
Computational Studies of Natural and Nonnatural Nucleic Acids: Formation and Repair of Thymine Dimers and Structure/Dynamics of GNA

Andrew Johnson

Publication Date

11-04-2011

License

This work is made available under a All Rights Reserved license and should only be used in accordance with that license.

Citation for this work (American Psychological Association 7th edition)

Johnson, A. (2011). *Computational Studies of Natural and Nonnatural Nucleic Acids: Formation and Repair of Thymine Dimers and Structure/Dynamics of GNA* (Version 1). University of Notre Dame.
<https://doi.org/10.7274/12579s17s9c>

This work was downloaded from CurateND, the University of Notre Dame's institutional repository.

For more information about this work, to report or an issue, or to preserve and share your original work, please contact the CurateND team for assistance at curate@nd.edu.

COMPUTATIONAL STUDIES OF NATURAL AND NONNATURAL NUCLEIC
ACIDS: FORMATION AND REPAIR OF THYMINE DIMERS AND
STRUCTURE/DYNAMICS OF GNA

A Dissertation

Submitted to the Graduate School
of the University of Notre Dame
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

by

Andrew T. Johnson

Olaf Wiest, Director

Graduate Program in Chemistry and Biochemistry

Notre Dame, Indiana

April 2011

© Copyright 2011

Andrew T. Johnson

COMPUTATIONAL STUDIES OF NATURAL AND NONNATURAL NUCLEIC
ACIDS: FORMATION AND REPAIR OF THYMINE DIMERS AND
STRUCTURE/DYNAMICS OF GNA

Abstract

by

Andrew T. Johnson

Exposure of DNA to UV-light (260-320 nm) leads to the formation of two major photolesions between adjacent pyrimidine nucleobases, the cyclobutane pyrimidine dimer (CPD) and the (6-4) photoproduct (6-4PP). Both originate from an ultrafast [2 + 2] cycloaddition and the process leads to the covalent linking between the two bases. This damage produces stalling of DNA replication and transcription and causes errors in the genome leading to phenotypic display of skin cancer. Many species contain enzymes capable of repairing these photolesions, however, humans do not and must rely on nucleotide excision repair to replace the bases.

DNA photolyase is an enzyme capable of binding to and repairing CPD photolesions. A redox active flavin cofactor (FADH[•]) found near the active site is capable of donating an electron into the CPD breaking orbital symmetry and leading to a breaking of the cyclobutane ring covalently linking the two bases. This repair mechanism is light dependant and highly efficient. As humans do not possess this

enzyme, there is much interest in developing small molecules which can mimic DNA photolyase. Our group has developed an artificial photolyase which is capable to binding to CPDs in water effecting repair. The work here builds upon previous studies by testing this artificial photolyase on duplex DNA.

(6-4) photolyase is an enzyme capable of repairing the 6-4PP. In contrast to the mechanism of CPD repair, little is known about mechanism employed by (6-4) photolyase. Much debate exists in the literature on what this mechanism is, however, each proposed mechanism is lacking in experimental evidence and support. In this work we use high level electronic structure methods to study the energetic landscape of the proposed mechanisms both in solution and in the (6-4) photolyase active site. By using these methods we conclude that none of the currently proposed mechanisms are energetically feasible and a new, two proton repair mechanism is proposed for the repair of 6-4PP.

Much work has been done by scientists in modifying the nucleic backbone of nucleic acids. However, only a few unnatural nucleic acids are known to form stable duplexes. The newly discovered glycol nucleic acid (GNA) is perhaps the simplest backbone scheme capable of duplex formation. It consists of a phosphodiester backbone connected by repeating propylene glycol subunits and is completely acyclic. GNA forms duplexes far exceeding the thermal stabilities of DNA with melting temperatures $\approx 20^{\circ}\text{C}$ greater. Thermodynamic analysis shows that GNA duplex formation is entropically less penalizing than the same process in DNA and RNA. This is a counterintuitive notion considering the acyclic nature of the GNA backbone. This study uses molecular dynamics (MD) simulations to study the structure and dynamics of dsGNA and uncovers

a coiling/ uncoiling mode which explains the decreased entropic penalty of annealing in comparison to natural nucleic acids. MD simulations are also used to study the arrangement of porphyrin base pairs incorporated into GNA and support the notion of a slipped cofacial geometry of adjacent porphyrin moieties which has been speculated from the experimental data.

This is for my family.

CONTENTS

Figures.....	v
Tables.....	x
Acknowledgments.....	xi
Chapter 1: Introduction	1
1.1 Thymine Dimer Formation and Biological Effects	1
1.2 Thymine Dimer Repair Mechanisms	5
1.3 Unnatural Nucleic Acid Analogs	11
Chapter 2: Structure and Dynamics of Poly(T) Single Strand DNA: Implications for CPD Formation	16
2.1 Introduction.....	16
2.2 Methods.....	22
2.3 Results and Discussion	24
2.3.1 Base Stack Interactions: Distances	24
2.3.2 Base Stack Interactions: Dihedrals	36
2.4 Conclusions.....	41
Chapter 3: Computational Studies of the (6-4) Photoproduct Repair Reaction.....	45
3.1 Introduction.....	45
3.2 Methods.....	54
3.3 Results.....	55
3.3.1 Calculations on 6-4PP.....	55
3.3.2 Calculations on 6-4PP + HIS365 + HIS369	58
3.3.3 Calculation on 6-4PP + HIS365 + HIS369 + TYR423.....	62
3.3.4 Calculations on the Expanded Theozyme Model	64
3.4 Conclusions.....	65
Chapter 4: Evaluation of a Biomimetic Artificial Photolyase on CPD Repair Within Duplex DNA	69
4.1 Introduction.....	69
4.2 Experimental	74
4.3 Results.....	80
4.4 Future Outlook	85
Chapter 5: MD Simulations of Glycol Nucleic Acids	87

5.1 Introduction.....	87
5.2 Methods.....	105
5.2.1 CuGNA, BrGNA, and GNA	105
5.2.2 Porphyrin Containing GNA	111
5.3 Instructions for Running GNA Simulations	112
5.4 Results for MD Simulations of GNA.....	115
5.4.1 8mer-CuGNA.....	115
5.4.2 16mer-CuGNA.....	117
5.4.3 6mer-BrGNA	117
5.4.4 8mer-BrGNA and 16mer-BrGNA	121
5.4.5 8mer-GNA	124
5.4.6 16mer-GNA	127
5.4.7 Simulations of Porphyrin Containing dsGNA	132
5.5 Conclusions.....	139
References.....	142

FIGURES

Figure 1.1- Formation of the two most common thymine dimers in DNA.	2
Figure 1.2- A) Scheme of the Diels-Alder reaction between 1,3-Butadiene and ethylene. B) Diagram of the electronic configuration and orbital symmetry present between the HOMO and LUMO orbitals in a [4 + 2] cycloaddition.	3
Figure 1.3- A) Scheme of the [2 + 2] cycloaddition between two ethylene molecules. B) Electronic configuration and orbital symmetry present in the ground state. C) Electronic and orbital symmetry present in the excited state.	4
Figure 1.4- Crystal structure of DNA photolyase from <i>Anacystis nidulans</i> . Orange- Bound DNA. Yellow- Repaired thymines in the flipped-out conformation in the active site. Red- FADH ⁻ cofactor. Blue- 8-HDF light harvesting cofactor. ¹¹	6
Figure 1.5- Catalytic cycle and intermediates involved in CPD repair by DNA photolyase. ¹²	8
Figure 1.6- Plot of C5-C5' (blue) and C6-C6' (green) distances versus time. Inset depicts the formation of the double bonds between C5-C6 (black) and C5'-C6' (red). ¹⁸ .10	10
Figure 1.7- Early mechanism proposed by Sancar for the repair of 6-4PP by (6-4) photolyase. ¹²	11
Figure 1.8- Structure of the PNA backbone.....	12
Figure 1.9- Backbone structure of LNA. ²¹	13
Figure 1.10- Structure of threose nucleic acid.	14
Figure 1.11- Structure of glycol nucleic acid.....	15
Figure 2.1- Formation of the CPD.	16
Figure 2.2- Coordinates of interest to CPD formation.....	21
Figure 2.3- Radial distribution functions for C6-C6' (A-D) and C5-C5' (E-H) distances from the simulation. Plots are colored by which base stack is represented (inset). Figure spans 3 pages.	26

Figure 2.4- Overall view of the globular structure fully formed at the end of the 50 ns MD simulation.....	30
Figure 2.5- Close up view of hairpin region of the poly(T) ssDNA containing duplex like hydrogen bonding arrangement.	31
Figure 2.6- Distance of 3% additive probabilities for each base stack.	32
Figure 2.7- Plot of hydrogen bond distances versus time for base pairs involved in duplex like hydrogen bonding interactions.....	35
Figure 2.8- Plot of populations of base stack dihedral angles with values of $27 \pm 3^\circ$	37
Figure 2.9- Histograms C5-C6-C6'-C5' dihedral angles between individual base stacks for the MD simulation. Figure spans 2 pages.	39
Figure 2.10- Sequences used to study context dependence on CPD formation. ⁵⁵	43
Figure 2.11- CPD yields versus time for the sequences studied by Hariharan and Lewis. ⁵⁵	43
Figure 3.1- Mechanism of CPD and 6-4PP formation.....	45
Figure 3.2- Original mechanistic proposal involving an oxetane intermediate put forth by Sancar and coworkers. ¹²	46
Figure 3.3- Water transfer mechanism proposed by Carell and coworkers along with the publishing of the first crystal structure of (6-4) photolyase. ⁶⁶	50
Figure 3.4- Hydroxyl transfer mechanism proposed by Domratcheva and Schlichting. Energies are relative to the 6-4PP radical anion. ⁶⁹	51
Figure 3.5- Reaction coordinate showing the nonadiabatic repair mechanism proposed by Domratcheva and Schlichting. The oxetane mechanism is shown on the left while the hydroxyl mechanism is shown on the right. ⁶⁹	52
Figure 3.6- Mechanistic proposal put forth by Sancar and coworkers based upon observations on their ultrafast spectroscopy study. ⁶⁸	53
Figure 3.7- Transition structure of hydroxyl transfer in 6-4PP model. Activation energy ≈ 45 kcal/mol.	57
Figure 3.8- Transition structure of water dissociation. Activation energy ≈ 7.2 kcal/mol.	58
Figure 3.9- Optimized ground state structure of 6-4PP with the active site histidines.....	59
Figure 3.10- Optimized structure of the active site following injection of an electron. Notice the rotation of H365 to accommodate hydrogen bonding to N3'.	59

Figure 3.11- Transition state structure for hydroxyl transfer in the theozyme model.	60
Figure 3.12- Transition state structure with H369 protonated. Transition state involves a synchronous breaking of C-O bond with formation of O-H bond. ≈ 32 kcal/mol endothermic from the radical cation system.	61
Figure 3.13- Hydrogen bonding network present in 6-4PP + HIS365 + HIS369 + TYR423 theozyme model.	63
Figure 3.14- Transition state structure of hydroxyl transfer in the theozyme model including TYR423. Activation energy =32.6 kcal/mol from N3' protonated species.	63
Figure 3.15- SOMO of 6-4PP in the theozyme active site.	64
Figure 3.16- Previously proposed intermediate along the 6-4PP repair pathway. ^{66,68,69} ...	66
Figure 4.1- Formation of the CPD in DNA.	70
Figure 4.2- Crystal structure of Zn(II) cyclen bound to thymine nucleotide. ⁸⁰	72
Figure 4.3- Scheme showing the function of the artificial photolyase. R- ⁿ Bu or (CH ₂) ₂ O(CH ₂) ₂ OCH ₃	73
Figure 4.4- HPLC trace of damage solution after irradiation. CPD containing strand elutes around 8 minutes while the undamaged strand elutes around 14 minutes. .	76
Figure 4.5- ESI-MS of the CPD containing ssDNA.	77
Figure 4.6- ESI-MS of the undamaged DNA.	78
Figure 4.7- HPLC trace of collected CPD containing DNA.....	79
Figure 4.8- Experimental setup for the repair reaction. Repair solution is contained in the vial on the right and is under an argon atmosphere.	81
Figure 4.9- Repair reaction containing artificial photolyase.	82
Figure 4.10- Repair versus time for repair reaction.....	82
Figure 4.11- Control reaction without artificial photolyase.....	83
Figure 4.12- Repair reaction using the Rayonet lamp.	83
Figure 4.13- Control reaction (no artificial photolyase) using the Rayonet lamp.	84
Figure 5.1- Structures of common natural and synthetic nucleic acids.	88

Figure 5.2- CD spectra of the melting of the a) GNA and b) DNA single strands. ³⁶	91
Figure 5.3- Crystal structure of 8mer-CuGNA. ³⁸	93
Figure 5.4- Descriptions of helicoidal parameters describing nucleic acid structure. ⁹²	95
Figure 5.5- View of interstrand base stacking found in GNA.	96
Figure 5.6- View of the dihedrals discussed in this work.	97
Figure 5.7- Crystal structure of 6mer-BrGNA. ³⁴	99
Figure 5.8- AMBER atom types used to describe the GNA backbone.	108
Figure 5.9- An overlay of the crystal structure (yellow) and average structure from the MD simulation (blue).	118
Figure 5.10- View of average backbone dihedral from MD (a) and crystal structure (b). Overlay of the experimental (red) and MD average (green) structures of 6mer-BrGNA.	120
Figure 5.11- Overlay of the 16mer-BrGNA crystal structure (yellow) and average MD structure (blue).	123
Figure 5.12- Average structure from simulation of 8mer-GNA.	126
Figure 5.13- Plots of backbone rmsd and average base pair slide versus time for the simulation of 16mer-GNA.	129
Figure 5.14- Snapshots from the 100 ns simulation of 16mer-GNA depicting the uncoiled (left) and coiled (right) conformations.	130
Figure 5.15- Plots of hydrogen bond distance versus time for a terminal base pair in 16mer-DNA (red) and 16mer-GNA (blue).	131
Figure 5.16- Average structure obtained from the MD simulation of GNA containing on porphyrin moiety.	133
Figure 5.17- Average structure obtained from the MD simulation of a GNA sequence containing two adjacent porphyrin moieties.	134
Figure 5.18- View of the π stacking between the porphyrin moiety and the adjacent base pairs of entry 3.	135
Figure 5.19- View of the slipped cofacial geometry of the adjacent porphyrins predicted from the MD simulation of entry 7.	136

Figure 5.20- Views of porphyrin base stacking from the convex (top) and concave (bottom) sides of the GNA duplex.....	137
Figure 5.21- Backbone rmsd versus time observed in the simulation of entry 7.....	138

TABLES

Table 5.1 Thermodynamic and thermal stabilities of (S)-GNA and DNA duplexes (in parenthesis). ³⁶	89
Table 5.2 Comparison of Average Helical Parameters for GNA Duplex Crystal Structures in to B-DNA and A-DNA. ^{90a}	94
Table 5.3 Thermal stabilities of porphyrin-containing 16mer GNA duplexes together with Watson-Crick reference duplexes. ¹⁰³	103
Table 5.4 List of abbreviations used and simulations performed in this study. ^{90a}	106
Table 5.5 Additional information on the systems studied.	109
Table 5.6 Average helicoidal parameters for GNA duplex structures obtained from experiment and MD simulations.	116
Table 5.7 Average helicoidal parameters of 6mer-BrGNA crystal structure and average MD structure. ⁹⁰	119
Table 5.8 Helicoidal parameters for 8mer and 16mer-BrGNA. ⁹⁰	122
Table 5.9 Comparison of helicoidal parameters of 8mer-DNA and GNA. ⁹⁰	125
Table 5.10 Comparison of helicoidal parameters of 8mer-DNA and 16mer-GNA.	127
Table 5.11 Results from <i>nmode</i> analysis of DNA and GNA annealing (kcal/mol). ⁹⁰	132

ACKNOWLEDGMENTS

I have been extraordinarily fortunate to work with and be taught by many great educators who, without them I would not be where I am today. In high school my chemistry teachers, Tim Taylor and Felishia Hill, first sparked my interest in the subject and challenged their students beyond what was typically expected at the school. In college Dr. Rodderick Macrae convinced me after taking his ‘Chemistry in Society’ to fulfill my science requirement to become a chemistry major, a tough sell at the time. He has also taught a staggering 50% of all chemistry classes I’ve taken in college and recruited me to the ‘dark side’ (aka computational chemistry) while I was at Marian. I must also thank Dr. Carl Lecher who introduced me to organic chemistry and taught me the basics of reactivity of which I will apply for the rest of my days.

In my graduate school classes I have learned more than I ever imagined I could. From physical organic chemistry to biophysics and thermodynamics, I have many professors to thank for a challenging and enlightening curriculum. The research presented in this work would not have been possible without the in depth understanding that my professors at the University of Notre Dame provided me through my courses.

Perhaps the greatest interest I found in graduate school is my love for teaching. In this respect I must thank Dr. Paul Helquist for allowing me the opportunity to assume real responsibility in my teaching assignment. While rather daunted the first time I lectured,

his words of encouragement and helpful comments led to me to really enjoy teaching. Teaching for him was perhaps the greatest learning experience I had in graduate school. Watching how he taught and interacted with the students has given me a model of dedication that I will attempt to follow in my future as an educator.

I would also like to thank everyone who contributed to this work. In particular I want to thank the Smith lab for their endless supply of HPLC grade water and accessibility to their equipment when needed. Dr. Smith for accepting me into his lab my first summer, it was a great experience. He also provided me with nice challenge on the golf course which should not be understated. Dr. Bill Boggess for the help determining the molecular weights of the nucleic acids referred to in this work. I would also like to acknowledge Dr. Meggers and Dr. Essen for their collaboration on the GNA project.

Most influential to me during my time at the University of Notre Dame was, of course, my advisor Dr. Olaf Wiest. Our daily discussions, whether about chemistry or any other topic under the sun, are always entertaining and enlightening. His knowledge is so vast that I am certain that I learned a great deal simply through diffusion. Working with him has both broadened and deepened my understanding of science. He taught me to think critically and not to believe everything you read which, at times, wasn't very difficult in our chosen field!

Last, but certainly not least, I thank my parents. As a child they made me appreciate the importance of a good education and have supported me in every endeavor. Whether it be taking an active role in teaching me as a child or driving across the country to watch me play golf they have always been there. I remember struggling with balancing chemical equations when I took my first chemistry class in high school. My

father took a little time to help me through some problems and I was on my way. Most of all this showed me that they cared about my education, a presence that is absent in many children's lives. I have been truly blessed with a great family and to them I am eternally grateful.

CHAPTER 1:

INTRODUCTION

1.1 Thymine Dimer Formation and Biological Effects

Extended exposure to UV light can lead to skin cancer. What is not widely understood is the molecular trigger leading to this disease. Irradiation of DNA with UV-A or UV-B light (260-320 nm) causes a [2 + 2] cycloaddition to take place between adjacent thymine in a strand of DNA leading to two major products; the cyclobutane pyrimidine dimer (CPD) and the (6-4) photoproduct (6-4PP).¹ These photoproducts can be seen in Figure 1.1. As the formation of thymine dimers is the likely first step in a cascade leading to skin cancer, the study of these photolesions is of high interest to scientists.

The formation of the two photolesions can be understood with a basic knowledge of the Woodward-Hoffmann rules and orbital symmetry.² Cycloadditions proceed through a HOMO-LUMO interaction between the π -systems on the adjacent bases. Diels-Alder [4 + 2] cycloadditions can proceed thermally due to orbital symmetry between the HOMO of the diene and the LUMO of the dienophile or vice versa (Figure 1.2). In Figure 1.3 a different scenario is observed for a [2 + 2] cycloaddition as is seen

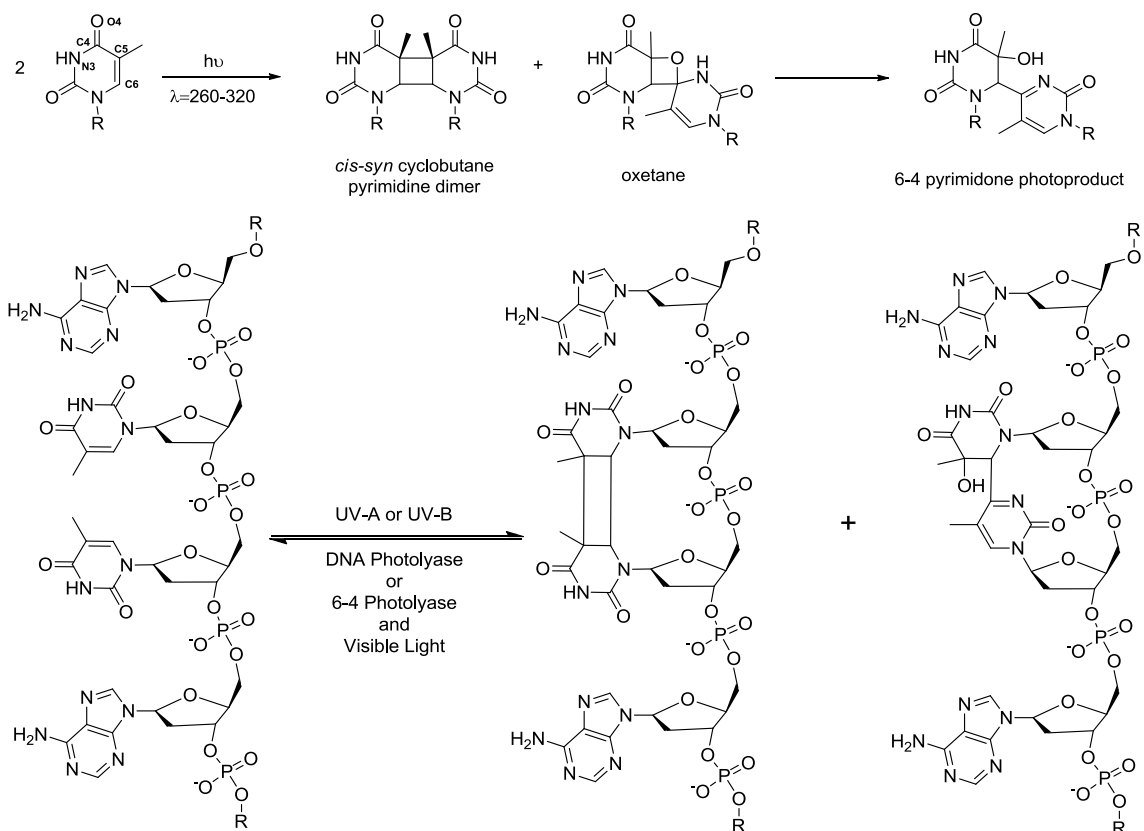


Figure 1.1- Formation of the two most common thymine dimers in DNA.

in the formation of thymine dimers. The HOMO and LUMO of two, two electron π systems are not symmetrical in the ground state, therefore the reaction does not proceed thermally. However, upon electronic excitation resulting in the promotion of an electron, orbital symmetry is achieved between the HOMO and LUMO of the two π systems allowing the reaction to proceed. This formation of thymine dimers is analogous to the formation of cyclobutane through the [2 + 2] cycloaddition of ethylene as has been shown in Figure 1.3.³

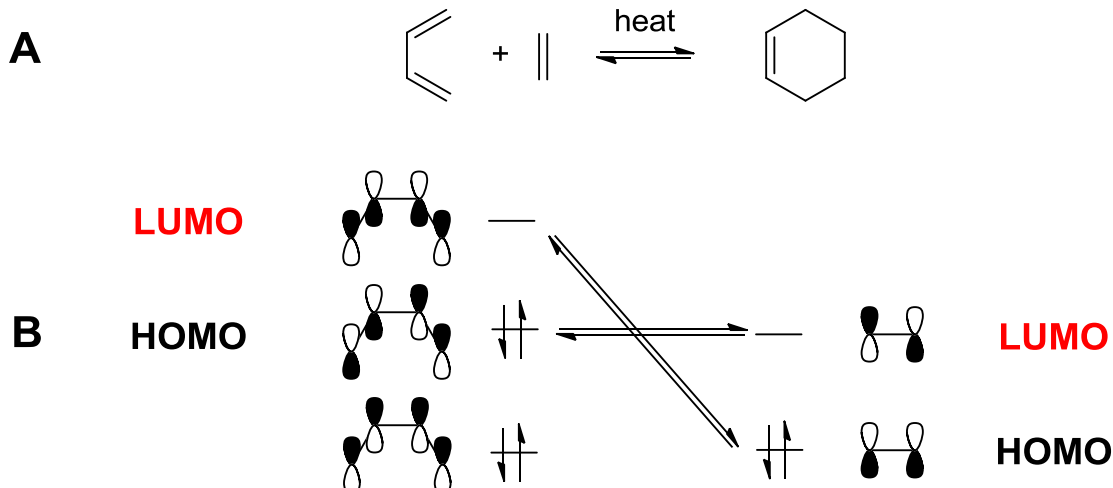


Figure 1.2- A) Scheme of the Diels-Alder reaction between 1,3-Butadiene and ethylene. B) Diagram of the electronic configuration and orbital symmetry present between the HOMO and LUMO orbitals in a $[4 + 2]$ cycloaddition.

As mentioned previously, two major products are formed upon irradiation of DNA with UV light. These originate from two separate $[2 + 2]$ cycloadditions. The most common, the CPD which accounts for approximately 70-80% of the yield, originates from the reaction between the C5-C6 atoms of both thymine bases.⁴ This results in the formation of a cyclobutane ring effectively linking the two bases together covalently. The less common (20-30% of the yield) 6-4PP is formed from the $[2 + 2]$ cycloaddition between C5-C6 of one thymine and C4-O4 of another thymine.⁵ This Paterno-Buchi reaction leads to an unstable oxetane species (Figure 1.1) which undergoes ring opening resulting in the 6-4PP. While this species is produced in lower quantities, its mutagenic effect may very well be greater than what is caused by the formation of CPDs.^{6,7}

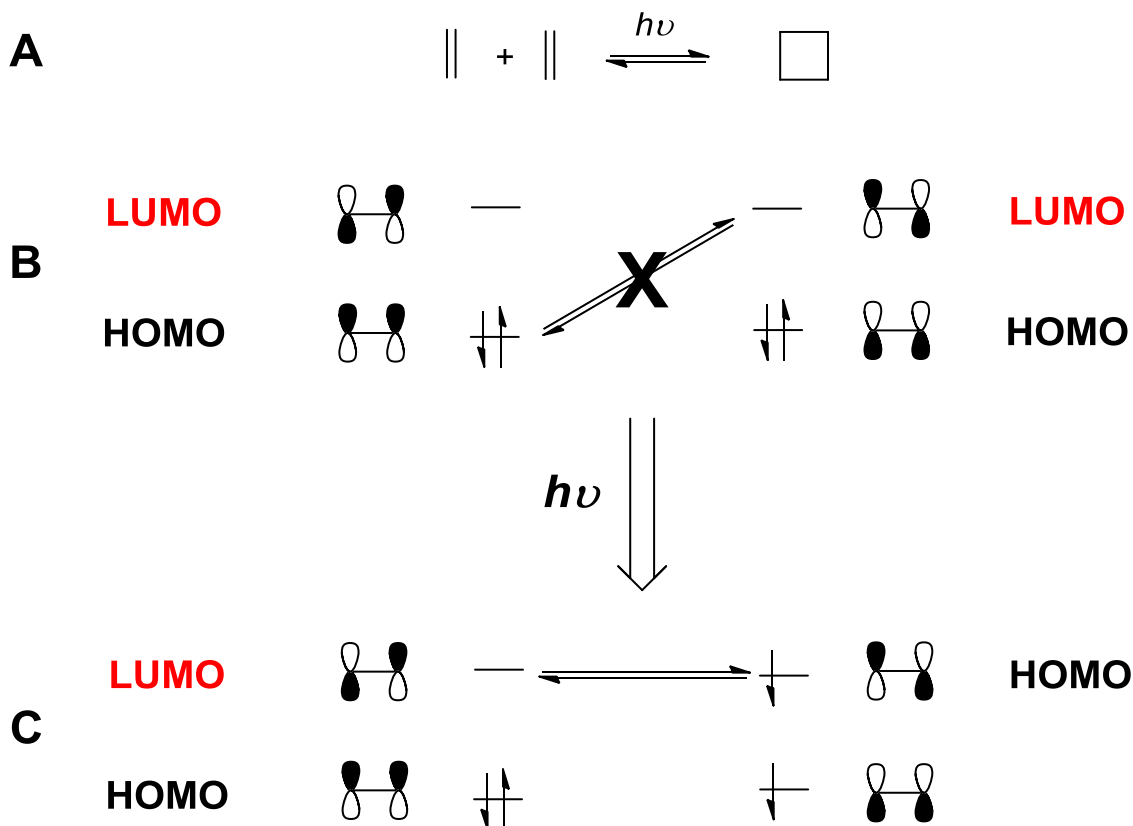


Figure 1.3- A) Scheme of the [2 + 2] cycloaddition between two ethylene molecules. B) Electronic configuration and orbital symmetry present in the ground state. C) Electronic and orbital symmetry present in the excited state.

Small changes in the structure of DNA can lead to problems in both DNA replication and transcription. CPDs and 6-4PPs in DNA have been shown to halt these molecular mechanisms and must first be repaired before the DNA can be read.⁸ In the repair of the CPD and 6-4PP humans rely on nucleotide excision repair (NER). While NER is not highly error prone, errors that do persist are passed through transcription and translation leading to changes in the proteome. Build up of thymine dimers cause an increase in the amount of single stranded DNA present in the cell which is caused by the stalling of DNA helicase at the damage sites.⁸ The importance of the NER mechanism of thymine

dimer repair is highlighted in the disease xeroderma pigmentosum. In this autosomal recessive genetic disorder, the NER mechanism is compromised. This leads to an inability to repair CPDs and 6-4PPs and to patients being extremely photosensitive, typically developing skin cancer by the age of 20-30.⁹ Due to these links between thymine dimers and skin cancer, the study of repair mechanism involved in repairing these photolesions is of great importance to scientists.

1.2 Thymine Dimer Repair Mechanisms

NER is the main defense for humans against the mutagenic effects of DNA photolesions. The basic mechanism of NER can be described in five steps.¹⁰ The first step in this process is the recognition of the damage site. From here, sequential action of DNA helicases and endonucleases open the double helix and cleave the DNA backbone 5' and 3' of the damage. The damaged segment is removed and a DNA polymerase enzyme fills in the gap left using the complementary strand as a template. Mutations of various proteins involved in the steps of this process manifest themselves as various diseases including xeroderma pigmentosum, Cockayne's syndrome and trichothiodystrophy. One common symptom of all three syndromes is photosensitivity which again displays the importance of NER in the repair of DNA photolesions.⁹

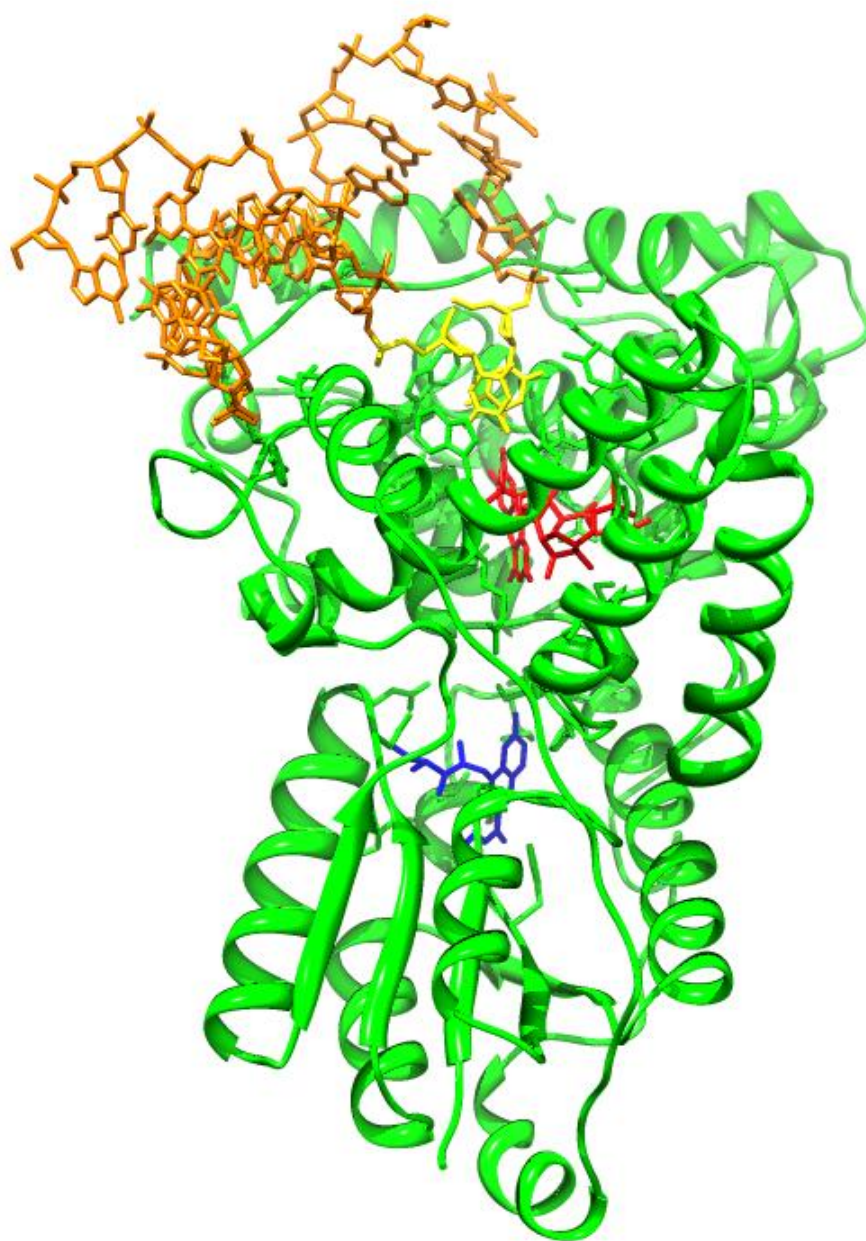


Figure 1.4- Crystal structure of DNA photolyase from *Anacystis nidulans*. Orange- Bound DNA. Yellow- Repaired thymines in the flipped-out conformation in the active site. Red- FADH⁻ cofactor. Blue- 8-HDF light harvesting cofactor.¹¹

Many organisms from the bacteria, eucarya, and archaea kingdom possess enzymes capable of repairing CPDs and 6-4PPs in a light dependant manner.¹² The first enzyme found to be capable of this activity is DNA photolyase (Figure 1.4). DNA photolyase is capable of binding to double stranded or single stranded DNA containing CPD photolesions. Recognition is thought to come from the significant kink of $\approx 15^\circ$ that is caused by the linkage of the two bases. The dissociation constants for CPD containing ssDNA and dsDNA are 10^{-8} and 10^{-9} M respectively.^{13,14} Two classes of DNA photolyase exist due to the use of two different light harvesting cofactors (LHC). The folate class DNA photolyase enzymes utilize 5,10-methenyltetrahydropteryl polyglutamate (5,10-MTHF) while the deazaflavin class uses 7-desmethyl-8-hydroxy-5-deazaflavin.¹² Both of these LHCs are capable of direct energy transfer though a Förster type mechanism to another chromophore in DNA photolyase, flavin adenine dinucleotide or FAD. In photolyases, this FAD is present in its two electron reduced form, FADH⁻. Upon electronic excitation, either directly or through energy transfer from the LHC, the FADH⁻ is able to donate an electron into the CPD effecting repair.¹² As the formation of CPDs are disallowed thermally due to orbital symmetry, so is the repair process. Upon the injection of an electron into the CPD, orbital symmetry is broken, driving the breaking of the C5-C5' and C6-C6' bonds of the cyclobutane ring. Following back electron transfer to FADH⁻, the catalytic cycle is complete and the two repaired thymine bases are reconstituted into the DNA. Figure 1.5 shows the intermediates along this repair pathway. In total, the repair of a CPD is ≈ 26.5 kcal/mol exothermic.¹⁵ This is to

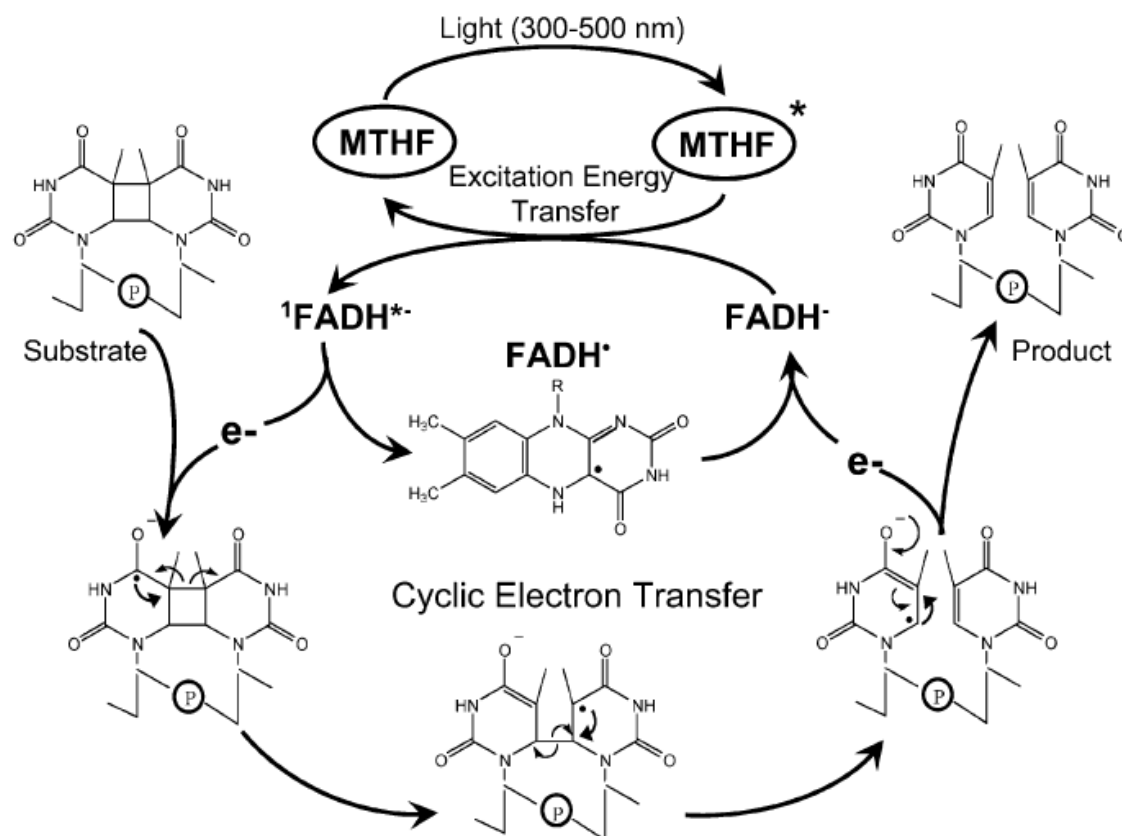


Figure 1.5- Catalytic cycle and intermediates involved in CPD repair by DNA photolyase.¹²

be expected due to the release of the ring strain present in the cyclobutane as well as the reestablishment of the aromaticity of the thymine nucleobases.

As experimental work has given us much information on the possible mechanism of CPD repair, computational work has given us atomistic detail into the intricacies of the repair mechanism employed by DNA photolyase.¹⁶ While many studies have been published on this mechanism, I will only present two as the mechanism has been settled upon. In an article published in 2000, Durbeej and Eriksson used high level density functional theory calculations to calculate the energies of the intermediates along the repair pathway of the CPD.¹⁷ Upon electron uptake by the CPD, a lengthening of the C5-

C5' bond from 1.59 Å to 2.76 Å is observed. This indicates that the cleavage of the C5-C5' bond is spontaneous on the radical anion potential energy surface. From here, the C6-C6' bond breaks with an activation energy of 2.3 kcal/mol with the T + T⁻ product lying 2.4 kcal/mol lower in energy than the ring opened species following C5-C5' fragmentation. Back electron transfer to the flavin cofactor closes the catalytic cycle resulting in two repaired thymine bases.

Another interesting study by Masson et al. uses Carr-Parrinello QM/MM techniques to follow the repair reaction over time.¹⁸ This technique allows the electronic structure of a system to evolve over time. In this study the enzyme is not included, but rather the self repair process is examined. An electron is injected into a CPD in dsDNA and the C5-C5' and C6-C6' distances are monitored along with the electron spin on each atom. C-C distances vs. time from this work are shown in Figure 1.6. This study displays the spontaneous C5-C5' bond breaking, along with the small activation energy required to break the C6-C6' bond. They estimate an upper-bound of 2.5 kcal/mol for an activation energy of C6-C6' bond cleavage, which is in close agreement with the study by Durbeej and Eriksson. An in depth understanding of how the DNA photolyase enzyme effects repair of CPDs is important if we are to mimic this function. The artificial repair of CPD containing DNA will be reported in Chapter 4.

While the repair mechanism employed by DNA photolyase is well understood, the mechanism of 6-4PP repair is hotly debated. The 6-4PP is not repaired by DNA photolyase, but rather is repaired by a separate photoreactivation enzyme, (6-4)

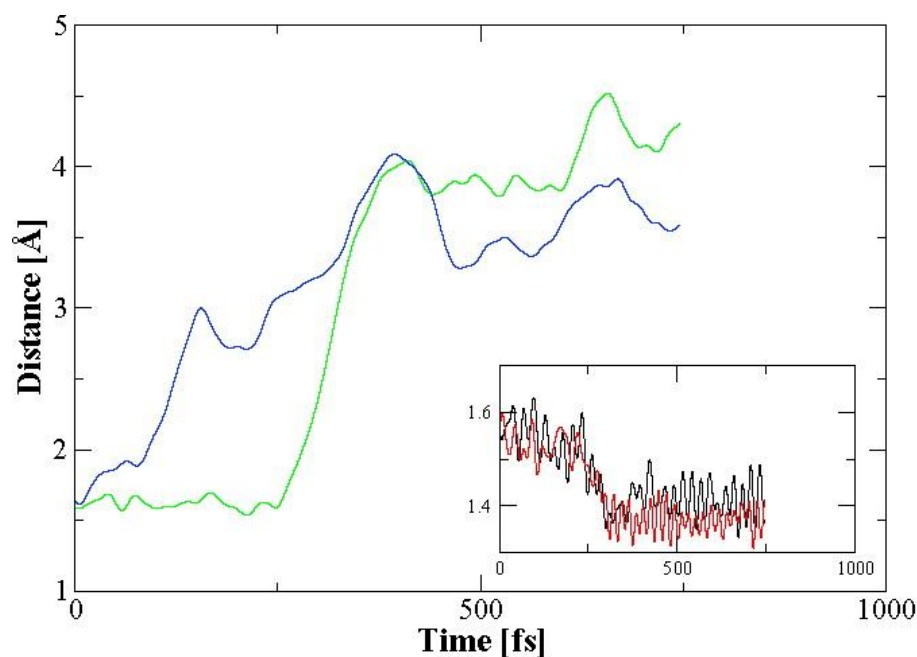


Figure 1.6- Plot of C5-C5' (blue) and C6-C6' (green) distances versus time. Inset depicts the formation of the double bonds between C5-C6 (black) and C5'-C6' (red).¹⁸

photolyase.¹² The overall catalytic scheme employed by (6-4) photolyase is nearly identical to that of DNA photolyase. In short, (6-4) photolyase enzymes use the same LHCs and reduced deprotonated flavin cofactors in the repair mechanism. However, the transfer of the O4' group from one base to the other in the formation of the 6-4PP undoubtedly complicates the repair mechanism. Early proposed mechanisms focused on the closing of O4' onto the C4' carbon forming the same oxetane intermediate formed in the [2 + 2] cycloaddition which leads to 6-4PP (Figure 1.7).¹⁹ From here, it is proposed that electron transfer to the oxetane intermediate leads to repair of the thymine dimer in analogy to the repair mechanism of DNA photolyase. While this mechanism explains the bond breaking and bond making needed to arrive at the correct products, it has been

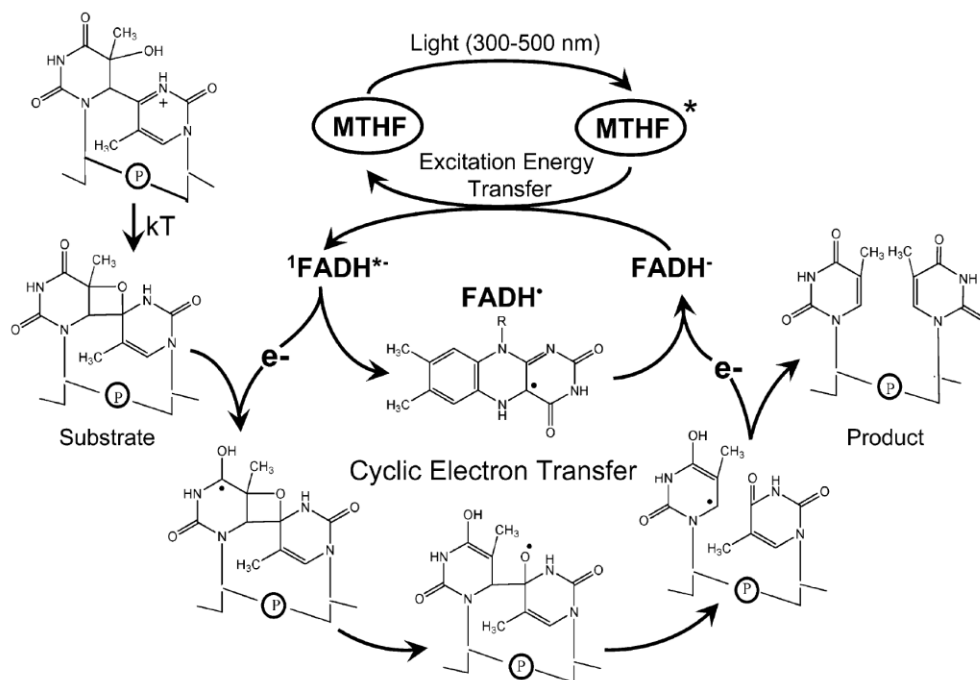


Figure 1.7- Early mechanism proposed by Sancar for the repair of 6-4PP by (6-4) photolyase.¹²

argued on the basis of energetics that it may not be a feasible mechanism. This mechanism of 6-4PP repair is discussed in detail in Chapter 3 along with three newly proposed mechanisms.

1.3 Unnatural Nucleic Acid Analogs

Scientists have been changing the backbone and nucleobases of nucleic acids either to alter their properties or to investigate the stability of DNA and RNA. With the successful synthesis of several different nucleic acid backbone analogs, it is interesting to think about why nature has chosen ribonucleic acids as the carrier of the information of life. Backbone analogs of note (that form stable duplexes) are peptide nucleic acids (PNA), threose nucleic acids (TNA), locked nucleic acids (LNA) and glycol nucleic acids

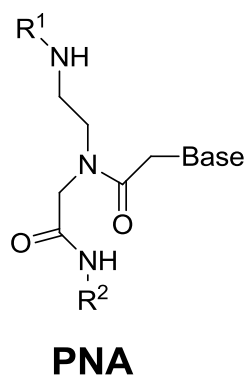


Figure 1.8- Structure of the PNA backbone.

(GNA). Each of these use different modifications to study the importance of concepts such as electrostatic repulsion between nucleic acid strands in the duplex form and entropic effects of annealing.

PNA was designed for use in antisense technology.²⁰ The structure of PNA can be seen in Figure 1.8. The peptide backbone of PNA is much different than that of DNA or RNA. Its backbone is acyclic and is uncharged at physiological pH. The idea behind this scaffold is the reduction of the charge on the backbone which prevents natural nucleic acids from easily passing the cell membrane. The phosphate backbone of naturally occurring nucleic acids is highly negatively charged due to the acidity of the proton on the phosphoric acid functional group. Since both strands contain the same negative charge, a large electrostatic repulsion is present between the two strands. When ssDNA is annealed with a complementary strand of PNA, an increase in melting temperature is observed when compared to the dsDNA counterpart.²⁰ This indicates that the PNA-DNA duplex is more stable than the DNA-DNA duplex and displays the advantage the reduction of electrostatic repulsion can have in antisense technology.

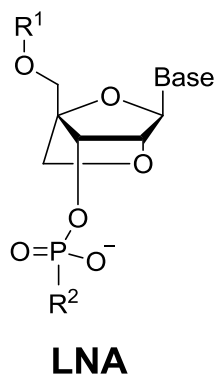


Figure 1.9- Backbone structure of LNA.²¹

LNA is a nucleic acid analog of RNA containing a bridge from the 2' oxygen to the 4' carbon (Figure 1.9).²²⁻²⁴ This effectively locks the ribose ring in the 3'-*endo* conformation. This restriction of the conformational flexibility in the nucleic acid backbone has the effect of increasing the preorganization of the single strand. This preorganization prior to annealing reduces the entropic penalty associated with annealing, effectively increasing the melting temperature in comparison to the corresponding DNA and RNA sequences. For example, a 13-mer LNA-DNA hybrid has been shown to have a $T_m=83^{\circ}\text{C}$, an increase of 36°C when compared to the corresponding DNA duplex.²³ LNA is a nice example of the effect of entropy on the formation of nucleic acid duplexes.

Threose nucleic acids (Figure 1.10) attempts to push the limits of the number of “bonds-per-backbone” required to form a duplex.²⁵⁻³¹ The Eschenmoser laboratory synthesized TNA and measured melting temperatures and thermodynamic parameters of ssTNA annealing. DNA and RNA have 6 bonds along the backbone (O5' to O5'). TNA, on the other hand, possesses only 5 bonds per backbone demonstrating that having 5 bonds per repeating unit along the length of the duplex is not a necessary requirement to form a stable duplex. Surprisingly, TNA is able to form more stable duplexes than both

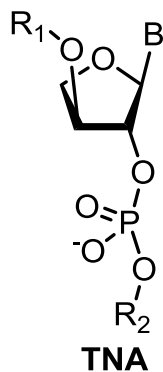


Figure 1.10- Structure of threose nucleic acid.

DNA and RNA.³² The Meggers group envisioned the removal of the 4'CH₂O group of TNA to form a further simplified nucleic acid consisting of repeating glycol units connected through phosphodiester bonds.³³⁻⁴⁰ This is shown in Figure 1.11. GNA is a relatively new, nonnatural nucleic acid and has only been studied in depth by the Meggers group over the past 5 years. GNA displays several interesting properties. Perhaps most striking is the fact that GNA duplex formation is entropically less penalizing than the corresponding process in DNA.³⁶ This idea is somewhat counterintuitive considering the backbone of GNA is acyclic and one would expect it to be highly flexible. Using the available crystal structures and molecular dynamics simulations we shed some light on the molecular recognition process leading to this counterintuitive observation in Chapter 5.

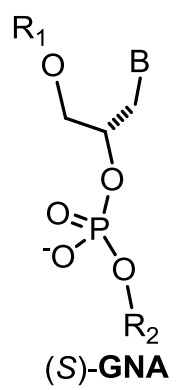


Figure 1.11- Structure of glycol nucleic acid.

CHAPTER 2:
STRUCTURE AND DYNAMICS OF POLY(T) SINGLE STRAND DNA:
IMPLICATIONS FOR CPD FORMATION

2.1 Introduction

As it is thought that CPD formation is the first step in a cascade eventually leading to skin cancer, it is of great interest to gain a deep understanding of how they form. The basis of CPD formation can be gleaned from early observations made by Woodward and Hoffmann.² Using orbital symmetry arguments it can easily be shown that [2 + 2] cycloadditions are symmetry forbidden in the ground state and must proceed via the excited state (Figure 2.1). In pyrimidines, upon absorption of a photon of UV-A or UV-B light (260- 320 nm) a [2 + 2] cycloaddition can occur.¹⁶

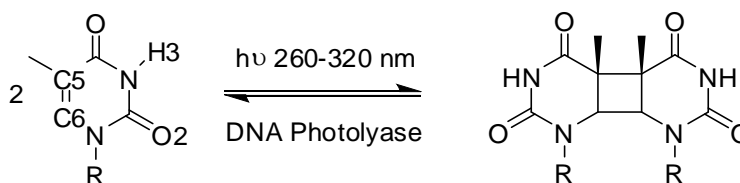


Figure 2.1- Formation of the CPD.

The scaffold of DNA provides an interesting framework for such reactions to take place. Due to geometrical constraints these reactions only take place between adjacent

pyrimidines along a single strand. In other words, only bases stacking with each other can react. Furthermore, a study published by Kohler and coworkers in 2007 demonstrated that this reaction proceeds in an ultrafast manner.⁴¹ By using femtosecond time-resolved IR spectroscopy it can be shown that CPDs in an 18-mer all thymine single strands are formed in under 1 ps of irradiation at 272 nm. The ultrafast nature of CPD formation indicates two things. First, the kinetics of the cycloaddition point towards a very small barrier for reaction. This has also been demonstrated computationally.⁴² Secondly, the reaction proceeds faster than conformational realignments. In a previous MD study it was shown that base stacking/ unstacking events proceed on a time scale of 15-20 ps.⁴³ It therefore follows that the base stack arrangement present at the instant of excitation is the determining factor in whether or not a given pyrimidine base stack dimerizes or not. Considering that only a 2-3% yield was reported for the femtosecond time resolved IR spectroscopy study, it was concluded that only a small percentage of thymine base stacks are aligned properly for reaction at the moment of excitation.

The low yields observed in the femtosecond excitation study could arise from a few different scenarios. One being that the base stacks interact in a dynamic manner with many stacking/ unstacking events leading to a statistical distribution of both reactive and unreactive conformers that result in the observed yield. Another possible explanation is that the single strand DNA used in the experiment is capable of folding in such a manner which cases some base stacks to become “locked” in a unreactive conformation while other base stacks are positioned in such a way which allows the reaction to proceed. In short, the secondary structure formed in the ssDNA strand may affect dimerization yields. A third, and less likely possibility is that at the moment of excitation there are a large

number of reactive conformers and the low yields result from a partitioning of reactive pathways and vibronic relaxation to the ground state. Since most MD force fields only describe molecular systems in the ground state, we can only distinguish between the first two possibilities using force field techniques. Because the speed of the reaction is much faster than the conformational rearrangements of the base stacks, it is reasonable to assume that base stacks that are properly aligned at the moment of electronic excitation will proceed to react and form the CPD. This allows for the use MD simulation of a sufficient timescale to be used to collect statistics to determine which base stacks are aligned for reaction and also to predict damage hot spots within DNA strands.

Several studies have been published on the dynamics of ssDNA and the agreement between experiment and theory looks promising. In 2005, Biyani and Nishigaki published a study comparing gel electrophoresis motilities to end to end distances calculated from MD simulation.⁴⁴ They are able to correlate the overall globular structure in solution to both theory and experiment. A correlation coefficient of $R^2=0.93$ was obtained between end to end distances and gel electrophoresis mobilities. Perhaps even more important for this work, they observed stable RMSDs of the MD trajectories within 5ns. This allows us to perform an equilibration within a reasonable amount of time and collect production statistics beyond the equilibration. Knowing this, we can expect to simulate the poly(T) ssDNA studied by Kohler and coworkers from an extended coil and form a globular structure (equilibration) within a few nanoseconds. From here we can collect statistics on the globular structure instead of the unnatural coiled or extended structure.

Computational research into structure and dynamics of ssDNA has also been performed. A 2001 study by Martinez, Elmroth, and Kloo used MD simulation to study the effects of counterions on the solution structure of ssDNA.⁴³ In analogy to Kohlers work, one sequence studied was a poly(T) single strand. In this simulation, a globular folded structure was obtained within 3.5 ns. Also, interesting intrastrand hydrogen bonding patterns were observed between distant (in terms of primary sequence) bases. However, their simulations were not on a sufficient time scale to determine if the hydrogen bonds formed are strong enough to hold the ssDNA in this globular conformation. The ability to predict the folded structure of poly(T) ssDNA is important to our study because the process of folding is likely to greatly perturb canonical base stacking interactions. We hypothesize that this perturbation will likely result in an unequal probability for CPD formation between base stacks. For example, some bases may be confined to a small conformational area due to a duplex like hydrogen bonding arrangement. Others may assume solvent exposed positions with no adjacent base to interact with. Another possibility is that distant nucleobases can intercalate between other base stacks, effectively pushing them apart and decreasing the probability of CPD formation between the two bases. While it is possible to speculate on the effect of such structural features on CPD yields, no study had looked at a particular sequence of DNA and studied how secondary structure can affect base stack alignment. This was the main purpose of this study.

In the presented work molecular dynamics simulations are used to probe base stacking interactions in the context of CPD formation. To do this, long time scale (50 ns) simulations are performed to allow the ssDNA strand to assume a globular structure and

base stacking interactions to be elucidated. These long time scale simulations afford a large sampling of the conformational space of the ssDNA strand. In the literature, very few studies have utilized MD simulations of this time scale to study ssDNA. This is an interesting observation because the importance of conformational sampling for the correct description of large biopolymers is widely accepted. If it indeed is true that there is a “strong link between conformation before light absorption and photodamage”⁴¹ then the results from MD simulation can be connected with the previously determined structural requirements that need to be achieved in order for a CPD to form in a [2 +2] cycloaddition. The first structural requirement for CPD formation is that the two bases must be near in space. Previous crystallographic studies can be used to determine the ground state distances needed to achieve the cycloaddition prior to irradiation (≈ 3.5 - 4.2 Å).^{45,46} By using these distances it will then be possible to monitor which base stacks are in close enough proximity to react at any given snapshot along the MD trajectory. The second parameter that must also be monitored is the alignment of the two bases of the base stack. Previous computational work has laid the groundwork for determining the correct dihedral angle that must be obtained between the reacting atoms for the reaction to proceed.⁴² For the [2 + 2] cycloaddition to take place between two thymines, the C5-C6-C6'-C5' dihedral angle must be properly aligned prior to excitation ($\approx 27^\circ$). Obviously, the final requirement is that a photon must be absorbed while the bases are properly aligned for the CPD to be formed. To date, there is still much debate as to whether this reaction proceeds through a singlet excitation, triplet excitation, or involves a triplet- singlet crossing along the PES.^{42,47} However, in this study since we are only

concerned with ground state structures, the electronic state of the reaction is of little importance. The coordinates important to this study are shown in Figure 2.2.

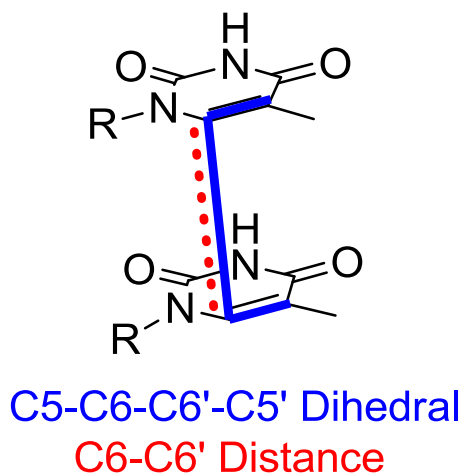


Figure 2.2- Coordinates of interest to CPD formation.

In this work we use a novel approach to study how secondary structure of ssDNA can affect CPD yields. Using MD each base stack (17 total) is analyzed by plotting radial distribution functions of base to base distances as well as histograms of the dihedral formed between reacting atoms (C5-C6-C6'-C5'). Through examination of this data it is possible to predict not only the overall structure of the folded poly-T ssDNA studied by Kohler, but also shed light upon the effects of secondary structure on base stack alignment and CPD formation. While several previous studies have looked at the folding and structure of ssDNA, this study is unique in that it is the first investigation using statistical methodology to elucidate the probabilities of CPD formation. The overall goal of this work is to study the effects of secondary structure around T-T steps in ssDNA on thymine base stack alignment. These alignment parameters are then related to probabilities of CPD formation.⁴⁸

2.2 Methods

All structure preparation, MD simulations and post-processing of trajectory data was performed with the AMBER 9 suite of programs.⁴⁹ *Nucgen* was used to prepare a canonical poly(T)- poly(A) B-DNA duplex. The single strand of poly(T) DNA was created by simply deleting the poly(A) strand from the pdb file. The resulting pdb file was then subsequently loaded into the *xleap* program where the system was then neutralized with Na⁺ counterions. TIP3P explicit waters were used to create a solvent box extending 8.0 Å beyond the solute. The final system contains 12967 atoms including 4125 TIP3P waters and 17 Na⁺ counterions.

Upon completion of the initial system setup, minimization was performed in two steps. First, the ssDNA was restrained while letting the solvent box and Na⁺ ions relax. Secondly, the entire system was allowed to relax into a local minimum. The newly (at the time) implemented Particle Mesh Ewald Molecular Dynamics (PMEMD) module of AMBER 9 was used to perform all MD simulation. PMEMD is basically a rewritten version of *sander* which has been shown to significantly increase the parallel performance of particle mesh ewald simulation. This allows us to use many processors (32 in this case) with essentially linear scaling, meaning a calculation on 32 processors gives the results approximately 32 times faster. Furthermore, results obtained using PMEMD and *sander* are the same within rounding errors. The force field chosen for this work was the Cornell *et al.* force field along with the adjustments by Wang *et al.* (parm99).^{50,51} Heating and equilibration of the system was performed using a constant volume (NVT) ensemble with 10 kcal/mol·Å restraint added to the solute. The system was heated from 0- 300°K over 20 ps. Next, the restraints were removed from the solute

and further equilibration and production runs were performed (50 ns total) in a constant pressure (NPT) ensemble. 50 ns trajectories are necessary to sufficiently sample the conformational space of the ssDNA which is more flexible than its dsDNA counterpart. The long simulation times allow the folding of the ssDNA into a globular structure where interactions with non-adjacent bases can cause large disturbances in individual base stack interactions. It should be noted that the sampling times in ssDNA simulations in previous studies in the literature have been much shorter and this study represents perhaps the longest time scale simulation of ssDNA. The effects of these disturbances can then be analyzed in their relation to CPD yields. One advantage in simulating ssDNA is that it is not subject to fraying in the same way dsDNA is when running multi-nanosecond simulations. Even with recent advances in force field techniques, fraying of duplex DNA is still seen as a problem.⁵² In our case since we are dealing with ssDNA, it is not necessary to apply any kind of artificial restraint to the system to prevent such behavior.

In order to draw conclusions from the simulations, post-processing functions of the AMBER program were used. In particular, the *ptraj* module of AMBER was used to extract distance and dihedral information. Microsoft Excel was used in creating and weighing both radial distribution functions of inter-base distances as well as plotting histograms of the dihedral formed between the reacting atoms of the [2 + 2] cycloaddition. The visualization program VMD was used to view trajectories and create figures depicting DNA conformations from the simulations.⁵³

2.3 Results and Discussion

2.3.1 Base Stack Interactions: Distances

As discussed in the introduction, the distance between the reacting atoms of adjacent bases is one of the coordinates that must be considered when determining if a reactive alignment of bases has been formed or not. It is thought that if a sufficiently small distance is obtained between the reacting atoms (C5-C5' and C6-C6') prior to irradiation that a CPD can be formed at that instant. To determine the maximum distance required for a [2 + 2] cycloaddition to proceed, we must look to experiment. Studies of olefin stacking in the solid phase prior to excitation and subsequent [2 + 2] cycloadditions are available. Ramamurthy and Venkatesan published a review that suggests that in most cases olefins must stack between 3.5 and 4.2 Å prior to irradiation for the reaction to proceed.⁵⁴ While these experiments take place in the solid phase, they are still useful in helping to describe CPD formation which takes place in preorganized systems in solution. This allows us to calculate a distance in which 3% (percent yield reported in Kohler's work⁴¹) of the population is under, we can estimate how often each base stack is held in a reactive conformation and furthermore predict which base stacks are most susceptible to CPD formation.

In order to compare the base stack distances found in simulations to experiment, the C5-C5' and C6-C6' are analyzed. Using the *ptraj* module of AMBER, radial distribution functions of the inter-base distances were created and imported into Microsoft Excel. From here, total probabilities are normalized to one and the distance of 3% additive probability is calculated for each base stack. The radial distributions can be found in Figure 2.3. It is readily apparent that a wide distribution of base stack distances

are present in the system. In base stacks 1-2 to 3-4 (Figure 2.3 A) a narrow distribution around 4 Å can be observed. Base stack 4-5 displays broadening towards longer distances up to 6 Å, which corresponds to a lower probability of obtaining close proximity base stacking. The reason behind this broadening becomes clearer by looking at the distribution of base stack 5-6 (Figure 2.3 B). In this base stack, there is a large distribution of distanced ranging between 5 and 7 Å. This indicates that base 5 is in an isolated position with its base stacking with both base 4 and 7 being disturbed. Looking through the structural data it can be seen that base 5 lies on a turn in the ssDNA structure. This results in base 5 being isolated and positioned in a solvent exposed manner (Figure 2.6).

A similar situation can be observed for base stack 10-11. In this case two peaks are observed centered about 5 and 7 Å (Figure 2.3 C). Analysis of the structural data shows that this distribution is caused by an intercalation of base 18 between bases 10 and 11, effectively forcing the two apart. Base stacks 13-14 and 14-15 also display shifted population distributions (Figure 2.3 D). Analysis of the structural data indicates that this deviation is caused again by its positioning on a hairpin in the ssDNA structure. The folding at this point of the ssDNA structure results in base 14 to be forced out of the duplex- like arrangement that is formed (Figures 2.4 and 2.5). The secondary structure formed in this poly(T) ssDNA sequence are comparable to what has been observed in MD studies of ssDNA by others.

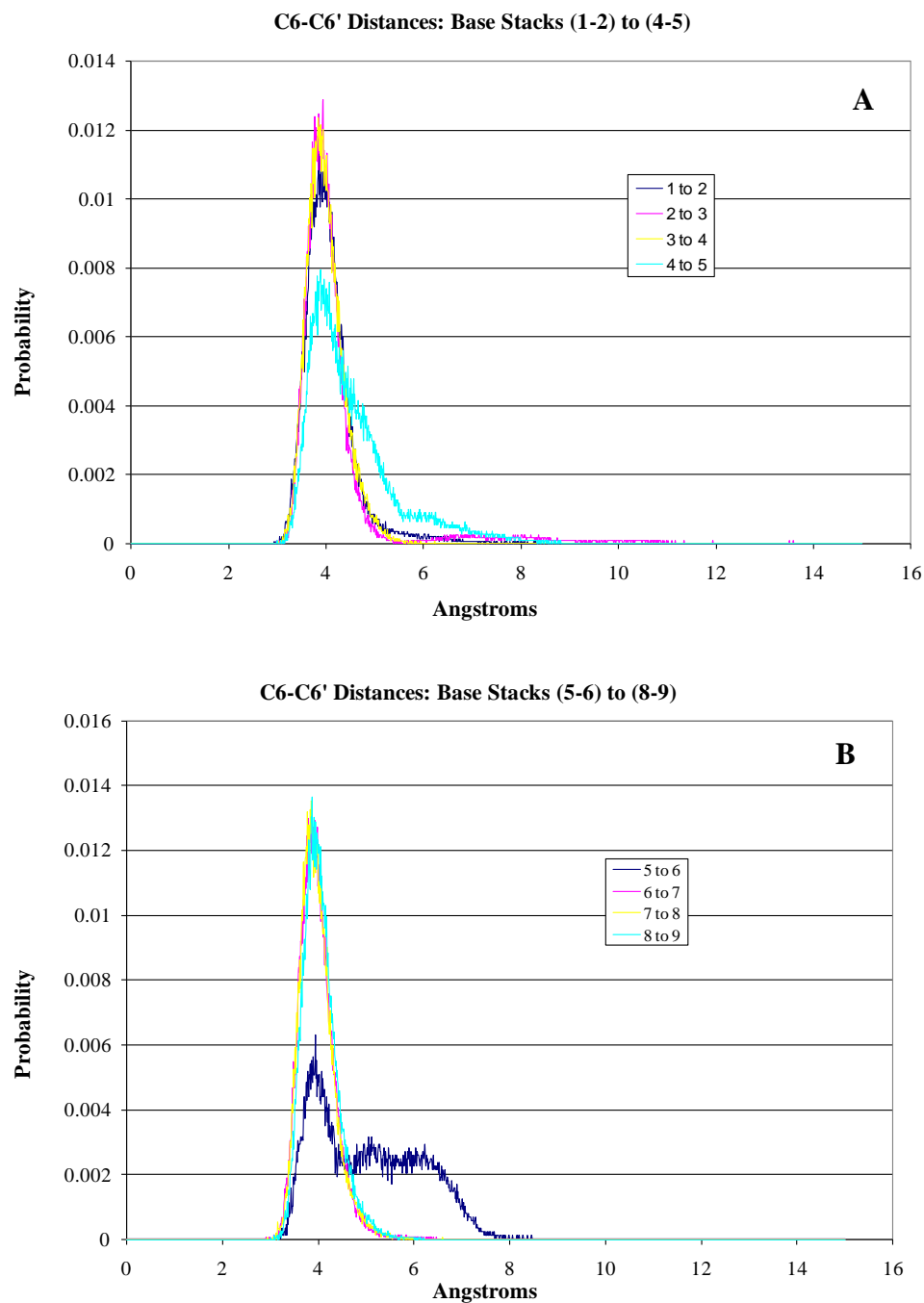


Figure 2.3- Radial distribution functions for C6-C6' (A-D) and C5-C5' (E-H) distances from the simulation. Plots are colored by which base stack is represented (inset). Figure spans 3 pages.

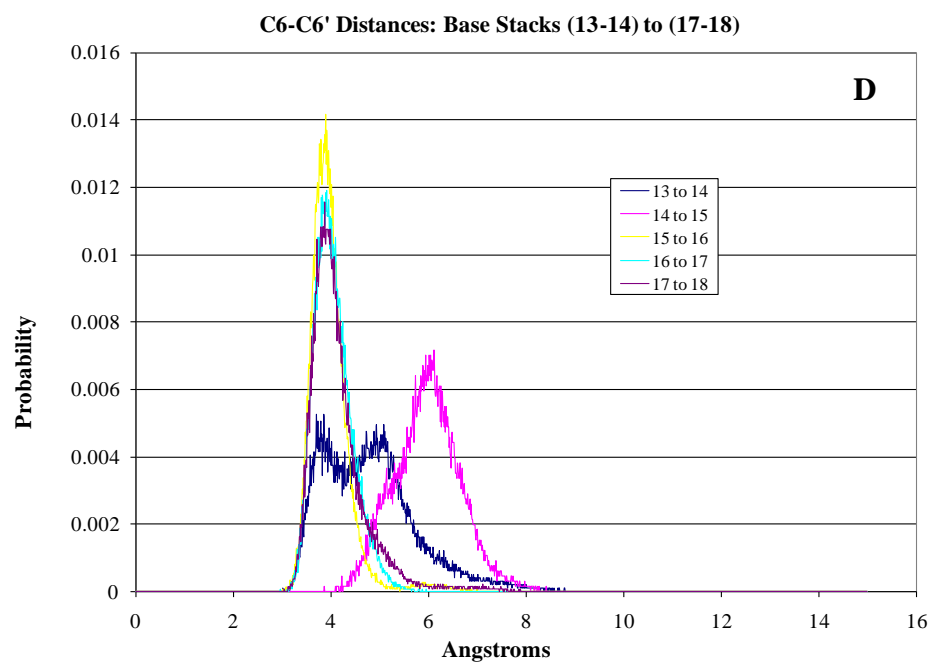
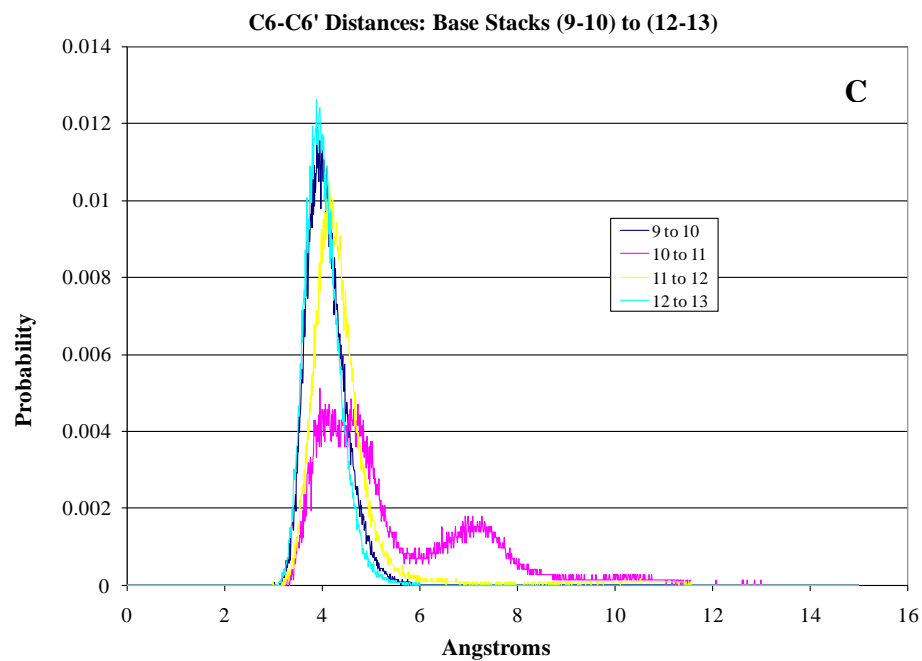


Figure 2.3- Continued.

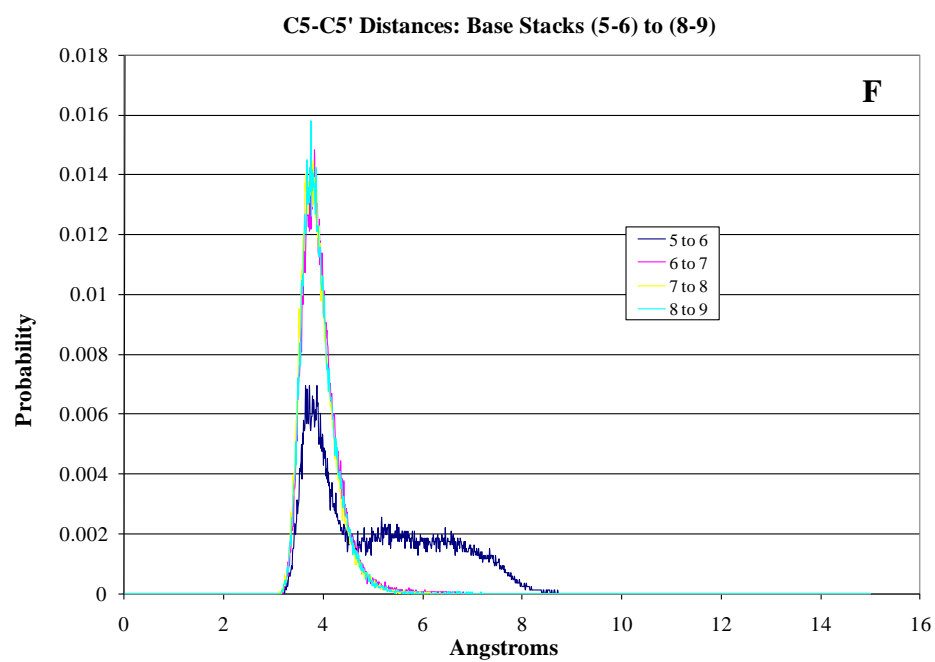
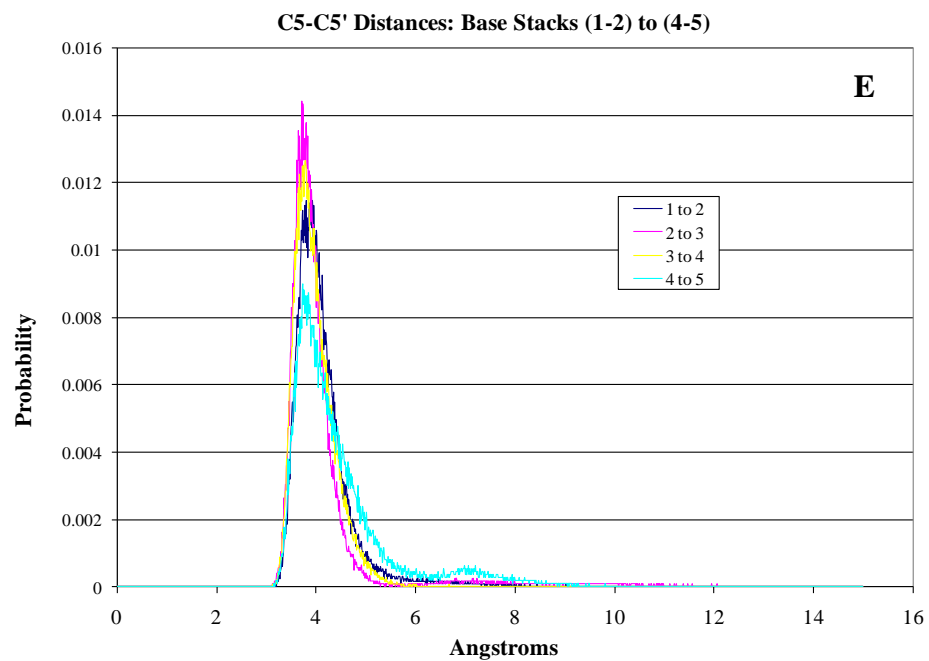


Figure 2.3- Continued.

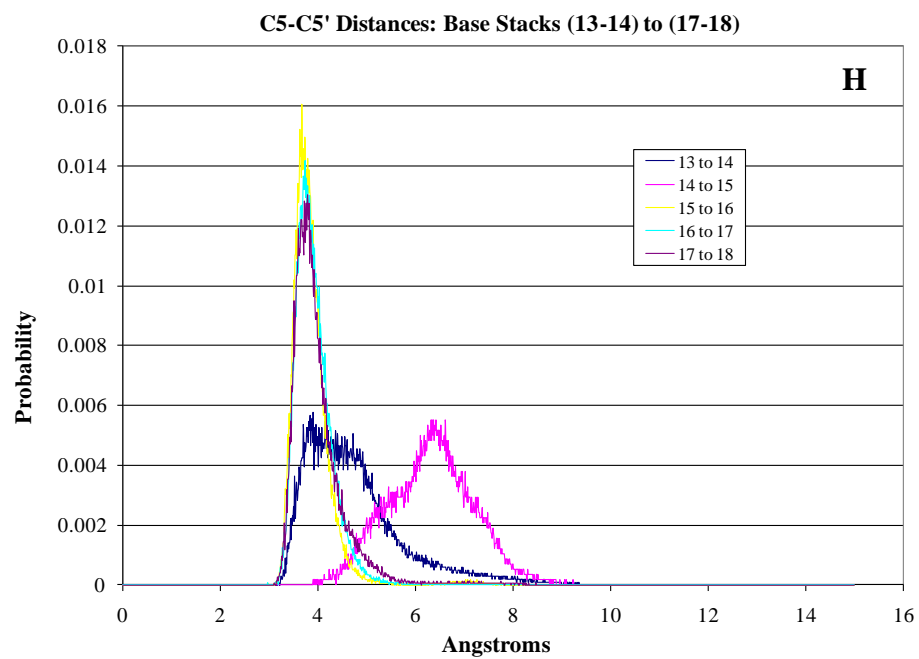
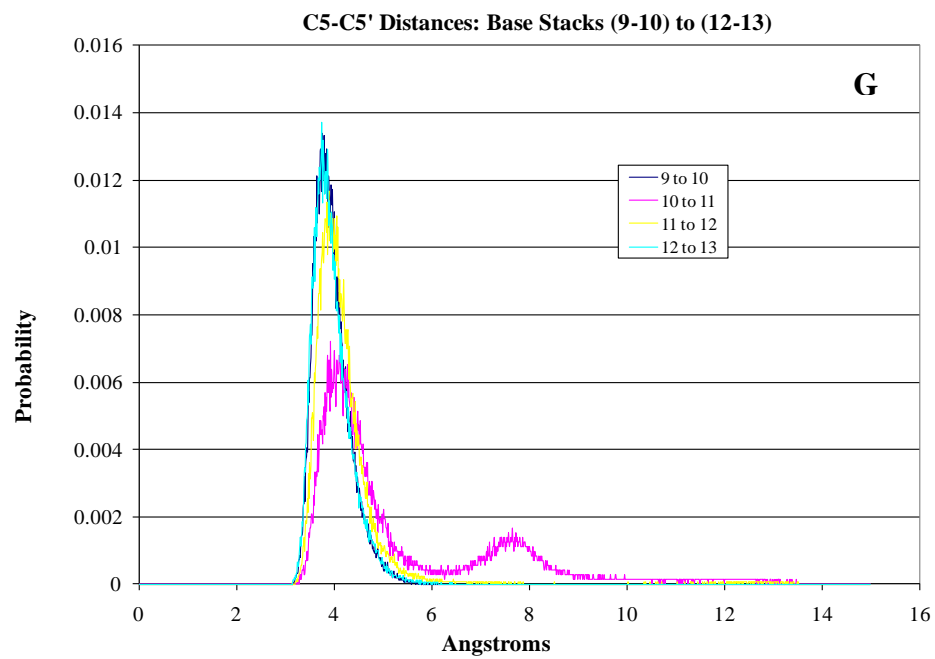


Figure 2.3- Continued.

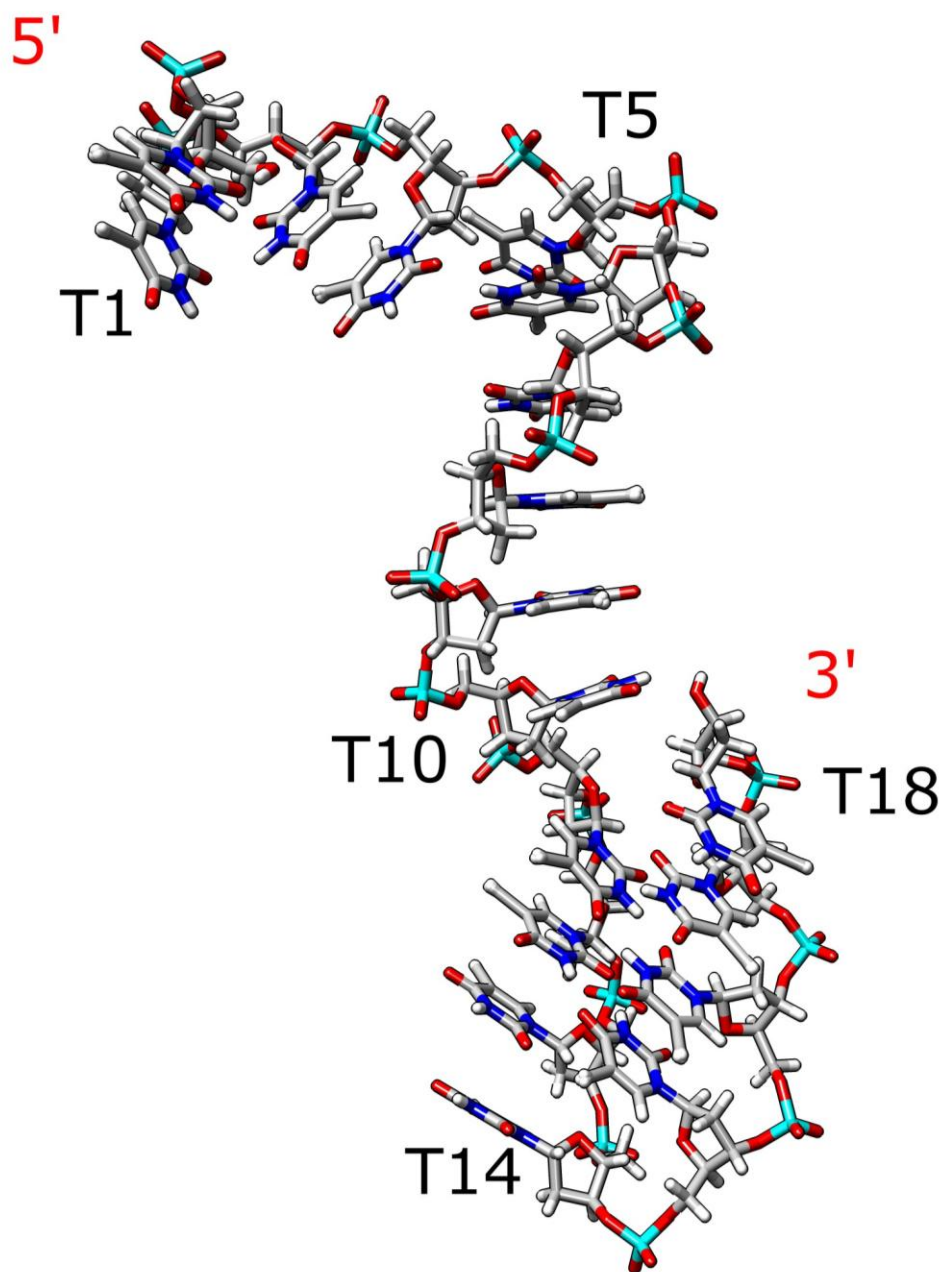


Figure 2.4- Overall view of the globular structure fully formed at the end of the 50 ns MD simulation.

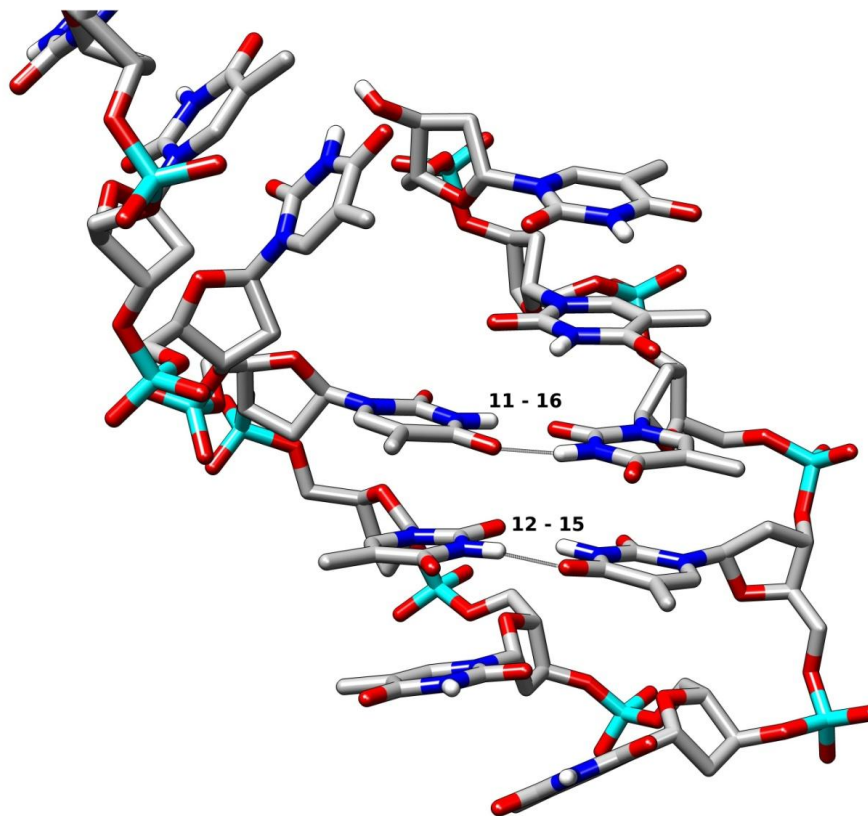


Figure 2.5- Close up view of hairpin region of the poly(T) ssDNA containing duplex like hydrogen bonding arrangement.

The plots of the normalized radial distribution show the data for the probabilities of a particular base stack to be at any given distance. By summing these probabilities we can obtain a distance in which 3% of the population has a lower C6-C6' (or C5-C5') distance. Through use of these additive probabilities we can compare the π - π stacking in the ground state of our ssDNA base stacks to the experimental findings by Kohler. The link between the two works being the crystallographic study of olefin stacking prior to [2 + 2] cycloadditions which indicate that distances in the 3.5- 4.2 Å range are necessary for the reaction to proceed.⁵⁴ The plot in Figure 2.6 shows the distance of 3% additive probability for each base stack in the poly(T) single strand. It

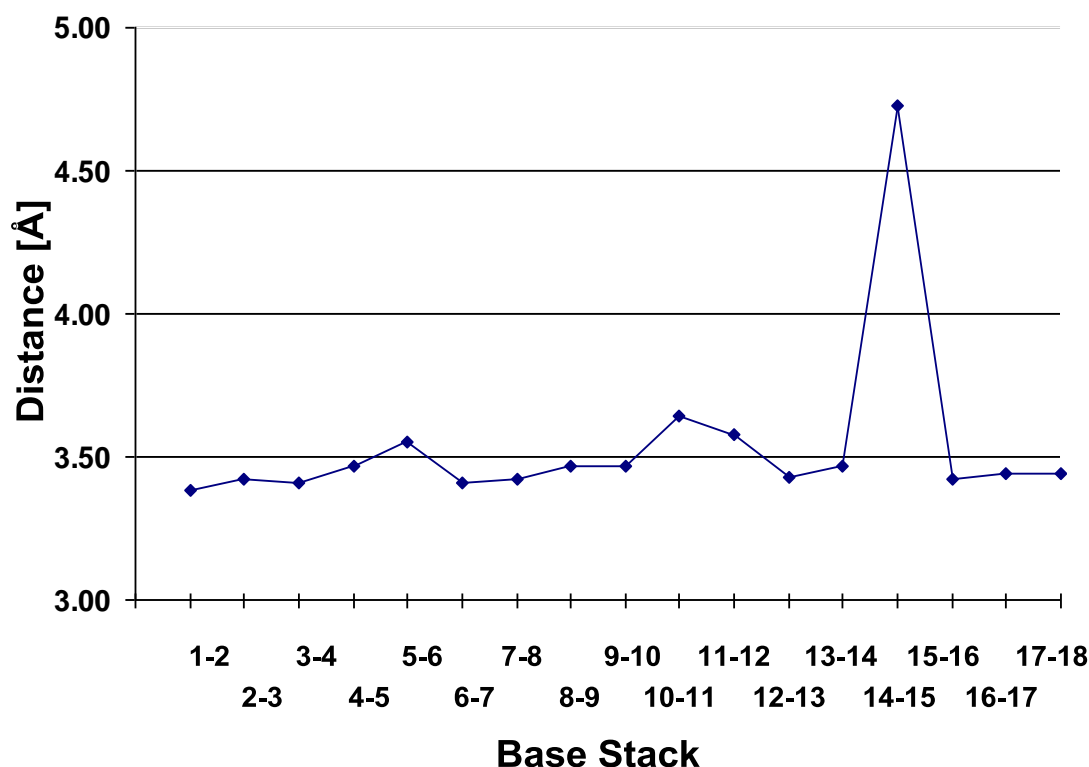


Figure 2.6- Distance of 3% additive probabilities for each base stack.

should be mentioned that this value is approximate and does not account for a situation where bases are in close proximity prior to irradiation but fail to react due to vibronic relaxation back to the ground state. However, it still holds that if a high population is stacked at a close distance, there will be more opportunities for the reaction to proceed, indicating that the particular base stack may be a hotspot for CPD formation.

Figure 2.6 indicates that base stack 14-15 has a much higher distance of 3% additive probability than the others. With its distance of 3% additive probability lying at 4.73 Å, it is unlikely that this base stack will form a CPD upon irradiation. Most importantly, this result indicates that the probability of CPD formation is not equally

distributed throughout the DNA strand, but rather some bases are likely held in more reactive conformations than others. For example, base stacks 3-4 (3.41 Å), 6-7 (3.41 Å), 12-13, (3.43 Å), and 15-16 (3.42 Å) are held much closer together than base stack 14-15 and likely are hotspots which are more susceptible to photodamage. Another feature of the closely stacked bases is that they all lie adjacent to a solvent exposed base and near a bend or hairpin on the folded poly(T) ssDNA structure. The most likely explanation for these phenomena is that with losing a base stacking partner the remaining bases stack within a closer proximity. By removing a base from the stacking interaction, there is less competition leading to a closer interaction with the remaining base. In the analysis of the plots in Figures 2.3 and 2.6, it appears that bases lying directly on the turn in the ssDNA are protected from reaction while those near the turn, who have lost a base stacking partner, are held more closely together likely leading to increased incidences of photodimerization.

Through analysis of the structural data from the MD simulation we can glean detailed information on the origins of these effects. Figure 2.4 shows that bases 14 and 15 form a hairpin on the poly(T) ssDNA. Formation of this hairpin in the structure results in a “duplex-like” arrangement of the bases in the 3’ region of the molecule. While T-T hydrogen bonding is not typically seen in the typical Watson-Crick description of the DNA duplex, in this case they provide some stability to the overall structure. The hydrogen bonding pattern shown in Figure 2.5 is likely weaker than the standard A-T base pair, however hydrogen bond analysis shows that the T-T hydrogen bonds formed in this “duplex-like” hairpin region are stable over the simulation time. A snapshot of the hydrogen bonding pattern between bases 11 and 16, and 12 and 15 are

shown in Figure 2.5. It can also be seen that due to the arrangements of bases around the hairpin, bases 14 and 15 are largely isolated from each other and likely do not undergo photodimerization upon irradiation. While our trajectory results in this particular hairpin arrangement, it should be noted that other combinations are conceivable. One can imagine that hairpins involving more of these T-T hydrogen bonding interactions will be more favored, however with current MD sampling techniques timescales on the order of hundreds of nanoseconds are possible which may not be sufficient to find such conformations. Obviously, heterogeneous ssDNA would be expected to form more defined and predictable (and likely more stable) hairpins due to the propensity to form canonical Watson-Crick base pairs. Investigation of the sequence dependence on hairpin formation is beyond the scope of this study and will not be further discussed.

Due to the importance of forming a globular structure for statistics collection the formation of the 3' hairpin is studied by plotting hydrogen bond distances between bases involved in “duplex-like” interactions. Figure 2.7 indicates that the hydrogen bonds are made and broken between 3 and 34 ns. After this, the hydrogen bonds are stable and remain throughout the rest of the simulation. While simulations by other groups displayed indications of such hydrogen bond formation on the 5 ns timescale, our results show that these are stable interactions.⁴⁴ This reemphasizes the importance of ample sampling when dealing with MD simulations. In particular, when dealing with flexible systems such as ssDNA, where many different microstates or conformations are possible, sampling on a long timescale is a must in order to find a stable local minimum for data collection.

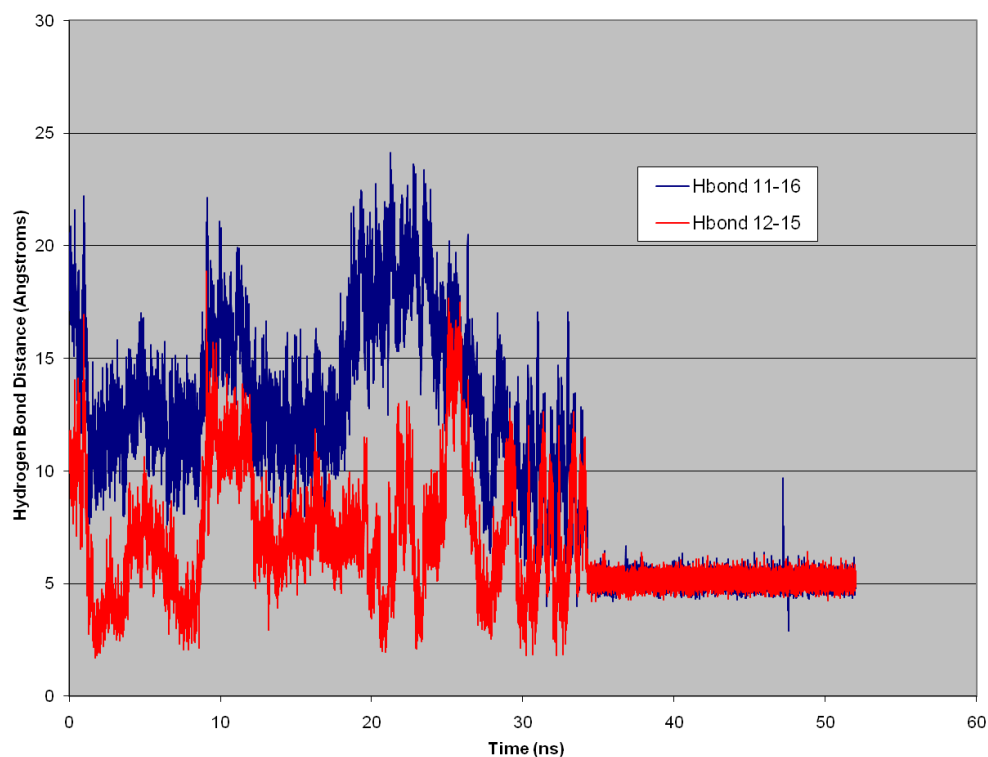


Figure 2.7- Plot of hydrogen bond distances versus time for base pairs involved in duplex like hydrogen bonding interactions.

Aside from the cases where nucleobases are left in a solvent exposed position, the distances of 3% additive probability are in good agreement with what was previously known about π - π stacking of olefins prior to irradiation and subsequent [2 + 2] cycloaddition.⁵⁴ In order for base stack distances to obtain values around 3.5 Å, it is clear that the alignment of the two bases must also be quite good. This suggests that using base stack distances as predictive tool in determining the susceptibility of CPD formation between a given base stack is promising.

2.3.2 Base Stack Interactions: Dihedrals

Along with base stack distance, the dihedral angle formed between the reacting atoms (C5-C6 double bonds) must also be properly aligned for CPD formation to occur at the moment of excitation. In order to do this the dihedral angles are analyzed in much the same way as the base stack distances were in the previous section. This will allow us to pair the two pieces of data and make predictions on which positions in the poly(T) ssDNA represent hot and cold spots for photodamage.

In 2006, Eriksson and Zhang published a computational study aimed at elucidating the reaction pathway leading to the CPD.⁴² Using hybrid DFT and TD-DFT methods they proposed that the reaction initially proceeds through a triplet excited subsequently followed by a triplet-singlet crossing at 2.34 Å on the coordinate forming the C6-C6' bond. Combining this with Kohler's prediction that the reaction is ultrafast and barrierless leads to the hypothesis that if the correct dihedral angle is obtained at the moment of excitation then a reactive conformation is formed and the [2 + 2] cycloaddition proceeds to product. In the reaction leading to CPD formation, the reacting atoms forming this dihedral are C5-C6-C6'-C5'. In the work by Eriksson, two possible pathways for the reaction on the triplet surface are proposed. The first possibility is that the reaction begins by attack of C5 on C5' followed by C6-C6' bond formation. The second possibility is simply the reverse, C6-C6' bond formation followed by C5-C5' bond formation. In each case the C5-C6-C6'-C5' dihedral formed at the TS is $\approx 27^\circ$. In analogy to the previous section dealing with base stack distances, a base pair will be considered to be in a reactive conformation when its dihedral approaches this value.

The *ptraj* module of AMBER was used to extract dihedral angle vs. time information from the trajectory. Because *ptraj* does not have the ability to calculate radial distribution functions from dihedral data, histograms were created using Microsoft Excel. All possible dihedral angles were considered in the histogram analysis and bins 1° in width were chosen. The histogram was subsequently normalized by the total counts resulting in a probability of the dihedral to be located in each particular bin. To estimate how much time each base stack spends properly aligned for reaction, the bins $\pm 3^\circ$ of 27° were summed. While the choice of looking at a 6° range along the dihedral is somewhat arbitrary, the trends seen will hold if the range is widened (or narrowed). This data can be seen in Figure 2.8.

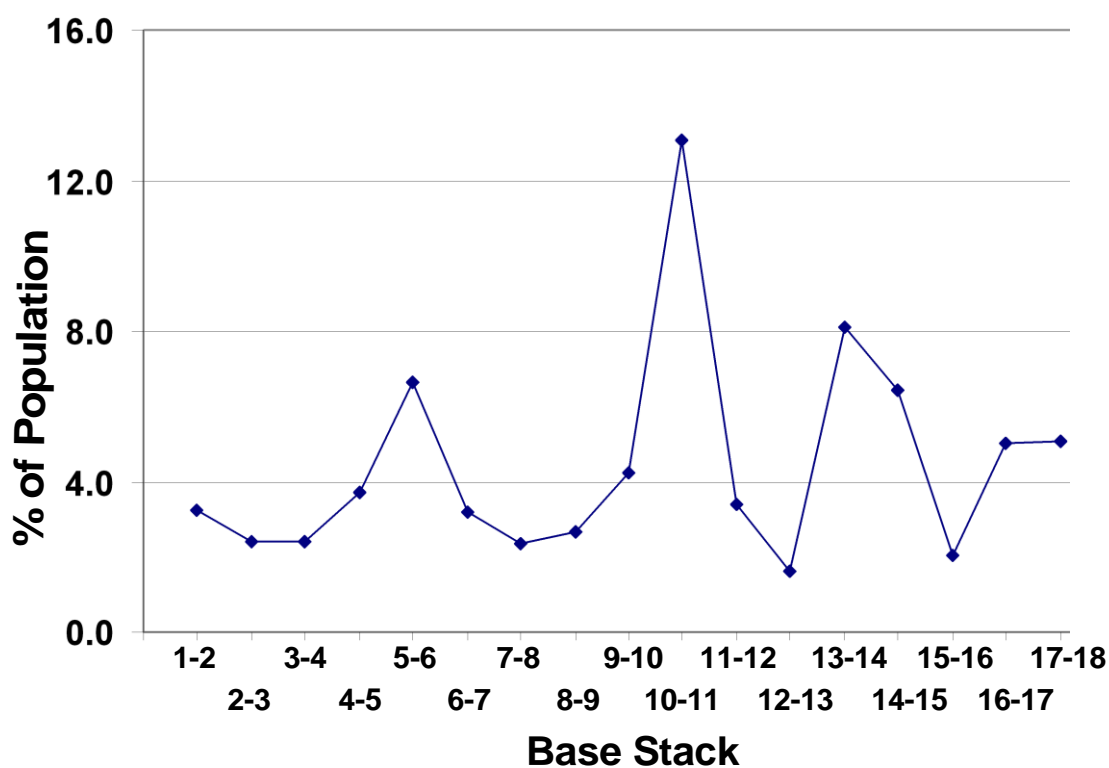


Figure 2.8- Plot of populations of base stack dihedral angles with values of $27 \pm 3^\circ$.

In Figure 2.9 it is clear that the alignment between base stacks in terms of dihedral angle vary widely throughout the ssDNA. Furthermore, the distribution observed is affected by the secondary structure that is formed in the globular folded state. The percentage of time that the C5-C6-C6'-C5' is within the $\pm 3^\circ$ of the TS value ranges from 1.6% for base stack 12-13 to 13.1% for base stack 10-11. In the structure shown in Figure 2.6 it can be seen that base stack 11-12 lie in the “duplex-like” region of the molecule while base stack 13-14 lie on the edge of the hairpin formed. This results in base stack 13-14 being largely isolated and solvent exposed leading to an exclusive base stacking. Towards the end of the simulation, these two bases are not stacking with any other bases, causing a dynamic stacking interaction and leading to an increased probability of crossing through the 24-30° dihedral range chosen. As might be expected from the observation of reduced CPD yields in dsDNA versus ssDNA, the bases in the “duplex-like” region of the DNA strand display a reduced probability of populating the chosen dihedral range. This “duplex-like” arrangement in the 3' region appears to effectively lock the included bases in an unreactive conformation. This seems to confirm the idea that conformational rigidity, much like what is seen in dsDNA, reduce the populations of properly aligned bases for CPD formation. While base stack 10-11 is calculated to have the largest population within chosen “reactive” range, it can be seen from both the radial distribution functions of distance and the snapshot in Figure 2.5 that T18 disrupts the base stack through intercalation between the two bases. Figure 2.6 shows that the distance of 3% additive probability for base stack 10-11 is the second highest of all base stacks with only base stack 14-15 being larger. As mentioned earlier,

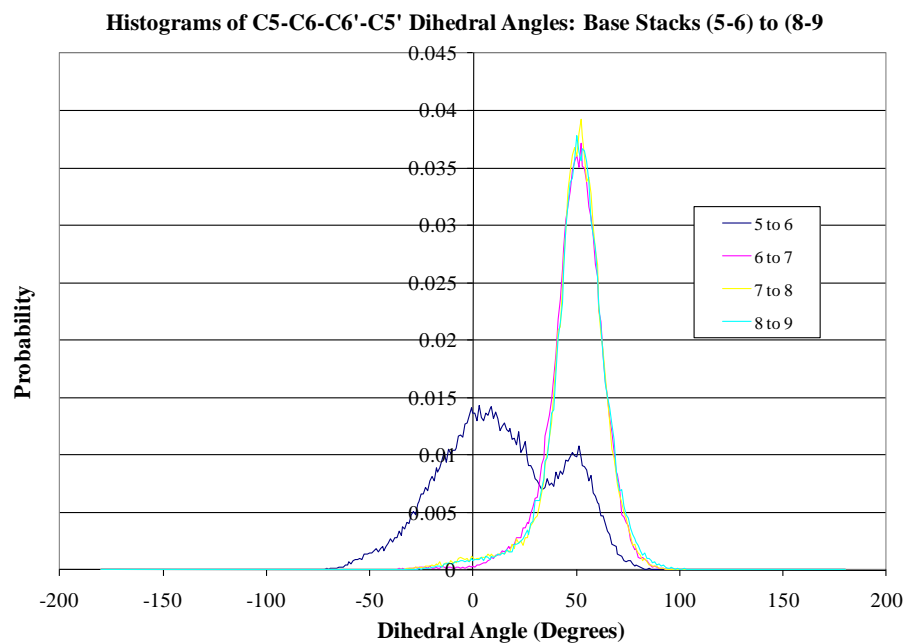
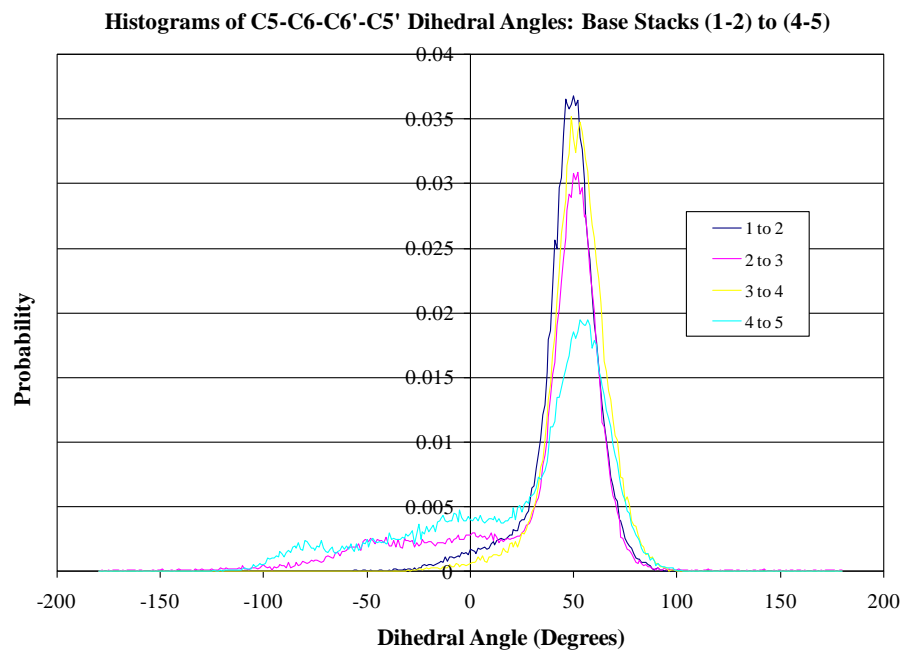


Figure 2.9- Histograms C5-C6-C6'-C5' dihedral angles between individual base stacks for the MD simulation. Figure spans 2 pages.

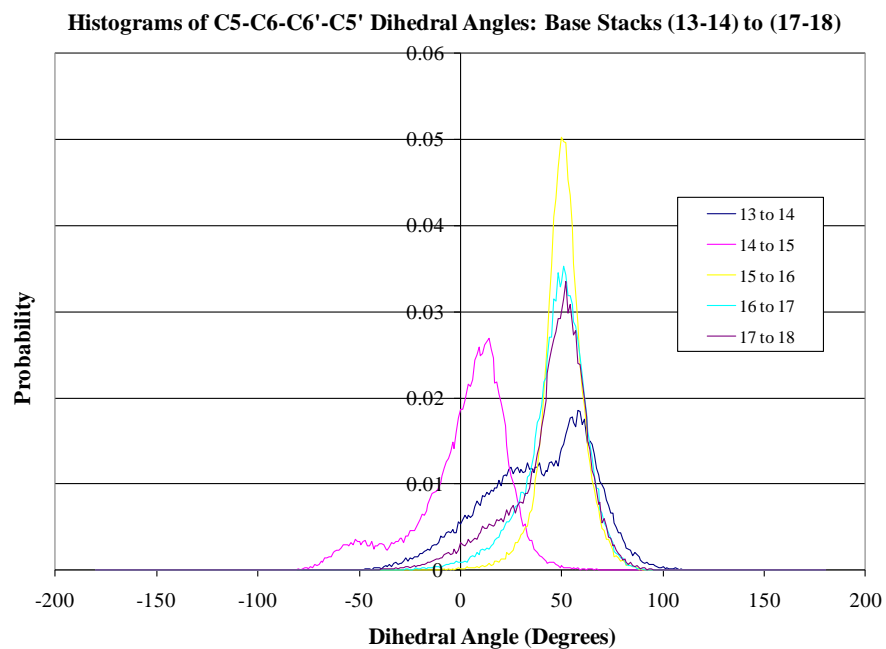
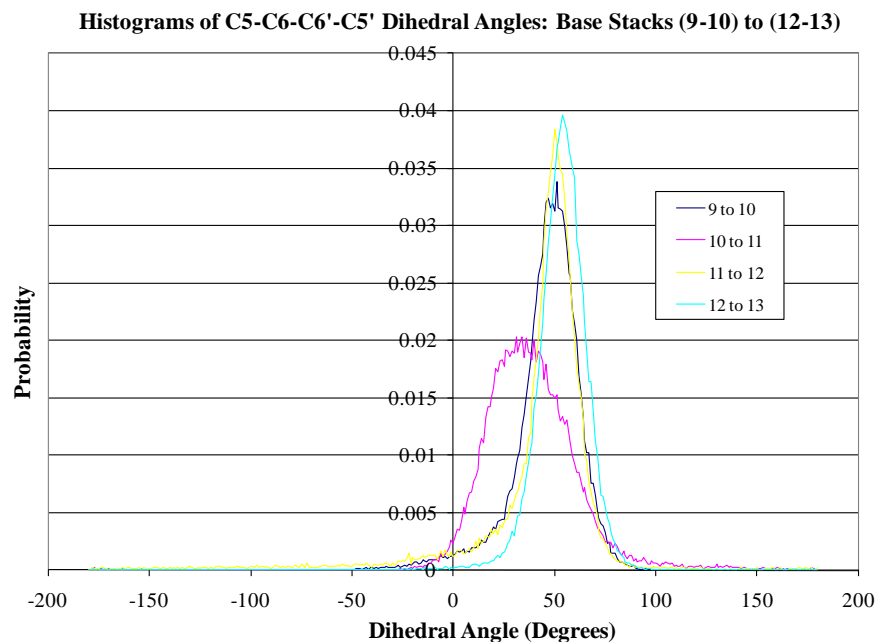


Figure 2.9- Continued.

the cause of the large base stack distance between bases 14 and 15 is the formation of the 3' hairpin.

2.4 Conclusions

As in many organic and biochemical reactions, the orientation of reactants is of utmost importance. From these studies, it appears that in pre-organized systems such as DNA, we may be able to predict excited state reactivity using ground state computational techniques. This is made possible by Kohler's finding that CPD formation is an ultrafast process that proceeds on a timescale much shorter than conformational rearrangement.⁴¹ Furthermore, as CPD formation is likely the first step in a cascade eventually leading to skin cancer, a full understanding of the mechanisms leading to their formation is highly relevant to development of treatments to prevent their formation. In this study, structural information obtained from MD simulation is used to rationalize experimental observations, which taken alone do not give information on an atomistic level. 50 ns MD simulations indicate that the incidences of CPD formation are not equally distributed in poly(T) ssDNA and that the formation of secondary structure features affect the conformations of base stacks along the strand. This can lead to hotspots for damage and also areas of the strand which are resistant to photodamage.

In order to find damage hotspots within poly(T) ssDNA, we selected the six best aligned bases in terms of C6-C6' distance and C5-C6-C6'-C5' dihedral for further analysis. Base stack 13-14 displayed the best combination of the two parameters having the second highest probability for proper dihedral alignment and one of the closest distance of 3% additive probability. By having a combination of both a well aligned

dihedral along with a close base stack distance indicates that base stack 13-14 may represent a hotspot for damage along the single strand. Similar dihedral alignments can be observed for base stacks 5-6 and 14-15, however, 14-15 has the largest distance of 3% additive probability while 5-6 is about average. Base stacks 16-17 and 17-18 are nearly identical with respect to both dihedral alignment and base stack distance. Both have favorable dihedral alignment (5% of the population having a dihedral in the range $27 \pm 3^\circ$) along with a close distance of 3% additive probability of 3.44 Å.

The trends seen from distance and dihedral analysis allow a few generalizations to be made. First, around each bend (5-6) or hairpin (13-14) formed, there is a base stack that is favored in both dihedral and distance terms. The best explanation for this observation is that where a turn in the ssDNA arises, a few bases end up isolated and interact with each other in a dynamic manner. This leads to more crossings through the reactive dihedral range as well as a closer base stack interaction due to the loss of a base stacking partner. This observation is expected to hold even when other possibilities for hairpin formations are possible. From this study, there also seem to be an increased probability for CPD formation to take place in the 3' end of poly(T) ssDNA. Finally, the base stacks that are involved in duplex-like hydrogen bonding tend to be protected from dimerization due to the fact that they are locked in unfavorable dihedral alignments. This is the most likely explanation for increased CPD yields in ssDNA versus dsDNA.

After publication of this study, Lewis and Hariharan studied the context dependence of CPD formation.⁵⁵ By creating defined hairpins through the use of a 12 carbon linker they placed T-T base stacks in different positions in the structure allowing the yields of CPD formation to be related to the surrounding structural elements. Figure

2.10 shows the sequences used in the study and Figure 2.11 shows CPD yields versus time of irradiation. From this study, it is clear that thymine steps lying adjacent to a hairpin produce the largest yields of CPDs which is in agreement with our predictions. Specifically they state “the enhanced reactivity of 6 and 7 suggests that the presence of

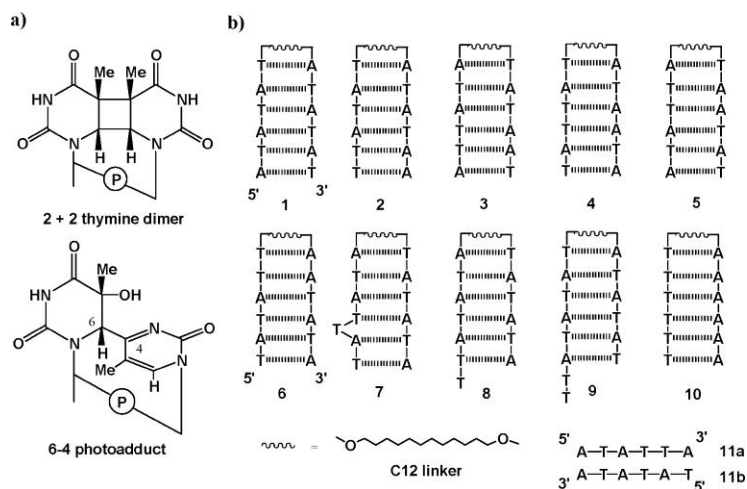


Figure 2.10- Sequences used to study context dependence on CPD formation.⁵⁵

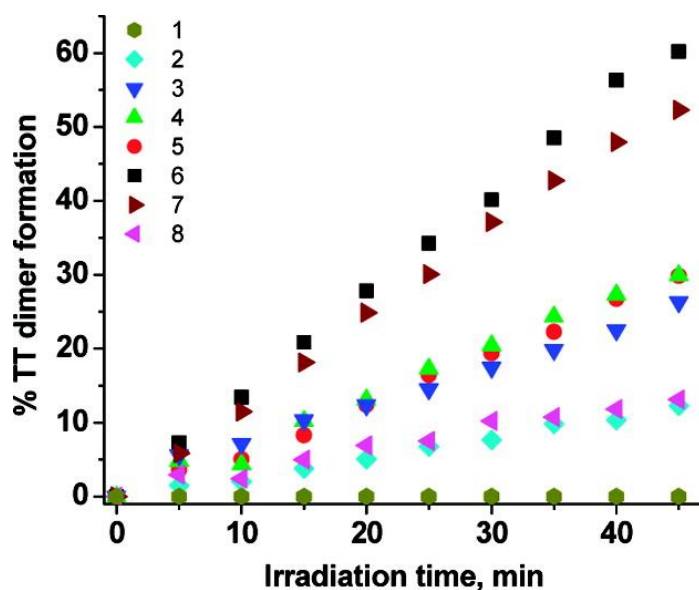


Figure 2.11- CPD yields versus time for the sequences studied by Hariharan and Lewis.⁵⁵

reactive conformations rather than flexibility is the best indicator of T-T dimerization efficiency, as suggested by a recent molecular dynamics simulations for single strand poly(T)".

While our work does not necessarily predict CPD yields from MD simulation, others have used it as the groundwork to attempt to do so. In 2008, Law et. al developed a method to predict CPD yields from MD simulations.⁵⁶ In order to do this, they measured quantum yields from CPD formation of a simple T-T dinucleotide and subsequently fit distance and dihedral parameters to reproduce their experimental results. This lead them to predict that a dimerizable conformers have a base stack distance less than 3.63 Å and a dihedral angle formed between C5-C6-C6'-C5' of less than 48.2°. It must be noted that in subsequent publications, our published work has been compared to that of Law et. al in predicting CPD yields.^{57,58} However, the chosen parameters for our study are derived from experiment and are only meant to be used to estimate what might represent a dimerizable conformation. Furthermore, we do not try to predict CPD yields, but rather simply point out that secondary structure likely has an effect on CPD formation.

CHAPTER 3:

COMPUTATIONAL STUDIES OF THE (6-4) PHOTOPRODUCT REPAIR

REACTION

3.1 Introduction

The (6-4) photoproduct (6-4PP) is one of two major photolesions formed between adjacent pyrimidines in DNA, the other being the cyclobutane pyrimidine dimer (CPD) which is discussed in detail in Chapters 2 and 4. It is formed by a [2 + 2] cycloaddition of the C5-C6 bond of one pyrimidine with the C4-O4 carbonyl (Figure 3.1). This Paterno-Büchi reaction results in the formation of an unstable oxetane species, which further breaks down into the 6-4PP. This net oxygen transfer between the thymine bases that results from the cycloaddition complicates the repair reaction in comparison to the CPD. A special photoreactivation enzyme, (6-4) photolyase, is capable of directly repairing this photolesion in a light dependant manner.⁵⁹ This enzyme, however, does not repair the CPD which reflects the likely difference between the two repair mechanisms. As the repair mechanism employed by DNA photolyase was discussed in detail in

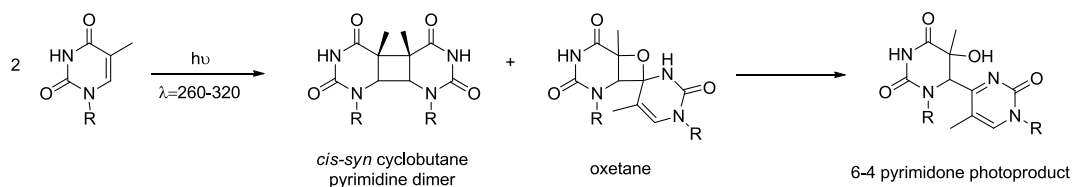


Figure 3.1- Mechanism of CPD and 6-4PP formation.

Section 1.2, it will not be discussed in here. While the relative yield of 6-4PP is only 20-30% compared to 70-80% for the CPD, it is commonly thought that the mutagenic effects of 6-4PP are greater.^{6,7,60,61} It should also be noted the quantum yield for repair for 6-4PP is 0.11, much lower than the reported quantum yield of 0.9 which is reported for CPD repair.^{62,63}

Early mechanistic proposals concerning the repair of 6-4PP focused on drawing analogies to CPD repair. The first, and at the time widely accepted, mechanism of repair involved a thermal closure of the O4' oxygen onto the C4' carbon reforming the oxetane intermediate originally formed in the Paterno-Büchi reaction which formed the 6-4PP (Figure 3.2).⁶⁴ From here, an electron is injected into the oxetane intermediate leading to

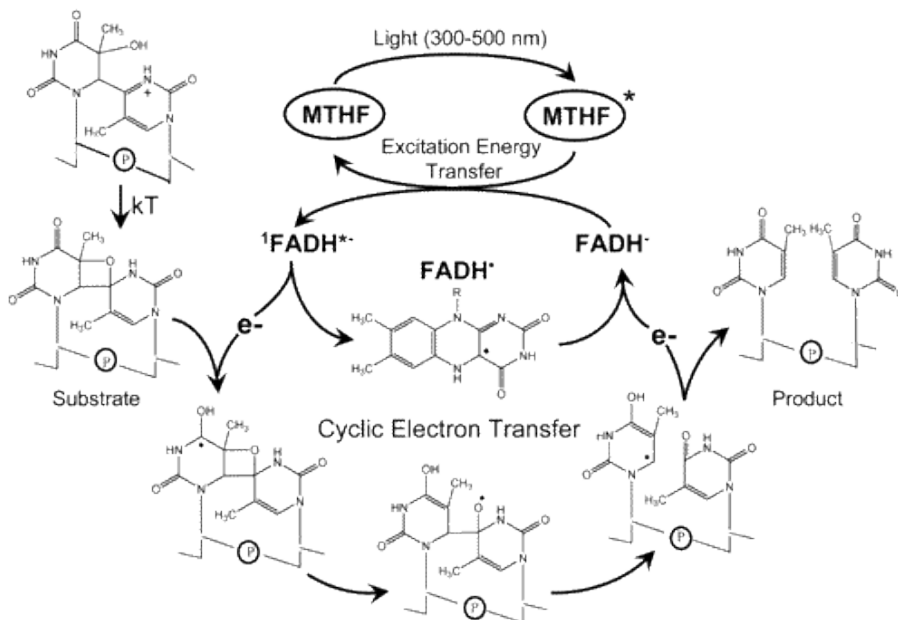


Figure 3.2- Original mechanistic proposal involving an oxetane intermediate put forth by Sancar and coworkers.¹²

a stepwise cycloreversion reaction in analogy to the proposed CPD repair mechanism. This mechanistic proposal by Sancar et al. has lead to over 10 years of research into either confirming or repudiating their proposal.^{6,19,60,63-69}

The thermal formation of the oxetane intermediate is highly endothermic. Calculations on model systems by Heelis and Liu⁷⁰ and also by Harrison and Wiest predict this species to be between 15 and 16.5 kcal/mol endothermic from the 6-4PP. This is in line with experimental data showing that the oxetane intermediate is not stable above -78°C.¹ The endothermic nature of this step of the reaction can be explained using very simple principles from organic chemistry. The oxetane species is of a higher energy due to ring strain and bond angle strain. Another interesting property of the proposed oxetane intermediate is its carbinolamine (hemiaminal) character. In this scenario the N3' nitrogen is capable of donating its lone pair of electrons forming a double bond with C4', breaking the oxetane ring. This "neighboring group effect" leads to a weakening of the C5-O4' bond and further increases the energy of the intermediate. In order for this mechanism to work, the (6-4) photolyase enzyme would have to selectively stabilize the oxetane intermediate and then facilitate electron transfer into the intermediate. The mechanism is in line with a few important experimental observations. First, it explains the transfer of the 4' oxygen from one base to the other.⁶⁴ Also, it is in line with what is known about thiarane formation in analogs of 6-4PP.⁶⁷ It should be noted however, that the sulfur analogues would be expected to be more stable due to the increased nucleophilicity of sulfur and also the involvement of d-orbitals resulting in a decrease in angle strain. Also, C-S bonds are approximately 0.4 Å longer than C-O bond further decreasing the strain of the thiatane analog. Other than being a mechanism that explains

the bond making and bond breaking required to arrive at the correct product, it is lacking supporting experimental evidence and the thermodynamics involved with oxetane formation do not bode well for the feasibility of the mechanism.

In an article published in *Science* Carell and coworkers published the first crystal structure of a (6-4) photolyase.⁶⁶ Two crystal structures bound to duplex DNA were disclosed, one bound to a sequence containing a 6-4PP and one containing the reconstituted thymine bases. However, no electron density was assigned to anything resembling an oxetane intermediate which would be expected to be formed if the Sancar mechanism was correct. The authors cite another piece of evidence that points away from the oxetane mechanism is the positioning of the two key histidine residues in the active site.¹⁹ Neither is hydrogen bonded to N3' which is essential to catalyze the closure into the punitive oxetane intermediate.¹² In order for the oxetane mechanism to proceed, the enzyme would have to selectively stabilize the oxetane for electron transfer. However, there is no structural evidence to support this claim as no contacts present in the crystal structure would confer the necessary stability to this intermediate. Furthermore, it has been shown that the electron affinity of 6-4PP (14.8 kcal/mol) is much greater than in the oxetane intermediate (3.6 kcal/mol).⁷¹ In this scenario, it is much easier to donate an electron into 6-4PP than the oxetane intermediate. The authors instead propose a new mechanism which offers an alternative to the highly strained oxetane intermediate.

The mechanism proposed by Carell and coworkers is shown in Figure 3.3. The first step of their proposal involves the electron transfer from FADH⁻ into the 5' base of 6-4PP. From here, the C5-O4' bond breaks coupled with a proton transfer from HIS365

resulting in dehydration. Next, the released water then attacks the C4' carbon forming a carbinolamine. Following C4'-C6 bond cleavage, a proton transfer and electron transfer the repaired thymines are obtained. While this mechanism is a likely improvement upon the oxetane mechanism in terms of ring and angle strain, it is not without its questions. The most obvious question that should be asked is why is the 4' OH group protonated over the α - β unsaturated imine? The real question here is where does the electron from the FADH⁻ go when injected into the 6-4PP? On the surface, it seems obvious that any electron added to 6-4PP would be localized on the 3' base due to the LUMO of the aromatic system being lower than the LUMO of the nonaromatic 5' base. If the electron is localized on the 5' base it may be possible to protonate the 4' OH group first. However, if the electron is localized on the 3' base's α - β unsaturated imine it is much more likely that N3' would be protonated in the first step deviating from the proposed mechanism. They argue that the fact that uracil dimer is repaired at a factor of 4-5 times slower is evidence of the more stable radical intermediate in the thymine example.⁶⁶ However, it may just be reflective of the leaving group ability of water being much better at a tertiary position in comparison to a secondary position.

In the two years following the publishing of the first (6-4) photolyase structure, at least two new mechanisms were proposed in the literature. In the first, published by Domratcheva and Schlichting, a nonadiabatic hydroxyl transfer was proposed (Figure 3.4).⁶⁹ In this proposal the first step involves electron transfer into the 6-4PP. From here, the loss of hydroxide anion occurs with a large transition state energy (24.5 kcal/mol). A shallow minima was found corresponding the dissociated hydroxyl species. From here,

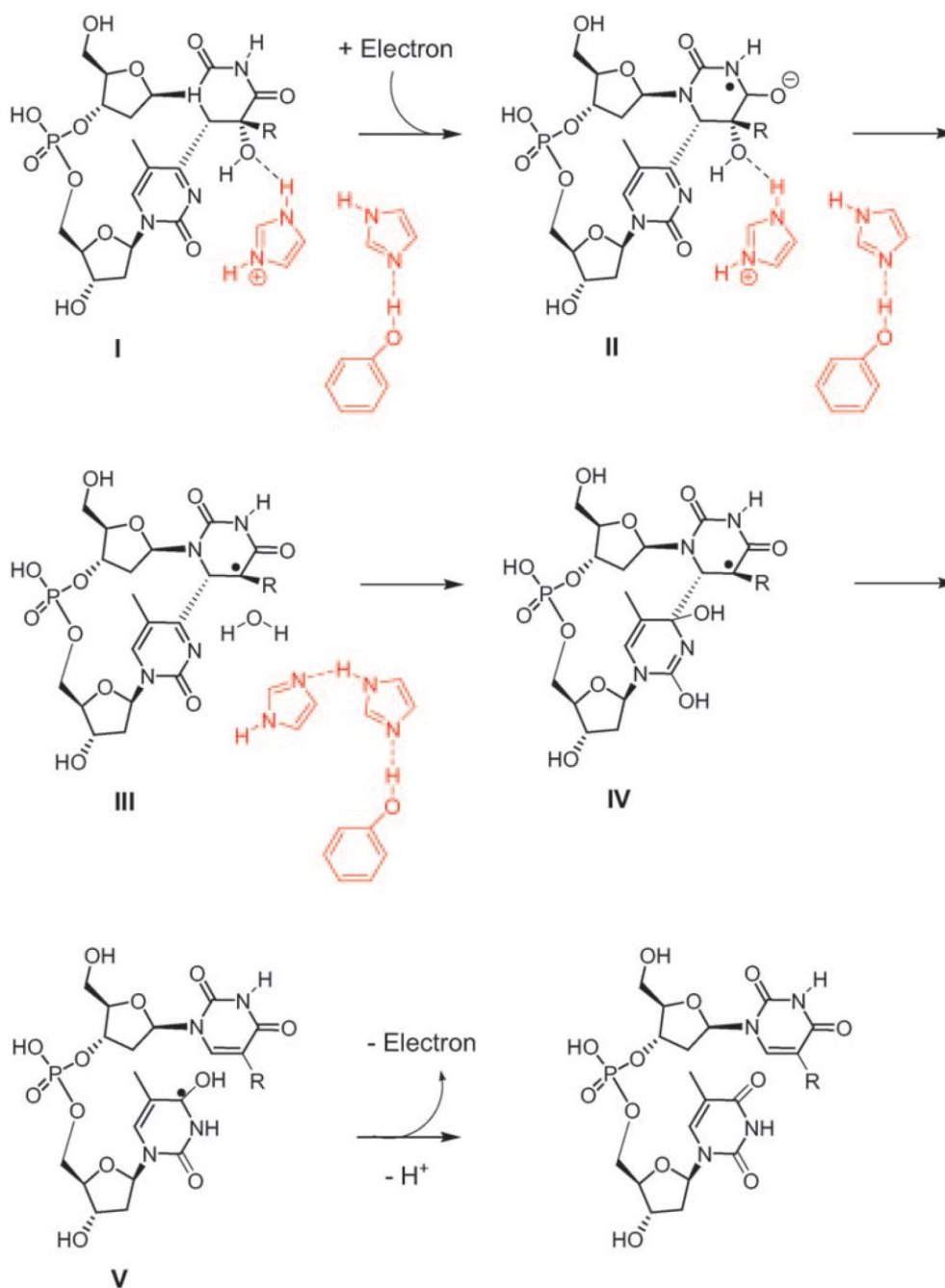


Figure 3.3- Water transfer mechanism proposed by Carell and coworkers along with the publishing of the first crystal structure of (6-4) photolyase.⁶⁶

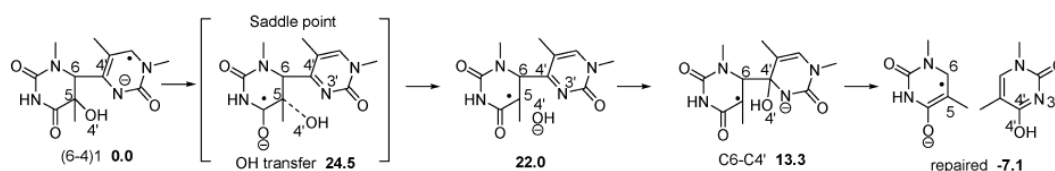


Figure 3.4- Hydroxyl transfer mechanism proposed by Domratcheva and Schlichting. Energies are relative to the 6-4PP radical anion.⁶⁹

the hydroxide anion attacks the C4' carbon followed by cleavage of the C6-C4' bond. The final step of the reaction is back electron transfer to FADH[•]. This proposal is also not without its questions. First, it has been shown that likely H365 is protonated in the active site through the use of pH titration experiments and mutation studies.⁶⁵ However, there is no proton transfer component to the mechanism they proposed. Another question that must be answered is if the nonadiabatic nature of this mechanism is achievable. For their reaction to take place there must be a photoexcitation from the $3'\pi_1^*$ orbital to the $5'\pi_2^*$ orbital must take place (Figure 3.5). This proposal predicts that the repair process requires two photons. The final question that must be answered is if the hydroxide anion is an acceptable leaving group? In the likely acidic environment of the active site it is unlikely that hydroxide would be lost, but rather the 4' OH group would need to be protonated in order for dehydration to occur.

Another mechanistic proposal was put forth again by Sancar et al.⁶⁸ Using femtosecond spectroscopy they studied the electron transfer reactions present in the repair mechanism of 6-4PP. They show that electron transfer from the excited flavin cofactor to the 6-4PP occurs in 225 ps, however, they also detect fast back electron transfer in 50 ps without repair. A proton transfer was found to be important to the

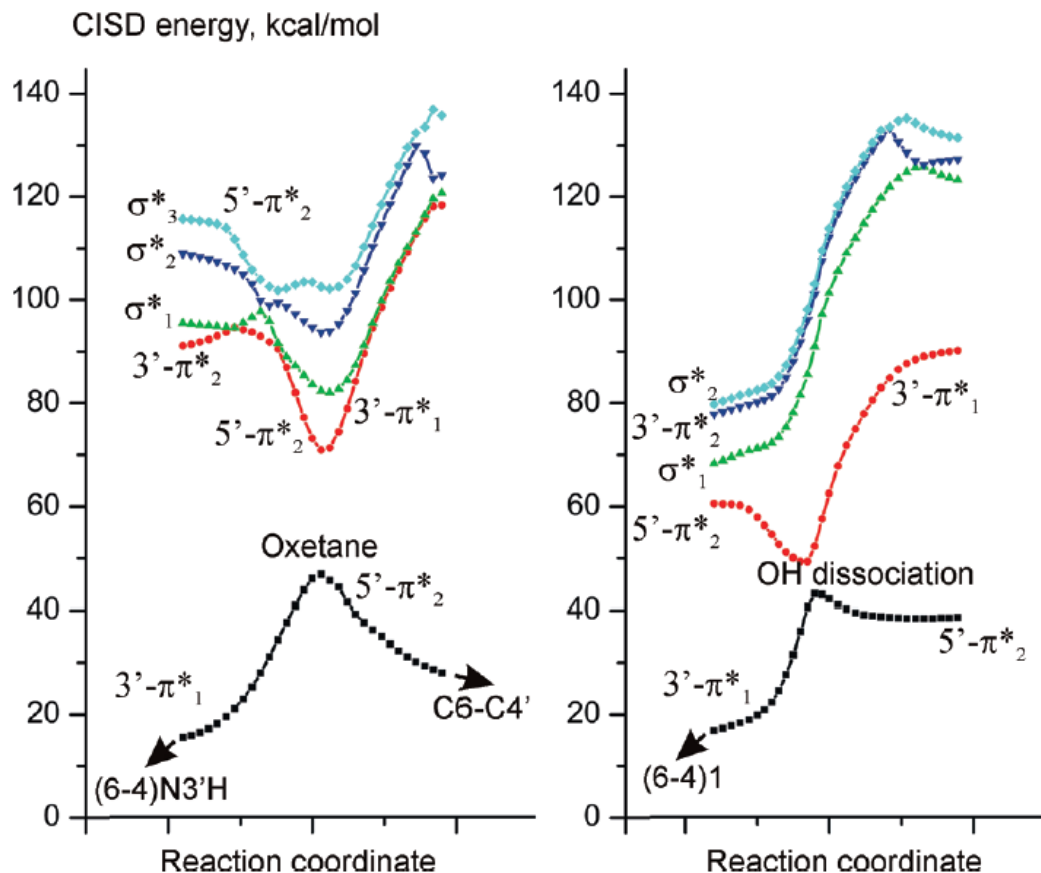


Figure 3.5- Reaction coordinate showing the nonadiabatic repair mechanism proposed by Domratcheva and Schlichting. The oxetane mechanism is shown on the left while the hydroxyl mechanism is shown on the right.⁶⁹

reaction mechanism and serves as a gate leading either to repair, or immediate back electron transfer. When proton transfer occurs in approximately 425 ps, the repair proceeds to products in tens of nanoseconds. The mechanism proposed by Sancar and coworkers can be found in Figure 3.6. In this mechanism, electron transfer from the

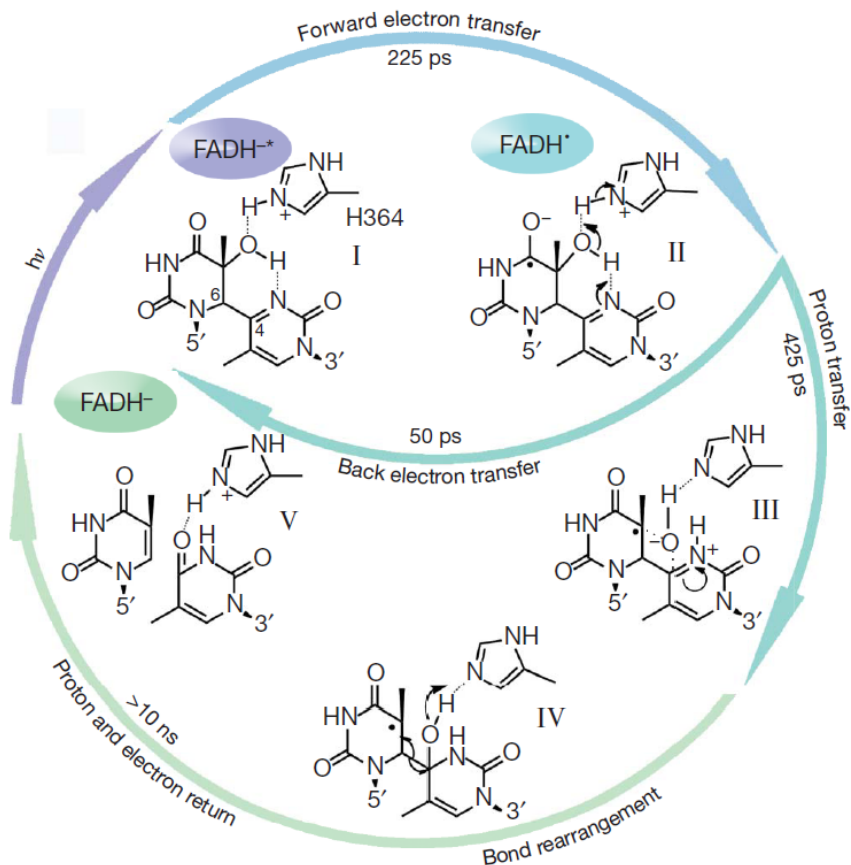


Figure 3.6- Mechanistic proposal put forth by Sancar and coworkers based upon observations on their ultrafast spectroscopy study.⁶⁸

flavin cofactor results in the protonation of N3'. From here, hydroxide is transferred from the 5' to the 3' base through an "oxetane-like" transition state. The final steps involve the cleavage of C6-C4' bond, proton transfer back to the histidine, and back electron transfer to the flavin cofactor. This mechanism also is not without any spurious intermediates. In the hydroxyl transfer step of the mechanism there is the presence of a hydroxide anion and a protonated α - β unsaturated imine. As water has a pKa of 15.7 and a protonated imine has a pKa of <6, it is hard to imagine that this species exists in any

appreciable amounts. However, it is possible that the radical character of this intermediate affects this equilibrium.

The purpose of this work is to explore these hydroxyl transfer mechanisms through the use of electronic structure methods. While the newly proposed mechanisms are likely an improvement over the original oxetane mechanism, they are not without their questions. The first question we plan to answer is what is the electronic structure of the 6-4PP radical anion? It is important where this electron is localized as it will increase the basicity of these groups and will likely result in protonation. The second, and perhaps most important piece of information we want to discover is what is the dissociating species? The two newly published mechanisms suggest that this species is a free hydroxide anion.^{68,69} However, many studies have demonstrated how poor of a leaving group the hydroxide anion is. An example of this can be seen in the E₁cB dehydration mechanism.⁷² Therefore, we also explore ways to activate this hydroxyl group for dissociation including hydrogen bonding to active site residues along with additional protonation steps.

3.2 Methods

Initial coordinates were obtained from the crystal structure of (6-4) photolyase bound to 6-4PP (pdb code 3CVV). Gaussview 5.0 was used to set up all calculations and Gaussian 09⁷³ was used to perform all calculations. Unresolved hydrogens were subsequently added to fill valencies. For theozyme calculations amino acids were terminated with methyl groups (histidine -> methyl imidazole, etc.). Calculations were performed at the UB3LYP/ 6-31G* level and zero point energies were obtained through

vibrational analysis. In order to preserve the approximate volume of the active site, the carbon of the terminal methyl groups of active site amino acids were frozen. A polarizable continuum model with dielectric constants of 20 and 80 were used to approximate the dielectric environment of the active site. Scans of the potential energy surface of C5-O4' bond elongation were performed by taking steps of 0.1 Å. From here, the highest energy structure was chosen and the transition state was optimized using the Berny algorithm. The identity of the transition state was confirmed through vibrational analysis and the presence of one negative frequency corresponding to C5-O4' bond cleavage.

Due to the large size of the expanded theozyme model used to the electronic structure of (6-4)PP[•] UAM1 was used for the optimization procedure. From here, single point calculations at the UB3LYP/ 6-31+G* level were performed. Subsequently, cubegen was used to create the SOMO plot seen in Figure 3.15. Three different models are utilized in this study. The first is a minimalistic model containing only 6-4PP. The second contains 6-4PP and both active site histidines while the third also contains TYR423.

3.3 Results

3.3.1 Calculations on 6-4PP

The first model used in this study contains only 6-4PP. This simplistic model not containing any active site residues allows for fast study of the reaction leading to 6-4PP repair. It also allows for comparison of the uncatalyzed reaction to that in the active site environment. In Sancar's second mechanistic proposal (hydroxyl transfer) it was shown

that electron transfer is the likely first step in 6-4PP repair.⁶⁸ Therefore the electronic structure of the 6-4PP radical anion was first calculated. From here the O4' oxygen was protonated. Upon minimization this proton was spontaneously transferred to the N3' imide nitrogen suggesting a barrierless pathway for this proton movement. This supports the idea that the localization of the electron will be on the α - β unsaturated imine upon electron donation from FADH⁻ upon excitation. This finding is in contrast to Carell's published mechanism in which water is able to leave upon protonation of O4'.⁶⁶ It is also not consistent with the mechanism proposed by Domratcheva and Schlichting which predicts that no proton transfer from the enzyme is utilized.⁶⁹ Next, we studied the hydroxyl mechanism proposed by Sancar. A scan of the C5-O4' bond was commenced by lengthening the bond by 0.1 Å steps while minimizing all other degrees of freedom. A synchronous transition state (Figure 3.7) was found corresponding to the breaking of the C5-O4' bond (1.86 Å) and the formation of the C4'-O4' bond (2.27 Å). This "oxetane-like" transition state is what Sancar and coworkers predict from their femtosecond spectroscopy studies. However, we predict this transition state to be \approx 42 kcal/mol endothermic from the protonated imine radical intermediate. This transition state is \approx 26 kcal/mol higher in energy than their previously proposed oxetane intermediate and is not likely to be involved in the enzymatic repair mechanism. This is to be expected by considering the leaving group ability of the hydroxide anion, a very poor leaving group that generally must be converted into better leaving group through either protonation or synthetic means.

A double proton transfer mechanism including protonation of O4' resulted in a much lower transition state leading to water dissociation. This transition state is shown

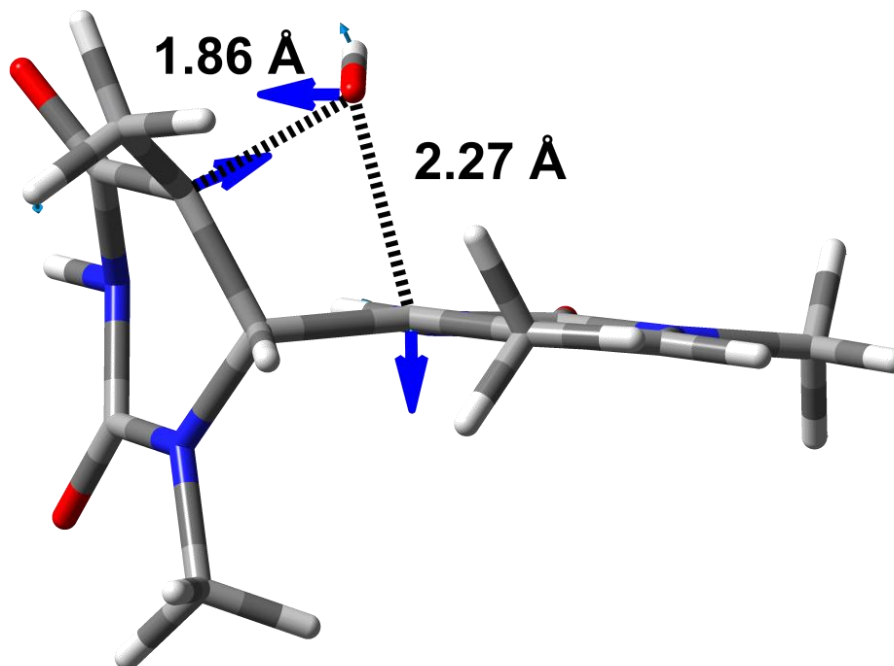


Figure 3.7- Transition structure of hydroxyl transfer in 6-4PP model. Activation energy ≈ 45 kcal/mol.

in Figure 3.8. The energy of this species is only 7.2 kcal/mol endothermic from the protonated alcohol radical cation intermediate, a much more reasonable energetic penalty more in line with the observation that the reaction is complete within 10 ns. There is however the question of where this second proton could come from. The first proton, as stated earlier, comes from a protonated histidine in the active site. The second could come from a tyrosine which is hydrogen bonded to H369. With a pKa of ≈ 9.9 it could be involved in a proton shuttling mechanism leading to protonation of O4'.

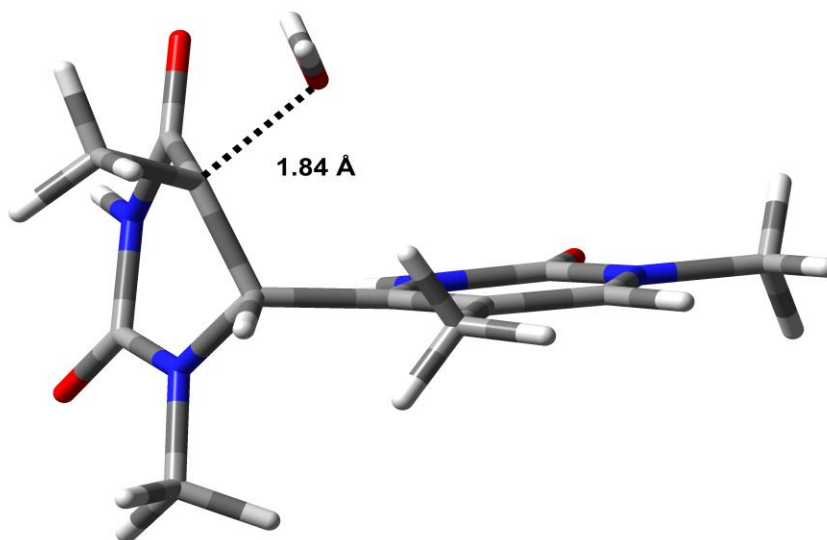


Figure 3.8- Transition structure of water dissociation. Activation energy ≈ 7.2 kcal/mol.

3.3.2 Calculations on 6-4PP + HIS365 + HIS369

In order to investigate the role of the active site histidines in the 6-4PP repair mechanism they were included in a theozyme model. As stated in the methods section, the terminal methyl groups of the histidine side chains were restrained to their initial coordinates. This allowed the sidechains to rotate but kept the overall volume of the active site close to that of the crystal structure. Figure 3.9 shows the structure of the active site containing 6-4PP and the two histidines (HIS365 is protonated). When an electron is injected into 6-4PP, a conformational rearrangement of the active site occurs with HIS365 forming a hydrogen bond with N3' (Figure 3.10). This again is consistent with the injected electron being localized on the α - β unsaturated imine. Proton transfer from HIS365 to N3' is calculated to be -14.7 kcal/mol exothermic. This indicates once

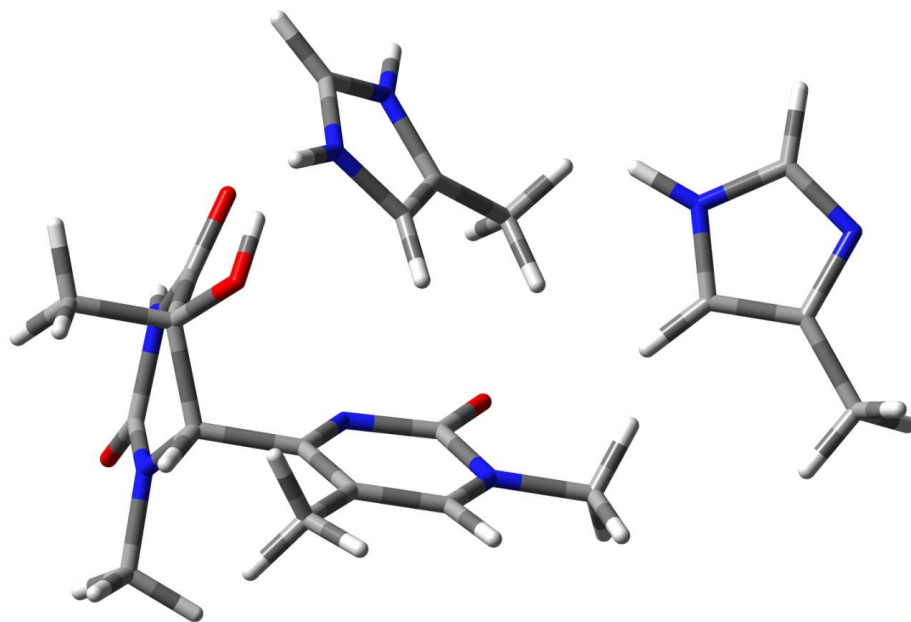


Figure 3.9- Optimized ground state structure of 6-4PP with the active site histidines.

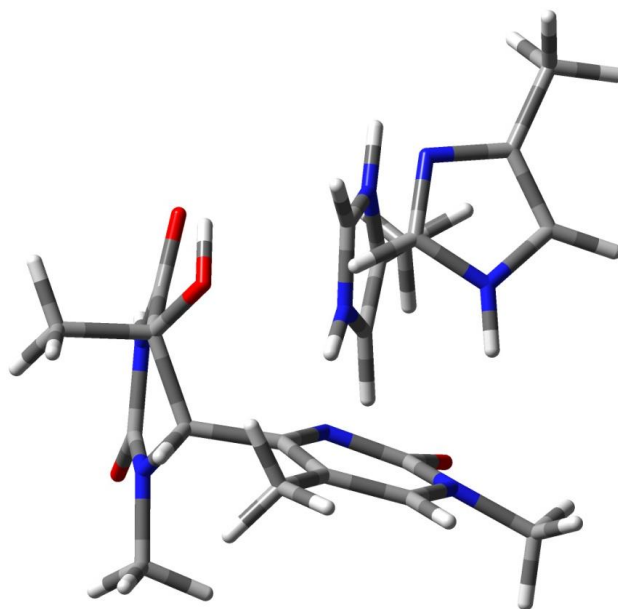


Figure 3.10- Optimized structure of the active site following injection of an electron. Notice the rotation of H365 to accommodate hydrogen bonding to N3'.

more that the likely first step following electron uptake by 6-4PP is protonation of N3'. Next, a scan is performed which lengthens the C5-O4' bond by 0.1 Å per step. In the optimized structures a later transition state is observed in respect to the reaction coordinate describing the C5-O4' bond. This transition state has a bond length between C5 and O4' of 2.08 Å which is 0.22 Å longer than what is observed without the active site histidines (Figure 3.11). However, this only results in a reduction of the transition state energy from 45 kcal/mol to 42 kcal/mol. This indicates that the histidines in their isoelectric state are not able to greatly reduce the activation energy leading to C5-O4' bond cleavage. Again, this leads us to believe that in some way this O4' hydroxyl must be activated, most likely through protonation.

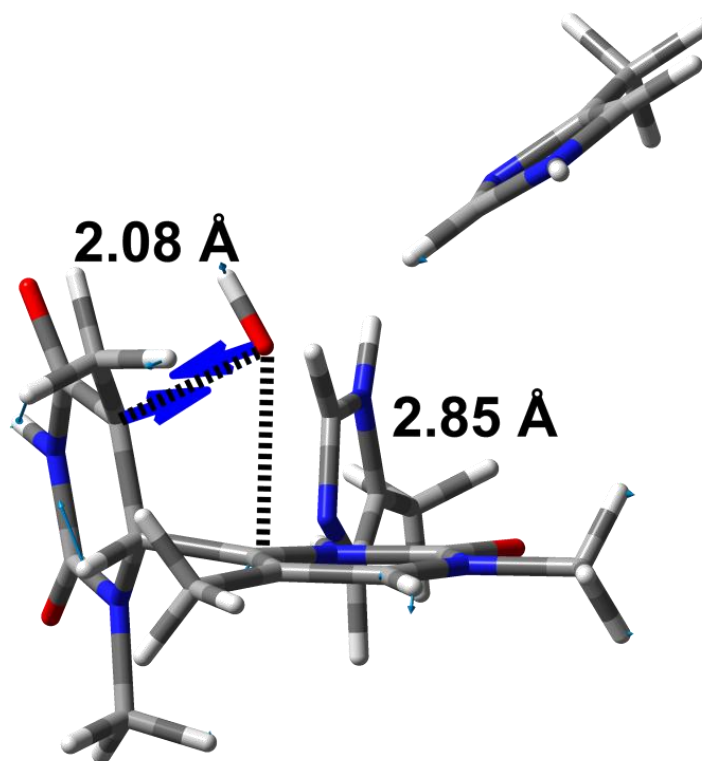


Figure 3.11- Transition state structure for hydroxyl transfer in the theozyme model.

If one assumes that H369 is able to shuttle a proton from TYR423 we can model the effect of having a protonated histidine in proximity to the departing hydroxyl group. A scan along the C5-O4' bond leads to a synchronous transition state with a large imaginary frequency cooresponding to both C5-O4' cleavage and protonation of the hydroxyl group by H369 (Figure 3.12). This transition state lies ≈ 32 kcal/mol higher in energy than the radical cation system. This seems to preclude a synchronous transition state where the C-O bond is broken as protonation also takes place. However, a stationary point was not able to be found with the hydroxyl group protonated. Rather, protonation of O4' leads to spontaneous proton transfer back to HIS369. Furthermore,

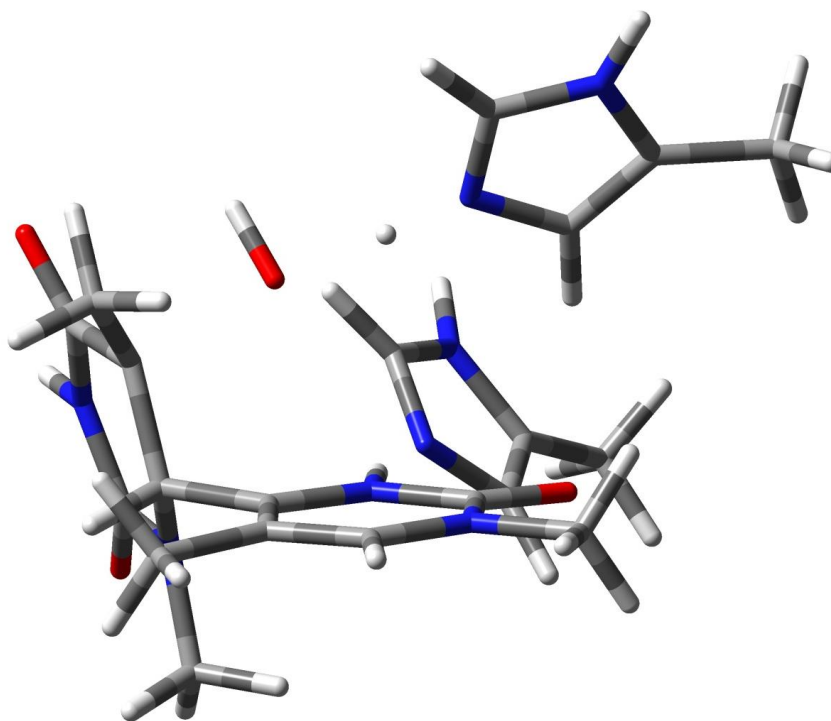


Figure 3.12- Transition state structure with H369 protonated. Transition state involves a synchronous breaking of C-O bond with formation of O-H bond. ≈ 32 kcal/mol endothermic from the radical cation system.

this transition state is much higher in energy than what is predicted by our model utilizing only 6-4PP. It does indicate that in some way discrete protonation, perhaps through a proton tunneling mechanism, is required for C5-O4' cleavage.

3.3.3 Calculation on 6-4PP + HIS365 + HIS369 + TYR423

In order to study whether TYR423 can be active in proton transfer to 6-4PP we studied a theozyme model in which it is included in the calculation. Being hydrogen bonded to H369 and also containing a somewhat acidic proton ($pK_a = 9.9$), it could be involved in proton shuttling to O4'. Just as in the previous example, when an electron is injected into the system, the protonated HIS365 rotates to hydrogen bond with N3'. Again, protonation of N3' by HIS365 is exothermic by 13.2 kcal/mol. The structure of the N3' protonated species including hydrogen bonding distances is shown in Figure 3.13. From here it can be seen that the contacts required for a proton shuttling mechanism are intact. A scan along the C5-O4' coordinate followed by an optimization of the transition state show that the activation energy required to break this bond is 32.6 kcal/mol (Figure 3.14). This is the same activation energy required in the previous example with a protonated HIS369. Again, this transition state corresponds to a synchronous breaking of the C5-O4' bond and the forming of the O4'-HIS369 proton transfer. No proton transfer from TYR423 is observed. The high energy of this transition state likely precludes this mechanism from being employed in the repair of 6-4PP.

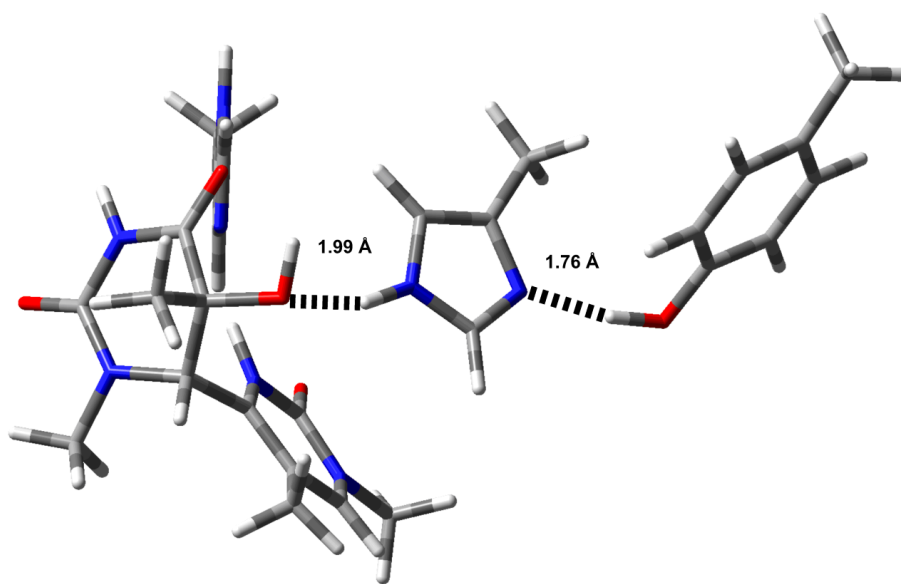


Figure 3.13- Hydrogen bonding network present in 6-4PP + HIS365 + HIS369 + TYR423 theozyme model.

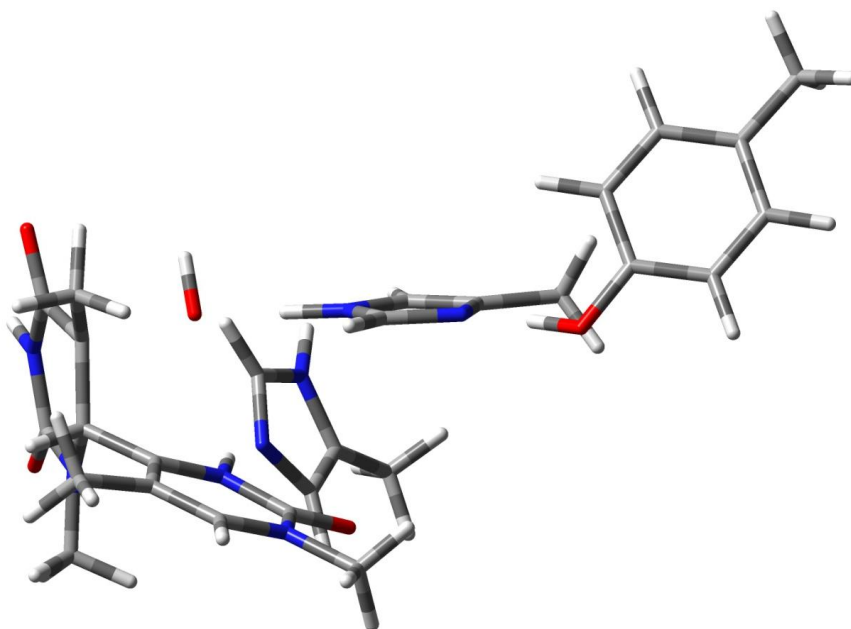


Figure 3.14- Transition state structure of hydroxyl transfer in the theozyme model including TYR423. Activation energy =32.6 kcal/mol from N3' protonated species.

3.3.4 Calculations on the Expanded Theozyme Model

In order to test whether hydrogen bonding interactions present in the active site are capable of stabilizing the radical anion formation on the 5' base instead of the α - β unsaturated imine present on the 3' base, a theozyme model was constructed in which all hydrogen bonds to the 5' base are included. Optimizations at the UAM1 level are used to calculate the geometry of the 6-4PP radical anion species. In analogy to the other theozyme models, upon uptake of an electron HIS365 rotates to hydrogen bond to N3'. Single point calculations of this 6-4PP species with N3' hydrogen bonding to HIS 365 at the UB3LYP/6-31+G* level show that the SOMO upon electron uptake lies on the 3' base (Figure 3.15). This again is a behavior that would be expected. As the 3' base is aromatic, one would expect its LUMO to be substantially lower in energy than the

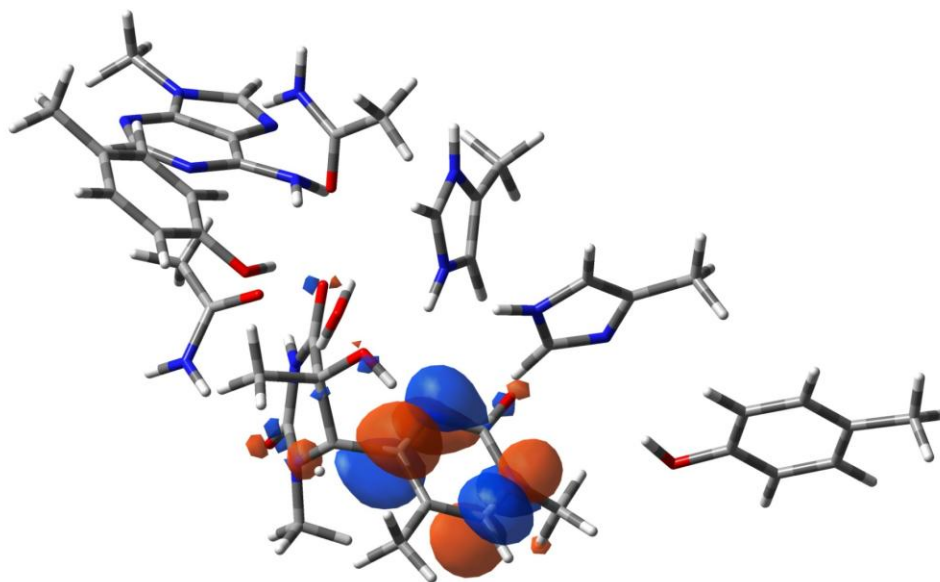


Figure 3.15- SOMO of 6-4PP in the theozyme active site.

orbitals on the 5' base meaning it is the likely destination for the transferred electron. Analysis of this species reaffirms this thinking and shows that the hydrogen bonding to the 5' base is not enough to stabilize radical anion formation on the 5' base. The negative charge formed on the 3' base increases the basicity of the imide nitrogen (N3') and is therefore the most likely position for protonation upon electron transfer. This point has been demonstrated in the previous theozyme calculations showing the proton transfer from HIS365 to the 6-4PP radical anion N3' nitrogen is ≈ 13.2 - 14.7 kcal/mol exothermic. Therefore, upon electron uptake from the flavin cofactor the most likely and energetically favorable next step of the reaction would be protonation of N3'. Any mechanistic proposal studied from here should take this finding into account as it both makes sense in terms of energetic derived from the calculation, but also make chemically intuitive sense unlike many other mechanisms which have been published on the mechanism of 6-4PP repair.

3.4 Conclusions

The calculations performed in this study provide insight into the three newly proposed mechanisms in the literature.^{66,68,69} The C5-O4' cleavages steps for each of these repair mechanisms are shown in Figure 3.16. Mechanism A⁶⁶ which was proposed by Carell and coworkers and involves the protonation of O4' and subsequent dehydration. Our calculations show that there is no stationary point on the potential energy surface corresponding to this intermediate. When O4' is protonated on the 6-4PP radical anion, spontaneous transfer of the proton to N3' (in the 6-4PP model) or HIS369 (in the theozyme models) occurs. This is consistent with the idea that the electron which

is donated from the flavin cofactor is localized upon the α - β unsaturated imine. Even in the presence of the hydrogen bonding moieties of the active site, electronic structure calculations show that this electron is indeed localized on the 3' base. Therefore, we hypothesize that the first step of the reaction following electron uptake is protonation of N3', not protonation of O4'.

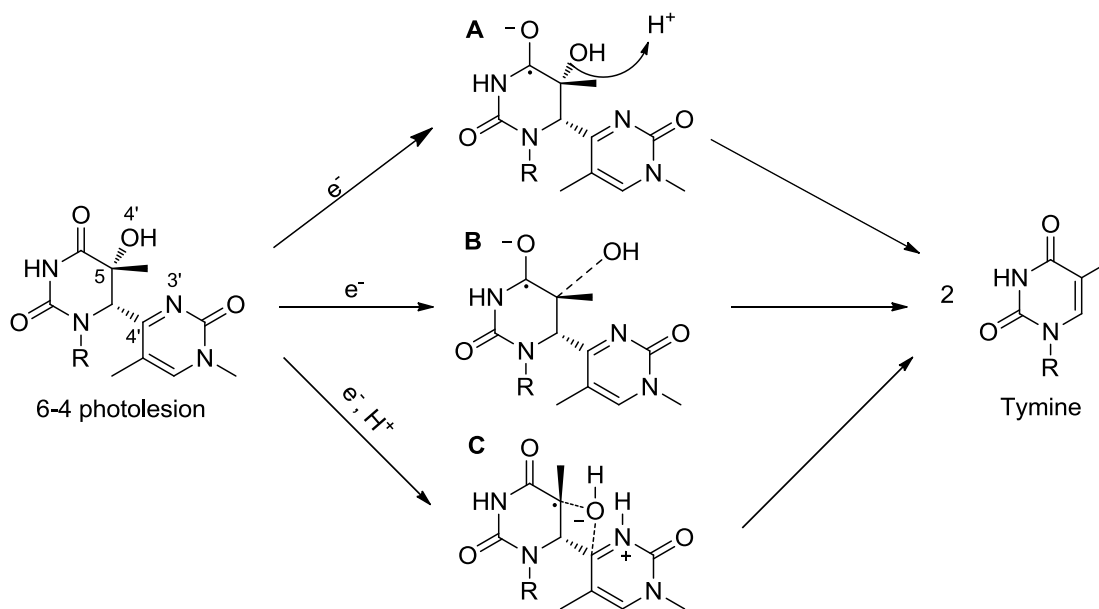


Figure 3.16- Previously proposed intermediate along the 6-4PP repair pathway.^{66,68,69}

The hydroxyl transfer mechanism **B**⁶⁹ can likely be ruled out by considering the experimental evidence. First, this mechanism requires both the excitation of the flavin cofactor as well as 6-4PP (a two photon process) for the nonadiabatic mechanism to occur. However, repair of the 6-4PP is detected with excitation wavelengths outside of the absorption range of 6-4PP. Also, in the ultrafast spectroscopy study, repair is detected upon excitation at 325 nm. While this wavelength is within the absorption band of 6-4PP, its excited state lifetime would need to be on the order of nanoseconds in order

for this mechanism to occur. In addition, the loss of hydroxide anion is a very endothermic process with an activation energy on the order of 24.5 kcal/mol. This process is more endothermic than oxetane formation and our theory calculations show that it is unlikely to be sufficiently stabilized by the 6-4 photolyase enzyme. It also does not include a proton transfer from the enzyme to the 6-4PP radical anion which has been shown to be a part of the repair reaction through both titration¹⁹ and ultrafast spectroscopy experiments.⁶⁸

Pathway C⁶⁸, which was calculated in detail in this study, is likely inaccessible due to similar energetic arguments. While it is reasonable that the first step of the repair mechanism is protonation of N3', however, our calculations show the loss of hydroxide anion is extremely endothermic. We have shown that loss of hydroxide requires ≈ 32 -45 kcal/mol in activation energy and is unlikely to be involved in the repair reaction of 6-4PP. Therefore, other possibilities need to be considered. In the crystal structure of 6-4 photolyase there is a water molecule present in the active site. It is possible that following electron and proton transfer to 6-4PP that attack of this external water could occur leading to carbinolamine formation. This type of mechanism would be able to avoid both the formation of an oxetane intermediate along with the loss of hydroxide following a proton transfer. However, experimental evidence shows that running the repair reaction in O¹⁸ labeled water that O¹⁶ is retained in the repaired thymine.⁶⁶ This suggests a direct transfer mechanism is involved in the reaction. It is an interesting question to debate why the water is present in the active site if it performs no obvious function? One possibility is that this water does attack the C3' position forming the carbinolamine but upon dehydration the water is not released from the active site. In

such a scenario the water formed from the dehydration reaction does not exchange with solvent leading to retention of O¹⁶ in the products. It should be noted that this proposal is completely speculative as there have been no experimental studies on the solvent exchange rates into and out of the active site of 6-4 photolyase. In an MD study by Christopher Harrison, a water was detected with 80% occupancy within 3.1 Å of C4'.⁷¹ This indicates that perhaps the 6-4 photolyase is able to “hold on” to this water for an extended period of time allowing for catalytic turnover of a water released in a dehydration mechanism.

The main idea behind this work is to alert those in the field that spurious mechanisms that are published without experimental or even computational data should be viewed with a critical eye. In the ten years following the publishing of the oxetane mechanism significant resources were spent trying to confirm a mechanism with little to no experimental evidence to back it. Furthermore, I believe we have shown that using simple concepts from organic chemistry one is able to rule all of these reactions out. Oxetanes are high energy species which are not likely able to be stabilized through enzymatic means. Hydroxide is a very poor leaving group requiring >30 kcal/mol in activation energy to depart. Therefore, I believe that people working in this field need to take a step back and reanalyze this repair process before the field spends another 10 years studying spurious mechanisms in which experimental evidence is absent.

CHAPTER 4:

EVALUATION OF A BIOMIMETIC ARTIFICIAL PHOTOLYASE ON CPD REPAIR WITHIN DUPLEX DNA

4.1 Introduction

As discussed in Chapter 1, the formation of pyrimidine dimers is likely the first step in a cascade that eventually leads to skin cancer. As the [2 + 2] cycloadditions leading to the two pyrimidine dimers, the cyclobutane pyrimidine dimers (CPD) and (6-4) photoproduct (6-4PP), are symmetry forbidden in the ground state, photoexcitation of the bases by UV-A or UV-B (260-320 nm) light is required (Figure 4.1). Several species in all three biological kingdoms are capable of cycloreversion of these photolesions to the original bases with enzymes specific to either type of photolesion. The first enzyme discovered was named DNA photolyase and is capable of cycloreversion of CPD photolesions.¹⁶ Just as the formation of the CPD photolesion is ground state forbidden, so is the repair. DNA photolyase circumvents this situation through the use of a redox active flavin adenine dinucleotide (FAD) which is able to inject an electron into the CPD effecting repair on the radical anion surface.

While many species possess CPD photolyase as a means of repair of CPD photolesions in DNA, humans do not. With the rates of skin cancer rising, it is of great interest to develop a biomimetic approach to repairing these photolesions in humans

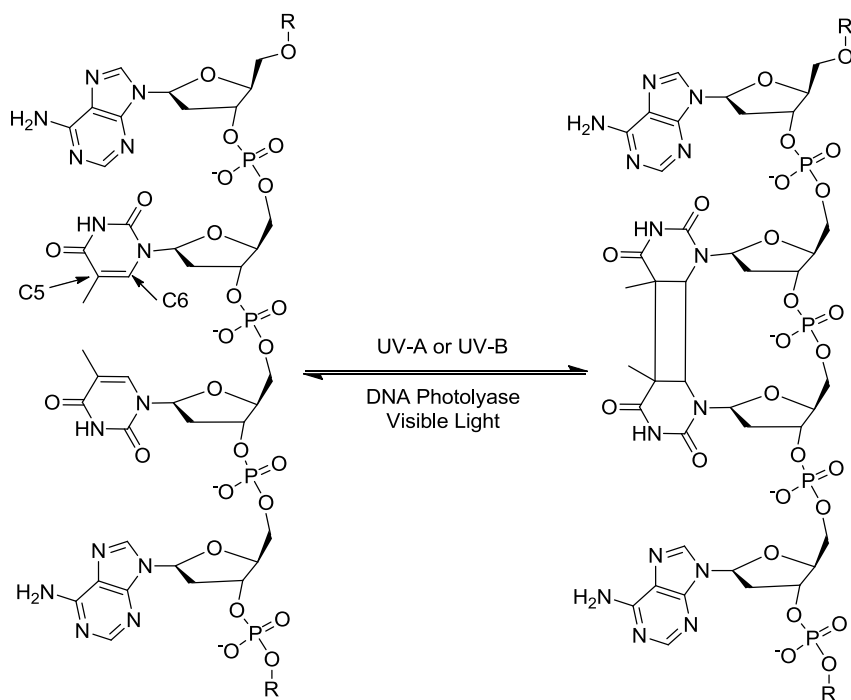


Figure 4.1- Formation of the CPD in DNA.

before it causes genomic damage. Several studies have looked into repairing CPDs with a noncovalent binding species. Selective recognition of thymine and uracil dimers in solution have been reported. Park et al. designed and synthesized a binder of a CPD derivative through the use of a xanthane-spaced diaminopyrimidine.⁷⁴ Association constants (K_a) of 4800 M^{-1} were achieved, however no repair of the CPD was reported. Goodman and Rose developed a macrocycle containing two preorganized diaminopyridine thymine binding moieties.^{75,76} K_a 's of 1.8×10^4 were observed with quantum yields of splitting of 0.11 in acetonitrile. However, when polar protic solvents such as methanol were used no repair is detected. The most likely explanation for this is that hydrogen bonding between the thymine dimer and methanol competes with binding to the recognition unit.

Carrell and coworkers have done much work in exploring the effects of placing flavin derivatives in the vicinity of formed CPDs. This is done by covalently linking the flavin groups at defined positions within a scaffold. Several important pieces of information can be gleaned from these studies. First, they investigated the effects of solvent polarity on splitting efficiency.⁷⁷ It was shown that splitting efficiency increases with increasing solvent polarity. Splitting in water was shown to proceed with an efficiency of 0.06 while in dioxane was shown to decrease splitting efficiency to 0.01. This result is consistent with the polar active site that was found in the complex between CPD containing DNA and DNA photolyase. They have also investigated the effect of covalently attaching a flavin chromophore directly to CPD containing DNA.^{78,79} In this study, they are able to demonstrate that electron transfer into the CPD through the DNA base stack. These studies lay the groundwork for artificial induction of CPD repair. However, as the flavin is covalently linked in all cases, no catalytic turnover is possible.

Work by Kimura et al. showed that through the use of a Zn(II) complexed macrocyclic tetramine (Figure 4.2), thymine and its derivatives can be selectively recognized in aqueous solution.⁸⁰ In this scheme, the N(3) of thymine is deprotonated and forms a close ion pair with the Zn(II) of the cyclen recognition unit. Two of the amine functionalities of the cyclen binding unit also form hydrogen bonds to imide carbonyl oxygens. Binding constants to thymine in water were found to be approximately 10^3 M^{-1} at pH 7.0.

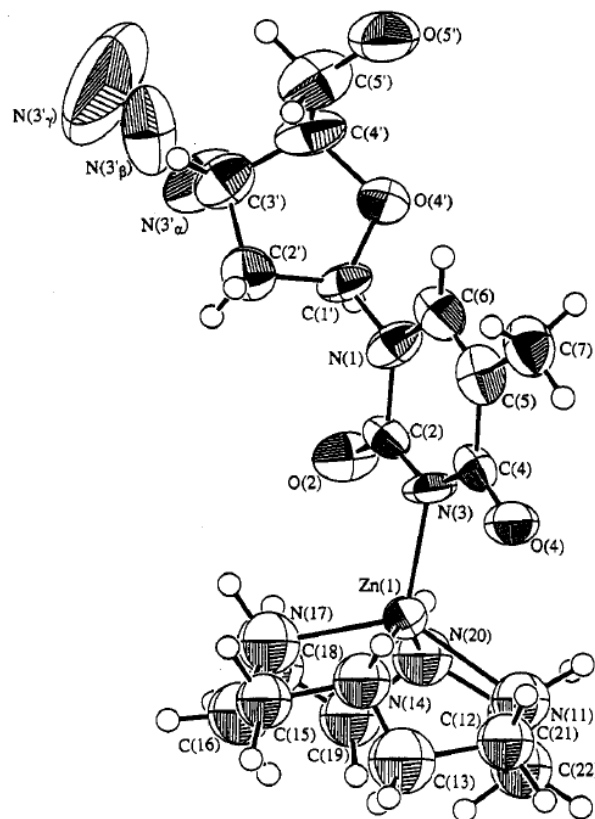


Figure 4.2- Crystal structure of Zn(II) cyclen bound to thymine nucleotide.⁸⁰

Our group envisioned covalently linking the Zn(II) cyclen moiety to a flavin derivative as a means to catalytically repair CPDs in solution. This biomimetic artificial photolyase can be seen in Figure 4.3. In a study by our group published in 2004 reported the repair of uracil dimer analogs at physiological pH.⁸¹ In water at pH 7.2 78% repair was observed upon irradiation with visible light for 10 seconds. As a load of 25 mol % of catalyst was used, catalytic turnover was also demonstrated. Beyond 10 seconds of irradiation, severe product inhibition was observed. This is thought to be caused by the differences in pK_a between the uracil dimer ($pK_a = 10.5$) and free uracil ($pK_a = 9.9$). As

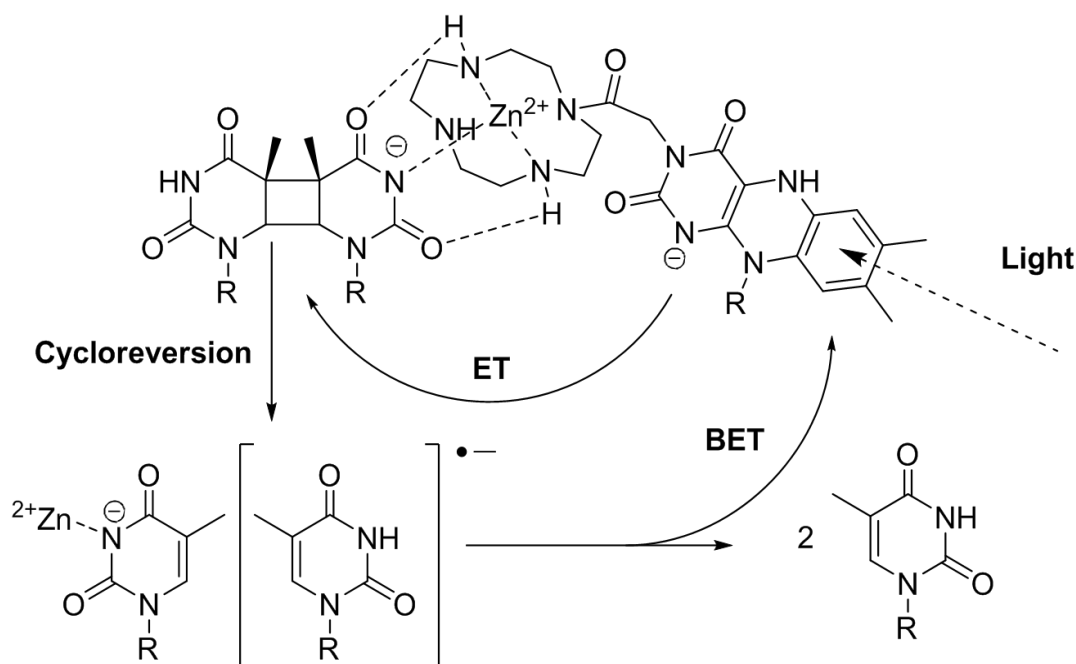


Figure 4.3- Scheme showing the function of the artificial photolyase. R- ⁿBu or (CH₂)₂O(CH₂)₂OCH₃.

the formation of the CPD breaks aromaticity of the bases, the imide hydrogen is less acidic. In work performed by Benjamin Hankeln, it was shown that the inclusion of a phosphate backbone between the two bases results in a slowing of the photoreversion reaction. This could be caused by one of two things. The inclusion of the phosphate backbone changes the electronics of the system due to the negative charge present. It follows that it is more difficult to inject an electron into a negatively charged molecule in comparison to a neutral one in analogy to the previous experiment. Also, the artificial photolyase could have some binding affinity to the negatively charged backbone leading to a greater distance between the flavin and the photodimer effectively slowing the reaction. However, some repair is observed and it was decided to test the effectiveness of the artificial photolyase on duplex DNA.

In the crystal structure of the complex of CPD containing DNA and DNA photolyase it can be seen that the CPD is in a flipped-out conformation.⁸² This finding shows that there is an energetically accessible conformation in which the thymine dimer is solvent exposed and available to bind our artificial photolyase. A computational study by O'Neil and Wiest predicted that the flipped out conformation of a CPD photolesion is approximately 5.25- 7.50 kcal/mol endothermic.⁸³⁻⁸⁵ It follows that if this conformation can be captured and repaired in solution, upon an unbinding event the reconstituted thymine can flip back into the duplex and effectively be removed from the equilibrium. This use of LeChatlier's Principle may alleviate the product inhibition which was evident in the uracil dimer model studies.

4.2 Experimental

As CPD containing DNA is not commercially available, it must be synthesized. Mu and coworkers developed a method to increase yields of thymine dimers upon irradiation of a T-T step containing ssDNA.⁸⁶ By adding acetone as a photosensitizer yields of 32% were obtained. This required 20% acetone concentration by weight, a ssDNA concentration of 150 μ M, and a irradiation time of one hour. This method of making CPD containing DNA has several advantages. Synthesis in the usual sense is not required from our side, but rather, the ssDNA is obtained from Integrated DNA Technologies and the solutions are made and subsequently irradiated. Also, separation of CPD containing from undamaged DNA is relatively straightforward using standard HPLC DNA separation techniques.

The single strand sequence 5'-ACGGCAGTTGATACGT-3' was obtained and stock solutions were made in Sodium Chloride-Tris-EDTA (STE) buffer. Buffer and acetone was added resulting in 150 μ M concentration of ssDNA and 20% acetone by weight. This solution was then placed into an NMR tube and degassed by sparging with argon gas. The NMR tube was then sealed and placed into the photolyzer for irradiation. A 400 W Hanovia medium-pressure mercury emersion lamp equipped with a pyrex filter was used and irradiation times of \approx 45 minutes produced the highest yields.

The photodamaged strand was isolated using preparative HPLC. A Waters Xterra column was used along with a Waters 600 controller, Waters 600 pump, Waters 996 PDA detector, and a Waters 717plus autosampler to perform the separation. Analysis of the HPLC traces was performed using the Empower software program. Conditions for the separation are as follows; mobile phase A- 5% acetonitrile/ 95% 0.1 M triethylammonium acetate buffer (TEAA), mobile phase B- 15% acetonitrile/ 85% 0.1 M TEAA. Both mobile phases are adjusted to pH 7.0 using acetic acid or triethylamine. It should be noted that the triethylamine was purified by distillation prior to making the TEAA buffer. A flow rate of 1mL/min was used and the gradient was changed from 80% A/ 20% B to 42.5% A/ 57.5 B over 50 minutes. A column temperature of 60°C was maintained through the use of a water or oil bath. Baseline separations of up to 5 minutes were possible using this technique. Yields typically varied between 20- 30% and provides us with ample amounts for study. A representative HPLC trace of this solution after irradiation can be seen in Figure 4.4. The strand containing the CPD elutes around 8 minutes and the undamaged strand around 14 minutes. Upon collection of this peak it was analyzed using ESI-MS. This analysis shows a molecular weight of 4917.8 AMU

(Figures 4.5 and 4.6) for both the damage and undamaged strands of DNA. This is to be expected as one would not expect a change in mass upon cyclization. Furthermore, we have shown that the strands not only have the same mass, but also possess different retention times on HPLC.

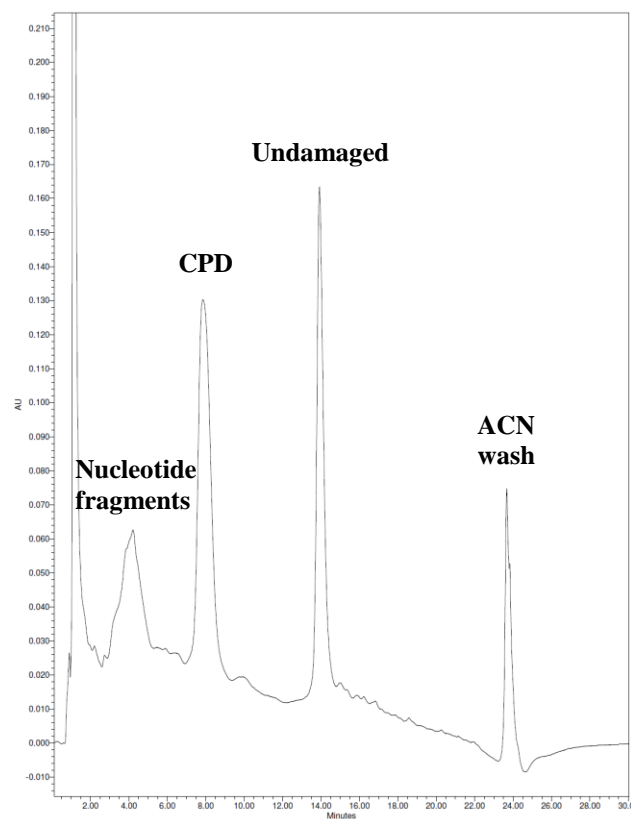


Figure 4.4- HPLC trace of damage solution after irradiation. CPD containing strand elutes around 8 minutes while the undamaged strand elutes around 14 minutes.

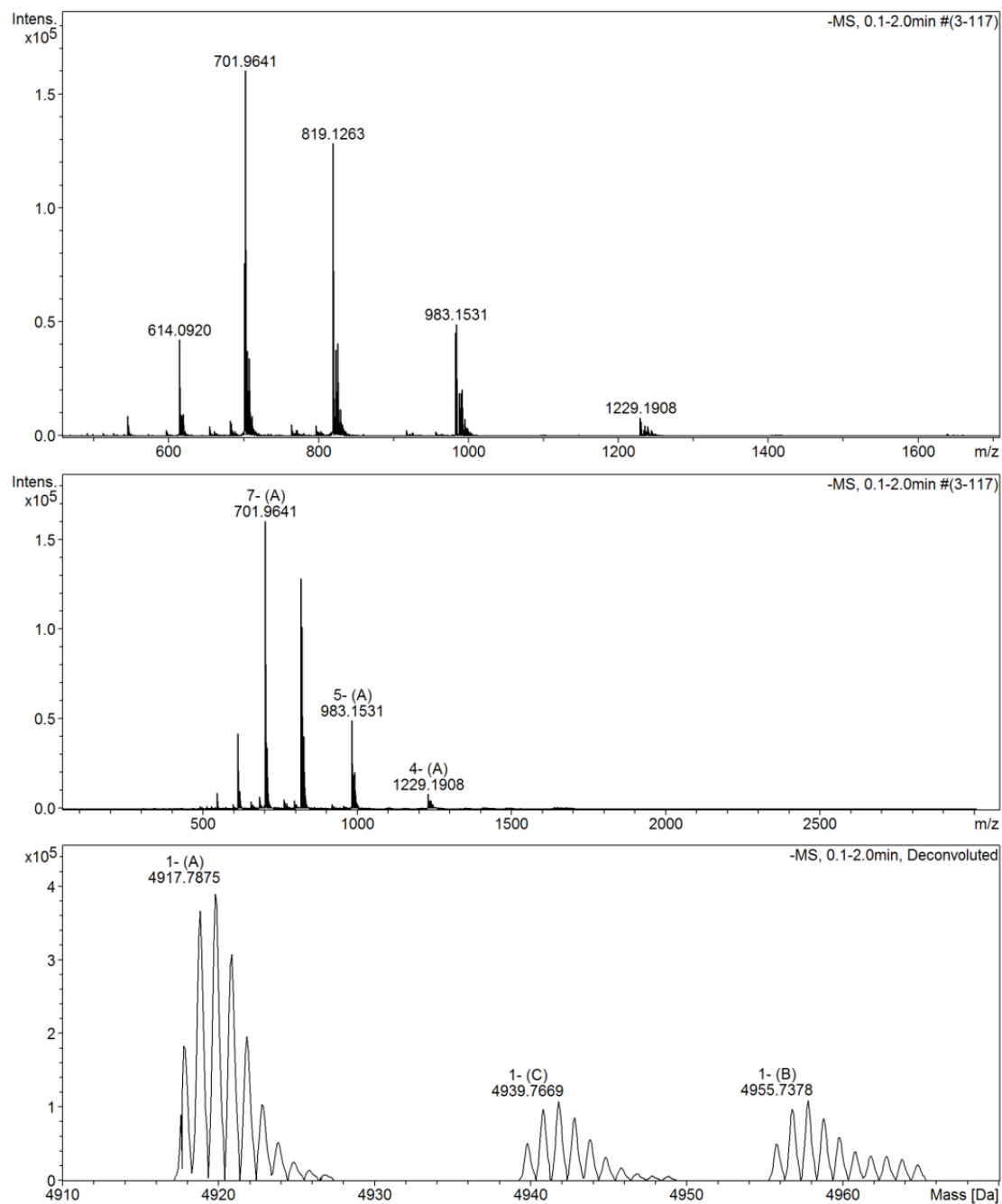


Figure 4.5- ESI-MS of the CPD containing ssDNA.

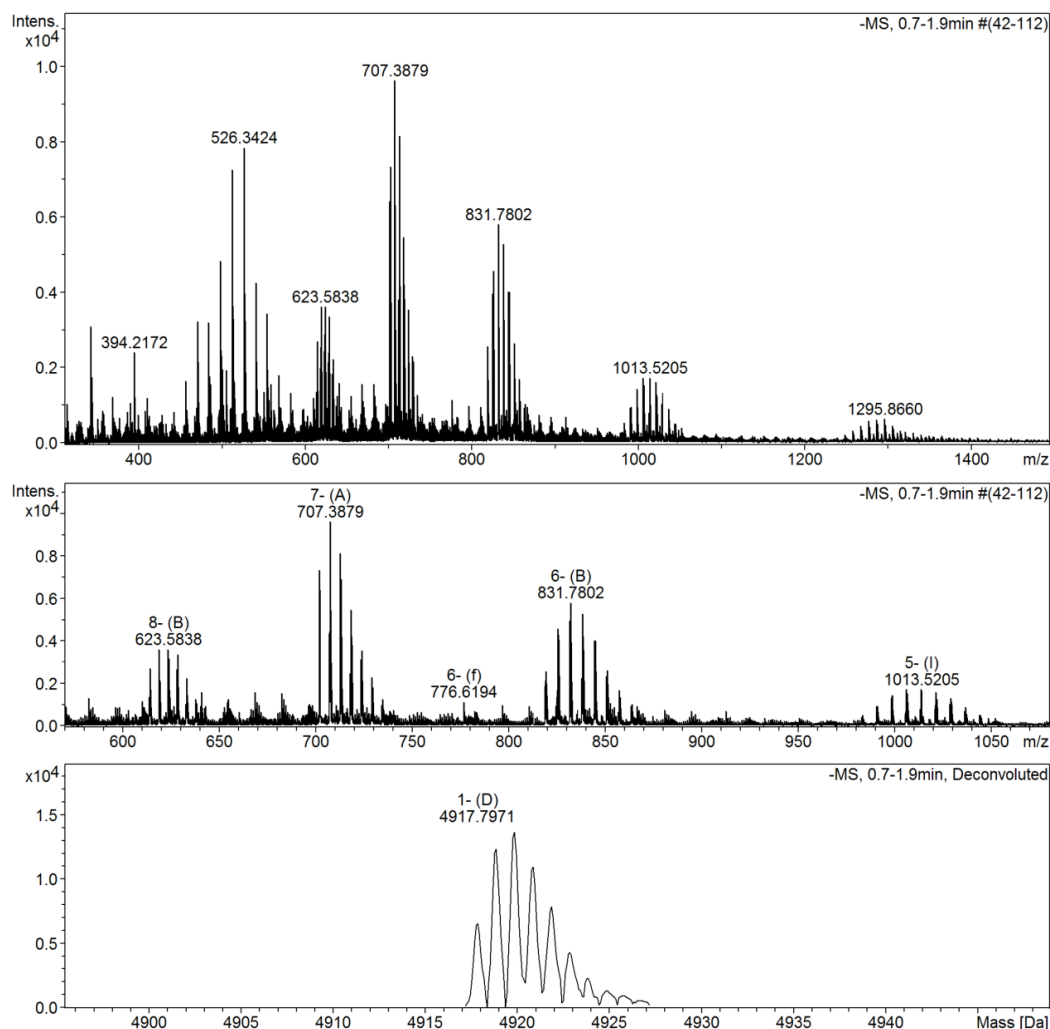


Figure 4.6- ESI-MS of the undamaged DNA.

As the solutions that are used in the damaging reaction are very dilute, there is a requirement to develop a method to take large amounts (≈ 10 mL) of our product mixture and concentrate it down prior to separation. To do this we used Amicon Ultra-4 centrifugal filters which retain high molecular weight molecules (>3000 AMU) while allowing solvent and salt ions to pass through. This allows us to concentrate a many mL solution down to <200 μL allowing for our separations to proceed much faster. One issue that must be considered is overloading the column. As more molecules of ssDNA

are interacting with the C-18 bed of the column, separation decreases and retention times decrease. To circumvent this, longer flatter gradients (ramped over 30 minutes instead of 15) were used in the separation program. The first large peak was identified as the CPD dimer by ESI-MS analysis and coinjection with the undamaged strand. A reinjection of the collected sample from this HPLC separation can be seen in Figure 4.7. Integration of the peaks in the HPLC trace show that the strand is >99% pure.

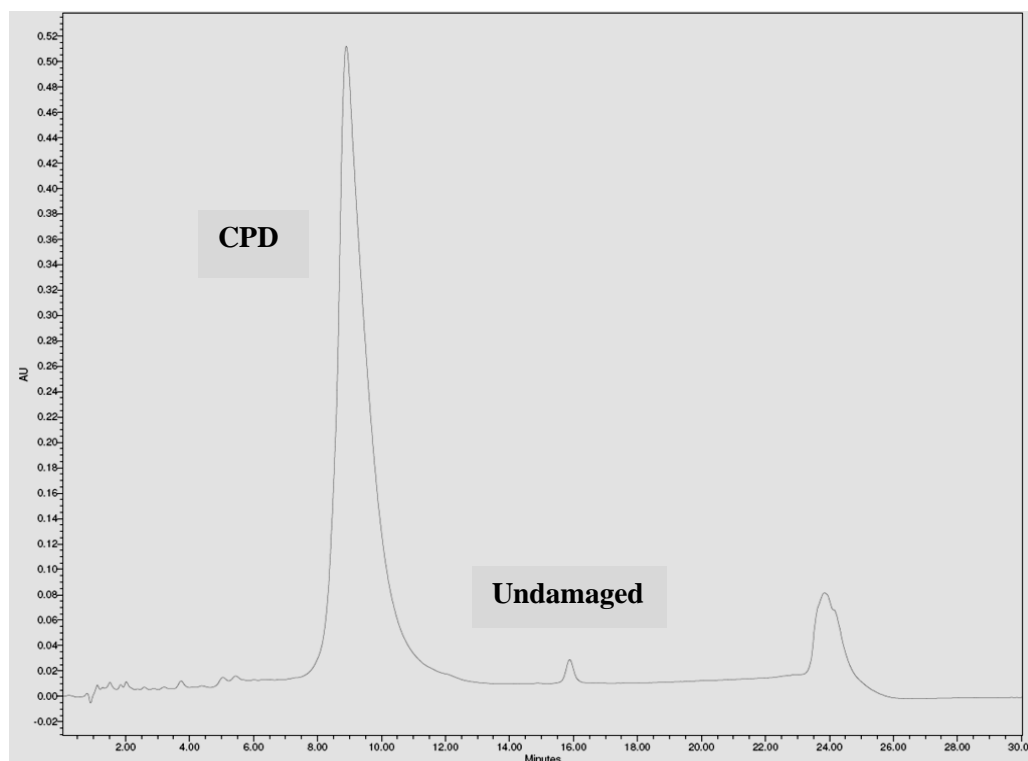


Figure 4.7- HPLC trace of collected CPD containing DNA.

With pure ssDNA containing a CPD in hand, we performed a solvent exchange in order to remove acetonitrile and TEAA buffer. The solution was spun down again using the Amicon Ultra-4 centrifugal filters and an ultracentrifuge at 5000 rpm. Once concentrated, the centrifuge tube was refilled with annealing/ storage buffer and spun down again. This is repeated several times to exponentially reduce the concentration of

our HPLC mobile phases. The end product was pure, concentrated ssDNA containing a CPD photolesion in annealing buffer (Figure 4.7).

Next, the concentration of the ssDNA was determined using UV-Vis spectroscopy and measuring the OD₂₆₀. From here, the CPD containing strand along with its complimentary strand are mixed in an equimolar manner. The solution was then heated slowly to 100°C and cooled to room temperature to facilitate annealing. This results in a stock solution of CPD containing duplex DNA which was used to test our artificial photolyase.

Repair of the photodamage was attempted in much the same way as the model systems.⁸¹ The repair solution consisted of 5 µM DNA, excess sodium dithionate, TEA to raise pH to ≈7.2, 25 mol % of the artificial photolyase, and water to normalize concentrations. This solution was degassed under argon for 30 minutes prior to photoreactivation. The experimental set-up for the repair reaction can be seen in Figure 4.8. Different lamps were tried in the repair reaction, which will be discussed along with the results.

4.3 Results

Upon irradiation with our 400 W Hanovia medium-pressure mercury lamp we saw almost complete repair (≈90%) of the CPD photolesions in 45 minutes. HPLC plots of this repair can be found in Figure 4.9. A plot of amount of repair versus time can be

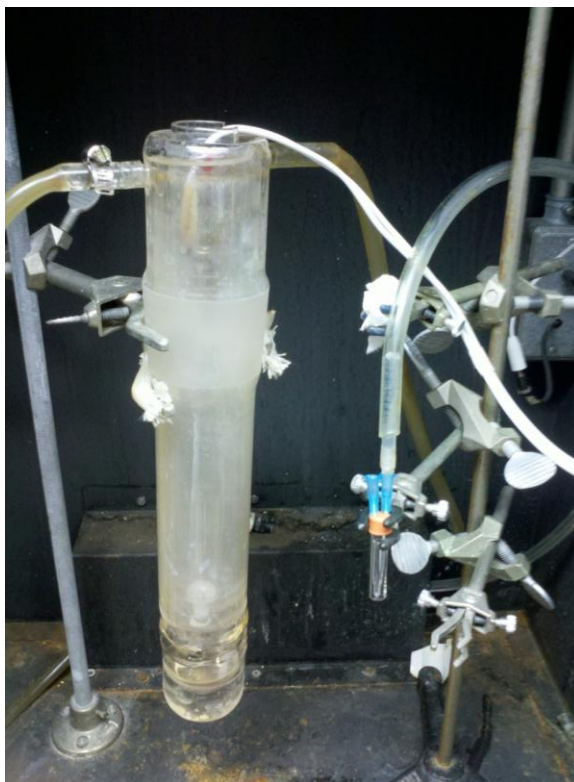


Figure 4.8- Experimental setup for the repair reaction. Repair solution is contained in the vial on the right and is under an argon atmosphere.

seen in Figure 4.10. To confirm these findings the same reaction was performed in the absence of the artificial photolyase. Interestingly, similar results were observed without our artificial photolyase (Figure 4.11). With the failure to show repair using this methodology, we attempted repair using a Rayonet lamp which emits light of the wavelength 350 ± 20 nm. This lamp is also much less powerful so longer irradiation times were utilized. Figures 4.12 and 4.13 depict the results from the irradiation of the repair solution for 2.5 hours with and without the artificial photolyase. In this case, we saw an increased repair of the CPD in the presence of the artificial photolyase relative to the

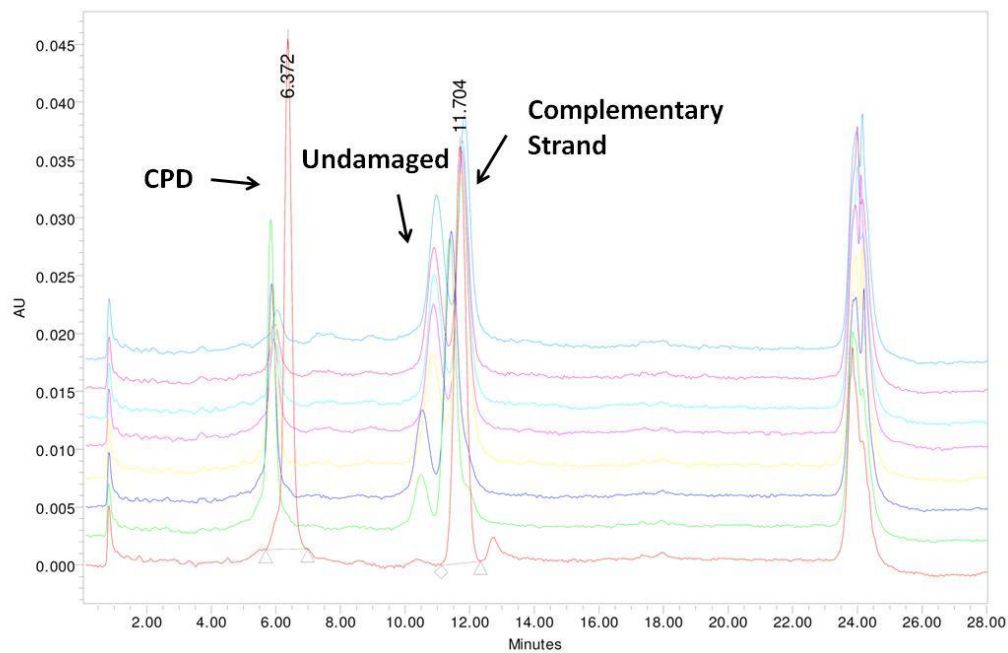


Figure 4.9- Repair reaction containing artificial photolyase.

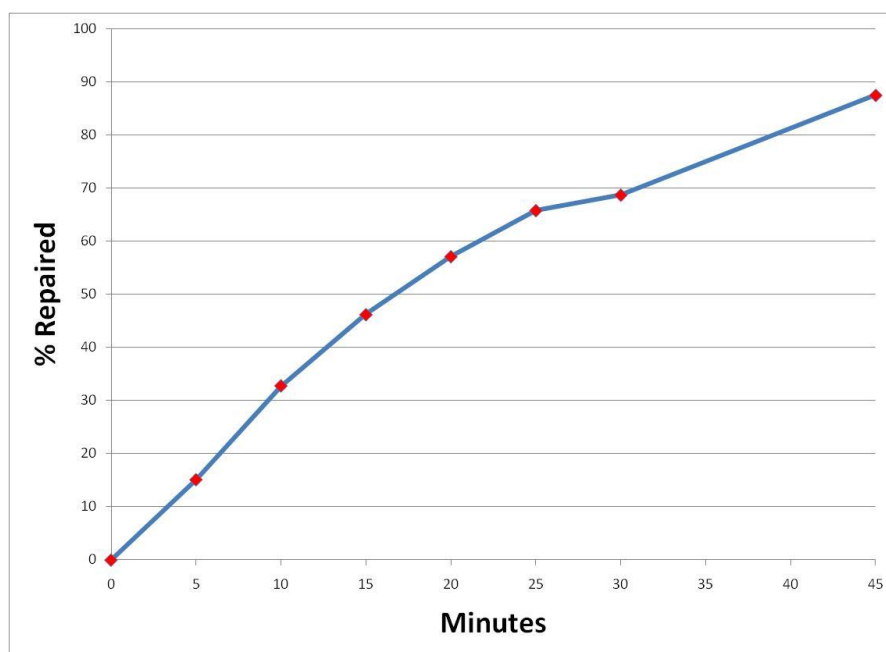


Figure 4.10- Repair versus time for repair reaction.

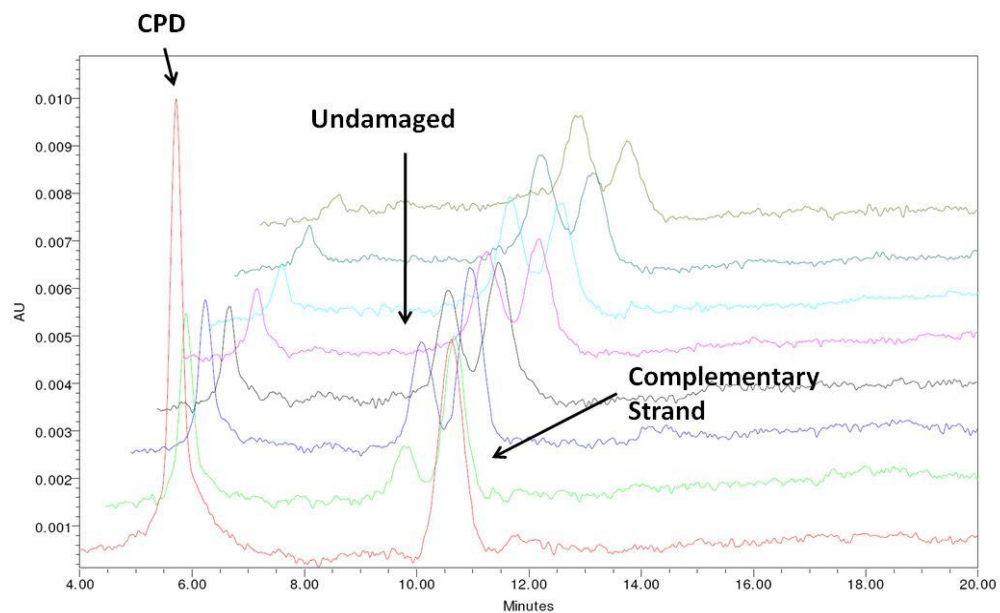


Figure 4.11- Control reaction without artificial photolyase.

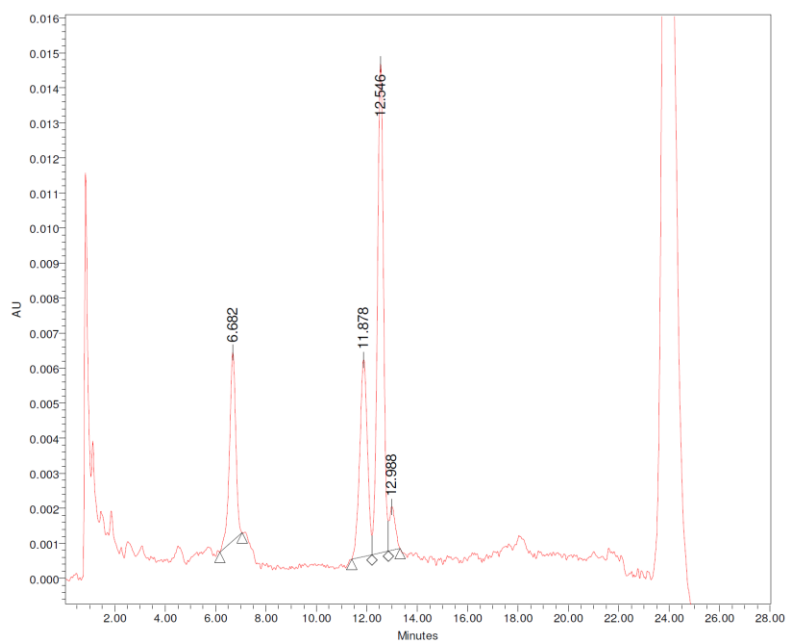


Figure 4.12- Repair reaction using the Rayonet lamp.

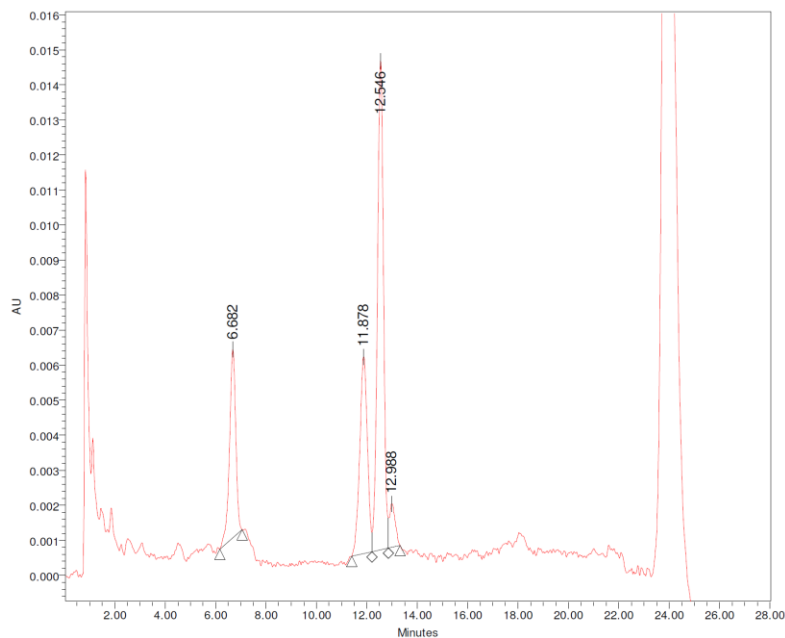


Figure 4.13- Control reaction (no artificial photolyase) using the Rayonet lamp.

control experiment. 56.3% repair was observed for the solution containing the artificial photolyase while the control reports 47.2% repair. This difference indicated that some binding of the artificial photolyase taking place along with the associated electron transfer leading to repair. However, one would expect no repair to be observed without the flavin cofactor present.

Holman et al. showed that guanine bases flanking a T-T step influenced the rate of formation of CPDs.⁸⁷ Their research showed that when in the excited state, a guanine may be capable of electron transfer through the base stack effecting repair. However, it is interesting to note that although our sequence contains two flanking guanines about the T-T step, this electron transfer does not seem to affect our CPD yields. This could be due to the added photosensitizer (acetone). In the absence of photosensitizers, the duplex containing the CPD does display self-repair in the control experiments. Therefore, future

experiments should focus on sequences without flanking guanines. A study by O'Neil and Wiest predicted the energies required for a CPD to flip out of the duplex.⁸³ It was shown that the energy required for this base flipping to occur was dependent on the flanking bases. CPDs flanked by thymines (5.25-5.5 kcal/mol) required the least amount of energy followed by adenine (6.25-6.5 kcal/mol), cytosine (6.5-6.75 kcal/mol), and guanine (7.2-7.5 kcal/mol). This is important because it gives us some insight on another possible problem with our experimental setup. Since our sequence has guanines flanking our formed CPD, it requires more energy to flip out of the duplex. This effectively reduces the number of CPDs flipped out in solution affecting the ability of the artificial photolyase to bind.

In conclusion, this study demonstrates an ability to form CPDs in ssDNA, concentrate and separate out the damaged strand, perform repair reactions and monitoring the reaction over time, and providing further evidence that self-repair inside the DNA duplex is likely affected by flanking guanine nucleobases. Further studies in the Wiest group will attempt to change the sequence to alter the base-flipping qualities and hinder the self-repair process seen in this study. This will allow for more binding of the artificial photolyase and reduce the noise in the experiment caused by self-repair.

4.4 Future Outlook

In the next setup, steps can easily be taken to see if we can decrease the energy of base flipping while minimizing self-repair within the duplex. Taking this into consideration adenines flanking the T-T step would be a good choice. This would provide the second lowest base flipping energy, the lowest rates of self-repair, and no

side reactions to complicate our separation as would be the case if thymine or cytosine were used. If flipping energies still prove too high to overcome CPDs could be placed across from abasic subunits eliminating hydrogen bonding across the duplex and reducing the base flipping energy. A possible drawback of this method would be increased rates of base flipping for the repaired duplex leading to the product inhibition observed in the early model studies by our group.

CHAPTER 5:

MD SIMULATIONS OF GLYCOL NUCLEIC ACIDS

5.1 Introduction

Glycol nucleic acid (GNA) is one of many synthetic nucleic acid analogs that have been successfully synthesized. Over the past 20 years, through the use of synthetic organic techniques chemists have extensively modified the backbone of DNA and RNA in an attempt to push the limits of nature's chosen information storage molecules. Examples of this include threose nucleic acid (TNA)²¹, locked nucleic acid (LNA)⁸⁸, peptide nucleic acids (PNA)⁸⁹, and now GNA. While many analogs have been reported in the literature, very few actually form duplexes. The structures of these analogs are shown in Figure 5.1. Each example is specifically synthesized in order to test our understanding of both the structure and/ or dynamics of nucleic acids. With each new nucleic acid that is studied, we learn more about why evolution chose DNA and RNA in particular as the information carriers for life as we know it. The ideas behind the synthesis of each of these nucleic analogs are discussed in Chapter 1.

GNA was specifically made in order to test how simple a nucleic acid can be while retaining base pairing common to DNA and RNA. For years, it was thought that six bonds per backbone (from 5'O to 3'P) were necessary to form a stable duplex.²¹ However, pioneering work from the Eschenmoser laboratory was able to show that the

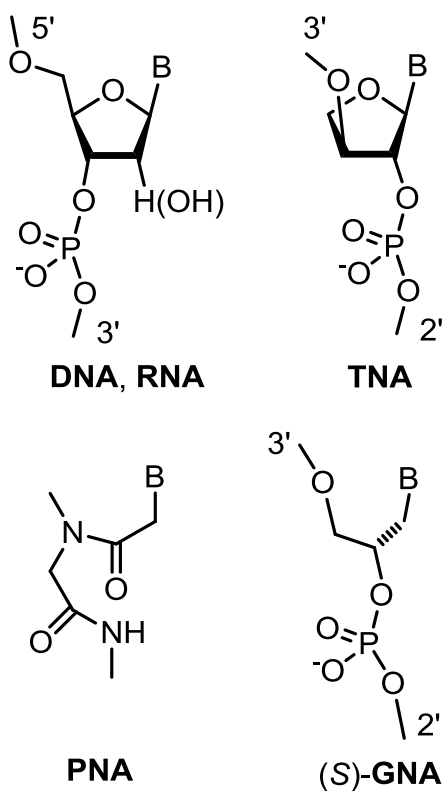


Figure 5.1- Structures of common natural and synthetic nucleic acids.

“six-bonds-per-backbone-rule” could be violated.²¹ Their synthesis of (L)- α -threofuranosyl oligonucleotide (TNA) demonstrated that having just five bonds between vicinal phosphate group can still result in a stable duplex. Inspired by Eschenmoser’s work, the Meggers group envisioned the removal of the 4’CH₂O group of TNA resulting in a simpler acyclic propylene glycol backbone.³⁵ Surprisingly, GNA has been shown to form highly stable, antiparallel duplexes far exceeding the thermal stabilities of other known nucleic acids, natural or otherwise.³⁹ The stabilities of (S)-GNA duplexes along with their DNA counterparts are shown in Table 5.1. It can be seen that in all cases, the

TABLE 5.1
THERMODYNAMIC AND THERMAL STABILITIES OF (S)-GNA AND DNA
DUPLEXES (IN PARENTHESIS).³⁶

Entry	Sequence ^[a]	T _m [°C]	ΔG (298 K) [kcal/mol]	ΔH [kcal/mol]	-TΔS (298 K) [kcal/mol]
1	CACATTATTGTTGTA	70	-21.1	-95.6	74.4
	GTGTAATAACAACAT	(47)	(-15.2)	(-103.3)	(88.1)
2	AATATTATTATTTTA	59	-16.2	-77.0	60.8
	TTATAATAATAAAAT	(41)	(-12.4)	(-86.5)	(74.1)
3	CGAATTCG	54	-12.2	-54.0	41.8
	GCTTAAGC	(36)	(-9.4)	(-59.4)	(49.9)

[a] Upper strand in 3' to 2' direction for GNA, 5' to 3' for DNA.

GNA analogs are more thermally stable than the corresponding DNA sequences. In many cases, the T_m measured for GNA exceeds that of the same sequence of DNA by more than 20°C. This is a surprising result considering the acyclic nature of the backbone. Previous attempts at creating stable non-natural nucleic acid duplexes focused on reducing the conformational space of the single strand (LNA) or reducing electrostatic repulsion of the backbone (PNA). GNA, on the surface, accomplishes neither of these goals. One would think by having an acyclic backbone that the single strand would be highly flexible resulting in a large entropic penalty upon duplex formation. Also, since it

is composed of a phosphodiester backbone, electrostatic repulsion of the negatively charged backbone should be comparable to DNA.

A 2009 study by Meggers et. al examined the thermodynamics of GNA duplex formation was published.³⁶ In this work three sequences of GNA and DNA were studied in terms of T_m , enthalpy, and entropy of annealing. The results from this study are shown in Table 5.1. While it was already known that GNA forms stable duplexes, the thermodynamic parameters behind this were unknown. Surprisingly, and perhaps counterintuitively, GNA achieves its thermal stability by reducing its entropic penalty associated with annealing. For example, entry 1 in Table 5.1 displays an entropic penalty of 74.4 kcal/mol for GNA while the same sequence of DNA has an entropic penalty of 88.1 kcal/mol for annealing. Duplex formation in GNA is, however, enthalpically less favorable than the same process in DNA. From entry 1 in Table 5.1 it can be seen that the enthalpy of duplex formation for GNA is -95.6 kcal/mol while in DNA it is -103.3 kcal/mol. In total, the free energy of duplex formation in this particular example is -21.1 and -15.2 kcal/mol, resulting in T_m 's of 70 and 47°C for GNA and DNA respectively.

The findings by Meggers et. al are certainly interesting. Perhaps the most striking results are found in the entropies associated with annealing of the GNA duplex. On the surface, one would predict that the GNA backbone would be considerably more flexible than the DNA backbone. As the backbone of GNA consists entirely of rotatable bonds while DNA contains cyclic, 5 membered deoxyribose rings along the backbone, it would be expected that a greater number of degrees of freedom would be locked out upon the annealing of GNA single strands when compared to DNA. Based on the experimental data, this is not the case. To test whether this may be caused by preorganization of the

single strand, circular dichroism spectra were obtained at different temperatures.³⁶ This is an effective way to test whether the conformation of the GNA single strand changes significantly upon heating. CD spectra of a representative single strand of GNA and DNA are shown in Figure 5.2. In this plot it can be seen that ssGNA displays significant Cotton effects upon heating from 15- 90°C. This indicates that ssGNA may have a defined helical structure which is lost upon heating. If this preorganized structure is similar in

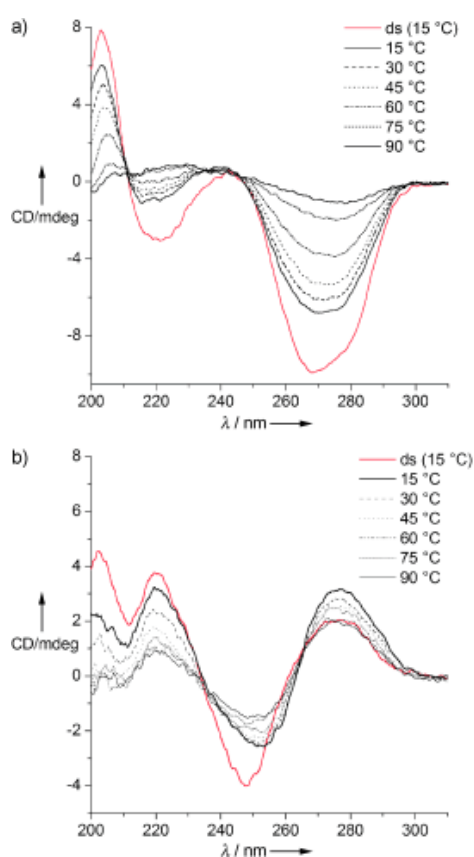


Figure 5.2- CD spectra of the melting of the a) GNA and b) DNA single strands.³⁶

conformation to individual strands in the duplex, this can explain the lesser entropic penalty upon annealing in comparison to DNA. However, several other possibilities could be possible. For example, the Cotton effects could be caused by the formation of a globular conformation of the ssGNA. If this is the case, one can imagine that the decreased entropic penalty may arise from retained flexibility of the duplex. To study this experimentally would be nontrivial. Theory, on the other hand, may provide an attractive avenue to study the dynamics of the GNA duplex.

Over the past several years, the Meggers group has developed a repertoire of structural data on GNA through the use of x-ray crystallography. While “native” GNA so far has not lent itself to crystallization, there has been success in obtaining structural information of several analogs. The first published structure of a GNA duplex was disclosed in 2008.³⁸ In this example, the self-complementary strand 3'-CGHATHCG-2' (referred to as 8mer-CuGNA) was used to obtain high quality crystals for X-ray crystallography. The GNA nucleotide H refers to hydroxypyridone nucleobase which forms highly stable homobase pairs upon chelation of copper(II) ions. This is done to provide two heavy atoms to the duplex to help in the crystallization and help in the phasing of the crystallographic data. This results in crystals which diffract to 1.3 Å.

The overall structure of the 8mer-CuGNA duplex can be seen in Figure 5.3 and is a unique structure. While the well known B-DNA duplex consists of a double helix type structure with major and minor grooves, the GNA duplex might best be described as a “helical ribbon loosely wrapped around the helical axis”.³⁸ There is one major groove which corresponds to the canonical minor groove and there are two distinct surfaces

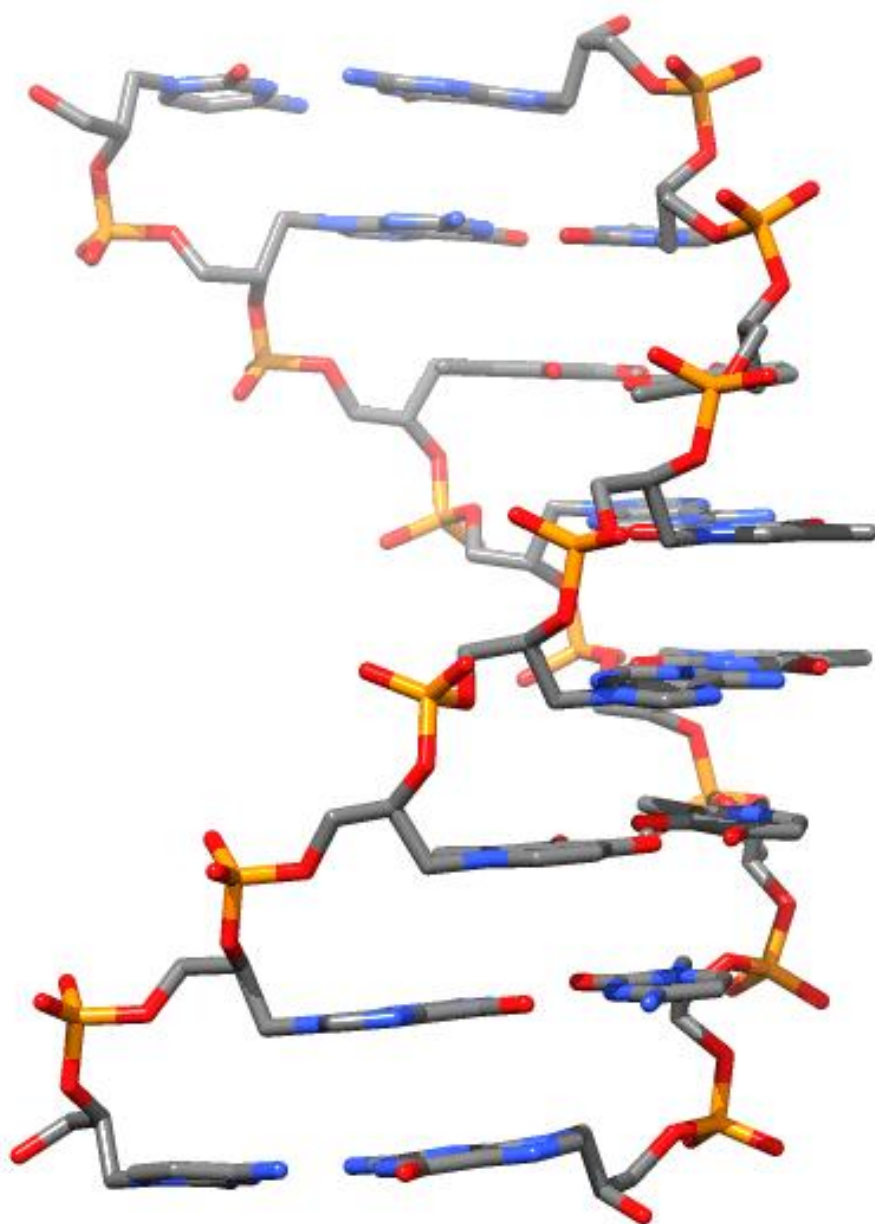


Figure 5.3- Crystal structure of 8mer-CuGNA.³⁸

TABLE 5.2
COMPARISON OF AVERAGE HELICAL PARAMETERS FOR GNA DUPLEX
CRYSTAL STRUCTURES IN TO B-DNA AND A-DNA.^{90a}

	8mer- BrGNA (N-type)	6mer- BrGNA (N-type)	8mer- CuGNA (M-type)	B-DNA	A-DNA
Helical sense	right	right	right	right	right
Residues per turn	10	10	16	10	12
Helical pitch (Å)	29	26	60	34	34
Helical rise (Å)	2.9	2.6	3.8	3.4	2.9
x-displacement (Å) ^b	-5.8	-6.0	-5.4	0.1	-4.2
Tilt (°) ^c	0.0	0.5	0.0	-0.1	0.1
Roll (°) ^c	3.6	6.4	-2.8	0.6	8.0
Twist (°) ^c	35.8	35.7	22.9	36	31
Slide (Å) ^c	-3.2	-3.4	-3.4	0.2	-1.5
P-P distance (Å) ^d	5.5	5.4	5.4	7.0	5.9

^a Data for GNA were calculated using the program CURVES.^{91,92} ^b Global inter-base pair parameter. ^c Global base pair-axis parameter. ^d Local inter-base pair step parameters. ^e Average intra-strand P-P distances.

present, concave and convex. All natural nucleobases are involved in base pairing in Watson-Crick fashion. While specific base pairing is comparable to DNA, that is where the structural similarities end. The helicoidal parameters derived from the 8mer-CuGNA as well as those from canonical B-DNA can be found in Table 5.2 (descriptions of helicoidal parameters can be found in Figure 5.4). The CuGNA duplex is significantly

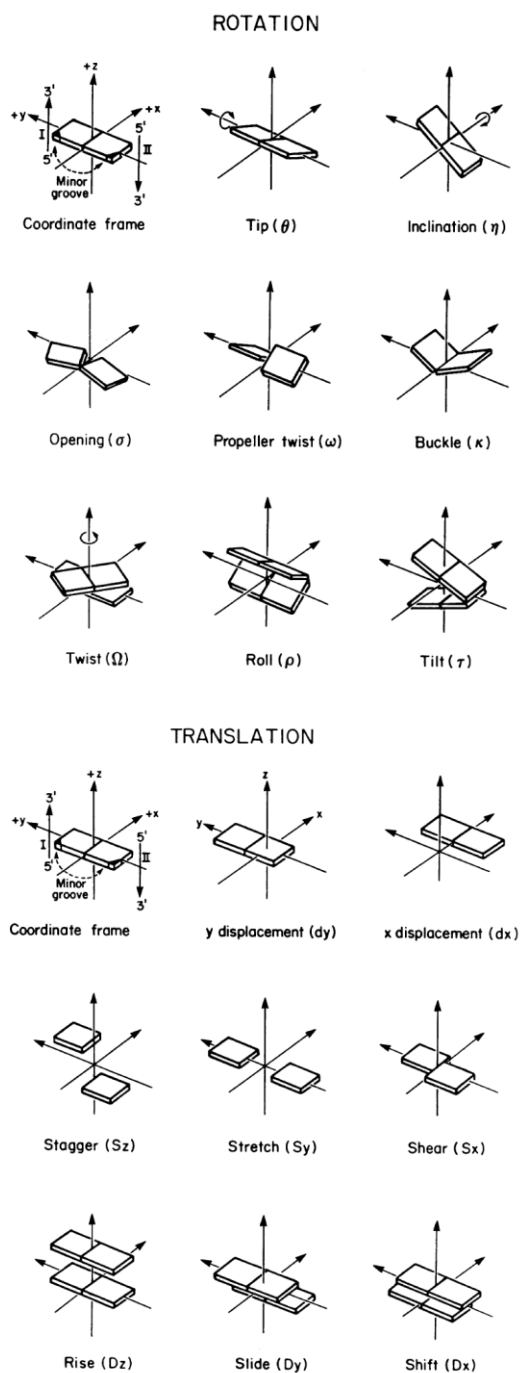


Figure 5.4- Descriptions of helicoidal parameters describing nucleic acid structure.⁹²

less twisted than B-DNA (22.9° and 36° respectively) leading to CuGNA possessing 16 base pairs per turn compared to 10 in B-DNA. This effect is most likely derived from the fact that GNA has fewer bonds per backbone than DNA causing an unwinding in order to retain a similar distance between the base stacks. Perhaps the most interesting difference between GNA and DNA is the large average slide between neighboring base pairs. GNA displays a large average base pair slide of -3.4 \AA compared to DNA's of 0.2 \AA . This is caused by the large backbone-base inclination of $42\text{--}50^\circ$. In turn, this results in GNA adopting a completely different mode of base stacking than what is seen in B-DNA. In

GNA interstrand base stacking is predominant (Figure 5.5) with significant π - π interactions being seen when both 2' bases involved are purines. This arrangement of base stacks likely contributes to the observation that GNA duplex formation is less exothermic than its DNA counterpart. In intrastrand base stacking as is seen in DNA, each nucleobase stacks with adjacent bases on both 3' and 5' side. With the large backbone-base inclination inherent to GNA, it is only possible for each base to stack with

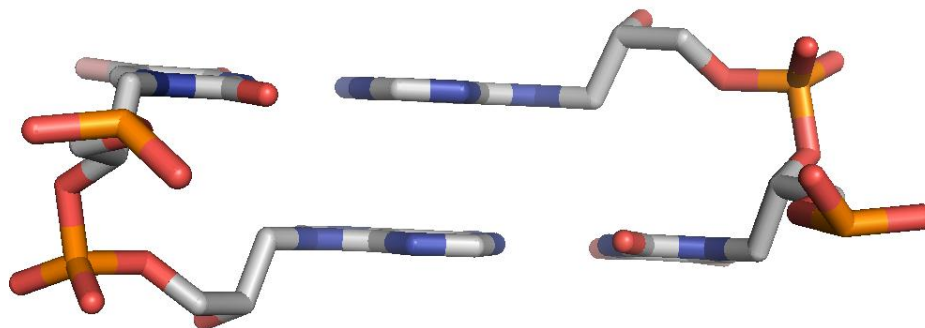


Figure 5.5- View of interstrand base stacking found in GNA.

one partner. Furthermore, when the two 2' bases in a base stack are pyrimidines, there is little to no hydrophobic overlap. The reduction of hydrophobic stacking therefore is the most likely cause of the decreased exothermicity of GNA annealing in comparison to DNA.

In addition to the helicoidal parameters previously discussed, the backbone of GNA also displays some interesting properties. In the 8mer-CuGNA crystal structure it can be seen that the conformations involving the vicinal C-O bonds (dihedral involving O2'-C2'-C3'-O3') vary. In the hydroxypyridone nucleotides this dihedral adopts an *anti* configuration while the nucleotides of Watson-Crick base pairs maintain a *gauche* (g- to be exact) configuration. A view of this dihedral can be found in Figure 5.6. This observation can be interpreted in different ways. One interpretation is that the differences are caused by the increased size of the hydroxypyridone base pairs, forcing the backbone out from the helical axis resulting in a rotation about the O2'-C2'-C3'-O3'. It also sheds some light upon the flexibility of the backbone. As would be expected, this indicates that

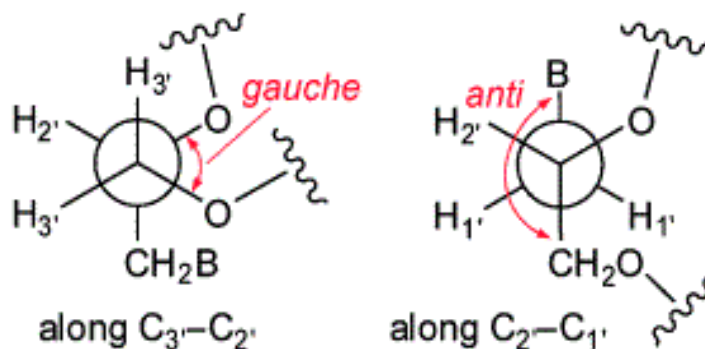


Figure 5.6- View of the dihedrals discussed in this work.

the backbone is indeed flexible and is capable of adopting different conformations. Later work would further confirm this hypothesis and will be discussed later.

While the 8mer-CuGNA structure provides some information about the structure of “native” GNA, there are some obvious drawbacks derived from the use of hydroxypyridone base pairs. First, the hydroxypyridone base pair is approximately 12.7 Å in width across the duplex from C1' to C1'. This is 2 Å wider than a G-C base pair within the same duplex. While this may not seem like a significant distance, it does cause significant changes in the backbone conformations around the hydroxypyridone base pairs as discussed previously. Furthermore, the introduction of the copper(II) chelating hydroxypyridone base pairs results in an increase in T_m from 40°C (replaced with A-T base pairs) to 78°C. Most likely this is caused by difference in energy between the hydrogen bonds typically found in Watson-Crick base pairs and the copper(II)-oxygen bonds found in the hydroxypyridone base pairs. The strength of such an interaction may have significant consequences for the structure and dynamics of the duplex. Also, the atomic radius of copper(II) is slightly larger than that of carbon, oxygen, and nitrogen so there is also an increase in size for these base pairs along the helical axis.

Further work by the Meggers group lead to a new way to crystallize GNA duplexes. In 2009, they reported the X-ray structure of the 6mer self-complementary sequence 3'G^{Br}CGCGC-2' (6mer-BrGNA) with ^{Br}C representing a 5-bromocytosine analog (Figure 5.7).³⁴ By using a bromocytosine analog they are able to obtain crystals which diffract at 0.956 Å with, in theory, little to no structural perturbation to the duplex.

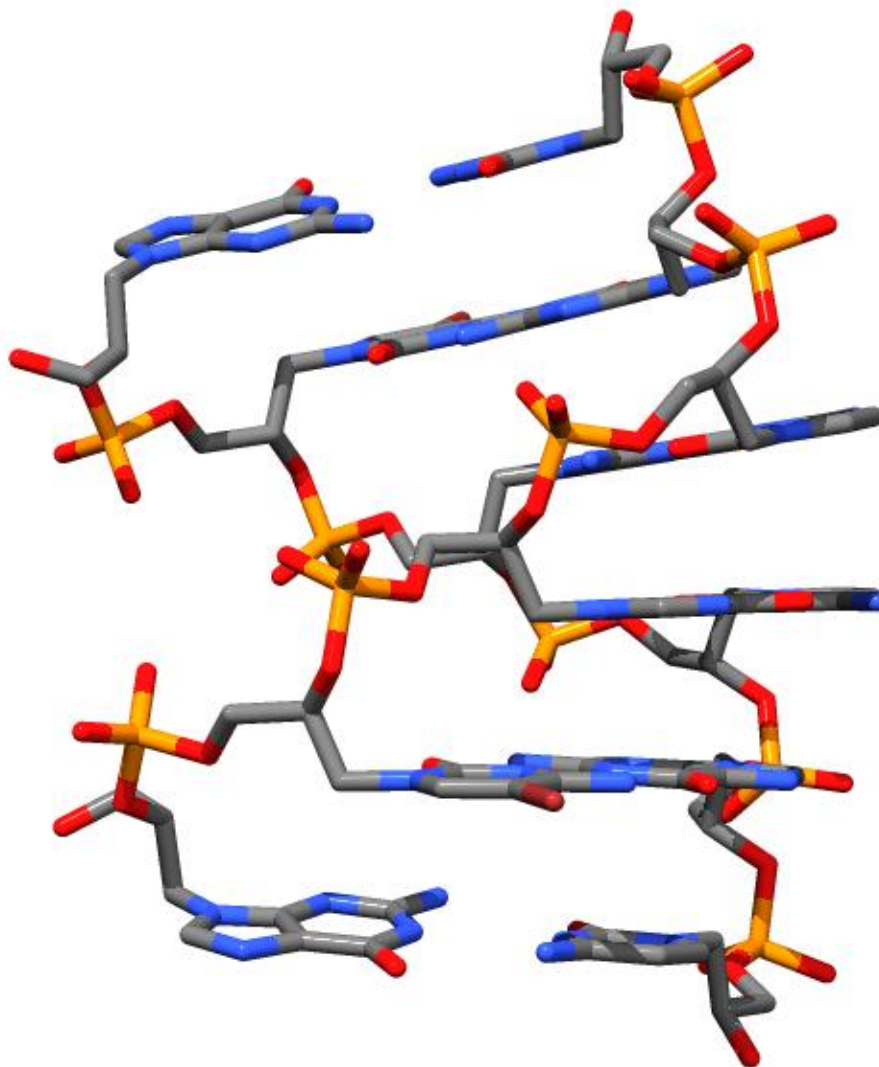


Figure 5.7- Crystal structure of 6mer-BrGNA.³⁴

Also, the new structure represents a structure in which Watson-Crick base pairing is used exclusively which is in stark contrast to the previous metallo-base pair example.

Structurally, there are similarities and differences between 8mer-CuGNA and 6mer-BrGNA. They both form a helical ribbon type structure wrapped around a helical axis. Furthermore, each also possesses a single large groove and two distinct curved surfaces (concave and convex). As far as base stacking is concerned, both display similar

properties due to the large, and nearly identical base pair slides (-3.5 \AA for 8mer-CuGNA and -3.4 \AA for 6mer-BrGNA). Again, an interstrand base stacking mode is utilized for both as a result of the large, negative base pair slide. Along with these similarities, there are also a few striking differences that can be observed. Most notably, 6mer-BrGNA is significantly compressed along the helical, or z-axis. Also, the average base pair twist shows that 6mer-BrGNA is significantly more tightly coiled than its 8mer-CuGNA counterpart (35.7° versus 23.5°) leading fewer base pairs per turn (10 versus 16) and a much shorter helical pitch (26 \AA versus 60 \AA). The conformations found in the backbone also differ significantly. As previously stated, 8mer-CuGNA nucleotides assume a *gauche* conformation of the vicinal C-O bonds of the backbone in the Watson-Crick nucleotides while the hydroxypyridone nucleotides assume an *anti* configuration. In 6mer-BrGNA, a significantly different situation is present. Interestingly, along the backbone of BrGNA this conformation alternates between *gauche* and *anti* through the entire length of the duplex. While the cause of this is not clear, similar phenomena have been observed in the crystallography of DNA. As short oligonucleotides have the shortcoming of possessing a high ratio of end-stacking base pairs vs. total bases, there is a prerequisite for the formation of crystals that some distortion of helical twist will be present compared to the same duplex in solution.⁹³ As there are no other readily apparent reasons for this behavior, this is the current working hypothesis.

In order to test whether the structural changes seen between the metallo-base pair GNA and the bromocytosine version was due to sequence dependence and/ or length, a new 8mer sequence ($3'\text{-CTC}^{\text{Br}}\text{UAGAG-2}'$, with $^{\text{Br}}\text{U}$ representing a 5-bromouracil nucleobase) was synthesized and subsequently crystallized (8mer-BrGNA).⁹⁰ These

crystals diffracted at 1.8 Å resolution and represents the third structure solved of GNA. Again, while this is not strictly the “native” form of GNA due to the inclusion of the 5-bromouracil nucleotide, one would expect little to no effects upon the overall structure of GNA to be caused by this inclusion. Comparing the structural data of 8mer-BrGNA to 6mer-BrGNA, only small differences can be gleaned. The two have nearly identical parameters across the board as can be seen in Table 5.2. Also, 8mer-BrGNA displays that same alternation between *gauche* and *anti* configurations of the vicinal C-O bonds present in its backbone. Having a stretch of four pyrimidine nucleotides compared to the alternating purine-pyrimidine sequence utilized in 6mer-BrGNA this demonstrates that there is not an intrinsic preference for purines to adopt the *gauche* conformation. This is further confirmed by examining the structure of 8mer-CuGNA where all Watson-Crick base pairs assume a *gauche* configuration regardless of makeup. Overall, GNA duplexes display a large degree of variability in backbone conformation that currently cannot be traced back to unique interactions or different crystal packing environments.

With three structures in hand, we were interested in studying both the structure and dynamics of “native” GNA containing only the natural nucleobases adenine, guanine, thymine, and cytosine. MD simulations of nucleic acids have been used to substantially improving our understanding of both natural^{44,52,94-96} and unnatural nucleic acids.^{83-85,97-102} MD simulations can be used to compliment experiment in a variety of ways. The most obvious advantage is in the study of species that do not lend themselves easily to crystallization. Secondly, MD provides scientists with structural data as a function of time. MD allows the study of the dynamics of a structure on the nanosecond timescale while X-ray crystallography provides a snapshot of an otherwise time-averaged structure.

In the absence of a crystal structure of “native” GNA, model building and MD simulation can be used to understand the structure and dynamics of these unique species.

We performed the first computational study of the structure and dynamics of GNA duplexes. The structures included in this study are the previously determined crystal structures 8mer-CuGNA, 6mer-BrGNA, and 8mer-BrGNA. In addition to these sequences, the hydroxypyridone nucleotides were removed from 8mer-CuGNA and replaced by an A-T base pair, resulting in a “native” GNA sequence. Also, due to the quasicontinuous packing present in the GNA crystals, we are able to build longer sequences by including additional duplexes from the crystal lattice. The structure and dynamics of these GNA models are compared to both to each other as well as DNA.

In a related study, GNA containing porphyrin bases was synthesized and tested for its photophysical properties.¹⁰³ Nucleic acids are extensively used as a scaffold for the predictable and well defined positioning of multiple chromophores and other functional units.¹⁰⁴⁻¹⁰⁸ For example, DNA has the advantages of the canonical Watson-Crick base pairing scheme which allows for the creation of well defined helices and synthesis is highly reliable allowing for modifications to be made at defined sites along the structure. This has allowed DNA to be used as a supramolecular scaffold for the controlled arrangement of porphyrins. Through the use of acetylene spacers it has been shown that DNA can be used to create helical multiporphyrin arrays.¹⁰⁹ As the GNA duplex has been shown previously to have a much different structure than DNA, it is of interest to study if porphyrins are able to be accommodated by the GNA duplex, and also, are any unique photophysical properties derived from the novel arrangement.

In the work by Zhou and Zhang, two series of 10 GNA sequences (16mer and 22mer) were synthesized.¹⁰³ The strands varied in how many porphyrins are contained as well as the metal ions present (Zn^{2+} and/ or Ni^{2+}). It can be seen from Table 5.3 that the

TABLE 5.3
THERMAL STABILITIES OF PORPHYRIN-CONTAINING 16MER GNA
DUPLEXES TOGETHER WITH WATSON-CRICK REFERENCE DUPLEXES.¹⁰³

Entry	Sequence	T_m (°C)
1	3'-TAAAAATAATAATATT-2' 2'-ATTTTTATTATTATAA-3'	54
2	3'-TAAAAATATAATATT-2' 2'-ATTTTTATATTATAA-3'	52
3	3'-TAAAAAT P ATAATATT-2' 2'-ATTTTTA H TATTATAA-3'	48
4	3'-TAAAAAT P ATAATATT-2' 2'-ATTTTTAT T TATTATAA-3'	44
5	3'-TAAAAAT P ^{Zn} ATAATATT-2' 2'-ATTTTTA H TATTATAA-3'	42
6	3'-TAAAAAT P ^{Ni} ATAATATT-2' 2'-ATTTTTA H TATTATAA-3'	49
7	3'-TAAAAAT P H TAATATT-2' 2'-ATTTTTA H P ATTATAA-3'	43
8	3'-TAAAAAT P ^{Zn} H TAATATT-2' 2'-ATTTTTA H P ^{Zn} ATTATAA-3'	34

incorporation of one ($T_m = 48^\circ\text{C}$) or two ($T_m = 43^\circ\text{C}$) porphyrins across from abasic sites in the duplex result only in a net destabilization effect of 4 and 6°C , respectively, compared to their reference sequence. This difference in T_m is the likely result of losing hydrogen bonds inherently present in the A-T base pairs of the reference sequence. Even

with this loss of stability, porphyrin containing GNA is still more stable than the corresponding DNA duplex. Furthermore, when Zn^{2+} is added, a large destabilization effect is observed lowering the T_m to 34°C for the sequence containing two porphyrin moieties. In contrast, when Ni^{2+} is added, a small increase (1°C) in stability is observed. This phenomena is thought to arise due to the preference of Zn^{2+} to recruit axial ligands whereas Ni^{2+} prefers a square planar coordination. In order for the Zn^{2+} chelating porphyrin base to stack within the duplex, we hypothesize that it would need to dissociate the axially coordinated ligands (most likely water).

Longer, more robust 22mer sequences were synthesized in order to study the photophysical properties associated with the arrangement of the porphyrin moieties within the GNA scaffold. Uv-Vis spectroscopy showed that duplexes containing adjacent non metalated porphyrins, Zn^{2+} chelating porphyrins, and mixed metal examples display a splitting of the Soret band which is thought to be caused by dipole-dipole interactions of the two porphyrin units. CD spectroscopy provides further evidence that π - π stacking of the porphyrins is present within the GNA duplex. This is an important finding as it indicates a slipped-cofacial geometry of the porphyrin complexes which is necessary to achieved charge separation and has important implications in energy research. However, in the absence of a crystal structure, the exact arrangement of the porphyrins is difficult to determine. In this study, 20 ns MD simulations of two of the sequences in Table 5.3 are performed (entries 3 and 7). This provides us with structural and dynamic information which describes the incorporation of porphyrin moieties within the GNA duplex along with specific interactions between them.

5.2 Methods

5.2.1 CuGNA, BrGNA, and GNA

In order to obtain atomic charges for the GNA nucleotides for use in MD simulations, the Gaussian03¹¹⁰ program along with the *antechamber* module of AMBER9 was used.⁴⁹ The same procedure used in calculating atomic charges that was used in the original development of the AMBER DNA library was used in this study.⁵¹ Geometry optimizations of individual nucleotides was performed using Gaussian03¹¹⁰ at the B3LYP/6-31G* level while restraining the backbone torsional angles to those found in the experimentally determined structures. Electrostatic potentials were subsequently calculated at the HF/6-31G* level of theory through the fitting of RESP charges using *antechamber* and setting intermolecular equivalencies as needed. By setting intermolecular equivalencies, we insure that common backbone atoms in each created residue are the same. Only charges on the C1' and nucleobase atoms are allowed to differ. Due to the ionic state typically observed on the phosphate group of nucleic acids, each nucleotide is restrained to a charge of -1. After completion of this process a library is created containing atomic charges for all nucleotides used in this study. AMBER .off files for all the GNA residues created can be found in the digital archive in the 'Chapter5/Residues' folder.

Sequences of GNA shown in Table 5.4 were built from the available crystal structures. For the examples involving “native” GNA containing only natural

TABLE 5.4
LIST OF ABBREVIATIONS USED AND SIMULATIONS PERFORMED IN THIS
STUDY.^{90A}

6mer-BrGNA	3'-GC ^{Br} GCGC-2'
8mer-BrGNA	3'-CTC ^{Br} UAGAG-2';
16mer-BrGNA	3'-CTC ^{Br} UAGAGCTC ^{Br} UAGAG-2'
8mer-GNA	3'-CGAATTTCG-2'
16mer-GNA	3'-CGAATTTCGCGAATTTCG-2'
8mer-CuGNA	3'-CGHATHCG-2'
16mer-CuGNA	3'CGHATHCGCGHATHCG-2'
8mer-DNA	5'-CGAATTTCG-3'
16mer-DNA	5'-CGAATTTCGCGAATTTCG -3'

^aH = hydroxypyridone, ^{Br}C = 5-bromocytidine, ^{Br}U = 5-bromouracil.

nucleobases some editing of the crystal structures are necessary. To edit the 8mer-CuGNA structure to give a structural model containing only natural nucleobases, the hydroxypyridone bases are removed from the pdb file leaving only the nitrogen linkage to the backbone along with flanking carbons. By leaving these atoms, it allows a plane to be defined for the addition of the A-T base pairs which are substituted in. The residue name along with the remaining atoms are renamed in the pdb file to correspond to those

belonging to A or T residues. From here, the pdb file is loaded into the *xleap* module of AMBER allowing it to insert the missing atoms.

For simulations of species for which a crystal structure is available, set up is much more straightforward. Since the asymmetric unit of the crystal structures contain only a single strand, the complementary strand is included using Pymol¹¹¹ and including an additional strand from the crystal lattice. This pdb is subsequently saved and loaded into *xleap*. Since parameters are not available for the hydroxypyridone, bromocytosine, and bromouracil nucleobases in the standard AMBER force field, additional parameters must be added. For the copper(II) chelating hydroxypyridone base pairs parameters concerning bonds, angles, and torsions involving the Cu-O bonds were fitted to the experimentally determined structure by manually adjusting the force field parameters until the minimized base pair structurally matched the crystal structure. The final parameters for the hydroxypyridone base pairs can be found in the electronic archive in the 'Chapter5/Parameters' folder. Missing parameters for the bromocytosine and bromouracil bases were simply taken from the Generalized Amber Force Field (gaff) and added to the traditional AMBER force field. These parameter files can also be found in the 'Chapter5/Parameters' folder. AMBER atom types used to describe the GNA backbone can be found in Figure 5.8.

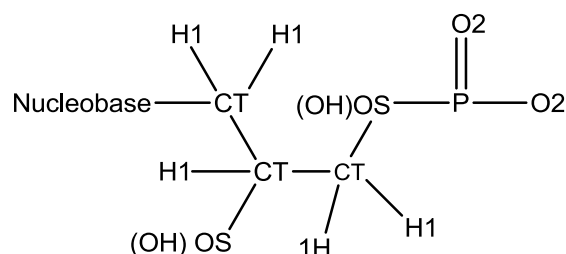


Figure 5.8- AMBER atom types used to describe the GNA backbone.

One advantage of having an end to end arrangement of duplexes in the crystal lattice is the quasi-continuous nature of the structures. This allows us to do is to include an additional duplex along the helical axis essentially creating a sequence of twice the length resulting in the structural models of the 16mer-GNA and 16mer-BrGNA sequences. Using the Pymol¹¹¹ program, it is possible to include an additional duplex from the crystal lattice which can be saved as a pdb. Residues then must be renumbered in the file to correspond to their new position along the strand. This edited pdb file can then be loaded into *xleap* in the same manner discussed for the previous examples. Pdb's for all the starting structures used in this study can be found in the 'Chapter5/Start' folder of the digital archive.

All simulations were set up in a similar manner. First, the system was neutralized by adding Na^+ counterions to balance the charge of the negatively charged phosphate backbone per common practice.⁹⁴ Then, the solute was solvated by placing it in a pre-equilibrated box of TIP3P waters extending 8 Å beyond the solute in each dimension. Detailed information on each system can be found in Table 5.5.

TABLE 5.5
ADDITIONAL INFORMATION ON THE SYSTEMS STUDIED.

Name	Sequence	Dimensions (Å)	Atoms	Waters	Na ⁺
8mer-GNA	3'-CGAATTCG-2'	43.5x47.9x50.6	7232	2264	14
16mer-GNA	3'-CGAATTCGCGAATTCG-2'	53.8x50.9x85.6	17372	5496	30
8mer-CuGNA	3'-CGHATHCG-2'	43.5x47.9x50.6	7234	2264	14
16mer-CuGNA	3'-CGHATHCGCGHATHCG-2'	50.2x46.0x81.1	13344	4152	30
6mer-BrGNA	3'-GC _(Br) GCGC-2'	46.6x46.0x43.8	6434	2036	10
8mer-BrGNA	3'-CTCU _(Br) AGAG-2'	44.7x45.1x50.1	6974	2180	14
16mer-BrGNA	3'-CTCU _(Br) AGAGCTCU _(Br) AGAG-2'	49.7x51.1x73.7	13484	4204	30
8mer-DNA	5'-CGAATTCG-3'	42.1x50.9x52.6	7948	2476	14
16mer-DNA	5'-CGAATTCGCGAATTCG-3'	46.0x47.1x79.6	12312	3756	30

After initial system set up, a sequential minimization was used to prepare the system for MD. First, the solute is restrained allowing the sodium counterions and solvent relax. In the case of “native” GNA structures which were created from the 8mer-CuGNA structure, an additional step was included where the hydrogen bonds of the replaced base pairs are restrained to their equilibrium distance. This allows for the slight perturbation of the structure caused by the replacement of the hydroxypryidone base pair to be corrected prior to MD. For all cases the final step involved the removal of all restraints allowing the entire system to relax. The *pmemd* module of AMBER was used to perform all minimizations and MD simulations to improve the multiprocessor, parallel

simulations. The base parameters for the simulations were adopted from the Cornell et al. force field⁵⁰ along with the adjustments by Wang et al.,⁵¹ commonly known as parm99. A bonded model is used to describe the Cu-O bonds in the hydroxypyridone base pairs. While this forcefield is designed the study of known biomolecules, the manner in which the forcefield was conceived leads us to believe it will perform well in describing the GNA backbone torsional behavior. The Cornell et al. forcefield is based on the fitting upon experimental and quantum mechanical results for methyl ethyl ether and similarly acyclic model systems. Thus, the GNA backbone is much more similar to the model systems used to create the forcefield than ribose based backbone of DNA and RNA, which are cyclic ethers.

MD simulations were started by first heating the system within a constant volume ensemble (NVT) by adding a 10 kcal/mol restraint on the solute while heating from 0-300°K over 20 ps. Solute restraints were then removed and the system was equilibrated for 200 ps in a constant pressure ensemble (NPT) at 1 atm of pressure. Production runs of 20 ns (100 ns for 16mer-GNA) were performed. 2 fs timesteps were used for DNA, GNA, and BrGNA examples. However, 1fs timesteps were required for all CuGNA models to maintain system stability. The SHAKE algorithm was used to restrain all bond to hydrogen. This allows a longer timestep to be taken due to elimination of the highest frequency molecular motions of the system.

All post processing of trajectory data was performed using the *ptraj* module of AMBER. Backbone rmsds, snapshot structures, and various interatomic distances were extracted. Pdb's were written every 20 ps to give a total of 1000 structures per simulation. The program Curves 5.1^{91,92} was used to calculate the helicoidal parameters from the

extracted snapshots. Results were then averaged or plotted over time. Additional information on setting up the simulations and analyzing the results is discussed in Section 5.3.

5.2.2 Porphyrin Containing GNA

The crystal structure of porphyrin containing GNA duplexes is not available. We therefore build a models of the 3'-TAAAAATPATAATATT-2' / 2'-ATTTTAAHTATTATAA-3' and 3'-TAAAAATPHTAATATT-2' / 2'-ATTTTAAHPATTATAA-3' duplexes (entry 3 and 7 from Table 5.3) based on our previously described GNA model.⁹⁰ In short, the crystal structure of an 8-mer duplex of an GNA analog was extended by including an additional end to end stacked duplex from the crystal lattice using the Pymol program and refined using MD simulations as described in Section 5.2.1. In order to create the models for the porphyrin containing GNA, the nucleobases from the parent GNA were removed, leaving only the nitrogen linkage to the backbone and the flanking carbons. This allows the plane of the nucleobase to be properly defined and the *xleap* module of AMBER9 was used to add the remaining atoms. As the porphyrin moiety is incorporated through an acetylene linkage, the proper definition of the plane is not possible and the porphyrin residues must be manually rotated into the GNA duplex.

Charges for all residues (2', 3', and interior adenine and thymine, abasic site, and porphyrin residue) were assigned by RESP fitting on HF/6-31-G* optimized structures. Intermolecular charge constraints were used ensure equivalent charges along the backbone, which were neutralized by adding 30 Na⁺ ions using *xleap*. The solute is

placed into a pre-equilibrated box of TIP3P water extending 10 Å beyond the solute in all directions.

The initial models were minimized first by restraining solute atoms while relaxing the sodium ions and solvent, followed by minimization using constraints to fix the base pair hydrogen bonds and to position the porphyrin moieties properly within the duplex. Finally, an unrestrained minimization was performed in preparation for MD. The Particle Mesh Ewald Molecular Dynamics (pmemd) module of AMBER was used to perform all MD simulations to improve the efficiency of the parallel simulations. Force field parameters for the simulations were adopted from the Cornell et al. forcefield⁵⁰, along with the adjustments by Wang et al⁵¹ (parm99). Parameters for unmodified GNA bases and glycol backbone were taken directly from the Cornell et al. force field. Parameters for the porphyrin moieties were adopted from the *gaff* force field. Heating of the system from 0 – 300° K was performed in the NVT ensemble over 20 ps with restraints on the solute. From here, equilibration and production runs were performed in the NPT ensemble resulting in 50 ns trajectories for both sequences studied. The ptraj module of AMBER was used for all post-processing of the trajectory data.

5.3 Instructions for Running GNA Simulations

All files needed to perform these simulations can be found in the digital archive in the folder named ‘Chapter5’. This folder contains all residue files (.off) ‘/Residues’, parameter files (.frcmod) ‘/Parameters’, starting geometries (.pdb) ‘/Start’, scripts for extracting CURVES data (.sh) ‘/Scripts’, ensemble average structures (‘/Average’), and structural overlays of crystal structures (‘/Overlay’). To perform this work as laid out

one needs access to the AMBER MD package and CURVES 5.1. The instructions provided here are not comprehensive and a basic knowledge of the AMBER program suite, Linux, and basic bash scripting is required.

To set up a calculation in *xleap* all of the residue files must be loaded corresponding to the sequence that is being simulated. For example, if the sequence 3'CGAATTCG-2' is to be loaded into *xleap*, the ZC3.off, ZG.off, ZA.off, ZT.off, ZC.off, and ZG2.off are loaded using the command 'loadoff *.off' with the wildcard corresponding to the needed residue file. The hydroxypyridone base pair is composed of two ZH residues and a single CU residue. In this case bonds need to be added between the hydroxypyridone bases and the copper atom. Bromouracil and bromocytosine residues can be found as ZBR and CYB respectively. If running a simulation including a nonnatural nucleobase, it is necessary to load the corresponding parameters (.frcmod) using the command 'loadamberparams *.frcmod'. From here the pdb file can be loaded (be sure atom and residue names match between AMBER residue files and pdb file) using the command 'loadpdb (unit name) *.pdb'. Counterions can be added to neutralize the system using the command 'addions (unit name) Na+ 0' and the solvent box can be added using the command 'solvatebox (unit name) TIP3PBOX (distance from solute)'. The command 'check (unit name)' should be run in order to check whether all the parameters have been properly loaded. Finally, the AMBER topology file and input coordinates can be written using the command 'saveamberparm (unit name) *.prmtop *.inpcrd' with the wildcard being whatever the user wishes to name the files.

In order to extract helicoidal parameters using the program CURVES, snapshots from the trajectory need to be extracted using the *ptraj* module of AMBER. When

writing pdb files, all solvent and counterions must be stripped from the trajectory file. Two example bash scripts are included in the digital archive within the ‘Chapter5/Scripts’ folder. Script1.sh reads in the extracted pdbs and edits each so it can be read by the CURVES program. The changes being made to the pdbs is converting ZA residues to A, ZH to U, etc. The *sed* portion of the script (line 23) is responsible for this and should be edited to match the sequence you are submitting for CURVES analysis. In short, the script renames the residues and bases to match that typically seen in DNA naming. Also, script1.sh creates a CURVES input file for each extracted pdb and runs the program. The command to run this script is ‘sh script1.sh’. Script2.sh is used to extract the desired lines of interest from the CURVES output (for example the line containing local average inter-base parameters). In this script, the *sed* portion (line 10) needs to be edited to extract the line of interest from the CURVES output from the previous step. This script is run by typing the command ‘sh script2.sh’. The provided scripts are only meant to be used as examples and will most certainly need to be edited to tailor the sequence of interest. However, with a basic knowledge of bash scripting the changes are very straightforward. Finally, the output files created by script2.sh need to be combined. This can be done by running the command ‘cat out* > output.out’. The file output.out will contain the CURVES helicoidal parameter data versus time which can be imported into graphing software for analysis.

5.4 Results for MD Simulations of GNA

5.4.1 8mer-CuGNA

To validate the performance of the AMBER force field sequences known crystal structures were studied first. The 8mer-CuGNA crystal structure was set-up per the methods discussed in the previous two sections. 20 ns production MD trajectories were performed and 1000 pdb's were written to be analyzed by the CURVES program. Overall, the helicoidal parameters reported from the experimental and simulated 8mer-CuGNA are in almost complete agreement (Table 5.6). The simulated 8mer-CuGNA duplex structure remains in the elongated form which is observed in the crystal structure. This is indicated by the nearly identical average base pair twist of the two structures (MD=24.1°, Experimental=23.5°) leading to 15-16 base pairs per turn. The two structures match up nearly perfectly with the backbone rmsd between the two structures being < 1 Å. From these results, it is clear that the inclusion of the copper(II) chelating hydroxypyridone base pairs (which make up 25% of all base pairs present in 8mer-CuGNA) exert a strong unwinding effect on the overall helical structure of the GNA duplex in comparison to the BrGNA examples. This phenomena has been observed recently in an NMR structure of a synthetic DNA hybrid which contained three consecutive Ag(I) imidazole base pairs.¹¹² In that work the unwinding results in the average base pair twist being reduced from 36° to 28°. Interestingly, the magnitude of the unwinding seen from the inclusion of Ag(I) imidazole base pairs in duplex DNA is similar to what is seen in the case going from CuGNA to BrGNA.

TABLE 5.6
AVERAGE HELICOIDAL PARAMETERS FOR GNA DUPLEX STRUCTURES
OBTAINED FROM EXPERIMENT AND MD SIMULATIONS.

	8mer-CuGNA	8mer-CuGNA	16mer-CuGNA
	Exp.	MD	MD
Shift (Å) ^b	0.0	-0.1	0.0
Rise (Å)	3.5	3.5	3.5
Tilt (°)	0.0	-0.8	0.1
Roll (°)	-2.7	-0.4	-2.4
Twist (°)	23.5	24.1	23.9
Slide (Å)	-3.5	-2.9	-2.9

Overall, the similarities seen between theory and experiment are promising. It should be noted that the overall goal of this work is not to simulate the structures of the obtained crystal structures, but rather learn about the structure and dynamics of “native” GNA. The simulations of CuGNA and BrGNA are mainly performed in order to evaluate the performance of the AMBER force field in describing GNA. From the results shown here for 8mer-CuGNA it appears that the forcefield performs admirably in

describing the GNA backbone. Therefore, model building and subsequent simulation of “native” GNA appears promising.

5.4.2 16mer-CuGNA

As discussed in section 5.2, simulations of longer sequences are possible by including an additional end-stacked duplex from the crystal lattice. In this case an additional duplex resulting in a 16mer-CuGNA of the sequence 3'-CGHATHCGCGHATHCG-2' was included. 20 ns MD production trajectories were calculated and CURVES analysis was performed on 1000 equally spaced snapshots throughout the simulation. Helicoidal parameters derived from this study are shown in Table 5.6. It can be seen, that much like what was observed in the 8mer-CuGNA sequence, the results are in almost complete agreement with the experimentally solved structure. An overlay of the two structures can be found in Figure 5.9.

5.4.3 6mer-BrGNA

As bromination of GNA bases in 6mer-BrGNA crystal structure was not found to induce obvious perturbation of the GNA duplex, MD simulations can be started from the available structure.³⁴ A 20 ns production MD trajectory was used to generate 1000 snapshots to be analyzed by the CURVES program for helicoidal parameters. In Table 5.7, it can be seen that the agreement between theory and experiment is quite good. For example, the local rise between neighboring base pairs is almost identical between the

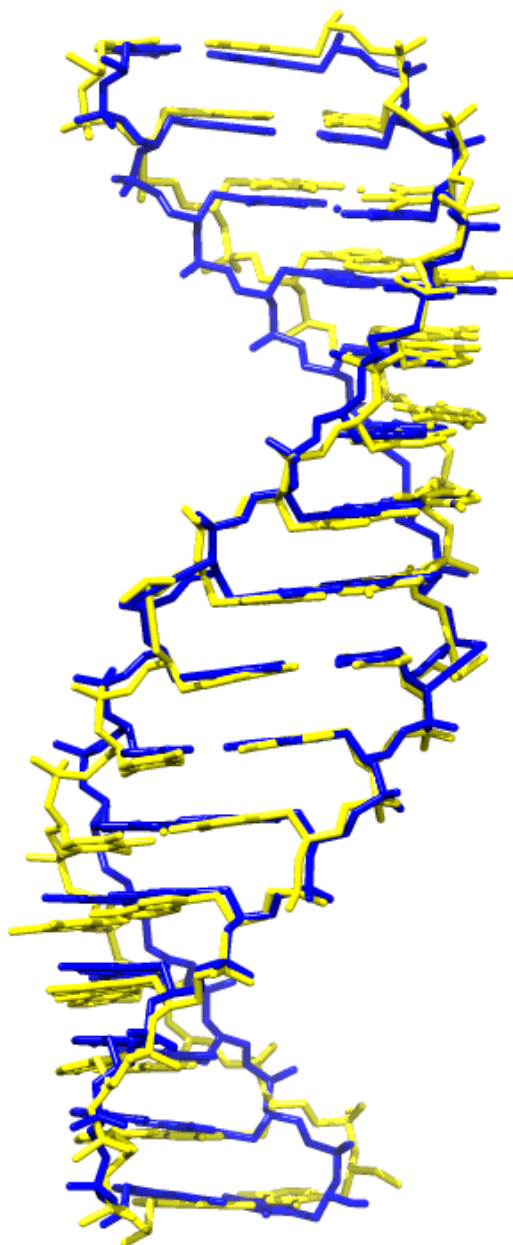


Figure 5.9- An overlay of the crystal structure (yellow) and average structure from the MD simulation (blue).

TABLE 5.7
AVERAGE HELICOIDAL PARAMETERS OF 6MER-BRGNA CRYSTAL
STRUCTURE AND AVERAGE MD STRUCTURE.⁹⁰

	6mer-BrGNA	6mer-BrGNA
	Exp.	MD
Shift (Å) ^b	0.0	0.0
Rise (Å)	3.2	3.3
Tilt (°)	0.0	0.0
Roll (°)	5.9	4.5
Twist (°)	37.2	31.8
Slide (Å)	-3.3	-3.2

experimental structure and simulation of 6mer-BrGNA in solution, along with base pair tilt, roll, and slide. The only major difference that can be gleaned from the helicoidal parameters is the difference in average base pair twist. However, it is not uncommon for MD simulations of nucleic acids to reports different values for this parameter as well as several other backbone parameters due to the different environments caused by crystal packing.⁹³ As stated previously, the crystallographic analysis of short oligonucleotides

has the shortcoming possessing a high ratio of end-stacking base pairs. Therefore, there is often a prerequisite for the formation of ordered crystal lattices for some deformation of base pair twist as compared to the same duplex in solution. An overlay of the two structures is shown in Figure 5.10.

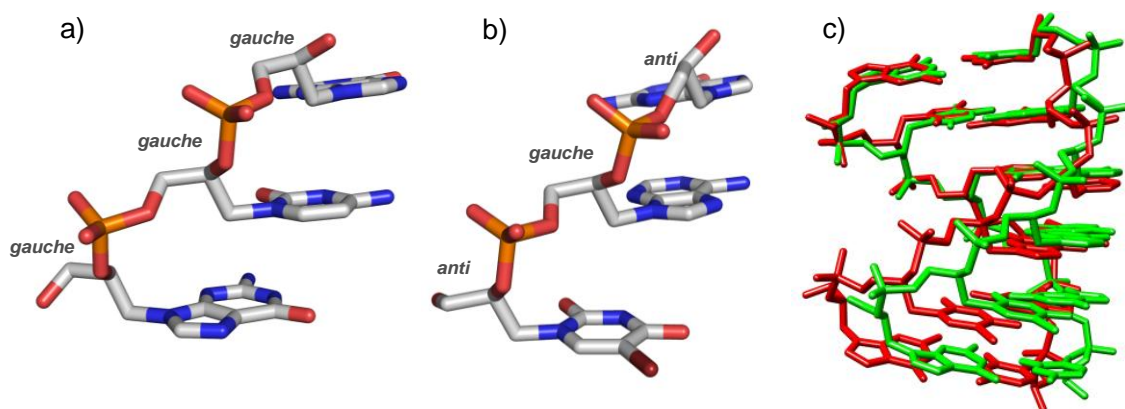


Figure 5.10- View of average backbone dihedral from MD (a) and crystal structure (b). Overlay of the experimental (red) and MD average (green) structures of 6mer-BrGNA.

In the crystal structure of 6mer-BrGNA, the average base pair twist is found to be 37.2° while the MD simulation reports 31.8° . This same phenomenon can be seen in the comparison of an 8mer-DNA crystal structure, where experiment reports an average base pair twist of 36.0° whereas MD reports 32.9° . The only other major structural difference between experiment and theory is the conformations adopted by O3'-C3'-C2'-O2' backbone dihedral. As discussed previously, this backbone torsional angle alternates between *gauche* and *anti* throughout the entire length of the BrGNA duplexes. However, in the MD simulations there is a strong preference for the *gauche* conformation and very little time is spent in the *anti* conformation ($\approx 1\%$). In a previous study by the Meggers

Group, crystal structures of a single nucleotide show that the gauche conformation is indeed favored.³⁶ Also, in the 8mer-CuGNA crystal structure, it can be observed that all Watson-Crick base pairs show a preference for the gauche conformation. Close inspection of the crystal structures does not reveal any specific interactions that rationalize this behavior of the O3'-C3'-C2'-O2' torsion. One piece of information that can be gleaned from this observation is that the GNA backbone, as expected, is likely to be very flexible and rotation of the backbone torsions may play a part in the duplex stability of native GNA. Other than this, the agreement between theory and experiment is excellent.

5.4.4 8mer-BrGNA and 16mer-BrGNA

As the results from the simulations of 8mer-BrGNA and 16mer-BrGNA do not deviate in any significant way from the 6mer-BrGNA simulation and the same conclusions apply, they will not be discussed in great detail. It should be noted again that the 16mer-BrGNA structure was created by including an additional end-stacked duplex from the crystal lattice. Just as in the 6mer-BrGNA simulation, the agreement between theory and experiment is excellent. All helicoidal parameters are again in close agreement, with the only major deviation being in the average base pair twist (Table 5.8). While the experimentally determined structure of 8mer-BrGNA displays an average base

TABLE 5.8
HELICOIDAL PARAMETERS FOR 8MER AND 16MER-BRGNA.⁹⁰

	8mer-	8mer-	16mer-
	BrGNA	BrGNA	BrGNA
	Exp.	MD	MD
Shift (Å)	0.0	0.0	0.0
Rise (Å)	3.3	3.3	3.3
Tilt (°)	0.0	0.2	0.1
Roll (°)	3.1	2.5	2.2
Twist (°)	36.8	30.9	30.6
Slide (Å)	-3.2	-3.0	-3.0

pair twist of 36.8° compared to 30.9° and 30.6° for the simulations of 8mer-BrGNA and 16mer-BrGNA respectively. An overlay of the experimental and theoretical structures can be found in Figure 5.11. With the success of reproducing the experimental helicoidal parameters through the use of MD simulation, we can now proceed to model building and

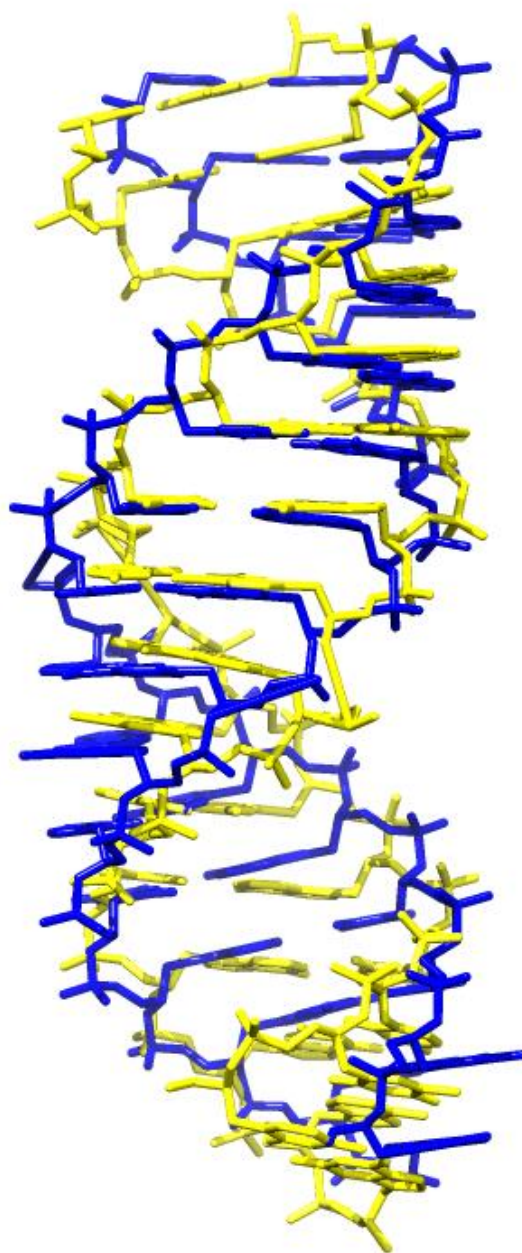


Figure 5.11- Overlay of the 16mer-BrGNA crystal structure (yellow) and average MD structure (blue).

analysis of the elusive native GNA duplex which to this point has, to this point, not been crystallized

5.4.5 8mer-GNA

In the absence of a crystal structure of this elusive species, MD provides an attractive avenue to study both the structure and the dynamics of GNA. As discussed previously, the “native” GNA model can be built by replacing the hydroxypyridone base pairs in 8mer-CuGNA with A-T base pairs resulting in the sequence 3'-CGAATTCG-2'. 20 ns production trajectories were obtained through the use of MD simulation in the AMBER9 program. As expected, the average helicoidal parameters of 8mer-GNA differ greatly from those inherent to B-DNA (Table 5.9). The most notable difference, which is also observed in all current crystal structures, is the large average base pair slide. 8mer-GNA is predicted to have an average base pair slide of -3.2 Å compared to -0.6 Å in dsDNA. This difference leads to 8mer-GNA to base stack in an interstrand fashion as opposed to DNA which base stacks in an intrastrand fashion. This hydrophobic overlap is most pronounced when the 2' bases involved are both purines. This phenomena has been demonstrated experimentally by the Meggers group through the study of the effects of adding overhanging nucleotides on either the 3' or 2' termini.³⁶ It was found that the addition of an overhanging adenine nucleotide on the 2' terminus leads to an increase of T_m by 14.6°C and an increase free energy of 4.0 kcal/mol associated with annealing. This overlap of 2' nucleobases is reproduced well by all of the MD simulations of GNA presented. It was also found that the addition of an overhanging nucleotide to the 3' terminus results in a much less pronounced effect leading to only a slight increase of T_m

TABLE 5.9
COMPARISON OF HELICOIDAL PARAMETERS OF 8MER-DNA AND GNA.⁹⁰

	8mer-DNA	8mer-DNA	8mer-GNA
	Exp.	MD	MD
Shift (Å)		0.1	0.0
Rise (Å)	3.4	3.5	3.5
Tilt (°)	-0.1	0.5	0.0
Roll (°)	0.6	4.3	0.9
Twist (°)	36.0	32.9	28.7
Slide (Å)	0.2	-0.6	-3.2

(1-1.2°C).³⁶ Performing the same experiments with DNA much less pronounced effects were observed.

CURVES analysis of the helicoidal parameters of 8mer-GNA can be found in Table 5.9. The structure of native GNA can best be described as a hybrid between the two types of experimental structures, hydroxypyridone metallo base pair containing and bromo nucleobase analogs. 8mer-GNA is predicted to possess a roll of 0.9° versus 4.5° and 3.1° for 6mer-BrGNA and 8merBrGNA respectively. 8mer-CuGNA possesses a roll

of -2.8° . 8mer-GNA also possesses an average base pair twist of 28.7° which lies approximately halfway between the two types of crystal structures presented. Along with the interesting structure present in 8mer-GNA (Figure 5.12), it also displays interesting properties in terms of dynamics. This will be discussed in detail in the next section with the 16-mer-GNA simulation.

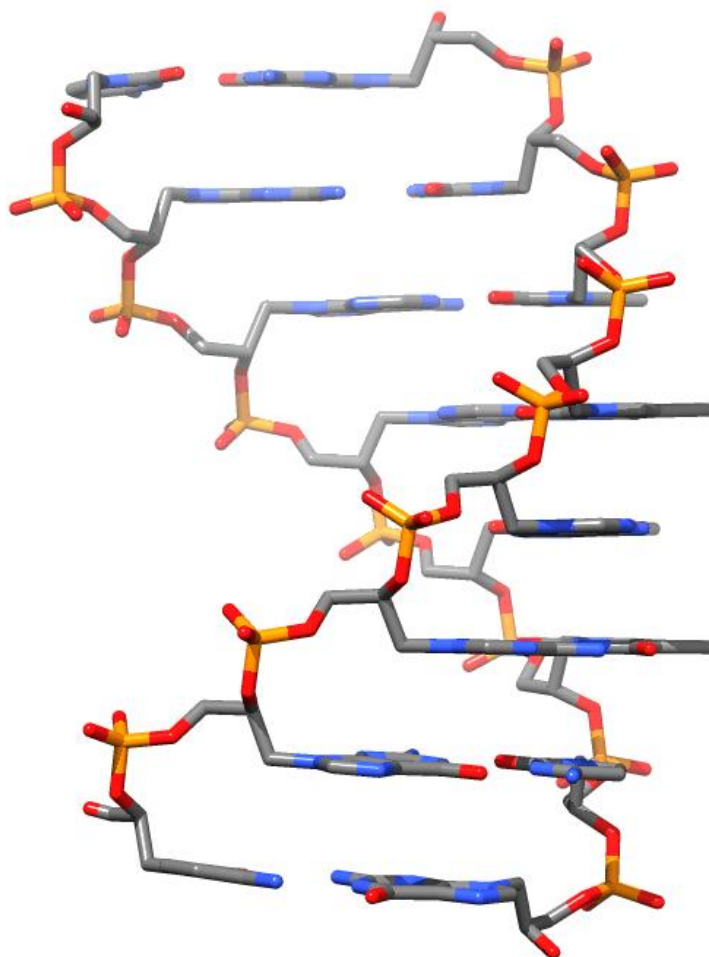


Figure 5.12- Average structure from simulation of 8mer-GNA.

5.4.6 16mer-GNA

A 16mer-GNA model was built from the previously created 16mer-CuGNA structure resulting in the sequence 3'-CGAATTCGCGAATTCG-2'. For this final study, a 100 ns simulation was performed. As the helicoidal parameters reported from this simulation are negligibly similar to 8mer-GNA, they will not be discussed in detail. The results of this analysis can be found in Table 5.10. Here, the dynamics of the 16mer-

TABLE 5.10
COMPARISON OF HELICOIDAL PARAMETERS OF 8MER-DNA AND 16MER-GNA.

	8mer-DNA	16mer-DNA	16mer-GNA
	Exp.	MD	MD
Shift (Å) ^b		0.0	0.0
Rise (Å)	3.4	3.5	3.5
Tilt (°)	-0.1	-0.7	0.0
Roll (°)	0.6	4.2	1.2
Twist (°)	36.0	33.3	29.5
Slide (Å)	0.2	-0.8	-3.5

GNA structure will be discussed. Per common practice in testing whether a simulation is “equilibrated”, the backbone rmsd versus time is plotted in Figure 5.13. It can be seen that throughout the simulation, the backbone rmsd fluctuates throughout the simulation. This can sometimes be an indication of system instability in nucleic acids such as base pair melting. However, in this case the fluctuations are periodic on an 8-10 ns timescale and correspond to large scale molecular motions of the 16mer-GNA helix. These fluctuations correspond to a helicoidal twisting and untwisting mode which is observed along the entire 100 ns trajectory. Figure 5.14 depicts representative coiled and uncoiled structures from the simulation. It is interesting to note that this motion is again periodic in nature. This is in line with the notion that other nucleic acids could be considered “molecular springs” while suggesting that GNA has an even larger amplitude of this motion. It follows that the conformational flexibility of the GNA duplex allows for a large number of microstates to be accessed and is consistent with the decreased entropic penalty upon annealing in comparison with DNA. Analogous twisting and untwisting modes are not observed in our simulations of dsDNA where fluctuations tend to coorespond to terminal base pair melting and disruptions of hydrogen bonds. This is in stark contrast to what is observed in 16mer-GNA. In Figure 5.15, hydrogen bond distance vs. time for the terminal base pairs of 16mer-DNA and 16mer-GNA are plotted. While the backbone rmsd fluctuates throughout the simulation, hydrogen bonding across the duplex remains intact for 16mer-GNA. This indicates that much of the enthalpic contribution to the free energy of duplex formation found in dsDNA is retained while the flexibility of the dsGNA backbones ability to adopt many conformations is also retained

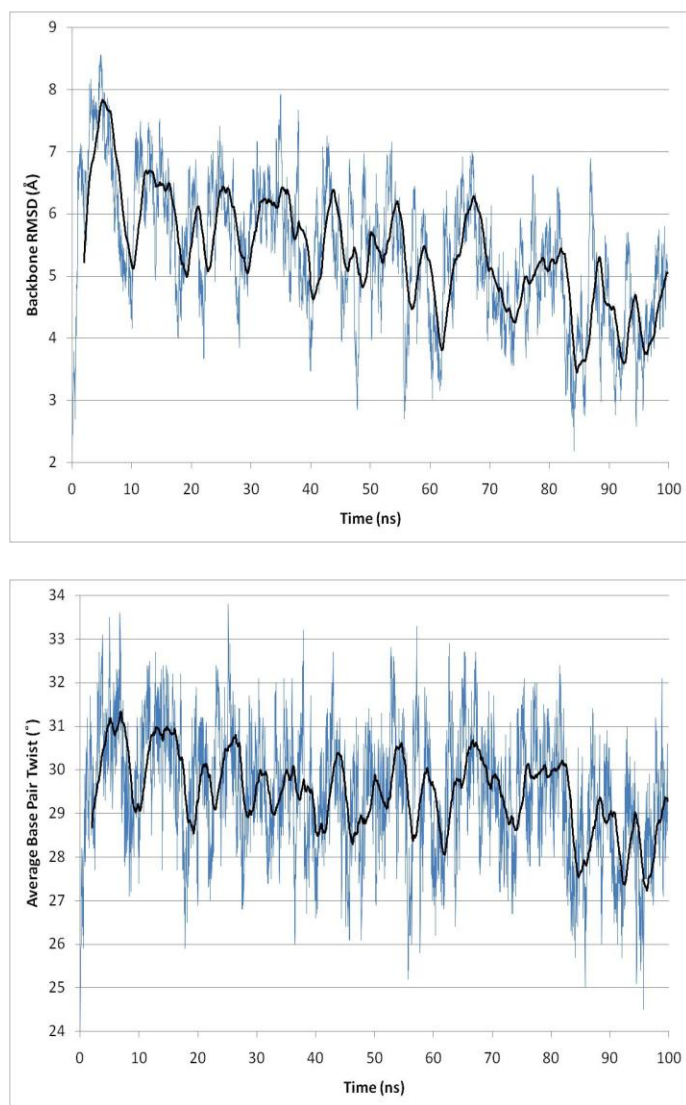


Figure 5.13- Plots of backbone rmsd and average base pair slide versus time for the simulation of 16mer-GNA.

leading to a greater entropy of the duplex. This leads to a smaller entropy difference between the single strand and the duplex for GNA in comparison to DNA.

In order to explore the relative entropic effects between the annealing of DNA and GNA duplexes, frequency analysis is performed to get a more quantitative picture of

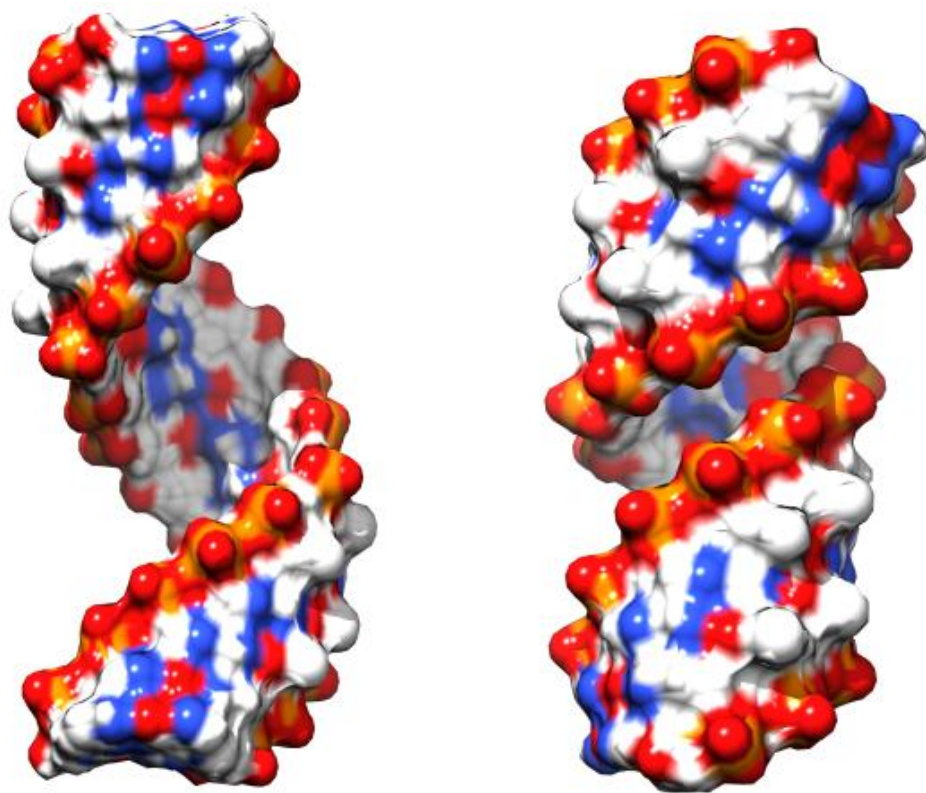


Figure 5.14- Snapshots from the 100 ns simulation of 16mer-GNA depicting the uncoiled (left) and coiled (right) conformations.

the situation. 20 ns MD trajectories of 16mer-DNA and 16mer-GNA in both their single stranded and duplex forms are obtained. 400 snapshots from each simulation are subjected to vibrational mode analysis using the *nmode* module of AMBER. Two values of the dielectric constant are employed in this analysis ($\epsilon=4$ and 80). The $\Delta\Delta S$ experimentally observed between DNA and GNA annealing is ≈ 14 kcal/mol. Results obtained through vibrational analysis predict a $\Delta\Delta S$ of ≈ 5.8 -6.7 kcal/mol (Table 5.11). This is in agreement with previous analyses of entropic contributions to dimerization with

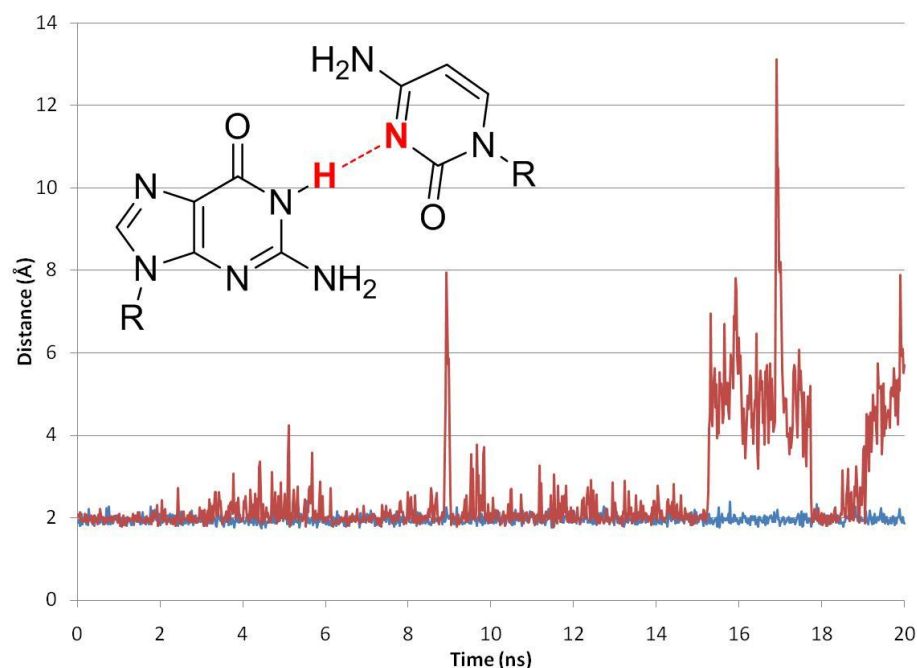


Figure 5.15- Plots of hydrogen bond distance versus time for a terminal base pair in 16mer-DNA (red) and 16mer-GNA (blue).

the absolute magnitude of $\Delta\Delta S$ predicted by *nmode* being underestimated by ~50%.¹¹³ One possible reason for this underestimation could be that counterions and solvent molecules are not included in the normal mode calculation. Nonetheless, the results clearly show that GNA duplex formation is entropically less unfavorable than the same process in DNA. The calculations also suggest that the largest contribution to $\Delta\Delta S$ is due to vibrational entropy. This is consistent with the hypothesis that the observed twisting/untwisting modes inherent to duplex GNA are likely responsible for the entropic effects previously explained. Other possibilities such as solvation effects, of course, cannot be excluded. However, this is beyond the scope of this study.

TABLE 5.11
RESULTS FROM *NMODE* ANALYSIS OF DNA AND GNA ANNEALING
(KCAL/MOL).⁹⁰

	Single Stranded		Double Stranded		Single Strand- Double Strand	
	TS _{vib}	TS _{total}	TS _{vib}	TS _{total}		
DNA ($\epsilon=4$)	405.1	435.8	786.1	818.2	24.0	53.3
GNA ($\epsilon=4$)	361.4	391.9	705.2	737.1	17.6	46.6
DNA ($\epsilon=80$)	402.6	433.3	778.5	810.5	26.7	56.1
GNA ($\epsilon=80$)	354.6	384.8	687.7	719.2	21.4	50.3

5.4.7 Simulations of Porphyrin Containing dsGNA

Figures 5.16 and 5.17 show the average structure of the two duplexes studied (entries 3 and 7 from Table 5.3). It can be seen that the porphyrin moiety is well incorporated into the GNA duplex across from an abasic site with only minor distortions of the overall structures, leading to overall structures that are essentially identical to the parent 16mer-CuGNA. For both systems studied, the average interstrand P-P distance for

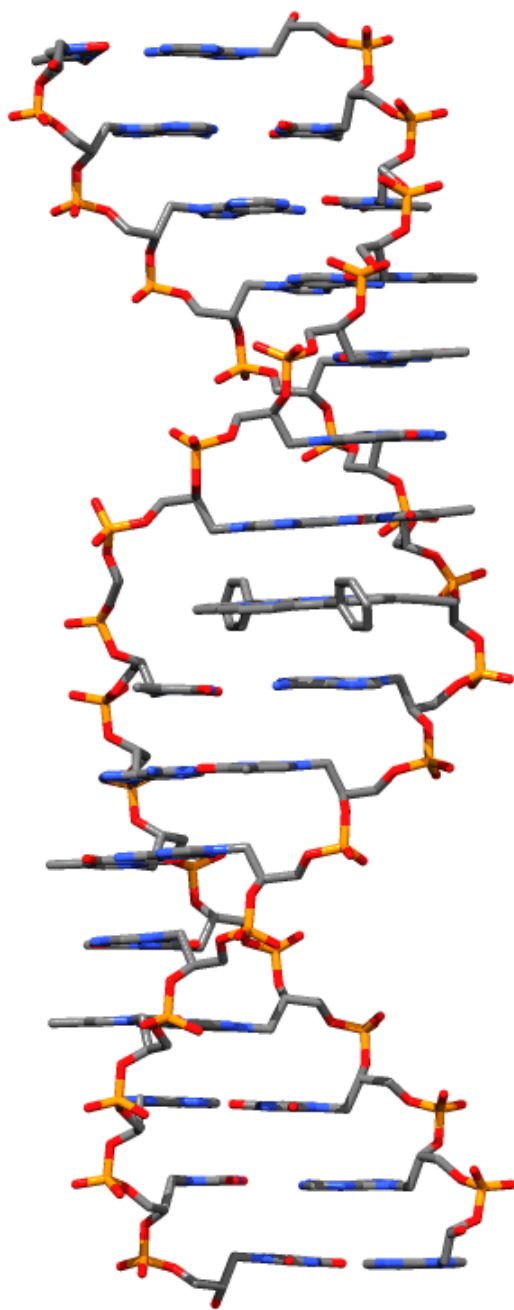


Figure 5.16- Average structure obtained from the MD simulation of GNA containing on porphyrin moiety.

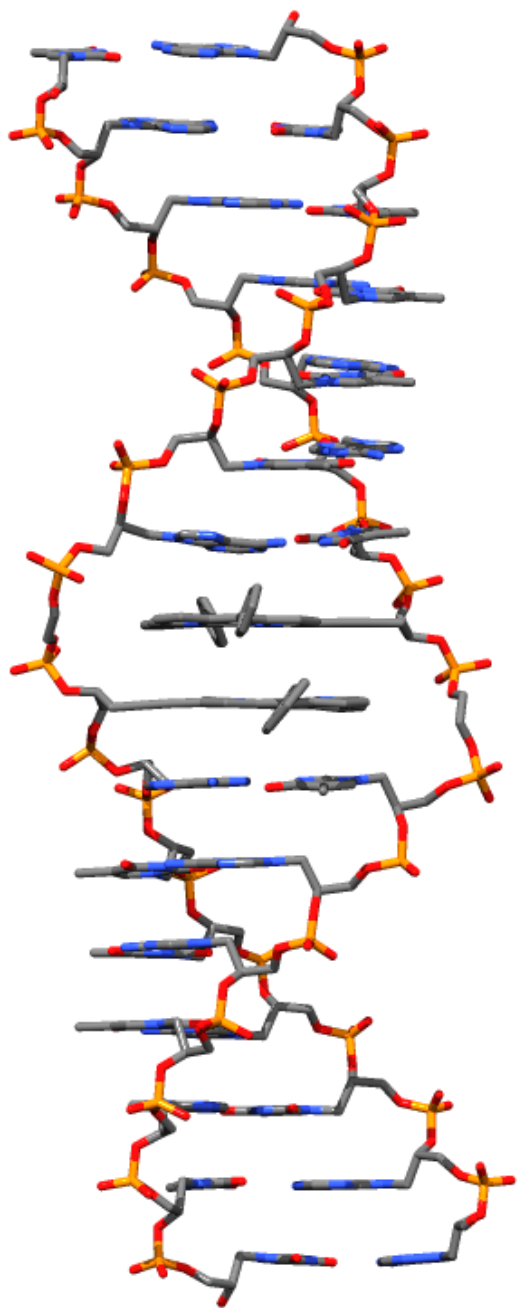


Figure 5.17- Average structure obtained from the MD simulation of a GNA sequence containing two adjacent porphyrin moieties.

the incorporation site is 17.0 Å versus 16.6 Å for a canonical GNA base pair, with the porphyrin protruding beyond the band formed by the GNA. In the case of entry 3, close π -stacking interactions are observed between the porphyrin base and the neighboring base pairs (Figure 5.18). In entry 7, the two porphyrin groups (center to center distance) stack at an average distance of 4.3 Å throughout the 50 ns trajectory. The π -stacking

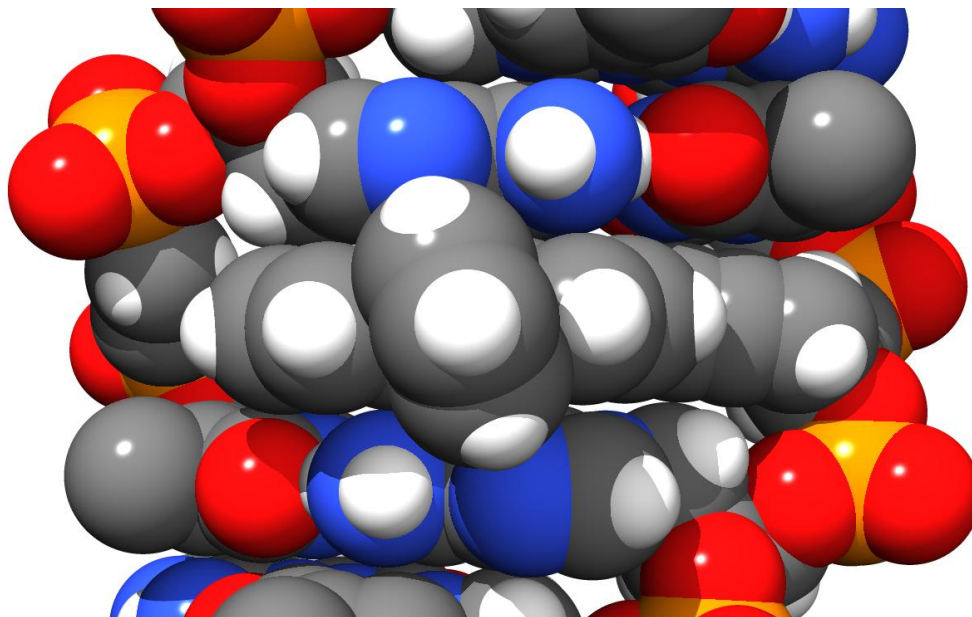


Figure 5.18- View of the π stacking between the porphyrin moiety and the adjacent base pairs of entry 3.

interaction positions the two porphyrin rings on top of each other, as shown in Figure 5.19. This offset of the stacked porphyrin centers in the calculated structures is in excellent agreement with the experimental observation of the split Soret band seen in the UV-Vis spectra of the metal complexes of the related entry 7.

The key differences in the calculated structures of the two systems studied here are the conformations of the phenyl rings of the porphyrin moiety on either the concave

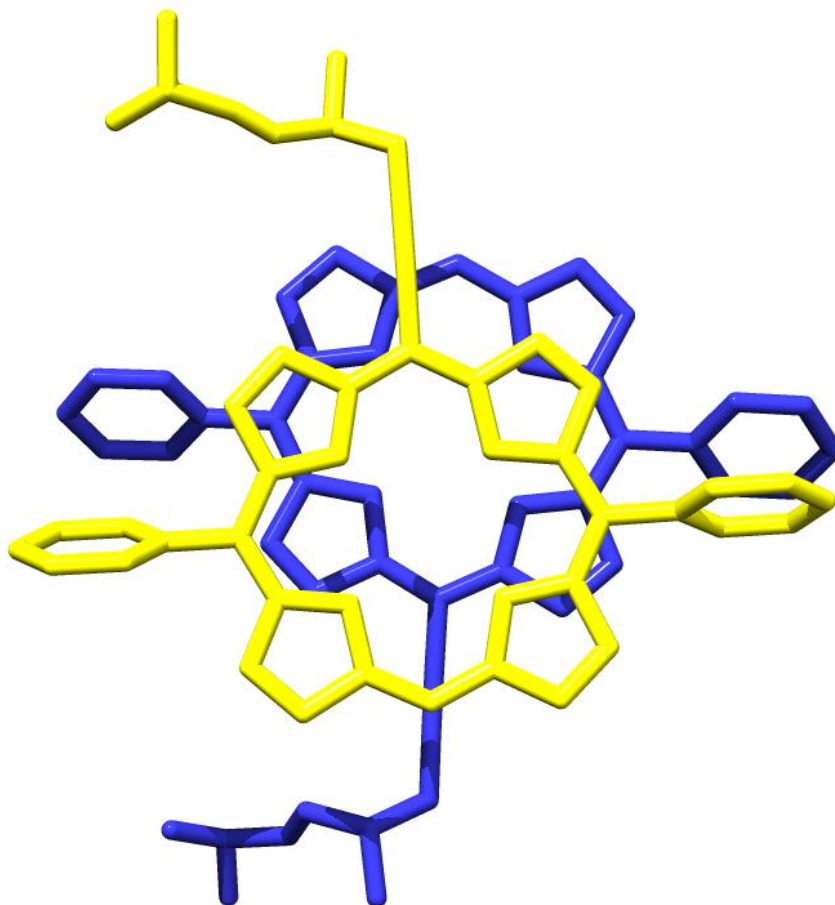


Figure 5.19- View of the slipped cofacial geometry of the adjacent porphyrins predicted from the MD simulation of entry 7.

or convex surface of the duplex. On the convex surface, which is more solvent exposed, the two phenyl rings π -stack with each other effectively reducing their solvent exposed surface (Figure 5.20a). In contrast, the two phenyl rings on the concave surface are arranged in an edge-to-face fashion (Figure 5.20b). This can be rationalized by the concave surface of the GNA duplex being more shielded from solvent, therefore decreasing the importance of the π - π interactions between the aromatic systems. It

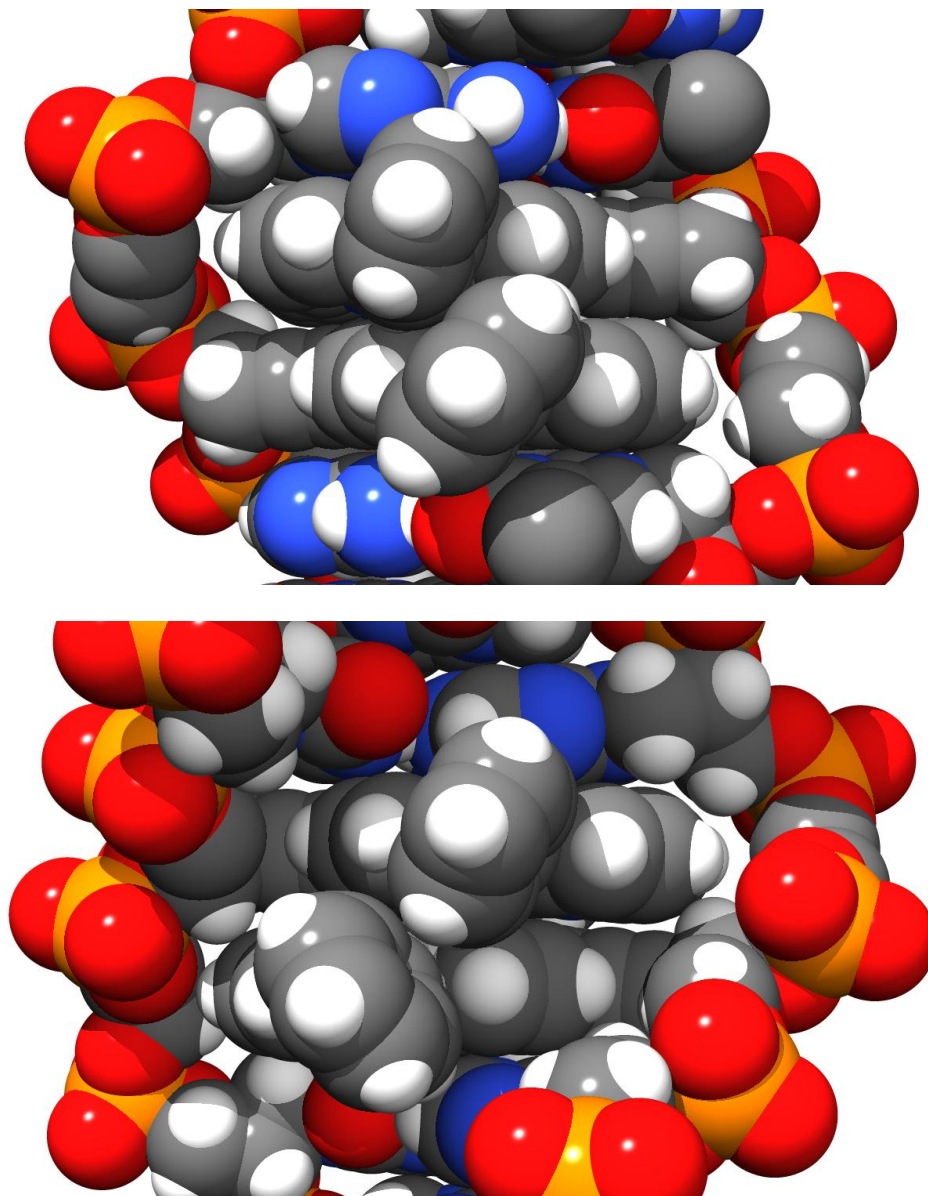


Figure 5.20- Views of porphyrin base stacking from the convex (top) and concave (bottom) sides of the GNA duplex.

should also be noted that the concave surface of the GNA duplex creates a hydrophobic pocket lined by the aliphatic carbons constituting the GNA backbone. This allows the phenyl ring to adopt conformations other than the face to face arrangement observed for the convex surface.

The dynamics of the GNA duplexes incorporating the porphyrin moiety as represented by the evolution of the backbone rmsd as a function of time (Figure 5.21) exhibit the characteristic periodic fluctuation corresponding to a twisting/ untwisting mode of the duplex. This mode inherent to duplex GNA is thought to explain, in part, the observation that duplex formation of GNA is entropically less penalizing than the corresponding process in DNA.

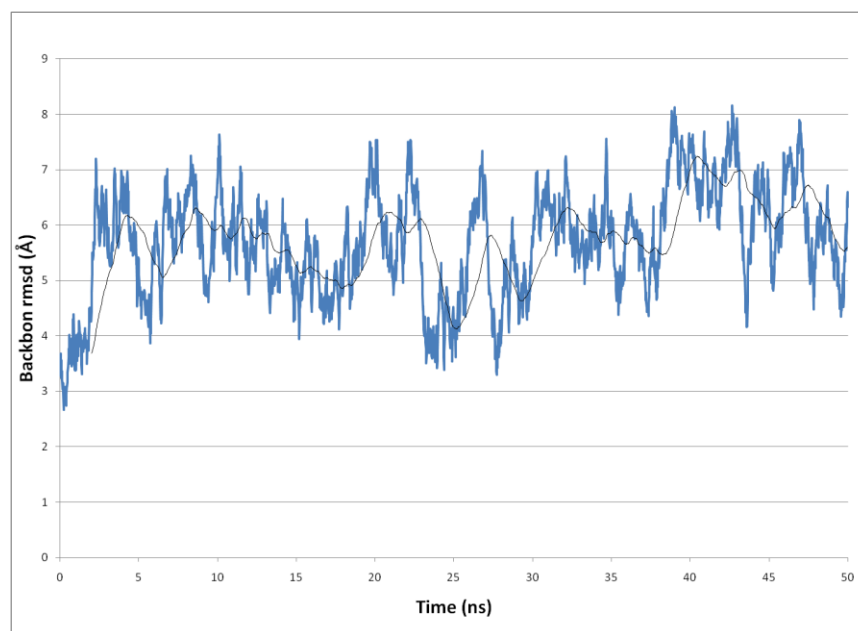


Figure 5.21- Backbone rmsd versus time observed in the simulation of entry 7.

The calculated structures also provide insights into the structural origin of the experimentally observed decrease in melting temperatures (Table 5.3) associated with porphyrin/ abase incorporation within the GNA duplex. It can be seen that in both entry 3 and entry 7, there appears to be little or no disruption of hydrogen bonding between the flanking A-T base pairs. This is another indication that structurally there is little

perturbation of the overall structure caused by incorporation of the porphyrin group across from an abasic site. Thus, the decrease in melting temperature can mostly be attributed to the loss of two (entry 3) or four (entry 7) hydrogen bonds resulting from the replacement of a nucleobase with a porphyrin, which is not compensated for by the stacking interactions.

Another interesting experimental observation is the additional decrease in melting temperature upon chelation of Zn^{2+} by porphyrin, while the formation of the Ni^{2+} complex has little effect on melting temperature. This can be rationalized by considering the coordination geometry preferred by the respective metals. Zn^{2+} prefers an octahedral arrangement¹¹⁴ which cannot be accommodated within the GNA duplex, thus destabilizing the position of the Zn^{2+} -porphyrin in the duplex due to the close π -stacking with either the canonical base pairs and/or another porphyrin base and potentially favoring a flip of the porphyrin out of the duplex to recruit water molecules to serve as axial ligands. In the case of Ni^{2+} chelation, a square planar arrangement is preferred.¹¹⁴ This allows the Ni^{2+} -porphyrin complex to be easily incorporated within the GNA duplex with little effect on melting temperature.

5.5 Conclusions

GNA is a nucleic acid analog which is synthetically simple while displaying interesting structural and thermodynamic properties. Through the use of MD simulation, we have shown that underlying flexibility of the GNA duplex is likely responsible for the decrease in entropic penalty associated with annealing when compared to DNA. This is counterintuitive to most of the efforts made in the synthetic nucleic acid field to decrease

this entropic penalty as a means to increase the stability of the duplex. The idea behind locked nucleic acids (LNA) is by restriction of the backbone in the single strand you effectively reduce the difference in microstates between the single stranded and duplex forms. Reduction of molecular flexibility is also a common practice in drug design, where the replacement of rotatable bonds leads to a decrease in the entropic penalty associated with active site binding. GNA appears to work in the opposite manner, trading off some of the enthalpic contributions to free energy inherent to DNA for a more flexible structure which resists melting. In GNA, thermal energy results in enhanced global motions rather than the rupturing of hydrogen bonds. Structures along the 100 ns trajectory of 16mer-GNA show differences up to 10° in their average base pair twist while the hydrogen bonding remains intact. In DNA, large changes in average base pair twist tend to correspond to terminal base pair melting, not large scale twisting/ untwisting modes. In short, this work has shown that two flexible nucleic acid strands can be created which retain this flexibility in its duplex form. This leads to the difference of the available microstates accessible in single stranded and duplex forms to decrease. The energy GNA gains through this decreased entropic penalty greatly outweighs what it loses enthalpically in comparison to DNA.

GNA also possesses an interesting structure which may prove useful as a scaffold for the arrangement of chromophores. A slipped cofacial geometry of adjacent porphyrins is detected in the 50 ns MD simulations. Evidence of such an arrangement had previously been predicted through the use of UV-Vis spectroscopy and CD spectroscopy.¹⁰³ Such an arrangement can could be useful in light harvesting and charge

transport applications. Furthermore, the arrangement of the porphyrin base stacks can be modulated by the inclusion of different coordinated metal ions.

REFERENCES

- (1) Rahn, R. O.; Hosszu, J. L. Photoproduct Formation in DNA at Low Temperatures. *Photochem Photobiol*, **1968**, 8, 53.
- (2) Woodward, R. B.; Hoffmann, R. Stereochemistry of Electrocyclic Reactions. *J Am Chem Soc*, **1965**, 87, 395.
- (3) Woodward, R. B.; Hoffmann, R. Conservation of Orbital Symmetry. *Angew Chem Int Edit*, **1969**, 8, 781.
- (4) Cadet, J.; Sage, E.; Douki, T. Ultraviolet radiation-mediated damage to cellular DNA. *Mutat Res-Fund Mol M*, **2005**, 571, 3.
- (5) Sancar, A. Photolyase and cryptochrome blue-light photoreceptors. *Adv Protein Chem*, **2004**, 69, 73.
- (6) Glickman, B. W.; Schaaper, R. M.; Haseltine, W. A.; Dunn, R. L.; Brash, D. E. The C-C (6-4) Uv Photoproduct Is Mutagenic in Escherichia-Coli. *P Natl Acad Sci USA*, **1986**, 83, 6945.
- (7) Mitchell, D. L. The Relative Cyto-Toxicity of (6-4) Photoproducts and Cyclobutane Dimers in Mammalian-Cells. *Photochem Photobiol*, **1988**, 48, 51.
- (8) Friedberg, E.; Walker, G.; Siede, W.; Wood, R.; Schultz, R.; Ellenberger, T. *DNA repair and mutagenesis*; 2nd ed.; ASM Press, 2006; Vol. 3.
- (9) Svoboda, D. L.; Briley, L. P.; Vos, J. M. H. Defective bypass replication of a leading strand cyclobutane thymine dimer in xeroderma pigmentosum variant cell extracts. *Cancer Res*, **1998**, 58, 2445.
- (10) Costa, R.; Chigancas, V.; Galhardo, R.; Carvalho, H.; Menck, C. The eukaryotic nucleotide excision repair pathway. *Biochemie*, **2003**, 85.
- (11) Mees, A.; Klar, T.; Gnau, P.; Hennecke, U.; Eker, A. P. M.; Carell, T.; Essen, L. O. Crystal structure of a photolyase bound to a CPD-like DNA lesion after in situ repair. *Science*, **2004**, 306, 1789.
- (12) Sancar, A. Structure and function of DNA photolyase and cryptochrome blue-light photoreceptors. *Chem Rev*, **2003**, 103, 2203.

- (13) Sancar, G. B.; Smith, F. W.; Sancar, A. Binding of Escherichia-Coli DNA Photolyase to Uv-Irradiated DNA. *Biochemistry-Us*, **1985**, *24*, 1849.
- (14) Husain, I.; Sancar, A. Binding of Escherichia-Coli DNA Photolyase to a Defined Substrate Containing a Single T Reversible T Dimer. *Nucleic Acids Res*, **1987**, *15*, 1109.
- (15) Diogo, H. P.; Dias, A. R.; Dhalla, A.; Dapiedade, M. E. M.; Begley, T. P. Mechanistic Studies on DNA Photolyase .4. The Enthalpy of Cleavage of a Model Photodimer. *J Org Chem*, **1991**, *56*, 7340.
- (16) Harrison, C. B.; O'Neil, L. L.; Wiest, O. Computational studies of DNA photolyase. *J Phys Chem A*, **2005**, *109*, 7001.
- (17) Durbeej, B.; Eriksson, L. A. Thermodynamics of the photoenzymic repair mechanism studied by density functional theory. *J Am Chem Soc*, **2000**, *122*, 10126.
- (18) Masson, F.; Laino, T.; Tavernelli, I.; Rothlisberger, U.; Hutter, J. Computational study of thymine dimer radical anion splitting in the self-repair process of duplex DNA. *J Am Chem Soc*, **2008**, *130*, 3443.
- (19) Hitomi, K.; Nakamura, H.; Kim, S. T.; Mizukoshi, T.; Ishikawa, T.; Iwai, S.; Todo, T. Role of two histidines in the (6-4) photolyase reaction. *J Biol Chem*, **2001**, *276*, 10103.
- (20) Wittung, P.; Nielsen, P. E.; Buchardt, O.; Egholm, M.; Norden, B. DNA-Like Double Helix Formed by Peptide Nucleic-Acid. *Nature*, **1994**, *368*, 561.
- (21) Eschenmoser, A. Chemical etiology of nucleic acid structure. *Science*, **1999**, *284*, 2118.
- (22) Kumar, R.; Singh, S. K.; Koshkin, A. A.; Rajwanshi, V. K.; Meldgaard, M.; Wengel, J. The first analogues of LNA (Locked Nucleic Acids): Phosphorothioate-LNA and 2'-thio-LNA. *Bioorg Med Chem Lett*, **1998**, *8*, 2219.
- (23) Koshkin, A. A.; Singh, S. K.; Nielsen, P.; Rajwanshi, V. K.; Kumar, R.; Meldgaard, M.; Olsen, C. E.; Wengel, J. LNA (Locked Nucleic Acids): Synthesis of the adenine, cytosine, guanine, 5-methylcytosine, thymine and uracil bicyclonucleoside monomers, oligomerisation, and unprecedented nucleic acid recognition. *Tetrahedron*, **1998**, *54*, 3607.
- (24) Singh, S. K.; Nielsen, P.; Koshkin, A. A.; Wengel, J. LNA (locked nucleic acids): synthesis and high-affinity nucleic acid recognition. *Chem Commun*, **1998**, 455.

- (25) Ebert, M. O.; Mang, C.; Krishnamurthy, R.; Eschenmoser, A.; Jaun, B. The Structure of a TNA-TNA Complex in Solution: NMR Study of the Octamer Duplex Derived from α -(L)-Threofuranosyl-(3' \rightarrow 2')-CGAATTCG. *J Am Chem Soc*, **2008**, *130*, 15105.
- (26) Ferencic, M.; Reddy, G.; Wu, X. L.; Guntha, S.; Nandy, J.; Krishnamurthy, R.; Eschenmoser, A. Base-pairing systems related to TNA containing phosphoramidate linkages: Synthesis of building blocks and pairing properties. *Chem Biodivers*, **2004**, *1*, 939.
- (27) Eschenmoser, A. The TNA-family of nucleic acid systems: Properties and prospects. *Origins Life Evol B*, **2004**, *34*, 277.
- (28) Pallan, P. S.; Wilds, C. J.; Wawrzak, Z.; Krishnamurthy, R.; Eschenmoser, A.; Egli, M. Why does TNA cross-pair more strongly with RNA than with DNA? An answer from X-ray analysis. *Angew Chem Int Edit*, **2003**, *42*, 5893.
- (29) Schoning, K. U.; Scholz, P.; Wu, X. L.; Guntha, S.; Delgado, G.; Krishnamurthy, R.; Eschenmoser, A. The α -L-threofuranosyl-(3' \rightarrow 2')-oligonucleotide system ('TNA'): Synthesis and pairing properties. *Helv Chim Acta*, **2002**, *85*, 4111.
- (30) Wu, X. L.; Guntha, S.; Ferencic, M.; Krishnamurthy, R.; Eschenmoser, A. Base-pairing systems related to TNA: α -threofuranosyl oligonucleotides containing phosphoramidate linkages. *Org Lett*, **2002**, *4*, 1279.
- (31) Wu, X. L.; Delgado, G.; Krishnamurthy, R.; Eschenmoser, A. 2,6-diaminopurine in TNA: Effect on duplex stabilities and on the efficiency of template-controlled ligations. *Org Lett*, **2002**, *4*, 1283.
- (32) Schoning, K. U.; Scholz, P.; Guntha, S.; Wu, X.; Krishnamurthy, R.; Eschenmoser, A. Chemical etiology of nucleic acid structure: The α -threofuranosyl-(3' \rightarrow 2') oligonucleotide system. *Science*, **2000**, *290*, 1347.
- (33) Meggers, E.; Zhang, L. Synthesis and Properties of the Simplified Nucleic Acid Glycol Nucleic Acid. *Accounts Chem Res*, **2010**, *43*, 1092.
- (34) Schlegel, M. K.; Essen, L. O.; Meggers, E. Atomic resolution duplex structure of the simplified nucleic acid GNA. *Chem Commun*, **2010**, *46*, 1094.
- (35) Schlegel, M. K.; Meggers, E. Improved Phosphoramidite Building Blocks for the Synthesis of the Simplified Nucleic Acid GNA. *J Org Chem*, **2009**, *74*, 4615.
- (36) Schlegel, M. K.; Xie, X. L.; Zhang, L. L.; Meggers, E. Insight into the high Duplex Stability of the Simplified Nucleic Acid GNA. *Angew Chem Int Edit*, **2009**, *48*, 960.

- (37) Schlegel, M. K.; Zhang, L. L.; Pagano, N.; Meggers, E. Metal-mediated base pairing within the simplified nucleic acid GNA. *Org Biomol Chem*, **2009**, *7*, 476.
- (38) Schlegel, M. K.; Essen, L. O.; Meggers, E. Duplex structure of a minimal nucleic acid. *J Am Chem Soc*, **2008**, *130*, 8158.
- (39) Schlegel, M. K.; Peritz, A. E.; Kittigowittana, K.; Zhang, L. L.; Meggers, E. Duplex formation of the simplified nucleic acid GNA. *Chembiochem*, **2007**, *8*, 927.
- (40) Zhang, L. L.; Peritz, A. E.; Carroll, P. J.; Meggers, E. Synthesis of glycol nucleic acids. *Synthesis-Stuttgart*, **2006**, 645.
- (41) Schreier, W. J.; Schrader, T. E.; Koller, F. O.; Gilch, P.; Crespo-Hernandez, C. E.; Swaminathan, V. N.; Carell, T.; Zinth, W.; Kohler, B. Thymine dimerization in DNA is an ultrafast photoreaction. *Science*, **2007**, *315*, 625.
- (42) Zhang, R. B.; Eriksson, L. A. A triplet mechanism for the formation of cyclobutane pyrimidine dimers in UV-irradiated DNA. *J Phys Chem B*, **2006**, *110*, 7556.
- (43) Martinez, J. M.; Elmroth, S. K. C.; Kloo, L. Influence of sodium ions on the dynamics and structure of single-stranded DNA oligomers: A molecular dynamics study. *J Am Chem Soc*, **2001**, *123*, 12279.
- (44) Biyani, M.; Nishigaki, K. Single-strand conformation polymorphism (SSCP) of oligodeoxyribonucleotides: An insight into solution structural dynamics of DNAs provided by gel electrophoresis and molecular dynamics simulations. *J Biochem*, **2005**, *138*, 363.
- (45) Coates, G. W.; Dunn, A. R.; Henling, L. M.; Ziller, J. W.; Lobkovsky, E. B.; Grubbs, R. H. Phenyl-perfluorophenyl stacking interactions: Topochemical[2+2] photodimerization and photopolymerization of olefinic compounds. *J Am Chem Soc*, **1998**, *120*, 3641.
- (46) Zhang, X. N.; Cote, A. P.; Matzger, A. J. Synthesis and structure of fused alpha-oligothiophenes with up to seven rings. *J Am Chem Soc*, **2005**, *127*, 10502.
- (47) 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*, **2007**, *111*, 7409.
- (48) Johnson, A. T.; Wiest, O. Structure and dynamics of poly(T) single-strand DNA: Implications toward CPD formation. *J Phys Chem B*, **2007**, *111*, 14398.

- (49) D.A. Case, T. A. D., T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R.; Luo, K. M. M., D.A. Pearlman, M. Crowley, R.C. Walker, W. Zhang, B. Wang, S.; Hayik, A. R., G. Seabra, K.F. Wong, F. Paesani, X. Wu, S. Brozell, V. Tsui, H.; Gohlke, L. Y., C. Tan, J. Mongan, V. Hornak, G. Cui, P. Beroza, D.H. Mathews, C.; Schafmeister, W. S. R., and P.A. Kollman *AMBER 9*, **2006**, University of California, San Francisco,
- (50) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. *J. Am. Chem. Soc.*, **1995**, *117*, 5179.
- (51) Wang, J. M.; Cieplak, P.; Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comp. Chem.*, **2000**, *21*, 1049.
- (52) Cheatham, T. E., 3rd; Young, M. A. Molecular dynamics simulation of nucleic acids: successes, limitations, and promise. *Biopolymers*, **2000**, *56*, 232.
- (53) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J Mol Graphics*, **1996**, *14*, 33.
- (54) Ramamurthy, V.; Venkatesan, K. Photochemical-Reactions of Organic-Crystals. *Chem Rev*, **1987**, *87*, 433.
- (55) Hariharan, M.; Lewis, F. D. Context-dependent photodimerization in isolated thymine-thymine steps in DNA. *J Am Chem Soc*, **2008**, *130*, 11870.
- (56) Law, Y. K.; Azadi, J.; Crespo-Hernandez, C. E.; Olmon, E.; Kohler, B. Predicting thymine dimerization yields from molecular dynamics simulations. *Biophys J*, **2008**, *94*, 3590.
- (57) McCullagh, M.; Hariharan, M.; Lewis, F. D.; Markovitsi, D.; Douki, T.; Schatz, G. C. Conformational Control of TT Dimerization in DNA Conjugates. A Molecular Dynamics Study. *J Phys Chem B*, **2010**, *114*, 5215.
- (58) 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*, **2010**, *132*, 12856.
- (59) Todo, T.; Takemori, H.; Ryo, H.; Ihara, M.; Matsunaga, T.; Nikaido, O.; Sato, K.; Nomura, T. A New Photoreactivating Enzyme That Specifically Repairs Ultraviolet Light-Induced (6-4)Photoproducts. *Nature*, **1993**, *361*, 371.
- (60) Cleaver, J. E.; Cortes, F.; Karentz, D.; Lutze, L. H.; Morgan, W. F.; Player, A. N.; Vuksanovic, L.; Mitchell, D. L. The Relative Biological

Importance of Cyclobutane and (6-4) Pyrimidine-Pyrimidone Dimer Photoproducts in Human-Cells - Evidence from a Xeroderma Pigmentosum Revertant. *Photochem Photobiol*, **1988**, 48, 41.

(61) Kamiya, H.; Iwai, S.; Kasai, H. The (6-4) photoproduct of thymine-thymine induces targeted substitution mutations in mammalian cells. *Nucleic Acids Res*, **1998**, 26, 2611.

(62) Kim, S. T.; Sancar, A. Effect of Base, Pentose, and Phosphodiester Backbone Structures on Binding and Repair of Pyrimidine Dimers by Escherichia-Coli DNA Photolyase. *Biochemistry-U.S.*, **1991**, 30, 8623.

(63) Hitomi, K.; Kim, S. T.; Iwai, S.; Harima, N.; Otsoshi, E.; Ikenaga, M.; Todo, T. Binding and catalytic properties of Xenopus (6-4) photolyase. *J Biol Chem*, **1997**, 272, 32591.

(64) Zhao, X. D.; Liu, J. Q.; Hsu, D. S.; Zhao, S. Y.; Taylor, J. S.; Sancar, A. Reaction mechanism of (6-4) photolyase. *J Biol Chem*, **1997**, 272, 32580.

(65) Schleicher, E.; Hitomi, K.; Kay, C. W. M.; Getzoff, E. D.; Todo, T.; Weber, S. Electron nuclear double resonance differentiates complementary roles for active site histidines in (6-4) photolyase. *J Biol Chem*, **2007**, 282, 4738.

(66) Maul, M. J.; Barends, T. R. M.; Glas, A. F.; Cryle, M. J.; Domratcheva, T.; Schneider, S.; Schlichting, I.; Carell, T. Crystal Structure and Mechanism of a DNA (6-4) Photolyase. *Angew Chem Int Edit*, **2008**, 47, 10076.

(67) Clivio, P.; Fourrey, J. L.; Gasche, J.; Favre, A. DNA Photodamage Mechanistic Studies - Characterization of a Thietane Intermediate in a Model Reaction Relevant to 6-4 Lesions. *J Am Chem Soc*, **1991**, 113, 5481.

(68) Li, J.; Liu, Z. Y.; Tan, C.; Guo, X. M.; Wang, L. J.; Sancar, A.; Zhong, D. P. Dynamics and mechanism of repair of ultraviolet-induced (6-4) photoproduct by photolyase. *Nature*, **2010**, 466, 887.

(69) Domratcheva, T.; Schlichting, I. Electronic Structure of (6-4) DNA Photoproduct Repair Involving a Non-Oxetane Pathway. *J Am Chem Soc*, **2009**, 131, 17793.

(70) Heelis, P. F.; Liu, S. B. Photoenzymic repair of the DNA 6-4 photoproduct - A density functional theory and semiempirical study. *J Am Chem Soc*, **1997**, 119, 2936.

(71) Harrison, C., University of Notre Dame, 2007.

- (72) Lukeman, M.; Scaiano, J. C. Carbanion-mediated photocages: Rapid and efficient photorelease with aqueous compatibility. *J Am Chem Soc*, **2005**, *127*, 7698.
- (73) M. J. Frisch, G. W. T., H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox *Gaussian 09*, **2009**, Gaussian, Inc., Wallingford, CT
- (74) Park, T. K.; Schroeder, J.; Rebek, J. New Molecular Complements to Imides - Complexation of Thymine Derivatives. *J Am Chem Soc*, **1991**, *113*, 5125.
- (75) Goodman, M. S.; Rose, S. D. Molecular Recognition of a Pyrimidine Dimer and Photosensitized Dimer Splitting by a Macrocyclic Bis(Diaminopyridine). *J Am Chem Soc*, **1991**, *113*, 9380.
- (76) Goodman, M. S.; Rose, S. D. Photosensitized Pyrimidine Dimer Splitting by a Methoxyindole Bound to a Dimer-Recognizing Macrocycle. *J Org Chem*, **1992**, *57*, 3268.
- (77) Epple, R.; Wallenborn, E. U.; Carell, T. Investigation of flavin-containing DNA-repair model compounds. *J Am Chem Soc*, **1997**, *119*, 7440.
- (78) Schwogler, A.; Burgdorf, L. T.; Carell, T. Self-repairing DNA based on a reductive electron transfer through the base stack. *Angew Chem Int Edit*, **2000**, *39*, 3918.
- (79) Schwogler, A.; Carell, T. Toward catalytically active oligonucleotides: Synthesis of a flavin nucleotide and its incorporation into DNA. *Org Lett*, **2000**, *2*, 1415.
- (80) Shionoya, M.; Kimura, E.; Shiro, M. A New Ternary Zinc(II) Complex with [12]Anen(4) (=1,4,7,10-Tetraazacyclododecane) and Azt (=3'-Azido-3'-Deoxythymidine) - Highly Selective Recognition of Thymidine and Its Related Nucleosides by a Zinc(II) Macrocyclic Tetraamine Complex with Novel Complementary Associations. *J Am Chem Soc*, **1993**, *115*, 6730.

- (81) Wiest, O.; Harrison, C. B.; Saettel, N. J.; Cibulka, R.; Sax, M.; Konig, B. Design, synthesis, and evaluation of a biomimetic artificial photolyase model. *J Org Chem*, **2004**, *69*, 8183.
- (82) Park, H. W.; Kim, S. T.; Sancar, A.; Deisenhofer, J. Crystal-Structure of DNA Photolyase from Escherichia-Coli. *Science*, **1995**, *268*, 1866.
- (83) O'Neil, L. L.; Wiest, O. Structures and energetics of base flipping of the thymine dimer depend on DNA sequence. *J Phys Chem B*, **2008**, *112*, 4113.
- (84) O'Neil, L. L.; Wiest, O. Sequence dependence in base flipping: experimental and computational studies. *Org Biomol Chem*, **2008**, *6*, 485.
- (85) O'Neil, L. L.; Grossfield, A.; Wiest, O. Base flipping of the thymine dimer in duplex DNA. *J Phys Chem B*, **2007**, *111*, 11843.
- (86) Mu, W. M.; Han, Q. K.; Luo, Z. F.; Wang, Y. Z. Production of cis-syn thymine-thymine cyclobutane dimer oligonucleotide in the presence of acetone photosensitizer. *Anal Biochem*, **2006**, *353*, 117.
- (87) Holman, M. R.; Ito, T.; Rokita, S. E. Self-repair of thymine dimer in duplex DNA. *J Am Chem Soc*, **2007**, *129*, 6.
- (88) Obika, S.; Nanbu, D.; Hari, Y.; Morio, K.; In, Y.; Ishida, T.; Imanishi, T. Synthesis of 2'-O,4'-C-methyleneuridine and -cytidine. Novel bicyclic nucleosides having a fixed C-3,-endo sugar puckering. *Tetrahedron Lett*, **1997**, *38*, 8735.
- (89) Nielsen, P. E.; Egholm, M.; Berg, R. H.; Buchardt, O. Sequence-Selective Recognition of DNA by Strand Displacement with a Thymine-Substituted Polyamide. *Science*, **1991**, *254*, 1497.
- (90) Johnson, A. T.; Schlegel, M. K.; Meggers, E.; Essen, L. O.; Wiest, O. On the Structure and Dynamics of Duplex DNA. *Submitted*.
- (91) Lavery, R.; Sklenar, H. Defining the Structure of Irregular Nucleic-Acids - Conventions and Principles. *J. Biomol. Struct. Dyn.*, **1989**, *6*, 655.
- (92) Lavery, R.; Sklenar, H. The Definition of Generalized Helicoidal Parameters and of Axis Curvature for Irregular Nucleic-Acids. *J. Biomol. Struct. Dyn.*, **1988**, *6*, 63.
- (93) Bevan, D. R.; Li, L. P.; Pedersen, L. G.; Darden, T. A. Molecular dynamics simulations of the d(CCAACGTTGG)(2) decamer: Influence of the crystal environment. *Biophys J*, **2000**, *78*, 668.
- (94) Cheatham, T. E.; Kollman, P. A. Molecular dynamics simulation of nucleic acids. *Annu Rev Phys Chem*, **2000**, *51*, 435.

- (95) Cheatham, T. E.; Miller, J. L.; Spector, T. I.; Cieplak, P.; Kollman, P. A. Molecular dynamics simulations on nucleic acid systems using the Cornell et al force field and particle mesh Ewald electrostatics. *Acs Sym Ser*, **1998**, 682, 285.
- (96) Cheatham, T. E.; Srinivasan, J.; Case, D. A.; Kollman, P. A. Molecular dynamics and continuum solvent studies of the stability of polyG-polyC and polyA-polyT DNA duplexes in solution. *J Biomol Struct Dyn*, **1998**, 16, 265.
- (97) De Winter, H.; Lescrinier, E.; Van Aerschot, A.; Herdewijn, P. Molecular dynamics simulation to investigate differences in minor groove hydration of HNA/RNA hybrids as compared to HNA/DNA complexes. *J Am Chem Soc*, **1998**, 120, 5381.
- (98) Lynch, S. R.; Liu, H. B.; Gao, J. M.; Kool, E. T. Toward a designed, functioning genetic system with expanded-size base pairs: Solution structure of the eight-base xDNA double helix. *J Am Chem Soc*, **2006**, 128, 14704.
- (99) He, W.; Hatcher, E.; Balaeff, A.; Beratan, D. N.; Gil, R. R.; Madrid, M.; Achim, C. Solution structure of a peptide nucleic acid duplex from NMR data: Features and limitations. *J Am Chem Soc*, **2008**, 130, 13264.
- (100) Ivanova, A.; Rosch, N. The structure of LNA : DNA hybrids from molecular dynamics simulations: The effect of locked nucleotides. *J Phys Chem A*, **2007**, 111, 9307.
- (101) O'Neil, L. L.; Wiest, O. Molecular dynamics simulations of a DNA strand containing two thymine dimers. *Abstr Pap Am Chem S*, **2005**, 230, U1362.
- (102) O'Daniel, P. I.; Jefferson, M.; Wiest, O.; Seley-Radtke, K. L. A Computational Study of Expanded Heterocyclic Nucleosides in DNA. *J Biomol Struct Dyn*, **2008**, 26, 283.
- (103) Zhou, H. L.; Johnson, A. T.; Wiest, O.; Zhang, L. Incorporation of Porphyrin Acetylides into Duplexes of the Simplified Nucleic Acid GNA. *Submitted*.
- (104) Malinovskii, V. L.; Wenger, D.; Haner, R. Nucleic acid-guided assembly of aromatic chromophores. *Chem Soc Rev*, **2010**, 39, 410.
- (105) Gao, J. M.; Watanabe, S.; Kool, E. T. Modified DNA analogues that sense light exposure with color changes. *J Am Chem Soc*, **2004**, 126, 12748.
- (106) Varghese, R.; Wagenknecht, H. A. DNA as a supramolecular framework for the helical arrangements of chromophores: towards photoactive DNA-based nanomaterials. *Chem Commun*, **2009**, 2615.

- (107) Wagenknecht, H. A. Helical Arrangement of Porphyrins along DNA: Towards Photoactive DNA-Based Nanoarchitectures. *Angew Chem Int Edit*, **2009**, 48, 2838.
- (108) Mayer-Enthart, E.; Wagner, C.; Barbaric, J.; Wagenknecht, H. A. Helical self-assembled chromophore clusters based on DNA-like architecture. *Tetrahedron*, **2007**, 63, 3434.
- (109) Nguyen, T.; Brewer, A.; Stulz, E. Duplex Stabilization and Energy Transfer in Zipper Porphyrin-DNA. *Angew Chem Int Edit*, **2009**, 48, 1974.
- (110) Frisch, M. J. T.; G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A. *Gaussian 03, Revision C.02*, **2004**, Wallingford, CT
- (111) DeLano, W. L.; Lam, J. W. PyMOL: A communications tool for computational models. *Abstr Pap Am Chem S*, **2005**, 230, U1371.
- (112) Johannsen, S.; Megger, N.; Bohme, D.; Sigel, R. K. O.; Muller, J. Solution structure of a DNA double helix with consecutive metal-mediated base pairs. *Nat Chem*, **2010**, 2, 229.
- (113) Li, L. W.; Uversky, V. N.; Dunker, A. K.; Meroueh, S. O. A computational investigation of allostery in the catabolite activator protein. *J Am Chem Soc*, **2007**, 129, 15668.
- (114) Pasterna.Rf; Francesc.L; Raff, D.; Spiro, E. Aggregation of Nickel(II), Copper(II), and Zinc(II) Derivatives of Water-Soluble Porphyrins. *Inorg Chem*, **1973**, 12, 2606.