DNA Repair by DNA with Visible Light: Investigations and Implications

by

Adam Barlev

B.Sc., University of British Columbia, 2008

Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in the Department of Chemistry Faculty of Science

© Adam Barlev 2017

SIMON FRASER UNIVERSITY

Fall 2017

Copyright in this work rests with the author. Please ensure that any reproduction or re-use is done in accordance with the relevant national copyright legislation.
Approval

Name: Adam Barlev

Degree: Doctor of Philosophy

Title: DNA Repair by DNA with Visible Light: Investigations and Implications

Examinining Committee: Chair: Dr. Hogan Yu

Assistant Professor

Dr. Dipankar Sen
Senior Supervisor
Professor

Dr. Peter Unrau
Supervisor
Professor

Dr. Gary Leach
Supervisor
Professor

Dr. Andrew Bennet
Supervisor
Professor

Dr. Edgar Young
Internal Examiner
Professor
Molecular Biology and Biochemistry

Dr. Cynthia Burrows
External Examiner
Distinguished Professor
Department of Chemistry
University of Utah

Date Defended/Approved: April 27th, 2017
Abstract

The DNAzyme UV1C was selected previously on the basis of its ability to utilize UV-B light to catalyze the repair of a cis-syn cyclobutane thymine dimer in which no phosphodiester linkage exists between the dimerized thymines. Systematic replacement of each of nine guanines in and around the active site by the guanine analog 6-MI allowed the expansion of the photocatalytic cross section throughout the UV-A and to the edge of the visible. The behaviour of these mutants fell into 3 classes. In one class, replacement of guanines in the quadruplex did not disrupt the wild-type activity. In another class, quadruplex positions, when replaced with 6-MI, led to a decrease in activity in the UV-B but new activity in the UV-A, providing strong evidence for exactly which guanine residues are catalytic in the DNAzyme. Most surprisingly, the G-23 position, thought to be near the active site but not catalytic in UV1C, when replaced with 6-MI, leads to a full retention of activity in the UV-B with the strongest gain of activity in the UV-A. Further modifications to the G-23 position pushed its activity to maximize in the visible, but also ultimately disrupted the quadruplex-dependent activity in the UV-B.

While selected against a model thymine dimer substrate, the DNAzyme is also shown to have photocatalytic activity on a bona fide DNA substrate. The continuity of the natural DNA substrate allows us to measure for the first time the effect of the UV1C DNAzyme on the rate of both thymine dimer formation as well as the rate of repair. When compared to double-stranded and single-stranded controls, at its photostationary state, UV1C leads to an overall reduction in fraction of dimerized thymines. Surprisingly, UV1C catalyzes both the repair and formation of thymine dimers in natural DNA, but more slowly than the model substrate that it was selected against. Together, these results shed further light on the emerging field of protein-independent thymine dimer repair.

Arguments connecting the self-repair properties of DNA to the RNA world and prebiotic chemistry are offered.

Keywords: DNA repair; Photocatalysis; DNAzyme; Thymine Dimer; G-quadruplex; RNA world
Dedication

I would like to dedicate this thesis to the science fiction author Neal Patrick Stephenson, in particular for his novel *The Diamond Age*. 
Acknowledgements

This thesis would not have been possible without the inspiration of Dr. Fabini, Dr. Ledbetter, Dr. Smith, Dr. McIntosh, and Dr. Nagata, for whom I’ll be forever grateful.
Table of Contents

Approval .......................................................................................................................... ii
Abstract ........................................................................................................................... iii
Dedication ...................................................................................................................... iv
Acknowledgements ...................................................................................................... v
Table of Contents .......................................................................................................... vi
List of Tables ................................................................................................................ viii
List of Figures .............................................................................................................. ix
List of Acronyms .......................................................................................................... x
Glossary ........................................................................................................................ xii

Chapter 1. Introduction .................................................................................................. 1
1.1. Nucleic Acid Structure ......................................................................................... 2
1.2. G-Quadruplex Structure ....................................................................................... 5
1.3. Synthetic Nucleic Acids ....................................................................................... 8
1.4. Catalytic nucleic acids ........................................................................................ 14
1.5. In-vitro selection .................................................................................................. 15
1.6. The RNA world Hypothesis ................................................................................ 17
1.7. Photochemistry and charge transfer in nucleic acids ........................................... 21
1.8. Pyrimidine dimer formation and structure .......................................................... 24
1.9. Photolyase repairs pyrimidine dimers ................................................................. 28

Chapter 2. Towards the visible ..................................................................................... 33
2.1. Introduction .......................................................................................................... 33
2.2. Results and Discussion ....................................................................................... 37
  2.2.1. Three Functional Classes of UV1C G→6-MI Point Mutants. ......................... 37
  2.2.2. Is Efficient Photoreactivation by the G23 G→6-MI DNAzyme Wholly a “Proximity” Effect? ................................................................. 42
  2.2.3. A Variety of Chromophores Can Functionally Substitute for Guanine at the G23 Position of UV1C. ............................................................... 44
2.3. Conclusions ......................................................................................................... 47
2.4. Methods ............................................................................................................... 48
  2.4.1. Oligonucleotides ............................................................................................ 48
  2.4.2. CD Spectroscopy. ........................................................................................ 51
  2.4.3. UV–Vis Spectroscopy. .................................................................................. 51
  2.4.4. Irradiation Sample Preparation ................................................................... 52
  2.4.5. Laser Irradiation. ........................................................................................ 53
  2.4.6. Gel Analysis. ................................................................................................ 54
  2.4.7. Kinetic Analysis. .......................................................................................... 54

Chapter 3. Repair of an authentic DNA substrate ....................................................... 56
3.1. Introduction .......................................................................................................... 56
3.2. Methods .......................................................................................................................... 60
  3.2.1. Oligonucleotides ......................................................................................................... 60
  3.2.2. Thymine Dimer Generation and HPLC Purification of Thymine Dimer Containing Oligonucleotides ................................................................. 60
  3.2.3. Ligation of LDP and LMP Substrates ......................................................................... 61
  3.2.4. Permanganate footprinting of LMP and LDP ................................................................. 62
  3.2.5. Laser Irradiation Experiments .................................................................................... 62
  3.2.6. Kinetic Analysis ........................................................................................................ 63
3.3. Results .................................................................................................................................. 64
  3.3.1. UV1C-Mediated Repair of an Ungapped DNA Substrate, LDP ....................................... 64
  3.3.2. UV1C Weakly Catalyzes Thymine Dimer Formation but Strongly Catalyzes Photoreactivation of the Thymine Dimer with 305 nm Irradiation .................................................... 70
  3.3.3. UV1C's Photoreactivation of the Ungapped Substrate, LDP, is ∼30 fold Slower than the Gapped Substrate, TDP ................................................................. 72
3.4. Discussion ............................................................................................................................. 73

Chapter 4. Conclusions and Future Work ................................................................................. 77
4.1. Discussion .............................................................................................................................. 77
  4.1.1. Possible Explanations for the Discrepancy between TDP and LDP repair rates ............ 77
  4.1.2. Formation of thymine dimers in UV1C ...................................................................... 81
  4.1.3. On UV1C, Prebiotic Chemistry and the RNA world .................................................. 82
4.2. Future work ........................................................................................................................... 88
  4.2.1. On time-resolved spectroscopy of the UV1C system ................................................ 88
  4.2.2. Modified bases in UV1C ............................................................................................ 88
  4.2.3. A Selection strategy for modifying the substrate specificity of UV1C ......................... 91
  4.2.4. Potential applications for the UV1C-TDP system ....................................................... 95

References .................................................................................................................................. 99

Appendix . Supplementary Figures ......................................................................................... 108
List of Tables

Table 2-1.  Rates of thymine dimer repair at 305 and 345 nm........................................ 39
Table 3-1.  Kinetic rate constants and photostationary states under irradiation at 305nm
....................................................................................................................... 71
Table 3-2.  Comparison of the photoreactivation kinetics of the TDP and LDP
substrates................................................................................................................ 72
List of Figures

Figure 1-1. DNA structure and canonical bonding.......................................................... 4
Figure 1-2. Guanine-quadruplex structures ..................................................................... 6
Figure 1-3. Non-standard purine bases ............................................................................ 9
Figure 1-4. MTHF and 6-MI ......................................................................................... 10
Figure 1-5. Artificial expanded genetic alphabets ........................................................... 13
Figure 1-6. Generalized scheme for in-vitro selection. .................................................. 16
Figure 1-7. Roadmap from the RNA world to the present day ...................................... 19
Figure 1-8. Light absorption process ............................................................................. 21
Figure 1-9. Thymine dimer formation ............................................................................ 26
Figure 1-10. Thymine dimer species ............................................................................. 27
Figure 1-11. The main types of cofactors found in photolyase ...................................... 30
Figure 1-12. Photolyase and photolyase-mimicking biomolecules ................................. 31
Figure 2-1. Comparison between photolyase and UV1C ............................................. 34
Figure 2-2. The G-quadruplex in UV1C-TDP and a G-quartet containing 6-MI ............. 36
Figure 2-3. The three functional classes of 6-MI mutants ............................................. 40
Figure 2-4. Photocatalytic effect of 6-MI alone compared to the G23 UV1C mutant ....... 44
Figure 2-5. Nucleotides with significant absorption in the visible ................................. 45
Figure 2-6. Sodium dependence of a selection of 6-MI mutants .................................. 50
Figure 2-7. Circular dichroism spectra of the UV1C 6-MI mutants ............................... 51
Figure 2-8. The UV-vis absorption spectra of G->6-MI point mutants ............................ 52
Figure 2-9. Laser irradiation apparatus and gel analysis .............................................. 53
Figure 2-10. The linear region of a typical time-course ................................................ 55
Figure 3-1. Comparison between TDP, LDP and LMP ................................................. 57
Figure 3-2. Short irradiations of LMP and LDP in three contexts ................................. 67
Figure 3-3. Longer time-course irradiations .................................................................. 68
Figure 3-4. Mixing Experiments .................................................................................. 70
Figure 4-1. Hypothetical backbone constructs capable of photoequilibrium ............... 79
Figure 4-2. Backbone constructs incapable of photoequilibrium ................................ 80
Figure 4-3. Analogy between self-replication and self-repair ......................................... 83
Figure 4-4. In-vitro evolution scheme for the improvement of LDP catalysis ............... 93
# List of Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>6-MI</td>
<td>6-Methyl Isoxanthopterin</td>
</tr>
<tr>
<td>2-AP</td>
<td>2-Aminopurine</td>
</tr>
<tr>
<td>8-OG</td>
<td>8-Oxoguanine</td>
</tr>
<tr>
<td>A</td>
<td>Adenosine</td>
</tr>
<tr>
<td>ALS</td>
<td>Amyotrophic lateral sclerosis</td>
</tr>
<tr>
<td>C</td>
<td>Cytosine</td>
</tr>
<tr>
<td>CD</td>
<td>Circular dichroism</td>
</tr>
<tr>
<td>CPD</td>
<td>Cyclobutane pyrimidine dimer</td>
</tr>
<tr>
<td>CPF</td>
<td>Cryptochrome photolyase family of proteins</td>
</tr>
<tr>
<td>DAP</td>
<td>2,6-Diaminopurine</td>
</tr>
<tr>
<td>DMS</td>
<td>Dimethylsulphate</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>Dss</td>
<td>7-(2,2′-bithien-5-yl)-imidazo[4,5-b]pyridine nucleoside</td>
</tr>
<tr>
<td>E</td>
<td>Perylene nucleoside</td>
</tr>
<tr>
<td>FAD</td>
<td>Flavin adenine dinucleotide</td>
</tr>
<tr>
<td>G</td>
<td>Guanine</td>
</tr>
<tr>
<td>HOMO</td>
<td>Highest occupied molecular orbital</td>
</tr>
<tr>
<td>HPLC</td>
<td>High pressure liquid chromatography</td>
</tr>
<tr>
<td>I</td>
<td>Inosine</td>
</tr>
<tr>
<td>K</td>
<td>Dicyanomethylene aminostyryl pyran nucleoside</td>
</tr>
<tr>
<td>LDP</td>
<td>Ligated dimer primer</td>
</tr>
<tr>
<td>LMP</td>
<td>Ligated monomeric primer</td>
</tr>
<tr>
<td>MTHF</td>
<td>5,10-Methylenetetrahydrofolate</td>
</tr>
<tr>
<td>PAGE</td>
<td>Polyacrilamide gel electrophoresis</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>PSS</td>
<td>Photostationary state</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>RNR</td>
<td>Ribonucleotide reductase</td>
</tr>
<tr>
<td>T</td>
<td>Thymine</td>
</tr>
<tr>
<td>T4PDG</td>
<td>T4 Endonuclease V</td>
</tr>
<tr>
<td>TDP</td>
<td>Thymine dimer primer</td>
</tr>
<tr>
<td>tRNA</td>
<td>Transfer RNA</td>
</tr>
<tr>
<td>U</td>
<td>Uracil</td>
</tr>
<tr>
<td>UCNP</td>
<td>Upconverting nanoparticle</td>
</tr>
</tbody>
</table>
Glossary

UV1C A 42 nucleotide long G-rich sequence shown to possess photorepair activity on thymine dimers in a substrate strand.

ABzyme A catalytic antibody.

Aptamer An oligonucleotide which binds a particular antigen.

Dimer A covalent compound between two chemically similar monomers.

DNAzyme A catalytic DNA sequence.

G-Quadruplex A secondary structure of nucleic acids which consists of a quadruple-helix of stacked G-quartets.

G-Quartet Four guanines in a planar hydrogen bonded pattern.

Oligonucleotide A polynucleotide between 10 and 100 bases in length.

Palimpsest A re-used parchment, on which the writing has been erased, but of which traces of the original words remain.

Parsimony A principle whereby explanations for evolutionary phenomena with the least number of independent events is most likely.

Ribozyme A catalytic RNA sequence.

UV-A Ultraviolet radiation with wavelengths ranging from 400 nm-315 nm.

UV-B Ultraviolet radiation with wavelengths ranging from 315 nm-280 nm.

UV-C Ultraviolet radiation with wavelengths ranging from 280 nm-200 nm.
Chapter 1.

Introduction

Nucleic acids have been studied extensively by some of the greatest minds in science, with good reason. The list of properties unique to nucleic acids, among all other materials both natural and artificial, is long. Yet still, we may find new properties of nucleic acids. New observations and new unimaginable uses have characterized the history of the study of nucleic acids. In this thesis, a new property, a new resilience, will be explored. Experiments old and new will together show evidence for an extraordinary, almost miraculous property of nucleic acids, not yet well known, but of which more and more evidence emerges to support. For while light can damage nucleic acids, perhaps with the help of light, nucleic acids can repair themselves.

In this thesis is discussed two published projects, supplemented with details and experiments not ready at the time of publication. First is the transformation of the DNAzyme UV1C from a photocatalyst that can only utilize UV-B light to a photocatalyst capable of using violet light. This transformation was carried out in the most minimal way, replacing only one residue at a time with the guanine analog 6-MI, and resulted in a photocatalyst with both the original activity in the UV-B and new activity throughout the UV-A. Unpublished results involving oligonucleotides where bases were replaced with organic molecules which did not have base-like hydrogen bond donors and acceptors gave new insights into the limits of the possible modifications to UV1C’s active site. While these oligonucleotides were effective photocatalysts, all of their effect could be attributed to the substituent, with no enhanced activity conferred by the DNAzyme UV1C. Subsequent work focused on the substrate specificity of UV1C, and demonstrates that while UV1C was selected against an unnatural DNA substrate, it in fact is also active with a bona-fide DNA substrate. In order to explain these results and put them in context, an introduction to the fundamental concepts of photochemistry, electron transfer, enzyme catalysis, nucleic acid structure, abiogenesis, and specifics of the G-quadruplex are presented.
1.1. Nucleic Acid Structure

Nucleic acids are among the most studied materials in history. The literature on DNA and RNA and their properties is incredibly vast and varied. Besides playing crucial roles in biology, of which new roles continue to be discovered, nucleic acids show tremendous promise as materials in their own right, suitable for production of nanostructures\(^1\), sensors\(^2\), and new classes of computational devices\(^3\). They play host to a suite of properties wholly unique in the realms of chemistry, most importantly for their biological function, their ability to self-template their own replication\(^4\). This incredible ability is unknown to all other polymers. Crucial in this property is the phosphodiester backbone, present in both DNA and RNA. Each phosphate gives a full negative charge for each base in an oligonucleotide, stabilized by being delocalized over two oxygen atoms. The repulsion between these negative charges in a single strand is partly responsible for the ability of a single-stranded oligonucleotide to exist in an extended, partially structured form in which its nitrogenous bases are oriented to make contacts with its complementary partner strand\(^5\). The repulsion between phosphates in the two strands of the duplex also contributes to orienting the bases towards each other. Always surrounding the negatively charged backbone are charge neutralizing cations, either the loosely bound monovalent cations like sodium or potassium, or the more tightly bound divalent cations, the most biologically relevant of which is magnesium.

The negative charge of nucleic acids also results in their difficulty in crossing plasma membranes. While this property is fundamental to how cells maintain their identities, it also makes nucleic acids poor candidates for drugs. The polyanionic nature of nucleic acids also leads to their exceptional water solubility among biopolymers. Attempts to replicate the properties of nucleic acids with artificial substitutes have sometimes omitted this property of the backbone, at the detriment of predictable water solubility. In many instances, changes to the phosphate backbone result in gross changes to a nucleic acid’s structure and physical properties. Consider the case of the Holliday junction, a structure in which four DNA helices converge. Magnesium leads to a conformational change from an open form in which all four arms are equally extended, and a closed form where two of the strands pass quite closely and make a tight turn to exchange helices. When the phosphodiester linkage at the junction are replaced uniformly with methylphosphonate linkages, the junction no longer closes properly in the
presence of magnesium\textsuperscript{6}. Replacing these phosphates two at a time, instead of on all four strands, returns the ability of the junction to close properly in magnesium, but only in the isoform in which the unmodified negatively charged phosphates are in the interior loops. This implies that in such tight turns and unusual secondary structures, such as might be found in DNAzymes, the backbone phosphates can create specific binding sites for particular cations, and that these binding sites may be crucial to maintaining particular structural elements.

Also fundamental to the properties of nucleic acids are the backbone furanose sugars. The entirety of the chirality of nucleic acids arises from the chirality of these backbone sugars. The nucleic acids currently in use in the cells of all forms of life are divided into two exclusive types, depending on their sugar; RNA refers to nucleic acids with a ribose sugar in their backbone, while DNA is missing a hydroxyl on the 2’ position of the ribose sugar. While only one atom difference, this leads to dramatically different properties of the two natural nucleic acids, including but not limited to the sugar’s preferred ring conformation. The extra hydroxyl in RNA is also implicated in many of the reactive properties of RNA. RNA also uses the base Uracil, the methylated version of which, Thymine, appears in DNA exclusively. In order to produce nucleic acids that are of potential medicinal use, a variety of new nucleic acids have been developed by chemical synthesis. Generically referred to as XNA’s\textsuperscript{7}, these nucleic acids have sugars in their backbone which are neither ribose or deoxyribose. Some of these XNA’s can be enzymatically replicated under conditions which require very permissive polymerases, but most enzymes which have evolved to recognize DNA and RNA have no effect on these modified backbones.
Figure 1-1. DNA structure and canonical bonding.
Comparison of DNA and RNA backbones, canonical G-C and A-T base pairs, indicating the Hoogsteen hydrogen bonding face of the purine bases.

All bases, purine or pyrimidine, have between 2 and 3 hydrogen bond donating or accepting functionalities on their base-pairing face. Besides the base pairing face, the purine bases have another region capable of forming hydrogen bonds. The Hoogsteen face consist of the N7, which is a hydrogen bond acceptor, and in the case of adenine, the exocyclic amine, a hydrogen bond donor, and in the case of guanine the ketone oxygen, another hydrogen bond acceptor. The hydrogen bonds supplied by the Hoogsteen face are crucial to forming higher-order structures in which more than two bases interact.
The nitrogenous bases are well known for their ability to use hydrogen bonds to base pair to each other, but pi-pi stacking interactions between adjacent bases account for much of the energy stabilizing DNA structure\(^8\). The degree of stacking is highly variable depending on the base sequence, with the larger purines contributing more to the pi-pi stacking stabilization. The presence or lack of stacking has several signatures. Relative to the UV-Vis absorbance of unstacked, unstructured bases, bases in a double helix show a pronounced hypochromicity. The CD spectroscopy signal observed from nucleic acids arises from the chirality of the overlapped pi orbitals of the bases, and is sensitive to the glycosidic conformation of the bases as well\(^9\).

While the double helix is the most common nucleic acid secondary structure, single stranded nucleic acids are capable of folding into a variety of other structures. Single-stranded RNA has long been known to fold back on itself into complex three dimensional shapes. The first nucleic acid to be crystallized was a tRNA, a molecule with a defined tertiary structure and many unusual base interactions extending deeply beyond the canonical Watson-Crick base pairs\(^10\). Besides the four standard RNA bases produced in a pre-tRNA transcript, mature tRNA’s contain a plethora of non-standard bases. The ancient evolutionary origin of the tRNA indicate that while these non-standard bases are now all introduced by protein enzymes, that in the original riboorganisms hypothesised to have given rise to the ribosomal peptide synthesis system, ribozymes could have caused some of the same modifications\(^11\). The structural diversity of single-stranded nucleic acids is not limited to RNA. Forty-three years after the first tRNA crystal structure, the first crystal structure of a catalytic DNA was reported. The structure proved unequivocally that DNA can also form highly varied structures, with intricate hydrogen bonding patterns not expected in double-stranded forms\(^12\).

### 1.2. G-Quadruplex Structure

Of all the non-canonical secondary structures available to single-stranded DNA, none has more biological relevance than the G-quadruplex. Many properties of the G-quadruplex are of significance to the activity of the DNAzyme UV1C. Guanine has the correct hydrogen bond pattern on its base-pairing face to make contacts with the Hoogsteen face of an adjacent guanine. Four guanines form a guanine quartet, with a cation filling the central gap, directly coordinated to the guanine ketone oxygens. Multiple
adjacent G-quartets stack well on each others’ all-purine pi-surfaces, further stabilizing the G-quadruplex structure. Cations, particularly the biologically relevant sodium and potassium monovalent cations, are crucial to forming the G-quadruplex fold\textsuperscript{13}. While these ions are usually considered inert, geometrical requirements and the requirement of the G-quadruplex for a fully dehydrated cation limits the choices which may coordinate with a G-quadruplex. Starting from the top of the column of the periodic table, the alkali metals, Lithium cations are hard, complexing strongly to 6 water molecules in an octahedral arrangement. Sodium and potassium ions have a slightly larger ionic radius, and waters on these ions are more labile. These are the cations most commonly associated with the stabilization of guanine quadruplexes, but there are many others.

![Figure 1-2. Guanine-quadruplex structures](image)

Figure 1-2. Guanine-quadruplex structures
(left) A small sample of the diversity of quadruplex structures and topologies, with examples of tetramolecular, bimolecular and unimolecular quadruplexes. Many more topologies are possible\textsuperscript{14}. Each stack of the quadruplex consists of a G-quartet (right).

Unlike the standard double helical DNA structure, which is always antiparallel, the four strands of an intramolecular DNA G-quadruplex have been reported in all possible combinations of orientations\textsuperscript{14}. In the case of the fully intermolecular G-quadruplex, only all-parallel quadruplexes have been observed. In some sequences, varying conditions, such as salt concentration, cation identity, organic solvents, and molecular crowding agents can promote different quadruplex topologies, sometimes leading to controversies. For example, the human telomeric repeat sequence d[AGGG(TTAGG)]\textsubscript{3} forms an antiparallel basket topology in sodium as determined by crystallography. The crystal structure of the same sequence in the more physiologically relevant potassium reveals a parallel propeller topology, while the NMR structure
determined in potassium gives a mixed conformation, with two propeller loops and one loop reversing direction. The discovery and characterization of some of the first quadruplex forming oligonucleotides in fact relied heavily on thymine dimerization. Because four repeats of the Tetrahymena telomeric sequence (TTGGG)₄ folds into an antiparallel basket topology in the presence of sodium, two of its loops, composed exclusively of thymine, are efficiently cross-linked by forming a thymine dimer when irradiated with UV light. G-quadruplexes and their polymorphism are implicated in a variety of diseases including cancer and ALS, and are an intense field of study. Relative to the same sequence in a double-helical form, quadruplex DNA shows a slight red shift and a hyperchromicity to its UV absorption band. The chirality of DNA manifests in a circular dichroism signal. Different quadruplex topologies show dramatically different peaks in their CD signals.

G-quadruplexes are synonymous with the base guanine, but other unusual and unnatural bases also form quadruplex-like higher order structures. Many telomeric sequences form quadruplexes with mixed quartets containing some non-guanine bases stacked upon the other fully G-quartets in a quadruplex. In order to test the effect of a particular guanine analog on quadruplex formation, Mergny systematically explored quadruplexes containing 4 guanines and a single guanine analog at all 5 positions in the sequence. Of many base analogs to guanine tested by Mergny et al, all show some negative effect on the stability of an intermolecular G-quadruplex. The least destabilizing are the oxidized form of guanine, 8-OG, and the G-analog 6-MI.

Ironically, G-quadruplex nucleic acids occur more frequently in the results of in-vitro selection experiments as perhaps they are known to occur in nature. The most well-known and characterized aptamer is the Thrombin binding aptamer, which consists of an antiparallel basket topology G-quadruplex, in which the binding is mediated through the top quartet face itself. Later, another selection for an independent aptamer for thrombin again resulted in an aptamer which contained a quadruplex, although this subsequent aptamer also had duplex regions, and did not use its quadruplex directly for binding. The quadruplex forming sequence PS2M, which has porphyrin metalation and peroxidase activity when bound to heme, is one of the only DNAzymes to have been selected based on binding to a metalation transition state analog, n-methyl mesoporphyrin. The propensity of quadruplexes to arise in selections is perhaps attributable to their high thermodynamic stability, large pi-stacking surface, and their...
ability to exist in such a diverse complement of topological conformations. The predisposition for g-rich sequences to form quadruplexes is an emergent property intrinsic to guanine, and must be considered if one is to imagine an abiogenesis based on nucleic acids\textsuperscript{27}.

\section*{1.3. Synthetic Nucleic Acids}

Research involving nucleic acids is simplified by the ease of solid phase DNA synthesis. This mature technology uses protected nucleosides with a reactive phosphoramidite functionality on the 3' hydroxyl to couple new nucleosides to the growing chain\textsuperscript{28}. Conditions used in DNA synthesis are optimized for the natural bases, but the efficiency of the phosphoramidite coupling allows unnatural chemical moieties to be incorporated in an oligonucleotide. Modifications are possible throughout the nucleic acid structure, in the phosphodiester backbone, sugar, or the base itself. Previous research into UV1C relied on a phosphorothioate modification to the backbone\textsuperscript{29}. Sugar modifications, such as FANA, can change preferred structures of the same sequence\textsuperscript{30}. Modified bases are often the most conservative changes that can be incorporated into chemically synthesized DNA.
Figure 1-3. Non-standard purine bases

Purine base modifications compared to guanine. Inosine, 6-methyl isoxanthopterin and 8-oxoguanine are known to participate in substituted G-quartets and quadruplex structures.

Even a cursory examination of the natural base pairs A-T and G-C yields an obvious question: If guanine forms 3 hydrogen bonds to cytosine, why doesn’t adenosine form 3 hydrogen bonds to thymine? Thymine has a hydrogen bond acceptor where adenosine has no donor. In fact the 3-hydrogen bond equivalent base, diaminopurine, (DAP), does in fact exist in nature in the DNA of the cyanophage 2-SL\textsuperscript{31}. If the exocyclic amine is restricted to the 2 position of the purine base, the most commonly used fluorescent base analog 2-aminopurine (2-AP) is obtained\textsuperscript{32}. While the change from adenine to 2-AP seems quite modest, the photo-physical properties of 2AP are dramatically different from adenine. Its absorbance spectrum shows a peak at 310 nm. While the natural bases have fluorescent quantum yields on the order of 1 in a hundred thousand, 2-AP has a quantum yield of 0.68. Despite looking very similar to a standard AT base pair, the pairing between 2-AP and thymine is much weaker, leading to a significant decrease in melting temperature. 2-AP’s fluorescence is quenched significantly in a duplex relative to its high fluorescence in single-stranded or unstructured regions. 2-AP is capable of forming a base pair with thymine, but also with cytosine, and an oligo containing 2-AP can be a substrate for DNA polymerase, acting to mutagenize the position complementary to the modification with either T or C.
Figure 1-4. MTHF and 6-MI
Comparison between structures of MTHF, an antenna cofactor found in type I photolyase enzymes and the fluorescent guanine analog 6-MI

6-Methylisoxanthopterin (6-MI) is another fluorescent base analog, engineered to stand in place for guanine\textsuperscript{33}. With the same arrangement of hydrogen bond donors and acceptors on its base pairing face and Hoogsteen face, 6-MI is a better mimic of guanine than 2-AP is a mimic of adenine. 6-MI is a pterin analog\textsuperscript{34}. Pterin is the backbone of the important folate family of cofactors used in a variety of atom-transfer reactions, as well as acting as the antenna cofactor in many photolyase enzymes. These pterin-based cofactors are in fact biosynthesized from guanine, and all retain the guanine base-pairing face hydrogen bond pattern. When compared to a duplex containing a G-C base pair, the 6-MI base pair results in only a negligible change to the melting temperature of the DNA duplex. In further contrast to 2-AP, 6-MI fluorescence is, in certain sequences, enhanced upon forming a duplex\textsuperscript{35}, while other sequences are quenched, with up to two bases away having an influence on the relative fluorescence yield\textsuperscript{36}. Unlike 2-AP, 6-MI allows a cytosine base to be inserted across from it by DNA polymerase, but following that single insertion, no further polymerization occurs.
In popular parlance, base pairing is synonymous with hydrogen bonding, by taking advantage of the equally important pi-pi stacking interactions, unnatural fluorescent bases without any hydrogen bond donors or acceptors may fill the hydrophobic core of a DNA duplex with an aromatic moiety. Developed by Erik Kool, these unusual nucleosides share the sugar-phosphate backbone with the natural bases, but have little else in common. Compared to the natural bases, which absorb around 260 nm, and most fluorescent base analogs, which absorb in the region from 300 nm-370 nm, some of the extended aromatic systems developed by the Kool lab have significant absorption above 400 nm. Besides aromatic systems, other fluorophores, such as the dicyanomethylene aminostyryl pyran nucleoside ("K"), derive an even longer wavelength absorption and high extinction coefficient from charge transfer within the fluorophore. The spectroscopic properties of these bases are dramatically altered when multiple fluorophores are placed in series in a single oligonucleotide. These bases were designed for fluorescent imaging in live cells, and much of their development has focused on improving photostability and reducing phototoxicity.

Fluorescent base analogs, when incorporated into oligonucleotides, often show dramatic changes in their emission spectrum on forming duplexes or other secondary structures. While single-stranded oligonucleotides have some structure, their bases are much less stacked and much more solvated by water than the same sequence participating in a secondary structure.

The above base modifications are meant to act as standalone units or in oligonucleotides. An ambitious goal of many biochemists is development of completely artificial sets of base pairs, that is, non-natural nucleosides which may be amplified enzymatically, forming specific interactions with each other, but not with the other bases or base analogs. Thanks to the diligent efforts of some very talented research groups, a scattering of these systems now exist. The first expanded genetic alphabet came out of the group of Steve Benner. These bases bear strong resemblance to the standard bases, featuring rings of the same sizes, and use the same system of hydrogen bond donors and acceptors, but in a different order. These artificial bases are amplified less efficiently by PCR than the standard bases, however they offer aptamers containing them such a selective advantage that they nonetheless survive in-vitro evolution. The nitro group on the Z base may offer hydrogen bonding possibilities not available to the standard bases. Another unusual expanded genetic alphabet, designed by the
Romesberg lab, eschews hydrogen bonding entirely, using hydrophobic interaction alone to align its artificial bases\textsuperscript{41}. Even stranger still are the artificial base pair developed by the Hirao lab. While built on a purine backbone, the fluorescent base Dss extends up into the major groove space with two linked thiophene rings. Its base pairing partner, Pa, looks nothing like either a purine or pyrimidine, and acts to quench the fluorescence of its pairing partner when annealed together into a duplex\textsuperscript{42}. Like the artificial base pairs developed by Benner, the Dss-Pa pair has been incorporated into in-vitro selection\textsuperscript{43}. Again, the additional diversity allowed by these unusual bases greatly enhanced the binding properties of aptamers which contained them. Artificial genetic alphabets have found a variety of practical applications, including a life-saving test for HIV viral load\textsuperscript{5}, but for the researchers developing them, they represent more philosophical explorations of alternative forms of life. The bases currently in use in DNA and RNA are highly optimized for a variety of properties, and as of yet, using alternative genetic alphabets remains more challenging compared to the routine use of natural nucleic acids and enzymes. Development of this exciting field is stimulated by the speculation that in an RNA-dependent model of abiogenesis, a larger set of nucleobases may have been present than have currently survived the progress of molecular evolution to the present day\textsuperscript{44}.
In addition to modifications introduced during chemical synthesis, a wide array of commercially available enzymes enable a researcher to label, cut, join, and modify DNA and RNA in a bewildering variety of ways. Working on the radiological scale is made possible by the enzyme polynucleotide kinase, which transfers the gamma-phosphate from 32P ATP onto the 5’ hydroxyl of the oligonucleotide of choice. The enzyme T4 endonuclease V (T4 PDG) is the first DNA repair protein to be crystallized bound to its substrate. T4 PDG is a pyrimidine dimer deglycosylase, highly specific for the cis-syn cyclobutane thymine dimer. Unlike most restriction endonucleases, T4 PDG does not catalyze phosphodiester hydrolysis. In order to cleave the dimer, T4 PDG uses its N-terminal amino group as a nucleophile, attacking the c1’ of the 5’ thymine involved in the dimer. This forms a covalent enzyme/substrate intermediate, in the form of an imine, or

**Figure 1-5. Artificial expanded genetic alphabets**
Completely unnatural base pairs compared with the canonical G-C base pair.
schiﬀ base, linkage to the c1’ carbon of the sugar. The covalent enzyme-substrate intermediate is resolved when the phosphate bridging the two thymines making up dimer is eliminated from the sugar. The elimination reaction occurs in the case of a phosphate leaving group, but in the case for which there is no phosphate in-between the dimerized thymines, the elimination does not occur. Hence TDP and other gapped dimer species shows no cleavage by T4PDG.

1.4. Catalytic nucleic acids

In his famous treatise on the nature of the chemical bond, Linus Pauling speculated that enzymes accelerated the rates of chemical reactions by binding to, and hence lowering the free energy of the transition state, the high energy species that reactants must go through before becoming products. Decades later, crystal structures of thousands of enzymes conﬁrmed that enzymes often adhere to this model of catalysis, holding substrates in their hydrophobic interiors, within an exquisitely placed network of hydrogen bond donors and acceptors which are tuned not to the substrate, but to the transition state. This form of catalysis is at play in many enzymes, but there are other ways in which enzymes accelerate rates. Proteins contain a hydrophobic core, a volume wherein, by excluding water, a phase is created with a much lower dielectric constant than the polar aqueous phase. Many reactions increase in rate in a less polarisable phase, on the benefit of having a lower reorganization energy around a charged transition state. On the other hand, in order to increase the rate of a particular reaction, many enzymes create an alternate reaction pathway, and entirely different mechanism than exists in the uncatalyzed case, which allows a reaction to side-step a high energy barrier in the uncatalyzed pathway. Of course, these descriptions are not mutually exclusive.

RNA, once thought to be simply a genetic messenger biopolymer, was shown by Tom Cech and Sidney Altman to be capable of enzymatic behavior as well. The ﬁrst catalytic RNA’s discovered in nature catalyze phosphodiester bond-making or bond-breaking by stabilizing the 5-coordinate phosphate intermediate in transesteriﬁcation. Just like protein enzymes, nucleic acid enzymes also show Michaelis Menten kinetics. Since that initial discovery, RNA catalysis has been implicated in peptide bond formation in the ribosome itself, among many other crucial catalytic roles. The fact that nucleic
Acids can catalyze reactions was once attributed solely to their ability to bind metal ions, in particular magnesium, which is itself obligatory for almost any protein which cuts or ligates nucleic acid backbones. Many catalytic nucleic acids do use metals for catalysis, but further evidence is emerging for catalytic nucleic acids which use general acid base catalysis, with the normally inert nucleobases acting as acids and bases. While the nucleobases on their own or in the case of the standard double helix have $pK_a$s far from neutrality, just as in the cases of proteins’ general acid base catalysis, structures exist which stabilize the charged, protonated versions of certain crucial catalytic bases, shifting their $pK_a$s into the catalytic range. With such a limited range of functional groups, nucleic acids use ingenious means to shift the $pK_a$ of a functional group. The famous HDV ribozyme maintains a fold in which a closely grouped cluster of negatively charged phosphates generates an electrostatic potential. This potential causes a shift in electron density in the catalytic cytosine, making the nitrogen on its base-pairing face as basic as a histidine. One could then imagine how such electrostatic potentials could be used to modulate redox potentials of bases as well as their $pK_a$s.

1.5. In-vitro selection

RNA is now known to be the central molecule of molecular biology, capable of carrying information, in certain situations acting as a catalyst, and fulfilling a huge repertoire of regulatory roles in the cell. The catalytic properties of RNA led researchers to investigate the catalytic potential of DNA. While DNA lacks a 2’ hydroxyl group, the similarities between DNA and RNA imply that if DNA has the functionality to catalyze a reaction, then RNA should, in principle, be capable of the same catalysis. In vitro selection is an iterative method by which new functional nucleic acid sequences may be identified from large combinatorial libraries. Such a library may be generated by chemical synthesis. The key feature of in-vitro selection is the selection step itself, a means by which sequences which have a particular function can be separated from the vast majority of sequences which do not. Several techniques are commonly used to achieve this separation. The first in vitro selections of nucleic acids yielded sequences which could bind unnatural dyes with high specificity. The separation step in this case was carried out by affinity column chromatography. Those sequences which could bind targets immobilized on the affinity column could thus be separated. Affinity based selections are sufficient to enrich for binding, but only very rarely resulted in catalytic
sequences. New selection techniques have been developed to allow the identification of catalytic DNA and RNA sequences capable of accelerating a wide variety of biochemical transformations, including but not limited to RNA cleavage and ligation, depurination and oxidation\textsuperscript{56}. Most of these selections used either a gel-shift or a biotin tag to separate those species which were catalytic from those which were not. This type of selection has the advantage of avoiding some of the possible off-target sequences which seem to sneak through in the case of affinity chromatography.

![Diagram of in-vitro selection](image)

**Figure 1-6. Generalized scheme for in-vitro selection.**

The DNAzyme UV1C emerged unexpectedly in a selection based on such a change in electrophoretic mobility\textsuperscript{57}. It relied on a synthetic oligonucleotide referred to as TDP, an acronym for “thymine dimer primer.” TDP is an unusual 35nt long oligonucleotide which consists of a 20mer T1 and a 15mer T2. T1 ends and T2 starts with a thymine. These thymines are joined covalently by a cis-syn cyclobutane thymine
dimer. When the thymine dimer is repaired, TDP reverts to the shorter T1 and T2. TDP was used as a primer for PCR on a random library, thus generating a library of random DNA sequences attached to a gapped thymine dimer. The complementary strand in the PCR cannot extend past the dimer in TDP, leading to a size difference between the strands containing dimers and their complement. In one step of the selection, this random library was folded in the presence of a potential cofactor, serotonin, and irradiated with a UV lamp. Those sequences which could hold the cofactor in the correct position to photosensitize the repair of the thymine dimer were hence separable on the basis of electrophoretic mobility, and this population led to the isolation of the sequence SERO1C. Negative rounds of selection were also carried out, by irradiating the folded pool in the absence of serotonin in an attempt to ensure that sequences which survived were able. A significant population survived these rounds which were meant to eliminate any sequence that couldn't bind serotonin. Of course, if a DNAzyme needs no cofactor to survive a selection round, it can survive the positive rounds as well.

1.6. The RNA world Hypothesis

The catalytic potential of RNA molecules is but one of many lines of evidence indicating that before the last universal common ancestor to all modern life evolved to use DNA for information storage, RNA for some catalysis and most information transfer, and protein for most catalysis, that life passed through a phase of evolution in which RNA was the only biomolecule. In the so-called RNA world, RNA must have acted as both a genetic material and all catalysts. In the prevailing accepted theory of abiogenesis, this RNA world is itself preceded by a pre-biotic world in which chemistry alone produces a variety of complex organic polymers, with RNA emerging as a biopolymer as the result of what has been termed “molecular evolution”. It is hypothesised that these first biopolymers, including short RNA strands, formed spontaneously but in small quantities. While the conditions on the early earth could have favored a variety of possible polymerization reactions, what makes RNA special is its ability to template its own replication non-enzymatically, using specific, activated monomers, albeit at very low rates. This non-enzymatic activity is thought to be an intrinsic property of RNA. Other potential polymers may have formed as well, but only polymers which could template their own replication were able to propagate themselves. As longer and longer RNA’s are formed, eventually, by sheer chance, some of these
RNA’s had sequences which lead to catalytic secondary structures, such as RNA polymerase capability. These first ribozymes would have been a breakthrough, allowing much faster growth of RNA populations which had them as compared to RNA which had to rely only on self-templating to replicate. The ability of certain RNA sequences to form catalytic species is considered essential to their rise above the many other potential molecules which could be polymerized spontaneously on the early earth. In this proto-RNA world, an intrinsic property of RNA, its ability to self-template, allowed it to emerge as the dominant species from the primordial soup.

Another indication of the nature of the RNA world is that some of the conserved cofactors in metabolism, used by almost all life forms, are RNA-like or have RNA bases still attached. The proteins that use these RNA-like cofactors may represent replacements for catalytic molecules which were once wholly RNA. While the RNA world hypothesis is largely accepted within the research community, there are still significant disagreements about the nature of the RNA world and the transition points to modern biochemistry. In one hypothesis, the riboorganisms in the RNA world developed the ability to catalyze the formation of peptide bonds, and eventually longer proteins, leading to a period referred to at the ribonucleoprotein (RNP) world\(^6\). During this hypothesized epoch, proteins evolved to bind and support their RNA enzymes, allowing the sequences of these ribozymes to evolve to decrease in length, until only their catalytic residues were left, in the form of a cofactor. In this model, DNA is not present as a genetic material until long after the development of ribosomal peptide synthesis. The catalytic complexity of the highly conserved enzyme ribonucleotide reductase (RNR), which now produces DNA monomers from RNA monomers, has, for quite some time led most to rule out DNA synthesis in the RNA world\(^6\). The reliance of RNR on cysteine radicals in its catalytic mechanism has been a major sticking point, since free radicals tend to cause nonspecific damage to the RNA backbone.
While the late emergence of DNA is accepted by the majority in the field, dissenters such as Benner and Lehman have proposed an alternative timeline in which RNA and DNA coexist or coevolve with ribosomal protein synthesis. While this view is certainly held by the minority of researchers, it has several advantages to the late advent of DNA. In the version of this model proposed by Lehman, ribozymes use a completely different pathway to synthesize deoxynucleotides directly, without having to reduce ribonucleotides. Instead of RNR being the first deoxynucleotide producing enzyme, the protein RNR is a more efficient replacement for an existing pathway mediated by ribozymes which assembles deoxynucleotides by an aldol condensation, a reaction for which RNA catalysis presents no obstacle. In contrast, recent work by the Silverman lab, the leader in selection of unusual new catalytic nucleic acids, has selected a DNAzyme which catalyzes a radical reaction without redox active metals, and which leads to a defined product without any of the non-specific breakdown as was expected to occur in an oligonucleotide in the presence of radicals. Based on this result, it is now conceivable that a radical RNR could have existed in the RNA world.

Figure 1-7. Roadmap from the RNA world to the present day
Comparison between (a) peptide-first and (b) DNA-first models of a transition from the RNA world to modern life.
The catalytic repertoire of ribozymes discovered thus far mainly consists of phosphodiester bond making and breaking reactions\textsuperscript{66}. DNA dependant RNA polymerase and RNA dependent DNA polymerase ribozymes have yet to be reported, but given the proper starting materials, the hurdle of polymerizing DNA or using it as a template is much lower than synthesis of the monomers itself. The advantages a DNA genome would impart in the RNA world are vast. The DNA phosphodiester backbone is inherently more stable to basic conditions. The larger genome sizes possible with a DNA genome compared to an RNA genome would have allowed greater complexity in the evolution of the RNA world, making the emergence of huge complexes like the ribosome much more plausible. It should also be noted that if DNA were used as a genetic material in the RNA world before proteins, that it would only be required in much lower quantities than the catalytic RNA’s themselves, and that safeguarding this material would be crucial to the survival of these organisms.
1.7. Photochemistry and charge transfer in nucleic acids

Figure 1-8. Light absorption process
Absorption of light promotes an electron from the ground state to a vibrational state in an excited singlet state in a very fast process. The second singlet state is pictured, but any excited singlet state would lead to the same result, because internal conversion quickly relaxes the electron to the lowest energy singlet excited state. Fluorescent emission of a photon results from the return of the singlet electron to its ground state. Intersystem crossing and phosphorescence are much slower.

As UV1C is a photochemical DNAzyme, a basic understanding of photochemistry is absolutely crucial to understanding its mechanism. Of course, most of the biochemistry of nucleic acids occurs in the ground state, where all electrons are paired in their lowest energy states. Excited electronic states, populated by absorption of appropriately energetic photons, allow unusual reactivity not seen in the ground state populations. For example, in the ground state, 2+2 cycloaddition reactions, such as between two adjacent thymines, are quantum mechanically forbidden. Also quantum mechanically forbidden are transitions between the initially formed singlet state, in which the excited electron remains in the same spin as it was initially excited in, and the triplet state, where the spins of the excited state electron and the electron left in its lower level are aligned. While intersystem crossing to the triplet state is slow, the use of a triplet photosensitizer, such as acetophenone, allows one to create a much greater population...
of triplet state excited species. Once formed, due to the very slow nature of triplet to singlet (ground state) transition, molecules in the triplet state will remain in an excited state for much longer than excited singlet state molecules, giving the excited state a longer period during which it is in a reactive high-energy state.

One property of these electronically excited states is that the high-energy electron has enough energy to jump to an orbital on an adjacent molecule. Reactions characterized by the transfer of one electron from an excited state molecule to some acceptor are crucial to life. Besides photolyase, this type of electron transfer is of course crucial to photosynthesis. Electron transfer is perhaps unique among chemical reactions because of its extremely well characterized kinetics. Marcus theory relates the energy barrier to electron transfer, and hence rate, to the thermodynamic driving force and the reorganization energy. A prediction of Marcus theory is that while for most systems the rate of electron transfer increases with increasing driving force, there exists a rate maximum, following which increasing driving force lowers the rate of the reaction\(^68\). This so-called “inverted region” is an unusual consequence of the parabolic potential wells of the electron donor and electron acceptors’ energy of solvent repolarisation. One wrinkle to Marcus theory is that while it describes each single-step electron transfer, in many proteins which rely on electron transfer for catalysis, there exist multiple hopping pathways to electron transfer, with competition possible between one-step electron transfer and a multi-step trajectory\(^69\).

Isolated bases, such as those in nucleotides, have sub-picosecond relaxation from their excited states. Such short relaxation times come about due to conical intersections between the excited state vibrational manifold and the ground state\(^70\). Specific distortions to nucleobases, typically ones with out of plane motions, allow non-radiative relaxation to the ground state. In contrast, time-resolved spectroscopy of UV-excited oligonucleotides in the IR region carried out by Carell and Zinth clearly show that oligonucleotides and duplexes have much longer lived excited states, on the order of hundreds of picoseconds\(^71\). Longer lived excited states of oligonucleotides were once explained by restrictions on out-of-plane motions in stacked bases, but newer evidence indicates stacking is actually responsible for generating new decay pathways for monomer excitation, pathways which themselves show longer lived transient absorptions. Many explanations have been proposed to explain the existence of these excited states, including charge-transfer between individual adjacent bases, neutral
excimers, or larger delocalized charge separated domains. The nature of charge-transfer states between bases are highly sequence dependent, with charges separating to place the excess electron on the more easily reduced base and the hole on the more easily oxidized base\textsuperscript{72}. Spectroscopy of small molecule mimics of base-paired nucleotides show that in addition to charge transfer, photoexcitation stimulates proton transfer across the Watson-Crick face\textsuperscript{71}. In the case of oligonucleotides participating in a duplex, both electron-transfer and proton-coupled electron transfer are invoked to describe excited states visible for hundreds of picoseconds. The relatively long lifetimes of the DNA charge-transfer states are a consequence of the Marcus inverted region. While a high thermodynamic driving force favours recombination, the solvent polarization about the exciton pair significantly slows charge recombination. Suffice it to say, light has the capability of inducing persistent excited states in oligonucleotides, in single strands, duplexes, and other more unusual structures.

In addition to the photo-induced excitons present in pure nucleic acids, the nitrogenous bases are capable of being reversibly oxidized and reduced by small molecules, covalently attached molecules or other adjacent bases\textsuperscript{73}. Studies of individual bases have shown that guanine is the easiest to oxidize, and this property remains when incorporated into oligonucleotides. Another important base found in tRNA is inosine, structurally related to guanine but lacking an exocyclic amine. Inosine has many properties in common with guanine in terms of hydrogen bonding, but is much less easily oxidized. Thymine is the base with the highest reduction potential, and thymine is the primary carrier of excess electrons injected into DNA\textsuperscript{74}. Uracil, differing from thymine by only a single methyl substituent lacking at the 5 position, is not as easily reduced. In order to introduce an electron hole into an oligonucleotide, a photooxidant may be tethered covalently to one of the ends, or in the case of ruthenium complexes, allowed to intercalate into a double helix\textsuperscript{75}. A common way to observe an electron hole in DNA is by the propensity of oxidized guanines to react further, forming 8-OG, which may then be cleaved and imaged on a denaturing gel\textsuperscript{76}. The interactions between neighbouring bases in an oligonucleotide modulate these redox potentials, such as in the case of multiple guanines. Once an oxidized base is introduced in a duplex or oligonucleotide, the electron hole can hop between adjacent bases with a relatively low barrier. Since the bases have increasing oxidation potentials relative to the NHE, with G>A>C,T, guanine is the most likely base to play host to an electron hole. The electron hole can also
delocalize itself onto multiple bases, owing to the close pi-pi stacking in a duplex. A similar strategy using tethered photoreductants has been used to observe excess electrons in a duplex\textsuperscript{77}. These have been shown to hop from thymine base to thymine base, although at lower rates than the holes.

Guanine is the most easily oxidized base, and adjacent pairs or triples of guanine show even lower redox potentials, in either duplex or single-stranded forms. The guanines involved in a G-quadruplex exist in an environment that modulates their redox potential as well. It has been noted by several authors that relative to G-containing duplex DNA, G-quadruplexes are better conductors of holes and are more easily oxidized, while at the same time undergo less of the type of permanent oxidative damage that occurs in the case of holes carried in the duplex\textsuperscript{78}. When compared to the pattern of oxidative guanine damage in a duplex, the guanines in the interior of the quadruplex show little damage, with damage concentrating on the exposed end guanines. Time-resolved observations of hole injection into G quadruplex DNA have shown longer lifetimes in the charge separated state, both when photo oxidized by a tethered riboflavin or via pulse radiolysis\textsuperscript{79}. Further stabilizing the charge separated state, the G quadruplex carrying a radical cation is slower to deprotonate than a radical cation of guanine in a single strand or double helix\textsuperscript{80}. The physical properties of the G quadruplex therefore make it an ideal site for reversible oxidation in a nucleic acid structure.

The spectroscopic properties of G-quadruplex DNA, manifest in a redshifted absorbance spectrum and increased fluorescence quantum yield, is best explained by delocalized excited states\textsuperscript{19}. The high degree of stacking in the quadruplex and lack of possibility of proton-coupled electron transfer extend the lifetimes of pi-pi\textsuperscript{*} excitations. These excited states tend to have more delocalization when compared to duplex guanine excitation.

### 1.8. Pyrimidine dimer formation and structure

The photochemical cycloaddition reaction between two adjacent thymines is a photochemical reaction, and a paragon of the importance of quantum mechanics to our everyday life. This reaction involves the pi bonds between the 5 and 6 position of
adjacent thymines. If molecular symmetry were not conserved, the pi bonds of any two adjacent thymines, held in close proximity, could spontaneously rearrange into a cyclobutane ring. The consequences of this reaction occurring unfettered would be disastrous towards all life. In fact, it only occurs when a thymine in an adjacent pair is excited. The reason for this forbidden reactivity is that the symmetry of the bonding HOMO in the case of isolated thymines is of the same symmetry as the antibonding orbital in the dimerized thymines. Since orbital symmetry is conserved, the formation of a thymine dimer from the ground state is impeded by a high barrier. However, should electrons from the thymines be promoted to their higher antibonding orbitals, such as by photoexcitation, these antibonding orbitals are correlated with bonding orbitals of the same symmetry.
Figure 1-9. **Thymine dimer formation**
Symmetry-labelled combinations of bonding and antibonding orbitals of the double-bonds in isolated thymines (left) and single-bonds in dimerized thymines (right). Dashed arrows indicate correlations between symmetries. Only pictured are the orbitals which form the cis-syn CPD.

Under irradiation by UV-C light, the chromophore which gives rise to the pyrimidine dimer is the pyrimidine itself, based on its absorbance peaking at 260 nm. The formation of thymine dimers has been shown to be an ultra-fast reaction, occurring in under a picosecond in the case of a T18 oligonucleotide. DNA conformational changes occur on a longer timescale, indicating that the ensemble of unstructured T18 oligonucleotides must contain a pre-organized pair of thymines positioned in such a conformation as to be immediately reactive once irradiated by UV light. This ultra short timescale also shows that the dimer must be able to form directly from the singlet state,
as it is not long enough for intersystem crossing to the triplet state. Subsequently, bases participating in a cyclobutane dimer lose aromaticity, and hence absorb less at 260 nm and more at shorter wavelengths. Cytosine-thymine, thymine-cytosine and cytosine dimers are formed at lower rates than pure thymine dimers, but are often more mutagenic, due to deamination of the cytosine exocyclic amine in the dimerized form. The most numerically abundant form of thymine dimers is the cis-syn cyclobutane dimer, and this type of dimer is the primary subject of this work. While the double-helical arrangement of adjacent thymines encourages the formation the cis-syn dimer, if one thymine is rotated relative to the other, the two trans-syn cyclobutane dimer isomers may form as well. Other types of photoproducts include the 6-4 dimer, which results from a photochemical reaction between a ketone and a double bond and subsequent rearrangement. This 6-4 product can itself undergo further photochemical reactions to give the Dewar isomer. In only very specific DNA conformations, such as occur in tightly packed DNA in bacterial spores, unusual 3-5 thymine dimers are observed. As much as possible, these types of lesions are excluded from our study.

Figure 1-10. Thymine dimer species
Comparison between monomeric thymines and isomers of dimerized thymines. The cis-syn cyclobutane thymine dimer is the subject of this study.
Direct excitation with 260 nm or shorter wavelength light also leads to repair of some dimers, but formation of others. This wavelength of light is mostly filtered out of solar irradiation by the ozone layer. Evidence is accumulating that thymine dimers are still formed by UV-B and UV-A light exposure, despite the very low absorbance of thymine at those wavelengths. One possible pathway to form thymine dimers from lower energy UV-A light is photosensitization. In this process, small molecules absorb light in the UV-A and undergo intersystem crossing to the triplet state. Collisions between these small molecules and adjacent pyrimidines would allow population of the triplet excited state, a state thought to lead to formation of thymine dimers. While many such possible photosensitizing molecules are present in the cell, even purified DNA forms thymine dimers when irradiated in the UV-A, despite having a very low absorbance in that regime. Authors explain this absorbance on the basis of the delocalized excited states, excitons, which have a lower energy than the excitations of the isolated bases.

The use of a triplet photosensitizer may be used to selectively induce a thymine pair into a triplet, and hence, longer lived, excited state. Among photosensitizers, acetophenone has favourable triplet transfer only to thymine, impeding the formation of CT and TC photoproducts, and reducing CC dimers to almost undetectable levels. The excited thymines have several possible reaction pathways. However, in a typical DNA duplex, different sequence contexts lead to either more or less alignment of adjacent thymines, and considerable variations in the rate of thymine dimer formation are observed. This variation in rate is at least in part due to the high degree of polymorphism inherent in the DNA double helix. Electronic properties of the adjacent bases also appear to play a significant role.

1.9. Photolyase repairs pyrimidine dimers

Thymine dimers are efficiently repaired by the protein Photolyase. Photolyase is part of the ancient cryptochrome/photolyase family (CPF) of proteins, present in eukaryotes, bacteria and archaea. In fact, of all the DNA repair proteins, photolyase is the most universally distributed, and based on phylogenetic analysis considered to be the most ancient. The family includes all types of photolyase enzymes as well as the light sensing cryptochromes, which regulate circadian rhythm and even detect magnetic
fields. Recent work suggests that cryptochromes have some vestigial photolyase activity on single-stranded DNA as well\textsuperscript{89}. Common to all these functions is the cofactor flavin adenine dinucleotide (FAD), along with a menagerie of variable antenna cofactors\textsuperscript{90}. Flavin’s role in photolyase is to transfer a high energy electron from an excited state on FAD to the thymine dimer. Flavin is suited to this role due it its facile reduction and oxidation reactions. While flavins are easily oxidized while free in solution, when bound to photolyase, their reduced and semiquinone forms are stabilized dramatically. In the photolyase catalytic cycle, flavin starts in the anionic reduced form. The highest energy electron is promoted to an excited state where it may jump to the dimer, leaving behind flavin in the neutral semiquinone form. Once an electron is transferred to the dimer, the reduced form of the dimer rapidly reverts to undimerized thymines. Following repair, the electron returns to the flavin cofactor, returning the enzyme to the resting state. Like their formation, the repair of thymine dimers by photolyase is an ultra-fast reaction, and time-resolved spectroscopy has elucidated its mechanism in exquisite detail\textsuperscript{91}. While the simplest explanation for the activity of photolyase is direct electron transfer to its substrate, the kinetics and energies involved are also consistent with a hopping pathway in which the electron is first transferred through the adenosine of the cofactor before hopping to the dimer\textsuperscript{69}. Such a pathway is reminiscent of the photolyase-like self repair pathway observed by Carell\textsuperscript{72}.

Photolyase is still catalytically competent in the absence of its antenna cofactor, but the action spectrum and quantum yield is greatly enhanced by its antenna\textsuperscript{92}. The function of the antenna cofactor is to absorb light which is not absorbed by the reduced flavin. Excitation of the antenna cofactor leads to radiationless energy transfer to the flavin. Phylogenetic analysis of the CPF indicates the oldest enzymes used the pterin MTHF as an antenna cofactor\textsuperscript{88}. These photolyases have a maximum quantum yield around 380 nm, still in the UV, but closer to the edge of the visible. Later photolyases evolving to use deazaflavin as an antenna, giving them a maximum quantum yield in the blue, at longer wavelengths, maximizing at 440 nm\textsuperscript{93}. The biosynthetic complexity of deazaflavin relative to MTHF and flavin itself is reflective of the long history of photolyase evolution. The apoenzyme is still capable of some basal level of photolyase activity even when stripped of both its flavin and antenna. This activity is attributed to a tryptophan residue in the active site, adjacent to the bound thymine dimer\textsuperscript{94}. Tryptophan
is the most redox-active amino acid residue, and has the longest wavelength absorption among amino acids.

Figure 1-11. The main types of cofactors found in photolyase
Top row: Antenna cofactors MTHF and Deazaflavin. Bottom: The reduced form of the flavin adenine dinucleotide active in all photolyase and cryptochromes. Note the peptide linkages present in deazaflavin and absent in flavin.

Besides the photolyase enzymes observed in nature, a catalytic antibody (ABzyme) for the repair of thymine dimers has been developed, simply by creating a monoclonal antibody for a cis-syn thymine dimer. The hydrophobicity of the dimer itself led the Schultz group to speculate that a tryptophan moiety would be a natural binding partner with the thymine dimer, and as in the case of natural photolyase, that the tryptophan alone would be sufficient for photoreactivation. This strategy was a success, leading to a thymine-dimer photorepair ABzyme which showed saturation kinetics, with a $k_{cat}$ of 1.2 min$^{-1}$ under irradiation by 300 nm light. Although the flux for this observation was not specified, over the flux range from $5 \times 10^{-8}$ to $1.6 \times 10^{-7}$ Einstein/min the rate scaled linearly with increased light.

While the repair efficacy of photolyase is profound, recent studies into thymine dimer repair have proposed that double-stranded DNA on its own has some intrinsic photolyase-like activity. The first hints of this activity were measured by the Rokita group,
who observed different photostationary states for thymine dimer formation in different sequence contexts. While dimer formation rates of the sequences they studied were found to be quite similar, G-containing sequences, especially on the 5’ side to a thymine dimer, increased the repair rate. Repair rates ranged from $4 \pm 1 \text{ min}^{-1}$ to $15 \pm 2 \text{ min}^{-1}$ at 0.03 mW irradiation with a UV-lamp having maximum emission at 254 nm. They explained variations in the repair rate by proposing electron transfer from an excited guanine to the dimer. Subsequent work by Carell refines this mechanism in terms of an exciton pair between an oxidized guanine and an adjacent reduced adenine base being the catalytic moiety. A key finding is that a single base adjacent to a dimer, such as would occur near an end of an oligonucleotide, doesn't have the ability to carry out this self-repair. Only in cases where the redox potentials of adjacent bases are tuned to place an excess electron adjacent to a thymine dimer is self-repair observed. Rates were not reported for this work on repair with 290 nm light at 1 mW.

Figure 1-12. Photolyase and photolyase-mimicking biomolecules
(Top left) Photolyase is capable of harnessing visible light both through its antenna, direct flavin excitation, and has some activity even in the absence of cofactors due to the absorption of UV light by tryptophan. The excited flavin may transfer an electron directly (green) or by a hopping pathway through the adenine of the dinucleotide cofactor (pink). (Top middle) A catalytic antibody uses tryptophan alone for its photorepair activity. (Top right) The DNAzyme SERO1C uses the peptide-derived small molecule serotonin as a cofactor. (Bottom left) UV1C is a cofactor-free DNAzyme. (Bottom middle) An exciton formed between G and A also has photolyase-like activity. (Bottom left) The oxidized nucleobase 8-OG extends this purely nucleotide-based repair to longer wavelengths.
These results indicate that DNA self-repair is a fundamental property of the unmodified DNA bases. From a different perspective, Burrows considers the oxidation product 8-oxoguanine (8-OG). 8-OG has an additional absorbance peak centered at 295 nm, extending much further into the UV-B than guanine or the other bases. Relative to G, 8-OG has an even lower oxidation potential. Using light from a UV-lamp with maximum emission at 313 nm and a polystyrene cuvette to cut off wavelengths shorter than 300 nm, Burrows showed that 8-OG was also capable of thymine dimer repair, proposing a pre-biotic role for 8-OG as a flavin mimic, or precursor to flavin from the perspective of molecular evolution. In the double-stranded systems studied, 8-OG’s photocatalytic effect was dependent on sequence content and orientation of 8-OG relative to the thymine dimer. The maximum rate was observed when 8-OG was directly adjacent to the 5’ side of the dimer, 0.03 min⁻¹, but just moving the 8-OG on the 3’ side cuts this rate in half. 8-OG functions most effectively in the same strand as the thymine dimer, but still catalyzes repair when in the strand complementary to the dimer-containing strand, with a maximum repair rate of 0.01 min⁻¹ when placed across from an adenosine on the 5’ side of the dimer. Repeated melting and reannealing cycles followed by irradiation demonstrated that 8-OG could carry out multiple turnovers, but as demonstrated was not capable of multiple turnovers under isothermal conditions.
Chapter 2. Towards the visible

2.1. Introduction

As previously discussed, thymine dimers are the most significant lesions formed in cellular DNA from exposure to solar ultraviolet B (290–320 nm) light. Two major classes of thymine dimers are most commonly formed in mammalian cells the cyclobutane (CPD) and the less abundant but more mutagenic 6-4 dimers (reviewed in refs 99, 100, 101, 102, and 103). Unrepaired, these lesions pose a formidable challenge to cellular DNA replication; indeed, defects in cellular thymine dimer repair machinery have been linked both with human skin cancers and such diseases as Xeroderma pigmentosum. For this reason, organisms have evolved diverse strategies for repairing thymine dimers and pyrimidine dimers in general. The most direct approach, for either the CPD or 6-4 product lesions, is to use light in the ultraviolet A-to-visible range (320–450 nm) to directly repair (photo-reactivate) the dimers back to thymines. Though photolyases are found in all kingdoms of life, mammals have lost the genes coding for photolyases. Nevertheless, it has recently been shown that mice transgenic for Arabidopsis CPD photolyase show “superior resistance to sunlight-induced Tumorigenesis”\textsuperscript{104}. Another remarkable study reports that the topical application of \textit{A. nidulans} photolyase-containing liposomes to human skin, followed by exposure to damaging UV light, leads to a significant level of repair of CPD thymine dimers in the DNA of epidermal cells\textsuperscript{105}.

The functional architecture of a typical CPD photolyase (protein) enzyme is schematized in Figure 2-1a. Two classes of light-absorbing cofactors augment the function of these proteins: first, an antenna pigment (variously, a folate derivative, MTHF; or deazaflavin) that harvests 350–450 nm wavelength light, and transfers this excitation energy, non-radiatively, to the second cofactor, a reduced flavin (FADH\textsuperscript{−}). The photoexcited FADH\textsuperscript{−} then transfers an electron to a thymine dimer recipient that has been flipped out of a distorted DNA duplex\textsuperscript{106} (thymine dimers stack poorly with their neighboring bases and also base pair poorly with complementary adenines, leading to
their relatively facile mobility in and out of the double helix). The resulting CPD thymine dimer radical anion now reverts rapidly to thymine bases.

Figure 2-1. **Comparison between photolyase and UV1C**
Schematic illustrations of the key functional components of a proteinaceous photolyase enzyme versus a photolyase DNAzyme. (a) A proteinaceous photolyase enzyme (shown in gray). Blue light is absorbed by both the flavin and antenna (MTHF or deazaflavin) chromophores. The photoexcited antenna chromophore transmits its energy by Forster resonance energy transfer to the flavin cofactor. The excited flavin then transfers an electron to a thymine dimer extruded out of a DNA duplex into the photolyase’s active site. Functionally, the flavin cofactor alone is sufficient for photo-reactivation. However, the antenna pigment substantially enhances the light-harvesting capability and hence the efficiency of photo-reactivation by the photolyase. (b) The UV1C DNAzyme (shown in black), bound to its single-stranded DNA substrate, TDP (shown in red). It is hypothesized that guanines within a G-quadruplex formed within UV1C absorb light in the 300−310 nm wavelength range, followed by transfer of an electron by one or more photoexcited guanines to the thymine dimer located within the TDP substrate.

The discovery of catalytic RNAs, or ribozymes, in the 1980s, encouraged the formulation of the RNA World hypothesis, which posits that prior to the evolution of RNA–DNA–protein-based life forms, RNA-containing cells constituted a primitive life, in which RNA served both as information carrier and catalyst. We have been interested to know, for the purpose of defining the range and limits of RNA catalysis, whether catalytic RNAs (ribozymes) or their DNA surrogates (deoxyribozymes or DNAzymes) are capable of catalyzing photochemical reactions, specifically, biochemical reactions that require light. In 2004 we initiated an in vitro selection experiment (SELEX), to see if we could select from a random-sequence library of single-stranded DNAs (∼10^{14} sequences) catalysts capable of harnessing UV-A light (>300 nm wavelength, and at the
time thought to be less damaging for DNA, although some evidence to the contrary has since been reported\textsuperscript{108, 109} to photo-reactivate CPD thymine dimers within a specified DNA substrate. Our selection strategy was inspired by the mechanism of proteinaceous bacterial photolyases, as well as that of a catalytic antibody reported for the photochemical repair of thymine dimers\textsuperscript{95}. In the latter a tryptophan residue, positioned close to the bound thymine dimer substrate, was sufficient for the lesion’s photoreactivation\textsuperscript{97}.

For our in vitro selection we created a special DNA substrate that incorporated a cis,syn-cyclobutane thymine dimer but lacked the linking phosphodiester between the dimer thymines\textsuperscript{57}. Photo-reactivation of such a substrate necessarily resulted in two shorter pieces of DNA, which enabled us to purify the catalytic sequences away from the larger pool. Serotonin was included in our selection as a flavin substitute\textsuperscript{57}. Unexpectedly, two quite distinct thymine-dimer reactivating DNAzymes were cloned from this in vitro selection: “Sero1C”, a DNAzyme that required serotonin for activity\textsuperscript{58}, and “UV1C”, which catalyzed photo-reactivation without the aid of any extraneous cofactor.

Sero1C, requiring serotonin for activity, had an action spectrum that stretched to \(\sim 340\) nm\textsuperscript{58}; by contrast, UV1C’s action spectrum extended only to \(\sim 315\) nm, with the most optimal activity (defined as \(k_{\text{catalyzed}}/k_{\text{uncatalyzed}}\)) occurring at 305 nm\textsuperscript{57}. UV1C had a requirement for sodium ions for catalytic activity (its folded structure incorporated a G-quadruplex, specifically stabilized by Na\textsuperscript{+}). UV1C was capable of multiple-turnover catalysis, and had a catalytic efficiency comparable to those of catalytic antibodies reported for the same activity. The only energetically reasonable catalytic mechanism possible for UV1C (Figure 2-1b) was that its guanine-quadruplex (G-quadruplex\textsuperscript{110}) functioned both as light antenna (G-quadruplexes have a higher absorbance in the 300–310 nm wavelength range than DNA duplexes) as well as electron source for thymine dimer photo-reactivation. Quantum yield measurements made in the 250–320 nm range showed a maximal value of 0.05 at 305 nm. While this number is substantially lower than the quantum yields typical of CPD photolyases, it matches those of other naturally occurring photolyases (such as those that repair 6-4 thymine dimers).

We have recently made strides in understanding the structure–function relationships of UV1C, and these add credence to our mechanistic hypothesis for this DNAzyme\textsuperscript{58, 111}. First, methylation protection experiments showed that eight specific
guanines within UV1C (guanines 12, 13, 17, 18, 21, 22, 32, and 33; Figure 2-2a) were involved in forming the G-quadruplex. G23, though part of a contiguous GGG stretch, was not a participant in the G-quadruplex. Second, two distinct kinds of contact cross-linking experiments confirmed the close spatial proximity (required for efficient electron transfer) between the substrate’s thymine dimer and the DNAzyme’s G-quadruplex (the key contact cross-links formed between the thymine dimer and DNAzyme bases are summarized in Figure 2-2a).

Figure 2-2. The G-quadruplex in UV1C-TDP and a G-quartet containing 6-MI
(a) The UV1C DNAzyme, complexed to the TDP substrate, folds to an intramolecular, wholly parallel-stranded G-quadruplex. Blue and green arrows indicate the loci of contact cross-links between the thymine dimer (or a phosphorothioate residue placed between the two thymidines forming the dimer) and various nucleobases of UV1C. Sites of the different G→6-MI point mutations in UV1C are indicated by yellow squares. (b) Canonical guanine quartet and (c) a base-substituted quartet composed of three guanines and one 6-MI residue.

Based on the above, and mutational studies, we were able to refine our structural/topological model for UV1C folded and complexed with its DNA substrate, TDP (Figure 2-2a), which incorporated the thymine dimer to be repaired. The action spectrum of UV1C extends up to ~315 nm; however, naturally occurring proteinaceous photolyases typically utilize light well into the visible spectrum (<450 nm). A DNAzyme that hypothesically finds utility in either the prophylaxis or therapeutics of UV damage to human skin, would benefit from the ability to harness light in the 400–450 nm range, because at such wavelengths there is little danger of new thymine dimer formation in DNA. We therefore considered if it might be possible to modify the structure and base composition of UV1C in minimal ways, so as to endow it with the ability to photo-reactivate CPD thymine dimers with UV-A and/or visible light. A significant proportion of proteinaceous photolyases utilize a folate/pterin cofactor, MTHF, as their antenna pigment. The chemical structure of MTHF reveals a striking similarity between its pterin
ring and the nucleobase guanine. On this basis, we explored the utility of a commercially available fluorophore, 6-methylisoxanthopterin (6-MI), which shares the pterin ring of MTHF and the complete set of hydrogen-bonding sites of guanine\(^\text{112}\). Indeed, 6-MI has been shown to be able to replace one or more guanine within G-quartets (Figure 2-2c), with relatively little structural perturbation\(^\text{23}\). Mergny and co-workers found that the rate of formation of a parallel, intermolecular quadruplex from 5’-TGGGCGT (where M stands for 6-MI) actually exceeds that of quadruplex formation by the unmodified DNA, 5’-TGGGGGT. Additionally, the stability of the resultant 6-MI-containing quadruplex is only modestly lower than that of the quadruplex formed by 5’-TGGGGGT. Most compelling of all, 6-MI, absorbs strongly in the near-UV to visible region of the spectrum, with an absorption maximum at \(\sim\)345 nm.

We generated nine G→6-MI point mutants of the “wild-type” UV1C DNAzyme (WT UV1C), with each mutant incorporating a single 6-MI into the G-quadruplex, or replacing G23 (which does not participate in the quadruplex yet contact cross-links strongly to the substrate’s thymine dimer). We wished to examine the catalytic as well as light usage properties of these mutants, relative to those of the original WT UV1C DNAzyme. The nine individual mutation sites are shown in yellow in Figure 2-2a.

### 2.2. Results and Discussion

#### 2.2.1. Three Functional Classes of UV1C G→6-MI Point Mutants.

We carried out two different sets of experiment to evaluate the action spectra and catalytic properties of the nine G→6-MI mutant DNAzymes. First, to obtain a broad overview of the properties of the mutant DNAzymes, they were complexed with the TDP substrate and irradiated at nine different wavelengths in the 305–400 nm range from a tunable laser. Irradiation at each wavelength was carried out at a constant power output of 10 mW (under these conditions repair rates varied linearly with laser power), for an invariant duration of 10 min each. All experiments were carried out under single turnover conditions, with the DNAzyme being present at three orders or magnitude higher concentration than the TDP substrate. Unmutated UV1C (“WT UV1C”) was used as a positive control, and BLD, a single-stranded DNA oligonucleotide incapable of folding to a G-quadruplex, as a negative control. As described above, photo-reactivation of the
TDP substrate leads to the formation of two smaller DNA strands, and the time-dependent formation of these strands was monitored using denaturing gel electrophoresis (Figure 2-5). The results (% TDP photo-reactivated following 10 min of irradiation at individual wavelengths in the 305–400 nm range) are shown in Figure 2-3. The positive control, WT UV1C, was expected to be active in the 305–315 nm range, at the red edge of DNA’s absorption spectrum. Indeed, by 325–330 nm, UV1C’s catalytic activity is very low, indistinguishable from that of the negative control (“SS DNA”).

The behavior of the six G→6-MI mutant DNAzymes falls into three distinct categories. First, the mutants G32 and G33 (which incorporate their G→6-MI mutations at the G32 and G33 sites, respectively) show action spectra and photo-reactivation rates notably similar to those of WT UV1C, as seen in Figure 2-3a. These mutants are catalytically inactive in the 340–400 nm wavelength range, where 6-MI absorbs, suggesting that in these DNAzymes the photo-excited 6-MI residues cannot serve as effective electron sources for TDP repair, either for steric or other reasons. A second category of DNAzymes, comprising six mutants (G12, G13, G17, G18, G21, and G22), shows a surprisingly homogeneous behavior of enabling a low level of photo-reactivation across the 305–400 nm spectral range (Figure 2-3b). The most interesting behavior, however, is shown by the G23 DNAzyme, which constitutes a functional category of its own. It shows a high level of TDP repair across the 305–400 nm wavelength range (Figure 2-3c). The above experiments, though broadly informative, yielded only a “percent TDP repaired” figure for each DNAzyme; such a figure did not necessarily correlate with a rate constant (such as \( k_{\text{obs}} \)), because it was uncertain whether in a given experiment we sampled a linear response range, or were close to activity saturation. To obtain authentic \( k_{\text{obs}} \) values from linear initial rates of photo-reactivation, we tested the activity of all nine DNAzyme mutants at two wavelengths: 305 nm (the optimal functional wavelength for WT UV1C) and 345 nm (the peak of 6-MI’s absorption spectrum). The \( k_{\text{obs}} \) values obtained (from two sets of independent experiments) are listed in Table 2-1.
Table 2-1. Rates of thymine dimer repair at 305 and 345 nm

<table>
<thead>
<tr>
<th>Sample</th>
<th>At 305 nm (hr⁻¹)</th>
<th>At 345 nm (hr⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Double-stranded control</td>
<td>1.8 ± 0.7</td>
<td>0.3 ± 0.1</td>
</tr>
<tr>
<td>Single-stranded control</td>
<td>2.1 ± 0.7</td>
<td>0.3 ± 0.1</td>
</tr>
<tr>
<td>WT UV1C</td>
<td>230 ± 30</td>
<td>0.4 ± 0.4</td>
</tr>
<tr>
<td>G12</td>
<td>2.7 ± 0.7</td>
<td>3.2 ± 0.7</td>
</tr>
<tr>
<td>G13</td>
<td>3.0 ± 0.7</td>
<td>5.0 ± 0.7</td>
</tr>
<tr>
<td>G17</td>
<td>3.0 ± 0.7</td>
<td>4.0 ± 0.7</td>
</tr>
<tr>
<td>G18</td>
<td>2.6 ± 0.7</td>
<td>4.0 ± 0.7</td>
</tr>
<tr>
<td>G21</td>
<td>4.0 ± 0.7</td>
<td>6.1 ± 0.7</td>
</tr>
<tr>
<td>G22</td>
<td>3.9 ± 0.7</td>
<td>4.0 ± 0.4</td>
</tr>
<tr>
<td>G23</td>
<td>344 ± 1.1</td>
<td>19.4 ± 0.4</td>
</tr>
<tr>
<td>G32</td>
<td>310 ± 25</td>
<td>0.4 ± 0.4</td>
</tr>
<tr>
<td>G33</td>
<td>370 ± 10</td>
<td>0.7 ± 0.4</td>
</tr>
</tbody>
</table>

Observed photo-reactivation rate constants, \( k_{obs} \), at 305 and 345 nm, for the TDP substrate complexed to the UV1C DNAzyme (WT UV1C), or different UV1C G→6-Mi point mutant DNAzymes.

At both 305 and 345 nm wavelength, the existence of TDP as either single-stranded oligonucleotide or component of a duplex, makes little difference to its low (background) rate of photo-reactivation. However, when TDP is complexed with WT UV1C, a >100-fold enhancement of repair occurs at 305 nm, while at 345 nm it remains at the background level.
The G32 and G33 mutants show a photo-reactivation profile similar to that of WT UV1C (Figure 2-3a). It is likely that at 305 nm these guanines (32 and 33) do not participate in thymine dimer repair, possibly because they do not lie sufficiently close to the thymine dimer to serve as efficient electron donors (indeed, earlier cross-linking studies suggested precisely such a lack of proximity). The six mutants DNAzymes of the second functional class all mutated within their G-quadruplexes (vide infra) are surprisingly poor at photo-reactivation at 305 nm. However, at 345 nm they are consistently 10–20-fold more effective at photo-reactivation than either of the negative controls, WT UV1C as well as the G32 and G33 DNAzymes. The comparable behavior of all six mutants in this class is consistent with the functional redundancy of these six quadruplex guanines. In particular, it suggests that no single guanine within WT UV1C or
the mutant DNAzymes may have the role of sole electron supplier to the thymine dimer (Figure 4).

How then to account for the poor photo-reactivation by these six DNAzymes at 305 nm? One possibility is that a G→6-MI substitution at any of these six positions either abolishes or destabilizes the DNAzyme’s G-quadruplex. To test this, we carried out irradiation experiments at 345 nm with WT UV1C and the G13, G17, and G23 DNAzymes (G13 and G17 belonging to this second functional class, above), and found that all the above DNAzymes photo-reactivated the substrate at 345 nm with far higher rates in the standard, sodium-containing, irradiation buffer (20 mM Tris, pH 7.4, 2 mM EDTA, and 200 mM NaCl), compared to in a buffer where all Na⁺ had been replaced with Li⁺ (Figure 2-6).

This result clearly indicates that G-quadruplexes do indeed form within all the above DNAzymes (Na⁺ stabilizes G-quadruplexes, while Li⁺ does not). Could it be that such quadruplexes differ structurally from, or fold with lower efficiency than, the catalytically optimal G-quadruplex formed within WT UV1C? To test these ideas, we first measured the CD spectra of WT UV1C, bound to its substrate analog LMP, as well as of the nine mutant DNAzymes (all bound to LDP). Figure 2-7 displays these data cumulatively, juxtaposed with the CD spectra of WT UV1C bound to its complementary oligonucleotide (to give a B-DNA duplex); as well as the spectrum of a standard, intermolecular parallel-stranded G-quadruplex, (T₇G₅A)₄. The parallel-stranded G-quadruplex standard shows a characteristic positive peak at 265 nm¹¹³, while the B-DNA duplex shows a characteristic peak at 280 nm²⁰. All 10 DNAzyme-LDP spectra show a broad positive peak in the 260–290 nm region; this peak is likely a composite of contributions from the predicted parallel-stranded G-quadruplex within each DNAzyme along with those from duplex elements within the DNAzyme-LDP complexes.

Figure 2-7 shows that, interestingly, the amplitudes of the 260–290 nm positive peak from the different DNAzyme-LDP complexes (shown are the spectra of WT UV1C, G12, G21, and G33) vary significantly (high for WT and G33; lower for G12 and G21) relative to the almost invariant duplex signals generated from all four DNAzymes. This may reflect the relative levels of G-quadruplex formed in the four DNAzyme–LDP complexes. An additional detrimental factor for the catalytic properties of six mutants DNAzymes of the second functional class may be the formation of charge transfer
complexes between their 6-MI residues and neighboring, stacking guanines (resulting in 6-MI− and G+)36. The UV−vis spectra attributable to the presence of 6-MI in different DNAzyme-substrate complexes (Figure 2-8) do indeed show modest, but definite, shifts in the absorption spectra, which is consistent with the above notion114.

In contrast to the mutant DNAzymes of the second class, the unique and efficient photo-reactivation profile of DNAzyme G23, shown in Figure 2-3c, is borne out by its $k_{obs}$ numbers (Table 2-1): at 305 nm its $k_{obs}$ values are, respectively, 1.5-fold higher than that of WT UV1C and ~170-fold higher than those of the negative controls; at 345 nm, they are ~70-fold higher than those of either negative control or WT UV1C. The G23 G→6-MI mutant, therefore, efficiently photo-reactivates the substrate thymine dimer throughout the $\lambda$= 305−400 nm range. As described above, cross-linking experiments on WT UV1C complexed with TDP analogues had established that the G23 residue lies in close proximity to the thymine dimer. However, the G23 residue has also been shown to be non-essential for the 305 nm photo-reactivation by WT UV1C, since it can be mutated to inosine (which has a higher ground state oxidation potential than guanine, 1.5 V compared to 1.29 V, both relative to NHE115) without any discernible reduction in the DNAzyme’s catalytic function111. How then does the G23 mutant DNAzyme achieve such a strong photo-reactivation function? Two possibilities, not mutually exclusive: (1) electron transfer from 6-MI at the G23 position occurs via direct contact with the thymine dimer (distinct from the likely more complex electronic path to the thymine dimer from 6-MI’s located in the distal of the two quartets of the quadruplex); and (2) being extrahelical, the photoexcited 6-MI residue at G23 likely does not form significant charge separation complexes with proximal guanines. If hypothesis (2) is true, the fluorescent lifetime of the G23 DNAzyme should exceed that of the G22 DNAzyme, for instance. Future experiments may investigate this question.

2.2.2. Is Efficient Photoreactivation by the G23 G→6-MI DNAzyme Wholly a “Proximity” Effect?

Given the strong catalytic performance of the G23 mutant DNAzyme across the 305−400 nm wavelength range, we asked whether this capability arose exclusively from the close spatial proximity of its 6-MI and the thymine dimer. If this is true, could the observed properties of the G23 DNAzyme be replicated in a structurally unrelated
system, where a 6-MI residue was deliberately positioned close to a thymine dimer? To investigate this possibility we constructed two test duplexes, both incorporating the TDP substrate as a component strand. The complementary strand in these duplexes was either one or other of the DNA oligonucleotides, DS1 and DS2:

DS1: 5’ ACTCGTACGCACAC(6-MI)TACATGTAG 3

DS2: 5’ ACTCGTACGCACAC(6-MI)ATACATGTAG 3

The oligonucleotide X offers a perfect complement to TDP, except for a single 6-MI residue (“M” in Figure 2-4a) substituting for the “AA” sequence that would normally base pair to TDP’s thymine dimer. This duplex, DS1, might be expected to have a kinked structure, such as shown in Figure 2-4a. The oligonucleotide Y incorporates an additional adenine, offering the sequence “6-MI-A” across from TDP’s thymine dimer (in duplex DS2). In both the DS1 and DS2 duplexes, the 6-MI residue and the thymine dimer are positioned proximal to each other. Figure 2-4b plots the $k_{obs}$ values, measured at 305 and 345 nm, of (i) a complex of TDP and the G23 DNAzyme; (ii) DS1; and (iii) DS2. It is evident that photo-reactivation $k_{obs}$ values of all three TDP-containing complexes at 345 nm are close, adding credence to the notion that photo-reactivation by the G23 DNAzyme at this wavelength results from the proximity of its 6-MI chromophore and the thymine dimer. However, the 305 nm $k_{obs}$ values for the three TDP complexes are strikingly different: the G23 DNAzyme photo-reactivates the thymine dimer 2 orders of magnitude faster than either DS1 or DS2. Clearly, at 305 nm, 6-MI’s proximity to the thymine dimer is not a determining factor. Most likely, in the G23 DNAzyme, electron transfer for photo-reactivation at 305 nm occurs from a source distinct from the 6-MI, i.e., one or more of the quadruplex guanines. These data help reinforce the notion that both the WT and mutant UV1C DNAzymes have multiple, likely redundant, options for transferring photo-reactivating electrons to the thymine dimer.
2.2.3. A Variety of Chromophores Can Functionally Substitute for Guanine at the G23 Position of UV1C.

Given our understanding, above, that (a) guanine 23 of WT UV1C 6-MI is not a participant in the DNAzyme’s G-quadruplex and (b) 6-MI placed at the 23 position can efficiently photo-reactivate at 345 nm, we reasoned it may be possible to place divergent chromophores at this locus of the DNAzyme, and so obtain UV1C variants capable of utilizing visible light more effectively. We therefore substituted a structurally and spectroscopically distinct dye nucleoside at position 23: 7-(2,2′-bithien-5-yl)-imidazo[4,5-b]pyridine, or Dss\textsuperscript{42} (Figure 2-5). The absorption spectrum of the G23→DSS mutant DNAzyme is shown in the Supplementary Figure S7. This G23→Dss mutant DNAzyme, complexed with TDP, was irradiated at 400 nm or at 420 nm. Following 30 min of irradiation at 400 nm, a G13→6-MI control DNAzyme showed ∼10% TDP repair; a
G23→6-MI control DNAzyme showed ~15% repair. However the G23→DSS DNAzyme showed ~43% repair. At 420 nm (where 6-MI no longer absorbs), the two control DNAzymes had zero activity; however, the G23→Dss DNAzyme showed 27% repair. These experiments confirmed that it was indeed possible to position a variety of chromophores at the G23 site of UV1C, and thereby gain significant flexibility over the desired action spectrum of the resulting DNAzyme.

![Chemical structures of DSS, Perylene, “E”, and Dicyanomethylene aminostyril pyran, “K”](image)

**Figure 2-5. Nucleotides with significant absorption in the visible**

The above nucleotides were incorporated into the UV1C sequence to extend its activity further into the visible. These bases all lack the ability to form hydrogen bonds with the canonical DNA bases.

The success of the incorporation of the unusual base-analog Dss into the G23 position of UV1C emboldened us, and led us to search for other potential chromophores that might extend the action spectrum even further into the visible. The lack of hydrogen bond donors or acceptors on DSS broadened our prospective search to include a greater range of nucleoside modifications, but focused on those which were still connected to a deoxyribose sugar at the 1’ position. Some of the most photostable chromophores for DNA labelling have been developed by the lab of Erik Kool\cite{39}. The compound he refers to as “E” is in fact a nucleoside of the polyaromatic hydrocarbon perylene, which shows two absorbance peaks between 400-450 nm, but little absorbance in the UV-C region. A nucleoside containing a dicyanomethylene aminostyril pyran moeity he refers to as “K”, shows a broad absorption band throughout the UV, reaching to 600 nm in the visible. Based on the extinction coefficient and
bandwidth of this absorption, it is most likely the result of intramolecular charge-transfer. E and K were synthesized into oligonucleotides at the G23 position with the intent that besides conferring new photoreactivation activity in the visible, these variants would retain the full quadruplex-dependent repair activity at 305 nm present in the 6-MI G-23 mutant. Where as DSS was only ever incorporated into the G-23 position, oligonucleotides incorporating E and K were prepared with E and K in place of G-17 as well as G-23. Our hope was that these bulky modifications in the G-17 position would disrupt the quadruplex and abolish all activity. As in the case of 6-MI, fully complementary oligonucleotides were designed to place E and K directly across from dimer in TDP. The spectroscopic properties of these oligos were investigated, with the hopes of observing changes upon formation of secondary structures, but in the case of all the E and K oligos, only small perturbations to their absorption and emission spectra were observed on annealing to complementary strands to form double-stranded species or the putative catalytic structures.

In terms of photoreactivation, the E and K fluorophores showed considerable repair at 305 nm in the double-stranded controls, irregardless of salt, although in all cases slower than the WT UV1C. The E and K UV1C mutants were then tested in buffers containing no salt, 200 mM Li+ and 200 mM Na+, in order to confirm that the G-quadruplex fold was in fact necessary for repair at 305 nm. The DNAzyme constructs all required a salt in the buffer for activity, showing only background activity in the case of no salt. This is to be expected, as only short stretches of complementarity exist between the UV1C and TDP sequence. It’s likely that in the cases of no-salt, the UV1C mutant can’t even bind its substrate. Surprisingly, the G-17 E and K mutants, when irradiated at 305 nm, showed equal, but modest, repair rates in the case of both lithium and sodium, in both cases lower than the double-stranded control. The equivalency between the rates in both lithium and sodium indicate that this repair process is wholly quadruplex independent. The G-23 mutants of E and K both showed a small increase in activity when in the presence of sodium as compared to lithium, but were again in all cases lower than the rates of the double-stranded controls. These results at 305 nm seem to indicate that the incorporation of a bulky nucleoside with no hydrogen bond donors or acceptors significantly changes the structure adopted by UV1C, with the modifications at the G17 position totally abolishing the quadruplex-dependant activity native to wild-type UV1C. The modifications at the G-23 position again showed an increased activity when
in quadruplex-forming salt conditions, but when compared to the activity present in the wild type and the G23 6-MI mutants, both E and K have much lower rates of photoreactivation. These results indicate that the quadruplex-dependant repair pathway is absent or significantly reduced in the case of the G23 mutants, despite their slight salt dependence, and that while a quadruplex may still exist in the structure, the bulky modifications introduced are of a completely different nature from the minimal strategy employed with 6-MI. These results imply a certain limit on the size of structural modifications or modified bases which may coexist with the quadruplex dependant photorepair active site of the UV1C structure.

2.3. Conclusions

The data reported here cumulatively show that WT UV1C, as well as a number of UV1C mutants that incorporate base-analogue dyes such as 6-MI, 2AP, DSS, E and K represent an outstanding and adaptable catalytic system, able to carry out a biocatalytic function in vitro comparable to that carried out by more complex proteinaceous photolyase enzymes in vivo. Figure 2-3 summarizes our current understanding of the multiple ways that UV1C and its variants are able to photo-reactivate the TDP thymine dimer. Regarding the precise mechanism of thymine dimer photoreactivation in these mutant DNAzymes, it is in principle possible that the photoexcited 6-MI moiety acts as a photooxidant rather than a photoreductant, repairing the thymine dimer via the formation of the latter’s radical cation rather than an anion (both the radical cation and anion are highly unstable and revert rapidly to thymine bases\textsuperscript{97, 116}). However, it is difficult at this point to carry out a reliable Rehm–Weller analysis on the likely direction of charge flow\textsuperscript{117}, given lack of information on such key parameters as the oxidation potential of the CPD thymine dimer, as well as the oxidation and reduction potentials of photoexcited 6-MI nucleotides within DNA (particularly, in a purine-rich sequence context, such as in our DNAzymes). Redox potentials of the 6-MI heterocyclic base, dissolved in organic solvents, have been reported in the literature; however, redox potentials of heterocyclic bases are highly sensitive to nearest neighbor effects (for instance, the oxidation potential of an isolated guanine in DNA is 1.20 V relative to NHE, while that of GGG is 0.64 V relative to NHE\textsuperscript{118}) as well as to solvation status.
Currently, no high-resolution structure exists for the UV1C-TDP catalytic complex. However, extensive cross-linking and chemical probing studies, as well as data from experiments described here enable us to create a low-resolution conception of the functional components of UV1C-TDP, in three dimensions. The experiments described throughout, whether measuring catalysis or spectra, represent ensemble averages of all the possible conformations of the DNAzyme system. Several methods exist to try and deconvolute this distribution of conformations. SAXS, or small-angle X-ray scattering, is a novel technique which gives solution-state low-resolution structural information, such as the overall size of a biomolecule complex, as well as an estimate of the variance around the center of this size distribution. If the wild type UV1C-substrate complex differs significantly from its 6-MI substituted mutants, SAXS might detect it, or provide evidence against such a criticism, or show evidence of structural changes upon absorbance of photons. The ultimate method to observe any heterogeneity in the catalysis of the DNAzyme system is to carry out single-molecule measurements of the repair process. If such measurements showed a single rate law for all species, it would be strong evidence that our structural hypothesis is correct and complete. However, if there were multiple rates observed at the single molecule level, it could only be explained by the DNAzyme-substrate complex occupying several slowly interconverting conformations.

2.4. Methods

2.4.1. Oligonucleotides.

Oligonucleotides containing 6-MI were purchased from Fidelity Systems. The DSS phosphoramidite was purchased from Glen Research. Both standard oligonucleotides and those containing DSS modifications were synthesized at the University of Calgary CORE DNA services. Oligonucleotides containing E and K were a generous gift from the Kool lab. Two key DNA sequences were used in this study, UV1C:

5′-GGA GAA CGC GAG GCA AGG CTG GGA GAA ATG TGG ATC ACG ATT-3′
and TDP:

5'-AGG ATC TAC ATG TAT=TGT GTG CGT ACG AGT ATA TG-3'

(T=T refers to a special thymine dimer, which lacks the intervening phosphodiester). The single-stranded control oligonucleotide was LDP, a continuous piece of DNA with the same sequence as TDP, but lacking the latter’s thymine dimer. The double-stranded control was TDP hybridized to a short splint oligonucleotide that was wholly complementary to the central portion of the TDP sequence. Oligonucleotides were size purified in 8% denaturing polyacrylamide gels run in 50 mM Tris borate–EDTA (TBE) buffer. Following elution into TE (10 mM Tris, pH 8.0, 1 mM EDTA) buffer, the DNA was ethanol precipitated, washed twice with 70% v/v EtOH, and air-dried. The DNA pellets were redissolved in 40 µL of 5% acetonitrile (ACN) and 50 mM triethylamine acetate (TEAA), then loaded into Thermo Scientific Pepclean C18 spin columns. The columns were washed twice with 200 µL of 5% ACN, 50 mM TEAA, and the bound DNA eluted with 70% ACN, 50 mM TEAA. The ACN and TEAA were removed by ethanol precipitation, and the DNA pellet washed twice with 70% (v/v) ethanol, air-dried, and dissolved in TE buffer. DNA concentrations were determined by absorbance measurements taken in a Nanodrop spectrophotometer. TDP was prepared as previously described, and 5'−[32P]-labeled with OptiKinase (Affymetrix) in 50 mM Tris-HCl, pH 7.5, 10 mM MgCl2, 5 mM DTT. The kinased DNA was ethanol precipitated and gel purified, eluted overnight into TE buffer, and recovered by ethanol precipitation. Figure 2-8 shows that the UV−vis spectra of all nine G→6-MI point mutants had the expected absorption peak at ∼345 nm wavelength, diagnostic of the presence of 6-MI within each DNA construct. We used circular dichroism (CD) spectroscopy to confirm that all nine point mutants folded to a G-quadruplex (the DNAzymes gave complex CD spectra, owing to their folding to, in addition to a short G-quadruplex, extensive duplex, and other secondary structure elements Figures 2-7), a functional test was carried out with the 6-MI-containing DNAzymes that they were able to photo-reactivate the TDP substrate with >320 nm light, and that this activity was notably higher in a sodium-containing buffer relative to a lithium-containing buffer (Figure 2-6).
Figure 2-6. Sodium dependence of a selection of 6-MI mutants
Observed rate constants ($k_{obs}$) for TDP single-turnover photo-reactivation, at 345 nm, in the presence of excess concentrations of DNAzymes and controls (SPL: a double-stranded control; WT UV1C: unmutated UV1C DNAzyme; G13, G17, G23: UV1C DNAzyme point-mutant with 6-MI at those guanine locations. Irradiations were carried on DNA dissolved in either the standard Irradiation Buffer (20 mM Tris, pH 7.4, 2 mM EDTA, and 200 mM NaCl), shown in red, compared to a buffer where all Na+ has been replaced with Li+ (shown in blue).
2.4.2. CD Spectroscopy.

Figure 2-7. Circular dichroism spectra of the UV1C 6-MI mutants

CD spectra of the substrate surrogate, LDP, complexed with WT UV1C, as well as the nine G->6-MI point mutant DNAzymes (legends shown on right). Two controls structures, whose CD spectra are also shown are: a parallel-stranded G-quadruplex, (5'-T7G5A)4 (shown as a continuous black line > 250 nm) and the duplex form of WT UV1C (shown as a broken black line).

Circular dichroism spectra of WT UV1C as well as of each mutant DNAzyme (all complexed with LDP), in addition to those of controls, were taken with a Jasco 810 CD spectrometer, using in a 0.1cm path-length quartz cuvette (Starna). The DNAs (10µM each) were dissolved in a solution of 20 mM Tris, pH 7.4, 2.0 mM EDTA at and 200 mM NaCl. Spectra were taken over the 200−400 nm wavelength range. The cuvette was rinsed thoroughly with ddH2O in between different samples.

2.4.3. UV−Vis Spectroscopy.

The UV−vis spectrum of each mutant DNAzyme, as well as of controls, was taken with a Cary spectrophotometer, using a 1 cm path-length quartz cuvette (Hellma). Spectra were taken from 230 to 450 nm.
Figure 2-8. The UV-vis absorption spectra of G->6-MI point mutants
Each of the 9 mutant DNAzymes complexed an equimolar amount of the substrate surrogate, LMP. Differences in the absorption maxima can be seen for the different DNAzyme-LMP complexes. These spectra were obtained in TE buffer at pH 7.4, 200 mM NaCl, and 10µM DNAzyme and substrate analog (LMP).

2.4.4. Irradiation Sample Preparation.

Each reaction took place in 100 µL of 20 mM Tris, 2 mM EDTA, pH 7.4, 200 mM NaCl (or LiCl), and 2 µM DNAzyme (or control DNA) and 2 nM 5'−32P-labeled TDP. At these low DNA concentrations, the absorption of light was negligible along the laser path length. After all reagents had been added, except for the salt solution, the mixture was heated to 95 °C for 1 min, and then allowed to cool for 10 min, after which the NaCl was added from a stock solution to aid in the folding of the G-quadruplex. DNA solutions were prepared away from direct bright light and kept under darkness as far as possible.
2.4.5. Laser Irradiation.

![Diagram of Laser Irradiation Apparatus](image)

**Figure 2-9. Laser irradiation apparatus and gel analysis**

*Above.* A schematic diagram for the irradiation apparatus, with a constant 10 mW light output from a Continuum Panther EX, at discrete wavelengths in the range of 305-420 nm. *Below.* Example of analysis of photo-reactivation yields for the $^{32}$P-5' labeled TDP substrate, upon irradiation in the presence of different DNAzymes or control DNAs. Because of the unique structure of TDP (which lacks the linking phosphodiester between the dimerized thymines) photo-reactivation leads to the creation of two pieces of DNA smaller than TDP, one of which maintains the $^{32}$P-5' label. The example gel shown here is a time course for an irradiation of the G23->6-MI mutant at 375 nm.

A Continuum Panther EX OPO laser, firing at 10 Hz, was first allowed to stabilize for 30 min, at each of the wavelength settings used. The laser power at each wavelength was measured with a NIST-calibrated power meter and adjusted to 10 mW. To maximize pulse-to-pulse consistency, the laser was adjusted to its highest possible output and brought back to 10 mW with one or two beam splitters (Thor Labs). An electromechanically driven shutter (Thor Labs SH05 beam shutter) was used to ensure that irradiation times were accurate. The power meter was fixed in place to measure the laser power after the light had passed through the cuvette (Figure 2-9). Laser power was monitored throughout the experiment, and upon observation of any irregularity, the sample was discarded and the irradiation repeated. Each sample for irradiation was...
pipetted into a quartz micro-scale fluorescence cuvette, (Starna). The cuvette was shaken thoroughly to mix its contents prior to the drawing of each aliquot. A null aliquot was removed prior to irradiation, mixed with gel-loading dye, and stored in the dark at −20 °C. Each irradiated aliquot was likewise removed from the cuvette, and mixed immediately with denaturing dye to dissociate the DNAzyme–substrate complex, and stored in the dark at −20 °C. For laser irradiation, the cuvette holder was positioned such that the laser passed through the greatest length of the sample. The cuvette was rinsed thoroughly with ddH₂O between experiments.

2.4.6. **Gel Analysis.**

Each DNA aliquot was brought to room temperature, mixed thoroughly, and run in 8% denaturing polyacrylamide gels run in 50 mM TBE buffer. Samples were loaded to ensure at least 10 counts per second in each lane. Gels were exposed to phosphor screens, which were then scanned with a Typhoon phosphorimager (GE). Peak intensities were quantitated using the ImageJ software. The integrated area of a “repaired” band was divided by the sum of the “repaired” and “unrepaired” areas of that particular lane, then multiplied by 100 to give a percentage value for repair. Any small initial amount of repaired DNA, at time = 0 (corresponding to no laser irradiation), was subtracted from each repair percentage in the irradiated samples.

2.4.7. **Kinetic Analysis.**

Initial rates were determined by regression through data points, measured as a function of time, that were in the linear range (i.e., represented <10% overall repair). Typically, five data points were measured within the linear response range. Duplicate samples were irradiated in every case, to estimate the error in each measurement. Figure 2-10 shows a typical plot of the fraction of thymine dimer repaired as a function of irradiation time.
Figure 2-10. The linear region of a typical time-course
A representative initial rate measurement, reporting the thymine dimer repair within the TDP substrate by the G13->6-MI DNAzyme, using irradiation at 375 nm with a flux of 10 mW.
Chapter 3. Repair of an authentic DNA substrate

3.1. Introduction

Cyclobutane thymine dimers or, more broadly, cyclobutane pyrimidine dimers (CPDs) are the major UV light-induced lesions found in genomic DNA. They are formed by a [2 + 2] photochemically allowed cycloaddition reaction that covalently links two adjacent thymine bases on a given DNA strand\textsuperscript{119}. When these photoproducts occur in living cells, the consequences is manifold, ranging from acute skin erythema to skin cancer\textsuperscript{120}. Different organisms use a multitude of strategies, often involving the removal and replacement of the affected base, nucleotide, or stretch of DNA, to repair these potentially dangerous lesions\textsuperscript{121,122}. A unique repair strategy, found in all kingdoms of life, involves photoreactivation of the dimers back to monomers by photolyase enzymes\textsuperscript{99}. Photoreactivation of CPDs by these enzymes occurs by way of light-induced charge separation and transfer phenomena. Typically, light of \textasciitilde 400–450 nm wavelength is absorbed by embedded antenna pigments and funneled via radiationless transfer to a fully reduced flavin cofactor. Electron transfer then occurs from the photoexcited flavin to the targeted dimer within the active site of photolyase-bound DNA. The resulting thymine dimer radical anion reverts rapidly to the individual thymine bases, and the catalytic cycle of the photolyase is completed with the return of the donated electron to the bound flavin cofactor.
Figure 3-1. Comparison between TDP, LDP and LMP
(a) Three single-stranded substrates for the UV1C DNAzyme. TDP is a gapped DNA substrate, incorporating a central cyclobutane thymine dimer but lacking entirely the phosphate moiety linking the two deoxynucleotides involved in the thymine dimer. LDP is an ungapped version of TDP, a continuous DNA oligonucleotide with an intact phosphodiester linking the dimerized thymidines. LMP has the same nucleotide sequence as LDP but lacks a thymine dimer. (b) Reaction schemes of UV1C with a gapped substrate (TDP) and ungapped substrates (LDP and LMP). Photoreactivation of TDP leads to two small DNA products, T1 and T2. LDP and LMP, can, in principle, interconvert with each other.
A number of successful chemical model systems that mimic photolyases have been reported\(^\text{123}\). Largely, these take advantage of charge transport from a photoexcited organic or organometallic moiety covalently appended to a CPD-containing duplex DNA\(^\text{124}\). Catalytic antibodies have also been described for CPD photoreactivation, in which tryptophan moieties functionally substitute for the flavin cofactor of photolyases\(^\text{95}\).

The first example of CPD photoreactivation by a purely nucleotide-based polymer was demonstrated by the UV1C DNAzyme, a catalytic DNA selected from a random DNA library via in vitro selection (SELEX)\(^\text{57}\). It was directly selected for its ability to photoreactivate a cyclobutane thymine dimer positioned within a gapped single-stranded DNA substrate, TDP (Figure 3-1a). UV1C has been shown to be an authentic enzyme, capable of multiple turnovers of the TDP substrate. It also obeys Michaelis–Menten kinetics. It photoreactivates most optimally (assessed by the \(k_{\text{cat}}/k_{\text{uncat}}\) ratio) in the 300–310 nm wavelength range, with an activity peak at 305 nm. Curiously, UV1C’s photoreactivation activity requires no extrinsic cofactor or photosensitizer (although serotonin had been provided as a potential sensitizer during the in vitro selection process). It is proposed that one or more photoexcited guanine bases within its present G-quadruplex-fold act as the photosensitizer/electron source.

Remarkably, purely nucleotide-mediated CPD photoreactivation has since been found to occur within standard duplex DNAs, with guanine bases\(^\text{96}\), 8-oxoguanine\(^\text{98}\) (an oxidized product of guanine found in cellular DNA), as well as guanine-adenine exciplexes\(^\text{72}\) invoked as the likely photosensitizers for these processes.

Although no high-resolution structure has been reported yet for the UV1C: TDP complex, a series of biochemical and biophysical experiments\(^\text{111}\), including contact cross-linking\(^\text{29}\), have shown that, as hypothesized, the dimerized thymines within TDP are proximal to key guanines of the UV1C G-quadruplex (Figure 3-1b). These studies highlight the likelihood that the UV1C G-quadruplex both mimics the light harvesting antenna pigments found in protein photolyases and also serves as the electron source for photoreactivation (the role played by the flavin moiety found in proteinaceous photolyases). The multiple-turnover capability of the DNAzyme and lack of observation of axidative damage in the DNAzyme itself is evidence that the electron donated to the dimer is returned to the quadruplex.
The primary sequence of UV1C is largely intolerant to mutation. Interaction with TDP occurs through, primarily, two short stretches of Watson–Crick complementarity (Figure 3-1b), one located within a loop of UV1C’s core G-quadruplex. With regard to substrate usage, UV1C strongly prefers the TDP substrate (Figure 3-1b, top), which is an artificial, gapped DNA, lacking the phosphate between the dimerizing thymines that would normally form the phosphodiester linkage of the DNA backbone111. Exploration of other nucleic acid substrates incorporating pyrimidine dimers at the gap have shown that while UV1C strongly prefers deoxyribose sugars, it is largely indifferent to whether the more distal parts of the substrate are DNA or RNA. UV1C photoreactivates with the following order of preference: dT=dT ≫ dT=dU ≈ dU=dT > dU=dU > rT=dT > dT=rT (where “d” stands for deoxyribose and “r” for ribose). All other pyrimidine dimer combinations tested have been exceedingly poor substrates, including dimers incorporating cytosine.

With consideration of the earlier studies, the UV1C DNAzyme has never been tested for its ability to reactivate CPDs within the biologically authentic, ungapped (continuous) DNA substrate. This was in part due to the methodology used to measure the activity of UV1C. Photoreactivation of TDP or any gapped analogue gives rise to two smaller DNA fragments (shown in the upper schematic figure of hypothetical gels in Figure S1, Supporting Information). Thus, the activity of UV1C, when using the gapped substrates, could be easily monitored using denaturing polyacrylamide gel electrophoresis57. An ungapped (continuous) variant of TDP, termed LDP (Figure 3-1a, middle) would require more complex methodologies, for both its creation and analysis of photorepair. Therefore, it was unaddressed whether UV1C, selected for photoreactivation of an artificial, gapped DNA substrate (as well as being, generally, highly selective in terms of its substrate usage), can accept a single-stranded DNA with an intact sugar–phosphate backbone as a substrate.

Another important question that could not be addressed using the gapped TDP substrate was: Does UV1C exclusively catalyze thymine dimer reactivation with 305 nm light, or does it also catalyze the de novo formation of thymine dimers? TDP is able to bind transiently and form a Michaelis complex with UV1C, with a $K_m$ of .58 µM125, but the two small reactivation products of TDP (T1 and T2-- Figure 3-1b) individually bind very poorly to UV1C and not necessarily in conformations that are productive for rejoining.
However, if an ungapped, thymine dimer-containing single stranded DNA, such as LDP (Figure 3-1a) were to be a bona fide substrate for UV1C, its photoreactivated product (LMP: Figure 3-1a) would be expected to bind UV1C with an affinity comparable to that of LDP, in addition to having its adjacent thymines in close juxtaposition for any potential de novo formation of dimer.

The motivation for this work was therefore 2-fold: (1) to test whether UV1C can photoreactivate a thymine dimer contained within an authentic, continuous, single-stranded DNA, LDP, and, (2) to test whether UV1C can catalyze not only photoreactivation but also de novo formation of thymine dimers.

3.2. Methods

3.2.1. Oligonucleotides.

All oligonucleotides were obtained from IDT DNA synthesis lab. The oligonucleotides required to create the LMP and LDP substrates consisted of the following sequences: L1: AGG ATC TAC AT, L2: pGTA TTG TGT, L3: pGCG TAC GAG TAT ATG (with “p” representing a 5'- phosphate modification). The oligonucleotides used to generate the TDP substrate were the following: T1: AGG ATC TAC ATG TAT and T2: TGT GTG CGT ACG AGT A TA TG. Oligonucleotides were also obtained complementary to the mentioned sequences in order to generate double-stranded DNA for DNA ligation and thymine dimer formation experiments. The nucleotide sequence for UV1C is 5'-GGA GAA CGC GAG GCA AGG CTG GGA GAA ATG TGG ATCACGA TT-3'. The sequence for LMP is 5'-AGG ATC TAC ATG TAT TGT GTG CGT ACG AGT ATA TG-3'. In LDP, the thymines underlined in LMP are dimerized. In TDP, the same thymines are dimerized; additionally, the phosphate moiety linking the dimerized deoxythymidines in LDP is wholly missing.126

3.2.2. Thymine Dimer Generation and HPLC Purification of Thymine Dimer Containing Oligonucleotides.

The thymine dimer containing substrates were synthesized following the methodologies mentioned previously. For LMP and LDP substrate preparation, acetophenone triplet sensitized thymine dimer formation was induced in the L2 DNA
hybridized to the splint (31-mer) according to the protocol used for TDP synthesis. Following irradiation, the monomer (L2M) and dimer (L2D) containing L2 DNA reactions were ethanol precipitated and reconstituted with 20 µL of TE. A C18 25 cm Vydac protein and peptide column was used for reverse phase separation of L2M, L2D, and Splint on an Agilent 1100 HPLC system with a fraction collector. Buffer A contained 5% CH3CN v/v in 50 mM TEAA, pH, 7.6, and Buffer B contained 70% CH3CN v/v. Samples were injected (20 µL) with a method of A: 95%/B: 5% at time 0 min to A: 80%/B: 20% at 20 min with a linear gradient. Denaturing conditions were applied with a temperature controller held at 60 °C. Fractions of L2M and L2D were separated and collected, with the dimerized species eluting 1 minute before the starting material. The splint DNA had a longer sequence length than the L2 DNA providing a significantly longer retention time than the monomer and dimer DNA. L2M was also collected for LMP construction in order to ensure that both substrates, LMP and LDP, were processed under the same conditions. Figure S5 (Supporting Information) shows typical HPLC traces.

3.2.3. Ligation of LDP and LMP Substrates.

In order to ensure purity of dimer and monomer containing substrates, shorter L2 sequences were dimerized and purified on HPLC (mentioned above) and ligated to flanking sequences to generate the full-length substrates. Ligation experiments were carried out in 50 µL reactions containing 2 µM concentrations of L1, L2M/L2D, L3, and Splint. To each reaction a small amount of 5'-32P-labeled L1 was added to identify and purify products on polyacrylamide gels. Final concentration of 1× of the supplied T4 Ligase buffer was added. Samples were heated to 90 °C for 2 min and then cooled to 20 °C over 5 min. Ten units of T4 ligase was added and reactions proceeded at room temperature for 1 h. Prior to purification, ligation reactions were ethanol precipitated. Following autoradiography, the 35 nt LMP and LDP were excised and eluted from gel using the standard crush and soak method. To investigate whether the prepared LDP does contain the designated thymines as CPD dimers, routine permanganate-footprinting was used to check the reactivity of the relevant thymines in LDP and LMP. The substrates LMP, LDP, and TDP were 5'-32P-end-labeled using procedures mentioned previously.
3.2.4. Permanganate footprinting of LMP and LDP

Solutions of 5'-32P-labeled LMP and LDP were made up to 2 µM unlabeled LMP in order to maintain a consistent DNA concentration. A 64 mM aqueous stock solution of KMnO4 was diluted 21× and 4× with TE buffer, pH 8.0. To 40 µL samples in TE buffer, cooled on ice, 5 µL of the freshly diluted KMnO4 solution (as above) was added. The reactions were quenched after 5 minutes by addition of 5 µL β-mercaptoethanol. The DNA solutions were then ethanol precipitated, and the pellets washed with 70% ethanol and dissolved in 70 µL of 10% aqueous piperidine. The solutions were heated in boiling water for 30 minutes, lyophilized to dryness, then dissolved in denaturing gel loading solution, and analyzed in 13% denaturing polyacrylamide gels. The DMS reaction ladders were generated by standard protocols110.

3.2.5. Laser Irradiation Experiments.

All experiments were conducted with a Continuum Panther EX nanosecond laser with frequency doubled optical parametric oscillator (OPO) tuned to 305 nm, firing at 10 Hz. The beam was allowed to stabilize for 30 min prior to the initiation of any experiment. To ensure stable beam energy, the doubled 355 pump laser was maintained at no less than one watt, and the 305 nm output was attenuated with two beam splitters. Prior to irradiation, the laser’s power output was determined to be 5 mW, as measured on an Ophir 10A powermeter. However, actinometry (Supporting Information) was carried out determine the photon flux through the irradiated solution, and this was calculated to be 1.44 ± 0.09 mEinstein/min. Each substrate was irradiated under the following contexts: SS: single-stranded conditions, in the presence of 2 µM of BLD, a non-G-quadruplex forming, nonradiolabeled DNA, that had no sequence complementarity with the substrate, DS: a double-stranded complex (with substrate bound to a 31 nucleotide splint sequence complementary to the substrate), and in the presence of an excess of UV1C (2 µM). In all cases, 130 µL reactions were set up containing 2 µM total DNA, ~50 nM substrate (5’-32P-labeled), in a solution of 50 mM Tris, pH, 7.2, and 1 M KCl. DNA solutions were heated to 90 °C for 2 min followed by cooling to 20 °C over 5 min. Irradiation durations were controlled using a Thor Laboratories SH05 solenoid shutter switch.
All DNA solutions to be irradiated were pipetted into quartz cuvettes. A null sample was removed from each cuvette without having undergone any irradiation. 32P-end labelled TDP, LMP, and LDP substrates (as single strands, double strands, or complexed to UV1C, see above) were subjected to irradiation at 305 nm. For TDP, in order to assay its repair, 6 µL time point aliquots were mixed to 10 µL of denaturing gel loading solution (formamide with 1 mM EDTA and loading dyes) and loaded directly in a 12% denaturing gel, in order to enable separation of the radiolabeled T1 (5’ product fragment) as well as unchanged TDP. In order to assay dimer repair or formation in the case of the wholly DNA LMP and LDP oligonucleotides, following irradiation, they were subjected to digestion with T4 endonuclease V (T4 PDG). Following irradiation, 5 µL time aliquots were taken. Prior to each time point collection from irradiated solutions, the solutions were thoroughly mixed within the cuvettes for 10 s prior to removal each aliquot. The aliquot so collected was ethanol precipitated following the addition 0.5 µg of glycogen (as carrier), and the pellet was reconstituted, after 70% ethanol washing, in 9 µL of 1.1× PDG reaction buffer (50 mM Tris, pH 7.4, 1 mM EDTA, and 5 µM Splint DNA). The irradiated DNA and splint were annealed by initial heating at 90 °C for 2 min, followed by slow cooling to 20 °C over 20 min in a thermal cycler. Three units (1 µL) of PDG enzyme was added to each sample, mixed, and allowed to incubated at 37 °C for 30 min. To stop digestion, 10 µL of denaturing gel loading solution was added and mixed by vortexing. Digests corresponding to all irradiation time points were then subjected to gel electrophoresis in 12% denaturing polyacrylamide gels and visualized by autoradiography on phosphor screens scanned on a Typhoon imager.

3.2.6. Kinetic Analysis.

Quantitation was carried out with imageJ software (NIH), using the lane box method to obtain histograms of entire lanes. The peak densities of each band in a lane were manually selected, background subtracted and quantified. For each condition, whether single-stranded, double-stranded or DNAzyme catalyzed, all the data for such a condition was simultaneously fit using a nonlinear least-squares routine in the program Prizm 5.0 to an equation of the form:

\[ [\text{LMP}]_t = [\text{LMP}]_\infty + [\text{LMP}]_0 \left( C_1 \times \exp(-k_1 \times t) + C_2 \times \exp(-k_2 \times t) \right) \]
The form of this equation assumes that as irradiation time progresses, the amount of thymines remaining monomeric will reach an asymptote, \([LDP]_\infty\). The second exponential was necessary to ensure a good fit to the data in the DNAzyme catalyzed case only, and can be justified by the existence of a slow and fast population of reactive conformations. Subsequent graphs were plotted using the program Grapher 10.5. The progress of a reaction was determined as a fraction of substrate, whether it be TDP, LMP, or LDP, that contains undimerized thymines. Assuming that the photoreaction converges to an equilibrium (photostationary state), at the photostationary state \([LDP]_\infty = [LDP]_{eq}\) is defined by \(K_{eq} = k_{\text{repair}}/k_{\text{formation}} = [LMP]/[LDP]\). Values of \(k_{\text{repair}}\) and \(k_{\text{formation}}\) were therefore calculated from \(K_{eq}\) and \(k_{obs}\) in each case. Fitting parameters and residuals are summarized in the appendix, table 7.

3.3. Results

3.3.1. UV1C-Mediated Repair of an Ungapped DNA Substrate, LDP.

The LDP substrate was synthesized in order to test the ability for the UV1C DNAzyme to accept an ungapped analogue of the original SELEX substrate. The protocol for generating TDP has been described earlier. In short, two small DNA oligonucleotides, T1 and T2 are positioned next to each other by hybridization to a complementary splint oligonucleotide so that the 3′-terminal thymine of T1 and the 5′-most thymine of T2 are positioned adjacent to one another for triplet sensitizer-mediated dimerization, which exclusively forms cyclobutane thymine dimers. The LDP substrate is synthesized in an quite different way (Figure S2, Supporting Information). First, a small DNA oligonucleotide, L2, which contains one set of adjacent thymines to be dimerized, is irradiated in the presence of a triplet-sensitizer, under the same conditions as the reaction used to generate TDP. Second, the dimerized product, L2D, is HPLC purified away from residual L2 (or L2M) containing undimerized thymines, in approximately 10% yield based on HPLC trace integration. Finally, oligonucleotides L1, L2D and L3 annealed to a splint oligo to produce a nicked DNA duplex. DNA ligase covalently joins L1, L2D and L3 to give LDP. LDP is separated from its reactants on a denaturing gel, in approximately 10% yield, based on gel quantitation. To confirm the existence of a CPD within LDP and verify the LDP sequence, the Maxam-Gilbert chemical sequencing reactions were used. Both LDP and LMP have bands in their DMS
G-ladders consistent with their intended sequences. The lack of bands for the two adjacent thymines seen in LMP is further evidence that LDP has been successfully synthesized.

**Figure 3-2  Permanganate and DMS Footprinting of LMP and LDP**

5'-32P-labeled LMP and LDP treated with KMnO4 (described below). The red-marked thymine bands correspond to the TT sequenced dimerized in LDP and undimerized in LMP. Blue marks indicate other thymines in the LMP-LDP nucleotide sequence that do not participate in CPD dimer formation, and whose reactivity with KMnO4 is comparable between LMP and LDP. "DMS" lanes show guanine sequence ladders, generated by treatment of the DNA with dimethyl sulfate.
In order to create a reliable methodology for distinguishing LDP from LMP, an oligonucleotide of the same sequence as LDP but containing no thymine dimer, a highly specific enzymatic assay involving T4 pyrimidine dimer glycosylase (T4 PDG), is used\textsuperscript{46}. First, the T4 PDG recognizes cyclobutane thymine dimers and binds to the double-stranded sequence. The enzyme uses its N-terminus as a nucleophile, displacing the glycosidic bond of the 5′-thymine base involved in the dimer, creating an abasic covalent enzyme-substrate intermediate. The enzyme then resolves the DNA backbone at this abasic site. The resulting enzyme treated products can be identified by polyacrylamide gel electrophoresis (PAGE), where the undimerized/repaired species will have a slower mobility (higher product band) within the gel, and the dimerized/unrepaired species will have a faster mobility (lower product band) within the schematic figure shown in Figure 3-3.
Figure 3-3. Short irradiations of LMP and LDP in three contexts

Three experimental contexts for the LMP and LDP substrates: as a single-strand (SS), as a component strand within a duplex (DS), and in the presence of excess UV1C (UV1C). Denaturing polyacrylamide gels show time courses for thymine dimer formation and repair in response to irradiation at 305 nm. In these gels, LMP (unreactive to the T4 PDG enzyme) runs as the upper band, whereas the much smaller cleavage product of LDP (cleaved by T4 PDG) runs as the lower band. The top panel a shows gels of LMP irradiated in its three contexts (as single strand, double strand, and in the presence of excess UV1C). The lower panel b shows gels of LDP, similarly irradiated and T4 PDG-treated. Lanes 1–9 for each substrate represent time points taken during irradiation with 305 nm laser light, at 22 °C, at 0, 1, 2, 3, 4, 5, 6, 7, 10, and 12 min, respectively.

Our experimental design (shown schematically in Figure 3-2) is as follows: both UV1C-mediated photoreactivation and de novo thymine dimer formation are studied under single turnover conditions, with [UV1C] > [LDP] or [LMP]. The control reactions are also tested, in the absence of UV1C, with the LDP (or LMP) in single-stranded form (“SS”) or in completely double-stranded duplex form (“DS”). Figure 3-2 also shows denaturing PAGE gels of LMP DNA (top panel) and LDP DNA (bottom panel) digested with T4 PDG enzyme following time-dependent irradiation with a 305 nm pulsed laser over a 10 min time frame. In these gels, LMP (unreactive to the T4 PDG enzyme) runs
as the upper band, whereas the repair/cleavage product of LDP (cleaved by T4 PDG) runs as the lower band. It can be seen, qualitatively, from the gel at lower right ("LDP/UV1C") that LDP is indeed substantially photorepaired to LMP by UV1C within the 10 min irradiation time frame of our experiment, whereas LDP (in the absence of UV1C) as a single strand ("LDP (SS)") or as a duplex ("LDP (DS)") shows far lower levels of repair. The gels in the upper panel show that the pure LMP substrate, under all three experimental/irradiation conditions (with LMP single-stranded, double-stranded, and single-stranded in the presence of excess UV1C) shows very low and mutually comparable levels of de novo thymine dimer formation (leading to formation of LDP) within the experimental time frame.

Figure 3-4. Longer time-course irradiations
Dimer formation and repair with 305 nm laser irradiation. Percent LMP is calculated according to the equation \((100[LMP]/(LMP + LDP))\), plotted against time of irradiation. Data are shown in the order of complexed to UV1C ("UV1C"), in the form of a double helix (DS), and, as a single strand (SS). Fit parameters and residuals are tabulated in appendix table S7.

Figure 3-3 plots time-dependent measurements, made from two independent experiments, on the levels of LMP following irradiation of solutions of pure LMP, and of an 80% LDP/20% LMP composite, in the three experimental situations (SS, DS, and in the presence of excess UV1C). In the simplest possible mechanistic scheme for LMP ⇌ LDP the measured \(k_{\text{obs}}\) for any one of its individual contexts (i.e., SS or DS or bound to UV1C) should have a consistent value, though, in the different contexts they may well be mutually different owing to the existence of distinct ground states. To extract the values of \(k_{\text{repair}}\) and \(k_{\text{formation}}\) from each experimental condition, the observed rate constant, \(k_{\text{obs}}\), for approach to equilibrium \((k_{\text{obs}} = k_{\text{repair}} + k_{\text{formation}})\) and \(K_{\text{eq}}\), the photostationary state \((K_{\text{eq}} = k_{\text{repair}}/k_{\text{formation}})\) can be used\(^{129}\).
The experimental data plotted in Figure 3-3, however, show that for the LMP:UV1C and LDP:UV1C solutions it takes >300 min to approach a stationary state/equilibrium. This represents a notably high level of 305 nm irradiation, which presents a danger of nucleobase bleaching and general structural damage to the DNA. The situation is even more severe for LMP/LDP solutions in their SS or DS contexts, given that even 300 min of irradiation does not lead to convergent photostationary states (Figure 3-3).

To obtain accurate values of $k_{\text{obs}}$ as well as of photostationary states with a minimal degree of irradiation, we adopted a new experimental approach. Solutions containing predetermined, varying molar mixtures of LMP and LDP were subjected to irradiation under the standard conditions. For any single context (e.g., SS, or DS), all LDP/LMP molar ratios should yield common $k_{\text{obs}}$ as well as stationary state values. We systematically subjected different LMP:LDP molar mixtures, in their different experimental contexts, to 305 nm irradiation. One advantage of this approach is that for a given context (e.g., SS or DS) an initial estimate of photostationary states can quickly be made by searching for the LDP/LMP starting ratio whose composition does not substantially change with the course of irradiation. For the UV1C complex (Figure 3-4), the approach to the stationary state does not follow a simple first-order dependence. Therefore, for the UV1C data, constraints are set such that for each LDP/LMP mixture there is a common two-exponential approach to a common equilibrium/photostationary state. For the single-stranded and the double-stranded (SS and DS) mixtures, however, a single exponential approach to equilibrium is sufficient (Figure 3-4).
Figure 3-5. Mixing Experiments
Percent LMP (100[LMP]/([LMP] + [LDP])) plotted against time of irradiation for different molar mixtures of LMP and LDP, in each of the three contexts (UV1C, DS, and SS). All experiments were run in duplicate, and fit to a double exponential decay model using the program Graphpad Prizm 5.0. Fits were constrained to share a single plateau (photostationary state) and decay constant (k_{obs}). Fit parameters and residuals are tabulated in appendix table S7.

3.3.2. UV1C Weakly Catalyzes Thymine Dimer Formation but Strongly Catalyzes Photoreactivation of the Thymine Dimer with 305 nm Irradiation.

Table 3-1 shows the values of k_{obs} and photostationary state (PSS or K_{eq}) obtained for LDP–LMP in the three experimental contexts (SS, DS, and UV1C). In the cases of the two controls (SS and DS) the respective k_{obs} values fit well to single exponentials, the rate constants for photoreactivation (k_{repair}) and de novo thymine dimer formation (k_{formation}) are then simply computed from the relationships: k_{obs} = k_{repair} + k_{formation}, and, K_{eq} or PSS = k_{repair}/k_{formation}. The UV1C reactions, however, show complex kinetic behavior, and here the data are best fit with a two-exponential function, with the first and dominant rate constant being ~20-fold larger than the second. Such two phase exponential behavior is consistent with either a small fraction of the LDP:UV1C complex being less reactive, and on the time scale of the reaction not interconverting with the fast reacting complex, or a slow interconversion of a nonreactive complex to the photolabile complex. Regardless, because of the errors associated with our measurements, we analyze the reactions of LDP:UV1C using only the larger rate constant and the derived PSS (K_{eq}) (Table 3-1).
Table 3-1. Kinetic rate constants and photostationary states under irradiation at 305 nm

<table>
<thead>
<tr>
<th>substrate</th>
<th>context</th>
<th>$k_{obs}$ (min$^{-1}$)</th>
<th>$k_{repair}$ (min$^{-1}$)</th>
<th>$k_{formation}$ (min$^{-1}$)</th>
<th>PSS or Keq ([LMP]/[LDP])</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDP-LMP</td>
<td>SS</td>
<td>0.027 ± 0.001</td>
<td>0.017 ± 0.002</td>
<td>0.010 ± 0.001</td>
<td>1.77 ± 0.01</td>
</tr>
<tr>
<td>LDP-LMP</td>
<td>DS</td>
<td>0.008 ± 0.001</td>
<td>0.005 ± 0.001</td>
<td>0.004 ± 0.001</td>
<td>1.22 ± 0.02</td>
</tr>
<tr>
<td>LDP-LMP</td>
<td>UV1C</td>
<td>0.146 ± 0.012</td>
<td>0.118 ± 0.024</td>
<td>0.028 ± 0.003</td>
<td>4.17 ± 0.03</td>
</tr>
</tbody>
</table>

For these measurements, the flux of photons through the 130 µL irradiation cuvette was measured to be 1.44 ± 0.09 mEinstein/min. Calculations of $k_{repair}$ and $k_{formation}$ were carried out using only the largest rate obtained from the double exponential dependence observed in the case of UV1C. Errors in the plateau and in $k_{obs}$ are correlated, hence, the relative error in the calculated $k_{repair}$ and $k_{formation}$ is obtained by summing the relative errors of the plateau and $k_{obs}$.

A key observation from these data is that there are quite distinct PSS (Keq) states for the three experimental contexts for LMP-LDP. This is not surprising given that the ground states for the three contexts are quite distinct from one another. The significantly higher PSS value (4.17 ± 0.20) reached in the presence of UV1C, relative to the SS (1.23 ± 0.03) and DS (1.16 ± 0.05) contexts, clearly reveals that relative to the controls the binding of LDP and LMP to UV1C perturb the ground states such that photoreactivation is preferred to the de novo formation of thymine dimers.

The half-lives ($t_{1/2}$) for approach of each reaction to the PSS are 4.7 min for (LDP-LMP) UV1C, 25.5 min for LDP-LMP (SS), and 83.4 min for LDP-LMP (DS). It is clear that the UV1C DNAzyme enables a swifter transition to the PSS compared to either control. Table 1 lists the calculated values of $k_{repair}$ and $k_{formation}$ in the three experimental contexts (in the case of UV1C, the “faster” of the two exponentials is the one referred to in the Table). It is evident that at least with 305 nm light irradiation, UV1C does promote a modestly higher rate of thymine dimer formation in LMP (0.028 ± 0.003 min$^{-1}$) than are seen with LMP alone in its single-stranded (0.010 ± 0.001 min$^{-1}$) or double-stranded (0.004 ± 0.00 min$^{-1}$) contexts. Repair/photoreactivation of LDP by UV1C, however, is many-fold higher (0.118 ± 0.024 min$^{-1}$) than of LDP alone in its single-stranded (0.017 ± 0.002 min$^{-1}$) or double-stranded (0.005 ± 0.001 min$^{-1}$) contexts. It is therefore demonstrated that UV1C, originally selected for precisely the photoreactivation of CPD thymine dimers using 305 nm light, does indeed do that. It does so, furthermore, with a significantly greater efficiency than it promotes the reverse reaction, the de novo formation of thymine dimers. In a nutshell, binding to UV1C perturbs the thymine dimer:monomer equilibrium of LDP-LMP toward the monomer.
3.3.3. UV1C’s Photoreactivation of the Ungapped Substrate, LDP, is \(\sim\) 30 fold Slower than the Gapped Substrate, TDP.

How good a substrate is LDP for photoreactivation by UV1C compared to the originally selected for TDP substrate? Does the presence and absence of the connecting phosphate moiety as well the inter-deoxythymidine gap between the two substrates make a difference? We measured the kinetics of photoreactivation carried out in the TDP substrate under identical experimental conditions in which the experiments with the LDP-LMP substrates, above, had been carried out. Figure S3 (Supporting Information) shows the time dependence of TDP photoreactivation in the presence of UV1C, and Figure S4 shows the photoreactivation time-courses of TDP in its single-stranded and double-stranded forms (in the absence of UV1C). The UV1C-mediated photoreactivation of TDP is, even by visual inspection, between 1 and 2 orders of magnitude faster than the photoreactivation of LDP. Accurate fitting of the relatively complex time course for TDP repair requires a function that includes two exponentials as well as one linear component (the linear component corresponding, presumably, to a very slow kinetic process, the uncatalyzed repair process). The time courses for TDP (SS) and TDP (DS) are, by contrast, satisfactorily fit with single exponentials. Table 2 lists the comparative data.

Table 3-2. Comparison of the photoreactivation kinetics of the TDP and LDP substrates

<table>
<thead>
<tr>
<th>context</th>
<th>TDP (k_{\text{obs}}) (min(^{-1}))</th>
<th>LDP (k_{\text{repair}}) (min(^{-1}))</th>
<th>((\text{TDP } k_{\text{obs}})/(\text{LDP } k_{\text{repair}}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>SS</td>
<td>0.008 ± 0.006</td>
<td>0.017 ± 0.002</td>
<td>0.5</td>
</tr>
<tr>
<td>DS</td>
<td>0.012 ± 0.001</td>
<td>0.005 ± 0.001</td>
<td>2.4</td>
</tr>
<tr>
<td>UV1C</td>
<td>3.67 ± 0.001</td>
<td>0.118 ± 0.024</td>
<td>31.1</td>
</tr>
</tbody>
</table>

In the case of TDP, \(k_{\text{obs}} \approx k_{\text{repair}}\), since \(k_{\text{formation}}\) is effectively negligible for TDP (TDP photoreactivation leads to two small DNA oligonucleotides, T1 and T2, with poor affinity for binding to UV1C). The comparative data indicate that while LDP is indeed a good substrate for UV1C, its \(k_{\text{repair}}\) value (0.118 ± 0.024 min\(^{-1}\)) is \(\sim\) 30-fold lower than that for TDP (3.67 ± 0.30 min\(^{-1}\)). Earlier results had shown the major determinants of interaction between UV1C and TDP, and those determinants are identical for UV1C and LDP\(^{111}\). By contrast, the \(k_{\text{repair}}\) values for TDP (SS) (0.017 ± 0.002 min\(^{-1}\)) and DS (0.005
± 0.001 min\(^{-1}\)) are not notably different from those measured for LDP (0.008 ± 0.006 min\(^{-1}\) and 0.012 ± 0.001 min\(^{-1}\), respectively). Thus, while the repair rate constants of LDP and TDP are not so different, in either single-stranded or double-stranded contexts, the UV1C-catalyzed repair of the two substrates is strikingly different in terms of kinetics. One possible explanation is that binding to UV1C has the effect of straining TDP’s thymine dimer toward the photoreactivation transition state to a much greater degree than happens to LDP’s thymine dimer. Another potential explanation is that the negatively charged backbone of the LDP dimer leads to a higher reorganization energy around the charge-transfer transition state, which Marcus theory would correlate with a lower rate of charge transfer and hence repair.

### 3.4. Discussion

The two key observations from our experiments, above, are first, the UV1C DNAzyme, although selected for the specific task of repairing a gapped DNA substrate, TDP, is also capable of repairing an authentic, ungapped, DNA substrate, LDP albeit with notably slower kinetics. The second observation is that UV1C perturbs the equilibrium constant between the thymine dimer and the pair of adjacent thymines by virtue of its binding to the LDP–LMP substrates.

Thus, UV1C is possibly the first in vitro selected catalytic DNA whose ability to catalyze both forward and backward reactions has been systematically investigated. As expected, natural phosphodiester-cleaving ribozymes such as the hammerhead, and in particular, the hairpin ribozyme, have been shown to catalyze both directions of their substrate cleavage/ligation reactions\(^{130}\). In the experiments reported above, because the reaction ground states are different than in the controls, with UV1C complexed to LDP and LMP, both the activation barrier for kinetics (\(k_{\text{repair}}\) and \(k_{\text{formation}}\)) and the thermodynamic driving force for reaction (\(k_{\text{repair}}/k_{\text{formation}}\)) have been perturbed. A component of the structural basis for UV1C’s ability to facilitate repair of the thymine dimer within TDP or LDP is likely the introduction of strain in the dimer, and this to a greater degree in TDP than in LDP. Such induced conformational strain would lower the energy level of the thymine dimer LUMO toward accepting an electron from a photoexcited guanine within UV1C. By contrast, in their pure single-stranded or double-stranded
stranded contexts, both TDP and LDP undergo thymine dimer repair with comparable efficiencies.

Induction of conformational strain in a bound substrate is, of course, a major catalytic strategy of enzymes. Inspection of the protein photolyase literature has not revealed any explicit identification for a role of thymine dimer strain in their mechanism. With the much smaller and structurally less complex UV1C DNAzyme, however, substrate strain could well constitute a plausible catalytic strategy. With regard to the differential ability of UV1C to photoreactivate the TDP and LDP substrates, it is worth noting that the difference between the two substrates is the lack and presence of, respectively, the phosphate bridging the dimerized thymines. Although, again, we have not found explicit prior data on the structural and dynamic differences between such gapped and ungapped CPD thymine dimers, the presence of a single phosphate would impart a full negative charge, as well as different sets of steric clashes and conformational restraints. It is plausible that such differences contribute materially to the differential effectiveness of TDP and LDP as substrates for UV1C.

With regard to de novo thymine dimerization, the quantum yield for adjacent thymines forming a CPD dimer has been shown to depend on their conformation, with conformers having greater thymine–thymine overlap showing a greater tendency to dimerize. The optimal overlapped geometry for the dimerization is closer in structure to the geometry of a cis-syn CPD itself than geometries typically found in either double-stranded or disordered DNA. The UV1C DNAzyme was selected for its ability to bind the gapped DNA substrate, TDP, and to repair a thymine dimer within it. We hypothesize that on binding the ungapped substrate, LMP, UV1C positions its undimerized thymines also in a more favorable geometry for dimerization than experienced by the thymines in the single-stranded and double-stranded controls. Currently, high-resolution structural information on UV1C is lacking entirely; the obtaining of which will undoubtedly enable a clearer view of all aspects of UV1C’s mechanism. In the absence of detailed mechanistic information, we proceed under the assumption that microscopic reversibility holds true with regard to UV1C’s promotion of the forward and back reactions, and that the DNAzyme promotes a common transition state for electron transfer.

The notable substrate discrimination (TDP versus LDP) shown by UV1C does raise the question of whether a gapped thymine dimer is intrinsically easier to repair than
an ungapped one. Some clue toward this may be obtained from the properties of intracellular protein enzymes that process them comparably or differentially. Different CPD photolyases seem to have differential capabilities to photoreactivate pyrimidine dimers incorporating a backbone nick\textsuperscript{133}. Thus, the E. coli PhrB photolyase is competent to photoreactivate nicked thymine dimers. However, PhrA photolyase, also from E. coli, is incapable of photoreactivating “nicked” thymine dimers lacking an intact phosphodiester backbone\textsuperscript{134}. In vitro experiments by Patrick have revealed that a DNA backbone nick present within a thymine dimer precludes photoreactivation by yeast CPD photolyase\textsuperscript{135}. Additionally, it has been observed that the nicked thymine dimer in TDP is not susceptible to deglycosylation by the T4 endo V\textsuperscript{134}. It has also been reported that enzymatic cleavage of the intradimer phosphate linkage likely precedes excision-repair of UV light-induced CPD dimers in the DNA of human cells\textsuperscript{136}. Given that mammals, with the exception of marsupials, no longer code for photolyases, it will be interesting to see whether in the metazoans as well as higher plants in which photolyases are expressed, such an initial intradimer phosphate cleavage routinely precedes photolyase-mediated photoreactivation\textsuperscript{137}.

One potentially negative aspect of the repair of the authentic DNA substrate (LDP) as well as the original, gapped, TDP substrate by UV1C is that the DNAzyme’s ability to promote multiple turnovers of the LDP substrate may be compromised somewhat by virtue of the slower dissociation of the repaired substrate, LMP, relative to the “cleaved” products of TDP repair. Thus, the conceptual and experimental advance of having the UV1C DNAzyme repair an “authentic” DNA may be partially negated by the slowing of substrate turnover. To address this issue, it might be possible to slightly reduce the complementarity with its substrate. This would lead to an increase in $K_m$ for the DNAzyme to its substrate and could potentially lead to multiple turnovers.

Are there larger implications in biology for purely DNA-mediated DNA photoreactivation, such as represented by the UV1C-LDP complex? Our initial report of UV1C and postulation of a likely catalytic mechanism, namely, electron transfer to the cyclobutane thymine dimer from one or more photoexcited guanine within the DNAzyme- - have foreshadowed progressively evolving ideas about purely nucleotide-based photoreactivation of pre-existing CPD lesions within double helical DNA in living cells. Thus, flanking guanines, 8-oxoguanine (8-oxo-7,8-dihydroguanine), as well as tandem guanine and adenine bases capable of forming exciplexes in the photoexcited state
have all been proposed to participate in a purely nucleotide-based photoreactivation of CPD dimers within DNA in living cells. These proposals hold in common the intriguing notion that nucleotide-mediated photoreactivation of CPD lesions is an intrinsic and evolved property of DNA and RNA, a net contributor to their overall suitability as high-fidelity repository (DNA) and transcript (RNA) of genetic information in living systems.
Chapter 4. Conclusions and Future Work

4.1. Discussion

4.1.1. Possible Explanations for the Discrepancy between TDP and LDP repair rates

UV1C was selected based on a single criterion: Its ability to repair a thymine dimer in the TDP oligo, a dimer without a bridging phosphodiester linkage. Experimental results showed that UV1C strongly prefers thymines to all other forms of dimers and is highly sensitive to sugar identity about the dimer. Nevertheless, UV1C was also shown to have the capability of repairing dimers in LDP, which contain a linked dimer, but at much lower rates, over an order of magnitude slower. There are many possible explanations for this behaviour. These explanations are all contingent on the assumption that the rate-limiting step in single-turnover catalysis is electron transfer to the substrate dimer.

The unlinked and linked dimer have many chemical similarities, but in the absence of structural data it’s impossible to know with any certainty how they differ. Perhaps the simplest possible reason for a lowering of catalytic rate is structural. If the reduced flexibility in the case of the linked dimer perturbs the UV1C catalytic complex, it could lead to a greater distance between the dimer and the active site of electron transfer. If the electron transferred from UV1C to its substrate jumps in a single step, its rate would be distance-dependent with an exponential decay constant. In other words, a very small change in distance could easily account for the large difference in observed rate constant. A way to test this assumption is by substituting the phosphodiester linkage in LDP with another structurally similar type of linkage. Such a linkage is available in the case of the several types of thymine dimer phosphoramidites. Reported dimer phosphoramidites include variants where the thymines in the dimer are connected by a methylene linker, as well as with a neutral methyl phosphotriester backbone. Should all of these variants show similar photoreactivation rates to LDP, it would stand to reason
that the rate discrepancy between TDP and LDP is a matter of conformation and structure, not an electronic effect.

Marcus theory predicts the electron transfer rate constant based on the thermodynamic driving force, the coupling between the donor and acceptor states, and the reorganization energy. A fair assumption is that the dimers in TDP and LDP have roughly the same orbital energies, and that the UV1C active site excited state electron energies are unchanged by the backbone about the dimer. Thus, if we assume the driving force is unchanged and the same coupling exists between states, then perhaps the reorganization energy about the dimer differs between TDP and LDP. The reorganization energy refers to the shuffling of dipoles in the solvent required to accommodate the transit of charge during and following electron transfer. As such, the reorganization energy depends on the dielectric constant of the medium. A more polar site of electron transfer, such as in the case of phosphodiester linked dimer, would be expected to have a higher reorganization energy than the unlinked dimer. A way to test this hypothesis is with comparison to the same artificial dimer substrates as discussed above. The natural, linked thymine dimer available in phosphoramidite form is protected with a methyl group at the bridging phosphate. Deprotection with thiophenol/triethylamine/THF (1:2:2) v/v/v is required to remove this methyl group from the linking phosphate, following which the standard ammonia hydroxide deprotection of the rest of the bases in the oligonucleotide occurs. Should this initial deprotection step be omitted, a dimer could be prepared which is phosphate linked but uncharged. Such a substrate could be termed ULDP. Using the same methods as LDP synthesis, but using a dimerizing strand with a methyl-phosphonate at the dimer backbone would be another uncharged, less polar variant, MPLDP. An even less polar dimer could be prepared with the methylene linked dimer phosphoramidite, termed MLDP. If reorganization energy is the primary cause of the reduced activity of TDP vs LDP, barring gross changes to the overall structure of the complex, these neutral phosphate-backbone dimer containing oligonucleotides would be expected to show a repair rates of the order TDP > MLDP > MPLDP > ULDP > LDP. Another wrinkle to this work is the analysis of dimer repair in these substrates. While T4 PDG is capable of cleaving LDP, it’s unlikely that MLDP or ULDP would make good substrates. Thus we would have no choice but to rely on HPLC analysis, which has previously been shown to discriminate between dimerized and monomeric forms of oligonucleotides containing these unusual dimer-containing
backbones. These substrates, like LMP and LDP, are likely to remain bound to the DNAzyme following photoreaction. It would thus be possible in each case to observe a unique photostationary state.

![Chemical structures](image)

**Figure 4-1. Hypothetical backbone constructs capable of photoequilibrium.** (Top) When compared to TDP, LDP has a full additional negative charge on its backbone. This electrostatic charge may account for some of the decrease in the photoreactivity of LDP under UV1C catalysis. (Bottom) To test this hypothesis are proposed some backbone modifications which might have similar structures to LDP, but no charge. Polarity increases from left to right in the bottom row.

The phosphoramidites required to create such oligos are incredibly expensive and their detection time-consuming. Perhaps a simpler set of oligonucleotides is feasible
which might provide initial justification for the synthesis of these more intensive substrates. Chemical synthesis can easily produce oligonucleotides with a 3’ or 5’ phosphate. Thus by synthesizing T1 with a 3’ phosphate and T2 with a 5’ phosphate, previously discussed methods would allow TDP-like constructs to be synthesized with a highly polar phosphate in the proximity of the dimer, on the 3’OH of the 5’ T’s sugar, (5PTDP), the 5’ OH of the 3’ T’s sugar, (3PTDP), or possibly both, (35DPTDP), although such an unusual arrangement about a thymine dimer has not been reported. The only one of these oligos tested thus far is 3PTDP, which leads to a 50-fold reduction in catalytic rate\textsuperscript{111}, putting it in the same ballpark as the 30-fold reduction seen in LDP. However it should be noted that this measurement was at 200 mM NaCl, not the 1 M KCl buffer used for all subsequent LDP work. It’s possible that in this buffer, the rate of repair of 3PTDP could recover to closer to the rate of LDP repair. Such analysis of 3PTDP, 5PTDP and 35DPTDP at identical conditions to LDP would be prudent before proceeding with the synthetically intensive oligos described above. An advantage to these types of species is that they would be amenable to multiple-turnover analysis. The Michealis-Menten behaviors of UV1C on 3PTDP, 5PTDP and 35DPTDP has never been observed. By measuring the \( k_{\text{cat}} \) and \( K_m \), this analysis would perhaps allow some deconvolution of the effect of the additional phosphates on both binding and catalysis.

![Figure 4-2. Backbone constructs incapable of photoequilibrium.](image)

Only 5PLDP has been tested under UV1C catalysis, and resulted in a similar rate reduction to LDP vs TDP.
4.1.2. Formation of thymine dimers in UV1C

A definitive feature of enzymes is their acceleration of the forward as well as the reverse rates of a reaction they catalyze without changing the position of equilibrium. Of course such a description doesn’t apply to photocatalysis because the forward and reverse reactions aren’t microscopic reversals of each other. Each direction requires a photon of light. Our investigations into the effect of UV1C on LDP and its partner LMP revealed for the first time information about the reverse reaction, photocatalytic formation of thymine dimers. When UV1C acts on LDP, it causes a much more rapid approach to a photostationary state, greatly favouring the monomeric form. A surprising finding is that when UV1C is bound to LMP and irradiated at 305 nm, that new cis-syn cyclobutane dimers emerge, as evidenced by cleavage by the highly specific T4 PDG. The rate of decay of dimerization to its photostationary state is also more rapid than the decays of either uncatalyzed control. A few factors could be responsible for this unusual behavior.

Formation of thymine dimers is an ultrafast photochemical reaction. In the time it takes a dimer to form, very little time elapses from the perspective of the motion of atoms. The conformations of pairs of thymines that spend the greatest amount of time overlapped should most rapidly form dimers. Another way of describing this effect is that the transition state of dimerization looks more like the dimer than an isolated pair of thymines. The selection of UV1C ensured that it developed a binding site for its substrate TDP. Much of the binding to TDP is the result Watson-crnick base pairing, but presumably there are weak non-covalent interactions between the cis-syn dimer and the hydrophobic bases of UV1C. Compared to double-stranded DNA, UV1C might hold the monomeric thymines of LMP in a more dimer-like, overlapped conformation which would more quickly dimerize. An inspiration for the 6-MI replacements discussed above was replacement of the same 9 guanines with inosine. Inosine has a higher oxidation potential than guanine. These mutants might retain much of the same structure as UV1C, but have a lower ability to transfer electrons to it. If the effect of UV1C on dimerization is mainly structural, it would be consistent that the same rate enhancement of the reverse reaction be observed in the case of inosine quadruplex mutants. Thymine has a vanishingly low absorbance at 305 nm, but a very low rate of dimerization is still observed even in the absence of any triplet photosensetizers. The thymine dimers formed from direct exposure to UV-B and UV-A light are explained in terms of an excimer of thymines. Should UV1C’s active site stabilize this excimer relative to the
double-stranded or single-stranded cases, this would be another mechanism by which UV1C could enhance the rate of dimer formation.

4.1.3. On UV1C, Prebiotic Chemistry and the RNA world

In the study of the origin of life on earth, an event that took place over 3.5 billion years ago, it is inherently difficult to make conclusions with certainty, or even propose hypotheses which are truly falsifiable. That being the case, it is within our bounds as philosophers to make these hypotheses, and hope that future scientists will find ways to verify or deny them. If nucleic acids are assumed to be the molecule from which life originated, it is worth considering what role, if any, UV1C and other photolyase-mimicking nucleic acids had in the early earth.

In order to have an RNA world, one must first have a sufficient concentration of RNA monomers. Demonstrating that prebiotic chemical conditions could generate these monomers in quantity has been a significant challenge for the chemistry research community. Leading research into this field, the Sutherland group has proposed a network of reactions of the very simple molecules hydrogen cyanide and hydrogen sulfide as catalyzed by UV light. Before life brought oxygen, and hence ozone, into the earth's atmosphere, much more of the solar spectrum shown on the planet's surface. Far from detrimental, the reactivity engendered by UV excitation may have been crucial to generating the molecular diversity from which self-replicating polymers could emerge. The RNA world is thought to have emerged from this primordial soup, based on its intrinsic ability to template its own replication, albeit at very slow rates.

Should these nucleic acid molecules, thought to be the origin of all life on earth, be restricted to hide in the cracks, in the darkness, or deep under the sea? The last universal common ancestor of all life was likely a resident of hydrothermal vents, but analysis of the highly reducing nature of hydrothermal vents on the early earth indicates that the RNA monomers required for abiogenesis could not survive in them for very long. I now propose another property of RNA which may have been crucial to its molecular evolution. Imagine the formation of a variety of potential polymers in the prebiotic earth, sometimes referred to as the prebiotic jumble. These oligomers are assembled randomly from whatever monomers exist at the time, either brought in on meteorites or assembled by reactions on earth. When exposed to sunlight, those
oligomers in the jumble which are highly photoreactive, forming internal lesions or cross-links easily, would themselves be unable to properly self-template. Those oligomers which consisted only of monomers which were resistant to photodegradation would have an advantage from the perspective of molecular evolution\textsuperscript{142}. Similarly, polymers which have self-repair capability, such as in the case of DNA-mediated DNA repair, would have the valuable advantage that when pyrimidine dimerization occurs, depending on the sequence context, further irradiation could repair the dimer, allowing the once damaged oligomer to again template its replication and live on. While quite slow relative to modern photolyase and inconsistent, depending heavily of certain sequences which repair faster than others, this self-repair property is analogous to the self-replicating property inherent in the RNA.

![Figure 4-3. Analogy between self-replication and self-repair.](image)

(Top) Self-templated replication gave way to polymerase ribozymes, which in turn were replaced by the protein polymerase enzymes we see today. (Bottom) The self-repair inherent in purely double-stranded nucleic acids would be surpassed by photolyase–like ribozymes, either cofactor dependant or quadruplex dependant. I hypothesize that photolyase enzymes are protein replacements for pre-existing ribozymes.

Theories of abiogenesis hypothesize that the prebiotic RNA world, where strands replicated in an uncatalyzed manner, eventually led, purely by chance, to the generation of catalytic ribozyme sequences. These ribozymes were able to greatly enhance the replication rates of the sequences around them. Unlike the large catalytic ribozymes
such as the ribosome or the self-splicing introns, statistics dictates that for these first ribozyme sequences to have emerged by chance, they must have been small and simple, more analogous to the self-cleaving hammerhead domain present in RNA viruses. Based on our demonstration of the exceptionally short and simple sequences of UV1C and Sero1C, it's plausible that the same random processes which created the first self replicating ribozymes could also have generated photorepair ribozymes.Ribo-organisms with photorepair ribozymes would have had an even greater advantage than those with mere self-repair. It's impossible to know for sure whether or not such repair existed, but it's hard to imagine how access to such facilities wouldn't have given early life forms an advantage. One could imagine how a community of ribo-organisms which could withstand the harsh UV-B falling on the surface of the early earth, able to repair their photodamaged sequences and regain function, would have had access to the richest feeding grounds of new monomers. At this point, this is pure speculation, but perhaps other logical arguments could link photorepair ribozymes to the prebiotic and RNA worlds.

Considering the transition from the early RNA world to the ribonucleoprotein world. The complexity and sheer scale of the protein synthesis machinery, as well as certain highly unlikely properties of the modern genetic code, seems to imply a long period of evolution in the RNA world. For RNA-based life to have existed for long enough to develop peptide synthesis, they must have been fairly sophisticated. From a biochemical perspective, its interesting to imagine what types of reactions were present in the RNA world and remain to this present day. Membership in this exclusive group of RNA world reactions is based on the principle of parsimony, a sort of Occam’s razor in which the least evolutionary ‘work’ possible is used to explain present circumstance. As noted by White, Gilbert and many others, perhaps modern metabolism, and the ubiquity of key nucleotide-like cofactors, holds some clues to identifying the types of reactivity which were available in the RNA world. In the aftermath of originating the field of paleogenomics, Benner observed that the very ancient, fundamental metabolic pathways would change like a “palimpsest”. A palimpsest is an ancient scroll, with some parts scraped clean and reused to save precious parchment. Some words have been erased and new words written in their place. The impressions of the more ancient writing are left behind, and as words change, the “sentences” or “paragraphs” they fit in remain. Using this logic, Benner argues that in cases where a
reaction could hypothetically be catalyzed by a non-nucleotide-like cofactor, yet in all domains of life a conserved, nucleotide-like cofactor still carries out the crucial steps of the reaction, (albeit buried inside a protien home), that the cofactor, and hence the reaction catalyzed by it, is a relic of the RNA world.

Another way of making this argument is that if a mechanism arose in the RNA world, it would have no choice but to use RNA-like cofactors for catalysis, as no peptide-synthesized molecules had yet emerged. However if a mechanism came about after the ribosome developed peptide synthesis, it would be equally likely to use either a nucleotide-like cofactor, or another unrelated molecule not reminiscent of the RNA world. The minimal list of reactions predicted by this method still includes a complex energy metabolism, and the ability to catalyze carbon-carbon bond formation. The enzyme photolyase is considered the oldest DNA repair protein to evolve\textsuperscript{87}, and is found in all domains of life, experiencing no less than 8 duplications before bacteria and eukaryotes diverged. The nucleotide-like antenna cofactor MTHF, present in the more ancient photolyases, has, during the long course of evolution, in some organisms been replaced by the thoroughly non-nucleotide like deazaflavin, in fact leading to photolyases which have an action spectrum extending to longer wavelengths than photolyases which still use MTHF. Based on the criterion of parsimony suggested by Benner, it is consistent that the RNA world had a ribozyme with photolyase activity. UV1C represents a paradigm for a potential photolyase ribozyme. Also consistent with the RNA world is the UV1C quadruplex topology. Unlike the topological flexibility seen in DNA quadruplexes, RNA has only been observed to form all-parallel quadruplexes, of which UV1C’s active site is one.

In our study intended to change the action spectrum of UV1C, it was found that UV1C is both an resilient photocatalyst and a versatile one, able to gain a new function, namely photorepair activity in the UV-A, while still maintaining all of its original function in the UV-B. This property makes quadruplex-dependent photorepair, such as is observed in UV1C, an ideal candidate for an RNA world progenitor to the modern photolyase family we see today. Consider an RNA organism which has evolved quadruplex-dependent photoreactivation RNAzymes, either UV1C-like or of some different structure of which we do not know of. If these sequences were intolerant of mutation and substitution, development of subsequent photorepair sequences which lead to modern day photolyases would have had to start again de novo, with quadruplex dependent
photorepair eventually lost and protein-based photolyase taking over the role. However, given the incredible tolerance of UV1C to base substitutions in the proximity of its active site, rather than starting again from scratch, it's the path of least evolutionary resistance, or as Benner put it, parsimony, that a gain of function in a quadruplex-dependant ribozyme would lead to photorepair ribozymes which could use the longer wavelengths available to modern photolyases. The full functionality of the quadruplex-dependent pathway with the modification of G23 to 6-MI shows that a ribo-organism could have undergone significant experimentation with modified bases while still maintaining a functional photorepair activity.

The many types of modified bases in tRNA, themselves considered as vestiges of the RNA world, seem to indicate that ribo-organisms would have had the necessary machinery to produce quadruplex-dependent ribozyme sequences with individual bases at particular locations modified to absorb at longer wavelengths. In this model, another hypothetical ribozyme recognizes a quadruplex-dependent photorepair sequence and chemically modifies one of its bases. An example of such a base is the highly fluorescent Wyosine, whose absorbance has a peak at 310 nm\textsuperscript{144}. An even simpler possibility is the guanine oxidation product 8-OG, itself already known to be capable of photorepair, independent of quadruplexes. Its been observed that while the interior guanines in a quadruplex are protected from oxidative damage, the bases at the edge of a quadruplex are often more susceptible to oxidation than those in a duplex\textsuperscript{78}. Thus a quadruplex-dependent ribozyme could more easily be oxidized to contain 8-OG in a catalytic location than a non-catalytic location. If 8-OG is postulated as a flavin precursor, its worth noting that when riboflavin is covalently attached to a quadruplex forming oligonucleotide, irradiation of the riboflavin leads to oxidation of the quadruplex and reduction of the flavin\textsuperscript{79}. This reactivity would allow flavin adjacent to a quadruplex to assume the same partially reduced form present in modern photolyase. Flavin itself is not incompatible with quadruplexes, in fact, the flavin binding aptamer generated by in-vitro selection consists of an antiparallel G-quadruplex\textsuperscript{145}.

As time passed and the RNA world developed into the RNP world, the new, non-quadruplex dependent repair activity could have evolved to gain more function, and eventually its activity could overtake the importance of the quadruplex dependent pathway. With a duplicated function, it's common in evolutionary history for the old function to disappear. Thus, I propose the hypothesis that, consistent with Benner’s
criterion of parsimony, the nucleotide like cofactors still used in photolyase are vestigial components of a photolyase ribozyme. Parsimony suggests that this original photolyase activity was purely nucleotide-mediated. If this is the case, its consistent that these photolyase ribozymes using chemically modified bases evolved from an original, purely quadruplex-dependent photolyase ribozyme. The chemical similarity between MTHF and Guanine is even further indication that Guanine was present, and in fact catalytic, in photorepair ribozymes that led to photolyase. While this hypothesis is at the moment simply that, a hypothesis, the fact is, photolyase is an enzyme universally distributed across all domains of life, and which uses two nucleotide like cofactors, the oldest one of which happens to be derived from guanine.

But why have DNA repair enzymes in the RNP world, if DNA had not yet been developed? This hypothesis seems to support a DNA first, or DNA-concurrent model of a transition to the ribonucleoprotein world. Consider the alternative: If, during the RNP, in which RNA was the only genetic material, but by which peptide synthesis had already developed, why would photolyase develop to use two nucleotide-like cofactors? The proteins available in the RNP world could have bound or synthesized many types of organic molecules and used them as cofactors to effect photorepair. However, had there been an existing photorepair ribozyme, like other ribonucleoproteins, a peptide might have more easily evolved to support an existing ribozyme than to have photolyase evolve again completely de-novo. This concept is again supported by parsimony.

If one postulates that some of the ribocytes in RNA world had, among the majority of their RNA, a precious DNA genome, having the capability to repair that genome should it be photodamaged would be an efficient use of resources, especially if the early RNA catalyzed reactions leading to deoxynucleotides were inefficient compared to the highly advanced RNR\textsuperscript{64}. In the context of the RNA/DNA world, the existence of the capability to catalyze efficient pyrimidine dimer repair is even more advantageous than in the pure RNA world.
4.2. Future work

4.2.1. On time-resolved spectroscopy of the UV1C system

While at this point, there can be little doubt that the DNAzyme UV1C carries out thymine dimer repair in a mechanism similar to photolyase, by transfer of an excited electron to the dimer, definitive, spectroscopic proof of electron transfer has not been shown. Several issues confound this direct observation. The UV1C*TDP system consists of a relatively large number of bases. The excitation wavelength of many of these bases are overlapping. With 8 hypothetical nucleotides in the G-quadruplex active site, it is not clear how the contributions of these nucleosides, which have identical absorption spectra, might be deconvoluted. The hypothesized intermediate species, the guanine radical cation, has a very small extinction coefficient in the visible. Despite these difficulties, recent developments in nucleic acid time-resolved spectroscopy, as lead by Carell and Zinth, offer inspiration as to how such a “smoking gun” might be observed.

Consider the six UV1C 6-MI substitutions which were found to be catalytic in the quadruplex. Each had a reduced rate, relative to both WT UV1C in the UV-B and in the UV-A compared to the non-quadruplex substitution G23. If this behavior is due to a greater charge-transfer quenching between the 6-MI and the guanines adjacent in the quadruplex, it might be possible to use these mutants in a transient IR experiment to observe the G-radical cation hypothesized to exist in UV1C during the photoreaction. Use of the pump wavelength at 340 nm, where guanine bases do not absorb, would allow specific excitation of each of the purported catalytic guanines. Similar excitation of the G23 control would not be expected to produce G-oxidation, as its not stacked to the same degree as the quadruplex guanines. Should excitation at 340 nm lead to the characteristic changes in IR absorbance consistent with the G-radical cation, that would provide evidence for the hypothesis that UV1C and its 6-MI mutants act by electron transfer to the thymine dimer and not any other means.

4.2.2. Modified bases in UV1C

The capabilities of modified, artificial DNA bases and backbones, as well as the modifications that exist in nature, are a theme of this work. While some of this chemical space has been explored, a huge potential range of future experiments still exists.
untested. Perhaps the simplest types are the single-base modifications with commercially available phosphoramidites. The 9 sequence locations tested in this work are ideal sites to test a variety of molecules. The most obvious first choice is 8-OG. Like 6-MI, 8-OG is compatible with the G-quadruplex secondary structure. Like 6-MI, 8-OG has been shown to photo catalyze thymine dimer repair in both the same and in opposite strands in a duplex. Like 6-MI, 8-OG is commercially available as a phosphoramidite, requiring no modifications to the normal coupling procedure, and only minimal changes to the deprotection scheme. Once 8-OG containing sequences of UV1C were prepared, their repair rates and action spectra could be compared to purely double-stranded sequences containing 8-OG as well as to wild-type UV1C. As in the case of 6-MI, CD spectroscopy of these oligonucleotides complexed with their substrate or a substrate analog would give evidence for the conservation of the UV1C quadruplex topology. I would hypothesize that as in the case of 6-MI, 8-OG would red-shift the action spectrum of UV1C, and that 8-OG in the G23 position would leave the native activity of UV1C intact.

By replacing single guanines in UV1C with the base analog 6-MI, DNAzymes were obtained with action spectrums which were a composite of the original UV1C action spectrum and the action spectrum of 6-MI on its own. While this was certainly a surprising finding, what would have been a much better mimic of photolyase would have been a UV1C variant in which there was a modified action spectrum, either of 6-MI or the UV1C native spectrum. In photolyase, two cofactors are employed, one to act as an antenna, transmitting its excitation by FRET to the electron-transfer cofactor. Could such a strategy be employed in the case of UV1C?

The assumption must be made that the core catalytic competency of UV1C comes from the electron transfer ability of the its excited state G-quadruplex. In order to expand the action spectrum of UV1C, it would be necessary to find a way to excite its quadruplex with light other than the light directly absorbed by the quadruplex. There are several reasons why this might be difficult or yield very low efficiency. UV1C is catalytic with light at the red-edge of the absorbance of guanine. A good FRET donor would have to have good emission spectrum overlap with this region, centered at 305 nm. Even with a conservative stokes shift, the remaining DNA bases in the UV1C system all absorb at the wavelengths intended for the antenna. Thus a good FRET partner for the G-
quadruplex would have to absorb in the same region as the natural DNA bases which make up the majority of the UV1C molecule.

The pyrimidine analog 5-methyl zebularine (5-MZ) is a close relative of the cytosine base from which it is derived. Its minimal changes give a new absorption peak at 310 nm, in fact the same wavelength at which UV1C is optimally active. If any base analog were to serve as an antenna for a 6-MI modified UV1C quadruplex, 5-MZ is the least-invasive possible choice. The UV1C active site has previously been shown to be a parallel propeller topology. The diagonal loops in this quadruplex region contain several cytosines which have been shown by chemical crosslinking to be quite close to the substrate dimer. Should these cytosines be replaced by 5-MZ, it might be possible that photons which are not absorbed by the quadruplex itself might in fact excite 5-MZ. The fluorescent properties of 5-MZ indicate it has a much longer excited state lifetime than any of the naturally occurring bases. With its close proximity to the G-quadruplex, this long-lived excited state would be capable of transferring energy to 6-MI substitutions in the active site, leading to an increase in quantum yield of the DNAzyme. Several controls and other experiments would be necessary to confirm the mechanism of such an effect.

It’s been well known that many fluorescent molecules are capable of photocatalyzing thymine dimer repair. If this is the case, an observed increase in quantum yield from incorporating 5-MZ in the loop regions could merely be direct repair. A control experiment, with a DNA duplex containing 5-MZ directly across from the thymine dimer would give an upper limit for such an effect. Luckily, previous work on modified UV1C has already shown that incorporation of Inosine in the quadruplex active site dramatically reduces photocatalysis. Measuring the photorepair rate of a modified UV1C sequence oligonucleotide containing both 5-MZ in a loop and inosine in the quadruplex might be a way to quantify the direct-repair effect of 5-MZ in a UV1C context. Should the “sum” of incorporating 5-MZ in a loop with a competent quadruplex exceed the constituent parts, it would be profound evidence to support an antenna-like behavior.

The above possibilities are all made using commercially available phosphoramidites. Flavin would be an excellent species to examine for its effect on the action spectrum of UV1C. Several strategies exist to incorporate Flavin into an oligonucleotide. In the simplest of these methods, Sugiyama demonstrates a synthesis
of a flavin phosphoramidite which consists of only a single ethylene linkage between the heterocyclic base and a phosphate, sufficient to incorporate a flavin moiety at the 5’ end of an oligonucleotide without too much trouble\textsuperscript{147}. Such an incorporation would be useless in the case of the UV1C sequence, but was used by Majima to create a quadruplex-forming oligonucleotide with an attached flavin\textsuperscript{79}. This was used to investigate the time-dependent reversible oxidation of the attached G-quadruplex as already discussed. The only reported flavin nucleotide which may be incorporated internally to an oligonucleotide was developed from Carell’s lab, and requires several synthetic steps and modifications to the standard coupling methods\textsuperscript{148}. While most oligonucleotide synthesis uses the phosphoramidite coupling, the oxidative fragility of flavin required a switch to the H-phosphonate coupling method for its insertion mid-oligo. This flavin nucleotide differs from standard nucleotides in the nature of its sugar and the linkage between the base and the sugar. Nevertheless, the flavin-containing oligonucleotides prepared by Carell showed surprising stability, and when flavin was incorporated adjacent to guanine, a profound quenching, attributed to charge-transfer with the adjacent G was reported. These oligonucleotides seemed destined to be used to study photorepair of thymine dimers in a double-stranded context, however no such activity was ever reported. It would be fascinating to synthesize a UV1C-like sequence with flavin at the G23 position. In such a species, its possible that both the quadruplex-dependent pathway activated in the UV-B and a new flavin-dependent activity in the UV-A and the visible would coexist, just as in the case of UV1C. Most interesting would be the comparison between the rate of photoreactivation in a UV1C-like construct and a duplex control. Should the rate be higher in the DNAzyme case, it would be further evidence for an antenna-like behavior, this time with the quadruplex itself acting as the antenna rather than the acceptor. However it is also possible, based on our experience with the Kool-designed aromatic bases, that the larger base, and the non-standard sugar attachment, would lead to a reduction in the capabilities of the wild type or a decrease in binding of the substrate. Regardless, such an experiment is the ultimate logical conclusion of the experiments with 6-MI.

4.2.3. A Selection strategy for modifying the substrate specificity of UV1C

UV1C, while a remarkable discovery in its own right, is a much more effective catalyst against its selection substrate TDP, which conspicuously lacks a phosphate
between the dimerized thymines, than a natural, phosphodiester linked dimer containing single stranded oligonucleotide such as LDP. With this in mind, it may be possible, by additional rounds of mutation and selection, to generate a modified UV1C DNAzyme which is a more effective catalyst against LDP. While the original selection relied on an increase in electrophoretic mobility upon DNAzyme-mediated thymine dimer repair, the specificity of the reaction between T4 PDG and the cis-syn cyclobutane thymine dimer found in LDP may allow a separation by maintenance of higher electrophoretic mobility. Consider the following schema: First, prepare pure LDP, as described above. Using the same reverse-template sequence as the single-turnover catalyst UV1A, but apply a 15% mutagenesis to the catalytic region using automated DNA synthesis. Perform PCR on this template using LDP as a primer, generating a diverse library of thymine dimer containing sequences. These would be easily separated from their complementary strands by denaturing gel. This library could then be incubated in quadruplex-forming conditions and irradiated with UV light. The sequences which could repair the dimer would thus be insensitive to T4 PDG, while the dimer-containing species would be cleaved. It has been observed that in the case of double-stranded and single-stranded irradiations, some dimer repair does take place, however the time-scale for the uncatalyzed reaction is an order of magnitude slower. By choosing an appropriate irradiation time, for example at 10 minutes, more than 50% of dimers are repaired in the case of UV1C catalysis, while only a negligible change is observed for the double-stranded or single-stranded cases. These full-length repaired sequences can be separated from the cleaved, unrepaired sequences with another denaturing gel, and amplified by PCR with a biotinylated primer. While an initial diversification could be obtained by automated synthesis, the PCR at this stage could introduce new mutations by using error-prone PCR in the presence of manganese ions. These double-stranded DNA’s contain both the template and catalytic sequences, but not the thymine dimer. The non-biotin strand could be washed off of magnetic streptavidin beads. A new round of PCR using LDP as a forward primer and the same biotinylated backwards primer would leave easily separable full-length UV1A-like sequences. This library could thus be subjected to another round of selection, with shorter irradiation times or lower salt conditions used to improve the fitness of the selected species.
The above selection strategy is the most conservative one possible, using only standard DNA synthesis, commercially available enzymes and previously published methods. Given the starting point of UV1C, it would likely arrive at a similar, but slightly modified sequence with better reactivity to the natural thymine dimer. But in recent years, new nucleic acid chemistries have become available which may allow completely new functionalities and radically increase the diversity of in-vitro selection libraries. Some of these artificial base pairs have already been explored, such as the unusual base pair between DSS and pyrole (Pa). In fact, the Dss base has already been incorporated to allow UV1C to catalyze photorepair with visible light. This artificial base-pair system has been shown to be capable of going through in-vitro selection. The only aptamer ever approved as a drug is an aptamer for VEGF, a growth factor associated with macular degeneration. When the DSS-Pa base pair was incorporated into an in-
vitro selection against VEGF, an aptamer was obtained with higher affinity for the target, better nuclease resistance and thermal stability, as well as less strict requirements of ions for binding. Despite only containing 2 modified bases, the additional diversity of functionalities possible with expanded genetic alphabets monumentally increases the possibilities of in-vitro selection.

However, all of these strategies would likely lead to another DNAzyme which acts on a single-stranded DNA substrate. UV1C has shown without a doubt that nucleic acids are capable of catalyzing the repair of thymine dimers in a substrate single-stranded oligonucleotide. In order to move towards a therapeutic DNAzyme, it would be far preferable to select for a DNAzyme which could act on a double-stranded substrate. There are a few ways to effect such a selection strategy. Like the catalytic antibody for thymine dimer repair, a possible first step towards such a species would be a selection for an aptamer specific for binding thymine dimers in a double-stranded oligonucleotide. One could imagine what features such an aptamer would look for. The thymine dimer is well-known to sharply kink the DNA double helix. This unusual arrangement of the backbone would be a unique feature that such an aptamer could hydrogen bond to. The thymine dimer sits across from a pair of adjacent adenosine bases. Without their usual hydrogen bonding partners, these adenines breathe and have unusual orientations in the double-helix relative to adjacent adenines not across from a thymine dimer. Such an aptamer may itself supply two thymines which may pair to the adenosines.

In order to carry out such a selection, it is necessary to consider several scenarios which could lead to enrichment of non-binding sequences. This first step of such a selection would be to generate a thymine dimer containing duplex immobilized to a column or beads. Using the already synthesized LDP would be one simple way to generate the dimer. An end-biotinylated complementary sequence would allow a small duplex, containing a thymine dimer, to be immobilized on streptavidin beads. Of course, such a selection would only be successful if appropriate negative selection steps were carried out, to eliminate aptamers for the streptavidin beads or an immobilized duplex without a thymine dimer. A random library might also contain sequences which are simply complementary to either strand of the LDP containing double helix. These sequences could affix themselves to the beads by strand displacement and not by formation of novel hydrogen bonds. In order to eliminate these types of sequences from a random library, RNA containing the LDP and LDP complement sequences could be
incubated with the random library. Those members of the pool capable of forming significant RNA/DNA chimera duplexes could be easily cleaved with RNase H and separated from the library by denaturing gel electrophoresis. Left with only sequences which would have to find an alternative strategy to bind the dimer, this library could be applied to the immobilized dimer-containing duplex under high salt conditions to favor intermolecular interactions. Washing first with this high-salt buffer would remove sequences which were not interacting with the dimer, and subsequent elution with buffer of a low salt concentration would allow separation of those sequences which had affinity for a thymine dimer in a duplex. Amplification could be followed by more stringent selection conditions approaching physiological salt concentrations. To ensure specificity, competitive elution with a non-biotinylated dimer containing DNA duplex, and negative selection with elution by non-dimer containing duplex DNA would ensure the pool only consisted of true thymine-dimer binding sequences.

4.2.4. Potential applications for the UV1C-TDP system

UV1C was the result of pure, curiosity driven research, but the techniques developed in studying UV1C may have potential impacts to other fields of nucleic acid research. Outlined below are some proposals for applications of the photoreactive oligonucleotides developed in the study of UV1C.

One such field is DNA computing, or logic circuits assembled from DNA components. While there are many ways to use molecular biology to perform computations, the most programmable and predictable utilize the method of strand-displacement\textsuperscript{149}. The signals representing logical states are therefore the presence or absence of specific DNA single-strands. The readout is often accomplished by molecular beacons sensitive to these DNA signals, converting a particular signal DNA’s concentration to a visible fluorescent signal. While light is used to detect the output of these signals, the input strands are often introduced by simply pipetting in an initiating strand to begin the strand displacement cascade. Perhaps by using the UV1C motif or its many variants, a rapid, light-triggered input method to DNA computers could be implemented. Such a method would consist of a library of TDP-like oligonucleotides, each containing a unique set of sequences containing a gapped phosphodiester thymine
dimer at a known position. Like TDP, members of this library could be easily conjugated to any arbitrary signal sequence necessary for DNA computing. Each sequence could have its own light-triggered cleaver sequence, complementary to a specific dimer, and containing a dye, such as the previously described 2-AP, 6-MI, pyrene or perylene, among many other heretofore untested possibilities. When irradiated at the proper wavelength, each thymine dimer could be selectively repaired, leading to a specific DNA signal being released into solution. Threshold levels for oligonucleotide signals could be maintained by complementary oligonucleotides so that a small amount of photorepair by an unintended wavelength would not lead to inadvertent triggering of the wrong DNA logic processes. The rapid release by a chromophore targeted at a specific wavelength could quickly surpass the concentration of a threshold oligonucleotide.

Many biological processes rely on molecular motors or walkers which transport substrates to specific locations in the cell. The design of an artificial walker system based on thymine dimer repair has been a long-standing goal in our research. Previous attempts by our lab to generate such a system were designed by linking a carpet of TDP-like molecules to a DNA origami surface were never published. The walker design, referred to as a spider due to its many legs, consisted of many biotinylated UV1C-like constructs, connected at the center by streptavidin. In theory, such a system would have allowed light to control the movement of a nanoscale walker. This preliminary work never took off due to the unexpected negative effect of biotinylation on UV1C’s ability to carry out multiple turnovers. Compared to unmodified UV1C, biotinylated UV1C has a higher Km and lower kcat for its TDP substrate. Such a system would be greatly simplified by switching from the complex UV1C repair mechanism, in which many aspects of its catalysis and structure are unknown, to instead use a more predictable, double-helical arrangement between the thymine dimer containing substrate and the arms of the spider. The multiple chromophores already explored in this work could also provide a more precise means of controlling the nanoscale movements of the spider. Consider a sheet of repeating DNA origami Wang tiles A, B, C and D, each augmented with their own 20 base single-strand region. Each single-strand would need to be designed complementary to the 3’ side of the gapped-dimer containing ATDP, BTDP, CTDP, and DTDP sequences. Following hierarchical assembly of the tile sequences into a 2D lattice, the gapped dimer species are added, hybridizing to consecutive tiles in the lattice. A spider to walk on this substrate could consist of a an assembly of
oligonucleotides linked together at the 3’ end with long, flexible peg linkers, each containing an orthogonal thymine-dimer cleaving chromophore. Once added to the primed tiles, individual walkers would be stuck to the surface via hybridization to their respective gapped dimer species. The PEG linkers would need to be long enough to ensure flexibility, but short enough that each walker can only reach across 4 tiles at a time. Photocleaving of the sequence ATDP, triggered by its wavelength, would release the A’ arms. If cleavage was complete, the A’ arm would have no more complementary ATDP to bind to on that tile, and the A’ arm would have no choice but to swing past its B’, C’, and D’ arms to move to the next A tile, moving the center of the spider one tile-length forward. Repeated applications of this cycle allow directed movement across the oriented wang-tile sheet.

The ability to use light to trigger the release of an oligonucleotide could also be of use in a drug delivery system. There has been intensive research into using nucleic acids as pharmaceutical in medicine, however very few active pharmaceuticals are nucleic acids. Their high negative charge makes it difficult for standard nucleic acids to cross membranes, hampering bioavailability. Enzymes also exist to cleave non-native nucleic acids, which are part of our natural defence against viral nucleic acids. One promising form of nucleic acids which overcome some of these problems are spherical nucleic acids. Developed by the Mirkin group, spherical nucleic acids are constructed by linking a carpet of single-stranded nucleic acids through an alkanethiol bridge to the surface of a gold nanoparticle. Chemical steps applied to modified oligonucleotides can cross-link the nucleic acids and completely remove the core nanoparticle without significantly altering the properties of the nucleic acids on the exterior of the sphere. Despite having a huge concentration of negative charge, these spherical nucleic acids are engulfed by cells by phagocytosis, and due to their unusual appearance, not resembling anything else in biology, they have excellent resistance to nucleases. Synthesis of an SNA could quite easily incorporate a gapped thymine dimer oligonucleotide. A complementary oligonucleotide could again position a chromophore, allowing light of a specific wavelength to trigger repair of the thymine dimer, and hence release of the free oligonucleotides from their spherical form. Light activated therapeutics also allow one to localize the position of delivery of a drug, for example only releasing the oligonucleotide therapy in a tumor.
Visible light shows only modest penetration of human tissue. Infrared light is transmitted much more readily. The ability to use IR light for photodynamic release of an oligonucleotide would be advantageous. While extending the action spectrum of a catalytic oligonucleotide into the visual was achieved, infrared radiation lacks the energy to generate the excited states necessary for photorepair. Upconverting nanoparticles (UCNP) such as the lanthanide-doped NaYF4, allow 980 nm light to cause multiple subsequent electronic excitations, resulting in emission of photons of 290 nm\textsuperscript{152}. Such UCNP can be decorated with molecules via a carboxylate moiety. The T1, the 5’ piece of TDP, could easily be synthesized with a 5’ carboxylate modification, without affecting its base-pairing ability or its ability to be incorporated into TDP. A UCNP coated in carboxy-modified TDP-like constructs would allow release of therapeutic oligonucleotides with IR light. Just the absorption of TDP on its own at 290 nm causes a basal amount of photorepair. However, when hybridized to a strand containing the fluorescent base analog 2-AP, it would be expected that a much faster photorelease would be possible, even when irradiated with light in the IR.
References


(10) Shi, H.; Moore, P. B. RNA 2000, 6 (8), 1091–1105.


(37) Matray, T., & Kool, E. JACS 1998 120(24), 6191-6192.


(87) Mei, Q.; Dvornyk, V. PloS One 2015, 10 (9), e0135940.


Appendix.

Supplementary Figures

Figure S1. Hypothetical gels illustrating TDP vs LDP analysis
(Top) Photoreactivation of TDP or its gapped analogues gives rise to two smaller pieces of DNA as products. Thus, the activity of UV1C, when using the gapped substrates, could be easily monitored using denaturing polyacrylamide gel electrophoresis.
(Lower) For separating LDP from LMP (an oligomer of the same sequence as LDP but containing no thymine dimer), a highly specific enzymatic assay involving T4 pyrimidine dimer glycosylase (T4 PDG), was used. T4 PDG specifically recognizes cyclobutane thymine dimers; upon binding, the enzyme cleaves the glycosidic bond of the 5′-thymine base involved in the dimer, creating an abasic site. The same enzyme then cleaves the DNA backbone at this abasic site. Consequently, a denaturing gel following LDP photoreactivation shows the repaired product (LMP) running slower than the substrate itself.
Figure S2.  LDP synthesis scheme
(a) The first step in the synthesis of LDP is annealing a small 5'-phosphate-modified DNA oligomer, L2, which contains only one set of adjacent thymines, to a complementary splint strand. Irradiation of this duplex in the presence of triplet sensitizer leads to the formation of cis-syn cyclobutane dimers. Dimerized species are separated by HPLC. (b) L2D is combined with 5'32P labelled L1 and 5' phosphorylated L3 and annealed to the mutually complementary splint. T4 DNA ligase leads to LDP, which is separated from its splint by denaturing gel electrophoresis.
Figure S3. UV1C-catalyzed TDP repair at 305 nm
TDP repair under UV1C catalysis was measured under identical circumstances and timescales to the LDP experiments. A very rapid first exponential is better depicted in the inset plot.
Uncatalyzed TDP repair at 305 nm
Uncatalyzed TDP repair was measured under identical circumstances and timescales to the LDP experiments.

Figure S4.
Figure S5. HPLC purification and validation of L2D
(1st trace) unirradiated L2. (2nd trace) Irradiated L2, showing the dimerized species L2D, undimerized L2, and the splint cleanly separated. L2D and irradiated L2 was collected and concentrated. (3rd trace) Re-injection of L2D shows no L2, and vis versa (4th trace).
Figure S6. Absorption spectrum of the G23-Dss DNAzyme compared with the G22-6MI DNAzyme
The absorbance of the Dss base extends significantly into the visible when compared to the G-analog 6-MI. Dss has no hydrogen bond donors or acceptors with which it might participate in a g-quadruplex. These spectra were obtained in TE buffer at pH 7.4, 200 mM NaCl, and 10µM DNAzyme and substrate analog (LMP). Note the significantly higher absorption from the Dss substitution.
<table>
<thead>
<tr>
<th></th>
<th>UV1C catalyzed</th>
<th>Single Stranded</th>
<th>Double Stranded</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Best Fit Values</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>k1</td>
<td>0.008677</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>k2</td>
<td>0.1458</td>
<td>0.02717</td>
<td>0.008311</td>
</tr>
<tr>
<td>Plateau</td>
<td>79.91</td>
<td>63.94</td>
<td>53.4</td>
</tr>
<tr>
<td><strong>Std Error</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>k1</td>
<td>0.002596</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>k2</td>
<td>0.01139</td>
<td>0.001252</td>
<td>0.0005491</td>
</tr>
<tr>
<td>Plateau</td>
<td>1.47</td>
<td>0.784</td>
<td>1.259</td>
</tr>
<tr>
<td><strong>95% CI</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>k1</td>
<td>0.03504 to 0.01385</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>k2</td>
<td>0.01231 to 0.1685</td>
<td>0.02462 to 0.02969</td>
<td>0.007212 to 0.009409</td>
</tr>
<tr>
<td>Plateau</td>
<td>76.98 to 82.84</td>
<td>62.35 to 65.53</td>
<td>50.88 to 55.92</td>
</tr>
<tr>
<td><strong>Goodness of fit</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R square</td>
<td>0.9886</td>
<td>0.994</td>
<td>0.9941</td>
</tr>
</tbody>
</table>

**Table S7. Non-linear fitting parameters**

Raw values and associated errors from fitting of photoequilibrium experiments to a biexponential decay model. While the fit is dramatically improved by the second exponential in the case of DNAzyme catalysis, its slow decay constant is more than 2 orders of magnitude below the decay constant of the fast component. Thus only the fast component is used for subsequent calculations. The uncatalyzed cases required only a single exponential.