA, Swanton MT, Donini P, Prescott DM. All gene-sized DNA molecules in four species of hypotrichs have the same terminal sequence and an unusual 3' terminus. *Proc Natl Acad Sci U S A*. 1981 May;78(5):3015–9.
31. Smith CA, Taylor JS. Preparation and characterization of a set of deoxyoligonucleotide 49-mers containing site-specific cis-syn, trans-syn-I, (6-4), and Dewar photoproducts of thymidylyl(3'→5')-thymidine. *Journal of Biological Chemistry*. 1993 May;268(15):11143–51.
32. Arichi N, Inase A, Eto S, Mizukoshi T, Yamamoto J, Iwai S. Mechanism of the alkali degradation of (6–4) photoproduct-containing DNA. *Organic & Biomolecular Chemistry*. 2012;10(11):2318–25.
33. Wang Y, Patel DJ. Solution Structure of the Oxytricha Telomeric Repeat d(G4(T4G4)3) G-tetraplex | Elsevier Enhanced Reader. *Journal of Molecular Biology*. 1995 Aug 4;251(1):76–94.
34. Choi KH, Choi BS. Formation of a hairpin structure by telomere 3' overhang. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*. 1994 Apr 6;1217(3):341–4.
35. Miura T, Benevides JM, Thomas GJ. A phase diagram for sodium and potassium ion control of polymorphism in telomeric DNA. *Journal of Molecular Biology*. 1995 Jan 1;248(2):233–8.

36. Lee JY, Yoon J, Kihm HW, Kim DS. Structural Diversity and Extreme Stability of Unimolecular Oxytricha nova Telomeric G-Quadruplex. *Biochemistry*. 2008 Mar 1;47(11):3389–96.
37. Abu-Ghazalah RM, Macgregor RB. Structural polymorphism of the four-repeat Oxytricha nova telomeric DNA sequences. *Biophysical Chemistry*. 2009 May 1;141(2):180–5.
38. Simonsson T, Pecinka P, Kubista M. DNA tetraplex formation in the control region of c-myc. *Nucleic Acids Res*. 1998 Mar 1;26(5):1167–72.
39. Phan AT, Modi YS, Patel DJ. Propeller-Type Parallel-Stranded G-Quadruplexes in the Human c-myc Promoter. *J Am Chem Soc*. 2004 Jul 1;126(28):8710–6.
40. Seenisamy J, Rezler EM, Powell TJ, Tye D, Gokhale V, Joshi CS, et al. The Dynamic Character of the G-Quadruplex Element in the c-MYC Promoter and Modification by TMPyP4. *J Am Chem Soc*. 2004 Jul 1;126(28):8702–9.
41. Mathad RI, Hatzakis E, Dai J, Yang D. 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*. 2011 Nov;39(20):9023–33.
42. Yu Z, Schonhofs JD, Dhakal S, Bajracharya R, Hegde R, Basu S, et al. ILPR G-Quadruplexes Formed in Seconds Demonstrate High Mechanical Stabilities. *J Am Chem Soc*. 2009 Feb 11;131(5):1876–82.
43. Wickhorst PJ, Ihmels H, Paululat T. Studies on the Interactions of 3,11-Difluoro-6,8,13-trimethyl-8H-quinol[4,3,2-kl]acridinium and Insulin with the Quadruplex-Forming Oligonucleotide Sequence a2 from the Insulin-Linked Polymorphic Region. *Molecules*. 2021 Jan;26(21):6595.
44. Catasti P, Chen X, Moyzis RK, Bradbury EM, Gupta G. Structure–Function Correlations of the Insulin-linked Polymorphic Region. *Journal of Molecular Biology*. 1996 Dec 6;264(3):534–45.

# Chapter 4: Future Studies and Conclusion

## 4.1 Conclusions

While there is much evidence suggesting that DNA adopts G-quadruplex conformations *in vivo*, unequivocal evidence for their formation and involvement in gene expression, telomere biology, and genome stability is lacking (1). The precise regulatory roles that these non-B conformations of DNA have is still very much unknown due to the very limited set of tools that are currently available for identifying and mapping non-B DNA conformations that do not disrupt the DNA or the cellular environment (2). To circumvent these problems it was proposed that DNA itself could be used as an intrinsic photoprobe for certain classes of G-quadruplex structures and other non-B DNA conformations *in vivo* (3). It was previously discovered by our group that UV irradiation of certain G-quadruplex and reverse Hoogsteen hairpin forming sequences results in photocrosslinks between adjacent DNA loops by formation of unique, non-adjacent T=T cyclobutane pyrimidine dimers (CPDs). The formation of these non-adjacent CPDs led to the proposal that they could serve as irrefutable stable molecular markers for the presence and location of non-B DNA structures *in vivo* (4–6). In this thesis we developed highly sensitive methods for identifying the formation of non-adjacent CPDs and used it to explore the idea that these products could be used to detect non-B DNA conformations in various types of G-quadruplexes and reverse Hoogsteen hairpin forming sequences.

In Chapter 2 we reported the development of radioactive post- and pre-labeling assays for the detection of non-adjacent *trans,anti* and *cis,anti* T=T CPDs that form in human telomeric DNA. We showed that the post-labeling assay is capable of detecting *anti* photoproducts at approximately

a 10 fmol level, which was estimated to be about eight times higher than the amount of *anti* TT CPDs that would be produced if one *anti* TT CPD were to form per telomere in 10 million human cells. We expect that it would be difficult to detect this amount of *anti* CPDs in human genomic DNA without some sort of prior chromatographic and/or affinity purification steps due to the high background interference from partially digested DNA and digestion products of adjacent DNA photoproducts. The pre-labeling assay enabled us to confirm the identity of the post-labeled products and verify the nucleotides in the sequence involved in *anti* CPD formation. The unique ability of 254 nm light to photorevert labeled *anti* CPD tetranucleotides to dinucleotides enabled *anti* CPD to be distinguished from partially degraded DNA and other types of non-photoreversible non-adjacent photoproducts such as 6-4PP or Dewar photoproducts. Furthermore, photoreversion could be used to improve sensitivity by isolating the radiolabeled tetranucleotide band prior to photoreversion and then quantifying the photoreverted band free from background bands. While the post-labeling assay was validated for *anti* T=T CPDs it is also expected to be able to assay non-adjacent *anti* T=U, U=T and U=U CPDs resulting from deaminated C-containing CPDs, as well as *anti* CPD photoproducts of RNA.

In Chapter 3 we used the post-radiolabeling assay to investigate the photochemistry of a number of biologically relevant non-B secondary structures of DNA in order to determine the scope and limitations of the method. We showed that the post-labeling assay can be used to study *anti* CPD formation under varying conditions of metal ions and ionic strength and could potentially be used for kinetic studies. We also showed that this assay is helpful at corroborating and complementing previous enzyme-coupled mass spectrometry assays, and that together with the NP1-MS assay enables for a more complete understanding of the photoreactivity of non-B secondary structures



of DNA. Using this assay, we were able to confirm that the c-MYC G-quadruplex in  $K^+$  is very resistant to *anti* CPD formation in accord with its short loops and a preference for a parallel structure, but in  $Li^+$  solution *anti* CPDs were detected suggesting the presence of a reverse Hoogsteen hairpin. We also found that irradiation of the Tet-4 and Oxy-4 G-quadruplex forming sequences in  $K^+$  solution produced a high yield of non-photoreversible tetranucleotide-like products that might be due to formation of non-adjacent non-photoreversible DNA photoproducts such as 64PP or Dewar products. Many of these observations warrant further study by enzyme-coupled MS/MS studies.

## 4.2 Current and Future Studies

### 4.2.1 Detection and Location of *anti* CPDs in HeLa cells

The original motivation for developing the post-labeling assay for *anti* T=T CPDs was to see if human telomeres that cap the ends of chromosomes were adopting the photoreactive basket-like G-quadruplex structures *in vivo*. To minimize the possible interference with the detection of *anti* T=T CPDs produced in the telomeres from adjacent *cis,syn* CPDs trimers and partially degraded DNA resulting from the much larger proportion of genomic DNA a pre-purification strategy was undertaken (Fig. 4.1). HeLa cells were chosen as a source of human cells because of their ease of growth and because they have almost double the number of telomeres. The workflow for the *anti* T=T CPD post-labeling assay for HeLa telomeres is based on published methods with some modifications (7). First HeLa cells grown in 150 mm cell culture dishes are exposed to UVB light for 2.5 h while on ice and in cold PBS media. After irradiation, the genomic DNA is harvested with a kit in the presence of proteinase-K to ensure removal of proteins that may have been photocrosslinked to the DNA. Once the irradiated DNA is recovered, the telomeres are released

from the genomic DNA by overnight digestion with the restriction enzymes AluI, HinfI, HphI, and MnlI that do not recognize or cleave sequences from the telomeric regions (8). To capture the undigested telomeric DNA, the digested DNA is then incubated with biotinylated 5'-(CCCTAA)<sub>3</sub>-3' oligonucleotides that anneal to the single strand d(TTAGGG)<sub>n</sub> tail of the HeLa telomeric sequences. The biotinylated DNA is then captured by streptavidin-coated magnetic beads which are immobilized by a magnet allowing the non-telomeric DNA and enzymes to be washed away. Once the telomeric DNA is eluted from the biotin-streptavidin complex, the post-radiolabeling assay will be performed. Because background interference from genomic *cis,syn* CPDs and partially degraded products of genomic DNA has been eliminated, we expect that the post-labeling assay will enhance the sensitivity of the method and enable the detection of at least one *anti*-CPD per telomere from ten million HeLa cells. If successful, we will also try the post-labeling assay on genomic DNA. We would also plan to develop monoclonal antibodies against non-adjacent dipyrimidine DNA photoproducts for ELISA and ChIP-seq assays to detect them and map their locations in the genome. Once the potential sequences that produce these photoproducts are determined, the sequences can be further studied *in vitro* by post-labeling and enzyme-coupled MS/MS to determine what type of non-B DNA structure may have been involved. Similarly, if non-adjacent products are found from the HeLa irradiation experiments, they would constitute a new type of DNA damage that could contribute to DNA mutagenesis and cancer that would pose interesting challenges for DNA repair systems and would need to be further studied.

#### **4.2.1 Photochemical Signatures of Different Non-B Structures of DNA**

To determine what type of non-B DNA structures could have led to *anti* CPD formation in sequences identified by ChIP-seq and other methods *in vivo*, we would carry out post- and pre-

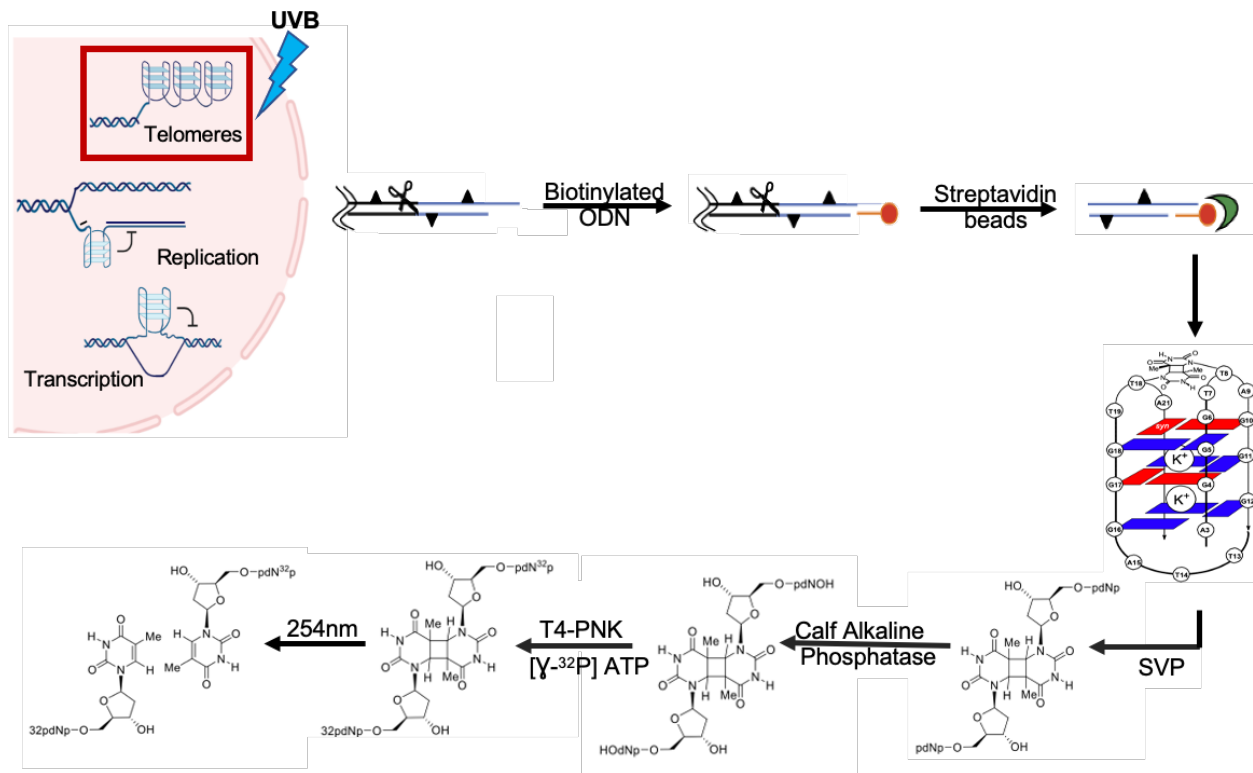
labeling assays and enzyme-coupled LC-MS assay of photoproduct formation in these sequences *in vitro*. These studies might reveal specific photoproduct signatures for specific types of non-B structures that would also be studied by chemical probes like DMS, permanganate, and ion mobility MS and NMR. We would then want to determine how loop lengths and pyrimidine placement within loops affect non-adjacent CPD formation in G-quadruplexes and reverse Hoogsteen hairpins, among other possible structures.

We could start by studying the G-quadruplex forming sequence found in the STARD3NL gene discussed in Chapter 3. The NP1-MS assay initially determined that a T=C and T=T *anti* CPD were the major non-adjacent photoproducts produced from UVB irradiation in the presence of K<sup>+</sup> and Na<sup>+</sup>, which could only narrow down their location to a set of possible intra- and inter-loop sites (6). By using an SVP-MS assay that identifies the base present at the 5'-side of the photoproduct, the flanking sequence information provided by these two enzymes would allow to determine if the photoproducts formed either from intra- or inter-loop reactions (Table 4.1& 4.2). Once the loops involved in the CPD formation is determined, the pre-labeling assay could then be used to identify the exact bases involved in the photoreactions by radiolabeling the various candidates (Table 4.3). As the T22 base found in loop 3 could be challenging to study through the pre-labeling assay due to possible difficulties that could be encountered in the ligation step needed to internally radiolabel that site, we could instead study its potential involvement with the SVP and Np1-MS assay by replacing it with a U base. As discussed in Chapter 3, the post-labeling studies done on this sequence could be complemented with a 2D gel using low pH or by 2D paper chromatography to distinguish the enzymatic photoproducts by sequence.

### **4.3. Concluding Remarks.**

We have been able to develop post- and pre-labeling assays for non-adjacent CPDs that can serve as intrinsic photoprobes for non-B DNA structures and may play a role in UV mutagenesis. Very little is still known about the structure activity relationships in non-adjacent CPD formation, and nothing is known about non-adjacent formation of other photoproducts such as (6-4), TA or spore photoproducts. Therefore, continued studies of this unique class of photoproducts and their formation in DNA is warranted and may lead to new insights of DNA structure and function.

### 4.3 Figures and Tables



**Figure 4.1. Workflow for the post-labeling assay for the detection of *anti* T=T CPDs photoproducts *in vivo*.** Once HeLa cells are irradiated with UVB light, the post-labeling assay will be performed on telomeric DNA that is purified through a biotin-streptavidin pull down method.

	<b>NP1</b>	<b>SVP</b>
<b>T<sub>13/22</sub>=C<sub>15/19</sub> CPD BASKET/CHAIR</b>		
Intra-Loop 2	pT <sub>13</sub> (pG <sub>14</sub> )=pU <sub>15</sub> (pG <sub>16</sub> )	(pG <sub>12</sub> )pT <sub>13</sub> =(pG <sub>14</sub> )pU <sub>15</sub>
Intra-Loop 3	pT <sub>22</sub> (pG <sub>23</sub> )=pU <sub>19</sub> (pG <sub>20</sub> )	(pG <sub>21</sub> )pT <sub>22</sub> =(pG <sub>18</sub> )pU <sub>19</sub>
<b>T<sub>6</sub>=C<sub>19</sub> CPD BASKET</b>		
Loop1-Loop3	pT <sub>6</sub> (pG <sub>7</sub> )=pU <sub>19</sub> (pG <sub>20</sub> )	(pC <sub>5</sub> )pT <sub>6</sub> =(pG <sub>18</sub> )pU <sub>19</sub>
<b>T<sub>9</sub>=C<sub>19</sub> CPD CHAIR</b>		
Loop1-Loop3	pT <sub>9</sub> (pG <sub>10</sub> )=pU <sub>19</sub> (pG <sub>20</sub> )	(pC <sub>8</sub> )pT <sub>9</sub> =(pG <sub>18</sub> )pU <sub>19</sub>

**Table 4.1. Predicted SVP and NP1 digestion products from the possible sites of non-adjacent T=C CPD formation in the STARD3NL.** Depending on the antiparallel structure adopted, two different intra-loop CPDs could have given rise to the non-adjacent T=C CPDs that were detected by NP1-MS assays to form in the STARD3NL sequence. The *anti* CPDs could have arisen from inter-loop reactions between C19 and T6 if in the basket form or C19 and T9 if in the chair form.

	<b>NP1</b>	<b>SVP</b>
<b>T<sub>6</sub>=T<sub>9</sub> CPD BASKET/CHAIR</b>		
Intra-Loop 1	pT <sub>6</sub> (pG <sub>7</sub> )=pT <sub>9</sub> (pG <sub>10</sub> )	(pC <sub>5</sub> )pT <sub>6</sub> =(pC <sub>8</sub> )pT <sub>9</sub>
<b>T<sub>9</sub>=T<sub>22</sub> CPD BASKET</b>		
Loop1-Loop3	pT <sub>9</sub> (pG <sub>10</sub> )=pT <sub>22</sub> (pG <sub>23</sub> )	(pC <sub>8</sub> )pT <sub>9</sub> =(pG <sub>21</sub> )pT <sub>22</sub>
<b>T<sub>6</sub>=T<sub>22</sub> CPD CHAIR</b>		
Loop1-Loop3	pT <sub>6</sub> (pG <sub>7</sub> )=pT <sub>22</sub> (pG <sub>23</sub> )	(pC <sub>5</sub> )pT <sub>6</sub> =(pG <sub>21</sub> )pT <sub>22</sub>

**Table 4.2. Predicted SVP and NP1 digestion products from the possible sites of non-adjacent T=T CPD formation in the STARD3NL sequence.** Depending on the antiparallel structure adopted, the T=T CPD could have arisen from intra-strand photoreaction between T6 and T9 in loop 1, or from T9 in loop 1 and T22 in loop 3 in the basket form or T6 and T22 if in the chair form.

STARD3NL	GGGG <u>C</u> <sub>6</sub> <u>G</u> <u>C</u> <u>T</u> <sub>9</sub> GGG <u>T</u> <sub>13</sub> <u>G</u> <u>C</u> <u>G</u> <u>G</u> <u>G</u> <u>C</u> <u>G</u> <u>G</u> <u>T</u> <sub>22</sub> GGG
STARD3NL-U <sub>22</sub>	GGGGCTGCTGGGTGCGGGCGG <u>U</u> GGG
STARD3NL- <sup>32</sup> pT <sub>6</sub>	GGGGC <sup>32</sup> <u>p</u> <u>T</u> GCTGGGTGCGGGCGGTGGG
STARD3NL- <sup>32</sup> pT <sub>9</sub>	GGGGCTGC <sup>32</sup> <u>p</u> <u>T</u> GGGTGCGGGCGGTGGG
STARD3NL- <sup>32</sup> pT <sub>13</sub>	GGGGCTGCTGGG <sup>32</sup> <u>p</u> <u>T</u> GCGGGCGGTGGG
STARD3NL- <sup>32</sup> pC <sub>15</sub>	GGGGCTGCTGGGTG <sup>32</sup> <u>p</u> <u>C</u> GGGGCGGTGGG
STARD3NL- <sup>32</sup> pC <sub>19</sub>	GGGGCTGCTGGGTGCGGG <sup>32</sup> <u>p</u> <u>C</u> GGTGGG

**Table 4.3. Proposed pre-labeled or T/U replacement sequences to determine the bases involved in *anti* T=T and T=C CPDs that formed in the STARD3NL sequence.** To determine which pyrimidine base is involved in the formation of the non-adjacent CPD photoproducts, any base in loop 1, 2, and 3 could be pre-radiolabeled or if in the case of T22 it can be replaced by a uracil base.

## 4.4 References

1. Robinson J, Raguseo F, Nuccio SP, Liano D, Di Antonio M. DNA G-quadruplex structures: more than simple roadblocks to transcription? *Nucleic Acids Res.* 2021 Sep 7;49(15):8419–31.
2. Kwok CK, Merrick CJ. G-Quadruplexes: Prediction, Characterization, and Biological Application. *Trends in Biotechnology.* 2017 Oct 1;35(10):997–1013.
3. Taylor JS. Adjacent and Nonadjacent Dipyrimidine Photoproducts as Intrinsic Probes of DNA Secondary and Tertiary Structure†. *Photochemistry and Photobiology* [Internet]. [cited 2022 Nov 23];n/a(n/a). Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/php.13694>
4. Su DGT, Fang H, Gross ML, Taylor JSA. Photocrosslinking of human telomeric G-quadruplex loops by anti cyclobutane thymine dimer formation. *Proc Natl Acad Sci USA.* 2009 Aug 4;106(31):12861–6.
5. Lu C, Smith-Carpenter JE, Taylor JSA. Evidence for Reverse Hoogsteen Hairpin Intermediates in the Photocrosslinking of Human Telomeric DNA Sequences. *Photochem Photobiol.* 2018 Jul;94(4):685–97.
6. Smith-Carpenter JE, Taylor JS. Photocrosslinking of G-Quadruplex-Forming Sequences found in Human Promoters. *Photochem Photobiol.* 2018 Aug 7;95(1):252–66.
7. Wright WE, Tesmer VM, Huffman KE, Levene SD, Shay JW. Normal human chromosomes have long G-rich telomeric overhangs at one end. *Genes Dev.* 1997 Nov 1;11(21):2801–9.
8. Kimura M, Stone RC, Hunt SC, Skurnick J, Lu X, Cao X, et al. Measurement of telomere length by the Southern blot analysis of terminal restriction fragment lengths. *Nat Protoc.* 2010 Sep;5(9):1596–607.



