Investigation of Nucleotide-Mediated CPD Thymine Dimer Photo-Repair in Duplex DNA: Development of a Novel Method to Investigate the Effect of Flanking Bases on the Formation and Repair of CPD Thymine Dimers in Duplex DNA

by

Hamza Ben Zeglam

M.Sc., University of Huddersfield 2012

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

in the
Department of Molecular Biology and Biochemistry
Faculty of Science

SIMON FRASER UNIVERSITY
Spring 2022

Declaration of Committee

Name:	**Hamza Ben Zeglam**
Degree:	**Doctor of Philosophy**
Title:	**Investigation of Nucleotide-Mediated CPD Thymine Dimer Photo-repair in Duplex DNA: Development of a Novel Method to Investigate the Effect of Flanking Bases on the Formation and Repair of CPD Thymine Dimers in Duplex DNA**
Committee:	**Chair: Mark Paetzel** Professor, Molecular Biology and Biochemistry **Dipankar Sen** Supervisor Professor, Molecular Biology and Biochemistry **Peter Unrau** Committee Member Professor, Molecular Biology and Biochemistry **William Davidson** Committee Member Professor Emeritus, Molecular Biology and Biochemistry **Nicholas Harden** Committee Member Professor, Molecular Biology and Biochemistry **Tim Audas** Examiner Assistant Professor, Molecular Biology and Biochemistry **Andrew MacMillan** External Examiner Professor, Biochemistry University of Alberta

Abstract

In a primordial "RNA world," survival and self-replicating nucleic acid species within the extremely harsh environment of early Earth, especially UV radiation, imposed this prebiotic nucleic acid to evolve and acquire certain desirable features to survive and reproduce. Eventually, DNA develops as a more durable nucleic acid analog to store genetic information. The cis-syn Cyclobutane Pyrimidine Dimers (CPDs) are the most common lesion formed in cellular DNA from exposure to solar light. Although many studies have been carried out to identify the influence of DNA sequence and structure on its photochemical and photophysical properties, the molecular bases of these observations are not yet been well understood.

This research has been conducted to comprehensively study the comparative impact of neighboring bases on either side of a pyrimidine pair on CPD formation and photo-repair of a pre-existing CPD in double-stranded DNA (dsDNA). The principle of our approach relies on precise blockage of Taq DNA polymerase at pyrimidine dimers (CPDs), which cause the disappearance of the sequences that contain a thymine dimer from the irradiated DNA library pool. High-throughput sequencing is then done to explore the formation and repair of CPDs over the time-course experiment at different wavelengths. We are using a random double-stranded DNA library consisting of a stretch of 10 randomized base pairs with a central thymine pair flanked by constant bases as primer binding sites for PCR amplification. The constant regions were designed to avoid two adjacent pyrimidines. The sources of UV irradiation being used in this research are a monochromatic 278 nm LED and a 365 nm LED with a triplet sensitizer. Our study is unique and unprecedented in terms of being done on quite long authentic dsDNA and many random nucleotides flanking the thymine pair or dimer, being studied at once. Besides, we use both UVC and UVA monochromatic LEDs to create/repair CPD dimers, which is more comparable to CPD formation due to solar exposure in biological systems. All possible sequences up to 5 nucleotides on both 5' and 3' sides are being studied and we compared our data with the known UV hot spots in the human genome.

Keywords: Photo-repair; DNA; Thymine dimers; Cyclobutane pyrimidine dimers; High-throughput sequencing

Dedication

To my best friend Doctor Rana Faryad Ali

Acknowledgements

I would like to thank my senior supervisor, Professor Dipankar Sen, for his kindness, guidance, teaching and training throughout my Ph.D. research journey. I am fortunate to have this opportunity to work with him, where he trained me to think critically, taught writing manuscripts and to communicate effectively.

I would like to thank my committee members, Professor Peter Unrau, Professor William Davidson and Professor Nicholas Harden for their valuable feedback and encouragement during my research journey and doctoral degree.

I would also like to thank past and current group members of the Sen group for filling the lab with energy and laughter every day. I am also grateful to Joshua Unrau for his kind assistance in statistical analyses through Python and ANOVA. I would also express my thanks to Adam Barlev for coordinating the design and fabrication of UV irradiation setup and his valuable contribution to this work. In particular, I thank Kun Liu, Prince Kumar Lat, Nisreen Shumayrikh, Owen Einarson, Hanadi Ibrahim, Janet Yu Chuan Huang, and Fiona Yutong Huang, for always having nice discussions related to research. I also want to thank the amazing staff at the Molecular Biology and Biochemistry, for all their guidance and support I have had during my time at SFU.

I also thank my parents, wife, brothers, and sisters for their gracious love and support.

Table of Contents

List of Tables

List of Figures

List of Acronyms

2′-Methoxyacetophenone	2-M
8-Oxo-2'-Deoxyguanosine	8-oxo-dG
8-Oxo-7,8-Dihydroguanine	8-oxoGua
Adenine	A
Charge Transfer	CT
Cyclobutane Pyrimidine Dimers	CPDs
Cytosine	C
Deoxyribonucleic Acid	DNA
Double Stranded	ds
Excision Repair Sequencing	XR-seq
First Excited State	S_1
Ground State	S_0
Guanine	G
High-Performance Liquid Chromatography	HPLC
Infrared	IR
Integrated DNA Technologies	IDT
Internal Conversions	IC
Intersystem Crossing	ISC
Ligodeoxynucleotide	dT
Mass Spectrometry	MS
Nanometer	nm
Norfloxacin	NFX
Pi	π
Polymerase Chain Reaction	PCR
Pyrimidine Dimer Glycosylase	T4-PDG
Reactive Oxygen Species	ROS
Ribonucleic Acid	RNA
Spore Photoproduct	SP
Systematic Evolution of Ligands by Exponential Enrichment	SELEX
Thymidine Monophosphate	TMP
Thymine	T
Translesion DNA Synthesis	TLS
Translesion Excision Repair Sequencing	tXR-seq
Triplet-Triplet Energy Transfer	TTET
Ultraviolet	UV
Unique Molecular Identity	UMI
Uracil	U

Chapter 1. Introduction

1.1. Thesis overview and the scope of the study

The "RNA world hypothesis" postulates that the first forms of life on Earth arose from RNA or RNA-like polymers, owing to their unique capability to encode genetic information and catalyze reactions required for life such as self-replication and self-preservation.[1,2] When metabolism became more complex, polypeptide enzymes eventually evolved to do more sophisticated biocatalysis, and DNA developed as a more chemically durable nucleic acid analog to store their genetic information.[3,4] In a primordial "RNA world," survival and self-replicating nucleic acid species within the extremely harsh environment of early Earth, especially UV radiation, imposed this prebiotic nucleic acid to evolve and acquire certain desirable features to survive and reproduce. Extended UV light exposure causes physical damage to the genetic information stored in DNA through the propagation of energy-rich excited states in the nucleobases that promote its photochemical reactions.[5] However, the high photostability of DNA has often been attributed to the ultra-short excited state lifetime (sub-picosecond range) of the individual nucleotides, which leads to ultrafast deactivation of the reactive excited states. This deactivation mechanism is based on single nucleotides that may be evolutionarily selected under intense UV irradiation.[6] Additional long-living excited state phenomena (lifetimes in the ~100 ps), however, have been identified upon exposure of single or double-stranded DNA to UV irradiation, which has been proposed to be excimer/charge transfer states.[7–9] This phenomenon has been extensively studied and has a significant impact on both DNA dimerization rates and DNA repair rates as well.[10,11] The incidence of the major product of photoexcited DNA, thymine dimer, has been shown to vary with neighboring bases,[12,13] DNA conformation,[14] and protein-dependent bending and looping of DNA.[15–17] Although many studies have been carried out to identify the influence of DNA sequence and structure on its photochemical and photophysical properties, the molecular bases of these observations are not well understood yet. Furthermore, the neighbouring deoxynucleotide sequence affects the energy and lifetime of singlet and triplet excited states through excimer formation and perhaps delocalization, but their effect on thymine dimer formation is not well known.[18,19]

My research has been conducted to comprehensively study the comparative impact of neighboring bases on either side of a pyrimidine pair on CPD formation and/or photorepair of a pre-existing CPD in double-stranded DNA. The principal of our approach relies on precise blockage of Taq DNA polymerase at pyrimidine dimers (CPDs)[20], which cause disappearance of the sequences that contain a thymine dimer from the irradiated DNA library pool. High-throughput sequencing is then done to explore both the formation and repair of CPDs over the time-course experiment, at different wavelengths. We are using a randomized double stranded DNA library consisting of a stretch of 10 randomized base pairs around a central thymine pair (the thymine pair has five random bases on either T-side) as shown in Figure 1.1. The above twelve base pairs are flanked by stretches of known sequence, which can serve as primer binding sites for primer extension or PCR amplification. These constant region stretches (51 bases on either side) were carefully designed to not contain two adjacent pyrimidines.

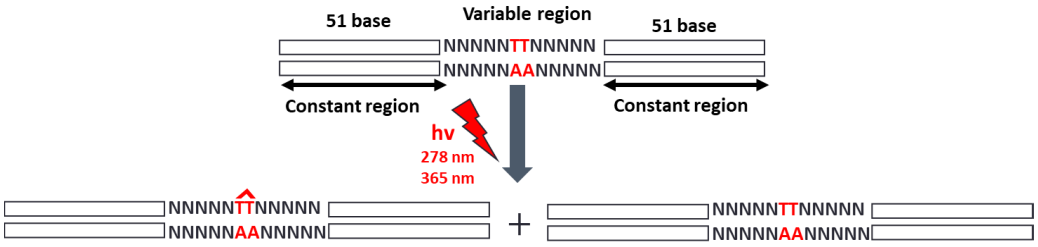


Figure 1.1. A randomized double stranded DNA library consisting of a stretch of 10 randomized base pairs around a central thymine pair (the thymine pair has five random bases on either T-side) The sources of UV irradiation being used in this research are a monochromatic 278 nm LED and a 365 nm LED.

The sources of UV irradiation being used in this research are a monochromatic 278nm LED and a 365 nm LED (for use in irradiation experiments with a triplet sensitizer). Our study is unique and unprecedented in term of it being carried out on authentic dsDNA that incorporates a relatively large number (five on each side) of random nucleotides flanking the central thymine pair or thymine dimer, all being studied at once. In addition, we have used both UVC at 278 nm and UVA at 365 nm monochromatic LEDs to create/repair CPD dimers, which is more comparable to CPD formation due to solar exposure in biological systems. All possible sequences up to 5 nucleotides on both 5' and

3' side being studied, this should enable comparison of our data with known UV hot spots in the human genome.

The goal of this project is to investigate intrinsic photochemical properties of CPD formation and DNA-dependent repair of the same postulated for double strand DNA: the effect of DNA sequence on the propensity to form dimers and the ability to self-repair such dimers under different experimental conditions. In addition, we have inquired into the still controversial role of UVA in CPD formation, which is weakly absorbed by the nucleobases. In this thesis, we have examined a fundamental physico-chemical property of DNA, which in every organism exposed to UV light, therefore likely has in common. The questions we attempt to answer in this work, and which have been asked in earlier literature on the subject is: are actual sequences, coding sequences in living organisms, optimized according to minimize the formation of damaging CPD dimers or not. In summary, the list of objectives for this thesis were:

- Investigate the photoinduced self-repair process of CPD in standard duplex DNA as a constituent of a nucleic acids' inbuilt mechanism for self-preservation.
- Inquire into the controversial role of UVA (weakly absorbed by the nucleobases) in CPDs formation.

1.2 Background

1.2.1 DNA Photophysical and photochemical properties

As demonstrated by Watson and Crick, a DNA single strand consists of a 2'-deoxyribose sugar-phosphate backbone to which glycosidic bonds attach four nucleobases carrying the genetic information.[21] There are two different classes of nucleobases forming DNA's structure: two different pyrimidines (single-ringed aromatic heterocycles): thymine (T), cytosine (C), and two purines (double-ringed aromatic heterocycles): adenine (A), guanine (G) (Figure 1.2).

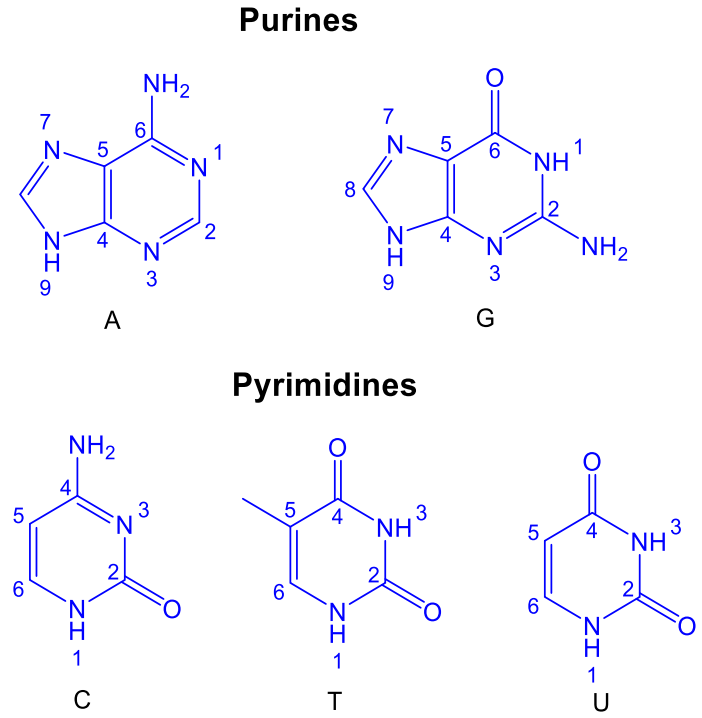


Figure 1.2. Both DNA and RNA contain the nitrogenous bases adenine (A) cytosine (C) and guanine (G) while thymine (T) exist only in DNA and uracil (U) in RNA molecules.

The double-helical structure of the DNA in water solution is thermodynamically enabled by two factors, vertical intrastrand base stacking and horizontal interstrand base pairing. The Watson-Crick base pairing rule states that guanine pairs cytosine involving three hydrogen bonds, whereas adenine pairs thymine with two hydrogen bonds (Figure 1.3). Compared with the weak hydrogen bond, the π-π stacking between the adjacent planar aromatic bases contributes much more to DNA structure stability than the hydrogen bonds.[22] The structure of the DNA backbone is made from repeating deoxyribose sugar-phosphate groups in which the sugar is defined to be on the 3' end, and the phosphate on the 5' end of each nucleotide. The phosphate group attached to the 5' carbon of the deoxyribose sugar on one nucleotide forms a phosphodiester bond with the free hydroxyl on the 3' carbon of the adjacent nucleotide. The asymmetric pattern of the phosphodiester bond imparts the DNA strands a specific orientation that causes the two strands within a DNA double-helix to run in opposite directions to each other and twist together in a helical shape (to form an anti-parallel duplex). A schematic diagram of DNA backbones is illustrated in (Figure 1.3). The sugar-phosphate backbone is strongly negatively charged and hydrophilic, which makes the DNA soluble in water.

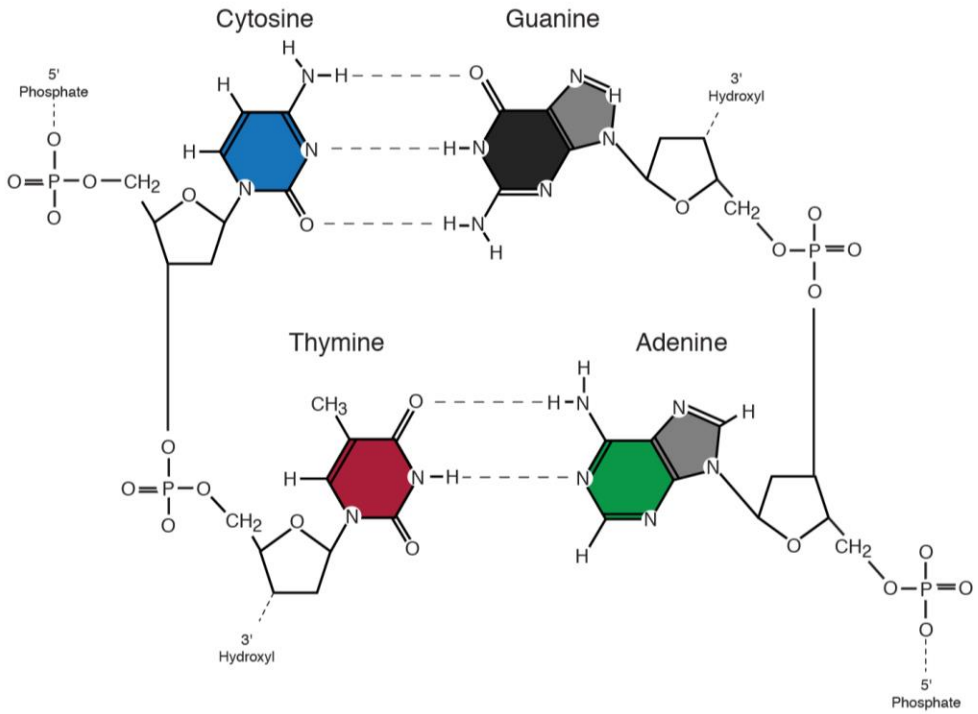


Figure 1.3. A representation of the DNA backbone strand composed of alternating phosphate and sugar groups. The two anti-parallel strands connected via two hydrogen bonds between the adenine base and the thymine base, and three hydrogen bonds between the guanine base and the cytosine base, which stabilize the double helix structure of DNA.

The DNA nucleobases, being aromatic, exhibit different absorbance spectra relative to sugar-phosphate backbone. DNA as a whole absorbs UV light at 260 nm due to heterocyclic rings of the nucleobases[23]; the sugar-phosphate backbone does not contribute to this absorption.[22] At the same time, the four nucleobases show absorbance peaks in the range 250-280 nm, while the DNA backbone absorbance occurs below 220 nm.

Solar UV light can be divided into three wavelength categories: UVA (315-400 nm), UVB (280-315 nm), and UVC (100-280 nm). UVC light is the shortest wavelength and, therefore, the most energetic and most harmful to any living organism. Fortunately, all UVC is absorbed by the ozone-rich layer of the stratosphere before it can reach the Earth.[24,25] The stratosphere of the Earth also shields the earth's surface from the majority of UVB, and only a tiny portion of UVB, constituting only 5% of terrestrial UV radiation, reaches the Earth's surface. In contrast to UVC and the bulk of UVB, the ozone layer is almost transparent to the longer wavelength UV-A light (315–400 nm). UVA rays can

penetrate the skin more deeply, but they are less intense than UVB and often cause indirect damage to DNA via endogenous sensitizing molecules.[25,26]

1.2.2 UV-light induced photolesions

Exposure to solar UV radiation initiates several photophysical and photochemical processes in our genetic material, which may eventually lead to mutagenesis and cancer. Different wavelengths of UV light induce various forms of DNA damage. The cis-syn cyclobutane pyrimidine dimers (CPDs), most commonly thymine CPD dimers, are the most abundant lesion formed in cellular DNA from exposure to solar UVB (280-315 nm) light. To a lesser extent, pyrimidine (6-4) pyrimidone photoproducts (6,4-PPs) are also formed (Figure 1.4).[27–30] These above-mentioned UV-light-induced damages will be discussed in more detail separately.

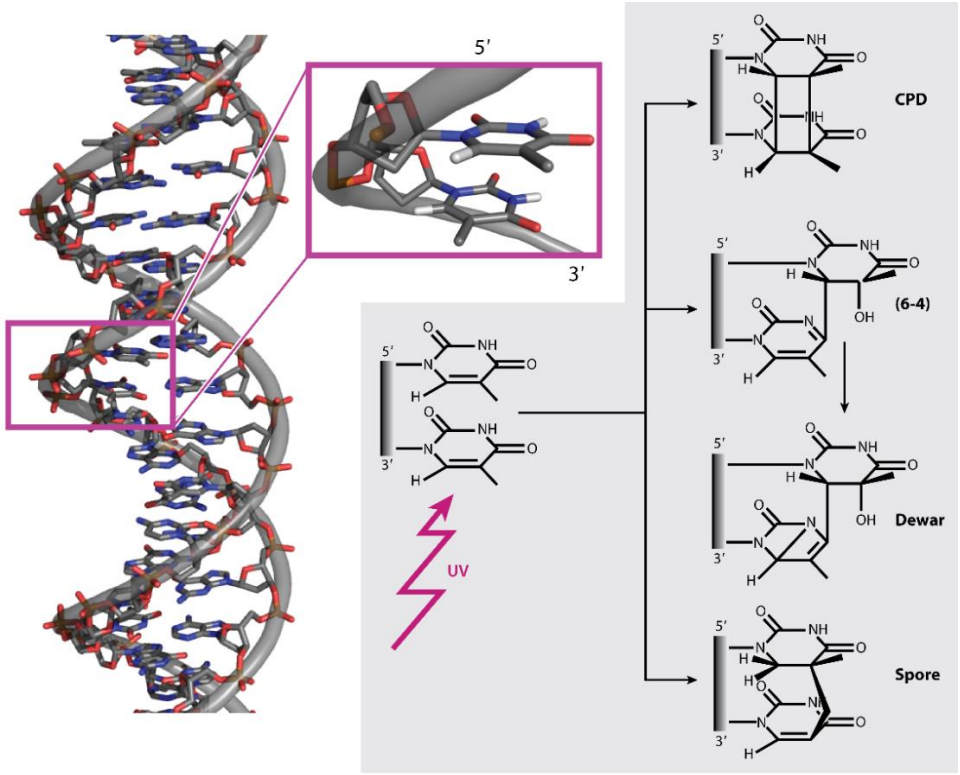


Figure 1.4. Molecular structures of the DNA photolesions depicted by stick representations provides the structure of double-stranded DNA in the B form. The gray tubes are overlaid on the sugar phosphate backbone. The violet box highlights two adjacent thymine bases on one strand. A simplified scheme of the precursor of the photolesions compared with the stick representation is shown in middle column. Chemical structures of the photolesions are represented in the right column. Note that the structures illustrate the chemical

connectivities, but the actual geometries differ substantially.[5] Copyright 2015 by Annual Reviews, and adapted with permission from ref (5).

By contrast to the above, exposure to UVA (315-400 nm) light is responsible for the exclusive formation of CPD lesions without significant formation of (6-4) photoproducts. There are also other, minor photoproducts formed in DNA from exposure to UVB or UVC (100-280 nm) radiation, such as the spore photoproduct, pyrimidine mono-adducts and purine dimers. Pyrimidine monoadducts comprise cytosine, uridine, and thymine photohydrates, as water additions, and thymine glycol.[31] The cytosine photohydrate is the most common UV-induced pyrimidine monoadduct, which is formed by addition of the H_2O to the 5,6-double bond of cytosine, catalysed by UV irradiation.[32,33]

1.2.2.1 Cyclobutane pyrimidine dimer

It is clearly established that CPD is the most prominent photoproduct in cellular DNA and DNA model systems after *in vivo* or *in vitro* UV radiation.[34–40] The cyclobutane thymine dimers are formed by UV-induced photocycloaddition [2π+2π] between two adjacent thymine bases on the same polynucleotide strand, leading to a direct covalent linkage between the two bases. The cycloaddition can produce four different CPD stereoisomers, identified as cis-syn, trans-syn, cis-anti, and trans-anti.[41,42] The various CPD stereoisomers define the conformation and orientation of the two participating pyrimidine bases. Cis/trans referring to whether the two pyrimidines are on the same side or opposite sides of the cyclobutane ring, and syn/anti describing the parallel or antiparallel orientations between the C5-C6 carbons (Figure 1.5). Whereas almost all these CPD isomers have been identified in thymine dimers formed under different experimental conditions,[35] the cis-syn isomer is formed exclusively in B-form double-stranded DNA (the occurrence of a trans-syn isomer is very low in single-stranded DNA).[43,44] The other stereoisomers, such as those with anti-conformations, cannot exist in double-stranded DNA owing to the steric constraints imposed by the DNA backbone. The quantum yield of CPD formation is highly dependent on the identity of the pyrimidine bases involved in both cellular and naked DNA.[44–46] Both UVB and UVA irradiation show a higher distribution of thymine dimer over other pyrimidine dimers, and the reported order is as follows T^T> T^C> C^T> C^C (e.g., ratio = 49 : 33 : 14: 4).[44,47,48] UVC even produce higher T^T ratio

than UVB and UVA and it leads to ratios of 55 : 34 : 11: 1 among T^T >T^C> C^T> C^C, respectively.

Even though the CPDs are very stable photolesions in an aqueous medium, they are prone to some reactions. Further UVC irradiation of CPDs can cause their photo-reversal back to the starting pyrimidines by the splitting apart of the cyclobutane ring.[49] Although the T^T-CPD photolesion is not strongly mutagenic in general, it can block DNA polymerases and results in DNA replication forks stalling and also transcription inhibition.[50] Cytosine-containing CPDs can, additionally, also lead to base substitutions (TC, CT, or CC to TT) in the daughter DNA strand during replication.

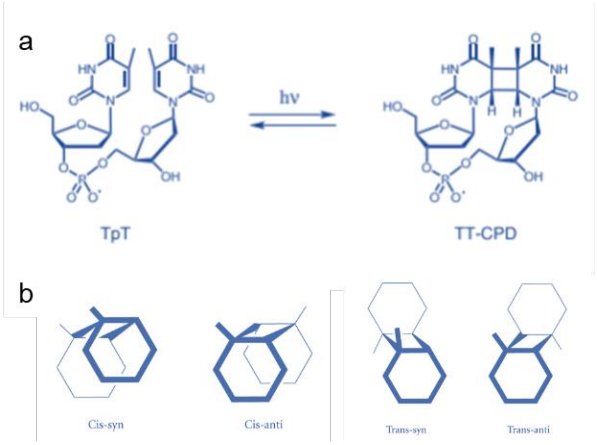


Figure 1.5. a) Schematic diagram illustrates [2π+2π] cycloaddition of C5=C6 double bonds between two adjacent thymine bases causes the CPD lesion. b) Stereoisomers of the CPD lesion.

Cytosine or methylcytosine bases within CPDs are chemically highly unstable and are easily deaminated to uracil (U) or thymine, respectively (Figure 1.5).[51–54] The half-life of deamination of cytosine in CPD is from 2 to 100 hours, which is much faster than cytosine in duplex DNA with a half-life of 30.000 years. This substantial difference results from the loss of aromatic stabilization in the CPD.[52,55] Studies have reported that C5 methylation of cytosine promotes the formation of CPD[56,57] but reduces the rate of deamination.[58] Deamination in trans/ syn is also slower than in cis/syn CPD.[54]

Different organisms in biology have adopted a variety of mechanisms to repair CPDs; the most distinctive pathway is to use visible light and special enzymes (photolyases) to photoreactivate the dimers directly back to thymine monomers. However,

photolyase enzymes are found in many but not all kingdoms of life: placental mammals lost photolyase genes during evolution. Therefore, placental mammals rely on DNA excision repair to remove and replace the affected base, nucleotides, or stretch of DNA to get rid of these potentially dangerous lesions.[5,59] Two processes that work at stalled replication forks to minimize the frequency of recombination and enhance cell survival following UV-induced DNA damage are nucleotide excision repair and translesion DNA synthesis (TLS). (1,2) While nucleotide excision repair is typically thought to be error-free, mutations can occur during translesion synthesis. TLS DNA polymerases that bypass CPDs in arrested fork belong to Y-family DNA polymerases, which include polymerase eta (Polη), REV1, and zeta (Polζ). The Y-family polymerase (Polη) is essential for the efficient and relatively error-free bypass of cyclobutane pyrimidine (TT) dimers (45). Polη, can only insert two adenines precisely opposite CPD lesions, necessitating the use of another TLS polymerase, Polζ, to extend beyond the insertion site (45). When compared to other TLS polymerases, Polζ is unique in that it is reasonably efficient at extending past mispaired primer termini and nucleotides inserted opposite a diversity of DNA lesions, even if this can be potentially mutagenic (45).

CPDs cause a conformational change that twists the B-form DNA axis by 7 to 9 degrees.[60] This slight distortion is enough to make it difficult for the human exonuclease DNA repair enzyme system to detect and repair CPDs, increasing the possibility that they would hinder DNA polymerase and cause mutations. Thus, defects in the CPD repair machinery predispose human skin to cancers and diseases such as Xeroderma pigmentosum and Cockayne syndrome.[27,28] Studies have shown that UVB-induced CPD is associated with skin cancer hot spots, such as in the p53 gene.[57,61]

There are two processes that work at stalled replication forks to minimize the frequency of recombination and enhance cell survival following UV-induced DNA damage are nucleotide excision repair and translesion DNA synthesis (TLS).[62,63] While nucleotide excision repair is typically thought to be error-free, mutations can occur during translesion synthesis. TLS DNA polymerases that bypass CPDs in ds DNA belong to Y-family DNA polymerases, which include polymerase eta (Polη), REV1, and zeta (Polζ). The Y-family polymerase eta ((Polη) is essential for the efficient and relatively error-free bypass of cyclobutane pyrimidine (TT) dimers.[64,65] Polη, can only insert two adenines precisely

opposite CPD lesions, necessitating the use of another DNA polymerase to extend beyond the insertion site. When compared to other TLS polymerases, Polζ is unique in that it is reasonably efficient at extending past mispaired primer termini and nucleotides inserted opposite a diversity of DNA lesions, even if this can be potentially mutagenic. [64]

1.2.2.2 (6-4) Photoproduct and Dewar lesion

The second main type of dipyrimidine photolesion is a pyrimidine (6-4) pyrimidinone (6-4PP) photoproduct. In 6-4PP, a single bond connects the two carbon atoms at the 6 and 4 positions of the two rings (Figure 1.6). The (6-4) photoproduct is generated by UVB or UVC excitation, (but not by UVA) through a Paterno-Büchi reaction. In this reaction, the C5=C6 double bond of one pyrimidine and the C4=O carbonyl group of the adjacent pyrimidine react to give an unstable four-membered ring intermediate.[66] In the case where the pyrimidine on the 3' side is a thymine, it forms an oxetane ring, and if it is a cytosine, it will forms an azetidine ring.[67,68] Both oxetane and azetidine rings are not stable; they rearrange through ring opening to form the final (6-4) product in each case.[43,69] The quantum yield for the 6-4PP formation is one order less than that of CPD.[26,70] The 6-4PP also causes less distortion of the B-DNA duplex than that caused by CPD formation, it causes only 44 degrees distortion of the B-DNA axis, and therefore, it is still easily detected and repaired by the human exonuclease, preventing it from blocking DNA polymerases during replication.[60]

The order of pyrimidine pair tendency toward 6-4PP formation is different from that found for CPD:6-4PPs are specially formed at TC and CC sites. They are less common in TT and rarely at the CT sites. [44,71,72] The mutagenicity of 6-4PP is extremely high, in particular for the TT derivative. Figure 1.7 depicts a T→C mutation. The 3' thymine of the (6-4) product can bypass the replication process by translesion DNA synthesis (TLS) and tends to pair with a guanine base.[73] A cytosine is incorporated as a complementary base during subsequent replication or repair, resulting in a T→C mutation.

6-4PP has a characteristic absorption band in the UV-A range at approximately 325 nm. When irradiated with UV light above 300 nm, it undergoes a secondary photoreaction, leading to a Dewar valence isomer formation.[74,75] It has also been reported that a comparatively high proportion of the 6-4PP formed rapidly converted into its Dewar valence isomer under solar UV radiation with a relatively high yield of 5–8% in dimeric

samples.[76,77] Dewar proposed in 1867 that the Dewar valence isomer is produced by a ring closure reaction of the (6-4) pyrimidone ring (Figure 1.7) and has a heterocyclic form of benzene bicyclic structure.[78] The Dewar lesion has less mutagenic potential compared to 6-4PP. When encountering a Dewar lesion, the (TLS) DNA polymerases tend to add an adenine base and read it as an abasic site.[79] This repairing mechanism is known as the A-rule and can potentially cause mutations.[80]

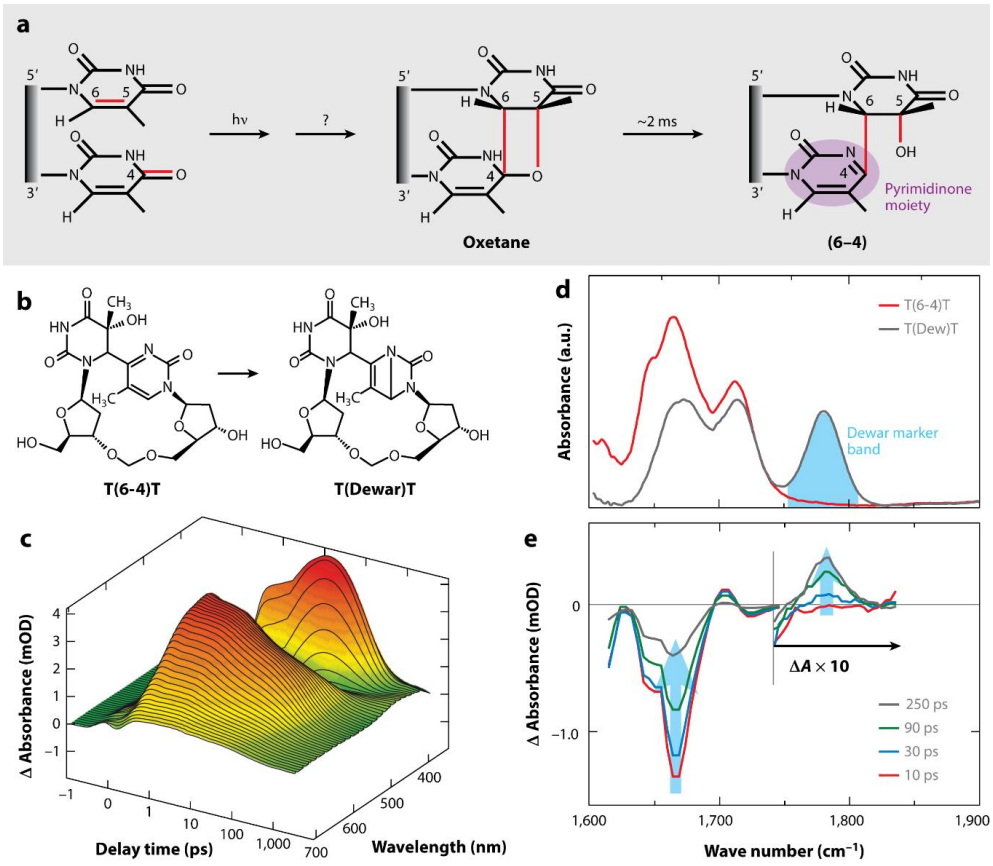


Figure 1.6. a) Formation of the (6-4) lesion between two thymine bases. Bonds broken and formed during the reaction are highlighted in red. An intermediate oxetane has been suggested. The pyrimidinone moiety (lower heterocyclic ring, highlighted in purple shading) of the (6-4) lesion allowed the monitoring of the final step in the reaction. (b) Structures of T(6-4)T and T(Dewar)T. (c) Absorption changes recorded in the UV/visible range after illumination of T(6-4)T by pulses at 325 nm. (d) Stationary IR absorption data for T(6-4)T before UV illumination (gray) and after illumination (red), highlighting the marker band for the Dewar form at 1,780 cm^{-1}. (e) Time-resolved IR experiment showing the decay of the excited electronic state (~1,670 cm^{-1}) and the formation of the T (Dewar)T band at 1,780 cm^{-1} on a timescale of 100 ps. Copyright 2015 by Annual Reviews, and adapted with permission from ref (5).

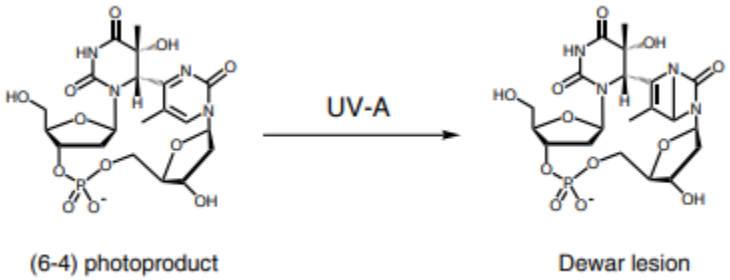


Figure 1.7. Scheme depicting the formation of Dewar lesion from a (6-4) photoproduct.

1.2.2.3 The spore photoproduct (SP)

The fourth type of dipyrimidine photoproduct induced upon UV-C radiation is established in bacterial spores.[44,81,82] The DNA structure in spores is organized in such a way that genetic information remains functional even in extreme UV adverse conditions for extended periods.[83] In bacterial vegetative cells, CPD and (6-4) photolesions are the dominant lesions formed by UV-C irradiation. However, these lesions are only detected in small amounts in spore DNA. In spore DNA, the only type of photolesion that occurs almost exclusively is 5-(a-thyminyl)-5,6-dihydrothymine, arising from two adjacent thymine residues on the same strand.[84] The spore lesion is made up of a bond between a methyl group on the 3'-end thymine and a C5 atom on the 5'-end thymine residue.[82,85] Similarly, when combined UV-B and UV-A irradiation is applied to Bacillus subtilis spores, spore photoproduct remains the primary lesion. References[77] and [86] provide detailed reviews of spore DNA's specific photoresistivity and photoreactivity.

1.2.2.4 Oxidative damages.

For formation of oxidative lesions in DNA, the underlying mechanism is different from those described for dipyrimidine photoproducts. Natural cellular metabolism can create reactive oxygen species (ROS), including the hydroxyl radical ($OH^{\cdot}$), peroxide (O_2^{-2}), superoxide (O^{-2}), and singlet oxygen ($^1O_2$), which eventually can induce oxidative damages. Typically, oxidative DNA damage in cells is considered as a background feature or a part of everyday cellular processes.[87] Under UV-A irradiation, however, cells can also produce higher levels of ROS, which can oxidize DNA nucleobases. The oxidation

potential of guanine is the lowest among the four DNA nucleobases, making it the primary target of ROS. Thus far, there have been more than 20 types of oxidative damage identified.[87] Of these compounds, 8-Oxo-2'-deoxyguanosine (8-oxo-dG) attracts the most attention and is being investigated in depth. This damage can also be represented by 8-OH-dG, in which the keto form 8-oxo-dG undergoes a tautomerization to the enol form 8-OH-dG (Figure 1.8). Researchers have found that approximately 14% of 8-oxo-dG undergo mutations with the G:C to T:A transversion occurring at a frequency of 5.9% being the most significant.[88] Further, 8-oxo-dG concentration is a useful indicator of oxidative stress in humans.[89,90] Oxidative stress plays a crucial role in many diseases and cancers.

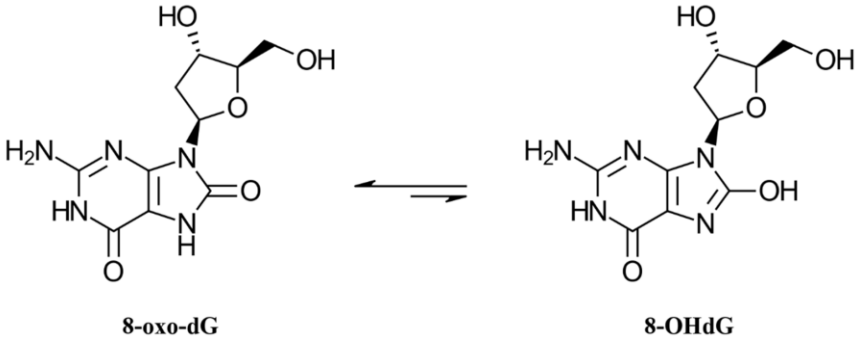


Figure 1.8. The structure of 8-oxo-dG (keto tautomer) and 8-OH-dG (enol tautomer).

1.2.3 Self-repair of DNA photolesions

Despite being strongly absorbed by DNA, UV rays rarely cause permanent damage. Single nucleobases are highly photostable because they can rapidly dispose of UV excitation energy, in the sub-picosecond range, from high-lying electronic states into heat, preventing harmful reactions.[6] On the other hand, there are longer-lived excited states, lifetimes in the 100 ps, found in biologically significant DNA strands.[5,91] These states have recently been identified as excimers, charge separation, or charge transfer states, which are delocalized along DNA strands. The redox potential of the DNA bases involved in this charge transfer or separation determines its direction.[8] In addition, charge transfer states have been observed between DNA lesions and neighboring nucleobases, which significantly influences photochemical reactivity.[92]

Although charge-transfer states are highly reactive and can result in DNA damage[93–101], the charge transfer states can also reverse pre-existing photolesions.[10,102–105]

The question of a related repair mechanism in purely nucleotide-based systems has been largely investigated. In 2004, the first experimental evidence of CPD photoreactivation by a purely nucleotide-based system was demonstrated by Chinnapen and Sen.[106] They discovered a catalytic DNA (or DNAzyme-the UV1C DNAzyme), which was capable of harnessing light of >300 nm wavelength to repair TT-dimers within a substrate DNA (Figure 1.9). The UV1C DNAzyme is a catalytic DNA selected from a random-sequence single-stranded DNA library SELEX (Systematic Evolution of Ligands by Exponential enrichment). Chinnapen and Sen showed that the most plausible mechanism for this unprecedented DNAzyme implies electron (rather than energy) transfer from photoexcited guanine(s) in its G-quadruplex fold to the thymine dimer to trigger self-repair.

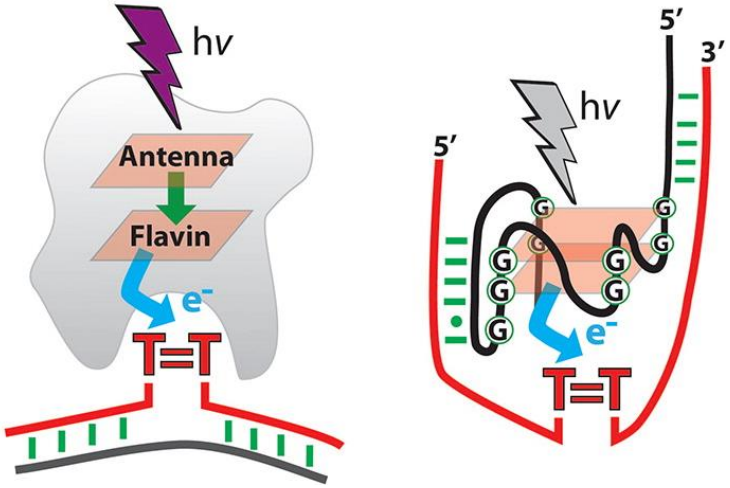


Figure 1.9. Depicts a typical photolyase enzyme versus UV1C DNAzyme.[107] Copyright 2013, American Chemical Society and adopted with permission from ref (103).

Curiously, purely nucleotide-based CPD photoreactivation has since been found to also operate within a standard double-helical DNA[11,108,109], with guanine bases[110], 8-oxoguanine[9,111,112] or guanine-adenine exciplexes[8,10], acting as presumptive photosensitizers for these processes. It has been shown that oxidatively damaged guanine (8-oxoG) transfers electrons to a nearby CPD lesion, causing its repair in small oligonucleotide model systems.[112,113] When (8-oxoG) is integrated into a DNA or RNA strand in proximity to a CPD, it mimics flavin's function in photorepair. 8-oxoG acts catalytically in a mechanism similar to that of photolyase in which a photoexcited purine

(8-oxoG) donates an electron to a pyrimidine dimer to initiate a bond cleavage; subsequent back electron transfer restores 8-oxoG.

An identical mechanism has been proposed for the self-repair of thymine dimers in duplex DNA.[110] Using a defined system of oligodeoxynucleotides, Rokita and his group investigated the effect of purines surrounding dimer sites within DNA duplexes, showing that neighboring guanines promote dimer repair by providing electrons for CPD repair. It was clearly established that 5'-GTTG-3' and 5'-ATTA-3' sequences resulted in the dimer's lowest and highest photostationary levels, respectively. For 5'-GTTA-3' and 5'-ATTG-3', intermediate levels were observed, although the 5'-G was more capable of suppressing CPD levels than the 3'-G. In this study, data indicate that DNA has a naturally occurring ability to promote dimer repair by transient charge transfer from a neighboring G.[110] In contrast, studies have shown that excitation of one nucleobase per se does not promote repair of an adjacent TT-dimer.[114,115]

Likewise, another study demonstrates that self-repair in DNA does exist and proposes a novel mechanism based on excimer (charge transfer) states in DNA.[10] In the beginning, they illustrate no effect of photoexcitation an adenine adjacent to a CPD. On the other hand, excitation of a guanine (G) adenine (A) motif results in photoreversal of CPD into intact thymine (TT) bases. The participation of two bases in the repair suggests a long-lasting charge transfer state between G and A in the repair process. Upon illumination by UV light, an electron migrates from excited guanine (G) to adenine (A), giving rise to the zwitterion $G^{\bullet+}A^{\bullet-}$ formation.[10] During the charge transfer state (300 ps), the negative charge on the A base may migrate to the adjacent thymine dimer. By donating an electron to the CPD, the negatively charged radical $A^{\bullet-}$ anion induces ring splitting and repair. As a result of an electron (charge) transfer, the cyclobutane ring opens, which induces a photolyse-like repair of the CPD lesion.

1.2.4 Photochemistry and charge transfer in nucleic acids.

In a photophysical process, photoexcitation retains the same molecules that were illuminated. In contrast, the photochemical process produces new molecules. Most molecules normally exist in a singlet state in their ground state. Hence, electrons with even numbers are oriented pairwise with opposite spin in the molecule's orbitals (total spin number S = 0, multiplicity 2S + 1 = 1, hence “singlet”) [1]. Upon absorbing light, electronic

excited states can become occupied. Transitions fall into two categories: transitions that maintain net spin angular momentum ("allowed") and those that change the above ("forbidden"). The total spin number stays the same after following a spin-allowed transition, but it flips/changes following a spin-forbidden transition. According to the quantum mechanical selection rule, the transition from the singlet state to the triplet state is forbidden. However, high-energy spin-orbital interactions can lead to spin-forbidden transitions. In the singlet ground state (S_0, μ 00), molecules can be excited optically to the higher electronic singlet states (S_1 or S_n, μ 0) [2]. Because the vibrational levels of the same atomic state are weakly coupled, the higher vibrational levels (μ 0) are often occupied before relaxing fast to the lower vibrational levels. In order to dissipate the vibrational energy, the energy can be rapidly redistributed intramolecularly within molecules, as well as it can be transferred from molecules as heat to the surrounding medium. This process cools the hot vibrational molecules. Kasha's rule states that the higher singlet states (S_n) are deactivated rapidly and radiationless via vibrational relaxation into the first excited state (S_1).[116] Afterward, either fluorescence emission or radiationless depopulation can return the S_1 state to the ground state S_0. Radiationless processes can be broken down into spin-allowed internal conversions (IC) and spin-forbidden cross-system processes (ISC). Internal conversion describes the transition between states with the same multiplicity, whereas intersystem crossing refers to a spin-forbidden transition from the singlet to the triplet state. Usually, the triplet state cannot be directly excited from the ground state; hence it emits the long-lasting phosphorescence or decomposes through radiationless decay to the ground state. A Jablonski diagram for the transitions mentioned above and the various states are shown in Figure 1.10. Those various deactivation processes play an important role in the high photostability of nucleobase and DNA molecules.

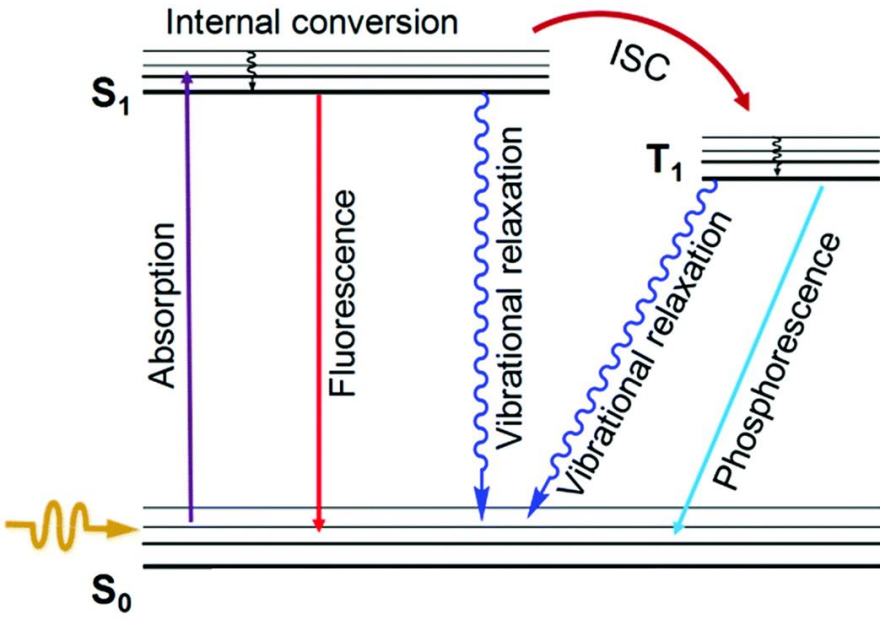


Figure 1.10. A typical Jablonski diagram showing the possible radiative and non-radiative transitions in a fluorescent system.

The excited state of isolated bases, such as nucleotides, is relaxed within a few picoseconds. There are conical intersections between the vibrational manifold of the excited state and the ground state that result in such ultrashort relaxation times.[117] Nonradiative relaxation can happen in nucleobases with specific distortions, such as out-of-plane motions. However, a time-resolved study of UV-excited oligonucleotides in the region of infrared (IR) reveals that the excited states of oligonucleotides and duplexes have much longer-lived excited states and last for hundreds of picoseconds.[8] Stacking of bases has been implicated in generating new decay pathways for monomer excitation, which show longer-lived transient absorptions of oligonucleotides with more extended excited states. Various mechanisms to explain these long-lived excited states have been proposed, including charge transfer between adjacent bases, neutral excimers, and delocalized charge-separated domains. Charge-transfer states among bases are based on the sequence of the nucleobases itself, with excess electrons being placed on more easily reduced bases and holes (absence of electrons) being placed on more easily oxidized bases.[10] Based on the spectra of small molecules mimicking base-paired nucleotides, it has been shown that an additional proton transfer can occur via photoexcitation across the Watson-Crick face.[8] When oligonucleotides participate in a duplex, both electron and proton-coupled electron transfers are implied to describe excited states seen over hundreds of picoseconds. There

is no doubt that light can induce persistent excited states in oligonucleotides, single strands, duplexes, and many other structures. Besides the photo-induced excitons in pure nucleic acids, the nucleobases can also be reversibly oxidized and reduced by small molecules, covalently attached molecules, or other adjacent bases.[118]

Studies on the individual bases have shown that guanine is the easiest nucleobase to oxidize, and this property also persists in oligonucleotides. Thymine has the highest reduction potential of all DNA bases, so it is the primary carrier of excess electrons injected into oligonucleotides or duplex**.**[119] Uracil differs from thymine only in lacking a methyl substituent at position 5, making it more difficult to reduce. DNA electron holes can often be monitored using the tendency of guanines to form oxidized 8-OG, which can then be chemically or enzymatically cleaved and visualized on a denaturing gel.[120] A given redox potential in an oligonucleotide is affected by the interactions between neighboring bases, such as when multiple guanines are present. An electron-hole can hop between adjacent bases more easily when an oxidized base is introduced in a duplex or an oligonucleotide. The oxidation potentials of the bases also increase relative to the NHE, with the order G > A > C > T, so guanine is the base most likely to host an electron-hole. As a consequence of pi-pi stacking, the electron-hole in a duplex can also delocalize across multiple bases. Researchers have used a similar strategy to observe excessive electrons in a duplex using tethered photoreductants.[121] Electron hopping from thymine base to thymine base has been shown for these, though at a lower rate than the migration of holes.[121]

1.2.5 Effect of DNA Methylation (5-Methylcytosine) on CPD formation.

Cytosine methylation is an epigenetic modification of DNA that controls gene expression, and it is required in vertebrates for normal embryonic development.[122,123] The heterocyclic aromatic ring of 5-methylcytosine contains a keto group at the C-2, an amine group at the C-4, and a methyl group at 5 positions. Therefore, there is a close structural similarity with the heterocyclic aromatic ring of thymine, which consists of two keto groups at C-2 and C-4 and a methyl group at C-5. The base pairing of 5-methylcytosine with guanine is not stable, and 5-methylcytosine can undergo spontaneous deamination to thymine.[124,125] The methylation of DNA enhances the formation of CPDs. It is well established that CPDs are more likely to occur at dipyrimidine sites containing methylated cytosine than at the

identical unmethylated sites after UVB exposure.[37,57,126–130] The reason for this is that methylated cytosine absorbs light at a wavelength of 273 nm and unmethylated cytosine at 267 nm.[29] According to Douki and Cadet, monophosphate dinucleosides containing a single 5-methylcytosine show a higher dimerization rate than that dinucleosides containing a single cytosine.[37] Similarly, Mitchell estimated that CPD occurred about two times more often in poly(dG): poly(m5dC) than in poly(dG): poly(dC).[127] Pfeifer and colleagues observed that irradiating human cultured cells with natural light significantly raised the yield of CPDs at mutational hot spots on the p53 gene by 15-fold when 5-methylcytosine bases were present.[126] A modeling study postulated that the existence of C5-methylation affected the confirmation equilibria, thereby, the electronic excited states of duplex DNA leading to an increase in the quantum yield of CPDs and reducing that for (6-4) photoproduct.[131]

1.2.6 Cyclobutane pyrimidine dimers Formation: singlet or triplet states involved?

1.2.6.1 Direct CPD formation

As early as in the 1960s, CPD lesions had already been discovered. For years, however, the mechanism of their formation remained dubious. For direct UV-C and UV-B excitation, early experimental studies supporting both the triplet and singlet pathways involved in CPD formation were reported.[132] Since thymine has the lowest triplet energy between the four DNA nucleobases, it has been suggested that the triplet states will be localized on thymines, and energy will be transferred from higher triplet states to lower triplet states. It has been often suggested that the triplet state is a precursor for the formation of the CPD lesions even though thymidine monophosphate (TMP) offers a low intersystem crossing efficiency upon UV excitation (1.4% at 266 nm). Thymine is the DNA base with the lowest triplet energy.[133] Moreover, the triplet energy of thymine base in dimer TpT or oligomer (dT)18 falls below that of monomer TMP due to the base stacking effect.[134] The thymine bases, therefore, become the primary target of photosensitization through triplet-triplet energy transfer (TTET) in DNA strands, which can lead to CPD lesions.[135,136] However, it has been demonstrated unambiguously that the formation of CPD lesions occurs within 1 ps primarily due to the excited singlet state.[137,138] Singlet pathways are strongly supported by recent advances in time-resolved methods and accurate

computational calculations. Yasui *et al.* conducted time-resolved infrared experiments to illustrate that single-strand oligodeoxynucleotide (dT)18 can form CPDs within 1 ps of being exposed to UVB light.[137] The reaction is believed to proceed through a singlet pathway during direct excitation with a minor contribution from triplet states (Path A, Figure 1.11).[139,140] UVC irradiation, primarily at 254 nm used most commonly in studies, can also photo-reverse the initially formed CPDs by a retro [2+2] reaction (Path B+C, Figure 1.11 A). It has been found that CPD distribution appears to depend on both time and wavelength of irradiation and flanking sequences.[114,141] Short irradiation times are characterized by CPD formation that is dominated by the forward rate constant, k_f at photostationary state. On the other hand, long irradiation times are characterized by photoproduct distributions corresponding to $k_f/(k_f + k_r)$ at photostationary state.[114] In addition, it has been suggested that CPDs might be photoreversed by electron transfer from a photo-excited adjacent base, mainly G, to produce an intermediate radical ion pair, which causes repair of the CPD (Path D, Figure 1.11 A). As an alternative, the excited state of the CPD can remove an electron from an adjacent base to produce the same radical ion pair (Path E, Figure 1.11 A). In particular conformations, photoinduced electron transfer from an adjacent base may also contribute to the photoreversal of CPDs, as observed in some G-quadruplexes (Path D, Figure 1.11 A).[106,142] Given the short time required to bring stacked pyrimidine bases into a photoreactive conformation, the ground state conformation of stacked thymine bases at the moment of photoexcitation is supposed to be the true determining factor for dimerization. Accordingly, the low quantum yield of CPD is a consequence of its low photoreactive conformations. Nevertheless, the process by which CPDs form under UVA irradiation is still not fully understood. DNA poorly absorbs UVA light, so it has been proposed that UVA-induced CPD may result from either direct excitation or photosensitization mediated by the excitation of endogenous or exogenous chromophores.[26,143–147]

With advances in spectroscopy and computational methods, a growing number of studies have shown that collective excited states (excited states located over two or more bases) are involved in the formation of CPDs.[148–151] A delocalization occurs as a result of their close proximity that leads to electrons interaction between the bases. The degree of delocalization depends heavily on the relative positions of the bases and the context of the

surrounding sequence. Frenkel excitons (delocalized π-π* states) and charge-transfer (CT) states are the two preconditions of collective excited states. Frenkel excitons are thought to be responsible for the formation of CPDs.[152] Moreover, studies have shown that the transition from Frenkel excitons to CT states provides a decay channel for escaping the formation of CPDs.[115]

Another significant finding concerns so-called "dark CPDs," in which CPDs are generated in melanocytes for three or more hours after exposure to UVA.[153] As we know, UVA is a component of the radiation in sunlight and tanning beds**.** The researchers also declared that the "dark CPDs" contained more C-containing CPDs than the CPDs induced by UVA alone. Dark CPDs reactions cannot be understood by direct excitation or photosensitization mechanisms, which happen immediately after UV photons are absorbed. Thus, the authors proposed a chemiexcitation mechanism in which peroxynitrite, produced by the reaction of UV-induced superoxide and nitric oxide, reacts with melanin fragments to form high-energy dioxetane (Path B Figure 1.11 B). After that, a triplet energy carbonyl is formed from the decomposition of dioxetane, transferring its triplet energy to DNA to produce CPDs.[25,153–155]

Various methods have been used to study *in vitro* the frequency of CPD formation in duplex DNA by direct irradiation. In an early study, Carrier *et al*. studied the DNA containing tritiated thymine or cytosine, irradiated at two UVC regions, was hydrolyzed by acid to release the dimerized bases in the form of T=T, T=U, U=U, and U=T. Then, it was separated onto paper chromatography and quantified, except for T=U and U=T, which are hard to differentiate.[156] When they were irradiated at 265 nm, their TT:TC+CT:CC CPD ratio was 62:25:13, and at 280 nm, the ratio increased to 72:18:9 after correction for dinucleotide frequency. Modern approaches use enzymes to digest DNA into photoproduct-containing dinucleotides followed by separation and quantification by high-performance liquid chromatography coupled with electrospray tandem mass spectrometry (HPLC-MS/MS).[44,48] In this method, at broadband UVB irradiation (280–370 nm), the pre-steady state ratio of CPD formation for TT, TC, CT, and CC was 35:39:19:7, but changed to 42:41:15:02 upon 254 nm irradiation.[48] Pyrimidine's CPD formation was also evaluated by analyzing data obtained from genomic DNA using broadband UVB light show an average ratio of 36 ± 4: 32 ± 3: 20 ± 5: 13 ± 2 for TT:TC:CT:CC CPD formation.[157] Studies

have also been conducted on CPD formation by analyzing cleavage sites of specific pyrimidine dimer glycosylases, such as *M. luteus* UV specific endonuclease[158] and T4 pyrimidine dimer glycosylase[159] using end-labeled DNA, NextGen sequencing, and ligation-mediated PCR. One study using T4 endonuclease V digestion has shown that the yield of photoproducts at specific sites depends on both the nucleotide composition and the wavelength of the UV light used to induce the damage.[13] Different doses of light, sequence coverage, and analysis methods probably contributed to the variation in CPD ratios among the various studies. Earlier literature reported an increase in CPD formation at C-containing sites upon using longer UVC wavelengths, which can be explained by the use of larger doses of UVC light that lead to a decrease in photoreversal efficiency.[114,141,160] It is also possible that the increased CPD formation at TC sites is due to a decline in the competitive formation of TC (6-4) photoproducts. The quantum yield of TC (6-4) decreases with increasing wavelength compared to CPDs.[139]

1.2.6.2 Formation of CPD through photosensitization

In addition to direct UV excitation, CPDs can also be produced by photosensitization through the triplet-triplet energy transfer pathway (TTET). In a TTET process, UV photons excite the photosensitizer (an organic molecule) to a singlet state, which undergoes intersystem crossing (ISC) to enter its triplet state. In the presence of sufficient excitation energy and close vicinity to a dipyrimidine, a triplet excited molecule transfers its energy to DNA by triplet-triplet energy transfer (TTET) to generate a CPD (Figure 1.11).[26,140,161–164]

A TTET process involves a double electron exchange known as the Dexter mechanism.[165–167] TTET transfer can proceed through collision complexes (Path C) or intercalated intermediates (Path D). TTET's efficiency correlates with the triplet energies between photosensitizers and DNA bases, the coupling interaction between them, and how much the photosensitizer binds to DNA (nature of photosensitizer complexation).[134,165,168–172] Several fluoroquinolones[162,170,173] and ketone derivatives[172,174] with triplet states higher than those of pyrimidines are commonly used to sensitize CPD formation.[134] An excited triplet pyrimidine can then develop a single bond from C6 to C6 with the neighboring pyrimidine, resulting in a triplet biradical intermediate (^{3}BR). Afterward, the triplet

biradical may undergo intersystem crossing to the singlet state (^{1}BR) and proceed to either the CPD formation or reverse back to the dipyrimidine pair.[175,176]

Multiple studies on TTET induced by various photosensitizers in DNA have shown that TT CPD remains by far the most common photoproduct, followed by TC and CT with at least one order of magnitude lower yield, and CC-CPD rarely formed.[135,165,177–180] It has been demonstrated that flanking bases impact the yield of CPD generated via TTET, and the ratio of CPDs formation between TT, TC, CT, and CC differs for different photosensitizers.[180] The results of these studies indicate that further investigation of the mechanism of TTET-induced CPD is needed. It is possible that there is a more complex mechanism at work rather than the simple transfer of triplet energy to a single thymine base.[165]

In a new study, Costalat *et al.* used gel electrophoresis and T4 endonuclease V digestion, all 64 possible NYYN sequences were quantified in a 129 bp DNA duplex.[140] With UVB light, CPD formation varied by 2-fold between dipyrimidines and by 12-fold with flanking sequences. It was most abundant at YYYR sites and less frequent at GYYN sites, consistent with charge transfer. Conversely, photosensitized CPD formation placed a considerable preference on TT over C-containing sites, particularly for norfloxacin (NFX) than acetone, in agreement with their differences in triplet energies. The pattern of sequence dependence for photosensitized TT CPD formation is similar to UVB light; however, there are notable differences between NFX and acetone that may be due to the DNA intercalation of NFX.

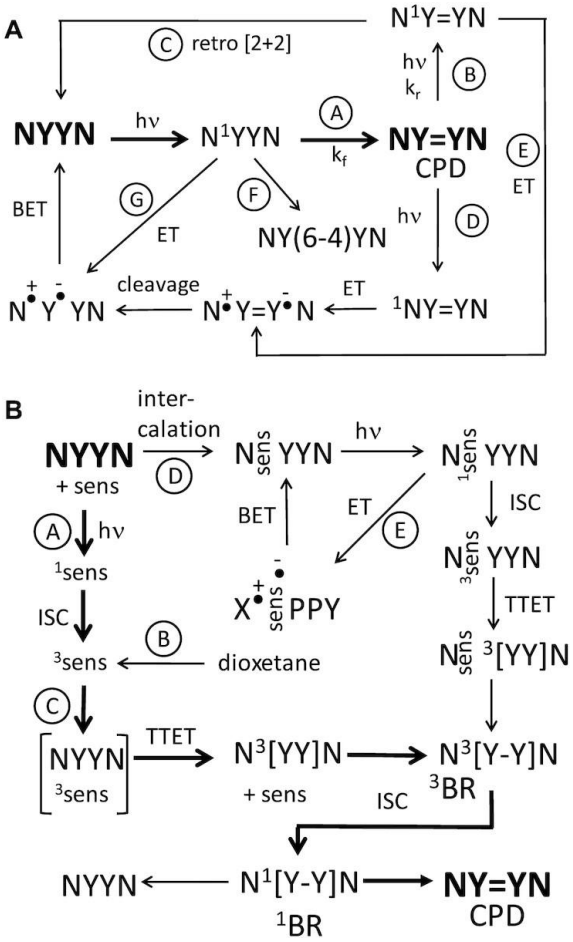


Figure 1.11. A depiction of CPD formation by the direct and photosensitized methods. Copyright 2021, Oxford University Press, and adapted with permission from ref (126).[140] A multi-step mechanism represents the formation and photorepair of CPDs, where a direct excitation (path A) yield a singlet state of a pyrimidine (Y), which undergoes a [2+2] cycloaddition ('=') with an adjacent pyrimidine. The CPD is unstable at 254 nm and could reverse back to monomers by its direct excitation (path B) followed by a retro [2+2] reaction (path C). A possible pathway (path D) is also proposed where the transfer of an electron through an excited flanking base could lead to the reversal of CPD. Another pathway (in which an electron transfer from a flanking base to an excited CPD may lead to a charge transfer intermediate (path E)). The rate of CPD formation can be diminished by competitive formation of a (6-4) photoproduct from the excited singlet state (path F) and the excited singlet state could also be deactivated by electron transfer from a flanking base pair (path G) followed by back electron transfer (BET) to the ground state.

1.2.7 Next-generation sequencing- based methods to detect DNA UV damage

Several radiolabel-based methods have been developed to detect DNA damage and repair with single-nucleotide molecular resolution. However, they are limited to small genomic regions and rough or approximate quantitative detection. Many DNA damages

can be at once identified and quantified using mass spectrometry-based DNA adductomic techniques, but these techniques do not include genome sequence information.[181–184]As a result of the advent of NGS technology, genome-related research has been transformed with superior data accuracy, reduced costs, and high throughput. Recently, a growing number of NGS-based DNA adductomic methods are being developed to detect, characterize, and quantitate environmental induced DNA damage and repair at the genome-scale.[185–188] In order to measure DNA damage and repair after exposure to environmental hazards, the standard NGS procedure must be altered since the covalent modification of DNA bases causes blockage of the PCR amplification step.[189] Currently, three fundamental strategies are employed to overcome this obstacle in the field of NGS-based DNA adductomics. In the first strategy, DNA fragments with DNA damage are captured by immunoprecipitation or biotin-streptavidin enrichment. In one round of primer extension before PCR amplification, a translesion DNA polymerase is used to directly repair or bypass DNA damage. The second method involves creating a nick at damage sites by enzymatic or chemical treatment and ligating a sequencing adaptor to prepare an NGS library. The third approach makes use of both DNA damage immunoprecipitation and high-fidelity DNA polymerase blockage during primer extension to enrich DNA damage and pinpoint its location in the genome.

1.2.7.1 Enriching DNA damage after damage reversal or using bypass techniques on damaged DNA

The strategy of this approach involves enriching DNA fragments containing DNA damage through immunoprecipitation or biotin-streptavidin binding. Then UV-induced DNA damage can be directly reversed by repair enzymes or chemical reactions. For other damage that cannot be reversed, they can be bypassed by translesion DNA polymerases in one round of primer extension before PCR amplification of the NGS library. CPDs were detected in the genome using damaged DNA immunoprecipitation and next generation sequencing (DDIP-seq),[190] as well as excision repair sequencing (XR-seq).[191] The (DDIP-seq) can detect DNA damage across the entire genome, but they produce low resolution are usually 100 to 1,000 base pairs in size. The excision repair sequencing (XR-seq) method relies on the fact that the excised oligonucleotide in the cell was released in complex with TFIIH and XPG, which cuts the damaged strand at the 30 nucleotides away

from the lesion. After ligation with adapters, the incised oligonucleotides are purified further using specific DNA damage immunoprecipitation. Photolyases reverse the UV-induced DNA damage on the adapter-ligated excised oligonucleotides before PCR amplification.[191,192] Sequencing reads from XR-seq are aligned with the genome, and damaged areas can be identified according to the dual incision method of nucleotide excision repair. An improved version of XR-seq called translesion excision repair sequencing (tXR-seq) shares the same DNA damage enrichment process, but they differ in pre-PCR step treatment.[193] During the one-cycle primer extension, tXR-seq bypasses CPD-DNA damage with human translesion synthesis polymerases η.

1.2.7.2 Nick creation and ligation-based approach

This strategy uses special endonucleases or chemical reactions to induce nicks at damage sites before or after fragmenting genomic DNA and then ligates an adapter to 5' or 3' ends of the nick site by using ligases prior to PCR amplification and sequencing. This approach has led to the development of a variety of methods to identify DNA damage, such as UV damage, alkylation damage, oxidative damage, ribonucleotide incorporation, and chloroplast-DNA adducts in different types of organisms. An example of UV damage detection by this approach is the Excision-seq, which uses E. coli UDG and T4 Endo IV to create a nick at uracils in DNA, CPD, or (6-4) PP sites.[194] Despite its high resolution, this method requires a high degree of DNA damage for a library building, limiting its usability. In this method, for instance, the UV dose is 10,000 J/m^2, which is much higher than has been usually used in most studies. Another method termed CPD-seq was developed to detect CPD damage sites across the entire genome with single-nucleotide resolution in yeast.[195] The principle of this method relies on creating a nick and a ligatable-OH group upstream of CPD sites using T4 PDG and APE1 enzymes. Afterward, the 5'-Phospate group is excised, and the second adapter is ligated to DNA fragments that contain 3'-OH groups. In the following steps, the ligated products are amplified by PCR and then sent for next-generation sequencing. The drawbacks of this method are the background signal that can be detected in nonirradiated DNA, which is produced due to DNA breakage during extraction and purification, and inefficient adaptor ligation.[195] Therefore, the CPD-seq method might be challenging to use for identifying damage with relatively low CPD levels.

1.2.7.3 Primer extension-based strategy for DNA damage enrichment

An example of this strategy is damage-seq, which employs damage-specific antibodies to enrich the products of sonicated DNA that have been UV damaged and ligated to the first adaptor. Afterward, biotin tagged primers are extended by Q5 DNA polymerase, which will stop before the damage site. Then sequencing will be done after ligating the primer extension products to a second adaptor and amplified by PCR. The (HS-Damage-seq) method improved the detection sensitivity of this technique by introducing a subtractive hybridization step in the original Damage-seq procedure to sequester the undamaged DNA strands prior to PCR amplification.[187]

Chapter 2. A novel method to investigate the effect of flanking bases on the formation and repair of thymine dimers in duplex DNA

2.1 Abstract

The cis-syn Cyclobutane Pyrimidine Dimers (CPDs) are the most common lesion formed in cellular DNA from exposure to solar light. Although many studies have been carried out to identify the influence of DNA sequence and structure on its photochemical and photophysical properties, the molecular bases of these observations have not yet been well understood. This research has been conducted to comprehensively study the comparative impact of neighboring bases on either side of a pyrimidine pair on CPD formation and photo-repair of a pre-existing CPD in double-stranded DNA. The principle of our approach relies on precise blockage of Taq DNA polymerase at pyrimidine dimers (CPDs), which cause the disappearance of the sequences that contain a thymine dimer from the irradiated DNA library pool. High-throughput sequencing is then done to explore the formation and repair of CPDs over the time-course experiment at different wavelengths. We are using a random double-stranded DNA library consisting of a stretch of 10 randomized base pairs surrounding a central thymine pair (5 random base pairs on either side), in turn flanked by constant bases as primer binding sites for PCR amplification. The constant regions were designed to avoid two adjacent pyrimidines. The sources of UV irradiation used in this research are a monochromatic 285nm LED and a 365 nm LED, the latter used in conjunction with a triplet sensitizer co-dissolved with the DNA to be irradiated. Our results show that the TTTT and ATTT are always the hottest spots for CPD formation and persistence in NTTN motif sequences, and GTTG is the coldest spot (i.e. least CPD persistence) in the two different irradiation setups. In agreement with previous studies, we observed a slightly higher suppressive effect of 5' flanking guanine on TT-CPD formation than 3' flanking guanine with all other bases rather than thymine. In addition, G-rich flanking motifs (GCG, GTG, GGG) show the lowest T^T CPD formation, while AT-rich motifs (AAT, TTT, ATT) show the highest formation of thymine dimers. There is no apparent direct correlation between the 5' side and 3' side on the formation and repair of TT-CPDs in dsDNA. It seems to be that the mechanism of both formation and repair of

CPDs from the 5' and 3' sides are identical, and fluctuation of different motifs found on the 5' side are almost a mirror image to that found on the 3' side.

2.2 Introduction

Solar ultraviolet light is well known to induce dipyrimidine photoproducts, which play a crucial role in cellular photodamage involved in skin cancer.[24,196] Dipyrimidine photoproducts fall into two main classes: cyclobutane pyrimidine dimers (CPDs) and the less abundant but more mutagenic pyrimidine (6-4) pyrimidone photoproducts (6-4PPs).[28,197] CPDs can occur at any dipyrimidine site by a [2π+2π] photocycloaddition between the C5-C6 double bonds of two neighboring pyrimidine bases. As a consequence of the stereochemical constraints imposed by the structure of native B-form DNA, only the cis-syn stereoisomer of the CPD has been observed.[44] Cyclobutane Pyrimidine Dimers (CPDs) are the most significant lesions in DNA brought about by exposure to solar ultraviolet B and A radiation. These lesions pose a severe challenge to DNA replication in the cells; defects in cellular thymine dimer repair machinery also seem to be associated with human skin cancers and Xeroderma pigmentosum. Although the T^T CPD photolesion is not strongly mutagenic in general, it can block DNA polymerases and result in DNA replication forks stalling and also transcription inhibition.[20,189] CPDs are particularly important to skin cancer because they are formed more rapidly than 6-4PP and repaired more slowly, thereby increasing their carcinogenic potential.[198] A recent study found that photoinduced CPDs can also be generated after exposure to UVA and UVB exposure for 2 to 3 h via chemisensitization pathway.[153]

In order to understand CPDs formation, it is crucial to understand their photochemistry. Many factors affect the effectiveness of CPDs formation. At first, the dimerization is determined by the wavelength of UV light, the intensity of the light, and the type of two pyrimidines involved. The UVC (190 to 290 nm) and UVB (290 to 320 nm) rays excite DNA directly to form CPDs.[13,48] Under UVB and UVC irradiation, Mitchell *et al.* quantified the CPDs formation at different pyrimidine bases using sequencing gel and T4 endonuclease V enzyme.[13] The authors estimated that the ratio of CPD formation at TT, TC, CT, CC sites were 52 : 21: 19 : 7 under UVB and 68: 16: 13: 3 under UVC irradiation, respectively. A comparable result is obtained by a study that used high-performance liquid chromatography coupled to electrospray tandem mass

spectrometry (HPLC-MS/MS) to investigate CPD distribution among naked DNA and *in cellulo*.[48] Cadet *et al.*, in 2001 found that UVC-induced CPD distribution in naked DNA was 2.970: 1.823: 0.573: 0.069 (55: 34: 11: 1) for TT, TC, CT and CC while UVB-induced CPDs in the following order: 1.023 : 0.694 : 0.289 : 0.094 (49 : 33 : 14 : 4) in naked DNA and 3.147 : 0.279 : 1.289 : 0.577 (60 : 24 : 11 : 5) *in cellulo*.[48] These studies have shown that CPDs form more frequently at TT than TC, CT, and CC. In addition, UVB causes more CPDs with cytosine than UVC. These studies have discrepancies because of the different DNA sequences, genomes, and UV sources used. Furthermore, there is no systematic information on the influence of bases beyond the two pyrimidines, even though numerous other studies have provided empirical evidence that the 5' and 3' bases flanking the dipyrimidine sites can play an important role in CPD formation. Accordingly, CPD formation has been observed to be augmented by flanking pyrimidines while adjacent guanines suppressed CPD formation.[13,115,141,199–202] A recent study by Taylor *et al.,* in 2021 examined all 64 possible NYYN sequences in a 129-bp DNA duplex using gel electrophoresis and T4 endonuclease V digestion. However, including all the possible 64 NYYN in one piece (design) of DNA can create bias from many aspects.[140] First, the dimer formation at a given site will prevent the formation of another dimer on the adjacent side, creating a bias due to predominant TT-CPD formation over the other dipyrimidine CPDs. Since the approach used T4 endonuclease V digestion of ^{32}P radiolabeled irradiated DNA, with the products run on a denaturing gel, only a rough quantitative estimation of CPDs is possible. Third, all the samples were irradiated at only one-time point, and the chosen time ensured that at least 60% of the DNA remains uncut. The chosen time may or may not reflect the photostationary state nor fluctuation of formation and repair of CPDs over time.

Studies of the formation of CPD from UVA (320 to 400 nm) irradiation are fewer because direct CPD formation is highly inefficient. CPDs were found to produce in more proportion than oxidative lesions like 8-oxo-7,8-dihydroguanine (8-oxoGua) and this to be the primary form of UVA damage found in both human skin and cultured cells.[144,146] The formation of CPD by UVA has a quantum yield at least two orders of magnitude lower than that by UVB and UVC and about five times as efficient as 8-oxoGua. Considering that DNA weakly absorbs UVA photons, it is possible that UVA-induced CPD formation could occur either by direct excitation or by photosensitization through the excitation of

endogenous cellular components.[26] UVA photoproduct distribution differs substantially from UVB and UVC, with TT CPD corresponding to 90% of all dipyrimidine products, followed by TC and CT, and a minimal CC.[48]

UV rays are absorbed strongly by DNA but seldom cause significant damage. The photostability of single nucleobases arises from their ability to dissipate the UV excitation energy in ultra-fast fashion, in the sub-picosecond range, from high-lying electronic states into heat, preventing harmful reactions.[6] There are also longer-lived excited states, lasting a few hundred picoseconds, found in biologically significant DNA.[5,203] These states have recently been identified as excimers, charge separations, or transfer states delocalized along DNA strands. The charge transfer or separation phenomena are governed by the redox potentials of the bases involved.[8] The charge transfer states between DNA photolesions and neighboring nucleobases have also been observed, significantly influencing its photochemical reactivity.[92] The charge-transfer states are highly reactive and can cause DNA damage, yet they can also reverse pre-existing photolesions.[10,102–105] In purely nucleotide-based systems, the question of a related repair mechanism has drawn considerable attention. It has been proposed by different labs (Rokita,[110] Burrows,[9] and Carell[10]) that the nucleotide-mediated photoinduced self-repair process of CPD may be important in standard duplex DNA, as a constituent of a DNA's 'in-built' mechanism for self-preservation. Rokita *et al.* have demonstrated the influence of purines surrounding dimer sites within DNA duplexes, showing that neighboring guanines promote dimer repair by acting as electron sources for CPD repair.[110] Recently, Carell *et al.* have shown that charge transfer/excimer states between two photoexcited bases (G-A tandems) can migrate to the adjacent CPD lesion, leading to its repair.[10] The high repair effectiveness of neighboring G-A tandems is attributed to a long-lived charge transfer state between these two bases, in other words, a guanine-adenine exciplex. Although many studies have been carried out to identify the influence of DNA sequence and structure on its photochemical and photophysical properties, the molecular bases of these observations are not well understood yet. Furthermore, the neighbouring deoxynucleotide sequence affects the energy and lifetime of singlet and triplet excited states through excimer formation and perhaps delocalization, but their effect on thymine dimer formation is not well known[18,19]

Our research has been conducted to comprehensively study the comparative impact of neighboring bases on either side of a thymine pair on CPD formation and/or photorepair of a pre-existing CPD in double-stranded DNA. The principle of our approach relies on precise blockage of Taq DNA polymerase at pyrimidine dimers (CPDs), which causes disappearance of the sequences that contain a thymine dimer from the irradiated DNA library pool. High-throughput sequencing is then done to explore both the formation and repair of CPDs as a time-course experiment, at chosen wavelengths. We used a randomized double stranded DNA library consisting of a stretch of 10 randomized base pairs around a central thymine pair (the thymine pair has five random bases on either side). The above twelve base pairs are flanked by stretches of known sequence, which can serve as primer binding sites for primer extension or PCR amplification. These constant region stretches (51 bases on either side) were carefully designed to not contain two adjacent pyrimidines. The sources of UV irradiation being used in this research are a monochromatic 285 nm LED and a 365 nm LED for use in irradiation experiments with 2′-methoxyacetophenone (2-M), a triplet sensitizer. Our study is unique and unprecedented in terms of it being carried out on authentic dsDNA that incorporates a relatively large number (five on each side) of random nucleotides flanking the central thymine pair or thymine dimer, all being studied at once. In addition, we have used both UVC and UVA monochromatic LEDs to create/repair CPD dimers, which is more comparable to CPD formation due to solar exposure in biological systems.

2.3 Materials and Methods

2.3.1 Materials

Single-strand random DNA library and reverse primer were purchased from Integrated DNA Technologies (IDT), [γ-^{32}P] ATP from Perkin Elmer, T4-PDG (pyrimidine dimer glycosylase), previously known as T4 endonuclease V, from New England Biolabs (NEB), AMPure XP beads from Beckman Coulter, Qubit™ 1X dsDNA HS Assay Kit from ThermoFisher Scientific, and 2′-methoxyacetophenone from Sigma Aldrich.

2.3.2 Preparation of cold and ^{32}P-labeled double strand random DNA libraries

A non-radiolabeled (cold) double-stranded random DNA library was prepared by one-round reverse primer extension as follows: Multiple tubes of 100 µL primer extension reaction mixture contained 1x KAPA Hifi Taq DNA polymerase buffer (from Kapa Biosystems), dNTP mix 0.2 mM, single-strand DNA library 0.2 µM, reverse primer 0.4 µM, KAPA Hifi DNA polymerase (5 U/µL) 1.25 Unit and PCR grade water to 100 µL. The amplification condition was one cycle as follows: initial denaturation at 95 ℃ for 2 min, denaturation at 95 ℃, annealing at 54 ℃ for 30 s and extension at 72℃ for 5 min. The double-stranded random DNA library so generated was then purified by (0.8x) AMPure XP beads and kept dissolved in (20 mM NaCl/ 10 mM Tris-HCl, pH 8.0) buffer. The 5' ^{32}P-radiolabeled double-stranded random DNA library was prepared under the same condition, except the 5' ^{32}P-radiolabeled single-stranded random DNA library was used in the primer extension reaction mixture. Also, the 5' ^{32}P-radiolabeled dsDNA randomized library was purified by 8% polyacrylamide native gel instead of AMPure XP beads. Library quality control: following purification the successful double-strand library amplification was checked by measuring DNA concentration of the purified libraries using Qubit, which precisely measure the concentration of double strand DNA and recommended for any library measurement before NGS.

2.3.3 Selection of an optimal photosensitizer at 365 nm irradiation

There are 4 main criteria to select an appropriate photosensitizer at UVA range. The high intersystem crossing (ISC) efficiency is the most crucial aspect. Second, it should have triplet energy higher than the thymine pair (TpT). In addition, UV-A absorption should be strong in the UV-A range, and it should be readily soluble in water.

We tested two photosensitizers for 365 nm irradiation, the commonly used acetophenone (AP) and newly characterized acetophenone derivative 2-methoxyacetophenone (2-M).

2’-Methoxyacetophenone (2-M) has been demonstrated as an efficient promising photosensitizer at the UVA range.[204,205] 2-M has a higher extinction coefficient over the UVA range than AP. 2-M undergoes very efficient intersystem crossing with quantum yield (φISC = 97. 4%) after excitation and has a low fluorescence yield η = 2. 6%, suggesting an efficient non-radiative decay process.

Our result demonstrates the superiority of 2-M over AP as an efficient photosensitizer with 365 nm irradiation. A lower concentration of 2-M (2.5 mM) is required for photosensitization compared to AP (10 mM), and the DNA library shows minor background damage when run into denaturing gel (Figure S1). Therefore, we carried out all 365 nm irradiation using 2-M as a photosensitizer (2.5 nM final concentration).

2.3.4 UV irradiation

The irradiation solutions were each 200 µL (20 mM NaCl/ 10 mM Tris-HCl, pH 8.0) buffer solution containing either cold or 5'-radiolabeled 114 bp randomized DNA library, along with 2.5 nM 2′-methoxyacetophenone if 365 nm irradiation used. All irradiation was carried out in a cold room at 4°C. The irradiation source consisted of one of two monochromatic UV light LEDs; the first was a 278 nm, 2.4W LED purchased from IRTronix LG. The second irradiation source was a monochromatic UV LED emitting peak wavelength centered at 365 nm (LEUVA33U70RL00 – 365 nm, 2.6 W LED) purchased from IRTronix LG. The two LEDs were fitted into a custom-built 3D-printed irradiation chamber, as shown in Figure 2.1. The UV exposure times ranged between 3 s and 12 h, depending on the experiment. All irradiations and analyses were performed in duplicates.

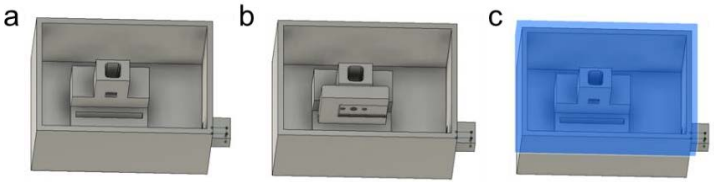


Figure 2.1. A three-dimensional (3D) printed setup without lid (a), containing an LED holder (b), and covered with a lid (c) to hold the sample in a cuvette during irradiation with UV light generated through an LED light.

2.3.5 Confirmation of CPD formation by T4-PDG digestion

The formation of the CPD in UV irradiated dsDNA library samples were confirmed by digesting the aliquots of the irradiated duplexes with T4 PDG. After UV irradiation, an aliquot of each ^{32}P-labeled DNA or cold DNA sample was incubated with 1µL (10 U) of T4-PDG (final volume 10 µL) for 30 min at 37 °C (different digestion times were tried and no differences in amount of cutting observed). All samples were then denatured by adding to 10 µL of denaturing dye containing 100% formamide and boiled in a water bath for 6

min, followed by immediate placement in ice. Subsequently, 1µL of phenol was added to each sample prior to loading onto the 8% Denaturing PAGE. Every time, a control sample was not irradiated with UV light but treated with the same succeeding steps. The ^{32}P-labeled DNA samples were visualized by autoradiography using a Typhoon phosphor imager (GE) while unlabeled (cold) DNA samples in gels were stained by SYBR® Gold Nucleic acid staining. Densitometry measurements were done using ImageJ software and graphs were plotted using Origin Pro 2021 Academic Version. Calculation of Kinetic (K_{obs}, k_r and k_f was done by Origin Pro 2021 Academic Version.

2.4 Results and Discussion

2.4.1 Design of the dsDNA randomized library

We designed a randomized, single-stranded DNA library consisting of a stretch of 10 randomized base pairs around a central thymine pair (the thymine pair has five random bases on either side) (Figure 2.2). The ten variable nucleotides will generate over one million different sequences (4^{10} = 1,048,576 sequences). The above single-stranded randomized DNA library was perfectly convenient to work with for subsequent PCR and Illumina sequencing. However, to ensure higher coverage for statistical calculation, we investigated only 3 variable nucleotides from each side. We also take into consideration the possibility of the formation of CPD in the lower strand after UV irradiation, so we use only the TT-strand (upper strand) for our calculation. In addition, the formation of CPDs in the lower strand can affect the formation of CPDs in the upper strand due to minor conformational changes in the DNA duplex.[206] However, this effect is equally distributed for all the sequences within the DNA library.

Single strand random DNA library

5'TACGCGTGTATACATACACACACGTATGTATATGTACACGCACGTGTATACNNNNNTTNNNNNGCACACATATGT
GTACACACACGTACGTATACACATAC ATACACGCACATG 3'

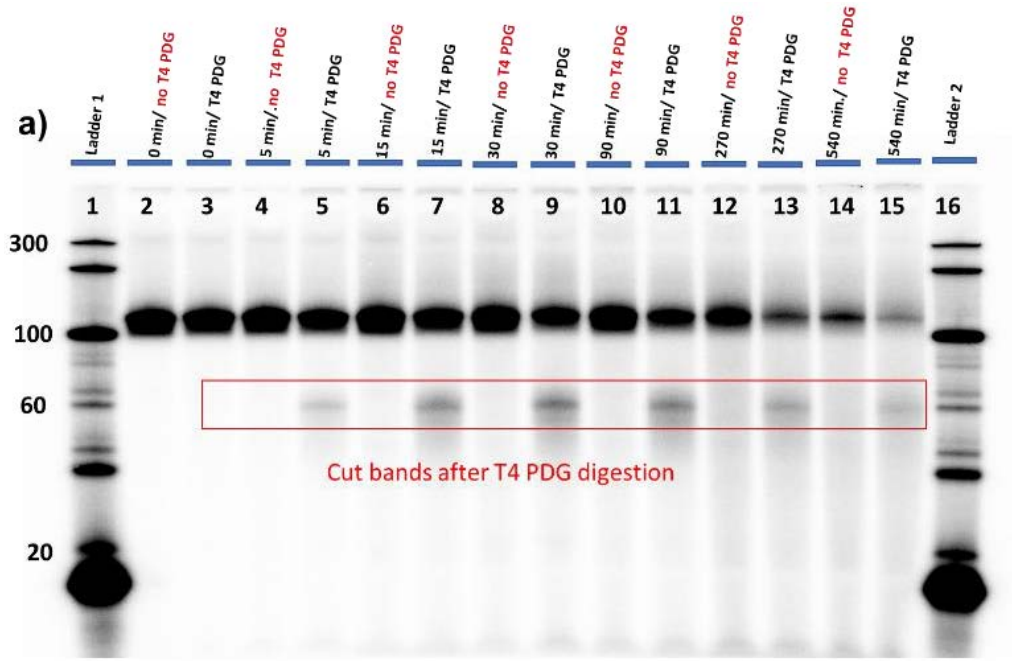
Reverse primer sequence

5' CATGTGCGTGTATGTATGTG 3'

Figure 2.2. Illustration of the single strand random DNA library and reverse primer sequences.

2.4.2 UV irradiation and designing the time-course experiments at 278 nm and 365 nm

As mentioned in the Methods section, UV irradiation was carried out using either a monochromatic 278 nm LED, or a 365 nm LED with the presence of 2′-methoxyacetophenone (2-M) as a triplet sensitizer (carried out under an anaerobic condition). Two time-dependent (time-course) experiments were performed by irradiating the purified dsDNA library with 365 nm light at eight-time points (0, 5, 15, 30, 90, 270, 540, and 720 min). For the 278 nm LED irradiation, nine different irradiation times, ranging from 0, 3, 10, 30, 90, 270, 540, 900, to 1800 s were selected (this irradiation was carried out under aerobic conditions and without added photosensitizer). The above time points were selected following preliminary experiments on T4 PDG digestion time courses on multiple ^{32}P-radiolabelled samples irradiated with either 278 the nm LED or 365 nm LED (the latter with triplet sensitizer). The cleaved DNAs were run on denaturing PAGE (Figure 2.3).



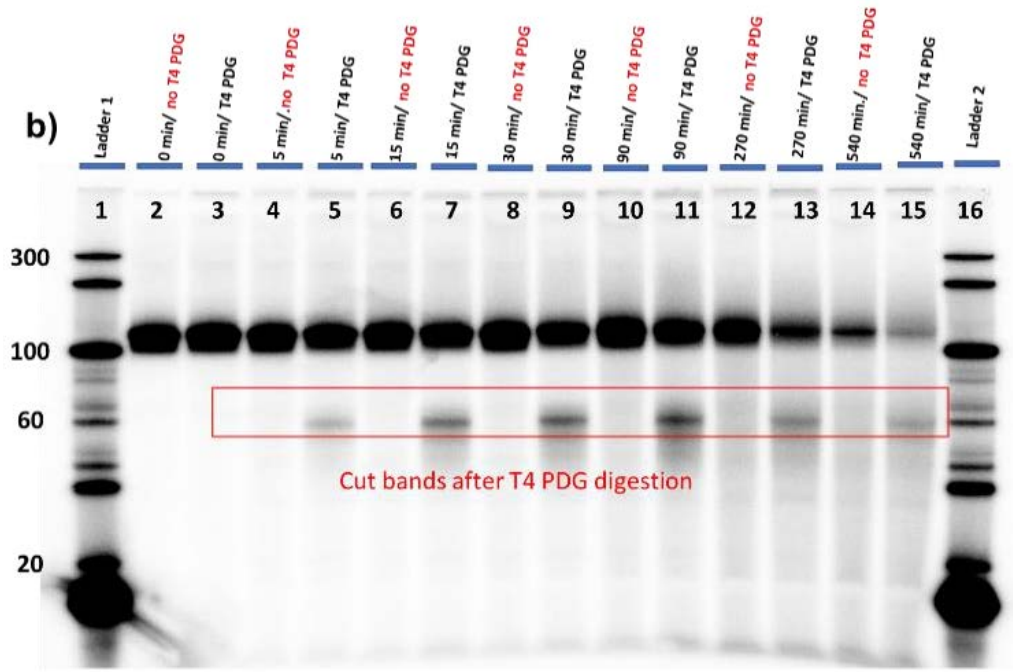


Figure 2.3. Two time-course experiments of ^{32}P-radiolabelled library DNA samples irradiated at 365 nm in the presence of 2.5 nm 2′-methoxyacetophenone (2-M). The experiments were done in replicate for consistency (**a, b**). All the samples were digested with T4 PDG and run onto 8% denaturing PAGE alongside the same irradiated DNA samples without T4 PDG digestion. Densitometry measurements show that the maximum level of cutting (~28%) occurs at 30 and 90 min.

The time-course experiment of 365 nm LED irradiation consists of eight time points, and this experiment was repeated twice. The purified dsDNA library was divided into 8 portions, which were irradiated in turn for: 0, 5, 15, 30, 90, 270, 540, and 720 min. All the sample for irradiation were adjusted to 200 µl by adding (20 mM NaCl/ 10 mM Tris-HCl, pH 8.0) buffer, and then degassed by freezing and thawing two times or until no further large bubbles were seen. Then to each time point DNA sample, 5% 2′-methoxyacetophenone (2-M) in acetone was added up to 2.5 mM final concentration and mixed by pipetting immediately prior to UV irradiation. After adding the 2-M, the mixed sample was transferred into a 1 ml UV quartz cuvette with a rubber stopper. Then, the air was removed from the stoppered cuvette by flushing the argon gas through a needle through the septum along with an exit needle; 2 to 3 min). All the time point samples were irradiated in the cold room (4°C) for the above indicated time except for the time = zero time point (which was not irradiated and was kept in the dark). Following irradiation, all the time point samples were purified using AMPURE XP beads and checked by Qubit for

their dsDNA concentration and quality and size of dsDNA libraries checked by Agilent Bioanalyzer. All the samples were tested for CPD formation by way of T4 PDG digestion by digesting a small aliquot of each time point sample with the T4 PDG enzyme, which was then analyzed in a denaturing gel. Since those samples were submitted for next generation sequencing, ^{32}P radiolabeling cannot be used. Therefore, we used SYBR® gold nucleic acid gel staining as a sensitive method to monitor and detect the DNA libraries, their quality, and the presence of CPDs within them. Following confirmation of this quality control, all samples were properly labeled and sent to the Sequencing and Bioinformatics Consortium (SBC) at UBC for sequencing using NextSeq 550 system.

The time-course experiment using the 278 nm LED irradiation consisted of nine-time points, and this experiment was also carried out in duplicate. The irradiation was done under aerobic conditions, and no photosensitizer was used. The nine-time points were prepared by dividing the purified dsDNA library into nine time-point portions: for 0, 3, 10, 30, 90, 270, 540, 900, and 1800 s. Each time point sample was adjusted to 200 µL volume by adding (20 mM NaCl/ 10 mM Tris-HCl, pH 8.0) buffer and transferred into a 1 mL UV quartz cuvette. All the time points irradiated for the above indicated time except for the time = zero sample, which was not irradiated and was kept in the dark. Following irradiation, all the time point samples were purified by AMPURE XP beads and checked by Qubit for dsDNA concentrations and quality and the size of the dsDNA libraries were checked by Agilent Bioanalyzer. All the samples were tested for CPD formation by digesting a small aliquot of each time point sample with the T4 PDG enzyme, and the results analyzed in a denaturing gel (Figure 2.4). Since these samples were attended to send for NGS Following this check for quality control, all samples were adequately labeled and sequencing using NextSeq 550 platform.

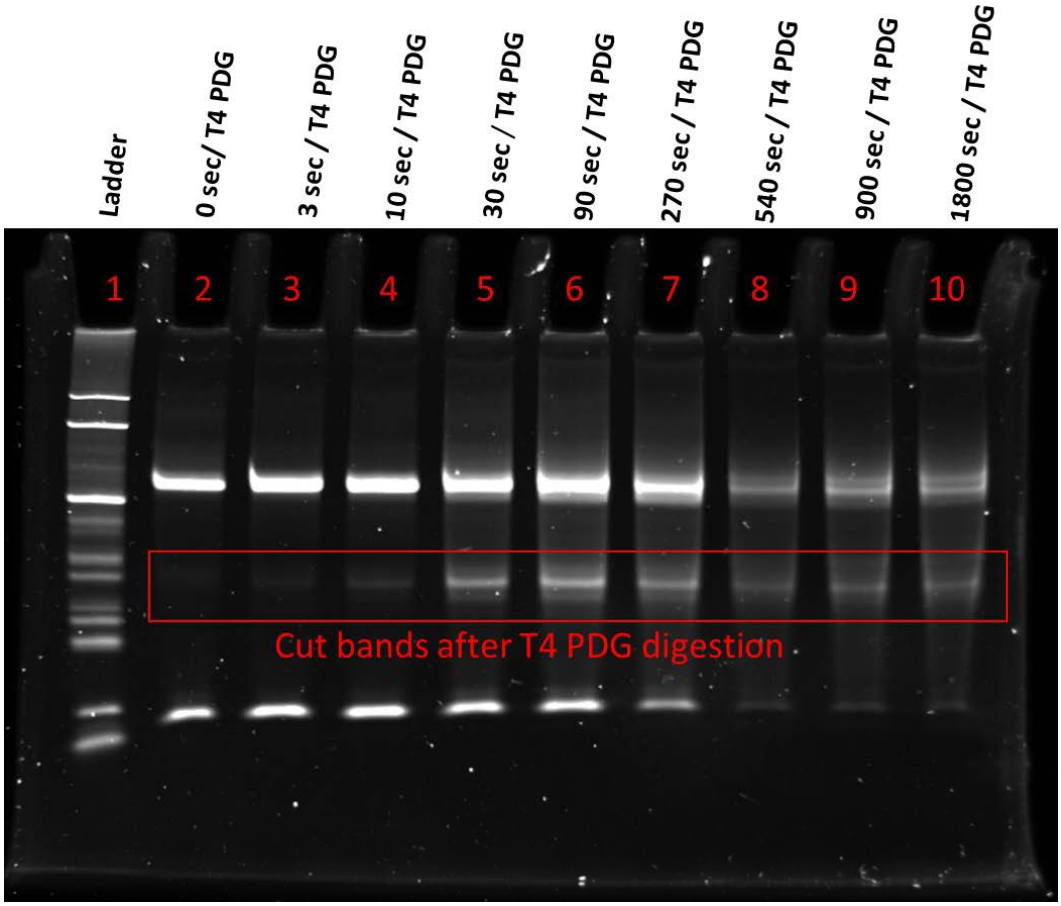


Figure 2.4. shows quality control testing the time points samples (278 nm irradiation) for CPD formation prior to sequencing. A small aliquot of each time point sample was digested with the T4 PDG enzyme and run in an 8% denaturing gel. The gel was stained by SYBR® gold nucleic acid gel stain and then visualized by ChemiDoc™ Imaging System from Bio-Rad.

2.4.3 DNA Polymerase blockage at pyrimidine dimers

UV light causes damage in cells that prevents DNA polymerase from replicating. Depending on which template strand bears the lesion, lesions that inhibit polymerases are able to either stop the replication machinery or cause nascent-strand gaps.[207–214] Several investigations utilizing plasmid substrates show that lesions in the leading-strand template stop the replication fork from progressing altogether, with the nascent lagging strand extending a short distance beyond the stalled leading strand.[212,213,215] Lesions in the lagging-strand template, on the other hand, are hypothesized to induce gaps in the nascent DNA strand at places opposite the lesion, likely because the lagging strand's discontinuous synthesis permits the halted polymerase to restart downstream of the lesion.[209,212,215] On

the chromosome of UV-irradiated Escherichia coli, events that are consistent with this may also be seen. The rate of DNA synthesis is transiently decreased after a modest dose of UV irradiation before quickly recovering in a period that correlates with lesion elimination.[211,216] Cyclobutene pyrimidine dimers (CPDs) and (6-4) photoproducts (6-4PP) are the most stable DNA-lesions formed by UV light (254 nm) irradiation that causes blockage of DNA polymerase.[217] This phenomenon has been used in the primer extension assay for photofootprinting[218] and for mapping CPDs at single nucleotide resolution, using Thermus aquaticus (Taq polymerase).[219] These tests rely on Taq DNA polymerase being blocked very well and very accurately at CPDS and (6-4) photoproduct. Recently, blockage of high-fidelity DNA polymerase has been used in a number of NGS-based techniques to map the UV and cisplatin damage positions.[192,220]

We adopted a next-generation sequencing-based method to detect thymine dimer (TT-CPD) formation and repair in DNA duplex. The principle of our approach relied on precise blockage of high-fidelity Taq DNA polymerase at pyrimidine dimers (CPDs) sites. The presence of TT-CPD in irradiated samples causes the disappearance of sequences that contain a thymine dimer from the irradiated DNA library pool during the PCR enrichment step (Figure 2.5). High-throughput sequencing using NextSeq 550 system was then done to quantitatively determine the formation and repair of CPDs over the time-course experiments at different time points and wavelengths. A NextSeq, mid-output flow cell, can generate up to 130 million paired-end clusters (260 million paired-end reads) (2 × 150 bp). Although paired-end reads have been done on all our irradiated samples, only the upper strand (TT-strand) reads have been used for all the statistical and bioinformatic calculations (Figure 2.6).

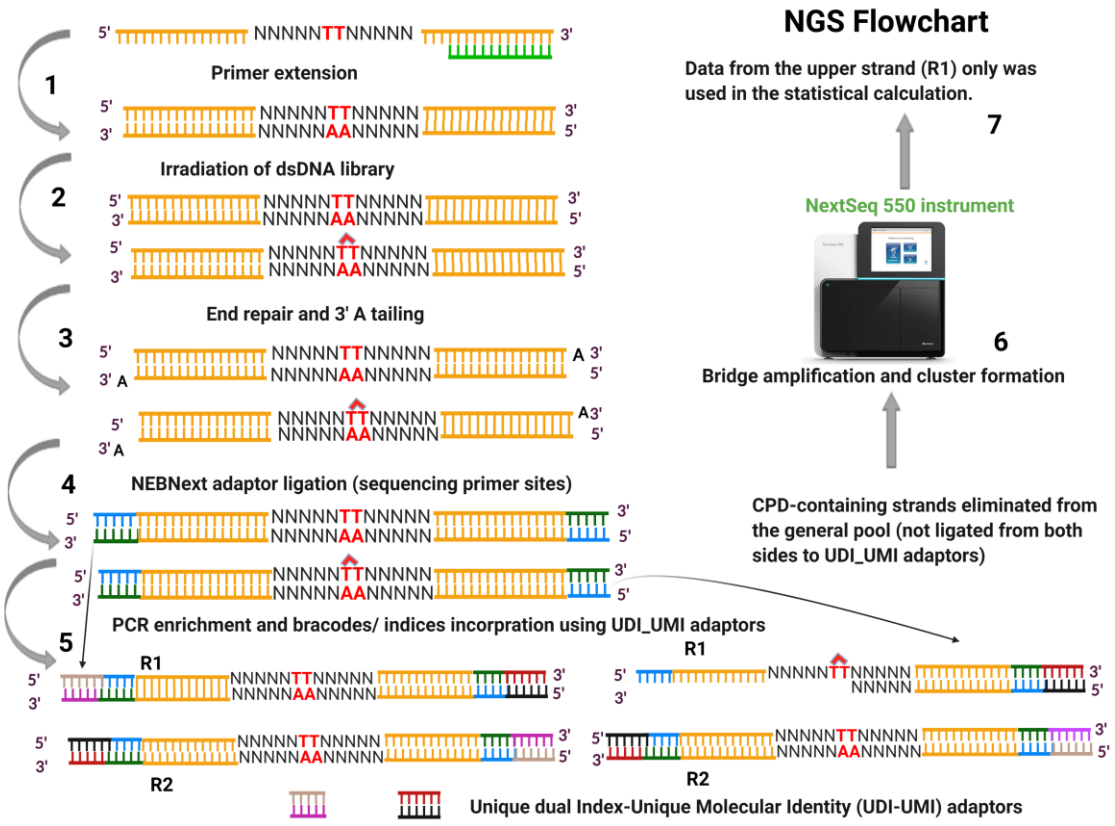


Figure 2.5. The above schematic diagram shows the steps of next-generation sequencing: 1) Reverse primer extension of the single-stranded DNA library to produce the dsDNA library. 2) The dsDNA library was irradiated at two different conditions, and time points were taken. 3) Following AMPURE XP beads purification, all DNAs were end-repaired and 3' A tailed to prepare them for subsequent NEBNext adaptor ligation. 4) NEBNext adaptors are short adaptors that contain sequences required downstream (complementary to Illumina primers). Here all the sequences (either containing CPD or not) would be able to ligate to NEBNext adaptor because no PCR is involved in this step. 5) All ligated libraries were then cleaned up and size selected before undergoing PCR enrichment to incorporate all necessary barcodes and indices required for multiplexing, thus giving a unique molecular identity (UMI) for each sequence. Here, only sequences *without* CPD would be amplified by PCR using Illumina indexed primers, while all the sequences containing CPD would not be amplified and thereby removed from the pool. 6) In this step, only complete dsDNA libraries ligated from both sides to the adaptors by PCR using Illumina indexed primers were selected for sequencing by size selection. 7) A NextSeq, mid output platform was used to generate (2 × 150 bp) paired-end reads. 8) The reads from the upper strand (TT-strand) were used in the statistical analysis.

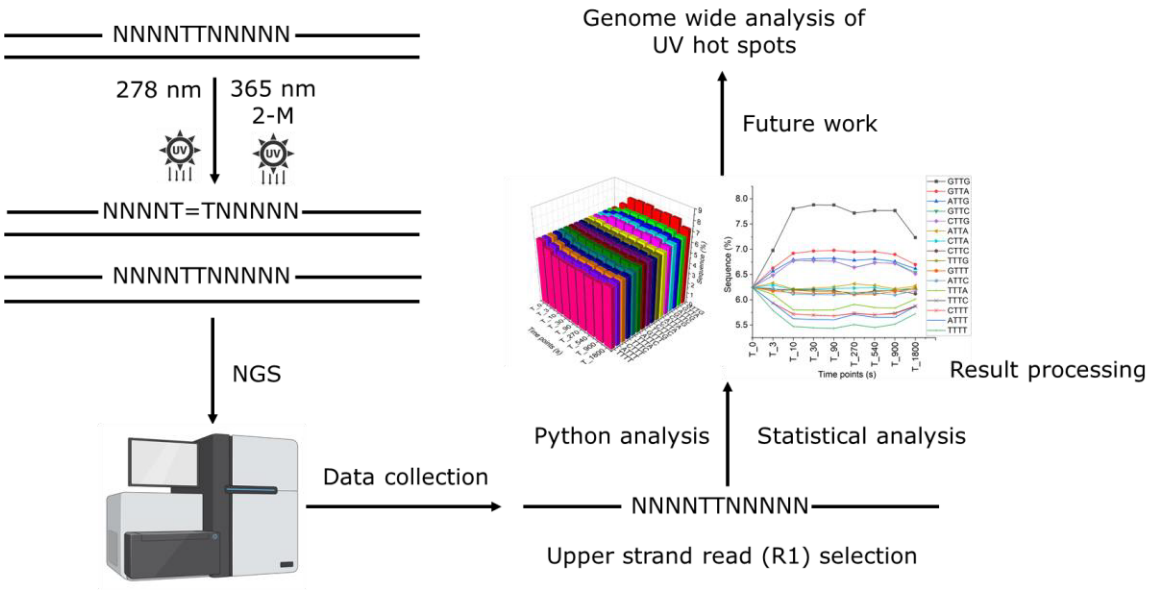


Figure 2.6. A schematic diagram to show the summary of the project and the selection of the upper strand (the "TT" strand) of the DNA library for statistical analysis.

2.4.4 Effect of flanking bases on the formation of TT-CPD

Python code was generated to analyze the resulting sequencing data and identify sequence fluctuations in the time-course experiments, at the single molecular level. The statistical analysis of the sequencing data using one-way analysis of variance (ANOVA), and Python code was done by collaboration with Joshua Unrau, University of Waterloo. The Python code and all the sequencing data are available in the online repository: https://github.com/Sen-lab-SFU. Regarding the statistical calculation and the applied cutoff: In our statistical calculation, we only relied on sequences and motifs with no missing or ambiguous bases. The data generated by the Python code was carefully examined to remove any sequences with missing or ambiguous nucleotides in the variable region. We have done the statistical calculation for 6 variable nucleotides, 3 bases from each side, which can generate up to 4096 unique sequences. Two confidence intervals were used: one was using student's t, the other is the population standard deviation to estimate the difference between the zero-time point (control sample and each time point). As for delta calculation, the formula is Dx = Tx - T0. All confidence intervals are present in Appendix C compressed file.

We first determined the frequency of TT-CPD fluctuation as a function of the time in response to irradiation from both UV light sources at all 16 possible sites (NTTN) within the randomized dsDNA library. The 16 total possible base tetrads of form NTTN were detected in two-replicate time-course experiments from each type of UV irradiation (UVC at 278 nm and UVA at 365 nm) (Figure 2.7). All time points for both irradiations were obtained with high coverage, and the number of copies for each time point is shown in Table 2.1 and Table 2.2.

Table 2.1 Total number of copies for the NTTN motif in two-replicate 365 nm irradiation.

Experiment 1 365 nm / min	**Total of copies for (NTTN) motif**	**Experiment 2 365 nm / min**	**Total of copies for (NTTN) motif**
Time point 0	8189385	Time point 0	11512034
Time point 5	77418005	Time point 5	80713405
Time point 15	17643825	Time point 15	21122272
Time point 30	49393126	Time point 30	52222344
Time point 90	103082383	Time point 90	106375127
Time point 270	37292182	Time point 270	40161382
Time point 540	71540696	Time point 540	74457803
Time point 720	84079406	Time point 720	87394896

Table 2.2 Total number of copies for the NTTN motif in two-replicate 278 nm irradiation.

Experiment 1 278 nm / s	**Total of copies for (NTTN) motif**	**Experiment 2 278 nm / s**	**Total of copies for (NTTN) motif**
Time point 0	1366576	Time point 0	4586339
Time point 3	14683393	Time point 3	62545829
Time point 10	55483470	Time point 10	58857086
Time point 30	43338165	Time point 30	46560947
Time point 90	96794428	Time point 90	100045418
Time point 270	31079461	Time point 270	34094071
Time point 540	65530858	Time point 540	68418033
Time point 900	90448900	Time point 900	93570451
Time point 1800	25135023	Time point 1800	28319566

The overall pattern of thymine dimer formation and repair across all 16 tetrads under the two different irradiation settings appears remarkably similar (Figure 2.7), with TTTT and ATTT always being the hottest spots, showing the highest negative selection over time, and GTTG being the coldest spot, showing the highest positive selection over time. In agreement with previous studies, we observed a slightly higher suppressive effect of 5' flanking guanine on TT-CPD formation than 3' flanking guanine with all other bases rather than thymine.[13,115,141,200–202] Guanine with all other bases showed more positive selection when G on 5' side, which supports the postulation of greater 5'-suppressed CPD formation more than of 3'-suppressed, owing to better pi-stacking of a G with the 5' pyrimidine".[115,202,221] For example, GTTA exhibits more positive selection than ATTG; and, the same is true for GTTC, which shows a higher selection than CTTG. However, the story is different when a thymine flanks the thymine pair on one side with guanine on the other side; in this case, guanine from the 3' side showing a higher suppressing effect than guanine from 5' side (TTTG > GTTT). This finding is entirely different from any earlier studies. In general, guanines directly flanking the central thymine pair show increased

DNA photostability and relatively higher levels of repair of the neighboring thymine dimer. Meanwhile, cytosines and adenines show slower CPD formation and repair rates and comparable fluctuation over time.

The NTTN pattern can generally be grouped into four groups (Table 2.3): the first group includes two guanines sandwiching the thymine pair (GTTG). The GTTG tetrad shows the highest positive selection over time, supporting the suggestion that neighboring guanines promote dimer repair by acting as electron sources for CPD repair.[110] The second group includes guanine on one side and adenine or cytosine on the other (GTTN, NTTG; N = A or C). GTTA sequence was the second selected sequence in both irradiation experiments, which was also expected. Adjacent purines have been shown to increasingly inhibit CPD formation by quenching of the pyrimidine excited state by electron transfer from the flanking base consistent with oxidation potential.[201] GTTC and CTTG show a very similar pattern, with the highest selection over time following GTTG and GTTA—either by direct 278 nm irradiation or photosensitized 365 irradiation. Group 2 has the following order from the highest to lowest selection: GTTA > ATTG > GTTC > CTTG. The third group contains cytosines or adenines on both sides, cytosine on one side, adenine on the other, guanine on one side, and thymine on the other (CTTC, ATTA, ATTC, CTTA, GTTT, and TTTG). This group shows a fairly constant or marginal 'negative' selection over time. Surprisingly, TTTG reveals less negative selection (or better UV protection) than GTTT, which is different from previous postulated studies.[115,202,221] The fourth group includes cytosine or adenine on one side and thymine on the other, as well as thymines on both sides (TTTC, CTTT, TTTA, ATTT, and TTTT). In this group, cytosine and adenine on the 3' side with thymine (TTTA and TTTC) exhibit slightly less thymine dimer formation than on the 5' side (ATTT and CTTT). It is not surprising that the TTTT tetrad has the highest negative selection compared to all other NTTN sequences because the T-tetrad maximizes the likelihood of CPD formation in that stretch of sequence.

Table 2.3 shows the pattern and fluctuation of the 5' NTTN 3' sequence in 278 nm irradiation and 365 nm. The 16 tetrads are grouped into four groups according to their selection tendency.

NTTN Classification	Selection order	Wavelength 278 nm	Wavelength 365 nm	Tendency of selection
Group 1	1	GTTG	GTTG	Highest positive selection
Group 2	2	GTTA	GTTA	Moderate selection
	3	ATTG	GTTC	
	4	GTTC	ATTG	
	5	CTTG	CTTG	
Group 3	6	ATTA	CTTC	Constant to marginal negative selection
	7	CTTA	ATTA	
	8	CTTC	TTTG	
	9	TTTG	CTTA	
	10	GTTT	ATTC	
	11	ATTC	GTTT	
Group 4	12	TTTA	TTTC	Highest negative selection
	13	TTTC	CTTT	
	14	CTTT	TTTA	
	15	ATTT	ATTT	
	16	TTTT	TTTT	

It is worthwhile to compare our NTTN pattern findings with the recent paper by Taylor. *et al*. 2021. The NTTN pattern fluctuation in this study is quite similar to both UV irradiation at 278 nm and 365nm, in which the TTTA has the highest thymine dimer formation among all NTTN tetrads. However, few NTTN pattern sequences show different order from our result and different from previous studies (Table 2.4). Unexpectedly, ATTT and TTTT show a low and an intermediated level of CPDs formation, respectively, in this study. The third main difference is that TTTG is the second-highest CPDs formation among all possible tetrads, while TTTG comes in the middle of fluctuation in our result (Table 2.5). **Table 2.4** Main NTTN pattern differences between our result and UVC irradiation of Taylor. et al. 2021 paper.

Table 2.4 Main NTTN pattern differences between our result and UVC irradiation of Taylor. et al. 2021 paper.

Selection order	278 nm Our data	UVC Taylor *et al. 2021*	
1	GTTG	GTTG	
2	GTTA	ATTT	←
3	ATTG	GTTC	
4	GTTC	GTTT	
5	CTTG	CTTG	
6	ATTA	GTTA	
7	CTTA	CTTC	
8	CTTC	TTTT	←
9	TTTG	ATTG	←
10	GTTT	ATTC	
11	ATTC	ATTA	
12	TTTA	TTTC	
13	TTTC	CTTT	
14	CTTT	CTTA	
15	ATTT	TTTG	←
16	TTTT	TTTA	

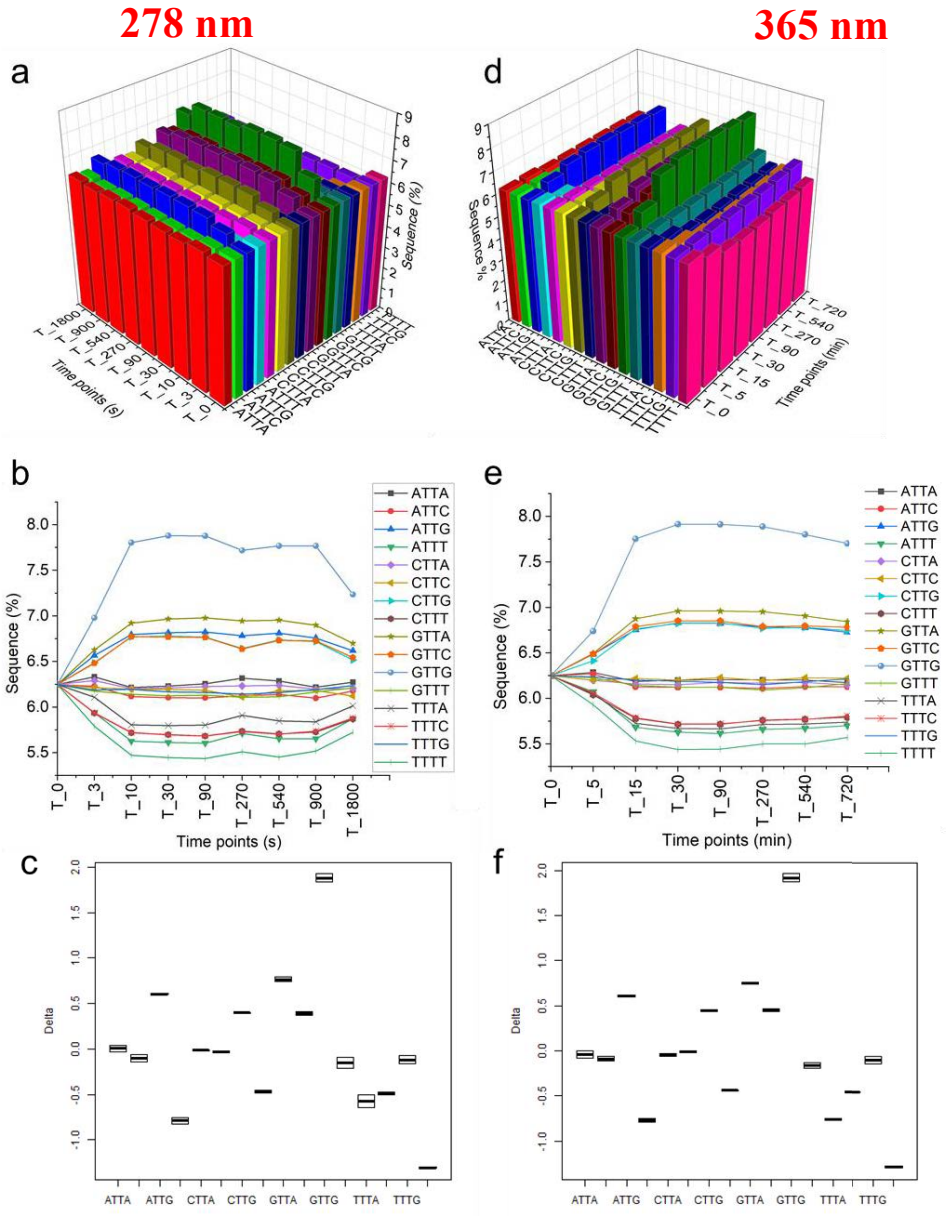


Figure 2.7. A representation of 16 possible tetrads (NTTN) fluctuation in a set of duplicate experiments at 278 nm (panels a, b, c) and 365 nm (panels d, e, f). The data points were normalized, so that the time point = 0 in all sequences starts from 6.25 % (which is the ratio of 1 tetrad (NTTN) out of 16). The sequence order from positive selection to negative selection is as follows: GTTG> GTTA> ATTG> GTTC> CTTG> ATTA> CTTA> CTTC> TTTG> GTTT> ATTC> TTTA> TTTC, CTTT> ATTT> TTTT. The fluctuation of the 16 possible tetrads (NTTN) in the two-replicate experiment at 365 nm. The selection order in 365 nm irradiation experiments is very similar to that found in 278 nm irradiation experiments. The tetrad sequences are ordered (from positive selection to negative selection in 365 nm experiments) as follows: GTTG> GTTA> ATTG> GTTC> CTTG> ATTA> CTTA> CTTC>TTTG> GTTT>ATTC>TTTA>TTTC, CTTT>ATTT>TTTT. Statistical analyses by ANOVA and student's t-test showing the statistical results of mean values and standard deviations in the data as depicted in panels (c) and (f) Confidence intervals for delta calculation, the formula is Dx = Tx - T0 for 278 and 365 nm, respectively.

Table 2.5. Average delta values for the percent fluctuation for each motif calculated by ANOVA and R statistics. Confidence intervals for delta calculation, the formula is (X time point - zero time) point Dx = Tx - T0 for 365 and 278 nm, respectively.

Motif (NTTN)	Delta (δ) values for 365 nm wavelength exposure for 90 min	Delta (δ) values for 278 nm wavelength exposure for 90 s
ATTA	-0.044	0.007
ATTC	-0.088	-0.101
ATTG	0.607	0.604
ATTT	-0.770	-0.786
CTTA	-0.049	-0.015
CTTC	-0.010	-0.032
CTTG	0.443	0.398
CTTT	-0.439	-0.471
GTTA	0.749	0.764
GTTC	0.450	0.389
GTTG	1.917	1.880
GTTT	-0.165	-0.151
TTTA	-0.764	-0.573
TTTC	-0.452	-0.489
TTTG	-0.101	-0.118
TTTT	-1.284	-1.305

The second step in our data analysis was determining individual nucleotides' fluctuation in the TT strand ($N_3N_2N_1TTN_4N_5N_6$) in the dsDNA library pool over irradiation time (Figures 2.8 and 2.9). The total number of copies at each position for each time point is given in Appendix C. This fluctuation would reflect the positive and negative selection of the nucleotides at these positions. Thereby we can have a general idea about which nucleobase promotes repairing the thymine dimer and which nucleobase favours the formation of the thymine dimer. The fluctuation of the four different nucleotides adjacent to the thymine pair in all sequences can be summarized in the following: Guanines show the highest *positive* (i.e. protective) selection in the two UV irradiation wavelengths. The intensity of selection increases as the thymine pair's proximity increases from both sides in all sequences. Cytosines show only a slight negative selection at the first position adjacent to the thymine pair. However, cytosines reveal the second positive selection after guanines at the second and third positions (C_2, C_3, C_6, and C_5) from the 5' side and 3' side,

respectively. Adenines and thymine show a considerable negative selection. The highest negative selection for adenine at the second position from thymine pair from both sides (A_2 and A_5). While thymine shows the highest negative selection at the first position to thymine pair from both sides (T_1 and T_4) (Tables 2.6 and 2.7). Therefore, we can argue that relatively positive adenine selection in the previous NTTN pattern (GTTA and ATTG) could be deceptive or misleading because of the presence of cytosine or thymine adjacent to the thymine pair increases the likelihood of CPDs formation. This argument is supported by the negative selection of adenine at all positions except directly adjacent to the thymine pair.

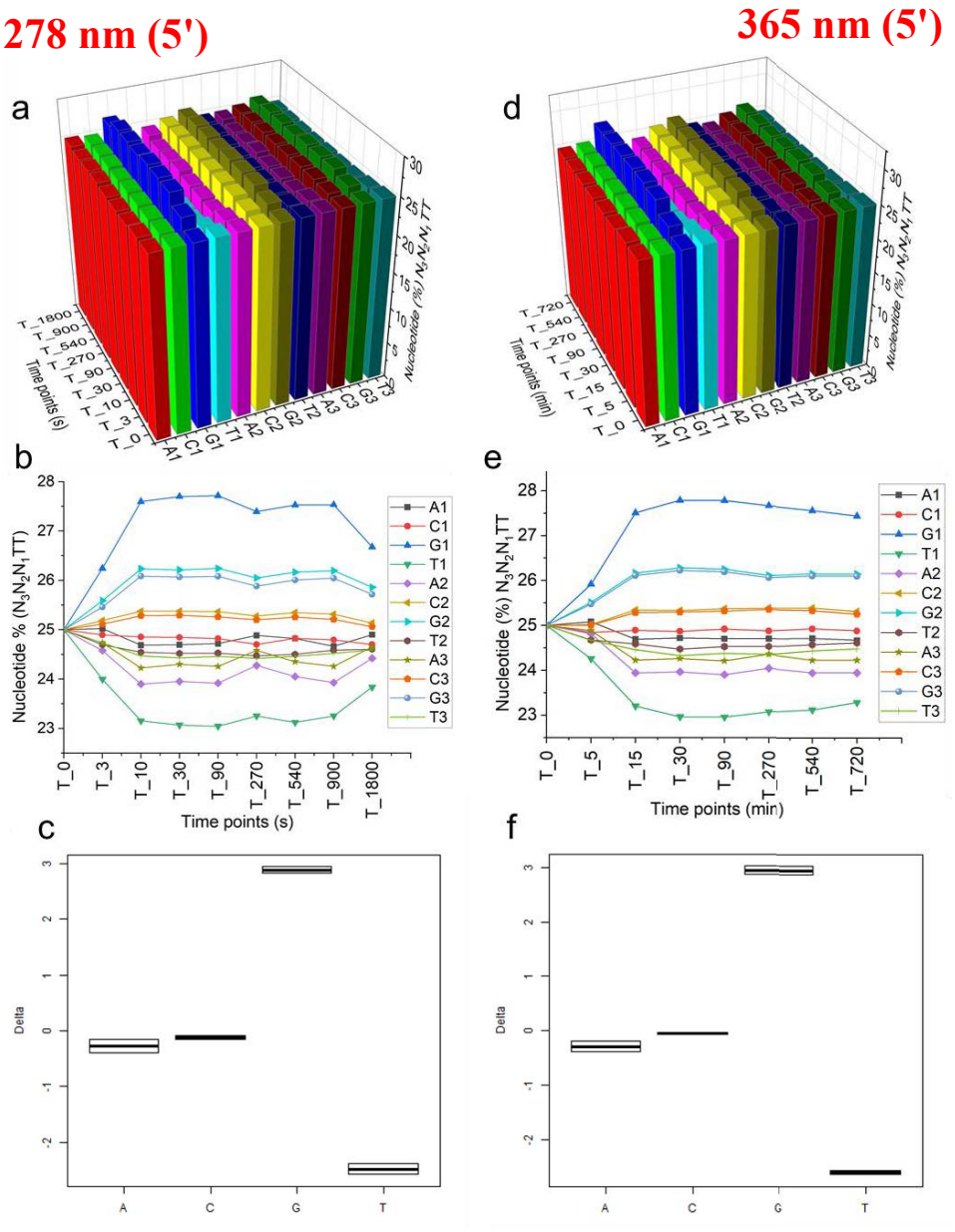


Figure 2.8. Fluctuation of individual nucleotides up to three positions at the 5' end of the central thymine pair (N3N2N1TT) in a duplicate experiment at 278 nm is represented in panels (a, b, c). Timepoint 0 in all sequences was normalized to start from 25% (which the percentage of 1 nucleotide out 4). The order of nucleotides' fluctuation from higher to lower selection follows as G1> G2> G3> C2> C3> C1> A1> T2> T3> A3> A2> T1. The individual nucleotides' fluctuation at three positions 5' to the central thymine pair (N3N2N1TT) in a duplicate experiment at 365 nm irradiation in the right-side panel (d, e, f). The data normalized, as above, so the timepoint 0 in all sequences starts from 25%. The order of nucleotides' fluctuation from higher to lower selection in 365 nm irradiation experiments is completely identical to 278 nm irradiation experiments. Statistical analyses by ANOVA and R studio showing the statistical results of mean values and standard deviations in the data as depicted in panels (c) and (f) for 278 and 365 nm, respectively. Values of delta (δ) for the selection of each nucleotide were calculated from the data sets to depict their fluctuation.

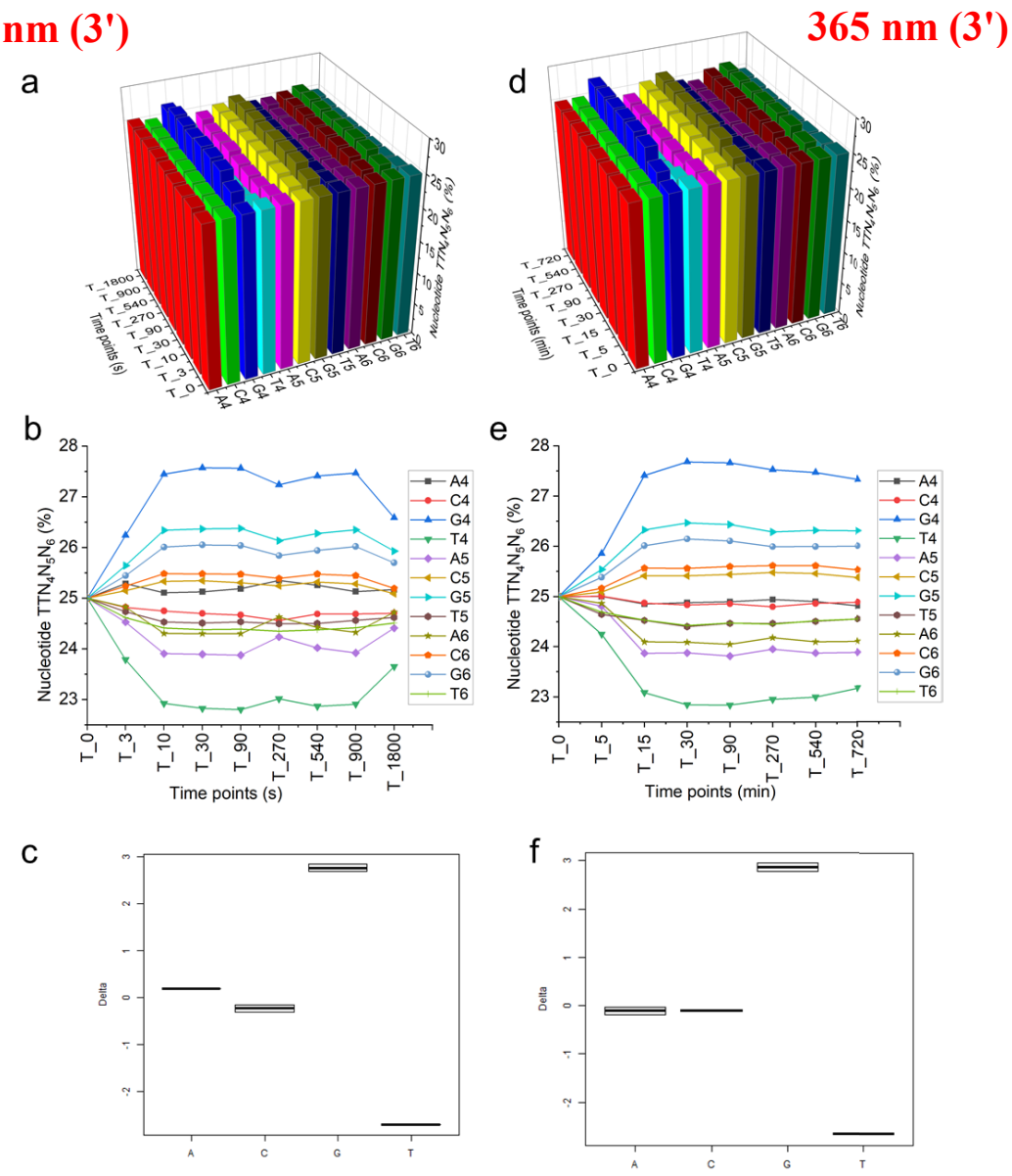


Figure 2.9. A depiction of individual nucleotides' fluctuation up to three positions at 3' of the central thymine pair ($TTN_4N_5N_6$) in duplicate experiments performed at 278 nm (a, b, c). Timepoint 0 in all sequences is normalized to start from 25% (which the percentage of 1 nucleotide out 4). The order of nucleotides' fluctuation from higher to lower selection as follows: G_4> G_5> G_6> C_6> C_5> A_4> C_4> T_5> T_6> A_6> A_5> T_4. The individual nucleotides' fluctuation at three positions 3' to the central thymine pair ($TTN_4N_5N_6$) in the two-replicate experiment at 365 nm irradiation in the right-side panel (c and d). The data normalized, as above, so the timepoint 0 in all sequences starts from 25%. The order of nucleotides' fluctuation from higher to lower selection in 365 nm irradiation experiments is also identical to 278 nm irradiation experiments. Statistical analyses by ANOVA and R studio showing the statistical results of mean values and standard deviations in the data as depicted in panels (c) and (f) for 278 and 365 nm, respectively.

Table 2.6. Average delta values for the percent fluctuation for each nucleotide at 5' side calculated by ANOVA and R statistics. Confidence intervals for delta calculation, the formula is (X time point - zero time) point Dx = Tx - T0 for 365 and 278 nm, respectively.

Nucleotide (5' side)	Delta (δ) values for 365 nm wavelength exposure for 90 min	Delta (δ) values for 278 nm wavelength exposure for 90 s
A	-0.294	-0.276
C	-0.055	-0.120
G	2.951	2.881
T	-2.601	-2.484

Table 2.7. Average delta values for the percent fluctuation for each nucleotide at 3' side calculated by ANOVA and R statistics. Confidence intervals for delta calculation, the formula is (X time point - zero time) point Dx = Tx - T0 for 365 and 278 nm, respectively.

Nucleotide (3' side)	Delta (δ) values for 365 nm wavelength exposure for 90 min	Delta (δ) values for 278 nm wavelength exposure for 90 s
A	-0.108	0.183
C	-0.100	-0.232
G	2.867	2.763
T	-2.659	-2.714

In the third kind of analysis, I investigated motif fluctuation adjacent to the central thymine pair from both 5' and 3' sides (5'NNTT, 5' NNNTT, TTNN, 3'and TTNNN 3').

Motifs fluctuation neighboring the central thymine pair would reflect the enrichment of a particular motif due to excimer formation or other charge transfer preferences. We first looked at the fluctuation of the two nucleotides motif adjacent to the thymine pair from both sides (Figures 2.10 and 2.11). Then we extended our investigation to three nucleotides motifs adjacent to the thymine pair from both sides (Figures 2.12 to 2.18). The total number of copies of for each time point is shown in Table 2.8 and Table 2.9.

Table 2.8 Total number of copies for the NNNTT and TTNNN motifs in two-replicate 278 nm irradiation.

Experiment 1 278 nm / s	**Total of copies for 5' NNNTT**	**Total of copies for TTNNN 3'**	**Experiment 2 278 nm / s**	**Total of copies for 5' NNNTT**	**Total of copies for TTNNN 3'**
Time point 0	1366578	1366576	Time point 0	4586335	4586231
Time point 3	14683370	14682876	Time point 3	62545586	62543263
Time point 10	55483251	55481138	Time point 10	58856857	58854657
Time point 30	43337990	43336358	Time point 30	46560763	46559014
Time point 90	96794031	96790447	Time point 90	100045001	100041300
Time point 270	31079350	31078173	Time point 270	34093938	34092636
Time point 540	65530603	65528166	Time point 540	68417767	68415208
Time point 900	90448548	90445234	Time point 900	93570086	93566647
Time point 1800	25134948	25133991	Time point 1800	28319473	28318401

Table 2.9 Total number of copies for the NNNTT and TTNNN motifs in two-replicate 365 nm irradiation.

Experiment 1 365 nm / min	**Total of copies for 5' NNNTT**	**Total of copies for TTNNN 3'**	**Experiment 2 365 nm / min**	**Total of copies for 5' NNNTT**	**Total of copies for TTNNN 3'**
Time point 0	8189370	8189134	Time point 0	11512014	11511644
Time point 5	17643796	17643181	Time point 5	21122224	21121454
Time point 15	77417693	77414833	Time point 15	80713086	80710102
Time point 30	49392925	49391057	Time point 30	52222135	52220126
Time point 90	103081949	103078145	Time point 90	106374685	106370724
Time point 270	37292039	37290627	Time point 270	40161227	40159698
Time point 540	71540405	71537742	Time point 540	74457491	74454717
Time point 720	84079087	84075993	Time point 720	87394541	87391332

Motifs containing guanines directly adjacent to the thymine pair from both 5' and 3' sides show the highest positive selection, particularly the following three motifs: GCG, GTG, and GGG. Motifs containing adenines and cytosines directly adjacent to the thymine pair from both 5' and 3' sides show mostly negative selection except for a few sequences, which show slight positive to constant selection. These slightly positively selected sequences are (GCA, GGA, GTA from 5' side and ACG, AGG, ATG from 3' side) (CGG, CGC, CAC from 5' side and CGG, CGC, CAC from 3' side). The common factor in these sequences is mainly the presence of CG, GC, GG, TG, and GT motifs. Motifs containing thymine directly adjacent to the thymine pair from both 5' and 3' sides have a negative effect on the photostability of DNA and promote CPDs formation, except when it present adjacent to 5' CG, GC 3', 5' GG, GG 3', 5' TG and GT 3' motifs. Thymine has the highest negative selection when followed by AA, TT or AT motifs (5'AAT, 5'ATT, 5'TTT) (TAA3', TTA3', TTT3'). The first observation is that adenines adjacent directly to the thymine pair have a mostly negative impact on the photostability of DNA and promote CPDs formation, except when it present adjacent to the motifs mentioned above. The second observation is when two adenines followed by thymine (5' TAA, AAT3') adjacent directly to the thymine pair exhibit the highest negative selection in all the sequences. Cytosine only had a negative impact when it is directly adjacent to the thymine pair (N_1 or N_4) or adjacent to another thymine, which leads to increased CPDs likelihood; other than that, cytosine shows positive selection at all other positions (N_3, N_2, N_5, and N_6). The presence of two consecutive pyrimidines can increase the likelihood of the formation of another CPD (to which our experimental assay is sensitive) rather than involving the central thymine pair. In accordance with the previous studies, the motifs containing two consecutive pyrimidines show less positive selection and tend to form the CPDs in the following order: 5’-NCC**TT** < 5’-NCT**TT** < 5’-NTC**TT** < 5’-NTT**TT**, and **TT**CCN-3’ < **TT**TCN-3’ < **TT**CTN-3’ < **TT**TTN-3' (Table 2.12 and 2.13).

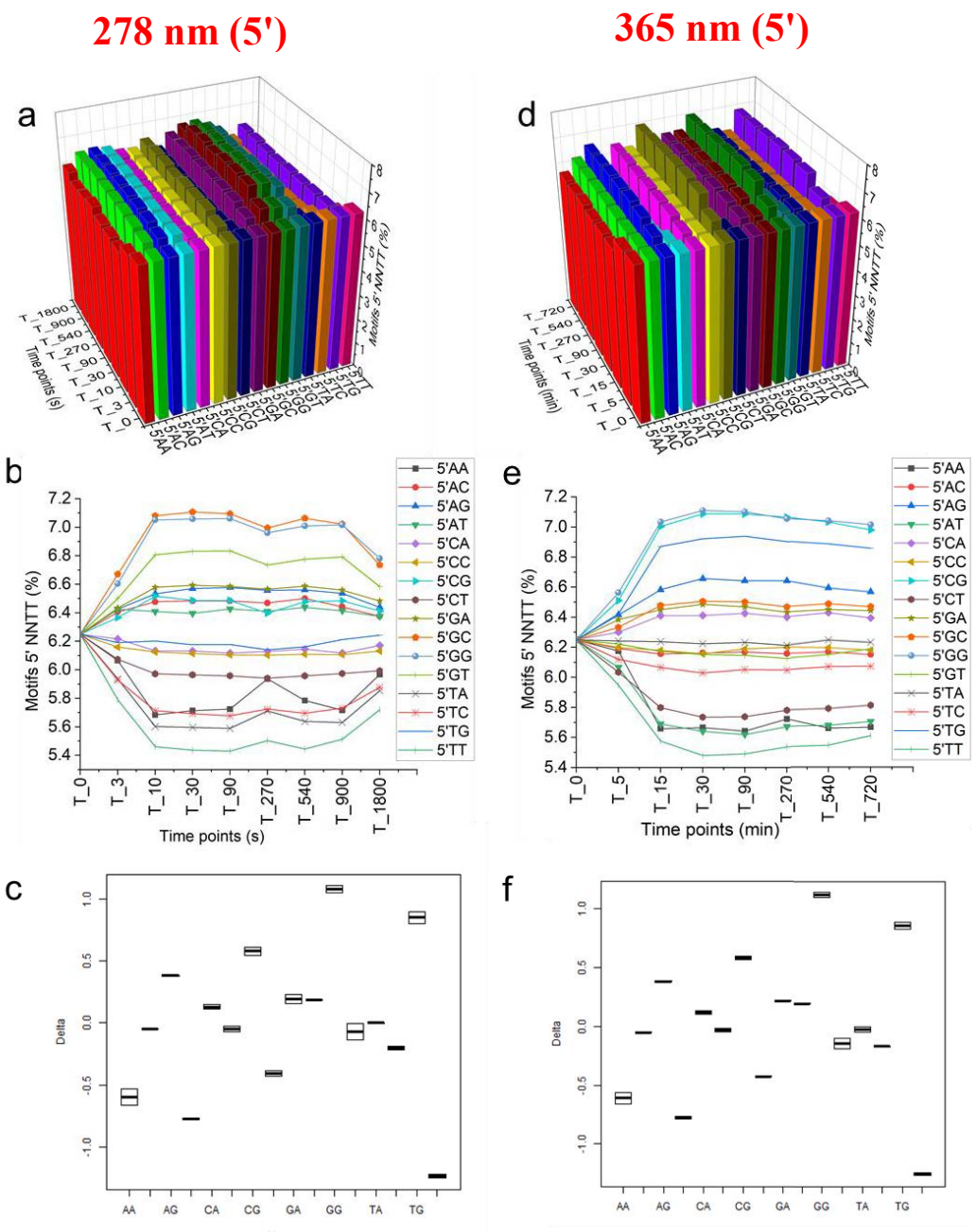


Figure 2.10. A top-down graph in the left-side panel (a and b) show motifs' fluctuation at two positions (5'NNTT) 5' adjacent to the central thymine pair in all sequences in the two-replicate experiment at 278 nm irradiation. Timepoint 0 in all motif sequences is normalized to start from 6.25%, making the percentage of 1 motif out 16 possible. The order of 5' dinucleotide motifs' fluctuation from higher to lower selection as follows: CG> GG>TG> AG> GC> GA> CA> TA> GT> AC> CC> TC> CT> AA> AT> TT. The 16 possible 5' dinucleotide motifs' (NNTT) fluctuation in the two-replicate experiment at 365 nm irradiation is shown in the right-side panel (c and d). The data normalized, so the timepoint 0 in all sequences starts from 6.25%. The order of 5' dinucleotide motifs' fluctuation from higher to lower selection in 365 nm irradiation experiments is almost identical to 278 nm irradiation experiments. Statistical analyses by ANOVA and R studio showing the statistical results of mean values and standard deviations in the data as depicted in panels (c) and (f) for 278 and 365 nm, respectively.

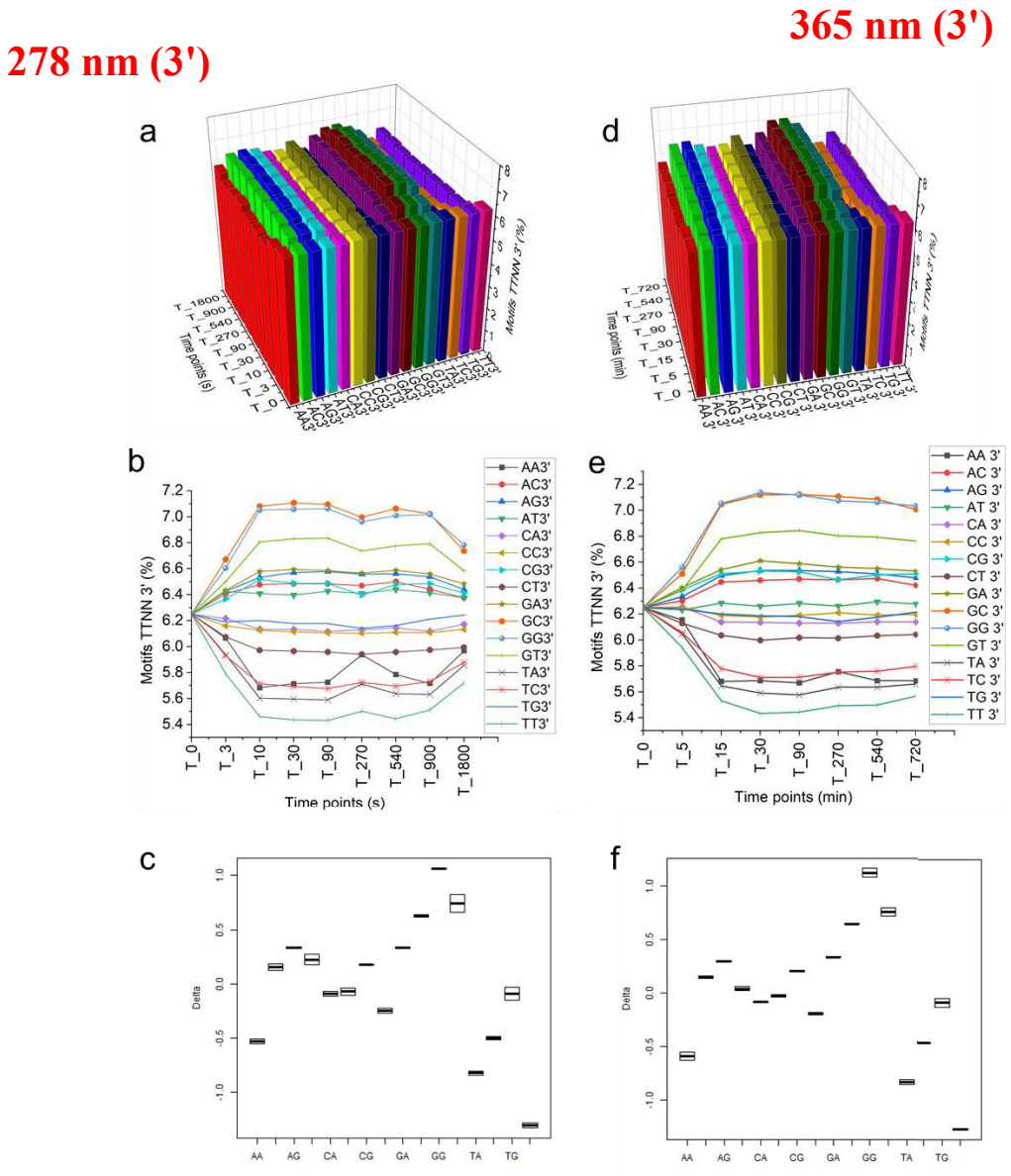


Figure 2.11. A top-down plot in the left panel (a and b) depicts the dinucleotide motifs' fluctuation (TTNN 3') 3' adjacent to the central thymine pair in all sequences in the two-replicate experiment at 278 nm irradiation. The timepoint 0 of every motif sequence is normalized to 6.25, giving a probability of 1 out of 16 motifs. Following is the breakdown of the 3' dinucleotide motifs' selection order from higher to lower: GC> GG> GT> GA> AG> CG> AC> AT> TG> CA> CC> CT> AA> TC> TA> TT. Right-side panels (c and d) show the fluctuation of the 16 different 3' dinucleotide motif (NNTT) in the two-replicate experiment at 365 nm irradiation. As in the 278nm experiment, the data normalized, so all timepoints 0 start from 6.25%. The order of 3' dinucleotide motifs' fluctuation from higher to lower selection in 365 nm irradiation experiments is similar to 278 nm irradiation experiments. The dinucleotide motifs' fluctuation in 278 nm and 365 nm experiments on 3' sides are almost a mirror image to 5' sides. Statistical analyses by ANOVA and R studio showing the statistical results of mean values and standard deviations in the data as depicted in panels (c) and (f) for 278 and 365 nm, respectively.

Table 2.10. Average delta values for the percent fluctuation for dinucleotide motifs at 5'side calculated by ANOVA and R statistics. Confidence intervals for delta calculation, the formula is (X time point - zero time) point Dx = Tx - T0 for 365 and 278 nm, respectively.

Dinucleotide Motif (5'NNTT)	Delta (δ) values for 365 nm wavelength exposure for 90 min	Delta (δ) values for 278 nm wavelength exposure for 90 s
AA	-0.611	-0.598
AC	-0.051	-0.050
AG	0.384	0.381
AT	-0.776	-0.771
CA	0.121	0.127
CC	-0.029	-0.051
CG	0.585	0.577
CT	-0.425	-0.408
GA	0.218	0.194
GC	0.195	0.184
GG	1.123	1.077
GT	-0.142	-0.070
TA	-0.022	0.001
TC	-0.169	-0.203
TG	0.859	0.847
TT	-1.258	-1.235

Table 2.11. Average delta values for the percent fluctuation for dinucleotide motifs at 3'side calculated by ANOVA and R statistics. Confidence intervals for delta calculation, the formula is (X time point - zero time) point Dx = Tx - T0 for 365 and 278 nm, respectively.

Dinucleotide Motif (TTNN3')	Delta (δ) values for 365 nm wavelength exposure for 90 min	Delta (δ) values for 278 nm wavelength exposure for 90 s
AA	-0.592	-0.531
AC	0.148	0.156
AG	0.295	0.334
AT	0.042	0.223
CA	-0.083	-0.093
CC	-0.029	-0.072
CG	0.206	0.177
CT	-0.194	-0.245
GA	0.338	0.334
GC	0.646	0.624
GG	1.125	1.060
GT	0.758	0.745
TA	-0.836	-0.821
TC	-0.463	-0.498
TG	-0.086	-0.090

TT	-1.273	-1.304

278 nm (5')

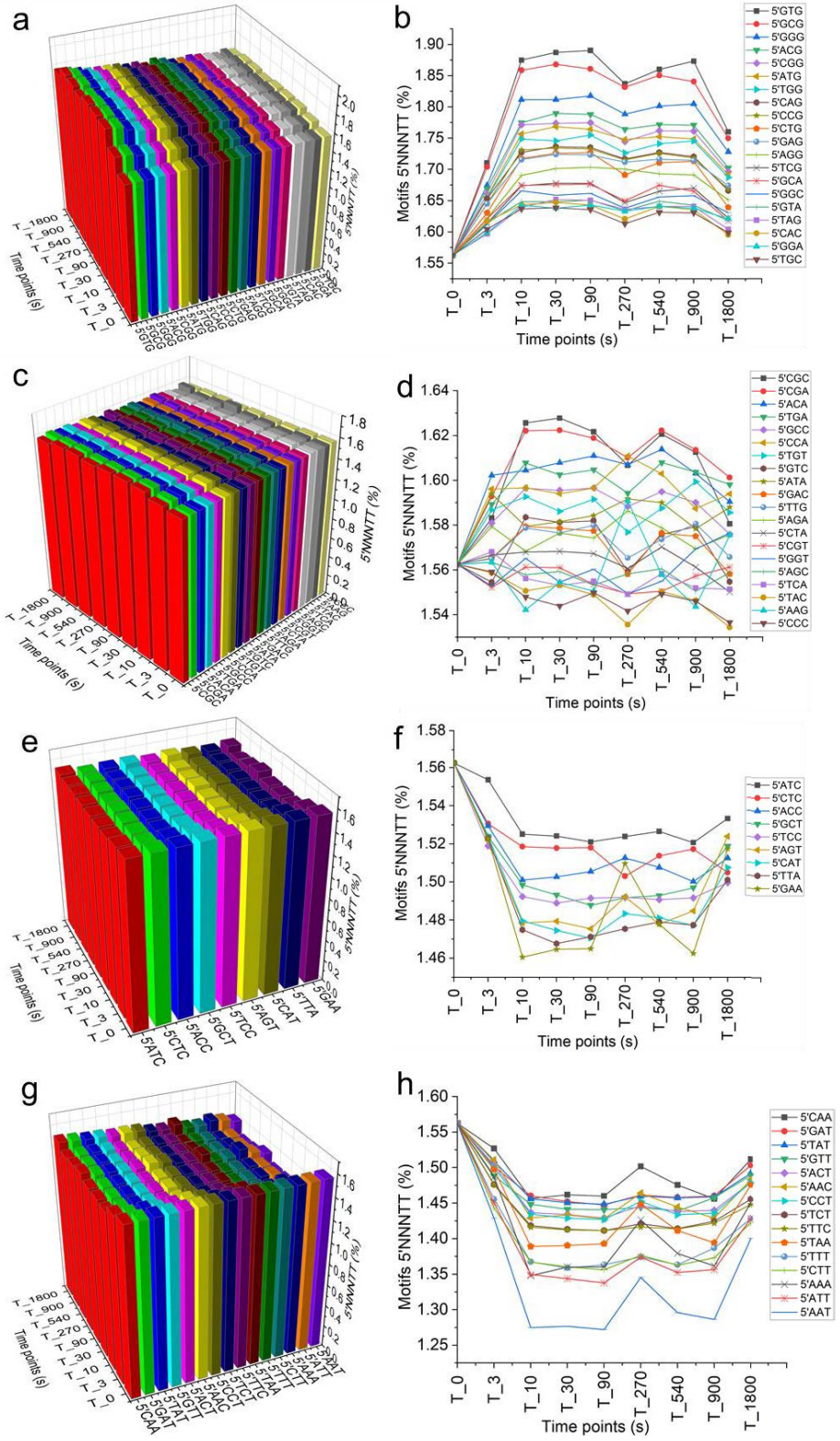


Figure 2.12. The fluctuation of 5' trinucleotide motifs' adjacent to the central thymine pair (5' NNNTT) in the two-replicate experiment at 278 nm irradiation. The data normalized, so the timepoints 0 in all motif sequences start from 1.56%, making the percentage of 1 motif out 64 possible. (a, b) The highest positive 5' trinucleotide motifs' selection is shown in this group. The main observation in this group is dominated by guanine adjacent directly to the thymine pair. (c, d) The s order of positive selection is shown in this group, which is dominated by cytosine or adenine adjacent directly to the thymine pair. The s group is characterized by a small positive selection to constant fluctuation. (e, f) The third group shows a negative selection over time and is dominated by two pyrimidines adjacent to the central thymine pair. (g, h) The fourth group shows the highest negative selection over time and is characterized by adenine and thymine on the two nearest positions to the central thymine pair.

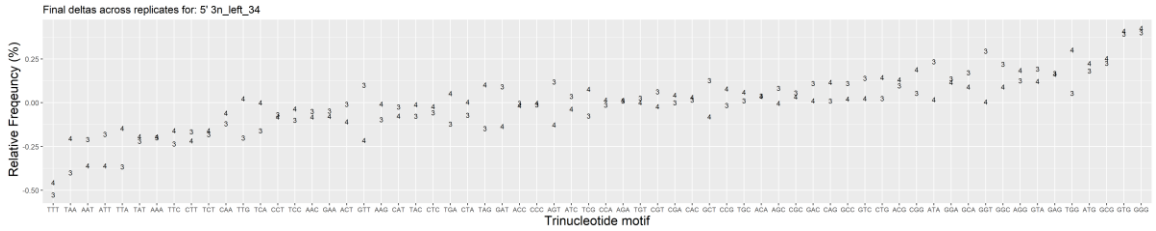


Figure 2.13. Statistical analyses by ANOVA and R studio showing the statistical results of mean values and standard deviations in the data for 278 nm irradiation at the 5’ motif site.

278 nm (3')

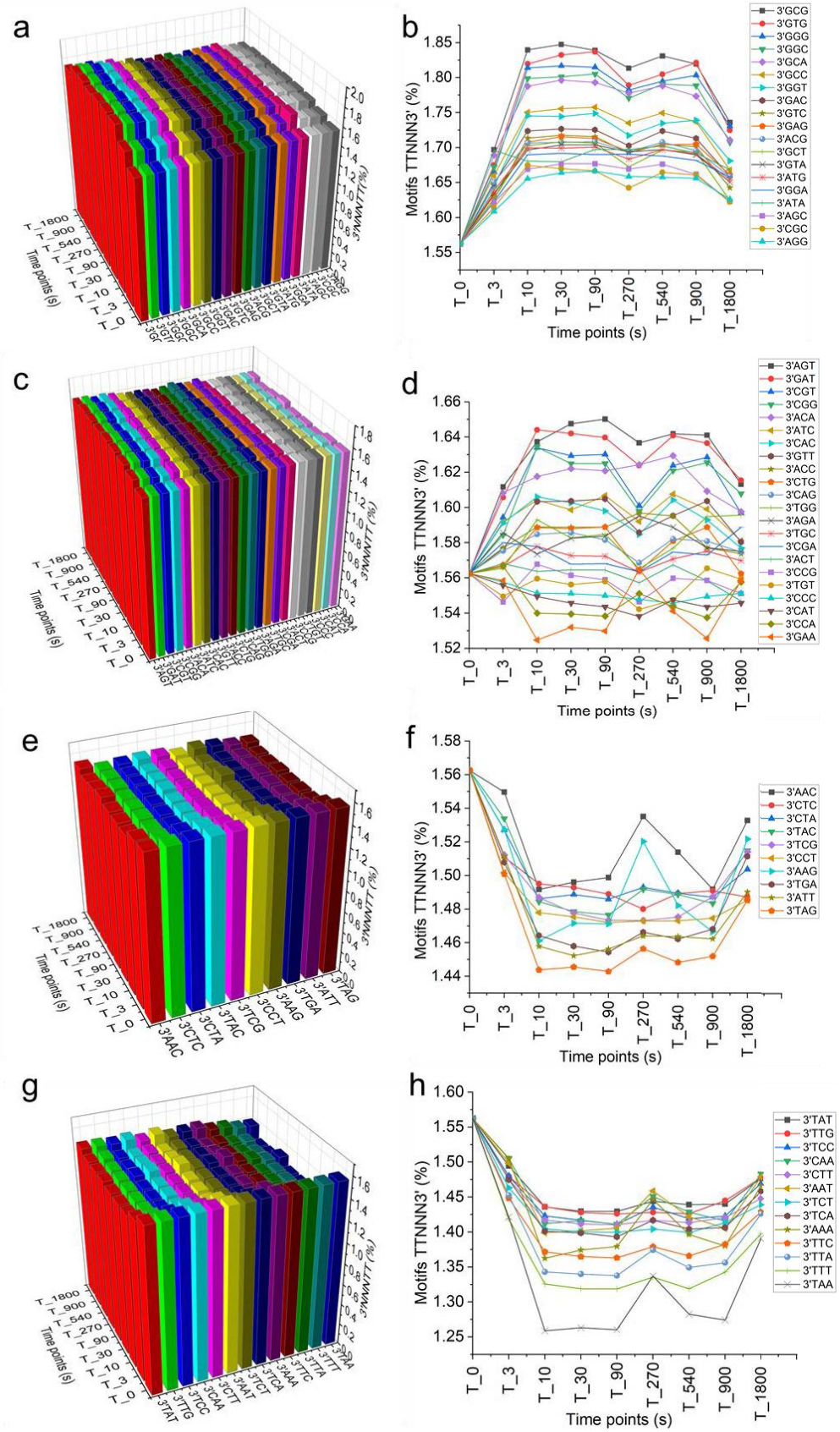

a
b
c
d
e
f
g
h
Time points (s)
Motifs TTNNN3' (%)
T_0
T_3
T_10
T_30
T_90
T_270
T_540
T_900
T_1800
3'GCG
3'GTG
3'GGG
3'GGC
3'GCA
3'GCC
3'GGT
3'GAC
3'GTC
3'GAG
3'ACG
3'GCT
3'GTA
3'ATG
3'GGA
3'ATA
3'AGC
3'CGC
3'AGG
3'AGT
3'GAT
3'CGT
3'CGG
3'ACA
3'ATC
3'CAC
3'GTT
3'ACC
3'CTG
3'CAG
3'TGG
3'AGA
3'TGC
3'CGA
3'ACT
3'CCG
3'TGT
3'CCC
3'CAT
3'CCA
3'GAA
3'AAC
3'CTC
3'CTA
3'TAC
3'TCG
3'CCT
3'AAG
3'TGA
3'ATT
3'TAG
3'TAT
3'TTG
3'TCC
3'CAA
3'CTT
3'AAT
3'TCT
3'TCA
3'AAA
3'TTC
3'TTA
3'TTT
3'TAA

Figure 2.14. This graph shows the fluctuation of 3' trinucleotide motifs' adjacent to the central thymine pair (TTNNN 3') in a two-replication experiment at 278 nm irradiation. Data normalized so that the timepoints 0 in all motif sequences begin at 1.56%, giving the percentage of 1 motif out of 64. (a, b) This group displays the highest positive 3' trinucleotide motif selection. This group manifests itself as a mirror image of 5' fluctuations, also dominated by guanine adjacent directly to the thymine pair. (c, d) We saw the s-order positive selection in this group, which is dominated by cytosine or adenine next to the thymine pair. The s group displays a small amount of positive selection to constant fluctuation. (e, f) A negative selection is evident in the third group, mainly consisting of two pyrimidines surrounding the central thymine pair. (g, h) The fourth group shows the highest negative selection over time and is characterized by adenine and thymine on the two nearest positions to the central thymine pair. The pattern of 3' trinucleotide motifs' fluctuation is obviously a mirror image for 5' fluctuation.

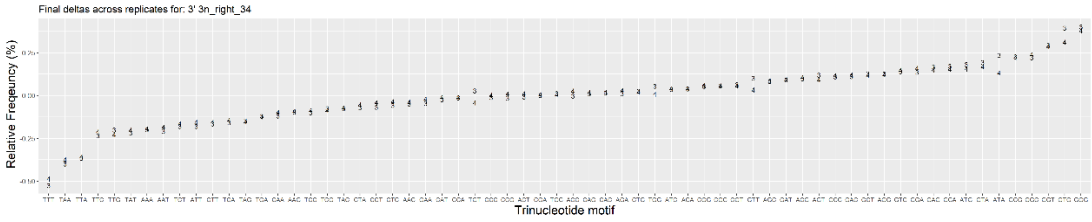


Figure 2.15. Statistical analyses by ANOVA and R studio showing the statistical results of mean values and standard deviations in the data for 278 nm irradiation at the 3’ motif site.

365 nm (5')

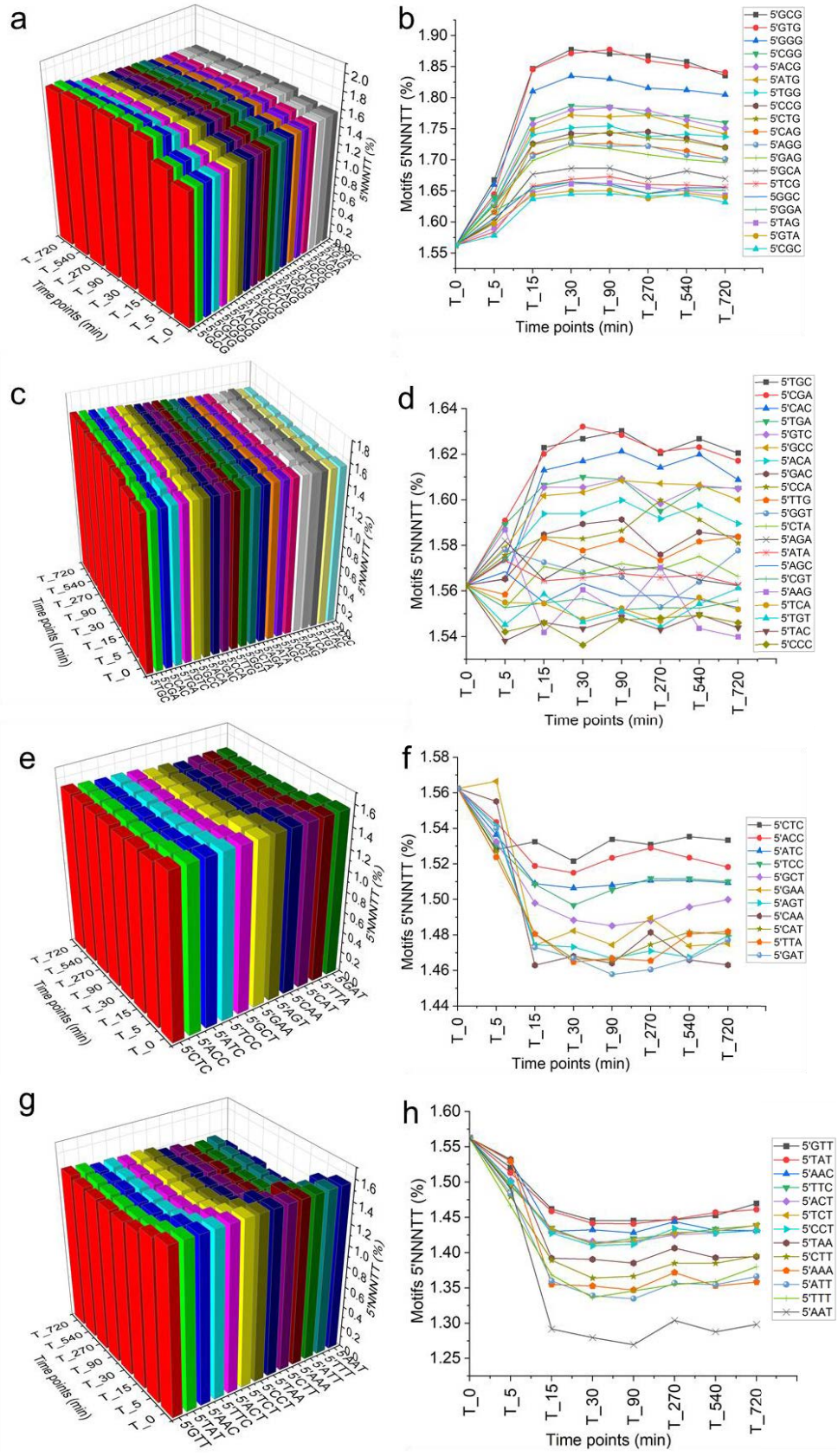

a
b
Motifs 5'NNNTT (%)
Time points (min)
T_0
T_5
T_15
T_30
T_90
T_270
T_540
T_720
5'GCG
5'GTG
5'GGG
5'CGG
5'ACG
5'ATG
5'TGG
5'CCG
5'CTG
5'CAG
5'AGG
5'GAG
5'GCA
5'TCG
5GGC
5'GGA
5'TAG
5'GTA
5'CGC
c
d
5'TGC
5'CGA
5'CAC
5'TGA
5'GTC
5'GCC
5'ACA
5'GAC
5'CCA
5'TTG
5'GGT
5'CTA
5'AGA
5'ATA
5'AGC
5'CGT
5'AAG
5'TCA
5'TGT
5'TAC
5'CCC
e
f
5'CTC
5'ACC
5'ATC
5'TCC
5'GCT
5'GAA
5'AGT
5'CAA
5'CAT
5'TTA
5'GAT
g
h
5'GTT
5'TAT
5'AAC
5'TTC
5'ACT
5'TCT
5'CCT
5'TAA
5'CTT
5'AAA
5'ATT
5'TTT
5'AAT

Figure 2.16. Fluctuation of 5' trinucleotide motifs adjacent to central thymine pair (5' NNNTT) in two replicates under 365 nm irradiation. The data normalized as in 278 nm experiments, so the timepoints 0 in all motif sequences begin at 1.56% (% of 1 motif out of 64). (a, b) These are the most favorable selected 5' trinucleotide motifs. This group is dominated by guanine adjacent directly to the thymine pair in almost identical order to 278 nm irradiation. (c, d) This group shows a s order of positive selection, most of which consists of cytosine or adenine adjacent directly to thymine. A small positive selection is present in the s group, but fluctuation tends to become constant at the end of this group. (e, f) The third group, consisting primarily of two pyrimidines adjacent to a thymine pair, has undergone a negative selection over time. (g, h) Adenine and thymine are found on the two closest positions to the central thymine pair in the fourth group, which shows the greatest negative selection over time.

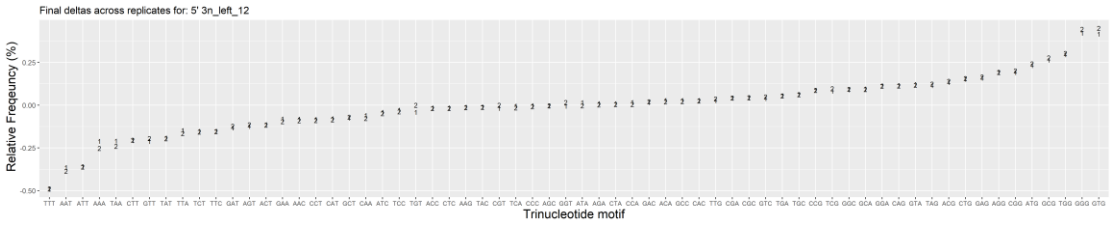


Figure 2.17. Statistical analyses by ANOVA and R studio showing the statistical results of mean values and standard deviations in the data for 365 nm irradiation at the 5’ motif site.

365 nm (3')

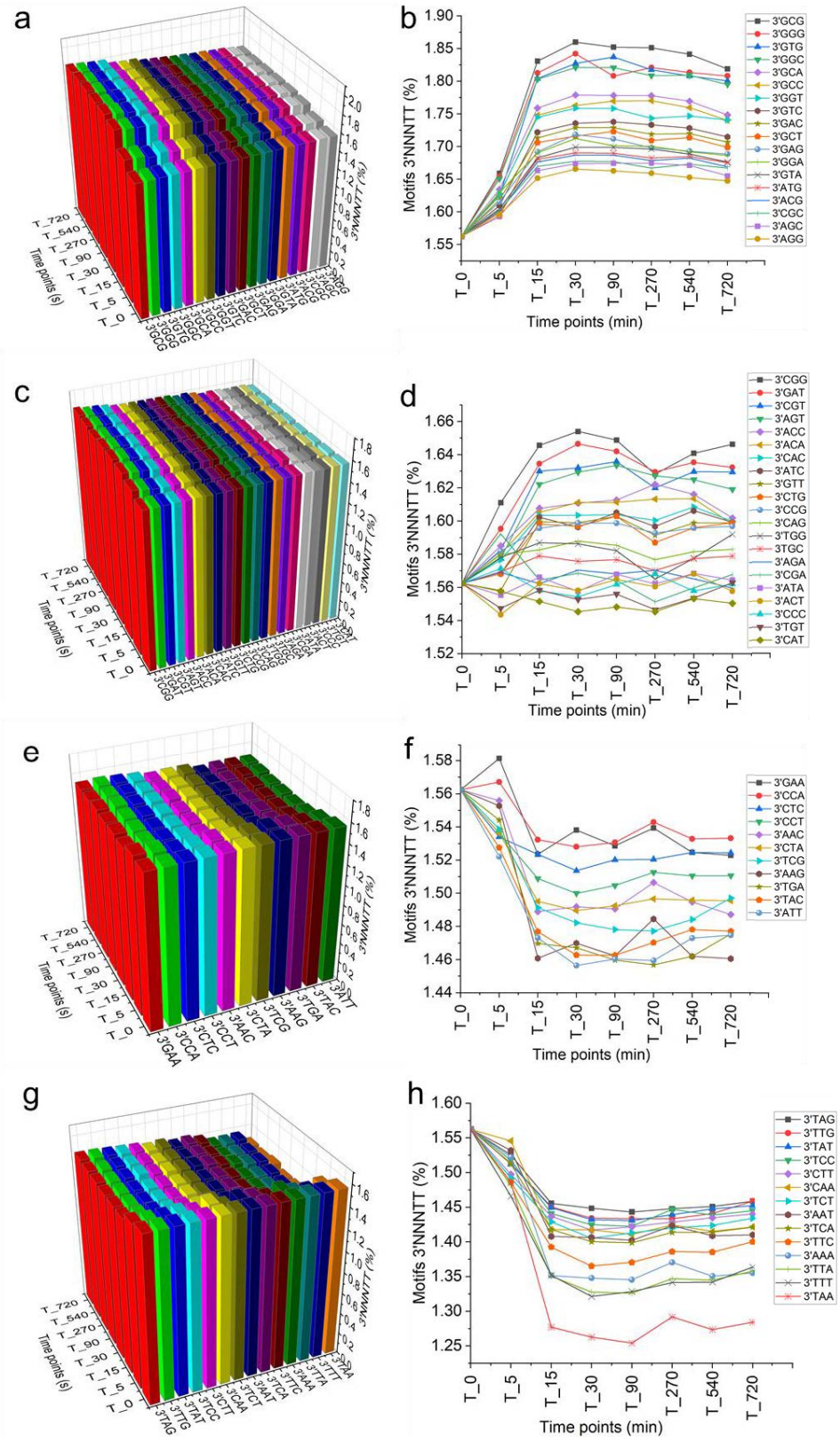

a
b
Motifs 3'NNNTT (%)
Time points (min)
T_0
T_5
T_15
T_30
T_90
T_270
T_540
T_720
3'GCG
3'GGG
3'GTG
3'GGC
3'GCA
3'GCC
3'GGT
3'GTC
3'GAC
3'GCT
3'GAG
3'GGA
3'GTA
3'ATG
3'ACG
3'CGC
3'AGC
3'AGG
c
d
3'CGG
3'GAT
3'CGT
3'AGT
3'ACC
3'ACA
3'CAC
3'ATC
3'GTT
3'CTG
3'CCG
3'CAG
3'TGG
3TGC
3'AGA
3'CGA
3'ATA
3'ACT
3'CCC
3'TGT
3'CAT
e
f
3'GAA
3'CCA
3'CTC
3'CCT
3'AAC
3'CTA
3'TCG
3'AAG
3'TGA
3'TAC
3'ATT
g
h
3'TAG
3'TTG
3'TAT
3'TCC
3'CTT
3'CAA
3'TCT
3'AAT
3'TCA
3'TTC
3'AAA
3'TTA
3'TTT
3'TAA

Figure 2.18. The graph above shows the fluctuation of 3' trinucleotide motifs adjacent to a thymine pair (TTNNN 3') in a two-replication experiment at 365 nm. The data were normalized so that all timepoint 0 in motif sequences began at 1.56%, which gave one motif out of 64 out of 1. (a, b) This group displays the highest positive 3' trinucleotide motif selection and shows a mirror image of 5' fluctuations, dominated by guanine adjacent directly to the thymine pair. (c, d) This group comes s in the positive selection, dominated by cytosine and adenine next to the central thymine pair. There is some positive selection among the s group, as well as a little fluctuation. (e, f) The third group exhibits a negative selection pattern, primarily composed of two pyrimidines adjacent to the central thymine pair. (g, h) Among the fourth group, adenine and thymine appear on the two nearest positions to the central thymine pair, leading to the highest degree of negative selection over time. The pattern of 3' trinucleotide motifs in 365 nm fluctuation is a mirror image for 5' fluctuation and almost the same as 3' trinucleotide motifs in 278 nm.

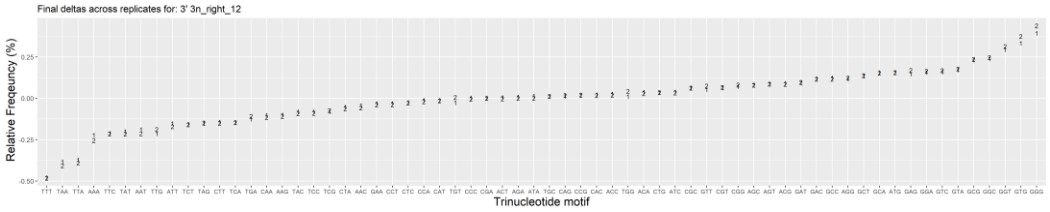


Figure 2.19. Statistical analyses by ANOVA and R studio showing the statistical results of mean values and standard deviations in the data for 365 nm irradiation at the 3’ motif site.

Table 2.12. The motifs fluctuation in 5'NNGTT, TTGNN3', 5'NNATT and TTANN3'.

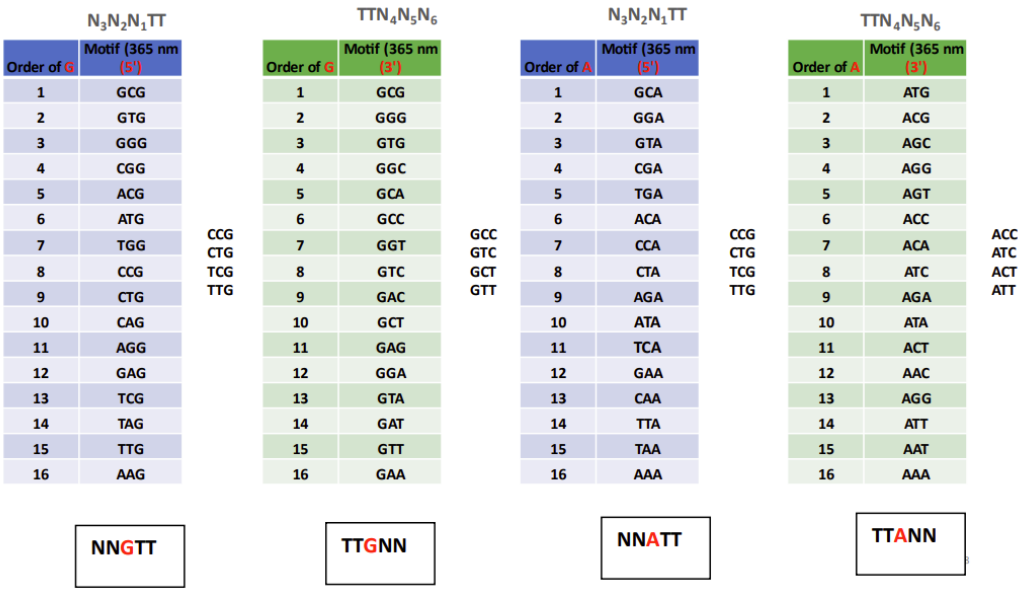
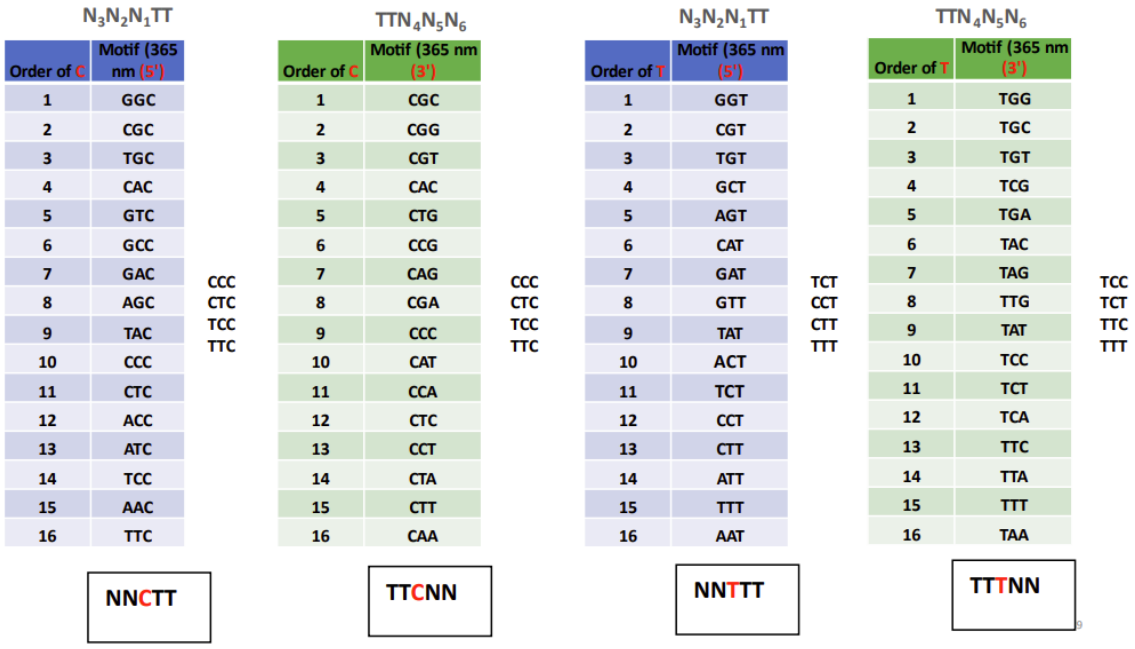


Table 2.13. The motifs fluctuation in 5'NNCTT, TTCNN3', 5'NNTTT and TTTNN3'.



Tables 2.7 and 2.8 show the order of directly adjacent G, A, C and T to thymine pair with the subsequent dinucleotide motif from 5' and 3' sides (5' NNNTT and TTNNN 3'). When the dinucleotide subsequent motif contains two pyrimidine, then the selection order will be as follows: 5' CCNTT > 5' CTNTT > 5' TCNTT > 5' TTNTT and TTNCC 3'> TTNTC 3'> TTNCT 3'> TTNTT 3'.

2.5 Conclusion

In summary, our study provides a novel comprehensive and systematic investigation of the effect of flanking sequence on TT-CPD formation and repair by direct UVC irradiation and UVA irradiation in the presence of photosensitizer (2′-Methoxyacetophenone). The overall pattern of NTTN motifs fluctuation under the two different irradiation settings appears similar, with TTTT and ATTT always the hottest spot (i.e. favoring CPD formation by the adjacent TT pair), showing the highest negative selection over time, and GTTG being the coldest spot (showing least formation of CPD formation by the adjacent TT pair). In agreement with previous studies, we observed a slightly higher suppressive effect of 5' flanking guanine on TT-CPD formation than 3' flanking guanine with all other bases rather than thymine.[13,115,141,200–202] Guanine with cytosine and adenine showed more positive selection when guanine on 5' side, which supports the postulation" 5' suppressed CPD formation more than a 3' due to better pi stacking of the G with the 5' pyrimidine".[115,202,221] However, when thymine flanking the thymine pair with guanine on the other side, in this case, guanine from 3' side shows an equal or even slightly higher suppressing effect than guanine from 5' side (TTTG≥ GTTT). This finding is entirely different from any previous studies. In general, guanines directly flanking the central thymine pair show increased DNA photostability and relatively higher levels of repair of the neighboring thymine dimer. Meanwhile, cytosines and adenines show slower formation and repair rates than thymine and comparable fluctuation over time.

We studied the motif fluctuation adjacent to the central thymine pair from both 5' and 3' sides to investigate a particular preference for charge transfer or exciplex formation. Motifs containing guanines directly adjacent to the thymine pair from both 5' and 3' sides show the highest positive selection, particularly these three motifs (GCG, GTG, GGG). Motifs containing adenines and cytosines directly adjacent to the thymine pair from both 5' and 3' sides show the mostly negative selection except for a few sequences. The pattern and fluctuation of the trinucleotide motifs imply the main factors for the formation and reversal of CPDs could be the following: First, better base stacking, guanine especially on 5' side suppressed CPD formation more than any other nucleotides due to better pi stacking and more favorable geometry overlap at 5' side.[115,202,221] The second factor is the free

energy of stacked bases, which apparently controls the tendency of motif fluctuation. Adenine has the highest stacking free energy[222] and shows the most negative selection in (5' NNNTT and TTNNN 3') motifs. The third factor is the likelihood of forming another CPD (presence of two adjacent pyrimidines). The motif fluctuation in the middle looks like it is governed to a large extent by the presence or absence of two consecutive pyrimidines, which leads to an increase in the likelihood of formation of another CPD rather than the central thymine pair. In accordance with the previous studies, the motifs containing two consecutive pyrimidines show less positive selection and tend to form the CPDs in the following order: NCC**TT** < NCT**TT** < NTC**TT** < NTT**TT** 5' side, and **TT**CCN < **TT**TCN< **TT**CTN < **TT**TTN 3' side.

Furthermore, there is no apparent direct correlation between the 5' side and 3' side on formation and repair CPDs in dsDNA. It seems to be the mechanism of formation and repair CPDs from 5,' and 3' sides are identical (fluctuation of different motifs from 5' side are almost mirror image to 3' side). The plausible mechanism could be photons or delocalized singlet state, with a minor contribution from triplet states, transfer from nearest 5' or 3' side bases to the central thymine pair to cause their dimer formation or repair. The 3' side shows an almost mirror image to the 5' side, suggesting possibly the exact mechanism for dimer formation and its photo-reversal.

Chapter 3. Kinetic Analysis of Selected Trinucleotide Motif Sequences to Evaluate the Role of Base Stacking in the Formation and Repair of Cyclobutane Pyrimidine Dimers

3.1 Abstract

We studied the kinetics of the formation and repair of thymine-thymine cyclobutane pyrimidine dimers in 10 distinct 114-base pair DNA sequences containing central thymine pairs irradiated at 278 nm. The ten different 114-base pair DNA sequences were selected from the random dsDNA library according to their trinucleotide motif tendency selection in two UV-irradiation time-course experiments. We pick the two highest winning trinucleotide motif sequences that offer protection against CPD formation, two mid-level sequences that neither favor nor disfavor CPD formation and one low-level sequence that favors CPD formation from the 5' and 3' flanks. The TT-CPD was assayed by T4 pyrimidine dimer glycosylase (T4 PDG) and separated by gel electrophoresis. There was a considerable variation in yields of the TT-CPD based on trinucleotide motif sequences, but the yields of a given motif were similar at both flanking sites (5' and 3' sides). The kinetic analysis of TT-CPD formation in the ten different sequences reveals that the formation and reversal of photoreactions depend remarkably on the identity of the nearest-neighbor bases on the 5' and 3' sides of the central TT bases. Our results show that the central TT pair within the (AAGTTGAA) motif reveals the lowest TT-CPD formation at the photostationary state. On the other hand, the ATATTAGA motif exhibits the highest TT-CPD formation at the photostationary state. In general, the results indicate that interactions with nearest-neighbor bases strongly influence CPD formation and photoreversal probabilities.

3.2 Introduction

Electron transfer (ET) is one of the simplest chemical reactions that plays a vital role in a number of life-sustaining processes (e.g., respiration and photosynthesis) in living organisms.[223] Detailed investigations of biomolecules and other σ-bonded chemical networks have provided an understanding of thermodynamics, reorganization energies, and distance for ETs.[224–226] It is worth noting that σ bond-mediated ET is mainly limited to distances less than 15–20 Å as a result of the sheer drop in the rate of electron transfer with distance. However, work over the last decade has shown that charge transport through extended molecular pi stacks can take place over distances much longer than 15–20 Å.[227] A need, therefore, exists to develop a conceptual framework to understand the new long-range charge transfer events.

Super-exchange interactions between the electron donor (D) and acceptor (A) are responsible for the long-range electron transfer in proteins and other σ-bonded frameworks.[227] The intensity of this interaction decreases with an increase in the number of atoms between the redox sites, causing the rate of electron transfer to decline rapidly with an increase in the donor-acceptor distance. On the other hand, it has been observed that charge transport through DNA occurs at distances of up to 200 Å, indicating that decay with distance is extremely shallow.[228] Understanding these long-range events in DNA, therefore, remains a major challenge. A number of studies suggest that inhomogeneities, base-pairs dynamics and energetics within π stacks of DNA play vital roles in distinguishing DNA-mediated charge transport.[227]

DNA has intrinsic evolutionary photostability due to the energetic cost of repairing DNA photolesion. A single base dissipates electronic energy non-radiatively in a subpicosecond timeframe, but base stacking and base pairing play crucial roles in conducting the decay of excess electronic energy in the double helix.[6] However, Kohler *et al.*, have shown that the fate of singlet electronic states in single and double-stranded DNA fragments made from adenine (A) and thymine (T) nucleobases is determined by vertical stacking rather than base pairing.[19] Quantum chemical calculations have revealed that the energetic contributions from base stacking to stabilize the DNA are comparable to those of H-bonding.[229,230] Several experiments have demonstrated that rather than base pairing, base stacking mainly determines the stability of the DNA duplex. The energetic contributions

to the stacked A//T and G//C dimers are 11.6 and 16.9 kcal mol^{-1}, while H-bonded A:T and G:C WC pairs contribute 15.4 and 28.8 kcal mol^{-1} to energy, respectively.[231] Additionally, molecular dynamics simulations on DNA dodecamers have shown that dispersion energy plays a crucial role in forming its double-helical structure.[232]

A variety of non-covalent interactions including base stacking affect the stability, structure, sequence-dependent physicochemical properties, and biological functionality of nucleic acid molecules.[233] The long-range electron transfer via nucleobase π-stacks is one of the most significant phenomena in this context. The electron transfer through π-stacks is an intrinsic (spontaneous) process. The ET processes proceed without the need for any charge injection into DNA that could result in oxidation or reduction of nucleobases.[234] Intrinsic ET processes in DNA are related to both base stacking and base pairing H-bonds, which run parallel and perpendicular to the helical axis of DNA, respectively.

It has been found that UV radiation can play a dual role in forming CPDs and photo-reversal them to the starting dipyrimidine bases. Direct excitation of a CPD with UV light results in its photoreversal with a quantum yield of ~100%. As CPDs absorb shorter wavelengths than undamaged nucleobases, photoreversal is expected to be more prominent at shorter wavelengths of UV (i.e., UVC).[235] However, CPDs containing cytosine (TC, CT, and CC) may have a greater absorption cross-spectrum of UVB wavelengths than TT-CPDs. This claim of direct photosplitting warrants further study to investigate the role of UVB, given that a major number of prior model studies have been performed with UVC irradiation.[236,237]

We want to test the effect of adjacent purines in the context of flanking sequences and test our hypothesis: base stacking and the redox potential of neighboring bases play the most critical roles in photorepair and photoformation of CPD thymine dimer.

3.3 Materials and Methods

3.3.1 Materials

A total of 10 different oligomers single-strand DNA sequences and forward and reverse primers were purchased from Integrated DNA Technologies (IDT), [γ-^{32}P] ATP from Perkin Elmer, T4-PDG (pyrimidine dimer glycosylase), from New England Biolabs

(NEB). We kept the DNA library length at 114 bp by adding GC and CG to 5' and 3' respectively.

3.3.2 Preparation of ^{32}P-labeled DNA duplexes

A forward primer was radiolabeled on the 5'-terminus with γ-[^{32}P]-ATP and T4 polynucleotide kinase following standard procedures. The ^{32}P-labeled DNA duplexes were prepared by 20 rounds PCR as follows: Multiple tubes of 100 µL PCR reaction mixture contained 1x KAPA Hifi Taq DNA polymerase buffer, dNTP mix 0.2 mM, forward primer 0.2 µM, reverse primer 0.4 µM, KAPA Hifi DNA polymerase (5 U/ µL) 1.25 Unit and PCR grade water to 100 µL. The amplification condition was for one cycle as follows: initial denaturation at 95 ℃ for 2 min, denaturation at 95 ℃, annealing at 54 ℃ for 30 s and extension at 72 ℃ for 5 min. The 10 different DNA duplexes were then purified by 8% native polyacrylamide gel and stored in (20 mM NaCl/ 10 mM Tris-HCl, pH 8.0) buffer.

3.3.3 UV irradiation

The irradiation solutions were each 200 µL (20 mM NaCl/ 10 mM Tris-HCl, pH 8.0) buffer solution containing either cold or 5'-radiolabeled 114 bp duplex DNA samples. All irradiation was carried out in a cold room at 4 °C using a 278 nm, 2.4 W LED purchased from IRTronix LG. The UV exposure times ranged between 3 s and 15 min (0 s, 3 s, 10 s, 30 s, 90 s, 270 s, 540 s, and 900 s). All irradiations and analyses were performed in duplicates.

3.3.4 T4 PDG digestion and denaturing PAGE analysis of CPD formation.

3.3.5 Quantification and data analysis

The oligodeoxynucleotides were digested with T4 PDG (NEB) after irradiation and separated by 20% denaturing PAGE. The separated oligodeoxynucleotides were analyzed by phosphorimagry using ImageQuant. The ImageJ plugin (NIH) was used for the quantitation of the data in which a lane box method yields the histograms of entire lanes. The quantity of T<>T formation was determined by dividing the amount of radioactivity associated with the DNA fragments generated by the endonuclease to the total radioactivity of this fragment and its remaining parent strand. The intensities of each band in a lane were selected and quantified.

3.3.6 Kinetics calculations

The obtained data points were fit using a nonlinear BoxLucas1 model in the Origin Pro software where the formation of the T<>T was plotted as a function of irradiation time (t). The concentration of the parent dipyrimidine site is noted as TT, while that of the photodimer is indicted as T<>T. The rate constants for the forward reaction (dimerization) and reverse reaction (monomerization) are represented as formation constant (k_f) and repair constant (k_r), respectively. The values of the k_f and k_r were corresponding to the "a" and "b" functions of the fit obtained for time vs concentration. Here, the decay constant (k_{obs}) was a sum of k_r and k_f, and the ratio of k_r to k_f impel photostationary state (k_{eq}) at the equilibrium.[110]

At the photostationary state,

$$TT \underset{k_r}{\overset{hv\ k_f}{\rightleftharpoons}} T\langle\rangle T \quad (hv)$$

$$[T\langle\rangle T]\cdot k_r = [TT]\cdot k_f,$$

therefore,

$$k_r/k_f = [TT]/[T\langle\rangle T]. \quad (a)$$

$$k_r = [TT]/[T\langle\rangle T].\ k_f \quad (b)$$

$$k_{obs} = k_r + k_f \quad (c)$$

By substitution equation (b) in (c) to get k_f and k_r.

$$k_{obs} = k_f + [TT]/[T\langle\rangle T].\ k_f$$

Values of k_f and k_r were calculated for each sequence by considering the corresponding values of K_{eq} (i.e., value of the term 'a' obtained after the fitting) and k_{obs} (i.e., value of the term 'b' obtained after the fitting).

3.4 Result and Discussion

A total of 10 distinct 114-base pair DNA sequences containing central thymine pairs were irradiated at 278 nm. The ten different 114-base pair DNA sequences were selected from the random dsDNA library according to their trinucleotide motif tendency selection in our two UV-irradiation time-course experiments. We pick the two highest winning trinucleotide motif sequences that offer protection against CPD formation (5' GGG, 5' GCG, GGG3', GCG 3'), two mid-level sequences that neither favor nor disfavor CPD formation (5' ATA, 5' AGA, ATA 3', AGA 3') and, and one low-level sequence that favors CPD formation from the 5' and 3' flanks (5' AAG, GAA 3'). A combination of all these motifs was designed to study the effect of these surrounding motifs on the formation and repair of T<>T by using T4 PDG digestion and denaturing gel-based separation and analysis. The design of the ten oligonucleotides is presented in Table 3.1.

A total of 10 duplexes was prepared by PCR using (forward primer: TACGCGTGTATACATACACA and reverse primer: CATGTGCGTGTATGTATGTG) as mentioned in the material and methods. All DNA duplexes were ^{32}P radiolabeled and irradiated at 278 nm as mentioned above and subjected to digestion with T4 PDG and separated onto 8% denaturing gel (Figures 3.1 and 3.2).

The rate constants, photostationary (steady-state) levels of T<>T were calculated for all different Duplexes as described in Kinetics calculation section (Figures 3.3, 3.4 and 3.5).

Table 3.1. Design of the group of 10 sequence motifs used to study the kinetics of T<>T formation and repair.

Sequence (motif)	**5' side**	**3' side**
GCGTTGCG	Good	Good
GGGTTAGA	Good	Medium
ATATTAGA	Medium	Medium
ATATTGCG	Medium	Good
GCGTTAGA	Good	Medium
GGGTTGGG	Good	Good
AAGTTGCG	Low	Good
AAGTTAGA	Low	Medium
AAGTTGAA	Low	Low
ATATTGGG	Medium	Good

Note: Two highest winning trinucleotide motif sequences that offer protection against CPD formation (Good) (5' GGG, 5' GCG, GGG3', GCG 3'), two mid-level sequences that neither favor nor disfavor CPD formation (Medium) (5' ATA, 5' AGA, ATA 3', AGA 3') and, and one low-level sequence that favors CPD formation from the 5' and 3' flanks (Low) (5' AAG, GAA 3').These motifs were selected according to motif fluctuation in Figures 2.12 and 2.14 in chapter 2.

The kinetics analysis by time-dependent UV irradiation gave insight into the dimer formation rate and monomerization rate (Table 3.2). These two rates were calculated from the K_{eq} (steady-state) and k_{obs} ($k_{obs} = k_f + k_r$). It is worth noting that the k_f and k_r for the ATATTAGA are relatively higher than all other sequences. Contrary to this, the k_r and k_f values for GGGTTGGG sequences show the lowest dimerization rate and monomerization rate. These observations imply that the surrounding bases affect the DNA photorepair. These kinetics analyses reveal that surrounding bases may have a role in determining the excited state dynamics of T<>T, affecting the transfer or delocalization of excited states.

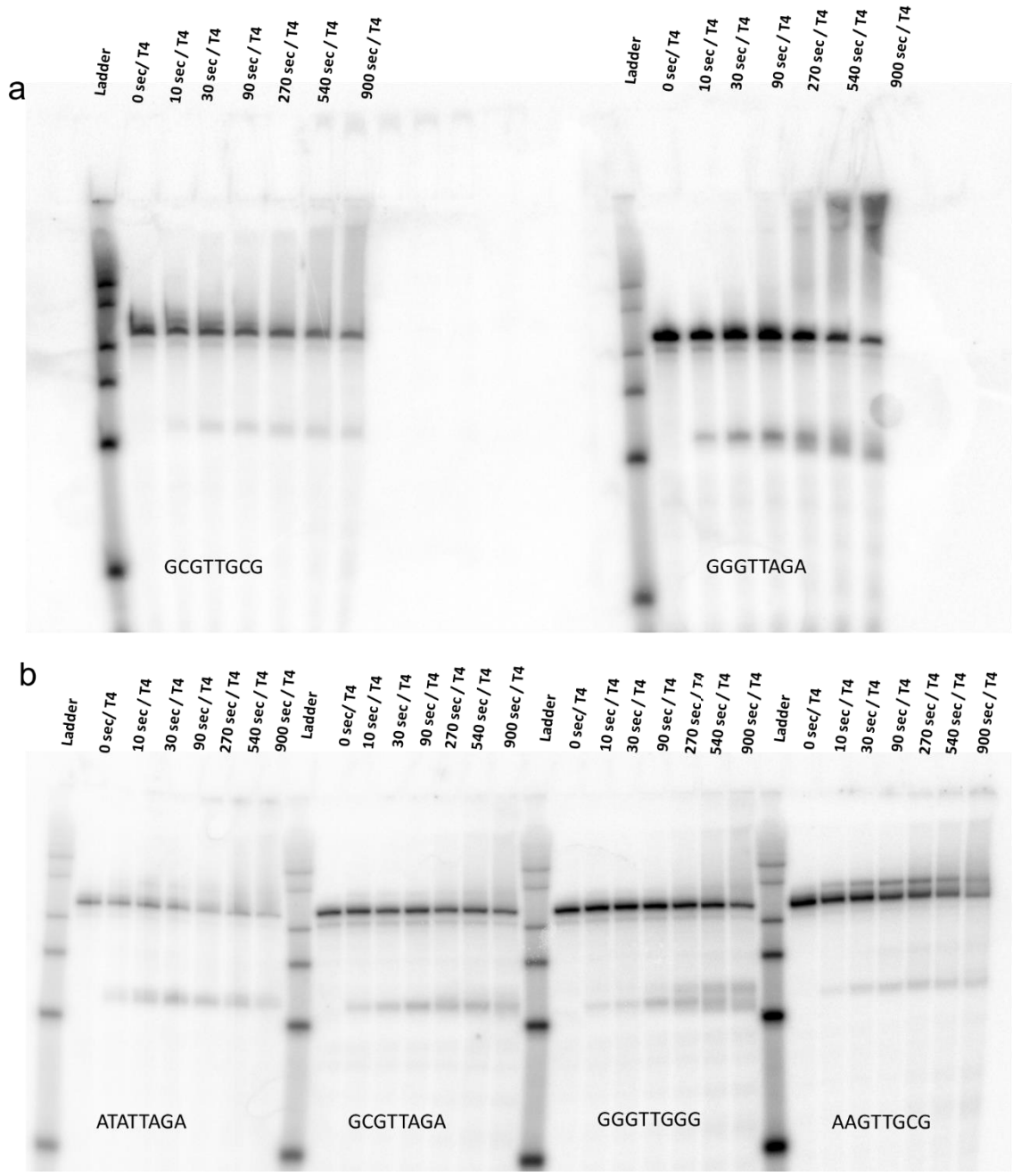


Figure 3.1. Electrophoretic separation and analysis of T4 PDG treated DNA. Time-dependent experiments and T4 PDG digestion of GCGTTGCG, GCGTTAGA, ATATTAGA, GCGTTAGA, GGGTTGGG and AAGTTGCG, respectively.

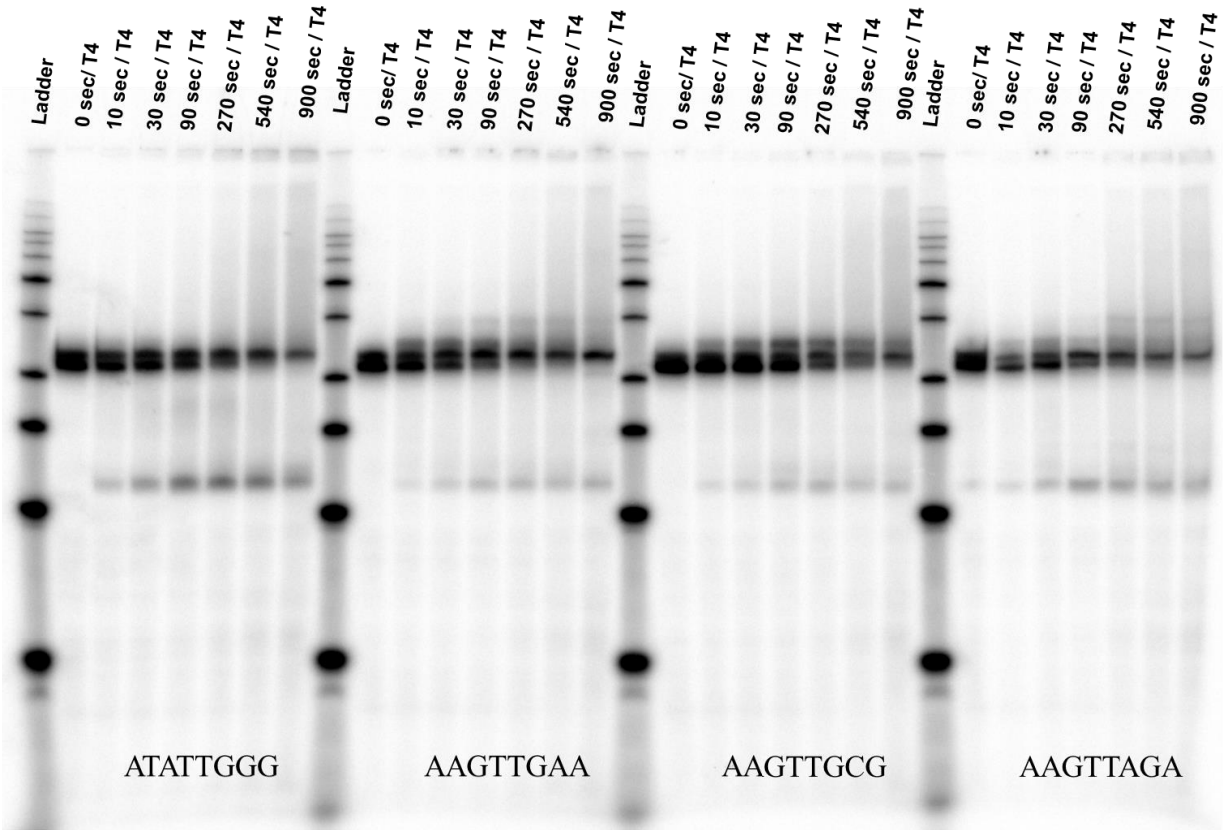


Figure 3.2. Electrophoretic separation and analysis of T4 PDG treated DNA. Time-dependent experiments and T4 PDG digestion of ATATTGGG, AAGTTGAA, AAGTTGCG and AAGTTAGA, respectively.

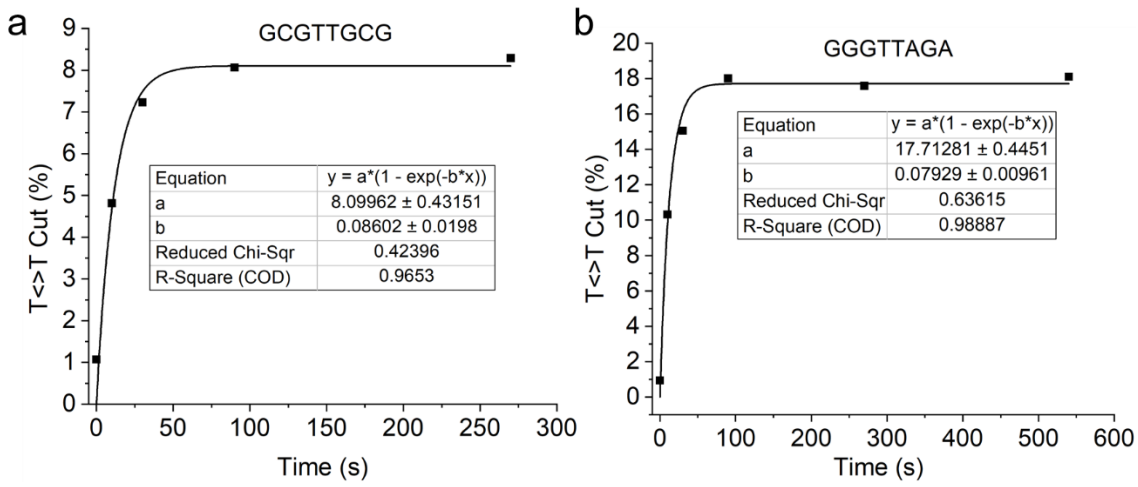


Figure 3.3. The representative kinetic analyses of the formation and repair of the T<>T where the percent fraction of T<>T containing DNA was quantified and plotted as a function of irradiation time. The data points were fit by considering a reversible synchronization between the rate constants of formation (k_f) and repair (k_r) for T<>T (Equation 3.1).

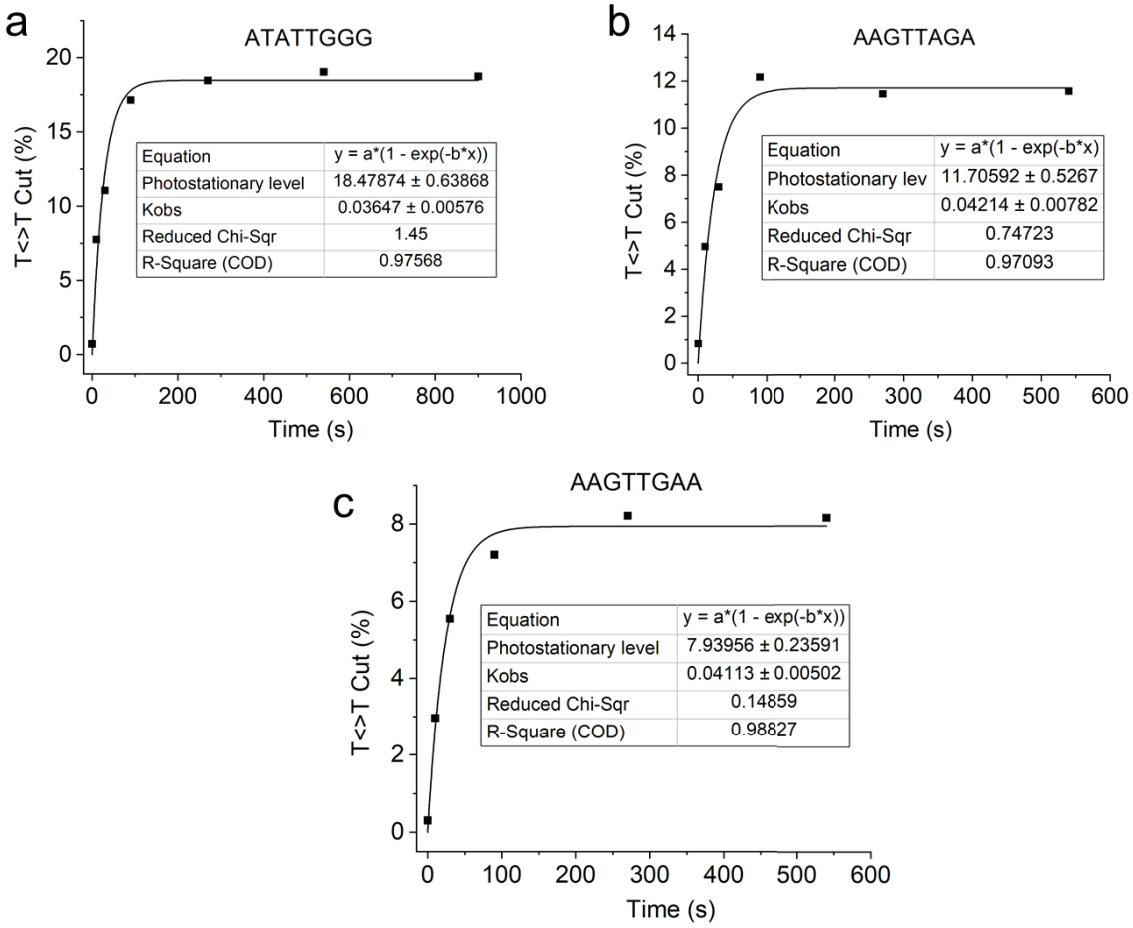


Figure 3.4. The representative kinetic analyses of the formation and repair of the T<>T where the percent fraction of T<>T containing DNA was quantified and plotted as a function of irradiation time. The data points were fit by considering a reversible synchronization between the rate constants of formation (k_f) and repair (k_r) for T<>T (Equation 3.1).

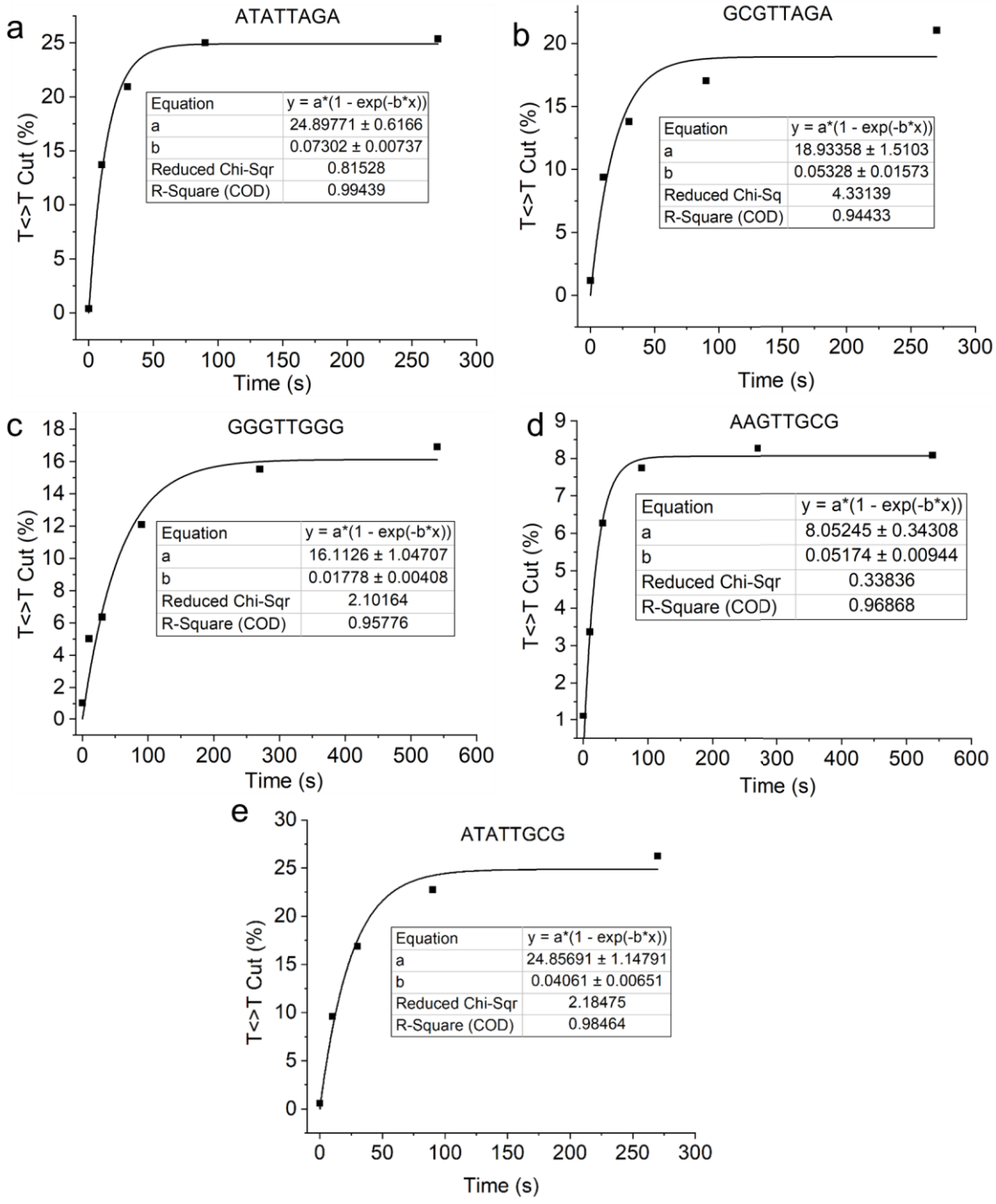


Figure 3.5. The representative kinetic analyses of the formation and repair of the T<>T where the percent fraction of T<>T containing DNA was quantified and plotted as a function of irradiation time. The data points were fit by considering a reversible

synchronization between the rate constants of formation (k_f) and repair (k_r) for T<>T (Equation 3.1).

Table 3.2. Kinetic rate constants and photostationary states of DNA duplexes by an irradiation at 278 nm. Here, as the value of k_r represents the monomerization of T<>T, which is backward reaction.

Sequence (motif)	k_f (s^{-1})	k_r (s^{-1})	k_{obs} (s^{-1})	k_r/k_f	Photostationary (steady-state) levels of T<>T (%)
AAGTTGAA	0.003	0.038	0.041	12.7	7.9
AAGTTGCG	0.004	0.048	0.052	12.0	8.0
GCGTTGCG	0.007	0.079	0.086	11.3	8.1
AAGTTAGA	0.005	0.037	0.042	7.4	11.7
GGGTTGGG	0.003	0.015	0.018	5.0	16.1
GGGTTAGA	0.014	0.065	0.079	4.6	17.7
ATATTGGG	0.007	0.029	0.036	4.1	18.5
GCGTTAGA	0.010	0.043	0.053	4.3	18.9
ATATTGCG	0.010	0.031	0.041	3.1	24.9
ATATTAGA	0.018	0.055	0.073	3.1	24.9

Note: In Table 3.2, the values of k_f and k_r represent the rate of forward (dimerization) and backward (monomerization) reactions, respectively. The values of k_{eq} are corresponding to the photostationary states. k_{obs} represents the reaction rate constant, which is the sum of k_r and k_f at the photostationary state.

The result of T<>T quantification and kinetic analysis of the ten different sequence duplexes were presented in Table 3.2. All the sequences were arranged from lowest to highest T<>T formation at the photostationary state. The percentage of T<>T formation at the photostationary state fluctuated from 7.9% for the AAGTTGAA sequence to 24.9% for the ATATTAGA sequence. The percentage of T<>T formation at the photostationary state for all sequences was obviously governed by the k_r/k_f ratio. The highest k_r/k_f ratio indicates the lowest T<>T formation at the photostationary state. The main observations of the data in this table are that the rate constants of repair (monomerization) (k_r) are higher than the rate constants of T<>T formation (dimerization) (k_f) in all sequences. Surprisingly, the AAGTTGAA sequence shows the highest k_r/k_f ratio and the lowest T<>T % at the photostationary state where 5'AAGTT and TTGAA 3' motifs were the lowest selection among 5'NNGTT and TTGNN3' motifs in the NGS data.

It is notable that, even though the GGGTTGGG sequence exhibits the lowest rate of T<>T formation, the flanking guanines sequence (GGGTTGGG) also exhibits the lowest rate of

repair and a moderate level of T<>T formation. This result is different from what we already found in our NGS time-course experiments, in which GCG, GGG, and GTG motifs exhibit the highest positive selection and disfavoring T<>T formation. The low constant formation rate (k_f) of T<>T for the GGGTTGGG sequence can be attributed to suppressing the thymine excited state by flanking guanines. At the same time, the lack of difference in oxidation potential between the tri-guanine motif inhibits the efficient charge transfer required for photorepair, which causes the lowest repair constant rate (k_r) among all the sequences.

According to the % of T<>T at the photostationay state of the different DNA sequences studied, the five different motifs can be classified from lowest forming (disfavoring T<>T formation) to highest forming (favoring T<>T formation) as follows: 5'AAG > 5'GCG > 5'GGG > 5'AGA> 5'ATA. We were expecting these results, as all previous literature indicates flanking guanines suppress thymine formation more than adenines. Nevertheless, we did not expect that the 5'AAG motif suppresses the T<>T formation and shows the highest k_r/k_f ratio. In the time-course irradiation experiments, 5'AAGTT and TTGAA3' motifs were the lowest selection among the NGS data for 5'NNGTT and TTGNN3' motifs (Table 2.7). We can explain the disagreement between the NGS results and the AAAGTTGAA sequence CPD quantification by the complexity of the random DNA library and the possibility that '5AAG and GAA3' motifs may occur in combination with other low selection motifs.

Adenine directly adjacent to thymine pair motifs (ATA, AGA) show more T<>T yield % at the photostationary state than when there is genuine directly adjacent to the thymine pair. Formation of T<>T is more evident when the adenine-containing motif flanking the thymine pair from 5' side (ATATTAGA > ATATTGCG> GCGTTAGA>ATATTGGG> GGGTTAGA> AAGTTAGA). Our result. is consistent with Rokita *et al.,* 2007 and Kholer *et al.,* 2013 finding that 5' GTTA suppresses dimer formation more than ATTG.[110,114]

Based on evidence from literature, neighboring bases may influence the lifetime and energy of singlet and triplet excited states through excimer formation and possibly delocalization. However, it is not well-known how neighboring bases impact thymine dimerization.[110,200] Hence, the observed levels of T<>T are determined by competition

between forward and reverse reactions. Previous studies have attributed the relationship between the TT dimer yield and the oxidation potential of flanking purines to excited state electron transfer processes.[110,201] The flanking purines effect observed by Rokita *et al.,* 2007*,* was attributed to purine-sensitized dimer repair when 254 nm irradiation was used.[110] Moreover, the electron transfer from the excited purine to the TT dimer was also postulated to result in the anion radical cleavage of the TT dimer in an analogy study of flavin repair mechanism.[113] Cannistraro and Taylor postulated that the purines repairing effect is caused by the quenching of the pyrimidines' excited state by the flanking purines through an electron transfer mechanism in which the purines act as electron donors and the pyrimidines as electron acceptors.[201] Their model proposes that excited state electron transfer produces an exciplex, which decays to the ground state through nonradiative decay, rather than TT dimer or fluorescence of exciplexes formation.

According to Cannistraro and Taylor, the greater quenching efficiency by 5′- versus 3′-purines may be explained by better geometry π-overlap between the flanking 5′-purine and the adjacent thymine comparing with 3'-purines.[201] In addition, Z Pan *et al.,* 2011, have shown that thymine photodimerization values for a group of short DNA single-strand and base-paired hairpin constructs having a pair of thymine with flanking purines depend on the flanking purine and its oxidation potential. The 5'-G directly adjacent to TT may affect photostationary levels of T<>T to a certain degree by inducing changes in excitation transfer or excited-state delocalization (3,12,17).[110,138,238] However, how these might affect dimerization or repair to alter T<>T photostationary level is not well understood.

Rokita *et al.,* 2007 reported that no difference in photostationary level had been found between 5'-GTTA- 3 'and 5'-ATTG-3' in single-strand DNA, which he attributed to the heterogeneity of these systems.[110] Nevertheless, lower T<>T photostationary levels were correlated with 5'-GTTA-3' short DNA duplexes, and higher levels correlate with 5'-ATTG-3' short DNA duplexes. Moreover, the photostationary level of T<>T declined by more than 50% by swapping the flanking from 5'-ATTG-3 to 5'-GTTA-3 in these DNA duplexes. Our result also shows a slightly higher suppression of thymine dimer formation from the 5' side than the 3' side, in agreement with this finding. In general, a relatively low T<>T level was maintained when 5'G was flanking the central thymine pair, except for the GCGTTAGA sequence. Contrarily, a relatively high T<>T level was maintained when 5′

A was flanking the central thymine pair. The ability of a nucleobase within duplex DNA to promote repair of a neighboring T<>T may also be related to its oxidation potential. Our results are consistent and agreeable to both Rokita's findings and this assumption. According to the oxidation potential assumption, ATATTAGA and ATATTGCG show the highest photostationary level, while GCGTTGCG and ATATTGCG show the lowest. GCGTTGCG also shows the highest rate of repair among all ten sequences. On the other hand, ATATTAGA shows the highest rate for dimer formation (k_f). The most noticeable is that GGGTTGGG shows the lowest rates for the formation and repair of thymine dimer. It has been demonstrated that the distance dependence of a DNA mediated electron transfer reaction is not only sensitive to the identity of donor and acceptor but is also affected by the stacking environment of the donor and acceptor.[110,221] Our result also indicates that thymine dimer formation is sensitive to the location of the flanking purine and their oxidation potentials. Guanine has the lowest oxidation potential among the nucleobase and, therefore, shows the most efficient charge transfer. In some simple model systems, ground-state conformational distributions obtained from molecular dynamics simulations have been used to explain CPD formation, but this method has not yet been applied to the long DNA duplexes.[239–242] Kundu *et al*. proposed that a decrease in duplex flexibility for a TT pair with flanking G-C versus A-T base pairs might prevent the substantial reorientation of the duplex required to accommodate the photodimer.[200]

In summary, the main factors affecting the efficiency of charge transfer states are either intrinsic features, such as differences in the redox potentials of neighboring bases, or environmental factors, such as conformational constraints imposed by the helical structure of DNA or electrostatic interactions with their surroundings.

3.5 Conclusion

Cyclobutane pyrimidine dimer formation in single-stranded DNA and duplex DNA is well known to depend upon the identities and locations of the flanking purine bases, with greater efficiency being detected for A than G flanking bases and lower efficiency for 5′ than 3′ G flanking bases.[110,200,201] Various mechanisms have been proposed to explain the effects of purine bases on dimerization efficiency. These include differences in conformation, electronic sensitization of dimer cleavage by G, and thymine excited state quenching by G. Trinucleotide dimerization efficiency is determined mainly by the

influence of the purine on the TT ground-state conformation rather than stacking interactions involving all three nucleotides or electronic interactions between the purine and adjacent thymine.[239] It has also been reported that in base-paired hairpins, in which TT contains flanking G versus A bases, dimerization is inhibited substantially due to a difference in ground-state conformation.[200] On the other hand, preliminary results from molecular dynamics simulations indicate a slight difference in the population of reactive TT conformations for flanking G compared to A bases.[239] Moreover, there is evidence that the efficiency of dimerization of dipyrimidines (2 + 2) in long single-stranded and duplex DNA depends more on flanking purines' identity and location than ground-state conformation as in the trinucleotide model.[110,201] Electron transfer from a nearby photoexcited base is the second proposed mechanism for photoinduced CPD splitting, which has attracted more attention than ground-state conformation.[110] A low oxidation potential electron-donor base such as guanine or 8-oxoguanine is thought to be needed for the photolyase-like repair mechanism to function.[106,110,113] In contrast with CPD formation, exciplex formation is believed to be an effective quenching mechanism.[19,115,201,202] A charge-transfer exciplex between the dipyrimidine and flanking bases is also proposed for deactivating the reactive excited state and the reduced rate of TT-CPD formation.

In this work, we have investigated the effect of neighboring purines on the formation and repair of thymine dimer in 10 different 114-base pair DNA sequences containing central thymine pairs. Our results indicate the importance of the neighboring bases up to 3 positions away from the thymine pair in controlling the levels of TT-CPD formation. Neighboring guanines suppress this level significantly, perhaps by acting as temporary electron donors to facilitate the cleavage of thymine dimer or by an effective quenching of the reactive excited state.[106,110] For instance, GGGTTGGG sequences show the lowest rate of CPD formation consistent with these plausible mechanisms. However, repairing through charge transfer or charge delocalization involving more than one adjacent base was also evident in our data. The GCGTTGCG sequence shows the highest repair rate (K_r), indicating a more efficient charge transfer. The research done by Kawai *et al*., 2013 shows that the charge transfer proceeds across GC repeat through delocalized charge over guanines instead of localized G-hoping system.[243] Other sequences such as GGGTTAGA and AAGTTGCG show a high rate of CPD repair, emphasizing the importance of the

existing difference in oxidation potential between the neighboring motif to thymine pair as a prerequisite for efficient charge transfer. This assumption was further confirmed by looking at the GGGTTGGG sequence; despite the lowest formation rate of TT-CPD, the flanking guanines sequence (GGGTTGGG) also shows the lowest repair rate and a moderate level of T<>T formation at the photostationary state. The ability of a particular motif within duplex DNA to promote repair of a neighboring T<>T may rationally be related to its oxidation potential. 5'AAG and GAA3' motifs show the lowest positive selection among 5' NNG and GNN 3' motifs in our NGS time-course experiments in direct and photosensitized irradiations. However, the AAGTTGAA sequence shows the lowest T<>T formation level at the photostationary state. The discrepancy between the NGS results and AAAGTTGAA sequence CPD quantification can be attributed to the complexity of the random DNA library and the possibility of the presence of these motifs with other low selection motifs. The T<>T level % at the photostationary state for all different sequences was governed by k_r/k_f ratio, the highest ratio the lowest T<>T formation at the photostationary state.

To conclude, our results clearly show that TT-CPD formation and photoreversal are not highly predictable based on the identity of the immediately flanking base only. The formation and repair of TT-CPD in long duplex DNA is controlled by more than one mechanism or factor, and the net result of these factors determines the final dimerization quantum yield. Consequently, steric and/or electronic effects contribute to CPD production in duplex DNA through interactions with more than one nearest-neighbor base. The nucleotide contribution from the second and third positions also plays an essential role in the formation and reversal of TT-CPD. It is safe to say that the formation and repair of TT-CPD within long double-stranded DNA cannot be attributed merely to a particular plausible mechanism. Differences in ground-state conformation, electronic sensitization of dimer by adjacent guanine or thymine excited state quenching by guanine, and delocalized or charge transfer. The mechanisms mentioned above solely or collectively can play an influential role in determining the quantum yield of thymine dimer formation.

Chapter 4 Conclusions and Future Work

4.1 Conclusions

Natural selection employs different mechanisms to ensure that genetic information remains intact when exposed to UV radiation. In detecting UV-induced DNA damage and repair, an array of methods can be employed to identify, visualize, and quantify the formation and repair of UV-induced DNA damage. Recent advances have been made by combining ultrafast spectroscopic techniques, particularly time-resolved infrared (TRIR) spectroscopy. With TRIR methods, we better understand the dynamic processes in DNA after UV photons are absorbed and are able to investigate the causes of photolesion. Thereby, this can provide us with an opportunity to examine the molecular mechanisms involved in UV-induced DNA damage formation, DNA repair pathways, mutagenesis, and carcinogenesis. Individual DNA bases are very efficient at dissipating the excitation, on a sub-picosecond time scale. In contrast, stranded DNA exhibits longer-lasting excited states. Single-stranded DNA is characterized by the process of charge transfer facilitated by DNA base stacking. Nonetheless, dsDNA presents a more complex situation, and more research is necessary to provide a thorough understanding of the nature of the long-lived excited states and their contribution to DNA integrity. Other unresolved issues include the role of flanking bases and UV-A in direct and photosensitized TT-CPD lesion formation, which we have tried to address in this work.

In the last decade, advancements in next-generation sequencing technologies have allowed many novel methods to be developed for detecting UV-induced damage and repair. In this thesis, we described a novel NGS-based approach to in vitro investigate the effect of flanking sequence on TT-CPD formation and repair by direct UVC irradiation and UVA irradiation in the presence of photosensitizer (2′-methoxyacetophenone). For the first time, we provide a comprehensive and systematic investigation of the effect of flanking bases on the actual formation and repair of TT-CPD over time.

Our results show that the T**TT**T and A**TT**T are the hottest spots for CPD formation and persistence in N**TT**N motif sequences, and G**TT**G is the coldest spot (i.e. least CPD persistence) in both the direct and photosensitized irradiations. Our results corroborate previous studies, in that we observed a somewhat greater suppressive effect from a 5'

flanking guanine on T^T-CPD formation than a 3' flanking guanine, especially if there is cytosine or adenine on the 3' flank (G**TT**C> C**TT**G and G**TT**A> A**TT**G).[13,115,141,200,201] However, when a guanine flanks the thymine pair on one flank with thymine on the other flank, in this case, guanine from the 3' side showing a slightly higher suppressing effect on T^T-CPD formation than guanine from 5' side (T**TT**G > G**TT**T). In comparison with previous studies, this finding is entirely different.[13,115,141,199–201] As a rule of thumb, the presence of guanines directly flanking the central thymine pairs substantially increases DNA's photostability and promotes the repair of neighboring thymine dimers. Meanwhile, cytosines and adenines show slower formation and repair rates than thymine and comparable fluctuation over time. The motif fluctuation adjacent to the central thymine pair was examined from both 5' and 3' sides to search for a particular preference for charge transfer or exciplex formation. The highest positive selection is evident for those motifs containing guanines directly adjacent to thymine pairs from both 5 'and 3 'sides, particularly these three motifs (GCG, GTG, GGG). Motifs containing thymine directly adjacent to the thymine pair from both 5' and 3' sides have a negative effect on the photostability of DNA and promote CPDs formation, except when it present adjacent to 5' CG, GC 3', 5' GG, GG 3' motifs. Thymine has the highest negative selection when followed by AA, TT or AT motifs (5'AAT, 5'ATT, 5'TTT) (TAA3', TTA3', TTT3'). The pattern and fluctuation of the trinucleotide motifs imply the main factors for the formation and reversal of CPDs could be the following: First, better base stacking-- guanine especially on 5' side-- suppressed CPD formation more than any other nucleotides due to better pi stacking and more favorable geometry overlap at 5' side.[115,202,221] The second factor is the free energy of stacked bases, which controls the tendency of motif fluctuation. Adenine has the highest stacking free energy in the context of a short DNA duplex [244] and shows the most negative selection in (5' NNNTT and TTNNN 3') motifs. The third factor is the likelihood of forming another CPD (presence of two adjacent pyrimidines). The fluctuation of the motif-containing two adjacent pyrimidine looks like it is mainly governed by their probability to form a second dimer. The presence of two consecutive pyrimidines can increase the likelihood of the formation of another CPD (to which our experimental assay is sensitive) rather than involving the central thymine pair. In accordance with the previous studies, the motifs containing two consecutive pyrimidines show less positive selection and tend to

form the CPDs in the following order: 5'-NCC**TT** < 5'-NCT**TT** < 5'-NTC**TT** < 5'-NTT**TT**, and **TT**CCN-3' < **TT**TCN-3' < **TT**CTN-3' < **TT**TTN-3'.

The fluctuation of the various sequences in the two irradiation experiments (278 nm and 365 nm) was almost identical. The majority of the variation in literature for UV irradiation findings arises mainly from different irradiation conditions. Therefore, this similarity was not surprising for many reasons. First, the similarity of the irradiation set up and environment, in which we used two different LEDs at two different wavelengths, but they are of comparable strengths. The source of irradiation is kept at a constant distance from each sample in the cuvette, and there is no variation in temperature by carrying all the experiments at 4 °C. The second is establishing the approximate photostationary state for each time-course experiment before choosing the time points for sequencing. Third, the formation of CPD lesions occurs within 1 ps, obviously due to the excited singlet state, as has been established unambiguously.[137,138] Recent developments in time-resolved methodologies and reliable computer computations strongly support singlet routes. The resemblance between direct and triplet-sensitized irradiation was also described in the recent Tylor publication,[140] except for Norfloxacin, which shows a different pattern due to its propensity to intercalate into DNA. Direct (singlet) mechanism involves direct light absorption followed by a picosecond concerted mechanism, while triplet sensitized mechanism involves a physical interaction with an excited triplet molecule, which results in a biradical intermediate that must then crosslinked to form the CPD.[175,176] One possibility is that both direct and sensitized CPD formation occur in the same excited state, which, if true, would be the lower energy triplet state. Some early research suggested that the triplet state is involved in CPD production after direct irradiation, but this was contradicted by some other study that indicated that the triplet state is not implicated.[132,245] If this is the case, the coordinated and stepwise methods should both be sensitive to sequence-dependent excited state localization and conformational effects on dimerization. Accordingly, the formation of CPD in dsDNA weather induced by direct or photosensitized irradiation occurs by the same mechanism. From all of these, we can argue that if the conditions of irradiation were very comparable, the outcome of direct and photosensitized irradiation would be similar.

Finally, what is the answer to the big picture question of this thesis and in earlier literature on the subject: are actual sequences, coding sequences in living organisms, optimized according to minimize the formation of damaging CPD dimers or not? The honest answer is that we are still far away from proving or disproving this postulation. For instance, there are well-documented selection biases in the codon usage tables of different microorganisms.[246–249] It is not clear, despite their relatively small effective population size, whether codon usage is also influenced by natural selection in large organisms.[246] A large body of evidence now demonstrates that the variation in synonymous codon usage is influenced by mutational bias along with both natural selection and genetic drift. However, it is still a mystery how these processes collectively control patterns of codon usage bias across entire lineages.[250] On the other hand, codon usage in mammals is mainly influenced by variation in GC content within a genome and is only weakly influenced by gene expression and tRNA content.[251–253]

4.2 Future Work

4.2.1 Extended CPD Study

The CPD lesions occur most frequently between two thymine bases. However, CPD can potentially occur between any dipyrimidine motif. Therefore, new DNA libraries containing central CT, TC, or CC instead of TT will be used to round out the whole picture. We designed the three random DNA libraries to be able to analyze them precisely in the same manner as we used to detect thymine dimer; except, in these new libraries, instead of the central thymine pair, we have introduced central CT, TC, or CC.

We used the same UV-irradiation setup for these newer experiments as mentioned in the Methods section. UV irradiation was carried out using either a monochromatic 285 nm LED or a 365 nm LED with the presence of 2′-methoxyacetophenone (2-M) as a triplet sensitizer (carried out under an anaerobic condition). Two time-dependent experiments were performed for each library by irradiating the purified dsDNA library with 365 nm light at six time points (0, 5, 15, 30, 90, and 270 min). For the 278 nm LED irradiation, seven different irradiation times, ranging from 0, 3, 10, 30, 90, 270, and 540 s, were selected. The above time points were selected following preliminary experiments on T4 PDG digestion time courses on multiple ^{32}P-radiolabelled library duplex samples irradiated

with the 278 nm LED or the 365 nm LED. The cleaved DNAs were analyzed on denaturing PAGE.

All the sequencing for all time points has already been carried out using a NextSeq, high-output flow cell that can generate up to 400 million paired-end clusters (2 × 150 bp). Sequencing was done in high quality for a total of 78-time points, and currently, we are working on Python analysis of the data.

4.2.2 Active CPD selection (CPD-seq)

Our method can be modified to work with the CPD-seq technique.[254] The principle of this method relies on creating a nick and a ligatable-OH group upstream of CPD sites using T4 PDG and APE1 enzymes. Afterward, the 5'-phosphate group is excised, and -a second adapter is ligated to DNA fragments that contain 3'-OH groups. In following steps, the ligated products are amplified by PCR and then sent for next-generation sequencing. The drawbacks of this method are background signal that can be detected in nonirradiated DNA, which is produced due to DNA breakage during extraction, purification, sonication, and inefficient adaptor ligation. However, it is not an issue with the approach and library, due to the lack of extraction, purification, and/or fragmentation steps. A single round primer extension results in the formation of libraries in the duplex form, which makes the library ready for irradiation and analysis. This combination of techniques will allow us to isolate and identify the CPD-containing sequences directly by active selection.

4.2.3 Bioinformatic study to compare our data with the known UV hot spot in human genome

Human DNA damage induced by UV light is formed and repaired based on the context of the cellular system. Several factors are involved in UV damage formation and repair, such as transcriptional factors,[191] transcription factor binding,[187,255–257] histone modification following transcription[258], nucleosome positioning[195], chromatin structure[255,258], and 3D genome structure[259]. UV susceptibility generally inversely correlates with chromatin accessibility when viewed from a 3D genome organization perspective[259]. For instance, CPDs prefer the outward-facing rotation setting in nucleosomes, whereas (6-4)PPs are commonly found near nucleosome linker regions.[195,260] Accessibility of repair machinery is crucial when repairing UV damage. It is known that

chromatin repair occurs earlier in unrepressed regions (euchromatin) than in repressed regions and that later repair sites, such as heterochromatic regions and transcription factor binding sites, have higher mutation rates.[257,258,261,262] Additionally, it has been demonstrated that repair super-hotspots were significantly enriched both in frequently interacting regions (FIREs) and in super-enhancers.[263] A study of UV-exposed hyperhotspots revealed 170-fold more CPDs than the genome average, with these sites mainly affecting melanocytes.[264] The most recurrent hyperhotspots are located within two motifs, primarily for RNA-binding proteins. The first motif is found at the UV-sensing ETS family transcription factor binding sites and at sites where mTOR/5′-terminal oligopyrimidine-trace transcription is regulated. The second motif occurs in A2–15TTCTY, which developed "dark CPDs" long after UV exposure, caused the CPDs to be repaired slowly, and accumulated CPDs prior to the experiment.

The sequencing of whole cancer genomes reveals the presence of an abundance of recurrent mutations in gene-regulatory promoter regions. It is found particularly in melanoma, where strong mutation hotspots are observed adjacent to ETS-family transcription factor (TF) binding sites. These mutation hot spots have a controversial role in the repair of DNA, but are commonly considered to locally inhibit the repair of DNA. Larsson et al., demonstrated that a lower dose of UV light induces mutations at a known ETS promoter hotspot in cultured cells even in the absence of global or transcription-coupled nucleotide excision repair (NER). A strong increase in the CPD formation was observed for the ETS-related mutation hotspots shortly after UV exposure, which was similar to the tumor mutation data at the single-base level. An increase in the somatic mutation burden in regulatory elements including ETS sites was observed by inhibiting NER. An elevated DNA damage was found at the specific genomic bases, and the prominent promoter mutation hotspots were observed as a key phenomenon in whole-genome cancer analyses.

A genome-wide analysis of DNA lesions and their repair efficiency at the single-nucleotide level is essential to better understanding UV-induced mutagenesis and carcinogenesis. After we finish analyzing the second set of data of DNA libraries with central CT, TC, or CC pair, we will have considerable information to construct our

bioinformatics study. These data will enable us to perform the first genome-wide analysis of UV hot spots based solely on nucleotide-mediated photorepair.

References

1. Gilbert, W. Origin of life: The RNA world. *Nature* **319**, 618 (1986).
2. Cech, T. R. The RNA worlds in context. *Cold Spring Harb. Perspect. Biol.* **4**, a006742 (2012).
3. Bartel, D. P. & Unrau, P. J. Constructing an RNA world. *Trends Cell Biol.* **9**, M9–M13 (1999).
4. Carter, C. W. What RNA world? Why a peptide/RNA partnership merits renewed experimental attention. *Life* **5**, 294–320 (2015).
5. Schreier, W. J., Gilch, P. & Zinth, W. Early events of DNA photodamage. *Annu. Rev. Phys. Chem.* **66**, 497–519 (2015).
6. Crespo-Hernández, C. E., Cohen, B., Hare, P. M. & Kohler, B. Ultrafast excited-state dynamics in nucleic acids. *Chem. Rev.* **104**, 1977–2020 (2004).
7. Doorley, G. W. *et al.* Tracking DNA excited states by picosecond-time-resolved infrared spectroscopy: signature band for a charge-transfer excited state in stacked adenine–thymine systems. *J. Phys. Chem. Lett.* **4**, 2739–2744 (2013).
8. Bucher, D. B., Pilles, B. M., Carell, T. & Zinth, W. Charge separation and charge delocalization identified in long-living states of photoexcited DNA. *Proc. Natl. Acad. Sci.* **111**, 4369–4374 (2014).
9. Zhang, Y. *et al.* Efficient UV-induced charge separation and recombination in an 8-oxoguanine-containing dinucleotide. *Proc. Natl. Acad. Sci.* **111**, 11612–11617 (2014).
10. Bucher, D. B., Kufner, C. L., Schlueter, A., Carell, T. & Zinth, W. UV-induced charge transfer states in DNA promote sequence selective self-repair. *J. Am. Chem. Soc.* **138**, 186–190 (2016).
11. Genereux, J. C. & Barton, J. K. Mechanisms for DNA charge transport. *Chem. Rev.* **110**, 1642–1662 (2010).
12. Yoon, J.-H., Lee, C.-S., O'Connor, T. R., Yasui, A. & Pfeifer, G. P. The DNA damage spectrum produced by simulated sunlight. *J. Mol. Biol.* **299**, 681–693 (2000).
13. Mitchell, D. L., Jen, J. & Cleaver, J. E. Sequence specificity of cyclobutane pyrimidine dimers in DNA treated with solar (ultraviolet B) radiation. *Nucleic Acids Res.* **20**, 225–229 (1992).
14. Lyamichev, V. I., Frank-Kamenetskii, M. D. & Soyfer, V. N. Protection against UV-induced pyrimidine dimerization in DNA by triplex formation. *Nature* **344**, 568–570 (1990).
15. Schieferstein, U. & Thoma, F. Modulation of Cyclobutane Pyrimidine Dimer Formation in a Positioned Nucleosome Containing Poly (dA⊙ dT) Tracts. *Biochemistry* **35**, 7705–7714 (1996).
16. Pehrson, J. R. & Cohen, L. H. Effects of DNA looping on pyrimidine dimer formation. *Nucleic Acids Res.* **20**, 1321–1324 (1992).
17. Gale, J. M., Nissen, K. A. & Smerdon, M. J. UV-induced formation of pyrimidine dimers in nucleosome core DNA is strongly modulated with a period of 10.3 bases. *Proc. Natl. Acad. Sci.* **84**, 6644–6648 (1987).
18. Markovitsi, D., Onidas, D., Gustavsson, T., Talbot, F. & Lazzarotto, E. Collective behavior of Franck− Condon excited states and energy transfer in DNA double helices. *J. Am. Chem. Soc.* **127**, 17130–17131 (2005).

19. Crespo-Hernández, C. E., Cohen, B. & Kohler, B. Base stacking controls excited-state dynamics in A· T DNA. *Nature* **436**, 1141–1144 (2005).
20. Wellinger, R.-E. & Thoma, F. Taq DNA polymerase blockage at pyrimidine dimers. *Nucleic Acids Res.* **24**, 1578–1579 (1996).
21. WATSON, J. D. & CRICK, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **171**, 737–738 (1953).
22. Yakovchuk, P., Protozanova, E. & Frank-Kamenetskii, M. D. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Res.* **34**, 564–574 (2006).
23. Berk, A., Zipursky, S. L. & Lodish, H. *Molecular cell biology*. (WH Freeman, 2008).
24. Pfeifer, G. P. & Besaratinia, A. UV wavelength-dependent DNA damage and human non-melanoma and melanoma skin cancer. *Photochem. Photobiol. Sci.* **11**, 90–97 (2012).
25. Coelho, M. M. V., Matos, T. R. & Apetato, M. The dark side of the light: mechanisms of photocarcinogenesis. *Clin. Dermatol.* **34**, 563–570 (2016).
26. Cadet, J., Mouret, S., Ravanat, J. & Douki, T. Photoinduced damage to cellular DNA: direct and photosensitized reactions. *Photochem. Photobiol.* **88**, 1048–1065 (2012).
27. Essen, L. O. & Klar, T. Light-driven DNA repair by photolyases. *Cell. Mol. Life Sci. C.* **63**, 1266–1277 (2006).
28. Sancar, A. Structure and function of DNA photolyase and cryptochrome blue-light photoreceptors. *Chem. Rev.* **103**, 2203–2238 (2003).
29. Pfeifer, G. P. Formation and processing of UV photoproducts: effects of DNA sequence and chromatin environment. *Photochem. Photobiol.* **65**, 270–283 (1997).
30. Li, J., Uchida, T., Todo, T. & Kitagawa, T. Similarities and differences between cyclobutane pyrimidine dimer photolyase and (6-4) photolyase as revealed by resonance Raman spectroscopy: Electron transfer from the FAD cofactor to ultraviolet-damaged DNA. *J. Biol. Chem.* **281**, 25551–25559 (2006).
31. Patrick, M. H. & Rahn, R. O. Photochemistry of DNA and Polynucleotides. *Photochem. Photobiol. Nucleic Acids* **2**, 35–91 (1976).
32. Fisher, G. J. & Johns, H. E. Pyrimidine photohydrates. in *Photochemistry and photobiology of nucleic acids* 169–224 (Elsevier, 1976).
33. Boorstein, R. J., Hilbert, T. P., Cunningham, R. P. & Teebor, G. W. Formation and stability of repairable pyrimidine photohydrates in DNA. *Biochemistry* **29**, 10455–10460 (1990).
34. Stief, L. J. & DeCarlo, V. J. Vacuum-Ultraviolet Photochemistry. III. Formation of Carbon Atoms in the Photolysis of Carbon Suboxide at 1470 Å. *J. Chem. Phys.* **43**, 2552–2553 (1965).
35. Varghese, A. J. Photochemistry of thymidine in ice. *Biochemistry* **9**, 4781–4787 (1970).
36. Douki, T. & Cadet, J. Far-UV photochemistry and photosensitization of 2′-deoxycytidylyl-(3′-5′)-thymidine: isolation and characterization of the main photoproducts. *J. Photochem. Photobiol. B Biol.* **15**, 199–213 (1992).
37. Douki, T. & Cadet, J. Formation of cyclobutane dimers and (6-4) photoproducts upon far-UV photolysis of 5-methylcytosine-containing dinucleoside

monophosphates. *Biochemistry* **33**, 11942–11950 (1994).
38. Johns, H. E., Pearson, M. L., LeBlanc, J. C. & Helleiner, C. W. The ultraviolet photochemistry of thymidylyl-(3′→ 5′)-thymidine. *J. Mol. Biol.* **9**, 503-IN1 (1964).
39. Lemaire, D. G. E. & Ruzsicska, B. P. Quantum yields and secondary photoreactions of the photoproducts of dTpdT, dTpdC and dTpdU. *Photochem. Photobiol.* **57**, 757–769 (1993).
40. Liu, F.-T. & Yang, N. C. Photochemistry of cytosine derivatives. 2. Photohydration of cytosine derivatives. Proton magnetic resonance study on the chemical structure and property of photohydrates. *Biochemistry* **17**, 4877–4885 (1978).
41. Cadet, J. *et al.* Characterization of thymidine ultraviolet photoproducts. Cyclobutane dimers and 5, 6-dihydrothymidines. *Can. J. Chem.* **63**, 2861–2868 (1985).
42. Rastogi, R. P., Kumar, A., Tyagi, M. B. & Sinha, R. P. Molecular mechanisms of ultraviolet radiation-induced DNA damage and repair. *J. Nucleic Acids* **2010**, (2010).
43. Ravanat, J.-L., Douki, T. & Cadet, J. Direct and indirect effects of UV radiation on DNA and its components. *J. Photochem. Photobiol. B Biol.* **63**, 88–102 (2001).
44. Douki, T. The variety of UV-induced pyrimidine dimeric photoproducts in DNA as shown by chromatographic quantification methods. *Photochem. Photobiol. Sci.* **12**, 1286–1302 (2013).
45. Cadet, J. Photochemistry and nucleic acids. *Bioorganic Photochem.* **1**, 1–272 (1990).
46. Garces, F. & Davila, C. A. Alterations in DNA irradiated with ultraviolet radiation—I. The formation process of cyclobutylpyrimidine dimers: cross sections, action spectra and quantum yields. *Photochem. Photobiol.* **35**, 9–16 (1982).
47. Friedel, M. G., Cichon, M. K. & Carell, T. DNA damage and repair: photochemistry. *CRC Handb. Org. Photochem. Photobiol.* **2**, (2004).
48. Douki, T. & Cadet, J. Individual determination of the yield of the main UV-induced dimeric pyrimidine photoproducts in DNA suggests a high mutagenicity of CC photolesions. *Biochemistry* **40**, 2495–2501 (2001).
49. Johns, H. E., Rapaport, S. A. & Delbrück, M. Photochemistry of thymine dimers. *J. Mol. Biol.* **4**, 104–114 (1962).
50. Thomas, D. C. & Kunkel, T. A. Replication of UV-irradiated DNA in human cell extracts: evidence for mutagenic bypass of pyrimidine dimers. *Proc. Natl. Acad. Sci.* **90**, 7744–7748 (1993).
51. Jiang, N. & Taylor, J. S. In vivo evidence that UV-induced C. fwdarw. T mutations at dipyrimidine sites could result from the replicative bypass of cis-syn cyclobutane dimers or their deamination products. *Biochemistry* **32**, 472–481 (1993).
52. Barak, Y., Cohen-Fix, O. & Livneh, Z. Deamination of Cytosine-containing Pyrimidine Photodimers in UV-irradiated DNA: SIGNIFICANCE FOR UV LIGHT MUTAGENESIS (∗). *J. Biol. Chem.* **270**, 24174–24179 (1995).
53. Peng, W. & Shaw, B. R. Accelerated deamination of cytosine residues in UV-induced cyclobutane pyrimidine dimers leads to CC→ TT transitions.

Biochemistry **35**, 10172–10181 (1996).
54. Lemaire, D. G. E. & Ruzsicska, B. P. Kinetic analysis of the deamination reactions of cyclobutane dimers of thymidylyl-3', 5'-2'-deoxycytidine and 2'-deoxycytidylyl-3', 5'-thymidine. *Biochemistry* **32**, 2525–2533 (1993).
55. Frederico, L. A., Kunkel, T. A. & Shaw, B. R. A sensitive genetic assay for the detection of cytosine deamination: determination of rate constants and the activation energy. *Biochemistry* **29**, 2532–2537 (1990).
56. Marks, R. Epidemiology of melanoma: Clinical dermatology• Review article. *Clin. Exp. Dermatology Clin. dermatology* **25**, 459–463 (2000).
57. Drouin, R. & Therrien, J. UVB-induced cyclobutane pyrimidine dimer frequency correlates with skin cancer mutational hotspots in p53. *Photochem. Photobiol.* **66**, 719–726 (1997).
58. Celewicz, L., Mayer, M. & Shetlar, M. D. The Photochemistry of Thymidylyl-(3′-5′)-5-methyl-2′-deoxycytidine in Aqueous Solution¶. *Photochem. Photobiol.* **81**, 404–418 (2005).
59. Pfeifer, G. P., You, Y.-H. & Besaratinia, A. Mutations induced by ultraviolet light. *Mutat. Res. Mol. Mech. Mutagen.* **571**, 19–31 (2005).
60. Kim, J., Patel, D. & Choi, B. Contrasting structural impacts induced by cis-syn cyclobutane dimer and (6–4) adduct in DNA duplex decamers: implication in mutagenesis and repair activity. *Photochem. Photobiol.* **62**, 44–50 (1995).
61. You, Y.-H. *et al.* Cyclobutane pyrimidine dimers are responsible for the vast majority of mutations induced by UVB irradiation in mammalian cells. *J. Biol. Chem.* **276**, 44688–44694 (2001).
62. Sancar, A., Lindsey-Boltz, L. A., Ünsal-Kaçmaz, K. & Linn, S. Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annu. Rev. Biochem.* **73**, 39–85 (2004).
63. Makridakis, N. M. & Reichardt, J. K. V. Translesion DNA polymerases and cancer. *Front. Genet.* **3**, 174 (2012).
64. Prakash, S., Johnson, R. E. & Prakash, L. Eukaryotic translesion synthesis DNA polymerases: specificity of structure and function. *Annu. Rev. Biochem.* **74**, 317–353 (2005).
65. Ohmori, H. *et al.* The Y-family of DNA polymerases. *Mol. Cell* **8**, 7–8 (2001).
66. Büchi, G., Inman, C. G. & Lipinsky, E. S. Light-catalyzed organic reactions. I. The reaction of carbonyl compounds with 2-methyl-2-butene in the presence of ultraviolet light. *J. Am. Chem. Soc.* **76**, 4327–4331 (1954).
67. Varghese, A. J. & Wang, S. Y. Thymine-thymine adduct as a photoproduct of thymine. *Science (80-.).* **160**, 186–187 (1968).
68. Wang, S. Y. & Varghese, A. J. Cytosine-thymine addition product from DNA irradiated with ultraviolet light. *Biochem. Biophys. Res. Commun.* **29**, 543–549 (1967).
69. Taylor, J. S. Unraveling the molecular pathway from sunlight to skin cancer. *Acc. Chem. Res.* **27**, 76–82 (1994).
70. Desnous, C., Guillaume, D. & Clivio, P. Spore photoproduct: A key to bacterial eternal life. *Chem. Rev.* **110**, 1213–1232 (2010).
71. Lippke, J. A., Gordon, L. K., Brash, D. E. & Haseltine, W. A. Distribution of UV light-induced damage in a defined sequence of human DNA: detection of alkaline-

sensitive lesions at pyrimidine nucleoside-cytidine sequences. *Proc. Natl. Acad. Sci.* **78**, 3388–3392 (1981).
72. Pfeifer, G. P., Drouin, R., Riggs, A. D. & Holmquist, G. P. In vivo mapping of a DNA adduct at nucleotide resolution: detection of pyrimidine (6-4) pyrimidone photoproducts by ligation-mediated polymerase chain reaction. *Proc. Natl. Acad. Sci.* **88**, 1374–1378 (1991).
73. LeClerc, J. E., Borden, A. & Lawrence, C. W. The thymine-thymine pyrimidine-pyrimidone (6-4) ultraviolet light photoproduct is highly mutagenic and specifically induces 3'thymine-to-cytosine transitions in Escherichia coli. *Proc. Natl. Acad. Sci.* **88**, 9685–9689 (1991).
74. Taylor, J. S. & Cohrs, M. P. DNA, light, and Dewar pyrimidinones: the structure and biological significance to TpT3. *J. Am. Chem. Soc.* **109**, 2834–2835 (1987).
75. Taylor, J., Lu, H. & Kotyk, J. J. Quantitative conversion of the (6–4) photoproduct of TpdC to its Dewar valence isomer upon exposure to simulated sunlight. *Photochem. Photobiol.* **51**, 161–167 (1990).
76. Douki, T. & Sage, E. Dewar valence isomers, the third type of environmentally relevant DNA photoproducts induced by solar radiation. *Photochem. Photobiol. Sci.* **15**, 24–30 (2016).
77. Haiser, K. *et al.* Mechanism of UV-induced formation of Dewar lesions in DNA. *Angew. Chemie Int. Ed.* **51**, 408–411 (2012).
78. Dewar, J. 5. On the Oxidation of Phenyl Alcohol, and a Mechanical Arrangement adapted to illustrate Structure in the Nonsaturated Hydrocarbons. *Proc. R. Soc. Edinburgh* **6**, 82–86 (1869).
79. Lee, J.-H., Bae, S.-H. & Choi, B.-S. The Dewar photoproduct of thymidylyl (3′→ 5′)-thymidine (Dewar product) exhibits mutagenic behavior in accordance with the "A rule". *Proc. Natl. Acad. Sci.* **97**, 4591–4596 (2000).
80. Taylor, J.-S. New structural and mechanistic insight into the A-rule and the instructional and non-instructional behavior of DNA photoproducts and other lesions. *Mutat. Res. Mol. Mech. Mutagen.* **510**, 55–70 (2002).
81. Donnellan, J. E. & Setlow, R. B. Thymine photoproducts but not thymine dimers found in ultraviolet-irradiated bacterial spores. *Science (80-.).* **149**, 308–310 (1965).
82. Varghese, A. J. 5-Thyminyl-5, 6-dihydrothymine from DNA irradiated with ultraviolet light. *Biochem. Biophys. Res. Commun.* **38**, 484–490 (1970).
83. Setlow, P. I will survive: DNA protection in bacterial spores. *Trends Microbiol.* **15**, 172–180 (2007).
84. Slieman, T. A., Rebeil, R. & Nicholson, W. L. Spore photoproduct (SP) lyase from Bacillus subtilis specifically binds to and cleaves SP (5-thyminyl-5, 6-dihydrothymine) but not cyclobutane pyrimidine dimers in UV-irradiated DNA. *J. Bacteriol.* **182**, 6412–6417 (2000).
85. Cadet, J., Anselmino, C., Douki, T. & Voituriez, L. New trends in photobiology: Photochemistry of nucleic acids in cells. *J. Photochem. Photobiol. B Biol.* **15**, 277–298 (1992).
86. Nicholson, W. L., Munakata, N., Horneck, G., Melosh, H. J. & Setlow, P. Resistance of Bacillus endospores to extreme terrestrial and extraterrestrial environments. *Microbiol. Mol. Biol. Rev.* **64**, 548–572 (2000).

87. Cooke, M. S., Evans, M. D., Dizdaroglu, M. & Lunec, J. Oxidative DNA damage: mechanisms, mutation, and disease. *FASEB J.* **17**, 1195–1214 (2003).
88. Yasui, M. *et al.* Tracing the fates of site-specifically introduced DNA adducts in the human genome. *DNA Repair (Amst).* **15**, 11–20 (2014).
89. Yoshida, R., Ogawa, Y. & Kasai, H. Urinary 8-oxo-7, 8-dihydro-2′-deoxyguanosine values measured by an ELISA correlated well with measurements by high-performance liquid chromatography with electrochemical detection. *Cancer Epidemiol. Prev. Biomarkers* **11**, 1076–1081 (2002).
90. Haghdoost, S., Czene, S., Näslund, I., Skog, S. & Harms-Ringdahl, M. Extracellular 8-oxo-dG as a sensitive parameter for oxidative stress in vivo and in vitro. *Free Radic. Res.* **39**, 153–162 (2005).
91. Middleton, C. T. K. d. L. Harpe, C. Su, YK Law, CE Crespo-Hernández and B. Kohler. *Annu. Rev. Phys. Chem* **60**, 217–239 (2009).
92. Bucher, D. B., Pilles, B. M., Carell, T. & Zinth, W. Dewar lesion formation in single-and double-stranded DNA is quenched by neighboring bases. *J. Phys. Chem. B* **119**, 8685–8692 (2015).
93. Marguet, S. & Markovitsi, D. Time-resolved study of thymine dimer formation. *J. Am. Chem. Soc.* **127**, 5780–5781 (2005).
94. Barbatti, M., Borin, A. C. & Ullrich, S. *Photoinduced phenomena in nucleic acids II: DNA fragments and phenomenological aspects*. vol. 356 (Springer, 2014).
95. Kufner, C. L., Zinth, W. & Bucher, D. B. UV-Induced Charge-Transfer States in Short Guanosine-Containing DNA Oligonucleotides. *ChemBioChem* **21**, 2306 (2020).
96. Kumar, S. *et al.* Adenine photodimerization in deoxyadenylate sequences: elucidation of the mechanism through structural studies of a major d (ApA) photoproduct. *Nucleic Acids Res.* **19**, 2841–2847 (1991).
97. Görner, H. Chromophore loss of uracil derivatives and polyuridylic acid in aqueous solution caused by 248 nm laser pulses and continuous UV irradiation: mechanism of the photohydration of pyrimidines. *J. Photochem. Photobiol. B Biol.* **10**, 91–110 (1991).
98. Gomez-Mendoza, M., Banyasz, A., Douki, T., Markovitsi, D. & Ravanat, J.-L. Direct oxidative damage of naked DNA generated upon absorption of UV radiation by nucleobases. *J. Phys. Chem. Lett.* **7**, 3945–3948 (2016).
99. Madugundu, G. S. *et al.* Generation of Guanine–Thymine Cross-Links in Human Cells by One-Electron Oxidation Mechanisms. *Chem. Res. Toxicol.* **26**, 1031–1033 (2013).
100. Bergeron, F., Auvré, F., Radicella, J. P. & Ravanat, J.-L. HO• radicals induce an unexpected high proportion of tandem base lesions refractory to repair by DNA glycosylases. *Proc. Natl. Acad. Sci.* **107**, 5528–5533 (2010).
101. M nzel, M., Szeibert, C., Glas, A. F., Globisch, D. & Carell, T. Discovery and synthesis of new UV-induced intrastrand c (4− 8) g and g (8− 4) c photolesions. *J. Am. Chem. Soc.* **133**, 5186–5189 (2011).
102. Szabla, R., Kruse, H., Stadlbauer, P., Šponer, J. & Sobolewski, A. L. Sequential electron transfer governs the UV-induced self-repair of DNA photolesions. *Chem. Sci.* **9**, 3131–3140 (2018).
103. Giese, B. *et al.* Excess Electron Transport Through DNA: A Single Electron

Repairs More than One UV-Induced Lesion. *Angew. Chemie Int. Ed.* **43**, 1848–1851 (2004).
104. Breeger, S., Hennecke, U. & Carell, T. Excess electron-transfer-based repair of a cis-syn thymine dimer in DNA is not sequence dependent. *J. Am. Chem. Soc.* **126**, 1302–1303 (2004).
105. Behrens, C. & Carell, T. Excess electron transfer in flavin-capped, thymine dimer-containing DNA hairpins. *Chem. Commun.* 1632–1633 (2003).
106. Chinnapen, D. J.-F. & Sen, D. A deoxyribozyme that harnesses light to repair thymine dimers in DNA. *Proc. Natl. Acad. Sci.* **101**, 65–69 (2004).
107. Barlev, A. & Sen, D. Catalytic DNAs that harness violet light to repair thymine dimers in a DNA substrate. *J. Am. Chem. Soc.* **135**, 2596–2603 (2013).
108. Kneuttinger, A. C. *et al.* Formation and direct repair of UV-induced dimeric DNA pyrimidine lesions. *Photochem. Photobiol.* **90**, 1–14 (2014).
109. Liu, Z. *et al.* Dynamics and mechanism of cyclobutane pyrimidine dimer repair by DNA photolyase. *Proc. Natl. Acad. Sci.* **108**, 14831–14836 (2011).
110. Holman, M. R., Ito, T. & Rokita, S. E. Self-repair of thymine dimer in duplex DNA. *J. Am. Chem. Soc.* **129**, 6–7 (2007).
111. Nguyen, K. Van & Burrows, C. J. Whence flavins? Redox-active ribonucleotides link metabolism and genome repair to the RNA world. *Acc. Chem. Res.* **45**, 2151–2159 (2012).
112. Anusiewicz, I., wierszcz, I., Skurski, P. & Simons, J. Mechanism for repair of thymine dimers by photoexcitation of proximal 8-oxo-7, 8-dihydroguanine. *J. Phys. Chem. A* **117**, 1240–1253 (2013).
113. Nguyen, K. Van & Burrows, C. J. A prebiotic role for 8-oxoguanosine as a flavin mimic in pyrimidine dimer photorepair. *J. Am. Chem. Soc.* **133**, 14586–14589 (2011).
114. Law, Y. K., Forties, R. A., Liu, X., Poirier, M. G. & Kohler, B. Sequence-dependent thymine dimer formation and photoreversal rates in double-stranded DNA. *Photochem. Photobiol. Sci.* **12**, 1431–1439 (2013).
115. Pan, Z. *et al.* Electron donor–acceptor interactions with flanking purines influence the efficiency of thymine photodimerization. *J. Am. Chem. Soc.* **133**, 20793–20798 (2011).
116. Kasha, M. Characterization of electronic transitions in complex molecules. *Discuss. Faraday Soc.* **9**, 14–19 (1950).
117. Bucher, D. B., Schlueter, A., Carell, T. & Zinth, W. Watson–Crick Base Pairing Controls Excited-State Decay in Natural DNA. *Angew. Chemie Int. Ed.* **53**, 11366–11369 (2014).
118. Kawai, K. & Majima, T. Photoinduced charge-separation in DNA. *Photoinduced Phenom. Nucleic Acids II* 165–182 (2014).
119. Park, M. J., Fujitsuka, M., Kawai, K. & Majima, T. Excess-electron injection and transfer in terthiophene-modified DNA: terthiophene as a photosensitizing electron donor for thymine, cytosine, and adenine. *Chem. Eur. J.* **18**, 2056–2062 (2012).
120. Schuster, G. B. *Long-range charge transfer in DNA I*. vol. 236 (Springer Science & Business Media, 2004).
121. Tainaka, K., Fujitsuka, M., Takada, T., Kawai, K. & Majima, T. Sequence dependence of excess electron transfer in DNA. *J. Phys. Chem. B* **114**, 14657–

14663 (2010).
122. Bird, A. The essentials of DNA methylation. *Cell* **70**, 5–8 (1992).
123. Kautiainen, T. L. & Jones, P. A. DNA methyltransferase levels in tumorigenic and nontumorigenic cells in culture. *J. Biol. Chem.* **261**, 1594–1598 (1986).
124. Rideout III, W. M., Coetzee, G. A., Olumi, A. F. & Jones, P. A. 5-Methylcytosine as an endogenous mutagen in the human LDL receptor and p53 genes. *Science (80-.).* **249**, 1288–1290 (1990).
125. Sved, J. & Bird, A. The expected equilibrium of the CpG dinucleotide in vertebrate genomes under a mutation model. *Proc. Natl. Acad. Sci.* **87**, 4692–4696 (1990).
126. Tommasi, S., Denissenko, M. F. & Pfeifer, G. P. Sunlight induces pyrimidine dimers preferentially at 5-methylcytosine bases. *Cancer Res.* **57**, 4727–4730 (1997).
127. Mitchell, D. L. Effects of cytosine methylation on pyrimidine dimer formation in DNA. *Photochem. Photobiol.* **71**, 162–165 (2000).
128. Ikehata, H. & Ono, T. Significance of CpG methylation for solar UV-induced mutagenesis and carcinogenesis in skin. *Photochem. Photobiol.* **83**, 196–204 (2007).
129. Rochette, P. J. *et al.* Influence of cytosine methylation on ultraviolet-induced cyclobutane pyrimidine dimer formation in genomic DNA. *Mutat. Res. Mol. Mech. Mutagen.* **665**, 7–13 (2009).
130. Esposito, L. *et al.* Effect of C5-methylation of cytosine on the photoreactivity of DNA: a joint experimental and computational study of TCG trinucleotides. *J. Am. Chem. Soc.* **136**, 10838–10841 (2014).
131. Banyasz, A. *et al.* Effect of C5-methylation of cytosine on the UV-induced reactivity of duplex DNA: conformational and electronic factors. *J. Phys. Chem. B* **120**, 4232–4242 (2016).
132. Kwok, W.-M., Ma, C. & Phillips, D. L. A doorway state leads to photostability or triplet photodamage in thymine DNA. *J. Am. Chem. Soc.* **130**, 5131–5139 (2008).
133. Gut, I. G., Wood, P. D. & Redmond, R. W. Interaction of triplet photosensitizers with nucleotides and DNA in aqueous solution at room temperature. *J. Am. Chem. Soc.* **118**, 2366–2373 (1996).
134. Bosca, F., Lhiaubet-Vallet, V., Cuquerella, M. C., Castell, J. V & Miranda, M. A. The triplet energy of thymine in DNA. *J. Am. Chem. Soc.* **128**, 6318–6319 (2006).
135. Lamola, A. A. & Yamane, T. Sensitized photodimerization of thymine in DNA. *Proc. Natl. Acad. Sci. U. S. A.* **58**, 443 (1967).
136. Greenstock, C. L. & Johns, H. E. Photosensitized dimerization of pyrimidines. *Biochem. Biophys. Res. Commun.* **30**, 21–27 (1968).
137. Schreier, W. J. *et al.* Thymine dimerization in DNA is an ultrafast photoreaction. *Science (80-.).* **315**, 625–629 (2007).
138. Schreier, W. J. *et al.* Thymine dimerization in DNA model systems: cyclobutane photolesion is predominantly formed via the singlet channel. *J. Am. Chem. Soc.* **131**, 5038–5039 (2009).
139. Banyasz, A. *et al.* Electronic excited states responsible for dimer formation upon UV absorption directly by thymine strands: joint experimental and theoretical study. *J. Am. Chem. Soc.* **134**, 14834–14845 (2012).

140. Lu, C., Gutierrez-Bayona, N. E. & Taylor, J.-S. The effect of flanking bases on direct and triplet sensitized cyclobutane pyrimidine dimer formation in DNA depends on the dipyrimidine, wavelength and the photosensitizer. *Nucleic Acids Res.* **49**, 4266–4280 (2021).
141. Gordon, L. K. & Haseltine, W. A. Quantitation of Cyclobutane Pyrimidine Dirner Formation in Double-and Single-Stranded DNA Fragments of Defined Sequence. *Radiat. Res.* **89**, 99–112 (1982).
142. Barlev, A., Sekhon, G. S., Bennet, A. J. & Sen, D. DNA repair by DNA: the UV1C DNAzyme catalyzes photoreactivation of cyclobutane thymine dimers in DNA more effectively than their de novo formation. *Biochemistry* **55**, 6010–6018 (2016).
143. Kuluncsics, Z., Perdiz, D., Brulay, E., Muel, B. & Sage, E. Wavelength dependence of ultraviolet-induced DNA damage distribution: involvement of direct or indirect mechanisms and possible artefacts. *J. Photochem. Photobiol. B Biol.* **49**, 71–80 (1999).
144. Douki, T., Reynaud-Angelin, A., Cadet, J. & Sage, E. Bipyrimidine photoproducts rather than oxidative lesions are the main type of DNA damage involved in the genotoxic effect of solar UVA radiation. *Biochemistry* **42**, 9221–9226 (2003).
145. Rochette, P. J. *et al.* UVA-induced cyclobutane pyrimidine dimers form predominantly at thymine–thymine dipyrimidines and correlate with the mutation spectrum in rodent cells. *Nucleic Acids Res.* **31**, 2786–2794 (2003).
146. Mouret, S. *et al.* Cyclobutane pyrimidine dimers are predominant DNA lesions in whole human skin exposed to UVA radiation. *Proc. Natl. Acad. Sci.* **103**, 13765–13770 (2006).
147. Jiang, Y. *et al.* UVA generates pyrimidine dimers in DNA directly. *Biophys. J.* **96**, 1151–1158 (2009).
148. Improta, R. & Barone, V. Interplay between "Neutral" and "Charge-Transfer" Excimers Rules the Excited State Decay in Adenine-Rich Polynucleotides. *Angew. Chemie Int. Ed.* **50**, 12016–12019 (2011).
149. Improta, R. & Barone, V. Excited states behavior of nucleobases in solution: Insights from computational studies. *Photoinduced Phenom. Nucleic Acids I* 329–357 (2014).
150. Plasser, F., Aquino, A. J. A., Lischka, H. & Nachtigallová, D. Electronic excitation processes in single-strand and double-strand DNA: a computational approach. *Photoinduced Phenom. Nucleic Acids II* 1–37 (2014).
151. Lu, Y., Lan, Z. & Thiel, W. Computational modeling of photoexcitation in DNA single and double strands. *Photoinduced Phenom. Nucleic Acids II* 89–122 (2014).
152. Markovitsi, D. UV-induced DNA damage: the role of electronic excited states. *Photochem. Photobiol.* **92**, 45–51 (2016).
153. Premi, S. *et al.* Chemiexcitation of melanin derivatives induces DNA photoproducts long after UV exposure. *Science (80-.).* **347**, 842–847 (2015).
154. Taylor, J.-S. The dark side of sunlight and melanoma. *Science (80-.).* **347**, 824 (2015).
155. Premi, S. & Brash, D. E. Unanticipated role of melanin in causing carcinogenic cyclobutane pyrimidine dimers. *Mol. Cell. Oncol.* **3**, e1033588 (2016).
156. Setlow, R. & Carrier, W. L. Pyrimidine dimers in ultraviolet-irradiated DNA's. *J.*

Mol. Biol. **17**, 237–254 (1966).
157. Matallana-Surget, S., Meador, J. A., Joux, F. & Douki, T. Effect of the GC content of DNA on the distribution of UVB-induced bipyrimidine photoproducts. *Photochem. Photobiol. Sci.* **7**, 794–801 (2008).
158. Haseltine, W. A. *et al.* Cleavage of pyrimidine dimers in specific DNA sequences by a pyrimidine dimer DNA-glycosylase of M. luteus. *Nature* **285**, 634–641 (1980).
159. Lloyd, R. S. Investigations of pyrimidine dimer glycosylases—a paradigm for DNA base excision repair enzymology. *Mutat. Res. Mol. Mech. Mutagen.* **577**, 77–91 (2005).
160. Pan, Z., Chen, J., Schreier, W. J., Kohler, B. & Lewis, F. D. Thymine dimer photoreversal in purine-containing trinucleotides. *J. Phys. Chem. B* **116**, 698–704 (2012).
161. Cadet, J., Grand, A. & Douki, T. Solar UV radiation-induced DNA bipyrimidine photoproducts: formation and mechanistic insights. *Photoinduced Phenom. nucleic acids II* 249–275 (2014).
162. Lhiaubet-Vallet, V., Bosca, F. & Miranda, M. A. Photosensitized DNA damage: the case of fluoroquinolones. *Photochem. Photobiol.* **85**, 861–868 (2009).
163. Cuquerella, M. C., Lhiaubet-Vallet, V., Bosca, F. & Miranda, M. A. Photosensitised pyrimidine dimerisation in DNA. *Chem. Sci.* **2**, 1219–1232 (2011).
164. Epe, B. DNA damage spectra induced by photosensitization. *Photochem. Photobiol. Sci.* **1**, 98–106 (2002).
165. Douki, T., Bérard, I., Wack, A. & Andrä, S. Contribution of Cytosine-Containing Cyclobutane Dimers to DNA Damage Produced by Photosensitized Triplet–Triplet Energy Transfer. *Chem. Eur. J.* **20**, 5787–5794 (2014).
166. Antusch, L., Gaß, N. & Wagenknecht, H. Elucidation of the Dexter-Type Energy Transfer in DNA by Thymine–Thymine Dimer Formation Using Photosensitizers as Artificial Nucleosides. *Angew. Chemie Int. Ed.* **56**, 1385–1389 (2017).
167. Strieth-Kalthoff, F., James, M. J., Teders, M., Pitzer, L. & Glorius, F. Energy transfer catalysis mediated by visible light: principles, applications, directions. *Chem. Soc. Rev.* **47**, 7190–7202 (2018).
168. Dumont, E. *et al.* Resolving the benzophenone DNA-photosensitization mechanism at QM/MM level. *J. Phys. Chem. Lett.* **6**, 576–580 (2015).
169. Nogueira, J. J., Oppel, M. & González, L. Enhancing intersystem crossing in phenotiazinium dyes by intercalation into DNA. *Angew. Chemie Int. Ed.* **54**, 4375–4378 (2015).
170. Cuquerella, M. C., Lhiaubet-Vallet, V., Miranda, M. A. & Bosca, F. Drug–DNA complexation as the key factor in photosensitized thymine dimerization. *Phys. Chem. Chem. Phys.* **19**, 4951–4955 (2017).
171. Jian, Y., Maximowitsch, E., Adhikari, S., Li, L. & Domratcheva, T. Indications of 5′ to 3′ interbase electron transfer as the first step of pyrimidine dimer formation probed by a dinucleotide analog. (2017).
172. Mu, W., Han, Q., Luo, Z. & Wang, Y. Production of cis–syn thymine–thymine cyclobutane dimer oligonucleotide in the presence of acetone photosensitizer. *Anal. Biochem.* **353**, 117–123 (2006).
173. Lhiaubet-Vallet, V., Cuquerella, M. C., Castell, J. V, Bosca, F. & Miranda, M. A.

Triplet excited fluoroquinolones as mediators for thymine cyclobutane dimer formation in DNA. *J. Phys. Chem. B* **111**, 7409–7414 (2007).
174. Rahn, R. O., Landry, L. C. & Carrier, W. L. Formation of chain breaks and thymine dimers in DNA upon photosensitization at 313 nm with acetophenone, acetone, or benzophenone. *Photochem. Photobiol.* **19**, 75–78 (1974).
175. Rauer, C., Nogueira, J. J., Marquetand, P. & González, L. Stepwise photosensitized thymine dimerization mediated by an exciton intermediate. *Monatshefte für Chemie-Chemical Mon.* **149**, 1–9 (2018).
176. Gontcharov, J. *et al.* Triplet-Induced Lesion Formation at CpT and TpC Sites in DNA. *Chemistry* **25**, 15164 (2019).
177. Trzcionka, J., Lhiaubet-Vallet, V. & Chouini-Lalanne, N. DNA photosensitization by indoprofen–is DNA damage photoinduced by indoprofen or by its photoproducts? *Photochem. Photobiol. Sci.* **3**, 226–230 (2004).
178. Lamola, A. A. Production of pyrimidine dimers in DNA in the dark. *Biochem. Biophys. Res. Commun.* **43**, 893–898 (1971).
179. Umlas, M. E., Franklin, W. A., Chan, G. L. & Haseltine, W. A. Ultraviolet light irradiation of defined-sequence DNA under conditions of chemical photosensitization. *Photochem. Photobiol.* **42**, 265–273 (1985).
180. Costalat, R. *et al.* Formation of cyclobutane thymine dimers photosensitized by pyridopsoralens: a triplet-triplet energy transfer mechanism. *Photochem. Photobiol.* **51**, 255–262 (1990).
181. Kanaly, R. A. *et al.* Development of the adductome approach to detect DNA damage in humans. *Antioxid. Redox Signal.* **8**, 993–1001 (2006).
182. Hemeryck, L. Y., Moore, S. A. & Vanhaecke, L. Mass spectrometric mapping of the DNA adductome as a means to study genotoxin exposure, metabolism, and effect. *Anal. Chem.* **88**, 7436–7446 (2016).
183. Chang, Y.-J., Cooke, M. S., Hu, C.-W. & Chao, M.-R. Novel approach to integrated DNA adductomics for the assessment of in vitro and in vivo environmental exposures. *Arch. Toxicol.* **92**, 2665–2680 (2018).
184. Balbo, S., Turesky, R. J. & Villalta, P. W. DNA adductomics. *Chem. Res. Toxicol.* **27**, 356–366 (2014).
185. Sloan, D. B., Broz, A. K., Sharbrough, J. & Wu, Z. Detecting rare mutations and DNA damage with sequencing-based methods. *Trends Biotechnol.* **36**, 729–740 (2018).
186. Panahi, Y. *et al.* Next-generation sequencing approaches for the study of genome and epigenome toxicity induced by sulfur mustard. *Arch. Toxicol.* **92**, 3443–3457 (2018).
187. Hu, J., Adebali, O., Adar, S. & Sancar, A. Dynamic maps of UV damage formation and repair for the human genome. *Proc. Natl. Acad. Sci.* **114**, 6758–6763 (2017).
188. Salk, J. J. & Kennedy, S. R. Next-generation genotoxicology: using modern sequencing technologies to assess somatic mutagenesis and cancer risk. *Environ. Mol. Mutagen.* **61**, 135–151 (2020).
189. Li, W. & Sancar, A. Methodologies for detecting environmentally induced DNA damage and repair. *Environ. Mol. Mutagen.* **61**, 664–679 (2020).
190. Alhegaili, A. S. *et al.* Genome-wide adductomics analysis reveals Heterogeneity in

the Induction and Loss of cyclobutane thymine Dimers across Both the Nuclear and mitochondrial genomes. *Int. J. Mol. Sci.* **20**, 5112 (2019).
191. Hu, J., Adar, S., Selby, C. P., Lieb, J. D. & Sancar, A. Genome-wide analysis of human global and transcription-coupled excision repair of UV damage at single-nucleotide resolution. *Genes Dev.* **29**, 948–960 (2015).
192. Hu, J., Lieb, J. D., Sancar, A. & Adar, S. Cisplatin DNA damage and repair maps of the human genome at single-nucleotide resolution. *Proc. Natl. Acad. Sci.* **113**, 11507–11512 (2016).
193. Li, W. *et al.* Human genome-wide repair map of DNA damage caused by the cigarette smoke carcinogen benzo [a] pyrene. *Proc. Natl. Acad. Sci.* **114**, 6752–6757 (2017).
194. Bryan, D. S., Ransom, M., Adane, B., York, K. & Hesselberth, J. R. High resolution mapping of modified DNA nucleobases using excision repair enzymes. *Genome Res.* **24**, 1534–1542 (2014).
195. Mao, P., Smerdon, M. J., Roberts, S. A. & Wyrick, J. J. Chromosomal landscape of UV damage formation and repair at single-nucleotide resolution. *Proc. Natl. Acad. Sci.* **113**, 9057–9062 (2016).
196. Narayanan, D. L., Saladi, R. N. & Fox, J. L. Ultraviolet radiation and skin cancer. *Int. J. Dermatol.* **49**, 978–986 (2010).
197. Müller, M. & Carell, T. Structural biology of DNA photolyases and cryptochromes. *Curr. Opin. Struct. Biol.* **19**, 277–285 (2009).
198. Mullenders, L. H. F. *et al.* UV-induced photolesions, their repair and mutations. *Mutat. Res. Toxicol.* **299**, 271–276 (1993).
199. Bourre, F., Renault, G., Seawell, P. C. & Sarasin, A. Distribution of ultraviolet-induced lesions in simian virus 40 DNA. *Biochimie* **67**, 293–299 (1985).
200. Kundu, L. M., Linne, U., Marahiel, M. & Carell, T. RNA is more UV resistant than DNA: The formation of UV-induced DNA lesions is strongly sequence and conformation dependent. *Chem. Eur. J.* **10**, 5697–5705 (2004).
201. Cannistraro, V. J. & Taylor, J.-S. Acceleration of 5-methylcytosine deamination in cyclobutane dimers by G and its implications for UV-induced C-to-T mutation hotspots. *J. Mol. Biol.* **392**, 1145–1157 (2009).
202. Lee, W. & Matsika, S. QM/MM studies reveal pathways leading to the quenching of the formation of thymine dimer photoproduct by flanking bases. *Phys. Chem. Chem. Phys.* **17**, 9927–9935.
203. Middleton, C. T. *et al.* DNA excited-state dynamics: from single bases to the double helix. *Annu. Rev. Phys. Chem.* **60**, 217–239 (2009).
204. Liu, L., Pilles, B. M., Reiner, A. M., Gontcharov, J. & Zinth, W. 2′-Methoxyacetophenone: An Efficient Photosensitizer for Cyclobutane Pyrimidine Dimer Formation. *ChemPhysChem* **16**, 3483–3487 (2015).
205. Liu, L., Pilles, B. M., Gontcharov, J., Bucher, D. B. & Zinth, W. Quantum yield of cyclobutane pyrimidine dimer formation via the triplet channel determined by photosensitization. *J. Phys. Chem. B* **120**, 292–298 (2016).
206. Svoboda, D. L., Smith, C. A., Taylor, J. S. & Sancar, A. Effect of sequence, adduct type, and opposing lesions on the binding and repair of ultraviolet photodamage by DNA photolyase and (A) BC excinuclease. *J. Biol. Chem.* **268**, 10694–10700 (1993).

207. Chan, G. L., Doetsch, P. W. & Haseltine, W. A. Cyclobutane pyrimidine dimers and (6-4) photoproducts block polymerization by DNA polymerase I. *Biochemistry* **24**, 5723–5728 (1985).
208. Carty, M. P., Lawrence, C. W. & Dixon, K. Complete Replication of Plasmid DNA Containing a Single UV-induced Lesion in Human Cell Extracts (∗). *J. Biol. Chem.* **271**, 9637–9647 (1996).
209. Higuchi, K. *et al.* Fate of DNA replication fork encountering a single DNA lesion during oriC plasmid DNA replication in vitro. *Genes to Cells* **8**, 437–449 (2003).
210. Mitchell, D. L. & Nairn, R. S. The biology of the (6-4) photoproduct. *Photochem. Photobiol.* **49**, 805–819 (1989).
211. Setlow, R. B. & Swenson, S. P. and Carrier, WL (1963). *Thymine dimers Inhib. DNA Synth. by Ultrav. Irradiat. cells. Sci.* **142**, 1464–1466.
212. Svoboda, D. L. & Vos, J. M. Differential replication of a single, UV-induced lesion in the leading or lagging strand by a human cell extract: fork uncoupling or gap formation. *Proc. Natl. Acad. Sci.* **92**, 11975–11979 (1995).
213. Veaute, X., Mari-Giglia, G., Lawrence, C. W. & Sarasin, A. UV lesions located on the leading strand inhibit DNA replication but do not inhibit SV40 T-antigen helicase activity. *Mutat. Res. Repair* **459**, 19–28 (2000).
214. Veaute, X. & Sarasin, A. Differential Replication of a SingleN-2-Acetylaminofluorene Lesion in the Leading or Lagging Strand DNA in a Human Cell Extract. *J. Biol. Chem.* **272**, 15351–15357 (1997).
215. McInerney, P. & O'Donnell, M. Functional uncoupling of twin polymerases: mechanism of polymerase dissociation from a lagging-strand block. *J. Biol. Chem.* **279**, 21543–21551 (2004).
216. Courcelle, J., Donaldson, J. R., Chow, K.-H. & Courcelle, C. T. DNA damage-induced replication fork regression and processing in Escherichia coli. *Science (80-.).* **299**, 1064–1067 (2003).
217. Friedberg, E. C., Walker, G. C., Siede, W. & Wood, R. D. *DNA repair and mutagenesis*. (American Society for Microbiology Press, 2005).
218. Axelrod, J. D. & Majors, J. An lmproved method for photofootprinting yeast genes in vivo using Taq polymerase. *Nucleic Acids Res.* **17**, 171–183 (1989).
219. Chandrasekhar, D. & Van Houten, B. High resolution mapping of UV-induced photoproducts in the Escherichia coli lacI gene: Inefficient repair of the non-transcribed strand correlates with high mutation frequency. *J. Mol. Biol.* **238**, 319–332 (1994).
220. Shu, X., Xiong, X., Song, J., He, C. & Yi, C. Base-resolution analysis of cisplatin–DNA adducts at the genome scale. *Angew. Chemie Int. Ed.* **55**, 14246–14249 (2016).
221. Lee, W. & Matsika, S. Role of charge transfer states into the formation of cyclobutane pyrimidine dimers in DNA. *Faraday Discuss.* **216**, 507–519 (2019).
222. Kool, E. T. Hydrogen bonding, base stacking, and steric effects in DNA replication. *Annu. Rev. Biophys. Biomol. Struct.* **30**, 1–22 (2001).
223. Cordes, M. & Giese, B. Electron transfer in peptides and proteins. *Chem. Soc. Rev.* **38**, 892–901 (2009).
224. Marcus, R. A. & Sutin, N. Electron transfers in chemistry and biology. *Biochim. Biophys. Acta (BBA)-Reviews Bioenerg.* **811**, 265–322 (1985).

225. Winkler, J. R. & Gray, H. B. Electron transfer in ruthenium-modified proteins. *Chem. Rev.* **92**, 369–379 (1992).
226. Isied, S. S., Ogawa, M. Y. & Wishart, J. F. Peptide-mediated intramolecular electron transfer: long-range distance dependence. *Chem. Rev.* **92**, 381–394 (1992).
227. Treadway, C. R., Hill, M. G. & Barton, J. K. Charge transport through a molecular π-stack: double helical DNA. *Chem. Phys.* **281**, 409–428 (2002).
228. Núñez, M. E., Hall, D. B. & Barton, J. K. Long-range oxidative damage to DNA: effects of distance and sequence. *Chem. Biol.* **6**, 85–97 (1999).
229. Riley, K. E. & Hobza, P. On the importance and origin of aromatic interactions in chemistry and biodisciplines. *Acc. Chem. Res.* **46**, 927–936 (2013).
230. Černý, J. & Hobza, P. Non-covalent interactions in biomacromolecules. *Phys. Chem. Chem. Phys.* **9**, 5291–5303 (2007).
231. Jurecka, P. & Hobza, P. True stabilization energies for the optimal planar hydrogen-bonded and stacked structures of guanine⊙⊙⊙ cytosine, adenine⊙⊙⊙ thymine, and their 9-and 1-methyl derivatives: complete basis set calculations at the MP2 and CCSD (T) levels and comparison with . *J. Am. Chem. Soc.* **125**, 15608–15613 (2003).
232. Cerny, J., Kabelác, M. & Hobza, P. Double-helical→ ladder structural transition in the B-DNA is induced by a loss of dispersion energy. *J. Am. Chem. Soc.* **130**, 16055–16059 (2008).
233. Siriwong, K. & Voityuk, A. A. Electron transfer in DNA. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2**, 780–794 (2012).
234. Karabıyık, H., Sevinçek, R. & Karabıyık, H. π-Cooperativity effect on the base stacking interactions in DNA: is there a novel stabilization factor coupled with base pairing H-bonds? *Phys. Chem. Chem. Phys.* **16**, 15527–15538 (2014).
235. Fisher, G. J. & Johns, H. E. Photochemistry and Photobiology of Nucleic Acids. *by SY Wang, Acad. Press. New York* **1**, 225–294 (1976).
236. Fenick, D. J., Carr, H. S. & Falvey, D. E. Synthesis and photochemical cleavage of cis-syn pyrimidine cyclobutane dimer analogs. *J. Org. Chem.* **60**, 624–631 (1995).
237. Varghese, A. J. Photochemical reactions of cytosine nucleosides in frozen aqueous solution and in deoxyribonucleic acid. *Biochemistry* **10**, 2194–2199 (1971).
238. Hariharan, M., McCullagh, M., Schatz, G. C. & Lewis, F. D. Conformational control of thymine photodimerization in single-strand and duplex DNA containing locked nucleic acid TT steps. *J. Am. Chem. Soc.* **132**, 12856–12858 (2010).
239. Pan, Z., McCullagh, M., Schatz, G. C. & Lewis, F. D. Conformational control of thymine photodimerization in purine-containing trinucleotides. *J. Phys. Chem. Lett.* **2**, 1432–1438 (2011).
240. Johnson, A. T. & Wiest, O. Structure and dynamics of poly (T) single-strand DNA: implications toward CPD formation. *J. Phys. Chem. B* **111**, 14398–14404 (2007).
241. Law, Y. K., Azadi, J., Crespo-Hernández, C. E., Olmon, E. & Kohler, B. Predicting thymine dimerization yields from molecular dynamics simulations. *Biophys. J.* **94**, 3590–3600 (2008).
242. McCullagh, M. *et al.* Conformational control of TT dimerization in DNA conjugates. A molecular dynamics study. *J. Phys. Chem. B* **114**, 5215–5221

(2010).
243. Osakada, Y., Kawai, K. & Majima, T. Kinetics of charge transfer through DNA across guanine–cytosine repeats intervened by adenine–thymine base pair (s). *Bull. Chem. Soc. Jpn.* **86**, 25–30 (2013).
244. Guckian, K. M. *et al.* Factors contributing to aromatic stacking in water: evaluation in the context of DNA. *J. Am. Chem. Soc.* **122**, 2213–2222 (2000).
245. Pilles, B. M. *et al.* Mechanism of the decay of thymine triplets in DNA single strands. *J. Phys. Chem. Lett.* **5**, 1616–1622 (2014).
246. Galtier, N. *et al.* Codon usage bias in animals: disentangling the effects of natural selection, effective population size, and GC-biased gene conversion. *Mol. Biol. Evol.* **35**, 1092–1103 (2018).
247. Sharp, P. M., Emery, L. R. & Zeng, K. Forces that influence the evolution of codon bias. *Philos. Trans. R. Soc. B Biol. Sci.* **365**, 1203–1212 (2010).
248. Akashi, H. Synonymous codon usage in Drosophila melanogaster: natural selection and translational accuracy. *Genetics* **136**, 927–935 (1994).
249. Li, Y. *et al.* Codon Usage Bias in Autophagy-Related Gene 13 in Eukaryotes: Uncovering the Genetic Divergence by the Interplay Between Nucleotides and Codon Usages. *Front. Cell. Infect. Microbiol.* 1063 (2021).
250. Labella, A. L., Opulente, D. A., Steenwyk, J. L., Hittinger, C. T. & Rokas, A. Variation and selection on codon usage bias across an entire subphylum. *PLoS Genet.* **15**, e1008304 (2019).
251. Rudolph, K. L. M. *et al.* Codon-driven translational efficiency is stable across diverse mammalian cell states. *PLoS Genet.* **12**, e1006024 (2016).
252. Pouyet, F., Mouchiroud, D., Duret, L. & Sémon, M. Recombination, meiotic expression and human codon usage. *Elife* **6**, e27344 (2017).
253. Doherty, A. & McInerney, J. O. Translational selection frequently overcomes genetic drift in shaping synonymous codon usage patterns in vertebrates. *Mol. Biol. Evol.* **30**, 2263–2267 (2013).
254. Mao, P. & Wyrick, J. J. Genome-wide mapping of UV-induced DNA damage with CPD-seq. in *The Nucleus* 79–94 (Springer, 2020).
255. Mao, P. *et al.* ETS transcription factors induce a unique UV damage signature that drives recurrent mutagenesis in melanoma. *Nat. Commun.* **9**, 1–13 (2018).
256. Poulos, R. C. *et al.* Functional mutations form at CTCF-cohesin binding sites in melanoma due to uneven nucleotide excision repair across the motif. *Cell Rep.* **17**, 2865–2872 (2016).
257. Sabarinathan, R., Mularoni, L., Deu-Pons, J., Gonzalez-Perez, A. & López-Bigas, N. Nucleotide excision repair is impaired by binding of transcription factors to DNA. *Nature* **532**, 264–267 (2016).
258. Adar, S., Hu, J., Lieb, J. D. & Sancar, A. Genome-wide kinetics of DNA excision repair in relation to chromatin state and mutagenesis. *Proc. Natl. Acad. Sci.* **113**, E2124–E2133 (2016).
259. García-Nieto, P. E. *et al.* Carcinogen susceptibility is regulated by genome architecture and predicts cancer mutagenesis. *EMBO J.* **36**, 2829–2843 (2017).
260. Mao, P., Wyrick, J. J., Roberts, S. A. & Smerdon, M. J. UV-induced DNA damage and mutagenesis in chromatin. *Photochem. Photobiol.* **93**, 216–228 (2017).
261. Roberts, S. A., Brown, A. J. & Wyrick, J. J. Recurrent noncoding mutations in skin

cancers: UV damage susceptibility or repair inhibition as primary driver? *Bioessays* **41**, 1800152 (2019).
262. Perera, D. *et al.* Differential DNA repair underlies mutation hotspots at active promoters in cancer genomes. *Nature* **532**, 259–263 (2016).
263. Jiang, Y. *et al.* Super hotspots and super coldspots in the repair of UV-induced DNA damage in the human genome. *J. Biol. Chem.* **296**, 100581 (2021).
264. Premi, S. *et al.* Genomic sites hypersensitive to ultraviolet radiation. *Proc. Natl. Acad. Sci.* **116**, 24196–24205 (2019).

Appendix A.

Supplementary figures

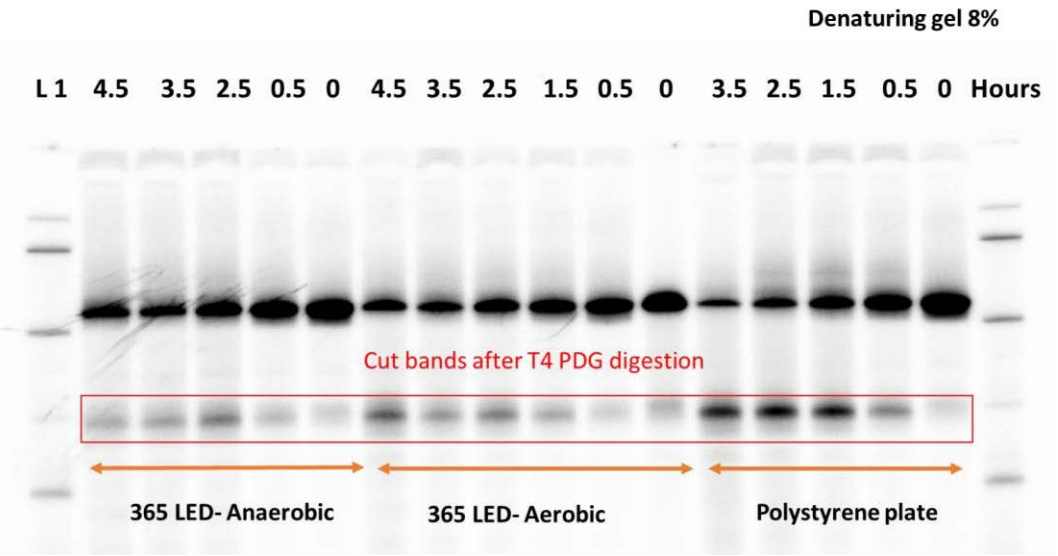


Figure S1: Three time-course experiments of ^{32}P-radiolabelled library DNA samples irradiated at 365 nm under anaerobic condition, aerobic condition and using polystyrene plate and UV transilluminator in the presence of 10 nm of 2′-acetophenone (AP).

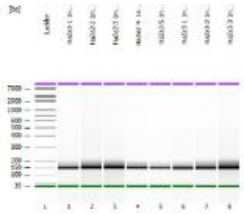
Appendix B.

BioAnalzyer Quality Control of all sequenced samples

Assay Class: High Sensitivity DNA Assay
Data Path: C:\Users\ss7\Desktop\HaZe2 and 3.xad

Created: 3/15/2021 3:56:42 PM
Modified: 3/16/2021 10:06:17 AM

Electrophoresis File Run Summary



Instrument Information:

Instrument Name:	DE13804608	Firmware:	C.01.069
Serial#:	DE13804608	Type:	G2939A

Assay Information:

Assay Origin Path: C:\Program Files\Agilent\2100 bioanalyzer\2100 expert\assays\dsDNA\High Sensitivity DNA.xsy

Assay Class: High Sensitivity DNA Assay

Version: 1.03

Assay Comments: Copyright © 2003-2010 Agilent Technologies

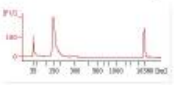
Chip Information:

Chip Lot #:

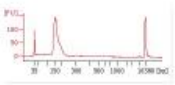
Reagent Kit Lot #:

Chip Comments:

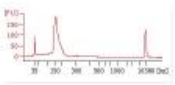
HaZe2-1 1ng/ul DNA (15/03/21)



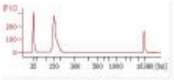
HaZe2-2 Undil DNA (15/03/21)



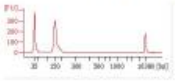
HaZe2-3 Undil I DNA (15/03/21)



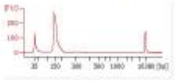
HaZe2-4- 1ng/ul DNA (15/03/21)



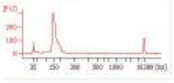
HaZe2-5 1ng/ul DNA (15/03/21)



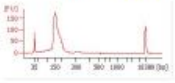
HaZe3-1 1ng/ul DNA (15/03/21)



HaZe3-2 1ng/ul DNA (15/03/21)



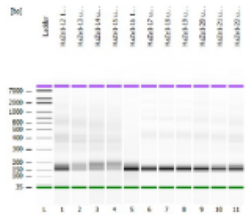
HaZe3-3 1ng/ul DNA (15/03/21)



Assay Class: High Sensitivity DNA Assay
Data Path: C:\Users\ss7\Desktop\HaZa1_2.xad

Created: 3/5/2021 3:10:46 PM
Modified: 3/9/2021 3:19:53 PM

Electrophoresis File Run Summary



Instrument Information:

Instrument Name: DE13804608 Firmware: C.01.069
Serial#: DE13804608 Type: G2939A

Assay Information:

Assay Origin Path: C:\Program Files\Agilent\2100 bioanalyzer\2100 expert\assays\dsDNA\High Sensitivity DNA.xsy
Assay Class: High Sensitivity DNA Assay
Version: 1.03
Assay Comments: Copyright © 2003-2010 Agilent Technologies

Chip Information:

Chip Lot #:
Reagent Kit Lot #:
Chip Comments:

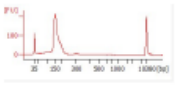
HaZe1-12 1ng/ul DNA (5/3/2021)

HaZe1-13 undil DNA (5/3/2021)

HaZe1-14 undil DNA (5/3/2021)

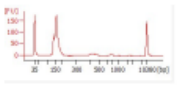
HaZe1-15 undil DNA (5/3/2021)

HaZe1-16 1ng/ul DNA (5/3/2021)



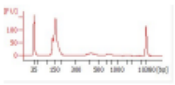
HaZe1-17 undil DNA (5/3/2021)

HaZe1-18 undil DNA (5/3/2021)

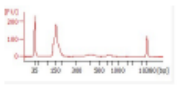


HaZe1-19 undil DNA (5/3/2021)

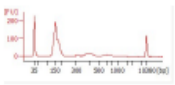
HaZe1-20 undil DNA (5/3/2021)



HaZe1-21 undil DNA (5/3/2021)



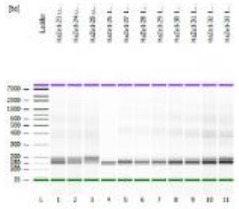
HaZe1-22 undil DNA (5/3/2021)



Assay Class: High Sensitivity DNA Assay
Data Path: C:\Users\ss7\Desktop\HaZa1_3.xad

Created: 3/5/2021 4:06:22 PM
Modified: 3/9/2021 3:20:58 PM

Electrophoresis File Run Summary



Instrument Information:

Instrument Name:	DE13804608	Firmware:	C.01.069
Serial#:	DE13804608	Type:	G2939A

Assay Information:

Assay Origin Path: C:\Program Files\Agilent\2100 bioanalyzer\2100 expert\assays\dsDNA\High Sensitivity DNA.xsy

Assay Class: High Sensitivity DNA Assay

Version: 1.03

Assay Comments: Copyright © 2003-2010 Agilent Technologies

Chip Information:

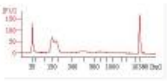
Chip Lot #:

Reagent Kit Lot #:

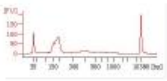
Chip Comments:

HaZe1-23 undil DNA (5/3/2021)

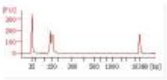
HaZe1-24 undil DNA (5/3/2021)



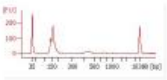
HaZe1-25 undil DNA (5/3/2021)



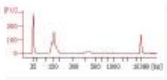
HaZe1-26 1ng/ul DNA (5/3/2021)



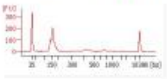
HaZe1-27 1ng/ul DNA (5/3/2021)



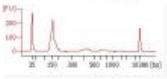
HaZe1-28 1ng/ul DNA (5/3/2021)



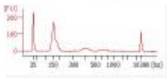
HaZe1-29 1ng/ul DNA (5/3/2021)



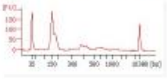
HaZe1-30 1ng/ul DNA (5/3/2021)



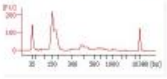
HaZe1-31 1ng/ul DNA (5/3/2021)



HaZe1-32 1ng/ul DNA (5/3/2021)



HaZe1-33 1ng/ul DNA (5/3/2021)



Appendix C

A number of relevant files containing the statistical analyses and calculations are added as supplemental files in the thesis submission system. These files are named as 278 nm Experiment, 375 nm Experiment, conf_p and conf_t. These supplemental files are pertinent to some results demonstrated in the thesis and are available in the SFU thesis database.

Appendix D

Method:

A combinatorial SELEX-based approach based on a gel mobility-shift induced by binding of an irradiated random duplex DNA library to inactive CPD photolyase enzyme (oxidized form).

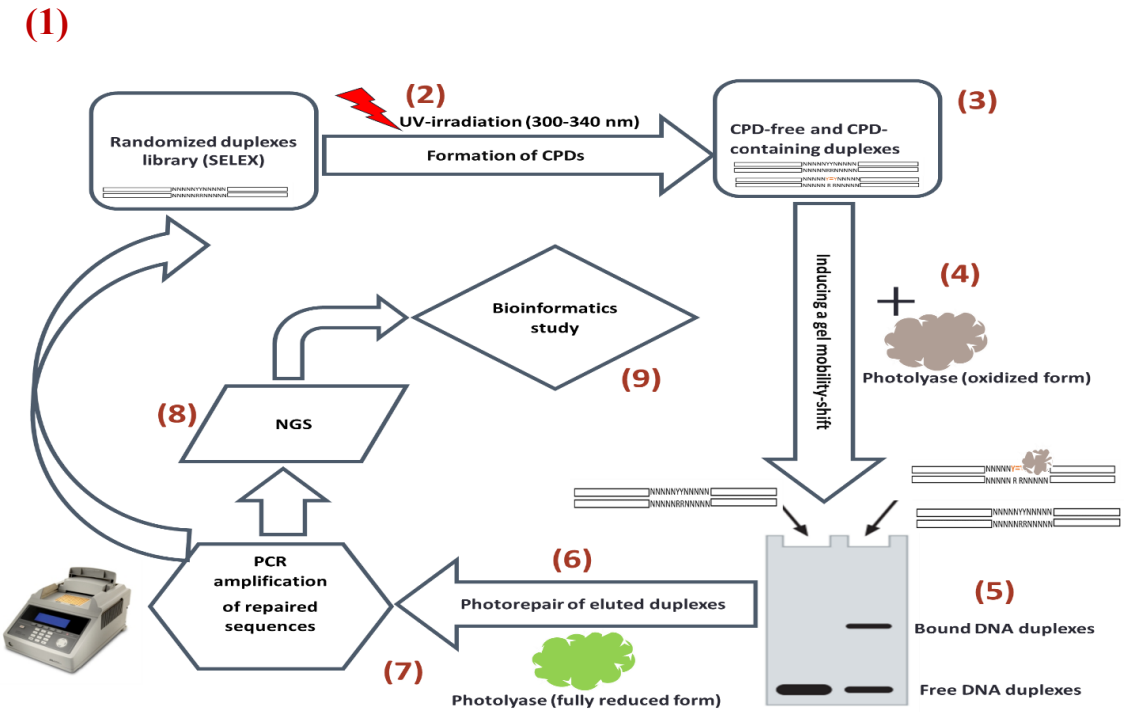


Figure S2: Flow chart illustrating the multi-round SELEX coupled with round-by-round quantitation of sequence enrichment by deep sequencing, starting from ds DNA library.

Result:

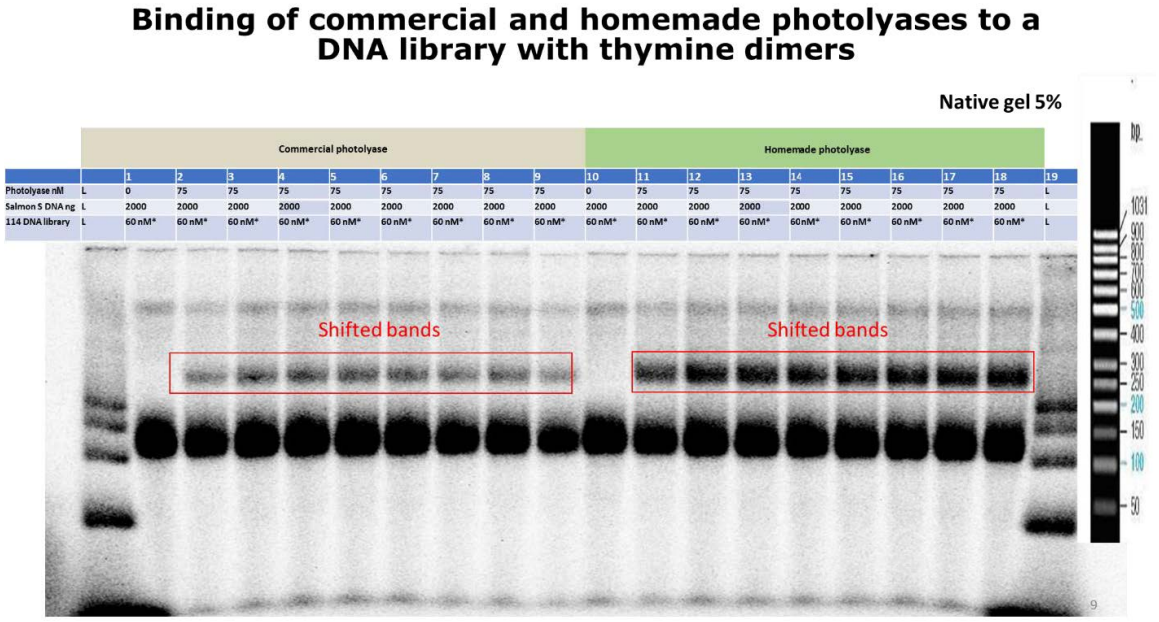


Figure S3: Electrophoretic mobility shift assay (EMSA) using home-made (550 H) and commercial (550 C) photolyase enzyme in the presence of increment amount of salmon sperm DNA (0, 20, 200 and 2000 ng): homemade photolyase showed higher activity than commercial one, but after adding 2000 ng of Salmon sperm both showed an identical result: lanes 16 and 17 (red box).

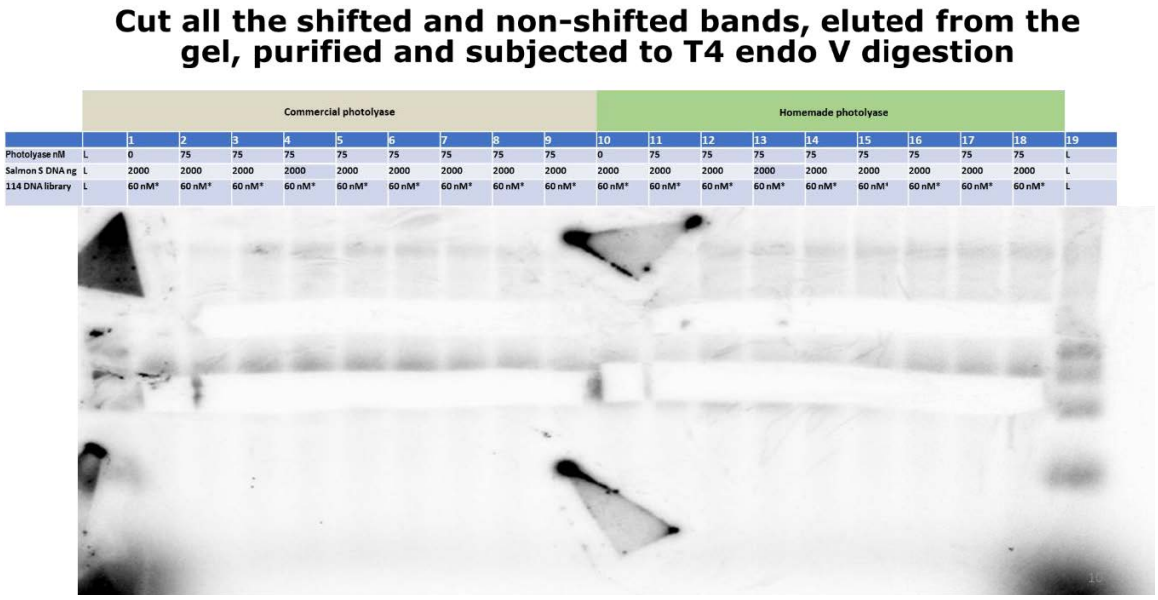


Figure S4: All shifted and non-shifted bands were eluted from the gel, subjected to chloroform: phenol ext., ethanol ppt, T4 endo V digestion and loaded onto denaturing gel.

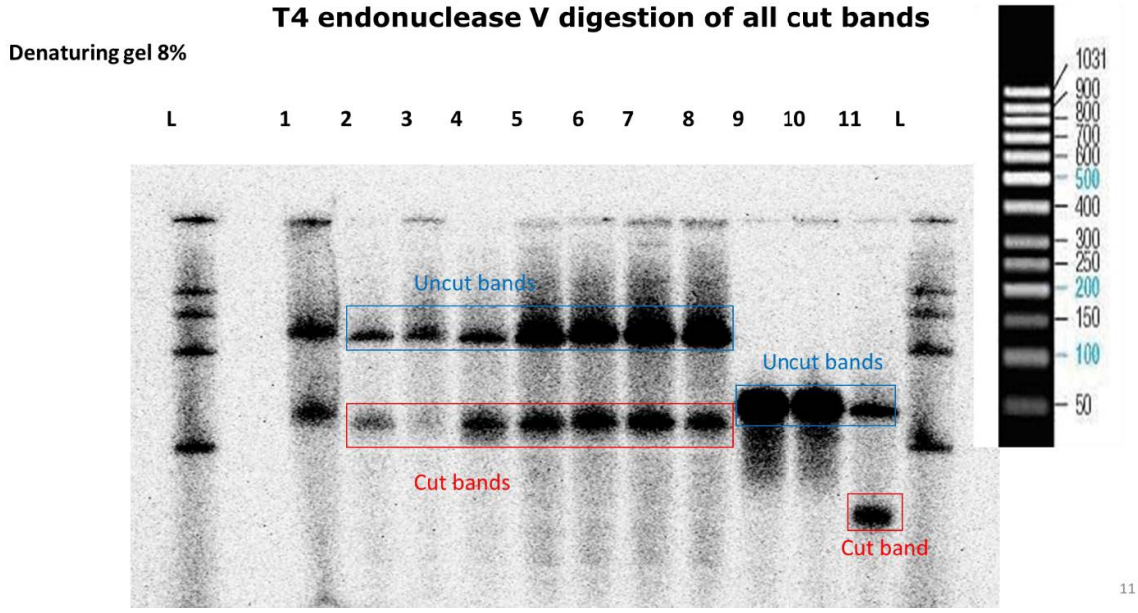


Figure S5: All shifted and non-shifted bands were eluted from the gel, subjected to chloroform: phenol ext., ethanol ppt, T4 endo V digestion and loaded onto denaturing gel.

Appendix E.

Supplementary Data Files

Filename:

278 nm Experiments.xlsx

Filename:

365 nm Experiment.xlsx

Filename:

conf_p.xlsx

Filename:

conf_t.xlsx