Effect of DNA Sequence Context, DNA structure, and Excitation Method on Cyclobutane Pyrimidine Dimer Formation

Chen Lu
Washington University in St. Louis

Follow this and additional works at: https://openscholarship.wustl.edu/art_sci_etds

Part of the Biochemistry Commons

Recommended Citation
https://openscholarship.wustl.edu/art_sci_etds/1559

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.
Effect of DNA Sequence Context, DNA structure, and Excitation Method on Cyclobutane Pyrimidine Dimer Formation
by
Chen Lu

A dissertation presented to
The Graduate School of
Washington University in
partial fulfillment of the
requirements for the degree
der of Doctor of Philosophy

May 2018
St. Louis, Missouri
# Table of Contents

List of Figures .......................................................................................................................... v
List of Tables ............................................................................................................................. vii
Acknowledgments ...................................................................................................................... viii
Abstract ...................................................................................................................................... xi

Chapter 1  Introduction ................................................................................................................ 1
  1.1 Sunlight and skin cancer .......................................................... ........................................... 1
  1.2 DNA and UV Irradiation ................................................................. ................................. 1
  1.3 DNA Photoproducts ......................................................................................... 3
    1.3.1 Dipyrimidine photoproducts ....................................................................................... 3
    1.3.2 Minor DNA photoproducts ......................................................................................... 5
  1.4 The discovery of CPDs ................................................................................. 5
  1.5 Factors on CPD formation ............................................................................. 7
    1.5.1 Effect of UV wavelength and dinucleotide sequences ........................................ 8
    1.5.2 Effect of Flanking sequences .................................................................................... 10
    1.5.3 Effect of Methylation .............................................................................................. 11
    1.5.4 Effect of DNA structure .......................................................................................... 12
      1.5.4.1 B-form DNA ........................................................................................................ 12
      1.5.4.2 A-form DNA ........................................................................................................ 14
      1.5.4.3 G-quadruplex ........................................................................................................ 15
    1.5.5 Effect of chromatin structure .................................................................................. 17
    1.5.6 Effect of protein binding ......................................................................................... 17
  1.6 Mechanism of CPD formation ................................................................. 18
    1.6.1 Singlet or Triplet? .................................................................................................. 18
    1.6.2 Photosensitization of CPD formation .................................................................... 19
  1.7 Conclusions and Subject of this Thesis .................................................... 21
  1.8 References ............................................................................................... 23
  1.9 Figures ................................................................................................... 38

Chapter 2  Evidence for Reverse Hoogsteen Hairpin Intermediates in the Photocrosslinking of Human Telomeric DNA Sequences ............................................................... 46
  2.1 Abstract ............................................................................................... 46
  2.2 Introduction ........................................................................................ 47
  2.3 Materials and Methods ..................................................................... 49
### Chapter 3

#### Results and Discussion
- **2.4.1 Methodology** .................................................. 52
- **2.4.2 Photoprocess formation in Tel26 as a function of metal ion** ........................................ 53
- **2.4.3 Effect of metal ions on the conformation and stability of Tel26** ........................................ 53
- **2.4.4 Kinetics of CPD formation in Tel26 and NF3** ............................................................. 54
- **2.4.5 Inhibitory effect of Na+ and higher temperature on anti CPD formation** .................... 55
- **2.4.6 Temperature dependence and kinetics of anti CPD formation in Tel26 in Li+ and K+ solution** 56
- **2.4.7 Photoprocess formation in 15-mer sequences** ............................................................ 57
- **2.4.8 Photoprocess formation in inosine and Watson Crick base pair containing 21-mer sequences** 59
- **2.4.9 Trans,anti CPD formation in non-telomeric sequences** ............................................. 62
- **2.4.10 Electrophoretic behavior of hairpin and G-quadruplex forming sequences as a function of metal ion** 63

#### Discussion
- **2.5 Discussion** .......................................................... 65

#### Acknowledgements
- **2.6 Acknowledgements** ............................................... 70

#### References
- **2.7 References** .......................................................... 70

#### Tables and Figures
- **2.8 Tables and Figures** .................................................. 75

### Chapter 3

**A DNA Library for Determining the Sequence Dependence of Cyclobutane Pyrimidine Dimer Formation** .................................................. 98

#### Abstract
- **3.1 Abstract** .............................................................. 98

#### Introduction
- **3.2 Introduction** .......................................................... 99

#### Materials and Methods
- **3.3 Materials and Methods** .......................................... 103
- **3.3.1 Materials** .......................................................... 103
- **3.3.2 Python code for generating sequence libraries** .......................................................... 103
- **3.3.3 Synthesis of 149-mer containing plasmid** ............................................................. 103
- **3.3.4 Preparation of 5'-32P-labeled DNA substrates** ......................................................... 104
- **3.3.5 UV irradiation** ...................................................... 104
- **3.3.6 Mapping of CPD formation by T4-pdg Assay** ......................................................... 105
- **3.3.7 Quantification and data analysis** ...................................................... 105

#### Results and Discussion
- **3.4 Results and Discussion** ............................................. 107
### 3.4.1 Design of the sequence ................................................................. 107
### 3.4.2 Gel assay for CPD formation ...................................................... 108
### 3.4.3 Effect of UV light source and dipyrimidine sequence on CPD formation ................................................. 109
### 3.4.4 Effect of flanking bases on the formation of CPDs ........................ 110
### 3.4.5 CPD distribution as a function of photosensitizer ........................... 111

### 3.5 Conclusion ...................................................................................... 114

### 3.6 Acknowledgements ........................................................................... 114

### 3.7 References ....................................................................................... 114

### 3.8 Figures .............................................................................................. 118

### Chapter 4 Conclusions and Future Studies .............................................. 132

#### 4.1 Conclusions .................................................................................. 132

#### 4.2 Future Studies ............................................................................... 133

##### 4.2.1 The sequence context effect on UVA-induced CPDs.................. 133

##### 4.2.2 The effect of sequence on deamination of C-containing CPDs ...... 134

##### 4.2.3 In vivo detection of anti CPDs .................................................. 135

#### 4.3 References .................................................................................... 135

#### 4.4 Figures .......................................................................................... 137
List of Figures

Figure 1.1: DNA structure. ................................................................. 38
Figure 1.2: The UV absorption spectra of DNA .................................................. 39
Figure 1.3: Structures of dipyrimidine photoproducts. ...................................... 40
Figure 1.4: Stereoisomers of CPD. The isomers are shown for TT CPDs. .......... 41
Figure 1.5: Deamination mechanism. ......................................................... 42
Figure 1.6: B-form and A-form of DNAs. ..................................................... 43
Figure 1.7: G-quadruplex structures. ......................................................... 44
Figure 1.8: Structures of G-quadruplex in Na$^+$ and K$^+$ ............................ 44
Figure 1.9: Two possible mechanism leading to CPD formation: Singlet Pathway and .... 45
Figure 2.1: Structures and photoproducts of G-quadruplex forming sequences. ..... 75
Figure 2.2: HPLC traces of NP1-digested UVB irradiated Tel26 in the presence of Li$^+$, Na$^+$ and K$^+$ ions. ................................................................. 77
Figure 2.3: CD spectra and G-quadruplex UV melting temperature curves .......... 78
Figure 2.4: Kinetics of photoproduct formation in Tel26 and NF3 ....................... 79
Figure 2.5: Effect of cation composition and temperature on cis,syn and trans,anti CPD formation. ................................................................. 80
Figure 2.6: Base pairing possibilities in hairpin structures .................................. 81
Figure 2.7: Temperature dependence and kinetics of trans,anti CPD formation in Tel26. ............... 82
Figure 2.8: Photoproduct formation in 15-mers with Li$^+$, Na$^+$, and K$^+$ .......... 83
Figure 2.9: Photoproduct formation in ionsine- and Watson Crick base pair containing 21-mers with Na$^+$ and K$^+$ ......................................................... 84
Figure 2.10: Photoproduct formation in non-telomeric 21-mers with K$^+$ and Na$^+$ .... 85
Figure 2.11: Gel electrophoresis of non-telomeric and telomeric sequences .......... 87
Figure 2.12: Proposed explanation for the variation in anti CPD formation as a function of sequence and metal ion. ................................................................. 88
Figure 2.13: Denaturing PAGE gels of 5'-$^{32}$P-endlabeled Tel26 and Tel15 before and after irradiation in the presence of various cations. ......................................................... 89
Figure 2.14: C2-symmetric homopurine base pairs. ........................................... 90
Figure 2.15: CD of sequences used in this study .............................................. 92
Figure 2.16: Irradiation of Tel21_3’con (5’-GGGTTAGGTTAIIITTAIII-3’). ............ 93
Figure 2.17: Irradiation of Tel21_3’alt (5’-GGGTTAIIITTAGGTTAIII-3’). ................ 94
Figure 2.18: Irradiation of Tel21_5’con (5’-IIITTAIIITTAGGTTAIII). ................. 95
Figure 2.19: Irradiation of Tel21_5’alt (5’-IIITTAGGTTAIIITTAGGG). .............. 96
Figure 2.20: Irradiation of Tel21_mix (5’-IGITTAGGTTAIIITTAGGG-3’). ........... 97
Figure 3.1: Structure of the CPDs and photosensitizers. .................................... 118
Figure 3.2: The sequence of the 149-mer containing all 64 possible NPyPyN tetrads. .... 119
Figure 3.3: PAGE analysis of CPD formation by UV irradiation ......................... 120
Figure 3.4: Ordered bargraphs of the relative frequency of CPD formation at XPyPyY by UV irradiation................................................................. 122
Figure 3.5: Heatmap of relative CPD yields at XPyPyY sites for UV irradiated DNA. ............ 123
Figure 3.6: PAGE analysis of NFX and acetone photosensitized CPD formation.................. 124
Figure 3.7: Barplot and heatmap of CPD formation patterns in XTYY by photosensitization. 125
Figure 3.8: Sequence verification for the 149-mer containing clone................................. 127
Figure 3.9: Absorption spectra of the DNA and photosensitizers (dashed lines), and emission spectra of the light sources (solid lines)............................................................ 129
Figure 3.10: Time dependent study of UV-induced CPD formation in the 149-mer. .......... 130
Figure 3.11: Time dependence of CPD formation in the 149-mer with different UV light sources. ......................................................................................... 131
Figure 4.1: New designed sequence.................................................................................... 137
Figure 4.2: Sequencing result of the new sequence containing clone. ................................. 137
Figure 4.3: Scheme for study of sequence context effect on deamination of C-containing CPDs ..................................................................................... 138
List of Tables

Table 1.1: Distribution of CPDs: Summary of previously published studies.......................... 42
Table 2.1: Sequences used in this study.................................................................................. 76
Table 2.2: Gel electrophoresis of non-telomeric and telomeric sequences.............................. 86
Table 3.1: Sequences used in this study.................................................................................. 126
Table 3.2: Examples of sequences of 129-mers containing all possible NPyPyN sites.......... 128
Acknowledgments

I must first and foremost thank my advisor Dr. John-Stephen A. Taylor for giving me the opportunity to do my doctorate in his lab and for his years of continued support and guidance. He is one of the smartest people I know. He has always inspired me with creative ideas, insightful discussion and passion for science. I am also very grateful to him for giving me the freedom to pursue various projects without objection. None of this research would have been possible without him.

I would like to thank Dr. Robert Blankenship and Dr. Michael L. Gross for serving my Doctoral Advisory Committee, offering me helpful suggestions, and letting me use MALDI-TOF and CD instrument in their labs. I thank Dr. Mikhail Berezin and Dr. Timothy Wencewicz for serving my Doctoral Examination Committee and bringing perspectives to my thesis.

I thank all former and current lab members! I appreciate their insightful discussion and unreservedly sharing of experience. I especially thank Dr. Jill Smith for teaching me the techniques and skills of biochemistry.

I thank faculty and staff members of Department of Chemistry for their kind help and support through my PhD studies. I thank all friends within Chemistry Department for their encouragement and help not only in the research but also in life!
Last but not least, I thank my parents! They always encourage me despite they are thousands of miles away and know very little about science. Thank you for always being there and supporting my dreams!

Research reported in this thesis was supported by the National Cancer Institute of the National Institutes of Health under award number R01CA40463.

Chen Lu

Washington University in St. Louis

May 2018
ABSTRACT OF THE DISSERTATION

Effect of DNA Sequence Context, DNA Structure, and Excitation Method on Cyclobutane Pyrimidine Dimer Formation

by

Chen Lu

Doctor of Philosophy in Chemistry

Washington University in St. Louis, 2018

Professor John-Stephen A. Taylor

The \textit{cis-syn} cyclobutane pyrimidine dimer (CPD) is the major photoproduct resulting from UV irradiation of duplex DNA that results in C to T mutations found in human skin. CPDs with the \textit{anti} stereochemistry were recently discovered to be formed in human telomeric DNA that adopts a quadruplex structure \textit{in vitro} and may also play a role in the effects of sunlight \textit{in vivo}. In this thesis, the effect of telomeric DNA structure on the formation of the \textit{anti}-CPDs is investigated, as well as the effect of sequence context on \textit{cis-syn} CPD formation which could help explain the origin of DNA mutation hot spots and cold spots.

Originally it was hypothesized that \textit{anti} CPDs formed from G-quadruplex structures, but I now show that \textit{anti} CPDs can also form in human telomeric DNA sequences when complexed with lithium ions that are known to disfavor G-quadruplex formation. I also show that \textit{anti} CPDs can also form in the presence of potassium ion when selected guanine bases are changed to inosine to interfere with G-quartet formation. Most significantly we show that \textit{anti}-CPDs form in sequences containing A’s in place of G’s that cannot form Hoogsteen hairpins, but can form reverse Hoogsteen hairpins. These results suggest that reverse Hoogsteen hairpins may play a hitherto
unrecognized role in the biology and photoreactivity of DNA in telomeres, and possibly in other purine-rich sequences found in regulatory regions.

To study sequence context effects, we designed 129-mer sequences containing all 64 possible NPyPyN tetrads (where PyPy is CC, CT, TC, or TT, and N is A, C, G or T), and used a T4 endonuclease gel electrophoresis assay to determine the relative yields of photoproduct formation at the different sites. The results show that CPD yields for different tetrads varied by as much as an order of magnitude. The yield of CPDs under UVC irradiation at a given site decrease in the order TT > TC > CT > CC, whereas the yield of CPDs under UVB irradiation decreased in the order TT > TC > CC > CT. The yield of CPD formation was lowest with a 5’-G and highest with a 5’-T, whereas the yield was lowest with either a 3’-C, G or T, and highest with a 3’-flanking A. We also studied the sequence context effect on photosensitized CPD formation in the presence of acetone and norfloxacin. The results not only show that the efficiency of photosensitized CPD formation depends on the flanking bases, but also indicates that the efficiency depends on the structure and properties of the photosensitizer. These results suggest that the photosensitizers may result in unique mutation spectra that can be used to identify endogenous photosensitizers such as those implicated in chemi-excitation pathways in melanoma induction.
Chapter 1 Introduction

1.1 Sunlight and skin cancer

The sun is undeniably essential to life on earth. The solar light reaching the earth is composed of 5.4% UV, 31.9% infrared and 62.7% visible light (1). Excessive exposure to UV rays leads to adverse consequences such as premature aging of the skin, immunosuppression and damage to different structures of the eye. However, the main detrimental effect is the development of skin cancer. Although UV light comprises only 5% of solar radiation reaching the Earth’s surface, it accounts for approximately 93% of skin cancer cases (2). Consequently, UV is classified as the group 1 carcinogens (i.e., carcinogenic to humans) (3).

Skin cancer is the most common form of cancer among light-skinned population in many nations around the world. Generally, there are three major types of skin cancer, squamous cell carcinoma (SCCs), basal cell carcinomas (BCCs) and melanomas (4). The two non-melanocytic skin cancers, BCCs and SCCs, often occur at an advanced age and represents 80% and 15% of skin cancer cases, respectively (5,6). If they are detected sufficiently quickly, they can be treated by surgical excision. On the other hand, melanoma is known as the most aggressive and lethal form of skin cancer. Studies proves that the mortality rate of skin cancer are constantly increasing (7). It has been reported that 1 out of 5 Americans will develop skin cancer in their lifetime (8). Recently, the increasing use of artificial UV sources like tanning beds have aggravated the problem (2).

1.2 DNA and UV Irradiation

It has long been determined that deoxyribonucleic acid (DNA) is the main molecular target for UV-induced skin carcinogenesis (9,10). As early as 1928, Frederick L. Gates studied bactericidal
effect of UV light and concluded that this effect was related to the absorption by DNA based on the wavelength dependency analysis (11). Twelve years later, Hollaender and Emmons discovered that UV-induced mutations in fungi also correlated with the UV absorption spectrum of DNA (12). Nowadays it is commonly known that DNA carries genetic information, and its integrity is vital to the survival of living organisms. The famous double helical structure of DNA was discovered by Watson and Crick (13) in 1953 using X-ray diffraction data collected by Franklin and Wilkins (14). The DNA backbone is composed of alternating deoxyribose sugar-phosphate groups to which the DNA bases are joined via glycosidic bonds (Figure 1.1). The 5' carbon of the deoxyribose sugar has a phosphate group attached to it and the 3' carbon is attached by a hydroxyl group, giving the DNA strand a specific orientation. The two strands of DNA coil around the same axis in opposite directions, giving DNA an anti-parallel orientation.

There are four DNA bases. Adenine and guanine are the two purines, and thymine and cytosine are the two pyrimidines. Base stacking and base pairing represents the two factors responsible for the stability of the DNA structure. The base pairs are called Watson-Crick base pairs where A-T is paired via two hydrogen bonds whereas G-C is paired via three hydrogen bonds (Figure 1.1). Studies showed that the base stacking between the adjacent planar aromatic bases contributes much more to the stability of DNA structure than do the hydrogen bonds (15).

The peak absorbance wavelength of the four DNA bases is in the range of 250–270 nm whereas the DNA backbone absorbs below ~220 nm (10). DNA has a maximum absorption at about 260 nm (Figure 1.2). UV radiation can be classified into three categories, UVA (315-400 nm), UVB (280-315 nm), and UVC (100-280 nm), based on its biological effects. UVC light is the most energetic and therefore most harmful to the biological systems as the maximum absorption of DNA falls into this range. Luckily, UVC light can be efficiently absorbed by the
ozone layer and other protective molecules in the atmosphere, and is thus not considered biologically relevant, except perhaps in regions of the planet where there is an ozone hole. The short-wave UVB (280-300nm) is also blocked by ozone layers. However, the long-wave UVB can reach the Earth and contributes nearly 5% of terrestrial UV radiation (16,17). It overlaps with the longer wavelengths of DNA absorption, and can therefore be absorbed by the nucleobases. Research shows that UVB light can penetrate the superficial layers of the skin and results in the direct excitation of DNA (17). The major type of photodamage to DNA induced by UVB light is the class of dipyrimididine photoproducts, among which the cyclobutane pyrimidine dimers (CPDs) and pyrimidine (6–4) pyrimidone photoproducts (6-4PPs) are the major products (18-23). On the other hand, UVA represents the vast majority of terrestrial solar UV radiation, and is able to penetrate to the deeper layers of the skin (17). UVA is weakly absorbed by DNA; therefore, it is commonly considered to be responsible for indirect photodamage in the presence of endogenous and exogenous photosensitizing agents and by the formation of free radicals (24). It has recently been shown that CPDs are also predominantly formed by UVA irradiation, but there is still disagreement as to whether UVA-induced CPD formation occurs by direct or indirect excitation (23,25-31). Most recently it was discovered that DNA CPDs can continue to arise via a dark pathway for 2 to 3 hours after UVA and UVB irradiation of melanocytes, presumably through a chemisensitization mechanism (32). In summary, UV-induced DNA damage is sum of the direct and indirect mechanisms.

1.3 DNA Photoproducts

1.3.1 Dipyrimididine photoproducts

As mentioned earlier, DNA damage induced by UV irradiation is dominated by dipyrimididine photoproducts. There are four main class of dipyrimididine photoproducts detected and identified in
isolated cells when exposure to UV radiation (10,22,23). Their structures are shown in Figure 1.3. The most abundant type of photoprodu

The most abundant type of photoprodu, also the main topic of this thesis, is called a CPD, which is short for cyclobutane pyrimidine dimer. CPDs occur when two pyrimidine bases are linked by two C-C bonds forming a four-membered ring structure (33,34). A CPD adopts a conformation that distorts the B DNA axis by 7 to 9 degrees (35). This minor distortion makes it difficult for the human excinuclease DNA repair enzyme system to detect and repair CPDs, increasing their likelihood of being encountered by a DNA polymerases and resulting in a mutation.

The second major type of dipyrimidine photoprodu is called the pyrimidine (6-4) pyrimidone photoprodu (6-4PP). In the 6-4PP, two pyrimidine bases are linked by a single bond between C6 of one base and C4 of the other base. The formation of this product occurs via a Paternò–Büchi reaction to first give an unstable four-membered ring intermediate (oxetane when 3’ base is thymine or azetidine when 3’ base is cytosine) that then decomposes to the 6-4PP (36). The quantum yield for formation of the 6-4PP is one order of magnitude lower than that of CPD. The 6-4PP distorts the B DNA axis by 44 degrees (35) which makes it easily detectable and repaired by the human exonuclease, preventing it from encountering DNA polymerases. 6-4PP are mainly formed at TC and CC sites. They are less frequent at TT and very rarely seen at the CT level (22,37,38).

The 6-4PP has a maximum absorption around 320 nm. When it is further excited by long-wave UVB light (wavelength longer than 290 nm), it will be converted into another type of photoprodu, the Dewar valence isomer (39,40). It has been reported that a high proportion of the 6-4PP can be converted into its Dewar valence isomer under solar UV radiation (41). Overall, the above three types of dipyrimidine photoproducts have been found in human skin as well as other living systems (10,23). The fourth type of dipyrimidine photoprodu, 5-thyminyl-5,6-
dihydrothymine, is specific to bacterial spores under UVC irradiation, also called spore photoproducts (SP) (42,43). In the SP, the two pyrimidine bases are joined by linking the allylic carbon of the thymine to C5 of the other base. SPs are also reported to form in A-form DNA, which is the common form for RNA-DNA duplexes and RNA-RNA duplexes, though RNA does not have the allylic carbon present in T, as it uses U instead (44).

1.3.2 Minor DNA photoproducts

It should be also mentioned that a few other photoproducts have been detected and characterized besides dipyrimidine photoproducts, including thymine-adenine photoproduct (45-47), adenine-adenine photoproduct (48), and thymine and cytosine photohydrates (49). In addition, UVB and UVA could oxidize DNA bases through an indirect mechanism giving rise to photoproducts(50). Although these photoproducts occur at a quiet low frequency in comparison with CPD and 6-4PP, these photoproducts may play an important biological role owing to their cytotoxicity and mutagenesis.

1.4 The discovery of CPDs

In 1958, Beukers and coworkers discovered and identified the UVC-induced formation of thymine dimers in a frozen aqueous solution of thymines (51). Subsequently, Setlow and coworkers found that CPDs formed between adjacent pyrimidine bases when exposed DNA strands to UVC or UVB light (52). These important early works initiated the development of numerous experimental and computational studies about UV radiation on cellular DNA and DNA model systems. Nowadays, it is well established that CPD is the most prominent photoproduct within DNA after in vivo or in vitro UV radiation. The formation of CPD can occur at any dipyrimidine site (thymine, cytosine,
5-methylcytosine) via the 2+2 photocycloaddition between the C5-C6 double bonds of two pyrimidine bases (53-59).

There are four possible isomers of CPD (60,61). The stereochemistry of CPD is dependent on the conformation of the two pyrimidine bases at the time of irradiation, with cis/trans indicating whether the two bases are on the same side or the opposite sides of the cyclobutane ring, and syn/anti representing the parallel or antiparallel orientations of the C5-C6 carbons (Figure 1.4). All of these isomers have been characterized, mostly for thymine dimers (54). In native B-form DNA, only the cis-syn CPD has been reported (22). To a much lesser extent, other conformations are also found for different DNA structures or under specific conditions, which will be addressed in the later sessions in this thesis. Extensive quantitative studies on isolated and cellular DNA have demonstrated that the quantum yield of CPD is strongly dependent on the pyrimidine bases involved in the cycloaddition, with cis-syn TT as the major photoproduct (62-64).

CPDs are very stable compounds in solution. However, they can undergo some reactions. First, CPD’s can photoreverse back to the starting pyrimidines by splitting the cyclobutane ring upon further UVC excitation (65). Second, the cytosine or methyl-cytosine base in CPDs is very unstable and prone to deamination into uracil (U) or T, respectively (66-69) (Figure 1.5). The deamination half-life of C in a CPD is about 2-100 hours and is strongly accelerated compared to cytosine, which has a half-life of 30,000 years in duplex DNA, due to the loss of aromatic stabilization in the CPD (70). The process causes frequent C to T or CC to TT transition mutations at dipyrimidine sites, which marks the most significant mutational event observed in skin cancer. Studies have revealed that UVB-induced CPDs are correlated with the skin cancer hotspots in p53 gene (71,72). Studies have also reported that C5-methylation of cytosine enhances CPD formation
(72) but decreases the deamination rate (73). Deamination is also slower in *trans-syn* than in *cis-syn* CPDs (69).

CPDs act as blocks to replication and transcription and can lead to mutations when encountered by a DNA damage bypass polymerase. Therefore, in order to maintain the integrity of genomic information, the cells have developed mechanisms to repair the DNA damage they suffer. Collectively, cells have developed three main mechanisms for repairing UV-induced damage: photoreactivation, base excision repair (BER) and nucleotide excision repair (NER) (20,61). Photolyases catalyze the repair of CPDs and 6-4PPs using a photo-induced electron transfer process (74-77). T4-pdg (pyrimidine dimer glycosylase, previously known as T4 endonuclease V) is a glycosylase capable of repairing CPDs. It functions by allowing the abnormal base to be placed outside the double helix and to be found in the active site of the enzyme. This method of repairing UV-induced damage is present in prokaryotes and several eukaryotes but is not present and functional in humans (76,78,79). In human and many other organisms, CPDs are removed by the NER process. In a typical NER process, a CPD-containing DNA fragment is recognized and cut on either side by a DNA repair complex to release a single strand DNA fragment containing the CPD. Subsequently, the gap is filled by DNA polymerase using the complementary strand as the template and is ligated. Oxidative damage to the bases is commonly repaired by BER (61).

### 1.5 Factors influencing CPD formation

Since the characterization of the first CPD photoproduct, a multitude of studies has been published for understanding the factors involved in the formation of CPDs. The formation of CPDs has been shown to vary with all kinds of factors, such as UV wavelength and dose, dinucleotide sequence, flanking sequence, DNA methylation, DNA structure, chromatin structure, as well as the bending
of the DNA and its interaction with proteins. We will discuss all these factors in the following sections.

1.5.1 Effect of UV wavelength and dinucleotide sequences

Studies have shown that the distribution of CPDs at the four types of dipyrimidine sites (TT, CC, TC and CT) is influenced by the wavelength of UV light (27,64,80-83). Herein, we summarize in vitro and in cellulo data from the previous studies in Table 1.1. Using DNA sequencing gel techniques, Mitchell et al. studied the CPD distribution in the Alu sequence in vitro, and estimated that the ratio of CPD formation at TT, TC, CT, CC sites was 68: 16: 13: 3 under UVC irradiation, and 52: 21: 19: 7 under UVB irradiation (81). Another study by Douki and Cadet used high performance liquid chromatography associated with electrospray tandem mass spectrometry (HPLC-MS/MS) to estimate CPD distribution within isolated and cellular DNA. They reported that in vitro the UVC-induced distribution is 2.970: 1.823: 0.573: 0.069 (55: 34: 11: 1) for TT, TC, CT and CC whereas UVB-induced CPD followed an order of 1.023: 0.694: 0.289: 0.094 (49: 33: 14: 4) in vitro and 3.147: 0.279: 1.289: 0.577 (60: 24: 11: 5) in cellulo (64). These works showed that CPDs predominantly form at TT sites followed by TC, CT and CC. Also, compared with UVC, UVB leads to more cytosine-containing CPDs. Another interesting observation worth pointing out is that isolated DNA and cellular DNA showed similar DNA photoproduct distribution under UVB irradiation, with the frequency order as TT > TC > CC > TC > CC. In a following study, it was found the similar photoproduct distribution for UVB-irradiated human skin samples (84).

However, other studies revealed that the yield at CC sites could be even higher under some conditions (83,85). A study using ligation-mediated polymerase chain reaction (LMPCR) to explore the CPD distribution within CHO cells showed a distribution of CPDs at the TT: TC: CT:
CC of 27: 26: 17: 30 after exposure to UVB and 26: 24: 14: 35 after exposure to UVC (27). Another study used the same technique and quantified CPD distribution in UVB-irradiated human fibroblasts and purified DNA to be 25: 29: 20: 26 \textit{in vitro} and 27: 25: 21: 27 \textit{in cellulo} at TT, TC, CT, CC sites respectively (83). The difference among these studies was attributed to the different sequence, method and UV settings used. Actually, it has been found that there is a strong correlation between GC content of the genome and the yield of C-containing pyrimidine dimers (the higher the GC\% the higher the percentage yield of CC is). At low \%GC, TT remains the main CPD whereas if the \%GC increases to 75\% CC CPD would actually be the major product (85). The correlation between GC content and CPD distribution was observed for UVC, UVB as well as UVA-irradiated DNAs (22,85,86).

Compared with well-established CPD distribution within UVC- or UVB-induced DNA, it is not fully understood how CPDs are formed by UVA irradiation. Not until recently did UVA-induced CPDs receive much attention because extensive evidence showed that CPDs rather than oxidative lesions like 8-oxoGua were the main type of DNA damage involved in UVA irradiation, in both cultured cells and human skin (26,28,87,88). With UVA, the quantum yield of CPD is two orders of magnitude lower than for UVB and UVC, but nearly five times as high as that for 8-oxoGua (23). Experiments revealed that the distribution of dipyrimidine photoproducts is quite different for UVA compared with UVB and UVC, with TT CPD always being the most predominant photoproduct (27,86). Mouret reported that 90\% of the UVA-induced dipyrimidine photoproducts are TT CPDs, followed by TC and CT in similar yields and CC below the detection limit (86). Rochette studied UVA-irradiated CHO cells and showed a distribution of CPD at the TT: TC: CT: CC of 57: 18: 11:14 (27). It is also worth noting that in contrast to UVB and UVC, neither 6-4PP nor Dewar isomers formed under UVA irradiation (23).
1.5.2 Effect of Flanking sequences

The quantum yield of CPD formation is also known to be significantly dependent on the flanking sequences. It has been reported that the presence of a 5’ pyrimidine enhances CPD formation (89,90), whereas 5’ guanine suppresses formation (81,89,91-96). Additionally, compared with 5’-flanking bases, the 3’ base is less important (90,95). Several studies have used gel electrophoresis to detect CPD formation in different sequence contexts because it is capable of single nucleotide resolution over long stretches of sequence. T4-pdg is used to specifically recognize the cis-syn CPDs and cleave the DNA strand at the 5’ end of the pyrimidine dimer. The yield for CPD formation at different sites therefore can then be quantified from the band intensities after resolving the cut bands on the electrophoresis gel. Haseltine and coworkers showed that UVC-induced CPD at photoequilibrium follows the order GTTA ≈ ATTG < ATTC < CTTA < CTTC < ATTA (97). Law and coworkers reached a similar conclusion that the order of quantum yield of UVC-induced CPDs was GTTG < GTTC < GTTA < ATTG ≈ ATTC ≈ CTGG < CTTC < CTTA ≈ ATTA (98). The study by Mitchell and coworkers provided a more general ranking of CPD frequencies as follows: under UVB irradiation, GCCG ≈ GCTG ≈ GTCG < ACCG < ATCA < ACTT < ATCT < CTTG < CTCA < GCTC < TTCG < CTTC, whereas under UVC irradiation, GCCG ≈ GCTG ≈ ACCG < ATCA < GTCG ≈ CTCA < ACTT < ATCT ≈ GCTC < CTGG < TTCG < CTTC (81).

The most extensively studied flanking effect is that of guanine because guanine has been shown by many studies to suppress TT CPD formation (81,89,91,92,94-96). The flanking sequence effect has been attributed to both conformational and electronic effects. In an early study, Kundu et al. believed that the higher base paring strength between G:C base pairs would make the duplex DNA less flexible, therefore hindering the DNA from adopting the photoreactive conformation (92). Later studies showed a correlation between the oxidation potential of flanking base and CPD
formation, and suggested that charge transfer between guanine to stacked thymine was the main reason for the low frequency of CPD formation. Holman et al. observed that a 5’ guanine could also promote the reversal of thymine dimers in UVC-irradiated duplex DNA, and proposed that the process was a consequence of electron transfer from guanine to thymine (93). The oxidized base, 8-oxoGua, was later demonstrated to have the same ability to reverse CPDs (99). Law et al. observed that the not only the formation rate, but also the photoreversal rate of CPD were dependent on the flanking bases under UVC irradiation (98). To explain the suppressive effect, Cannistraro and Taylor proposed the formation of charge-transfer exciplexes that would deactivate the excited thymines (94). Pan et al. proposed the suppression resulted from a combination of ground state electron donor-acceptor interactions, excited state exciplex formation as well as ground state conformation (95). A recent computational study examined the conformational and electronic effects of the flanking sequences separately, and concluded that a charge transfer state (CT) between flanking bases and the pyrimidine base provided a decay pathway for excited thymines to escape from dimer formation (96).

Despite all the work on this subject, the frequency of CPD formation at all possible combinations of flanking sequence, XPYPyY, where Py refers to the pyrimidines T and C, is yet to be established. The effect of flanking sequence on CPD formation at TC, CT and CC sites as a function of UVA, UVB and UVC is still poorly understood, and is one of the foci of this thesis.

1.5.3 Effect of Methylation

Methylation of DNA also has an impact on the formation of CPDs. In fact, CPDs are more commonly formed at dipyrimidine sites containing methylated cytosine compared with the same unmethylated site following exposure to UVB (72,100-106). This is due to the fact the maximum absorption wavelength is 273 nm for methylated cytosine and 267 nm for unmethylated cytosine.
Therefore, methylated cytosine absorbs more UVB than unmethylated cytosine, and is likely to be more damaged when exposed to UVB. Douki and Cadet found an enhancement of dimerization in dinucleoside monophosphates containing a single 5-methylcytosine (106). Mitchell observed that CPD formed at approximately two-fold greater frequency in poly(dG) : poly(m5dC) compared to poly(dG) : poly(dC) (103). Pfeifer and coworkers discovered that the presence of 5-methylcytosine bases strikingly increased the yield of CPDs at mutational hot spots on p53 gene up to 15-fold when irradiating human cultured cells to natural light (105). Drouin and Therrien observed a similar enhancement of CPD formation under UVB but not UVC irradiation (72). In a more recent study, Rochette et al. observed that a 1.7-fold of increase in UVB-induced CPD formation due to methylation of cytosine when mapping the photoproduct formation at the X-linked PGK1 (phosphoglycerate kinase 1) and FMR1 genes (101). A recent modeling study discovered that the existence of C5-methylation reduced the confirmation motions of duplex DNA thus facilitating the transfer of $\pi\pi^*$ excitation from populated exciton states to reactive pyrimidines (107).

1.5.4 Effect of DNA structure

DNA is a conformationally flexible polymer and its structure is not limited to B-form. The structure of DNA is shown by numerous studies to have an impact on the CPD formation, mostly in terms of CPDs’ stereochemistry. In this section, the formation of CPDs in three different DNA structures (B-form, A-form and G-quadruplex) are discussed.

1.5.4.1 B-form DNA

The native form of mixed sequence DNA in solution is the B form and is the major conformation found in solution under physiological conditions. It is a right-handed double-helical structure with
10–10.5 base pairs per turn. It contains a minor groove and wider major groove where the majority of DNA binding proteins bind (Figure 1.6A). The stability of the double-helical structure is attributed to base pairing between complementary strands via hydrogen bonding and base stacking between adjacent bases within the same strand. The close distance and π-stacking geometry between the bases (around 3.4Å for B-DNA) facilitates photoreactions and energy and charge transfer (108). As we discussed in the previous section, the efficiency of CPD formation within B-form DNA depends on its sequence-context and the location of dipyrimidine site within the DNA. In B-form DNA, the adjacent bases adopt a parallel orientation linked by the sugar-phosphate backbone, therefore, disfavors anti CPDs and restricts the formation of CPDs to be in cis-syn confirmation (62–64,109). No trans-syn CPD has been reported in double-stranded DNA. Trans-syn CPDs are generally regarded as a sign of flexibility or single strand character (22). They have been observed to form in single-stranded DNA or in denatured DNA resulting from lowering the ionic strength or increasing the temperature (110-112).

The local structure of DNA will change once CPD forms in B-form DNA. Crystal structures, in agreement with earlier theoretical and experimental studies, show that CPD results in 30° bending toward the major groove, helical unwinding of 9°, and widening of both minor and major grooves (113). The change in structure could enable the repair proteins to repair DNA as well as interfere with normal protein-DNA interactions essential to replication and transcription. It is interesting to note that compared with CPDs, 6-4PPs bend DNA even more (44°), which may explain why 6-4PPs are more mutagenic and removed more efficiently than CPDs in cells (114,115).
1.5.4.2  \textit{A-form DNA}

A-form DNA is similar to B-form, as it is a right-handed double-helical structure, but it has a more compact conformation than the B-form since its bases are tilted with disordered stacking rather than perpendicular to the helix-axis with parallel stacking (Figure 1.6B). A-form is also the conformation that RNA adopts (118). DNA tends to adopt A-form structures in dry states or in the presence of large amounts of ethanol and other solvents (22). The photochemistry of A-form DNA has been shown to be significantly different from that of B-form DNA. Douki and colleagues reported that irradiation of A-form DNA duplex with UVC light results in significant yield of interstrand DNA damage, which consists of predominant spore photoprod (SP) as well as CPDs and 6-4PPs. The interstrand CPDs occurred with all four possible stereochemistries, with a large fraction of \textit{anti} isomers that would result from bases with antiparallel orientation on opposite strands (44). Another study replaced UU in an A-form RNA hairpin stem with TT dinucleotides and revealed that dipyrimidine photoprod are almost completely abolished in this case, suggesting that the A-form conformation prevents photodimerization. They attributed the UV resistance to the unfavorable orientation of adjacent pyrimidine bases and the stiffness of compact A-form structure that makes structural rearrangement energetically costly (92). This has been suggested to explain the UV resistance of bacterial spores that allows such bacterial to survive extreme conditions. Bacterial spores are characterized by a low degree of hydration and the DNA is complexed with \textit{a/\beta} type small, acid soluble proteins (SASPs) in the spore core. It has been shown that exposure of spores to UVC irradiation will exclusively give rise to spore photoproducts (SPs). A crystal structure of spore DNA complexed to the SASP protein revealed that it adopts an “A-B-DNA” conformation (119).
To summarize, in B-form DNA, parallel base stacking occurs mainly on the same strand, allowing *cis-syn* CPDs to be formed, whereas in A-form DNA, the compact structure promotes π stacking on the opposite strand, giving rise to *anti* CPDs. Recently, there has been numerous intensive photophysical and computational studies on the excited states involved in CPD formation in duplex DNA. Compared with single-stranded DNA, the base pairing in duplex DNA complicates the calculations and questions remain as to whether excitation and charge transfer exist between base pairs.

1.5.4.3 **G-quadruplex**

Guanine-rich DNA strands, such as those encountered at telomeres, have been shown to form secondary structures called G-quadruplexes (120,121). The basic building block of G-quadruplexes is called a G-quartet, where four guanines are arranged in a circular, planar arrangement via Hoogsteen base pairing. The vertical stacking of G-quartets leads to the formation of G-quadruplex structure (Figure 1.7), with the coordination of monovalent or divalent cations to the carbonyl oxygens of the guanines that is essential to the stability of G-quadruplex structures.

The specific folding of the G-quadruplex structure is strongly dependent on the sequence context and the stabilizing ions (122). Human telomeric DNA consists of TTAGGG repeats. It has been found that d(AGGG(TTAGGG)_3) (Tel22) adopts a basket structure in Na\(^+\) (123) but favors a mixture of hybrid 1 and hybrid 2 structures in K\(^+\) (Figure 1.8) (124-126). An early study in our group provided unambiguous evidence that non-adjacent *anti* TT CPDs were specifically generated upon UVB irradiation of d[AGGG(TTAAGGG)_3] in the presence of K\(^+\) (127). Using Nuclease P1 digestion coupled with HPLC analysis, we were able to separate the adjacent TT CPD from nonadjacent TT CPD because the adjacent thymines are digested into trimers whereas the non-adjacent ones are digested into tetramers. Mass spectrometry and comparison of hydrolysis
products to authentics permitted the structure and stereochemistry of the product formed in the highest yield to be determined as a trans-anti TT CPD from the dimerization of thymines between loop1 and loop3. This was puzzling because the basket form, which contains two proximate loops, was expected to lead to photocrosslinking rather than the hybrid structure. However, the irradiation of the Tel22 sequence in Na\(^+\) solution, which favors a basket structure, did not lead to significant amounts of anti CPDs, whereas irradiation in K\(^+\) solution, which favors the hybrid-1 and hybrid-2 structures did lead to the anti CPDs. To account for this unanticipated observation, it was proposed that the photoreactive intermediate was the two G-tetrad basket-like form 3 conformation (NF3) that was shown to be formed in a related sequence in K\(^+\) solution (128,129). In a subsequent study, a series of site-specific substitutions and insertions of d(A\(_2\)G\(_3\)(TTAGGG)\(_3\)A\(_2\)) (Tel26) were made in order to stabilize the form 3 structure. The tel26 sequence was chosen because it almost exclusively gives rise to trans-anti T(A)=T(A) CPD. In spite of these mutations, however, the anti CPD product was still inhibited in the presence of Na\(^+\), and the formation of the trans,anti product rate was slower in K\(^+\) compared to that of Tel26. These results suggested that NF3 is not the photoreactive conformation but some other conformation may be involved (130).

Although our study was the first to characterize DNA photoproducst in human telomeric sequences, evidence telomeric DNA can be photocrosslinked was observed by others (131,132). It is also possible that anti CPD may form in other high order DNA structures. In an early study, our group discovered the formation of nonadjacent cis-anti TT CPD between T2 and T7 d(GTATCATGAGGTGC) when it was irradiated in aqueous solution at pH 5 with UVB light. It was suggested that there must be a higher-order folding structure that brings the two thymines close together to facilitate the photocrosslinking (133).
1.5.5 Effect of chromatin structure

The structure of chromatin also affects the distribution of CPDs at the DNA sequence level (135). The effect is of biological significance since the formation of photoproducts within genomic DNA takes place in the context of chromatin. When a cell is not in cell division, the DNA is arranged in the form of "beads on a string". The beads represent the nucleosomes, and the string separating the beads is linker DNA. Nucleosome is composed of 147bp DNA wrapping almost two times around a histone octamer, where DNA is bent and not in an ideal B-form confirmation (136). The presence of nucleosomes influences the flexibility of DNA and hence the formation of CPDs. In general, CPDs form almost randomly between linker DNA and nucleosome core DNA (137), but CPD formation within nucleosome core DNA shows a 10.3 base periodicity (138,139). Interestingly, it was revealed that the yield of CPD in nucleosome core DNA strongly depends on the orientation of DNA relative to the histone surface, with CPDs forming in 10-fold higher yield when the dipyrimidine backbone is furthest away from histone surface compared with a dipyrimidine in contact with the histones (138). On the other hand, the formation of photoproducts promotes the unwrapping of nucleosomes, presumably facilitating DNA repair process (140).

1.5.6 Effect of protein binding

DNA-protein interactions lead to bending and unwinding of DNA, thus giving rise to different photoproduct distributions compared with free DNA in solution. Studies compared the formation of CPDs in isolated DNA and cells upon exposure to UV light. The results showed a difference in the distribution of CPDs in promoter regions but little difference in exons (141-143). The most significant variations were observed at sites where a protein binds to DNA (inhibiting CPD formation) or at the boundaries of a protein binding site to DNA (enhancing CPDs formation).
short, the presence of DNA-bound proteins creates zones of DNA distortion that promote or inhibit the formation of CPDs.

1.6 Mechanism of CPD formation

1.6.1 Singlet or Triplet?

Even though CPDs have been detected and characterized for more than 60 years, a fundamental question remains the underlying reaction mechanism of CPD generation. Early experimental studies regarding the quantum yield of CPD support both singlet and triplet pathways (Figure 1.9). Because thymine has the lowest triplet energy out of all four bases, it has been proposed that the triplet states may be localized on thymines, and the energy transferring from higher triplet states to lower triplet states is favored. However, recent ultrafast time-resolved techniques and accurate computational calculations strongly support the singlet pathway. Femtosecond time-resolved infrared experiments showed that CPDs are formed within less than 1 ps upon exposure of single-stranded oligodeoxynucleotide (dT)$_{18}$ to UVB light (144). The reaction is thought to proceed via a singlet channel upon direct excitation with minor contribution from triplet states (145). Since the CPD formation process is significantly faster than the backbone motions that would bring the stacked pyrimidine bases into photoreactive conformation, the ground state conformation of stacked thymine bases at the instant of photoexcitation is assumed to be the key factor in the dimerization reaction. The low quantum yield of CPD is, therefore, attributed to the low amounts of the photoreactive conformations. On the other hand, it is not fully understood how CPDs are formed under UVA irradiation. DNA weakly absorbs UVA photons; therefore it has been proposed that UVA-induced CPD could arise either through direct excitation or by photosensitization through the excitation of the endogenous or exogenous chromophores (25-29,31).
More and more studies support the involvement of collective excited states (excited states delocalized over at least two bases) in the formation of pyrimidine dimers, thanks to the advances in spectroscopy (146-148) and computational methods (149-152). The delocalization is a result of electronic interaction among the bases because of their close proximity and strongly depends on the relative position of bases and sequence context. Frenkel excitons (delocalized ππ* states) and charge-transfer (CT) states are the two limiting cases of collective excited states, for which Frenkel excitons are believed to be involved in the formation of CPDs (153). In addition, studies demonstrated that the transition from Frenkel excitons to CT states provided a decay pathway for the population to escape from CPD formation (95).

Another striking finding deals with so-called “dark CPDs”, which were discovered by Premi et al. in 2015 (32). They showed that CPDs could continue to be produced 2 to 3 hours after UV excitation was terminated via a dark pathway. In addition, it was claimed that the “dark CPDs” had a higher fraction of C-containing CPDs than CPDs induced by UVA alone. The dark reaction could not be explained by direct excitation or photosensitization mechanism which happens in less than a second upon absorption of UV photons. They, therefore, proposed that a chemisensitization mechanism was involved, in which peroxynitrite, formed from the reaction of UV-induced superoxide and nitric oxide, reacts to form a high energy dioxetane from melanin. Then, an excited state carbonyl is generated from the decomposition of dioxetane, which transfers its triplet energy to DNA to form CPDs (17, 32, 154, 155).

1.6.2 Photosensitization of CPD formation

CPDs can also form by photosensitization via the so-called triplet-triplet energy transfer pathway (TTET). In a TTET process, the photosensitizer is excited by the absorption of UV photons and converted into its triplet state via intersystem crossing (ISC). If the energy is high enough and it is
close to a dipyrimidine, it will transfer its triplet energy to the DNA to generate a CPD (Figure 1.9) (23,31,156-158). The efficiency of the TTET process is strongly related to the triplet energy of photosensitizer and the DNA bases, the coupling interaction between them, the DNA, photosensitizer complexation, photosensitizer environment, as well as the distance between the photosensitizer and the bases (159-165).

TTET has been found in both isolated DNA (166-172) and cells (173,174). Early studies used ketones to photosensitize DNA because of their high triplet energy and high quantum yield of intersystem crossing (nearly 1) (175-178). Other popular exogenous photosensitizers include pyridopsoralens (PyPs), non-steroidal anti-inflammatory drugs (NSAIDs), fluoroquinolones (FQs), amino acids and derivatives or cosmetic agents (157). FQs are broad spectrum antibacterial drugs and have been the focus of many studies (156,173,174,179,180). Using FQs with known excited triplet state energies, Bosca et al. estimated the triplet energy of thymine in duplex DNA to be approximately 267 kJ/mol, which is 40 kJ/mol lower than that of free thymine (Thy) or thymidine (Thd) (159). Among all kinds of photosensitizers, one of FQs, Norfloxacin (NFX) was reported to be the most efficient photosensitizer for TTET (160). NFX has a triplet energy of approximately 269 kJ/mol (181) and has been previously used as a reference photosensitizer to establish the $E_T$ threshold needed to produce TT CPDs in double-stranded DNA (159).

According to TTET studies of DNA in the presence of various photosensitizers (160,172,182-185), TT CPD remains the most predominant photoproduct, followed by TC and CT with at least one order of magnitude lower yield, and CC CPD rarely (157). The ratio of CPDs formed between TT, TC, CT, and CC is different for different photosensitizers, and it has been shown that flanking bases influence the yield of CPD generated via TTET (185). These studies suggest a need to further investigate the mechanism involved in TTET-induced CPD. It may be,
that a more complex mechanism than the simple scheme of the transfer of triplet energy to the single thymine base is involved (160),

1.7 Conclusions and Subject of this Thesis

The *cis-syn* CPD is the major photoprodut resulting from UV irradiation that results in C to T mutations found in human skin. The *anti* CPD, as an unusual type of DNA photoprodut, that was recently discovered by our group to form upon UV irradiation of human telomeric DNA may also be of significance to the toxicity and mutagenicity of UV light. The general objective of this thesis is to investigate the formation of *anti* and *cis-syn* CPDs, and to elucidate the effect of conformation and sequence context on CPD formation.

In Chapter 2, we tried to elucidate which DNA conformations are involved in *trans-anti* T(A)=TA CPD photoprodut formation in human telomeric DNA. The photochemistry of Tel26 was studied in the presence of various metal ions, some of which like Li⁺ are known to disfavor G-quadruplex formation, along with base substitutions designed to interfere with G-quartet formation. The results suggest that a reverse Hoogsteen base-paired hairpin conformation with a T•T mismatch as the photoreactive conformation leads to the *trans-anti* T(A)=TA CPD and that this conformation may be in equilibrium with G-quadruplex structures.

In Chapter 3, we investigated the effect of sequence context on CPD formation and determined the photoprodut yield at all possible NPyPyN (N indicates A, T, C, or G, and PyPy represents TT, TC, CT or CC) sites in one short sequence (129-mer) as a function of wavelength and photosensitizers. A Python script was written to generate all possible sequence libraries, and one of shortest sequences was then synthesized, cloned, and PCR amplified for the photochemical studies. We used T4 endonuclease gel electrophoresis assay to determine the relative yields of photoprodut formation and showed a dramatic effect of sequence on CPD yield. We also studied
the different sequence context effect on the TTET-induced CPDs formation in the presence of either acetone or norfloxacin. The quantification data addressed here will provide some general rules for estimating the CPD distribution in duplex DNA. In addition, these results suggest that the photosensitizers may result in unique mutation spectra that can be used to identify endogenous photosensitizers such as those implicated in chemi-excitation pathways in melanoma induction.

Chapter 4 summarizes the overall conclusions of the thesis. We will also discuss several considerations for future studies based on the conclusions.
1.8 **References**


23


thymine dipyrimidines and correlate with the mutation spectrum in rodent cells. *Nucleic acids research*, 31, 2786-2794.


1.9 Figures

Figure 1.1: DNA structure. The DNA backbones consist of alternating phosphate and sugar groups with bases attached via glycosidic bonds. Adenine pairs with thymine via two hydrogen bonds and guanine pairs with cytosine via three hydrogen bonds. The two strands are in antiparallel orientation and the base pairing as well as base stacking stabilize the structure.
Figure 1.2: The UV absorption spectra of DNA. DNA has a maximum absorption at around 260 nm. As the diagram shown, UVC light and near UVB light is blocked by ozone layer. The longer-wave UVB light can penetrate the epidermis leading to predominant formation of dipyrimidine photoproducts. UVA light makes up 95% of terrestrial solar UV and can penetrate into the dermis.
Figure 1.3: Structures of dipyrimidine photoproducts. CPDs, 6-4PPs, Dewar valence isomers and SP’s are the four main types of dipyrimidine photoproducts. The structures are shown for a TT dimer. However, pyrimidine dimers also form at other dipyrimidine sites.
Figure 1.4: Stereoisomers of CPD. The isomers are shown for TT CPDs. The cis/trans stereochemistry reflects the relative position of the two pyrimidines with respect to the cyclobutane ring. A syn/anti stereochemistry corresponds to the parallel or antiparallel orientation of the bases.
Figure 1.5: Deamination mechanism.

Table 1.1: Distribution of CPDs: Summary of previously published studies.
Figure 1.6: B-form and A-form of DNAs. (A) PDB 1TTD B-form DNA duplex containing a cis-syn TT CPD (116) (B) PDB 440D: A-DNA crystal structures of d(AGGGGCCCT) (117)
Figure 1.7: **G-quadruplex structures.** Four guanines form the G-quartet via hoogstee bonding. The stacking of G-quartets combined with coordination from the cations stabilized the G-quadruplex.

Figure 1.8: **Structures of G-quadruplex in Na⁺ and K⁺.** (A) Tel22 in Na⁺ (B) Human telomeric sequences in K⁺. (A) PDB 143D: d(AGGG(TTAGGG)_3), Basket (123) (B) (1) PDB PDB 2JSM: d(TAGGG(TTAGGG)_3) (3+1) Hybrid 1 (134); (2) PDB: 2JSL: d(TAGGG(TTAGGG)_3TT) (3+1) Hybrid 2 (134); (3) PDB 2KF8: d((GGGTTA)_3GGGT) two-tetrad basket Form 3 (128).
Figure 1.9: Two possible mechanism leading to CPD formation: Singlet Pathway and TTET
Chapter 2 Evidence for Reverse Hoogsteen Hairpin Intermediates in the Photocrosslinking of Human Telomeric DNA Sequences

2.1 Abstract

UVB irradiation of human telomeric (GGGTTA)$_3$GGG sequences in potassium ion solution crosslinks the first and third TTA segments through \textit{anti} cyclobutane pyrimidine dimer (CPD) formation. The photocrosslinking reaction was first proposed to occur through a form 3 G-quadruplex in which the lateral 4-nucleotide GTTA loop can interact with an adjacent TTA loop. Curiously, the reaction does not occur in the presence of sodium ions, which was explained by the formation of a basket structure which only has 3-nucleotide TTA loops that cannot interact. Sequences known or expected to favor the two-tetrad basket did not show enhanced photocrosslinking, suggesting that some other structure was the reactive intermediate. Herein, we report that \textit{anti} CPDs form in human telomeric DNA sequences complexed with lithium ions that are known to disfavor G-quadruplex formation, as well as with potassium ion when the bases are modified to interfere with G-quartet formation. We also show that \textit{anti}-CPDs form in sequences containing A’s in place of G’s that cannot form Hoogsteen hairpins, but can form reverse Hoogsteen hairpins. These results suggest that reverse Hoogsteen hairpins may play a hitherto

---


unrecognized role in the biology and photoreactivity of DNA in telomeres, and possibly in other purine-rich sequences found in regulatory regions.

2.2 Introduction

Anti cyclobutane pyrimidine dimers (CPDs) (Figure 2.1A) were first discovered in the ultraviolet (UV) light irradiation products of thymidine in frozen aqueous solutions (1). Later, these products were detected in the irradiation products of dehydrated or alcoholic calf thymus DNA (2). A cis,anti cyclobutane thymine dimer was then accidentally discovered to form in high yield in aqueous solution during the irradiation of a 14-mer oligodeoxynucleotide at acidic pH but not neutral pH (3). The photocrosslinked T’s in the 14-mer were separated by four nucleotides, suggesting the involvement of an as yet unknown folded structure. This result prompted a study of the photochemistry of human G quadruplex forming telomeric sequences (Figure 2.1B) which can fold in aqueous solution at biologically relevant pH’s and cations to form basket structures. Basket structures formed from the human telomere sequence have adjacent lateral loops containing T’s that were expected to be able overlap and form anti CPD formation (4). Indeed, anti thymine dimers were discovered to be produced in the human G-quadruplex forming telomeric sequence Tel22 (Table 2.1) under neutral pH, although not in an entirely expected manner. What was puzzling was that irradiation of the Tel22 sequence in Na+ solution, which favors a basket structure (5), did not lead to significant amounts of anti CPDs, whereas irradiation in K+ solution, which favors the hybrid-1 (6) and hybrid-2 (7) structures (Figure 2.1B), did lead to the anti CPDs, in spite of their lack of adjacent lateral loops. Formation of anti CPDs has implications for the biological effects of sunlight and for the structure of telomeric DNA that is therefore important to understand.
We proposed that the hybrid structures favored in K\textsuperscript{+} solution were also in equilibrium with a more recently discovered two tetrad basket-like form 3 structure that the sequence NF3 adopts (8). The form 3 structure has one lateral loop (GTAA) that is one nucleotide longer than the corresponding TTA loop in the three tetrad basket structure adopted in Na\textsuperscript{+} solution, and we expect it to overlap better with the adjacent TTA loop to enable formation of the \textit{anti} CPD (4). This hypothesis was tested recently (9) by comparing the photochemistry of the Tel26 sequence (Table 2.1), which exists primarily in a non-photoreactive hybrid-1 conformation (10), and produces almost exclusively the \textit{trans,anti} T(A)=T(A) CPD, with that of the NF3 sequence, which exists primarily in the two G-tetrad form 3 conformation (8). Although both sequences produced the \textit{trans,anti} T(A)=T(A) CPD in about the same yield upon UVB irradiation, \textit{anti} CPD formation was faster and more selective in Tel26 than in NF3 suggesting that the form 3 structure was not the major photoreactive conformation\textsuperscript{3}. In addition, irradiation of a number of mutants specifically designed to stabilize a four-nucleotide loop-containing basket structure in both K\textsuperscript{+} and Na\textsuperscript{+} solution, while producing the \textit{trans,anti} T(A)=T(A) CPD in K\textsuperscript{+} solution, failed to produce any \textit{anti} CPDs in Na\textsuperscript{+} solution. The conclusion of our study was that some other conformation must also be involved, such as a chair or hairpin structure, in spite of QM/MM modeling calculations suggesting that the \textit{trans,anti} CPD could arise from the form 3 structure (11).

To gain further insight into the photoreactive conformations involved in \textit{anti} CPD formation, we now report the photochemistry of G-quadruplex forming sequences in the presence of various metal ions known to either stabilize or destabilize G-quadruplexes. We also report on the photochemistry of truncated sequences and sequences in which inosine and adenosine are used in place of guanosine to disrupt G-quartet formation in Na\textsuperscript{+} and K\textsuperscript{+} solution. From these studies

\textsuperscript{3} The detailed data was shown in Section 2.4.4 and 2.4.5
we discovered that photocrosslinking of T’s can occur in structures other than quadruplexes, and
provide evidence that implicates antiparallel reverse-Hoogsteen base-paired hairpins as the
photoreactive conformation leading to the trans,anti \( \text{T(A)} = \text{T(A)} \) CPD. The results suggest that
reverse Hoogsteen hairpin structures co-exist with G-quadruplexes, and possibly play a hitherto
unrecognized role in DNA and RNA structure and function.

2.3 Materials and Methods

2.3.1 Materials

Oligodeoxynucleotides were purchased from Integrated DNA Technologies, Inc. (IDT)
(Coralville, Iowa). Nuclease P1 (NP1) from \textit{Penicillium citrinum}, was from Sigma (St. Louis,
MO). Ammonium citrate and 3-hydroxypicolinic acid (3-HPA) for use as MALDI matrices were
purchased from Fluka (Milwaukee, WI). HPLC solvents were from Fisher (Fair Lawn, NJ).
HPLC separation and analysis were carried out on System Gold HPLC system with a binary
gradient Model 125 pump and a Model 168 diode array detector (Beckman Coulter, Inc.,
Fullerton, CA). An X-Bridge column (C18, 4.6 × 75 mm, 2.5 \( \mu \)m, 135 Å) from Waters
Corporation (Milford, MA) was used for reverse-phase HPLC. UVB (280-320 nm) irradiation
was carried out with two Spectroline XX-15B UV 15W tubes (312 nm) with peak UV intensity
of 1150 \( \mu \)W/cm\(^2\) at 25 cm filtered through a Longlife filter glass from Spectronics Corporation
(Westbury, NY). CD experiments were carried out on a J-810 spectropolarimeter (Jasco). UV
melting curves were obtained on a Cary 100 Bio UV-VIS Spectrometer (Varian).
2.3.2 Preparation and characterization of G-quadruplexes.

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

2.3.3 Melting temperature measurements.

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

Samples were first denatured, then annealed, and then denatured again. The T\textsubscript{m} was calculated from the intersection of the annealing curve and the mean of linear fits to the upper and lower limits of the annealing curve.

2.3.4 UVB Irradiation and product characterization.

UVB irradiation was carried out immediately at 4 °C after sample preparation or after storage at 4° C for overnight to days with similar results. G-quadruplex samples (50 µM) were irradiated in a polyethylene microcentrifuge tube and irradiated on a bed of ice for 2.5 h at approximately a distance of 1 cm from the UVB lamp. Irradiated samples (100 µL of 50 µM DNA) were digested with 1 µL of 1 U/µL aqueous NP1 from Penicillium citrinum (Sigma) and 1 µL of 10 mM ZnCl\textsubscript{2} at 37 °C for >36 h. Reverse-phase HPLC was carried out with an X-Bridge column (C18, 4.6 x 75 mm, 2.5 µm, 135 Å; Waters Corporation) NP1 digestion products were eluted with 1 mL/min of 100% solvent A (50 mM triethylammonium acetate, pH 7.5) for 3 min followed by a linear gradient of 0-20% B (50% acetonitrile in 50 mM triethylammonium acetate, pH 7.5) in solvent A.
for 3-53 min and detected at 260 nm. The yield of each photoproduct-containing digestion product was calculated from the integral of the corresponding HPLC peak at 260 nm by using molar extinction coefficients for the appended undamaged nucleotide or nucleotides at 260 nm relative to that for one of the undamaged nucleotides. HPLC fractions corresponding to photoproduct-containing NP1 digestion fragments were dried by *in vacuo*. Milli-Q water was added to each sample to yield a final ODN concentration of approximately 50 µM. One microliter of sample was mixed with 1 mL of 3-HPA with 10% ammonium citrate and spotted on a stainless-steel MALDI plate. Spectra were taken in reflectron positive ion mode and were the average of 20 accumulations on MALDI-TOF MS.

### 2.3.5 Denaturing PAGE analysis of photoproducts.

The 5’ termini of the ODNs were labeled by a forward reaction with 1× reaction buffer (500 mM Tris•HCl, pH 7.6, 100 mM MgCl2, 50 mM DTT, 1 mM spermidine, 1 mM EDTA; Fermentas Life Science) with 10 U T4 polynucleotide kinase and 20 pmol [α-32P]-ATP at 37 °C for 1 h. The reaction was quenched by adding to an equal volume of 2× loading buffer (98% formamide, 10 mM EDTA), boiling for 2 min, and cooled to room temperature. The radiolabeled samples were then analyzed on a 15% denaturing polyacrylamide gel and scanned with a Personal Molecular Imager (Bio-Rad Laboratories) and analyzed with Quantity One software.

### 2.3.6 Native PAGE analysis of DNA structures.

Native PAGE was performed as generally described (12). Briefly, 5’-32P-end labeled oligonucleotides were purified by 15% Denaturing PAGE. Then, each sample was prepared by mixing 5 µL 20 µM radiolabeled DNA, 1 µL 100 mM Tris (pH 7.5), 1 µL respective 1 M monovalent ion (1M LiCl, NaCl or KCl), then adding dd H2O to reach 10 µL volume. The samples
were boiled for 5 min and cooled immediately on ice for 2 h. An equal volume of 2× loading dye containing 46% glycerol and 5% xylene cyanol were added to the samples and 10 µL of each sample was loaded on 15% Native PAGE containing 100 mM the respective monovalent cations. Metal ion-free 5× TBE buffer was prepared with 0.0125 M EDTA (ethylenediaminetetraacetic acid, acid form), 0.455 M Tris and 0.445 M boric acid in dd H2O. The 15% native PAGE gels were prepared using 29:1 acrylamide:bisacrylamide, in 1× TBE buffer containing the indicated concentration of cation. The gels were pre-equilibrated with ice-cold 1× TBE running buffer containing the monovalent cation and run at 400 V for at about 30 min at 4°C in a cold room prior to loading the samples. Then the gel was run at 400 V at 4°C until xylene cyanol moved 14~15 cm on the gel.

2.4 Results and Discussion

2.4.1 Methodology

The method used for assaying CPD formation in DNA makes use of an enzyme-coupled HPLC assay in which nuclease P1 degrades adjacent photoproducts to pN=N(N) trinucleotides, (13) and non-adjacent photoproducts to pT(N)=pT(N) tetranucleotides, hereafter referred to as T(N)=T(N). (3) (Figure 2.1A). The identity of the nucleotides N can be deduced from MALDI MS, and the stereochemistry of the CPD from acid hydrolysis to release the base portion and comparison to authentics by HPLC. The TTA loops that are crosslinked can be deduced by replacement of specific T’s with U’s, changing the retention times of CPD-containing tetranucleotides.

4 Section 2.4.2 and 2.4.3 data was done by Dr. Jill E. Smith
2.4.2 Photoprocess formation in Tel26 as a function of metal ion

Tel26 was irradiated with UVB light in the presence of the monovalent alkalai metal ions, Li\(^+\), Na\(^+\), and K\(^+\). In Li\(^+\) and K\(^+\) solutions, the *trans,anti* T(A)=T(A) CPD was observed to be the major photoprocess, whereas in the presence of Na\(^+\), *anti* CPD formation was insignificant, (Figure 2.2). Most surprisingly, *trans,anti* T(A)=T(A) CPD formation was higher with Li\(^+\) (23%) than with K\(^+\) (15%). To determine whether or not intermolecular base-paired structures might be intermediates in the formation of any of the products, the irradiated samples of 5'-\(^{32}\)P-endlabeled Tel26 were electrophoresed (Figure 2.13). In all cases only a band corresponding to a monomeric species was detected, suggesting that the *anti* CPDs were arising from an intramolecularly folded structure.

2.4.3 Effect of metal ions on the conformation and stability of Tel26.

The formation of the *trans,anti* CPD in equal or higher yield in Li\(^+\) solution than in K\(^+\) solution was quite surprising as this ion is well known to inhibit G-quadruplex formation (14-16). To verify that G-quadruplexes were not forming in the presence of the Li\(^+\) ion and to establish the types of structures that might be present in the presence of the other ions, we obtained the CD spectra and thermal denaturation curves at 295 nm, a wavelength known to be characteristic for antiparallel G-quadruplexes (17-19) (Figure 2.3). The structurally well characterized Tel26 showed a large positive peaks at around 265 and 290 nm in the CD spectrum taken in K\(^+\) solution, characteristic of a hybrid quadruplex structure containing parallel and antiparallel strands, respectively. In Na\(^+\) solution, there was a single large positive peak at 300 nm characteristic of an antiparallel quadruplex. In contrast, the CD spectrum in Li\(^+\) had a weak positive peak at 275 nm, and the absence of the 265 or 295 nm peaks characteristic of parallel or antiparallel G-quadruplexes. Whereas the T\(_m\) values for Tel26 in Na\(^+\) and K\(^+\) were about 52 °C and 60 °C respectively when
monitored at 295 nm, no melting behavior was observed in Li\(^+\). These results are consistent with the results of previous studies showing decreasing stability for intramolecular quadruplexes of d(TTAGGG)\(_4\) in the order K\(^+\) > Na\(^+\) >> Li\(^+\) (20) and intermolecular quadruplexes of d(CGCG\(_3\)GCG), in the order K\(^+\) > Na\(^+\) > Li\(^+\) (14).

### 2.4.4 Kinetics of CPD formation in Tel26 and NF3

To gain more insight into the mechanism of \textit{anti} CPD photoproduct formation, the rate of formation in Tel26 and NF3 was investigated by NP1 digestion as a function of UVB irradiation time (Figure 2.4). For both sequences, \textit{trans,anti} T(A)=T(A) photoproduct formation appeared to fit a simple first order process that was projected to approach a maximum of 35%. Surprisingly, the initial rate of formation of the \textit{trans,anti} T(A)=T(A) CPD was about two times faster for Tel26 than for NF3 (0.0029 versus 0.0015 min\(^{-1}\)), even though NF3 was proposed to adopt primarily the photoreactive form 3 conformation, whereas Tel26 was not. In comparison, the initial rate of \textit{cis,syn} T=TA CPD formation was greater than \textit{trans,anti} CPD formation for both Tel26 and NF3 (0.0047 min\(^{-1}\)), but was projected to reach a lower maximum yield for NF3 than for Tel26 (19 versus 29\%). It is worth noting that the maximum theoretical yield of T=TA per ODN is 3 compared with 1 for T(A)=T(A), as there are three independent T=TA photoproduct-forming sites per sequence, whereas only one T(A)=T(A) photoproduct can form per sequence.

One possible explanation for the limiting yield of \textit{cis,syn} and \textit{trans,anti} CPD formation is that it represents a steady state value resulting from competitive photoproduct reversal. To test this hypothesis, a sample of \textit{trans,anti} T(A)=T(A) CPD-containing Tel26 was desalted and irradiated in 100 mM Na\(^+\) solution to prevent further photoproduct formation, but presumably allow for photoreversal. On irradiation for 30 min in Na\(^+\) solution, the \textit{trans,anti} CPD appeared to increase in yield from 10 to 15\%, but on further irradiation of an additional 30 and 60 min, no change in
yield was observed, suggesting that photoreversal was not taking place, or at least not in Na\(^+\) solution (data not shown). It was also possible that oxidized Gs that might form during UVB irradiation could photocatalyze photoreversal of the cyclobutane thymine dimer, as has been previously observed 8-oxo-7,8-dihydroguanine (47). However, there was no evidence for the presence of 8-oxoGMP in the NP1 digestion samples by HPLC based on its distinct absorption maximum at 293 nm (48).

### 2.4.5 Inhibitory effect of Na\(^+\) and higher temperature on anti CPD formation

The original hypothesis for why the anti photoproducts formed in low yield in Tel26 in Na\(^+\) solution was that Na\(^+\) was stabilizing the basket conformation, which only had 3-nt loops, and prevented formation of the form 3 structure with a two-tetrad basket and a 4-nt loop 1. Therefore, it was expected that adding an extra nucleotide to loop 1 of Tel26 or a third G to NF3 would result in a basket structure with a 4-nt loop 1, which would enhance anti CPD formation in Na\(^+\): Contrary to expectation, adding an extra nucleotide to loop 1 did not increase anti CPD formation compared with the parent Tel26 or NF3 in Na\(^+\) solution, although it did seem to alter the distribution of the products.

To determine the relative ability of sodium to inhibit anti CPD formation, a competitive reaction was carried out between Na\(^+\) and K\(^+\) ions. To detect small changes in photoproduct yield, Tel26 was irradiated at a higher intensity in the presence of an increasing mole fraction of K\(^+\) in a 100 mM Na\(^+\) solution. As can be seen from Figure 2.5A, even a small concentration of K\(^+\) was sufficient to produce the same yield of trans,anti CPD as observed in 100 mM potassium solution, with a half maximal yield being observed at a concentration of 3.6 mM K\(^+\): The yield of cis,syn CPD decreased slightly on going from 100 mM Na\(^+\) to 100 mM K\(^+\) probably in part due to
competition with \textit{trans,anti} CPD formation, which would remove two sites for \textit{cis,syn} CPD formation.

To determine the effect of temperature on the yield of the \textit{trans,anti} CPD formation under more biologically relevant buffer conditions, Tel26 was irradiated from 4 °C to 37 °C in 100 mM KCl, 1 mM MgCl$_2$ in pH 7 buffer (Figure 2.5B). Although temperature did not have a significant effect on \textit{cis,syn} CPD formation, which remained relatively constant at 11%, it had a much greater effect on \textit{trans,anti} CPD formation, reducing the yield from 10% at 4 °C to 1% at 37 °C.

\textbf{2.4.6 Temperature dependence and kinetics of \textit{anti} CPD formation in Tel26 in Li$^+$ and K$^+$ solution.}

We found that the yield of the \textit{trans,anti} T(A)=T(A) CPD formed from the irradiation of Tel26 in K$^+$ solution dropped from 12% to about 1% on raising the temperature from 0° to 37 °C (9). The decrease in yield with increasing temperature had been explained by either a decrease in the population of the reactive intermediate, or to a decrease in lifetime of the photoreactive intermediate. The interpretation of this temperature dependence is complicated by the presence of the non-photoreactive hybrid-1 as the major conformation in K$^+$ solution (10, 7). Because quadruplex conformations are not stable in Li$^+$ solution, it was expected that the temperature dependence of \textit{trans,anti} CPD formation in Tel26 in Li$^+$ solution might be more directly related to the stability of the conformation present under these conditions. A likely candidate is a hairpin structure composed of reverse Hoogsteen G•G base pairs (Figure 2.6) that has been previously been shown by NMR to form in the \textit{Oxytricha} telomere sequence, G$_4$T$_4$G$_4$, in Li$^+$ solution (21), though more recent studies have suggested the formation of symmetric G•G base pairs (Figure 2.6, Figure 2.14) (22).
Analysis of the temperature dependence of the UV absorption of Tel26 at 260 nm in Li\textsuperscript{+} solution showed that it had a T\textsubscript{m} of about 40 °C (Figure 2.7A). The decrease in the fraction of the helical form with an increase in temperature roughly paralleled the decrease in the relative yield of \textit{trans,anti} T(A)=T(A) CPD from 0 °C to 40 °C (Figure 2.7A). The temperature dependence of \textit{trans,anti} CPD formation in Li\textsuperscript{+} was similar to what we previously observed in K\textsuperscript{+} solution (9), supporting the idea that the reverse Hoogsteen hairpin might be the photoreactive conformation. The kinetics of \textit{trans,anti} CPD formation in Li\textsuperscript{+} and K\textsuperscript{+} also supported this notion. The initial rate of \textit{trans,anti} T(A)=T(A) CPD formation was 2.4 times greater in Li\textsuperscript{+} than in K\textsuperscript{+} solution at 0 °C (Figure 2.7B) as would be expected if the reverse Hoogsteen hairpin conformation were the photoreactive conformation, because there would be no competing non-photoreactive hybrid-1 quadruplex conformation in Li\textsuperscript{+} solution. It was also found that the initial rate of \textit{cis,syn} T=T CPD formation was 4.8-fold faster in Li\textsuperscript{+} than in K\textsuperscript{+} solution (Figure 2.7C). The greater rate with Li\textsuperscript{+} could be explained by the fact that all of the TTA CPD sites in the quadruplex forms are in loops that may inhibit \textit{cis-syn} CPD formation, whereas only one TTA is in a loop in a hairpin structure, and the other two are in the stem where the T’s can stack on each other.

2.4.7 Photoproduct formation in 15-mer sequences.

Because the \textit{trans,anti} photoproduct was observed to be formed in the presence of Li\textsuperscript{+} that does not stabilize G-quadruplex structures (14-16), but is known to stabilize reverse Hoogsteen hairpins (21), further attention was focused on hairpin structures as potential photoreactive intermediates. To this end, the photochemistry of Tel15, which consists only of loop 1 through loop 3 of Tel26, was investigated, and while too short to form an intramolecular quadruplex or triplex, it could still form a reverse or normal Hoogsteen hairpin, or a bimolecular quadruplex. To assess the importance of Hoogsteen-type base pairing on \textit{anti} CPD formation, the photochemistry of
HP15WC, which can only form a Watson Crick base paired hairpin, was also investigated. UVB irradiation of Tel15 gave the \textit{trans,anti} T(A)=T(A) product in about the same yield in Li\textsuperscript{+} and K\textsuperscript{+} solutions as observed for Tel26 in solutions of these same ions (Figure 2.8). Some of the \textit{trans,anti} CPD formed in Na\textsuperscript{+} and K\textsuperscript{+} may arise, however, from an bimolecular G-quadruplex, as a denaturing gel of the irradiated duplexes showed some slower moving bands that would be consistent with dimers in addition to the faster moving intramolecular crosslink product (Figure 2.13B).

The dimer bands observed in the gel may arise from either through photocrosslinking of a bimolecular G-quadruplex with adjacent lateral loops, or a reverse or normal Hoogsteen duplex. The dimer bands could not, however, have arisen from dimeric quadruplex forms of the types proposed for the Oxytricha d(GGGGTTTTGGGG) that have loops on opposite ends (23). The CD of Tel15 in K\textsuperscript{+} is very similar to that of Tel26 in K\textsuperscript{+} solution, with maxima at 265 and 290 nm suggesting the formation of a bimolecular hybrid type quadruplex (Figure 2.15). The CD of Tel15 in Na\textsuperscript{+} solution had a maximum at 290 nm, indicating formation of an antiparallel quadruplex as observed for Tel26 in Na\textsuperscript{+} solution, but much weaker in intensity. The Li\textsuperscript{+} sample, however, showed almost no discernable CD from 260-300 nm, indicating the lack of possible involvement of quadruplex type structures in \textit{trans,anti} CPD formation. Most interestingly, \textit{trans,anti} CPD formation was now observed in Na\textsuperscript{+} solution, which had not been the case for the intramolecular quadruplex forming sequences, indicating that the DNA was now able to adopt a photoreactive conformation in the presence of Na\textsuperscript{+}. In contrast to the G-quadruplex forming sequences, UVB irradiation of HP15WC, which had a Watson Crick base paired stem, gave very little in the way of \textit{anti} CPD products, indicating that a B DNA stem does not have the proper geometry for \textit{anti
CPD formation. The presence of a B type conformation in HP15WC was confirmed from the CD (Figure 2.15) showing a positive band at about 280 nm and a negative band at 250 nm (24).

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

To test further the hypothesis that anti CPDs can arise from hairpin conformations of Tel26 in K⁺, five 21-mer sequences (Table 1.1) were prepared in which G’s were strategically replaced with inosine to interfere with intramolecular G-quadruplex formation, but not with reverse or normal Hoogsteen hairpin formation (25). In addition, a sequence was prepared in which the two terminal G3-tracts were replaced by complementary Watson-Crick sequence to prevent intramolecular quadruplex formation and enforce hairpin formation. Replacing G by inosine reduces the ability of G to serve as a Hoogsteen or reverse Hoogsteen H-bond donor because of the absence of the N2-NH₂ group to form a second hydrogen bond while still maintaining its ability to serve as a Hoogsteen H-bond acceptor (Figure 2.6). In Tel21_3’con and Tel21_5’con, the two consecutive (con) G₃-tracts at the 3’-end or 5’-end of the 21-mer were replaced with I’s so that the entire 3’-side or 5’-side of the hairpin would function as the Hoogsteen accepting strand. In Tel21_3’alt and Tel21_5’alt, alternating (alt) sets of G₃-tracts were replaced with I’s so that a conformation with a two-fold axis of symmetry (disregarding the hairpin loop) would be produced in which the Hoogsteen accepting strand switches to the opposite strand following the T•T crosslink site at the center. With Tel21_mix, every other G was replaced with inosine so that hairpins formed from this sequence would have the Hoogsteen donor alternating with the Hoogsteen acceptor. To obtain a trans,anti CPD stereochemistry from any of the sequences in a hairpin form, one of the T’s of the T•T mispair must be in a syn glycosyl conformation. As shown in (Figure 2.6), only for the reverse Hoogsteen conformation do the two T’s come close enough, and with the proper alignment
for the required [2+2] cycloaddition to take place. The two TA base pairs flanking the T•T mispair site could in principle form either a reverse or normal Hoogsteen base pair as shown.

Of the five inosine-containing sequences, irradiation of only Tel21_3’con and Tel21_3’alt produced the trans,anti T(A)=T(A) CPD with the greatest yield (8% and 10% respectively) and selectively (Figure 2.9 & Figure 2.16 & Figure 2.17). The common feature of these sequences is that they share the same 3’-I3 tract. Although the CD for these two sequences in K+ are similar as far as having a positive peak at about 290 nm, there is another positive peak at 255 nm of almost equal amplitude for the 3’alt sequence that is much less pronounced in the 3’con sequence (Figure 2.15). The peak at 290 nm is indicative of an antiparallel quadruplex that may have arisen from dimerization of the two intact G3 tracts in each sequence. One cannot, therefore, rule out on the basis of the CD that some of the trans,anti CPDs observed may have arisen from such a bimolecular duplexes. On the other hand, the quadruplexes that could form would have quite different structures owing to the different relative positions of the G3-tracts in the two sequences, which would have been expected to lead to different patterns of anti CPD products. Furthermore, although Tel21_5’con had a very similar CD to that of Tel21_3’con in K+ (Figure 2.15), it gave a much more complex mixture of products (Figure 2.18), suggesting that the positioning of the G3 and I3-tracts is important. Tel21_5’alt also led to a complex mixture of products (Figure 2.19) and had a sizable positive peak at 260 nm in the CD spectrum, and a weaker one at 300 nm, indicative of a parallel quadruplex with some antiparallel quadruplex. The higher selectivity and yield with sequences having the 3’-I3 tract can be explained by a hairpin structure in which trans,anti T(A)=T(A) CPD formation is favored with the 3’-terminal I3-tract is serving as the Hoogsteen or reverse Hoogsteen acceptor strand. Replacing the 3’-I3 sequence with a Watson Crick accepting GCG sequence in HP21WC abolished formation of the trans,anti product (Figure 2.9). The CD
of this sequence in K+ solution (Figure 2.15) shows a weak positive peak at about 270 nm and a weak negative peak at 240 nm that is similar to that of HP15WC and supportive of the presence of a B DNA duplex and an absence of a quadruplex. Replacing the 3’-I3 sequence with an alternating 3’-GIG sequence in Tel21mix also greatly reduced the yield to 3% (Figure 2.9 & Figure 2.20). This sequence cannot form either an intramolecular or intermolecular quadruplex, and the CD spectrum in K+ only shows a weak positive peak at 270 nm (Figure 2.15).

In Na+ solution, the highest apparent yield and selectivity for trans,anti T(A)=T(A) CPD formation was for Tel21_3’alt (18%) and Tel21_mix (14%) (Figure 2.9& Figure 2.20), though Tel21_3’con also gave approximately the same yield (6%) and selectivity as was observed in K+ solution. Co-injection experiments with authentic trans,anti product suggested, however, that what appears to be the trans,anti product in Tel21_3’alt in Na+ may actually be a mixture of the trans,anti product with a slightly earlier eluting compound (Figure 2.17), which may be the cis,anti T(A)=T(A) based on the prior work(4). The high yield of the trans,anti T(A)=T(A) CPD in Tel21_mix in Na+ (Figure 2.9) contrasts with the low yield observed in K+-containing solutions indicating that Na+ and K+ can also affect the conformation of non-quadruplex forming sequences. Sequences with 5’-I3 arrangements of inosine (5’-alt and 5’con) led to much lower yields and poor selectivity (Figure 2.18-Figure 2.19). The HP21WC sequence containing a Watson-Crick stem produced very little trans,anti CPD (Figure 2.9).

The results with the inosine substituted sequences are more consistent with the involvement of a hairpin intermediate in the formation of the trans,anti T(A)=T(A) CPD than with an intermolecular G-quadruplex intermediate, especially for Tel21_mix, which cannot form any type of G-quadruplex. Furthermore, the preference for a 3’-terminal I3 sequence indicates that the photoreactive hairpin involves normal or reverse Hoogsteen (Figure 2.6) rather than C2-symmetric
base pairing (Figure 2.6, Figure 2.14) which would not expected to be affected by the presence of inosine. The observation of large quantities of the \textit{trans,anti} T(A)=T(A) CPD in the inosine-modified sequences in Na$^+$ solution, especially Tel21_mix, indicates that it must be the chair quadruplex conformation that forms in Na$^+$ solution, and not Na$^+$ by itself, that inhibits \textit{anti} CPD formation in the canonical sequence. Also of interest is the increased yield of the \textit{cis-syn} T=TA CPD products in the inosine-containing sequences. This may be due to the CPD enhancing effect of inosine relative to guanine (26) that was proposed to be due to the higher ionization potential of I than G that prevents competitive exciplex formation and deactivation (27).

\textbf{2.4.9 \textit{Trans,anti} CPD formation in non-telomeric sequences.}

At this point, it is not possible to conclude unambiguously whether the photoreactive conformation leading to the \textit{trans,anti} T(A)=T(A) CPD is a Hoogsteen or reverse Hoogsteen hairpin or some other type of hairpin. To determine whether or not a reverse Hoogsteen hairpin can be the photoreactive conformation, three 21-mer sequences, HP21GAG, HP21GGA, and TelAAG, were designed in which multiple G$\cdot$G base pairs of the human telomere sequence were mutated to A$\cdot$A base pairs (Table 2.1). A cannot form Hoogsteen base pairs or quartet structures with itself, although both A and G can form reverse Hoogsteen base pair with themselves (28, 29), and in principle, symmetric base pairs (Figure 2.6, Figure 2.14) (30). All three mutated sequences, led to the formation of the \textit{trans,anti} T(A)=T(A) CPD photoprodct upon irradiation with UVB in both K$^+$ and Na$^+$ solution (Figure 2.10). To determine what structures might be present, CD spectra were obtained for all three sequences in Na$^+$ and K$^+$ as well as in Li$^+$ in which G-quadruplexes would be inhibited.

The CD spectra of HP21GAG in Li$^+$, Na$^+$ and K$^+$ solution are almost identical with a minimum at about 285 nm and a positive peak at 260 nm (Figure 2.15) which is the opposite of B
DNA and very similar to left-handed or Z form DNA (24). A slightly longer version of this sequence, (TTAGAG)$_4$ has previously been shown not to adopt a G-quadruplex structure, and has a similar CD spectrum to what we observe (31). One can, therefore, conclude that no G-quadruplexes are involved in the formation of the trans,anti CPD in this sequence, and that the sequence is adopting a similar distribution of conformations in the presence of each ion. On the other hand, the CD spectra of HP21GGA in both K$^+$ and Na$^+$ solutions showed a positive peak at about 300 nm (Figure 2.15) that could indicate the presence of an antiparallel, two tetrad G-quadruplex that may have formed from the four intact G$_2$-tracts. The spectrum in Li$^+$, however, was quite different, showing a negative peak at about 245 nm and a positive peak close to 270 nm. As a result, one cannot rule out that the formation of the trans,anti T(A)$=\overline{\text{T}}$(A) product in TelGGA isn’t due to a G-quadruplex. The CD spectra of HP21AAG in the presence of all three ions was very weak. The CD spectra in K$^+$ and Li$^+$ closely resemble each other with a positive peak at about 270 nm and a negative peak at 250 nm, whereas the spectrum in Na$^+$ was weaker and showed a negative peak at about 285 nm (Figure 2.15). The CD spectra establish that HP21GAG and HP21AAG are able to form the trans,anti T(A)$=\overline{\text{T}}$(A) CPD although some conformation other than a G quadruplex, most likely a reverse Hoogsteen (Figure 2.6) or possibly a symmetric base paired hairpin (Figure 2.6, Figure 2.14).

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

To see if we could obtain evidence for the presence of a hairpin intermediate in the human telomeric Tel26 sequence in K$^+$ solution, we carried out gel electrophoresis experiments at 4 °C (12) on Tel26 and two Tel26 mutants in which selected G•G pairs were replaced with A•A (HP26GAG and HP26AAG) in the presence of Tris•H$^+$, Li$^+$, Na$^+$, and K$^+$ (Figure 2.11).
HP26GAG and HP26AAG correspond to HP21GAG and HP21AAG with two additional A’s at the 5’-end, and one at the 3’-end to make them the same length as Tel26. Neither sequence showed evidence of G quadruplex formation. The addition of the A’s weakened the CD spectrum of HP26GAG compared to HP21GAG, but strengthened the CD spectrum of HP26AAG compared to HP21AAG, which now showed a prominent positive peak at approximately 270 nm and a negative peak at 250 nm (Figure 2.15). As controls, we used triplex- and quadruplex-forming sequences d(TTAGGG)$_3$ (Tel18) and d(TTAGGG)$_4$ (Tel24) and scrambled versions of all the sequences denoted by S as a prefix.

In the absence of metal ions, Tel26 moved the slowest of all, and significantly slower than its scrambled version, while the other sequences had about the same mobility as their scrambled counterparts (Figure 2.11A). In Li$^+$, Tel26 went slightly slower than its scrambled counterpart, whereas d(TTAGGG)$_4$ went slightly faster. The G→A mutated sequences migrated together and more slowly than Tel26, and with the average mobility of their scrambled counterparts and not faster as might have been expected for hairpin formation (Figure 2.11B). It may be that there was significant unintended secondary structure in the scrambled sequences. In Na$^+$ and K$^+$ solutions, both Tel26 and d(TTAGGG)$_4$ moved much more quickly than their scrambled counterparts and with an equal or greater mobility than for the d(TTAGGG)$_3$ 21-mer or its scrambled counterpart (Figure 2.11C & D). The G→A mutated sequences, however, moved faster than they did in Li$^+$ compared to their scrambled counterparts. The electrophoretic behavior of Tel26 and d(TTAGGG)$_4$ in Na$^+$ and K$^+$ is highly indicative of G-quadruplex formation (12). No band was observed in Na$^+$ or K$^+$ for d(TTAGGG)$_4$ or Tel26 that corresponded to the bands observed in the presence of Li$^+$, suggesting that a hairpin was either not present, or if present, was only present in a very low amount and in rapid equilibrium with the quadruplex structure.
2.5 Discussion

We have shown that the trans-anti T(A)=T(A) CPD can form in both Na\(^+\) and K\(^+\) solution in A-rich sequences only capable of forming reverse-Hoogsteen hairpin or possibly hairpins with symmetric base pairs. It is not possible to prove conclusively that a hairpin is the sole or contributing photoreactive conformation of human telomeric sequences in K\(^+\) solution because of the preferential formation of more stable G-quadruplexes that make it difficult to independently detect independently the presence of a hairpin conformation. Under conditions where G-quadruplex formation is inhibited by Li\(^+\), it is likely that a reverse Hoogsteen hairpin is the photoreactive conformation given previous NMR data showing the formation of this conformation in the *Oxytricha* d(G\(_4\)T\(_4\)G\(_4\)) sequence(21). Although the trans,anti CPD also forms in Na\(^+\) or K\(^+\) solution when half the G’s are replaced with I’s to prevent intramolecular G quadruplex formation, we cannot be certain if this is due to a reverse or normal Hoogsteen hairpin, or a bimolecular quadruplex, except the Tel21_mix, which cannot form quadruplexes. The photoreactive conformation is unlikely to be due to a triplex conformation, which could also form under these conditions, as we have previously found that sequences that can only form a triplex give a low yield of trans,anti T(A)=T(A) (4). None-the-less, there is evidence that suggests that a non-Hoogsteen hairpin structure, and in particular, the reverse Hoogsteen structure, is a major photoreactive conformation leading to the formation of the trans-anti T(A)=T(A) CPD in K\(^+\) solution.

The first line of evidence for a photoreactive reverse Hoogsteen hairpin structure is that the yield of the trans,anti T(A)=T(A) photoproduct is greater for Tel26 than for Tel22 in K\(^+\) solution. This can easily be explained by the fact that Tel26 results from the addition of two A’s to both the 5’- and 3’-end of Tel22 that can form two additional reverse Hoogsteen base pairs at the end of
the hairpin, which would further stabilize a reverse Hoogsteen hairpin structure and increase its relative concentration. This is assuming that the two additional reverse Hoogsteen base pairs could only form on the end of a reverse Hoogsteen hairpin. It is also possible that the A’s and G’s of the non-Hoogsteen hairpin form C2-symmetric base pairs (Figure 2.6, Figure 2.14), but these would all require one of the glycosyl bonds of each pair to be in the higher energy syn form. The symmetric G•G base pair involving the sugar face are highly unlikely owing to a large steric interaction between the amino group and the sugar ring that is in a syn glycosyl conformation (base pair #4, Figure 2.14). Also, replacement of either G of the other symmetric G•G base pair (Figure 2.6, Figure 2.14) with inosine should not affect base pairing since the NH₂ group is not involved, and hence should not affect the photochemical outcome, which is contrary to what was observed in some of the inosine substitution experiments where the position of the substitution makes a big difference (Figure 2.9). The symmetric base pairs are also less likely because they would not have the same conformation as the normal or reverse Hoogsteen conformation needed to form the T•A base pairs.

A second line of evidence for a reverse Hoogsteen hairpin is that the major anti CPD photoproduct of Tel26 in both Li⁺ and K⁺ is the trans,anti T(A)=T(A) CPD, whereas three additional products are produced in significant yield in Tel22 in K⁺ solution. The reverse Hoogsteen structure affords a simple explanation for the selective formation of the trans,anti T(A)=T(A) product as shown in Figure 2.6. It predicts that the second T in each TTA segment will be opposite each other and lead to the trans,anti T(A)=T(A), while the other T’s would only be able to form adjacent T=TA CPDs. Furthermore, the trans stereochemistry for the T(A)=T(A) CPD can be explained by the requirement for one of the thymidines to adopt the syn gycosyl conformation for the 5,6-double bonds to be in close enough proximity to photodimerize. The
possibility that the cis,anti T(A)=T(A) CPD may also forming in the Tel21_3’alt sequence could be further explained by the T’s being flanked by two reverse Hoogsteen TA base pairs which would pinch the DNA duplex and bring the two T’s even closer together. In the normal Hoogsteen conformation, the two double bonds on the two T’s are much farther apart and would be significantly less able to photodimerize (Figure 2.6).

The third line of evidence is that the yield of the trans,anti T(A)=T(A) CPD in Tel26 linearly decreased in Li\(^+\) with an increase in temperature and was almost zero at 37 °C. The temperature dependence closely parallels the melting behavior of the hairpin form in Li\(^+\) (Figure 2.7A) which has a T\(_m\) of 40 °C. We had observed a virtually identical linear decrease in yield in K\(^+\) solution, which does not parallel the much sharper melting transition at a T\(_m\) of around 60 °C for the G-quadruplex form (Figure 2.3B). The melting temperature of the reverse Hoogsteen hairpin conformation of Tel26 is also expected to be low in K\(^+\) solution based on some recent data with A and G-containing sequences (32). The transition temperature (reported as the inflection point in a melting curve) of a 13 bp reverse-Hoogsteen hairpin containing 69% G’s and 31% A’s and no mismatches is 33 °C in Na\(^+\), but could not be measured in K\(^+\) solution due to the preferential formation of a G-quadruplex. This mixed sequence has a similar composition to a Tel26 reverse Hoogsteen hairpin, which would form an 11-mer with 60% G and 40% A.

We cannot rule out that some portion of the trans,anti T(A)=T(A) CPD is indeed coming from a basket type G-quadruplex structure in K\(^+\) solution, as recent QM/MM calculations have concluded that there is a viable path to this product in the form 3 quadruplex conformation (11). There are two arguments, however, against the involvement of the form 3 quadruplex as the principal photoreactive intermediate. The first is that the rate of formation in Tel26, which has been shown by NMR to adopt primarily non-photoreactive type I and type II quadruplexes, is two
times faster than for the NF3 sequence, which has been shown to adopt the putative form 3 quadruplex conformation (9). The second is that all attempts to make mutations to stabilize further the form 3 structure in Tel26 by adding nucleotides to loop 1, with the exception of insertion of a G, reduced the yield of the trans, anti CPD photoproduct (9). The negative effects of such mutations could perhaps be better explained if the photoreactive structure was the reverse Hoogsteen hairpin, in which case adding nucleotides to loop 1 would disrupt the photocrosslinking site. The insertion of the G, however, would not have been expected to affect the structure as much, because the G-tract could slip and restore the photocrosslinking conformation.

At this point, we believe that we can explain the photoreactivity of human telomeric G-quadruplex forming sequences in Na\(^+\), K\(^+\) and Li\(^+\) solution as illustrated in Figure 2.12. First, the lack of anti CPD formation in Na\(^+\) solution may be explained by a greater thermodynamic preference and kinetic stability for the basket form (Figure 2.12B) than the reverse Hoogsteen hairpin form or Form 3 basket, and the inability of the T’s in the two lateral loops to interact due to steric constraints. It is also possible that Na\(^+\) in the outer G-quartet coordinates with a thymine in one of the loops and holds it in such a way as to prevent anti CPD formation, as well as preventing strand slippage to make the form 3 basket structure. Coordination of a thymine to Na\(^+\) has been suggested to explain the orientation of a thymine in a diagonal loop in the dimeric d(G\(_4\)T\(_4\)G\(_4\)) Oxytricha quadruplex (33). In K\(^+\) solution, the DNA is more conformationally labile and the various hybrid and basket forms are in equilibrium with a small amount of weakly photoreactive Form 3 basket (Figure 2.12C) and the more photoreactive reverse Hoogsteen hairpin (Figure 2.12D). Indeed, a recent NMR study showed the presence of a non-quadruplex ensemble (referred to as unfolded) for d(TT(GGGTTA)\(_3\)GGGA) that equilibrates between hybrid 1 and hybrid 2 structures (34). The small relative amount of a photoreactive reverse Hoogsteen hairpin
form of Tel22, however, would then be expected to increase by the addition of two additional A’s to each end to form Tel26, which shows a higher yield and specificity for the \textit{trans,anti} T(A)=T(A) CPD. The reverse Hoogsteen structure most favoring the \textit{trans-anti} and potentially \textit{cis, anti} T(A)=T(A) product in K\textsuperscript{+} appears to be the one with the same Watson to Hoogsteen orientation as in Tel21_3’alt and Tel21_3’con (Figure 2.9) with the 3’-terminal G\textsubscript{3}-tract functioning as the Hoogsteen acceptor. In Li\textsuperscript{+} solution, G-quadruplex formation is inhibited, thereby increasing the fraction of the photoreactive reverse Hoogsteen structure, which would explain why the rate of the \textit{trans,anti} product of Tel26 is faster in Li\textsuperscript{+} than K\textsuperscript{+}.

Reverse Hoogsteen hairpins are well known and studied in the context of triplex DNA formed between two polypurine strands and a polypyrimidine strand and when formed in duplex DNA have been referred to as H\textsuperscript{*} DNA (35, 36). One polypurine strand base pairs to the purine strand of a Watson-Crick duplex in an antiparallel orientation via reverse Hoogsteen base pairs with all nucleotides in the anti glycosyl conformation as established by NMR on an intramolecular triplex (37). Triplex formation by reverse Hoogsteen base pairing was first exploited for targeting polypurine tracts in duplex DNA by a third polypurine strand (38). It has also been employed for targeting poly-pyrimidine strands through the use of poly-purine reverse Hoogsteen hairpins and circular DNA (39-41). Given the prevalence of the reverse Hoogsteen base pair in triplexes, it would not be unexpected for a reverse Hoogsteen hairpin to form in human telomeric DNA, although the stability of such hairpins may be compromised by the mismatches resulting from the TTA loop sequence. It is still possible, however, that other, less well established base pairs are involved, such as C2-symmetric Pu•Pu base pairs (Figure 2.6, Figure 2.14) proposed for a parallel duplex of d(GA)\textsubscript{n} (30) and an antiparallel hairpin of \textit{Oxytrichia} telomeric DNA (22).
In conclusion, it would appear that the photochemistry of G-quadruplex forming human telomeric sequences is rather complex, and may involve both reverse and normal Hoogsteen hairpins, as well as G-quadruplex conformations that are highly dependent on the cation and DNA sequence. Surprisingly, the reverse Hoogsteen hairpin is rarely considered in experiments and simulations of G-quadruplex folding (42-44), but perhaps should be considered as a conformation that may precede, and may be present in equilibrium with the quadruplex structures. It would be the first structure expected to form when single strand (GGGTTA)_n is released from B form duplex DNA that has all G’s in anti glycosyl conformations, or when released from hPot1, which binds single strand telomeric DNA and holds three of four G’s, in anti glycosyl conformations (45). Reverse Hoogsteen hairpins may thus play an otherwise unrecognized role in transactions involving human telomeric DNA, and possibly purine-rich regulatory sequences, in vivo, and the formation of trans-anti CPDs might be an indication of such intermediates. In addition, anti CPD formation may be a useful internal UV probe in addition to iodouracil (46) for following folding events in DNA in vitro or in vivo.

2.6 Acknowledgements

We thank Dr. Blankenship for use of his CD instrument, and the Washington University NIH Mass Spectrometry Resource (Grant No. P41GM103422). Research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health under award number R01CA40463.

2.7 References


2.8 Tables and Figures

A. Structures and photoproducts of G-quadruplex forming sequences.

- **cis,syn T=T dimer**
  pT=pT(N): R₁=H, R₂=dN

- **trans,anti T=T dimer**
  pT(N)=pT(N): R₁=H, R₂=dN

- **cis,anti-T=T dimer**
  pT(N)=pT(N): R₁=H, R₂=dN

B. Known structures of human telomeric G-quadruplex forming sequences in Na⁺ and K⁺ solution. The nucleotide numbering system is for the Tel26 sequence, and the numbers 1-3 refer to the TTA loop numbers as shown below.

**Figure 2.1: Structures and photoproducts of G-quadruplex forming sequences.** A) Structures of the *cis,syn* and *trans,anti* CPDs of TT and their nuclease P1 degradation products. B) Known structures of human telomeric G-quadruplex forming sequences in Na⁺ and K⁺ solution. The nucleotide numbering system is for the Tel26 sequence, and the numbers 1-3 refer to the TTA loop numbers as shown below.
<table>
<thead>
<tr>
<th>X</th>
<th>Loop 1</th>
<th>Loop 2</th>
<th>Loop 3</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>7</td>
<td>10</td>
<td>13</td>
</tr>
<tr>
<td>Tel26</td>
<td>AAA GGG TTA</td>
<td>GGG TTA GGG TTA GGG TTA GGG AA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tel22</td>
<td>A GGG TTA</td>
<td>GGG TTA GGG TTA GGG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NF3</td>
<td>GGG TTA</td>
<td>GGG TTA GGG TTA GGG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tel15</td>
<td>TTA GGG TTA GGG TTA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tel15U²</td>
<td>TTA GGG TUA GGG TTA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HP15</td>
<td>TTA CGG TTA GCG TTA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tel21_3’alt</td>
<td>GGG TTA III TTA GGG TTA III</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tel21_3’con</td>
<td>GGG TTA GGG TTA III TTA III</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tel21_5’alt</td>
<td>III TTA GGG TTA III TTA GGG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tel21_5’con</td>
<td>III TTA III TTA GGG TTA GGG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tel21_alt</td>
<td>IGI TTA GIG TTA IGI TTA GIG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tel21GGA</td>
<td>GGA TTA GGA TTA AGG TTA AGG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tel21GAG</td>
<td>GAG TTA GAG TTA GAG TTA GAG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tel21AAG</td>
<td>AAG TTA AAG TTA GAA TTA GAA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tel21WC</td>
<td>GCG TTA GGG TTA GGG TTA CGC</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.1: Sequences used in this study.
Figure 2.2: HPLC traces of NP1-digested UVB irradiated Tel26 in the presence of Li$^+$, Na$^+$ and K$^+$ ions.
Tel26 was irradiated with UVB light on ice and then degraded overnight with NP1 at 37°C.
Figure 2.3: CD spectra and G-quadruplex UV melting temperature curves.
A) CD spectra of Tel26 in the presence of 100 mM of the indicated metal ion at 4 °C. B) Temperature dependent absorption curves at 295 nm for Tel26 in the presence of the indicated concentration of metal ion. The curves were normalized to the 5 °C value.
Figure 2.4: Kinetics of photoprodut formation in Tel26 and NF3. 
Cis,syn and trans,anti CPD formation in (A) Tel26 and (B) NF3 determined by the nuclease P1 assay.
Figure 2.5: Effect of cation composition and temperature on cis,syn and trans,anti CPD formation.
(A) Effect of the mole fraction of KCl in 100 mM NaCl on cis,syn T=TA and trans,anti T(A)=T(A) CPD formation. (B) Effect of temperature on cis,syn and trans,anti CPD formation in 150 mM KCl solution.
Figure 2.6: Base pairing possibilities in hairpin structures.
Top panel: Possible base pairing arrangements for hairpin forms of Tel22 and Tel26 with symbols: •, reverse or normal Hoogsteen, or symmetric; ♦, reverse Hoogsteen or symmetric; ↑, allowed Watson face to Hoogsteen face base pairing. Tel26 is related to Tel22 by the addition of the underlined A’s. Lower panels, the different base pairing possibilities for G•G, G•I, A•T and A•A showing the pairs of glycosyl conformations in bold and dashed lines required to have an antiparallel helix. Only reverse Hoogsteen pairing allows for all anti glycosyl conformations. Also shown is the reverse Hoogsteen conformation of the T•T mispair site that results in the closest approach of the circled bonds required for the formation of the trans,anti CPD stereochemistry.
Figure 2.7: Temperature dependence and kinetics of trans,anti CPD formation in Tel26.
A) Temperature dependence of the relative fraction of trans,anti T(A)=T(A) CPD formed in Li$^+$ solution vs fraction helix form. Kinetics of B) trans,anti T(A)=T(A) CPD formation and C) cis,syn T=T(A) formation in Li$^+$ and K$^+$ solution.
**Figure 2.8: Photoproduct formation in 15-mers with Li⁺, Na⁺, and K⁺.**

HPLC traces of NP1-digested UVB irradiated Tel15 and the Watson Crick HP15WC hairpin in 100 mM metal ion. Possible hairpin conformations for the 15-mers where base pairing is denoted by: •, Hoogsteen; reverse Hoogsteen, or symmetric; ↑, allowed Watson to Hoogsteen face base pairing; and |, Watson Crick. Because both Tel15 and HP15WC have a 5′-terminal TTA sequence lacking a 5′-phosphate, T=TA will arise from the 5′-terminal TTA, and pT=TA will arise from the internal TTA site following NP1 digestion.
Figure 2.9: Photoproduct formation in ionsine- and Watson Crick base pair containing 21-mers with Na\(^+\) and K\(^+\).

HPLC traces of NP1-digested UVB irradiated 21-mers in 100 mM Na\(^+\), and 100 mM K\(^+\).

Possible hairpin conformations for the 21-mers where base pairing is denoted by: •, reverse or normal Hoogsteen, or symmetric; ↑, allowed Watson to Hoogsteen face base pairing, and |, Watson Crick. The asterisk indicates a possible mixture of cis,anti and trans,anti isomers, see Figure 2.17.
Figure 2.10: Photoproduct formation in non-telomeric 21-mers with K$^+$ and Na$^+$. HPLC traces of NP1-digested UVB irradiated HP21GAA, HP21GAG and HP21AAG in 100 mM K$^+$ or Na$^+$ solution. Possible hairpin conformations for the 21-mers where base pairing is denoted by: ♦, reverse Hoogsteen or symmetric; and ↑, allowed Watson face to Hoogsteen face base pairing.
<table>
<thead>
<tr>
<th></th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C_HP26GAG</td>
</tr>
<tr>
<td>2</td>
<td>C_HP26AAG</td>
</tr>
<tr>
<td>3</td>
<td>C_Tel26</td>
</tr>
<tr>
<td>4</td>
<td>C_4Repeats</td>
</tr>
<tr>
<td>5</td>
<td>C_3Repeats</td>
</tr>
<tr>
<td>1'</td>
<td>HP26GAG</td>
</tr>
<tr>
<td>2'</td>
<td>HP26AAG</td>
</tr>
<tr>
<td>3'</td>
<td>Tel26</td>
</tr>
<tr>
<td>4'</td>
<td>4Repeats</td>
</tr>
<tr>
<td>5'</td>
<td>3Repeats</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ATTGATAGAGAAAGAAGTGTAAGTG</td>
</tr>
<tr>
<td>2</td>
<td>AGTAAAAGAATAAAATAATGTAAGTA</td>
</tr>
<tr>
<td>3</td>
<td>GAAGAGAGAGATTGTGTGAGTGTGAG</td>
</tr>
<tr>
<td>4</td>
<td>TGTGAGACTGTGACTGTGACTGTGAG</td>
</tr>
<tr>
<td>5</td>
<td>GAGTGATGTGAGTGTGAGTG</td>
</tr>
<tr>
<td>1'</td>
<td>AAAGAGTTAGAGTTAGAGTTAGAGA</td>
</tr>
<tr>
<td>2'</td>
<td>AAAAGTTAAAGTTAGAATTAGAAAA</td>
</tr>
<tr>
<td>3'</td>
<td>AAAGGTTAGGTTAGGTTAGG</td>
</tr>
<tr>
<td>4'</td>
<td>TTAGGGTTAGGTTAGGTTAGG</td>
</tr>
<tr>
<td>5'</td>
<td>TTAGGGTTAGGTTAGGG</td>
</tr>
</tbody>
</table>

Table 2.2: Gel electrophoresis of non-telomeric and telomeric sequences.
Figure 2.11: Gel electrophoresis of non-telomeric and telomeric sequences. The sequences shown at the bottom were electrophoresed at 400 V in a cold room at 4°C on a 15% native polyacrylamide gel with Tris-borate EDTA buffer in the absence (A) or presence (B-D) of 100 mM of the indicated metal ion. The bands in the dotted boxes correspond to Tel26 and (TTAGGG)$_4$. 

Figure 2.12: Proposed explanation for the variation in anti CPD formation as a function of sequence and metal ion.

A) General scheme showing the effect of sequence and cation on the DNA conformations present for Tel22 and Tel26, in which the additional underlined A’s on Tel26 serve to further stabilize the reverse Hoogsteen hairpin base pairing. Based on the inosine and adenine substrates, the preferred pairing is denoted by the symbol ↑ for Watson face to Hoogsteen face reverse Hoogsteen base pairing, and ♦, for reverse Hoogsteen or symmetric base pairing. B) Na$^+$ is proposed to stabilize the basket conformation and prevent the T’s from photoreacting, possibly by chelation by the thymine. C) Intercalation of K$^+$ and the larger loop size of the form 3 quadruplex as a possible explanation for a lack of an inhibitory effect of K$^+$ on anti CPD formation. D) How a hairpin containing reverse Hoogsteen sections flanking a TT mismatch could facilitate trans,anti CPD formation, where the reverse Hoogsteen T•A base pairs are shorter than the G•G pairs, leading to a pinching effect that would bring the T T’s close together.
Figure 2.13: Denaturing PAGE gels of 5'-32P-end labeled Tel26 and Tel15 before and after irradiation in the presence of various cations. 
A) Tel26. There were no new species with higher molecular weight than Tel26 detected, indicating that all photoproducts resulted from intramolecular structures. S is untreated Tel26. B) Tel15 was irradiated in the presence of 10 mM Tris-HCl, 100 mM Na+, K+, 1 mM Ca^{2+}, Mg^{2+}. The only new species with higher molecular weight than Tel15 were detected for Na+ and K+ solution, suggesting that some of the anti CPDs formed under these conditions resulted from intramolecular structures, whereas for the other conditions they arose intramolecularly. The lower band in unirradiated Tel15 is an impurity. Substrates were photoirradiated with UVB light for 2 h and electrophoresed on a denaturing 15% polyacrylamide gel.
Figure 2.14: C2-symmetric homopurine base pairs.
Base pairs 1 and 3 have the same conformation and would make a regular helix, but to be antiparallel, one of the glycosyl angles in each pair would have to be syn. In addition, replacing guanosine with inosine would not change the hydrogen bonding and would not be expected to make a substantial change in the structure. Strangely, it was previously proposed that a parallel helix formed with A-A and G-G base pairs would best form with base pair 3 (both G’s in syn conformations) and the conformationally unrelated base pair 2 with both A’s in anti conformations, even though the more likely helix would be formed from base pairs 1 and 3 with all nucleotides in anti conformations. Base pair 4 is unrelated to any of the other base pairs, and would not be expected to form in a hairpin because one of the glycosyl conformations would have to be syn, which would cause steric interference with the amino group.
Figure 2.15: CD of sequences used in this study. CD’s were obtained in a 10 mm pathlength cell at 4 °C with 5 μM ODN in 100 mM of the indicated metal ion chloride salt in 10 mM Tris-HCl, pH 7.5. The spectra were the average of three scans, and then smoothed with a running average of 5 points/2 nm and zeroed at 320 nm.
Figure 2.16: Irradiation of Tel21_3’con (5’-GGTATAGGTTAIIITTAIII-3’).
Irradiation of Tel21_3’con in Na+ and K+ solutions and co-injection of K+ solution with authentic product from Tel26 in K+. 
Figure 2.17: Irradiation of Tel21_3’alt (5’-GGGTTAIIITTAGGTTAIII-3’).
Co-injection of irradiated Tel21_3’alt in Na\textsuperscript{+} and K\textsuperscript{+} solutions with authentic product from Tel26 in K\textsuperscript{+}. It appears that the major product in the Na\textsuperscript{+} solution may be a mixture of cis,anti and trans,anti isomers. It is more difficult to determine whether there is only one product in K\textsuperscript{+} solution. The peak at 46 min was not observed in the irradiation shown in Fig. 7 and may be annealing time dependent.
Figure 2.18: Irradiation of Tel21_5’con (5’-IIITAIITTAGGGTTAGGG-3’).
This substrate also did not lead to significant or selective trans,anti T(A)=T(A) product in either Na\(^+\) or K\(^+\) solution. The arrow shows authentic product from Tel26 in K\(^+\).
Figure 2.19: Irradiation of Tel21_5’alt (5’-IIITTAGGTTAIITTAGGG). This substrate did not lead to significant or selective trans,anti T(A)=T(A) product in either Na\(^+\) or K\(^+\) solution, possibly because it would require reverse AT base pairing. The arrow shows authentic product from Tel26 in K\(^+\).
Figure 2.20: Irradiation of Tel21_mix (5'-IGITTAGITGTAIGITTAGIG-3').
This substrate did not lead to significant trans,anti T(A)=T(A) product in K⁺ solution, but did lead to substantial product in Na⁺ solution. The arrow shows authentic product from Tel26 in K⁺.
Chapter 3 A DNA Library for Determining the Sequence Dependence of Cyclobutane Pyrimidine Dimer Formation\(^1\)

3.1 Abstract

The *cis-syn* cyclobutane pyrimidine dimer (CPD) is the major photoproduc\^\textsuperscript{t} induced by UV irradiation of duplex DNA that results in C to T mutations at dipyrimidine sites found in human skin cancers. The frequency of CPD formation at a given dipyrimidine site by direct irradiation depends on the dipyrimidine sequence, the flanking sequence, and the wavelength of light, as well as the conformation of the DNA, and its interaction with proteins. If a photosensitizer is involved, the structure and photophysical properties of the photosensitizer may also be important. Although there are many studies reporting the frequency of photoprod\^\textsuperscript{uct} formation in genomic sequences, there has been no systematic study of sequence context dependence of directly excited or photosensitized CPD formation and how this may relate to mutation hotspots and cold spots. We have designed 129-mer sequences containing all 64 possible NPyPyN tetrads (where PyPy is CC, CT, TC, or TT, and N is A, C, G or T), and used a T4 endonuclease gel electrophoresis assay to determine the relative yields of photoproduc\^\textsuperscript{t} formation. The results show that CPD yields for different tetrads varied by as much as an order of magnitude. The yield of CPDs under UVC irradiation at a given site decrease in the order TT > TC > CT > CC, whereas the yield of CPDs under UVB irradiation at a given site decreased in the order TT > TC > CC > CT. The yield of CPD formation was lowest with a 5’-G and highest with a 5’-T, whereas the yield was lowest with either a 3’-C, G or T, and highest with a 3’-flanking A. We also studied the sequence context effect

\(^1\) All contents in this chapter are in manuscript preparation and have not been published.

98
on photosensitized CPD formation in the presence of acetone and norfloxacin. The results not only show that the efficiency of photosensitized CPD formation depends on the flanking bases, but also indicates that the efficiency depends on the structure and properties of the photosensitizer. Analysis of the data provides some general rules for estimating the CPD distribution in native duplex DNA.

3.2 Introduction

Dipyrimidine photoproducts induced by solar UV radiation are well established to be the major type of cellular photodamage involved in the development of skin cancer (1),(2). Cyclobutane pyrimidine dimers (CPDs) and pyrimidine (6–4) pyrimidone photoproducts (6-4PPs) are the two principal classes of dipyrimidine dimers. CPDs can form at any dipyrimidine site by a [2+2] photocycloaddition between the C5-C6 double bonds of two adjacent pyrimidine bases (Figure 3.1A). In native B-form DNA, only the cis-syn stereochemistry of the CPD has been observed owing to the stereochemical constraints imposed by the DNA structure (3). CPDs are of major relevance to skin cancer because they not only form at a faster rate than 6-4PP, but also are repaired in slower rate thereby increasing their mutagenic potential (4). CPDs are the most predominant photoprodut induced by all types of UV light. A recent study showed that CPDs can also be generated via a dark pathway for 2 to 3 hours after UVA and UVB irradiation, possibly through a chemisensitization mechanism (5). Additionally, C-containing CPDs are not stable and spontaneously deaminate to U (or T at 5-methylC sites) resulting in the frequently observed C to T or CC to TT transition mutations at dipyrimidine sites in UV irradiated DNA. Studies have revealed that UVB-induced CPDs at 5-methylC containing dipyrimidines are correlated with C to T mutation hotspots in the p53 tumor suppressor gene in skin cancers (6,7). Therefore, it is of great significance to understand the photochemistry of CPDs in detail.
The efficiency of CPD formation is affected by many factors. First, it is dependent on the wavelength of UV and dose as well as the sequence of the two pyrimidines involved in the dimerization. UVC (190 – 290nm) and UVB (290 – 320nm) give rise to CPDs via direct excitation of DNA (8,9). Using DNA sequencing gel techniques, Mitchell et al. studied the CPD distribution of the Alu sequence in vitro, and estimated that the ratio of CPD formation at TT, TC, CT, CC sites was 68: 16: 13: 3 under UVC irradiation, and 52: 21: 19: 7 under UVB irradiation (8). Another study by Douki and Cadet used high performance liquid chromatography associated with electrospray tandem mass spectrometry (HPLC-MS/MS) to estimate CPD distribution within isolated and cellular DNA. They reported that in vitro UVC-induced CPD distribution was 2.970: 1.823: 0.573: 0.069 (55: 34: 11: 1) for TT, TC, CT and CC whereas UVB-induced CPDs followed the order of 1.023: 0.694: 0.289: 0.094 (49: 33: 14: 4) in vitro and 3.147: 0.279: 1.289: 0.577 (60: 24: 11: 5) in cells (9). These studies showed that CPDs predominantly form at TT sites followed by TC, CT and CC. Also, compared with UVC, UVB leads to more cytosine-containing CPDs. However, discrepancies exist between studies owing to the different DNA sequences or genomes used, as well as the UV lamps used. In addition, none of the studies provides systematic information on the influence of bases beyond the two pyrimidines, although numerous other studies have revealed that the 5’ and 3’ bases flanking the dipyrimidine sites have an impact on the efficiency of CPD formation. For example, it has been reported that the presence of a flanking pyrimidine enhances CPD formation (10), whereas a flanking guanine suppresses CPD formation (8,10-15).

There have been far fewer studies of CPD formation by UVA (320 – 400 nm) irradiation because of the extremely low efficiency of CPD formation. Some early studies of UVA induced DNA damage found that CPDs are produced in greater yield than oxidative lesions like 8-oxo-
7,8-dihydroguanine (8-oxoGua), and is the main type of DNA UVA damage detected in both cultured cells and human skins (16,17). The quantum yield of CPD formation by UVA is at least two orders of magnitude lower than those in UVB and UVC, but approximately five times as high as that of 8-oxoGua. Because DNA weakly absorbs UVA photons, it was proposed that UVA-induced CPD could arise either through direct excitation or by photosensitization via the excitation of endogenous chromophores (18). Experiments revealed that the distribution of dipyrimidine photoproducts is quite different for UVA compared with UVB and UVC, with TT CPD representing 90% of the dipyrimidine photoproducts, followed by TC and CT, and essentially no CC (19).

The mechanism of CPD formation by various forms of UV light has attracted considerable attention. Ultrafast time-resolved experiments showed that CPDs are formed within 1 ps upon exposure of single-stranded oligodeoxynucleotide (dT)\textsubscript{18} to UVB light (20). The reaction is thought to proceed via a singlet state with minor contribution from triplet states (21). On the other hand, the triplet state plays the principal role in the photosensitized formation of CPDs (22). Photosensitization occurs via a triplet-triplet energy transfer (TTET) process in which the excited state of the photosensitizers, such as ketones (23,24) and fluoroquinolones (25,26) transfer triplet energy to the DNA through a Dexter electron exchange process. The efficiency of the TTET process is related to the triplet energy of photosensitizer and the DNA bases, the coupling interaction between them, as well as complex formation between the DNA and the photosensitizer, and the photosensitizer environment (27,28). Although TTET has been investigated for many years, the effect of sequence context on CPD formation has not been systematically studied. Only one early study reported that TTTA and TTAAT sites are the hotspots for pyridopsoralen – photosensitized CPD formation (29).
To gain more insight into the effect of flanking sequence on direct and photosensitized CPD formation, we designed DNA sequence libraries containing all possible flanking sequences. To do so, we developed a computer program to generate the shortest possible sequences containing all possible flanking sequences of the type NPyPyN where PyPy represents a dipyrimidine site, and N represents all four possible flanking bases. The frequency of CPD formation at each site was then determined by high resolution gel electrophoresis of the products resulting from cleavage of an irradiated, 5’-end-labeled sequence with the CPD-specific T4 pyrimidine dimer glycosylase (T4-pdg). We determined the effect of flanking sequence on CPD formation with 254 nm UVC light, narrowband UVB light (NB UVB), Broadband UVB light (BB UVB) or Broadband UVB light with thin glass filter (Filt UVB). The yield of CPDs under UVC irradiation at a given site decrease in the order TT > TC > CT > CC, whereas the yield of CPDs under UVB irradiation at a given site decreased in the order TT > TC > CC > CT(). The yield of CPD formation was lowest with a 5’-G and highest with a 5’-T, whereas the yield was lowest with either a 3’-C, G or T, and highest with a 3’-flanking A. We also determined the effect of flanking sequence on CPD formation by photosensitization with acetone and norfloxacin (NFX) (Figure 3.1B). Acetone is among the first compounds shown to photosensitize DNA (30). The $E_{\text{triplet}}$ energy of acetone is around 337kJ/mol, which is able to transfer triplet energy to all four DNA bases. Norfloxacin (NFX) is one of the most efficient fluoroquinolones reported to photosensitize DNA (31). It exhibits an $E_{\text{triplet}}$ value of approximately 269 kJ/mol (25) and has previously been used to establish the $E_T$ threshold to produce TT CPDs in double stranded DNA (27). In the present study, we found that photosensitized CPD formation is highly dependent on the nature of flanking bases. Furthermore, the flanking sequence effect on photosensitized CPD formation by NFX was found to be dramatically different from that
induced by acetone. The data obtained provides some general rules for estimating the CPD
distribution in duplex DNA as a function of flanking sequence.

3.3 Materials and Methods

3.3.1 Materials

Oligodeoxynucleotides (ODNs) were purchased from Integrated DNA technologies (IDT), [γ-
$^{32}$P] ATP from Perkin Elmer, and T4-pdg (pyrimidine dimer glycosylase, previously known as
T4 endonuclease V) was prepared from a clone provided by Stephen Llyod.

3.3.2 Python code for generating sequence libraries

A python script seq_generator.py was written to minimize the length of a sequence containing all
possible NPyPyN sequences once by randomly shuffling all 64 NPyPyN sites and then overlapping
them so as to result in no more than four contiguous pyrimidines. The shortest sequences obtained
were 129-mer in length and some examples are shown in Table 3.2. One of them was randomly
chosen for study. The python code for seq_generator.py is included in online repository

3.3.3 Synthesis of 149-mer containing plasmid

The DNA sequence used for the studies was synthesized by ligating 4 short DNA sequences (48-
mer, 42-mer, 41-mer and 40-mer ODNs) using 3 complementary ligation scaffold sequences (39-
mer, 38-mer, 34-mer ODNs) (Table 3.1). All ODNs were added in equal molar amounts followed
by slowly cooling from 95°C to room temperature in 50 mM NaCl. The ligation reaction was
carried out overnight at 23°C with T4 DNA ligase (Promega) in 30 mM Tris-HCl (pH 7.8 at 25°C),
10 mM MgCl$_2$, 1 mM dithiothreitol (DTT) and 1 mM adenosine triphosphate (ATP). An aliquot
of the ligated single-stranded DNA (ssDNA) was then PCR amplified to obtain the double-
stranded DNA (dsDNA) in 100 µl of PCR reaction mix (1x LongAmp Taq DNA polymerase buffer, 0.5 µM of forward and reverse primers, 300 µM of dNTPs, 4 µl of 1 unit/µl LongAmp Taq (New England Biolabs) for 19 cycles followed by phenol extraction and ethanol precipitation. Subsequently, the dsDNA was purified by 10% Native PAGE (10% acrylamide, 0.33% bisacrylamide, 1 × TBE). The gel band containing the cold dsDNA was located with through the use of the two radiolabeled dsDNAs, then was cut out, crushed and shaken in 3 mL ddH₂O overnight. The eluted cold ssDNA was then concentrated followed by phenol extraction and ethanol precipitation. After dsDNA stock was obtained, it was cloned into EcoRI site of pBlueScript II SK-vector DNA (Agilent). A clone containing the desired 149-mer was verified by DNA sequencing (Figure 3.8). The plasmid stock was then harvested from the clone using Promega PureYield plasmid miniprep system.

### 3.3.4 Preparation of 5'-32P-labeled DNA substrates

In a typical experiment, 149-mer DNA or 79-mer was prepared from the plasmid by PCR amplification using LongAmp Taq DNA polymerase. The 100 µL PCR reaction mixture contained 1x LongAmp Taq DNA polymerase buffer, 30 ng plasmid, 0.5 µM of 5'-32P-labeled forward primer (Table 3.1), 0.5 µM of reverse primer, 300 µM of dNTPs, 4ul of 1 unit/µL LongAmp Taq.

### 3.3.5 UV irradiation

The irradiation solutions contained 5'- radiolabeled 149-mer or 79-mer DNA substrate and sensitizers if present. The final 20 µL solutions also contained T4-pdg buffer (50 mM NaCl, 5 mM EDTA, 10 mM Tris–HCl buffer (pH 7.5)). All irradiation was carried out at 0 °C. The UVC light source consisted of two UVC tubes (XX-15F, Spectroline). The broadband UVB irradiation was
carried out with two UVB tubes (XX-15B, Spectroline) delivering 4 mW/cm$^2$. The narrowband UVB irradiation was provided by a 311 nm UVB Lamp (PLS-9W/01/2P, Philips) delivering 2.5 J/m$^2$/s. The UVA irradiation was carried out with a UVA lamp (PL-L 36W UVA, Philips). The emission spectra of the light sources are provided in Figure 3.9. The UV exposure times ranged between 4 s and 30 min, depending on the experiment. All irradiations and analyses were performed at least in triplicate. For photosensitized samples, deaeration was achieved by bubbling with nitrogen.

### 3.3.6 Mapping of CPD formation by T4-pdg Assay

After UV irradiation, each end-labeled DNA sample was incubated with 1 µg of T4-pdg for 30 min at 37°C before treating with 1M piperidine at 90 °C for 5 min to ensure complete elimination of sugar ring. The samples were then dried in speedvac followed by two times of washing by ddH$_2$O. Subsequently, each sample was resuspended in 20 µL of formamide-dye before loading onto a 10% Denaturing PAGE. A control sample that was exempt from UV irradiation was treated with the same steps. In addition to the control and UV samples, a Maxam and Gilbert G-sequencing ladder specific to the 149-mer DNA was prepared according to standard procedures (38) and loaded on the gel for aligning the bands. We did triple loading for each condition to ensure all 64 sites of the 149-mer could be clearly visualized on the same gel.

### 3.3.7 Quantification and data analysis

For directly excited UV samples, the lane and band tools in Quantity One was used to determine quantitatively the relative amounts of CPD formation as a function of nucleotide position on the gel. The relative quantity of each band is calculated by Quantity One when the band within the certain lane was chosen. Bands were well separated from position 1 to 52 of 129mer on the first
loading, from 72 to 129 on the second loading, and from 42 to 86 on the third loading. Therefore, first set of loading were overlapped with third set of loading at GTTT, TTTC, TTCA sites, whereas the second set of loading and the third set of loading overlapped at GCTT, CTTT, TTTA, ATTA, ACCA, ACTG, GCCG sites. Consequently, for each kind of UV irradiation conditions, we multiplied all band intensity for the first set of loading lane by the ratio of (total intensity of overlapping bands in the third loading / total intensity of overlapping bands in the first loading), and also multiplied all band intensity for the second set of loading lane by the ratio of (total intensity of overlapping bands in the third loading / total intensity of overlapping bands in the second loading). After the band intensities were obtained and normalized for all 64 sites, the final percentage yield was calculated by dividing the intensity of each band by the total intensity of all 64 bands.

For photosensitized samples, the volume tool in Quantity one was used to quantitatively determine the relative amounts of CPD formation at the NTTN sites. The TC, CT or CC sites were not quantified due to the low intensity. The intensity of each band was measured by bounding a box around the band, and the size of bounding box was held constant for a given band across all lanes of a given gel. The amounts of photosensitized CPD at each site was determined by subtracting the band intensity for UV only lane from the band intensity for photosensitizer-UV lane. Bands were well separated from position 1 to 52 of 129-mer on loading 1 lane, 42 to 110 on loading 2 lane, and 86 to 129 on loading 3 lane. The final values for all 16 sites were obtained by combining data from all three loading lanes. Since loading lane 1 and loading lane 2 were overlapped at ATTG, GTTT, TTTC, CTTC sites, we multiplied all band intensity for loading lane 1 by the ratio of (total intensity of overlapping sites for lane 2 / total intensity of overlapping sites for lane 1). Similarly, loading lane 2 and loading lane 3 were overlapped at ATTA, ATTT, TTTT,
TTTG sites, so we multiplied all band intensity for loading lane 3 by the ratio of (total intensity of overlapping sites for lane 2/ total intensity of overlapping sites for lane 3). After the band intensities were obtained and normalized for all 16 TT sites, the final percentage yield was calculated by dividing the intensity of each band by the total intensity of all 16 bands. For all setting, three independent experiments were conducted to obtain three set of data, and the average was used to draw the heatmap graph.

3.4 Results and Discussion

3.4.1 Design of the sequence

There are 64 possible dipyrimidine flanking sequences of the type NPyPyN, four possible bases for N, and two for each pyrimidine, or \(4 \times 2 \times 2 \times 4 = 64\) sites (Figure 3.2A). If these were all strung together, they would create a sequence that is \(4 \times 64\) or 256 nucleotides long. Although high resolution sequencing gels and capillary electrophoresis sequencers may be capable of resolving the individual bands of such a length that would be produced upon T4-pdg cleavage at the CPD sites, in practice they are hard to quantify. Therefore, we decided to make a more economical sequence by overlapping all the 64 possible tetranucleotide sequences, with the restriction that there would be no more than four pyrimidines in a row, to minimize band overlap problems. To this end, we wrote a computer program in Python to generate sequences fulfilling these criteria and discovered the shortest sequence to be 129 nucleotides in length. To minimize end effects and be able to visualize bands resulting from cleavage of the CPDs at the ends of the 129-mer sequence on gel, we added 10-mer sequences that contained no dipyrimidine sites to both the 5’- and 3’- ends, resulting in a 149-mer DNA (Figure 3.2B). Finally, in order to insert the designed sequences into a plasmid to preserve the sequence and allow it to be retrieved by PCR, we engineered GATTC EcoRI sites into the ends of the sequence which were flanked by an extra
6 base pairs to ensure efficient digestion. The designed sequence was synthesized by ligating four short DNA sequences with the help of scaffold sequences, and then PCR-amplified to obtain the duplex. Then the duplex was cloned into the EcoRI site of a pBlueScript SK- vector. Individual clones were screened by enzyme digestion and verified by DNA sequencing to ensure successful insertion of the designed 149-mer (Figure 3.8).

3.4.2 Gel assay for CPD formation

To determine the sequence specificity of CPD formation, a 5’- radiolabeled 149-mer was prepared by PCR amplification of the clone with a 5’-radiolabeled forward primer and a cold reverse primer. The radiolabeled sequence was then irradiated, treated with the cis-syn-dimer-specific T4-pdg (pyrimidine dimer glycosylase, previously known as T4 endonuclease V), and then analyzed by a high resolution denaturing acrylamide gel electrophoresis. The UV irradiation times were adjusted to ensure close to single hit kinetics, by limiting the amount of uncut DNA to at least 70%, but allowing for sufficient cleavage to be able to quantify the individual bands. The cleavage bands were mapped onto the DNA sequence alongside the Maxam-Gilbert G reaction bands to localize precisely and quantify CPD yields. To better visualize and quantify the bands towards 5’-end of the directly irradiated sequence, we also synthesized a 79-mer, which started from position 61 of 129-mer to the end of 149-mer (Figure 3.2B), by PCR-amplification with a different 5’- radiolabeled forwarding primer (Table 3.1). For photosensitized-samples, most cleavage sites corresponded to TT sites we could be easily resolved by sequentially triple loading of the 149-mer substrate (Figure 3.6). To control for DNA cleavage occurring at pre-existing abasic sites in the DNA that would be cleaved by the AP lyase activity of T4-pdg or hot piperidine, unirradiated samples were also treated in the same way as irradiated samples, but no extra band were observed at the dipyrimidine sites.
3.4.3 Effect of UV light source and dipyrimidine sequence on CPD formation

We first determined the frequency of CPD formation as a function of the UV light source at all 64 dipyrimidine sites within the 149-mer. Each site represents one of 64 possible base tetrads of the form NPyPyN. We used four types of UV light, UVC (254 nm), broadband UVB (BB UVB, 276 – 400 nm, centered at 312 nm), broadband UVB with thin glass filter (FiltBB UVB, 302 – 400 nm, centered at 330 nm) and narrowband UVB (NB UVB, 310 – 320 nm, centered at 311 nm). The UVA light (322 – 400 nm, peak at 365 nm) did not give rise to sufficient CPDs, therefore was excluded from the data quantification (data not shown). The emission spectra of all the UV lamp are shown in Figure 3.9. Figure 3.3 showed all 64 cleavage bands clearly resolved on a 10% denaturing PAGE gel. The band intensities for sites at positions 1 to 52 were obtained from the first set of lanes (first loading of T4-pdg digested 149-mer), at positions 72 to 129 on second set of lanes (first loading of T4-pdg digested 79-mer), and positions 72 to 129 on third set of lanes (second loading of T4-pdg digested 149mer). The following hierarchy in the induction of CPDs was found: the ratio of TT, CT, TC and CC was 48: 24: 14: 14 for UVC, 34: 29: 16: 21 for BB UVB, 29: 29: 18: 24 for FiltBB UVB, and 30: 29: 18: 23 for NB UVB (Figure 3.7). In agreement with earlier studies (8,9), the distribution of UVC-induced CPDs at the dipyrimidine level is characterized by the preferential formation of TT over TC and CT, with CC being the least frequent. In the case of UVB light, the result is consistent with previous report that there is an increase in CPDs at most of the dipyrimidine sites containing cytosine as compared to TT. However, what is different from the previous studies was that the efficiency for UVB-induced TT CPDs was not dramatically higher than TC CPDs, especially for UVB light in the longer wavelength range where we can see the yield of TT CPDs is almost the same as that of TC CPDs for NB UVB light and FiltBB UVB light. It is also worth noting that in our study of UVC and
UVB light, we observed a much more efficient formation of CC CPDs. In the case of UVB light, the percentage yield of CC CPDs is even higher than that of CT CPDs. The reason for the discrepancy may be the failure to take into account differences in flanking sequence frequencies between DNA substrates.

### 3.4.4 Effect of flanking bases on the formation of CPDs

Although the total yield of CPDs for each of four types of dipyrimidines are different upon exposure of 149-mer to different UV light, the overall pattern of CPD distribution across all 64 sites under different irradiation settings appears similar (Figure 3.4), with TTTA always being the hottest spot and GCCG or GTCC being the coldest spot. Upon exposure to UVC light, the percentage of CPD formation for all 64 tetrads varies from 0.2% (GCCG) to 6.8% (TTTA). In the case of UVB light, the range is from 0.4% (GTCG) to 5.2% (TTTA) for BB UVB light, from 0.6% (GTCG, GTCC) to 4.1% (TTTA) for FiltBB UVB light, and from 0.5% (GTCG) to 4.7% (TTTA) for NB UVB light.

Figure 3.5 shows frequency of CPD formation at each type of dipyrimidine site as a function of the nature of adjacent bases under different UV conditions, and clearly shows that CPD formation frequency is highly sensitive to flanking bases. Consistent with previous studies, we observed a suppressing effect of 5’ flanking guanine on CPD formation for any dipyrimidine site (8,10-15), however, there is no apparent suppressive effect for a 3’ flanking guanine. Pan et al. proposed that the sequences containing 5’-G exhibited stronger long wavelength absorption because of the formation of an electron donor-acceptor complex (14), leading to more extensive quenching of excited states for a 5’-G versus a 3’- G. A recent computational study supported Pan’s conclusions. They modeled the conformational and electronic effects of the 5’ flanking sequences separately, and concluded that a charge transfer state (CT) between flanking bases and
the pyrimidine base provided a decay pathway for excited thymine to escape from photodimer formation (15). In addition, we observed that the flanking bases exhibit different levels of influence on TT, TC, CT and CC. For example, the suppressing effect of a 5’ flanking guanine is most pronounced at TC site. Also, dipyrimidine sites sandwiched between two guanines shows almost the lowest yield. Contrary to the suppressive effect of a 5’ flanking guanine, we observed higher photoproduct yields when the dipyrimidine sites were flanked by 5’ thymine, especially for TT and TC dimers.

As for the 3’ flanking base effect, the most significant one is the enhancing effect of a 3’ adenine, especially when flanked by a 5’ guanine. The enhancing effect of 3’ adenine is most pronounced for TT dimers. In contrast, TT flanked by 3’ thymine or a CC site flanked by a 3’ cytosine leads to a lower CPD yield.

To conclude, the overall effect of flanking base for all the irradiation conditions can be summarized as follows: T > C > A > G for 5’ flanking base, whereas the effect for 3’ flanking bases is less apparent, with A > T, C, G. The enhancing effect of a 3’ adenine and a 5’ thymine make TPyPyA the highest yield out of 16 combinations for each of four kinds of dipyrimidine sites.

### 3.4.5 CPD distribution as a function of photosensitizer

To investigate the effect of flanking sequence on photosensitized CPD formation, we irradiated the 149-mer in the presence of either acetone plus NB UVB or NFX plus UVA. In this set of experiments, the concentration of the sensitizers and the length of UV irradiation time were adjusted to reach similar levels of damage in the absence of significant DNA damage in the absence of the photosensitizer. In agreement with previous studies (32), both acetone and NFX led to predominant formation of CPDs at TT sites compared to TC, CT or CC (Figure 3.6).
predominance of CPD formation at TT sites is consistent with the fact that thymine has the lowest triplet energy among all the bases, and that triplet energy could migrate to the lowest energy sites. Under both photosensitization conditions, we observed significant CPD formation at all possible XTTY sites except for ATTT and GTTT sites. This can be attributed to our observation that at T tracts, CPD formation is favored at the 3’-end of the T-tract (33). The relative frequency of CPD formation at the different TT sites depended, however, on the photosensitizer, as can be readily seen by comparing the band intensities at GTTC, ATTC, CTTA, TTTC, CTTC, CTTG, ATTA sites, which were much weaker with NFX than with acetone (Figure 3.6). The barplot and heatmap graphs of integrated band densities (Figure 3.7) clearly show the strong flanking bases effect on TT dimers formation for both acetone and NFX, as well as the dramatic difference in CPD distribution pattern between them. The quantification data were normalized based on the total yield of all XTTY sites under each condition. Although the three hottest sites for acetone-sensitized CPD formation are TTTG (15%) > ATTA (13%) > TTTA (11%), they are TTTA (18%) > TTTG (13%) > ATTG (13%), for NFX. There is better agreement with the two coldest sites for photosensitized CPD formation which are GTTT and ATTT formed in 2% and 1% yield respectively for acetone, and 1% and 2% for NFX.

The results not only show that the efficiency of photosensitized CPD formation is also significantly dependent on the flanking bases, but also shows that the two photosensitizers, have different sequence dependence, indicating that they may interact differently with nucleobases. Since acetone is a very small molecule that does not have any appreciable binding affinity for DNA, the transfer of triplet energy must be occurring through a simple collision between the triplet-excited state acetone and the DNA. Because acetone has a higher energy triplet state than all the bases it could transfer triplet energy to any base in the DNA, after which the triplet energy
could migrate to a lower energy site. We also observed that the presence of cytosine as flanking base enhanced the yield of acetone-mediated NTTN CPD formation compared with that of NFX-mediated CPDs. The observation could be attributed to the fact that acetone has higher triplet energy capable of sensitizing all bases in the nucleic acid, whereas NFX can only sensitize thymine residues.

In contrast, previous studies revealed that NFX was able to bind to DNA, preferentially at GC sites, and it has greater affinity for purines than for pyrimidines (34). The binding mode of NFX to DNA is controversial. Two contradictory mechanisms were proposed, intercalation (35) or perpendicular to the helical axis (36). Recently, Cuquerella et al. showed that the NFX and DNA complexation is able to tune the energy transfer between them (28). And Antusch et al. proved that TT CPD exhibited shallow exponential distance dependency on the photosensitizers (37). Therefore, it can be argued that the different flanking base effect between acetone and NFX is the result of their different interaction with DNA. And the flanking base effect for NFX-mediated CPDs represents a combined effect of flanking bases and NFX binding specificity to DNA. Therefore, the higher yield of ATTG and GTTA in the case of NFX could be attributed to its higher affinity for purines.

Another intriguing observation is the low but detectable yield at some of TC and CT sites for photosensitization with acetone (TTCT, CTCC, CTCA, GTCC, GTCA, GCTA, GCTC, CCTC, CCTA, ACTA, GTCT, TCTC, CTCG, ACTC, CTCT, TCTA) but not for NFX (Figure 3.6). Although weak bands were observed at TC, CT sites for the NFX plus UVA, the band intensities were no higher than for control lane (UVA 30 min + no sensitizer). Thus, it appears that these bands were the result of direction UVA excitation. In addition, CPDs at CC sites were not observed in our study. The formation of CPDs at some TC and CT sites in the presence of acetone but not
NFX indicates that the $E_{\text{triplet}}$ value of thymine may be higher when stacked with C than with T. Therefore, acetone, exhibiting much higher triplet energy than NFX, would be able to sensitize these sites.

### 3.5 Conclusion

In summary, our study provides the first complete and systematic investigation of the effect of flanking sequence on CPD formation by direct UV irradiation and in the presence of photosensitizers. The quantitative data presented in this paper indicate that there is a general suppressive effect of a 5’-G on CPD formation, especially at TC sites and an enhancing effect of a 3’-A especially at TT sites. Furthermore, we find that the effect of flanking sequence is different for triplet photosensitized reactions, and that the effect also depends on the structure and/or triplet energy of the photosensitizer. The data presented suggest that photosensitizers may have unique photochemical signatures that might lead to unique signature mutations that can be used to identify them in skin cancers.

### 3.6 Acknowledgements

Research reported in this chapter was supported by the National Cancer Institute of the National Institutes of Health under award number R01CA40463.

### 3.7 References


3.8 Figures

A.

![Structures of the cis-syn TT CPD, TC CPD, CT CPD and CC CPD.]

B.

![Structure of Norfloxacin (NFX) and Acetone]

Figure 3.1: Structure of the CPDs and photosensitizers. A) Structures of the cis-syn TT CPD, TC CPD, CT CPD and CC CPD. B) Structure of Norfloxacin (NFX) and Acetone
Figure 3.2: The sequence of the 149-mer containing all 64 possible NPyPyN tetrads.  A) The 64 unique XPyPyY tetrad sites (where PyPy is CC, CT, TC, or TT, and N is A, C, G or T), B) A 149-mer sequence containing all 64 tetrads: the text in black is the 129-mer containing the 64 tetrad sites, the text in color are the two 10-mer sequences that contain no dipyrimidine sites used to minimize end effects and enable cleavage bands at the ends of the 129-mer to be visualized on the denaturing PAGE gels. The underlined sequence is a 79-mer subsequence.
Figure 3.3: PAGE analysis of CPD formation by UV irradiation. Lane 1: T4-pdg-treated UVC-irradiated DNA, lane 2: T4-pdg-treated BB UVB-irradiated DNA, lane 3: T4-pdg-treated FiltBB UVB-irradiated DNA, and lane 4: T4-pdg-treated NB-irradiated DNA. To resolve all 64 T4-pdg cleavage bands, the 149-mer substrates were sequentially loaded, while only one loading was needed for the 79-mer. The 79-mer comprises positions 61-129 of the 149-mer. Lane G: Maxam Gilbert G ladder, lane C: unirradiated controls treated with T4-pdg.
Figure 3.4: Ordered bargraphs of the relative frequency of CPD formation at XPpyPyY by UV irradiation. Data represent the average of three experiments.
Figure 3.5: Heatmap of relative CPD yields at XPyPyY sites for UV irradiated DNA. Data represent the average of three experiments.
Figure 3.6: PAGE analysis of NFX and acetone photosensitized CPD formation. Lane 1: 149-mer treated with T4-pdg and hot piperidine; lane 2: 149-mer treated with UV light in the absence of photosensitizer followed by T4-pdg and hot piperidine; lane 3: 149-mer treated with UV light in the presence of photosensitizer followed by the digestion with T4-pdg and hot piperidine. The red arrows point to all 16 XTTY sites, whereas the green arrows point to XTCY and XCTY sites. The samples were sequentially loaded a three different times to resolve all 64 cleavage sites.
Figure 3.7: Barplot and heatmap of CPD formation patterns in XTTY by photosensitization. Data represent the average of three experiments.
Table 3.1: Sequences used in this study. A) Oligodeoxynucleotides used to synthesize the 149-mer. B) Primers used to synthesize the 149-mer and 79-mer by PCR (F= forward primer, R= reverse primer).
Figure 3.8: Sequence verification for the 149-mer containing clone.
**Table 3.2:** Examples of sequences of 129-mers containing all possible NPyPyN sites. Each 129-mer sequence contains all 64 XPyPyY sites and each site only appears once. The blue colored sequence is the one studied in the paper.
Figure 3.9: Absorption spectra of the DNA and photosensitizers (dashed lines), and emission spectra of the light sources (solid lines).
Figure 3.10: Time dependent study of UV-induced CPD formation in the 149-mer. Samples were irradiated for the stated time, then treated with T4-pdg, followed by hot piperidine. UVC (lane 1: 15s; lane 2: 30s; lane 3: 60s), BB UVB (lane 1: 15s; lane 2: 30s; lane 3: 60s); FiltBB UVB(lane 1: 30min; lane 2: 60min; lane 3: 1h), NB UVB(lane 1: 5min; lane 2: 10min; lane 3: 20min), UVA(lane 1: 1h; lane 2: 2h; lane 3: 4h). Lane 4 is the same as lane 1 except without T4-pdg digestion.
Figure 3.11: Time dependence of CPD formation in the 149-mer with different UV light sources.
Chapter 4 Conclusions and Future Studies

4.1 Conclusions

UV radiation leads to acute and chronic effects on skin, from sunburn to skin cancer (1,2). CPDs are widely recognized as the most important environmentally produced DNA damage due to the abundance of UVB and UVA light in sunlight, and to the C to T or CC to TT mutations that they cause (3-6). The formation of CPDs is strongly dependent on the UV light and dose, sequence context, DNA structure, and protein binding. Advanced spectroscopy and computation studies support the involvement of the singlet pathway in the UVB-induced CPD, whereas the triplet pathway plays an important role in photosensitized CPDs. In this thesis, we focused on the effect of DNA sequence context, DNA structure, and excitation method on CPD formation.

In the first project, we continued the study of the structure-activity relationships of anti CPD formation in human telomeric DNA sequences that was previously discovered in our group (7). We first investigated our original proposal that NF3 was the photoreactive intermediate leading to the observed trans,anti T(A)=T(A) CPDs. A kinetic study of trans,anti T(A)=T(A) CPD formation revealed, however, that irradiation of the Tel26 sequence produces the trans,anti T(A)=T(A) CPD at twice the initial rate of NF3, and with a steady-state level of the CPD than that of NF3. This result indicated that NF3 may not be the major photoreactive intermediate (8). To gain more insight into what might be the photoreactive conformation leading to the anti CPD, we irradiated the Tel26 in the presence of the various metal ions. Surprisingly, Tel26 leads to higher yield of trans,anti T(A)=T(A) CPD in Li⁺ than in K⁺. However, Li⁺ is known not be able to stabilize the quadruplex structure, but may be able to stabilize a hairpin structure (9). We then studied the photochemistry of various sequences that could form the hairpin structure, such as
truncated telomeric sequences or sequences in which various G’s are replaced with inosine and adenosine to disrupt the quadruplex structure and favor the hairpin structure. The results are consistent with the involvement of reverse Hoogsteen hairpin structure as the photoreactive conformation, and may be in equilibrium with G-quadruplexes. We concluded that Hoogsteen hairpins and might therefore play a hitherto unrecognized role in the recognition and reactivity of DNA in telomeres and in other G-quadruplex forming sequences found in regulatory regions of both DNA and RNA.

In the second project, we designed a 129-mer sequence that contained all 64 NPyPyN sites, and compared the CPD formation among these sites at single nucleotide resolution under either direct excitation or photosensitizating conditions. We determined that UV wavelength, dinucleotide and flanking sequence, as well as well as photosensitizer structure have an effect. This information should be of use for theoretical studies intended to elucidate the mechanism of pyrimidine dimer formation. Also, the sequence libraries developed could be deployed for the study of the formation of other dipyrimidine dimers such as 6-4PP and Dewar photoproducts. Furthermore, in the study of the TTET process, we found that acetone and NFX produced different CPD distributions, indicating that photosensitizer structure and photophysical properties are also important, such as the mode of binding and the triplet state energy of the sensitizer. Sequence libraries could also be designed for the study of other reactions of DNA, such as 8-oxodG formation.

4.2 Future Studies

4.2.1 The sequence context effect on UVA-induced CPDs

The UVA light is of particular importance considering that it is the most abundant type of UV in the terrestrial sunlight radiation. CPDs were recently found to be the major DNA damage resulting
from UVA irradiation not only in cells but also in human skins. However, the reaction mechanism in the UVA range is still under discussion. It would be interesting to see the roles of flanking bases upon UVA irradiation, and how the distribution correlates with our current results. In one of our new studies, we observed apparent band cleavages at CPD sites if we extended UVA irradiation time from 1h to 4h (Figure 3.12). Unlike UVB and UVC, UVA leads predominantly to TT and TC CPDs with very small amounts CT and CC CPDs (Figure 3.13).

To improve our methodology for the study of dipyrimidine photoproduct formation, we designed new 129-mer sequences that have the 16 four consecutive pyrimidine runs (each of which includes two three pyrimidine runs) toward 5’-end for better gel separation and resolution, with the 16 two consecutive pyrimidine runs at the 5’-end (Figure 4.1). The sequence has already been synthesized, cloned and sequenced (Figure 4.2). It would be interesting to see if the relative yields from this sequence would be the same as for the original sequence. If not, it may mean that the context effects extend beyond the immediately flanking sequences.

4.2.2 The effect of sequence on deamination of C-containing CPDs

The CPD lesions occur most frequently between two thymine bases. However, cytosine- or 5-methylcytosine(\textsuperscript{m}C)- containing CPD lesions are much more mutagenic because C or \textsuperscript{m}C is prone to deamination into uracil (U) or T if the C5=C6 double bond is saturated (10). There is a need for the systematic and sensitive study of the effect of sequence context on the deamination of C or \textsuperscript{m}C and its contribution to mutagenesis at these hotspots. It is possible to use the NextGen sequencing methods to carry out these studies with our 129-mer as template. Figure 4.3 shows the general scheme. It may be necessary to optimize the scheme to reduce the sequencing errors (11).
4.2.3 In vivo detection of anti CPDs

Further studies to determine the formation of anti CPD formation in G-quadruplex structures in cellulo and in vivo is needed to better understand the structure-activity properties of this novel class of photoproducts. It has been conducted in our group a post-labeling assay for detecting anti photoproducts from irradiated cells. However, there is a strong need for the development of more sensitive methods than the current assay.

4.3 References


4.4 Figures

ACGTGCACGC ACTCTATTTTTGTCTCGTCCCGCCCTATCTTTACCCCA
ATTCTGTTCGCTTTTACCTTGCTCCCATCCTGCTCT
CATTTACACTGCGTTACTACCGTGCTGTGCTATCATTTGTCA
TCGCCCA CACGCGTACA

Figure 4.1: New designed sequence.
The black part is the 129-mer containing all 64 NPtyPyN sites, and the blue and green parts are the two 10-mers on both 5’ and 3’ ends to help visualize the end of sequences on the Denaturing PAGE gels.

Figure 4.2: Sequencing result of the new sequence containing clone.
Figure 4.3: Scheme for study of sequence context effect on deamination of C-containing CPDs