Synthesis of Modified Thymidine and Intrastrand Cross-Linked DNA Probes to Investigate Repair by O^6-Alkylguanine DNA Alkyltransferases and Bypass by *Human* DNA Polymerase *η*

Derek Kyle O'Flaherty

A Thesis in

the Department of

Chemistry and Biochemistry

Presented in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy at

Concordia University

Montreal, Quebec, Canada

January 2016

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: Derek K. O'Flaherty

Entitled: Synthesis of Modified Thymidine and Intrastrand Cross-Linked DNA Probes to Investigate Repair by O^6-Alkylguanine DNA Alkyltransferases and Bypass by *Human* DNA Polymerase η

and submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy Chemistry (Chemistry)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final Examining Committee:

_____ Chair

_____ External Examiner
Dr. Richard J. Wagner

_____ External to Program
Dr. Laszlo Kalman

_____ Examiner
Dr. Pat Forgione

_____ Examiner
Dr. Sébastien Robidoux

_____ Thesis Supervisor
Dr. Christopher J. Wilds

Approved by: _____
Dr. Peter Pawelek, Graduate Program Director

Janurary 2016 _____
Dean of Faculty

Abstract

Synthesis of Modified Thymidine and Intrastrand Cross-Linked DNA Probes to Investigate Repair by O^6-Alkylguanine DNA Alkyltransferases and Bypass by *Human* DNA Polymerase *η*

Derek K. O'Flaherty, Ph.D.

Concordia University, 2016

O^6-Alkylguanine-DNA alkyltransferases (AGTs) are responsible for genomic maintenance by repairing O^6-alkyl-2'-deoxyguanosine (O^6-alkyl-dG) and O^4-alkyl-thymidine (O^4-alkyl-dT) adducts. AGT-mediated repair was investigated against DNA intrastrand cross-links (IaCL). A variety of cross-linked dimers linking the O^6-atom of dG or O^4-atom of dT were prepared synthetically to produce precursors for IaCL DNA that either lack or containing an intradimer phosphodiester group in the oligonucleotide backbone. Studies with AGTs demonstrated that: 1) O^6-dG-alkylene-O^6-dG flexible IaCL DNA (lacking the phosphodiester linkage) were efficiently repaired by hAGT. Repair of the model IaCL DNA occurred more efficiently in comparison to similar ICL DNA; 2) O^6-dG-alkylene-O^6-dG IaCL DNA containing the intradimer phosphodiester were moderately repaired by hAGT. Efficiency of the hAGT-mediated repair was contingent on the presence of the intradimer phosphate, which suggest conformational flexibility may be a requirement for repair by AGTs; 3) Flexible O^4-dT-alkylene-O^4-dT IaCL DNA evaded repair from all AGTs tested, whereas the flexible IaCL 5'-O^4-dT-alkylene-O^6-dG were efficiently repaired by hAGT. Interestingly, the 5'-O^6-dG-alkylene-O^4-dT was not proficiently repaired by hAGT supporting the importance of the 3'-phosphate group of the target dG nucleotide. 4) Flexible IaCL can be employed to generate DNA-protein cross-links (DPCs), with good conversions, as observed with repair of O^6-dG-alkylene-O^6-dG and 5'-O^4-dT-alkylene-O^6-dG by

hAGT. The use of such cross-linking experiments may be useful for elucidating substrate discrimination across AGTs by X-ray crystallography.

Translesion synthesis (TLS) may be activated by the cell as a coping mechanism when DNA damage evades repair or remains otherwise irreparable by repair mechanisms. Human DNA polymerase η (hPol η) is a key TLS Pol involved in the bypass of certain UV-induced DNA damage, and lesions formed by platinum-containing chemotherapeutics. Bypass experiments were conducted to determine if conformational freedom of the lesion impacted hPol η processivity. Towards this end, bicyclic pyrimidines linking the *C*5-atom to the O^4-atom, by an ethylene or a propylene bridge, were synthesized as conformationally locked mimics of the biologically relevant DNA damage O^4-methyl thymidine (O^4-MedT) and O^4-ethyl thymidine (O^4-EtdT), respectively. Bypass studies revealed that: 1) The conformationally locked pyrimidyl analogues described above were bypassed by hPol η with different profiles, relative to O^4-MedT and O^4-EtdT. All thymidinyl modifications evoked an error-prone behavior from hPol η, with insertion of dGMP being incorporated most-frequently in the growing strand. 2) IaCL bypass profiles of O^6-dG-alkylene-O^6-dG containing the intradimer phosphodiester group behaved significantly different relative to those IaCL lacking it. hPol η inserted the correct nucleotide (dCMP) across the 3'-dG residue for all IaCL studied, whereas an error-prone behavior was observed across the 5'-dG residue. While the lack of the intradimer phosphodiester caused frameshift adduct formation across the 5'-dG, hPol η inserted the incorrect dTMP across the 5'-dG of the canonical IaCL DNA. More studies are required to elucidate whether this dependence is shared for other types of lesions.

Acknowledgments

For everyone that has helped me along the way, I thank you. I would like to begin by thanking my wife Meena Kathiresan for her undivided support. Meena, you have been an inspiration to me to become better on both a professional and personal level. It was with you that I found my true passion for science. I want to sincerely thank you for being in my life and having spent the better part of my graduate degree with me. You have motivated me in ways that words cannot describe. You also kicked me into gear for writing this document! I want to acknowledge my future children, whom I know will be an immense inspiration in my life. I cannot wait to meet you and start our very-own family. My deep appreciation go out to my mother Dominique Allaire and father Timothy O'Flaherty that have supported me throughout my entire studies, for as long as I can remember. You are the ones that made this all possible. I'll always cherish our movie nights and our family diners that brought so much joy and laughter to my life. I want to sincerely thank my brother Shawn O'Flaherty for his support and constant input. He has been an inspiration for me to strive for the best.

I am immensely grateful to Dr. Christopher Wilds for his teachings and most-of-all his mentorship. I have interacted with Dr. Wilds since my first year of undergraduate studies at Concordia University as he was my supervisor for a Co-op work term during the summer of 2008. Dr. Wilds was the first person to teach me how to conduct organic chemistry in a research laboratory. Dr. Wilds' had my best interest at heart and I have never been let down by him. Thinking of it at the moment, I cannot imagine how many reference letters Dr. Wilds had to write for me over the years; I count more than forty. He has guided me in so many ways and has been an excellent research supervisor. It has been an honor to work in Dr. Wilds' research group and I will never forget it.

My sincerest gratitude is extended to Dr. Martin Egli and Dr. F. Peter Guengerich for allowing me to conduct research in their laboratory at Vanderbilt University. Both visits were an eye-opening experience that I truly cherish. Thank you to Dr. Amritraj Patra, Dr. Pradeep Pallan, Dr. Yan Su, Dr. Qianqian Zhang, Dr. Ewa Kowal and Dr. Li Lei for all of your help during my stay. Many thanks to Ann Caldwell for lodging and helping me out on a daily basis.

I would like to thank the current and past members of the lab, first and foremost Dr. Anne Noronha. Dr. Noronha was the first person I met from the lab, she actually introduced me to the lab premises in 2008. Thank you for having the patience to constantly answer my questions in the lab, to listen when I had a spontaneous idea for a project and to review my work and presentations. Thank you for taking the time to explain and teach me solid-phase synthesis. I know I'll always double (better still, triple) check all reagents bottle amounts, waste levels, argon pressure, amidite bottles, insoluble materials in the bottles, sequences, DNA cycle, coupling times, sequences, DNA cycle, sequences, DMT-off and CPG columns. Your meticulous approach to solid phase is what got me to where I am today. I want to extend many thanks to Dr. Francis McManus for teaching me the biochemistry, despite my chemistry background. Protein work initially intimidated me. I also really enjoyed our project brainstorming sessions, which really helped with the creativity aspect of my research. I would like to thank Dr. Gang Sun who shared a fumehood beside mine. You have always helped me out when I was dealing with chemistry issues. I would like to thank past lab members Jack Cheong and David de Bellefeuille for all the support during the early stages of my graduate studies. Special thanks go out to more recent lab colleagues William Copp, Lauralicia Sacre, Gianna Di Censo, and Chris Liczner for constant support and an enjoyable atmosphere in the lab.

I want to extend appreciation for my committee members Dr. Sébastien Robidoux and Dr. Pat Forgione for their guidance. Further thanks to Dr. Robidoux for teaching me NMR. I would like to express gratitude to Dr. Joanne Turnbull for allocating space in her laboratory for radioactivity research and her helpful discussions. Thank you Dr. Alexey Denisov for helpful (NMR) discussions over the years. Thank you to Alain Tessier, Jean-Pierre Falgueyret, and Marcos Di Falco for MS analysis of small molecules and oligonucleotides.

I would also like to extend thanks to members of other labs, past and present: Marc Ouellet, Andrew Barber, Dirk Ortgies, Mohammad Sharif, Stéphane Sévigny, Amanda Gabriel, Dylan McLaughlin and Harrison Saulnier. A quick shout-out to my little companions Ruby and Saffire. Little Saffy was always there to cheer me up and present throughout this writing experience.

Many thanks to Félix Lefebvre, Victor Brochu, Ronica Plouffe, Audrey Murray Lefebvre, Martin Pauzé, and Alexandre Laurin for being ever so supportive over the many years. I will never forget how awesome our times together have been!

I am thankful to the funding agencies that have supported me over the graduate degree, including; NSERC, FQRNT, MESRST, GRASP, Bionano (CREATE), Hydro-Québec and the Concordia University graduate awards program.

Lastly, I would like to dedicate my thesis to my grandmother Edith O'Flaherty (*Nana*). She is an incredible person and such a sweetheart, and has supported me in everything I wanted to pursue. Nana took such good care of all our family. Your perseverance is inspiring.

"A failure is not always a mistake. It may simply be the best one can do under the circumstances. The real mistake is to stop trying."
B. F. Skinner

Table of contents

List of Figures

List of Schemes

List of Tables

List of Abbreviations

MeCN	- Acetonitrile
AGT	- O^6-alkylguanine-DNA alkyltransferase
Alloc	- Allyloxy carbonyl
Alloc-OBt	- Allyl 1-benzotriazolyl carbonate
AMBER	- Assisted Model Building with Energy Refinement
AP	- Apurinic/Apyrimidinic
BER	- Base excision repair
CD	- Circular dichroism
CHO	- Chinese hamster ovary
CPD(s)	- Cyclobutane-pyrimidine dimer(s)
CPG	- Controlled pore glass
dA	- 2′-Deoxyadenosine
DBU	- 1,8- Diazabicyclo[5.4.0]undec-7-ene
dC	- 2′-Deoxycytidine
DCM	- Dichloromethane
dG	- 2′-Deoxyguanosine
DIAD	- Diisopropylazodicarboxylate
DIPEA	- Diisopropylethylamine
DMAP	- *N*,*N*-Dimethylaminopyridine
DMF	- *N*,*N*-Dimethylformamide
DMTr	- 4,4′-Dimethoxytrityl
DNA	- 2′-Deoxyribonucleic acid
dNMP	- 2′-Deoxyribonucleoside monophosphate
dNTP	- 2′-Deoxyribonucleoside triphosphate
DPC	- DNA-protein cross-link
DR	- Direct Repair
dsDNA	- Double-stranded 2′-deoxyribonucleic acid
DSB	- Double strand break
dT	- Thymidine
dU	- 2′-Deoxyuridine
E. coli	- *Escherichia coli*
EDTA	- Ethylenediaminetetraacetic acid
FA	- Fanconi anemia
Fapy	- Formamido pyrimidine
hAGT	- Human O^6-alkylguanine-DNA alkyltransferase
HPLC	- High performance liquid chromatography
hPol η	- Human DNA polymerase η
hPol κ	- Human DNA polymerase κ
hPol ι	- Human DNA polymerase ι
HR	- Homologous recombination
IaCL(s)	- Intrastrand cross-link(s)
ICL(s)	- Interstrand cross-link(s)
IPTG	- Isopropyl β-D-thiogalactopyranoside
IR	- Infrared

LB	- Luria broth
Lev	- Levulinyl
MMR	- Mismatch repair
MMS	- Methyl methanesulfonate
MNNG	- *N*-methyl-*N*'-nitro-*N*-nitrosoguanidine
MNU	- *N*-methyl-*N*'-nitrosourea
MP	- Median product
MS	- Mass spectrometry
NER	- Nucleotide excision repair
NHEJ	- Nonhomologous end joining
NMR	- Nuclear magnetic resonance
O^4-MedT	- O^4-methyl-thymidine
O^4-dT-alkylene-O^4-dT	- O^4-thymidine-alkylene-O^4-thymidine
O^6-BndG	- O^6-benzyl-2'-deoxyguanosine
O^6-dG-alkylene-O^4-dT	- O^6-2′-deoxyguanosine-alkylene-O^4-thymidine
O^6-dG-alkylene-O^6-dG	- O^6-2′-deoxyguanosine-alkylene-O^6-2′-deoxyguanosine
O^6-MedG	- O^6-methyl-2′-deoxyguanosine
Pac	- Phenoxyacetyl
PAGE	- Polyacrylamide gel electrophoresis
PNK	- Polynucleotide kinase
Pol(s)	- Polymerase(s)
RNA	- Ribonucleic acid
RP	- Repair product
SAM	- *S*-adenosine methionine
SAX	- Strong anion exchange
SDS-PAGE	- Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SDS	- Sodium dodecyl sulphate
S_N2	- Nucleophilic substitution, second order
SSB	- Single strand break
ssDNA	- Single-stranded DNA
TBAF	- *n*-Tetrabutylammonium fluoride
TBDPS	- *t*-butyldiphenylsilyl
TBS	- *t*-butyldimethylsilyl
TCA	- Trichloroacetic acid
TEA	- Triethylamine
THF	- Tetrahydrofuran
TLC	- Thin layer chromatography
TLS	- Translesion DNA synthesis
T_m	- Thermal melt
UV	- Ultraviolet
yPol η	- Yeast DNA polymerase η
yPol Rev1	- Yeast DNA polymerase Rev1

Contribution of Authors

Shown below are the contributions of each the authors cited in this dissertation.

CHAPTER II

" Lesion Orientation of O^4-Alkylthymidine Influences Replication by Human DNA Polymerase η"

Patra A.: Performed crystallization and X-ray crystallography of ternary complexes between DNA, DNA polymerase η, and dNTP. Analyzed data pertaining to crystallographic structures. These results constitute **Figures 2.3** and **2.4** of the manuscript. Prepared the figures mentioned previously and other figures and tables found in the supporting information of the manuscript pertaining to crystallography. Wrote the crystallization and X-ray crystallography experimental, results, and discussion sections of the manuscript.

Su Y.: Performed DNA polymerase η protein over-expression and purification.

Egli M.: Project supervisor for Patra A. and host project supervisor for O'Flaherty D.K. at Vanderbilt University (visit abroad). Wrote sections pertaining to the X-ray crystallography experimental, results, and discussion. Reviewed and corrected the manuscript.

Guengerich F.P.: Project supervisor for Su Y., and reviewed / corrected the manuscript.

Wilds C.J.: Project supervisor for O'Flaherty D.K. at Concordia University. Responsible for reviewing and correcting the manuscript.

O'Flaherty D.K.: Performed remainder of experiments: small molecule synthesis, oligonucleotide synthesis, purified oligonucleotides; snake venom digestion analysis of oligonucleotides, circular dichroism and thermal denaturation experiments of

oligonucleotides; protein over-expression and purification, steady-state kinetic analysis of DNA polymerase η single-nucleotide incorporation products; liquid chromatography-mass spectrometry analysis of DNA polymerase η full extension reaction products. Wrote the first draft of the manuscript, excluding the portions mentioned above, assembled the manuscript and corrected the manuscript in later versions.

CHAPTER III

"Synthesis, Characterization and Repair of a Flexible O^6-2'-Deoxyguanosine-alkylene-O^6-2'-deoxyguanosine Intrastrand Cross-link"

Wilds C.J.: Project supervisor responsible for reviewing and correcting the manuscript.

O'Flaherty D.K.: Performed all experiments. Wrote the first draft of the manuscript and corrected the manuscript in later versions.

CHAPTER IV

"O^6-Alkylguanine DNA Alkyltransferase Repair Activity Towards Intrastrand Cross-Linked DNA is Influenced by the Internucleotide Linkage"

Wilds C.J.: Project supervisor responsible for reviewing and correcting the manuscript.

O'Flaherty D.K.: Performed all experiments. Wrote the first draft of the manuscript and corrected the manuscript in later versions.

CHAPTER V

"Backbone Flexibility Influences Nucleotide Incorporation by Human Translesion DNA Polymerase η Opposite Intrastrand Cross-linked DNA"

Egli M.: Host project supervisor for O'Flaherty D.K. at Vanderbilt University (visit abroad), and reviewed and corrected the manuscript.

Guengerich F.P.: Host project co-supervisor for O'Flaherty D.K. at Vanderbilt University (visit abroad), and corrected the manuscript.

Wilds C.J.: Project supervisor for O'Flaherty D.K., responsible for reviewing and correcting the manuscript.

O'Flaherty D.K.: Performed all experiments. Wrote the first draft of the manuscript and corrected the manuscript in later versions.

CHAPTER VI

"Site-specific conjugation of human O^6-Alkylguanine-DNA-Alkyltransferase to DNA strands using Intrastrand Cross-Linked DNA mimics"

Wilds C.J.: Project supervisor responsible for reviewing and correcting the manuscript.

O'Flaherty D.K.: Performed all experiments. Wrote the first draft of the manuscript and corrected the manuscript in later versions.

CHAPTER I
General Introduction

1.1 Preamble

2'-Deoxynucleic acid (DNA) is the medium of genetic information storage for all organisms. It is used by organisms to encode ribonucleic acids (RNA), which will ultimately be translated in the biosynthesis of functional proteins. Another important biological function of DNA is to serve as the template to generate daughter DNA strands, a process known as DNA replication carried out by DNA polymerase enzymes. The preservation of the chemical integrity of this macromolecule is paramount and any chemical alteration may lead to detrimental events felt at the DNA replication and transcription levels. These insults can consequently be translated to the protein level and disrupt cellular processes.

The chemical structure of the DNA macromolecule is composed of four smaller subunits known as nucleotides. Each nucleotide is in turn composed of a phosphate group, a 2'-deoxyribose sugar and a nucleobase. 2'-deoxyadenosine (dA) and 2'-deoxyguanosine (dG), also known as purines, pair with their pyrimidine partners, thymidine (dT) and 2'-deoxycytidine (dC) respectively. Generally, two complementary DNA polymers assemble into a double-helix where the strands are anti-parallel to one another.

DNA is susceptible to chemical alterations by endogenous and / or exogenous agents. Activity of proteins towards the DNA macromolecule may be impaired by the altered DNA structure, consequently resulting in a variety of undesirable cellular consequences such as mutations, chromosomal instability, teratogenic effects and cellular apoptosis. These consequences are dependent on a multitude of parameters such as the chemical composition of

the lesion, the location of the lesion on the DNA scaffold, and the protein interacting with the DNA, to name a few.

Certain cancer chemotherapeutic regimens target cellular DNA in order to eradicate the malignant neoplasm. Of these, the DNA alkylating agents functionalize one or multiple sites on the DNA duplex, disrupting vital processes involved in DNA replication. Small appendages such as methyl groups transferred to DNA are sufficient to result in significant toxicity. Bifunctional alkylating drugs, which by definition have two reaction centers, have the added facet of generating covalent linkages between nucleotides on the same strand (intrastrand cross-link) and to the complementary strand (interstrand cross-link).

Organisms have evolved a range of different mechanisms to cope with the constant formation of DNA insults in the cell. These repair pathways have the objective to eliminate or bypass the DNA damage, or have an intricate role in activating other cellular mechanisms such as apoptosis. The activation of a given repair pathway is dependent on the type of DNA damage and together these cellular repair mechanisms have been shown to reverse a wide variety of DNA damage. Conversely, cellular repair mechanisms have been associated with resistance to certain therapeutic drugs which act directly on DNA, such as alkylating agents of the anti-cancer armamentarium. This dichotomy highlights the importance of studying repair pathways involved in cell survival for both healthy and cancerous cells.

1.2 DNA Alkylation Damage

1.2.1 Sources of DNA chemical modifications

Estimations of cellular DNA modifications can be on the order of one million events per day per cell.[1] DNA damage can be defined as a DNA chemical modification, harmful to the organism, which can generally arise from either radiation or chemical sources. For instance,

chemical moieties in the DNA can undergo excitation when exposed to specific radiation, and in the process, can react with another chemical moiety in proximity such as (other) DNA nucleotides and proteins to name a few. Common examples include cyclobutane pyrimidine dimers (CPDs) and pyrimidine (6-4) pyrimidone photoproducts (**Figure 1.1**) which are produced by ultraviolet (UV) radiation. This is particularly important given that solar UV light contains the appropriate wavelength range to induce such DNA lesions. High-frequency radiation such as X-rays and γ-rays are capable of causing ionization, thus producing a free-radical on the DNA scaffold.[2–4] The production of such radicals can lead to the formation of various DNA modifications such as DNA strand breaks, nucleobase covalent attachment and oxidation products. Similar lesions can be produced from metabolism by-products as these can also generate free-radicals.[5]

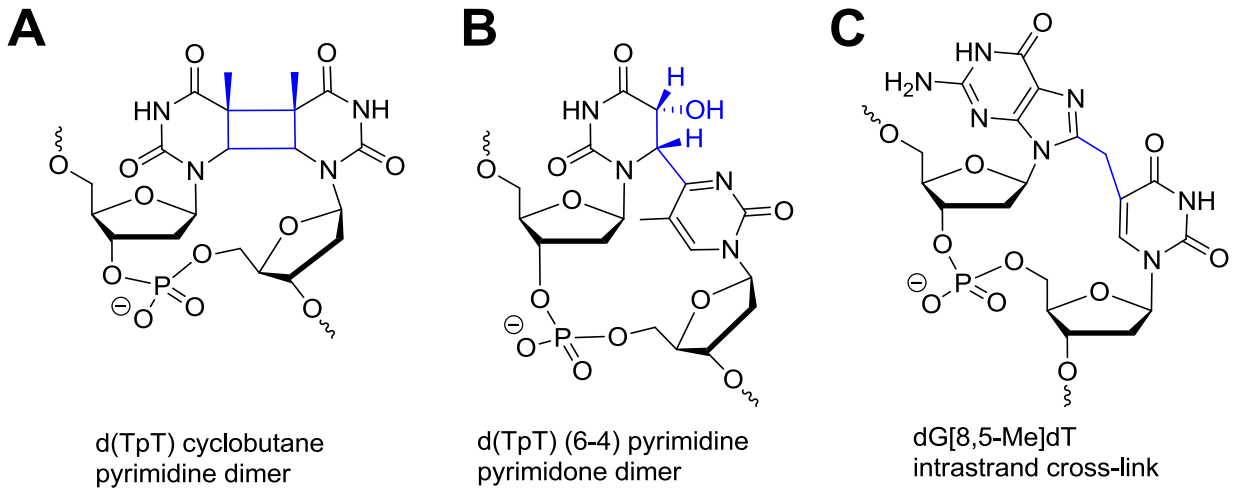


Figure 1.1: Selected example of DNA modifications from (**A**, **B**) UV radiation and (**C**) ionizing radiation. These were selected on the basis of the probes synthesized herein. Modifications are shown in blue.

The thesis is centered on alkylation of the DNA, which constitutes the installment of one or more alkyl appendages on the DNA strand. For the sake of this document, the term "alkylation" will encompass well-defined coordination of a metal to DNA, such as cross-link modifications

introduced by the chemotherapeutic agent cisplatin (cis-diamminedichloroplatinum(II)). Numerous cases of other chemical modifications have been documented,[6] such as spontaneous hydrolysis of the nucleobase or oxidation adducts, yet only prevalent types and those relevant to the chapters below will be discussed in this document.

One of the most common types of DNA modification is depurination and depyrimidation, with the former occurring more frequently compared to the latter.[7] These lesions are estimated on the order of 10^4 events per cell per day. The adverse effects of such a lesion are not surprising given that DNA nucleobases are utilized for encoding biological information (*e.g.* protein recognition and cellular instructions). However, the rate of nucleobase hydrolysis is quite slow in physiological conditions ($t_{1/2}$ ~ 730 years and $t_{1/2}$ ~ 14700 years for purines and pyrimidines respectively) but can be accelerated by a number of factors such as DNA alkylation.[8–10] Another frequent type of modification results from oxidation damage events, such as those that produce 8-oxo-7,8-dihydroguanine (**Figure 1.2**). These lesions are a result of aerobic respiration generating reactive oxygen species (ROS) which can act on DNA. Like alkyl groups, the introduction of oxygen atoms on the DNA chemical entity changes the steric and electronic properties of the DNA, ultimately disrupting vital processes like DNA replication and transcription.

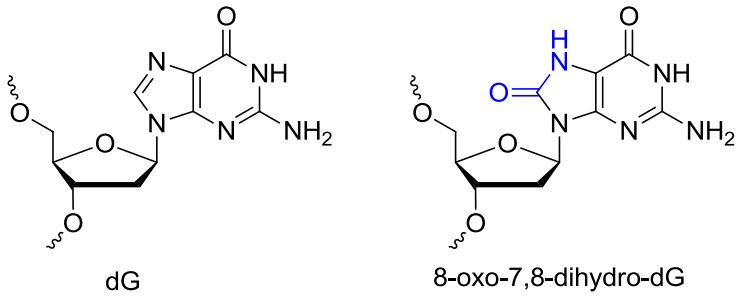


Figure 1.2: Chemical structures of 2'-deoxyguanosine (dG; left structure) and 8-oxo-7,8-dihydro-dG (right structure). Modifications are shown in blue.

1.2.2 DNA Damage by Alkylation

Historically, it was at the end of World War I, that the chemical bis(2-chloroethyl)sulphide (so-called "mustard gas") was recognized as an agent capable of generating long-term systemic malfunction, in addition to immediate acute damage.[11] This would later be attributed, at least in part, to its ability to act on the genetic material DNA.[12,13] Since then, a number of environmental agents both produced endogenously and / or exogenously have been discovered that assault the chemical identity of DNA. For instance, *S*-adenosine methionine (SAM) is produced endogenously and utilized as a cosubstrate for many methyl group transfer reactions. *In vitro* experiments have shown the methylation of DNA when exposed to SAM, and it has been estimated that the SAM produces over 4500 DNA lesions in a 24h time period per cell.[14] Lipid peroxidation metabolites are also a source of endogenous alkylating agents.[15,16] On the other hand, exogenous agents such as *N*-nitroso-*N*-methylurea (NMU) and methyl methanesulfonate (MMS) have been investigated for their ability to transfer an alkyl group to DNA.[17] Interestingly, agents targeting DNA displayed potent anti-mitotic (cell division) properties, a valuable attribute for treating groups of malignant cells. Yet, the elevated systemic toxicity prevented any viable use of such compounds in the clinic, and consequently prompted research on the development of optimized compounds that would retain anti-mitotic potency, while having reduced toxicity. Among the first chemotherapeutic drugs were chlorambucil and busulfan (**Figure 1.3**).[11]

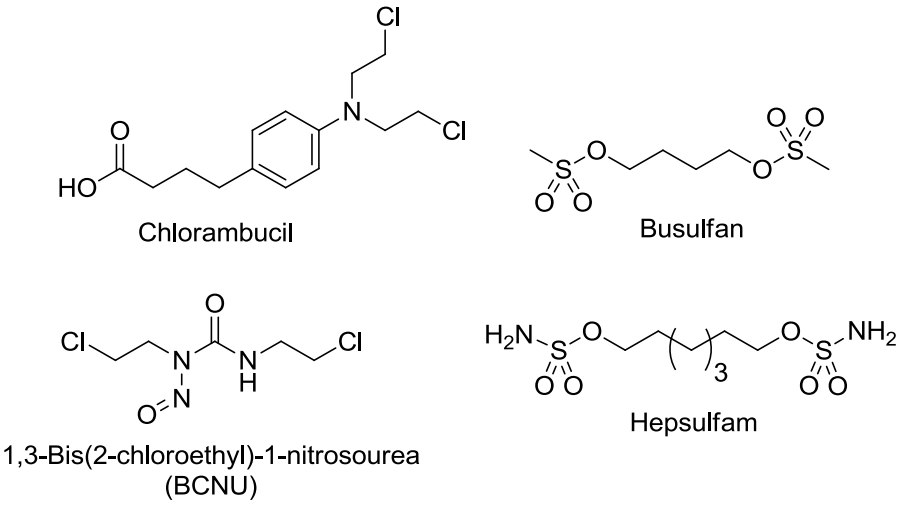


Figure 1.3: Selected alkylating agents. Busulfan and hepsulfam are particulaly relevant to the thesis.

Compounds capable of causing DNA lesions impede important cellular processes such as DNA replication and transcription. It should be noted that alkylating agents are capable of causing harmful damage to other biomolecules like RNA and proteins, however these topics exceed the scope of this thesis. Only lesions directly involving DNA will be discussed. Alkylating agents can be categorized further by the number of reactive moiety inherent to the compound. Monofunctional alkylating agents have one reactive center, whereas bifunctional alkylating agents have two reactive centers. The former class is limited to producing mono-adducts on the DNA scaffold (see **Scheme 1.1** for mono-alkylation by nitrosourea compounds). These added appendages can disrupt the physical and chemical properties important to DNA structure and function.[18] These include disruption of base stacking interactions, base pairing properties and the overall secondary DNA structure. For example, the O^6-methyl-2'-deoxyguanosine (O^6-MedG) and O^4-methyl-thymidine (O^4-MedT) lesions change the tautomeric state of the nucleobase (**Figure 1.4**), hence causing discrepancies in base pairing preferences.[19,20] This is particularly problematic given that DNA polymerases (Pol) generally incorporate the incorrect nucleotide in the newly synthesized strand, across O^6-MedG or O^4-MedT, during DNA

replication.[21,22] The presence of a methyl group, albeit small in size, can also severely hinder the DNA Pol kinetics (*i.e.* DNA Pol blockage).

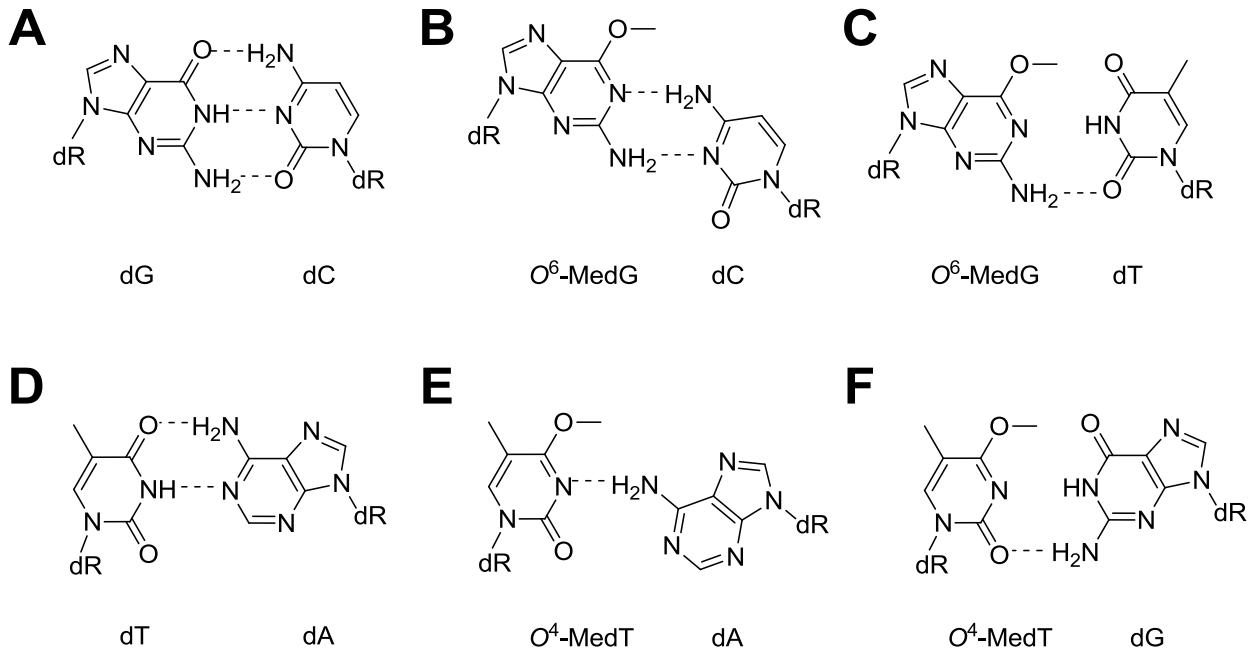
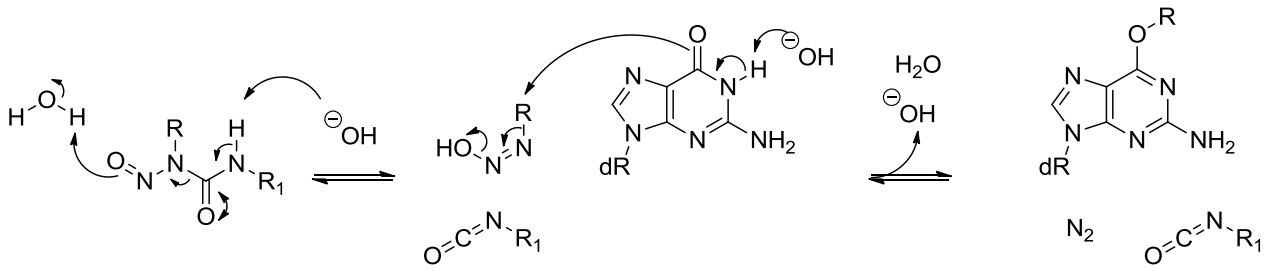


Figure 1.4: Base pairing discrepancies observed for O^6-MedG (**B** and **C**) and O^4-MedT (**E** and **F**) relative to dG (**A**) and dT (**D**), respectively. To be noted is the wobble type base pairing present with the alkylated nucleobase and its correct base pairing partner (**B** and **E**), which is not observed for its mismatch (O^6-MedG•dT in panel **C** and O^4-MedT•dG in panel (**F**).It should also be noted the tautomeric differences between O^6-MedG and dG, as well as O^4-MedT and dT.[23–26]



Scheme 1.1: Mono-adduct formation by nitrosourea compounds (scheme adapted from the following reference).[27] In this example the O^6-position of dG is alkylated. Many chemotherapeutics utilize the nitrosourea moiety (1 3-bis(2-chloroethyl)-1-nitrosourea (BCNU), 1-(2-chloroethyl)-3-cyclohexyl-1-nitrosourea (CCNU), and 1-(2-chloroethyl)-3-(4-amino-2-methylpyrimidin-5-yl)methyl)-1-nitrosourea (ACNU), for instance).[27]

In addition to mono-alkylation, the bifunctional alkylating agents are capable of generating more-pronounced types of DNA damage such as DNA cross-links.[28,29] These include intrastrand cross-links (IaCL) which attaches two sites on the same DNA strand, interstrand cross-links (ICL) where alkylation events occur on complementary strand, and DNA-protein cross-links (DPC) which attaches a protein molecule to the DNA strand (**Figure 1.5**). The heightened potency of bifunctional compounds is largely attributed to these types of lesions and will be discussed in further detail below.

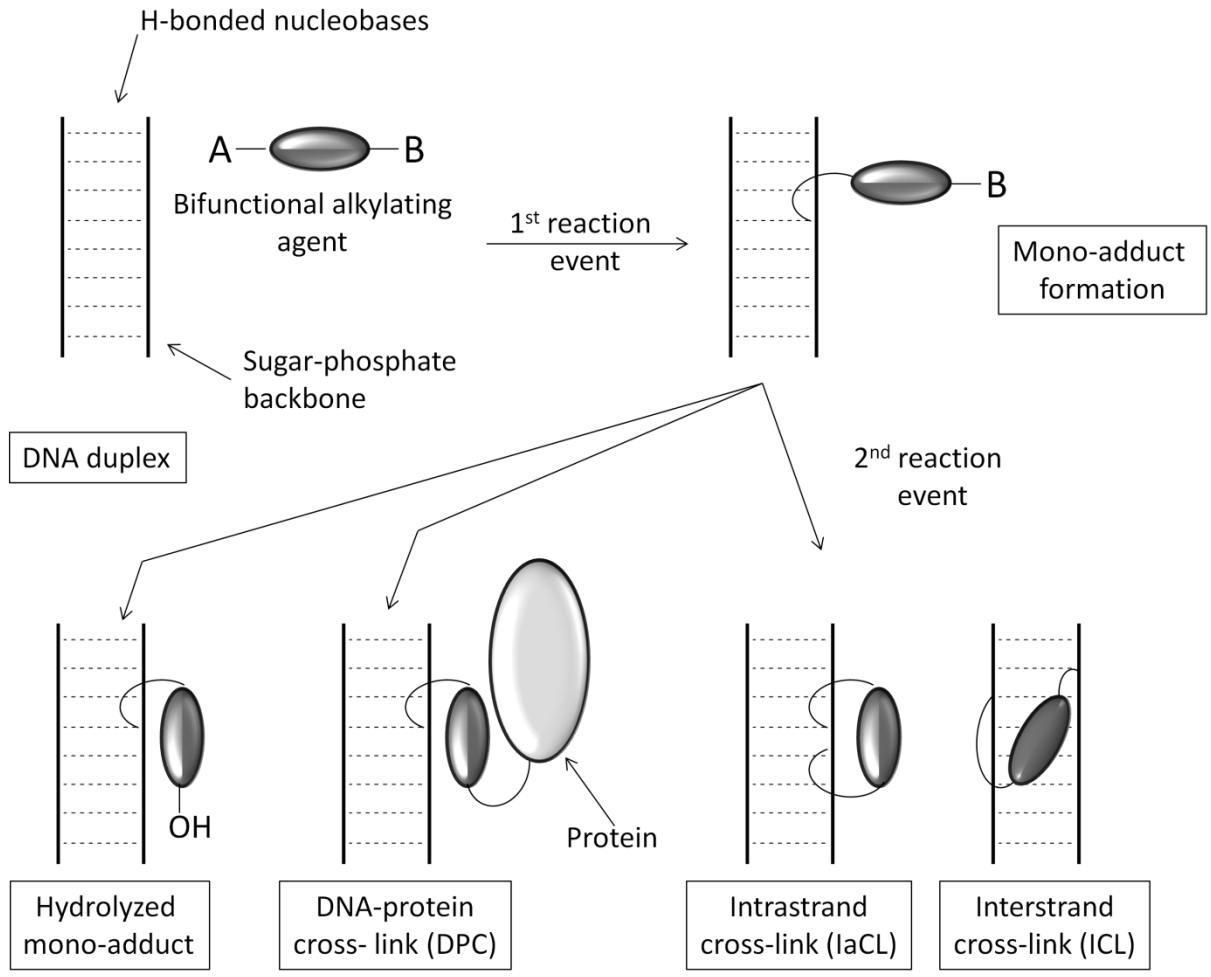


Figure 1.5: Selected reaction fates for bifunctional alkylating agents.[30] Note that the first reaction event is shown only with DNA, whereas the DNA protein cross-link could have been generated by a first reaction with the protein followed by a second with DNA. The same could be said for the hydrolyzed mono-adduct. The scheme was adapted from the following reference.[30]

1.2.3 Adverse Effects of O^4-Alkyl-dT and O^6-Alkyl-dG Mono-Adducts and Various Cross-Links

Depending upon their chemical structure, DNA lesions may lead to a number of adverse cellular effects.[31,32] The DNA damage may significantly slow DNA replication and / or transcription by interfering with the activity of polymerases. Downstream effects could lead to slowing of cell division and chromosomal instability. If capable of bypassing the damaged site, DNA polymerases could cause nucleotide misinsertions in the newly synthesized (daughter) DNA strand, consequently generating mutagenesis.[31] Certain DNA adducts are chemically unstable, which upon degradation increases the probability of aberrant global DNA lesions such as DNA strand breaks.[31] The understanding of DNA damage and its biological consequences has greatly advanced from tracing the chemical structure of a DNA adduct to the critical cellular event it causes.

With the exception of intercalators, DNA alkylators generally possess electrophilic centers. Nucleophilic sites found within the DNA can thus attack the electrophiles following the course of a single replacement reaction (substitution). Research using benchmark DNA alkylators have revealed that the distribution of alkylation sites on the DNA molecule depends on a variety of factors such as the nature of the nucleophile and electrophile.[33,34] *In vitro* and *in vivo* studies demonstrated that electrophilic alkylators, such as NMU and MMS, largely target the N7-atom of dG. This atom is the most susceptible site of alkylation given it has the highest negative electrostatic potential (hence being most nucleophilic) compared to other atoms found within the DNA nucleobases.[35] N7-alkyl-dG lesions larger than an ethyl group are detrimental to cellular activity.[8] Interestingly, N7-methyl-dG and N7-ethyl-dG are well-tolerated by cells, perhaps because they are small and do not interfere with the H-bonding face of the nucleobase.[8,32,36] Yet,

these adducts are prone to chemical degradation and can lead to various products harmful to the cell, such as abasic sites and formamidopyrimidine (Fapy) lesions.[8] An aspect to be highlighted is that N7-alkyl-dG lesions are particularly difficult to investigate in biochemical assays due to their tendency to degrade.

Potency of a given alkylating agent is governed by more than the number of lesions generated on the DNA per dose. Rather, the cytotoxicity heavily depends on the types of lesions that amass on the DNA (*e.g.* sites of alkylation), and the persistence of such DNA adducts. Lesions such as O^6-alkyl-dG and O^4-alkyl-dT are produced in relatively low levels, with the former occurring more frequently compared to the latter. For instance, subjecting DNA to NMU (*in vitro*) generates 6.3 % of O^6-MedG compared to only 0.4 % of O^4-MedT.[33] Albeit being minor sites of DNA alkylation, these types of damage are chemically stable, and if left unattended disrupt cellular activity. It has been estimated that approximately 6600 O^6-MedG lesions are sufficient to cause cell death in repair-deficient mammalian tumor cell lines.[37] Moreover, correlations have linked both O^6-MedG and O^4-MedT to carcinogenesis.[20,38–40] The O^4-MedT lesions, on the other hand, are inefficiently processed by mammalian cells compared to O^6-MedG. As a result, this lesion is considered more cytotoxic when present.[41]

Harmful effects of O^6-alkyl-dG and O^4-alkyl-dT are largely a results of DNA Pol interference during DNA replication, in both prokaryotes and eukaryotes. For instance, O^6-MedG is capable of blocking a number of DNA Pol such as eukaryote DNA Pol α,[42] β,[43] δ,[44,45] *Escherichia coli* (*E. coli*) Klenow fragment,[46] and T7 bacteriophage DNA polymerase.[42] Even specialized DNA polymerases from the translesion synthesis (TLS) pathway (human DNA Pol η, κ, ι, for instance) also suffer from enzymatic rate reduction when faced with O^6-MedG.[45] The latter DNA Pol, particularly Pol η, will be discussed in greater detail below. Likewise, O^4-MedT

stalls a variety of different DNA polymerases (both standard and TLS Pol).[22,47] Perhaps more-importantly, both of these alkylation sites cause significant misinsertions by DNA polymerase if acted upon.[48] This mutagenesis can have profound effects on the organism given that newly synthesized genetic information may be compromised, thus enabling deleterious events at any point downstream.

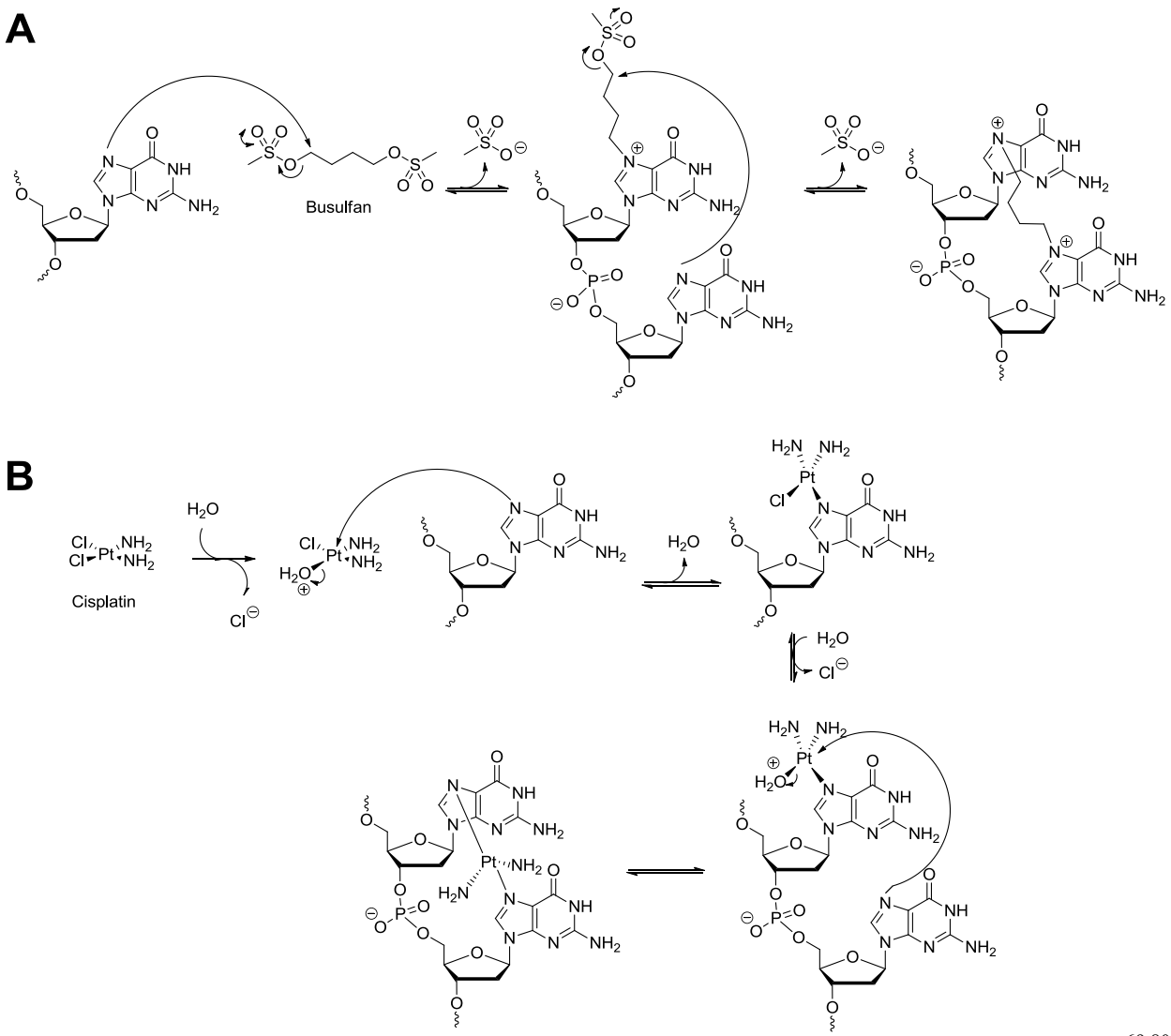
Organisms have, fortunately, evolved a wide variety of mechanisms to cope with the constant attack on DNA. Yet, a somewhat paradoxical event is triggered upon activation of the mismatch repair (MMR) pathway when faced with O^6-MedG or O^4-MedT.[49–52] The MMR machinery is capable of detecting mismatches incorporated into the primer strand (*i.e.* the one being grown). Upon excision of a DNA fragment containing the mismatch, DNA polymerases are capable of re-synthesis enabling the installment of the correct sequence identity. In the case where the lesion is found in the primer strand, the MMR pathway is capable of correcting the damage. In contrast, a lesion-containing template strand (*i.e.* the one being read by the replication machinery) will be detected and efforts to repair the DNA damage are futile. MMR recognizes the damage but its attempts at repair are only focused on the primer strand, allowing the lesion to remain unaffected in the template strand. This process is known as a futile excision-resynthesis cycle, where the MMR is repeatedly activated, which will ultimately lead to cellular lethality.[53]

Thus far, the biological consequence of mono-adducts found at the O^6-atom and O^4-atom of dG and dT, respectively, has been overviewed. However, bifunctional alkylating agents have the capability of extending the DNA damage profile, from simple mono-adducts, to an assortment of cross-links (ICL, IaCL and DPC illustrated in **Figure 1.5**). Historically, much focus has been bestowed to ICL lesions given their exceptional potency.[54] It is estimated that

only 40 ICL lesions are required to cause cell death in repair deficient mammalian cells, and single ICL lesion is required to kill bacterial cell.[55,56] This type of damage is exerted from compromising both DNA strands, and preventing DNA strand separation required for DNA replication and transcription, hence severely stalling proper cellular activity.[57] The introduction of ICL in the DNA duplex can have profound, or minimal, effects on the global structure of the duplex. Higher levels of structure distortion normally correlate with higher levels of detection and repair.[57–59] ICL are generated from a variety of agents including psoralens, mitomycin C, platinum compounds, nitrogen mustards and nitrosoureas, many of which have found use in the clinic.[57] ICL are also generated by furocoumarins, which are produced naturally in certain plants.[60] More recently, it was discovered that the presence of abasic sites in DNA can lead to the production of ICL lesions.[61–63] Part of the reason why ICL have been extensively investigated is due to their poorly understood cellular repair.[54,57] ICL DNA lesions are not repaired by only one pathway, but rather, the cell makes use of a combination of several repair mechanisms to restore the damaged DNA. Pathways such as nucleotide excision repair (NER), TLS, homologous recombination (HR), MMR, and Fanconi anemia (FA)-proteins have all been associated with ICL repair, which partly accounts for the high toxicity of ICL.[54] Our group revealed the possible involvement of the direct repair pathway in the repair of ICL DNA linking the O^6-atoms of two 2'-deoxyguanosine residues, either in a directly opposing fashion or in a GNC motif.[64,65]

In comparison, the genetic information of only one strand is compromised for IaCL-containing DNA and certain of these lesions, like UV-induced CPDs, are readily removed by the NER machinery.[66–68] IaCL lesions are produced from bifunctional agents, much like ICL lesions, but with the connectivity within one DNA strand (**Figure 1.5**). These are produced from cross-linking agents like alkyl sulphones,[69] mitomycin C,[70] platinum compounds,[71] and nitrogen

mustards.[72] As shown in **Scheme 1.1**, the nitrosourea can form DNA mono-adducts *via* a reactive 1-hydroxy-2-alkyldiazene intermediate. If the alkyl(ene)-modification contains a second good leaving group (terminal chloro group, for instance), another nucleophile may attack the mono-adduct generating a cross-link (IaCL, for instance). The mechanism is slightly different in the case of the alkyl sulphone busulfan (**Scheme 1.2A**) or cisplatin (**Scheme 1.2B**). The N7-position of dG is the preferred site of alkylation in these examples and, as mentioned above, the N7-alkylene-dG appendages are short lived ($t_{1/2}$ = 1h for N7-dG alkylene ICLs).[73] In contrast, the Pt-adduct are not prone to depurination.[54,74] In addition, IaCL are also generated in DNA from oxidation,[4] UV radiation,[75,76] and ionizing-radiation,[77–79] all of which are important and persistent environmental factors.



Scheme 1.2: Proposed mechanisms of IaCL production from (**A**) busulfan and (**B**) cisplatin.[69,80] Note that other mechanisms are possible and these constitute potential pathways to generate IaCL DNA.

Examples of IaCL have demonstrated *in vivo* toxicity in repair-deficient cells.[81,82] Although NER is thought to readily remove most IaCL lesions,[67] examples of IaCL adducts (certain cisplatin-induced, for instance) have shown to be shielded from the NER machinery by eliciting proteins containing high-mobility group (HMG) domains, both *in vitro* and *in vivo*.[83–85] NER displayed little to no capacity of repair because of binding of the protein to the damaged DNA.[86–88] As a result, the biological consequences of IaCL warrants further investigation. The focus of **CHAPTERS III-VI** revolve around the study of IaCL lesions.

The bifunctional nature of the IaCL adduct within one DNA strand, generally leads to global structural distortions in the DNA duplex. In the case of UV-induced CPD lesions, the IaCL linkage is sterically constrained and leads to significant helical axis bend of about 30^{o}.[89] Similar trends are observed with the (6-4) pyrimidine-pyrimidone photodimer (chemical structure shown in **Figure 1.1**), which drastically influences both the local and global duplex structure.[90,91] The *cis*-diammineplatinum(II) IaCL between adjacent dG residues, at their N7-atoms, causes unwinding of the DNA at the platinated site and a considerable bend of the DNA towards the major groove, due to the adduct.[92] In this example, it has been argued that the platinum(II)-mediated impact facilitates binding of proteins containing a HMG by opening up the minor groove of the duplex.[92] Molecular dynamics simulations of the oxidatively generated dG[8,5-Me]dT IaCL have suggested bends of approximately 20^{o} with respect to the helical axis.[93] More-recent, investigation has shown, however, that the dG[8,5-Me]dT IaCL may adopt a more favorable conformation, where the dT moiety is flipped out of the DNA duplex, rendering the overall structure more conical-like.[94] To be noted is the aforementioned IaCL are all composed of rigid systems (minimal flexibility inherent to the cross-link linkage). The structural distortions induced by the IaCL can drastically interfere with the biological events as bending of DNA plays a critical role in tuning protein-DNA interactions.[95] By the same token, it is these helix deformations that can alert the cellular repair response mechanisms.[59,96] One example of a compound that generates flexible DNA lesions is BBR3464 (triplatin tetranitrate), which has undergone clinical trials.[97] This drug generates long range (1,5) IaCL, in addition to (1,4) ICL, due to its two flexible alkylene linkers and proves to be very potent.[98,99] Computational approaches suggest minimal global structural perturbations induced by this adduct and may account for, in part at least, its elevated toxicity.[98] These bifunctional adducts can also interfere

with replication and transcription by blocking and inducing DNA Pol misinsertions. It may be that the bridge between the two nucleotides, residing on the same DNA strand, prevents malleability of the individual nucleotide units. This malleability may be necessary for proper activity of Pols (both kinetics and reading of the genetic information).

Another example of deleterious cross-links are DPCs.[100] DPCs are naturally occurring during biological processes and they are thought to be transient when enzymatically catalyzed. However, topoisomerases 1 and 2 (TOP1 and TOP2) are examples that can remain trapped as a DPC species.[101] In this case, the enzyme nicks the DNA phosphodiester backbone, generating the DPC (between the Tyr and the 3'-phosphate group) and a single-strand break (SSB) containing a 5'-hydroxyl. The free DNA can thus unwind for other biological process (like DNA replication), and after doing so, TOP1 ligates the SSB DNA, returning it back to its original state. Inhibition of the ligation steps, from DNA damage for instance,[102] stalls the TOP1-mediated event to a stable DPC and a SSB. Formation of DPC can also arise from non-enzymatic sources such as bifunctional alkylating agents. One very simple example is formaldehyde, which can be produced endogenously.[103] Although this compound may not normally be considered bifunctional in nature, it can undergo Schiff base formation with an initial amino group (from either the protein or DNA). This intermediate is the target of another nucleophile, such as another amino group, to produce a DPC bridged by a single methylene group. This process is so efficient that it has found applications in chemical biology to identify DNA-protein interactions.[104] As with other type of DNA damage, dG seems to be the most reactive cross-linking residue in DNA.[105] Many amino acids have been identified as part of the cross-link, with Lys and Cys being the most prevalent.[105] Chemotherapeutic agents, such a nitrogen mustards[106] and platinum-containing compounds,[107] are also capable of generating DPC adducts.

Interestingly, ionizing radiation is efficient at generating DPCs in hypoxic conditions, relative to other types of damage like double strand breaks (DSB).[108] DPCs are particularly cytotoxic because of their large size. They thus efficiently inhibit replication and transcription.[109] Depending on its site of attachment, their presence can interfere with unwinding of the DNA by stalling DNA helicase.[110,111] DPCs have been identified as strong blockers of DNA polymerases.[112–114]

1.3 Direct repair and O^6-Alkylguanine DNA Alkyltransferase

1.3.1 General repair mechanisms

There exist a plethora of types of DNA lesions, with only a selected few examples briefly discussed above. While repair processes can be sophisticated, no single repair pathway can process all the forms of DNA damage. Rather, organisms have evolved a multitude of interwoven DNA repair systems. These can deal with most of the assaults inflicted on the genetic material. The focus of the thesis is on two such repair systems. The first is direct repair (also known as direct reversal) and the second is translesion synthesis. Other repair mechanisms, such as base excision repair (BER), NER, MMR, HR, and nonhomologous end joining (NHEJ) may be highlighted briefly (mostly for comparison). These repair pathways have been reviewed elsewhere and more information concerning these may be found in the following references.[6,115–124]

1.3.2 Direct repair

Direct repair is unique since it does not involve a re-synthesis step, as observed for other repair pathways. The proteins act directly on the DNA scaffold to remove the aberrant chemical bonds of additional appendages. Other than photolyases, all known direct DNA repair proteins

are involved in the removal of DNA alkylation.[125] Photolyases, which have not been identified and are most-likely lacking in humans,[126] utilize visible light to restore UV-induced DNA damage like TT CPD.[127,128] These CPD photolyase enzymes function by clamping the DNA at the site of damage, inducing a gross conformational change. Using light of wavelength 350-450 nm, the enzyme can undergo excitation and transfers the energy to the fully reduced flavin adenine dinucleotide ($FADH^-$) cofactor. The excited co-factor in turn reacts with the CPD, ultimately causing a collapse of the CPD into its respective canonical pyrimidyl nucleobases.[127]

The AlkB protein family is another set of repair enzymes of the direct repair pathway.[129–131] These proteins make use of mononuclear Fe(II) as a cofactor, and α-ketoglutarate as a cosubstrate to repair alkyl lesions found on various DNA heteroatoms such as N1-methyl- and N1-ethyl-2'-deoxyadenosine (N1MedA and N1EtdA), etheno-2'-deoxyadenosine (εdA), and N3-methyl-2'-deoxycytidine (N3-MedC).[130] By using molecular oxygen as the second cosubstrate, alkyl lesions are oxidized to a hydroxyalkyl intermediate, which undergo spontaneous dealkylation (forming the corresponding aldehyde by-product). Interestingly, one of the sources for intracellular formaldehyde is this pathway given that formaldehyde can generate DPCs.

1.3.3 O^6-Alkylguanine DNA Alkyltransferase

O^6-alkylguanine-DNA alkyltransferases (AGTs) are suicide repair proteins that require no co-factor, or co-substrate, to remove DNA alkyl lesions at the O^6-atom of dG or the O^4-atom of dT.[132–134] This class of proteins irreversibly transfer the alkyl group from the DNA onto the thiolate of an activated Cys residue found in the active site, and in doing so inactivates the protein (**Figure 1.6**). The stoichiometric reaction proceeds *via* nucleophilic attack onto the α-carbon of the alkyl lesion. Once alkylated, the protein is rapidly degraded by the ubiquitin pathway.[135] The AGT proteins have a critical role in maintaining genomic integrity and span

over all kingdoms of life except plants.[132] All proteins have a conserved -P**C**HRV- motif (**Figure 1.7**) part of the active site (**C** denotes the position of the reactive Cys-residue).[136] They share an Asn-hinge used to the induce flippling of the damaged nucleotide into the active site. Lastly, there is absolutely conserved helix-turn-helix (HTH) DNA binding motif consisting of -RAV(A/G)- residues.

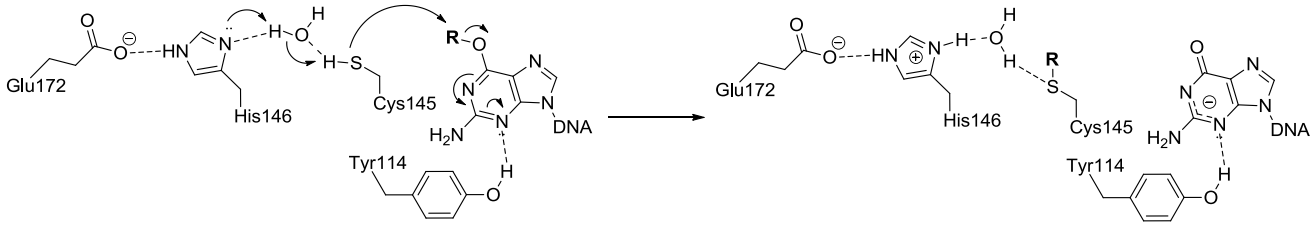


Figure 1.6: Proposed mechanism of repair by hAGT. The network is similar to the catalytic triad of serine proteases, which deprotonates the active site Cys145. Cys145 gets deprotonated by water-mediated His146, to allow for the dealkylation reaction.[137] The **R** group represents the alkyl group.

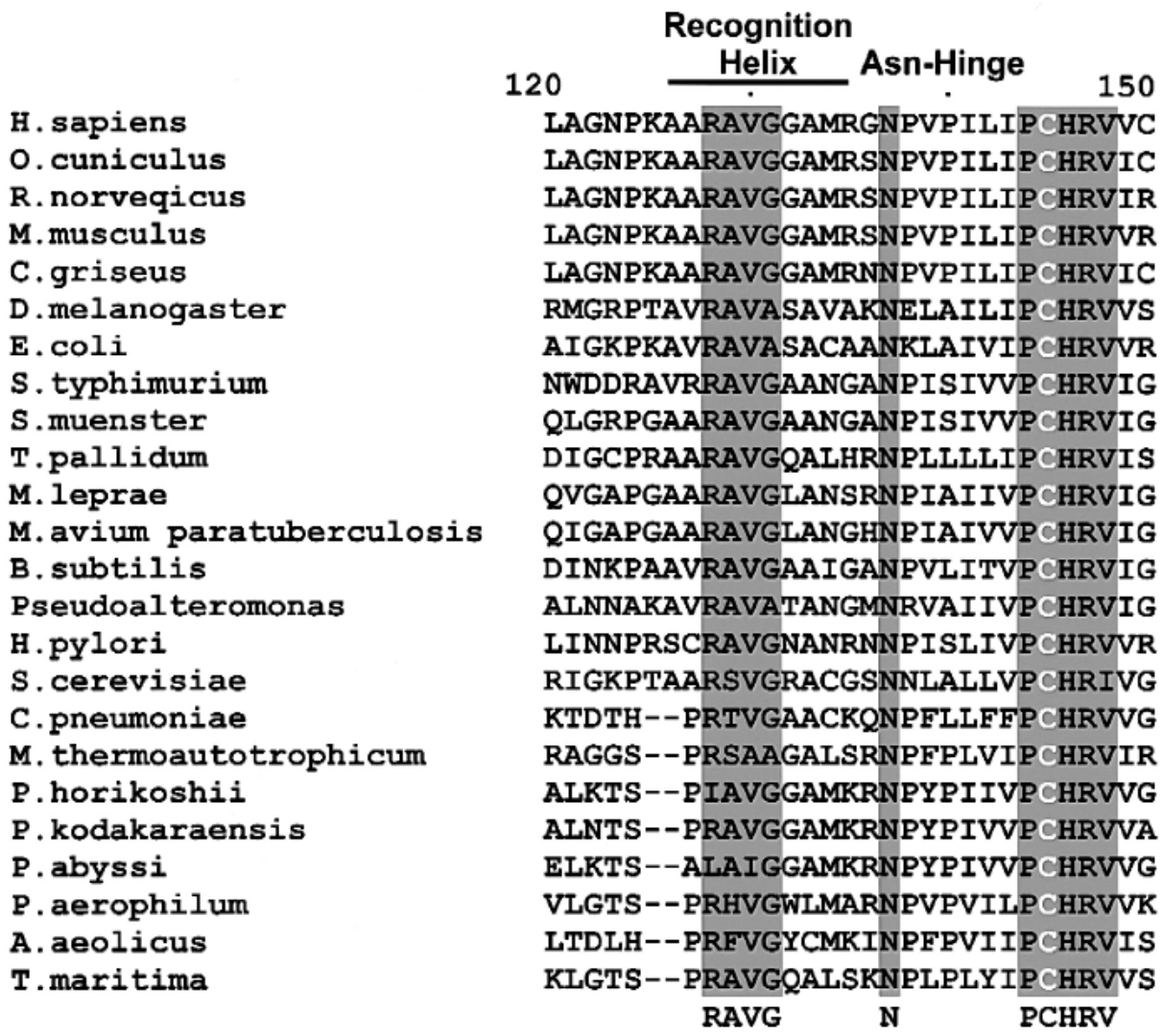


Figure 1.7: Sequence alignment of active site, asparagine hinge and recognition helix of AGT homologues (reproduced from the following reference).[136]

1.3.4 AGT homologues of interest

Different AGT homologues have evolved to act on a number of different substrates. Their substrate specificity is of interest to correlate a certain type of DNA damage to a, or many, repair pathway(s). This may eventually lead to the development of better chemotherapeutic regimens. Identifying a particular repair pathway responsible for restoring DNA damage that was induced by a specific drug can guide the next generation of drugs, or combination therapies. This has

been accomplished previously using O^6-benzylguanine as a pseudosubstrate inhibitor of human AGT (hAGT).[132,138,139]

Historically, the first AGT to be identified and characterized was the *E. coli* adaptive-reponse protein (Ada).[140,141] The Ada protein was found to protect *E. coli* from methylating agent *N*-methyl-*N*-nitroso-*N'*-nitroguanidine (MNNG)[142] and highly inducible.[143] Ada has two distinct domains responsible for removal of alkyl lesions. The first, the N-domain, can act on the phosphodiester backbone to remove the methyl group of S_p-methyl-phosphotriester and is in charge of regulating expression (adaptive response).[144,145] The C-terminal domain (also known as Ada-C) is capable of removing alkyl groups at the O^6-position of dG (Me and Et) and O^4-position of dT (Me and Et).[146–148] As of now, larger alkyl groups at these positions have shown no reaction in the presence of Ada-C.[65,146,149–151]

Later, it was discovered that *E. coli* possesses another AGT (OGT). Unlike Ada, this protein was not inducible and served as a frontline defense against DNA alkylation.[152] It accommodates a wider scope of substrates in comparison to Ada, as shown by our group.[150] OGT is capable of repairing O^4-MedT and O^6-MedG with approximately the same efficiency.[153] Interestingly, larger adducts at the O^4-atom of dT, such O^4-*n*-hydroxybutyl-dT, and O^4-*n*-hydroxyheptyl-dT, can undergo repair by OGT, whereas Ada-C and human AGT are inactive.[149,150] To date, OGT has been reported to repair a wide scope of O^6-alkylated-dG adducts such as: O^6-MedG, O^6-EtdG, O^6-benzyl-dG (O^6-BndG), O^6-*n*-propyl-dG, O^6-*n*-butyl-dG, O^6-*n*-hydroxyheptyl-dG, an O^6-heptylene-dG nucleoside cross-linked with an O^6-dG nucleotide part of an oligonucleotide.[143,150,154–156] Recent work by McManus and Wilds showed that OGT had enhanced repair capability towards O^4-alkyl-2'-deoxyuridine (O^4-alkyl-dU) compared to the dT series.[149] However, OGT was incapable of removing O^4-dT-alkylene-O^4-dT ICL lesions.[150]

The human variant, hAGT, is the most-widely characterized AGT. One reason stems from its potential participation in chemotherapeutic resistance from drugs such as the methylating-agent temozolomide[157] and chloroethylating-agents[158,159] (BCNU for instance).[132] hAGT, like OGT, is not inducible in the cell. To date, this AGT variant has the largest library of known substrates (**Figure 1.8**). It displays activity towards O^6-alkyl-dG mono-adducted analogues, whereas it is limited to only O^4-MedT for the pyrimidine series. It can repair cyclic anologues that link the O^6-atom with the N1-atom by an ethylene bridge (**eX** in **Figure 1.8**). In doing so, the product is a stable DNA-hAGT DPC. The attachement to the N1-position prevents another hAGT molecule from reacting.[160] The DPC species was later crystallized and played an integral part of the current understanding of the hAGT repair mechanism. Our group has shown that hAGT can repair an O^6-dG-alkylene-O^6-dG ICL.[64,65] In contrast, this substrate generates an DPC linked at the O^6-atom of dG after the first reaction, which in turn, is a substrate for a second repair event by another hAGT molecule. The final products are a hAGT-hAGT cross-link and the completely repaired DNA strands.

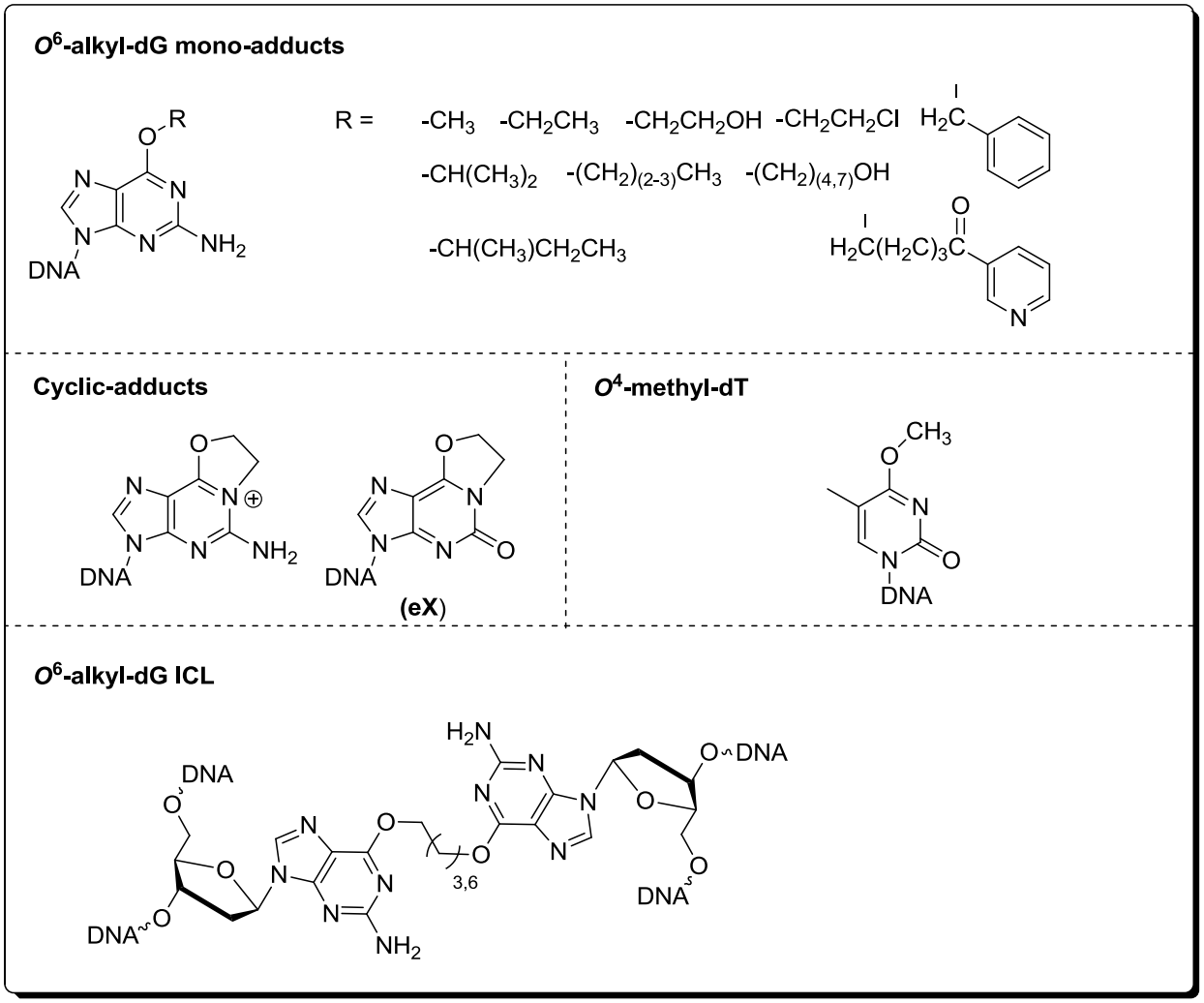


Figure 1.8: Current substrate scope of hAGT (adapted from following reference).[132]

1.3.5 AGT repair mechanism

One aspect that has remained elusive, despite structural investigation, is the exceptional rate of repair of DNA containing O^6-MedG by AGTs ($k > 10^8$ $M^{-1}s^{-1}$ for Ada-C, for instance) to overcome the energy barrier of severing the ether bond of O^6-MedG, without the help of cofactors.[161] The AGT-mediated repair mechanism was first devised using structural investigation on Ada-C, given that this protein was the first to be crystallized (manuscript published in 1994).[162] It demonstrated that the reactive Cys residue was buried into the active site and that the protein would have to undergo a conformational rearrangement upon binding to its

DNA target. In doing so the activated thiolate would be situated in close proximity to the lesion allowing for a S_N2-type attack. One challenge in using the Ada-C protein as the representative to illuminate the properties of the entire AGT family stems from the limited substrate range of Ada-C relative to OGT and hAGT. There is currently no crystal structures of *E. coli* OGT and in the year 2000 two separate studies were reported on hAGT crystal structures.[163,164] The structure of the alkylated version of hAGT provided insights on how the protein releases the repaired DNA substrate, and is subsequently destabilized to promote protein degradation (*via* ubiquitination).[164] It should be noted that crystal structures of the different AGT proteins have similarity despite having low sequence identity.[132] Taking a closer look at hAGT, the protein is divided into two major domains. The C-terminal domain hosts the active site and the HTH motif. Through biochemical analyses, the N-terminal domain fosters the structural aspects that are crucial for the repair activity of hAGT.[165] Deleting this domain from the protein removes activity. Interestingly, the N-terminal of hAGT contains a Zn binding site, which is not essential for repair to occur, but its presence increases the rate of repair.[166]

A study by Daniels *et al.* published in 2004, revealed many important aspects concerning the hAGT repair mechanism since previous crystal structures were of the native protein, or the protein (alkylated) product (*i.e.* post-repair).[137] The report described two crystal structures involving hAGT complexes with DNA containing modifications, which provided insights on the potential intermediate states of repair. The first structure was a non-covalent complex between the impotent C145S hAGT variant and DNA containing an O^6-MedG insert, and the second was a covalent DPC complex generated from the repair of DNA containing **eX** (**Figure 1.8**).[137] These structures revealed that the mode of binding of hAGT with the alkylated DNA was through the minor groove, as opposed to the major groove, *via* the HTH motif. This observation was unique

in proteins containing a HTH that bind B-form DNA.[137] Upon binding, the protein undergoes a minor conformational rearrangement, with one of the helicies from the HTH buried deep in the minor groove of the DNA (**Figure 1.9**). The binding widens the minor groove and allows for an arginine (Arg 128) to enter the DNA helix. Arg128 interacts with the base pair containing the damaged nucleotide to promote flipping of the alkylated nucleotide into the active site of the protein, where extrahelical repair may take place. Nucleotide flipping is aided by a tyrosine (Tyr114) residue which interacts with the 3'-phosphate of the damaged nucleotide. hAGT variants with this residue mutated to a phenylalanine (Y114F) or alanine (Y114A) have shown a reduction in repair rates by 30- and 600-fold, respectively.[167] Computational studies have suggested that the Tyr114 interaction may have electrostatic character,[168] as opposed to solely a steric contribution hypothesized by Daniels *et al*.[137] Once flipped into the active stie of the protein, the extrahelical base is oriented inside the active site by a number of amino acids (shown **Figure 1.9C** and **D**). The lowering of the C145 pKa, and thus increase in reactivity, is attributed to the catalytic triad (Glu-His-water-Cys) hydrogen bond network.[136,137,164]

Understanding the mechanism of detection continues to be of interest for AGT. AGTs are capable of binding both single stranded and double stranded DNA, in a cooperative fashion.[169] This may serve as a type of recruitment mechanism to scan for specific regions of DNA that have undergone DNA alkylation. One hypothesis of DNA damage recognition is AGTs binding to the DNA and scanning *via* one-dimensional sliding.[125,137] Alternatively, AGTs may be scanning DNA by a three-dimensional hopping mechanism.[137] It is however, unclear how this exactly occurs.

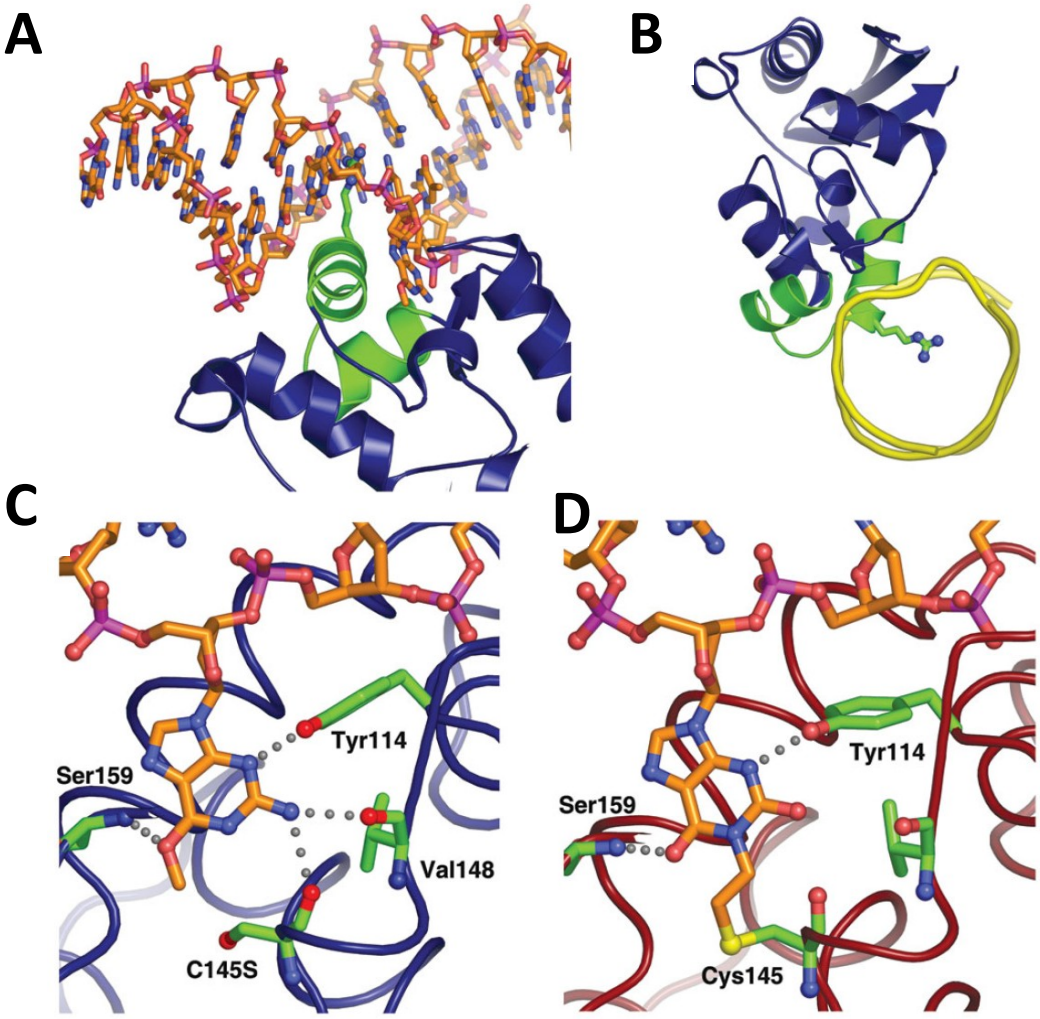
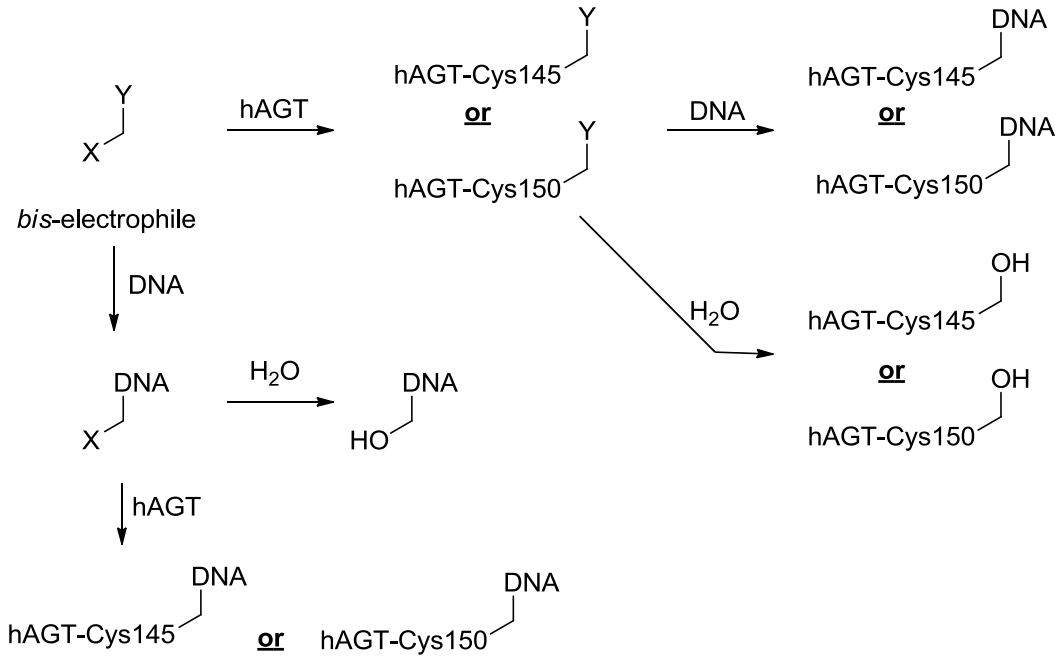


Figure 1.9: (**A**) Binding mode of hAGT to DNA by the HTH motif. Note the minor groove binding. (**B**) Arg finger (Arg128) that protrudes the DNA helix and H-binds with orphaned nucleobase (dC in this case). DNA backbone is illustrated as as a yellow tube. (**C**) Cutaway view C145S–O^6-MedG active site. (**D**) Cutaway view hAGT– $N1$,O^6-ethanoxanthine active site. Images are reproduced from the following reference.[137]

1.3.6 Cytotoxic effects of hAGT

hAGT can have aberrant cellular effects, despite having a critical role in genomic maintenance. In the presence of bifunctional alkylating agents, it can follow a series of pathways to generate other types of DNA damage like DPCs and abasic sites.[132] A general pathway is shown in **Scheme 1.3**. Because of the numerous reaction centers on the DNA, many DPCs can

be produced at various positions on the DNA scaffold. More importantly, these products may be difficult to identify and isolate in biological settings, as these are most-likely transient (degraded by the cell). Procedures to generate chemically stable DNA-hAGT DPC species is thus of value. One approach was developed by McManus and coworkers from our laboratory using ICL DNA.[151]



Scheme 1.3: Formation of DNA-AGT cross-links in the presence of bis-electrophiles (where X and Y represent the reaction centers and / or leaving groups). The chemical reacts with either the DNA or the protein first and subsequently reacts a second time with its counterpart to generate the DPC species. Since hAGT contains two Cys residues, two general DPC can be generated. Depending on the location of the linkage onto the DNA, multiple degradation pathways can occur if the DPC is chemically unstable. For instance, the linkage at the N7-atom leads to abasic site formation and hAGT functionalized with a guanine nucleobase.[170] Mono-adduct formation would arise from hydrolysis of one of the electrophilic centers.

1.3.7 Application of hAGT in Biotechnology

Because hAGT is capable of repairing a wide scope of substrates, including the free guanine nucleobase, more-recent developments have allowed for the generation of hAGT proteins fused with functional groups. O^6-benzylguanine is efficiently repaired by hAGT,[171] and

derivatives functionalized at the *para*-position have shown to be well-tolerated by hAGT.[172] This allows for the specific tagging of hAGT with virtually any *para*-functionalized benzyl adduct, which has found biotechnological applications and is known as the SNAP-tag.[172–178] In essence, an hAGT (or hAGT variant) is fused to a target protein and subsequent reaction with a labeled O^6-benzylguanine derivative supplies a final labeled covalent complex. This technology offers a variety of benefits including low off-target effects and exceptional orthogonality given that O^6-benzylguanine derivates are specific to hAGT.

1.4 Translesion Synthesis (TLS) and DNA Polymerase *η*

1.4.1 DNA Pols and replication

DNA polymerases catalyze the duplication of DNA by incorporating portions of reactive building blocks known as 2'-deoxyribonucleoside triphosphates (dNTPs, **Figure 1.10**) into a growing (primer) DNA strand. It discriminates which nucleotide to incorporate in the primer strand by reading the template strand. There are currently 17 known human DNA Pols. Of these only five are involved in the standard genomic and mitochondrial DNA replication (Pol α, δ, ε, and γ and telomerase).[179] DNA Pols heavily involved in DNA replication, also known as replicative DNA Pols, have exceptional kinetics and fidelity associated with their activity.[180] Moreover, these Pols are extremely diligent in incorporating the correct nucleotide in the primer strand, meaning they have high fidelity.[181,182] In the case of Pol δ for instance, the fidelity is approximately 1 error for every 10^5 nucleotides read. In addition, this polymerase (and others) possesses an exonuclease domain that can excise the misinserted nucleotide and allow for correction by the Pol (and contributes another 10-60 fold increase in fidelity).[181,182] In spite of their high fidelity, these Pols have an extreme intolerance for modification to the DNA. Even small modifications such as a methyl group (at the O^6-atom or O^4-atom of dG and dT,

respectively, for instance) can severely stall DNA replication. TLS is activated when repair mechanisms are unable to act on certain DNA damage, or when the damage evades these repair pathways. It had been hypothesized that the cell had to have coping mechanisms to accommodate irreparable DNA damage, yet, it was only within the past two decades that specialized DNA Pol were characterized as playing an integral part *via* TLS.[183,184] This specific pathway does not repair the damage *per se*, but rather, bypasses it in order to alleviate the replication stall. Organisms from all kingdoms of life have evolved many Pols capable of carrying out TLS and part of the focus of this dissertation is on one Y-family DNA Pol (Pol η).[185] Currently, four known Y-family DNA Pols have been identified in humans; Pol η, Pol ι, Pol κ, and Rev1. Pols from the Y-family are characterized as having low fidelity and processivity (nucleotide incorporation rates) for unmodified DNA. In addition, these proteins lack a proofreading domain, which contributes to their lowered fidelity. Their active sites are generally larger and is one of the contributors as to why they can accept a wider range of substrates during catalysis. Once thought to be "jacks-of-all-trades", evidence suggests that these proteins have specifically evolved to cope with defined types of DNA damage.[186] This remains an ongoing investigation using *in vitro* DNA polymerization experiments.

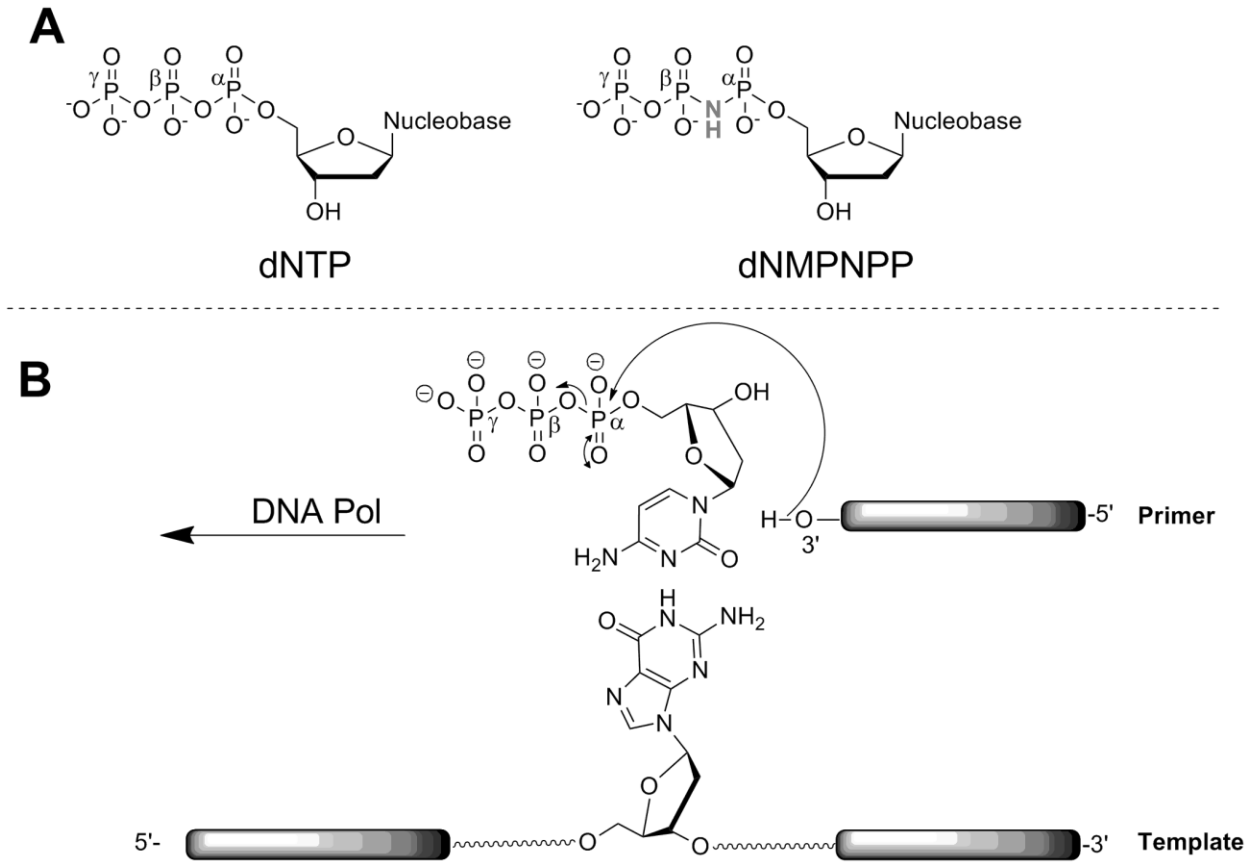


Figure 1.10: (**A**) The general chemical structure of dNTPs is shown, as well as a non-hydrolyzable analogue dNMPNPP. The latter has an (α-β) imido moiety which prevents cleavage of this bond by DNA Pols. These analogues have found applications in X-ray crystallography of ternary complexes.[187] (**B**) Simplified illustration of DNA primer extension from a DNA polymerase reading the template strand. In this case, the polymerase would have to incorporate dC mono-phosphate into the primer for the correct insertion. For simplicity, no general base and acid are shown although they play a critical role in the catalysis. Note the bond severing of the α- and β-phosphate group, as the 3'-OH attacks the α-phosphate group. The reactant is a dNTP, the 2'-deoxynucleoside monophosphate (dNMP) gets incorporated into the primer strand.

DNA Pols have three ubiquitous subdomains; the thumb, palm, and finger. The first crystal structure of these proteins, Klenow fragment of *E. coli*, was reported in 1985.[188] Since then many crystal structures have been reported and, despite varying sequence identity, there are strikingly common structural features across all cases.[189,190] They all share a similar catalytic domain architecture, resembling that of a right hand. Subdomains were named accordingly

(thumb, palm and finger domains).[191] Each of these has specific roles in catalyzing DNA extension. Briefly, the DNA lies in the palm domain, where the chemical bond formation occurs. Any of the four nucleotides can bind to the binary complex. Yet, it is with the finger domain that the correct or wrong dNTP is detected. If incorrect, the polymerase will remain in an open form and allow for dissociation and re-association of another dNTP. There are a number of structural reasons as to why the DNA polymerase remains in the open form, such as interference with the optimal geometry of the template or primer strands, wobble base pairing, or fraying of the DNA at the insertion site.[192,193] If the selection is detected as correct, the polymerase rearranges into a "closed" form.[194] The closed state generates the binding site in the palm domain for the newly-formed base pair with the incoming nucleotide and the template nucleotide. This allows for the subsequent chemical bond formation.

Catalysis is dependent on the binding of two divalent metal ions (normally Mg^{2+}, but others have been shown to work like Mn^{2+}, but not Ca^{2+}). The chemistry occurs on the primer DNA *via* a nucleotidyl-transfer reaction. A new bond is formed between the 3'-OH of the primer DNA and the α-phosphate of the dNTP, whereby the phosphodiester bond between the α- and β-phosphates is cleaved (**Figure 1.10B**). The reaction is assumed to go through a pentavalent phosphate intermediate *via* an addition-elimination mechanism.[195] DNA and RNA polymerases and many nucleases seem to follow this two-metal-ion-dependent mechanism.[196] Kinetic measurements, computational, and structural studies have all guided the understanding of the mechanism involved in DNA chain extension. That is, crystal structures of ternary complexes (E • D_n • dNTP; **Figure 1.11**) in the presence of Ca^{2+} (which prevents efficient activity by DNA Pols) have illustrated the geometric configuration of the active site prior to the reaction. Kinetic studies, such as the pre-steady-state, suggest that gross conformational rearrangement occurs by

substrate binding, followed by the rate limiting step which is a second conformational change in the active site.[193,197,198] It was only until recent that the process of DNA replication was visualized *via* multiple static snapshot images during bond formation.[198] This study was accomplished using DNA Pol η, the model Y-family translesion synthesis Pol described in this thesis (**CHAPTERS II** and **VI**), and its mechanism of bypass will be described in greater detail below.

The mechanism by which DNA gets extended is an assembly of six or more steps (**Figure 1.11**).[193,199] Referring to **Figure 1.11**, the first step is complexation of the enzyme with the DNA duplex. The binding occurs in a given directionality such that the shorter DNA fragment is oriented for 5'→3' extension catalysis. The binding of the dNTP generates the ternary complex (step 2), which entices the finger domain to close if the geometric and energetic "price-is-right". The enzyme reaches an excited state (step 3), where the chemistry occurs generating an extension of the DNA primer by one unit and pyrophosphate (step 4). The ternary complex then relaxes to ground state (step 5) followed by release of the pyrophosphate by-product. The cycle can halt by releasing of the DNA (step 7), or enter a second round of catalysis (step 8). Step 8 can only be achieved after movement of the polymerase along the oligonucleotide.

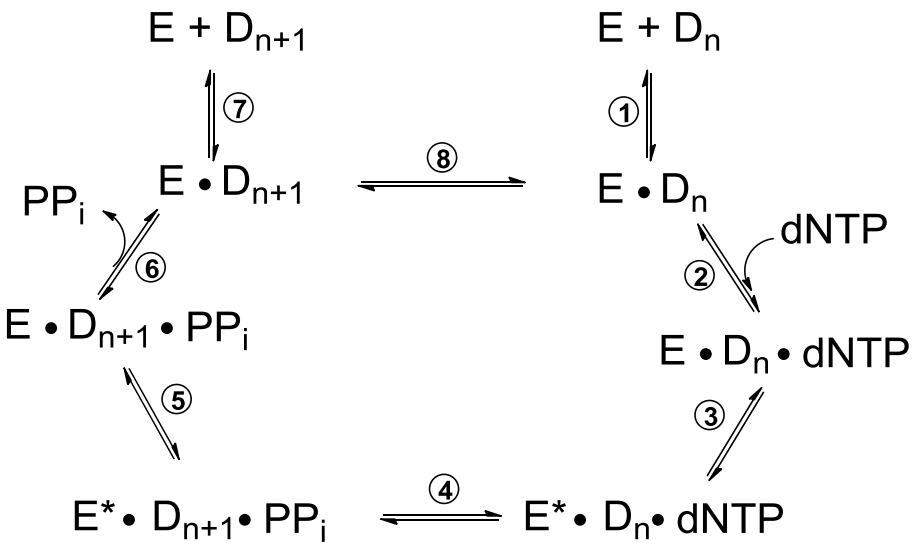


Figure 1.11: General catalytic mechanism for DNA polymerases. The figure illustrates a minimal mechanism required for DNA polymerization. E, D_n, and dNTP represent the enzyme, DNA and 2'-deoxynucleoside triphosphate, respectively. E* stands for the excited state and figure was adapted from the following reference.[199] Note that the each intermediate can have a conformational rearrangement associated with it or possess another state that is inactive. One such inactive state has been proposed for the binary complex (E •D_n) and ternary complex (E • D_n • dNTP).[199]

1.4.2 Y-family DNA Pols

The basic mechanism of TLS is illustrated in **Figure 1.12**. The first step is a stall occurring in face of a replicative DNA Pol. The TLS pathway is activated and the current Pol is removed and replaced by a TLS Pol. This stall-rescue mechanism allows the TLS Pol to add directly across the modified template nucleotide, or slightly past the damaged site. The activity of TLS Pol will be error-free or error-prone (*i.e.* high or low fidelity, respectively) depending on the type of DNA damage and the specific TLS Pol activated. Note that a second TLS Pol may be called upon after the extension from an initial Pol. Once canonical DNA is restored past the damage site, the original replicative Pol is reactivated.[200] The are numerous proposed interactions between different DNA Pols within cellular context and this remains an ongoing investigation.[199]

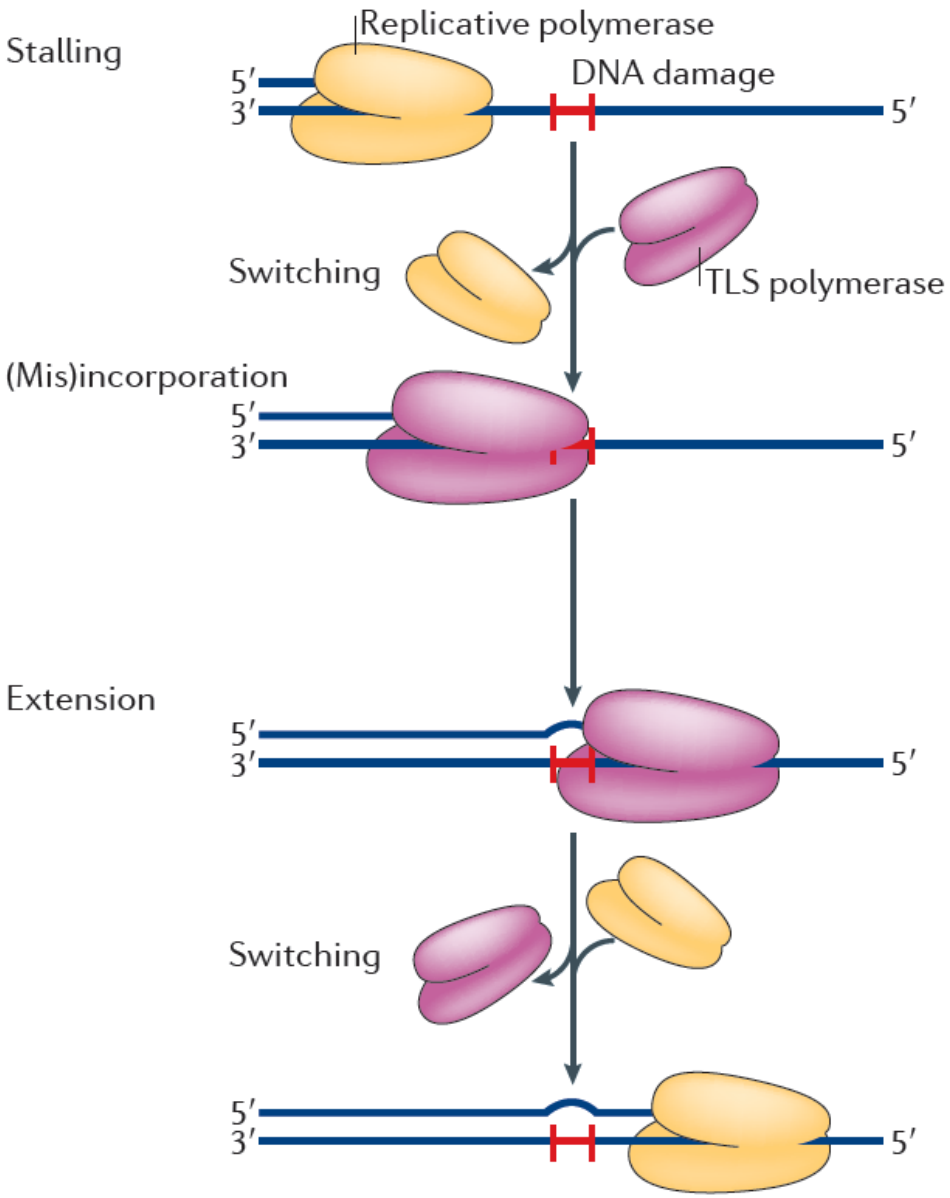


Figure 1.12: Simplified steps involved in TLS adapted from the following reference.[200]

The Y-family DNA Pols have the largest class of TLS Pols. There are two members of the Y-family DNA polymerase in *Saccharomyces cerevisia* (yeast; yRev1 and yPol η). As mentioned above, human Y-family Pols have four current members (Rev1, hPol ι, hPol κ, and hPol η), each having their own substrate profile. Rev1 is said to be unique across all Pols given its exceptional capability of incorporating dC across dG templates.[201] It can also bypass abasic

sites[202] and *N*2-dG adducts.[203] It was later discovered that the preference for dC was a feature specific to REV1 since this protein flipped out the template nucleotide and used its own active site (Arg residue) as the reading template, which was true for both yRev1 and human Rev1.[204,205] Pol *ι* is capable of correctly incorporating dT across dA and is also capable of extending across *N*2-dG adducts.[200,206] However, this Pol has difficulty in extending templates containing dT stretches, which was attributed to its mechanism of extension.[206] It turns out Pol *ι* utilizes the Hoogsteen face of the nucleobase, as opposed to the Watson-Crick face, to read the template. In this manner, it is capable of proficient extension across dA nucleotides (dAMP).[206] Pol *κ* has shown to generate frameshift adduct formation (skipping a template nucleotide, generating a product that is *n*-1 in length), when carrying out DNA replication and is also capable of extending across *N*2-dG adducts.[200] Moreover, hPol *κ* is capable of extending past the damaged site with fidelity (*e.g.* once the nucleotides have been incorporated directly across the damage).

1.4.3 Human DNA Pol *η*

The most widely studied TLS Pol is DNA Pol *η*, primarily because improper function of the *POLH* gene leads to the disease xeroderma pigmentosum variant (XPV).[207] These patients suffer from sunlight hypersensitivity and higher incidents of cancer, particularly of the skin.[207,208] Interestingly, the XPV cells are NER-proficient, whereas conventional xeroderma pigmentosum (XP) is characterized by having dysfunctional NER machinery. XPV are especially sensitive to UV radiation, which epitomized the importance of the hPol *η* in the protection against UV-induced DNA damage. In 2004, it was identified that hPol *η* was capable of bypassing the *cis-syn* TT CPD (**Figure 1.1**), which accounted for experimental evidence of hPol *η*'s involvement in TLS against UV-induced insults.[209] It was only later, in 2010, the first crystal structure of hPol *η* was reported in the presence of non-hydrolyzable dNMPNPP (**Figure 1.10**), undamaged DNA or

DNA containing a *cis-syn* TT CPD insert.[187] It was noted that, like all other Y-family DNA Pol, hPol η contained an additional domain called the little finger (LF).[187] The structure revealed that the protein acts as a splint, by enforcing an almost-canonical B-form structure on the DNA duplex containing the CPD damage. Alone, the CPD-containing duplexes have gross structural perturbations.[89] Moreover, other structures of repair protein complexed with DNA containing the TT CPD also didn't force the duplex back into canonical shape.[210,211] The larger active site of the protein allows for the entire CPD to be present and enables selectivity for the correct incoming nucleotide to undergo catalysis. Residues Gln38 and Arg61, uniquely conserved across all Pol η homologues, were found to have a crucial role in guiding the incoming nucleotide.[187] The bypass mechanism was extrapolated from four crystal structures with DNA containing modifications in varying positions along the template strand (**Table 1.1**), along with an unmodified duplex. This enabled static snapshots at different "simulated" stages of insertion (*e.g.* before the first insertion, at the first insertion, second insertion, and post insertion) (**Figure 1.13**). From TT1 and TT2, the study revealed that both nucleotides of CPD retained H-bonding with the incoming nucleotide, which was different from a replicative Pol (bacterial T7) complexes with DNA containing a CPD insert.[187,212] In the latter case, only one nucleotide part of the IaCL was H-bonding. Both cross-linked residues are interacting with Gln38 *via* the *O*2-atoms. The importance of Gln38 was verified using the Q38A variant, which lowered overall catalytic efficiency and promoted stall of the enzyme at the stage equivalent to TT3. In TT3, Gln38 aids the template base to stack with the CPD, consequently allowing for proper alignment necessary for primer extension. The Arg61 residue, on the other hand, seems to coordinate the incoming nucleotide by interacting with the *N*7-atom of dG (or dA) and the phosphate groups of the triphosphate moiety. Polymerase efficacy was also reduced for the R61A variant, and this variant caused purine misinsertion into

the primer (dG compared to dA) across the TT CPD. All crystal structures revealed that hPol η keeps the DNA aligned and rigid because of a positively charged surface in contact with a span of four nucleotides upstream (towards the 5'-end) of the active site. The little finger domain possesses a β-strand, which runs along the DNA template and heavily interacts with it *via* H-bonding of the protein amide backbone, in addition to side chain interactions with the sugar-phosphate DNA backbone.

Table 1.1: Structure name, protein, DNA sequence and incoming nucleotide studied by Biertümpfel *et al.*[187] The DNA sequence displays the varying positions of the modification (TT in red).

Structure name	DNA sequence	Incoming non-hydrolyzable dNTP
Nrm	3' - TCGCAGTATTACT- 5' 5'- TAGCGTCAT	dAMPNPP
TT1	3' - TCGCAGTA**TT**AC- 5' 5'- TAGCGTCAT	dAMPNPP
TT2	3' - CGCAGTA**TT**CAAT- 5' 5'- TGCGTCATA	dAMPNPP
TT3	3' - CGCAGTA**TT**CAAT- 5' 5'- ACGTCATAA	dGMPNPP
TT4	3' - CGAGCA**TT**ACTAC- 5' 5'- TCTCGTAAT	dGMPNPP

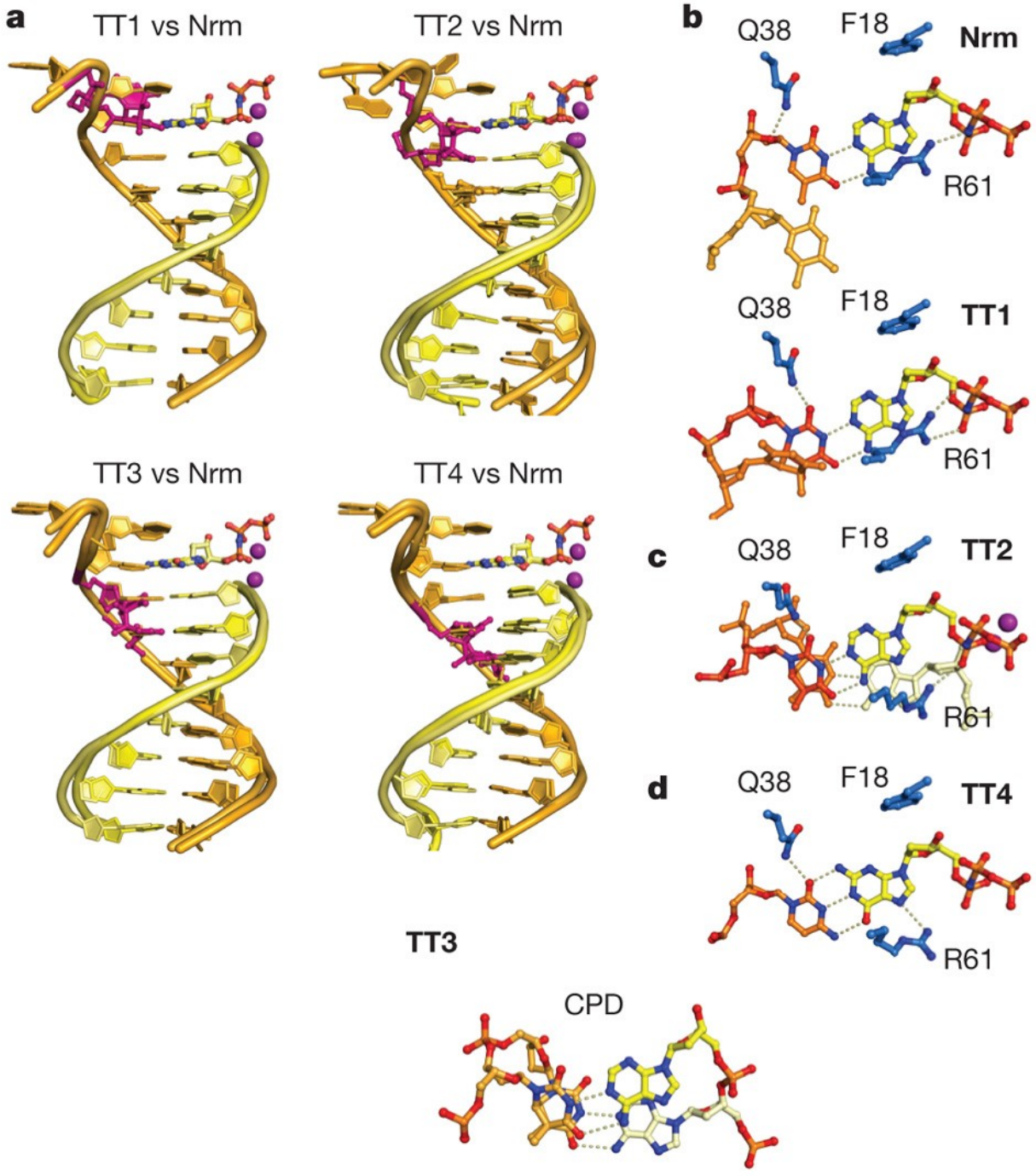


Figure 1.13: (**a**) Crystal structure overlay of TT1, TT2, TT3 and TT4 DNA with the DNA in Nrm (undamaged). The CPD modifications are shown as red sticks. Spheres are Mg^{2+} ions. (**b**–**d**) The replicating base pairs in the four ternary complexes hPol η residues (note Q38 and R61) that contact nucleobase and incoming dNMPNPP. H-bonds are shown as dashed lines. The figure was adapted from Figure 2 in the following reference.[187]

Since its discovery, a multitude of different DNA lesions have been tested against bypass by hPol η. Studies are normally carried out using *in vitro* DNA primer extension as a representative model. The identification of plausible substrates is difficult to predict and, more often than not, is carried out by trial and error.

1.4.4 Bypass of O^6-alkyl-dG, O^4-alkyl-dT and IaCL

The objective of this section is to briefly review a few key studies that have been reported for modifications pertaining to those reported in the thesis. These include O^6-alkyl-dG, O^4-alkyl-dT and IaCL since **CHAPTERS II** and **V** focus on O^4-alkyl-dT mono-adducts and O^6-dG-alkylene-O^6-dG IaCL, respectively. This will allow to situate these two studies in the context of the current literature. Although literature directly pertaining to these studies are reviewed in each individual chapter, this section will elaborate further.

A thorough investigation of O^6-alkyl-dG mono-adducts bypass by hPol η, hPol κ, and hPol ι was reported in 2006 by the Choi *et al.*[45] The study focussed on the effect of bulk at the O^6-atom of dG by using three adducts; the Me, benzyl (Bn) and 4-oxo-4-(3-pyridyl)butyl (Pob) groups. These were chosen on the basis that O^6-MedG and O^6-BndG could be bypassed by replicative bacterial DNA Pols, whereas the O^6-PobdG stalled the enzymes.[213] Moreover, the O^6-PobdG is a well-known tobacco-derived carcinogen.[214–216] hPol κ and ι mainly produced frameshift adduct formation (-1) by skipping the O^6-MedG and O^6-BndG insert in the template. These two TLS Pols were both blocked by the O^6-PobdG modification, suggesting that steric bulk influences the activity of hPol κ and ι. Yet, hPol η was capable of bypassing all three modifications with varying efficiencies (all reduced compared to unmodified DNA). Steady-state analysis revealed hPol η and κ inserted the pyrimidines (the correct dC (dCMP) and incorrect dT (dTMP) nucleotides) across the O^6-alkyl-dG adducts, while hPol ι exclusively misinserted

dTMP. hPol η caused frameshift adduct formation (-1) in the case of O^6-PobdG, and in doing so stalled after adding across the subsequent template nucleotide.

In terms of O^4-alkyldT, two recent studies both carried out by Anderson *et al.* were reported on O^4-MedT bypass (2012)[47] and O^4-EtdT (2013)[217] by various Pols including some from the Y-family. The O^4-MedT cause blockage of the yPol η and hPol κ, and upon bypass, both enzymes caused dG nucleotide (dGMP) misinsertions across the modification. A similar profile was observed for O^4-EtdT, tested against hPol η, hPol κ, and hPol ι with an error-prone behavior for all three Pols (dGMP preferantial misinsertion). It was noted that hPol ι was capable replicating one nucleotide across the O^4-EtdT modification but was stalled subsequently. hPol κ proved inefficient in bypassing the O^4-EtdT, whereas hPol η was capable of bypassing and extending past the lesion. Together, these results support the mutagenicity O^4-alkyl-dT lesions despite action by TLS Pols. It is unclear however, how this effect was exerted by such small lesions.

hPol η is proficient in bypassing the UV-induced d(TT) CPD IaCL, but is also capable of efficiently bypassing other CPD adducts d(TU)[218] CPD and other derivatives.[219] The d(CT) and d(TC) CPDs are chemically unstable and undergo deamination to generate the 2'-deoxyuridine (dU) analogue.[220] A CPD of dT and N^4-methyl-dC is chemically stable but induces misinsertions by hPol η and prevents fully extended products past the misinsertion.[221] Strikingly, hPol η is incapable of pyrimidine (6-4) pyrimidone IaCL adduct bypass.[82] For DNA containing the d(G[8,5-Me]T) IaCL, hPol η performed mostly error-free DNA replication, with more difficulty across the dG portion of the IaCL.[222,223] Evidence suggests that the inverted IaCL, d(T[5-Me,8]G), is more problematic for DNA replication machinery *in vivo*.[222] Studies using yPol η

demonstrated capability of bypass of the d(G[8,5]C) IaCL with error-prone profiles for the dG portion, as with the d(G[8,5-Me]T) IaCL.[78]

Another noteworthy aspect of hPol η is its potential role in chemotherapeutic resistance by efficiently bypassing lesions generated from drugs in the clinic, such as cisplatin and oxaliplatin.[71] hPol η is capable of bypassing the most-abundant Pt-dGG lesion adduct, but inefficient at extending past the lesion site.[224–227] Structural investigation of the ternary complex revealed that the enzyme conformationally rearranges the DNA backbone of the template to catalyze efficient insertion opposite the modification, much like the d(TT) CPD IaCL.[227] The conformation of the Pt-dGG lesions in the active site is different from the CPD, most-likely due to its larger size. The short loop R61-M63 is shifted to accommodate the entire IaCL dimer. The 5'-dG part of the lesion is held in a *syn* conformation about the glycosidic bond during insertion across the 3'-dG. This is primarily due to the steric constraints imposed by the Pt-lesion (≈30° roll angle). During the second insertion, this roll is forced onto the 3′-dG part of the lesion, allowing proper geometry of the 5'-dG residue to H-bond with the incoming nucleotide. Once insertion across both nucleotides part of the IaCL is carried out, the lesion protrudes the major groove of the DNA causing an entire shift of the LF domain. The molecular splint imparted from the protein is thus compromised and severely hinders Pol activity by allowing the damaged nucleotides to clash with the incoming dNMPNPP. The enzyme remains capable of inserting one nucleotide past the damage site, but is then stalled.[227]

1.5 Objectives of the thesis and organization

1.5.1 Objectives

A major objective of the thesis was the synthesis of DNA containing site-specific modifications, on scales and purities amenable to structural and biochemical investigation. The DNA probes were utilized to shed light on different cellular response mechanisms such as AGT-mediated repair and bypass by hPol η. Moreover, certain probes were designed to efficiently generate DPC between DNA and AGTs, which may find applications in structural biology or chemical biology.

There are two general methodologies that can be used in the production of DNA modifications. In one approach, unmodified DNA can be treated with certain exogenous chemicals that will generate the desired lesion on the DNA. This approach has been fruitful in identifying a number of different adducts that have formed at specific sites of alkylation (IaCL formed at the *N*7-atom of neighbouring dG residues,[69] for instance). However, this methodology is inefficient in producing large quantities of a desired site-specific modification, particularly for adducts part of the minor products, given that a number of unwanted side products are formed simultaneously in the DNA. Moreover, isolating the desired product from by-products may be virtually impossible in high yield and purity with current separation techniques. The second methodology, used by a number of research groups including ours, employs a combination of solution and solid-phase synthesis (SPS) to generate DNA with site-specific and well-defined modifications. This avenue is particularly effective given that the DNA strand is specifically tailored using chemical synthesis entirely.

The O^4-alkyl-dT series have been investigated to a lesser extent compared to their O^6-alkyl-dG counterpart families. Previous studies, from our group and others, have established that AGTs can efficiently discriminate against these two types of alkylation damage. More-recently, evidence using a conformationally-locked analogue of O^6-alkyl-dG has suggested that the

orientation of the alkyl group affects repair efficiency by hAGT.[228] Interestingly, the latter analogue locked the α-carbon (where nucleophilic attack occurs) in an *anti* conformation with respect to N1-atom, and was a poor substrate of hAGT. In contrast, the **eX** modification (**Figure 1.8**), which locks the α-carbon in *syn* conformation with respect to the *N*1-atom, was efficiently repaired by hAGT.[137,160] The repair of DNA containing **eX** by hAGT generated the DPC, which was later crystallized and the X-ray crystal structure was determined.[137] We initially developed analogous constructs for the O^4-alkyl-dT series to further probe repair by various AGTs. To our dismay, these novel constructs remained inert to AGT-mediated repair (mentioned in **CHAPTER IV**). This was not surprising for hAGT given that O^4-MedT is poorly processed by this protein. It is also known that O^4-alkyl-dT are poor substrates of mammalian repair pathways and have been proposed as potential substrates for TLS. A few studies have investigated the bypass of O^4-MedT[47] and O^4-EtdT[217] using recombinant translesion DNA Pols. However, no structural insights have elucidated why these lesions cause certain DNA Pols to miscode. In addition, our conformationally locked analogues (bicyclic systems) could offer insights on the impact of the lesion restriction on the bypass by DNA Pol η. The latter protein was utilized as the model translesion DNA Pol.

The occurrence of IaCL damage has been identified and sources for their production have been well-established and characterized. There has been growing evidence, however, that attributing their repair solely to the NER machinery may not have been unequivocally accurate. As mentioned above, translesion synthesis, particularly by Pol η, plays a pivotal role in genomic maintenance against DNA damage (including IaCL DNA) in eukaryotes.[186] It would thus appear that exact cellular processing mechanism of IaCL DNA is not fully elucidated and most-likely depends on the nature of the IaCL DNA and perhaps the frequency of occurrence. Our group has

suggested a potential role for AGTs in the repair of ICL DNA[64,65,150,151] and this may hold true as well for similar adducts, in an IaCL motif. This was a recurring objective in **CHAPTERS III**, **IV**, and **VI**.

Moreover, the strategy we devised to introduce IaCL in DNA allowed for the production of two sets of similar adducts; one lacking the phosphodiester linkage between the cross-linked residues, and another set containing this linkage. The former set of IaCL DNA would be inherently more-flexible. We could then probe the influence of the phosphodiester linkage on the AGT-mediated repair of IaCL DNA (**CHAPTER IV**). A similar approach was taken to identify the influence of the phosphodiester group on the bypass of IaCL DNA by hPol η (**CHAPTER V**).

Using the conclusions drawn from **CHAPTER III** and **VI**, and the substrate specificity of various AGTs, novel IaCL DNA probes were designed, prepared, and tested against AGTs (**CHAPTER VI**). As hypothesized, new DNA-hAGT conjugates were prepared with excellent conversions and increased versatility compared to previously developed ICL probes.[151]

1.5.2 Arrangement of the thesis

CHAPTER I

This chapter supplies highlights concerning DNA damage and the consequences of such damage on cellular activity. DNA alkylation is briefly overviewed with selected examples. The biological consequences of O^6-alkyl-dG and O^4-alkyl-dT is highlighted. The biological consequences of ICL, IaCL, and DPC lesions are briefly described. The direct repair pathway is overviewed with a focus of AGT-mediated repair. Finally, TLS by Y-family DNA Pols is described in general and a focus is placed on the bypass mechanism of hPol η.

CHAPTER II

This chapter presents the synthesis of conformationally restrained analogues of O^4-MedT and O^4-EtdT. These adducts link the *C*5-atom of the pyrimidine ring with the O^4-atom by an ethylene or an *n*-propylene bridge. Once incorporated into DNA oligomers, biophysical studies demonstrated minor differences between the open-chain analogues, with respect to those that were conformationally constrained. Bypass studies using DNA Pol η on the templates containing the adducts demonstrated a clear variance between all the modifications. Crystal structures of ternary complexes of hPol η, incoming dA or dG nucleotide, and DNA containing O^4MedT or O^4EtdT supplied additional insights concerning the error-prone behavior of hPol η towards these adducts. That is, the structures involving DNA containing O^4MedT revealed an unprecedented conformational geometry of the nucleobase in the active site of hPol η.

CHAPTER III

DNA containing O^6-dG-alkylene-O^6-dG IaCL mimics, lacking the phosphodiester linkage the cross-linked site, were synthesized and AGT activity towards them was evaluated. The biophysical studies demonstrated a reduction in the thermal stability due to the presence of the IaCL, while retaining the overall B-form duplex DNA. Finally, hAGT was capable of efficiently repairing the model IaCL DNA, with greater efficacy towards the heptylene IaCL DNA. This demonstrated the potential capability of hAGT-mediated repair of IaCL DNA

CHAPTER IV

DNA containing the O^6-dG-alkylene-O^6-dG IaCL containing the phosphodiester backbone were prepared synthetically. Duplexes containing the IaCL DNA demonstrated a decrease in thermal stability compared to the unmodified duplex. Interestingly, the thermal

profiles were in agreement with the identical IaCL lacking the phosphodiester linkage. The global DNA structure was not substantially affected, as assessed by circular dichroism. Repair assays of these IaCL demonstrated only moderate repair by hAGT, with much lower efficiency compared to the model IaCL lacking the phosphodiester group, which suggested that repair may be dependent on the lesion's orientation and/or flexibility. This chapter also reported the efficient repair of the single-stranded IaCL DNA mimic synthesized in **CHAPTER III**.

CHAPTER V

The IaCL DNA prepared in the previous two chapters were evaluated for bypass processivity by hPol η. The bypass profiles were significantly different for IaCL containing the phosphodiester backbone linkage (**GpG4** and **GpG7**), in comparison to those lacking it (**GG4** and **GG7**). Steady-state kinetic evaluation revealed blockage of hPol η by the presence of the IaCL in all cases, relative to the unmodified control. The enzyme incorporated the correct dCMP across the first adducted O^6-alkylene-dG nucleotide (3'-dG) but demonstrated error-prone behavior across the second O^6-alkylene-dG (5'-dG). Frameshift adduct (-1) formation was a major product observed in the replication of the 5'-dG of **GG4** and **GG7**, whereas dTMP was incorporated at the highest frequency for **GpG4** and **GpG7**. An increase in IaCL linker size, from butylene to heptylene, resulted in an increase in the rate of misinsertions by hPol η.

CHAPTER VI

Efficient production of site-specific DPC species may find use for a variety of different biotechnological applications. Towards this end, DNA linking two thymidine residues, or linking a thymidine residue to a 2'-deoxyguanosine residue (either in a 5'→3' or 3'→5' fashion), lacking a phosphodiester linkage at the cross-linked site, were synthetically prepared. In light of

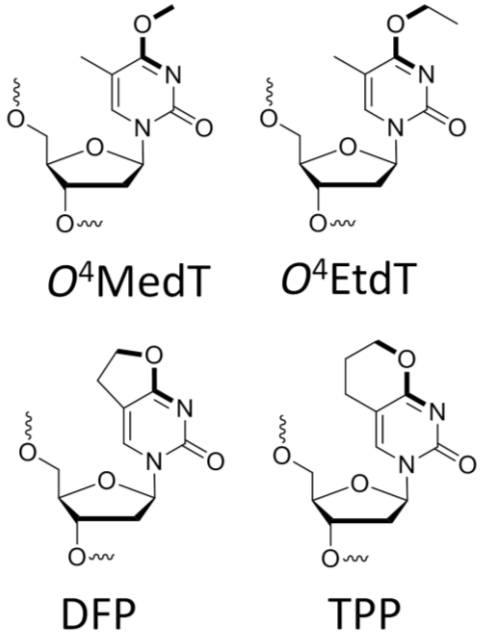
the results acquired throughout the previous chapters, we anticipated efficient repair by AGTs towards the newly prepared IaCL. All IaCL DNA displayed similar UV thermal denaturation profiles, with a decrease in thermal stability relative to the unmodified duplex DNA. Circular dichroism signatures revealed little global structural distortions induced in the duplex by the IaCL, relative to the control. The O^4-thymdine-alkylene-O^4-thymdine (**TT4** and **TT7**) and the (5'→3') O^6-2'-deoxyguanosine-alkylene-O^4-thymdine (**GT4** and **GT7**) IaCL DNA adducts were inert against any of the AGT tested (hAGT, OGT, and Ada-C). The (5'→3') O^4-thymdine-alkylene-O^6-2'-deoxyguanosine (**TG4** and **TG7**) IaCL DNA containing butylene or heptylene tethers were efficiently repaired by the human variant, whereas Ada-C was capable of modestly repairing the heptylene IaCL adduct. The IaCL DNA were utilized to conjugate hAGT to the 3'-end of the DNA strand in excellent yields with short reaction times using single stranded **TG4** and **TG7**, as well as **GG4** and **GG7** (single-strand repair of **GG4** and **GG7** by hAGT was first reported in **CHAPTER IV**). DPC products were characterized by SDS-PAGE and LC-MS. These IaCL expand the toolbox of strategies to site-specifically conjugate hAGT with DNA. hAGT was functionalized with a fluorescently-labelled DNA IaCL sequence as an example to show the feasibility of this method.

CHAPTER VII

This final chapter summarizes the results and general conclusions resulting from **CHAPTER II - CHAPTER VI**, as well as suggestions for future directions concerning the research presented in this dissertation.

CHAPTER II

Lesion Orientation of O^4-Alkylthymidine Influences Replication by Human DNA Polymerase η



Submitted as:

O'Flaherty DK[1], Patra A[2], Su Y[2], Guengerich FP[2], Egli M[2*] and Wilds CJ[1*] (2016) *Submitted to Chemical Science.*

[1]Department of Chemistry and Biochemistry, Concordia University, Montréal, Québec H4B1R6, Canada
[2]Department of Biochemistry, Vanderbilt Institute of Chemical Biology, Vanderbilt Ingram Cancer Center, Center in Molecular Toxicology, and Center for Structural Biology, School of Medicine, Vanderbilt University, Nashville, Tennessee 37232, United States

*corresponding authors.

Abstract

DNA lesions that elude repair may undergo translesion synthesis catalyzed by Y-family DNA polymerases. O^4-Alkylthymidines, persistent adducts that can result from carcinogenic agents, may encounter DNA polymerases. The influence of lesion orientation around the C4-O^4 bond on processing by human DNA polymerase η (hPol η) was studied for oligonucleotides containing O^4-methylthymidine, O^4-ethylthymidine, and analogs restricting the O^4-methylene group in an *anti*-orientation. Primer extension assays revealed that the O^4-alkyl orientation influences hPol η bypass. Crystal structures of hPol η•DNA•dNTP ternary complexes with O^4-methyl- or O^4-ethylthymidine in the template strand showed the nucleobase of the former lodged near the ceiling of the active site, with the *syn*-O^4-methyl group engaged in extensive hydrophobic interactions. This unique arrangement for O^4-methylthymidine with hPol η, inaccessible for the other analogs due to steric/conformational restriction, is consistent with differences observed for nucleotide incorporation and supports the concept that lesion conformation influences extension across DNA damage. Together, these results provide mechanistic insights on the mutagenicity of O^4-MedT and O^4EtdT when acted upon by DNA polymerase η.

2.1 Introduction

DNA alkylation results from a variety of endogenous and/or exogenous agents that can interfere with vital cellular processes, *i.e.* replication and transcription.[31] The addition of alkyl appendages on the DNA scaffold can have adverse consequences such as DNA polymerase (Pol) blockage, nucleotide misincorporation, chromosomal instability, and activation of the cellular apoptotic pathway.[18,31] However, organisms have various repair pathways to restore damaged DNA. In the event that a lesion evades the process of DNA repair, translesion synthesis (TLS) by Y-family DNA Pols can occur, allowing bypass of the DNA lesion in an error-free or error-prone manner.[200] Y-family DNA Pols are described as more "promiscuous" given their larger active sites when compared with replicative DNA Pols, which accounts for their ability to bypass damaged nucleotides that induce blockage. DNA Pol η in humans (hPol η) plays a pivotal role in the bypass of certain UV-induced DNA damage, which impedes DNA replication.[209] hPol η activity has also been correlated with chemotherapeutic resistance to platinum-based agents such as cisplatin and the efficient bypass of the oxidative DNA lesion 7,8-dihydro-8-oxo-2′-deoxyguanosine.[229,230]

The O^4-position of thymidine is susceptible to alkylation by agents such as *N*-nitroso alkylamines, in certain foods, water, air, and particularly tobacco products.[231,232] Albeit a minor site of alkylation, lesions such as O^4-methylthymidine (O^4-MedT) and O^4-ethylthymidine (O^4-EtdT) are poorly processed by mammalian repair pathways, making them persistent in the genome.[21,233] O^4-MedT and O^4-EtdT hinder high fidelity replicative DNA Pol activity, resulting in misinsertion of dGTP in the daughter DNA strands.[21,22,234] Correlations between the mutagenicity of O^4-MedT and cancer have been established,[20,40] highlighting the importance of

investigating the structural properties and biological outcomes associated with this type of DNA damage.

The current understanding of the mechanism of Y-family DNA Pol misincorporation during TLS depends on a number of factors, including the nature of the DNA damage, the DNA Pol and the incoming nucleoside triphosphate. The geometrical array of the ternary complex formed (involving DNA, protein and nucleoside triphosphate) is the key characteristic that governs efficient bypass of a DNA lesion. Structural investigation by NMR and X-ray crystallography of duplexes containing an O^4-MedT insert has revealed that the methyl group preferentially adopts a *syn* conformation around the C4-O^4 bond (**Figure 2.1**).[25,235] We hypothesized that the conformation of the O^4-alkyl lesion could affect the base pair geometry during the primer extension reaction catalyzed by DNA Pol η. To address this possibility, we probed hPol η processivity with thymidine analogs that link the C5 and O^4 atoms by a dimethylene or trimethylene group, which limits the O^4-lesion to adopt an *anti*-conformation (**Figure 2.1**) to relate the structural features of O^4-alkylated dT with the bypass activity of hPol η.

We investigated bypass profiles opposite all four lesions by hPol η (steady-state single nucleotide incorporation and LC-MS/MS analysis of full-length extension products). Crystal structures of ternary hPol η•DNA•dATP and hPol η•DNA•dGTP with template strands containing O^4-MedT or O^4-EtdT reveal a distinct orientation of the former lesion that stacks atop a tryptophan residue near the ceiling of the active site instead of pairing with the incoming nucleotide. Conversely, O^4-EtdT pairs with both incoming A and G *via* bifurcated H-bonds in the insertion complexes and displays the same configuration opposite primer dG in the crystal structure of an extension complex adjacent to the nascent dG:dCTP pair. The structures provide a

better understanding of the different behavior of the O^4-MedT or O^4-EtdT lesions in hPol η-catalyzed error-prone bypass reactions and suggests a unique intermediate step in the bypass of O^4-MedT.

2.2 Results

2.2.1 Synthesis and characterization of modified oligonucleotides

The structures of O^4-MedT, O^4-EtdT and the modified pyrimidyl nucleosides 3-(2′-deoxypentofuranosyl)-5,6-dihydrofuro[2,3-*d*]pyrimidin-2(3*H*)-one (DFP) and 3-(2′-deoxypentofuranosyl)-3,5,6,7-tetrahydro-2*H*-pyrano[2,3-*d*]pyrimidin-2-one (TPP) are shown in **Figure 2.1** (methods describing the preparation of nucleosides and oligonucleotides can be found in the section **2.6**). UV thermal denaturation studies of duplexes containing single inserts of the DFP or TPP modification revealed a comparable destabilizing effect to O^4-MedT and O^4-EtdT with complementary strands containing adenine or any mismatched base pairing partner (**Figure S2.6**). Circular dichroism spectra of duplexes containing the DFP or TPP inserts revealed little deviation from a B-form structure (see **Figure S2.7**).

2.2.2 Steady-state kinetics

Steady-state kinetic assays of individual nucleotide incorporations opposite O^4-MedT, O^4-EtdT, DFP, TPP and unmodified dT were carried out with the catalytic core construct of hPol η (amino acids 1-432). In all cases, these pyrimidyl modifications blocked DNA synthesis by the polymerase relative to the unmodified control (**Figure 2.2a**, values shown in **Table S2.1**). Incorporation of the correct dAMP nucleotide by hPol η opposite to O^4-MedT, O^4-EtdT, DFP, and TPP was reduced approximately 6.5-, 12-, 4.5-, and 5-fold, respectively, relative to dT (see **Figure 2.2a**).

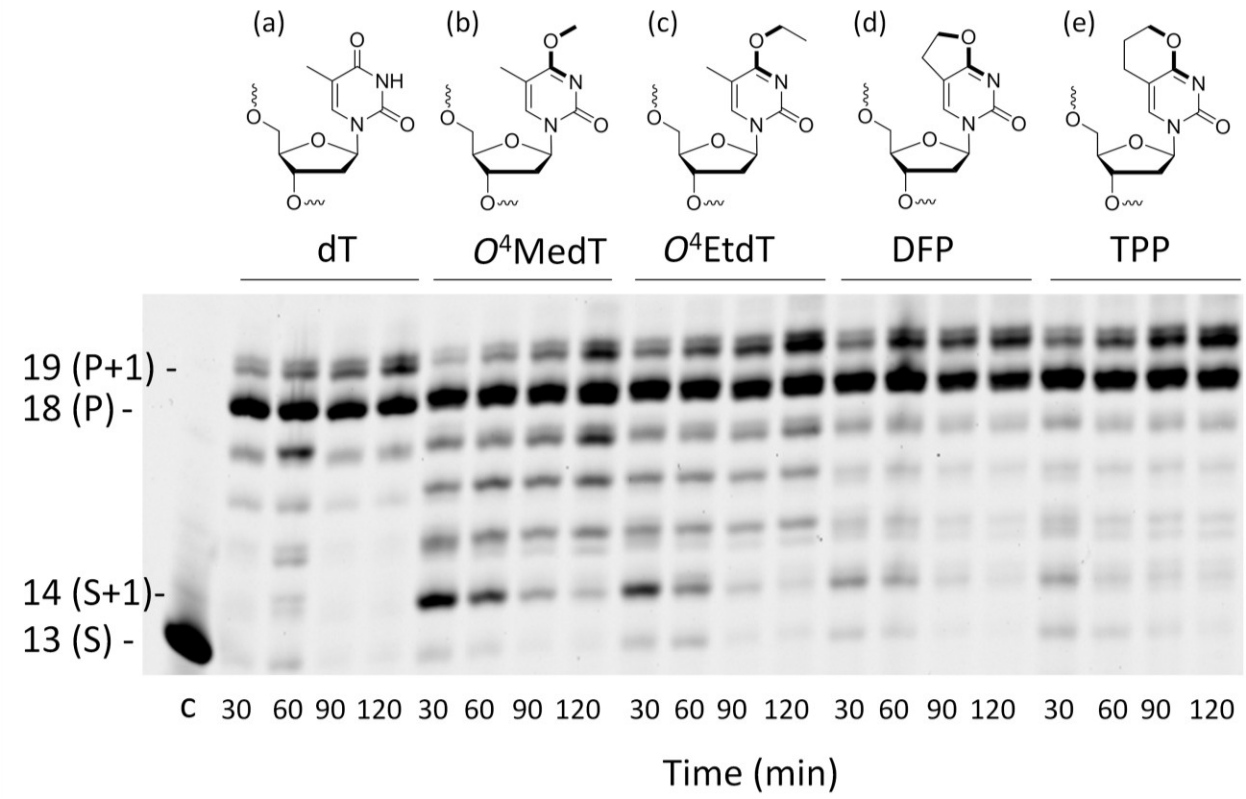


Figure 2.1: Structures of the (a) unmodified dT control, (b) O^4-MedT, (c) O^4-EtdT, (d) DFP and (e) TPP adducts. In bold are the bonds between the atoms shown to have a *syn* (b and c) or *anti* (d and e) orientation around the O^4-C4 bond. Full primer extension assays for the control undamaged substrate (dT), O^4-MedT, O^4-EtdT, DFP, and TPP-bearing substrates with hPol η. The template strand sequence is 3'-AGCATTCGCAGTAXTACT-5' where X denotes the modification and the 5'-FAM labeled primer strand sequence is 5'-TCGTAAGCGUCAT-3'.

hPol η incorporated dCMP and dTMP opposite all the pyrimidyl modifications and dT with similar catalytic efficiencies (ranging from approximately 0.002-0.015 μM^{-1} s^{-1}). However, a strong preference for either dAMP or dGMP incorporation opposite the modified pyrimidines was observed. Other than O^4-MedT, hPol η preferentially incorporated the correct dAMP nucleotide opposite all the pyrimidyl modifications. The significant incorporation of dGMP when hPol η encountered these pyrimidyl modifications, compared to the unmodified control, indicates a clear loss in substrate specificity by the polymerase (**Figure 2.2a**). In the case of O^4-

MedT, dGMP was slightly preferred as the nucleotide incorporated by hPol η (0.19 $\pm$ 0.01 *vs.* 0.18 $\pm$ 0.03 μM^{-1} s^{-1} for dGMP and dAMP, respectively).

2.2.3 LC-MS/MS Analysis of Full-length Extension Products Produced by hPol η

Analysis of single insertions by a DNA polymerase is useful for kinetic evaluation but may not reflect incorporation fidelity in the presence of all four dNTP across the damage and beyond this site. The fidelity of hPol η and its processivity past the damage site was addressed by the use of a full extension assay coupled with LC-MS/MS analysis.[45,229,236]

The optimal reaction times to observe the full extension products from the template strands containing the modifications and the unmodified control were evaluated (**Figure 2.1**). Full extension was achieved for the unmodified control at 30 min, whereas templates containing the modifications required longer reaction times (60-90 min). The time course assay revealed that hPol η had difficulty in extending past O^4-MedT and O^4-EtdT and displayed a significant "S+1" band at reaction times of 30 and 60 min. UPLC separation of the cleaved products and mass spectrometry analysis of their sequence identities revealed that dGMP was incorporated most efficiently opposite all the modifications (see **Figure 2.2b**).

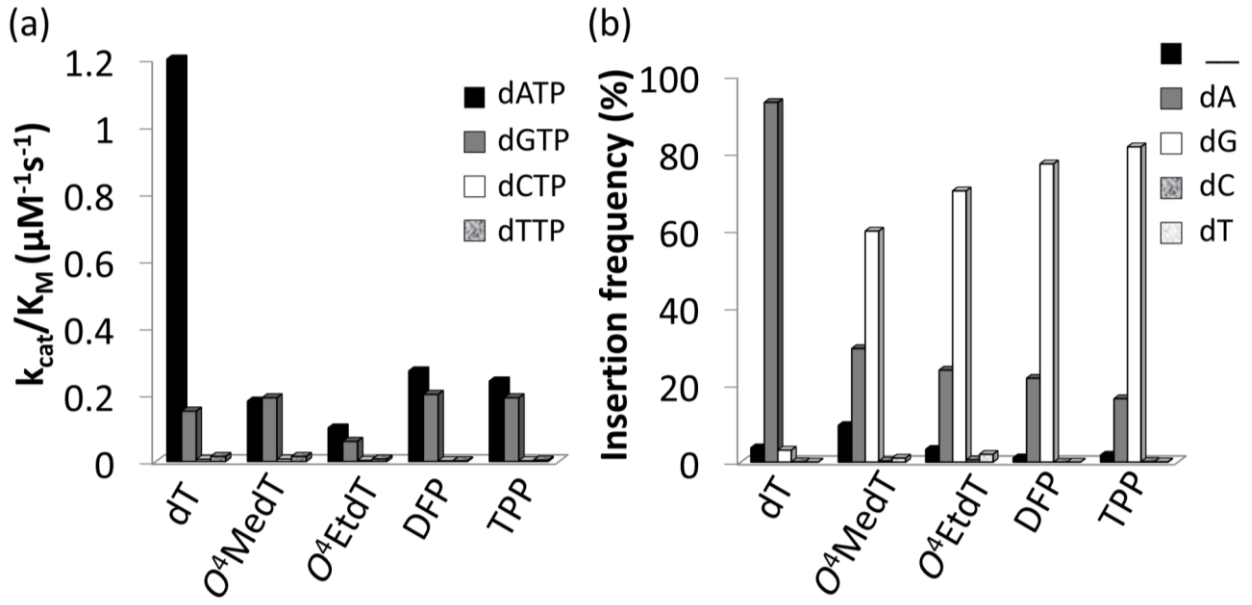


Figure 2.2: (a) Steady-state incorporation efficiencies opposite dT, O^4-MedT, O^4-EtdT, DFP, and TPP by hPol η with individual dNTPs. (b) Incorporation frequencies based on ESI-MS/MS analysis of primer extension products opposite the dT, O^4-MedT, O^4-EtdT, DFP, and TPP containing templates ("_" indicates frameshift adduct formation. Tabulated fragments identified by LC-MS/MS analysis of full-length extension products can be found in Appendix II).

The incorporation frequency opposite dT, O^4-MedT, O^4-EtdT, DFP, and TPP for the full extension products was evaluated (see **Figure 2.2b** and **Table S2.2**). The presence of O^4-MedT increased the level of frameshift formation by hPol η relative to the control (9.5 vs. 3.5%). Comparable levels of frameshifts were observed opposite O^4-EtdT and the dT control. However, the templates containing the bicyclic pyrimidine adducts did not induce a similar increase in frameshift formation with levels that were approximately one-half, relative to the control. The correct dAMP nucleotide was incorporated by hPol η at a frequency of approximately 30, 24, 22, 16, and 93% opposite O^4-MedT, O^4-EtdT, DFP, TPP, and dT, respectively. Out of the lesions investigated, hPol η exhibited the highest fidelity opposite O^4-MedT and lowest opposite the TPP. Incorporation of dGMP was observed to occur in the extension products with overall frequencies of 60, 70, 72, 82, and 3% opposite O^4-MedT, O^4-EtdT, DFP, TPP, and dT,

respectively. The accuracy of bypass varied for the O^4-alkylthymidine modifications by approximately 2:1 in favor of dGMP opposite O^4-MedT to 5:1 in favor of dGMP opposite TPP. An increased adduct size, from O^4-MedT to O^4-EtdT and DFP to TPP, resulted in a 10% increase of dGMP misinsertion at the expense of a 10% decrease of the correct dAMP incorporation. Similarly, the conformationally restrained analogues (DFP and TPP) induced an increase in dGMP misinsertion (10%) by hPol η compared to O^4-MedT and O^4-EtdT, respectively.

2.2.4 Crystal Structures of Ternary hPol η•DNA•dNTP Complexes with Templates Containing O^4-MedT or O^4-EtdT at the Insertion Stage

To visualize the O^4-MedT and O^4-EtdT lesions at the active site of hPol η trapped at the insertion stage, we determined four crystal structures of ternary complexes with the Pol bound to a 12mer template strand with the incorporated lesion and paired to an 8mer primer and incoming purine nucleoside triphosphate. For details regarding the crystallization, data collection and structure determination and refinement procedures please see the section **2.6**. Selected crystal data, data collection and refinement parameters and examples of the quality of the final electron density for all structures are summarized and depicted in the **Table S2.3**. The two complexes with O^4-MedT-containing templates and incoming dATP or dGMPNPP reveal similar orientations of the lesion (**Figure 2.3**, PDB ID codes 5DLF and 5DLG, respectively). Instead of pairing with the incoming nucleotide, O^4-MedT is lodged near the ceiling of the active site. Thus, its base portion is nestled against Trp-64 (base stacking interaction), Met-63 and Ser-62 (hydrophobic contacts between O^4-Me and both Cα and C(=O) from the two residues) from a loop region in the finger domain and Gly-46 from an adjacent β-strand that together form the roof of the active site (**Figure 2.3a,c**). In addition, O^2 of O^4-MedT and the amino group of Asn-38 are H-bonded (**Figure 2.3b,d**). This position of the O^4-MedT lesion at the entrance of the

active site places it quite far away from the incoming nucleotide triphosphate. The distance between its O^4-atom and N^6 of dATP is 9 Å (8.2 Å between O^4 of O^4-MedT and O^6 of dGMPNPP).

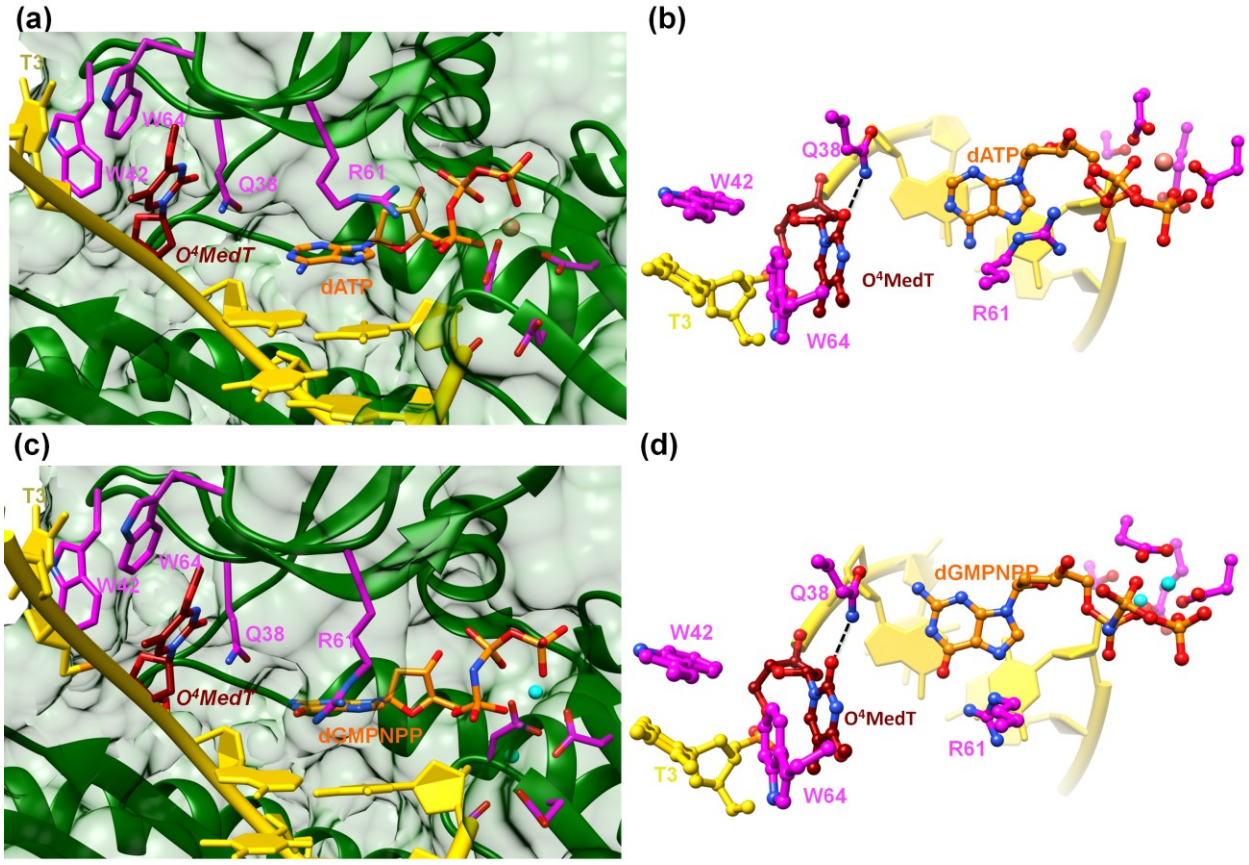


Figure 2.3: Detached arrangement of incoming purine nucleotide triphosphate and O^4-MedT in two hPol η insertion-stage complexes. (**a**), Active site conformation in the complex with dATP opposite O^4-MedT viewed into the DNA major groove, and (**b**), rotated by 90° and viewed perpendicular to the adenine plane. (**c**), Active site conformation in the complex with dGMPNPP opposite O^4-MedT, viewed into the DNA major groove, and (**d**), rotated by 90° and viewed perpendicular to the guanine plane. Carbon atoms of O^4-MedT, the incoming nucleotide triphosphate, and selected hPol η amino acid side chains are colored in maroon, orange and magenta, respectively. Mg^{2+} and Ca^{2+} ions are cyan and pink spheres, respectively, and selected H-bonds are shown as dashed lines.

Closer inspection of the positions of the incoming nucleotides shows that adenine stacks on the adjacent t(emplate)A:p(rimer)T pair, with the side chain of Arg-61 from the finger domain

hovering closely above the adenine plane and engaged in H-bonds to the α- and β-phosphates of dATP via its guanidino moiety (**Figure 2.3b**). By comparison, guanine is shifted into the minor groove and the stacking interaction with the adjacent tA:pT pair is slightly less favorable. The shift is likely a consequence of the altered orientation of the Arg-61 side chain that is extended in the structure of the complex with dGMPNPP, resulting in formation of H-bonds between its guanidino moiety and guanine O^6 and N7 (**Figure 2.3d**). The particular orientation of the incoming dG brings it closer to Asn-38 from the finger domain, but the distance of 3.64 Å between N2 of the former and the Oε1 oxygen of asparagine is slightly too long for formation of a H-bond.

A surface rendering of the hPol η active site in the O^4-MedT insertion-stage complex with the incoming dATP indicates that the nucleobase moiety of the lesion fits snugly into the gap between Trp-64 and Ser-62 (**Figure S2.16**). It is clear that O^4-EtdT (with the ethyl group in the *syn* conformation) cannot be accommodated in the same fashion, as the longer substituent would clash with residues from the finger domain. Instead, the O^4-EtdT lesion has been pulled inside the active site and pairs with incoming dAMPNPP or dGMPNPP *via* bifurcated H-bonds in the two crystal structures of insertion-stage complexes with this lesion (**Figure 2.4a-d**, PDB ID codes 5DQG and 5DQH, respectively). As in the complexes with O^4-MedT, Asn-38 forms a H-bond to O^2 of O^4-EtdT in the minor groove. However, unlike in the O^4-MedT complex with incoming dGMPNPP, the side chain of Arg-61 in the corresponding complex with O^4-EtdT does not adopt an extended conformation to contact the major groove edge of guanine. As can be seen in **Figure 2.4**, Arg-61 is directed toward the triphosphate moiety and forms a salt bridge with the α-phosphate group in both insertion-stage complexes. The methylene group (C1) of the O^4 substituent in O^4-EtdT adopts an *anti* conformation in the complex with dAMPNPP (torsion

angle C1-O4-C4-N3 = -142°) and a *syn* conformation in the complex with dGMPNPP (torsion angle C1-O4-C4-N3 = +62°). This is a clear difference to the structures of complexes with O^4-MedT, both of which show the lesion adopting a *syn* conformation (torsion angles C1-O4-C4-N3 of +33° and +25° in the dATP and dGMPNPP complexes, respectively). A further difference between the O^4-MedT and O^4-EtdT complexes is constituted by the orientations of template nucleotides 5′-adjacent to the lesions. In the former, A2 and T3 form a stack with Trp-42 outside the active site (**Figure 2.2**). In the O^4-EtdT complexes, A2 is located outside the active site and forms a stacking interaction with Trp-42. However, T3 sits inside the active site and stacks onto O^4-EtdT (dAMPNPP complex; **Figure 2.4a**). In the complex with dGMPNPP, T3 juts into the major groove (**Figure 2.4c**). Thus, neither orientation adopted by T3 in these complexes resembles that of O^4-MedT, lodged near the ceiling of the active site and stacked onto Trp-64.

2.2.5 Crystal Structure of a Ternary hPol η•DNA•dCTP Extension-Stage Complex with O^4-EtdT Paired Opposite Primer dA

The structure of a complex with O^4-EtdT paired to dA at the -1 position followed by template dG opposite incoming dCTP was determined at 2.05 Å resolution (**Figure 2.4e,f**, PDB ID codes 5DQI). The geometry of the O^4-EtdT:dA pair replicates that seen in the insertion complex with O^4-EtdT opposite incoming dAMPNPP (**Figure 2.4a,b** and **2.4e,f**). As in the case of the latter, the ethyl group has moved outside the thymine plane and adopts an *anti* orientation (torsion angle C1-O4-C4-N3 = -135°). The base pair itself adopts a Watson-Crick like geometry with formation of a single H-bond; the adjacent dG:dCTP pair displays a standard geometry with three H-bonds. Arg-61 is directed toward the phosphate moieties of the incoming nucleotide and forms two salt bridges with the α- and β-phosphate groups, and Asn-38 is engaged in two H-bonds with N3 and O4′ of template dG.

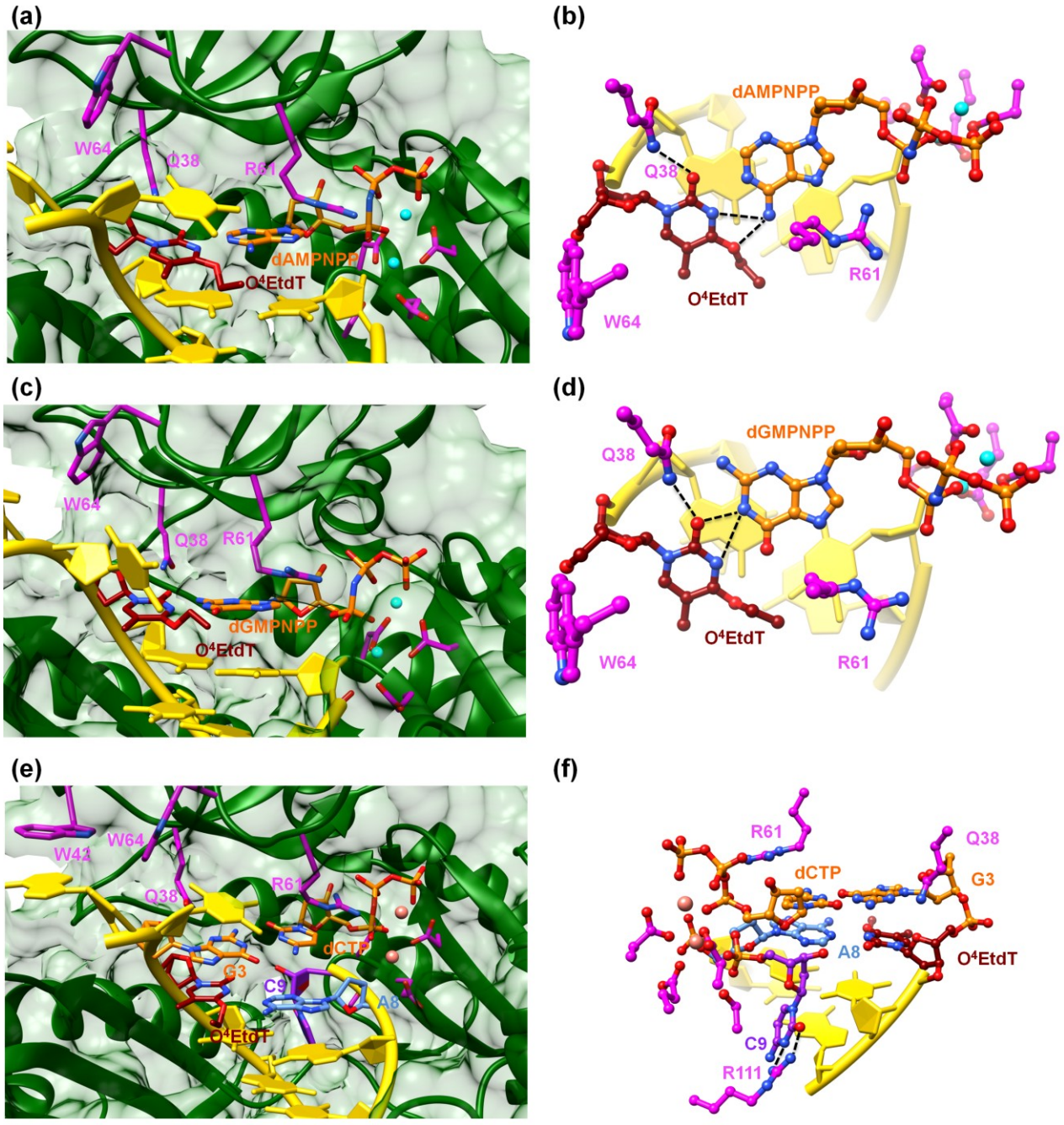


Figure 2.4: Pairing between incoming purine nucleotide triphosphate and O^4-EtdT in two hPol η insertion-stage complexes (**a** - **d**). (**a**), Major groove view of the Watson-Crick pair between dAMPNPP and O^4-EtdT with formation of bifurcated H-bonds. (**b**), The active site rotated by 90° relative to *A*, with the base pair viewed perpendicular to the adenine plane. (**c**), Major groove view of the sheared configuration between dGMPNPP and O^4-EtdT, with formation of bifurcated H-bonds. (**d**), The active site rotated by 90° relative to C, with the base

pair viewed perpendicular to the guanine plane. Carbon atoms of O^4-EtdT, the incoming nucleotide triphosphate, and selected hPol η amino acid side chains are colored in maroon, orange and magenta, respectively. Mg^{2+} ions are cyan spheres and selected H-bonds are shown as dashed lines. Active site configuration in a ternary hPol η extension-step complex with O^4-EtdT opposite primer dA, followed by template dG opposite incoming dCTP (**e - f**). (**e**), Major groove view of the active site, with carbon atoms of O^4-EtdT and the paired dA colored in maroon and light blue, respectively. Carbon atoms of the nascent dG:dCTP pair are colored in orange, carbon atoms of selected hPol η amino acid side chains are colored in magenta, and those of the 3'-terminal primer dC are colored in purple. (**f**), The active site rotated by 180° relative to *A* and viewed into the minor groove.

The most unusual feature of the extension-stage structure is the presence of an additional nucleotide at the 3′-end of the primer (**Figure 2.4f**). Because the crystallization solutions contained dCTP and the residual electron density is consistent with a pyrimidine, we extended the primer by dC (**Figure S2.17**). We suspect that hPol η possesses weak catalytic activity with Ca^{2+} as the cofactor or that traces of Mg^{2+} present in the crystallizations led to primer extension *in situ* (even a very low activity could result in extension by a single nucleotide over the course of two weeks). We showed earlier that the translesion DNA polymerase Dpo4 from *Sulfolobus solfataricus* is able to catalyze nucleotide insertion with Ca^{2+}, although the activity is far below that seen with Mg^{2+} as the prosthetic group.[237] The additional dC stacks against the backbone of the template strand in the minor groove and its position is further stabilized by two H-bonds between N3 and O2 and the guanidino moiety of Arg-111 (**Figure 2.4f**).

2.3 Discussion

The known toxicity of alkylated adducts at the O^4-position of thymidine prompted us to explore the influence of restricting orientation of the alkyl group around the C4-O^4 bond to an *anti* conformation in translesion synthesis catalyzed by hPol η. In these studies, the bicyclic pyrimidine analogs DFP and TTP, which link the C5 and O^4 atoms with a di- or trimethylene

linker, were evaluated in addition to O^4-MedT and the bulkier O^4-EtdT lesion. Conformationally locked analogs of damage that can occur at the nucleobase have been previously synthesized and employed in studies which have provided insights into the requirements for DNA repair processes.[137,228] In addition, DNA containing DFP and TPP were evaluated for AGT-mediated repair and experiments revealed no activity towards these modifications (data shown in **CHAPTER VI**, **Figure S4.17**).

UV thermal denaturation studies of oligomers containing DFP and TTP revealed similar influences on duplex stability to both complementary and mismatched nucleobases compared to O^4-MedT and O^4-EtdT. The most stable pairing of either DFP or TTP was with dG, also observed with O^4-MedT and O^4-EtdT. NMR studies of a duplex containing an O^4-MedT•dG pair revealed, in addition to the O^4-methyl group adopting a *syn*-conformation, that the base formed a Watson-Crick "like" pairing with a single hydrogen bond.[26] In this structure, the *syn*-orientation of the O^4-Me group influences the hydrogen bond between the imino proton of dG and the N3 atom of O^4-MedT by increasing the distance between the O^6 and O^4 atoms of dG and O^4-MedT, respectively. Limiting the orientation of the methylene group at the O^4-atom to the *anti*-conformation, in the case of the DFP and TTP modifications, appears to have a minimal impact on the interaction with dG and duplex stability. In pairing with dA, a similar drop in duplex stability compared with dT is observed for oligonucleotides containing the O^4-MedT, O^4-EtdT, DFP, and TTP modifications. The NMR structure of a duplex containing a O^4-MedT•dA pair indicated that the O^4-Me group is *syn* and that the bases adopt a wobble alignment with one hydrogen bond formed between the imino nitrogen of O^4-MedT and the amino group of dA.[25] The restricted *anti*-orientation of the methylene group for DFP or TPP modifications does not significantly impact duplex stability compared to O^4-MedT or O^4-EtdT.

Steady-state kinetics of individual nucleotide incorporation opposite the DFP and TTP modifications by hPol η demonstrated preferred insertion of purine nucleotides relative to the pyrimidines, similar to O^4-MedT and O^4-EtdT. The efficiency of nucleotide insertion (k_{cat}/K_m) for the correct nucleotide (dAMP) across the lesions followed the order DFP > TPP > O^4-MedT > O^4-EtdT. For dGMP, a similar efficiency of nucleotide insertion occurred for DFP, TPP, and O^4-MedT whereas a drop was observed for O^4-EtdT. In agreement with studies involving *Saccharomyces cerevisiae* DNA polymerase η (yPol η), a reduction in incorporation efficiency due to the presence of an O^4-MedT insert was observed.[47] However, whereas the yeast homolog revealed a significant preference for dGMP, which was incorporated approximately 80 times more efficiently than dAMP,[47] hPol η displayed almost equal selectivity at incorporating dAMP (f = 0.94) as dGMP opposite O^4-MedT. A comparable 80-fold preference for dGMP over dAMP was exhibited by yPol η for the bulkier O^4-carboxymethylthymidine lesion.[238] The rationale for the preferred incorporation of dGMP opposite O^4-MedT by yPol η was attributed to a dG•O^4-MedT wobble base pairing. Differences observed for nucleotide incorporation opposite O^4-MedT by the yeast and human homologs of Pol η may be influenced in part by different sequence contexts, as previously observed.[229,239] In addition, homologs of Pol η have exhibited differences in nucleotide incorporation across some types of DNA damage. For example, yPol η accurately inserts dCMP across 8-oxodG whereas hPol η is less accurate inserting some dAMP across this lesion as well.[229,240] Interestingly, similar misinsertion profiles have been observed in bypass experiments of hPol η and yPol η with O^6MedG, a lesion which also protrudes in the major groove of the DNA duplex.[241]

For the bulkier O^4-EtdT and conformationally restricted analogs DFP and TTP, a preference for nucleotide incorporation of dAMP over dGMP was observed. For O^4-EtdT, hPol η

was more proficient at incorporating dAMP over dGMP with catalytic efficiencies of 0.10 and 0.06 μM^{-1} s^{-1}, respectively. These values are approximately two-fold lower compared to those observed for the O^4-MedT-containing template, but can be rationalized by the increased bulk of the ethyl group, which may influence dNTP incorporation in the hPol η active site. Ethylation of the O^4-position of dT has been shown to stall the human Y-family DNA polymerases hPol κ and hPol ι but not hPol η (although steady-state analysis was not reported for oligonucleotides containing O^4-EtdT).[217] Bypass of O^4-EtdT by hPol η revealed dGMP misincorporation at 55 % compared to 19 % for dAMP insertion, in agreement with our data despite different sequence contexts.

For the conformationally restricted DFP and TPP modifications, incorporation efficiency was observed to be ~1.5- fold higher for dAMP (0.27 and 0.24 μM^{-1} s^{-1}, respectively) and comparable for dGMP relative to O^4-MedT. These results demonstrate that hPol η is more proficient at incorporating both the correct (dAMP) and incorrect (dGMP) nucleotides across from these more conformationally restricted lesions. In addition, the increase in steric bulk from the DFP to TPP slightly decreases incorporation efficiency. Exposure of the hydrogen bonding face of the DFP or TPP modifications may have a greater influence on stabilizing the wobble alignment geometry that has been suggested for the O^4-MedT•dA pairing. The conformational restriction of the alkyl group to an *anti*-orientation around the C4-O^4 bond, as in the DFP and TPP modifications, would direct the O^4-methylene group away from the amino group of dA, which could account for the enhancement of incorporation of the correct nucleotide (dAMP) compared to O^4-MedT. Incorporation of dGMP may not be as influenced by orientation of the alkyl group around the C4-O^4 bond as the proposed hydrogen bonding interaction, based on the NMR structure of the duplex containing the O^4-MedT•dG, which occurs between the amino

group of dG and O^2-atom of O^4-MedT.[26] In the case of O^4-EtdT, the combination of the *syn*-orientation and the size of the ethyl group may both contribute to the reduced efficiency of nucleotide insertion of dAMP and dGMP in this series.

Primer extension reactions in the presence of all four dNTPs for templates containing the O^4-MedT, O^4-EtdT, DFP, and TTP modifications demonstrated that hPol η was proficient at incorporating nucleotides across and past the adducted site. However, both O^4-MedT and O^4-EtdT exhibited a greater accumulation of non-full length oligonucleotide products at reduced reaction times (30 and 60 min), which was not observed for bicyclic DFP or TPP analogs. The LC-MS/MS analyses of the extension products from the *in vitro* primer bypass studies revealed that dGMP incorporation across the lesion was preferred over dAMP in all cases except the control (dT). The ratio of dGMP:dAMP incorporation by hPol η, assessed from the extension products, was found to decrease in the series TPP (4.6:1) > DFP (3.3:1) ≈ O^4-EtdT (3.2:1) > O^4-MedT (2.1:1). The presence of the larger alkyl group for O^4-EtdT or the analogs with the O^4-methylene group in an *anti*-conformation (TPP and DFP) clearly promotes dGMP misincorporation in the presence of all four nucleotides. In addition, the DFP and TPP modification were not found to induce a significant amount of frameshifts in the products compared to O^4-MedT. Differences in fidelity observed between the steady-state kinetic and LC-MS/MS full-length experiments have been observed previously.[45,47] The variance may be attributed to accommodation of the incoming dGTP relative to dATP for these modifications, highlighting that adduct size and the conformation of the O^4-methylene group can influence interactions in the active site of hPol η. In the case of the analogs investigated, hPol η continued extension of the primer in an error-free manner after incorporation of dATP or dGTP across from the damaged site on the template.

Several observations based on nucleotide incorporation profiles attest to the distinct effects on hPol η bypass synthesis exerted by the O^4-MedT and O^4-EtdT lesions. These concern (i) the more error-prone bypass caused by O^4-MedT, *i.e.* dGTP is favored relative to dATP (**Figure 2.2a**), (ii) increased accumulation of the +1 product in the full-length extension reaction for O^4-MedT (**Figure 2.1**), and (iii) significantly more frameshift products caused by the O^4-MedT lesion (**Figure 2.2b**). Interestingly, the structural data for insertion-stage hPol η complexes with either O^4-MedT or O^4-EtdT in the template strand reveal starkly different orientations of the two adducted nucleotides at the active site. O^4-MedT is trapped in an orientation that keeps it at a considerable distance from the incoming purine nucleotide triphosphates. Conversely, O^4-EtdT pairs opposite both dATP and dGTP with formation of bifurcated H-bonds (whereby the latter pair features a sheared orientation of the two partners, with G being pushed toward the minor groove). The increased proclivity for insertion of dG opposite O^4-MedT compared to O^4-EtdT is not surprising if one considers the strict preference by the O^4-methyl substituent for a *syn* conformation. The *syn* conformation precludes adoption of an O^4-MedT:dA pair with standard Watson-Crick geometry, but the sheared pairing mode seen in the case of O^4-EtdT:dG(MPNPP) (**Figure 2.4c,d**), also presumably adopted by the O^4-MedT:dG pair, is compatible with a *syn* conformation of the substituent. This conclusion is borne out by the observations that the ethyl moiety in the O^4-EtdT:dA(MPNPP) pairs assumes an *anti* conformation (**Figs. 2.4a,b,e,f**), whereas its conformation is *syn* in the O^4-EtdT:dGMPNPP pair. Furthermore, the TPP adduct opposite dGMPNPP was modeled from the O^4-EtdT:dG(MPNPP) ternary crystal structure coordinates. The configuration of the adduct seen in the model is consistent with the enhanced incorporation of dG observed in the primer extension experiments since the constrained *anti* conformation of the bicyclic system does not hinder the guanine

nucleobase from shifting towards the major groove and potentially form two H-bonds with TPP (**Figure S2.18**).

On one hand, one could argue that the higher fraction of frameshifts for O^4-MedT relative to O^4-EtdT is consistent with the structural data that show the former is not engaged opposite the incoming nucleotide but trapped adjacent to the 'entrance' of the active site. Perhaps the more pronounced accumulation of the +1 product in the case of the full-length extension reactions opposite O^4-MedT compared to the other O^4 adducts tested here are the result of non-templated insertion. Thus, purine nucleoside triphosphates would be favored and their incorporation would not be affected by the particular conformation of the O^4-methyl group, *syn* or *anti*. This scenario is certainly not inconsistent with the structural data that reveal no interaction between the O^4-MedT lesion and the incoming dATP or dGMPNPP. Clearly, it is intriguing that both activity and structural data show distinct consequences of the O^4-MedT and O^4-EtdT lesions for bypass by hPol η. However, it is important to note that the position of the O^4-MedT lesion at the active site, unique among all crystal structures of hPol η complexes analyzed to date, represents one state during bypass. Perhaps other orientations and interactions of the adducted nucleotide occur during bypass, which precludes all steps involved in the mechanism of O^4-alkyl bypass synthesis by hPol η.Yet, we have provided intriguing insights on a potentially different bypass mechanism of O^4-MedT in comparison to the larger O^4-EtdT adduct.

2.4 Conclusions

Oligonucleotides containing DFP and TPP, designed as analogs of O^4-alkylated thymidine, were synthesized to explore the influence of limiting the O^4-alkyl lesion to an *anti*-orientation on nucleotide incorporation by hPol η. These modifications were shown to destabilize the DNA duplex, based on UV thermal denaturation studies, regardless of the base-pairing

partner (A, G, T, or C), similar to O^4-MedT and O^4-EtdT. Primer extension assays demonstrated that these pyrimidyl modifications hindered nucleotide incorporation by hPol η. Single nucleotide incorporation studies revealed increased selectivity towards dAMP over dGMP that followed the order O^4-EtdT > DFP ≈ TPP. A slight preference for dGMP over dAMP incorporation was observed for O^4-MedT. LC-MS/MS analysis of primer extension studies (in the presence of all four dNTP) revealed that hPol η incorporated dGMP over dAMP across the lesions in the order TPP > DFP ≈ O^4-EtdT > O^4-MedT. These trends suggest that limiting the orientation of the O^4-alkylene group enhances the proficiency of dNTP incorporation by hPol η across O^4-alkylated dT damage. In the presence of all four dNTP, error-prone nucleotide incorporation by hPol η is enhanced by restricting the O^4-lesion to an *anti*-orientation. This study exemplifies how restricting a lesion's conformational freedom impacts bypass profile by hPol η. Moreover, our results supply mechanistic insights on the mutagenicity of the biologically relevant DNA damage O^4-MedT and O^4-EtdT.

2.5 Acknowledgements

We thank the Natural Sciences and Engineering Research Council of Canada (Grant No. 299384-2011 to C.J.W.), the Canada Research Chair Program (Grant No. 950-213807 to C.J.W.) and the US NIH for financial support (Grants No. ES010375 to F.P.G. and M.E., Grant No. CA160032 to M.E., and ES010546 to F.P.G.). D.K.O. thanks NSERC for a Canada Graduate Scholarship (CGS) and Michael Smith Foreign Study Supplement Scholarship, and the NSERC CREATE Program in Bionanomachines for support of an exchange with Vanderbilt University. Vanderbilt University is a member institution of the Life Sciences Collaborative Access Team at sector 21 of the Advanced Photon Source, Argonne, IL. Use of the Advanced Photon Source at

Argonne National Laboratory was supported by the United States Department of Energy, Office of Basic Energy Sciences (grant DE-AC02-06CH11357).

2.6 Supporting information

2.6.1 Supplementary Discussion

2.6.1.1 Synthesis and characterization of nucleosides.

The incorporation of O^4-MedT and O^4-EtdT in DNA oligomers has been described previously.[242] The chemical synthesis for phosphoramidites **2.6a** and **2.6b** is shown in **Scheme S2.1**. Precursors **2.1a** and **2.1b** were prepared according to published procedures.[243–247] Modified heterocycles **2.2a** and **2.2b** were obtained by a base-catalyzed intramolecular reaction under protic conditions.[244,245] Attachment of heterocycles **2.2a** or **2.2b** to commercially available 1-(α)-chloro-3,5-di-*O*-(*p*-toluoyl)-2-deoxy-D-ribose (Berry and Associates) by an uncatalyzed Vorbrüggen-type coupling yielded compounds **2.3a** and **2.3b** in moderate yields.[248] Resolution of the mixture of anomers for the latter coupling was challenging. Column chromatography using solvent systems such as CH_3OH:CH_2Cl_2, EtOAc:hexanes, and CH_3CN:CH_2Cl_2 with shallow gradients afforded low, if any, resolution of the anomeric mixture. The desired *β*-anomer was resolved by recrystallization from boiling EtOAc (1:1 EtOAc:hexanes, v/v for **2.3a** and 100% EtOAc for **2.3b**). Resolution of the anomers could also be accomplished by first protecting the free alcohol functionality with a *tert*-butyldimethylsilyl (TBS) group, followed by column chromatography (EtOAc:hexanes mobile phase) to separate the anomers. Subsequent removal of the TBS group supplied pure **2.3a** and **2.3b**. The bicyclic pyrimidines **2.4a** and **2.4b** were synthesized by first mesylating the terminal hydroxyl group followed by intramolecular cyclization in dilute basic conditions under reflux. A saturated solution of ammonia in methanol

removed the toluoyl groups followed by protection of the 5′-hydroxyl group with DMTr to yield mono-protected nucleosides **2.5a** and **2.5b**, which were then phosphitylated with *N*,*N* diisopropylaminocyanoethylphosphonamidic chloride in the presence of diisopropylethylamine to produce phosphoramidites **2.6a** and **2.6b**. ^{31}P NMR signals were observed at 148.27, 148.11 ppm for **2.6a** and 148.30, 148.03 ppm for **2.6b**, indicating the presence of the non-hydrolyzed phosphoramidite group.

2.6.1.2 Synthesis of DNA containing single modification inserts.

The solid-phase syntheses of **DFP** and **TPP** containing oligonucleotide sequences were performed on a 1.5 μmol scale using either phosphoramidite **2.6a** or **2.6b**, respectively. Given the potential lability of the lesions, as seen with other O^4-alkyl pyrimidine adducts,[150] "fast-deprotecting" 3′-*O*-2′-deoxyphosphoramidites were employed (dissolved to a concentration of 0.1 M in anhydrous acetonitrile) with phenoxyacetic anhydride used as the capping agent to prevent an undesired N-acetylation reaction that could occur with acetic anhydride.[249] Phosphoramidites **2.6a** or **2.6b** were dissolved in acetonitrile to a concentration of 0.12 M with a coupling time of 600 s (as opposed to 120 s) to ensure optimal production of the desired full-length oligonucleotides.

An ultra-mild oligonucleotide deprotection protocol was required due to the labile nature of the O^4-alkylpyrimidine adducts.[150] These conditions employed treatment of the support bound oligonucleotide with a 0.05M solution of K_2CO_3 in anhydrous CH_3OH for 3.5 h at room temperature in the dark. The resulting crude DNA solution was neutralized with acetic acid to prevent oligonucleotide damage during the evaporation of the solvent in the centrifugal concentrator. The resulting pellet was taken up in NaOAc (0.1M), desalted, and reconcentrated. The DNA sequences were purified by SAX-HPLC (see **Figure S2.1** for the HPLC

chromatographs) and subsequently desalted. ESI mass spectrometry of representative oligonucleotides $\mathbf{S_{DFP}}$ and $\mathbf{S_{TPP}}$ (with the sequence 5'-GGCTXGATCACCAG-3' where X is **DFP** and **TPP** for $\mathbf{S_{DFP}}$ and $\mathbf{S_{TPP}}$, respectively) revealed they had molecular weights of 4276.3 and 4290.8 Da (expected 4276.8 and 4289.9 Da), consistent with the expected values (see **Figures S2.2** and **S2.3** for the mass spectra). For longer deprotection time under the same conditions, ESI-MS revealed the formation of a product with a mass 32 Da higher than expected. This was most-likely due to the attack of the methoxide nucleophile on the bicyclic pyrimidine moieties during the deprotection step. The composition of $\mathbf{S_{DFP}}$ and $\mathbf{S_{TPP}}$ was further confirmed by enzymatic digestion to their constituent nucleosides with a mixture of snake venom phosphodiesterase and calf intestinal phosphatase followed by C_{18} RP-HPLC analysis (see **Figures S2.4** and **S2.5** for the chromatographs). The presence of the **DFP** or **TPP** nucleosides was established by the appearance of one additional peak to the four standard nucleosides, which eluted from the HPLC column with a retention time of 7.7 min for the **DFP** nucleoside and 9.1 min for **TPP** nucleoside. The later elution of the **TPP** relative to the **DFP** nucleoside can be attributed to the presence of additional methylene linkage in the former which increases its hydrophobicity. UV scans of **DFP** and **TPP** revealed that they had a reduced absorbance at 260 nm (the wavelength at which absorbance is monitored by the HPLC detector to detect the elution of nucleosides from the C_{18} column) relative to the four other nucleosides. Thus the C_{18} RP-HPLC experiments were simultaneously monitored at both 260 and 300 nm for the modified and standard nucleosides.

2.6.1.3 UV thermal denaturation of DNA duplexes and mismatch studies.

The influence of the **DFP** and **TPP** modifications on duplex stability was assessed by UV thermal denaturation experiments involving oligonucleotides $\mathbf{S_{DFP}}$ and $\mathbf{S_{TPP}}$ hybridized to

complementary DNA (see **Figure S2.6**). The melting curves of the duplexes containing **DFP** or **TPP** were sigmoidal with a comparable hyperchromicity between them that was slightly higher relative to the control duplex. The melting temperatures observed for duplexes containing **DFP** or **TPP** with adenine as the base pairing partner on the complementary strand were 48 and 49 ^{o}C, respectively, a significant reduction relative to the control duplex (62^{o}C). Duplexes containing O^4MedT and O^4EtdT were found to have melting temperatures of 51 and 52 ^{o}C, respectively, comparable to the destabilizing influence of O^4-alkylated dT in DNA duplexes that has been reported previously.[250,251] The similar T_m values of duplexes containing **DFP** or **TPP** versus O^4-MedT or O^4-EtdT suggest that the rigidity of the bicyclic pyrimidine system bears little impact on thermal stability relative to adducts with free rotation around the C4-O^4 bond. O^4-MedT has been shown, through NMR[25] and computational studies,[252] to adopt a wobble base pairing motif when paired with a dA residue. Given the similar trends observed in the T_m values, it is reasonable to expect that **DFP** and **TPP** can engage in similar hydrogen bond interactions when paired with adenine.

The influence of other base pairings involving the **DFP** and **TPP** modifications were also studied by UV thermal denaturation experiments. The T_m values of duplexes containing these modifications were equivalent to O^4-MedT or O^4-EtdT when mispaired against cytosine or thymine. An increase in T_m was observed for all the O^4-alkylated thymidine analogs with guanine as the base pairing partner relative to the other pairings. These findings for duplexes containing the bicyclic **DFP** and **TPP** modifications are consistent with studies involving O^4-alkylated dT·dG mispairs.[242,250] Interestingly, previous studies with duplexes containing O^4-MedT and O^4-EtdT modifications reported biphasic thermal profiles with purine mispairs.[250,253] However, this behavior was not observed in this study, perhaps due to differences with the

sequence investigated. Cruzeiro-Hansson and Goodfellow's computational analysis showed that the methyl group preferred to adopt the *syn* conformation for the O^4-MedT·dG pair.[252] In this orientation, there are three hydrogen bonds between guanine and methylated thymine, namely that of the N^2H-O^2 (strong), N1H-O^2 (weak), and N1H-N3 (weak), respectively. The bicyclic constructs of **DFP** and **TPP** inherently lock the methylene group in the *anti* conformation, however, hydrogen bonding interactions between the same atoms as the O^4-MedT·dG pair is possible.

2.6.1.4 Circular dichroism spectroscopy and molecular modeling of DNA duplexes.

CD spectra of duplexes composed of **DFP** and **TPP** paired against adenine on the complementary strand were acquired at 15 ^oC. The spectra (see **Figure S2.7**) displayed typical B-form DNA signatures with a wavelength maximum of 280 nm, an intercept at approximately 260 nm, and a minimum around 250 nm, similar to the DNA duplex control (containing a dT·dA base pair). This suggests that the presence of the bicyclic pyrimidine moiety had a minimal influence on the global DNA structure, similar to what has been observed previously in NMR studies of duplexes containing O^4-MedT.[25]

Hybridized oligomers containing a dT·dA, O^4-MedT·dA, O^4-EtdT·dA, **DFP**·dA or **TPP**·dA pair were constructed using Hyperchem® 7.5 software package starting from the canonical B-form duplexes. These were subject to geometry optimization using AMBER force field (see **Figures S2.8** and **S2.9** for structures of the duplexes with the water molecules omitted for clarity and views of the modified and flanking base pairs from the top). Both optimized structures containing the O^4-MedT·dA and O^4-EtdT·dA base pairs had very similar features with the alpha carbon adopting a *syn* conformation with respect to the N3-atom, in agreement with

published structural data by NMR and results from computational studies.[25,252] The **DFP** and **TPP** modifications behaved similarly to the O^4-MedT and O^4-EtdT. The duplexes exhibited minor global structural perturbations from the control duplex, namely retaining key features of B-form DNA. On a local level, the **DFP** and **TPP** modifications engaged with only one hydrogen bond contact between the N6H of adenine and O^4 of the modified pyrimidine at the Watson-Crick face. The modified pyrimidine bases were slightly shifted towards the major groove of the duplex to accommodate the bulk of the bicyclic construct. These slight distortions could explain the decrease in T_m values and increase in hyperchromicity changes observed. In general, the molecular model of **DFP** and **TPP** modifications do not seem to induce any gross geometrical distortions on the global structure, as corroborated by the circular dichroism experiments.

2.6.2 Supporting Methods

2.6.2.1 General

1-(*α*)-Chloro-3,5-di-*O*-(p-toluoyl)-2-deoxy-D-ribose was purchased from Berry & Associates, Inc. (Dexter, MI). *N*,*N*-Diisopropylaminocyanoethylphosphonamidic chloride was purchased from ChemGenes Corporation (Wilmington, MA). 5-*O*-Dimethoxytrityl-2'-deoxyribonucleoside-3-*O*-(*β*-cyanoethyl-*N*,*N*-diisopropyl)phosphoramidites and protected 2′-deoxyribonucleoside-CPG supports were purchased from Glen Research (Sterling, Virginia). All other chemicals and solvents were purchased from the Aldrich Chemical Company (Milwaukee, WI) or EMD Chemicals (Gibbstown, NJ). Flash column chromatography was performed using silica gel 60 (230-400 mesh) obtained from Silicycle (Quebec City, QC). Thin layer chromatography (TLC) was carried out using precoated TLC plates (Merck, Kieselgel 60 F_{254}, 0.25 mm) purchased from EMD Chemicals Inc. (Gibbstown, NJ). NMR spectra were recorded

on a Varian 500 MHz NMR spectrometer at room temperature. ^{1}H NMR spectra were recorded at a frequency of 500.0 MHz and chemical shifts were reported in parts per million (ppm) downfield from tetramethylsilane. ^{13}C NMR spectra (^{1}H decoupled) were recorded at a frequency of 125.7 MHz and chemical shifts were reported in ppm with tetramethylsilane as a reference. ^{31}P NMR spectra (^{1}H decoupled) were recorded at a frequency of 202.3 MHz and chemical shifts were reported in ppm with H_3PO_4 used as an external standard. High resolution mass spectrometry of modified nucleosides were carried out using an LTQ Orbitrap Velos - ETD mass spectrometer (Thermo Scientific) at the Concordia University Centre for Biological Applications of Mass Spectrometry (CBAMS) or using a 7T-LTQ FT ICR mass spectrometer (Thermo Scientific) at the Concordia University Centre for Structural and Functional Genomics. The mass spectrometer was operated in full scan, positive ion detection mode. ESI mass spectra for oligonucleotides were obtained at CBAMS using a Micromass Qtof2 mass spectrometer (Waters) equipped with a nanospray ion source. The mass spectrometer was operated in full scan, negative ion detection mode.

2.6.2.2 Chemical synthesis of nucleosides.

Compounds **2.1a** and **2.1b** were prepared according to published procedures.[243–247]

5-Hydroxyethyluracil (**2.2a**)

To a solution of NaOEt (1.23 g, 18.1 mmol) in EtOH (90 mL) was added compound **2.1a** (7.55 g, 48.4 mmol). After 16h, the solvent was removed in vacuo and the crude was taken up in a minimal amount of boiling water. The solution was acidified (to a pH ~ 2-3) on an ice bath and the precipitate was isolated by vacuum filtration and washed with cold EtOH (3 × 40 mL). To maximize the yield, the filtrate was evaporated and the solid was resuspended in EtOH (50 mL). The solid was isolated and washed once more according to the procedure described above to give

2.2a in an overall yield of 7.16 g (94.8%) as a colorless powder. $\lambda_{max(MeCN)}$ 262 nm. ^{1}H NMR (500MHz, d_6-DMSO, ppm): 10.99 (broad s, 1H, NH), 10.64 (broad s, 1H, NH), 7.19 (s, 1H, H6), 4.52 (t, 1H, OH, J = 5.5 Hz), 3.44 (m, 2H, $\underline{CH_2}OH$), 2.30 (t, 2H, CH_2Ar, J = 6.5 Hz). ^{13}C NMR (125.7 MHz, d_6-DMSO, ppm): 164.6, 151.4, 138.8, 109.1, 59.44, 29.7. IR (thin film); ν_{max} (cm^{-1}) 3680, 2922, 2849, 2382, 2349, 1755, 1674, 1451, 1244, 1032. HRMS (ESI-MS) *m/z* calculated for $C_6H_8N_2O_3Na^+$ 179.0427: found 179.0435 $[M+Na]^+$.

5-Hydroxypropyluracil (**2.2b**)

To a solution of NaOEt (1.52 g, 22.4 mmol) in EtOH (225 mL) was added compound **2.1b** (4.76 g, 28.0 mmol). After 16h, the solvent was removed in vacuo and the crude was taken up in a minimal amount of boiling water. The solution was acidified (to a pH ~ 2-3) and cooled, and the precipitate was isolated by vacuum filtration and washed with cold EtOH (3 × 50 mL). To maximize the yield, the filtrate was evaporated and the solid was resuspended in EtOH (50 mL). The solid was isolated and washed once more according to the procedure described above to give **2.2b** in an overall yield of 4.75 g (>99%) as a colorless powder. $\lambda_{max(MeCN)}$ 261 nm. ^{1}H NMR (500 MHz, d_6-DMSO, ppm): 10.92 (broad s, 1H, NH), 10.54 (broad s, 1H, NH), 7.14 (s, 1H, H6), 4.63 (t, 1H, OH, J = 5.0 Hz), 3.35 (m, 2H, $\underline{CH_2}OH$), 2.17 (t, 2H, CH_2Ar, J = 8 Hz), 1.54 (m, 2H, $CH_2\underline{CH_2}CH_2$). ^{13}C NMR (125.7 MHz, d_6-DMSO, ppm): 164.9, 151.7, 138.1, 111.9, 61.7, 31.3, 23.1. IR (thin film); ν_{max} (cm^{-1}) 3680, 2922, 2849, 2381, 2349, 1755, 1674, 1451, 1244, 1032. HRMS (ESI-MS) *m/z* calculated for $C_7H_{10}N_2O_3Na^+$ 193.0584: found 193.0594 $[M+Na]^+$.

3′5′-O-bis-Toluoyl-5-hydroxyethyl-2′-deoxyuridine (**2.3a**)

To a solution of compound **2.2a** (2.4 g, 15.4 mmol) in HMDS (60 mL, 164 mmol) was added TMS-Cl (0.4 mL, 3.14 mmol) which was then stirred vigorously at 140 ^{o}C. After 5 h the

excess HMDS was removed in vacuo and the resulting gum was left on the high vacuum for 2 h. The gum was then dissolved in 1,2-dichloroethane (60 mL) and to this was added a solution of 1-(α)-chloro-3,5-di-*O*-(*p*-toluoyl)-2-deoxy-D-ribose (2.0 g, 5.1 mmol) in 1,2-dichloroethane (80 mL) while stirring at room temperature. After 16 h the solvent was removed in vacuo and the residue was taken up in CH_2Cl_2 (60 mL) then washed with 3% (aq) $NaHCO_3$ (w/v, 2 × 50 mL). The organic layer was dried over anhydrous Na_2SO_4, decanted and the solvent evaporated. Then, 1 M TBAF (in THF, 5 mL, 5 mmol) was added and the reaction stirred for 15 min. The solvent was removed and the mixture of anomers was purified *via* flash column chromatography using a CH_3OH:CH_2Cl_2 solvent system (2% → 5%, v/v). The *β*-anomer was isolated by precipitation from EtOAc: hexanes (1:1, v/v) at a concentration of 31 mM with stirring at room temperature for 10 min. The resulting suspension was filtered under vacuo to afford 1.15 g (44%) of **2.3a** as a colorless foam. R_f (SiO_2 TLC): 0.33 CH_3OH:CH_2Cl_2 (5%, v/v). $\lambda_{max(MeCN)}$ 241 nm. ^{1}H NMR (500MHz, d_6-DMSO, ppm): 11.37 (s, 1H, NH), 7.92-7.87(m, 4H, Ar), 7.48 (s, 1H, H6), 7.36-7.30 (m, 4H, Ar), 6.28 (t, 1H, H1', J = 8 Hz), 4.61 (m, 1H, H3'), 4.55 (m, 1H, H5'), 4.51 (m, 1H, H5''), 4.46 (m, 1H, H4'), 3.37 (m, 2H, $\underline{CH_2}OH$), 2.95 (m, 1H, H2'), 2.53 (m, 1H, H2''), 2.39 (s, 3H, CH_3Ph), 2.38 (s, 3H, CH_3Ph), 2.24 (t, 2H, CH_2Ar, J = 7.0 Hz). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 166.6, 166.5, 163.6, 150.8, 145.1, 145.0, 136.8, 130.4, 130.1, 130.0, 129.8, 127.1, 126.8, 112.9, 85.5, 83.2, 75.3, 64.7, 60.8, 38.5, 30.6, 22.3, 22.2. IR (thin film); ν_{max} (cm^{-1}) 3204, 3064, 2955, 2925, 1719, 1685, 1612, 1510, 1464, 1269, 1178, 1100, 1020. HRMS (ESI-MS) *m/z* calculated for $C_{27}H_{29}N_2O_8^+$ 509.1918: found 509.1935 $[M+H]^+$.

Alternative method for resolution of the anomers: To a solution of the purified mixture of anomers (3.7 g, 7.3 mmol) and imidazole (1.6 g, 23 mmol) in CH_2Cl_2 (80 mL) was added TBS-Cl (1.8 g, 12 mmol) under stirring. After 4h the solvent was removed in vacuo and the content

was taken up in DCM (50 mL), washed with 3% (aq.) $NaHCO_3$ (2 × 50 mL), dried over anhydrous Na_2SO_4, decanted and the solvent evaporated. Purification was performed by flash column chromatography using an EtOAc: Hex solvent system (1:4, then 1:3, v/v) to afford 3.1 g (65%) of the β-anomer and 0.82 g (17 %) of the α-anomer, both as colorless foams. To a solution of the *β*-anomer **2.3a** (2.6 g, 4.1 mmol) in THF (40 mL) was added 1 M TBAF (in THF, 4.9 mL, 4.9 mmol) which was heated to 45^{o}C under stirring. After 1 h the solvent was removed in vacuo and the residue was taken up in DCM (50 mL) then washed with 3% (aq., w/v) $NaHCO_3$ (2 × 50 mL). The organic layer was dried over anhydrous Na_2SO_4, decanted and the solvent was evaporated. Purification was achieved by flash column chromatography using a CH_3OH:CH_2Cl_2 solvent system (1% → 3%, v/v) to afford 2.0 g (96 %) of **3a** as a colorless foam.

3′,5′-O-bis-Toluoyl-5-hydroxypropyl-2′-deoxyuridine (**2.3b**)

To a solution of compound **2b** (4.5 g, 26.4 mmol) in HMDS (95 mL) was added TMS-Cl (0.73 mL, 5.74 mmol) while stirred at 135^{o}C. After 5h, the excess HMDS was removed in vacuo and the resulting gum was placed on the high vacuum for 1 h. To a solution of the gum in 1,2-dichloroethane (225 mL) was added 1-(*α*)-chloro-3,5-di-*O*-(*p*-toluoyl)-2-deoxy-D-ribose (3.5 g, 9.0 mmol) and the atmosphere was exchanged with Ar while stirring at room temperature. After 16h the solvent was removed in vacuo and the crude was taken up in CH_2Cl_2 (125 mL), washed with cold brine (50 mL) followed by cold 3% (aq., w/v) $NaHCO_3$ (2 × 50mL). The organic layer was dried over anhydrous Na_2SO_4, decanted and the solvent was evaporated. The product was purified *via* flash column chromatography (short column) using an isocratic CH_3OH:CH_2Cl_2 solvent system (4%, v/v). A mixture of silylated and desilylated products were isolated to which 1 M TBAF (in THF, 5 mL, 5 mmol) was added under stirring. After 15 min, the solvent was removed in vacuo and the mixture of desilylated anomers were purified by flash column

chromatography using a CH_2Cl_2:CH_3OH solvent system (2 % → 4 % by increments of 0.5 %, v/v). The *β*-anomer was isolated by precipitation from EtOAc at a concentration of 31 mM with stirring at room temperature for 10 min. The resulting suspension was filtered under vacuo to afford 2.05 g (44 %) of **2.3b** as a colorless foam. R_f (SiO_2 TLC): 0.41 CH_3OH:CH_2Cl_2 (1:19, v/v). $\lambda_{max(MeCN)}$ 242 nm. ^{1}H NMR (500MHz, $CDCl_3$, ppm): 7.92-7.89 (m, 4H, Ar), 7.28 (s, 1H, H6), 7.25-7.23 (m, 4H, Ar), 6.41 (dd, 1H, H1′, *J* = 8.5, 8.75 Hz), 5.61 (m, 1H, H3′), 4.78 (dd, 1H, H5′, *J* = 3.5, 12.5 Hz), 4.63 (dd, 1H, H5′, *J* = 3.5, 12 Hz), 4.52 (m, 1H, H4′), 3.44 (m, 2H, C$\underline{H_2}$OH), 2.43 (s, 3H, CH_3Ph), 2.42 (s, 3H, CH_3Ph), 2.29 (m, 1H, H2′), 2.21 (m, 2H, CH_2Ar), 2.13 (m, 1H, H2′′), 1.52 (m, 2H, $CH_2C\underline{H_2}CH_2$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 166.14, 166.12, 164.1, 150.4, 144.61, 135.4, 129.8, 129.7, 129.6, 129.5 , 129.4, 129.3, 126.5, 126.3, 115.1, 85.0, 82.3, 74.9, 64.2, 60.8, 60.4, 38.0, 32.0, 31.8, 22.7, 21.73, 21.67, 21.0, 14.2. IR (thin film); ν_{max} (cm^{-1}) 3495, 3191, 300, 2926, 1718, 1611, 1466, 1377, 1270, 1178, 1103, 1020. HRMS (ESI-MS) *m/z* calculated for $C_{28}H_{31}N_2O_8^+$ 523.2075: found 523.2099 $[M+H]^+$.

Alternative method for the resolution of anomers: To a solution of a purified mixture of anomers (2.0 g, 3.8 mmol) and imidazole (1.2 g, 17 mmol) in CH_2Cl_2 (38 mL) was added TBS-Cl (1.3 g, 8.4 mmol) while stirring. After 4h the solvent was removed in vacuo and the content was taken up in CH_2Cl_2 (50 mL) and washed with 3% (aq., w/v) $NaHCO_3$ (2 × 50 mL). The organic layer was dried over anhydrous Na_2SO_4, decanted and the solvent removed in vacuo. Purification was achieved by flash column chromatography using an EtOAc:hexanes solvent system (35:65, v/v) to afford 1.2 g (49 %) of the *β*-anomer and 0.68 g (28 %) of the *α*-anomer, both as colorless foams. To a solution of the *β*-anomer of **2.3b** (0.50 g, 0.79 mmol) in THF (4 mL) was added 1 M TBAF (in THF, 1.2 mL, 1.2 mmol) which was heated to 45 ^{o}C with stirring. After 4 h the solvent

was removed in vacuo and the content purified by flash column chromatography using a solvent system (1% → 3%, v/v) to afford 0.33 g (81 %) of **2.3b** a colorless foam.

3-(2′-Deoxy-3′-5′-O-bis(toluoyl)-β-D-ribofuranosyl)-5,6-dihydrofuro[2,3-d]pyrimidin-2(3H)-one (**2.4a**)

To a solution of compound **2.3a** (1.2 g, 2.3 mmol) in pyridine (25 mL) was added MsCl (240 μL, 3.1 mmol) dropwise at 0°C while stirring. After 5 h, the reaction was diluted with THF (300 mL) and DBU (1.2 mL, 8.0 mmol) was added and heated to a gentle reflux. After 16 h the solvent was removed in vacuo and the content was taken up in CH_2Cl_2 (100 mL), washed with brine (2 50 mL) and 3% (aq.) $NaHCO_3$ (2 x 50 mL). The organic layer was dried over anhydrous Na_2SO_4, decanted and the solvent was removed in vacuo to produce a yellow gum. The product was purified *via* flash column chromatography using a gradient of CH_2Cl_2:CH_3OH (1 % → 4 % increased by increments of 0.5 %, v/v) to afford 1.0 g (91%) of **2.4a** as a colorless foam. R_f (SiO_2 TLC): 0.45 CH_3OH:CH_2Cl_2 (5% , v/v). $\lambda_{max(MeCN)}$ = 241 nm. ^{1}H NMR (500 MHz, $CDCl_3$, ppm): 7.95 (d, 2H, Ar, J = 8.5 Hz), 7.87 (d, 2H, Ar, J = 8.5 Hz), 7.72 (s, 1H, H6), 7.31 - 7.23, (m, 4H, Ar), 6.41 (dd, 1H, H1′, J = 6.0, 14 Hz), 5.61 (m, 1H, H3′), 4.83 (m, 1H, H5′), 4.66 - 4.60 (m, 3H, H4′ and CH_2O), 4.55 (m, 1H, H5′′), 3.06 (m, 1H, H2′), 2.91 (m, 1H, CH_2Ar), 2.69 (m, 1H, CH_2Ar), 2.44 (s, 3H, CH_3Ph), 2.43 (s, 3H, CH_3Ph), 2.22 (m, 1H, H2′′). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 166.1, 166.0, 162.8, 150.2, 144.61, 144.60, 136.9, 129.9, 129.7, 129.6, 129.5, 129.39, 129.37, 129.3, 126.6, 126.3, 111.4, 85.0, 82.9, 74.9, 64.1, 42.6, 38.1, 30.4, 21.75, 21.73. IR (thin film); ν_{max} (cm^{-1}) 3196, 3061, 2923, 2851, 1718, 1611, 1577, 1465, 1377, 1270, 11178, 1101, 1020. HRMS (ESI-MS) *m/z* calculated for $C_{27}H_{27}N_2O_7^+$ 491.1813: found 491.1797 $[M+H]^+$.

3-(2′-Deoxy-3′-5′-O-bis(toluoyl)-β-D-ribofuranosyl)-3,5,6,7-tetrahydro-(2H)-pyro[2,3-d]pyrimidin-2-one (**2.4b**)

To a solution of compound **2.3b** (1.5 g, 2.9 mmol) in CH_2Cl_2 (30 mL) was added NEt_3 (0.80 mL, 5.74 mmol) which was allowed to stir at 0 ^oC for 15 min. MsCl (0.30 mL, 3.9 mmol) was added dropwise over 10 min and cooled to -20 ^oC for 1 h without stirring. The reaction was diluted with CH_2Cl_2 (100 mL) and washed with ice/water (50 mL), ice/brine (50 mL) and ice/3% aq. (w/v) $NaHCO_3$ (50 mL). The organic layer was then dried over anhydrous Na_2SO_4, decanted and the solvent was removed in vacuo to produce a yellow gum. The gum was dissolved in THF (400 mL), DBU (0.40 mL, 2.7 mmol) was added and the reaction was gently refluxed under stirring. After 3 h the solvent was removed in vacuo and the crude was taken up in CH_2Cl_2 (150 mL), washed with ice/brine (50 mL) and 3% aq. $NaHCO_3$ (w/v, 2 × 50mL). The organic layer was dried over anhydrous Na_2SO_4, decanted and the solvent was removed in vacuo to produce a yellow gum. Purification was achieved by flash column chromatography using a CH_3OH:CH_2Cl_2 solvent system (3 → 4 %, v/v) to afford 1.27 g (88%) of **2.4a** as a colorless foam. R_f (SiO_2 TLC): 0.33 CH_3OH:CH_2Cl_2 (6 %). $\lambda_{max(MeCN)}$ 239 nm. ^1H NMR (500 MHz, $CDCl_3$, ppm): 7.96 - 7.94 (d, 2H, Ar, J = 8.5 Hz), 7.87 - 7.85 (d, 2H, Ar, J = 8.5 Hz), 7.67 (s, 1H, H6), 7.28 - 7.24 (m, 4H, Ar), 6.41 (dd, 1H, H1′, J = 5.5, 8.0 Hz), 5.61 (m, 1H, H3′), 4.84 (dd, 1H, H5′, J = 4, 13 Hz), 4.64 - 4.60 (m, 2H, H4′ and H5′′), 4.33 (m, 1H, CH_2O), 4.26 (m, 1H, CH_2O), 3.10 (dd, 1H, H2′, J = 5, 14.5 Hz), 2.44 (s, 3H, CH_3Ph), 2.42- 2.38 (m, 4H, 1H from CH_2Ar and CH_3Ph), 2.23 (m, 1H, H2′′), 2.15 (m, 1H, CH_2Ar), 1.87 (m, 1H, $CH_2\underline{CH_2}CH_2$), 1.79 (m, 1H, $CH_2\underline{CH_2}CH_2$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 170.0, 166.2, 166.0, 155.3, 144.55, 144.52, 140.4, 129.8, 129.5, 129.41, 129.37, 129.3, 129.2, 126.5, 126.3, 101.4, 87.5, 83.7, 75.3, 68.5, 64.3, 39.6, 21.72,

21.69. 21.6, 21.5. IR (thin film); ν_{max} (cm^{-1}) 3357, 2923, 2851, 2362, 2336, 1719, 1667, 1611, 1520, 1337, 1271, 1103, 1020. HRMS (ESI-MS) *m/z* calculated for $C_{28}H_{29}N_2O_7^+$ 505.1969: found 505.1985 $[M+H]^+$.

3-(2′-Deoxy-5′-O-(4,4′-dimethoxytrityl)-β-D-ribofuranosyl)-5,6-dihydrofuro[2,3-d]pyrimidin-2(3H)-one (**2.5a**)

Compound **2.4a** (0.59 g, 1.2 mmol) was allowed to stir at 0 ^oC in a solution of saturated methanolic ammonia (250 mL) and gradually allowed to warm up to room temperature. After 5 h the solvent was removed in vacuo and the resulting residue was washed with Et_2O (2 × 25mL). The deprotected nucleoside was partially purified by flash column chromatography (using a short column) with a gradient of CH_3OH:CH_2Cl_2 (4 % →15 %, v/v) to elute the product as a colorless powder. This powder was taken up in pyridine (15 mL) and stirred at 0 ^oC for 15 min. To this solution was added DMAP (cat) and DMT-Cl (0.38 g, 1.1 mmol) in small portions over 20 min under stirring. After 16 h, the solvent was removed in vacuo and the content was taken up in CH_2Cl_2 (100 mL), washed with 3 % (w/v) aqueous solution of $NaHCO_3$ (2 × 100 mL) and brine (100 mL). The organic layer was dried over anhydrous Na_2SO_4, decanted and the solvent was removed in vacuo to produce a yellow gum. The product was purified *via* flash column chromatography using a gradient of CH_3OH:CH_2Cl_2 (2% → 4%, v/v) to afford 0.32 g (48%) of **2.5a** as a colorless foam. R_f (SiO_2 TLC):0.18 CH_3OH:CH_2Cl_2 (5%, v/v). $\lambda_{max(MeCN)}$ 283 nm. ^1H NMR (500 MHz, $CDCl_3$, ppm): 8.08 (s, 1H, H6), 7.35 (d, 2H, Ar, *J* = 7.5 Hz), 7.28 - 7.19 (m, 7H, Ar), 6.81 (d, 4H, Ar, *J* = 9 Hz), 6.28 (t, 1H, H1′, *J* = 6.0 Hz), 4.61 - 4.50 (m, 3H, H3′ and CH_2O), 4.07 (m, 1H, H4′), 3.78 (s, 6H, OCH_3), 3.48 (dd, 1H, H5′, *J* = 3, 11 Hz), 3.42 (dd, 1H, H5’, *J* = 3, 10.5 Hz), 2.69 - 2.62 (m, 2H, H2′ and 1H from CH_2Ar), 2.49 - 2.41 (m, 1H, CH_2Ar), 2.32 (m, 1H, H2’′), 2.08 (d, 1H, OH, *J* = 4 Hz). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 177.9,

158.68, 158.67, 156.6, 144.5, 137.4, 135.3, 130.09, 130.06, 128.1, 128.0, 127.0, 113.29, 113.27, 103.9, 87.2, 86.9, 86.3, 71.7, 71.2, 63.0, 55.3, 42.3, 24.0. IR (thin film); ν_{max} (cm^{-1}) 3361, 2922, 2650, 1673, 1608, 1570, 1509, 1481, 1437, 1322, 1251, 1178, 1154, 1034. HRMS (ESI-MS) *m/z* calculated for $C_{32}H_{32}N_2O_7Na^+$ 579.2102: found 579.2100 $[M+Na]^+$.

3-(2′-Deoxy-5′-O-(4,4′-dimethoxytrityl)-β-D-ribofuranosyl)-3,5,6,7-tetrahydro-(2H)-pyro[2,3-d]pyrimidin-2-one (**2.5b**)

Compound **2.4b** (0.95 g, 1.9 mmol) was allowed to stir in a solution of saturated methanolic ammonia (500 mL) at 0°C and gradually allowed to warm to room temperature. After 5 h, the solvent was removed in vacuo and the resulting residue was washed with Et_2O (3 × 5 mL). The resulting gum was diluted with pyridine (18 mL) and stirred at 0 °C for 15 min. To this solution was added DMAP (cat) and DMT-Cl (0.38 g, 1.1 mmol) in pyridine (10 mL) dropwise over 25 min under stirring. The reaction was allowed to gradually warm up to room temperature. After 16h the solvent was removed in vacuo and the content was taken up in CH_2Cl_2 (50 mL) then washed with ice/brine (50 mL) followed by ice/3% (aq.) $NaHCO_3$ (w/v, 2 × 50 mL). The organic layer was then dried over anhydrous Na_2SO_4, decanted and the solvent was removed in vacuo to produce a yellow-orange gum. The product was purified *via* flash column chromatography using a gradient of CH_3OH:CH_2Cl_2 (3% → 5%, v/v) to afford 0.39 g (36%) of **2.5b** as a colorless foam. R_f (SiO_2 TLC): 0.18 CH_3OH:CH_2Cl_2 (5%, v/v). $\lambda_{max(MeCN)}$ 282 nm. ^1H NMR (500 MHz, $CDCl_3$, ppm): 8.05 (s, 1H, H6), 7.38 (d, 2H, Ar, *J* = 7.5 Hz), 7.31 - 7.21 (m, 7H, Ar), 6.83 (d, 4H, Ar, *J* = 9 Hz), 6.32 (t, 1H, H1′, *J* = 6.0 Hz), 4.49 (m, 1H, H3′), 4.30 (m, 2H, CH_2O), 4.11 (m, 1H, H4′), 3.80 (s, 6H, OCH_3), 3.50 (dd, 1H, H5′, *J* = 3, 10.5 Hz), 3.38 (dd, 1H, H5′′, *J* = 3, 10.5 Hz), 2.70 (m, 2H, H2′), 2.35 (m, 1H, H2′′), 2.21 (d, 1H, OH, *J* = 3.5 Hz), 2.15 (m, 1H, CH_2Ar), 1.87 (m, 1H, CH_2Ar), 1.81 (m, 2H, $CH_2\underline{CH_2}CH_2$). ^{13}C NMR (125.7 MHz,

d_6-DMSO, ppm): 170.0, 158.7, 155.0, 144.1, 142.1, 135.9, 135.7, 130.22, 130.16, 128.4, 128.1, 127.3, 113.8, 101.7, 86.4, 86.3, 86.2, 70.5, 68.7, 63.8, 55.54, 55.46, 41.5, 21.5, 21.2. IR (thin film); ν_{max} (cm^{-1}) 3360, 2921, 2850, 1661, 1632, 1522, 1509, 1454, 1336, 1250, 1177, 1117, 1033. HRMS (ESI-MS) *m/z* calculated for $C_{33}H_{35}N_2O_7^+$ 571.2439: found 571.2455 $[M+H]^+$.

3-(2'-Deoxy-3′-O-(β-cyanoethyl-N,N′ -diisopropylphosphite)-5′-O-(4,4'-dimethoxytrityl)-β-D-ribofuranosyl)-5,6-dihydrofuro[2,3-d]pyrimidin-2(3H)-one (**2.6a**)

To a stirred solution of compound **2.5a** (0.25 g, 0.45 mmol) and DIPEA (0.147 mL, 0.844 mmol) in THF (4.5 mL) was added dropwise Cl-POCEN*i*Pr_2 (0.149 mL, 0.674 mmol). After 30 min, the solvent was evaporated in vacuo and the content was diluted with EtOAc (50 mL) then washed with 3% (aq., w/v) $NaHCO_3$ (2 × 50 mL) and brine (50 mL). The organic layer was dried over anhydrous Na_2SO_4, decanted and the solvent was removed in vacuo to produce a yellow gum. Purification was achieved *via* short flash column chromatography using EtOAc:hexanes (4:1, v/v) (with 0.1% NEt_3, v/v) as eluent to afford 0.25 g (74 %) of **2.6a** as a colorless foam. R_f (SiO_2 TLC): 0.40, 0.31 EtOAc/hexanes (4:1, v/v). $\lambda_{max(MeCN)}$ 283 nm. ^{1}H NMR (500 MHz, d_6-acetone, ppm): 8.14-8.09 (m, 1H, H6), 7.53-7.50 (m, 2H, Ar), 7.40 - 7.25 (m, 7H, Ar), 6.95-6.92 (m, 4H, Ar), 6.33-6.26 (m, 1H, H1′), 4.83-4.77 (m, 1H, H3′), 4.64-4.56 (m, 2H, CH_2O), 4.26-4.20 (m, 1H, H4′), 3.94-3.47 (m, 12H, CH_2OP, NCH, H5′, H5′′ and OCH_3), 2.92-2.58 (m, 5H, CH_2CN, CH_2Ar & H2′), 2.43-2.36 (m, 1H, H2′′), 1.33-1.14 (m, 12H, 4 × CH_3). ^{13}C NMR (125.7 MHz, d_6-acetone, ppm): 177.89, 177.88, 158.9, 155.7, 144.95, 144.93, 136.9, 136.8, 135.7, 136.6, 135.5, 135.4, 130.2, 130.1, 128.2, 128.1, 127.90, 127.89, 126.91, 126.88, 118.1, 118.0, 113.2, 103.9, 103.8, 86.74, 86.71, 86.5, 86.4, 85.5, 85.2, 73.2, 71.5, 62.9, 62.7, 58.72, 58,65, 58.6, 58.5, 58.2, 54.7, 45.0, 44.9, 43.1, 43.0, 40.7, 40.5, 24.1, 24.02, 24.00, 23.92, 23.89, 22.29, 22.29, 22.2, 19.91, 19.85, 19.8, 19.5, 19.4. ^{31}P NMR (202.3 MHz, d_6-acetone, ppm):

148.27, 148.11. IR (thin film); ν_{max} (cm^{-1}) 2967, 2929, 2362, 2337, 1678, 1570, 1509, 1322, 1250, 1179, 1034. HRMS (ESI-MS) *m/z* calculated for $C_{41}H_{50}N_4O_8P^+$ 757.3361: found 757.3385 $[M+H]^+$.

3-(2′-Deoxy-3′-O-(β-cyanoethyl-N,N′-diisopropylphosphite)-5′-O-(4,4′-dimethoxytrityl)- β-D-ribofuranosyl)-3,5,6,7-tetrahydro-(2H)-pyro[2,3-d]pyrimidin-2-one (**2.6b**)

To a stirred solution of compound **2.5b** (0.25 g, 0.44 mmol) and DIPEA (0.143 mL, 0.816 mmol) in THF (4.3 mL) was added dropwise Cl-POCEN*i*Pr_2 (0.146 mL, 0.656 mmol). After 30 min the solvent was evaporated in vacuo and the content was diluted in EtOAc (50 mL) then washed with 3 % (w/v) aqueous solution of $NaHCO_3$ (2 × 50 mL) and brine (50 mL). The organic layer was dried over anhydrous Na_2SO_4, decanted and the solvent was removed in vacuo to produce a yellow gum. Purification was achieved *via* short flash column chromatography using EtOAc (with 0.1 % NEt_3, v/v) as eluent to afford 0.25 g (74%) of **2.6b** as a colorless foam. R_f (SiO_2 TLC): 0.20, 0.14 EtOAc. $\lambda_{max(MeCN)}$ 283 nm. ^{1}H NMR (500 MHz, d_6-acetone, ppm): 8.04-8.00 (m, 1H, H6), 7.54-7.48 (m, 2H, Ar), 7.41 - 7.25 (m, 7H, Ar), 6.96-6.89 (m, 4H, Ar), 6.33-6.27 (m, 1H, H1′), 4.80-4.73 (m, 1H, H3′), 4.40-4.11 (m, 4H, CH_2O, H4′, 1H from CH_2OP), 3.94-3.47 (m, 11H, 1 H from CH_2OP, NCH, H5′, H5′′and OCH_3), 2.96-2.38 (m, 4H, CH_2CN, H2', H2''), 2.29-2.24 (m, 1H, CH_2Ar), 2.09-1.97 (m, 1H, CH_2Ar), 1.87-1.79 (m, 2H, $CH_2\underline{CH_2}CH_2$), 1.34-1.11 (m, 12H, 4 × CH_3). ^{13}C NMR (125.7 MHz, d_6-acetone, ppm): 169.9, 169.82, 169.80, 158.93, 158.92, 154.81, 154.79, 144.9, 144.77, 144.76, 141.5, 141.4, 141.3, 135.69, 135.66, 135.6, 135.5, 135.4, 130.20, 130.18, 103,16, 130.1, 128.22, 128.18, 128.0, 127.91, 127.90, 127.0, 126.9, 113.23, 113.18, 101.4, 101.2, 101.1, 87.0, 86.72, 86.69, 86.3, 86.1, 85.6, 68.3, 68.2, 62, 9, 54.7, 43.1, 43.0, 24.1, 24.0,23.9, 22.31, 22.29, 21.50, 21.48, 21.3, 19.9, 19.4, 19.30, 19.27. ^{31}P NMR (202.3 MHz, d_6-acetone, ppm): 148.30, 148.03. IR (thin film); ν_{max}

(cm^{-1}) 2968, 2927, 2849, 1736, 1669, 1510, 1455, 1337, 1250, 1179, 1118, 1035. HRMS (ESI-MS) *m/z* calculated for $C_{42}H_{52}N_4O_8P^+$ 771.3517: found 771.3542 $[M+H]^+$.

2.6.2.3 Solid phase synthesis and purification of oligonucleotides.

All DNA sequences were synthesized using an Applied Biosystems Model 3400 synthesizer on a 1.5 μmol scale employing standard *β*-cyanoethyl phosphoramidite cycles supplied by the manufacturer with modifications to certain coupling times described below. Commercially available 3′-*O*-2′-deoxynucleoside phosphoramidites, containing "fast" deprotecting groups, were dissolved in anhydrous acetonitrile at a concentration of 0.1 M. The modified 3′-*O*-2′-deoxynucleoside phosphoramidites (**2.6a** and **2.6b**) were dissolved at a concentration of 0.12 M. Assembly of sequences began with detritylation (3% trichloroacetic acid (TCA) in CH_2Cl_2, v/v), followed by phosphoramidite coupling: Commercially available 3′-*O*-2′-deoxyphosphoramidites for 120 s and modified phosphoramidites (**2.6a** and **2.6b**) for 600 s. Capping with phenoxyacetic anhydride-pyridine-tetrahydrofuran (1:1:8; solution A) and 1-methyl-1*H*-imidazole-tetrahydrofuran (16:84 w/v; solution B) then oxidation (0.02 M iodine in tetrahydrofuran-water-pyridine, 2.5:2:1, v/v/v) followed every coupling. Removal of the 5′-terminal trityl group on the oligonucleotide was carried out on the synthesizer.

The oligomer-bound CPG support was removed from the column and transferred into screw cap microfuge tubes fitted with Teflon lined caps. The oligonucleotides containing the modifications 3-(2′-deoxypentofuranosyl)-5,6-dihydrofuro[2,3-*d*]pyrimidin-2(3*H*)-one (**DFP**) and 3-(2′-deoxypentofuranosyl)-3,5,6,7-tetrahydro-2*H*-pyrano[2,3-*d*]pyrimidin-2-one (**TPP**) were deprotected and cleaved from the solid support with mild deprotection conditions using freshly prepared anhydrous K_2CO_3 in CH_3OH (0.05 M) for 3.5 h at room temperature in the dark. The K_2CO_3 was neutralized with an equimolar amount of acetic acid prior to being

transferred into clean vials to separate the solution from the CPG. The CPG was rinsed twice with 250 μL aqueous CH_3CN (50%, v/v). The crude DNA was dried down in a centrifugal lypophilizer ("Speed vac") and then desalted using C-18 SEP PAK cartridges (Waters) prior to purification.

All oligonucleotide sequences were purified from pre-terminated products by strong anion exchange (SAX) HPLC using a Dionex DNAPAC PA-100 column (0.4 cm × 25 cm) purchased from Dionex (Sunnyvale, CA) with a linear gradient of 0-52% buffer B, v/v, over 24 min (buffer A: 100 mM Tris HCl, pH 7.5, 10% acetonitrile and buffer B: 100 mM Tris HCl, pH 7.5, 10% CH_3CN, 1 M NaCl) at 55 ^oC. The columns were monitored at 260 nm for analytical runs and/or 280 nm for preparative runs. The purified oligomers were desalted using C-18 SEP PAK cartridges (Waters).

2.6.2.4 Oligonucleotide characterization by ESI-MS and nuclease digestion.

Mass spectra for the oligonucleotides were acquired at the Concordia University Centre for Biological Applications of Mass Spectrometry (CBAMS) using a Micromass Qtof2 mass spectrometer (Waters) equipped with a nanospray ion source. The mass spectrometer was operated in full scan, negative ion detection mode and the raw data were deconvoluted.

The oligomers (0.05 A_{260} units / 0.38 nmol) were characterized by enzymatic digestion (snake venom phosphodiesterase: 0.28 units and calf intestinal phosphatase: 5 units) in buffer consisting of 10 mM Tris HCl, pH 8.1, and 2 mM $MgCl_2$ for 48 h at 37 ^oC. The resulting mixture of nucleosides was analyzed by reversed phase HPLC using a Symmetry® C-18 5 μm column (4.6 mm × 150 mm, Waters). A linear gradient of 0-70% buffer B (v/v) over 30 min was used to elute the analytes (buffer A: 50 mM sodium phosphate, pH 5.8, 2% CH_3CN (v/v) and buffer B: 50 mM sodium phosphate, pH 5.8, 50% CH_3CN (v/v)). The identity of the nucleosides was

verified by co-injection with the corresponding standards and eluted at the following times: dC (4.3 min), dG (6.7 min), dT (7.4 min), dA(8.5 min), modified bicyclic pyrimidyl nucleosides (7.7 min for the **DFP,** and 9.1 min for the **TPP** nucleosides), and the ratio of unmodified nucleosides was determined. The bicyclic pyrimidyl nucleosides showed poor absorbance at 260nm and improved absorbance at 300 nm. As a result, absorption of the modified nucleosides was monitored at 300 nm and their presence confirmed by an independent injection of the deprotected analogues of **2.4a** and **2.4b**, respectively. The molecular weights of the modified oligomers were determined by ESI-QToF MS and these were in agreement with the calculated values.

2.6.2.5 UV thermal denaturation studies of DNA duplexes.

Molar extinction coefficients for the modified oligonucleotides were calculated using the nearest-neighbour approximations (M^{-1} cm^{-1}) of the mononucleotides and dinucleotides. The molar extinction coefficients for thymidine were used for the modified nucleotides. All duplexes were prepared by mixing equimolar amounts of the interacting strands (0.5 A_{260} unit of the strand containing the modification / 3.8 nmol) and lyophilizing the mixture to dryness in a centrifugal concentrator under vaccum. The resulting pellet was dissolved in 90 mM NaCl, 10 mM sodium phosphate, and 1 mM EDTA buffer (pH 7.0) to give a final concentration of 3.8 μM duplex. Before the thermal run, samples were degassed in a centrifugal concentrator for 1 min. Samples were held at 90 ^{o}C for 5 min to ensure duplex melting. Annealing profiles were acquired at 260 nm at a rate of cooling of 0.5 ^{o}C min^{-1}, from 90 to 15 ^{o}C, using a Varian CARY Model 3E spectrophotometer equipped with a 6-sample thermostated cell block and a temperature controller. Samples were held at 15 ^{o}C for 2 min and re-heated at 0.5 ^{o}C min^{-1} to 90 ^{o}C showing

reversibility (data not shown). The data were analyzed according to the published procedure from Puglisi and Tinoco[254] and transferred to Microsoft Excel™ software.

2.6.2.6 Circular dichroism (CD) spectroscopy of DNA duplexes.

Circular dichroism spectra were acquired on a Jasco J-815 spectropolarimeter equipped with a Julaba F25 circulating bath as previously reported.[150] The spectra are an average of 5 scans acquired at a rate of 20 nm min^{-1}, with a bandwidth of 1 nm and a sampling wavelength of 0.2 nm in fused quartz cells (Starna 29-Q-10). Scans were performed between 350 and 220 nm at 15 °C. The molar ellipticity [θ] was calculated from the equation, $\theta = \varepsilon/Cl$, where ε is the relative ellipticity (mdeg), C is the concentration of the DNA duplex (M), and l is the path length (cm).

2.6.2.7 Molecular modeling of DNA duplexes.

Molecular modeling was performed with the Hyperchem 7.5 software package from Hypercube utilizing the AMBER force field. Hybridized oligomers containing a dT·dA, ***O*^{4}-MedT**·dA, ***O*^{4}-EtdT**·dA, **DFP**·dA, and **TPP**·dA base pair were constructed from the nucleic acid template option using a B-form duplex. Duplexes were solvated with water using a periodic box occupying about 3 times the volume of the duplex alone. Standard Amber99 parameters were used with the dielectric set to constant. "One to four scale factors" non-bonded interactions were set to 0.5 (both electrostatic and van der Waals). Cutoffs were applied to "switched" to an outer and inner radius of 13.5 and 9.5 Å, respectively. All structures were geometry optimized using Polak-Ribiere conjugate gradient until the RMS gradient was less than 0.1 kcal/(Å mol) using the periodic boundary condition option.

2.6.2.8 Steady-state kinetics with hPol *η*.

The 5′-FAM-labeled primer (5′-TCGTAAGCGTCAT-3′) and template DNA (3′-AGCATTCGCAGTAXTACT-5′, where X denotes the modified nucleotide) were mixed in a 1:1

molar ratio and annealed by heating to 90 ^{o}C, followed by slow cooling. Steady-state experiments were conducted with typical hPol η concentrations of 1.9-7 nM and 5 μM template-primer duplex substrates in 40 mM Tris HCl buffer (pH 7.5) containing 100 mM KCl, 5mM $MgCl_2$, 10 mM DTT, 5% glycerol (w/v), and 100 μg/ml bovine serum albumin (BSA), to which was added varying concentrations of a single dNTP (added last to initiate the reaction). Experiments were carried out at 37 ^{o}C and reactions were typically run for 5 to 15 min in order to keep product formation below 20% of the oligonucleotide substrate concentration.

Reactions were terminated with a quench solution containing formamide, EDTA, bromophenol blue and xylene cyanol, and aliquots were applied to an 18% (w/v) acrylamide/7.5 M urea gel and separated by electrophoresis.[255] Fluorescence in the substrate and product primer bands was monitored and quantified using a Typhoon system (GE Healthcare Life Sciences, Pittsburgh, PA) and the data were fit to hyperbolic plots (Michaelis-Menten equation) using the program GraphPad Prism (GraphPad, La Jolla, CA).

2.6.2.9 LC-MS/MS Analysis of Full-length Extension Products.

The 5′-FAM-labeled primer (5′-TCGTAAGCG*U*CAT-3′) and template (3′-AGCATTCGCAGTAXTACT-5′ where X denotes the modified nucleotide) were annealed as described above. 2′-Deoxyuridine (*U*) was included in the primer for facile cleavage of the product to a shorter oligonucleotide (by treatment with uracil DNA glycosylase followed by hot piperidine) that could be analyzed with an ion-trap mass spectrometer, as previously described.[229,236,256–258] DNA primer extension was accomplished by mixing hPol η (76 pmol, 0.95 μM) with template-primer duplex (2 nmol, 10 μM) and a mixture of 1 mM each of dATP, dCTP, dGTP, and dTTP at 37 °C for 1-1.5 h in 50 mM Tris-HCl buffer (pH 7.5), 50 mM NaCl, 5 mM DTT, 5 mM $MgCl_2$ and 50 μg/mL bovine serum albumin (BSA). The reactions were terminated

by spin column separations to extract the dNTPs and Mg^{2+}. The extent of the reaction was determined by electrophoresis/fluorography prior to LC-MS analysis. The resulting product was then treated with 25 units of uracil DNA glycosylase and 0.25 M piperidine.[229,236,256–258] To identify the products, the resulting reactions were analyzed by LC-MS/MS using an Acquity UPLC system (Waters) interfaced to a Thermo-Finnigan LTQ mass spectrometer (Thermo Scientific, San Jose, CA) equipped with a negative ion electrospray source. Chromatographic separation was achieved with an Acquity UPLC BEH octadecylsilane (C18) column (2.1 mm × 100 mm, 1.7 μm). The LC solvent system was as follows: Mobile phase A, 10 mM $CH_3CO_2NH_4$ in 98% H_2O; mobile phase B, 10 mM $CH_3CO_2NH_4$ in 90% CH_3CN (v/v). The following gradient (v/v) was used with a flow rate of 300 μL min^{-1} at a temperature of 50 ^oC: Linear gradient from 0-3% B (v/v) in 3 min, followed by a linear increase to 20% B (v/v) from 3-5 min, then 20-100% B (v/v) from 5-6 min which is held for 2 min. The column was re-equilibrated for 3 min with 100% A (v/v). MS conditions were as follows: Source voltage, 4 kV; source current 100 μA; capillary voltage,- 49 V; capillary temperature, 350 °C; tube lens voltage, -90 V. Product ion spectra were recorded over the range *m/z* 300-2000 and the most abundant species (-2 charge) was used for collision-induced dissociation (CID) analysis. The calculation for the oligonucleotide sequence CID fragmentation was carried out using the Mongo Oligo Mass Calculator v2.06 from The RNA Institute (College of Arts and Science, University at Albany State University of New York). The relative yields of various products were calculated based on the peak areas of extracted ion chromatograms from LC-MS analyses. The sum of the peak areas was used for multicharged species.

2.6.2.10 Crystallization

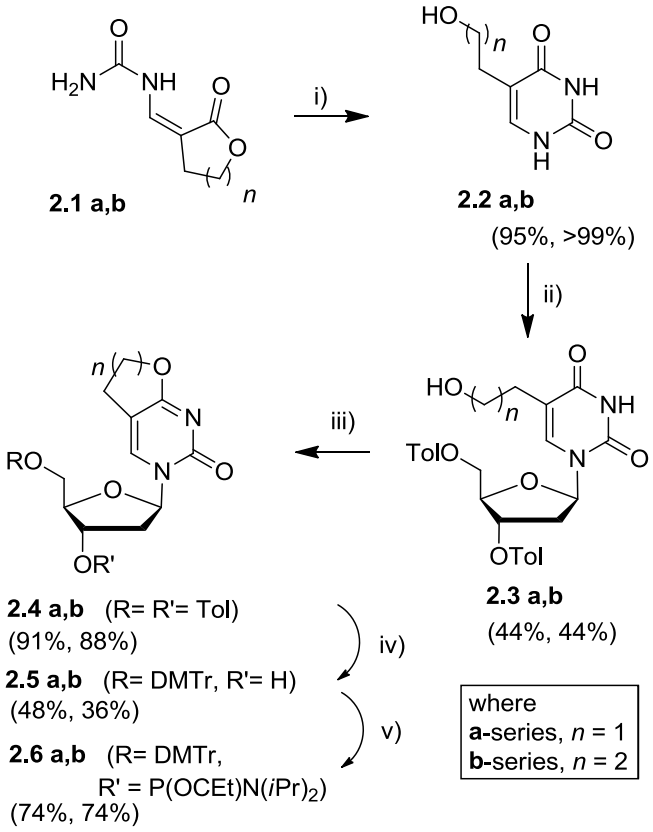
Crystals were obtained by the hanging drop vapor diffusion technique at 18 °C. O^4-Alkylthymidine modified DNA solutions were prepared by mixing template and primer strands in a 1:1 molar ratio and annealing the mixture in the presence of 10 mM sodium HEPES buffer (pH 8.0), 0.1 mM EDTA, and 50 mM NaCl at 85 °C for 5-10 min, followed by slow cooling to room temperature. hPol η protein was mixed with the DNA duplex in a 1:1.2 molar ratio in the presence of 50 mM Tris-HCl, pH 7.5, containing 450 mM KCl, and 3 mM DTT, followed by addition of either 5 μL of 100 mM $MgCl_2$ or 5 μL of 100 mM $CaCl_2$. Using a spin concentrator with Amicon cutoff filter (Millipore, Billerica, MA), the complex was concentrated to a final concentration of ~2 mg/mL. Either dTTP or one of the non-hydrolyzable nucleoside triphosphate analogs was added to the concentrated mixtures containing Ca^{2+} or Mg^{2+}. The ternary complex solution was mixed with equal volume of reservoir solution containing 0.1 M MES (pH 5.5), 5 mM $MgCl_2$, and 16-22% (w/v) PEG 2000 monomethyl ether (MME) and equilibrated against 500 μL reservoir solutions. Crystals typically appeared after overnight incubation and were allowed to grow for one to a few weeks. They were transferred to cryoprotectant solution containing reservoir solution along with 25% glycerol (v/v), and then frozen in liquid nitrogen for data collection.

2.6.2.11 X-ray Diffraction Data Collection, Structure Determination and Refinement

The atomic coordinates and structure factors (codes 5DLF, 5DLG, 5DQG, 5DQH and 5DQI) have been deposited in the Protein Data Bank (http://wwpdb.org/). X-ray diffraction data were collected at 100 K either on the 21-ID-F or the 21-ID-G beamline of the Life Sciences Collaborative Access Team (LS-CAT) at the Advanced Photon Source (APS), Argonne National

Laboratory (Argonne, IL). All data were processed with the program HKL2000.[259] Data collection statistics are summarized in **Table S2.3**. All structures were determined by Molecular Replacement in PHASER,[260,261] using the coordinates of the complex between hPol η and native DNA (PDB ID code 4O3N)[229] as the search model. Structures were refined either using PHENIX[262] or Refmac[260,263] and model building was carried out in COOT.[264] Model statistics and geometric parameters are summarized in **Table S2.3**. Illustrations were generated with the program UCSF Chimera.[265]

2.6.3 Supporting Schemes, Figures, and Tables



Scheme S2.1: Reagents and conditions for **a**-series: i) NaOEt, EtOH, reflux. ii) 1. HMDS, TMS-Cl. 2. 1-(α)-chloro-3,5-di-*O*-(p-toluoyl)-2-deoxy-D-ribose, 1,2-dichloroethane. 3. TBAF (1 M in THF). iii) 1. Ms-Cl, Py 2. DBU, THF, reflux. iv) NH_3/CH_3OH (sat. solution). 2. DMTr-Cl, DMAP, pyridine. v) Cl-P(OCEt)N(iPr)$_2$, DIPEA, THF. Reagents and conditions for **b**-series: Identical to a-series except step iii) 1. Ms-Cl, NEt_3, CH_2Cl_2. 2. DBU, THF, reflux.

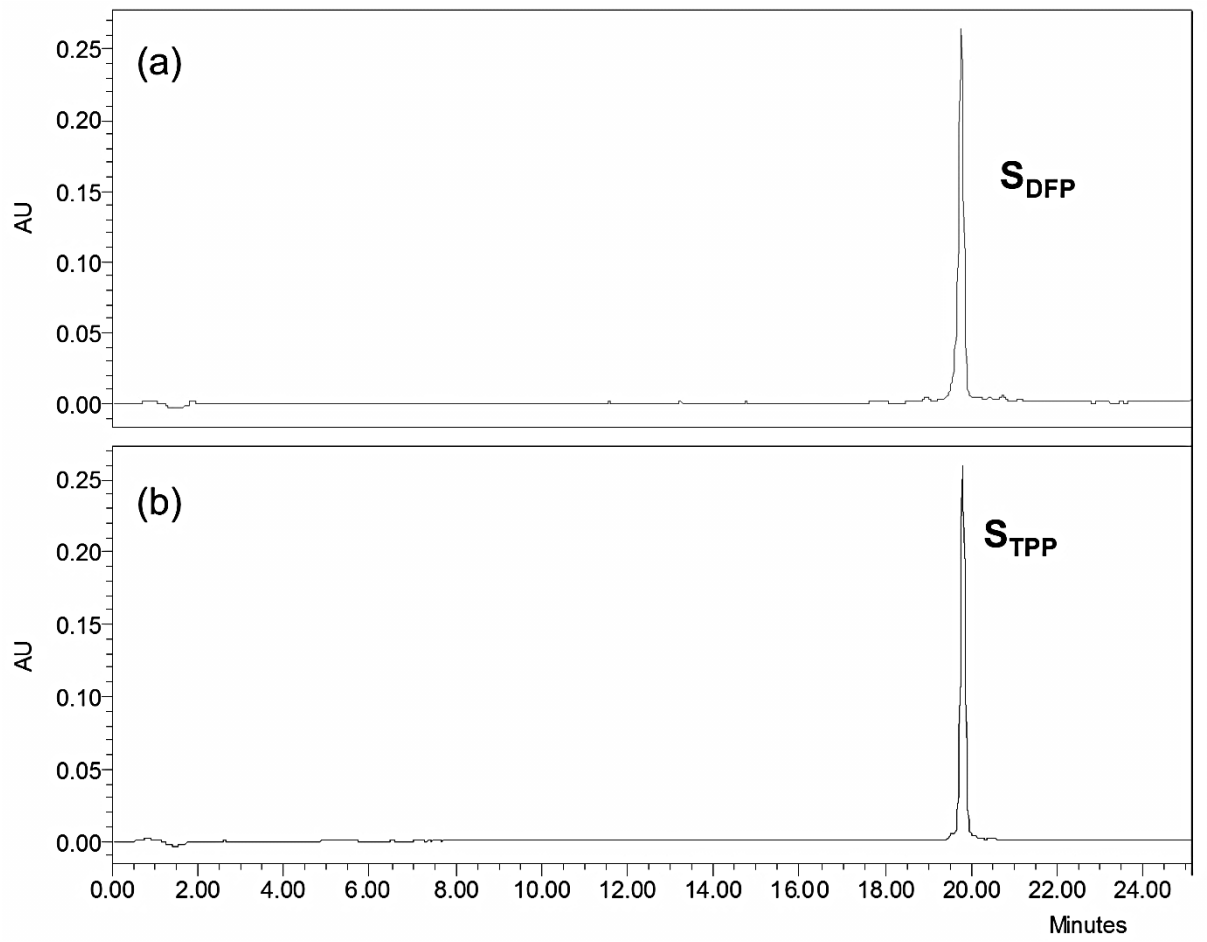


Figure S2.1: SAX-HPLC profile of purified (a) $\mathbf{S_{DFP}}$ and (b) $\mathbf{S_{TPP}}$. The column was eluted using a linear gradient of 0-52% buffer B over 24 min (buffer A: 100 mM Tris HCl, pH 7.5, 10% acetonitrile and buffer B: 100mMTris HCl, pH 7.5, 10% acetonitrile, 1 M NaCl) at 55^{o}C.

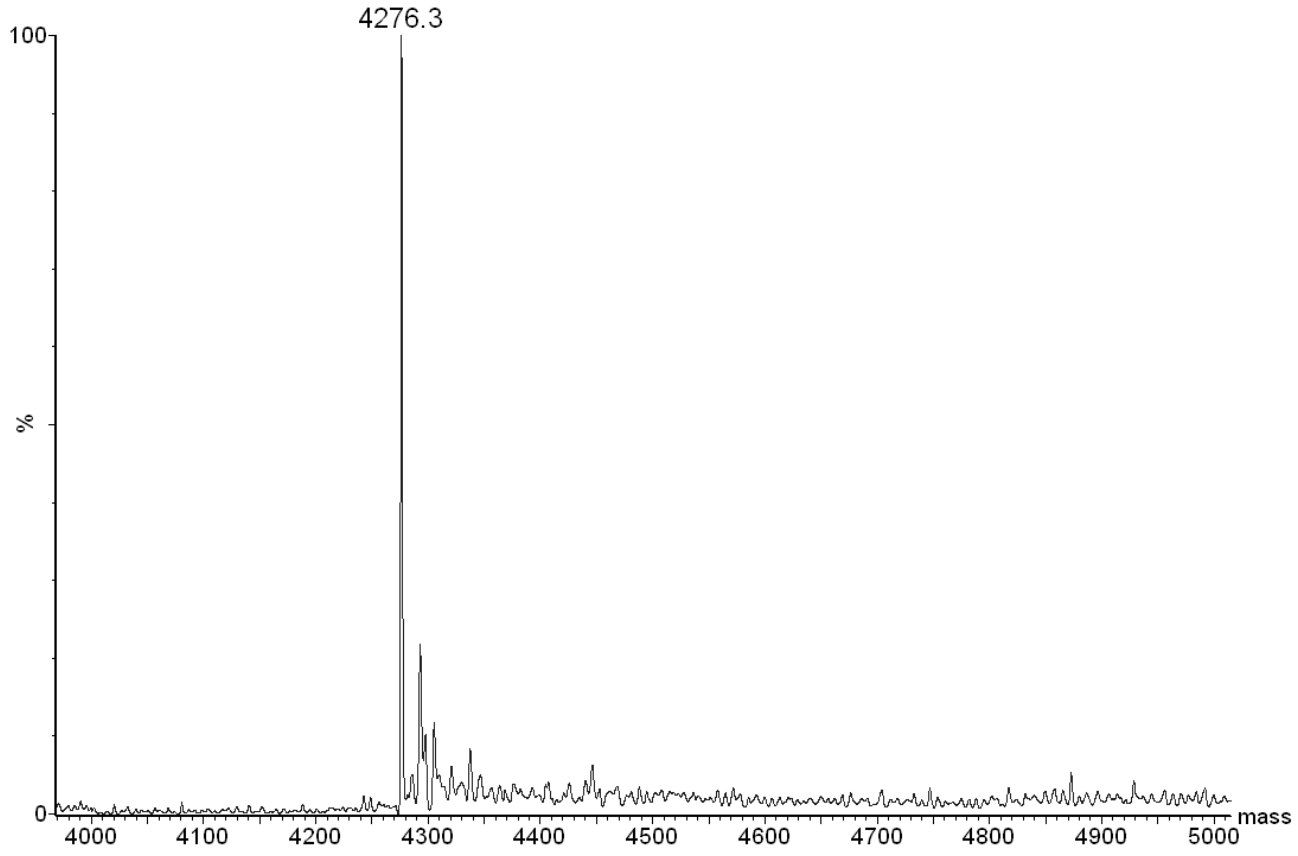


Figure S2.2: ESI MS of oligonucleotide $\mathbf{S_{DFP}}$

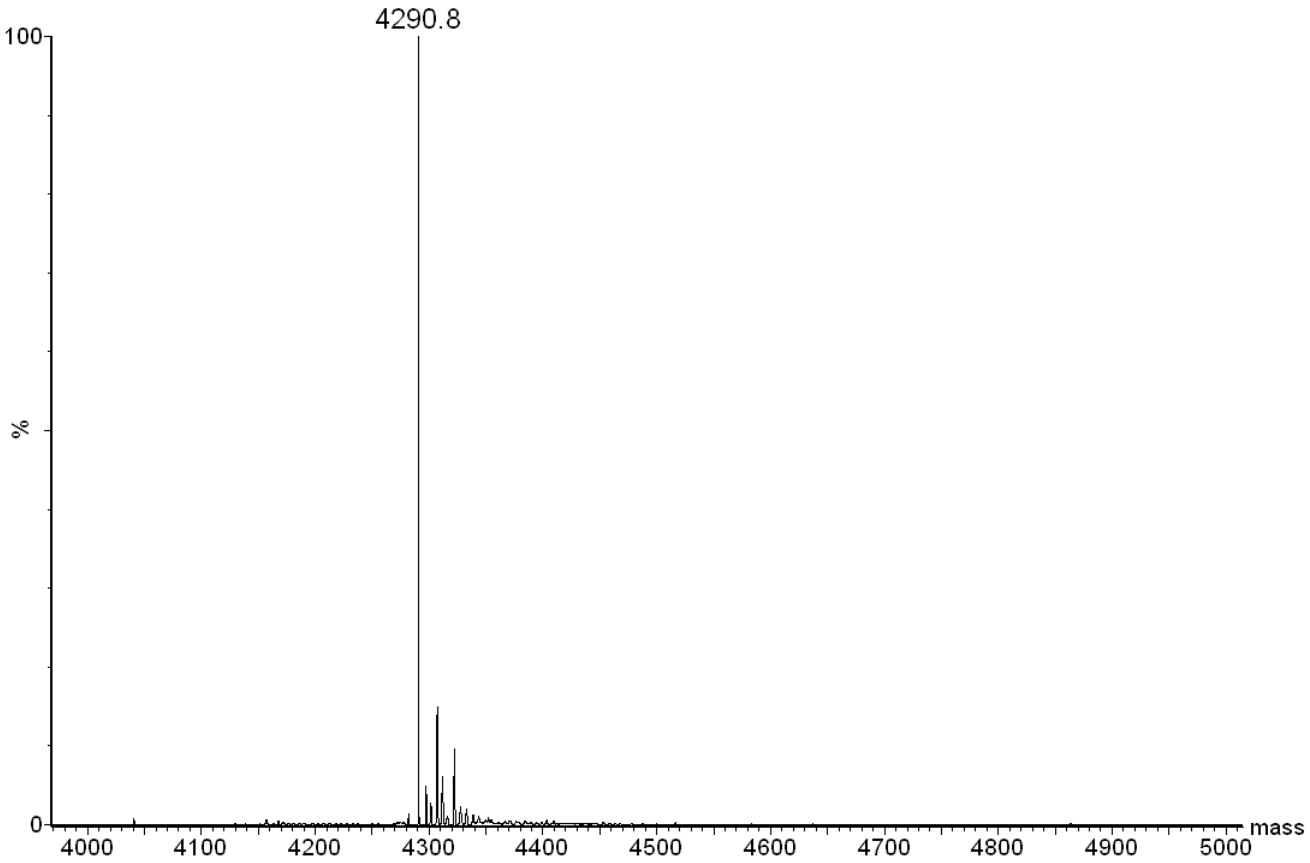


Figure S2.3: ESI MS of oligonucleotide $\mathbf{S_{TPP}}$

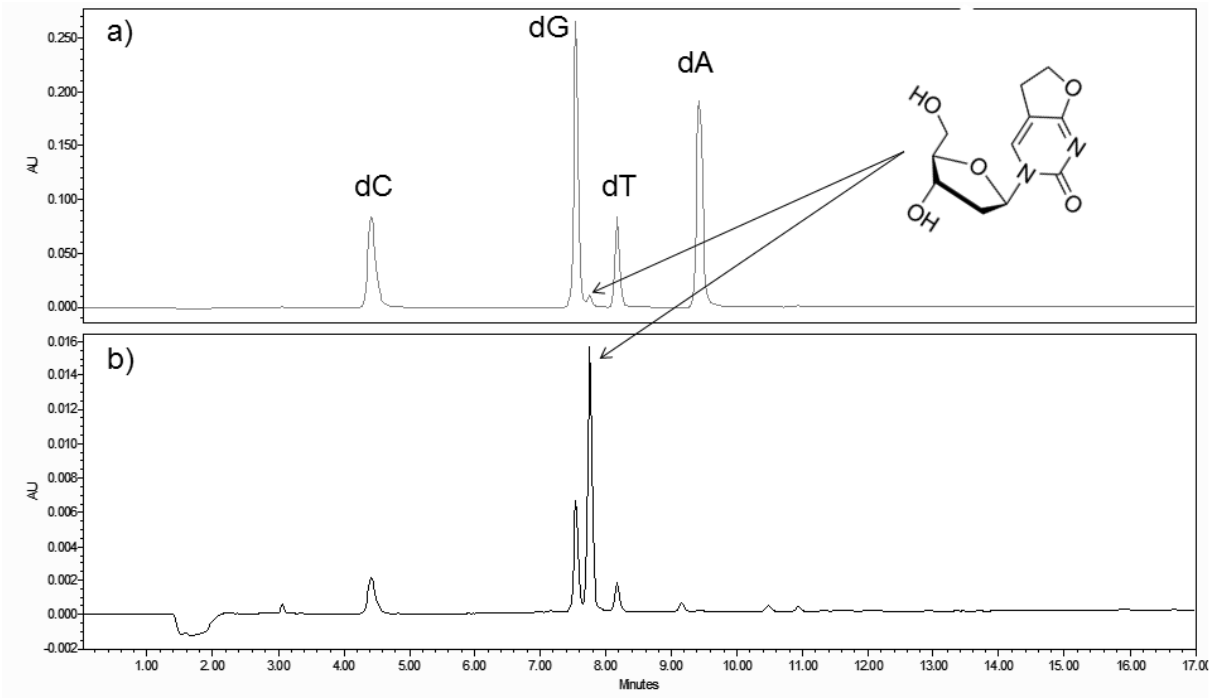


Figure S2.4: C-18 HPLC profile of digested oligomer $\mathbf{S_{DFP}}$ at (a) 260 nm and (b) 300 nm. The column was eluted with a linear gradient of 0-70% buffer B over 30 min was used to elute desired analytes (buffer A: 50 mM sodium phosphate, pH 5.8, 2% acetonitrile and buffer B: 50 mM sodium phosphate, pH 5.8, 50% acetonitrile).

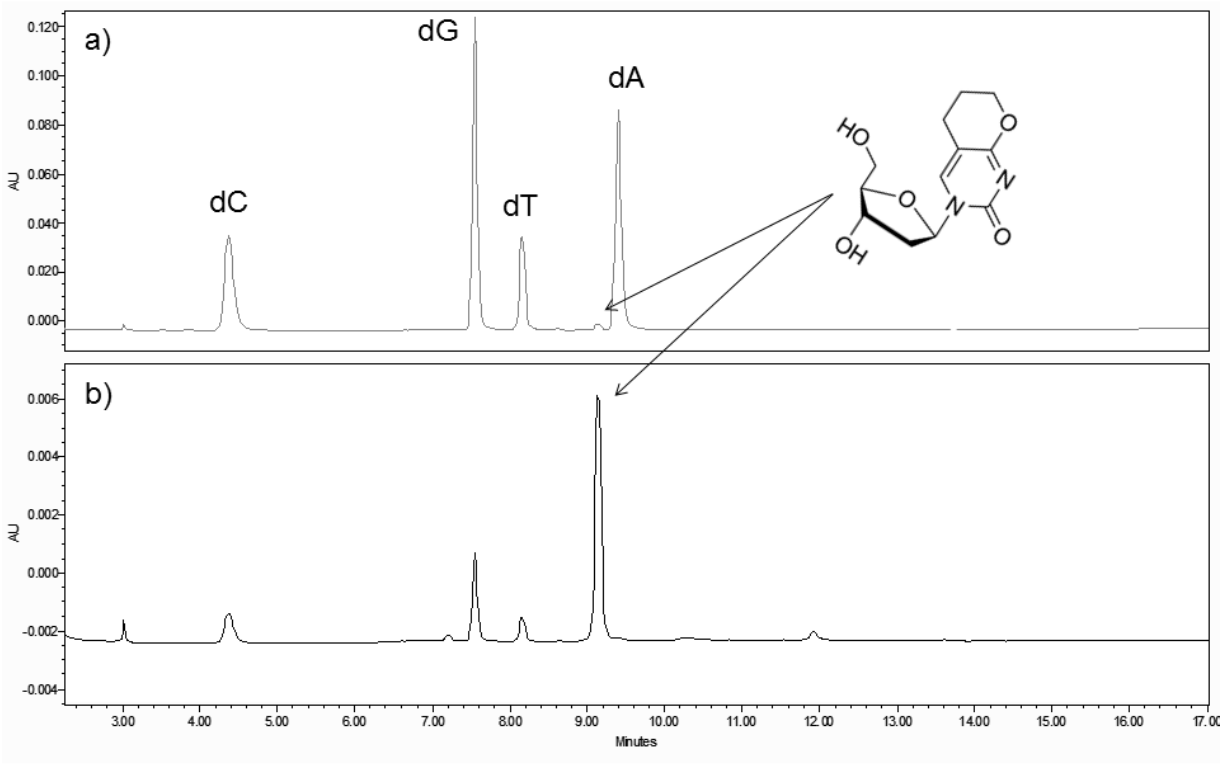


Figure S2.5: C-18 HPLC profile of digested oligomer $\mathbf{S_{TPP}}$ at (a) 260 nm and (b) 300 nm. The column was eluted with a linear gradient of 0-70% buffer B over 30 min was used to elute desired analytes (buffer A: 50 mM sodium phosphate, pH 5.8, 2% acetonitrile and buffer B: 50 mM sodium phosphate, pH 5.8, 50% acetonitrile).

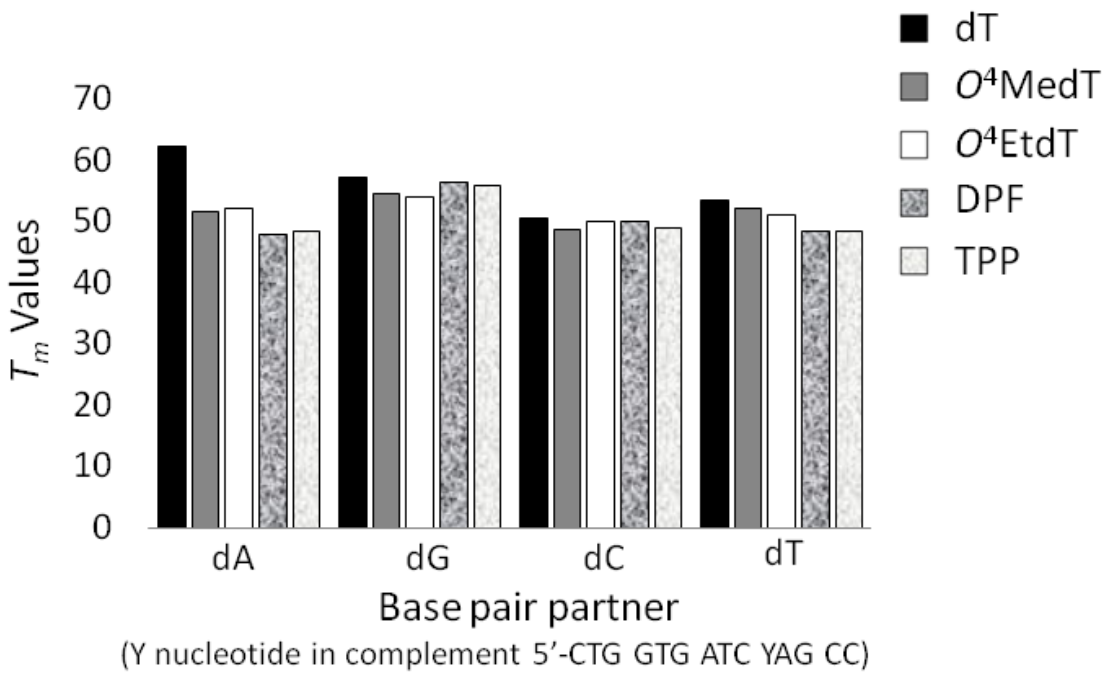


Figure S2.6: T_m values (^oC) of duplexes containing dT, O^4-MedT, O^4-EtdT, DFP or TPP across different base pairing partners. The sequence of the DNA duplex was 5′GGCTXGATCACCAG 3′/5′ CTGGTGATCYAGCC 3′ where X is dT, O^4-MedT, O^4-EtdT, DFP or TPP and Y is dA, dG, dC or dT. The solutions contain a total strand concentration of 3.8 μM for the duplexes in a buffer consisting of 90 mM sodium chloride, pH 7.0, 10 mM sodium phosphate, and 1 mM EDTA buffer. Data was acquired at 0.5 °C/min from 90 ^{o}C to 15 ^{o}C monitoring UV absorption at 260 nm. Samples were held at 15 ^{o}C for 2 min and re-heated at 0.5 ^{o}C/min to 90 ^{o}C showing reversibility (data not shown).

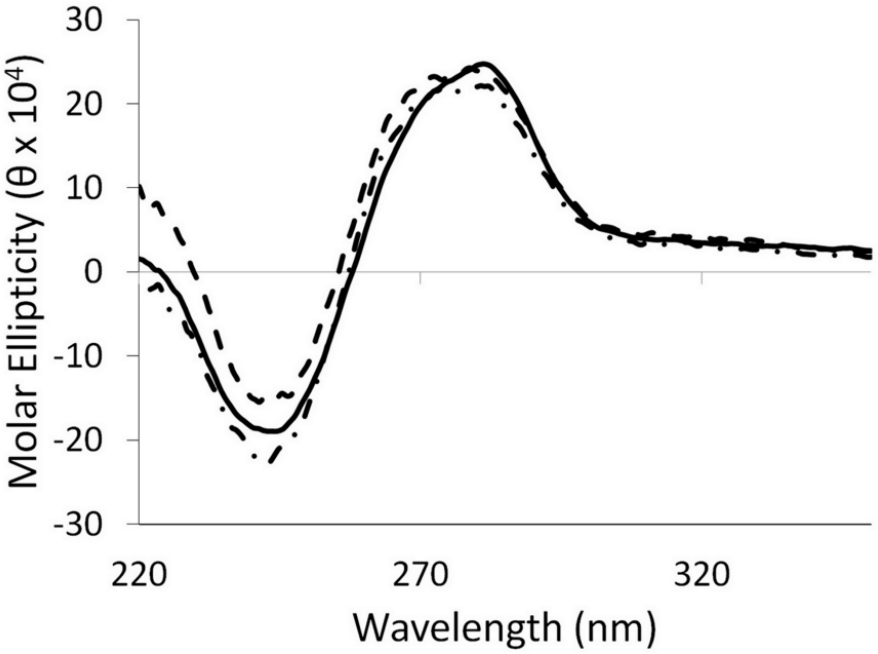


Figure S2.7: Far-UV circular dichroism spectra of DNA sequences containing the modified adducts in duplexes 5′ GGCTXGATCACCAG 3′ / 5′ CTGGTGATCAAGCC 3′ where X is dT (—), **DFP** (– – –) or **TPP** (– • –). Solutions containing a total strand concentration of 3.8 μM for the duplexes were prepared in buffer containing 90 mM sodium chloride, pH = 7.0, 10 mM sodium phosphate, and 1 mM EDTA buffer.

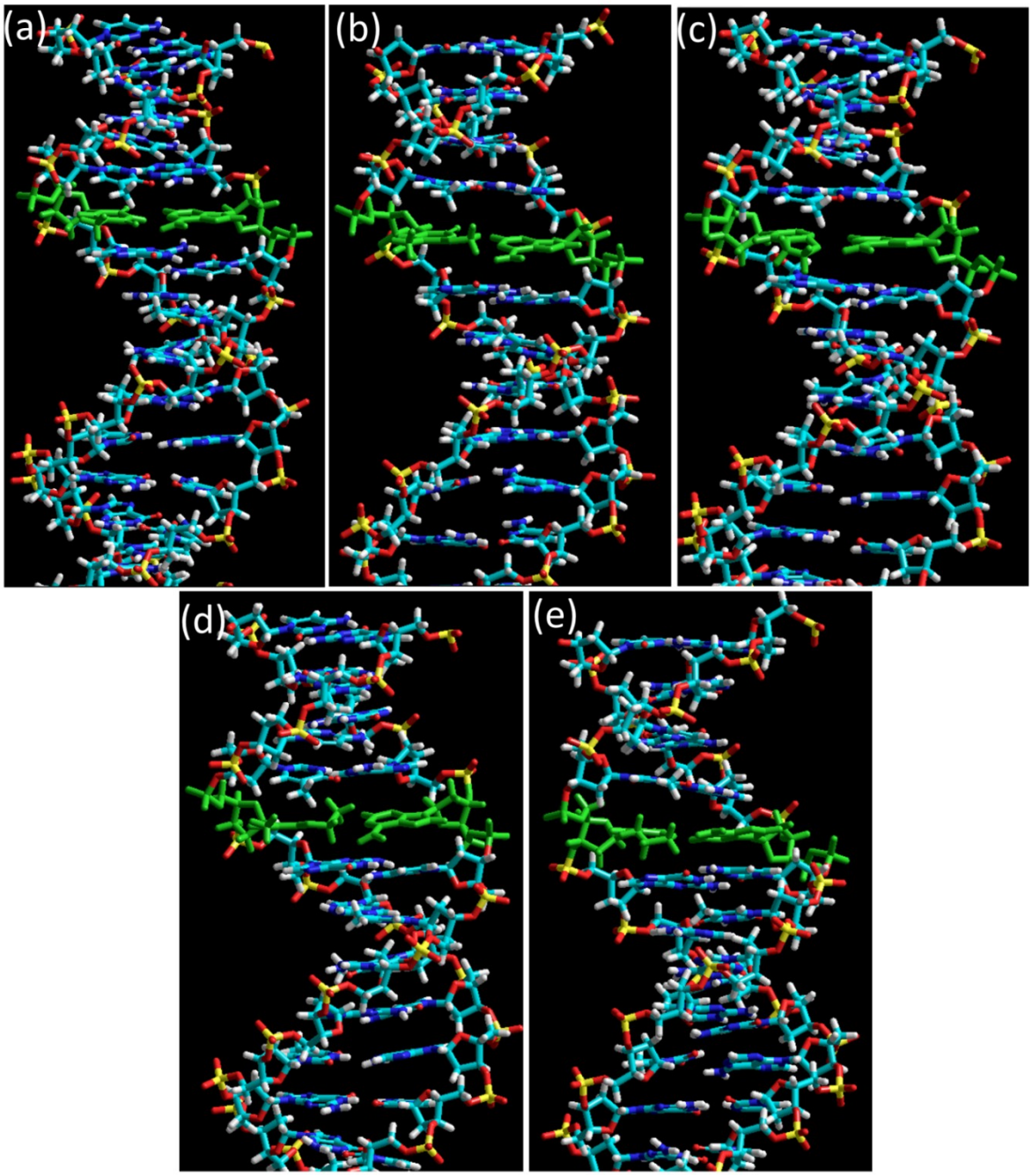


Figure S2.8: Side view of geometry optimized models of DNA duplexes 5' GGCTXGATCACCAG 3' / 5' CTGGTGATCAAGCC 3' where X is (a) dT, (b) O^4-MedT, (c) **DFP**, (d) O^4-EtdT, and (e) **TPP**.

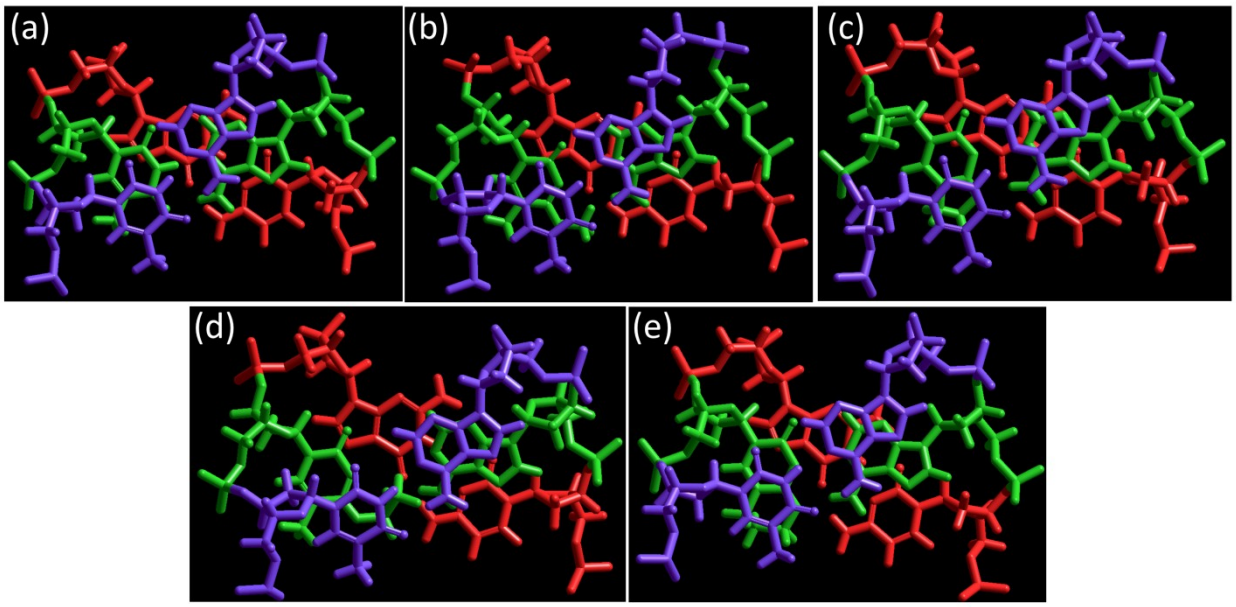


Figure S2.9: Top view of geometry optimized models of the DNA duplexes 5' GGCTXGATCACCAG 3' / 5' CTGGTGATCAAGCC 3' where X is (a) dT, (b) O^4-MedT, (c) **DFP**, (d) O^4-EtdT, and (e) **TPP**. All X · A base pairs are colored in green and flanking base pairs are red and purple. The modified thymidinyl nucleotides were found to adopt a wobble conformation similar to those described experimentally by NMR and computationally for O^4-MedT and O^4-EtdT.

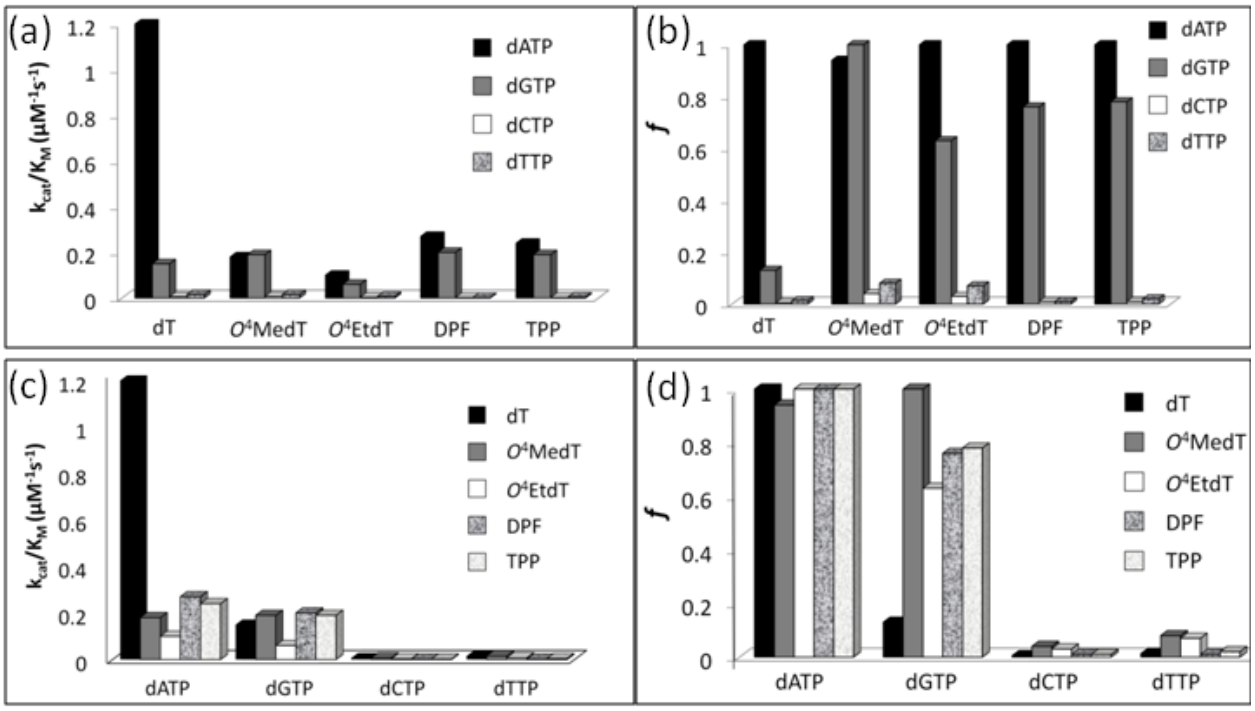


Figure S2.10: Steady-state incorporation efficiencies opposite dT, O^4-MedT, O^4-EtdT, DFP, and TPP by hPol η with individual dNTPs. Values are reported in **Table S2.1.** Panels (a-d) are different representations of steady-state kinetic data. Note that Panel (a) is re-depicted here, despite being shown in the main text, for ease of comparison with panels (b-d).

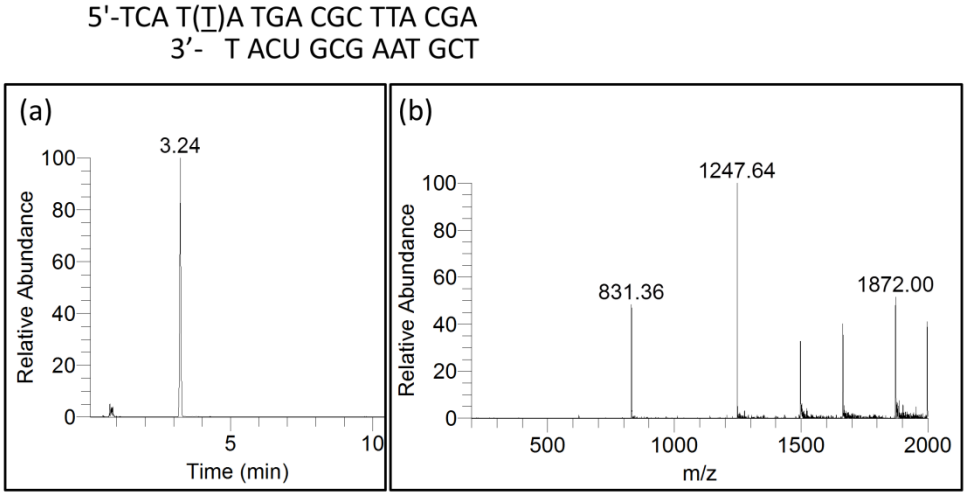


Figure S2.11: LC-MS analysis of the most abundant full-length extension products opposite unmodified dT in the DNA template by hPol η in the presence of all four dNTPs. (a) Sample reconstructed extracted ion chromatogram for *m/z* 1247.8 for product with sequence 5'-pCAT AATGA and (b) mass spectrum of peak at retention time 3.24 min. See **Table S2.2** in the manuscript for the full list of products and respective *m/z* assignments.

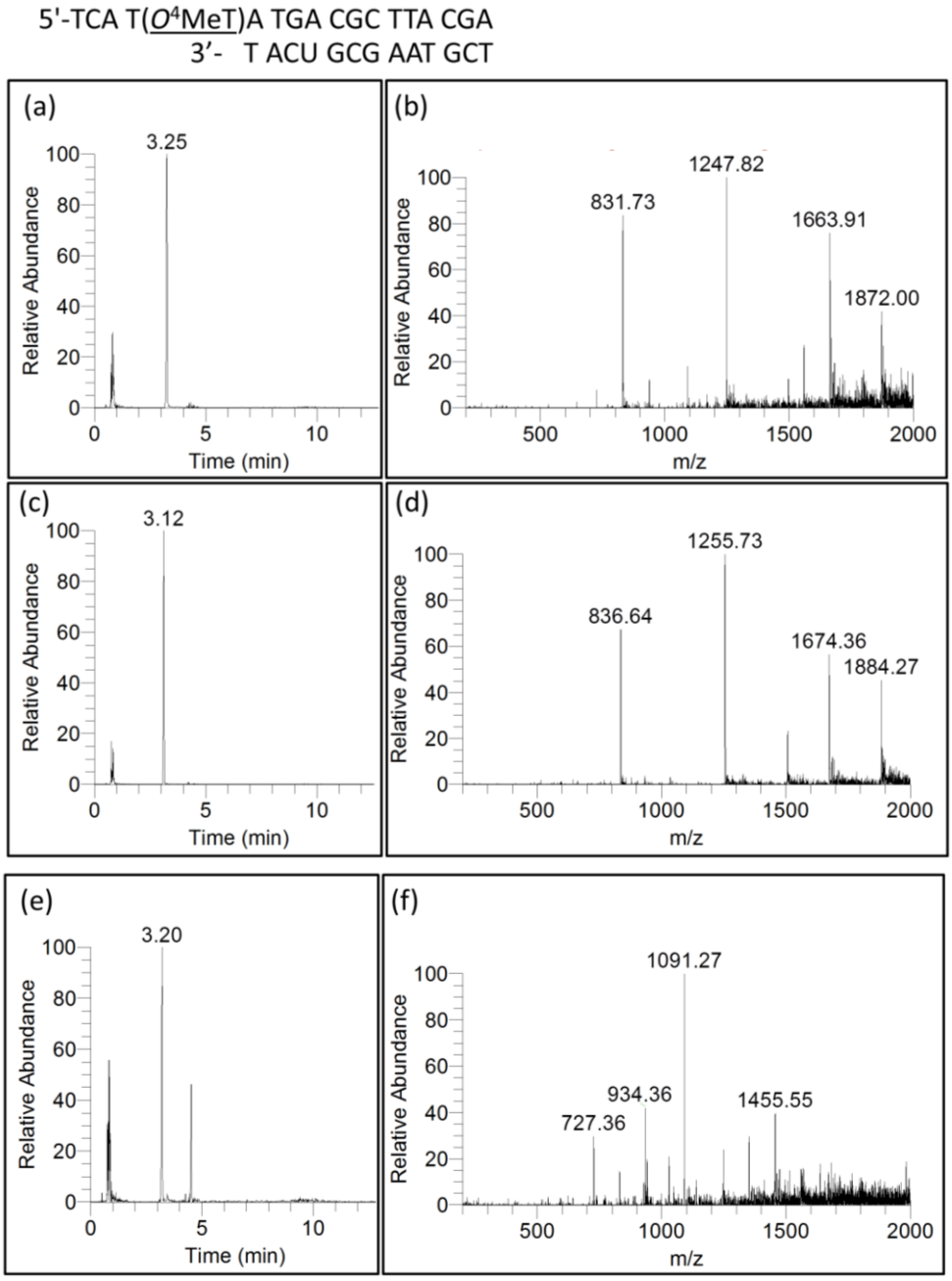


Figure S2.12: LC-MS analysis of the most abundant full-length extension products opposite O^4-MedT in the DNA template by hPol η in the presence of all four dNTPs. (a) Sample reconstructed extracted ion chromatogram for *m/z* 1247.8 for product with sequence 5'-pCAT AATGA and (b) mass spectrum of peak at retention time 3.25 min. (c) Sample reconstructed extracted ion chromatogram for *m/z* 1255.8 for product with sequence 5'-pCAT GATGA and (d) mass spectrum of peak at retention time 3.12 min. (e) Sample reconstructed extracted ion chromatogram for *m/z* 1091.2 for product with sequence 5'-pCAT _ATGA and (f) mass spectrum of peak at retention time 3.20 min. See **Table S2.2** in the manuscript for the full list of products and respective *m/z* assignments.

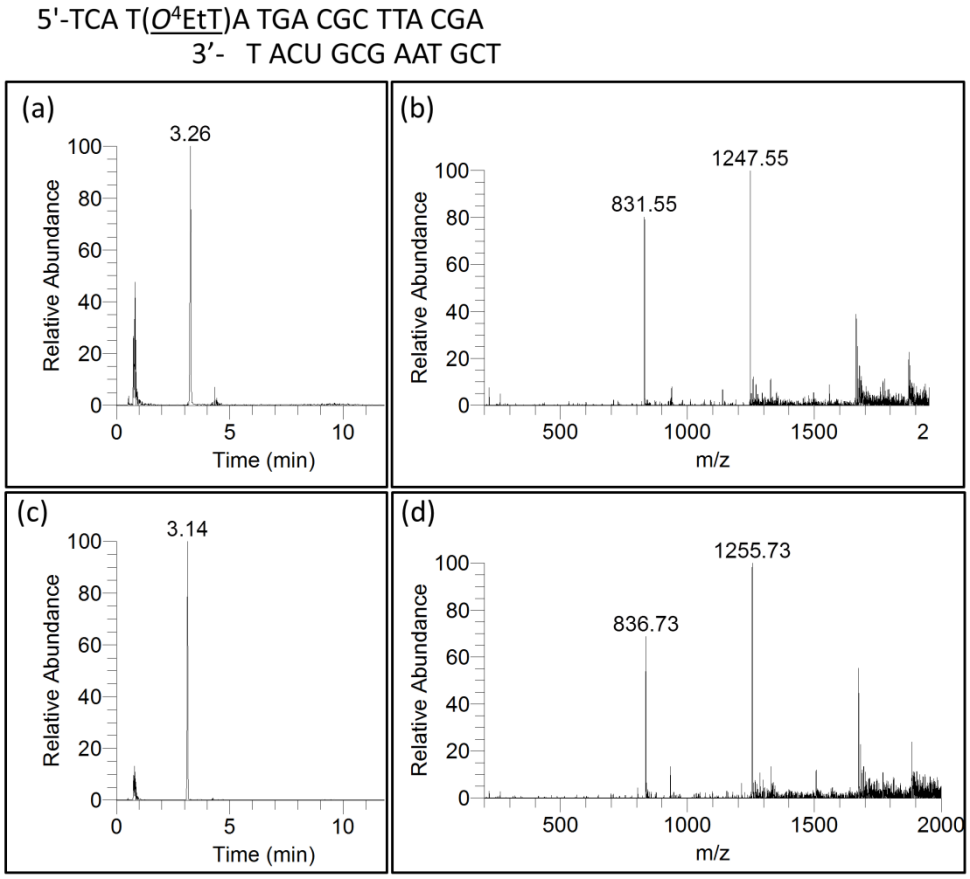


Figure S2.13: LC-MS analysis of the most abundant full-length extension products opposite O^4-EtdT in the DNA template by hPol η in the presence of all four dNTPs. (a) Sample reconstructed extracted ion chromatogram for *m/z* 1247.8 for product with sequence 5'-pCAT AATGA and (b) mass spectrum of peak at retention time 3.26 min. (c) Sample reconstructed extracted ion chromatogram for *m/z* 1255.8 for product with sequence 5'-pCAT GATGA and (d) mass spectrum of peak at retention time 3.14 min. See **Table S2.2** in the manuscript for the full list of products and respective *m/z* assignments.

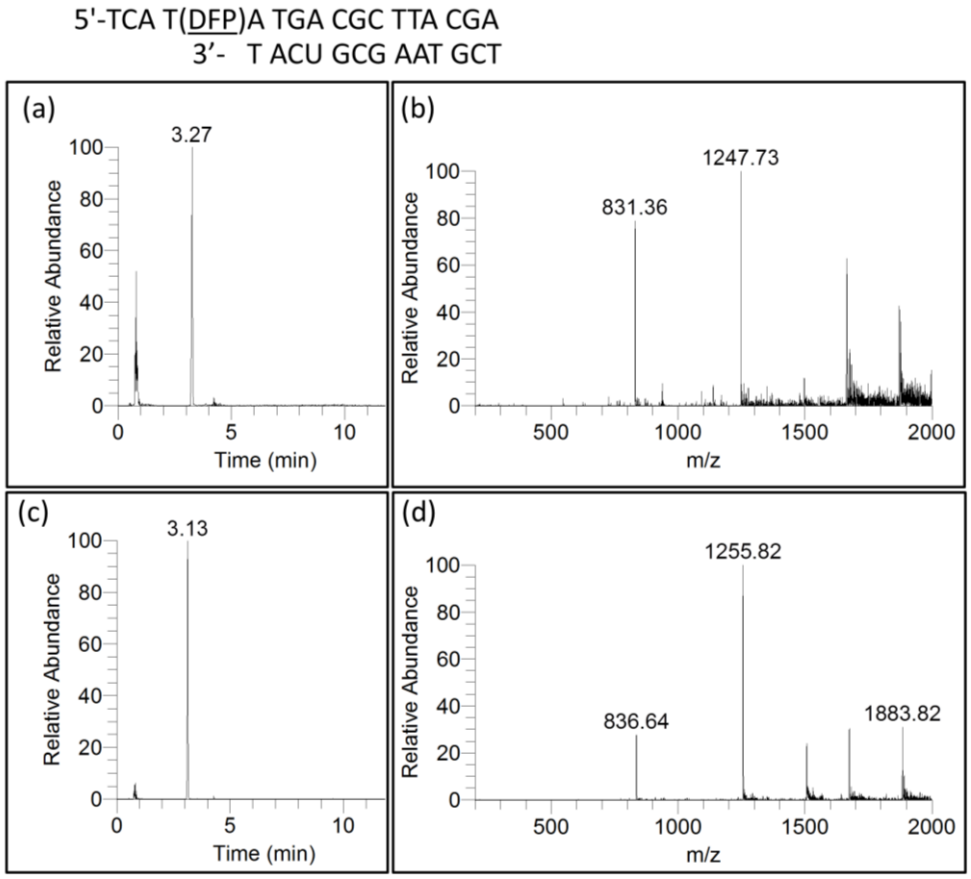


Figure S2.14: LC-MS analysis of the most abundant full-length extension products opposite DFP in the DNA template by hPol η in the presence of all four dNTPs. (a) Sample reconstructed extracted ion chromatogram for *m/z* 1247.8 for product with sequence 5'-pCAT AATGA and (b) mass spectrum of peak at retention time 3.27 min. (c) Sample reconstructed extracted ion chromatogram for *m/z* 1255.8 for product with sequence 5'-pCAT GATGA and (d) mass spectrum of peak at retention time 3.13 min. See **Table S2.2** in the manuscript for the full list of products and respective *m/z* assignments.

.

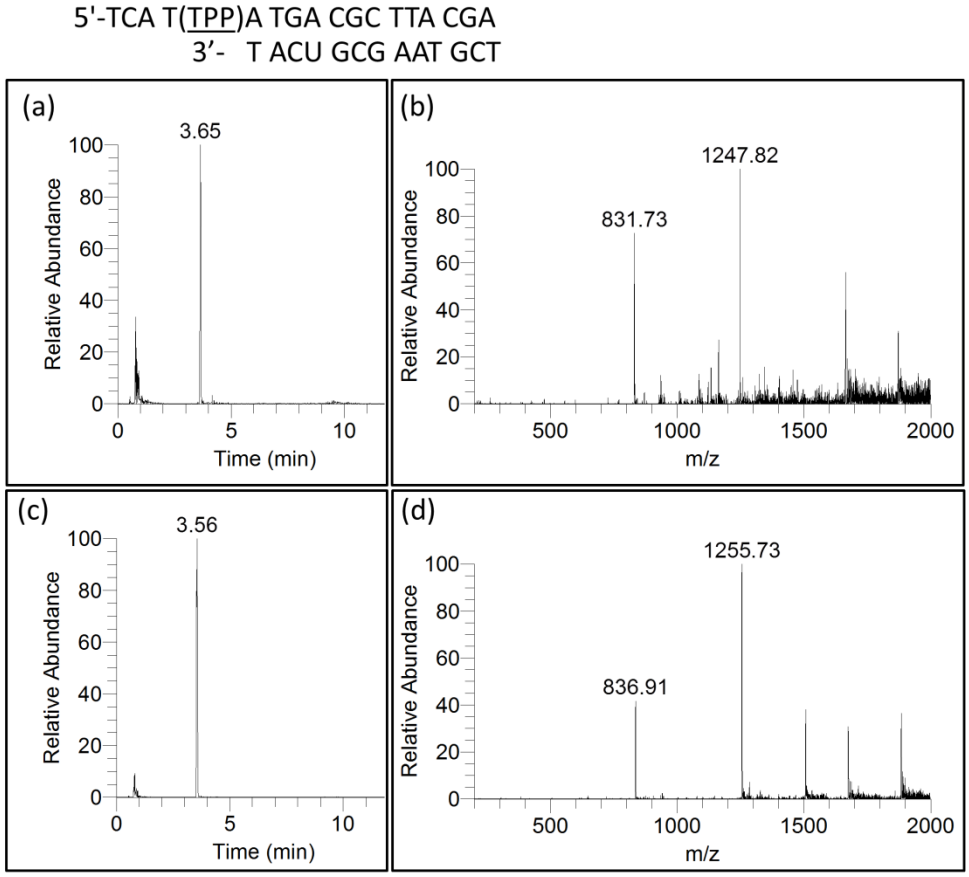


Figure S2.15: LC-MS analysis of the most abundant full-length extension products opposite TPP in the DNA template by hPol η in the presence of all four dNTPs. (a) Sample reconstructed extracted ion chromatogram for *m/z* 1247.8 for product with sequence 5'-pCAT AATGA and (b) mass spectrum of peak at retention time 3.65 min. (c) Sample reconstructed extracted ion chromatogram for *m/z* 1255.8 for product with sequence 5'-pCAT GATGA and (d) mass spectrum of peak at retention time 3.56 min. See **Table S2.2** in the manuscript for the full list of products and respective *m/z* assignments.

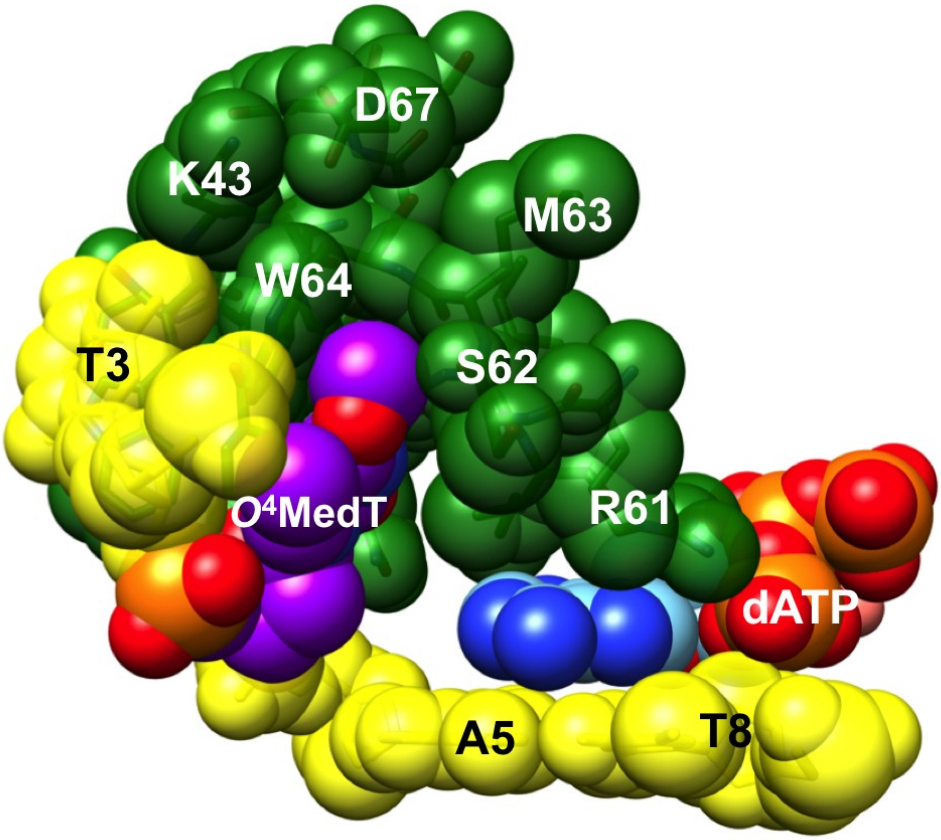


Figure S2.16: Space filling model of the hPol η active site with O^4-MedT (magenta carbon atoms) wedged between Trp-64 and Ser-62 (protein residues are green), allowing for extensive hydrophobic interactions between the O^4-MedT methyl group and surrounding residues from the polymerase finger domain. Carbon atoms of the incoming dATP are colored in light blue and template dT 3′ to the lesion and the 5′-adjacent T:A base pair are colored in yellow.

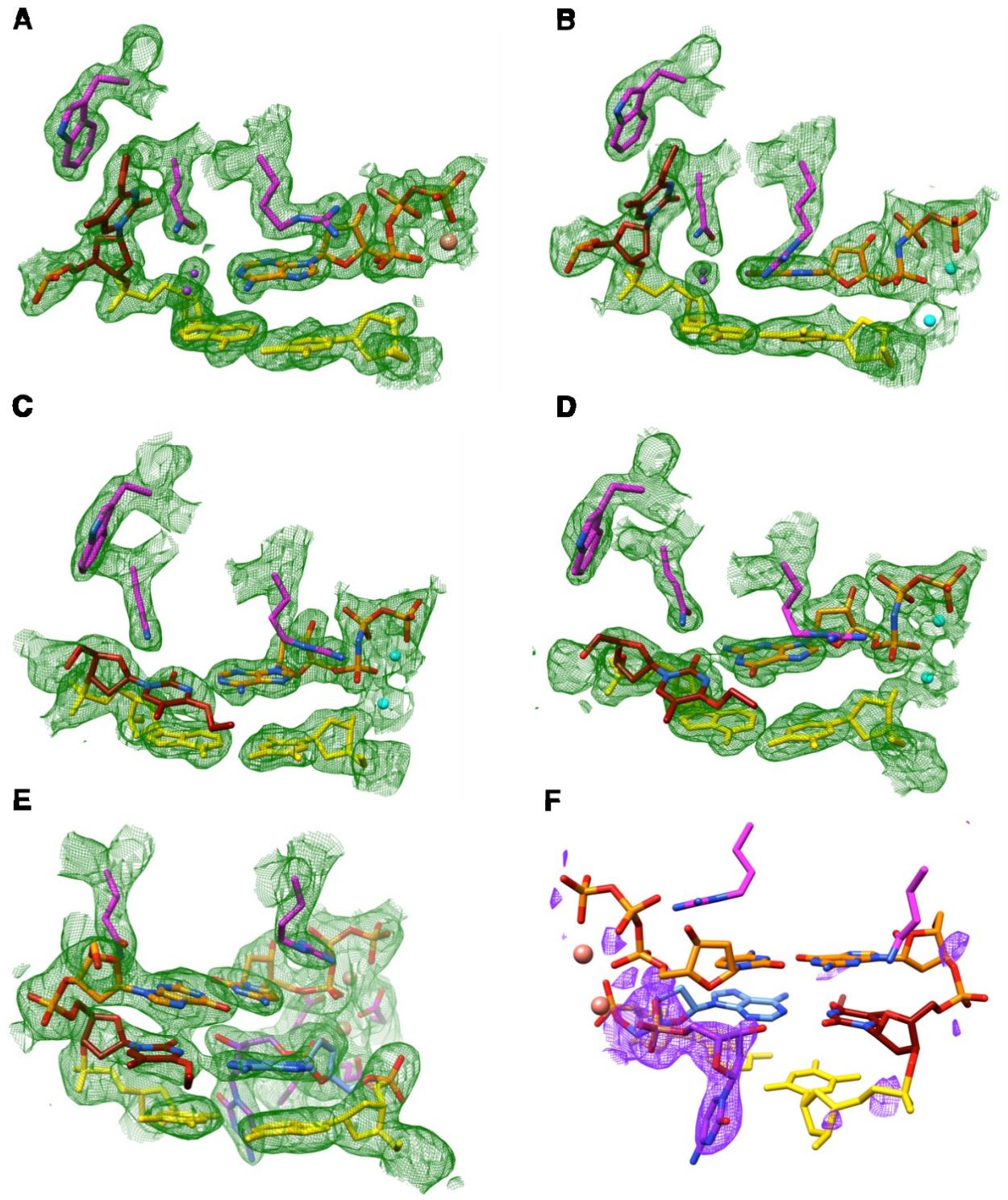


Figure S2.17: Quality of the final models of ternary hPol η complexes with O^4-MedT/O^4-EtdT-adducted DNA template strands. Fourier (2F*o*-F*c*) sum electron density drawn at the 1σ threshold (green meshwork) around the active site region. *A*, O^4-MedT and dATP, insertion stage. *B*, O^4-MedT and dGMPNPP, insertion stage. *C*, O^4-EtdT and dAMPNPP, insertion stage. *D*, O^4-EtdT and dGMPNPP, insertion stage. *E*, O^4-EtdT opposite primer dA, followed by dCTP opposite template dG; extension stage. *F*, Omit (F*o*-F*c*) electron density drawn at the 3σ threshold (purple meshwork) around the 3′-terminal primer dC that was added during crystallization. Carbon atoms of O^4-MedT and O^4-EtdT, the incoming nucleotide triphosphate, and selected hPol η amino acid side chains are colored in maroon,

orange and magenta, respectively, and Mg^{2+} and Ca^{2+} ions are cyan and pink spheres, respectively. Carbon atoms of the template dA paired to O^4-EtdT and those of the 3′-terminal primer dC are colored in light blue and purple, respectively.

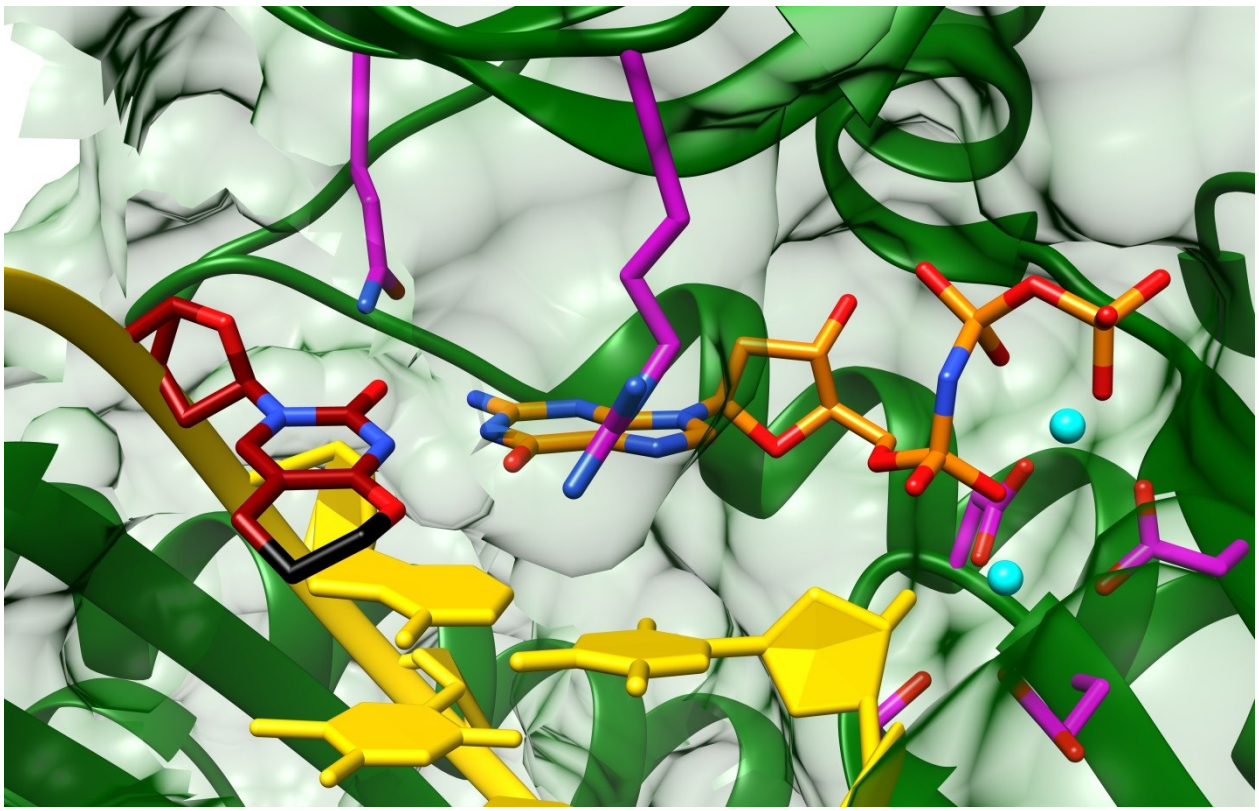


Figure S2.18: Model of the active site configuration in the ternary hPol η insertion-step complex with dGMPNPP opposite TPP (black carbon atoms). The TPP moiety was modeled in place of the ethyl adduct (O^4-EtdT) seen in the crystal structure of the complex. The bicyclic ring protrudes into the major groove of the template-primer duplex, and can be accommodated without clashing with incoming nucleotide triphosphate or amino acid side chains in its vicinity.

Table S2.1: Steady-state kinetics of incorporation of dATP, dGTP, dCTP and dTTP opposite dT, O^4-MedT, O^4-EtdT, DFP, and TPP by hPol η.

Template base	dNTP	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M (μM^{-1} s^{-1})	f^1
dT	dATP	1.60 ± 0.03	1.4 ± 0.1	1.2 ± 0.1	
	dGTP	0.71 ± 0.03	4.6 ± 0.7	0.15 ± 0.02	0.13
	dCTP	0.16 ± 0.01	29 ± 7	0.006 ± 0.001	0.005
	dTTP	0.37 ± 0.01	24 ± 3	0.015 ± 0.002	0.013
O^4-MedT	dATP	0.69 ± 0.03	3.9 ± 0.6	0.18 ± 0.03	0.94
	dGTP	1.18 ± 0.02	6.2 ± 0.4	0.19 ± 0.01	
	dCTP	0.11 ± 0.01	15 ± 3	0.007 ± 0.002	0.04
	dTTP	0.25 ± 0.02	17 ± 4	0.015 ± 0.004	0.08
O^4-EtdT	dATP	0.83 ± 0.03	9 ± 1	0.10 ± 0.01	
	dGTP	0.61 ± 0.02	10 ± 1	0.06 ± 0.01	0.63
	dCTP	0.12 ± 0.01	43 ± 10	0.003 ± 0.001	0.03
	dTTP	0.19 ± 0.02	27 ± 7	0.007 ± 0.002	0.07
DFP	dATP	1.18 ± 0.05	4.4 ± 0.7	0.27 ± 0.04	
	dGTP	1.15 ± 0.02	5.7 ± 0.4	0.20 ± 0.02	0.76
	dCTP	0.12 ± 0.01	55 ± 14	0.002 ± 0.001	0.01
	dTTP	0.20 ± 0.02	84 ± 21	0.002 ± 0.001	0.01
TPP	dATP	1.33 ± 0.04	5.4 ± 0.6	0.24 ± 0.03	
	dGTP	1.18 ± 0.03	6.2 ± 0.6	0.19 ± 0.02	0.78
	dCTP	0.12 ± 0.01	54 ± 12	0.002 ± 0.001	0.01
	dTTP	0.21 ± 0.01	51 ± 11	0.004 ± 0.001	0.02

[1] $f = (k_{cat}/K_M)_{dNTP} / (k_{cat}/K_M)_{max}$

Table S2.2: LC-MS/MS analysis of full-length extension products opposite dT, O^4-MedT, O^4-EtdT, DFP and TPP by hPol η (values are reported in %).

Name	Sequence	dT	O^4-MedT	O^4-EtdT	DFP	TPP
7Del	5'-pCAT _ATGA		6	2	< 1	< 1
8A	5'-pCAT AATGA	89	27	22	21	16
8_A	5'-pCAT _ATGAA	< 1	1	< 1	< 1	< 1
8G	5'-pCAT GATGA	3	45	50	69	72
8_G	5'-pCAT _ATGAG	3	3	1		< 1
8C	5'-pCAT CATGA	< 1		< 1		< 1
8T	5'-pCAT TATGA		1	2		
9AA	5'-pCAT AATGAA	3	1	1	< 1	
9AG	5'-pCAT AATGAG	2	1	1	< 1	< 1
9GA	5'-pCAT GATGAA	< 1	8	10	4	4
9GG	5'-pCAT GATGAG		7	10	3	3

Table S2.3: Crystal data, data collection parameters and structure refinement statistics.

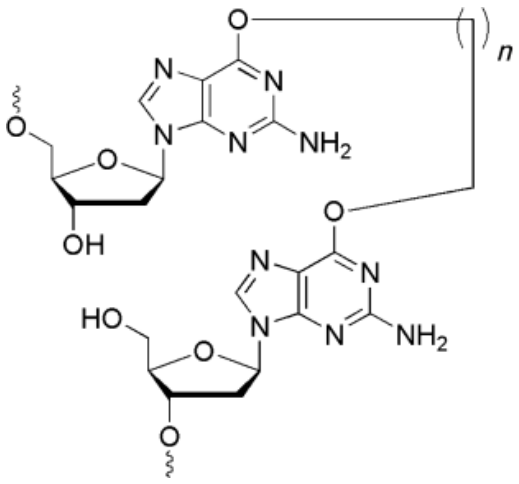
Complex	O^4-MedT: dATP	O^4-MedT: dGMPNPP	O^4-EtdT: dAMPNPP	O^4-EtdT: dGMPNPP	O^4-EtdT:dA Extension
Data Collection					
Wavelength [Å]	0.97856	0.97856	0.97856	0.97872	0.97872
Space group	$P6_1$	$P6_1$	$P6_1$	$P6_1$	$P6_1$
Resolution [Å]	50 - 1.97 (2.00 - 1.97)[a]	50 - 2.35 (2.39 - 2.35)	50 - 2.29 (2.35 - 2.29)	50 - 1.99 (2.02 - 1.99)	50 - 2.30 (2.34 - 2.30)
Unit cell *a*=*b*, *c* [Å]	99.03, 82.17	98.66, 82.13	98.82, 82.02	98.47, 82.23	98.64, 81.54
Unique reflections	32,497 (1,624)	18,606 (901)	20,600 (1,504)	30,841 (1,534)	20,077 (941)
Completeness [%]	100 (100)	97.9 (96.5)	100 (100)	98.9 (98.5)	99.4 (94.5)
I/σ(I)	14.30 (2.13)	10.46 (1.75)	8.40 (3.20)	19.92 (2.31)	22.72 (2.05)
Wilson B-factor [$Å^2$]	16.0	30.9	26.1	22.8	43.9
R-merge	0.165 (0.834)	0.127 (0.941)	0.140 (0.539)	0.139 (0.997)	0.093 (0.897)
Redundancy	5.7 (5.7)	4.2 (3.5)	5.6 (5.6)	11.4 (10.0)	7.2 (5.6)
Refinement					
R-work	0.165 (0.213)	0.171 (0.238)	0.176 (0.156)	0.161 (0.198)	0.204 (0.289)
R-free	0.216 (0.282)	0.247 (0.301)	0.227 (0.261)	0.210 (0.274)	0.266 (0.363)
Number of atoms Protein/DNA	3,402/384	3,366/387	3,367/389	3,399/394	3,323/410
dNMPNPP/ Mg^{2+}	30/1 (Ca^{2+})	31/2	30/2	31/2	28/2 (Ca^{2+})
Water/Solute	494/1	222/1	212/1	362/1	103/0

Protein residues	439	430	430	438	424
B-factor [Å^2]					
Average	24.0	34.0	31.0	28.0	55.5
Protein/DNA	23.2/26.5	34.2/39.8	30.9/36.9	27.0/34.2	55.3/60.1
dNMPNPP/ Mg^{2+}	20.5/12.8	28.3/28.6	19.3/17.8	23.6/18.9	39.3/41.6
Water/ glycerol	30.4/13.4	35.2/27.1	30.1/20.9	32.5/ 19.5	47.9/-
R.m.s. deviations					
bonds [Å]	0.016	0.009	0.012	0.008	0.011
angles [deg.]	1.6	1.1	1.4	1.1	1.5
Ramachandran					
Favored (%)	98	97	97	98	96
Allowed (%)	1.5	2.3	2.8	1.8	3.0
Outliers (%)	0.5	0.7	0.2	0.2	1.0
PDB ID Code	5DLF	5DLG	5DQG	5DQH	5DQI

[a] Statistics for the highest-resolution shell are shown in parentheses.

CHAPTER III

Synthesis, Characterization, and Repair of a Flexible O^6-2'-Deoxyguanosine-Alkylene-O^6-2'-Deoxyguanosine Intrastrand Cross-Link



Published as:

O'Flaherty DK, Wilds CJ (2015) *Chemistry - A European Journal. 21*, 10522–10529.

Department of Chemistry and Biochemistry, Concordia University, Montréal, Québec, Canada, H4B1R6.

Abstract

Oligonucleotides tethered by an alkylene linkage between the O^6-atoms of two consecutive 2'-deoxyguanosines, which lack a phosphodiester linkage between these residues, have been synthesized as a model system of intrastrand cross-linked (IaCL) DNA. UV thermal denaturation studies of duplexes formed between these butylene and heptylene linked oligonucleotides with their complementary DNA sequences revealed ca. 20^{o}C reduction in stability relative to the unmodified duplex. Circular dichroism spectra of the model IaCL duplexes displayed a signature characteristic of B-form DNA suggesting minimal global perturbations are induced by the lesion. The model IaCL containing duplexes were investigated as substrates of O^6-alkylguanine DNA alkyltransferase (AGT) proteins from human and *E. coli (*Ada-C and OGT). Human AGT was found to repair both model IaCL duplexes with greater efficiency towards the heptylene versus butylene analog adding to our knowledge of substrates this protein can repair.

3.1 Introduction

DNA is susceptible to modification by various endogenous metabolites and environmental agents.[31,266] Any chemical or structural change to this important biomolecule may have a profound effect on an organism since genetic instructions may be compromised. Lesions that can form in DNA include interstrand cross-links (ICL), intrastrand cross-links (IaCL), DNA-protein cross-links (DPCs) and various mono-alkylated adducts. The presence of ICL affects both strands of DNA and can block vital cellular processes that require unwinding such as replication and transcription.[54,57,267] The introduction of IaCL in DNA modifies only one of the strands but can lead to a distortion of DNA structure which can result in the interference of critical cellular processes.[69,95,268,269] These modifications can have an impact on DNA structural features such as helical bending that can affect the affinity of DNA-binding proteins which may consequently impair cellular functions that involve DNA-protein complex formation.[95,270–272]

DNA IaCL can arise from numerous sources. For example, γ-radiation can induce a G[8,5-Me]T IaCL lesion inducing structural perturbations in the DNA that has been shown to lead to blockage of Pol IV (dinB) of *E. coli* and Taq polymerase *in vitro*.[273] UV radiation is the main cause of photocarcinogenesis inducing high levels of p53 mutations at bipyrimidine sites in skin tumors.[75,274–277] Cyclobutane-pyrimidine dimers (CPDs) and pyrimidine (6-4) pyrimidone photoproducts (in addition to their Dewar isomers) are two major classes of IaCL DNA lesions introduced by UV radiation.[75,278,279] Structural investigation of oligomers containing different UV induced IaCL lesions revealed variation in global and local structures, all of which impair the proper functioning of DNA polymerases.[280]

Some chemical agents can introduce both DNA ICL and IaCL lesions and have applications in various chemotherapeutic regimens. One of the most-widely studied and

administered chemotherapeutic cancer drug, cisplatin, has an estimated lesion distribution of 90-95% IaCL and the remaining 5-10% is attributed to ICL, mono-adducts and DPCs [107,268,281–283]. Bifunctional electrophilic drugs such as busulfan,[69] mitomycin C,[70,284] and mechlorethamine[285] have all been shown to introduce IaCL lesions in DNA. A common cross-linking site on DNA occurs between the N7-atoms of adjacent purines.[69,286] For example, busulfan's main IaCL lesion occurs at 5'-d(GA) sites linking the N7-atoms.[69]

A major mechanism of chemotherapeutic resistance involves DNA damage recognition and repair. IaCL DNA is thought to be repaired primarily by nucleotide excision repair (NER)[66,287] and translesion synthesis (TLS).[71,288] Studies with an oxidatively generated IaCL between guanine and pyrimidine bases have revealed recognition and, in some cases, incision by the NER machinery in *E coli*.[81,289] Interestingly, proteins containing a high mobility group (HMG) domain DNA-binding motif showed inhibition of mammalian NER factors from acting on IaCL DNA.[83] Certain platinum IaCL in DNA, 1,2 d(GG) and 1,3 d(GNG) in particular, are readily recognized and excised by the NER pathway.[290,291] Y-family DNA polymerases, such as Pol η, are efficient and accurate in bypassing the platinum-based IaCL between guanine residues.[224–226] Cell lines with mutations in the POLH gene, responsible for encoding Pol η, are hypersensitive to both UV-radiation[292] and platinum-based drugs[230,293] despite proper functioning of NER.

The preparation of site-specific chemical modifications within DNA has been an important tool for investigating structural and biochemical relationships related to DNA synthesis and repair.[54,287,294] Kowalczyk and coworkers introduced an IaCL of various alkylene chain lengths between the exocyclic N^2 atoms using a post-synthetic methodology.[269] This elegant approach to introduce site-specific IaCL requires access to the appropriate convertible nucleosides that are stable during the reactions encountered during the solid-phase oligonucleotide synthesis.

Using a combination of solution and solid phase synthesis, we and others have developed methods to introduce site specific ICL and IaCL in DNA.[295–299] However, it can be challenging to generate adducts at certain positions in the DNA scaffold. For example, N7-methylated 2'-deoxyguanosine (dG) can undergo depurination[300] or ring opening to form the formamidopyrimidine (Fapy) derivative.[301] Albeit to a lesser extent relative to the N7 position, the O^6-atom of dG can undergo alkylation from exposure to different electrophiles such as the chemotherapeutic drugs temozolomide[157] and the chloroethylating and cross-linking agent carmustine.[302] Alkylation at the O^6-site of dG is mutagenic since the disruption of proper Watson-Crick base pairing leads to misinsertions of thymine across the modified residue.[19] Moreover, *in vivo* experiments have determined the direct association of O^6-methyl-dG in carcinogenesis[39] and tumor production, [303–305] making the study of this alkylation site invaluable.

Previously, DNA duplexes containing ICL attaching the O^6-atoms of dG by alkylene linkers in various sequence motifs have been prepared and it has been shown that such cross-links between the two strands of duplex DNA can undergo repair by human O^6-alkylguanine-DNA alkyl-transferase (hAGT).[64,65] AGT proteins are found in various organisms contributing to the conservation of genomic integrity by removing the alkyl group at the O^6-atom of dG.[132,306] The mechanism of AGT repair involves flipping of the alkylated nucleotide out of the duplex and into the protein's active site, allowing the activated C145 residue to attack the carbon adjacent to the O^6-atom of dG. A tyrosyl residue (Tyr114 in hAGT) promotes the flipping of the alkylated dG residue into the active site *via* steric and / or electronic interactions. It is, however, uncertain how Tyr114 interacts with the 3'-phosphodiester linkage to aid in repair of the alkylated nucleobase.[137,168] The irreversibly alkylated protein is then rapidly degraded by the ubiquitin pathway. hAGT has demonstrated the ability to repair a number of bulky lesions at the O^6-atom of dG.[132] The substrate

specificity of hAGT has been explored for protein tagging[178] and an approach to analyze hAGT activity monitoring conformational changes on DNA origami has been reported. [307] Recently, it has been shown that hAGT can also repair O^6-methyl-dG adducts in G-quadruplex DNA adding to our knowledge of DNA structures that this enzyme can function upon.[308]

In the present study, a method to synthesize an O^6-2′-deoxyguanosine-alkylene-O^6-2′-deoxyguanosine IaCL (O^6-dG-alkylene-O^6-dG, **Figure 3.1**) has been developed. This IaCL DNA lacks a phosphodiester linkage between the connected residues which confers added flexibility of the modified position and a simplified synthetic strategy to prepare the substrates, relative to one containing a phosphodiester linkage at this modified position in the chemically synthesized oligonucleotide. The influence of this IaCL modification was assessed by UV thermal denaturation and circular dichoism experiments. Finally, repair of these novel IaCL by various AGTs (from human and *E. coli*) was evaluated to address whether these enzymes can act upon this specific O^6-alkylated substrate. Repair of DNA containing ICL with similar alkylene tethers (butylene and heptylene) by AGTs have been observed previously. Moreover, the IaCL DNA probes aid in understanding the importance of the phosphodiester linkage at the 3' and 5' of O^6-alkylated-dG for AGT-mediated repair. For this study, IaCL containing butylene and heptylene linkers were selected on the basis that butylene linked IaCL can be introduced into DNA by the action of the drug busulfan which has been demonstrated to link the N7-positions of 5'-d(GA) and 5'-d(GG) sequences.[69] It should be noted, however, that a specific IaCL between the O^6-positions of 5'-d(GG) sequences has not been identified. Hepsulfam, an alkyl sulfamate, has been shown to be proficient at producing heptylene linked ICL at the N7-positions of guanine at 5'-d(GNC) sites.

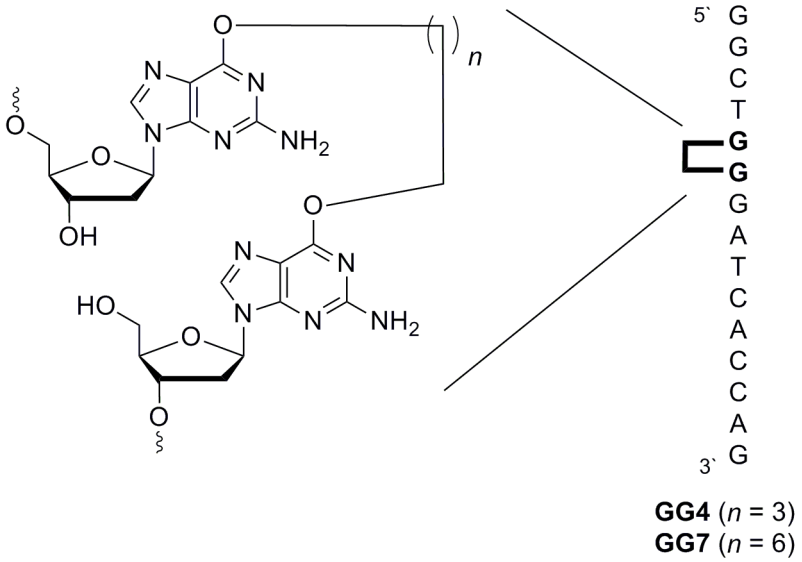


Figure 3.1: Structure of the O^6-2′-deoxyguanosine-alkylene-O^6-2′-deoxyguanosine IaCL mimic lacking a phosphodiester linkage between these two residues. The position of this IaCL mimic is shown by **GG** (where **GG4** has a butylene linker and **GG7** a heptylene linker).

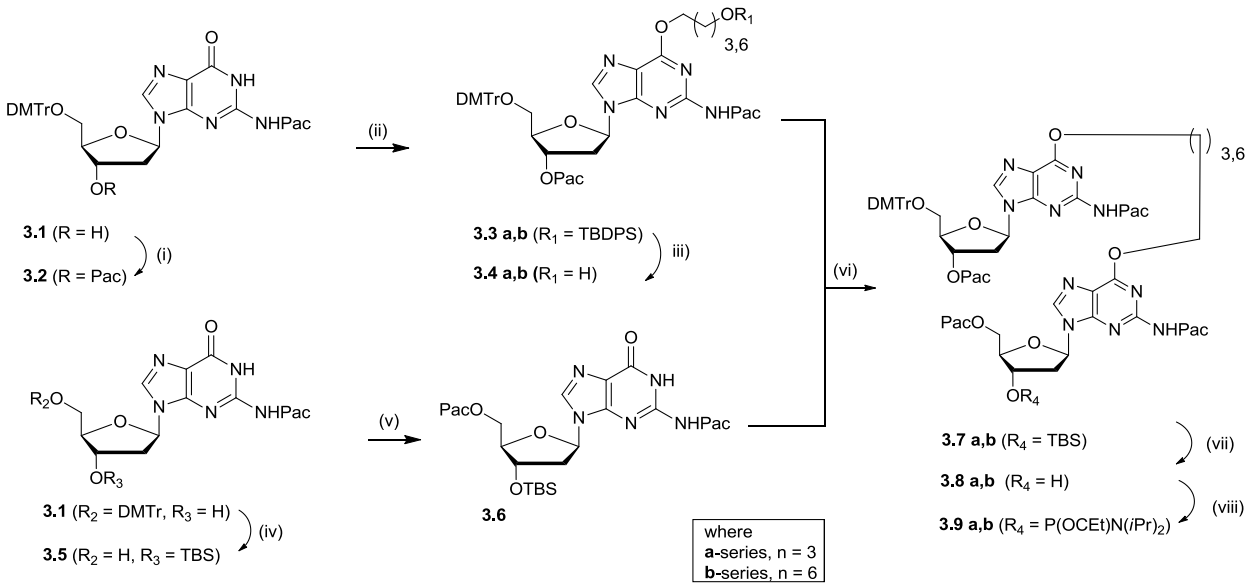
3.2 Results and Discussion

3.2.1 Synthesis and characterization of GG IaCL DNA

The investigation of the various AGTs to repair an O^6-dG-alkylene-O^6-dG IaCL required the synthesis of oligonucleotides containing this modification in a defined position. It has been demonstrated that hAGT is proficient at repairing butylene and heptylene linked O^6 -dG ICL.[64,65,151]

The structure of the O^6-dG-alkylene-O^6-dG IaCL (**GG4** and **GG7**) is shown in **Figure 3.1**. The synthesis of the cross-linked amidites **3.9a** and **3.9b**, required for the synthesis of **GG4** and **GG7**, is shown in **Scheme 3.1**. Starting with commercially available 5'-*O*-dimethoxytrityl-N^2-phenoxyacetyl-2'-deoxyguanosine **3.1**, the 3'-OH position was protected with a phenoxyacetyl group to furnish **3.2** in decent yield. The phenoxyacetyl, rather than an acetyl group, was selected for protection at the 3'-OH position due to possible transamidation observed at the exocyclic N^2 during oligonucleotide assembly and for its facile removal under mild conditions which is ideal for the synthesis of O^6-alkylated adducts.[249,309,310] Coupling of

compound **3.2** at the O^6-position with the appropriate *tert*-butyldiphenylsiloxy mono-protected diol (1,4-butanediol or 1,7-heptanediol) was accomplished by the Mitsunobu reaction. This reaction is known for its selectivity at introducing alkyl groups at the O^6-position of protected dG and has been utilized previously to generate O^6-alkylene-linked dimers containing dG for the preparation of ICL duplexes.[65,151,311,312] Slightly diminished yields were observed for this reaction relative to those previously reported,[65,151,311,312] which may be attributed to some loss of the Pac group under the mild basic conditions used to prepare compounds **3.3a** and **3.3b**. Desilylation to unmask the terminal hydroxyl of the linker produced nucleosides **3.4a** and **3.4b** for the dimerization step. In parallel, a second monomer was prepared from **3.1** starting with the introduction of a *tert*-butyldimethylsilyl (TBS) group at the 3'-OH and removal of the 5'-*O*-dimethoxytrityl group, respectively, followed by phenoxyacetylation to yield compound **3.6**, required to form the dimer. A second Mitsunobu reaction was performed to couple the protected monomers **3.4a** or **3.4b** with **3.6** to produce butylene or heptylene linked dimers **3.7a** or **3.7b**. Removal of the *tert*-butyldimethylsilyl group resulted in compounds **3.8a** and **3.8b** which were then phosphitylated to yield the phosphoramidites **3.9a** and **3.9b**, respectively. The masses observed by HRMS analysis of phosphoramidites **3.9a** and **3.9b** were in agreement with the expected values. ^{31}P NMR analysis revealed two sharp signals for compounds **3.9a** (148.62 ppm and 148.57 ppm) and **3.9b** (148.60 ppm and 148.55 ppm).



Scheme 3.1: Reagents and conditions for **a**-series: (i) Pac-Cl (1.5 eq), NEt_3 (2 eq), Py, 21°C, 30 min. (ii) 4-(*tert*-butyldiphenylsilyloxy)butan-1-ol (1.1 eq), Ph_3P (2.1 eq), DIAD (2.1 eq), dioxane, 21°C, 16h. (iii) TBAF (1.6 eq), THF, 21°C, 30min. (iv) 1. TBS-Cl (2.2 eq), Im (4.4 eq), DMF, 21°C, 30 min, 2. *p*-TsOH•H_2O (1.2 eq), MeOH/DCM (1:4), 21°C, 30 min. (v) Pac-Cl (1.5 eq), NEt_3 (1.8 eq), THF, 21°C, 30 min. (vi) **3.6** (0.91 eq), Ph_3P (2 eq), DIAD (2 eq), dioxane, 21°C, 16h. (vii) TBAF (1M in THF) (1.2 eq), THF, 21°C, 30 min. (viii) Cl-P(OCEt)N(*i*Pr)$_2$ (1.2 eq), DIPEA (1.5 eq), THF, 21°C, 30 min. Reagents and conditions for **b**-series: Identical to **a**-series except step (ii) 7-(tert butyldiphenylsilyloxy)heptan-1-ol (1.2 eq), Ph_3P (2.0 eq), DIAD (2.0 eq), dioxane, 21°C, 16h.

IaCL DNA assembly employed either dimer phosphoramidite **3.9a** or **3.9b**, which differed only in their respective linker length. The presence of one DMTr group and phosphoramidite moiety on the dimer scaffolds of **3.9a** and **3.9b** rendered the linear construction of the IaCL oligomers **GG4** and **GG7** straightforward requiring only minor modifications to the standard solid phase synthesis methodology for unmodified DNA.[312] Wait times were extended for amidites **3.9a** and **3.9b** in order to ensure complete coupling of the dimer. **GG4** and **GG7** were deprotected and cleaved from the solid support with aqueous NH_4OH : EtOH (3 : 1 v/v) for 4h at 55°C with no detectable degradation of the cross-link observed by SAX-HPLC (see **Figure S1.1**). Analysis of the **GG4** and **GG7** oligomers revealed a deconvoluted mass in agreement with the expected mass (values shown in **Table**

3.1). Further characterization by enzymatic digest to the constituent nucleosides followed by RP-HPLC demonstrated the appearance of a new peak (in addition to the other four nucleosides) with retentions of 16.4 min and 24.8 min for the butylene and heptylene containing dG dimers, respectively.

Table 3.1: Amounts, retention times, nucleoside ratios and MS data for cross-linked oligonucleotides

Oligo	A_{260} of Crude and Purified Oligo[a]	Nucleoside	Nucleoside Ratio		Mass	
			Exp	Obs	Exp	Obs
		dC	2	2.1		
		dG	2	2.2		
GG4	60 (31)	dT	1	1.0	4610.1	4611.1
		dA	1.5	1.5		
		dG-4C-dG	0.5	0.5		
		dC	2	1.9		
		dG	2	2.1		
GG7	55 (25)	dT	1	1.0	4652.1	4653.1
		dA	1.5	1.4		
		dG-7C-dG	0.5	0.4		

[a] A_{260} of purified oligonucleotide is shown in brackets.

3.2.2 UV thermal denaturation and circular dichroism studies of IaCL DNA

Impact of this IaCL mimic on duplex stability with the complementary DNA sequence, relative to the native duplex, was assessed by UV thermal denaturation experiments. The thermal denaturation profiles of the IaCL containing DNA duplexes were monophasic with T_m values of 46°C, 48°C and 67°C for **GG4**, **GG7** and the unmodified duplex, respectively (**Figure 3.2**). Duplexes containing a single O^6-methyl-dG insert leads to a reduction in T_m by approximately 18°C and two inserts reduce duplex stability by 40°C

relative to unmodified sequences.[313–315] However, duplexes containing the butylene and heptylene linked O^6-dG-alkylene-O^6-dG IaCL (**GG4** and **GG7**) were found to decrease the T_m by 19-21°C relative to the control, which was lower than expected. The decreased stability is most-likely attributed to the disruption of H-bonding between the alkylated 2`-deoxyguanosines and base paired 2`-deoxycytidines particularly with the bulky alkylene linkers which could induce local and global structural perturbations. However, flexibility at this site due to the lack of the phosphodiester linkage may reduce this effect.

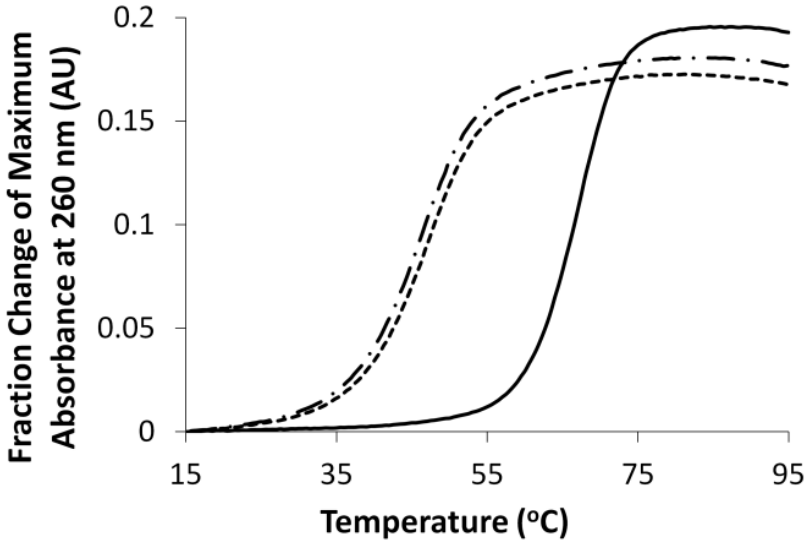


Figure 3.2: Fraction change of maximum absorbance at 260 nm (A_{260}) *versus* temperature (°C) profiles of duplexes containing **GG4** (- • -), **GG7** (- - -) and unmodified DNA (—).

Circular dichroism spectroscopy was performed to determine the global structural influence of the IaCL mimics on the duplexes. CD profiles of duplexes containing **GG4** and **GG7** displayed signatures consistent with the B-DNA family with maxima near 280 nm, cross-overs at about 260 nm and minima around 240 nm (shown in **Figure 3.3**). These results indicate minor global structural distortions were induced by the presence of the IaCL. Similar findings were observed for mono-adducts at the O^6-atom of dG (data not shown). Structures of unmodified, **GG4** and **GG7** were geometry optimized using the AMBER force field. In both cases, the alkylene linker protruded into the major groove, with a slight widening

observed. The effect was more pronounced for **GG4** than **GG7**, relative to the control. The molecular models suggest that the presence of the cross-link does not greatly distort the global structure of the duplex (see **Figure S3.5** for models).

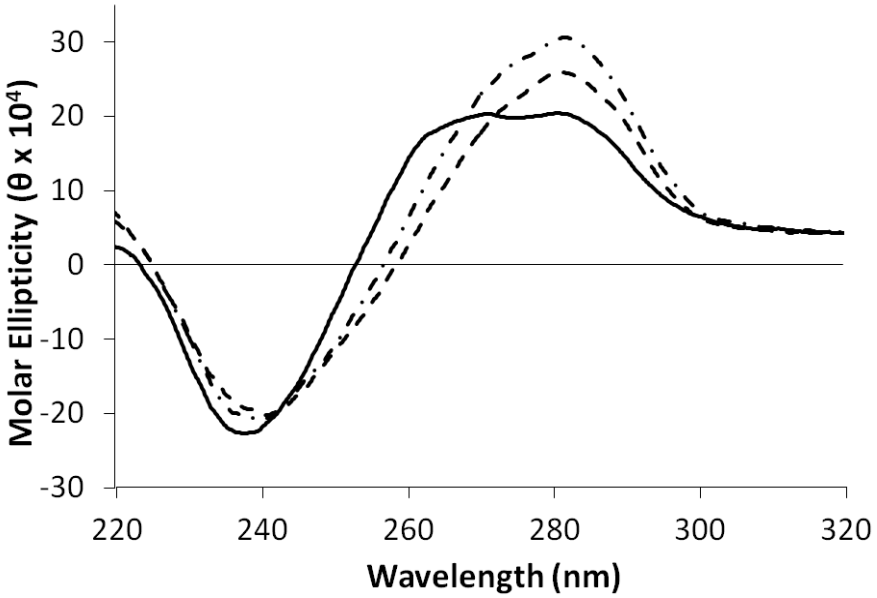


Figure 3.3: Circular dichroism spectra of IaCL duplexes **GG4** (- • -), **GG7** (- - -) and unmodified DNA (—).

3.2.3 AGT repair of IaCL DNA substrates

Four AGT proteins (hAGT, OGT, a S134P OGT variant, and Ada-C) were evaluated for their ability to repair duplexes **GG4** and **GG7**. The damaged strand was radio-labeled using γ-[^{32}P]-ATP, annealed with an equimolar amount of the complement followed by incubation overnight at 37^{o}C with the AGT (2 pmol DNA and 10 pmol protein). The repair reactions were quenched in stop reaction buffer to prevent complexation before loading onto the denaturing gel. Repair assays conducted on **GG4** and **GG7** revealed that only hAGT was capable of repair (lanes 2 and 7 in **Figure 4**). The lack of OGT and Ada-C repair was not surprising given their low proficiency in repairing lesions larger than a methyl group at the O^6-atom of dG. Despite being capable of repairing larger O^4-alkylated thymidine mono-adducts,[150] OGT was incapable of processing the IaCL even at higher AGT concentration

ratios and for elongated reaction times (data not shown). The OGT S134P variant,[149] presumably containing a larger active site, was also tested against the **GG4** and **GG7** duplexes and demonstrated no repair activity.

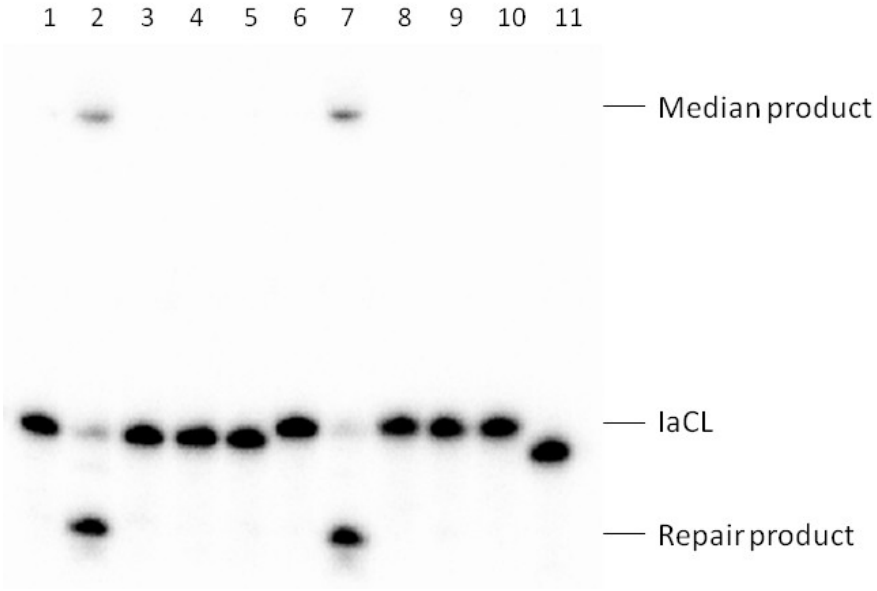


Figure 3.4: Repair of **GG4** and **GG7** by hAGT, OGT, Ada-C, and OGT S134P. Denaturing PAGE of repair reactions as described in the experimental section: Lane 1, 2 pmol **GG4**; lane 2, 2 pmol **GG4** + 10 pmol hAGT; lane 3, 2 pmol **GG4** + 10 pmol OGT; lane 4, 2 pmol **GG4** + 10 pmol Ada-C; lane 5, 2 pmol **GG4** + 10 pmol OGT S134P; lane 6, 2 pmol **GG7**; lane 7, 2 pmol **GG7** + 10 pmol hAGT; lane 8, 2 pmol **GG7** + 10 pmol OGT; lane 9, 2 pmol **GG7** + 10 pmol Ada-C; lane 10, 2 pmol **GG7** + 10 pmol OGT S134P; lane 11, 2 pmol control unmodified DNA + 10pmol hAGT.

The repair of **GG4** and **GG7** by hAGT is proposed to proceed by a first reaction to generate a DNA-protein cross-link, which can then undergo a second repair event by another AGT molecule ultimately forming a protein-protein cross-link (**Figure 3.5**). Given the lack of the phosphodiester linkage at the cross-link site, repair of the IaCL DNA would produce two smaller DNA fragments, which are resolvable on PAGE and detected by autoradiography. Trials were initially carried out overnight at 37^{o}C and hAGT displayed proficient repair of the O^6-dG-alkylene-O^6-dG IaCL.

Lanes 1 and 6 contain the **GG4** and **GG7** ssDNA, respectively. Lane 11 contains the unmodified ssDNA strand (with a phosphodiester linkage at the corresponding site relative to the IaCL ssDNA) of identical nucleotide sequence as **GG4** and **GG7**. As shown in **Figure 3.4**, the presence of the alkylene appendage and having one less phosphodiester linkage reduces the DNA migration on the denaturing gel with the unmodified ssDNA migrating fastest and **GG7** the slowest. Lanes 2 and 7 display the repair reaction of **GG4** and **GG7** (2 pmol), respectively, with 10 pmol of hAGT. In both cases the band corresponding to the intact IaCL DNA disappears and two new bands appear, one which migrates much slower compared to the substrate and another that migrates faster. The slow migrating band most-likely corresponds to the hAGT covalently attached to the DNA species, as observed previously with AGT repair of O^6-dG-alkylene-O^6-dG ICL DNA.[65,151] The faster migrating species is likely the smaller [^{32}P]-labeled repair product.

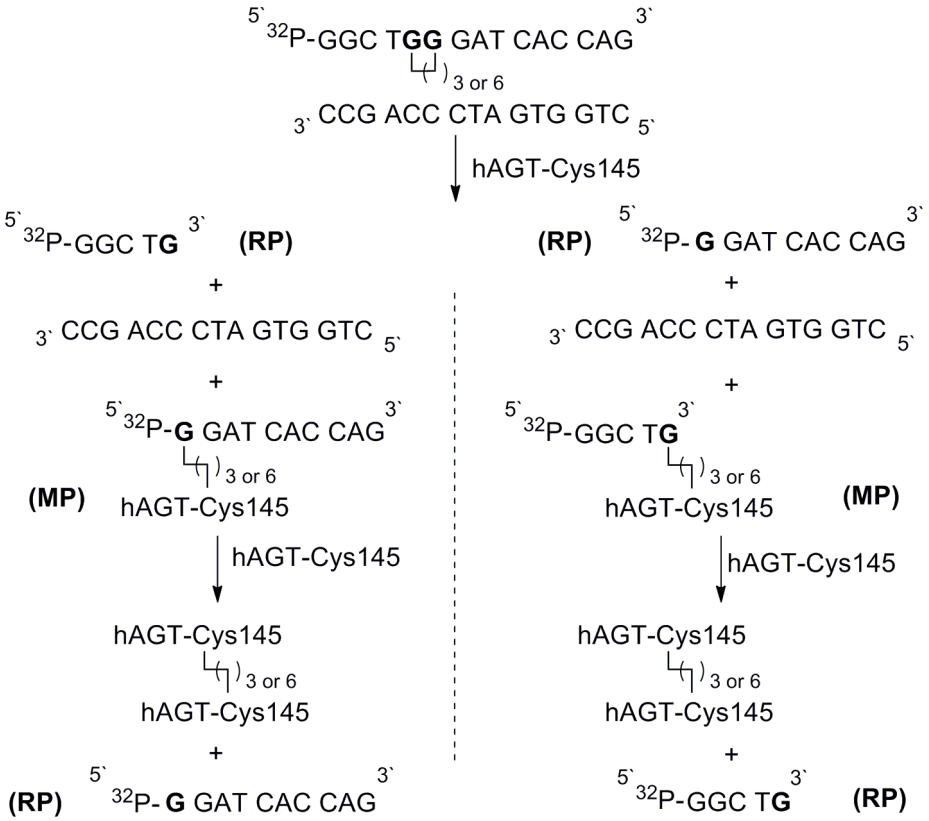


Figure 3.5: Proposed repair pathway of IaCL species by wild-type hAGT (**MP** - median product; **RP**- repair product).

The repair assays revealed that hAGT was more proficient in repairing **GG7** compared to **GG4**; leaving approximately 2 % and 15 % of unrepaired **GG7** and **GG4** substrates, respectively. Median products (**MP**) consisted of 6 % and about 9 % of the total counts for **GG7** and **GG4** experiments. Finally, repair products (**RP**) were generated at levels of 91 % and 76 % for **GG7** and **GG4**, respectively. Interestingly, these results follow a similar trend as repair of ICL DNA containing identical-length alkylene linkers. That is, the size of the linker affected hAGT repair with higher efficiency observed for heptylene linkers compared to butylene linkers.[65] Repair of the **GG7** and **GG4** by hAGT was greater compared to the corresponding ICL DNA by about 41 % and 54 %, respectively. These results suggest that the O^6-dG-alkylene-O^6-dG IaCL DNA are more-readily repaired by hAGT compared to ICL DNA.

Time course assays were performed using 10 pmol hAGT and 2 pmol of either **GG4** or **GG7** substrates by quantifying the amounts of unrepaired, **MP** and **RP** products (see **Figure S3.7** for denaturing PAGE). **GG7** was repaired faster with respect to **GG4** (**Figure 3.6**). The formation of the **RP** reached the plateau faster in the case of **GG7** compared to **GG4**. After repair reaction times of 15 min and 8h, 80% and 18 % of **GG4** remained, respectively, whereas only 21% and 2 % of **GG7** remained, respectively. The greater efficiency of repair by hAGT was also observed for heptylene versus butylene O^6-dG-alkylene-O^6-dG ICL DNA.[65] The current understanding of the AGT-mediated repair mechanism involves flipping the O^6-modified nucleotide into the active site followed by alkyl transfer to the activated thiol of C145.[137] The enhanced repair by hAGT observed for O^6-dG-alkylene-O^6-dG IaCL relative to ICL DNA may be attributed to additional strain involved in flipping the O^6-alkylated nucleotide tethered to the complementary strand of the ICL, relative to the IaCL lesion which resides on one strand only. Protein binding on the complementary alkylated strand of the ICL DNA may also be a contributing factor to reduced repair relative

to the IaCL. The faster repair of **GG7** compared to **GG4** is most-likely because the greater flexibility of the larger linker, which facilitates rotation of the nucleotide out of the duplex. The geometry optimized structures displayed an O^6-O^6 distance for adjacent dG nucleotides of 3.1 Å, 3.3 Å and 4.0 Å for **GG4**, **GG7** and unmodified duplex. Increased tilting of the cross-linked guanine bases was seen in the case of **GG4** relative to **GG7**, which suggests that more strain is observed with the shorter alkylene linker. As a result, the guanine bases in **GG7** are presumably more easily rotated into the active site of the protein for repair. Additional strain observed in **GG4** could explain why this substrate is not fully repaired with extended repair assay times, and the **GG7** substrates are almost fully consumed.

Formation of **RP** and **MP** occurs at different rates with **RP** forming faster compared to **MP**. This would suggest, as in the case of ICL DNA repair, that hAGT preferentially repairs the **MP** over the initial substrate since no accumulation of **MP** is observed at any time. **GG7** repair shows a crossover between the substrate consumption curve and **MP** formation curve. This was not observed in the case of **GG4** nor any ICL DNA previously tested.[64,65] Moreover, **GG7** was virtually completely consumed by a fivefold excess of hAGT protein after extended reaction times (8h), which was not observed in the case of the interstrand cross-linked probes. As for ICL DNA, **GG4** repair was not observed to proceed to completion over time which presumably was due to reacted hAGT binding to another DNA containing lesion.[169] Another possible explanation is that the alkylene linker adopts a less reactive conformation, which has been proposed for other O^6-alkylated-dG lesions.[316]

Using *in vitro* assays, these results demonstrate that IaCL DNA containing four and seven carbon atoms linking two adjacent O^6-dG residues, lacking the phosphate group, are efficiently repaired by hAGT adding to our knowledge of the range of substrates that this repair protein can act upon. Intrastrand cross-links can be formed in DNA by the bifunctional alkylating drug busulfan and other chemotherapeutic drugs, although to our knowledge no

IaCL linking the O^6-atoms of dG has been identified. In a previous study, it was shown that AGT-deficient CHO cells were sensitive to killing by hepsulfam.[65] These results suggest that expression of AGT protects, to a certain extent, CHO cells from exposure to hepsulfam but not busulfan. These preliminary results are also consistent with the more efficient repair observed for the heptylene versus butylene IaCL adducts by hAGT.

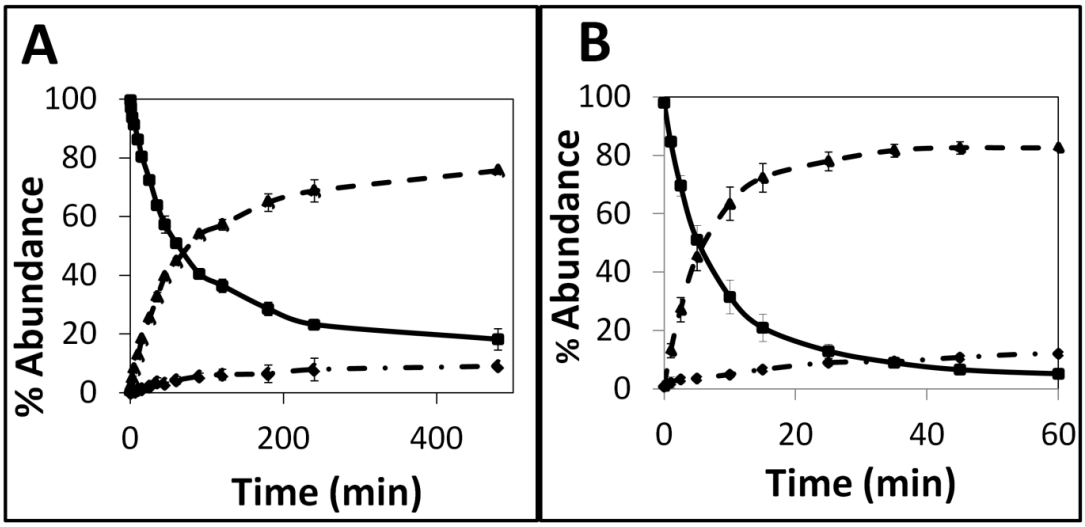


Figure 3.6: Time course repair assay of **GG4** (A) and **GG7** (B) by hAGT displaying significant faster repair of **GG7** by about a 14-fold factor. Graphical illustrations display abundances (%) of **MP** (- • -), **RP** (- - -) and substrate (—) over time (min).

One feature of the O^6-dG-alkylene-O^6-dG IaCL is the lack of a 5', 3'-phosphodiester linkage between the nucleotides at the modified site. Previous crystallographic studies have suggested that Tyr114 is necessary for flipping of the damaged nucleotide mediated by steric repulsion between this residue of AGT with the 3'-phosphate group of the target nucleotide.[137] X-ray crystallography studies of hAGT and O^6-MedG or N1,O^6-ethanoxanthine-containing DNA complexes have illustrated the role of this steric interaction between Tyr114 and the 3'-phosphate group.[137] This was further supported by studies using Y114F and Y114A mutants, which showed significant reduction in activity towards O^6-alkylated-dG DNA adducts. A computational approach to elucidate the AGT repair

mechanism pathway deciphered a different role for the Tyr114, whereby interactions would be mediated solely *via* electronics, rather than sterics.[168]

In the current study, the O^6-dG-alkylene-O^6-dG IaCL which lacks a 3'-phosphate group between the two O^6-modified-dG residues is observed to undergo repair by hAGT. Comparison of AGT repair of this flexible O^6-dG-alkylene-O^6-dG IaCL with the corresponding lesion containing a phosphodiester linkage at this site may contribute to probing the steric versus electronic role that the Tyr114 residue plays in AGT-mediated repair. Investigation of the synthetic probes herein sheds light on AGT repair of O^6-alkylated-dG lacking the 3'-phosphate group further contributing to our understanding of AGTs. Further work on IaCL DNA and AGT mutants will provide insights on the effect of the 3'-phosphate group on repair by AGTs as well.

3.3 Conclusions

DNA duplexes containing an O^6-2'-deoxyguanosine-alkylene-O^6-2'-deoxyguanosine IaCL were prepared using a nucleoside dimer phosphoramidite, followed by solid-phase synthesis assembly. IaCL DNA were deprotected and cleaved off solid support using standard protocols and purification was achieved by SAX-HPLC to obtain the probes in high yields and purity. Thermal denaturation analysis of duplexes containing IaCL of butylene and heptylene linkages were found to have similar T_m values; both of which led to a reduction in stability by 20^{o}C relative to the unmodified duplex. Circular dichroism and molecular modeling suggested minimal global perturbation in the structure. Both **GG4** and **GG7** were repaired by hAGT, with a much higher efficiency for IaCL containing a heptylene versus butylene linker, despite lacking the phosphate group at the cross-link site. These results serve an additional basis to ongoing studies related to AGT-mediated resistance to bifunctional chemotherapeutic agents.

3.4 Experimental Section

3.4.1 Purification and characterization of oligomer containing the phosphateless IaCL

The cross-linked duplexes, whose sequences are shown in **Figure 3.1**, were assembled on an Applied Biosystems Model 3400 synthesizer on a 1.5 μmol scale using standard β-cyanoethylphosphoramidite chemistry supplied by the manufacturer with slight modifications to coupling times. The nucleoside phosphoramidites containing standard protecting groups were prepared in anhydrous MeCN at a concentration of 0.1 M for the 3'-*O*-deoxyphosphoramidites, and 0.15 M for the cross-linked 3'-*O*-deoxyphosphoramidite. Oligomer sequence assembly was carried out as follows: (1) detritylation, 3% trichloroacetic acid in dichloromethane; (2) nucleoside phosphoramidite coupling time of 2 min for commercial 3'-*O*-deoxyphosphoramidites, and 10 min for the cross-linked phosphoramidites **3.9a** or **3.9b**; (3) capping, phenoxyacetic anhydride/pyridine/ tetrahydrofuran 1:1:8 (v/v/v; solution A) and 1-methyl-imidazole/ tetrahydrofuran 16:84 (w/v; solution B); (4) oxidation, 0.02 M iodine in tetrahydrofuran/water/pyridine 2.5:2:1.

Protecting groups and cleavage from the solid support was carried out by treatment with aqueous (28% v/v) ammonium hydroxide in EtOH (1 mL of a 3 : 1 v/v solution) for 4h at 55 ^{o}C in 2 mL screw cap microfuge tubes fitted with Teflon lined caps. Crude oligomers were transferred and evaporated in a speedvac concentrator followed by purification by strong-anion exchange HPLC using a Dionex DNAPAC PA-100 column (0.4 cm × 25 cm) purchased from Dionex Corp, Sunnyvale, CA using a linear gradient of 0-52% buffer B over 24 min (buffer A: 100 mM Tris HCl, pH 7.5, 10% acetonitrile and buffer B: 100mM Tris HCl, pH 7.5, 10% acetonitrile, 1 M NaCl) at 55^{o}C. The columns were monitored at 260 nm for analytical runs or 280 nm for preparative runs. The purified oligomers were desalted using C-18 SEP PAK cartridges (Waters Inc.) as previously described.[317] The amounts of

purified oligomers obtained are shown in **Table 3.1** (crude A_{260} of 60 and 55 and pure A_{260} of 31 and 25 for **GG4** and **GG7**, respectively).

Modified oligomers (0.05 A_{260} units) were characterized by digestion with snake venom phosphodiesterase (0.28 units) and calf intestinal phosphatase (5 units) in a buffer containing 10 mM Tris (pH 8.1) and 2 mM magnesium chloride for 2 days at 37 °C as previously described.[298] The resulting mixture of nucleosides was analyzed by reversed phase HPLC using a Symmetry C-18 5 μm column (0.46 × 15 cm) purchased from Waters Inc., Milford, MA. A linear gradient of 0–30% buffer B over 30 min (buffer A, 50 mM sodium phosphate, pH 5.8 and buffer B, 50 mM sodium phosphate, pH 5.8, 50% acetonitrile) was used to elute the analytes. The peaks in the elution profile were identified by coinjection with the corresponding standards with the following elution times: dC (4.5 min), dG (7.4 min), dT (8.0 min), dA (9.0 min), and cross-linked dimers (16.4 min and 24.8 min for the butylene and heptylene dimers, respectively), and the ratio of nucleosides was determined. The results are given in **Table 3.1**. The molecular mass of the modified oligomers were identifed by ESI-MS and the measured values were in agreement with the expected masses (see **Figure S3.3** and **Figure S3.4** for MS spectra).

3.4.2 Protein expression and purification

All AGT homologues were expressed under the promoter of the pQE30 vector in XL-10 Gold *E. coli* cells as previously described.[65,150,151]

3.5 Acknowledgement

We are grateful to Dr. Anthony E. Pegg (Pennsylvania State University) for the plasmid encoding the wild-type hAGT, OGT and Ada-C genes. We are also grateful to Dr. Francis McManus for helpful discussions concerning the repair assays and Dr. Anne Noronha for assistance with solid phase synthesis. This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC), the Canada

Foundation for Innovation (CFI) and the Canada Research Chair (CRC) program. D.K.O. is the recipient of a Canada Graduate Scholarship (CGS) from NSERC.

3.6 Supporting information

3.6.1 Supporting Methods

3.6.1.1 General

5'-*O*-(4,4'-Dimethoxytrityl)-N^2-phenoxyacetyl-2'-deoxyguanosine and N,N-diisopropylaminocyanoethylphosphonamidic chloride were purchased from ChemGenes, Inc. (Wilmington, MA). 5'-*O*-(4,4'-Dimethoxytrityl-2'-deoxyribonucleoside-3'-*O*-(β-cyanoethyl-N,N-diisopropyl)phosphoramidites and protected 2'-deoxyribonucleoside–CPG supports were purchased from Glen Research (Sterling, Virginia). All other chemicals and solvents were purchased from the Aldrich Chemical Company (Milwaukee, WI) or EMD Chemicals Inc. (Gibbstown, NJ). Flash column chromatography was performed using silica gel 60 (230–400 mesh) purchased from Silicycle (Quebec City, QC). Thin layer chromatography (TLC) was carried out with precoated TLC plates (Merck, Kieselgel 60 F_{254}, 0.25 mm) purchased from EMD Chemicals Inc. (Gibbstown, NJ). NMR spectra were recorded on a Varian 500 MHz NMR spectrometer at room temperature. ^{1}H NMR spectra were recorded at a frequency of 500.0 MHz and chemical shifts were reported in parts per million (ppm) downfield from tetramethylsilane. ^{13}C NMR spectra (^{1}H decoupled) were recorded at a frequency of 125.7 MHz and chemical shifts were reported in ppm with tetramethylsilane as a reference. ^{31}P NMR spectra (^{1}H decoupled) were recorded at a frequency of 202.3 MHz and chemical shifts were reported in ppm with H_3PO_4 used as an external standard. High resolution mass spectrometry of modified nucleosides were obtained using an LTQ OrbitrapVelos – ETD (Thermo Scientific) at the Concordia University Centre for Biological Applications of Mass Spectrometry (CBAMS). The mass spectrometer was operated in full scan, positive ion detection mode. ESI mass spectra for oligonucleotides were obtained at CBAMS using a

Micromass Qtof2 mass spectrometer (Waters) equipped with a nanospray ion source. The mass spectrometer was operated in full scan, negative ion detection mode. Ampicillin, isopropyl β-D-thiogalactopyranoside (IPTG), and most other biochemical reagents as well as polyacrylamide gel materials were purchased from Bioshop Canada Inc (Burlington, ON). Ni-NTA Superflow Resin was purchased from Qiagen (Mississauga, ON). Complete, Mini, EDTA-free Protease Inhibitor Cocktail Tablets were obtained from Roche (Laval, QC) Nitro-cellulose filters (0.20 μm) were obtained from Millipore. XL-10 Gold and Bl21(DE3) *E. coli* cells were obtained from Stratagene (Cedar Creek, TX). T4 polynucleotide kinase (PNK) was obtained from Fermentas (Burlington, ON). γ-[^{32}P]-ATP was purchased from Amersham Canada Ltd (Oakville, ON). Phusion Polymerase was obtained from New England Biolabs (Ipswich, MA). DNA primers for site directed mutagenesis and cloning were purchased from Biocorp (Montreal, QC).

3.6.1.2 Chemical Synthesis of nucleosides

The synthesis of phosphoramidites **3.9a** and **3.9b** are shown in **Scheme 3.1**

3'-O-N^2-bis(phenoxyacetyl)-5'-O-(4,4'-dimethoxytrityl)- 2'-deoxyguanosine **(3.2):**

To a solution of 5'-*O*-(4,4'-dimethoxytrityl)- *N*^{2}-phenoxyacetyl-2'-deoxyguanosine (2.5 g, 3.6 mmol) and TEA (1.8 mL, 12.9 mmol) in THF (25 mL) was added phenoxyacetyl chloride (1.1 mL, 8.2 mmol) dropwise while stirring at 0^{o}C. After 30 min, the solvent was removed in vacuo and the content taken up in DCM (50 mL), washed with $NaHCO_3$ (aq, 3%) (2 x 50mL), dried over anhydrous Na_2SO_4, decanted and the solvent removed to produce a yellow gum. The product was purified *via* flash column chromatography using MeOH : DCM (1 % → 3 % by increments of 0.5 %) to afford 2.4 g (79 %) of a yellow foam. R_f (SiO_2 TLC):0.23 MeOH : DCM (5 %). $\lambda_{max\ (ACN)}$ = 276 nm. ^{1}H NMR (500MHz, $CDCl_3$, ppm): 11.77 (s, 1H, NH), 9.05 (s, 1H, NH), 7.78 (s, 1H, H8), 7.40-7.17(m, 14H, Ar), 7.09-7.07 (m, 1H, Ar), 7.04-6.97 (m, 2H, Ar), 6.93-6.90 (m, 2H, Ar), 6.79-6.76 (m, 4H, Ar), 6.17 (dd, 1H,

H1', *J* = 7 Hz), 5.69 (m, 1H, H3'), 4.67 (s, 2H, $PhOCH_2CO$), 4.64 (s, 2H, $PhOCH_2CO$), 4.24, (m, 1H, H4'), 3.76 (s, 6H, 2 x OCH_3), 3.41 (m, 1H, H5'), 3.34 (m, 1H, H5''), 2.91 (m, 1H, H2'), 2.57 (m, 1H, H2''). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 169.5, 168.4, 158.63, 158.62, 156.4, 147.6, 146.2, 144.2, 137.3, 135.35, 135.31, 130.03, 130.01, 129.7, 128.0, 127.9, 127.0, 123.0, 122.2, 122.0, 114.9, 114.6, 113.2, 86.8, 83.9, 83.7, 75.7, 66.9, 65.3, 63.3, 55.2, 37.7. IR (thin film); ν_{max} (cm^{-1}) =3217, 3059, 2931, 2836, 1762, 1698, 1609, 1557, 1508, 1495, 1250, 1175, 1086, 1032. HRMS (ESI-MS) *m/z* calculated for $C_{47}H_{44}N_5O_{10}^+$ 838.3083: found 838.3122 $[M+H]^+$.

3'-O-N^2-bis(phenoxyacetyl)-5'-O-(4,4'-dimethoxytrityl)-O^6-(4-(tert-butyldiphenylsilyloxy)butyl)-2'-deoxyguanosine **(3.3a):**

To a solution of **3.2** (2.2 g, 2.7 mmol), 4-(*tert*-butyldiphenylsilyloxy)butan-1-ol (1.0 g, 3.2 mmol), and Ph_3P (1.7 g, 6.4 mmol) in dioxane (11 mL) was added DIAD (1.1 mL, 6.4 mmol) dropwise while stirring. After 16 h, the solvent was removed in vacuo and the content taken up in DCM (50 mL), washed with $NaHCO_3$ (aq, 3%) (2 x 75 mL), dried over anhydrous Na_2SO_4, decanted and the solvent removed to produce a yellow gum. The product was purified *via* flash column chromatography using EtOAc : hexanes (3 : 7, 2 : 3) to afford 1.0 g (66 %) of a colorless foam. R_f (SiO_2 TLC):0.53 EtOAc : Hex (1 : 1). $\lambda_{max\ (MeCN)}$ = 270 nm. ^1H NMR (500MHz, $CDCl_3$, ppm): 8.64 (s, 1H, NH), 7.98 (s, 1H, H8), 7.67-7.65 (m, 4H, Ar), 7.41-7.16 (m, 19H, Ar), 7.06-7.00 (m, 4H, Ar), 6.93-6.91 (m, 2H, Ar), 6.80-6.75 (m, 4H, Ar), 6.39 (dd, 1H, H1', *J* = 7 Hz), 5.68 (m, 1H, H3'), 4.74 (s, 2H, $PhOCH_2CO$), 4.69 (s, 2H, $PhOCH_2CO$), 4.60 (t, 2H, $ArOCH_2$, *J* = 7 Hz), 4.26 (m, 1H, H4'), 3.75-3.70 (m, 8H, 2 x OCH_3 & CH_2OSi), 3.51 (m, 1H, H5'), 3.39 (m, 1H, H5''), 3.00 (m, 1H, H2'), 2.63 (m, 1H, H2''), 2.02 (m, 2H, CH_2CH_2), 1.77 (m, 2H, CH_2CH_2), 1.05 (s, 9H, $(CH_3)_3CSi$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 168.4, 161.3, 158.59, 158.58, 157.6, 152.5, 151.2, 144.4, 139.8, 135.6, 135.48, 135.45, 133.9, 130.04, 130.02, 129.8, 129.7, 129.6, 127.0, 122.4, 122.0, 118.9,

115.0, 114.7, 113.2, 86.8, 84.3, 84.0, 76.2, 68.0, 67.9, 65.3, 63.7, 63.4, 55.2, 38.0, 29.0, 26.9, 25.5, 19.2. IR (thin film); ν_{max} (cm^{-1}) = 3409, 3069, 2931, 2856, 1764, 1724, 1696, 1605, 1510, 1496, 1427, 1248, 1176, 1110, 1087, 1034. HRMS (ESI-MS) *m/z* calculated for $C_{67}H_{70}N_5O_{11}Si^+$ 1148.4836: found 1148.4893 $[M+H]^+$.

3'-O-N^2-bis(phenoxyacetyl)-5'-O-(4,4'-dimethoxytrityl)-O^6-(7-(tert-butyldiphenylsilyloxy)heptyl)-2'-deoxyguanosine **(3.3b):**

To a solution of **3.2** (2.1 g, 2.4 mmol), 7-(*tert*-butyldiphenylsilyloxy)heptan-1-ol (1.1 g, 2.9 mmol), and Ph_3P (1.3 g, 4.9 mmol) in dioxane (8.2 mL) was added DIAD (1.0 mL, 4.9 mmol) dropwise while stirring. After 16 h, the solvent was removed in vacuo and the content taken up in EtOAc (50 mL), washed with $NaHCO_3$ (aq, 3%) (2 x 75 mL) and brine (75 mL), dried over anhydrous Na_2SO_4, decanted and the solvent removed to produce a yellow gum. The product was purified *via* flash column chromatography using EtOAc : hexanes (3 : 7, 2 : 3) to afford 2.0 g (67 %) of a colorless foam. R_f (SiO_2 TLC):0.56 EtOAc : Hex (1 : 1). λ_{max} (ACN) = 269 nm. ^{1}H NMR (500MHz, $CDCl_3$, ppm): 8.69 (s, 1H, NH), 7.98 (s, 1H, H8), 7.67-7.66 (m, 4H, Ar), 7.42-7.16 (m, 19H, Ar), 7.07-6.99 (m, 4H, Ar), 6.93-6.91 (m, 2H, Ar), 6.77-6.76 (m, 4H, Ar), 6.38 (dd, 1H, H1', J = 6 Hz), 5.68 (m, 1H, H3'), 4.74 (s, 2H, $PhOCH_2CO$), 4.69 (s, 2H, $PhOCH_2CO$), 4.57 (t, 2H, $ArOCH_2$, J = 6.75 Hz), 4.26 (m, 1H, H4'), 3.75 (s, 6H, 2 x OCH_3), 3.65 (t, 2H, CH_2OSi, J = 6.5 Hz), 3.51 (m, 1H, H5'), 3.37 (m, 1H, H5''), 3.01 (m, 1H, H2'), 2.63 (m, 1H, H2''), 1.87 (m, 2H, CH_2), 1.57 (m, 2H, CH_2), 1.47 (m, 2H, CH_2), 1.36 (m, 4H, CH_2CH_2), 1.04 (s, 9H, $(CH_3)_3CSi$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 168.4, 161.4, 158.6, 157.6, 152.5, 151.2, 144.4, 139.8, 135.6, 135.48, 135.45, 134.2, 130.04, 130.02, 129.8, 129.7, 129.5, 128.1, 127.9, 127.6, 127.0, 122.4, 122.0, 115.0, 114.7, 113.2, 86.8, 84.3, 84.0, 76.2, 68.0, 65.3, 63.9, 63.7, 55.2, 38.0, 32.5, 29.7, 29.2, 28.8, 26.9, 25.8, 19.2. IR (thin film); ν_{max} (cm^{-1}) = 3407, 3054, 2931, 2856, 2361, 1764, 1723,

1606, 1509, 1496, 1111, 1086, 1035. HRMS (ESI-MS) *m/z* calculated for $C_{70}H_{76}N_5O_{11}Si^+$ 1190.5305: found 1190.5358 $[M+H]^+$.

3'-O-N^2-bis(phenoxyacetyl)-5'-O-(4,4'-dimethoxytrityl)-O^6-(hydroxybutyl)-2'-deoxyguanosine **(3.4a)**:

To a solution of **3.3a** (1.0 g, 0.87 mmol) in THF (5 mL) was added TBAF (1.4 mL of a 1M solution in THF) dropwise over 30 seconds. After 30 min, the solvent was removed in vacuo and purified *via* flash column chromatography using MeOH : DCM (2 %, 3 %) to afford 0.50 g (63 %) of a colorless foam. R_f (SiO_2 TLC):0.27 MeOH : DCM (4 %). $\lambda_{max\ (MeCN)}$ = 269 nm. ^{1}H NMR (500MHz, $CDCl_3$, ppm): 8.76 (s, 1H, NH), 7.97 (s, 1H, H8), 7.37-7.16 (m, 13H, Ar), 7.07-6.99 (m, 4H, Ar), 6.93-6.91 (m, 2H, Ar), 6.80-6.75 (m, 4H, Ar), 6.37 (dd, 1H, H1', *J* = 6 & 8 Hz), 5.68 (m, 1H, H3'), 4.72 (s, 2H, $PhOCH_2CO$), 4.68 (s, 2H, $PhOCH_2CO$), 4.64 (t, 2H, $ArOCH_2$, *J* = 3 Hz), 4.26 (m, 1H, H4'), 3.75-3.71 (m, 9H, 2 x OCH_3 & CH_2OH), 3.50 (m, 1H, H5'), 3.38 (m, 1H, H5''), 3.01 (m, 1H, H2'), 2.63 (m, 1H, H2''), 2.01 (m, 2H, CH_2CH_2), 1.77 (m, 2H, CH_2CH_2). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 168.4, 161.2, 158.58, 158.57, 157.6, 157.1, 152.5, 151.2, 144.4, 139.9, 135.5, 135.4, 130.03, 130.01, 129.9, 129.7, 129.6, 128.1, 127.9, 127.0, 122.4, 122.0, 118.9, 115.0, 114.7, 113.2, 86.8, 84.3, 84.0, 76.1, 68.0, 67.6, 65.3, 63.7, 62.2, 55.2, 38.0, 29.0, 25.2. IR (thin film); ν_{max} (cm^{-1}) = 3359, 3192, 3057, 2955, 2922, 2851, 2298, 1766, 1723, 1606, 1510, 1496, 1421, 1249, 1175, 1084, 1033. HRMS (ESI-MS) *m/z* calculated for $C_{51}H_{52}N_5O_{11}^+$ 910.3658: found 910.3699 $[M+H]^+$.

3'-O-N^2-bis(phenoxyacetyl)-5'-O-(4,4'-dimethoxytrityl)-O^6-(hydroxyheptyl)-2'-deoxyguanosine **(3.4b)**:

To a solution of **3.3b** (1.0 g, 0.84 mmol) in THF (5 mL) was added TBAF (1.5 mL of a 1M solution in THF) dropwise over 30 seconds. After 30 min, the solvent was removed in vacuo and purified *via* flash column chromatography using MeOH : DCM (2 %, 3.5 %) to

afford 0.54 g (68 %) of a colorless foam. R_f (SiO_2 TLC):0.18 MeOH : DCM (2 %). $\lambda_{max\ (MeCN)}$ = 269 nm. ^{1}H NMR (500MHz, $CDCl_3$, ppm): 8.72 (s, 1H, NH), 7.98 (s, 1H, H8), 7.37-7.16 (m, 13H, Ar), 7.07-6.99 (m, 4H, Ar), 6.93-6.91 (m, 2H, Ar), 6.79-6.77 (m, 4H, Ar), 6.37 (dd, 1H, H1', *J* = 5.5 & 8 Hz), 5.67 (m, 1H, H3'), 4.75 (s, 2H, $PhOCH_2CO$), 4.69 (s, 2H, $PhOCH_2CO$), 4.60 (t, 2H, $ArOCH_2$, *J* = 7.25 Hz), 4.25 (m, 1H, H4'), 3.76 (s, 6H, 2 x OCH_3), 3.64 (t, 2H, CH_2OH, *J* = 6.5 Hz), 3.50 (m, 1H, H5'), 3.38 (m, 1H, H5''), 3.00 (m, 1H, H2'), 2.63 (m, 1H, H2''), 1.90 (m, 2H, CH_2CH_2), 1.58 (m, 2H, CH_2CH_2), 1.53 (m, 2H, CH_2CH_2), 1.41 (m, 4H, CH_2CH_2). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 168.4, 161.3, 158.6, 157.6, 157.2, 152.4, 151.2, 144.4, 139.8, 135.5, 135.4, 130.03, 130.02, 129.8, 129.7, 128.1, 127.9, 127.0, 122.4, 122.0, 118.9, 114.9, 114.7, 113.2, 86.8, 84.3, 84.0, 77.3, 76.2, 67.99, 67.95, 65.3, 63.7, 62.9, 55.2, 38.0, 32.7, 29.0, 28.6, 25.9, 25.5. IR (thin film); ν_{max} (cm^{-1}) = 3408, 3058, 2931, 2855, 1765, 1700, 1607, 1510, 1496, 1175, 1085. HRMS (ESI-MS) *m/z* calculated for $C_{54}H_{58}N_5O_{11}^+$ 952.4127: found 952.4176 $[M+H]^+$.

3'-O-(tert-butyldimethylsilyl)-N^2-phenoxyacetyl-2'-deoxyguanosine **(3.5):**

To a solution of **3.1** (1.3 g, 1.8 mmol), imidazole (0.53 g, 7.8 mmol) and DMAP (cat.) in DCM (18 mL) was added TBS-Cl (0.59 g, 3.9 mmol) while stirring. After 4h, the reaction was diluted with DCM (100 mL total volume), washed with $NaHCO_3$ (aq, 3%) (2 x 50 mL) and brine (50 mL), dried over anhydrous Na_2SO_4, decanted and the solvent was evaporated to produce a colorless gum. To a solution of the resulting content in DCM : MeOH (30 mL : 10 mL) was added *p*-TsOH•H_2O (0.37 g, 1.94 mmol) while stirring. After 10 min, the solvent was removed in vacuo and the content was taken up in EtOAc (75 mL), washed with $NaHCO_3$ (aq, 3%) (2 x 75 mL) and brine (75 mL), dried over anhydrous $Na_2SO_{4)}$, decanted and the solvent was once more removed to produce a yellow gum. The product was purified *via* flash column chromatography using MeOH : DCM (4 %, 5 %) to afford 0.73 g (81 %) of a colorless powder. R_f (SiO_2 TLC):0.15 MeOH : DCM (6 %). $\lambda_{max\ (MeCN)}$ = 253 nm. ^{1}H NMR

(500MHz, $CDCl_3$, ppm): 11.78 (br. s, 1H, NH), 9.38 (br. s, 1H, NH), 7.88 (s, 1H, H8), 7.38-7.34 (m, 2H, Ar), 7.10-7.07 (m, 1H, Ar),7.04-7.01 (m, 2H, Ar), 6.24 (dd, 1H, H1', J = 8.5, 5.5 Hz), 4.94 (br. s, 1H, OH), 4.71 (s, 2H, $PhOCH_2CO$), 4.68 (m, 1H, H3'), 4.11 (m, 1H, H4'), 3.97 (m, 1H, H5'), 3.78 (m, 1H, H5''), 2.76 (m, 1H, H2'), 2.29 (m, 1H, H2''), 0.94 (s, 9H, $(CH_3)_3CSi$), 0.13 (s, 6H, $(CH_3)_2Si$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 169.6, 156.3, 155.0, 146.8, 146.1, 138.6, 129.9, 122.8, 114.8, 89.5, 86.4, 73.2, 66.7, 62.6, 41.7, 25.7, 18.0, -4.72, -4.79. IR (thin film); ν_{max} (cm^{-1}) = 3371, 3238, 3055, 2956, 2930, 2856, 1704, 1613, 1600, 1560, 1497, 1443, 1265, 1212, 1141, 1038, 947. HRMS (ESI-MS) *m/z* calculated for $C_{24}H_{34}N_5O_6Si^+$ 516.2273: found 516.2295 $[M+H]^+$.

3'-O-(tert-butyldimethylsilyl)-5'-O-N^2-bis(phenoxyacetyl)- 2'-deoxyguanosine **(3.6):**

To a solution of **3.5** (0.80 g, 1.6 mmol) and TEA (0.37 mL, 2.7 mmol) in THF (15 mL) was added Pac-Cl (0.31 mL, 2.2 mmol) dropwise while stirring at 0°C. After 30 min, the solvent was removed in vacuo and the content taken up in DCM (50 mL), washed with $NaHCO_3$ (aq, 3%) (2 x 50mL), dried over anhydrous Na_2SO_4, decanted and the solvent was evaporated to produce a yellow gum. The product was purified *via* flash column chromatography using MeOH : DCM (1 % → 4 % by increments of 1 %) to afford 0.87 g (86 %) of a yellow foam. R_f (SiO_2 TLC):0.36 MeOH : DCM (2 %). $\lambda_{max\ (MeCN)}$ = 260 nm. ^1H NMR (500MHz, $CDCl_3$, ppm): 11.79 (br. s, 1H, NH), 9.17 (br. s, 1H, NH), 7.88 (s, 1H, H8), 7.37-7.33 (m, 2H, Ar), 7.26-7.22 (m, 2H, Ar), 7.10-7.07 (m, 1H, Ar), 7.00-6.97 (m, 3H, Ar), 6.87-6.84 (m, 2H, Ar), 6.26 (dd, 1H, H1', J = 6.75 Hz), 4.64 (m, 4H, 2 x $PhOCH_2CO$), 4.46 (m, 1H, H5'), 4.40-4.36 (m, 2H, H3' & H5''), 4.20 (m, 1H, H4'), 2.61 (m, 1H, H2'), 2.28 (m, 1H, H2''), 0.91 (s, 9H, $(CH_3)_3CSi$), 0.08 (s, 6H, $(CH_3)_2Si$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 169.6, 169.0, 157.4, 156.5, 155.2, 147.5, 146.1, 137.6, 129.9, 129.6, 122.9, 122.5, 122.0, 114.9, 114.3, 84.9, 84.4, 72.5, 67.0, 65.0, 64.3, 40.2, 25.7, 17.9, -4.77, -4.92. IR (thin film); ν_{max} (cm^{-1}) = 3414, 3061, 2953, 2930, 2857, 2362, 1689, 1608, 1556, 1509, 1463,

1431, 1329, 1252, 1178, 1109, 1035. HRMS (ESI-MS) *m/z* calculated for $C_{32}H_{40}N_5O_8Si^+$ 650.2641: found 650.2671 $[M+H]^+$.

1- {O^6-[3'-O-N^2-bis(phenoxyacetyl)-5'-O-(4,4'-dimethoxytrityl)- 2'-deoxyguanidinyl]}-4-{O^6-[3'-O-(tert-butyldimethylsilyl)-5'-O-N^2-bis(phenoxyacetyl)- 2'-deoxyguanidinyl]}-butane **(3.7a):**

To a solution of **3.4a** (0.47 g, 0.52 mmol), **3.6** (0.31 g, 0.47 mmol) and Ph_3P (0.25 g, 0.94 mmol) in dioxane (1.4 mL) was added DIAD (0.185 mL, 0.94 mmol) dropwise while stirring. After 3 days, the solvent was removed in vacuo and the content taken up in DCM (50 mL), washed with $NaHCO_3$ (aq, 3%) (2 x 75 mL), dried over anhydrous Na_2SO_4, decanted and the solvent removed to produce a yellow gum. The product was purified *via* flash column chromatography using EtOAc : hexanes (4 : 1, 9 : 1) to afford 0.51 g (70 %) of a colorless foam. R_f (SiO_2 TLC):0.47 MeOH : DCM (6 %). $\lambda_{max\ (MeCN)}$ = 269 nm. ^{1}H NMR (500MHz, $CDCl_3$, ppm): 8.86-8.85 (m, 2H, NH), 8.03 (s, 1H, H8), 7.98 (s, 1H, H8), 7.37-7.16 (m, 18H, Ar), 7.04-6.86 (m, 11H, Ar), 6.77-6.75 (m, 4H, Ar), 6.40-6.33 (m, 2H, H1'a & H1'b), 5.67 (m, 1H, H3'a), 4.77-4.63 (m, 13H, 4 x $PhOCH_2CO$, H3'b & $ArO\underline{CH_2}(CH_2)_2\underline{CH_2}OAr$), 4.56 (m, 1H, H5'b), 4.45 (m, 1H, H5''b), 4.25 (m, 1H, H4'a), 4.17 (m, 1H, H4'b), 3.75 (s, 6H, 2 x OCH_3), 3.51 (m, 1H, H5'a), 3.38 (m, 1H, H5''a), 3.00 (m, 1H, H2'a), 2.93 (m, 1H, H2'b), 2.63 (m, 1H, H2''a), 2.35 (m, 1H, H2''b), 2.12 (m, 4H, CH_2CH_2), 0.90 (s, 9H, $(CH_3)_3CSi$), 0.09 (s, 6H, $(CH_3)_2Si$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 168.7, 168.4, 161.2, 158.6, 157.7, 157.6, 152.5, 152.3, 151.0, 144.4, 139.9, 135.47, 135.45, 130.04, 130.02, 129.8, 129.7, 129.6, 128.1, 127.9, 127.0, 122.3, 122.0, 121.8, 115.0, 114.9, 114.7, 114.5, 113.2, 86.8, 85.1, 85.0, 84.3, 84.0, 76.2, 72.6, 68.0, 67.2, 65.3, 65.2, 64.7, 63.7, 55.2, 39.8, 38.0, 30.9, 25.7, 25.5, 18.0, -4.74, -4.88. IR (thin film); ν_{max} (cm^{-1}) = 3406, 3055, 2955, 2928, 2854, 1766, 1722, 1606, 1510, 1496, 1442, 1420, 1383, 1265, 1250, 1175, 1085, 1034. HRMS (ESI-MS) *m/z* calculated for $C_{83}H_{89}N_{10}O_{18}Si^+$ 1541.6120: found 1541.6211 $[M+H]^+$.

1- {O^6-[3'-O-N^2-bis(phenoxyacetyl)-5'-O-(4,4'-dimethoxytrityl)- 2'-deoxyguanidinyl]}-7-{O^6-[3'-O-(tert-butyldimethylsilyl)-5'-O-N^2-bis(phenoxyacetyl)- 2'-deoxyguanidinyl]}-heptane **(3.7b)** :

To a solution of **3.4b** (0.54 g, 0.57 mmol), **3.6** (0.34 g, 0.52 mmol) and Ph_3P (0.27 g, 1.0 mmol) in dioxane (1.5 mL) was added DIAD (0.205 mL, 1.0 mmol) dropwise while stirring. After 3 days, the solvent was removed in vacuo and the content taken up in DCM (50 mL), washed with $NaHCO_3$ (aq, 3%) (2 x 75 mL), dried over anhydrous Na_2SO_4, decanted and the solvent was evaporated to produce a yellow gum. The product was purified *via* flash column chromatography using MeOH : DCM (0.75 % → 3% by increments of about 0.5 %) to afford 0.62 g (76 %) of a colorless foam. R_f (SiO_2 TLC):0.54 MeOH : DCM (5 %). $\lambda_{max\ (MeCN)}$ = 269 nm. ^{1}H NMR (500MHz, $CDCl_3$, ppm): 8.79-8.71 (m, 2H, NH), 7.89 (s, 1H, H8), 7.88 (s, 1H, H8), 7.41-7.17 (m, 18H, Ar), 7.09-6.88 (m, 11H, Ar), 6.77-6.74 (m, 4H, Ar), 6.40-6.32 (m, 2H, H1'a & H1'b), 5.68 (m, 1H, H3'a), 4.74-4.63 (m, 8H, 4 x $PhOCH_2CO$ & H3'b), 4.60-4.51 (m, 5H, H5'b & $ArO\underline{CH_2}(CH_2)_2\underline{CH_2}OAr$), 4.47 (m, 1H, H5''b), 4.26 (m, 1H, H4'a), 4.17 (m, 1H, H4'b), 3.75 (s, 6H, 2 x OCH_3), 3.51 (m, 1H, H5'a), 3.37 (m, 1H, H5''a), 3.02 (m, 1H, H2'a), 2.95 (m, 1H, H2'b), 2.63 (m, 1H, H2''a), 2.32 (m, 1H, H2''b), 1.90 (m, 4H, CH_2CH_2), 1.57-1.43 (m, 6H, $CH_2CH_2CH_2$), 0.91 (s, 9H, $(CH_3)_3CSi$), 0.10 (s, 3H, $(CH_3)Si$), 0.09 (s, 3H, $(CH_3)Si$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 168.6, 168.3, 161.3, 161.2, 158.4, 157.5, 157.1, 152.3, 152.1, 151.1, 150.9, 144.3, 140.7, 139.7, 135.35, 135.33, 129.90, 129.88, 129.69, 129.68, 129.6, 129.54, 129.48, 129.3, 129.2, 127.9, 127.85, 127.76, 126.8, 122.2, 121.8, 121.7, 119.3, 118.8, 114.82, 114.80, 114.54, 114.47, 114.37, 114.25, 113.15, 113.06, 86.2, 85.0, 84.9, 84.2, 83.9, 76.1, 72.4, 67.88, 67.86, 67.8, 67.7, 65.1, 65.0, 64.6, 63.6, 55.1, 39.6, 37.8, 29.02, 28.96, 28.6, 25.7, 25.6, 25.5, 22.0, 21.9, 17.8, -4.86, -5.00. IR (thin film); ν_{max} (cm^{-1}) = 3411, 3061, 2929, 2855, 1765, 1723, 1602, 1501, 1496,

1440, 1249, 1175, 1085. HRMS (ESI-MS) *m/z* calculated for $C_{86}H_{95}N_{10}O_{18}Si^+$ 1583.6590: found 1583.6692 $[M+H]^+$.

1- {O^6-[3'-O-N^2-bis(phenoxyacetyl)-5'-O-(4,4'-dimethoxytrityl)- 2'-deoxyguanidinyl]}-4-{O^6-[5'-O-N^2-bis(phenoxyacetyl)- 2'-deoxyguanidinyl]}-butane **(3.8a)**:

To a solution of **3.7a** (0.58 g, 0.37 mmol) in THF (2 mL) was added TBAF (0.44 mL of a 1M solution in THF) dropwise over 30 seconds. After 30 min, the solvent was removed in vacuo and the crude compound purified *via* flash column chromatography using MeOH : DCM (2 %, 2.5 %) to afford 0.29 g (54 %) of a colorless foam. R_f (SiO_2 TLC):0.30 MeOH : DCM (6 %). $\lambda_{max\ (MeCN)}$ = 269 nm. ^{1}H NMR (500MHz, d_6-acetone, ppm): 9.47 (s, 1H, NH), 9.37 (s, 1H, NH), 8.26 (s, 1H, H8), 8.02 (s, 1H, H8), 7.40-7.39 (m, 2H, Ar), 7.33-7.13 (m, 16H, Ar), 7.02-6.89 (m, 11H, Ar), 6.80-6.69 (m, 4H, Ar), 6.49-6.44 (m, 2H, H1' & H1'b), 5.72 (m, 1H, H3'a), 5.05-5.04 (m, 4H, 2 x $PhOCH_2CO$), 4.82-4.80 (m, 3H, $PhOCH_2CO$ & H3'b), 4.75-4.67 (m, 6H, $PhOCH_2CO$ & $ArO\underline{CH_2}(CH_2)_2\underline{CH_2}OAr$), 4.51-4.47 (m, 2H, H5'b & H5''b), 4.35 (m, 1H, H4'a), 4.22 (m, 1H, H4'b), 3.74 (s, 6H, 2 x OCH_3), 3.65 (m, 1H, H5'a), 3.43 (m, 1H, H2'a), 3.34 (m, 1H, H5''a), 2.98 (m, 1H, H2'b), 2.84 (s, 1H, OH), 2.66 (m, 1H, H2'), 2.50 (m, 1H, H2''b), 2.10 (m, 4H, CH_2CH_2). ^{13}C NMR (125.7 MHz, d_6-acetone, ppm): 168.5, 168.3, 160.9, 158.64, 158.56, 158.2, 158.14, 158.05, 152.7, 152.6, 151.7, 151.5, 145.2, 141.4, 141.1, 135.9, 135.7, 130.1, 30.0, 129.9, 129.48, 129.45, 129.4, 128.0, 127.5, 126.6, 121.4, 121.3, 121.23, 121.20, 118.7, 118.5, 114.8, 114.7, 114.6, 114.6, 112.9, 112.8, 86.1, 84.9, 84.8, 84.6, 84.5, 76.0, 71.5, 68.0, 67.9, 66.9, 66.8, 64.9, 64.73, 64.70, 64.6, 54.59, 54.58, 39.3, 36.1, 25.3. IR (thin film); ν_{max} (cm^{-1}) = 3409, 3059, 2926, 2852, 2362, 1761, 1722, 1606, 1510, 1496, 1442, 1420, 1383, 1248, 1175, 1086, 1034. HRMS (ESI-MS) *m/z* calculated for $C_{77}H_{75}N_{10}O_{18}^+$ 1427.5255: found 1427.5338 $[M+H]^+$.

1- {O^6-[3'-O-N^2-bis(phenoxyacetyl)-5'-O-(4,4'-dimethoxytrityl)- 2'-deoxyguanidinyl]}-7-{O^6-[5'-O-N^2-bis(phenoxyacetyl)- 2'-deoxyguanidinyl]}-heptane **(3.8b)**:

To a solution of **3.7b** (0.51 g, 0.33 mmol) in THF (2 mL) was added TBAF (0.36 mL of a 1M solution in THF) dropwise over 30 seconds. After 30 min, the solvent was removed in vacuo and the crude compound purified *via* flash column chromatography using MeOH : DCM (2 %, 4 %) to afford 0.13 g (27 %) of a colorless foam. R_f (SiO_2 TLC):0.26 MeOH : DCM (5 %). $\lambda_{max\ (MeCN)}$ = 269 nm. ^{1}H NMR (500MHz, d_6-acetone, ppm): 9.44 (s, 1H, NH), 9.33 (s, 1H, NH), 8.24 (s, 1H, H8), 8.18 (s, 1H, H8), 7.34-7.14 (m, 18H, Ar), 7.01-6.87 (m, 11H, Ar), 6.77-6.68 (m, 4H, Ar), 6.48-6.43 (m, 2H, H1'a & H1'b), 5.71 (m, 1H, H3'a), 5.02-4.99 (m, 4H, 2 x $PhOCH_2CO$), 4.80-4.79 (m, 3H, $PhOCH_2CO$ & H3'b),4.71-4.68 (m, 3H, $PhOCH_2CO$ & H5'b) 4.61-4.53 (m, 4H, $ArO\underline{CH_2}(CH_2)_2\underline{CH_2}OAr$), 4.47 (m, 1H, H5''b), 4.33 (m, 1H, H4'a), 4.21 (m, 1H, H4'b), 3.73 (s, 6H, 2 x OCH_3), 3.64 (m, 1H, H5'a), 3.42 (m, 1H, H2'a), 3.33 (m, 1H, H5''a), 2.98 (m, 1H, H2'b), 2.83 (s, 1H, OH), 2.64 (m, 1H, H2''a), 2.48 (m, 1H, H2''b), 1.86 (m, 4H, CH_2CH_2), 1.48 (m, 6H, $CH_2CH_2CH_2$). ^{13}C NMR (125.7 MHz, d_6-acetone, ppm): 168.5, 168.3, 166.8, 161.03, 161.01, 158.64, 158.56, 158.2, 158.14, 158.06, 152.7, 152.6, 151.7, 151.6, 151.55, 151.47, 145.3, 141.3, 141.0, 135.9, 135.8, 130.1, 129.9, 129.5, 129.4, 128.0, 127.5, 126.6, 121.4, 121.3, 121.2, 118.8, 118.6, 114.8, 114.7, 114.6, 114.5, 112.9, 112.8, 86.1, 84.9, 84.81, 84.77, 84.6, 84.5, 76.0, 71.5, 71.4, 67.93, 67.90, 67.3, 67.2, 64.9, 64.74, 64.72, 64.6, 54.59, 54.58, 54.1, 39.21, 39. 16, 36.1, 30.5, 25.70, 25.67. IR (thin film); ν_{max} (cm^{-1}) = 3409, 3264, 3061, 2934, 2856, 2362, 2335, 2057, 1933, 1762, 1724, 1606, 1510, 1496, 1383, 1174, 1086. HRMS (ESI-MS) *m/z* calculated for $C_{80}H_{81}N_{10}O_{18}^+$ 1469.5725: found 1469.5822 $[M+H]^+$.

1- {O^6-[3'-O-N^2-bis(phenoxyacetyl)-5'-O-(4,4'-dimethoxytrityl)- 2'-deoxyguanidinyl]}-4-{O^6-[3'-O-(2-cyanoethoxy(diisopropylamino)-phosphino)-5'-O-N^2-bis(phenoxyacetyl)- 2'-deoxyguanidinyl]}-butane **(3.9a)** :

To a solution of compound **3.8a** (0.14 g, 0.098 mmol) and DIPEA (28 µL, 0.16 mmol) in THF (1.5 mL) was added Cl-POCEN*i*Pr_2 (28 µL, 0.13 mmol) dropwise while

stirring. After 30 min, the solvent was evaporated in vacuo and the content was diluted with EtOAc (50 mL), washed with $NaHCO_3$ (aq, 3%) (2 x 35 mL) and brine (35 mL), dried over anhydrous Na_2SO_4, decanted and the solvent was evaporated to produce a yellow gum. Purification was achieved *via* short flash column chromatography using EtOAc as eluent (with 0.1 % NEt_3) to afford 0.13 g (80 %) of a colorless foam. R_f (SiO_2 TLC):0.36, 0.28 EtOAc. $\lambda_{max\ (MeCN)}$ = 269 nm. ^{1}H NMR (500MHz, d_6-acetone, ppm): 9.31 (s, 1H, NH), 9.22 (s, 1H, NH), 8.11 (s, 1H, H8), 8.04 (s, 1H, H8), 7.24-7.22 (m, 2H, Ar), 7.16-6.97 (m, 16H, Ar), 6.86-6.73 (m, 11H, Ar), 6.64-6.53 (m, 4H, Ar), 6.35-6.27 (m, 2H, H1'a & H1'b), 5.56 (m, 1H, H3'a), 4.87-4.80 (m, 5H, H3'b & 2 x $PhOCH_2CO$), 4.65 (s, 2H, $PhOCH_2CO$), 4.57-4.50 (m, 6H, $PhOCH_2CO$ & $ArO\underline{CH_2}(CH_2)_2\underline{CH_2}OAr$), 4.45-4.33 (m, 2H, H5'b & H5''b), 4.26 (m, 1H, H4'b), 4.19 (m, 1H, H4'a), 3.82-3.66 (m, 2H, CH_2OP), 3.60-3.42 (m, 9H, 2 x CHN, H5'a & 2 x OCH_3), 3.25 (m, 1H, H2'a), 3.18 (m, 1H, H5''a), 2.98 (m, 1H, H2'b) 2.66-2.42 (m, 4H, CH_2CN, H2''a & H2''b), 1.95-1.89 (m, 4H, CH_2CH_2), 1.09-1.03 (m, 12H, 4 x CH_3). ^{13}C NMR (125.7 MHz, d_6-acetone, ppm): 170.01, 168.49, 168.42, 168.24, 160.92, 160.88, 158.62, 158.54, 158.17, 158.12, 158.03, 152.62, 151.69, 151.52, 145.23, 141.39, 141.17, 135.85, 135.74, 130.11, 129.91, 129.48, 129.46, 129.45, 129.42, 128.03, 127.54, 126.57, 121.36, 121.29, 121.23, 118.72, 118.63, 118.23, 114.76, 114.74, 114.61, 114.49, 112.86, 112.80, 86.11, 84.92, 84.65, 84.58, 84.05, 75.99, 73.74, 67.92, 66.89, 66.84, 65.23, 64.91, 64.77, 64.75, 64.57, 64.44, 64.33, 59.66, 58.74, 58.58, 54.61, 54.60, 43.22, 43.17, 43.12, 43.07, 38.31, 36.09, 25.30, 24.04, 24.01, 23.98, 21.93, 19.97, 19.92, 19.89, 19.87, 19.84, 14.75, 13.64. ^{31}P NMR (202.3 MHz, d_6-acetone, ppm): 148.62, 148.57. IR (thin film); ν_{max} (cm^{-1}) =. HRMS (ESI-MS) *m/z* calculated for $C_{86}H_{92}N_{12}O_{19}P^+$ 1627.6334: found 1627.6328 $[M+H]^+$.

1- {O^6-[3'-O-N^2-bis(phenoxyacetyl)-5'-O-(4,4'-dimethoxytrityl)- 2'-deoxyguanidinyl]}-7-{O^6-[3'-O-(β-Cyanoethyl-N,N'-diisopropyl phosphite)-5'-O-N^2-bis(phenoxyacetyl)- 2'-deoxyguanidinyl]}-heptane **(3.9b)**:

To a solution of compound **3.8b** (0.20 g, 0.14 mmol) and DIPEA (38 µL, 0.22 mmol) in THF (1.5 mL) was added Cl-POCEN*i*Pr$_2$ (39 µL, 0.18 mmol) dropwise while stirring. After 30 min, the solvent was evaporated in vacuo and the content was diluted with EtOAc (50 mL), washed with $NaHCO_3$ (aq, 3%) (2 x 35 mL) and brine (35 mL), dried over anhydrous Na_2SO_4, decanted and the solvent evaporated to produce a yellow gum. Purification was achieved *via* short flash column chromatography using an EtOAc : hexanes (9 : 1) solvent system (with 0.1 % NEt_3) to afford 0.16 g (73 %) of a colorless foam. R_f (SiO_2 TLC):0.69, 0.64 EtOAc. $\lambda_{max\ (MeCN)}$ = 269 nm. ^{1}H NMR (500MHz, d_6-acetone, ppm)): 9.26 (s, 1H, NH), 9.17 (s, 1H, NH), 8.10 (s, 1H, H8), 8.04 (s, 1H, H8), 7.25-7.24 (m, 2H, Ar), 7.18-6.98 (m, 16H, Ar), 6.88-6.74 (m, 11H, Ar), 6.63-6.54 (m, 4H, Ar), 6.36-6.29 (m, 2H, H1'a & H1'b), 5.58 (m, 1H, H3'a), 4.87 (m, 5H, H3'b & 2 x $PhOCH_2CO$), 4.66 (s, 2H, $PhOCH_2CO$), 4.59-4.58 (m, 2H, $PhOCH_2CO$), 4.47-4.33 (m, 6H, 2 x H5'a & $ArO\underline{CH_2}(CH_2)_5\underline{CH_2}OAr$), 4.26 (m, 1H, H4'b), 4.19 (m, 1H, H4'a), 3.82-3.68 (m, 2H, CH_2OP), 3.62-3.49 (m, 9H, 2x CHN, H5'b & 2 x OCH_3), 3.29 (m, 1H, H2'a), 3.19 (m, 1H, H5''b), 3.01 (m, 1H, H2'b), 2.67-2.46 (m, 4H, CH_2CN, H2''a & H2''b), 1.76-1.68 (m, 4H, $\underline{CH_2}(CH_2)_3\underline{CH_2}$), 1.40-1.32 (m, 6H, $(CH_2)_3$), 1.10-1.03 (m, 12H, 4 x CH_3). ^{13}C NMR (125.7 MHz, d_6-acetone, ppm): δ 169.99, 168.46, 168.39, 168.22, 161.06, 161.01, 158.62, 158.54, 158.13, 158.05, 152.63, 151.53, 145.25, 141.33, 141.13, 135.87, 135.75, 130.11, 129.90, 129.46, 129.46, 129.41, 128.03, 127.52, 126.54, 121.34, 121.27, 118.79, 118.21, 114.76, 114.74, 114.60, 114.49, 112.84, 112.78, 86.08, 84.93, 84.68, 84.60, 84.01, 76.01, 73.76, 67.88, 67.24, 67.16, 64.90, 64.76, 64.74, 64.59, 64.42, 64.33, 59.64, 58.73, 58.58, 54.58, 54.57, 43.21, 43.16, 43.10, 43.06, 38.18, 36.05, 25.70, 25.68, 24.03, 23.98, 19.94, 19.90,

19.85, 13.61. ^{31}P NMR (202.3 MHz, d_6-acetone, ppm): 148.60, 148.55. IR (thin film); ν_{max} (cm^{-1}) =. HRMS (ESI-MS) *m/z* calculated for $C_{89}H_{98}N_{12}O_{19}P^+$ 1669.6803: found 1669.6863 $[M+H]^+$.

3.6.2 Supporting Figures

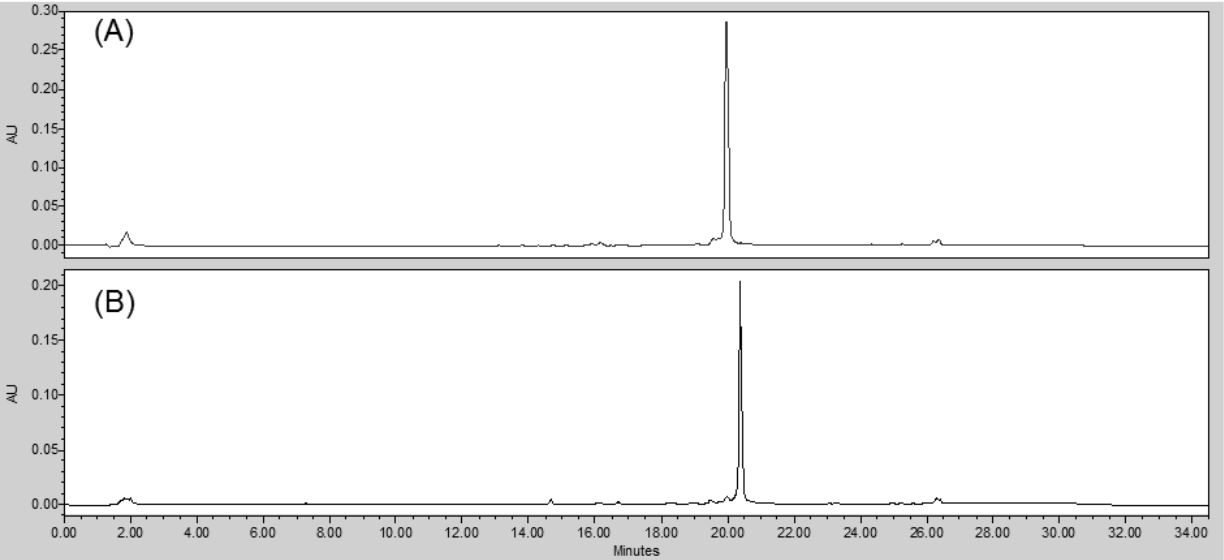


Figure S3.1: IEX-HPLC profile of **GG4** (A) and **GG7** (B). The column was eluted using a linear gradient of 0-52% buffer B over 24 min (buffer A: 100 mM Tris HCl, pH 7.5, 10% acetonitrile and buffer B: 100mMTris HCl, pH 7.5, 10% acetonitrile, 1 M NaCl) at 55^{o}C and flow rate of 1 mL/min. Monitoring was performed at 260 nm.

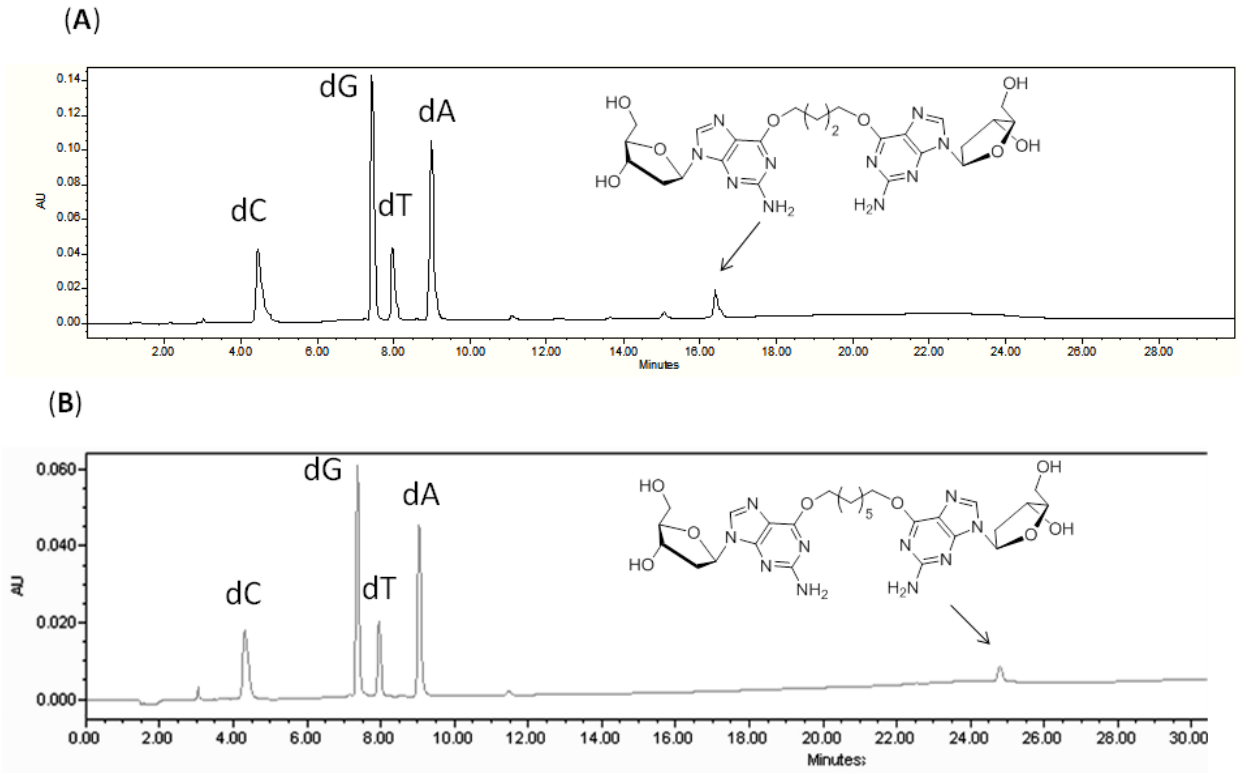


Figure S3.2: C-18 HPLC profile of nuclease digested **GG4** (A) and **GG7** (B). The column was eluted with a linear gradient of 0-60% buffer B over 30 min (buffer A: 50 mM sodium phosphate, pH 5.8 and buffer B: 50 mM sodium phosphate, pH 5.8, 50% acetonitrile) at a flow rate of 1.0 mL/min over 30 min, monitored at 260 nm.

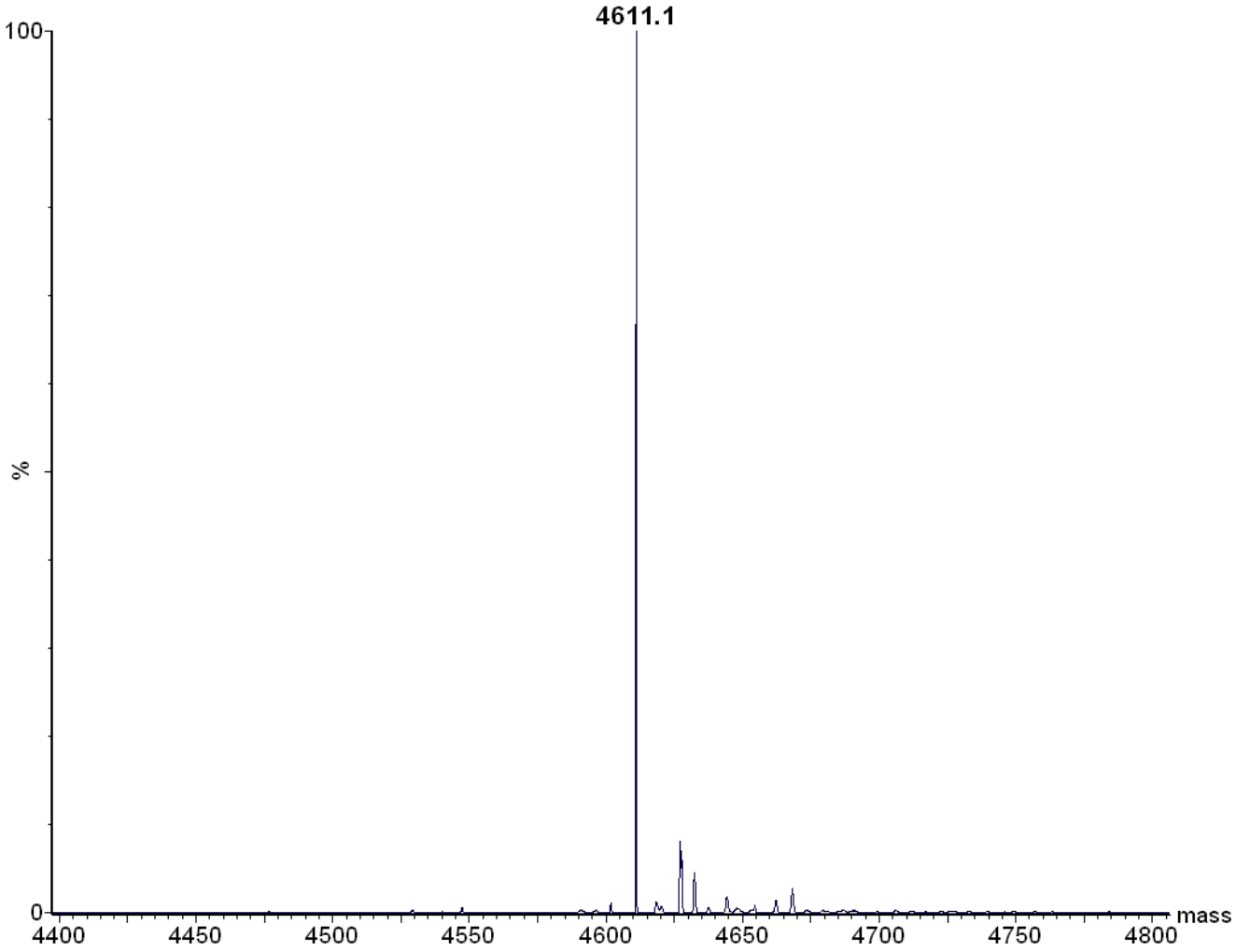


Figure S3.3: Deconvoluted ESI MS spectrum of oligonucleotide **GG4**

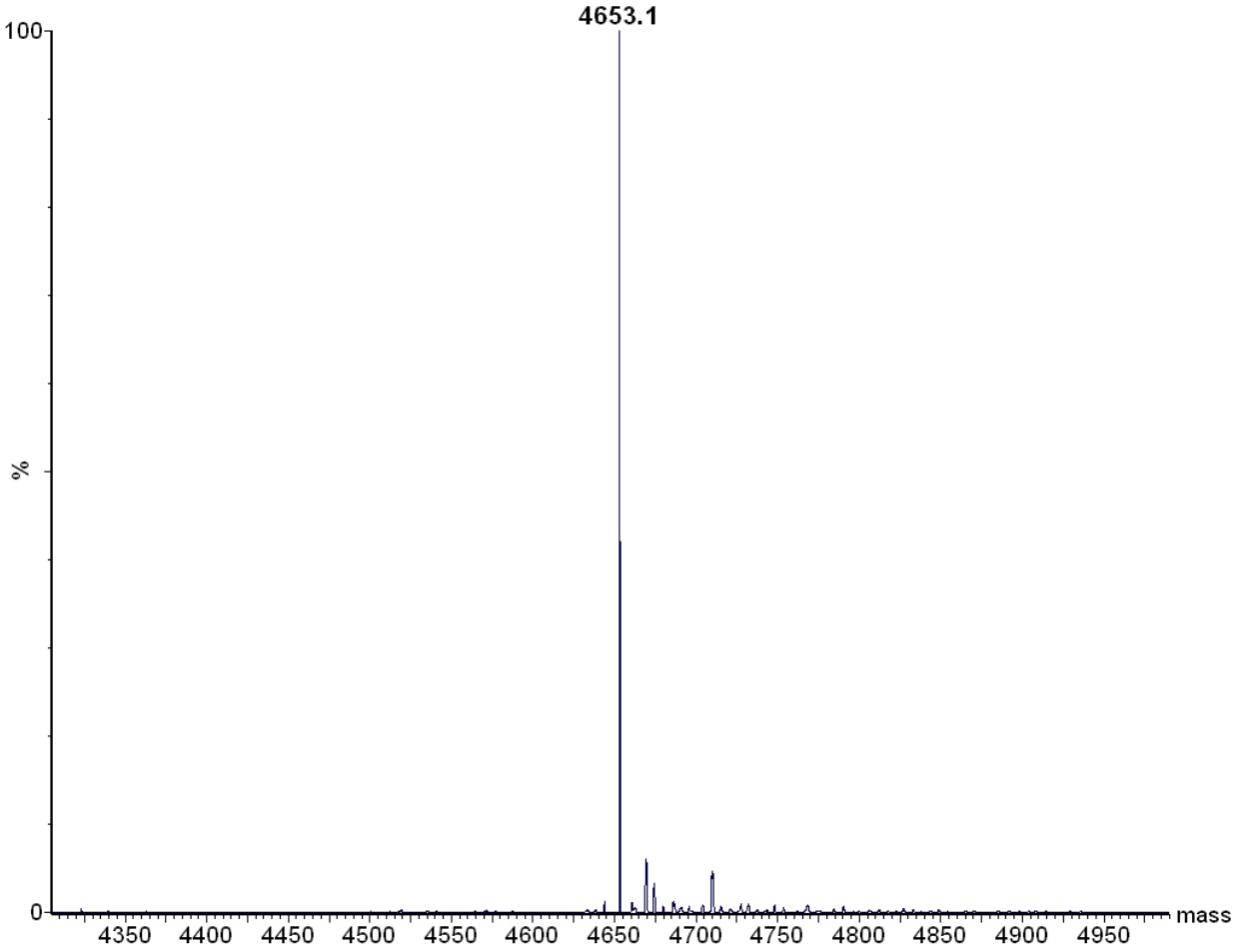


Figure S3.4: Deconvoluted ESI MS spectrum of oligonucleotide **GG7**

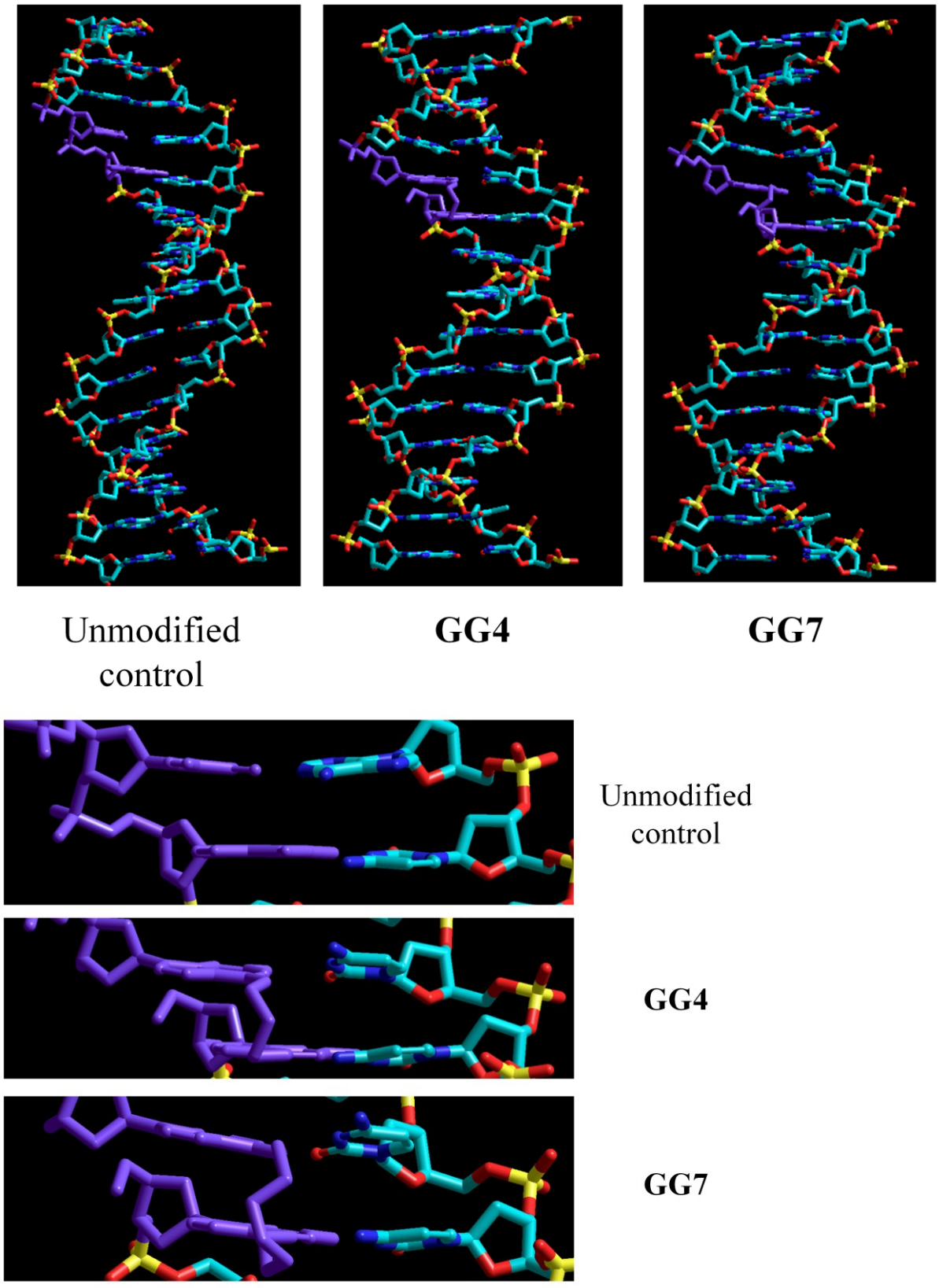


Figure S3.5: Molecular models of unmodified control duplex and duplexes containing **GG4** and **GG7** that were geometry optimized using the AMBER forcefield

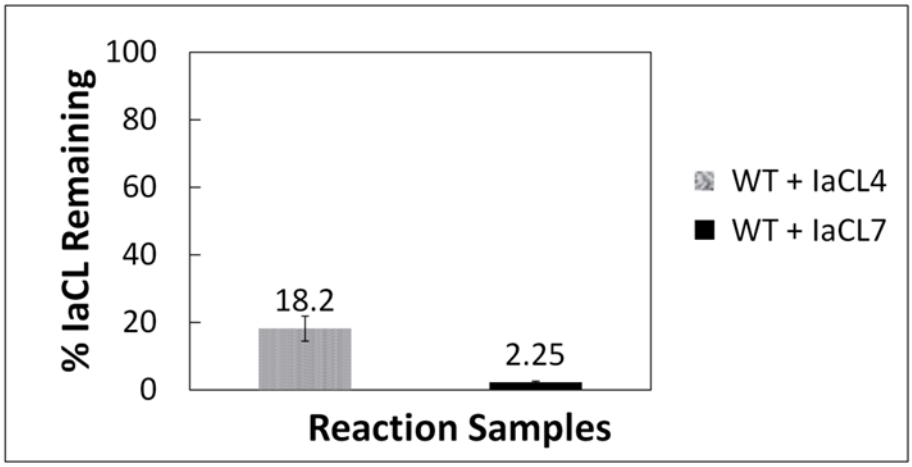


Figure S3.6: Graphical representation of the % IaCL of **GG4** and **GG7** remaining in the reaction tube after 8h obtained using ImageQuant™

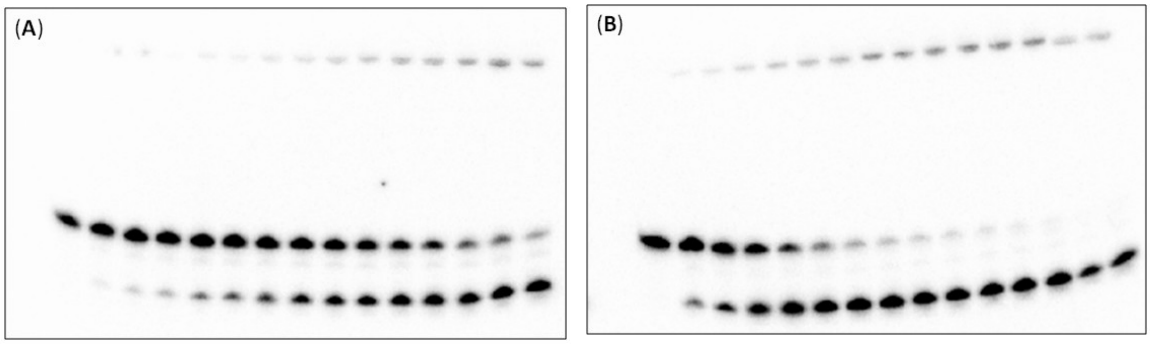


Figure S3.7: Time course repair gel of **GG4** (left) and **GG7** (right) by hAGT. (A) Denaturing gel of the repair of 2 pmol of **GG4** by 10 pmol hAGT as a function of time: lane 1, 2 pmol Control; lanes 2-15, 2 pmol **GG4** + 10 pmol hAGT incubated for 1, 2.5, 5, 10, 15, 25, 35, 45, 60, 90, 120, 180, 240 and 480 min, respectively (B) Denaturing gel of the repair of 2 pmol of **GG7** by 10 pmol hAGT as a function of time: lane 1, 2 pmol Control; lanes 2-15, 2 pmol **GG7** + 10 pmol hAGT incubated for 1, 2.5, 5, 10, 15, 25, 35, 45, 60, 90, 120, 180, 240 and 480 min, respectively.

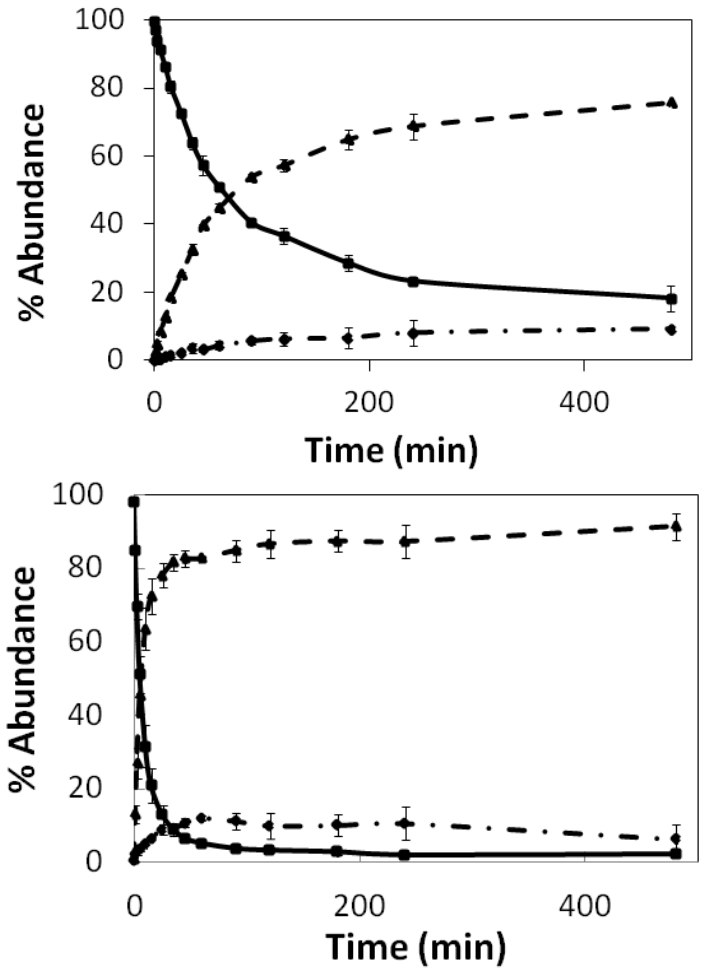
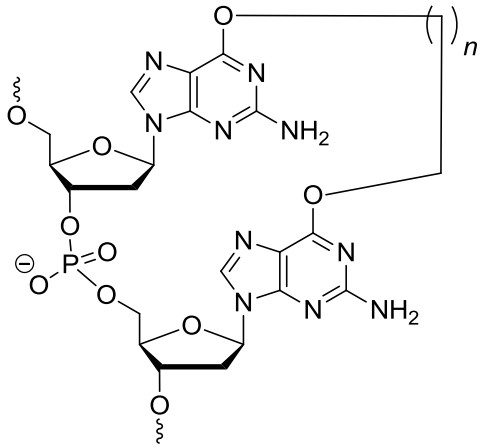


Figure S3.8: Time course repair assay of **GG4** (top) and **GG7** (bottom) by hAGT showing the full length assay (480 min). Graphical illustrations display abundances (%) of **MP** (—•—), **RP** (- - -) and substrate (—) over time (min) obtained using ImageQuant™.

CHAPTER IV

O^6-Alkylguanine DNA Alkyltransferase Repair Activity Towards Intrastrand Cross-Linked DNA is Influenced by the Internucleotide Linkage



Published as:

O'Flaherty DK and Wilds CJ. (2015) *Chemistry – An Asian Journal*. *DOI: 10.1002/asia.201501253*.

Department of Chemistry and Biochemistry, Concordia University, Montréal, Québec, Canada, H4B1R6.

Abstract

Oligonucleotides containing an alkylene intrastrand cross-link (IaCL) between the O^6-atoms of two consecutive 2'-deoxyguanosines (dG) were prepared by solid-phase synthesis. UV thermal denaturation studies of duplexes containing butylene and heptylene IaCL revealed a 20 ^oC reduction in stability compared to the unmodified control duplexes. Circular dichroism profiles of these IaCL DNA duplexes exhibited signatures consistent with B-form DNA, suggesting low global perturbations induced by these IaCL. Repair of DNA duplexes containing the IaCL modifications by O^6-alkylguanine DNA alkyltransferase (AGT) proteins from human and *E. coli* (OGT and Ada-C) was evaluated. Human AGT was capable of repairing both IaCL duplexes with slightly greater efficiency towards the heptylene analogue. Interestingly, repair efficiencies of hAGT towards these IaCL were lower compared to O^6-alkylene linked IaCL lacking the 5'-3'-phosphodiester linkage between the connected 2'-deoxyguanosine residues. These results demonstrate that the proficiency of hAGT activity towards IaCL at the O^6-atom of dG is influenced by the backbone phosphodiester linkage between the cross-linked residues.

4.1 Introduction

DNA insults may be incurred by environmental agents, endogenous metabolic processes and chemotherapeutic drugs.[31,266] Some chemotherapy regimens exploit the use of bi-functional electrophilic agents, which act on DNA to produce lesions and interfere with vital processes such as DNA replication.[54] Stalling these processes or modification of the information encoded in DNA can have severe effects on the cell. Bi-functional electrophilic agents can form adducts in DNA which link the atoms of two nucleotides on the same (intra) or opposing (inter) strands. Interstrand cross-links (ICL) in DNA are particularly cytotoxic given that the unwinding of the individual strands is prevented, which impedes cellular proliferation.[54,57,267] Although damage occurs on one strand for intrastrand cross-linked (IaCL) DNA, information content and structure can be affected. Cisplatin, a platinum-based drug used for the treatment of some forms of cancer, predominantly introduces lesions that are IaCL in nature.[114,268,281–283] Busulfan,[69] and mitomycin C,[70,284] are examples of bis-electrophilic agents which can introduce IaCL lesions in DNA, and can lead to the blockage of DNA replication and activation of apoptosis.[318,319]

O^6-Alkyl-2'-deoxyguanosine adducts can be formed by exposure to carcinogenic agents such as nitrosamines and chemotherapeutic alkylating agents such as temozolomide.[139] This class of lesions can induce transitional mutations *via* proficient base pairing with thymidine during DNA replication[19] and the activation of mismatch repair can result in futile excision-resynthesis ultimately leading to apoptosis.[53,320] Our group has shown that human O^6-alkylguanine DNA alkyltransferase (hAGT) is capable of repairing certain ICL.[64,65,150,151,321] The primary role of AGTs is to remove alkyl lesions found at the O^6-position of dG and, to a lesser extent, at the O^4-position of dT.[132] In the active site of the protein, the alkyl lesion becomes transferred to an

activated cysteine, rendering the AGT protein irreversibly inactivated and ultimately degraded *in vivo* by the ubiquitin pathway.[135,322]

The human AGT (hAGT) variant is capable of repairing butylene and heptylene ICL linking the O^6-positions of dG residues in a directly opposed fashion, as well as in a 5'-GNC motif.[64,65] Recently, we demonstrated that hAGT could also act on a DNA duplex containing an IaCL on one of the strands consisting of an alkylene functionality attaching the O^6-atoms of two dG residues that lack a phosphodiester linkage in the backbone between them (**Figure 4.1A**).[323] hAGT was proficient towards removing the lesion, with almost complete consumption observed for a heptylene-linked substrate within 60 min. These results prompted us to explore AGT activity towards the IaCL containing a phosphodiester linkage at this site (**Figure 4.1B**).

In the current investigation, a methodology to introduce this O^6-2'-deoxyguanosine-alkylene-O^6-2'-deoxyguanosine IaCL (O^6-dG-alkylene-O^6-dG, **Figure 4.1B**) was developed. The influence of this IaCL modification in a DNA duplex was assessed by UV thermal denaturation and circular dichroism. Repair of this IaCL DNA by a variety of AGTs (*human* and *E. coli*) was evaluated. Given the efficient action of hAGT on the IaCL DNA probes lacking the phosphodiester linkage between the 3'- and 5'-*O* groups of the O^6-alkylene linked dG residues, these new IaCL probes may aid in the understanding of backbone flexibility versus rigidity imposed by the phosphodiester linkage. Butylene and heptylene IaCL linkages were chosen on the basis that busulfan can react with DNA to produce butylene linked IaCL at the N7-positions of 5'-d(GA) and 5'-d(GG) sequences.[69] To be noted is that specific IaCL between the O^6-positions of 5'-d(GG) sequences have not been identified. The heptylene linkage was also prepared as the sulfamate hepsulfam has been demonstrated to introduce an ICL at the N7-

positions of 2'-deoxyguanosine in a 5'-d(GNC) motif.[324] The processing of this IaCL by hAGT will contribute to our understanding of the substrate range that can undergo repair by this protein.

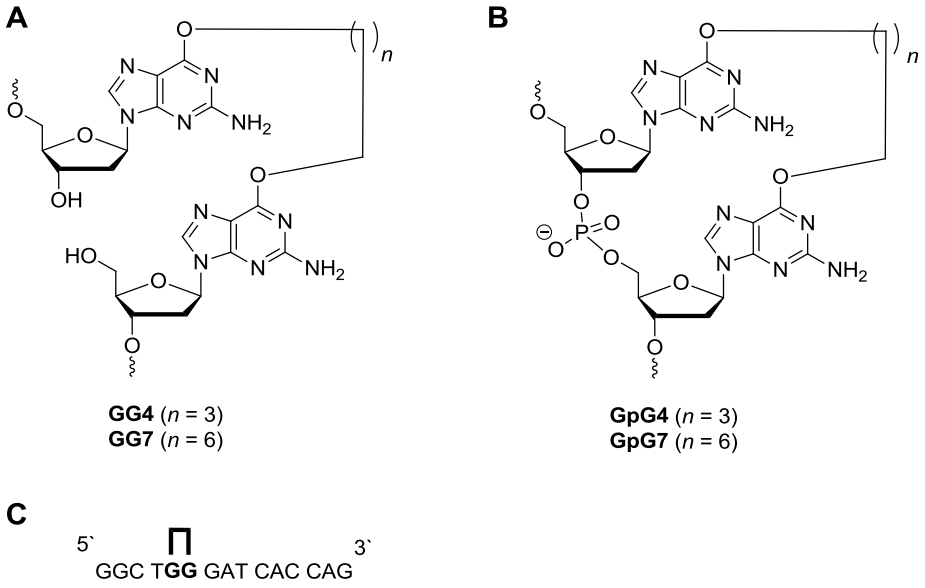


Figure 4.1: Structures of O^6-2′-deoxyguanosine-alkylene-O^6-2′-deoxyguanosine intrastrand cross-link (**A**) lacking or (**B**) containing the phosphodiester linkage and (**C**) the oligonucleotide sequence indicating the position of the IaCLs (shown by **GG**).

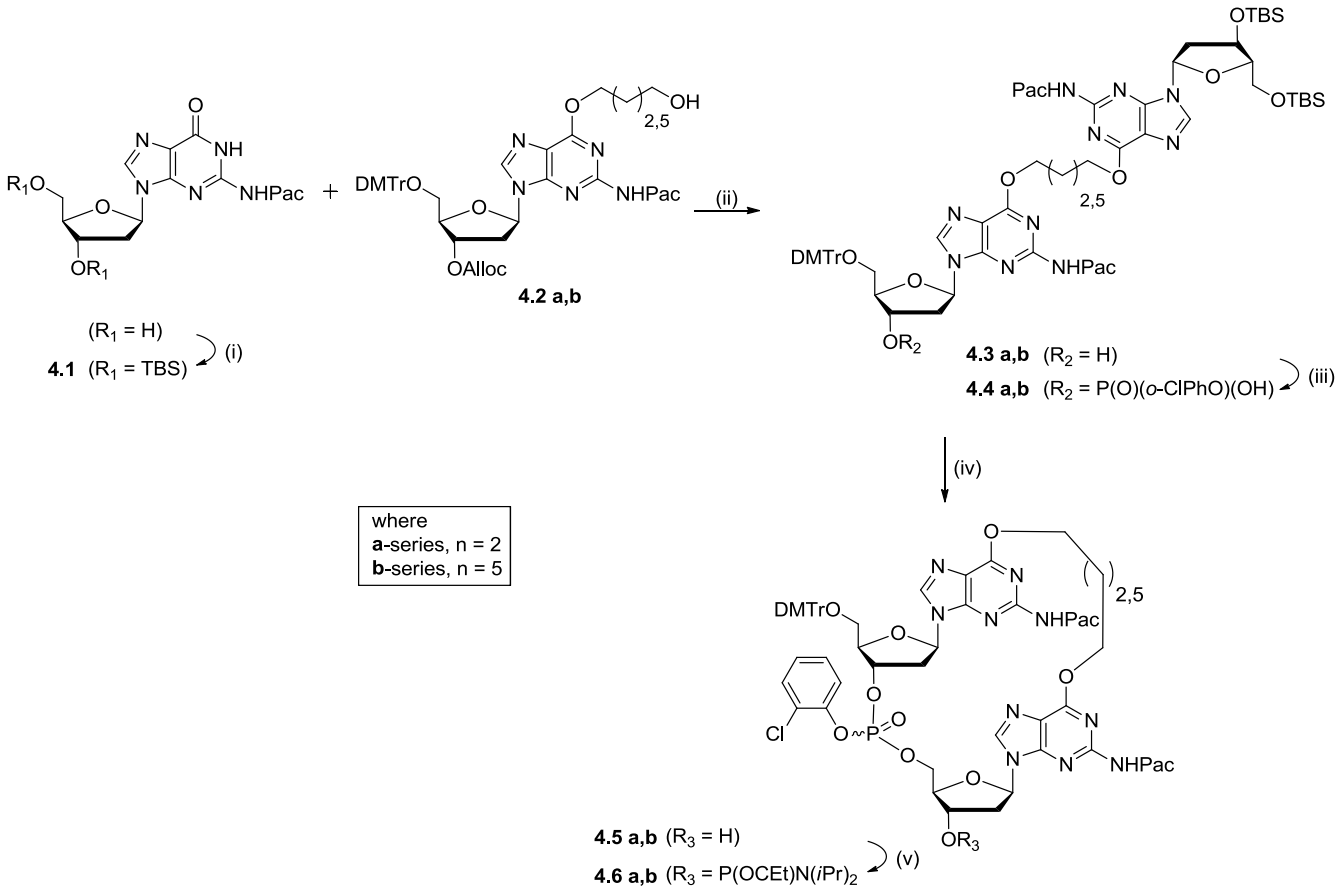
4.2 Results and Discussion

4.2.1 Synthesis and characterization of GpG IaCL DNA

The structure of the O^6-dG-alkylene-O^6-dG cross-links containing a phosphodiester linkage between the 5'- and 3'-*O* atoms (**GpG4** and **GpG7**) and their position in an oligonucleotide is shown in **Figure 4.1**. The synthesis approach for cross-linked amidites **4.6a** and **4.6b** is shown in **Scheme 4.1** and began with commercially available N^2-phenoxyacetyl-2'-deoxyguanosine. This material was azeotropically dried with anhydrous pyridine, followed by bis-silylation at the 3' and 5' hydroxyl functionalities. Compounds **4.2a** and **4.2b** were prepared according to procedures described previously by our group.[65] Dimers **4.3a** and **4.3b** were

produced using the Mitsunobu reaction of mono-adducts **4.2a** and **4.2b**, respectively, with compound **4.1** followed by the $Pd^{(0)}$ catalyzed removal of the allyloxycarbonyl group. Fully-protected versions of **4.3a** and **4.3b** were particularly difficult to isolate in high purity due to co-elution with mixtures of triphenylphosphine / triphenylphosphineoxide. The subsequent removal of the allyloxycarbonyl group, deprotecting the 3'-hydroxyl functionality, facilitated the chromatographic purification of the desired dimers (**4.3a** and **4.3b**). Phosphorylation of **4.3a** and **4.3b** was accomplished by adapting reported approaches which produced dimers **4.4a** and **4.4b**, respectively.[297,325,326] Solubility proved to be an issue during the washing (work-up) stages of the latter intermediates, particularly for **4.4a**. The introduction of NEt_3 in the work-up solvent alleviated this issue as the corresponding triethylammonium phosphate salts are generally more soluble in organic solvents. The ring-closing reaction to access **4.5a** and **4.5b** was accomplished according to a procedure described previously.[325] The use of MSNT (1-(2-Mesitylenesulfonyl)-3-nitro-1*H*-1,2,4-triazole) as the condensing reagent favors attack by the 5'-OH, as opposed to the 3'-OH, of the desilylated residue.[297,325,326] We hypothesized that these transformations would occur more quickly for our system, which we tested on starting material **4.4a**. First, the desilylation reaction was reduced to 1 h (as opposed to 6 h) and the cyclization reaction time was reduced to 4 h compared to 16 h with no significant yield differences observed (51 % and 52 % for shorter and longer reaction times, respectively). For the heptylene analogue **4.4b**, a cyclization reaction time of 16 h was used. Phosphoramidites **4.6a** and **4.6b** were synthesized according to previously described procedures.[323] In our hands, phosphoramidite products of greater purity are generally attained using column chromatography, relative to hexanes precipitation. However, we initially resorted to hexanes precipitation for isolating phosphoramidites **4.6a** and **4.6b** given the low mobility observed for these dimers on TLC. ^{31}P

NMR analysis revealed sharp signals in the region of 148-150 ppm characteristic of phosphoramidites. Compounds **4.6a** and **4.6b** were further characterized by HRMS and the masses observed were in agreement with the expected values.



Scheme 4.1: Reagents and conditions: (i) TBS-Cl (2.5 eq), Im (5.0 eq), DMAP (cat), DCM, 21°C, 16 h. (ii) 1. **4.1** (1.0 eq), **4.2a** or **4.2b** (1.1 eq), Ph_3P (1.2 eq), DIAD (1.2 eq), dioxane, 21 °C, 16 h. 2. $Pd(Ph_3P)_4$ (0.2 eq), Ph_3P (0.4 eq), *n*-$BuNH_2$ (7 eq), HCOOH (7 eq), THF, 21 °C, 30 min. (iii) 2-chlorophenyl dichlorophosphate (5 eq), 1,2,4-triazole (10 eq), NEt_3 (10 eq), THF, Py, 21 °C, 2 h. (iv) 1. TBAF (10 eq), THF, 21°C, 6h. 2. MSNT (5 eq), Py, 21 °C, 16 h. (v) Cl-P(OCEt)N(*i*Pr)$_2$ (1.2 eq), DIPEA (1.5 eq), THF, 21 °C, 30 min.

The IaCL DNA synthesis employed either cyclized dimer phosphoramidite **4.6a** or **4.6b**, which differed only in their respective alkylene linker length. The presence of a single DMTr group and single phosphoramidite moiety on the dimer scaffolds rendered a straightforward construction of the IaCL DNA. Assembly of oligonucleotides **GpG4** and **GpG7** by automated

solid-phase synthesis was carried out according to previous published procedures used to prepare ICL and IaCL DNA containing similar modifications.[65,323] Coupling wait times for phosphoramidites **4.6a** and **4.6b** were extended to 10 min, relative to 2 min for standard 3'-*O*-phosphoramidites, in order to ensure efficient coupling of the dimers to the nascent oligomer. **GpG4** and **GpG7** were deprotected and cleaved from the solid support using a protocol described by Glen Research (aqueous NH_4OH (28 %) for 17 h at room temperature with gentle rocking and an additional 4 h at 55^{o}C) with no detectable degradation of the cross-link observed by SAX-HPLC (**Figure S4.1**). Failure sequences due to incomplete couplings of **4.6a** or **4.6b** were well-resolved from desired products by SAX-HPLC. This was also observed for shorter sequences that were prepared (12-mer, data not shown). MS analysis of the **GpG4** and **GpG7** revealed deconvoluted masses in agreement with the expected masses (values shown in **Table 4.1** and spectra illustrated in **Figures S4.3** and **S4.4**). Further characterization by enzymatic digest followed by RP-HPLC showed the appearance of a new peak with retentions of 9.2 min and 14.7 min for the butylene and heptylene 2'-deoxyguanosine adducts, respectively (**Figure S4.2**). The retention of these dimers were significantly lower compared with those observed for the enzymatic digestion of **GG4** and **GG7** (16.4 min and 24.8 min),[323] respectively, suggesting incomplete digestion near the cross-linked site, as previously observed for other DNA modifications.[327,328]

Table 4.1. Retention time and mass spectrometry data for **GpG4** and **GpG7**.

Oligomer	Retention time (minutes)	Mass	
		Expected	Observed
GpG4	21.1	4672.1	4673.0
GpG7	21.6	4714.2	4715.1

4.2.2 UV thermal denaturation and circular dichroism studies of IaCL DNA

The influence of the **GpG4** and **GpG7** IaCL on duplex stability with the complementary DNA sequence, relative to the native duplex and the more "flexible" IaCL **GG4** and **GG7** (reported previously), was assessed by UV thermal denaturation experiments. The thermal denaturation profiles of the IaCL containing DNA duplexes were monophasic with T_m values of 45, 48 and 68 ^{o}C for **GpG4**, **GpG7** and the unmodified duplex, respectively (**Figure 4.2**). Presence of a single O^6-methyl-dG insert in a DNA duplex resulted in a T_m reduction of approximately 18 ^{o}C and two inserts dropped the value by 40 ^{o}C, compared to unmodified sequences.[313,315] The influence of two O^6-alkylated dG residues for the butylene- and heptylene linked IaCL (**GpG4** and **GpG7**) resulted in an overall reduction in T_m by 20 - 23 ^{o}C relative to the control, which was lower than expected. The reduction in stability is most likely attributed to the disruption of hydrogen bonding between the alkylated 2'-deoxyguanosines, containing the flexible alkylene linkers, with their paired 2'-deoxycytidines that may result in local and global structural perturbations. Interestingly, these values were comparable to the IaCL analogs which lack the phosphodiester linkage (46 ^{o}C and 48 ^{o}C for **GG4** and **GG7**, respectively). We had originally hypothesized that the added flexibility and lack of the anionic phosphodiester linkage at the cross-linked site for the **GG4**- and **GG7**-containing duplexes may have reduced repulsion effects relative to **GpG4** and **GpG7**, resulting in higher T_m values, which was not the case. This would suggest that the electrostatic repulsion involved in duplexes containing **GpG4** and **GpG7** is compensated for by other contributing factors. Towards this end, we performed Van't Hoff experiments to determine the thermodynamic parameters in order to evaluate if the decrease in T_m with respect to the unmodified control could be attributed to an enthalpic or entropic cost (see

Figure S4.7 for T_m profiles and **Table S4.1** for T_m values). The data suggest that the reduction in stability due to the presence of the flexible IaCL and phosphodiester linkage was generally attributed to an enthalpic cost. There seemed to be an entropic cost for **GG7**- and **GpG4**-containing duplexes, whereas **GG4**- and **GpG7**-containing duplexes revealed an entropic gain relative to the unmodified control.

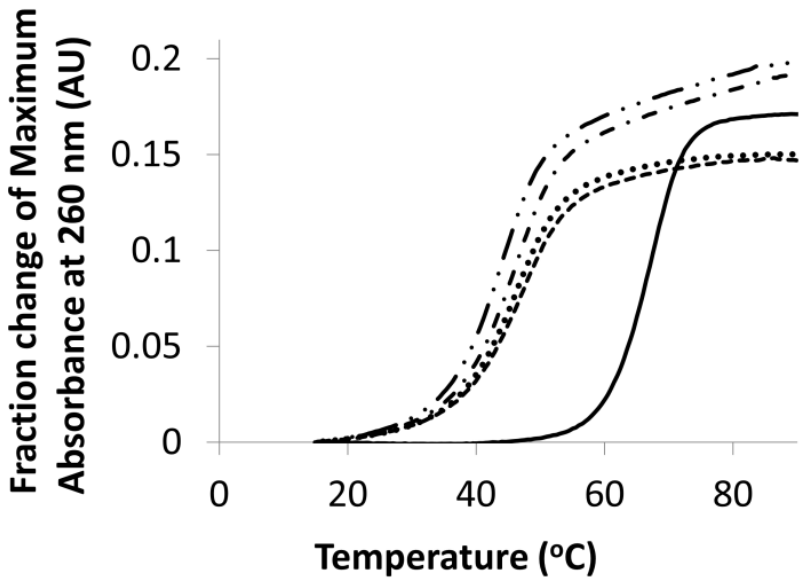


Figure 4.2 Fraction change of maximum absorbance at 260 nm (A_{260}) versus temperature [°C] profiles of duplexes containing **GG4** (•••),[323] **GG7** (- - -),[323] **GpG4** (—••), **GpG7** (–•–) and unmodified DNA (—).

Circular dichroism (CD) spectroscopy was performed to determine the global structural influence of the **GpG4** and **GpG7** IaCL in the DNA duplex. CD profiles of these modified duplexes displayed signatures consistent with the B-DNA form with maxima near 280 nm, cross-overs near 250-260 nm and minima around 240 nm (shown in **Figure 4.3**). However, the profile for duplexes containing the **GpG4** IaCL displayed a slight blue shift of the characteristic profile relative to the unmodified control. Nonetheless, these results indicate minor global structural distortions were induced by the presence of the alkylene IaCL in DNA duplexes. Similar

findings were observed for the "flexible" IaCL DNA (**GG4** and **GG7**) and duplexes containing monoadducts at the O^6-atom of dG (data shown only for **GG4** and **GG7** in **Figure 4.3**). Duplexes containing the **GpG4** and **GpG7** IaCL (and unmodified control) were geometry optimized using the AMBER force field (**Figure S4.8**). In both cases, the alkylene linker protruded into the major groove, with a slight widening observed in addition to buckling of the 5'-end O^6-alkylated dG nucleobase. The buckling effect was more pronounced for **GpG4** compared to **GpG7**, relative to the "flexible" **GG4**, **GG7** and control. The molecular models suggest that the presence of the cross-link does not greatly distort the global structure of the duplex consistent with the small variation in the CD profiles observed for **GpG4** and **GpG7**. High resolution structures of duplexes containing **GpG4** or **GpG7** are currently being investigated *via* by a combination of molecular dynamics and high-field NMR experiments.

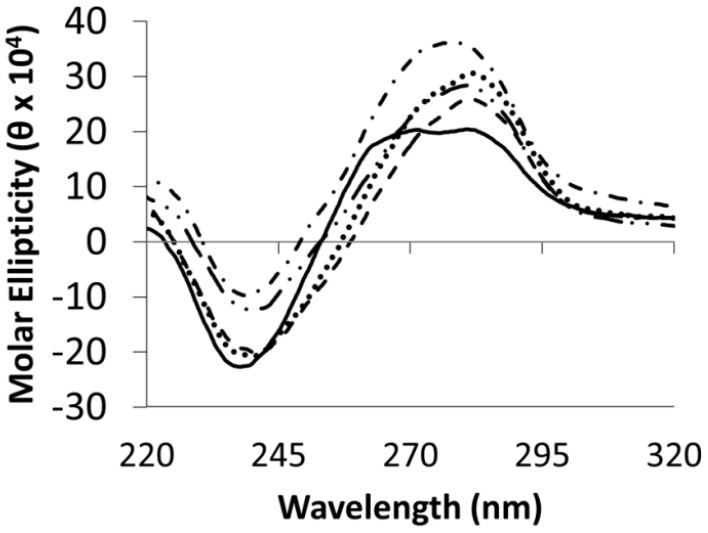


Figure 4.3. Circular dichroism spectra of IaCL duplexes, **GG4** (- - -),[323] **GG7** (•••),[323] **GpG4** (–•–), **GpG7** (—••) and unmodified DNA (—).

4.2.3 AGT-mediated repair of IaCL DNA

Four AGT proteins (hAGT, OGT, an S134P OGT variant, and Ada-C) were interrogated against the single strands **GG4** and **GG7**, as well as duplexes containing **GpG4** and **GpG7**. The repair of duplexes containing **GG4** and **GG7** has been described previously.[323] The modified strand was radiolabeled using γ-[^{32}P]ATP and either used as is or annealed with a 10% molar excess of the complement followed by incubation for 16 h, at 37 ^{o}C with the AGT protein of interest (2 pmol DNA and either 10 pmol or 60 pmol protein). Repair reactions were quenched and boiled to prevent complexation before loading onto the denaturing gel. Analysis of the AGT activity upon single stranded **GG4** and **GG7** revealed that only the human variant was capable of removing the lesion at 5-fold protein equivalence (**Figure S4.9-S4.11**). Interestingly, repair efficiencies of the **GG4** and **GG7** single strands by hAGT were in agreement with repair of duplexes containing the IaCL DNA, respectively.[323] No repair was observed, however, for other AGTs tested at 5-fold protein equivalence, whereas minimal repair (< 15 %) was observed by OGT, Ada-C and OGT S134P at 30-fold protein equivalence (see **Figure S4.9-S4.11**). Repair studies conducted on double stranded **GpG4** and **GpG7** revealed that only hAGT reacted with these IaCL, with slightly higher efficiency towards the **GpG7** (lane 4 and 10 in **Figure 4.4**). Surprisingly, the levels of repair were much lower for the IaCL DNA containing the phosphodiester linkage compared to both the single strands and duplexes containing the flexible **GG4** and **GG7** IaCL.[323] The **GG4** and **GG7** duplex substrates were virtually entirely consumed (4 and 5 % IaCL substrate remaining, respectively) compared to 74 % and 64 % substrate remaining for **GpG4** and **GpG7** duplexes, respectively, at thirty fold excess protein. OGT and Ada-C were incapable of repairing the IaCL DNA studied (10 pmol protein and 2 pmol DNA), which was not surprising given their inefficient repair of lesions larger than a methyl group at the

O^6-atom of dG. Although OGT is capable of repairing larger O^4-alkylated thymidine mono-adducts (but not ICL DNA),[150] OGT demonstrated no repair activity against **GpG4** and **GpG7**, even at higher AGT concentration ratios (data not shown). The OGT S134P variant[321] demonstrated no activity towards these IaCL DNA duplexes as well.

Complete repair of these IaCL DNA (**GpG4** and **GpG7**) by hAGT would begin with a reaction to generate a median product (**MP** in **Figure 4.4**) consisting of a hAGT-DNA cross-linked species. Repair of this hAGT-DNA **MP** by a second AGT protein would generate the completely repaired product (**RP** in **Figure 4.4**). Repair assays of various AGTs with **GpG4** and **GpG7** were performed and the products analyzed by denaturing gel electrophoresis (**Figure 4.4**). Lanes 2 and 8 contain the **GpG4** and **GpG7** ssDNA, respectively. Lane 1 contains the unmodified control DNA strand of identical sequence as **GpG4** and **GpG7**. The control DNA strand migrated slightly faster compared to the IaCL DNA, which indicated that resolution of the IaCL DNA and repaired DNA by electrophoresis was feasible. Lanes 3 and 9 display the product of the repair reaction for **GpG4** and **GpG7** (2 pmol) with hAGT (10 pmol), respectively. Lanes 4 and 10 display similar reactions with higher hAGT concentrations (60 pmol), for **GpG4** and **GpG7** respectively. Two new bands are observed for the reaction of **GpG4** and **GpG7** with hAGT; one which migrates much slower and another that migrates slightly faster. The slowly migrating band can be attributed to the formation of hAGT covalently bound to the damaged DNA (**MP**), as observed previously with AGT repair of O^6-dG-alkylene-O^6-dG IaCL and ICL DNA.[64,150,151,323] The faster migrating band corresponded to the **RP**.

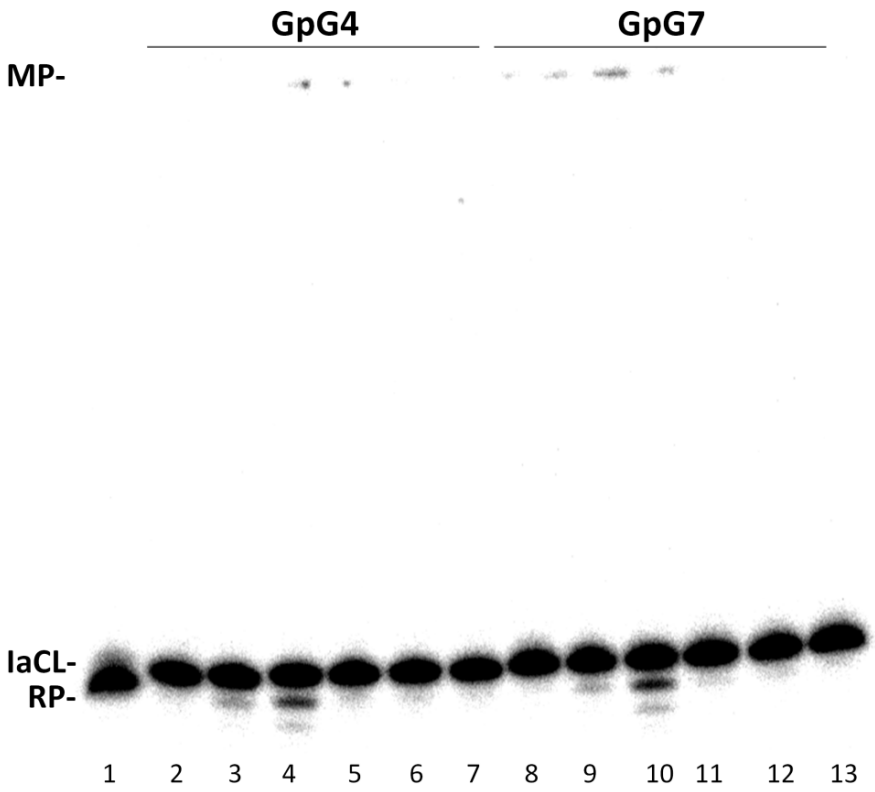


Figure 4.4. Repair of **GpG4** and **GpG7** by hAGT, OGT, Ada-C, and OGT S134P. Denaturing PAGE of repair reactions as described in the Experimental Section. Lane 1, 2 pmol control unmodified DNA + 10 pmol hAGT; lane 2, 2 pmol **GpG4**; lane 3, 2 pmol **GpG4** +10 pmol hAGT; lane 4, 2 pmol **GpG4** +60 pmol hAGT; lane 5, 2 pmol **GpG4** +10 pmol OGT; lane 6, 2 pmol **GpG4** +10 pmol Ada-C; lane 7, 2 pmol **GpG4** +10 pmol OGT S134P; lane 8, 2 pmol **GpG7**; lane 9, 2 pmol **GpG7** +10 pmol hAGT; lane 10, 2 pmol **GpG7** +60 pmol hAGT; lane 11, 2 pmol **GpG7** +10 pmol OGT; lane 12, 2 pmol **GpG7** +10 pmol Ada-C; lane 13, 2 pmol **GpG7** +10 pmol OGT S134P.

The amount of repair observed for the **GpG4** IaCL was similar to that of duplexes containing an O^6-dG-butylene-O^6-dG ICL (**XL4**, with approximately 65-70% of ICL DNA left unrepaired), whereas repair of the duplex containing a heptylene ICL (**XL7**) surpassed that of the **GpG7** (approximately 43% and 65% remaining of unrepaired **XL7** and **GpG7**, respectively). Previously, Abdu and coworkers described poor hAGT-mediated repair of DNA containing a conformationally locked *anti* analogue insert of O^6-methyl-2′-deoxyguanosine.[228] Our group has also observed that various AGTs were incapable of repairing DNA containing conformationally

locked analogues of O^4-alkyl-thymidine (**Figure S4.17**) where a methylene group at the O^4-position is *anti* relative to the N3-atom. Together, the latter two examples suggest that repair by AGTs may be dependent on lesion conformation, whereby attack by the active site Cys thiolate anion only proceeds if the α-carbon of the lesion is positioned correctly. The orientation of the alpha carbons of the O^6-alkylene linkers in **GpG4** and **GpG7** may not have the conformational freedom to adopt an optimal orientation for successful repair, which may explain the large discrepancies observed in the repair of the IaCL DNA containing the phosphodiester linkage. The alkylene linkers in these IaCL, lacking the phosphodiester linkage, may be more flexible and are thus capable of adopting multiple different conformations relative to the IaCL DNA investigated in this study. Fang and coworkers generated a molecular model of hAGT in complex with an O^6-dG-heptylene-O^6-dG ICL DNA, which suggested that the preferential repair of longer ICL DNA by hAGT is due to the shape of the hAGT active site once the DNA is bound (similar to a "tunnel").[64] The longer heptylene linker in **GpG7** may be accommodated more readily into the "tunnel" of the active site, allowing for positioning of the α-carbon near the C145 residue, compared to the more strained butylene linker in **GpG4**. This may account for the reason why the heptylene adducts are repaired more efficiently by hAGT. However, the length of the linker affected hAGT repair more drastically in the case of ICL DNA (**XL7** versus **XL4**) and IaCL lacking the phosphodiester linkage (**GG7** versus **GG4**), with much greater repair efficiency observed for the heptylene analogues. Much of our understanding of the AGT repair mechanism is derived from crystal structures of hAGT bound to damage-containing DNA (hAGT C145S with DNA containing an O^6-MedG insert and another of hAGT covalently cross-linked with DNA containing a $N1$,O^6-ethanoxanthosine insert). Both structures reveal conformational changes occurring in the DNA during the repair process with minimal changes in the protein

structure. Similar findings were observed for *E. coli* Ada.[329] The alkylated nucleotide is flipped into the protein active site allowing for attack of the activated thiolate anion of Cys145. In the case of **GpG7** and **GpG4**, the alkylene linkers could be thought of as part of a larger ring system, which may prevent the necessary entry of the alpha carbon into the active site in order for alkyl group transfer. This may be a result from rotation of the O^6-alkylated linked dG residue, which is covalently attached to the O^6-atom of the adjacent dG residue.

Time course assays were performed using 60 pmol hAGT and 2 pmol of either **GpG4** or **GpG7** duplex substrates by quantifying the amounts of unrepaired, **MP** and **RP** products (**Figure S4.16**). The data, summarized in **Figure 4.5**, reveals that **GpG4** was depleted by hAGT at a slower rate with approximately 20% repair occurring in 4 h whereas similar levels of repair for **GpG7** required 1 h. The formation of the **RP** reached a plateau after similar reaction times for **GpG4** and **GpG7** with values within 20-25 %. The formation of the **MP**, on the other hand, plateaued at a higher value for duplexes containing **GpG7** (14%) compared to **GpG4** (3%). The faster repair of **GpG7** compared to **GpG4** is probably due to the added flexibility imparted from the longer alkylene linker, which may allow for more optimal rotation of the damaged nucleotide into the protein active site. The geometrically optimized molecular models of **GpG4**, **GpG7** and unmodified duplexes revealed an O^6-O^6 distance of 3.0, 3.2, and 3.5 Å, respectively. An increase in base tilting observed for **GpG4** relative to **GpG7** suggests that the butylene adduct is more strained relative to the heptylene adduct, which could account for the greater repair observed for **GpG7** relative to **GpG4**. Comparison of the **GpG4** and **GpG7** structures to **GG4** and **GG7** revealed an increase in buckling for IaCL DNA duplexes containing the phosphodiester linkage. Greater repair was also observed for **XL7** versus **XL4** ICL DNA by hAGT.[65] Repair of **GpG4** duplexes by hAGT proceeded slightly faster compared to **XL4**, whereas repair of **XL7** occurred

more quickly compared to **GpG7** duplexes, showing a clear distinction in the efficiency of hAGT towards processing these forms of DNA modifications despite the alkylene linker of the IaCL residing on one strand only.

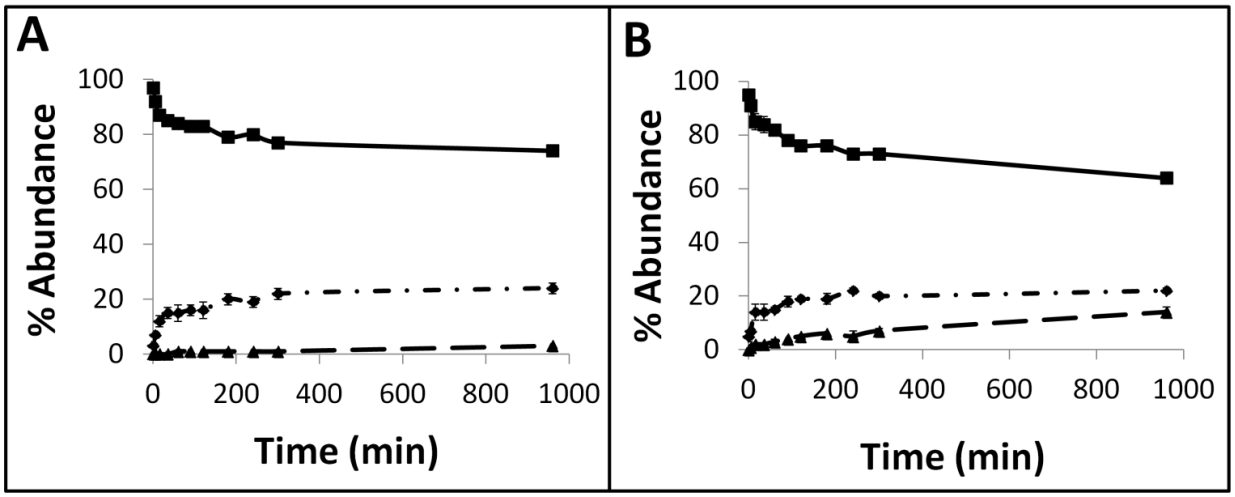


Figure 4.5. Time course repair assay of **GpG4** (left) and **GpG7** (right) by hAGT displaying a faster repair of **GpG7**. Graphical illustrations display abundances [%] of **MP** (- - -), **RP** (—•—) and substrate (—) over time (min).

The time course assay revealed reduced reaction rates for processing of **GpG4** and **GpG7** IaCL duplexes compared to those containing **GG4** and **GG7** by hAGT. After a repair reaction time of 4 h, 80% and 73% of **GpG4** and **GpG7** remained, respectively, whereas virtually all of **GG4** and **GG7** substrates were depleted by hAGT (at 30-fold excess of protein). As stated previously, this drastic difference imparted by the presence of the phosphodiester linkage may likely be the result of the alkylene linkers found in **GpG4** and **GpG7** adopting less reactive conformations, which has been proposed for other O^6-alkylated-dG lesions.[316]

In our previous report of the repair of duplexes containing the **GG4** and **GG7** IaCL by hAGT, we proposed that the phosphorylation of the damaged strand occurred at the 5'-terminal of the oligomer. A more-detailed analysis of the repair event revealed that T4 PNK is capable of

phosphorylating the internal 5'-hydroxyl group as well (**Figure S4.9 – S4.11**). To validate that T4 PNK could in fact phosphorylate both positions, T4 PNK and a three-fold excess of ATP was incubated with either **GG4** or **GG7**. Mass spectral analysis revealed the presence of mono- and bis-phosphorylated products (**Figure S4.5 – S4.6**). To help decipher the repair pathway of **GG4** and **GG7**, a mixture of γ-[^{32}P]ATP (1 uL, 10 μCi) and non-radioactive ATP (3-fold excess) was used to radiolabel the damaged strand DNA. The total repair assay of **GG4** and **GG7** resulted in different ratios of **RP** and **MP**, with substrate consumption levels in agreement with those previously observed (**Figure S4.11**). The variance observed in the **RP** and **MP** levels may have resulted from the mixture of mono- and bis-phosphorylated products of **GG4** and **GG7**.

The radioactivity-based assay has the limitation of only monitoring the DNA species, which prompted us to perform an SDS-PAGE analysis of the reaction to analyze the various proteinaceous products formed (**Figure 4.6**).[151] Lane 1 contains the hAGT protein alone, whereas lanes 2-6 contains hAGT with unmodified DNA, **GG4**, **GG7**, **GpG4**, and **GpG7**, respectively (600 pmol protein incubated with 600 pmol DNA for 16h at 37 ^{o}C). Lanes 3 and 4 both show the presence of two bands, one is minor with a faster migration similar to the unreacted hAGT protein. The major, slower migrating band is presumably the AGT-DNA covalent complex, and is in agreement with our radioactivity-based repair assay and previous preparations of AGT-DNA complexes using ICL DNA.[151] Interestingly, the reaction between a 1 : 1 molar equivalence of **GG4** or **GG7** with hAGT yielded almost quantitative conversion to the DNA-protein covalent complex. Lane 5 displays the lack of **GpG4** repair by hAGT whereas lane 6 reveals the appearance of a faint band with lower mobility compared to the hAGT protein (major band), consistent with the formation of a covalently linked AGT-DNA species.

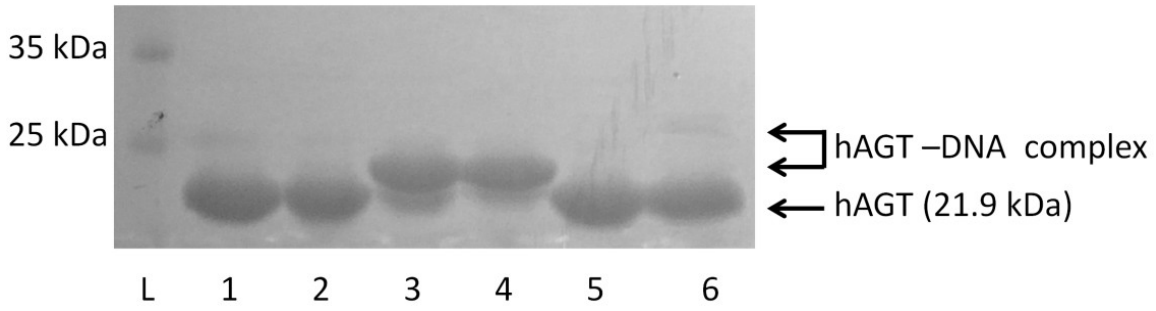


Figure 4.6. 12% SDS-PAGE of hAGT mediated repair of **GG4**, **GG7**, **GpG4**, and **GpG7**. Repair of 600 pmol IaCL DNA by 600 pmol hAGT for 16 h at 37 °C: lane L, unstained protein molecular weight marker; lane 1, hAGT; lane 2 unmodified control DNA + hAGT reaction; lane 3 GG4 + hAGT reaction; lane 4 GG7 + hAGT reaction; lane 5 GpG4 + hAGT reaction; lane 6 GpG7 + hAGT reaction.

The *in vitro* assay results demonstrate that IaCL DNA containing butylene and heptylene linkers are repaired by hAGT adding to our knowledge of the substrate range that this protein can act upon. IaCLs can be introduced in DNA by different bifunctional alkylating agents such as busulfan, however the O^6-position of dG is not the major site of adduct formation. McManus and coworkers previously showed that AGT-deficient CHO cells are sensitive to killing by hepsulfam. It was also shown that hAGT can protect CHO cells, in part, against hepsulfam exposure, but not busulfan. Our preliminary results are consistent with this finding as more efficient hAGT repair observed for the heptylene versus butylene IaCL adducts.

X-ray crystallography studies of hAGT bound to damaged DNA have shown that Tyr114 was necessary for flipping of the damaged nucleotide into the active site of the protein.[137] Y114F and Y114A hAGT mutants have reduced activity towards O^6-alkylated-dG lesions highlighting the important role of Tyr114. In the current study, the presence of the phosphodiester linkage between the O^6-linked IaCL (**GpG4** and **GpG7**) reduced the repair efficiency by hAGT compared to those lacking the phosphodiester group (**GG4** and **GG7**). The basis for the interaction of Tyr114 with the 3'-phosphate group of the target nucleotide has been proposed as

being steric[137] or electronic[168] in nature. The interaction of the phosphate group of **GpG4** and **GpG7** with Tyr114 may be unfavorable, which contributes to lower repair efficiencies observed by hAGT. Investigation of AGT repair of O^6-dG-alkylene-O^6-dG IaCL containing other backbone bio-isosteric linkages, such as formacetal group,[330] may contribute to probing the contributions of the sterics and electronics of Tyr114 towards AGT-mediated repair.

One highlight feature of this study is the efficient repair of single stranded **GG7** by hAGT with only 5-fold excess protein (and similar efficiencies for **GG4** at 60-fold excess hAGT). Given the lack of the phosphodiester linkage at the cross-link site, repair of the IaCL cleaves the DNA into two smaller DNA fragments. This system may be used as a functional irreversible switch for novel molecular nanotechnology devices. Recently, there have been a number of reports describing the activity of AGT towards various DNA structure and applications in protein tagging.[178,307,308] The integration of these IaCL modifications into a device would potentially benefit from orthogonality, and little crosstalk with other switch mechanisms such as pH, UV-radiation and toehold-mediated strand displacement mechanisms. Other types of flexible IaCL such as those linking the O^6-atom of dG to the O^4-atom of dT in DNA sequences are currently being investigated for such applications.

4.3 Conclusions

Cyclic nucleoside dimers containing an alkylene linkage between the O^6-atoms of two 2'-deoxyguanosines were synthesized and the corresponding phosphoramidites were incorporated into DNA by solid-phase synthesis to produce IaCL DNA. Thermal denaturation analysis of duplexes containing these butylene and heptylene IaCL were found to result in a reduction in the T_m by about 20 ^{o}C relative to the control. Circular dichroism and molecular modeling suggested minimal global perturbation in the duplex structure. Both **GpG4** and **GpG7** were repaired

moderately by hAGT, with slightly greater efficiency for the heptylene versus butylene linker. IaCL DNA lacking the phosphodiester group at the cross-link site (**GG4** and **GG7**) were repaired with much greater efficiency. These results contribute an ongoing investigation of AGT to act upon modified DNA structures.

4.4 Acknowledgements

The authors are grateful to Dr. Anthony E. Pegg (Pennsylvania State University) for the plasmid encoding the wild-type hAGT, OGT and Ada-C genes. We are also grateful to Dr. Francis McManus and Lauralicia Sacre for helpful discussions concerning the repair assays and Dr. Anne Noronha for assistance with solid phase synthesis. This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC, Grant No. 299384-2011) and the Canada Research Chair Program (Grant No. 950-213807). D.K.O. is the recipient of a Canada Graduate Scholarship (CGS) from NSERC.

4.5 Supporting information

4.5.1 Supporting Methods

4.5.1.1 General

5'-*O*-(4,4'-dimethoxytrityl)-*N*2-phenoxyacetyl-2'-deoxyguanosine and N,N-diisopropylaminocyanoethylphosphonamidic chloride were purchased from ChemGenes, Inc. (Wilmington, MA). Compounds **2a** and **2b** were prepared according to previously published procedures.[65,311,312] 5'-*O*-Dimethoxytrityl-2'-deoxyribonucleoside-3'-*O*-(β-cyanoethyl-N,N-diisopropyl)phosphoramidites and protected 2'-deoxyribonucleoside-Controlled Pore Glass (CPG) supports were purchased from Glen Research (Sterling, Virginia). All other chemicals and solvents were purchased from the Aldrich Chemical Company (Milwaukee, WI) or EMD

Chemicals Inc. (Gibbstown, NJ). Flash column chromatography was performed using silica gel 60 (230–400 mesh) purchased from Silicycle (Quebec City, QC). Thin layer chromatography (TLC) was carried out with precoated TLC plates (Merck, Kieselgel 60 F_{254}, 0.25 mm) purchased from EMD Chemicals Inc. (Gibbstown, NJ). NMR spectra were recorded on a Varian 500 MHz NMR spectrometer at room temperature. ^{1}H NMR spectra were recorded at a frequency of 500.0 MHz and chemical shifts were reported in parts per million (ppm) downfield from tetramethylsilane. ^{13}C NMR spectra (^{1}H decoupled) were recorded at a frequency of 125.7 MHz and chemical shifts were reported in ppm with tetramethylsilane as a reference. ^{31}P NMR spectra (^{1}H decoupled) were recorded at a frequency of 202.3 MHz and chemical shifts were reported in ppm with H_3PO_4 used as an external standard. High resolution mass spectrometry of modified nucleosides were obtained using a 7T-LTQ FT ICR mass spectrometer (Thermo Scientific) at the Concordia University Centre for Structural and Functional Genomics. The mass spectrometer was operated in full scan, positive ion or negative ion detection mode. ESI mass spectra for oligonucleotides were obtained at the Centre for Biological Applications of Mass Spectrometry (Concordia University) using a Micromass Qtof2 mass spectrometer (Waters) equipped with a nanospray ion source. The mass spectrometer was operated in full scan, negative ion detection mode.

4.5.1.2 Chemical synthesis of nucleosides

The synthesis of phosphoramidites **6a** and **6b** are shown in **Scheme 4.1**.

3',5'-O-bis(tert-butydimethylsilyl)-N^2-phenoxyacetyl-2'-deoxyguanosine (**4.1**):

N^2-phenoxyacetyl-2'-deoxyguanosine (2.0 g, 5.0 mmol) was dried *via* co-evaporation with anhydrous Py (3 x 8 mL) and dissolved in DCM (50 mL). To this solution was added

imidazole (1.7 g, 25 mmol), DMAP (cat.) and TBS-Cl (1.9, 12.5 mmol). A precipitate was immediately formed after the addition of the silylating reagent. After 16 h of stirring, the solvent was partially evaporated in vacuo (to about 50 mL), washed with a saturated aqueous solution of $NaHCO_3$ (2 x 75 mL) and brine (50 mL), dried over anhydrous Na_2SO_4, decanted and the solvent was evaporated in vacuo to produce a yellow gum. The product was purified *via* flash column chromatography using EtOAc : hexanes (1 : 1 to 100% EtOAc) to afford 2.7 g (86 %) R_f (SiO_2 TLC):0.15 EtOAc : hexanes (1 : 1). $\lambda_{max(MeCN)}$ = 260 nm and 285 nm. ^{1}H NMR (500MHz, $CDCl_3$, ppm): 11.77 (s, 1H, NH), 9.19 (s, 1H,NH), 8.00 (s, 1H, H8), 7.39-7.35 (m, 2H, Ar), 7.12-7.08 (m, 1H, Ar), 7.01-6.99 (m, 2H, Ar), 6.28 (t, 1H, H1', J = 6.5 Hz), 4.71 (s, 2H,PhOCH2CO), 4.60 (m, 1H, H3'), 3.90 (m, 1H, H4'), 3.80-3.74 (m, 2H, 2 x H5'), 2.48 (m,1H, H2'), 2.40 (m, 1H, H2'), 0.92 (s, 9H, $(CH_3)_3CSi$), 0.90 (s, 9H, $(CH_3)_3CSi$), 0.11 (s, 6H, $(CH_3)_2Si$), 0.078 (s, 3H, CH_3Si), 0.070 (s, 3H, CH_3Si). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 169.5, 156.4, 155.4, 147.5, 146.1, 137.1, 130.0, 123.0, 122.0, 114.9, 88.0, 83.6, 71.7, 67.0, 62.7, 41.5, 26.0, 25.8, 18.4, 18.0, -4.66, -4.77, -5.39, -5.52. IR (thin film); ν_{max} (cm^{-1}) = 3220, 2954, 2929, 2857, 1701, 1612, 1557, 1496, 1472, 1404, 1257, 1070, 837, 780. HRMS (ESI-MS) *m/z* calculated for $C_{30}H_{48}N_5O_6Si_2^+$ 630.3138: found 630.3141 $[M+H]^+$.

1-{O^6-[5'-O-(4,4'-dimethoxytrityl)-N^2-phenoxyacetyl-2'-deoxyguanidinyl]}-4-{O^6-[3',5'-O-bis(tert-butyldimethylsilyl)-N^2-phenoxyacetyl-2'-deoxyguanidinyl]}-butane (**4.3a**):

To a solution of **4.1** (2.1 g, 2.4 mmol), **4.2a** (1.4 g, 2.2 mmol) and Ph_3P (0.69 g, 2.6 mmol) in anhydrous dioxane (21.5 mL) was added DIAD (0.52 mL, 2.6 mmol) dropwise while stirring. After 16 h, the solvent was evaporated in vacuo and the content taken up in DCM (70 mL), washed with a saturated aqueous solution of $NaHCO_3$(2 x 75 mL), dried over anhydrous Na_2SO_4, decanted and the solvent was once more evaporated to produce a yellow gum. The

product was partially purified *via* flash column chromatography using EtOAc : hexanes (2 : 3, 3 : 2) (R_f (SiO_2 TLC): 0.66 EtOAc), fractions containing the product were pooled, concentrated in vacuo, and used without further purification for the subsequent step. To the residue, Ph_3P (0.22 g, 0.83 mmol) and $(Ph_3P)_4Pd$ (0.44 g, 0.39 mmol) in anhydrous THF (42 mL) was added in a solution of n-$BuNH_2$ (1.5 mL, 15 mmol) and HCOOH (0.58 mL, 15 mmol) in THF (16 mL). After 30 min of stirring, the solvent was evaporated in vacuo and the crude taken up in EtOAc (75 mL), washed with a saturated aqueous solution of $NaHCO_3$ (2 x 100 mL) and brine (50 mL), dried over anhydrous Na_2SO_4, decanted and the solvent was evaporated in vacuo to produce a yellow gum. The product was purified *via* flash column chromatography using EtOAc : hexanes (1 : 1 → 100 % EtOAc) to afford 2.4 g (86 % over 2 steps) of a colorless foam. R_f (SiO_2 TLC): 0.34 EtOAc : hexanes (17 : 1). $\lambda_{max(MeCN)}$ = 269 nm. ^{1}H NMR (500MHz, $CDCl_3$, ppm): 8.85 (m, 2H, NH), 8.17 (s, 1H, H8), 8.03 (s, 1H, H8), 7.39-7.12 (m, 13H, Ar), 7.06-6.98 (m, 6H, Ar), 6.79-6.76 (m, 4H, Ar), 6.58 (dd, 1H, H1'a, J = 6.5 Hz), 6.44 (dd, 1H, H1'b, J = 6.5 Hz), 4.80-4.78 (m, 3H, $PhOCH_2CO$, H3'a), 4.72-4.67 (m, 6H, $PhOCH_2CO$ & $ArO\underline{CH_2}(CH_2)_2\underline{CH_2}OAr$), 4.63 (m, 1H, H3'b), 4.21 (m, 1H, H4'a), 4.01 (m, 1H, H4'b), 3.88 (dd, 1H, H5'b, J = 4.5 & 11.5 Hz), 3.79 (dd, 1H, H5''b, J = 3.5 & 11.5 Hz), 3.75 (s, 6H, 2 x OCH_3), 3.46 (dd, 1H, H5'a, J = 5 & 10.5 Hz), 3.34 (dd, 1H, H5''a, J = 5 & 10 Hz), 2.76 (m, 1H, H2'a), 2.68-2.59 (m, 2H, H2''a & H2'b), 2.44 (m, 1H, H2''b), 2.13 (m, 4H, CH_2CH_2), 1.84 (bs, 1H, OH), 0.924 (s, 9H, $(CH_3)_3CSi$), 0.919 (s, 9H, $(CH_3)_3CSi$), 0.119 (s, 3H, $(CH_3)Si$), 0.115 (s, 3H, $(CH_3)Si$), 0.093 (s, 6H, $(CH_3)_2Si$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 161.1, 158.5, 157.3, 152.4, 151.1, 151.0, 144.6, 140.3, 135.8, 135.7, 132.1, 132.04, 131.96, 131.93, 130.03, 130.01, 129.78, 129.76, 128.6, 128.5, 128.1, 127.8, 126.9, 122.3, 122.2, 118.9, 114.92, 114.91, 113.1, 88.0, 86.5, 86.4, 84.2, 84.2, 72.5, 71.9, 68.1, 67.9, 67.12, 67.05, 64.1, 62.8, 55.2, 41.2, 40.6, 26.0, 25.8, 25.40, 25.36,

18.4, 18.0, -4.64, -4.78, -5.35, -5.47. IR (thin film); ν_{max} (cm^{-1}) = 3410, 3255, 2954, 2929, 2856, 1700 1607, 1510, 1496, 1249, 1176, 1074, 1033, 836, 780. HRMS (ESI-MS) *m/z* calculated for $C_{73}H_{91}N_{10}O_{14}Si_2^+$ 1387.6249: found 1387.6260 $[M+H]^+$.

1-{O^6-[5'-O-(4,4'-dimethoxytrityl)-N^2-phenoxyacetyl-2'-deoxyguanidinyl]}-7-{O^6-[3',5'-O-bis(tert-butyldimethylsilyl)-N^2-phenoxyacetyl-2'-deoxyguanidinyl]}-heptane (**4.3b**):

To a solution of **4.1** (2.4 g, 2.7 mmol), **4.2b** (1.5 g, 2.4 mmol) and Ph_3P (0.77 g, 2.9 mmol) in anhydrous dioxane (26 mL) was added DIAD (0.58 mL, 2.9 mmol) dropwise while stirring. After 16 h, the solvent was evaporated in vacuo and the content taken up in DCM (75 mL), washed with a saturated aqueous solution of $NaHCO_3$ (2 x 80 mL) and brine (70 mL), dried over anhydrous Na_2SO_4, decanted and the solvent was once more evaporated to produce a yellow gum. The product was partially purified *via* flash column chromatography using EtOAc : hexanes (1 : 4 → 3 : 2) (R_f (SiO_2 TLC): 0.73 EtOAc), fractions containing the product were pooled, concentrated in vacuo, and used without further purification for the subsequent step. To the resulting residue, Ph_3P (0.24 g, 0.92 mmol) and $(Ph_3P)_4Pd$ (0.50 g, 0.43 mmol) in anhydrous THF (48 mL) was added in a solution of n-$BuNH_2$ (1.7 mL, 17 mmol) and HCOOH (0.65 mL, 17 mmol) in THF (18 mL). After 15 min of stirring, the solvent was evaporated in vacuo and the crude taken up in EtOAc (75 mL), washed with a saturated aqueous solution of $NaHCO_3$ (2 x 80 mL) and brine (70 mL), dried over anhydrous Na_2SO_4, decanted and the solvent was evaporated in vacuo to produce a yellow gum. The product was purified *via* flash column chromatography using EtOAc : hexanes (1 : 1 → 7 : 3) to afford 2.8 g (82 % over 2 steps) of a colorless foam. R_f (SiO_2 TLC): 0.54 EtOAc : hexanes (17 : 1). $\lambda_{max(MeCN)}$ = 269 nm. ^{1}H NMR (500MHz, $CDCl_3$, ppm): 8.79 (bs, 1H, NH), 8.74 (bs, 1H, NH), 8.15 (s, 1H, H8), 8.03 (s, 1H, H8), 7.42-7.16 (m, 13H, Ar), 7.05-6.98 (m, 6H, Ar), 6.79-6.76 (m, 4H, Ar), 6.66 (dd, 1H, H1'a, J = 6.5 Hz), 6.44

(dd, 1H, H1'b, J = 6.5 Hz), 4.82-4.80 (m, 3H, PhOCH_2CO, H3'a), 4.66-4.55 (m, 7H, PhOCH_2CO, H3'b & ArO$\underline{CH_2}(CH_2)_2\underline{CH_2}$OAr), 4.25 (m, 1H, H4'a), 4.02 (m, 1H, H4'b), 3.88 (dd, 1H, H5'b, J = 4.5 & 11 Hz), 3.79 (dd, 1H, H5''b, J = 3.5 & 11.5 Hz), 3.75 (s, 6H, 2 x OCH_3), 3.46 (dd, 1H, H5'a, J = 5 & 10.5 Hz), 3.34 (dd, 1H, H5''a, J = 4 & 10 Hz), 2.74 (m, 1H, H2'a), 2.69-2.63 (m, 2H, H2''a & H2'b), 2.44 (m, 1H, H2''b), 1.89 (m, 4H, $\underline{CH_2}CH_2CH_2CH_2\underline{CH_2}$), 1.56-1.44 (m, 6H, $CH_2\underline{CH_2CH_2CH_2}CH_2$), 0.925 (s, 9H, $(CH_3)_3CSi$), 0.920 (s, 9H, $(CH_3)_3CSi$), 0.119 (s, 3H, $(CH_3)Si$), 0.115 (s, 3H, $(CH_3)Si$), 0.094 (s, 6H, $(CH_3)_2Si$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 161.1, 158.5, 157.3, 152.4, 151.1, 151.0, 144.6, 140.3, 135.8, 135.7, 132.1, 132.04, 131.96, 131.93, 130.03, 130.01, 129.78, 129.76, 128.6, 128.5, 128.1, 127.8, 126.9, 122.3, 122.2, 118.9, 114.92, 114.91, 113.1, 88.0, 86.5, 86.4, 84.23, 84.16, 72.5, 71.9, 68.1, 68.0, 67.12, 67.05, 64.1, 62.8, 55.2, 41.2, 40.6, 26.0, 25.8, 25.40, 25.36, 18.4, 18.0, -4.64, -4.78, -5.35, -5.47. IR (thin film); ν_{max} (cm^{-1}) = 3410, 3057, 2953, 2931, 2856, 1717, 1607, 1510, 1496, 1381, 1248, 1175, 1073, 836, 780. HRMS (ESI-MS) *m/z* calculated for $C_{76}H_{97}N_{10}O_{14}Si_2^+$ 1429.6719: found 1429.6737: found $[M+H]^+$.

1-{O^6-[3'-O-(2-chlorophenyl-phosphate)-5'-O-(4,4'-dimethoxytrityl)-N^2-phenoxyacetyl-2'-deoxyguanidinyl]}-4-{O^6-[3',5'-O-bis(tert-butyldimethylsilyl)-N^2-phenoxyacetyl-2'-deoxyguanidinyl]}-butane (**4.4a**):

To a solution of 1,2,4-triazole (0.25 g, 3.6 mmol) and NEt_3 (0.50 mL, 3.6 mmol) in anhydrous THF (19 mL) was added Cl_2(*o*-ClPhO)PO (0.30 mL, 1.8 mmol) dropwise at room temperature. A precipitate was formed immediately, which slightly dissipated over 10 - 20 min. After 40 min of vigorous stirring, the reaction was transferred by canula and filtered through a sintered glass funnel into another flask containing a solution of **4.3a** (0.50 g, 0.36 mmol) in anhydrous Py (8 mL). The transfer / filtration was accelerated under reduced pressure. After 1 h

of stirring, the reaction was quenched using a saturated solution of $NaHCO_3$ (10 mL), where a precipitate was observed to form. The solvent was partially removed in vacuo at 40 ^oC and the resulting residue was taken up in DCM (50 mL) containing NEt_3 (0.3 mL), washed with water (2 x 80 mL) followed by a saturated solution of NH_4Cl (70 mL) then dried over anhydrous Na_2SO_4 (≈ 4 g). A small amount of MeOH (1 mL) was added to the dried organic phase to help with solubility. The organic layer was filtered and the solvent was evaporated to produce a yellow gum. The product was purified *via* flash column chromatography using MeOH : DCM (5 %, 9 %) containing a constant 0.12 % (v/v) of NEt_3 to afford 0.53 g (93 %) of a colorless foam. R_f (SiO_2 TLC): 0.44 MeOH : DCM (10 %). $\lambda_{max(MeCN)}$ = 270 nm. ^{1}H NMR (500MHz, $CDCl_3$, ppm): 8.84 (bs, 1H, NH), 8.81 (bs, 1H, NH), 8.16 (s, 1H, H8), 7.95 (s, 1H, H8), 7.63 (d, 1H, Ar, *J* = 8 Hz), 7.35-7.10 (m, 15H, Ar), 7.01-6.90 (m, 7H, Ar), 6.73-6.70 (m, 4H, Ar), 6.47 (dd, 1H, H1'a, *J* = 7 Hz), 6.42 (dd, 1H, H1'b, *J* = 6.5 Hz), 5.22 (m, 1H, H3'a), 4.83-4.81 (m, 4H, 2 x $PhOCH_2CO$), 4.69 (m, 4H, $ArO\underline{CH_2}(CH_2)_2\underline{CH_2}OAr$), 4.62 (m, 1H, H3'b), 4.45 (m, 1H, H4'a), 3.99 (dd, 1H, H4'b, J = 3.5 & 7.5 Hz), 3.85 (dd, 1H, H5'b, *J* = 4 & 11.5 Hz), 3.77 (dd, 1H, H5''b, *J* = 3.5 & 11.5 Hz), 3.73 (s, 6H, 2 x OCH_3), 3.39 (dd, 1H, H5'a, *J* = 5 & 10.5 Hz), 3.29 (dd, 1H, H5''a, *J* = 3 & 10.5 Hz), 2.84-2.81 (m, 2H, H2'a & H2''a), 2.65 (m, 1H, H2'b), 2.42 (m, 1H, H2''b), 2.10 (m, 4H, CH_2CH_2), 0.902 (s, 9H, $(CH_3)_3CSi$), 0.897 (s, 9H, $(CH_3)_3CSi$), 0.098 (s, 3H, $(CH_3)Si$), 0.094 (s, 3H, $(CH_3)Si$), 0.071 (s, 6H, $(CH_3)_2Si$). Residual trimethylamine peaks are observed at 2.97 ppm and 1.26 ppm. ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 161.04, 161.02, 158.4, 157.4, 157.3, 152.6, 152.4, 151.2, 151.1, 149.2, 149.1, 144.5, 140.4, 140.1, 135.7, 135.5, 130.1, 129.90, 129.85, 129.73, 129.70, 128.2, 127.8, 127.6, 127.5, 126.8, 124.93, 124.87, 123.8, 123.6, 122.2, 122.0, 121.7, 121.41, 121.39, 118.8, 118.6, 114.90, 114.87, 113.1, 88.0, 86.5, 85.8, 85.7, 84.2, 84.1, 71.9, 68.1, 68.0, 67.0, 64.2, 62.8, 55.2, 41.2, 39.5, 26.0, 25.8, 25.4,

25.3, 18.4, 18.0, -4.65, -4.79, -5.36, -5.48. Residual trimethylamine peaks are observed at 45.8 ppm and 8.68 ppm. ^{31}P NMR (202.3 MHz, $CDCl_3$, ppm): -6.68. IR (thin film); ν_{max} (cm^{-1}) = 3410, 3063, 2954, 2930, 2856, 2635, 2486, 1700, 1606, 1509, 1496, 1383, 1247, 1176, 1081, 1032, 837, 780. HRMS (ESI-MS) *m/z* calculated for $C_{79}H_{93}ClN_{10}O_{17}PSi_2^-$ 1575.5690: found 1575.5634 $[M-H]^-$.

1-{O^6-[3'-O-(2-chlorophenyl-phosphate)-5'-O-(4,4'-dimethoxytrityl)-N^2-phenoxyacetyl-2'-deoxyguanidinyl]}-7-{O^6-[3',5'-O-bis(tert-butyldimethylsilyl)-N^2-phenoxyacetyl-2'-deoxyguanidinyl]}-heptane (**4.4b**):

To a solution of 1,2,4-triazole (0.25 g, 3.6 mmol) and NEt_3 (0.50 mL, 3.6 mmol) in anhydrous THF (19 mL) was added Cl_2(*o*-ClPhO)PO (0.30 mL, 1.8 mmol) dropwise at room temperature. A precipitate was formed immediately, which slightly dissipated over 10 - 20 min. After 40 min of vigorous stirring, the reaction was transferred by canula then filtered through a sintered glass funnel into another flask containing a solution of **4.3b** (0.50 g, 0.36 mmol) in anhydrous Py (8 mL). The transfer / filtration was accelerated under reduced pressure. After 1 h of stirring, the reaction was quenched using a saturated solution of $NaHCO_3$ (10 mL), where a precipitate was observed to form. The solvent was partially removed in vacuo at 40^{o}C and the resulting residue was taken up in DCM (50 mL) containing NEt_3 (0.3 mL), washed with water (2 x 80 mL) followed by a saturated solution of NH_4Cl (70 mL) then dried over anhydrous Na_2SO_4 (≈ 4 g). A small amount of MeOH (1 mL) was added to the dried organic phase to help with solubility. The organic layer was filtered and the solvent was evaporated to produce a yellow gum. The product was purified *via* flash column chromatography using MeOH : DCM (4 %, 8%) containing a constant 0.12 % (v/v) of NEt_3 to afford 0.50 g (88 %) of a colorless foam. R_f (SiO_2 TLC):0.46 MeOH : DCM (10 %). $\lambda_{max(MeCN)}$ = 269 nm. ^{1}H NMR (500MHz, $CDCl_3$, ppm): 8.73

(bs, 1H, NH), 8.64 (bs, 1H, NH), 8.15 (s, 1H, H8), 7.96 (s, 1H, H8), 7.64 (d, 1H, Ar, *J* = 8.5 Hz), 7.38-7.12 (m, 15H, Ar), 7.05-7.01 (m 6H, Ar), 6.96-6.93 (m, 1H, Ar), 6.74-6.72 (m, 4H, Ar), 6.49 (dd, 1H, H1'a, *J* = 6 & 8.5 Hz), 6.44 (dd, 1H, H1'b, *J* = 6.5 Hz), 5.22 (m, 1H, H3'a), 4.83 (m, 4H, 2 x $PhOCH_2CO$), 4.64-4.57 (m, 5H, H3'b & $ArO\underline{CH_2}(CH_2)_5\underline{CH_2}OAr$), 4.46 (m, 1H, H4'a), 4.01 (dd, 1H, H4'b, J = 3.5 & 7 Hz), 3.87 (dd, 1H, H5'b, *J* = 4.5 & 11.25 Hz), 3.78 (dd, 1H, H5''b, *J* = 3 & 11.25 Hz), 3.75 (s, 6H, 2 x OCH_3), 3.40 (dd, 1H, H5'a, *J* = 5 & 10 Hz), 3.31 (dd, 1H, H5''a, *J* = 3 & 10 Hz), 2.90-2.80 (m, 2H, H2'a & H2''a), 2.66 (m, 1H, H2'b), 2.43 (m, 1H, H2''b), 1.90 (m, 4H, CH_2 & CH_2), 1.45 (m, 6H, $CH_2CH_2CH_2$), 0.915 (s, 9H, $(CH_3)_3CSi$), 0.910 (s, 9H, $(CH_3)_3CSi$), 0.112 (s, 3H, $(CH_3)Si$), 0.108 (s, 3H, $(CH_3)Si$), 0.086 (s, 6H, $(CH_3)_2Si$). Residual trimethylamine peaks are observed at 3.10 ppm and 1.39 ppm ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 161.2, 158.4, 152.3, 151.1, 140.2, 135.6, 130.1, 129.8, 128.1, 127.7, 122.2, 114.9, 113.1, 88.0, 84.2, 71.9, 68.1, 67.8, 62.8, 55.1, 41.2, 29.2, 28.8, 26.0, 25.9, 25.8, 25.7, 18.4, 18.0, -4.64, -4.78, -5.35, -5.48. Residual trimethylamine peaks are observed at 45.7 ppm and 8.54 ppm. ^{31}P NMR (202.3 MHz, $CDCl_3$, ppm): -6.73. IR (thin film); ν_{max} (cm^{-1}) = 3409, 3057, 2953, 2931, 2856, 2485, 1717, 1607, 1510, 1496, 1443, 1419, 1382, 1247, 1175, 1081, 1032, 836, 754. HRMS (ESI-MS) *m/z* calculated for $C_{82}H_{99}ClN_{10}O_{17}PSi_2^-$ 1617.6160: found 1617.6145$[M-H]^-$.

Compound **4.5a**

To a solution of **4.4a** (0.96 g, 0.61 mmol) in THF (61 mL) was added TBAF (1M in THF) (6.1 mL, 6.1 mmol) dropwise while stirring. After 1h, the solvent was evaporated and the content taken up in DCM (75 mL), washed with a saturated solution of aqueous $NaHCO_3$ (2 x 75 mL) and brine (50 mL), then dried over anhydrous Na_2SO_4 (≈ 4 g), decanted and the solvent was once more evaporated to produce a yellow gum. The crude intermediate was further dried by co-

evaporation with anhydrous Py (3 x 10 mL), and then dissolved in anhydrous Py (130 mL) followed by the addition of MSNT (0.90 g, 3.0 mmol). After 4 h of stirring, the reaction was quenched with H_2O (10 mL) and the solvent evaporated in vacuo. The resulting residue was taken up in DCM (80 mL), washed with a saturated solution of aqueous $NaHCO_3$ (2 x 80 mL) and brine (50 mL), dried over anhydrous Na_2SO_4 (≈ 4 g), decanted and the solvent was evaporated to produce a yellow gum. The product was purified *via* flash column chromatography using MeOH : DCM (2 %→ 4 %) to afford 0.41 g (51 %) of a colorless solid. R_f (SiO_2 TLC):0.35 MeOH : DCM (6 %). $\lambda_{max(MeCN)}$ = 265 nm. ^{31}P NMR (202.3 MHz, $CDCl_3$, ppm): -8.51, -9.19. IR (thin film); ν_{max} (cm^{-1}) = 3406, 3251, 3058, 2932, 2837, 1700, 1608, 1510, 1496, 1419, 1384, 1250, 1176, 1085, 1032, 957, 833, 792, 755. HRMS (ESI-MS) *m/z* calculated for $C_{67}H_{65}ClN_{10}O_{16}P^+$ 1331.4001: found 1331.4008 $[M+H]^+$.

Compound **4.5b**

To a solution of **4.4b** (0.50 g, 0.31 mmol) in THF (31 mL) was added TBAF (1M in THF) (3.1 mL, 3.1 mmol) dropwise while stirring. After 6h, the solvent was evaporated and the content taken up in DCM (75 mL), washed with a saturated solution of aqueous $NaHCO_3$ (2 x 75 mL) and brine (50 mL), then dried over anhydrous Na_2SO_4 (≈ 4 g), decanted and the solvent was once more evaporated to produce a yellow gum. The crude intermediate was further dried by co-evaporation with anhydrous Py (3 x 10 mL), then dissolved in anhydrous Py (65 mL) followed by the addition of MSNT (0.48 g, 1.5 mmol). After 16 h of stirring, the reaction was quenched with H_2O (10 mL), and the solvent evaporated in vacuo. The resulting residue was taken up in DCM (70 mL), washed with a saturated solution of aqueous $NaHCO_3$ (2 x 80 mL) and brine (50 mL), dried over anhydrous Na_2SO_4 (≈ 4 g), decanted and the solvent was once more evaporated to produce a yellow gum. The product was purified *via* flash column chromatography using

MeOH : DCM (2 %→ 3.3 %) to afford 0.22 g (52 %) of a colorless solid. R_f (SiO_2 TLC):0.41 MeOH : DCM (6 %). $\lambda_{max(MeCN)}$ = 267 nm. ^{31}P NMR (202.3 MHz, $CDCl_3$, ppm): -7.82, -8.50. IR (thin film); ν_{max} (cm^{-1}) =3407, 3257, 3058, 2930, 2853, 1716, 1607, 1510, 1496, 1443, 1420, 1383, 1247, 1176, 1084, 1014, 953, 832, 755. HRMS (ESI-MS) *m/z* calculated for $C_{70}H_{70}ClN_{10}NaO_{16}P^+$ 1395.4290: found 1395.4245 $[M+Na]^+$.

Compound **4.6a**

To a solution of **4.5a** (0.28 g, 0.21 mmol) and DIPEA (62 µL, 0.40 mmol) in anhydrous THF (2.1 mL) was added Cl-POCEN*i*Pr_2 (70 µL, 0.32 mmol) dropwise while stirring. After 10 min, a precipitate was formed and after 30 min the solvent was evaporated in vacuo and the crude taken up in EtOAc (70 mL), washed with a saturated solution of aqueous $NaHCO_3$ (2 x 75 mL) and brine (50 mL), then dried over anhydrous Na_2SO_4 (≈ 4 g). The organic layer was filtered and the solvent was evaporated to produce a yellow gum. The product, a slightly yellow oil, taken up in EtOAc (2 mL) and was precipitated from hexanes (1 L) to afford 0.30 g of a yellow foam. R_f (SiO_2 TLC):0.095, 0.214, 0.310 EtOAc. $\lambda_{max(MeCN)}$ = 265 nm. ^{31}P NMR (202.3 MHz, d_6-acetone, ppm): 149.35, 148.87, 148.81, 148.72, -6.90, -6.93, -9.54, -9.71. IR (thin film); ν_{max} (cm^{-1}) = 3579, 3410, 2967, 2837, 2253, 1711, 1607, 1511, 1445, 1419, 1383, 1250, 1177, 1084, 955, 831, 756. HRMS (ESI-MS) *m/z* calculated for $C_{76}H_{81}ClN_{12}NaO_{17}P_2^+$ 1553.4899: found 1553.4860 $[M+H]^+$.

Compound **4.6b**

To a solution of **4.5b** (0.22 g, 0.16 mmol) and DIPEA (52 µL, 0.30 mmol) in anhydrous THF (2.0 mL) was added Cl-POCEN*i*Pr_2 (53 µL, 0.24 mmol) dropwise while stirring. After 10 min, a precipitate was formed and after 30 min the solvent was evaporated in vacuo and the

crude taken up in EtOAc (60 mL), washed with a saturated solution of aqueous $NaHCO_3$ (2 x 70 mL) and brine (50 mL), then dried over anhydrous Na_2SO_4 (≈ 4 g). The organic layer was filtered and the solvent was evaporated to produce a yellow gum. The product, a yellow gum, taken up in EtOAc (2 mL) and was precipitated from hexanes (1 L) to afford 0.24 g of a yellow foam. R_f (SiO_2 TLC):0.354, 0.430 NEt_3 : EtOAc (3 %). $\lambda_{max(MeCN)}$ = 269 nm. ^{31}P NMR (202.3 MHz, d_6-acetone, ppm): 148.97, 148.80, 148.74, 148.61, -7.40, -7.48, -8.77, -8.88. IR (thin film); ν_{max} (cm^{-1}) = 3411, 2965, 1710, 1607, 1510, 1496, 1445, 1421, 1383, 1248, 1177, 1084, 1031, 957, 831, 756. HRMS (ESI-MS) *m/z* calculated for $C_{79}H_{87}ClN_{12}NaO_{17}P_2^+$ 1595.5368: found 1595.5359 $[M+H]^+$.

Abbreviations: DCM, dichloromethane; DIAD, diisopropyl azodicarboxylate; DMAP, 4-dimethylaminopyridine; EtOAc, ethyl acetate; MeCN, acetonitrile; MeOH, methanol; MSNT, 1-(mesitylene-2-sulfonyl)-3-nitro-1,2,4-triazole; Py, pyridine; TBS-Cl, *tert*-butyldimethylsilyl chloride; THF, tetrahydrofuran .

4.5.1.3 Preparation, purification and characterization of the IaCL oligonucleotide

The cross-linked duplexes, the sequences for which are shown in **Figure 4.1**, were assembled with an Applied Biosystems Model 3400 synthesizer on a 1.5 mmol scale using β-cyanoethylphosphoramidite chemistry supplied by the manufacturer with slight modifications to coupling times. The nucleoside phosphoramidites protected with "fast-deprotecting" groups were prepared in anhydrous MeCN at a concentration of 0.1 M for the 3'-*O*-deoxyphosphoramidites, and 0.15 M for the cross-linked 3'-*O*-deoxyphosphoramidite. Oligomer sequence assembly was carried out as previously described.[323] The capping step of the assembly was carried out using

phenoxyacetic anhydride/pyridine/tetrahydrofuran 1:1:8 (v/v/v; solution A) and 1-methyl-imidazole/ tetrahydrofuran 16:84 (w/v; solution B). Coupling wait times for phosphoramidites **4.6a** and **4.6b** were extended to 10 min (compared to 2 min for the commercially available phosphoramidites). Protecting group removal and cleavage from the solid support was carried out by treatment with aqueous (28% v/v) ammonium hydroxide for 17 h at room temperature with mild rocking in 2 mL screw-cap microfuge tubes fitted with Teflon lined caps, followed by an additional 4 h at 55 ^{o}C. Crude oligomers were transferred and lyophilized in a Speedvac concentrator followed by purification by strong anion exchange HPLC using a Dionex DNAPAC PA-100 column (0.4 cm x 25 cm) purchased from Dionex Corp, Sunnyvale, CA using a linear gradient of 0–52% buffer B over 24 min (buffer A: 100 mm Tris HCl, pH 7.5, 10% MeCN and buffer B: 100 mm Tris HCl, pH 7.5, 10% MeCN, 1M NaCl) at 55 ^{o}C. The columns were monitored at 260 nm for analytical runs or 280 nm for preparative runs. The purified oligomers were desalted using C-18 SEP PAK cartridges (Waters Inc.) as previously described.[317] Modified oligomers (0.05 A_{260} units) were characterized by digestion with snake venom phosphodiesterase (0.28 units) and calf intestinal phosphatase (5 units) in a buffer containing 10 mM Tris (pH 8.1) and 2 mm magnesium chloride for 3 days at 37 ^{o}C. The resulting mixture of nucleosides was analyzed by reversed phase HPLC using a Symmetry C-18 5 mm column (0.46 x 15 cm) purchased from Waters Inc., Milford, MA. A linear gradient of 0–30% buffer B over 30 min (buffer A, 50 mm sodium phosphate, pH 5.8 and buffer B, 50 mm sodium phosphate, pH 5.8, 50% MeCN) was used to elute the analytes. The peaks in the elution profile were identified by co-injection with the corresponding standards with the following elution times: dC (4.5 min),dG (7.4 min), dT (8.0 min), dA (9.0 min), and the ratio of nucleosides was determined. The results are given in Table 1 of the main text. The molecular mass of the modified oligomers were

identified by ESI-MS and the measured values were in agreement with the expected masses (Supporting Information **Figure 4.3** and **4.4** for MS spectra).

4.5.1.4 UV thermal denaturation

Molar extinction coefficients for the unmodified and cross-linked oligonucleotides were calculated from those of the mononucleotides and dinucleotides using the nearest-neighbor approximations (M^{-1} cm^{-1}). All duplexes were prepared by mixing equimolar amounts of the interacting strands and lyophilizing the mixture to dryness. The resulting pellet was then re-dissolved in 90mM sodium chloride, 10 mM sodium phosphate, 1 mM EDTA buffer (pH 7.0) to give a final concentration of 3.5 μM duplex. Prior to the thermal run, samples were degassed by placing them in a speed- vac concentrator for 2 min. Annealing curves were acquired at 260 nm starting at 95 ^{o}C and decreasing temperature at a rate of cooling of 0.5 ^{o}C min^{-1} until 15 ^{o}C, using a Varian CARY Model 3E spectrophotometer fitted with a 6-sample thermostated cell block and a temperature controller. The samples were then denatured by heating from 15 ^{o}C to 95 ^{o}C at an increasing temperature rate of 0.5 ^{o}C min^{-1} to show reversibility. Denaturing data processing was carried out as described by Puglisi and Tinoco[254] and transferred to Microsoft Excel™ for viewing. Thermodynamic parameters of the control duplex and IaCL-containing duplexes were obtained by performing thermal denaturation studies in triplicate at concentrations of 0.7, 2.8, 13.8 and 69 μM duplex DNA in cuvettes with path length of 1, 0.5, 0.2 and 0.1 cm, respectively. A plot of $\ln(C_{tot}/4)$ as a function of $1/T_m$, used for a non- self-complementary bimolecular systems, was employed to obtain ΔH° and ΔS° for the system as described by Puglisi and Tinoco.[254]

4.5.1.5 Circular dichroism (CD) spectroscopy

Circular dichroism spectra were obtained on a Jasco J-815 spectropolarimeter equipped with a Julaba F25 circulating bath. Samples were allowed to equilibrate for 5 – 10 min at 15 ^{o}C in 90 mM sodium chloride, 10 mM sodium phosphate, 1 mM EDTA (pH 7.0), at a final concentration of 3.5 μM. Each spectrum was an average of 5 scans, collecting at a rate of 50 nm min^{-1} , with a bandwidth of 1 nm and sampling wavelength of 0.2 nm using fused quartz cells (Starna 29-Q-10). The CD spectra were recorded from 350 to 220 nm at 15 ^{o}C. The molar ellipticity (φ) was calculated from the equation $\varphi = \varepsilon/Cl$, where ε is the relative ellipticity (mdeg), C is the molar concentration of oligonucleotides (moles/L), and l is the path length of the cell (cm). The data were processed using software supplied by the manufacturer (JASCO, Inc.) and transferred into Microsoft Excel^{TM} for viewing.

4.5.1.6 Molecular modelling

Molecular modeling was performed by using the Hyperchem 7.5 software package from Hypercube utilizing the AMBER force field. Hybridized oligomers containing a dGG·dCC, dGpG4·dCC, dGpG4·dCC base pair were constructed from the nucleic acid template option using a B-form duplex. Sequence contexts were truncated to the following 5'-GGC TXX GAT C and 3'-CCG ACC CTA G for proper solvation. Duplexes were solvated with water using a periodic box occupying at least 3 times the volume of the duplex alone. Standard Amber99 parameters were used with the dielectric set to constant. "One to four scale factors" non-bonded interactions were set to 0.5 (both electrostatic and van der Waals). Cutoffs were applied to "switched" to an outer and inner radius of 15.3 and 11.3 Å. All structures were geometry optimized using Polak-Ribiere conjugate gradient until the RMS gradient was less than 0.1 kcal/(Å mol) using the periodic boundary condition option.

4.5.1.7 Protein expression and purification

Ampicillin, isopropyl β-D-thiogalactopyranoside (IPTG), and most other biochemical reagents as well as polyacrylamide gel materials were purchased from Bioshop Canada Inc (Burlington, ON). Ni-NTA Superflow Resin was purchased from Qiagen (Mississauga, ON). Complete, Mini, EDTA-free Protease Inhibitor Cocktail Tablets were obtained from Roche (Laval, QC). Nitro-cellulose filters (0.20 μm) were obtained from Millipore. XL-10 Gold and Bl21(DE3) *E. coli* cells were obtained from Stratagene (Cedar Creek, TX). T4 polynucleotide kinase (PNK) was obtained from Fermentas (Burlington, ON). [γ-^{32}P]ATP was purchased from PerkinElmer (Woodbridge, ON). Phusion Polymerase was obtained from New England Biolabs (Ipswich, MA). DNA primers for site directed mutagenesis and cloning were purchased from Biocorp (Montreal, QC). All AGT homologues were expressed under the promoter of the pQE30 vector in XL-10 Gold *E. coli* cells, as previously described.[149–151,331]

4.5.1.8 AGT repair assay of GpG4 and GpG7

DNA substrates were labeled at the using γ-[^{32}P]-ATP as previously described.[331] Briefly, a 20 μM solution of DNA was made in 1X PNK buffer along with 1 μL γ-[^{32}P]-ATP (10 μCi μL^{-1}) and 5 units of T4 PNK. The labelling reaction was conducted for 1 h at 37 °C after which the reaction was terminated by boiling the sample for 5 min. 100 pmol of labeled DNA was added to 110 pmol of the complement strand in a total volume of 50 μL of water making a 2 μM dsDNA solution with 10% excess of the non-damaged strand. The solution was boiled for 2 min, cooled slowly to room temperature and kept in a refrigerator at 4 °C overnight to ensure proper annealing of the duplex. The repair reaction mixtures were constituted of 2 pmol of the DNA duplex and 10 pmol of AGT in a total volume of 15 μL of Activity Buffer [10 mM Tris–HCl (pH 7.6), 100 mM NaCl and 1 mM DTT] and allowed to react at 37 °C overnight. The reaction was terminated by the addition of 18.2 μL of stop buffer [81 mM Tris–HCl, 81 mM boric acid, 1.8

mM EDTA and 1% SDS (sodium dodecyl sulfate) (pH 8.0) in 80% formamide] followed by boiling for 2 min. Samples (10 μL) were loaded on a 14 cm × 16 cm, 20% 7 M urea denaturing polyacrylamide gel (19 : 1) for separation. The gels were run using 1X TBE for 4h at 475 V and the gels exposed to a storage phosphor screen. The image was captured on a Typhoon 9400 (GE Healthcare, Piscataway, NJ) and the autoradiography counts obtained by Image-Quant™ (Amersham Biosciences). For the repair time course assay of **GpG4** and **GpG7** duplexes, master mixes of 120 μL composed of 400 pmol of hAGT and 13.3 pmol of DNA substrate were prepared. Each sample was placed at 37 ^{o}C and at each time point, 7.5 μL was removed from the master mix and placed in a tube with 9.1 μL of stop reaction buffer and analysis was achieved as described above. Repair experiments were repeated with single stranded IaCL DNA **GG4** and **GG7** as previously described with slight modifications.[323] Labelling of the DNA strands was carried out as above with an additional 300 pmol of non-radioactive ATP. All **GG4** and **GG7** samples were analyzed on polyacrylamide gel electrophoresis (using 1X TBE for 50 min at 425 V) and visualized as described above.

4.5.2 Supporting Figures and Tables

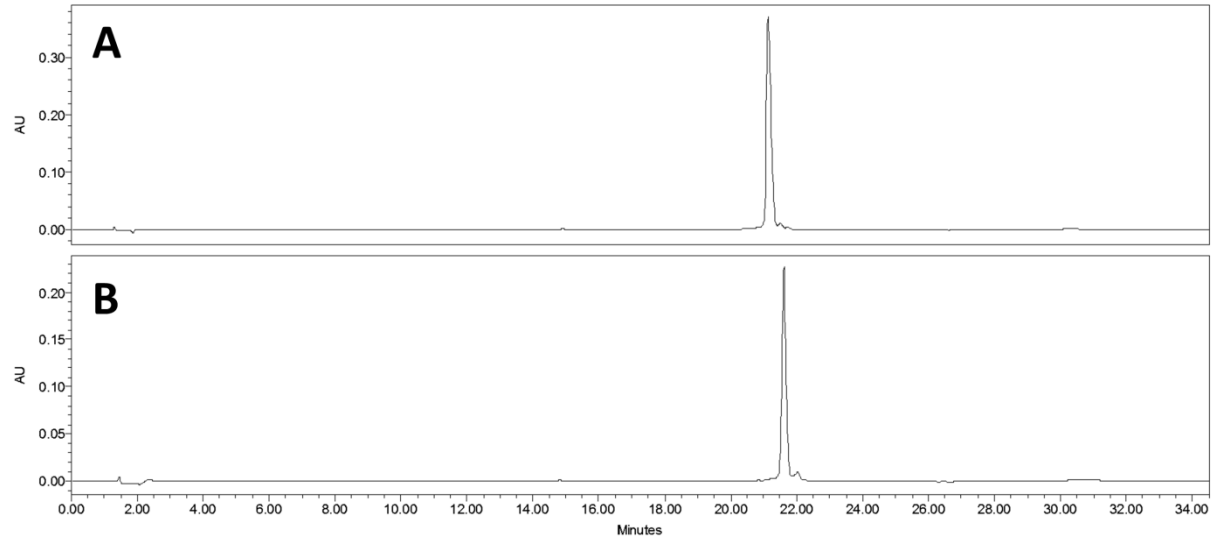
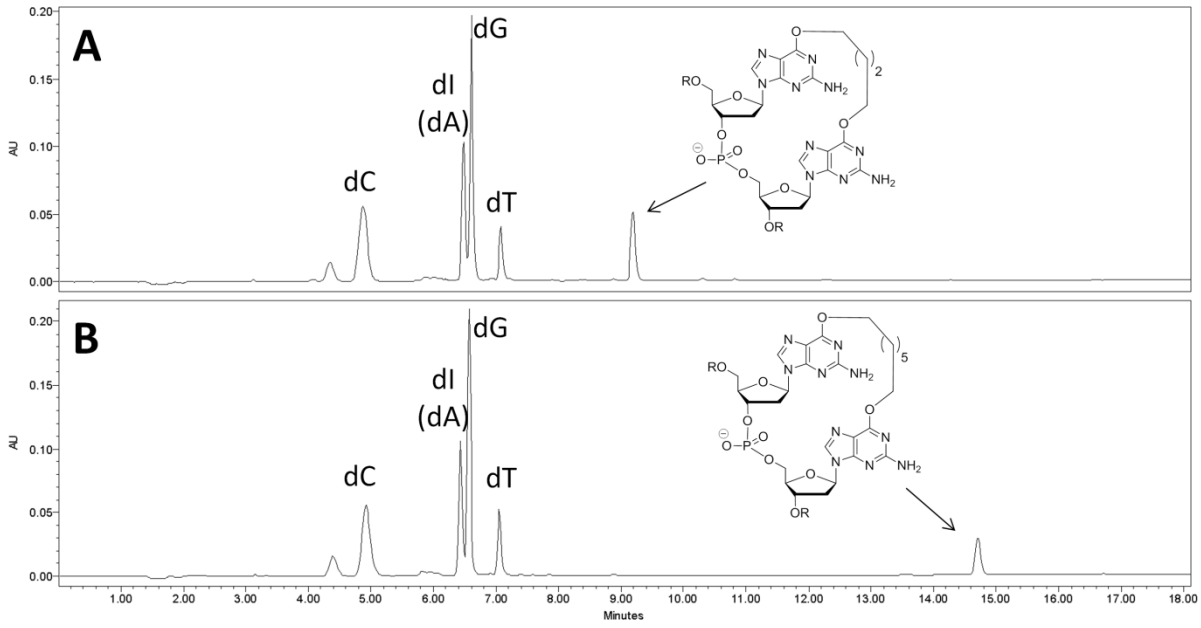


Figure S4.1 -IEX-HPLC profile of **GpG4** (**A**) and **GpG7** (**B**). The column was eluted using a linear gradient of 0-52% buffer B over 24 min (buffer A: 100 mM Tris HCl, pH 7.5, 10% MeCN and buffer B: 100mM Tris HCl, pH 7.5, 10% MeCN, 1 M NaCl) at 55 ^{o}C and flow rate of 1 mL/min. Monitoring was performed at 260 nm.



Oligomer	*Nucleoside composition*	*Nucleoside Ratio*	
		Expected	*Observed*
GpG4	dC	4	4.0
	dI (dA)	3	2.8
	dG	4	3.6
	dT	2	1.1
	dG-dG 4C	1	1.0
GpG7	dC	4	4.0
	dI (dA)	3	2.9
	dG	4	4.0
	dT	2	1.4
	dG-dG 7C	1	1.2

Figure S4.2- C-18 HPLC profile of nuclease digested **GpG4** (**A**) and **GpG7** (**B**). The column was eluted with a linear gradient of 0-60% buffer B over 30 min (buffer A: 50 mM sodium phosphate, pH 5.8 and buffer B: 50 mM sodium phosphate, pH 5.8, 50% MeCN) at a flow rate of 1.0 mL/min over 30 min, monitored at 260 nm. The appearance of a new peak which does not correlate with elution profiles of corresponding dimers lacking the phosphodiester group, suggests incomplete digestion of **GpG4** and **GpG7**.[323] Nucleotide ratios are reported below the figure.

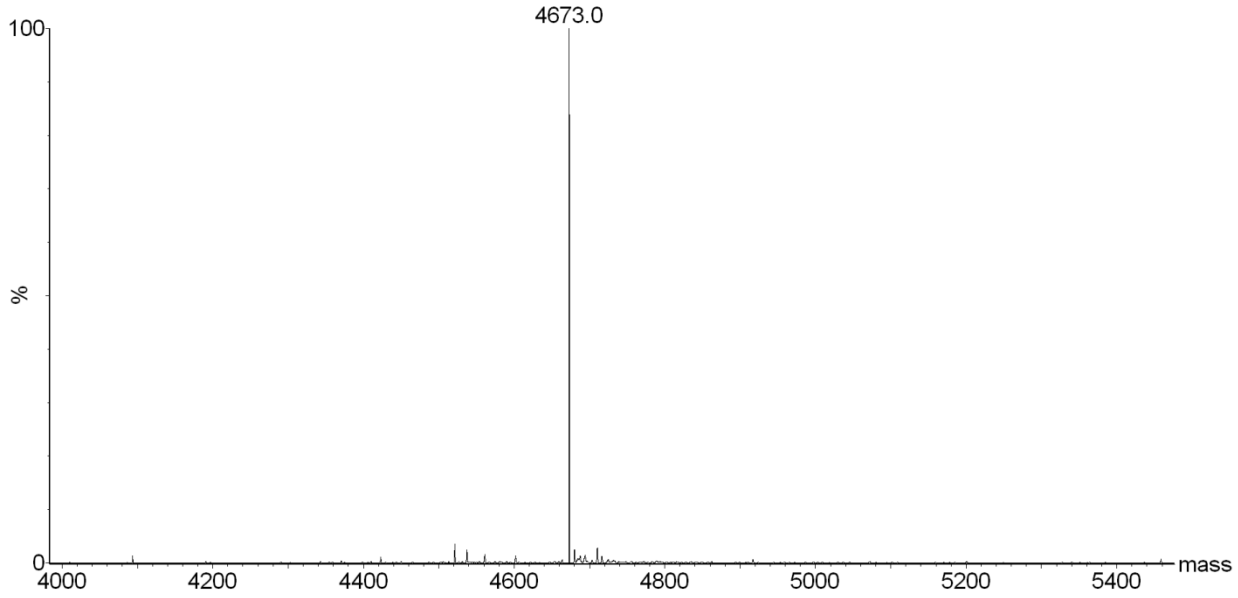


Figure S4.3 - Deconvoluted ESI MS spectrum of oligonucleotide **GpG4**

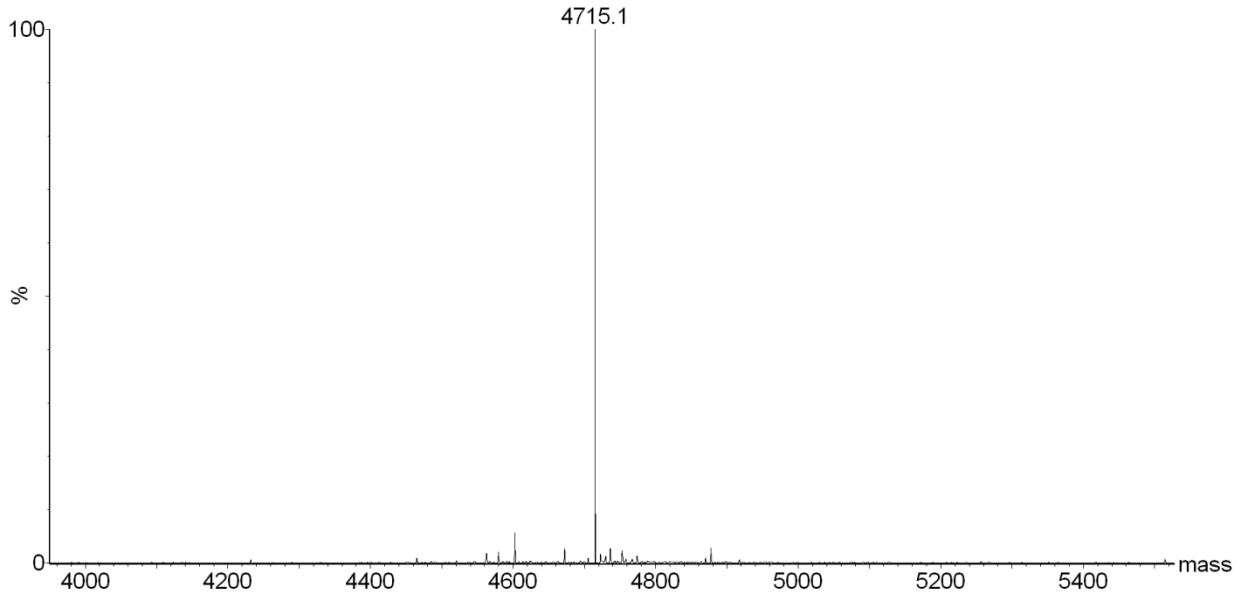


Figure S4.4 - Deconvoluted ESI MS spectrum of oligonucleotide **GpG7**

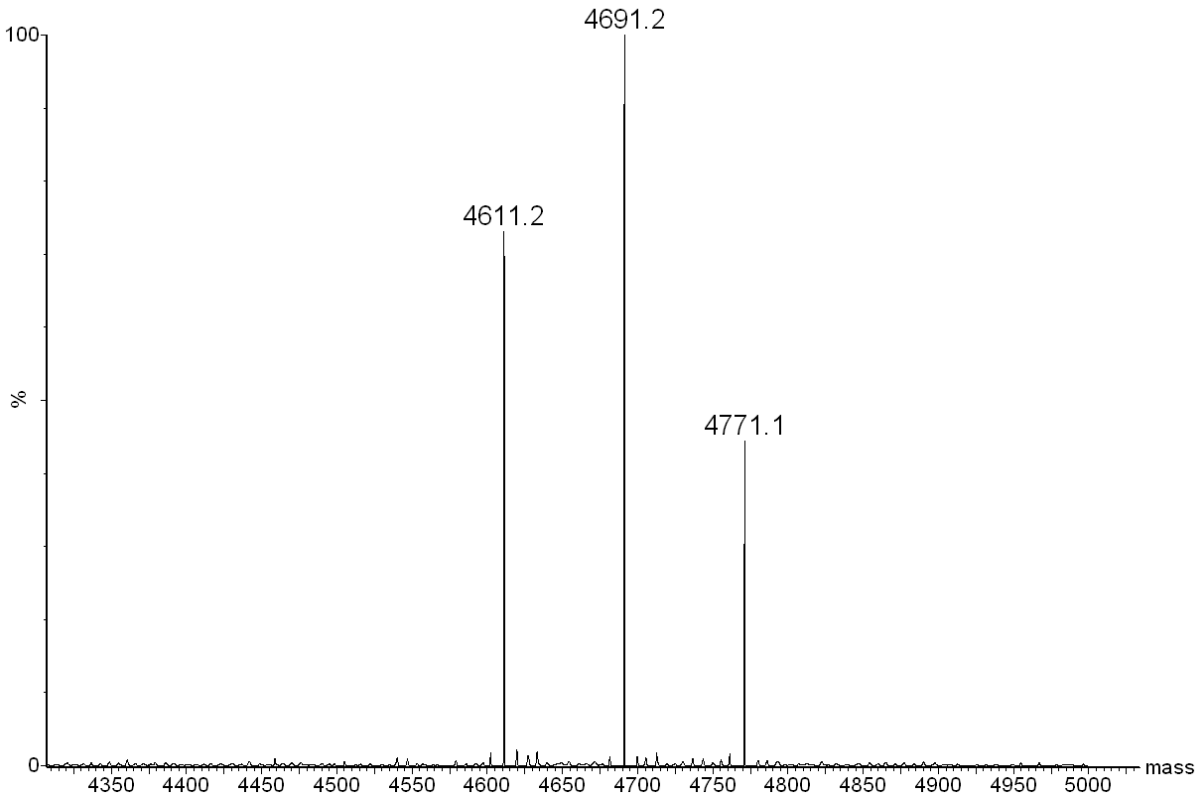


Figure S4.5 - Deconvoluted ESI MS spectrum of oligonucleotide **GG4** (400 pmol) phosphorylated using 20 units of T4 PNK in 1X PNK buffer with 3 molar equivalence of ATP.

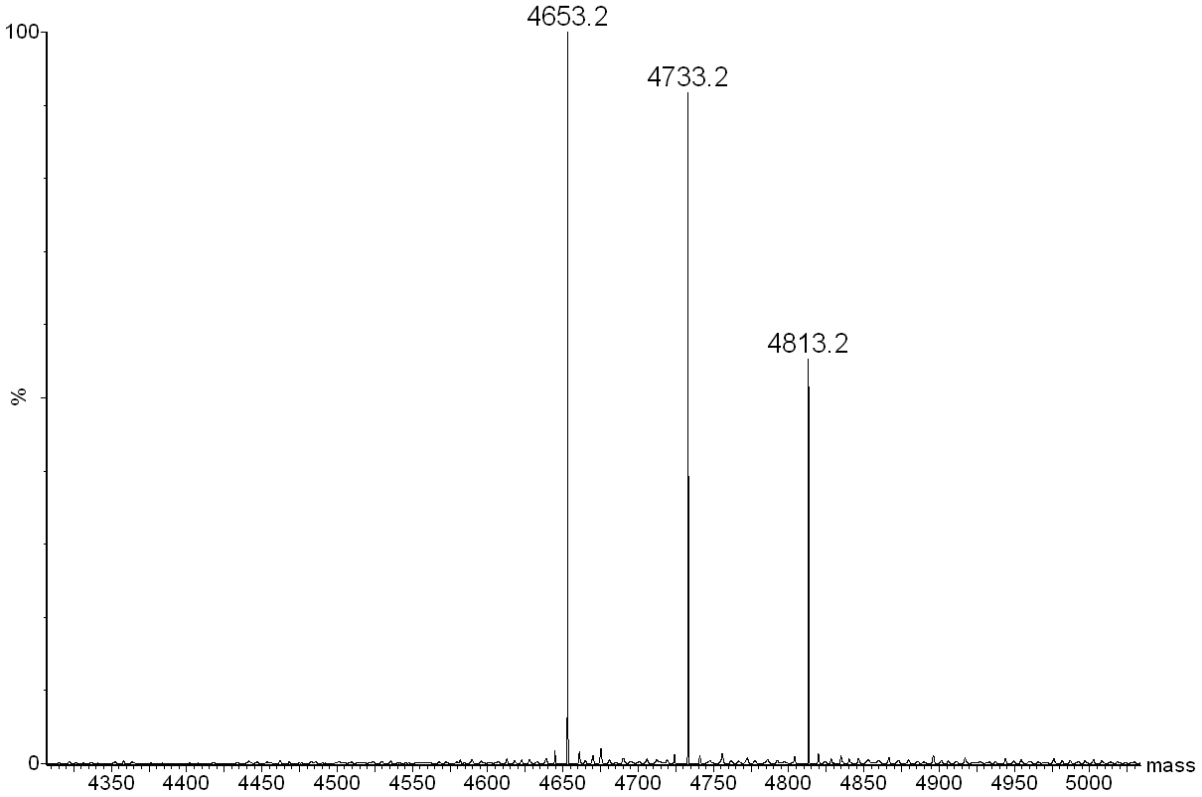


Figure S4.6 - Deconvoluted ESI MS spectrum of oligonucleotide **GG7** (400 pmol) phosphorylated using 20 units of T4 PNK in 1X PNK buffer with 3 molar equivalence of ATP.

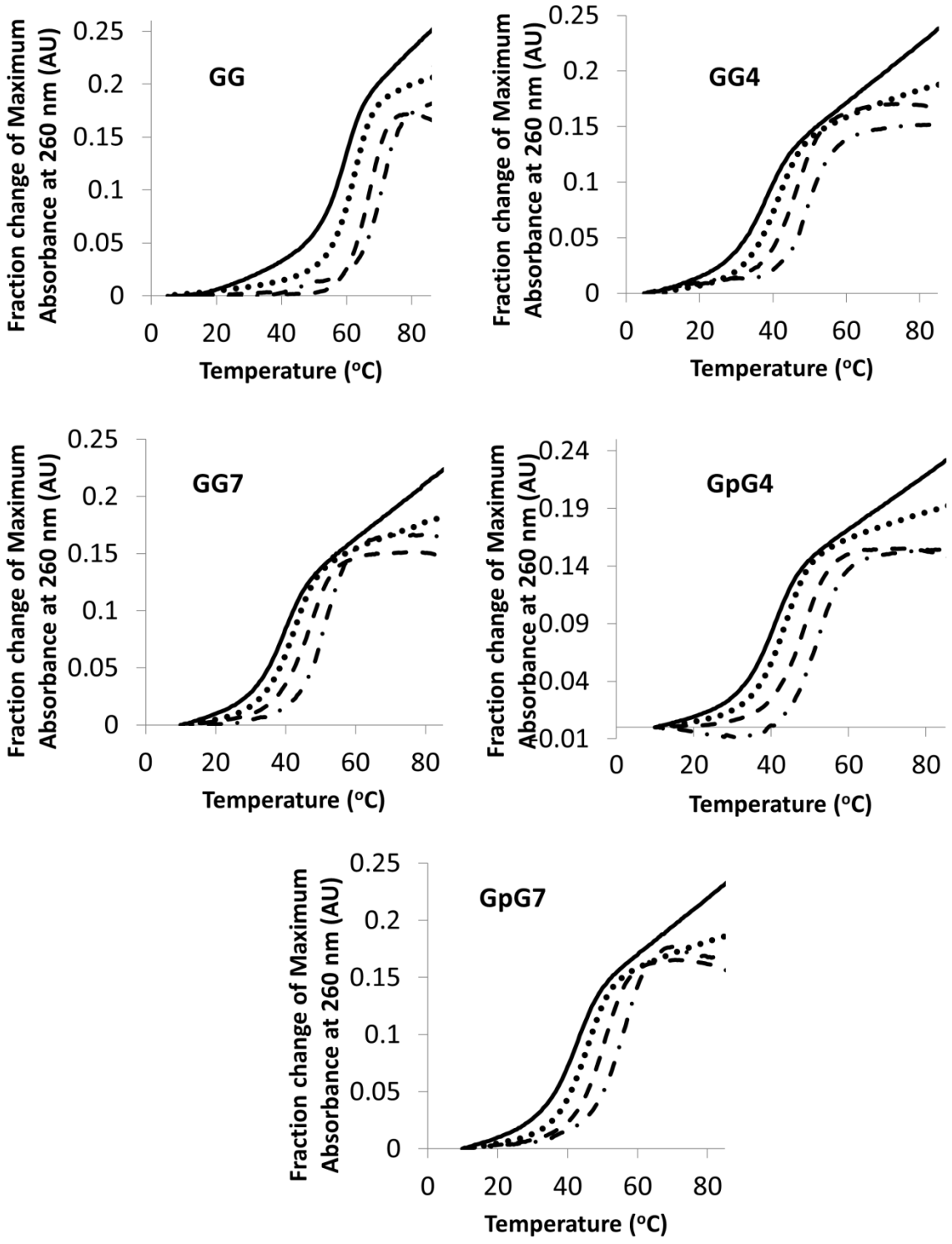


Figure S4.7 - UV thermal denaturation experiments for Van't Hoff experiments for unmodified DNA duplex (**GG**), **GG4**, **GG7**, **GpG4**, and **GG7** at 0.69 μM (—), 2.8 μM (•••), 14 μM (– – –), 69 μM (–•–). T_m values are reported in **Table S4.1** and all plots of T_m^{-1} versus $\ln(C_t/4)$ exceeded correlation values of 0.95.

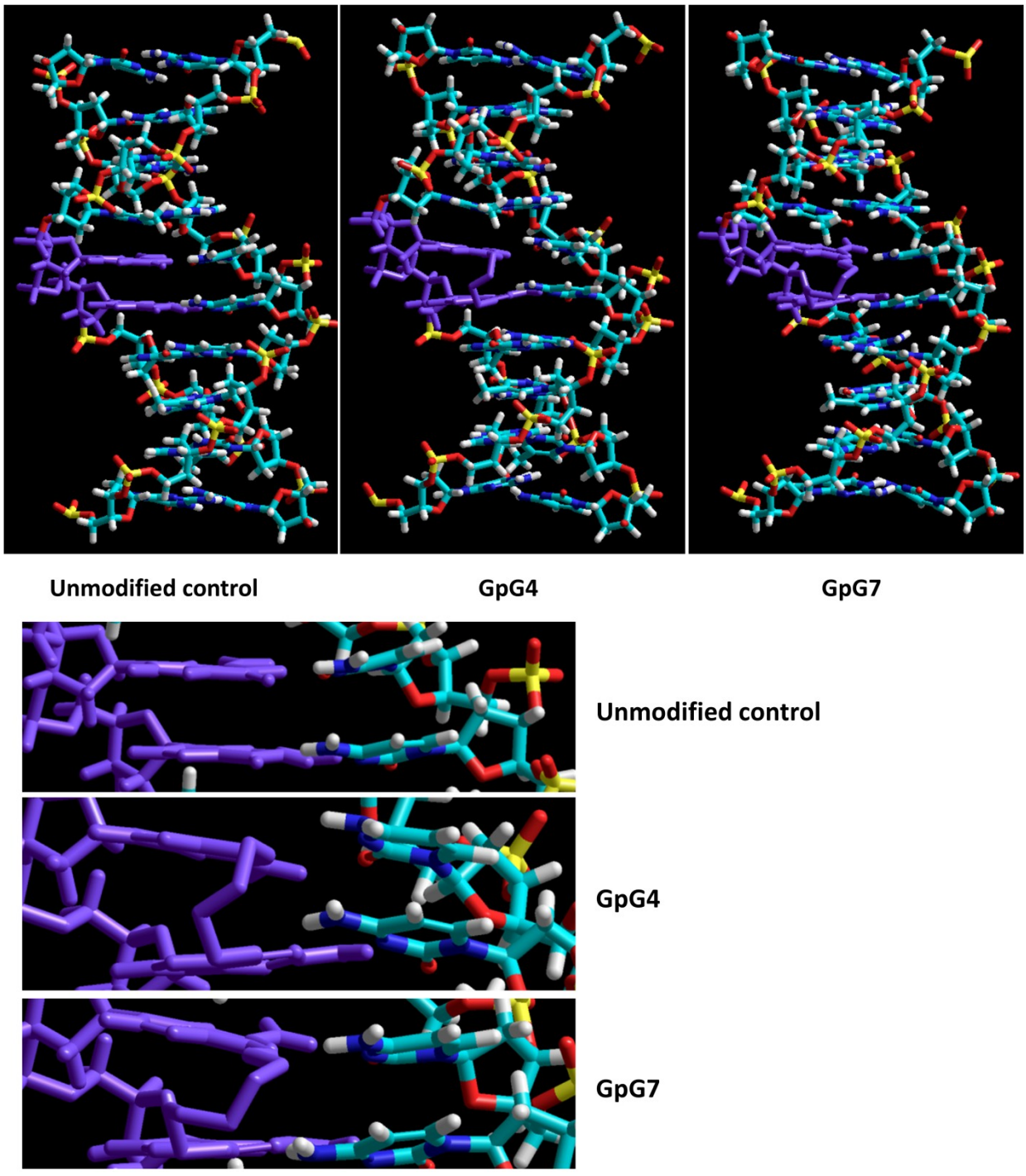


Figure S4.8 - Molecular models of unmodified control duplex and duplexes containing **GpG4** and **GpG7** that were geometry optimized using the AMBER forcefield.

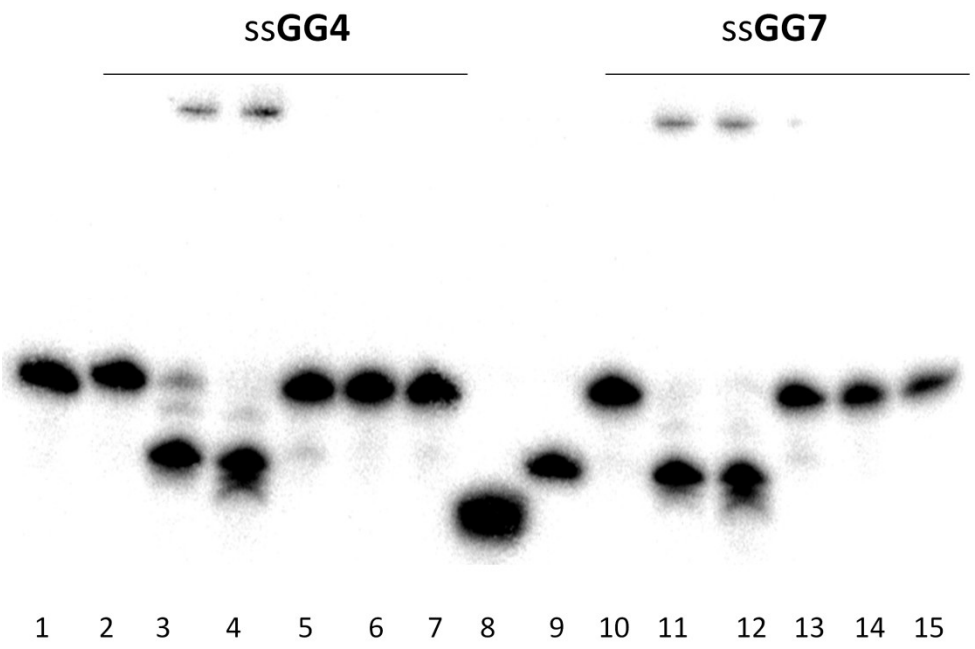


Figure S4.9 - Total repair gel of single stranded **GG4** and **GG7** by AGTs. Lane 1, 2 pmol control unmodified DNA; lane 2, 2 pmol **GG4**; lane 3, 2 pmol **GG4** +10 pmol hAGT; lane 4, 2 pmol **GG4** +60 pmol hAGT; lane 5, 2 pmol **GG4** +10 pmol OGT; lane 6, 2 pmol **GG4** +10 pmol Ada-C; lane 7, 2 pmol **GG4** +10 pmol OGT S134P; lane 8, 2 pmol 5'-pGGC TG; lane 9, 2 pmol 5'- pG GAT CAC CAG, lane 10, 2 pmol **GG7**; lane 11, 2 pmol **GG7** +10 pmol hAGT; lane 12, 2 pmol **GG7** +60 pmol hAGT; lane 13, 2 pmol **GG7** +10 pmol OGT; lane 14, 2 pmol **GG7** +10 pmol Ada-C; lane 15, 2 pmol **GG7** +10 pmol OGT S134P.

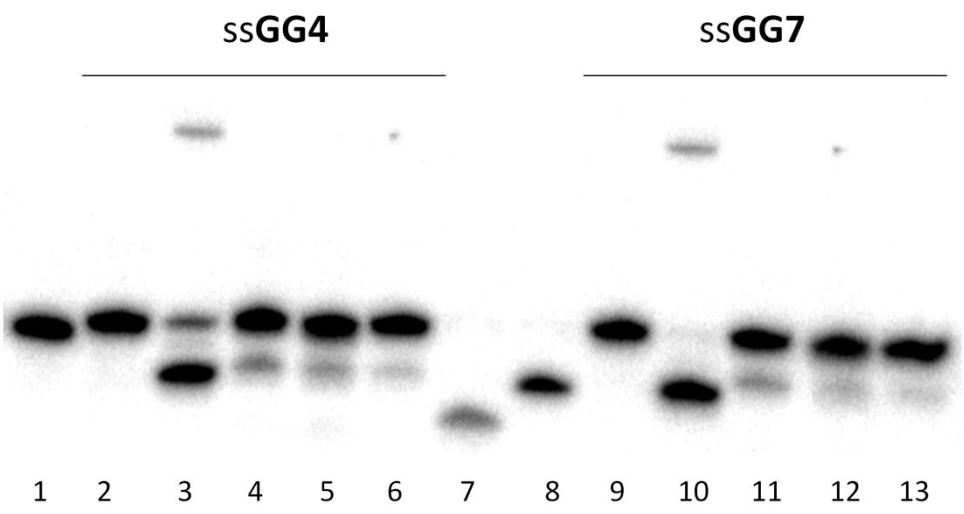


Figure S4.10 - Total repair gel of single stranded **GG4** and **GG7** by AGTs. Lane 1, 2 pmol control unmodified DNA; lane 2, 2 pmol **GG4**; lane 3, 2 pmol **GG4** +10 pmol hAGT; lane 4, 2 pmol **GG4** +60 pmol OGT; lane 5, 2 pmol **GG4** + 60 pmol Ada-C; lane 6, 2 pmol **GG4** + 60 pmol OGT S134P; lane 7, 1 pmol 5'-GGC TG; lane 8, 2 pmol 5'- G GAT CAC CAG, lane 9, 2 pmol **GG7**; lane 10, 2 pmol **GG7** +10 pmol hAGT; lane 11, 2 pmol **GG7** +60 pmol OGT; lane 12, 2 pmol **GG7** + 60 pmol Ada-C; lane 13, 2 pmol **GG7** + 60 pmol OGT S134P.

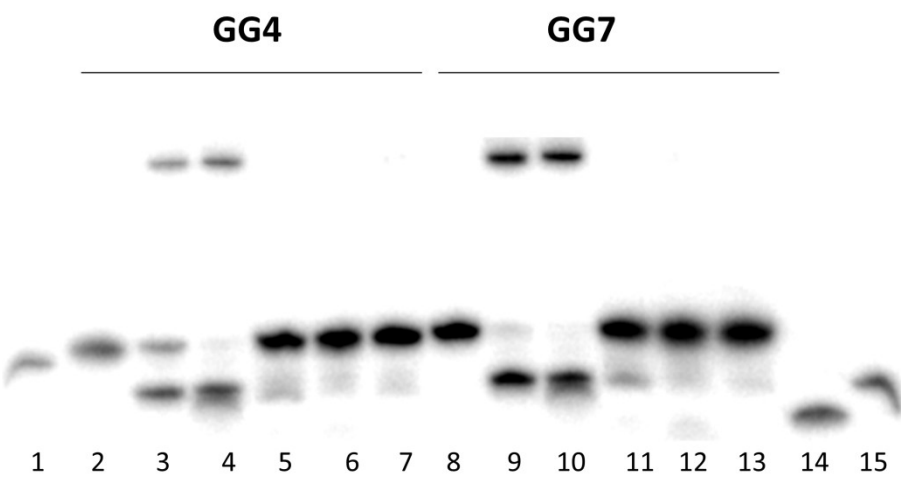


Figure S4.11 - Total repair gel of double stranded **GG4** and **GG7**, labelled with a mixture of γ-[^{32}P]ATP (1 uL, 10 μCi) and non-radioactive ATP (3-fold excess), by AGTs. Lane 1, 2 pmol control unmodified DNA; lane 2, 2 pmol **GG4**; lane 3, 2 pmol **GG4** +10 pmol hAGT; lane 4, 2 pmol **GG4** +60 pmol hAGT; lane 5, 2 pmol **GG4** + 60 pmol OGT; lane 6, 2 pmol **GG4** + 60 pmol Ada-C; lane 7, 2 pmol **GG4** + 60 pmol OGT S134P; lane 8, 2 pmol **GG7**; lane 9, 2 pmol **GG7** +10 pmol hAGT; lane 10, 2 pmol **GG7** + 60 pmol hAGT; lane 11, 2 pmol **GG7** + 60 pmol OGT; lane 12, 2 pmol **GG7** + 60 pmol Ada-C; lane 13, 2 pmol **GG7** + 60 pmol OGT S134P; lane 14, 2 pmol 5'-GGC TG; lane 15, 2 pmol 5'- G GAT CAC CAG

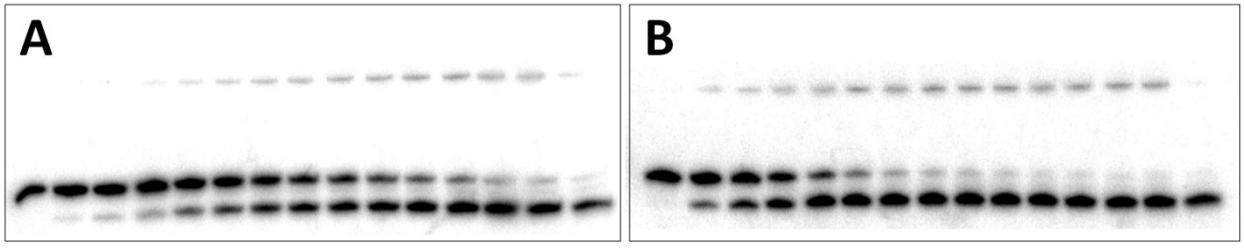


Figure S4.12 - Time course repair gel of single stranded **GG4** and **GG7** by hAGT. (**A**) Denaturing gel of the repair of 2 pmol of **GG4** by 10 pmol hAGT as a function of time: lane 1, 2 pmol Control; lanes 2-15, 2 pmol **GG4** + 10 pmol hAGT incubated for 1, 2.5, 5, 10, 15, 25, 35, 45, 60, 90, 120, 180, 240, 1020 min, respectively. (**B**) Denaturing gel of the repair of 2 pmol of **GG7** by 10 pmol hAGT as a function of time: lane 1, 2 pmol Control; lanes 2-15, 2 pmol **GG7** + 10 pmol hAGT incubated for 1, 2.5, 5, 10, 15, 25, 35, 45, 60, 90, 120, 180, 240 and 480 min, respectively.

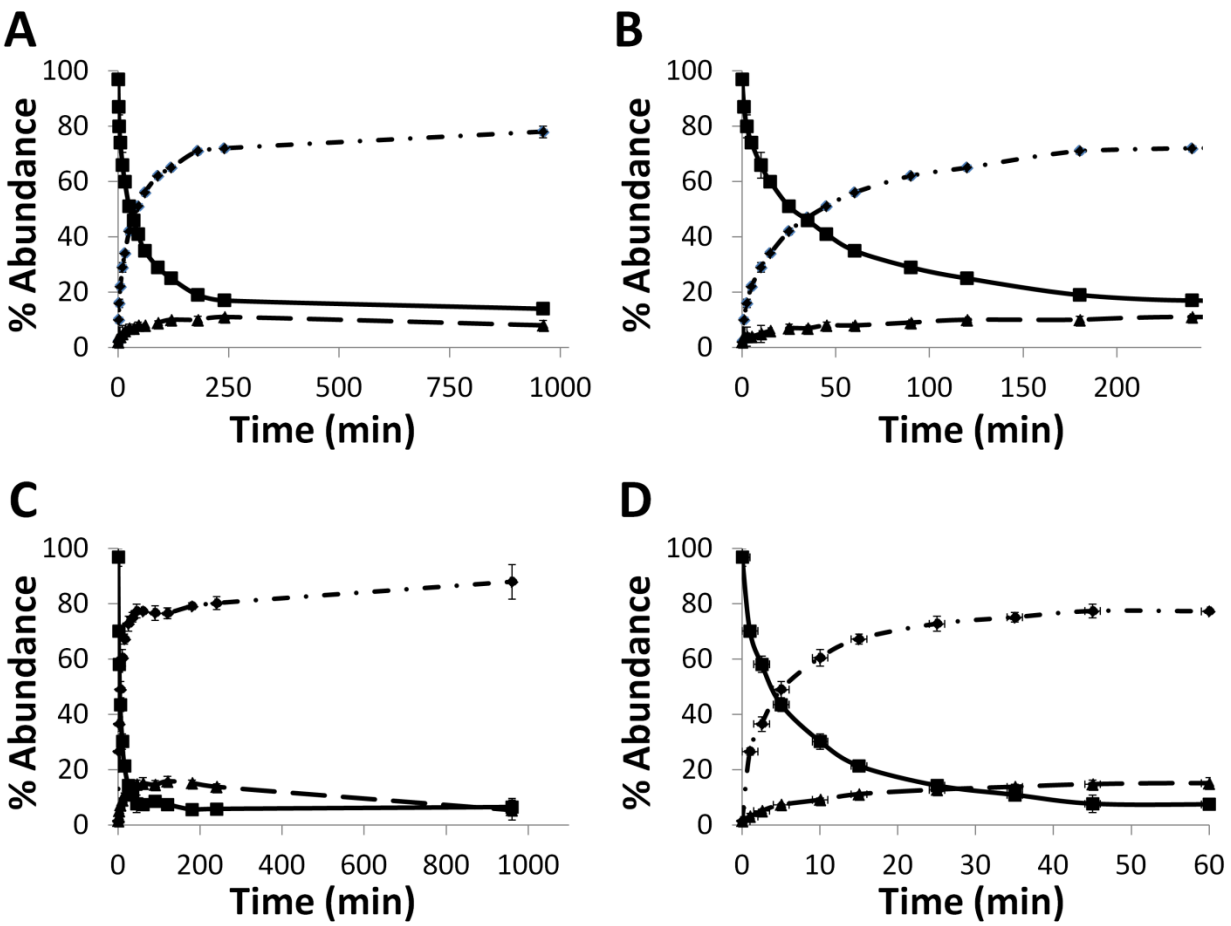


Figure S4.13 - Time course repair assay plots of **GG4** (**A** and **B**) and **GG7** (**C** and **D**) by hAGT displaying significant faster repair of **GG7** as previously observed for repair of corresponding duplexes, respectively. Graphs on the left show the full length assay (960 min) whereas those on the right are truncated. Graphical illustrations display abundances [%] of MP (- - -), RP (—•—) and substrate (—) over time (min) obtained using ImageJ software (National Institutes of Health).

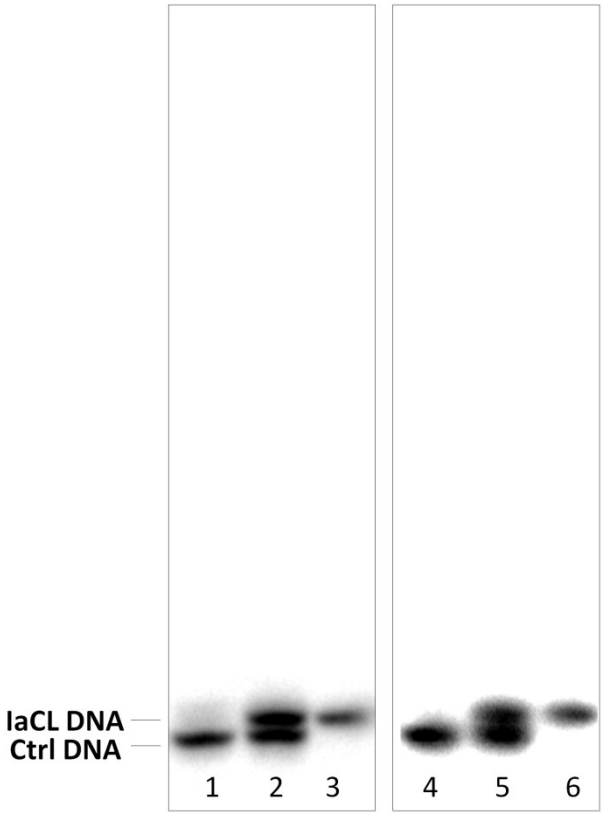


Figure S4.14 - Control experiment displaying feasibility of denaturing gels in resolving repair products and substrates: lane 1, 2 pmol control (5'-GGC TGG GAT CAC CAG) ; lane 2, 1 pmol control + 1 pmol **GpG4**; lane 3, 2 pmol **GpG4**; lane 4, 2 pmol control; lane 5, 1 pmol control + 1 pmol **GpG7**; lane 6, 2 pmol **GpG7**.

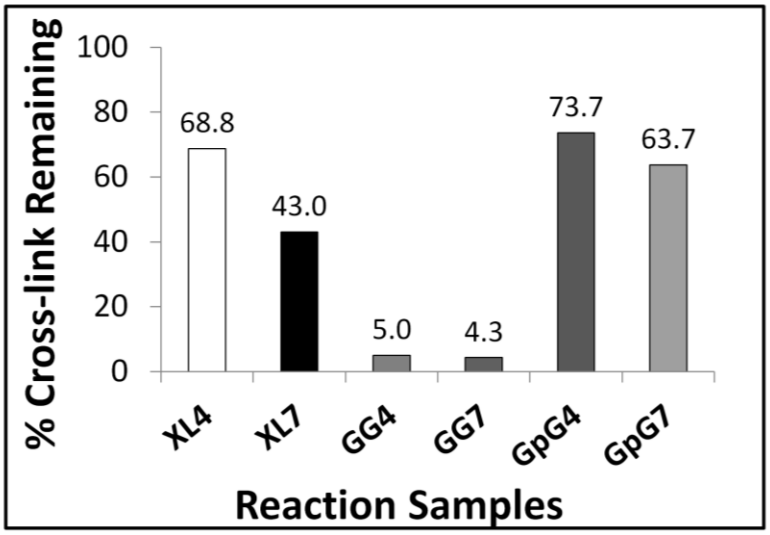


Figure S4.15 - Graphical representation of the % IaCL (2 pmol) of **GG4**, **GG7**, **GpG4** and **GpG7** remaining in the reaction tube after treatment with hAGT (30-fold protein, 60 pmol) for 16h at 37 ^{o}C obtained using ImageJ software (National Institutes of Health).

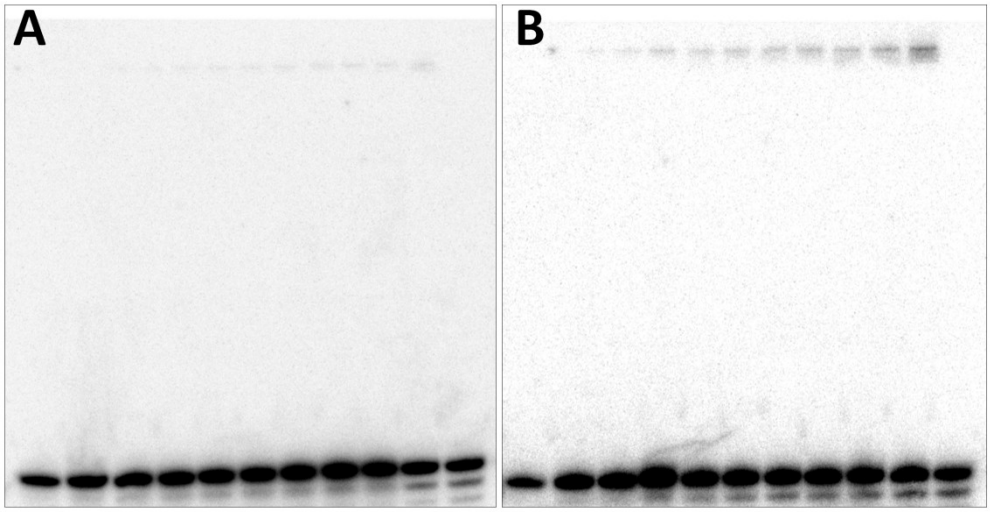


Figure S4.16 - Time course repair gel of **GpG4** and **GpG7** by hAGT. (**A**) Denaturing gel of the repair of 2 pmol of **GpG4** by 60 pmol hAGT as a function of time: lane 1, 2 pmol Control; lanes 2-11, 2 pmol **GpG4** + 60 pmol hAGT incubated for 5, 15, 35, 60, 90, 120, 180, 240, 300 and 960 min, respectively (**B**) Denaturing gel of the repair of 2 pmol of **GpG7** by 60 pmol hAGT as a function of time: lane 1, 2 pmol Control; lanes 2-11, 2 pmol **GpG7** + 60 pmol hAGT incubated for 5, 15, 35, 60, 90, 120, 180, 240, 300 and 960 min, respectively.

Figure S4.17 - Total repair gel of duplexes containing bicyclic pyrimidine adducts (**DFP** and **TPP**, structures shown below). The preparation of DNA oligomers containing these modifications have been described in **CHAPTER II** (5`-GGC TX GAT CAC CAG, where X denotes the site of modification). Lane 1, 2 pmol **DFP**; lane 2, 2 pmol **DFP** + 10 pmol hAGT; lane 3, 2 pmol **DFP** + 10 pmol OGT; lane 4, 2 pmol **DFP** + 10 pmol Ada-C; Lane 5, 2 pmol **TPP**; lane 6, 2 pmol **TPP** + 10 pmol hAGT; lane 7, 2 pmol **TPP** + 10 pmol OGT; lane 8, 2 pmol **TPP** + 10 pmol Ada-C; lane 9, 2 pmol unmodified control DNA.

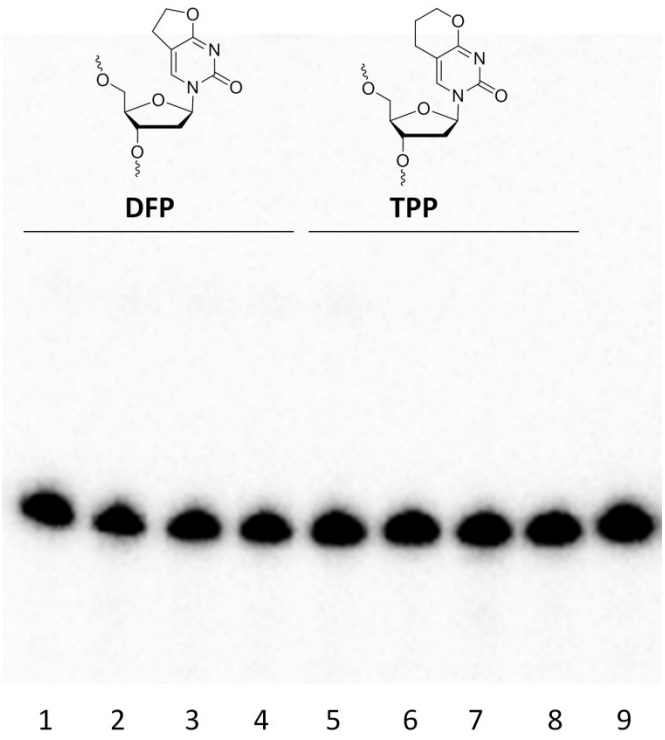


Table S4.1 - T_m values (^oC) with respect to DNA concentration (μM). All duplexes were resuspended in 90mM sodium chloride, 10 mM sodium phosphate, 1 mM EDTA buffer (pH 7.0). Annealing curves were acquired at 260 nm starting at 95 ^oC and decreasing temperature at a rate of cooling of 0.5 ^oC min^{-1} until 15 ^oC. Denaturation curves were acquired in a similar fashion (15 ^oC – 95 ^oC) followed by another round of annealing experiment (95 ^oC – 15 ^oC)

[DNA] (μM)	T_m values (^oC) in triplicates														
	GG			**GG4**			**GG7**			**GpG4**			**GpG7**		
0.69	59.0	60.0	61.0	37.0	38.1	38.0	39.2	40.1	41.0	42.0	43.0	41.0	43.2	44.0	43.0
2.8	62.0	62.6	64.3	41.0	41.5	41.9	42.0	43.5	43.0	44.1	45.5	43.9	46.1	47.4	46.9
14	68.0	67.9	69.0	45.4	45.6	46.0	47.5	47.8	48.5	48.9	50.0	49.2	51.0	52.3	52.9
69	73.0	70.7	71.0	48.7	49.0	50.0	50.0	49.9	51.4	52.0	53.2	52.9	56.1	55.8	55.9

CHAPTER V

Backbone Flexibility Influences Nucleotide Incorporation by Human Translesion DNA Polymerase *η* Opposite Intrastrand Cross-linked DNA

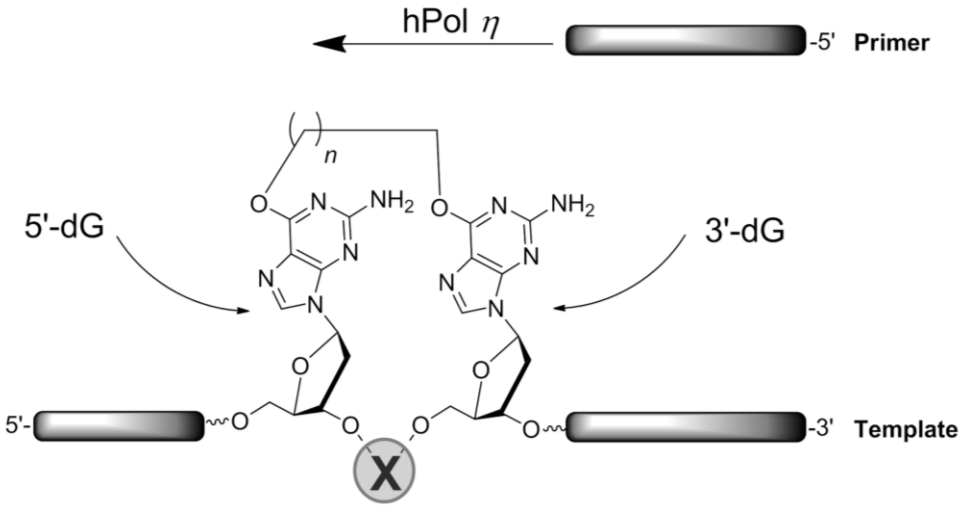


Table of contents Figure (Figure 5.2)

Published as:

Adapted with permission from O'Flaherty DK[1], Guengerich FP[2], Egli M[2*] and Wilds CJ[1*].

(2015) *Biochemistry 54*, 7449-7456.

[1]Department of Chemistry and Biochemistry, Concordia University, Montréal, Québec, Canada, H4B1R6.

[2]Department of Biochemistry, Vanderbilt Institute of Chemical Biology, Vanderbilt Ingram Cancer Center, Center in Molecular Toxicology, and Center for Structural Biology, School of Medicine, Vanderbilt University, Nashville, Tennessee 37232, United States

*corresponding authors.

Abstract

Intrastrand cross-links (IaCL) connecting two purine nucleobases in DNA pose a challenge to high fidelity replication in the cell. Various repair pathways or polymerase bypass can cope with these lesions. The influence of the phosphodiester linkage between two neighbouring 2′-deoxyguanosine (dG) residues attached through the O^6-atoms by an alkylene linker on bypass with human DNA polymerase η (hPol η) was explored *in vitro*. Steady-state kinetics and mass spectrometric analysis of products from nucleotide incorporation revealed that although hPol η is capable of bypassing the 3′-dG in a mostly error-free fashion, significant misinsertion was observed for the 5′-dG of the IaCL containing a butylene or heptylene linker. The lack of the phosphodiester linkage triggered an important increase in frameshift adduct formation across the 5′-dG by hPol η, in comparison to the 5′-dG of IaCL DNA containing the phosphodiester group.

5.1 Introduction

Chemotherapeutic agents such as the platinum-containing drugs used in the treatment of cancer exert their therapeutic effect mainly *via* the formation of cytotoxic DNA damage. The lesions that these agents produce have been identified primarily as intrastrand cross-links (IaCL) between the N7-atoms of purines with the distribution of these IaCL determined to be 65% 1,2(GpG), 25% 1,2(ApG), and 5-10% 1,3(GpTpG).[281,332–338] In addition, formation of minor products including interstrand cross-links (ICL), mono-adducts and DNA-protein cross-links occur.[281] The presence of these adducts on the DNA scaffold impedes vital cellular processes such as DNA replication and transcription ultimately leading to cell death. Drugs used in cancer regimens, other than platinum-containing agents, such as mechlorethamine,[72,285] mitomycin C[70,284] and busulfan[69] have also been shown to introduce IaCL in DNA, in particular between adjacent purine nucleobases. Using drugs that act directly on DNA to treat cancer have intrinsic and acquired drug resistance as a major limitation, which is mediated by the cellular response processes of DNA repair and translesion DNA synthesis (TLS).

The four TLS DNA polymerases identified in humans are Pol η, Pol κ, Pol ι, and Rev1. The most widely studied of these is Pol η given its crucial involvement in bypassing UV-induced intrastrand cross-linked DNA lesions. Disruption in the proper function of the *POLH* gene leads to xeroderma pigmentosum variant (XPV), a disease characterized by hypersensitivity to UV-radiation and an increased incidence of skin cancer.[207]As suspected, *POLH* knockout mice demonstrated heightened incidences of skin cancer compared to the control group when exposed to UV-radiation.[339] XPV cell extracts displayed replication inhibition of plasmid DNA containing a single (6-4) pyrimidone photoproduct lesion.[82] Moreover, human cells deficient in Pol η revealed greater cell death events when treated with platinum-based chemotherapeutic

agents.[71,226,230,340] Exposure of DNA to γ-radiation leads to the formation of a mixture of the IaCL lesions G[8,5]C and G[8,5]T, among others, formed *via* a radical mechanism.[4] Their bypass by yeast and/or human Pol *η* demonstrated reduced fidelity and processivity, in particular across the dG portion of the lesion.[78,223,341] Accounts of Pol *η* bypass are numerous and the search for other biologically relevant DNA damage, or mimics thereof, is ongoing.

DNA alkylating agents such as *N*-nitroso-*N*-methylurea readily modify the N7-atom of dG and, to a lesser extent, the O^6-position. Lesions at the O^6-atom of dG have also been detected after exposure to the methylating and chloroethylating chemotherapeutic drugs temozolimide and carmustine. Endogenous formation of O^6-MedG by *S*-adenosylmethionine is estimated at 10–30 damage events per cell per day.[14] The introduction of covalent appendages at the N7-atom of a purine nucleotide in DNA reduces chemical stability, leading to modifications such as the introduction of abasic sites and/or formation of formamidopyrimidine lesions. Alkyl modifications introduced at the O^6-position of dG are chemically stable and if left uncorrected, lead to stalls or significant G:C to A:T transitions by DNA polymerases including those of the Y-family.[45,342–346] The importance of this site of alkylation has been demonstrated by *in vivo* studies which revealed the direct link between O^6-MedG and carcinogenesis.[39] The disruption of high fidelity DNA polymerase activity in the presence of O^6-MedG and other O^6-alkyldG lesions has been attributed to disruption of Watson-Crick base pairing.

Much less is known about the bypass of IaCL containing more flexible lesions such as an alkylene linker, in comparison to more rigid systems including platinum-based IaCL DNA. In the present study, we investigated whether Pol *η* was capable of efficiently bypassing an O^6-dG-alkylene-O^6-dG IaCL containing butylene or heptylene tethers (**Figure 5.1A**). Our phosphoramidite synthetic strategy used to generate such IaCL DNA probes (**CHAPTER III**

and **IV**) allowed us to engineer identical sets of adducted DNA lacking the phosphodiester linkage at the cross-linked site (**Figure 5.1B**), which were also evaluated.[323] The absence of the phosphodiester linkage presumably confers increased flexibility to the IaCL DNA. This study set out to (i) investigate the ability of hPol η to bypass a malleable IaCL lesion that can disrupt the fidelity of Watson-Crick base pairing (at the O^6-position of dG as opposed to the N7-position). The second objective was to (ii) investigate the effect of IaCL linker length (butylene vs. heptylene) on the processivity of hPol η. The chain lengths of the IaCL linkers were inspired from DNA adducts formed from exposure to the alkyl sulfones busulfan and hepsulfam, known to generate butylene and heptylene DNA adducts, respectively.[69,324] It should be noted, however, that a specific IaCL connected at O^6-atoms of 5'-d(GG) sequences has not been identified to date. Our model IaCL DNA serve as chemically stable probes that can be prepared in scales and purity amenable for biochemical studies. The final objective of the study was to (iii) investigate the effect of phosphodiester linkage deletion on IaCL DNA bypass by hPol η.

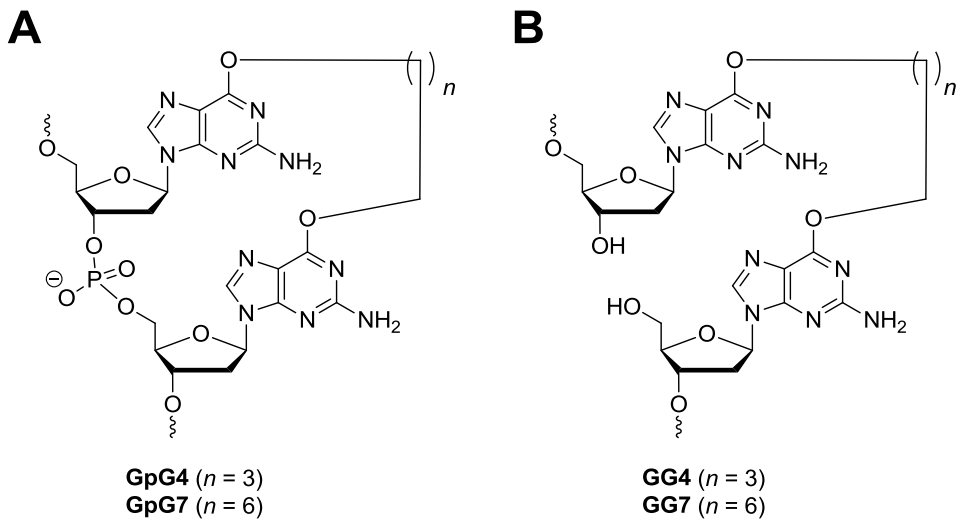


Figure 5.1. Structures of the O^6-dG-alkylene-O^6-dG IaCL containing (**A**) and lacking (**B**) the phosphodiester linkage between the O^6-linked nucleotides.

5.2 Materials and Methods

5.2.1 Steady-state Kinetics

All primer extension assays were carried out using a template strand with the sequence 5′- AC **XX** CT CAC ACT (where **XX** denotes the cross linked dGG residues or dGG for the unmodified control) and a fluorescently labelled primer 5′-(FAM)TAG TG*U* GAG (where FAM is 6-carboxyfluorescein and *U* is 2′-deoxyuridine). Steady-state kinetic experiments were conducted as previously described.[229,236,347,348] Briefly, assays were generally performed at 37 °C in 40 mM Tris-HCl buffer (pH 7.5) containing 100 mM KCl, 5% glycerol (v/v), 10 mM dithiothreitol (DTT), 5 mM $MgCl_2$ and 100 μg mL^{-1} bovine serum albumin (BSA). The 5′-fluorescently labelled (FAM) primer-template (9-/13-mer) duplex (5 μM) was extended using 1.9 to 500 nM concentrations of hPol η in the presence of various concentrations of a single dNTP (0 to 1 mM, at 7-10 different dNTP concentrations) at 37 °C for 5-20 min. Reactions were quenched using a solution containing 20 mM EDTA (pH 8.0), 95% formamide (v/v), bromphenol blue, and xylene cyanol dyes. Substrates and products were resolved on 18% (w/v) polyacrylamide electrophoresis gels containing 7.5 M urea. Gels were monitored by a Typhoon Scanner (GE Healthcare) and analyzed by fluorescence intensity using ImageJ software (National Institutes of Health). The values of k_{cat} and K_m were estimated by non-linear regression analysis (hyperbolic fit) with the program Graphpad Prism (La Jolla, CA).

5.2.2 LC-MS/MS Analysis of Fully Extended Products

All primer extension assays were carried out using identical sequences as described in the above section. DNA Primers were extended in the presence of all four dNTP followed by analysis *via* mass spectrometry. Primer sequences were specifically designed to contain a 2′-deoxyuridine (*U*) insert in order to produce shorter oligonucleotide fragments that are formed by

treatment with uracil DNA glycosylase followed by hot piperidine). These were subsequently analyzed by an LC-MS/MS method (ion-trap mass spectrometer), as previously described.[229,236,256–258] DNA primer extension was accomplished by combining hPol η (95 pmol, 0.95 μM for unmodified duplexes and 340 pmol, 3.4 μM for IaCL-containing duplexes) with template-primer duplex (2 nmol, 20 μM) and a mixture of 1 mM each of dATP, dCTP, dGTP, and dTTP at 37°C for 0.5-1.5h in 50 mM Tris-HCl buffer (pH 7.5), 50 mM NaCl, 5 mM DTT, 5 mM $MgCl_2$ and 50 μg/ml bovine serum albumin (BSA). The reactions were terminated by spin column separations (Micro Bio-Spin™ 6 Columns from BIO-RAD) to extract the dNTPs and Mg^{2+}. The extent of the extension was monitored by electrophoresis/fluorography prior to LC-MS/MS analysis (**Figure S5.6**). The extension products were treated with 25 units of uracil DNA glycosylase followed by 0.25 M piperidine, and subjected to LC-MS/MS analysis using an Acquity UPLC system (Waters) interfaced to a Thermo-Finnigan LTQ mass spectrometer (Thermo Scientific, San Jose, CA) equipped with a negative ion electrospray source.[229,236,256–258] Chromatographic separation was carried out with an Acquity UPLC BEH octadecylsilane (C18) column (2.1 x 100 mm, 1.7 μm). The LC solvent system was as follows: Mobile phase A, 10 mM $CH_3CO_2NH_4$ in 98% H_2O; mobile phase B, 10 mM $CH_3CO_2NH_4$ in 90% MeCN (v/v). The following gradient (v/v) was used with a flow rate of 300 μL min^{-1} at a temperature of 50 ^oC: Linear gradient from 0-3% B (v/v) in 3 min, followed by a linear increase to 20% B (v/v) from 3-5 min, then 20-100% B (v/v) from 5-6 min which was held for 2 min. The column was re-equilibrated for 3 min with 0 % B (v/v) for subsequent analysis. Mass spectrometry conditions were as follows: Source voltage, 4 kV; source current 100 μA; capillary voltage,- 49 V; capillary temperature, 350 °C; tube lens voltage, -90 V. Product ion spectra were recorded over the range *m/z* 300-2000 and the most abundant species (-2 charge) was used for collision-induced

dissociation (CID) analysis. The calculation for the oligonucleotide sequence CID fragmentation was carried out using Mongo Oligo Mass Calculator v2.06 from The RNA Institute (College of Arts and Science, University at Albany State University of New York). The relative yields of various products were calculated based on the peak areas of extracted ion chromatograms from LC-MS analyses. The sum of the peak areas was used for multi-charged species.

5.3 Results and Discussion

Bypass studies with hPol η have revealed its ability to process a wide variety of DNA modified at the nucleobase including O^6-alkyl-dG adducts,[45] 1,N^2-etheno-2′-deoxyguanosine;[349] and 8-oxo-7,8-dihydro-2′- deoxyguanosine.[229] Bulkier forms of DNA damage including intrastrand lesions induced by cisplatin,[227,350] UV radiation,[207,209] and the dG[8,5-Me]dT IaCL,[223] have also been the subject of bypass studies with hPol η (structures of selected examples are illustrated in **Figure S5.1**). Bis-alkylating drugs such as busulfan can introduce IaCL that may have increased flexibility at the modified site relative to the bulkier lesions described above. Given the versatility of hPol η to bypass a variety of lesions by TLS, we decided to explore the influence of flexibility at the modified site with O^6-dG-alkylene-O^6-dG IaCL containing either a butylene or heptylene linkage between the bases. To further probe the influence of flexibility, template strands lacking and containing a phosphodiester linkage between the linked nucleobases have been prepared (see **Figures S5.2-S5.5** for ESI-MS of the template strands). DNA duplexes containing an O^6-dG-alkylene-O^6-dG IaCL and lacking a phosphodiester linkage between the tethered nucleotides have been shown to be substrates of the human DNA repair protein O^6-alkylguanine DNA alkyltransferase (hAGT), which repairs the linkage restoring dG.[323] In this process, the first reaction with AGT forms two products, a DNA-protein cross-linked (DPC) species and a repaired oligonucleotide. The DPC is a very poor substrate for a second repair

event which releases the second unmodified (repaired) DNA fragment. Butylene linked IaCL were observed to be resistant to the action of hAGT using extended reaction times (8h) and five molar equivalents of the protein. However, the heptylene linked IaCL DNA analogues were almost completely consumed under these conditions.[323] The template strands employed in this study were designed based on cisplatin-adducted DNA sequences studied by Zhao and coworkers.[227]

Steady-state kinetic evaluations of nucleotide incorporations opposite IaCL lacking a phosphodiester linkage between the attached nucleobases (**GG4** and **GG7**, where 4 and 7 are butylene and heptylene linked), containing a phosphodiester linkage (**GpG4** and **GpG7)** and unmodified **GG** were carried out with the catalytic core construct of hPol η (amino acids 1-432). The primer extension assay is illustrated in **Figure 5.2**. In all cases, the IaCL modification blocked DNA synthesis by hPol η relative to the unmodified control. Incorporation of the correct dCMP nucleotide by hPol η opposite the first 3′-alkylated dG of **GG4**, **GG7**, **GpG4**, and **GpG7** was reduced by approximately 320-, 280-, 320-, and 320-fold, respectively, relative to **GG** (see **Figure 5.3** and **Table S5.1**). These results differ significantly from the cisplatin-adducted GG templates where a 1.2 fold decrease was observed,[227] which suggests that modification of the Watson-Crick hydrogen-bonding face poses a challenge to the incorporation efficiency of hPol η. Moreover, they correlate with studies of hPol η bypass with an O^6-MedG-containing template, which demonstrated a significant decrease (10-fold) in efficiency for incorporation of the correct dCMP nucleotide.[45] The reduced efficiency of incorporation may be attributed to hindrance due to the presence of the alkylene tether. Bulkier lesions at the O^6-atom such as the benzyl and 4-oxo-4-(3-pyridyl)butyl] lesion have been shown to reduce incorporation efficiency by a factor of ~ 65 and 250, respectively, for hPol η.[45] The alkylene linkage to a subsequent nucleotide, found

in IaCL DNA, could also contribute to the reduced relative kinetic efficiencies observed. Significant misinsertions by *Saccharomyces cerevisiae* Pol η have also been observed across the dG nucleotide of the dG[8,5-Me]dT IaCL, whereas hPol η mostly incorporated the correct dNMPs across both nucleotides of this IaCL.[222,223,341] Steady-state data for our IaCL DNA showed a preference for dCMP insertion across the first 3′-dG alkylated residue. However, an overall decrease in selectivity (*f* coefficients are shown in **Figure 5.3** and supplied in **Table S5.1**) was observed, particularly for the **GG4** and **GG7** templates. Relative insertions (*f* coefficients) of dAMP, dGMP and dTMP increased to 0.37, 0.34, and 0.24 for **GG4** and 0.40, 0.23, and 0.47 for **GG7** compared to 0.001, 0.001, and 0.003 for **GG**, respectively. Lower *f* coefficients were generally observed for **GpG4** and **GpG7** compared to **GG4** and **GG7**, respectively, suggesting the added flexibility inherent to **GG4** and **GG7** posed an issue for hPol η fidelity. It was interesting to observe an increased efficiency and selectivity for the butylene linkers (**GG4** and **GpG4**) in comparison to heptylene linkers (**GG7** and **GpG7**), substantiating that increased steric bulk hinders processivity by hPol η. It was noted that in all cases, except **GG4**, dTMP was incorporated most-efficiently from the misinsertion pool across the 3′-dG position.

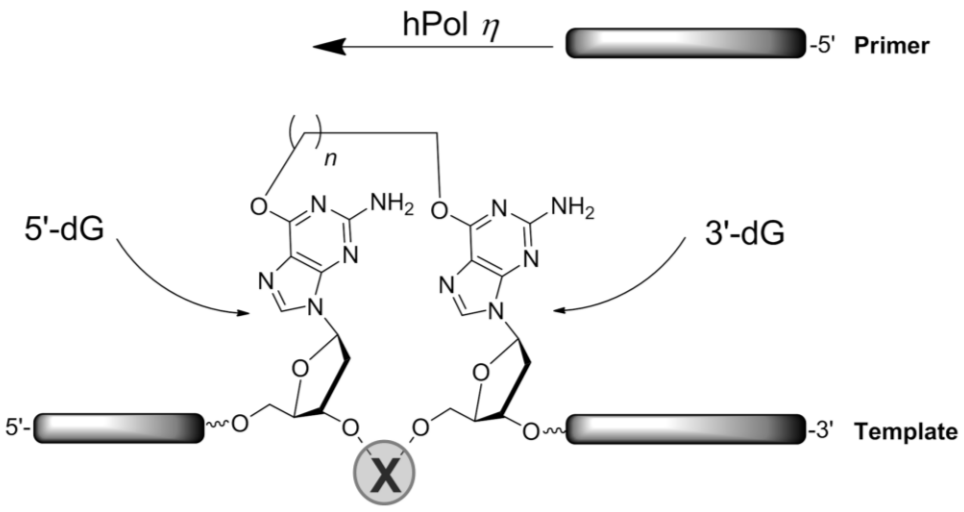
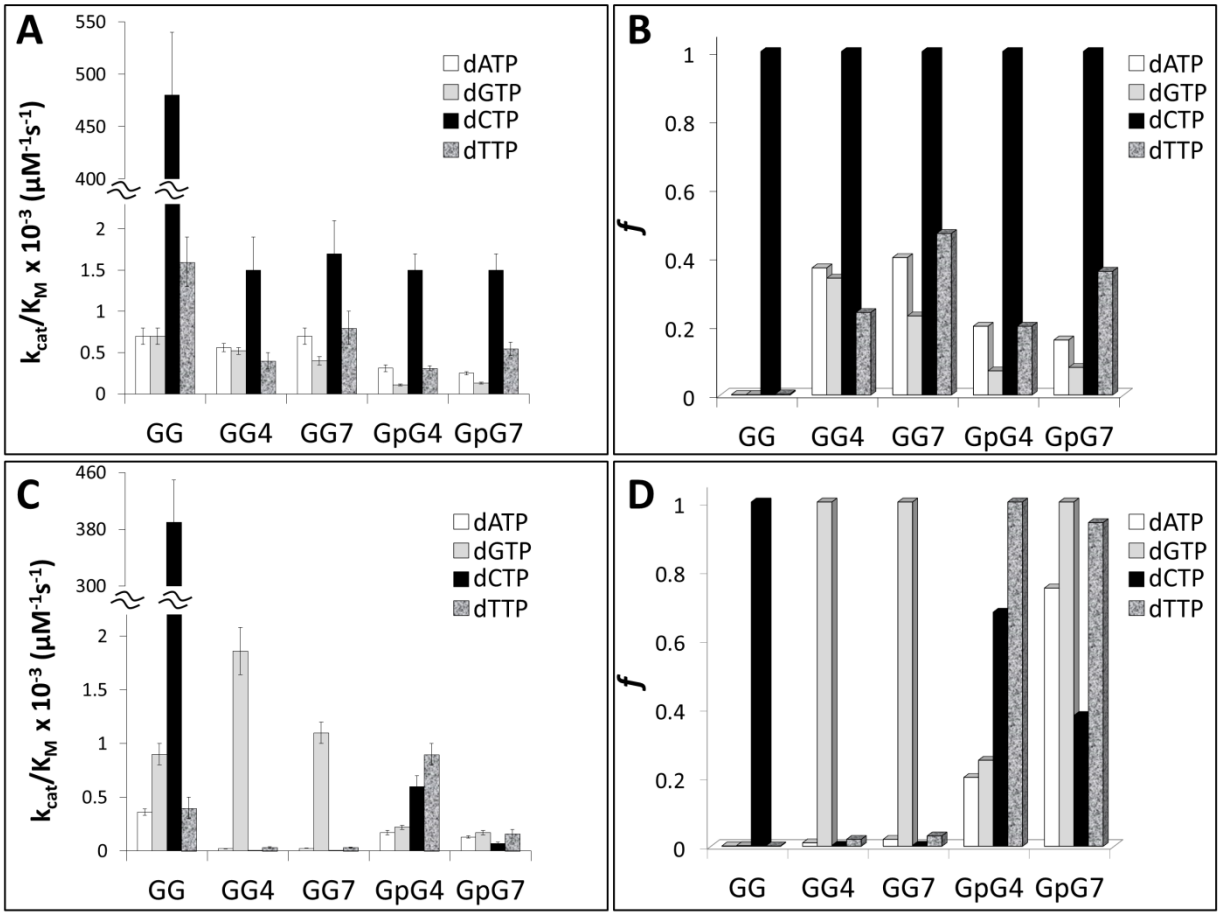


Figure 5.2: Pictorial representation of primer extension assay using hPol η, where **X** represents the presence, or lack, of a phosphodiester linkage. Frameshift adduct formation (-1), observed primarily at the 5'-dG residue for **GG4** and **GG7**, is characterized by hPol η skipping the template nucleotide.

Post-insertion extension across the second alkylene linked site (5′-dG) using primers that contain dC or dT directly across from the 3′-dG of the IaCL containing templates, as well as the unmodified **GG**, was studied. The results are summarized in **Figure 5.3** and tabulated in **Tables S5.2** and **S5.3**. Extension across the second O^6-alkylene linked dG was generally slower compared to the first regardless of the identity of the additional nucleotide (dC or dT) in the primer. An interesting feature for the **GG4** and **GG7** templates was the apparent proficient misinsertion of dGTP across the O^6-alkylated 5′-dG residue. It was suspected that frameshift adduct formation had occurred, which was confirmed by full extension analysis by LC-MS/MS, given that dC is the subsequent nucleotide in the template strand. dNTPs other than dGTP were not efficiently incorporated in the case of **GG4** and **GG7**. The post-extension profiles for **GpG4** and **GpG7** displayed a clear decrease in selectivity compared to that of insertion across the first (3′) O^6-alkylated-dG, with an almost complete loss in dNTP preference for **GpG7**. To be noted is the preference of dTTP over dCTP for **GpG4** and **GpG7**. The 3′-end O^6-alkylated-dG • dT

mismatch significantly lowered incorporation efficiency across the subsequent O^6-alkylated-dG for all IaCL DNA except for insertion of dCMP in the case of **GpG4**. The steady-state results showed an overall reduction in incorporation efficiency and fidelity, particularly across the 5′-dG of the IaCL. This occurrence is not observed for hPol η bypass of platinum-based dGG adducts nor the dG[8,5-Me]dT adduct.[224] The reduction of incorporation efficiency and fidelity across the 5′-dG may be the result of the alkylene linker adopting an orientation that disrupts efficient DNA primer extension.



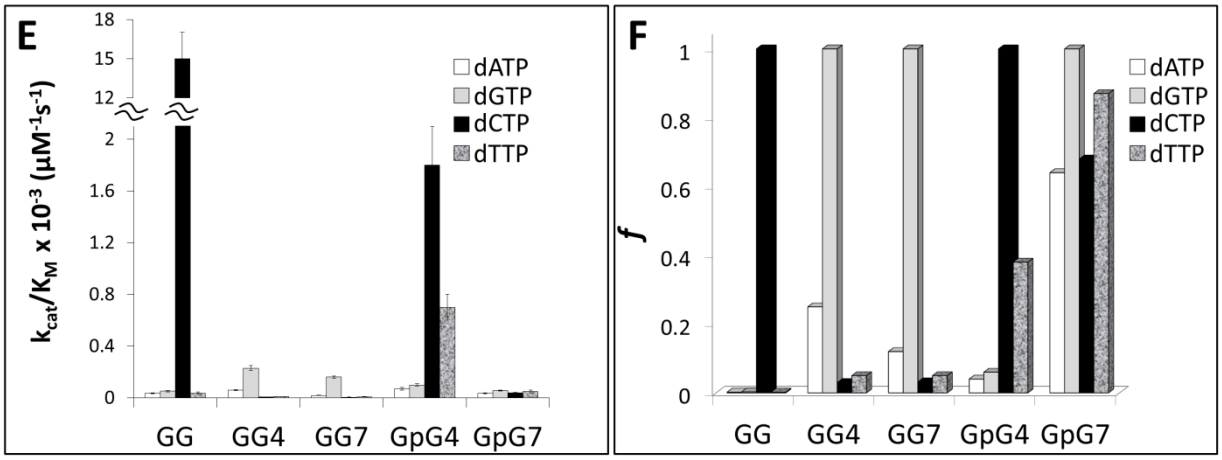


Figure 5.3: Summary of steady-state kinetics of incorporation of dNTP opposite IaCL-containing template **GG4**, **GG7**, **GpG4**, **GpG7** (5′- AC **XX** CT CAC ACT) and unmodified template (**GG**) by hPol *η*. DNA primer sequence identity were 3′- GA GUG TGA T(FAM)-5′ (Panel **A** and **B**), 3′- **C** GA GUG TGA T(FAM)-5′ (Panel **C** and **D**), or 3′- **T** GA GUG TGA T(FAM)-5′ (Panel **E** and **F**). Tabulated values are reported in **Table S5.1-S5.3**.

Analysis of single insertions by a DNA polymerase is useful for kinetic analysis in terms of identifying the extent of blockage by a given modification in a quantitative manner. However, this may not reflect insertion profiles in the presence of all four dNTPs as well as extension past the damaged sites. Fidelity of hPol *η* and processivity across and past the damage site was assessed *via* analysis of extended products by LC-MS/MS, using a reported methodology.[45,229,256–258] This method provides insight regarding hPol *η*'s preference to incorporate dNTP across the IaCL site (**Tables S5.4** and **S5.5**). The PAGE analysis of fully extended products (**Figure 5.4**) revealed blockage of hPol *η* after insertion across the **GG4** and **GG7** IaCL modification (*e.g.* extension was stalled after incorporation of the second dNTP). Similar results were observed for replication across and past DNA templates containing a single abasic site insert and may represent slippage of hPol *η* during replication.[236] Our study suggests that hPol *η* bypass requires an intact phosphodiester DNA backbone for processivity, particularly for IaCL. In contrast, hPol *η* replicated past the IaCL of **GpG4** and **GpG7** to reach an accumulation of the fully extended (13 nt) and over-extended (14 nt) product (**Figure 5.4**). hPol

η exhibited similar bypass extension profiles for other O^6-alkyl-dG lesions, which suggested that the IaCL containing a phosphodiester linkage may be processed as two adjacent mono-adducted dG inserts.

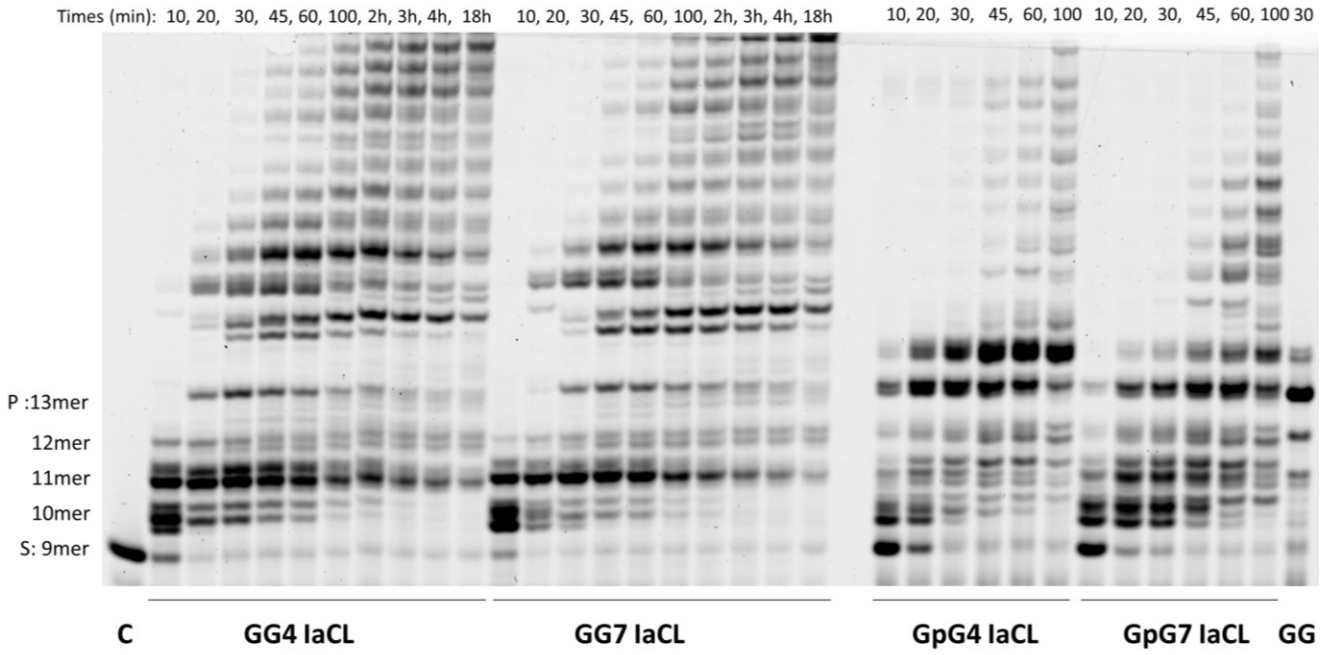


Figure 5.4: Time course assay of primer full extension (25 µM annealed DNA) by hPol η (3.4 µM for damaged and 1µM undamaged DNA) in presence of all four dNTPs (1 mM mixture). DNA duplexes were composed of IaCL-containing template **GG4**, **GG7**, **GpG4**, **GpG7** (5′- AC **XX** CT CAC ACT) or unmodified template (**GG**), annealed with DNA primer (3′- GA GUG TGA T(FAM)-5′). DNA template used is indicated at the bottom, product length identified on the left-hand side, reaction times are indicated at the top (in minutes unless stated otherwise). Note that only the 30 min reaction time is shown for the undamaged control (**GG**) (data for other time points not shown). Extended products were resolved by 17 % denaturing PAGE (19:1) and visualized *via* fluorography.

Insertion profiles opposite the 3′-dG and 5′-dG of **GG4**, **GG7**, **GpG4** and **GpG7** and the unmodified control are shown in **Figure 5.5**. Insertion of dCMP across the 3′-end of the O^6-dG-alkylene-O^6-dG was preferred by hPol η in all cases, which was in agreement with the steady-state kinetic data. Although the DNA replication machinery has been shown to misinsert dTMP opposite O^6-MedG[45,342–344], only a slight increase (3 - 12 %) in dTMP insertion by hPol η across the 3′-dG relative to the unmodified control was observed. Interestingly, the MS-based analysis

revealed that dAMP incorporation occurs at levels of 24% and 14% for **GG4** and **GG7**, respectively, and in only negligible amounts for **GpG4**, **GpG7** and the unmodified control. Similar erroneous dAMP insertions by hPol η have been observed opposite abasic sites (also known as the "purine-rule") in addition to frameshift adduct formation.[236,351–353] Minor products (<3 %) corresponded to frameshift adduct formation opposite 3′-dG lesions.

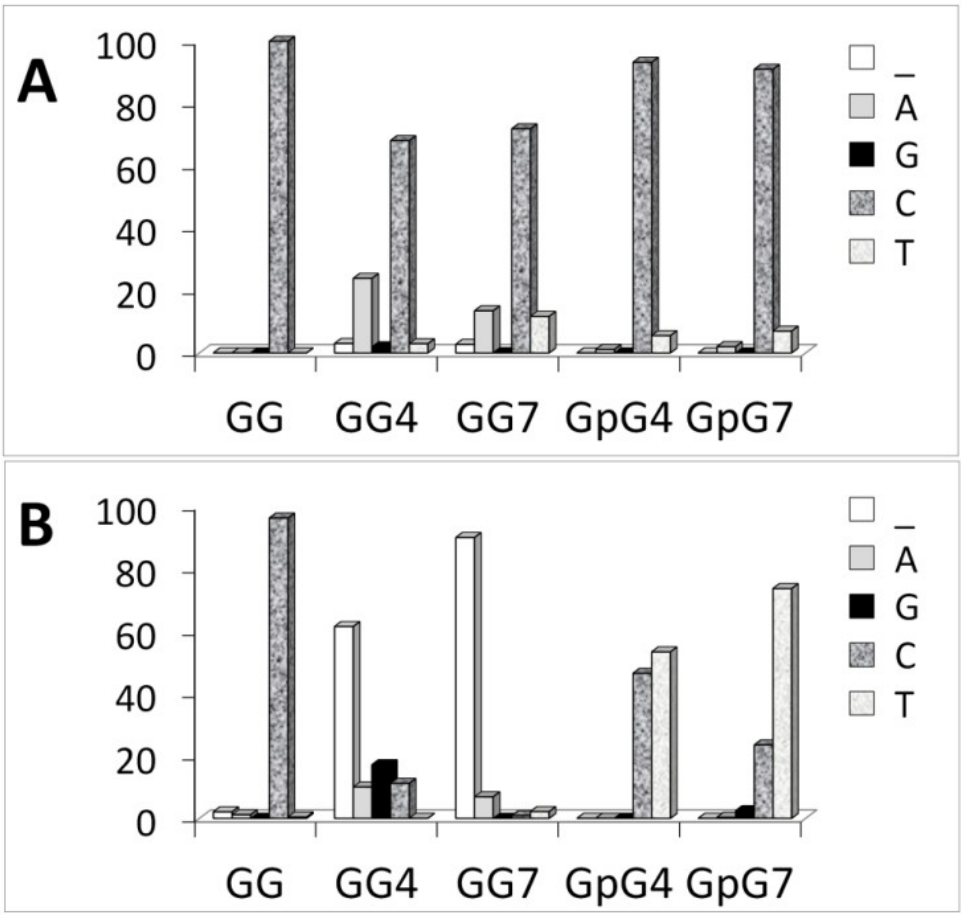


Figure 5.5. Insertion profiles by hPol η opposite the (**A**) 3′-dG and the (**B**) 5′-dG of the O^6-dG-alkylene-O^6-dG IaCL and unmodified control. DNA sequence contexts were as follows: IaCL-containing template (**GG4**, **GG7**, **GpG4**, **GpG7** or unmodified template **GG**) had sequence 5′-AC **XX** CT CAC ACT), and DNA primer sequence was 3′- GA GUG TGA T(FAM)-5′. Tabulated LC-MS extension products and percentages can be found in **Table S5.4-S5.5**.

Nucleotide insertion profiles for the 5′-end of the dG lesion by hPol η revealed proclivity for the frameshift adduct formation in the case of **GG4** and **GG7,** whereas a pyrimidine insertion

was preferred for **GpG4** and **GpG7**. These results were in agreement with the steady-state kinetic evaluation of dGMP insertion across the 5′-end of the dG lesion of **GG4** and **GG7**. We suspect that the added flexibility incurred from the lack of a phosphodiester linkage is responsible for large discrepancies observed between the IaCL DNA. hPol η incorporated little to no purines opposite the 5′-end of the dG lesion in **GpG4** and **GpG7**, but levels of 10 % and 17 % for dAMP and dGMP, respectively, were measured for **GG4**. Only 7 % dAMP was detected for extension across the O^6-alkylated 5′-dG in **GG7**, with no dGMP insertion detected. Misinsertion of dTMP was preferred by hPol η in the case of **GpG4** and **GpG7** (53% and 74%, respectively). Correct insertion of dCMP accounted for 47% and 24% opposite the O^6-alkylated 5′-dG in **GpG4** and **GpG7**, respectively, and only a minor dCMP insertion was observed for **GG4** and **GG7** (11% and 1%, respectively).

5.4 Conclusion

In conclusion, both the steady-state primer extension assays and LC-MS analysis of fully extended product demonstrated that extension opposite the 5′-end of the O^6-dG-alkylene-O^6-dG was more problematic compared to the 3′-end. An increase in frameshift adduct formation by hPol η was observed at 5′-dG of IaCL DNA lacking the phosphodiester linkage, compared to those IaCL containing this functional group. To be noted was the decrease in selectivity by hPol η for longer heptylene IaCL *versus* the butylene analogue, regardless of the presence of the phosphodiester group at the cross-linked site. Structural insights on the ternary complexes (DNA, hPol η and incoming nucleoside triphosphate) by X-ray crystallography are being initiated and will shed light on the error-free bypass and error-prone behavior of hPol η towards these IaCL.

5.5 Acknowledgement

The authors are grateful to Dr. Amritraj Patra and Dr. Anne Noronha for helpful discussions.This work was supported by grants from the Natural Sciences and Engineering Research Council (NSERC) of Canada (299384-2011 to C.J.W.) and Canada Research Chair Program (950-213807 to C.J.W.) as well as the US National Institutes of Health (R01 ES010375 to F.P.G. and M.E. and P01 CA160032 to M.E.). D.K.O. was a recipient of a postgraduate fellowship (CGS-D) and the Michael Smith Foreign Study Supplement Program from NSERC.

5.6 Supporting Information

5.6.1 Supporting methods

5.6.1.1 Chemical Synthesis and Characterization of Modified DNA Oligonucleotides

The chemical synthesis of the dimer phosphoramidites required for **GG4** and **GG7** have been described in previously published work.[323] Those required for **GpG4**, and **GpG7** are described **CHAPTER IV** with syntheses adapted from previously described procedures.[297] The cross-linked duplexes, whose sequences are 5′- AC **XX** CT CAC ACT (where **XX** denotes the cross linked dGG residues or dGG for the unmodified control), were assembled on an Applied Biosystems Model 3400 synthesizer on a 1.5 μmol scale using standard β-cyanoethylphosphoramidite chemistry supplied by the manufacturer with slight modifications to coupling times. The nucleoside phosphoramidites containing fast deprotecting groups were prepared in anhydrous MeCN at a concentration of 0.1 M for the 3′-*O*-deoxyphosphoramidites, and 0.15 M for the cross-linked 3′-*O*-deoxyphosphoramidite. Oligomer sequence assembly was carried out according to published procedures by our group.[323]

Protecting groups and cleavage from the solid support was carried out by treatment with aqueous (28% v/v) NH_4OH in EtOH (1 mL of a 3 : 1 v/v solution) for 4h at 55 ^oC in 2 mL screw cap microfuge tubes fitted with Teflon lined caps for **GG4** and **GG7**. Modified oligomers **GpG4** and **GpG7** were first deprotected for 16 h in aqueous (28% v/v) NH_4OH at room temperature with gentle rocking, followed by an additional 4 h at 55 ^oC. Crude oligomers were transferred to new screw cap microfuge tubes and the solvent removed using a Savant SC110A SpeedVac Concentrator (Thermo) followed by purification by strong-anion exchange HPLC with a Dionex DNAPAC PA-100 column (0.4 cm × 25 cm) purchased from Dionex, Sunnyvale, CA using a linear gradient of 0-52% buffer B (v/v) over 24 min (buffer A: 100 mM Tris HCl, pH 7.5, 10% MeCN and buffer B: 100 mM Tris HCl, pH 7.5, 10% MeCN, 1 M NaCl) at 55 ^oC. The columns were monitored at 260 nm for analytical runs or 280 nm for preparative runs. The purified oligomers were desalted using C-18 SEP PAK cartridges (Waters) as previously described.[317] ESI mass spectra for oligonucleotides were obtained at the Concordia University Centre for Biological Applications of Mass Spectrometry using a Micromass Qtof2 mass spectrometer (Waters) equipped with a nanospray ion source. The mass spectrometer was operated in full scan, negative ion detection mode. The molecular mass of the modified oligomers were identifed by ESI-MS and the measured values were in agreement with the expected masses (see **Figure S5.2-S5.4** for MS spectra).

5.6.2 Supporting Figures and Tables

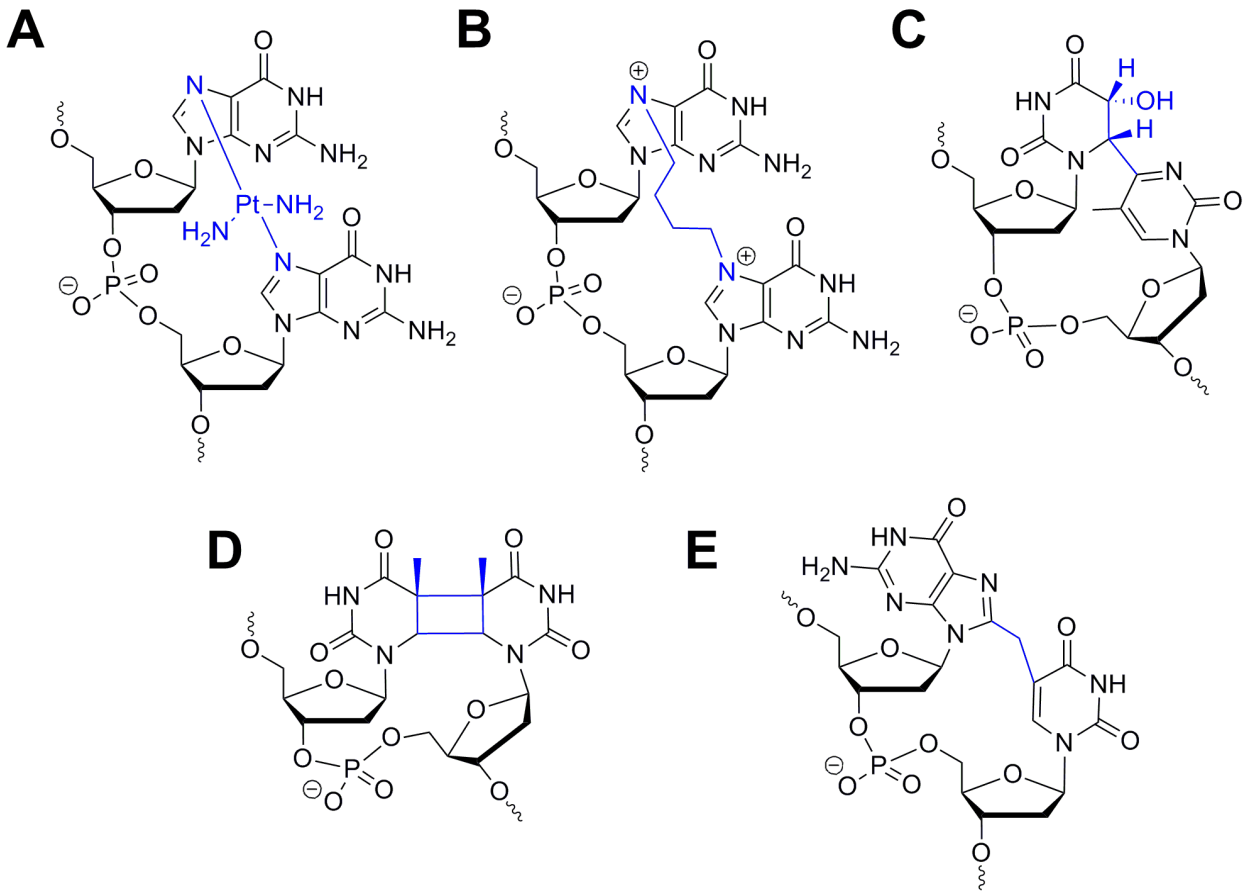


Figure S5.1 - Examples of different IaCL DNA adducts mentioned in the main text. **A** An example of an IaCL formed from exposure of DNA to the chemotherapeutic drug cisplatin.[227,350] **B** An example of an IaCL formed from exposure of DNA to the chemotherapeutic drug busulfan.[69] **C** An example of a UV-induced (6-4) pyrimidone photoproduct lesion.[82] **D** An example of a UV-induced cyclubutane pyrimidine dimer. [207] **E** The structure of dG[8,5-Me]dT IaCL.[223] Modifications are highlighted in blue.

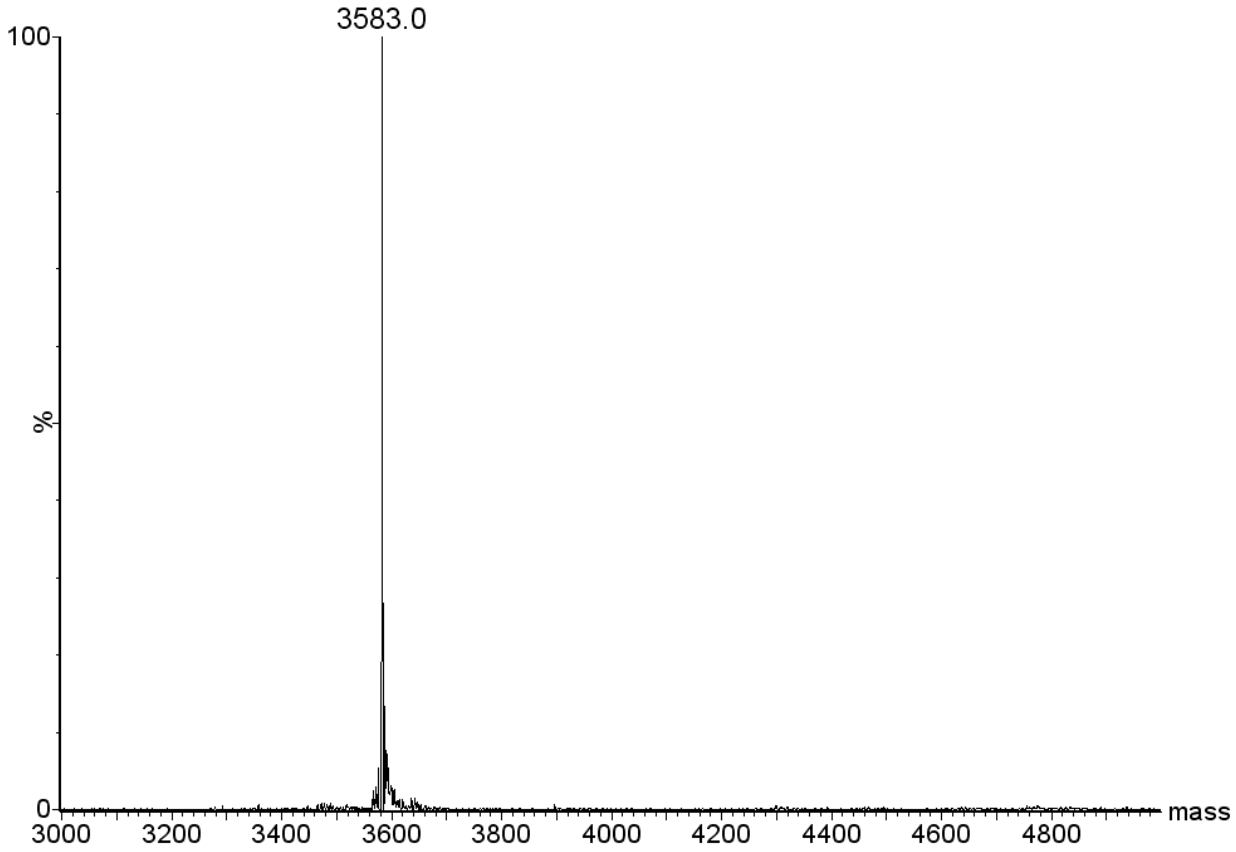


Figure S5.2 - Deconvoluted ESI MS spectrum of oligonucleotide **GG4** (expected mass of 3582.5).

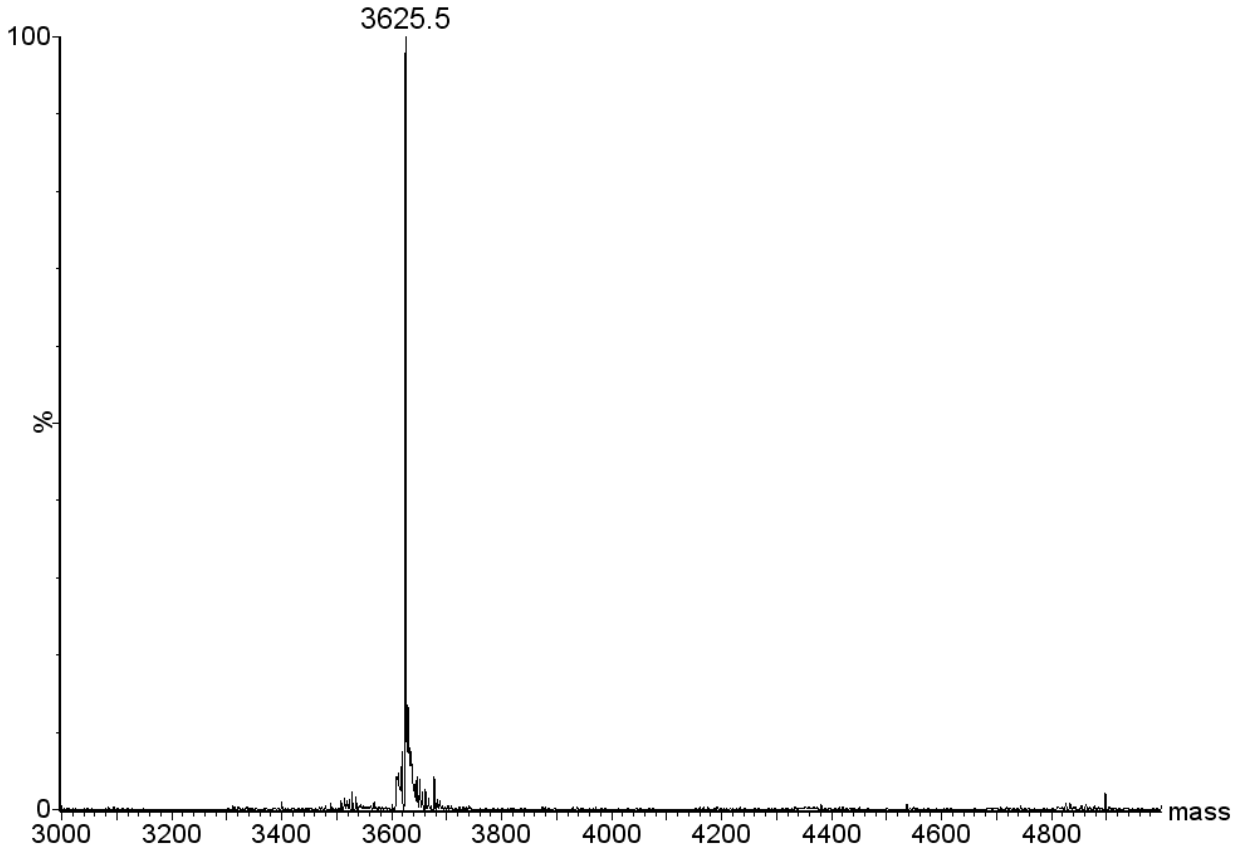


Figure S5.3 - Deconvoluted ESI MS spectrum of oligonucleotide **GG7** (expected mass of 3624.6).

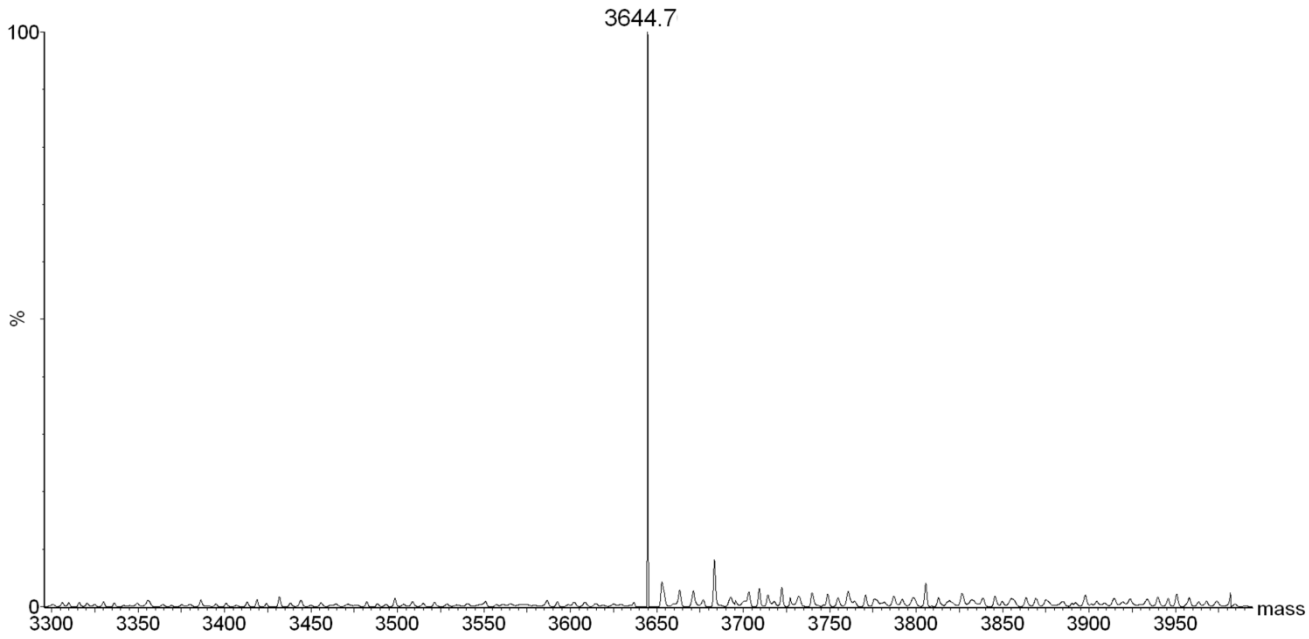


Figure S5.4 - Deconvoluted ESI MS spectrum of oligonucleotide **GpG4** (expected mass of 3644.5).

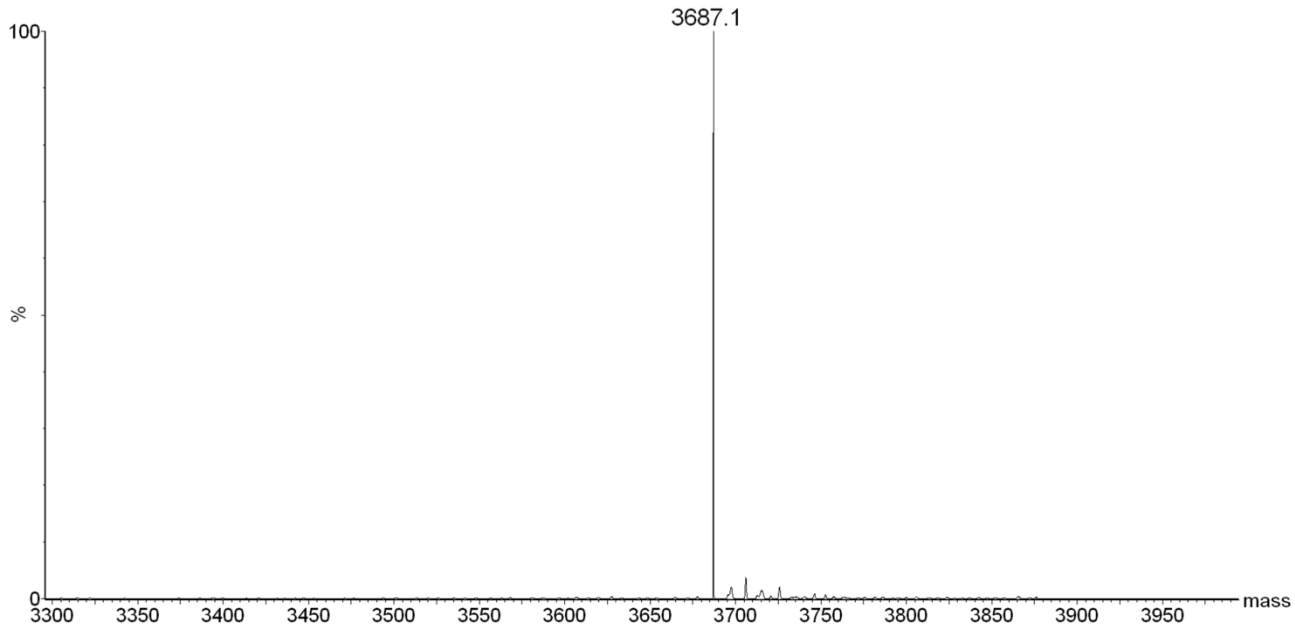


Figure S5.5 - Deconvoluted ESI MS spectrum of oligonucleotide **GpG7** (expected mass of 3686.6).

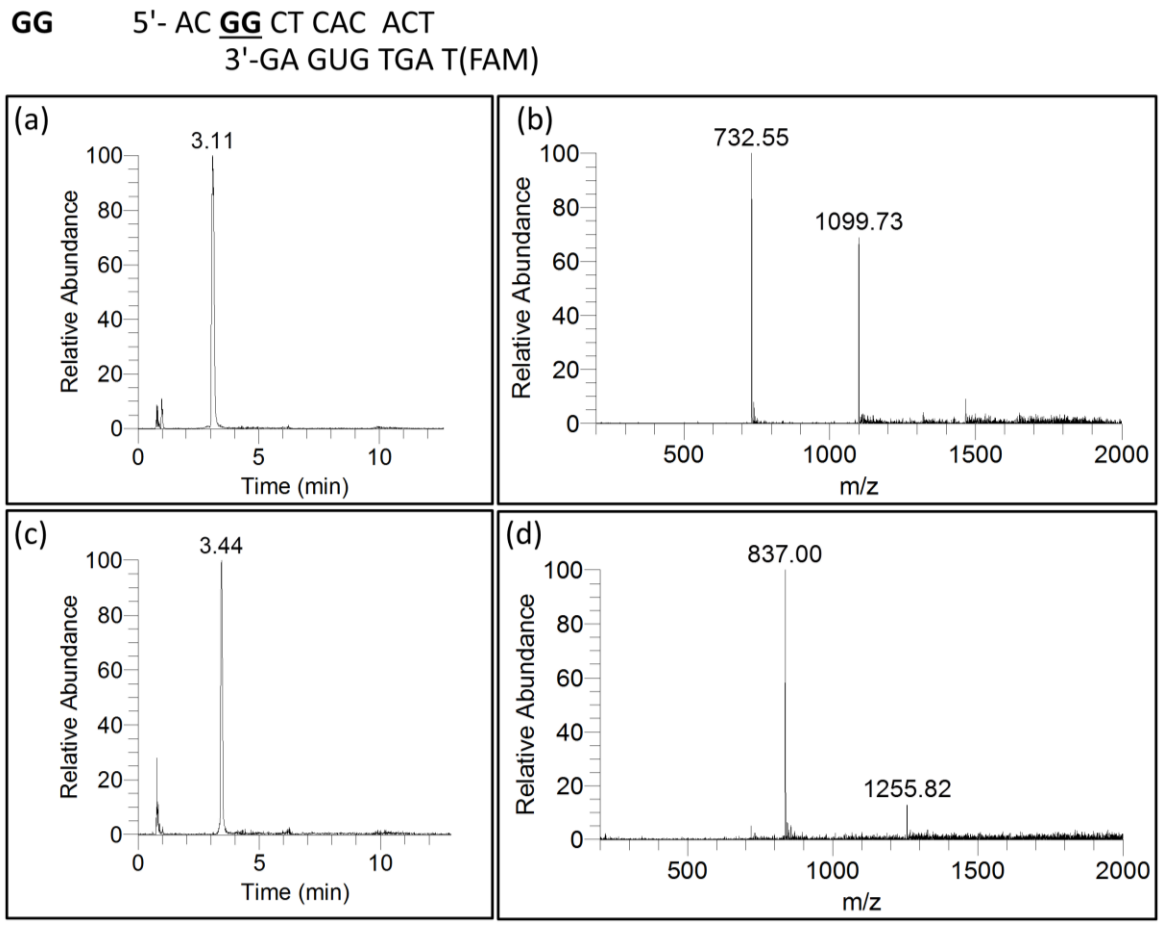


Figure S5.6 - LC-MS analysis of most abundant full-length extension products opposite **GG** in DNA template by hPol η in the presence of all four dNTPs. (a) Sample reconstructed extracted ion chromatogram for m/z 732.80 for product with sequence 5′-pGAGCCGT and (b) mass spectrum of peak at retention time 3.11 min. (c) Sample reconstructed extracted ion chromatogram for m/z 837.20 for product with sequence 5′-pGAGCCGTA and (d) mass spectrum of peak at retention time 3.44 min. See **Table S5.4** for full list of products and respective m/z assignments. See above for methodology.

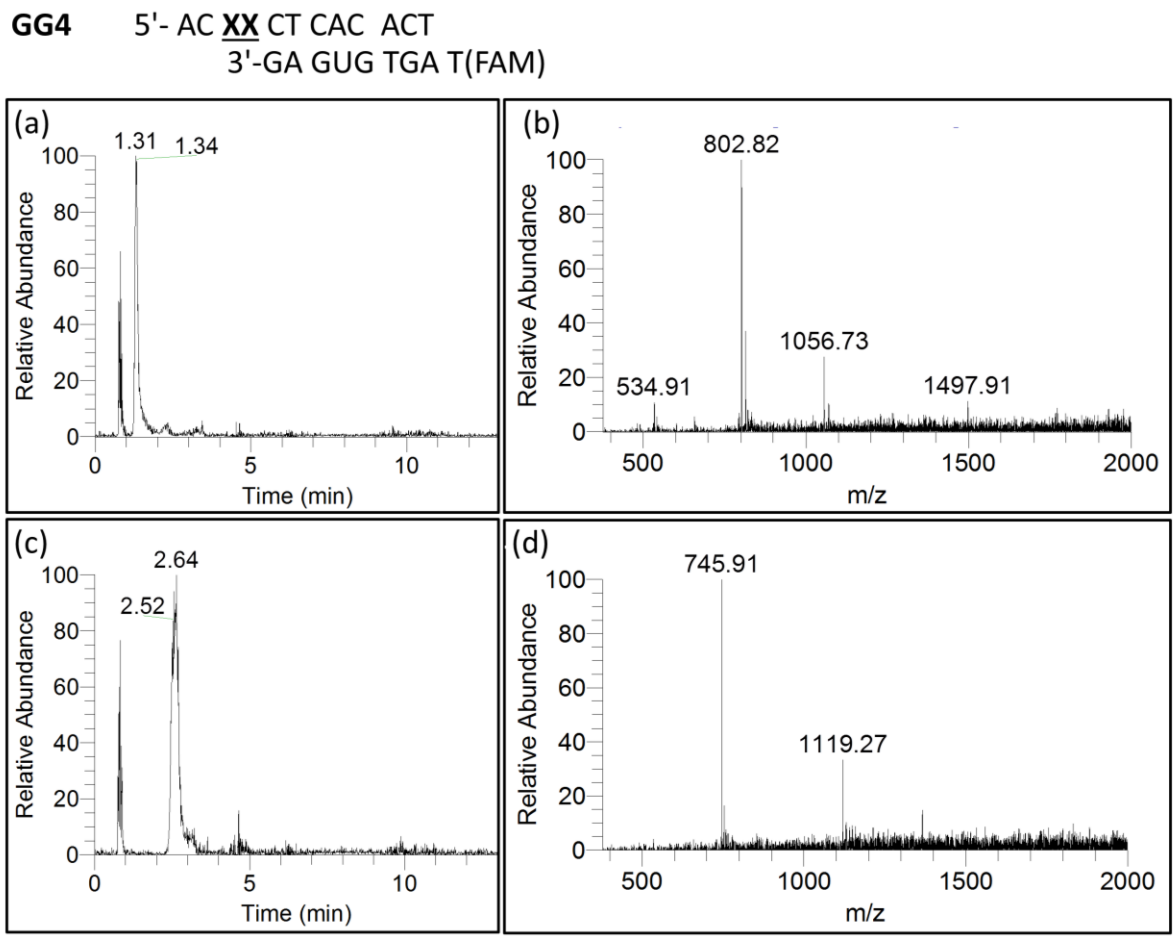


Figure S5.7 - LC-MS analysis of most abundant full-length extension products opposite **GG4** in DNA template by hPol η in the presence of all four dNTPs. (a) Sample reconstructed extracted ion chromatogram for m/z 803.0 for product with sequence 5′-pGAGCG and (b) mass spectrum of peak at retention time 1.31 min. (c) Sample reconstructed extracted ion chromatogram for m/z 746.14 for product with sequence 5′-pGAGC_GTG and (d) mass spectrum of peak at retention time 2.64 min. See **Table S5.4** for full list of products and respective m/z assignments. See above for methodology.

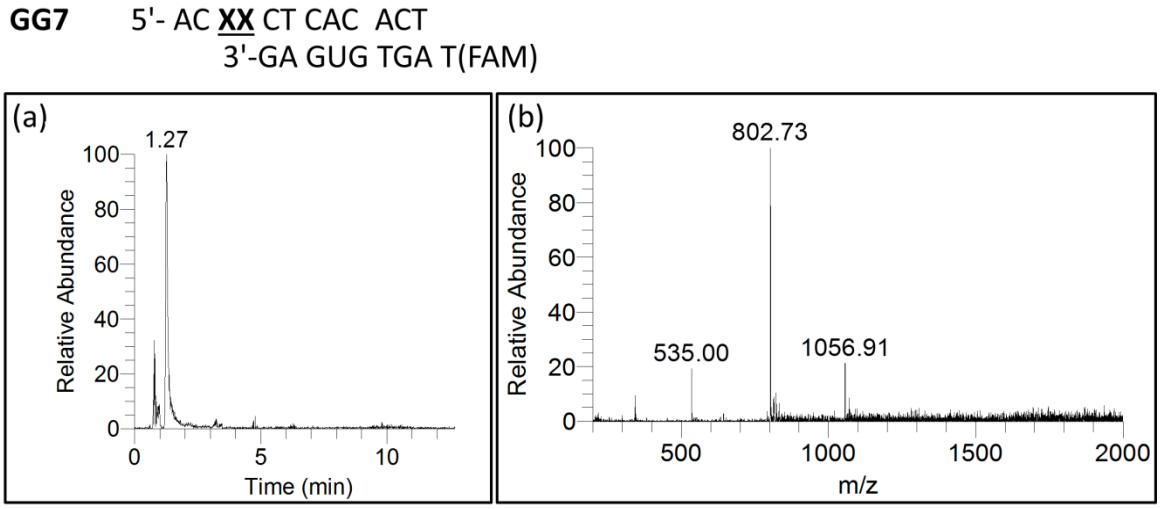


Figure S5.8 - LC-MS analysis of most abundant full-length extension products opposite **GG7** in DNA template by hPol η in the presence of all four dNTPs. (a) Sample reconstructed extracted ion chromatogram for m/z 803.0 for product with sequence 5′-pGAGCG and (b) mass spectrum of peak at retention time 1.27 min. See **Table S5.4** for full list of products and respective *m/z* assignments. See above for methodology.

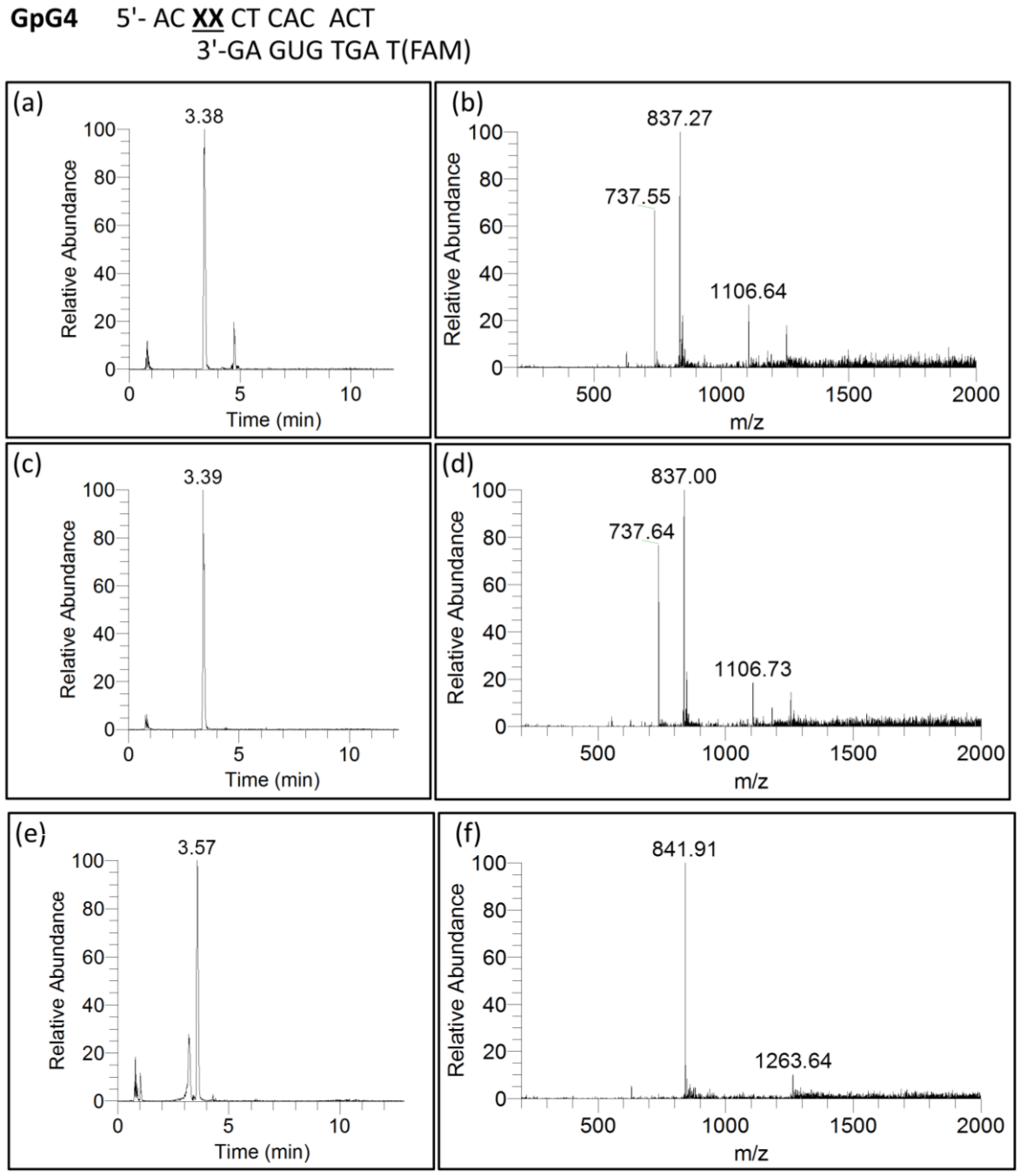


Figure S5.9 - LC-MS analysis of most abundant full-length extension products opposite **GG** in DNA template by hPol η in the presence of all four dNTPs. (a) Sample reconstructed extracted ion chromatogram for m/z 737.80 for product with sequence 5′-pGAGCTGT and (b) mass spectrum of peak at retention time 3.38 min. (c) Sample reconstructed extracted ion chromatogram for m/z 837.20 for product with sequence 5′-pGAGCCGTA and (d) mass spectrum of peak at retention time 3.39 min. (e) Sample reconstructed extracted ion chromatogram for m/z 842.21 for product with sequence 5′-pGAGCTGTA and (f) mass

spectrum of peak at retention time 3.57 min. See **Table S5.4** for full list of products and respective *m/z* assignments. See above for methodology.

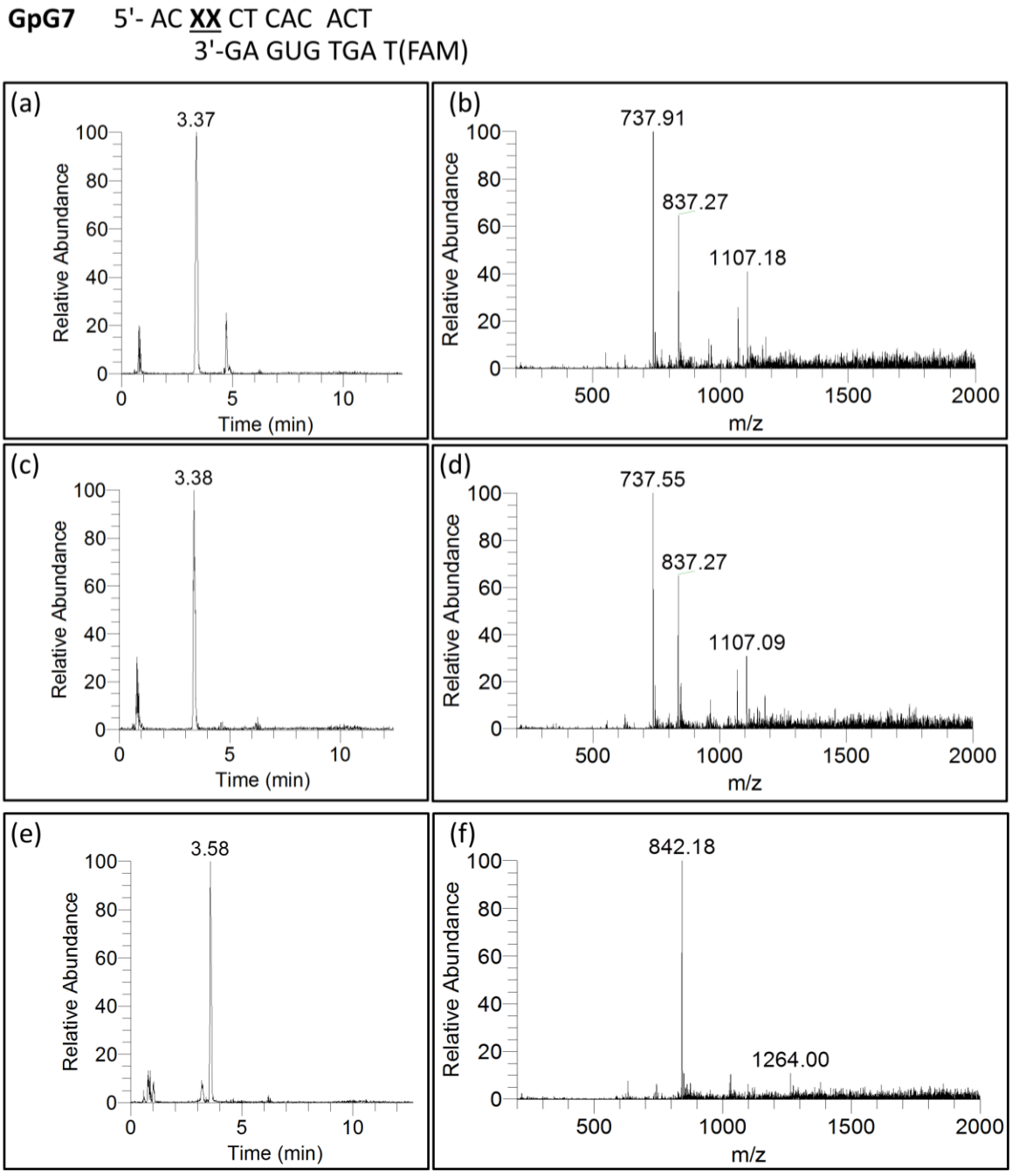


Figure S5.10 - LC-MS analysis of most abundant full-length extension products opposite **GG** in DNA template by hPol η in the presence of all four dNTPs. (a) Sample reconstructed extracted ion chromatogram for m/z 737.80 for product with sequence 5′-pGAGCTGT and (b) mass spectrum of peak at retention time 3.37 min. (c) Sample reconstructed extracted ion

chromatogram for m/z 837.20 for product with sequence 5′-pGAGCCGTA and (d) mass spectrum of peak at retention time 3.38 min. (e) Sample reconstructed extracted ion chromatogram for *m/z* 842.21 for product with sequence 5′-pGAGCTGTA and (f) mass spectrum of peak at retention time 3.58 min. See **Table S5.4** for full list of products and respective *m/z* assignments. See above for methodology.

Table S5.1 - Steady-state kinetics of incorporation of dNTP opposite 3′-end O^6-alkylated-dG of IaCL-containing template **GG4**, **GG7**, **GpG4**, **GpG7** (5′- AC **XX** CT CAC ACT) and unmodified template (**GG**) by hPol η. DNA primer sequence identity was 3′- GA GUG TGA T(FAM)-5′.

Template base	dNTP	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M x 10^{-3} ($\mu M^{-1}s^{-1}$)	f^1
G	dATP	0.045 ± 0.002	62 ± 9	0.7 ± 0.1	0.001
G	dGTP	0.052 ± 0.003	74 ± 13	0.7 ± 0.1	0.001
G	dCTP	2.32 ± 0.09	4.8 ± 0.5	480 ± 60	
G	dTTP	0.18 ± 0.01	111 ± 22	1.6 ± 0.3	0.003
GG4	dATP	0.037 ± 0.001	66 ± 5	0.56 ± 0.05	0.37
GG4	dGTP	0.029 ± 0.001	56 ± 4	0.52 ± 0.04	0.34
GG4	dCTP	0.12 ± 0.01	76 ± 17	1.5 ± 0.4	
GG4	dTTP	0.040 ± 0.003	111 ± 22	0.36 ± 0.08	0.24
GG7	dATP	0.031 ± 0.001	45 ± 5	0.70 ± 0.08	0.40
GG7	dGTP	0.025 ± 0.001	63 ± 7	0.40 ± 0.05	0.23
GG7	dCTP	0.12 ± 0.01	69 ± 16	1.7 ± 0.4	
GG7	dTTP	0.07 ± 0.01	92 ± 21	0.8 ± 0.2	0.47
GpG4	dATP	0.025 ± 0.001	80 ± 11	0.31 ± 0.04	0.20
GpG4	dGTP	0.0074 ± 0.0002	69 ± 5	0.107 ± 0.009	0.07
GpG4	dCTP	0.078 ± 0.003	51 ± 5	1.5 ± 0.2	
GpG4	dTTP	0.046 ± 0.006	150 ± 36	0.30 ± 0.08	0.20
GpG7	dATP	0.021 ± 0.001	83 ± 8	0.25 ± 0.02	0.16
GpG7	dGTP	0.0103 ± 0.0002	83 ± 5	0.13 ± 0.01	0.08
GpG7	dCTP	0.066 ± 0.003	43 ± 6	1.5 ± 0.2	
GpG7	dTTP	0.040 ± 0.002	72 ± 9	0.55 ± 0.08	0.36

[1] $(k_{cat}/K_M)_i / (k_{cat}/K_M)_{max}$

Table S5.2 - Steady-state kinetics of incorporation of dNTP opposite 5′-end O^6-alkylated-dG of IaCL-containing template **GG4**, **GG7**, **GpG4**, **GpG7** (5′- AC **XX** CT CAC ACT) and unmodified template (**GG**) by hPol η. DNA primer sequence identity was 3′- **C**GA GUG TGA T(FAM)-5′.

Template base	dNTP	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M x 10^{-3} ($\mu M^{-1}s^{-1}$)	f^1
G	dATP	0.028 ± 0.001	76 ± 5	0.36 ± 0.03	0.0009
G	dGTP	0.033 ± 0.001	38 ± 3	0.87 ± 0.07	0.002
G	dCTP	2.19 ± 0.11	5.7 ± 0.8	390 ± 60	
G	dTTP	0.09 ± 0.01	260 ± 34	0.35 ± 0.05	0.001
GG4	dATP	0.0021 ± 0.0001	93 ± 9	0.022 ± 0.002	0.01
GG4	dGTP	0.059 ± 0.002	32 ± 4	1.9 ± 0.2	
GG4	dCTP	0.0014 ± 0.0001	236 ± 26	0.0059 ± 0.0007	0.003
GG4	dTTP	0.009 ± 0.001	274 ± 42	0.033 ± 0.006	0.02
GG7	dATP	0.00176 ± 0.00004	76 ± 6	0.023 ± 0.002	0.02
GG7	dGTP	0.029 ± 0.001	27 ± 3	1.1 ± 0.1	
GG7	dCTP	0.00075 ± 0.00005	195 ± 35	0.0038 ± 0.0007	0.003
GG7	dTTP	0.0057 ± 0.0003	177 ± 24	0.032 ± 0.005	0.03
GpG4	dATP	0.0139 ± 0.0004	81 ± 8	0.17 ± 0.02	0.20
GpG4	dGTP	0.0057 ± 0.0001	26 ± 2	0.22 ± 0.02	0.25
GpG4	dCTP	0.060 ± 0.006	103 ± 20	0.6 ± 0.1	0.68
GpG4	dTTP	0.047 ± 0.003	55 ± 8	0.9 ± 0.1	
GpG7	dATP	0.0070 ± 0.0002	54 ± 5	0.13 ± 0.01	0.75
GpG7	dGTP	0.0055 ± 0.0001	32 ± 3	0.17 ± 0.02	
GpG7	dCTP	0.025 ± 0.002	375 ± 57	0.07 ± 0.01	0.38
GpG7	dTTP	0.049 ± 0.006	307 ± 60	0.16 ± 0.04	0.94

[1] $(k_{cat}/K_M)_i / (k_{cat}/K_M)_{max}$

Table S5.3 - Steady-state kinetics of incorporation of dNTP opposite 5′-end O^6-alkylated-dG of IaCL-containing template **GG4**, **GG7**, **GpG4**, **GpG7** (5′- AC **XX** CT CAC ACT) and unmodified template (**GG**) by hPol η. DNA primer sequence identity was 3′- **T**GA GUG TGA T(FAM)-5′.

Template base	dNTP	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M x 10^{-3} ($\mu M^{-1}s^{-1}$)	f^1
G	dATP	0.00151 ± 0.00003	43 ± 3	0.035 ± 0.003	0.002
G	dGTP	0.00247 ± 0.00005	49 ± 4	0.051 ± 0.004	0.003
G	dCTP	0.168 ± 0.004	11 ± 1	15 ± 2	
G	dTTP	0.0034 ± 0.0003	91 ± 19	0.037 ± 0.008	0.002
GG4	dATP	0.00218 ± 0.00003	37 ± 2	0.059 ± 0.004	0.25
GG4	dGTP	0.0168 ± 0.0003	72 ± 5	0.23 ± 0.02	
GG4	dCTP	0.00134 ± 0.00005	196 ± 18	0.0068 ± 0.0007	0.03
GG4	dTTP	0.0016 ± 0.0001	144 ± 19	0.011 ± 0.001	0.05
GG7	dATP	0.00155 ± 0.00002	80 ± 4	0.019 ± 0.001	0.12
GG7	dGTP	0.0092 ± 0.0001	56 ± 2	0.162 ± 0.007	
GG7	dCTP	0.00103 ± 0.00004	198 ± 23	0.0052 ± 0.0006	0.032
GG7	dTTP	0.0016 ± 0.0001	189 ± 21	0.008 ± 0.001	0.05
GpG4	dATP	0.0048 ± 0.0002	65 ± 7	0.07 ± 0.01	0.04
GpG4	dGTP	0.0036 ± 0.0001	34 ± 2	0.104 ± 0.007	0.06
GpG4	dCTP	0.14 ± 0.01	78 ± 13	1.8 ± 0.3	
GpG4	dTTP	0.066 ± 0.006	98 ± 18	0.7 ± 0.1	0.38
GpG7	dATP	0.00206 ± 0.00004	57 ± 4	0.036 ± 0.003	0.64
GpG7	dGTP	0.00259 ± 0.00003	46 ± 2	0.056 ± 0.003	
GpG7	dCTP	0.0053 ± 0.0002	140 ± 13	0.038 ± 0.004	0.68
GpG7	dTTP	0.0057 ± 0.0003	117 ± 16	0.049 ± 0.007	0.87

[1] $(k_{cat}/K_M)_i / (k_{cat}/K_M)_{max}$

Table S5.4 - LC-MS analysis of products of hPol η replicating template **GG4**, **GG7**, **GpG4**, **GpG7** (5′- AC **XX** CT CAC ACT) and unmodified **GG**. DNA primer sequence identity was 3′- GA GUG TGA T(FAM)-5′.

Name	Fragment	*m/z* calculated	GG	GG4	GG7	GpG4	GpG7
4A	GAGA	(-2) 650.42		(-2) 650.18	(-2) 650.27		
4G	GAGG	(-2) 658.42		(-2) 658.27			
4T	GAGT	(-2) 645.92			(-2) 645.91		
5AA	GAGAA	(-2) 807.02 (-3) 537.68		(-2) 806.82 (-3) 537.73	(-2) 807.00		
5A_G	GAGAG	(-2) 815.02 (-3) 543.01		(-2) 814.73 (-3) 543.09	(-2) 814.82		(-2) 815.36 (-3) 543.18
5GA	GAGGA						
5CA	GAGCA	(-2) 795.01 (-3) 529.67	(-2) 794.73 (-3) 529.36	(-2) 794.91 (-3) 529.45	(-2) 794.73		
5AC	GAGAC						
5C_G	GAGCG	(-2) 803.01 (-3) 535.01	(-2) 802.73	(-2) 802.82 (-3) 534.64	(-2) 802.82 (-3) 534.68		
5TA	GAGTA	(-2) 802.52 (-3) 534.68			(-2) 802.82 (-3) 534.68		
5T_G	GAGTG	(-2) 810.52 (-3) 540.01		(-2) 810.36 (-3) 540.09	(-2) 810.36		
5__	GAGGT						
5CT	GAGCT	(-2) 790.51			(-2) 790.27		
6AGG	GAGAGG	(-2) 979.63		(-2) 979.45			
6TAG	GAGTAG	(-2) 967.12		(-2) 967.55			
6ATT	GAGATT	(-2) 954.62					(-2) 954.36
6TAT	GAGTAT						
6C_	GAGCGT	(-2) 955.11		(-2) 954.82	(-2) 954.91		
6C_G	GAGCGG	(-2) 967.62		(-2) 968.00			
6CTA	GAGCTA	(-2) 947.11 (-3) 631.07					(-2) 946.82 (-3) 631.09
6T_	GAGTGT	(-2) 962.62		(-2) 962.45			
7CC	GAGCCGT	(-2) 1099.7 (-3) 732.80	(-2) 1099.73 (-3) 732.55	(-2) 1099.70 (-3) 732.64		(-2) 1099.36	(-3) 733.09

						(-3) 733.00	
7CT	GAG CTG T	(-2) 1107.21 (-3) 737.80				(-2) 1106.73 (-3) 737.55	(-2) 1107.18 (-3) 737.91
7TC	GAG TCG T						
7TT	GAGTTGT	(-2) 1114.72 (-3) 742.80					(-2) 1114.73 (-3) 742.82
7C_A	GAGCGTA	(-2) 1111.72 (-3) 740.81	(-2) 1111.91 (-3) 740.36	(-2) 1111.91 (-3) 740.73			
7C_G	GAGC_GTG	(-2) 1119.72 (-3) 746.14		(-2) 1119.82 (-3) 745.91			
7AT	GAGATGT	(-2) 1119.22 (-3) 745.81				(-2) 1118.73 (-3) 745.36	
8CCA	GAG CCG TA	(-2) 1256.31 (-3) 837.20	(-2) 1255.82 (-3) 837.00	(-3) 836.91		(-2) 1256.36 (-3) 837.00	(-2) 1256.36 (-3) 837.27
8CCG	GAG CCG TG	(-2) 1264.31 (-3) 842.54	(-2) 1264.45 (-3) 842.36	(-3) 842.45		(-2) 1264.31 (-3) 842.27	(-2) 1264.00 (-3) 842.18
8CCC	GAG CCCG T	(-3) 829.19	(-3) 828.82				
8CCT	GAG CCG CT	(-2) 1244.30 (-3) 829.19				(-2) 1244.45 (-3) 829.36	
8CTA	GAG CTG TA	(-2) 1263.81 (-3) 842.21	(-2) 1264.27 (-3) 842.36			1263.64 841.91	(-2) 1264.00 (-3) 841.82
8TCA	GAG TCG TA						
8CTG	GAG CTG TG	(-2) 1271.81 (-3) 847.54				(-2) 1272.27 (-3) 847.54	(-2) 1272.18 (-3) 847.55
8TCG	GAG TCG TG						
8CTC	GAG CTG TC	(-2) 1251.80 (-3) 834.20				(-2) 1251.91 (-3) 834.18	(-2) 1251.55 (-3) 834.09
8CCT	GAG CCG TT						
8CTT	GAG CTG TT	(-2) 1259.30 (-3) 839.20				(-3) 839.09	(-3) 839.27
8TCT	GAG TCG TT						
8TTC	GAG TTG TC						
8TTA	GAG TTG TA	(-2) 1271.32 (-3) 847.21				(-2) 1271.18 (-3) 847.00	(-2) 1271.18 (-3) 846.91
8TTT	GAG TTG TT	(-2) 1266.81 (-3) 844.21				(-2) 1266.64 (-3) 844.64	

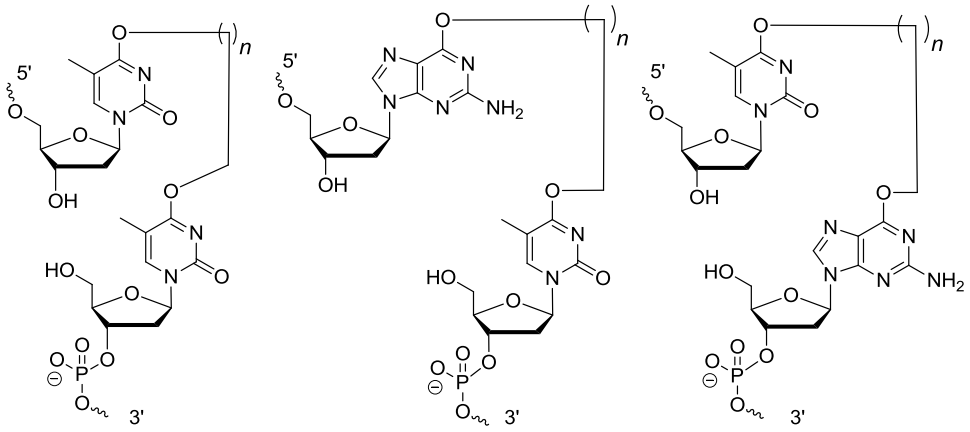
Table S5.5 - Percentages of full-length extension products determined by LC-MS/MS analysis.

Name	Fragment	GG	GG4	GG7	GpG4	GpG7
4A	GAGA		3	4		
4G	GAGG		1			
4T	GAGT			6		
5AA	GAGAA		7	3		
5A_G	GAGAG		12	6		2
5GA	GAGGA		< 1	< 1		
5CA	GAGCA	1	1	< 1		
5AC	GAGAC		1	< 1		
5CT	GAGCT			2		
5C_G	GAGCG	< 1	21	64		
5TA	GAGTA			4		
5T_G	GAGTG			3		
5__	GAGGT		3	3		
6AGG	GAGAGG		< 1			
6TAG	GAGTAG		1			
6ATT	GAGATT					< 1
6TAT	GAGTAT					< 1
6C_G	GAGCGG		5			
6CTA	GAGCTA					1
6C_	GAGCGT		2	5		
6T_	GAGTGT		2			
7CC	GAGCCGT	79	8		2	2
7CT	GAG CTG T				21	36
7TC	GAG TCG T				< 1	
7TT	GAGTTGT					2
7C_A	GAGC_GTA	1	< 1			
7C_G	GAGC_GTG		28			
7AT	GAGATGT				1	
8CCA	GAG CCG TA	11	2		31	16
8CCG	GAG CCG TG	6	1		9	3
8CCC	GAG CCCG TC	< 1				
8CCT	GAG CCG CT				1	
8CTA	GAG CTG TA	< 1			18	24
8TCA	GAG TCG TA				< 1	< 1
8CTG	GAG CTG TG				8	6
8TCG	GAG TCG TG				< 1	< 1
8CTC	GAG CTG TC				1	1
8CCT	GAG CCG TT				1	< 1
8CTT	GAG CTG TT				< 1	1

8TCT	GAG TCG TT					1
8TTC	GAG TTG TC				< 1	
8TTA	GAG TTG TA				2	2
8TTT	GAG TTG TT				1	

Chapter VI

Site-Specific Conjugation of Human O^6-Alkylguanine DNA Alkyltransferase to DNA Strands Using Intrastrand Cross-Linked DNA Mimics



TT4 (*n* = 3)
TT7 (*n* = 6)

GT4 (*n* = 3)
GT7 (*n* = 6)

TG4 (*n* = 3)
TG7 (*n* = 6)

To be submitted as:

O'Flaherty DK, Wilds CJ (2016) *Bioconjugate Chemistry*

Abstract

A facile methodology is reported to conjugate human O^6-alkylguanine-DNA-alkyltransferase to the 3'-end of the DNA strand in excellent yields with short reaction times by using intrastrand cross-linked (IaCL) DNA probes. This strategy exploited the substrate specificity of hAGT to generate the desired DNA-protein covalent complex. IaCL DNA linking two thymidine residues, or linking a thymidine residue to a 2'-deoxyguanosine residue (either in a 5'→3' or 3'→5' fashion), lacking a phosphodiester linkage at the cross-linked site, were prepared using a phosphoramidite strategy followed by solid-phase synthesis. All duplexes containing the model IaCL displayed a reduction in thermal denaturation stability relative to unmodified control duplexes. The O^4-thymdine-alkylene-O^4-thymdine and the (5'→3') O^6-2'-deoxyguanosine-alkylene-O^4-thymdine IaCL DNA adducts were unrepaired by any of the AGTs tested (human AGT and *Escherichia coli* homologues, OGT and Ada-C). The (5'→3') O^4-thymidine-alkylene-O^6-2'-deoxyguanosine IaCL DNA containing butylene or heptylene tethers were efficiently repaired by the human variant, whereas Ada-C was capable of modestly repairing the heptylene IaCL adduct. The IaCL strategy has expanded the toolbox for hAGT conjugation to DNA strands. Finally, hAGT was functionalized with a fluorescently-labelled DNA sequence to demonstrate the feasibility of this method.

6.1 Introduction

Bifunctional alkylating agents have the capability of causing pronounced DNA damage by generating inter- (ICL), intrastrand (IaCL) cross-links, mono-adducts, and DNA-protein cross-links (DPC).[28,132] Extensive research has been carried out on ICL given their exceptional cytotoxicity.[54] The introduction of ICL in DNA prevents the unwinding of the DNA strands, a necessary step for vital processes such as replication and transcription. IaCL can impede cellular proliferation despite residing on only one strand.[69,268] For instance, cyclobutane pyrimidine dimers, pyrimidine (6-4) pyrimidone dimers (and their Dewar isomers) produced by UV-radiation are prime examples of non-drug-induced IaCL DNA, which can lead to carcinogenesis if left unattended.[75,274,354] Exposure of DNA to gamma radiation produces IaCL DNA between neighboring 2'-deoxyguanosine (dG) and thymidine (dT) residues, which accounts at least in part, to the cytotoxicity of this type of radiation.[273] Moreover, certain cancer chemotherapeutic regimens exploit the formation of DNA alkyl lesions to treat certain diseases. Platinum-based chemotherapeutics such as cisplatin are widely utilized for cancer treatment and exert their efficacy mainly *via* the formation of IaCL DNA lesions.[268] Understanding the impact of IaCL on DNA structure, as well as how cells process such DNA damage is an ongoing investigation. As such, it is of values to develop methodologies for the preparation of species resembling intermediates, which could provide insights on their exerted cytotoxic effects.

Minor alkylation at the O^6-dG and O^4-dT sites are observed when exposing DNA to various alkylating agents, with O^6-alkyl dG occurring in higher frequency relative to O^4-alkyl dT.[17,20,355] Regardless of their low occurrence, these lesions are chemically stable and are particularly cytotoxic, if left unrepaired, as they severely hinder DNA polymerase activity.[342]

Both lesions are presumed to cause misinsertions by disrupting the Watson-Crick H-bonding properties with their natural complements leading to misreading by DNA polymerases during DNA replication. The O^6-MedG · dT and O^4MedT · dG adopt nonwobble base pair motifs, which may account for the error-prone behaviors of DNA polymerases.[48]

O^6-Alkylguanine DNA alkyltransferases (AGTs) proteins of the direct repair pathway are found in all kingdoms of life and have shown a large variance in substrate specificity in terms of lesion type and site of alkylation.[132] They play a pivotal role in maintaining genomic integrity by removing mutagenic lesions found at the O^6-position of dG residues, and to a lesser extent the O^4-position of dT. It is estimated that about 6650 O^6-methyl dG (O^6-MedG) lesions are sufficient in causing cell death in human AGT (hAGT) deficient tumour cells.[37] The O^4-methyl dT (O^4MedT) lesion has superior mutagenic properties in mammalian cell lines, relative to O^6-MedG, since this lesion is inefficiently repaired by AGTs or the mismatch repair.[233] Moreover, the increased binding affinity of hAGT to DNA containing O^4MedT can inhibit NER pathways from successfully removing these lesions.[356]

The human homologue (hAGT) displays superior repair activity against O^6-alkyl-dG lesions, including mono-adducts such as methyl, benzyl, and 4-oxo- 4-(3-pyridyl)butyl, to name a few.[132] AGT-mediated repair has also been associated with cancer resistance mechanisms to a number of chemotherapeutic regimens, such as 1,3-bis(2-chloroethyl)-1-nitrosourea[357,358] and temozolomide[139,359], known to alkylate the O^6-position of dG.[138] Unveiling the scope of repair for AGTs could aid in the development of more potent treatment strategies. Our group has reported the repair by hAGT of ICL linking the O^6-atoms of directly opposing 2'deoxyguanosine residues, as well as in a GNC motif.[64,65] Orientation of the alkylene bridge caused little variance in the processing by hAGT. More pronounced repair was observed for ICL DNA with a

heptylene linkage, in comparison to the butylene analogue. OGT, however, was incapable of removing the ICL between 2'-deoxyguanosine residues. In addition, no repair was observed with cross-links between O^4-atoms of thymidine residues, despite the preference of OGT for alkylation at the O^4-position of dT compared to the O^6-position of dG.[150] Recently, we have demonstrated the capability of hAGT to efficiently repair IaCL DNA lacking a phosphate group at the cross-link site. As with the ICL DNA, hAGT was more efficient in repairing the heptylene linkage compared to the butylene IaCL. Repair rates were also superior for the IaCL containing the longer heptylene linker compared to the butylene. Interestingly, the IaCL of greater length was virtually entirely consumed by hAGT, whereas repair of the heptylene ICL by a larger excess of protein amounted to 57% only.[65]

Despite having the typical role of maintaining the genomic material integrity, hAGT has shown to enhance cytotoxicity of certain bifunctional alkylating agents by an efficient production of DPC adducts. 1,2-dibromoethane and 1,2,3,4-Diepoxybutane, for instance, are capable of inducing AGT-DNA cross-links.[360–362] The paradoxical cytotoxicity was attributed to the formation of either a DNA or protein intermediate, capable of readily reacting with its counterpart.[170] Given the importance of such conjugates, efficient methodologies for the production of AGT-mediated DPC warrant further investigation.

The current understanding of the AGT-mediated repair mechanism involves flipping of the damaged nucleotide into the active site, whereby an activated cysteine residue attacks the carbon next to the O^6-dG or O^4-dT.[137] DPC covalent complexes between DNA and hAGT have shown to be especially insightful in providing much of the AGT-mediated repair mechanism. The N1,O^6-ethanoxanthine modification generates a DPC after hAGT repair and crystallographic studies of this covalent complex have identified critical interactions involved in the repair

event.[137,160] However, this approach generated a DPC tethered by an ethylene linker and it would be of interest to investigate longer linker tethers given that AGTs are capable of repairing a wide scope of substrates (butylene and heptylene adducts, for instance).[65,150] Towards this end, Mcmanus and co-workers developed a strategy of generating DPCs *via* the repair of a mixed ICL between the O^6-dG and O^4-dT by hAGT, whereby repair solely occurred at the dG unit.[151] Unfortunately, the repair reaction conversions for the heptylene mixed ICL substrate was fairly low (25 %) after 16 h, using a 30-fold excess of hAGT protein, allowing for improvements in the methodology of DPC species production. No appreciable amount of DPC was observed for repair of the butylene ICL substrate analogue. While hAGT was capable of low repair activity, no other AGT variant tested positive for repair, further limiting the versatility of the ICL probe repair.

Reaction of AGTs with an IaCL DNA, lacking the phosphodiester linkage, would potentially allow for an efficient production of DPCs tethered at either the 3'- or 5'-end of the DNA strand. The overall approach would benefit from improved atom economy and greater reaction yields compared to ICL DNA. One main hurdle of this strategy is the ability of a second AGT molecule reacting with the DPC, as observed for the hAGT-mediated repair of the O^6-dG-alkylene-O^6-dG IaCL (**CHAPTER IV**) and ICL DNA probes.[64,65] To circumvent this issue, we engineered a dT-dG mixed IaCL, where AGT substrate specificity could dictate the products generated and potentially provide a versatile and an efficient method to produce DPCs tethered at either the 3'- or 5'-end of the DNA strand.

6.2 Results and Discussion

6.2.1 Synthesis and Characterization of TT, GT, and TG IaCL DNA

Bis-alkylating agents, busulfan and hepsulfam, are known to generate butylene IaCL and heptylene ICL adducts, respectively, between purine residues (**Figure S6.1**). This served as the basis for the linker lengths chosen in the study herein. Although the specific O^6-alkylene-dG or O^4-alkylene-dT IaCL lesions presented herein have not been identified, formation of other various pyrimidine-pyrimidine and pyrimidine-purine adducts in DNA have been previously identified. For instance, pyrimidine(6-4)pyrimidone photoproducts (**Figure S6.1**) are examples of lesions linking the C4 position of one pyrimidine residue with a neighbouring pyrimidine nucleotide at the 6-position, and is said to play a significant role in mutagenesis and in inducing human skin cancer.[91,363] DNA treated with mechlorethamine has revealed the formation of ICL linking the N3-atoms of mispaired dC residues.[364] γ-radiation is capable of producing IaCL DNA between adjacent purine and pyrimidine residues, regardless of their orientation (5'-GT or 5'-TG, for instance, **Figure S6.1**).[2,3,77,78,341] These examples prompted us to engineer biologically relevant IaCL mimics to help understand repair by AGTs. Moreover, generating such DNA modifications allowed us to study duplexes containing flexible IaCL lesions that interfere with the Watson-Crick face of the nucleobases.

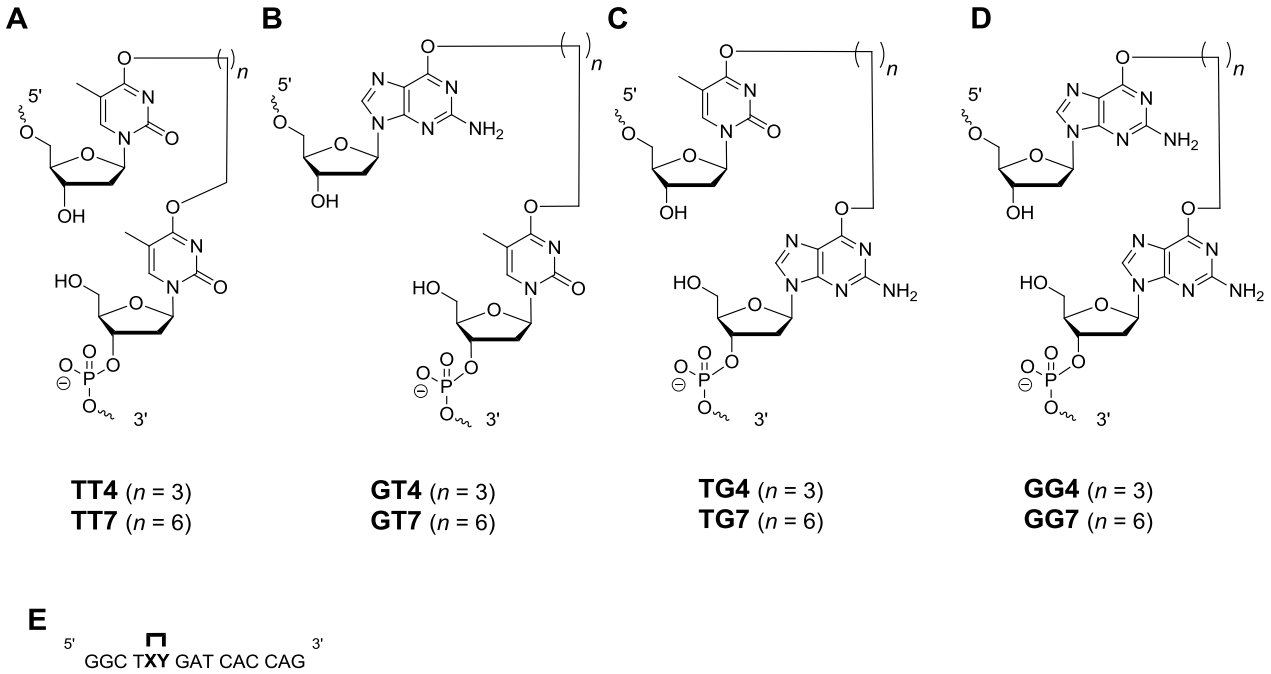
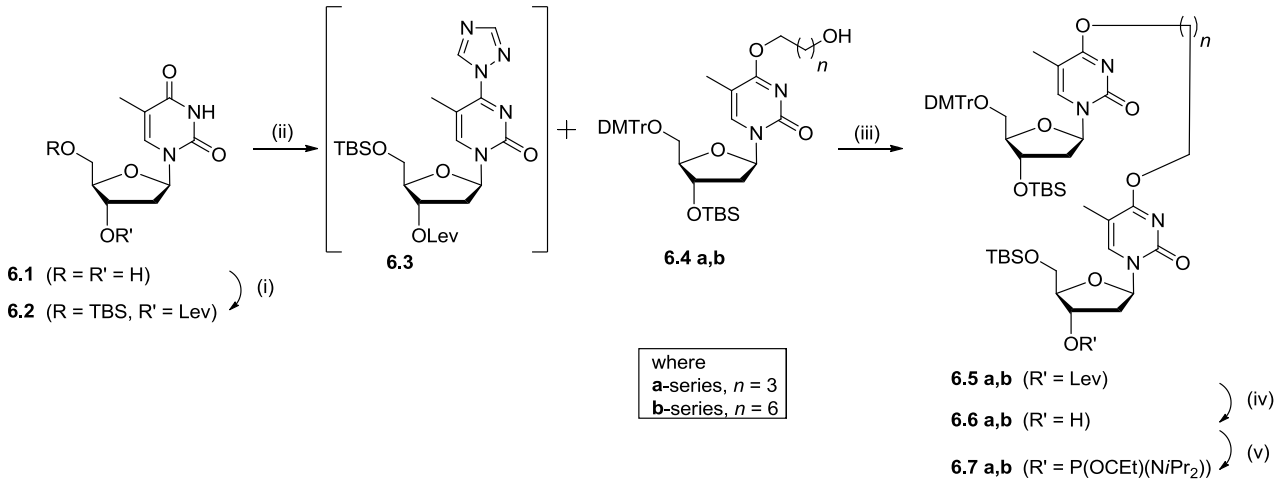


Figure 6.1: Structure of the (**A**) O^4-thymidine-alkylene-O^4-thymidine, (**B**) O^6-2'-deoxyguanosine-alkylene-O^4-thymidine, (**C**) O^4-thymidine-alkylene-O^6-2'-deoxyguanosine, (**D**) O^6-2'-deoxyguanosine-alkylene-O^6-2'-deoxyguanosine IaCL mimics lacking a phosphodiester linkage between the two residues. The probes showed in (**D**) have been investigated by our group previously (**CHAPTERS III, IV, V**)[323] (**E**) The position of the IaCL mimics in the DNA sequence is shown by the **XY** (where **XY4** has a butylene linker and **XY7** a heptylene linker).

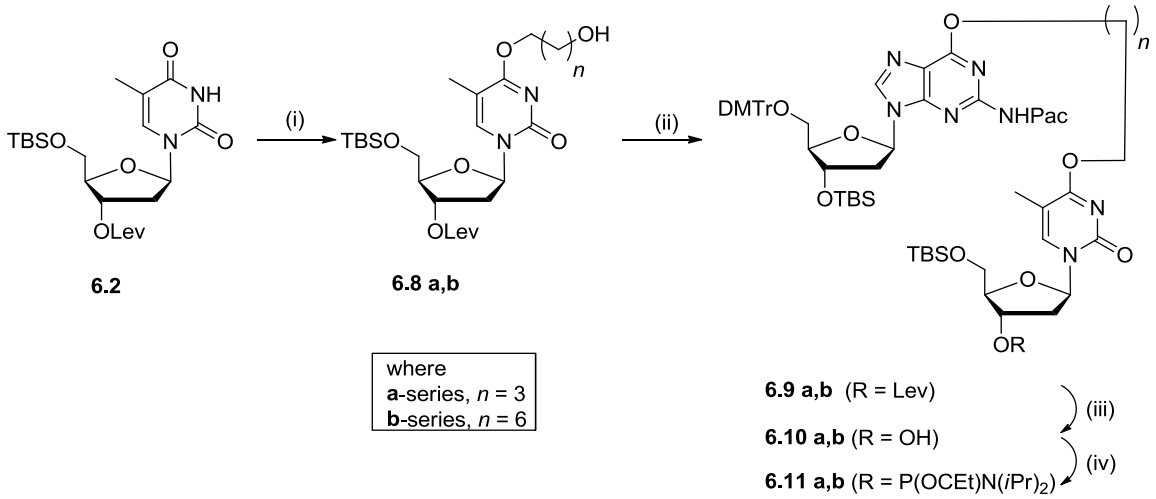
The current study aimed (i) at investigating the ability of certain AGTs to repair a variety of different IaCL DNA, with structures and corresponding labels shown in **Figure 6.1** (**6.1A-6.1C**). Since the initial report of IaCL DNA between two dG residues demonstrated efficient repair by hAGT (**Figure 6.1D**),[323] we envisioned an analogous set of probes suitable for repair by other AGTs such as OGT and Ada-C. The second aim of this study (ii) was to generate appreciable amounts of DPC species with greater versatility compared to the previously generated ICL DNA probes.

The preparation of the modified DNA scaffolds in scales and purity amenable for biochemical studies was accomplished by a combination of small molecule and automated solid-phase synthesis. The preparation of cross-linked phosphoramidites **6.7a** and **6.7b**, required for the preparation of **TT4** and **TT7**, respectively, is shown in **Scheme 6.1**. The synthesis was adapted from the procedures developed by Swann's group and our group.[150,242] Yet, the introduction of an additional protecting group was necessary to engineer **TT4** and **TT7**. The levulinyl protecting group was chosen based on its compatibility with various other nucleoside constructs.[299] Protected monomer **6.2** was prepared by selectively introducing the TBS group at the 5'-hydroxyl of commercially available thymidine (**6.1**), followed by esterifcation of the 3'-hydroxyl with levulinic acid, in good yield over two steps. The preparation of the convertible nucleoside required optimization since the previously utilized procedure for 3'-O-(tert-butyldimethylsilyl)-5'-O-(4,4'-dimethoxytrityl)-thymidine led to a significant production of side-product, presumably occurring with the carbonyl groups found in the levulinyl moiety. The production of the desired convertible nucleoside was accomplished by the addition of $POCl_3$ in two portions (1.5 equivalences followed by another 1.1 after 5 min of stirring on ice). The reaction time was not extended beyond 20-30 min to minimize side-product formation and, once worked-up, used without further purification for the subsequent step. Dimerization was achieved in the presence of DBU and a slight excess of the corresponding alcohol **6.4a** and **6.4b**, which have been synthesized using previously published procedures by our group.[150] Removal of the levulinyl group was achieved using a standard protocol,[299] with slight modifications, to furnish intermediates **6.6a** and **6.6b**.



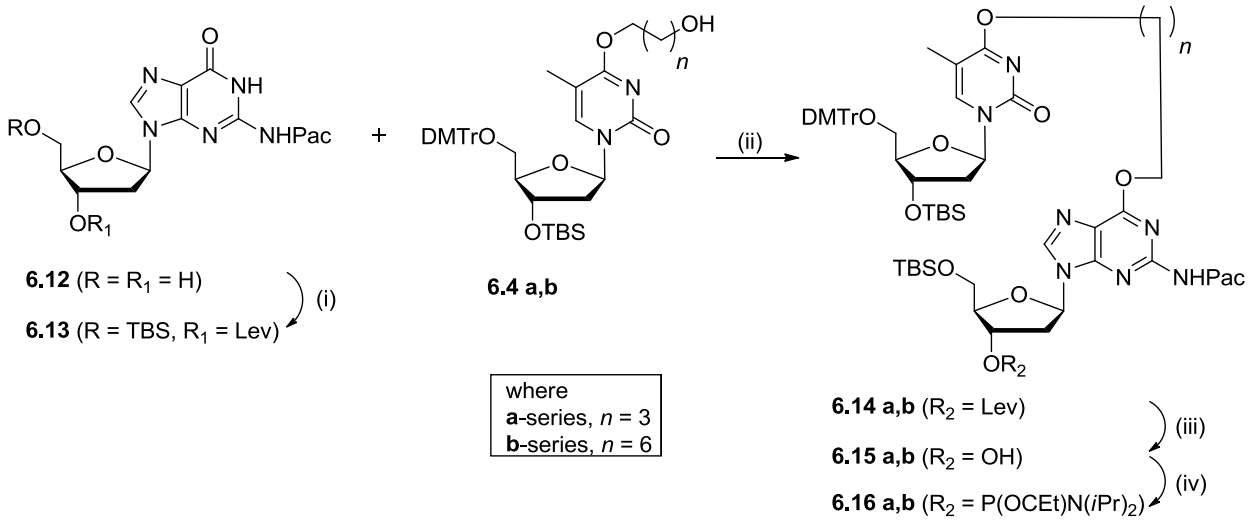
Scheme 6.1: Reagents and conditions where **a**-compounds are *n* = 3 and **b**-compounds are *n* = 6: (i) TBS-Cl (1.1 eq), Py, 0°C, 4 h, work up, Levulinic acid (2 eq), EDC (2 eq), dioxane, 21°C, 16h. (ii) Triazole (22.5 eq), NEt_3 (23eq), $POCl_3$ (2.1 eq), MeCN:DCM (1:1 v/v), 0°C, 2 h. (iii) DBU (2.3 eq), MeCN, 21°C, 16 h. (iv) N_2H_4 • H_2O (0.5 M) in pyridinium acetate buffer (1:1 v/v), 21°C, 10 min. (v) Cl-P(OCEt)(N*i*Pr_2) (1.5 eq), DIPEA (1.9 eq), THF, 21°C, 30 min.

The synthesis of cross-linked phosphoramidites **6.11a** and **6.11b**, required for the synthesis of **GT4** and **GT7**, respectively, is shown in **Scheme 6.2** and began with the preparation of hydroxybutyl and hydroxyheptyl mono-adducts **6.8a** and **6.8b**, respectively, from intermediate protected monomer **6.2**. Only a slight excess of the corresponding diol was utilized to minimize side-product formation. The Mitsunobu reaction was performed to couple **6.8a** or **6.8b** with a protected dG analogue[65] to produce dimers **6.9a** or **6.9b**. The Mitsunobu reaction was selected on the basis that it can selectively introduce alkyl groups at the O^6-atom of dG.[65,151,311,312,323] Removal of the levulinyl groups was achieved as described above to produce dimers **6.10a** and **6.10b**.



Scheme 6.2: Reagents and conditions where **a**-compounds are *n* = 3 and **b**-compounds are *n* = 6: (i) 1. Triazole (22.5 eq) NEt_3 (23 eq), $POCl_3$ (2.5 eq), DCM : MeCN (1:1 v/v), 0°C, 30 min, 2. 1,4-butanediol or 1,7-heptanediol (1.1 eq), DBU (1.2 eq), MeCN, 21°C, 16 h. (ii) 3'-*O*-(*tert*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-*N*2-phenoxyacetyl-2'-deoxyguanosine (1.1 eq), Ph_3P (2 eq), DIAD (2.5 eq), 21 °C, 18 h. (iii) N_2H_4 • H_2O (0.5 M) in pyridinium acetate buffer (1:1 v/v), 21°C, 15 min. (iv) Cl-P(OCEt)(N*i*Pr_2) (1.5 eq), DIPEA (1.9 eq), THF, 21°C, 30 min.

Phosphoramidites **6.16a** and **6.16b**, required for the preparation of **TG4** and **TG7**, was prepared in a similar fashion as described for **6.11a** and **6.11b** and shown in **Scheme 6.3**. Commercially available N2-phenoxyacetyl-2'-deoxyguanosine was converted to the fully protected monomer **6.13** using an analogous procedure to the thymidinyl intermediate **6.2**. Dimers **6.14a** or **6.14b** were prepared using the Mitsunobu coupling of **6.4a** or **6.4b** with monomer **6.13**. Finally, levulinyl deprotection of **6.14a** and **6.14b** furnished the corresponding alcohols **6.15a** and **6.15b**, respectively.



Scheme 6.3: Reagents and conditions where **a**-compounds are *n* = 3 and **b**-compounds are *n* = 6: (i) 1. TBS-Cl (1.05 eq), Py, 0°C →22°C, 16 h, 2. Levulinic acid (2 eq), EDAC • HCl (2 eq), DMAP (cat) dioxane, 22°C, 16 h. (ii) **6.4a** or **6.4b** (1.2 eq), Ph_3P (2.2 eq), DIAD (2.6 eq), dioxane, 22°C, 18 h. (iii) N_2H_4 • H_2O (0.5 M) in pyridinium acetate buffer (1:1 v/v), 21°C, 30 min. (iv) Cl-P(OCEt)(N*i*Pr$_2$) (1.5 eq), DIPEA (1.9 eq), THF, 21°C, 30 min.

All phosphoramidites (**6.7a**, **6.7b**, **6.11a**, **6.11b**, **6.16a**, and **6.16b**) were prepared in moderate to good yield using similar procedures. Purification was achieved using column chromatography, which resulted in pure products as demonstrated by ^{31}P NMR spectra. In all cases, two sharp signature peaks were observed in the region of 147-149 ppm of the ^{31}P NMR spectra, which are characteristic of phosphoramidites.[65,150,323]

The IaCL DNA assembly employed either dimer phosphoramidite **6.7a**, **6.7b**, **6.11a**, **6.11b**, **6.16a**, or **6.16b**, which differed in both their nucleobase composition and linker length (butylene compared to heptylene). Linear IaCL DNA assembly was conducted in accordance with previous published procedures on duplexes containing similar modifications.[150,151,323] The use of "fast-deprotecting" commercially available 3'-*O*-phosphoramidites was utilized due to the labile nature of O^4-alkyl-dT adducts.[150] Coupling times were extended for cross-linked amidites

to 600s as opposed to 120s for standard 3'-*O*-phosphoramidites, in order to ensure proper coupling of the dimers. Importantly, the addition of 20% (v/v) THF was used in order to solubilize amidites **6.11a**, **6.11b**, **6.16a**, or **6.16b**. IaCL DNA oligomers were deprotected and cleaved off the CPG solid support *via* treatment with an anhydrous solution of 0.05 M K_2CO_3 in methanol for 4 h under gentle rocking. Excess base was neutralized with an equimolar amount of acetic acid and lyophilized, without desalting, in a speed-vacuum concentrator. The resulting pellets were treated with triethylamine trihydrofluride to remove the remnant silyl protecting groups at 40^{o}C for 16h. Initial sonication was required to break the pellet apart for two rounds of 15s. Precipitation of IaCL DNA was achieved with cold *n*-butanol, as opposed to methanol, and allowed to cool at -20 ^{o}C for 10 min. IaCL oligomers proved slightly soluble in methanol, which was generally more-pronounced for heptylene adducts compared to the butylene IaCL DNA presumably because of the increase hydrophobicity of longer alkylene linkers. Also, solubility of the oligomers increased with shorter sequences (tested for O^4-dT-alkylene-O^4-dT IaCL DNA with sequences of 12 nucleotides in length), especially in cases where the modification was engineered in the central region of the sequence (data not shown). The precipitated material was washed with another cold aliquot of *n*-butanol before purification by SAX-HPLC. Characterization of IaCL DNA by ESI-MS revealed deconvoluted masses in agreement with all expected masses (**Supporting Figures S6.2-S6.7**)

6.2.2 UV thermal denaturation, circular dichroism, and molecular modeling studies of IaCL DNA

The impact of the different IaCL on DNA duplex stability was assessed by UV thermal denaturation experiments. Results of the experiments are summarized in **Figure 6.2** (see **Supporting Figures S6.8-S6.10** for all T_m curves). As previously observed for the O^6-dG-

alkylene-O^6-dG IaCL DNA, a general decrease in T_m value of approximately 16-19 ^oC relative to the unmodified control was observed in all cases.[323] The linker size had little effect on the thermal denaturation profiles across the various duplexes containing the IaCL DNA. All curves displayed a characteristic sigmoidal transition. Interestingly, we suspect the lack of the phosphodiester linkage at the cross-linked site has minimal influence on the duplex stability as observed in the case of the O^6-dG-alkylene-O^6-dG IaCL (**CHAPTER IV**).[323] Yet, the decreased stability is likely a result of H-bonding disruption between the alkylated dT (or dG) and its base-paired 2'-deoxyadenosine (or 2'-deoxycytidines) counterpart. In addition, the decrease in T_m may have been a consequence of the bulky alkylene linkers causing local or global perturbations.

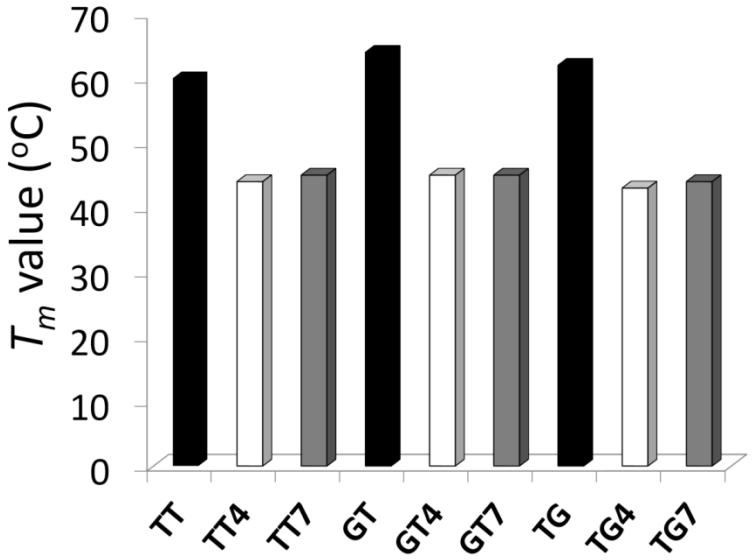


Figure 6.2. T_m values (^oC) of duplexes containing **TT**, **TT4**, **TT7**, **GT**, **GT4**, **GT7**, **TG**, **TG4**, **TG7**. Bar in black, colorless and grey represent the unmodified controls, butylene adducts, and heptylene adducts, respectively. Hyperchromicity change (A_{260}) *versus* temperature (^oC) profiles are shown in **Supporting Figures S8-S10**

Circular dichroism (CD) spectroscopy was carried out to evaluate the global structural influence of the IaCL mimics on the DNA duplexes. CD profiles of duplexes containing **TG4**, **TG7** and the unmodified (**TG**) control is provided in **Figure 6.3** as an example (see **Figure**

S6.11-S6.12 for other CD spectra). In all cases, CD profiles displayed characteristic signatures pertaining to the B-DNA family with maxima near 280 nm, cross-overs near 255 nm and minima near 240 nm. More importantly, all the CD profiles of duplexes containing the IaCL were similar to the corresponding unmodified DNA duplexes, which suggested that the presence of the IaCL modification minimally perturbed the DNA duplex, as previously reported.[323]

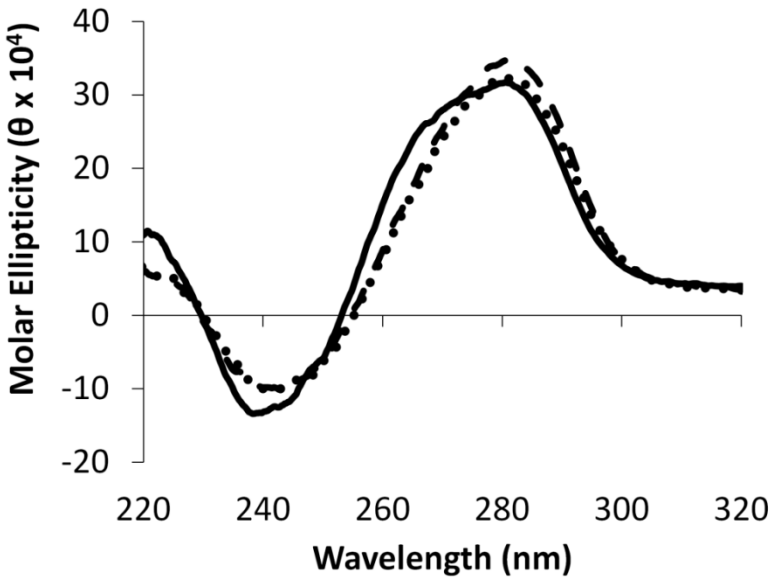


Figure 6.3. Circular dichroism spectra of IaCL duplexes **TG4** (•••), **TG7** (— — —) and unmodified control **TG** DNA (—).

Basic molecular modeling was conducted on duplexes containing the different IaCL adducts or the unmodified control nucleotides using AMBER force field (see **Figures S6.13-S6.15**). Duplexes containing the various IaCLs displayed the alkylene linkers residing in the major groove channel. The presence of the alkylene tethers seemed to generally disrupt the base pairing between the alkylated nucleobases' complements, relative to the unmodified controls which may account for the decrease in T_m observed. Yet, molecular models suggested minimal global distortions induced by the presence of the IaCL tether, which was in agreement with observation from CD spectroscopy.

6.2.3 AGT repair of IaCL DNA substrates

Four different AGT variants, hAGT, OGT, Ada-C, and an S134P OGT variant were tested for their ability to repair the DNA duplexes containing the different IaCL (**TT4**, **TT7**, **GT4**, **GT7**, **TG4**, and **TG7**). The repair assays were conducted as described in a prior report[323], whereby the successful repair of the IaCL mimic generates a smaller DNA repair product (**RP**) species and a DPC median product (**MP**) (as shown in **Figure 6.4**). Denaturing PAGE experiments are utilized to resolve the different reaction species after reaction times of 16h at 37^{o}C.

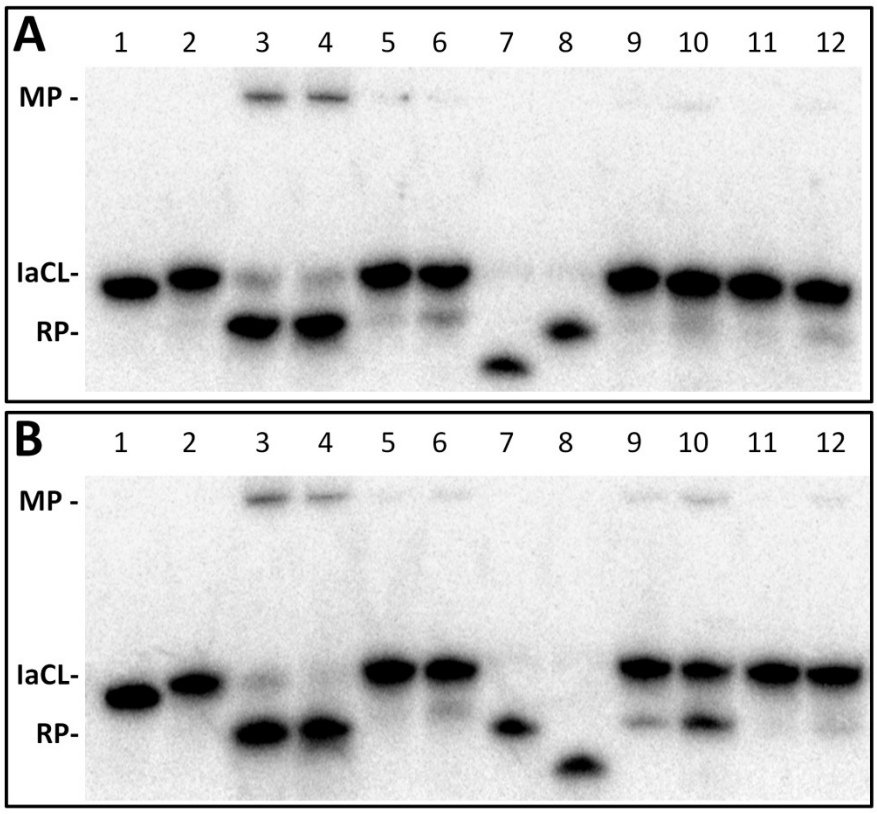


Figure 6.4. Repair of (**A**) **TG4** and (**B**) **TG7** by hAGT, OGT, Ada-C, and OGT S134P. Denaturing PAGE of repair reactions as described in the experimental section with varying protein concentration (either 5- or 30-fold protein). Panel **A**; Lane 1, 2 pmol control unmodified **TG** DNA; lane 2, 2 pmol **TG4**; lane3 **TG4** + 10 pmol hAGT; lane4 **TG4** + 60 pmol hAGT; lane5 **TG4** + 10 pmol OGT; lane6 **TG4** + 60 pmol OGT; lane7 unmodified 5-mer control DNA 5'-pGGCTT; lane8 unmodified 10-mer control 5'-pGGC ATC ACC AG; lane9 **TG4** + 10 pmol

Ada-C; lane10 **TG4** + 60 pmol Ada-C; lane11 **TG4** + 10 pmol OGT S134P; lane12 **TG4** + 60 pmol OGT S134P. Panel **B**; identical to panel **A** except with **TG7** instead of **TG4**.

The radioactivity-based repair assay revealed no appreciable repair (< 20 % repair) of **TT4**, **TT7**, **GT4**, and **GT7** at both five-molar equivalence and thirty-molar equivalence of AGT (**Supporting Figures S6.16-S6.17**). Lack of repair by Ada-C was not surprising as this protein is generally inactive to adducts larger than a methyl group at the O^4-position and O^6-position of dT and dG, respectively.[146,365] It was, however, surprising to observe the lack of OGT-mediated repair of these IaCL DNA, particularly **TT4** and **TT7**, given that OGT is proficient in repairing larger mono-adducts, such as hydroxyl-*n*-butyl and hydroxy-*n*-heptyl, at the O^4-position of dT.[150] The OGT S134P variant,[149] which is said to contains a larger active site, was also incapable of IaCL repair. Moreover, hAGT proved inefficient at repairing duplexes containing **GT4** and **GT7** despite being capable of efficiently repairing **GG4** and **GG7** (see **Figure 6.1D**). Our group has also previously demonstrated repair by hAGT toward an O^6-dG-heptylene-O^4-dT ICL occurring solely at the O^6-atom of dG, with no repair event at the O^4-atom of dT.[151] The lack of hAGT repair towards **GT4** and **GT7** could thus suggest that the 3'-phosphodiester linkage is playing a critical role in the repair at the dG moiety of the IaCL. This hypothesis is further supported by the efficient repair of **TG4** and **TG7** by hAGT depicted in **Figure 6.4**, which both possess the 3'-phosphodiester linkage at the dG residue (**Figure 6.1**). Crystallographic structures of covalent complexes between modified DNA and hAGT, as well as computational approaches, have revealed a critical interaction of Tyr114 with the 3'-phosphodiester in order for efficient repair of O^6-alkyl-dG.[137,168] We suspect that deleting this functional group severely cripples hAGT repair, at least for IaCL-type lesions.

Repair profiles of **TG4** and **TG7** are illustrated in the denaturing gel shown in **Figure 6.4**. Only hAGT was capable of **TG4** repair (lanes 3 and 4 in **Figure 6.4A**) with almost complete substrate consumption at five-fold protein-DNA ratio (10 pmol protein and 2 pmol DNA). As suspected, hAGT was also capable of efficiently repairing the heptylene analogue **TG7** (lanes 3 and 4 in **Figure 6.4B**). We were surprised to observe repair of **TG7** by Ada-C at thirty-fold protein equivalence (lane 10 in **Figure 6.4B**). The results revealed repair occurring only at the O^6-alkylene-dG part of the IaCL, due to the production of the larger DNA **RP** (**Figure 6.4B**). To the extent of our knowledge, this marks the first substrate adduct larger than O^6-ethyl-dG capable of undergoing repair by Ada-C,[146] albeit at a mediocre efficiency (30 % substrate consumption reactions of 16h at 37^{o}C). No repair at thymidinyl moiety of **TG7** was observed, which was in agreement with a lack of repair of larger O^4-hydroxyalkylene-dT mono-adducts Ada-C.[150] The IaCL strategy may find applications in efficiently generating DNA-Ada cross-linked species that have otherwise been impossible using the ICL avenue.[151] Our approach would, however, require optimization given the large protein excess required to initiate repair of **TG7**.

Time course assays were used to gain insights on the kinetics of the reaction with AGTs and the results are summarized in **Figure 6.5** (see **Supporting Figure S6.18** for gels). Substrate **TG4** (**Figure 6.5A**) was processed by hAGT significantly slower compared to **TG7** (**Figure 6.5B**), which was in agreement with repair trends observed with other IaCL[323] DNA as well as ICL[65] DNA repair by hAGT. Approximately 25 % of **TG4** IaCL substrate remained after a reaction period of 60 min with five-fold excess hAGT, whereas similar levels were reached after only 10 min for **TG7**. Interestingly, the **TG4** IaCL substrate was repaired with enhanced rates in comparison to the **GG4** IaCL DNA. For instance, **TG4** reached reaction plateau (20 % substrate remaining) after approximately 90 min, whereas it took hAGT approximately 480 min to reach

similar levels for **GG4**. [323] This trend was not shared for the heptylene analogues (**TG7** versus **GG7**), with repair levels reaching approximately 20% after 15 min. This observation may be a consequence of the energetic requirement involved in rotating the α-carbon to the O^6-alkylted-dG in a favorable fashion in order to initiate nucleophilic attack by the Cys145 thiolate anion. There are a variety of different accounts that suggest orientation of the alkyl group at the O6-atom of dG influences the repair by hAGT. Perhaps the alkylated dG nucleobase is more easily rotated in the case of **TG4** compared to **GG4**. Pyrimidines are capable of adopting a variety of different conformations, which may confer added flexbility.[366] The added flexibility may aid in the repair of the more constrained butylene IaCL DNA, consistent with repair kinetics of **TG4** *versus* **GG4**. The same effect may not necessarily cause the same benefit for the longer heptylene tether, as longer linkers are inherently more flexible (due to the additional rotatable bonds groups). This may explain why similar hAGT repair rates are observed for **TG7** and **GG7**. We have previously solved the NMR solution structure of an O^4-2`-deoxuridine-heptylene-$O^4$-2`-deoxyuridine ICL DNA duplex, which revealed that the α-carbon of the O^4-alkylated pyrimidine adopted an unexpected *E* conformation between the α-carbon and theC4-O^4 bond. Perhaps such a conformation is contributing to the effects observed with **TG4** *versus* **GG4**. We were please to observe the formation of a median product (**MP**) in the case of **TG4** and **TG7**, similar to observations previously made for **GG4** and **GG7**. The presence of such species would suggest that reaction conditions could be optimized for the efficient production of DPCs.

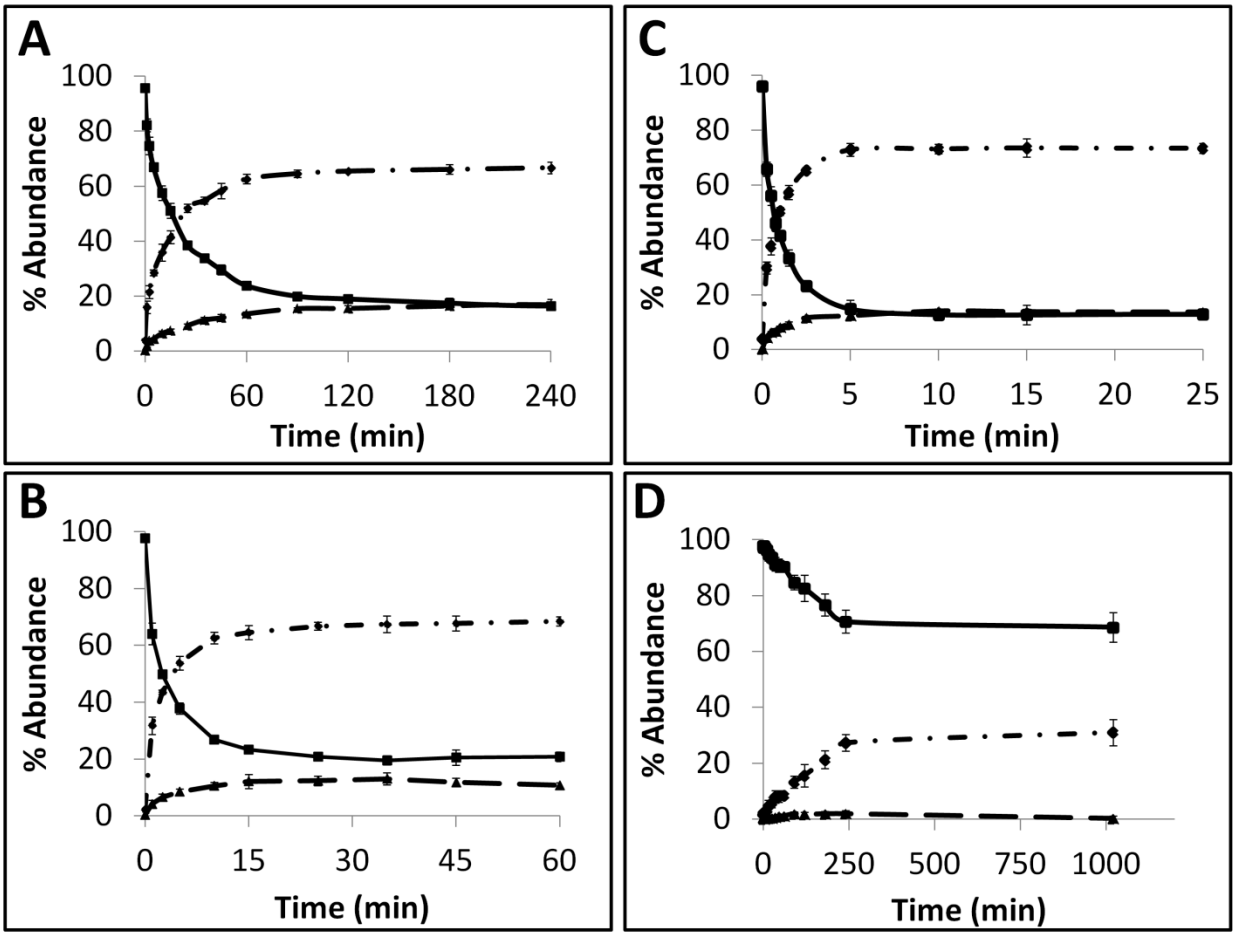


Figure 6.5. Time course repair assay plots of **TG4** (**A**) and **TG7** (**B**) by hAGT, at five molar equivalence protein, displaying significant faster repair of **TG7** by hAGT. Time course repair assay of **TG7** by hAGT (**C**) and Ada-C (**D**) at thirty molar equivalence protein, displaying significant faster repair of **TG7** by hAGT compared to Ada-C. Graphical illustrations display abundances [%] of MP (———), RP (—•—)and substrate (—) over time (min).

Time course assays revealed that Ada-C repaired **TG7** in a much slower fashion relative to hAGT (**Figure 6.5C** and **Figure 6.5D**, respectively, see **Supporting Figure S6.19** for gels). At thirty-fold protein equivalences, **TG7** levels reached about 70 % after reaction times of 4 h by Ada-C, whereas similar amounts were detected after only 15 s with hAGT. These results were not surprising bearing in mind the narrow substrate scope for Ada-C (O^6-methyl-dG, O^6-ethyl-dG, and O^4-methyl-dT)[146] relative to the wide substrate range previously reported for

hAGT.[132,323] Unfortunately, no considerable amounts of **MP** were detected for Ada-C mediated repair (**Figure 6.4B** and **Figure 6.5D**).

Studies from our group and others have suggested that repair by AGTs may be dependent on conformations of the alkyl group in order for efficient group transfer.[150,151,228,321] This effect may not be as pronounced in freely rotating adducts or smaller appendages that have a narrower range of orientations. The DNA-protein complex must be such that the α-carbon of the DNA lesion resides in a favorable orientation to allow attack by the cysteine thiolate anion.[137] It should be noted, however, that substrate discrimination due to conformational factors may be pertaining to more than simply the alkyl appendage found in the DNA oligomer. Notably, a prime example is that of O^6-methyl-riboguanosine, which is a poor substrate of hAGT when incorporated into DNA oligomers.[367] The culprit was accredited to a potential clash between the 2'-hydroxyl group and an amino acid near the active site of hAGT.[137] The lack of the phosphodiester group in our IaCL mimic presumably confers added flexibility to the nucleobases, sugar moiety and the alkylene appendage, which may explain the proficient repair observed by hAGT and mediocre repair by Ada-C. Yet, it is clear that the lack of a 3'-phosphodiester linkage prevents efficient repair by hAGT of O^6-alkylene-dG (such as those found in **GT4** and **GT7**). This may simply be because the Tyr114 residue is incapable of efficiently rotating the damaged nucleobase given the lack of the 3'-phosphodeister group. Yet, it is unclear if the alkylene linker orientation, presumable pointing towards the 3'-end of the sequence, in **GT4** and **GT7** adopts an adverse conformation.

6.2.4 Identification of GG4, GG7, TG4 and TG7 repair products by SDS-PAGE and ESI-MS

The radioactivity-based assay described above mainly monitors DNA species, thus prompting us to monitor proteinaceous species by SDS-PAGE followed by mass spectrometric characterization. We conducted repair reactions of AGT protein (600 pmol) and IaCL DNA (600 pmol), either single-stranded or part of a duplex. **GG4** and **GG7** were utilized as positive controls. As previously demonstrated for **GG4** and **GG7** (**CHAPTER IV**), hAGT-mediated repair of **TG4** and **TG7** was not significantly affected by the lack, or presence, of the complementary DNA strand (data not shown). The Coomassie blue-stained gel is illustrated in **Figure 6.6** and displays almost quantitative conversion of hAGT to a slower migrating species (presumably the DPC species) when reacted with single-stranded **GG4**, **GG7**, **TG4**, and **TG7** (lane 2, 3, 5, and 6, respectively) for 16h at 37°C. Furthermore, we were please to observe excellent reaction conversions when shortening the reaction times to only one hour (**Figure S6.20**). DPC formation occurred in a more efficient fashion compared to AGT-cross-linking using the N1,O^6-ethano-2′-deoxyxanthosine modification, which suffered from slower reaction kinetics under higher protein-to-DNA ratio.[160] No appreciable amount of DPC, however, was detected by reacting Ada-C (600 pmol) with **TG7** (600 pmol) as shown in lane 8 of **Figure 6.6**.

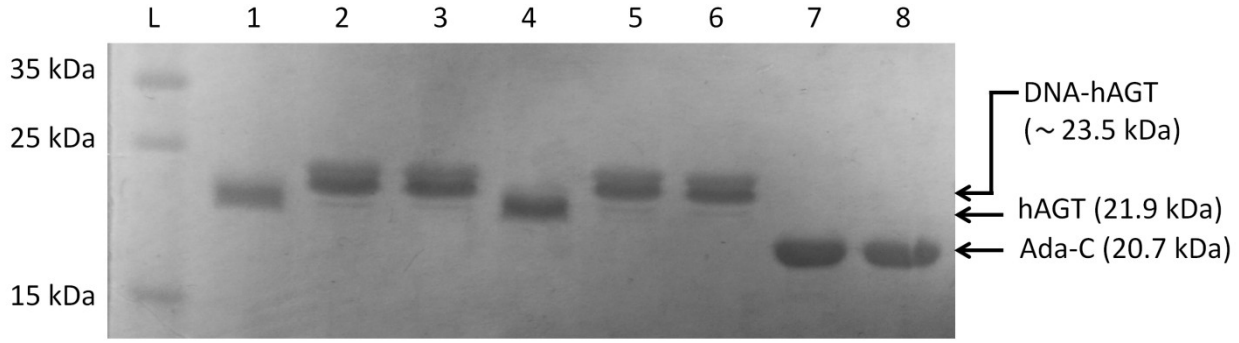


Figure 6.6. 15% SDS-PAGE (Coomassie stained) of hAGT mediated repair of **GG4**, **GG7**, **TG4**, and **TG7**, and Ada-C mediated repair of **TG7**. Repair of 600 pmol single-stranded IaCL DNA by 600 pmol protein for 16 h at 37 °C: lane L, unstained protein molecular weight marker; lane 1 unmodified control (**GG**) DNA + hAGT; lane 2 **GG4** + hAGT reaction; lane 3 **GG7** + hAGT reaction; lane 4 unmodified control (**TG**) DNA + hAGT; lane 5 **TG4** + hAGT reaction;

lane 6 **TG7** + hAGT reaction; lane 7 unmodified control (**TG**) + Ada-C; lane 8 **TG7** + Ada-C. Note that there is no appreciable amounts of DNA-Ada-C cross-link formation (lane 8).

The DPC species were characterized by ESI-MS and the spectra of complexes produced from reacting hAGT with **GG4**, **GG7**, **TG4**, and **TG7** for 30 min at 37^{o}C are shown in **Figure 6.7** (Panels **A**, **B**, **C**, and **D**, respectively). The spectra of butylene products display a minor peak attributed to unreacted hAGT (21 876 Da) and major peak corresponding to the butylene DPC. The heptylene analogues, on the other hand, reveal the presence of only one major peak attributed to the heptylene DPC species. It should be noted that the absence of another peak with higher mass values (~ 25 kDa) precludes the efficient formation of the other DPC reaction products (*i.e.* hAGT reacting with the 5'-residue of the IaCL). These results support the importance of the 3'-phosphodiester linkage in the repair mechanism by hAGT, as previously hypothesized.[137,168] Moreover, these results are consistent with previous studies that demonstrate repair of an O^6-dG-alkylene-O^6-dG ICL DNA, but no repair towards an O^4-dT-alkylene-O^4-dT ICL DNA, by hAGT.[64,65,150] The mixed O^6-dG-heptylene-O^4-dT ICL DNA was only repaired by hAGT at the dG residue[151] similarly to **TG4** and **TG7,** but not **GT4** and **GT7** which both remained inert. The repair by hAGT of the novel modification **TG4** and **TG7** is illustrated in **Figure 6.8**.

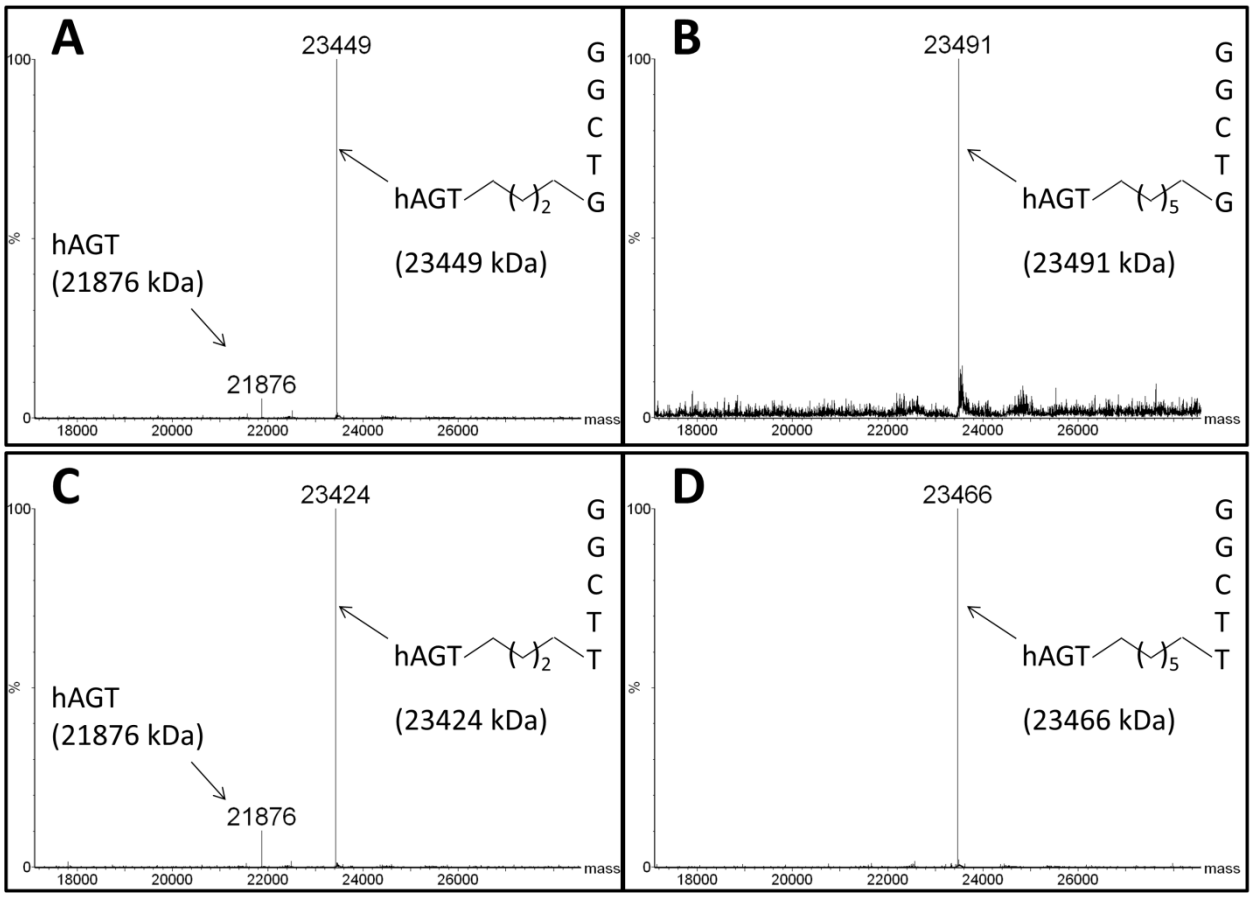


Figure 6.7: Identification of hAGT-DNA covalent complex by ESI-MS. Product identified from reaction of hAGT with single-stranded (**A**) **GG4**, (**B**) **GG7**, (**C**) **TG4**, (**D**) **TG7**. Values in brackets represent the calculated masses.

Cellular DPC production has become an increase concern due to their diversity and elevated cytotoxicity.[100] Our IaCL approach has demonstrated a facile methodology to generate hAGT cross-linked to the DNA with excellent reaction conversions, particularly compared to repair of ICL adducts. We have expanded the versatility of the current method by the introduction of different size alkylene tethers in DPCs. In addition, hAGT cross-linking sites were possible to either the O^6-atom of dG or the O^4-atom of dT, a commodity possible due to the necessity of the 3'-phosphodiester group for efficient hAGT-mediated repair of O^6-alkyl-dG adducts. There are no current methodologies, to the extent of our knowledge, to efficiently

generate a DPC between hAGT and an O^6-alkyl-dG residue. The N1,O6-ethanoxanthosine avenue generates an ethylene bridge between Cys145 and the N1-atom of the xanthosine nucleotide.[137,160] As proof-of-concept, we synthesized an oligomer with fluorescein tags at the 5'- and 3'-end of the **GG4** or **GG7** DNA oligomer. Cross-linking experiments yielded a fluorescently labelled DPC species, generated with similar yields (**Figure S6.21**).

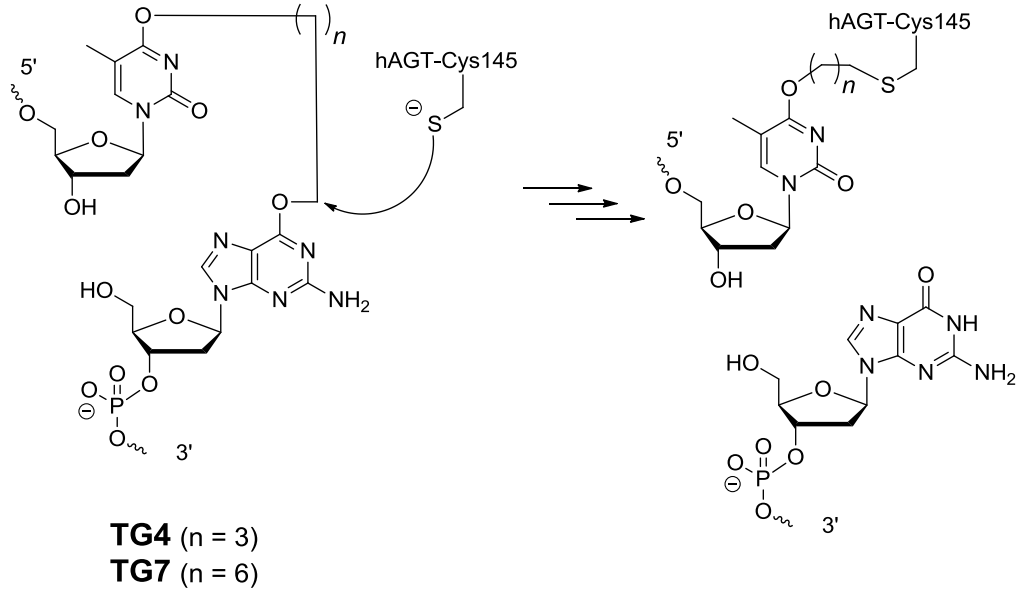


Figure 6.8: hAGT-mediated repair of **TG4** and **TG7** IaCL.

The IaCL may find applications in molecular nanotechnology devices as novel functional irreversible switches. There has been growing interest in exploiting the activity of AGTs for applications such as nanosensors in DNA origami,[307] protein tagging,[178] and molecular imaging[176] to name a few. The proposed IaCL switch, once inserted into a system, would benefit from orthogonality with virtually no response to other switching mechanisms like pH, UV-radiation and toehold-mediated mechanisms. Moreover, we have shown that the kinetics of **TG** IaCL cleavage are fast and reliable in producing only one DNA-protein complex. Our IaCL may

be an efficient avenue to conjugate a hAGT molecule to a device or cleaving a particular DNA segment part of the system, all depending on the directionality of the **TG** IaCL.

6.3 Conclusion

The introduction of a variety of different IaCL DNA mimics were prepared using a straightforward synthetic methodology in purity and scales sufficient for biophysical and biochemical studies. Our model IaCL adducts all decreased the thermal stability of duplexes by about 15-19 ^{o}C, relative to unmodified control, as verified by UV thermal denaturation experiments. Circular dichroism spectroscopy revealed minimal global perturbation of the DNA structure relative to control duplexes. A series of well-defined hAGT-DNA covalent complexes were generated *via* either a butylene or heptylene linker, to either an O^4-alkylated thymidinyl or an O^6-alkylated 2'-deoxyguanidyl residues. We identified a novel substrate for Ada-C (**TG7**), with no evident amount of DPC observed using the assays described herein. Unfortunately, we were unable to generate the 5'-tethered DPC complexes using the engineered **GT4** and **GT7** probes with the AGTs tested. The reason for the substrate discrimination by hAGT is hypothesized to arise from (1) the lack of the 3'-phosphodiester linkage of the O^6-alkylated dG residue or (2) the α-carbon adopting a non-favourable conformation hence preventing successful repair by hAGT. Nonetheless, the use of IaCL DNA substrates supplies an attractive method for the production of site-specific hAGT-DNA covalent complex with increased versatility compared to the ICL DNA avenue.

6.4 Acknowledgements

We are grateful to Dr. Anthony E. Pegg (Pennsylvania State University) for the plasmids encoding the wild-type hAGT, OGT and Ada-C genes. We are also grateful to Dr. Francis

McManus, Lauralicia Sacre, and Dr. Meena Kathiresan for helpful discussions concerning the repair assays and protein over-expression. We thank Dr. Anne Noronha for helpful discussions and Alain Tessier for assistance concerning LC-MS experiments. This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC), the Canada Foundation for Innovation (CFI) and the Canada Research Chair (CRC) program. D.K.O. is the recipient of a Canada Graduate Scholarship (CGS) from NSERC and a Graduate Award from the Groupe de Recherche Axé sur la Structure des Protéines (GRASP).

6.5 Supporting Information

6.5.1 Supporting Methods

6.5.1.1 General

*N*2-phenoxyacetyl-2'-deoxyguanosine , 5'-*O*-(4,4'-dimethoxytrityl)-*N*2-phenoxyacetyl-2'-deoxyguanosine, 5'-*O*-(4,4'-dimethoxytrityl)-thymidine, thymidine, and N,N-diisopropylaminocyanoethylphosphonamidic chloride were purchased from ChemGenes, Inc. (Wilmington, MA). 3'-*O*-(*tert*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-*N*2-phenoxyacetyl-2'-deoxyguanosine and was prepared according to previously reported procedures.[65] Mono-adduct **6.4a** and **6.4b** were prepared accordingly to previous published procedures.[150] 5'-*O*-Dimethoxytrityl-2'-deoxyribonucleoside-3'-*O*-(β-cyanoethyl-N,N-diisopropyl)phosphoramidites and protected 2'-deoxyribonucleoside–CPG supports were purchased from Glen Research (Sterling, Virginia). 6-FAM phosphoramidites and 6-FAM lcaa CPG (DMT) 500Å was purchased from ChemGenes (Wilmington, MA) and used according to the manufacturer's protocols. All other chemicals and solvents were purchased from the Aldrich Chemical Company (Milwaukee, WI) or EMD Chemicals Inc. (Gibbstown, NJ). Flash column

chromatography was performed using silica gel 60 (230–400 mesh) purchased from Silicycle (Quebec City, QC). Thin layer chromatography (TLC) was carried out with precoated TLC plates (Merck, Kieselgel 60 F_{254}, 0.25 mm) purchased from EMD Chemicals Inc. (Gibbstown, NJ). NMR spectra were recorded on a Varian 500 MHz NMR spectrometer at room temperature. ^{1}H NMR spectra were recorded at a frequency of 500.0 MHz and chemical shifts were reported in parts per million (ppm) downfield from tetramethylsilane. ^{13}C NMR spectra (^{1}H decoupled) were recorded at a frequency of 125.7 MHz and chemical shifts were reported in ppm with tetramethylsilane as a reference. ^{31}P NMR spectra (^{1}H decoupled) were recorded at a frequency of 202.3 MHz and chemical shifts were reported in ppm with H_3PO_4 used as an external standard. High resolution mass spectrometry of modified nucleosides were obtained using an 7T-LTQ FT ICR instrument (Thermo Scientific), at the Concordia University Centre for Structural and Functional Genomic. The mass spectrometer was operated in full scan, positive ion detection mode. ESI mass spectra for oligonucleotides were obtained at the Concordia University Centre for Biological Applications of Mass Spectrometry (CBAMS) using a Micromass Qtof2 mass spectrometer (Waters) equipped with a nanospray ion source. The mass spectrometer was operated in full scan, negative ion detection mode. Ampicillin, isopropyl β-D-thiogalactopyranoside (IPTG), and most other biochemical reagents as well as polyacrylamide gel materials were purchased from Bioshop Canada Inc. (Burlington, ON). Ni-NTA Superflow Resin was purchased from Qiagen (Mississauga, ON). Complete, Mini, EDTA-free Protease Inhibitor Cocktail Tablets were obtained from Roche (Laval, QC) Nitro- cellulose filters (0.20 μm) were obtained from Millipore. XL-10 Gold and Bl21(DE3) *E. coli* cells were obtained from Stratagene (Cedar Creek, TX). T4 polynucleotide kinase (PNK) was obtained from Fermentas (Burlington, ON). [γ-^{32}P]ATP was purchased from PerkinElmer (Woodbridge, ON). Phusion

Polymerase was obtained from New England Biolabs (Ipswich, MA). DNA primers for site directed mutagenesis and cloning were purchased from Biocorp (Montreal, QC).

6.5.1.2 Chemical synthesis of nucleosides

The synthesis of phosphoramidites **6.7a** and **6.7b** are shown in **Scheme 6.1**

5'-O-tert-butyldimethylsilyl-3'-O-levulinyl- thymidine (**6.2**)

Thymidine (2.0 g, 8.3 mmol) was co-evaporated in vacuo with anhydrous Py (2 x 5 mL). To a solution of dried thymidine in Py (50 mL) at 0^{o}C was added TBS-Cl (1.3 g, 8.6 mmol) and allowed to stir. After 3h, the solvent was removed in vacuo and the crude was taken up in DCM (50 mL) forming a white precipitate in suspension. The content was washed with aqueous $NaHCO_3$ (3% w/ v, 2 x 50 mL) and once with brine (50 mL), which removed the precipitate, followed by drying over anhydrous Na_2SO_4 (~ 4 g). The solvent was removed in vacuo and analysis by TLC revealed that the crude product was of sufficient purity to proceed to the subsequent step. To a solution of the 5'-*O* mono-silylated thymidine adduct (~2.9 g, ~8.1 mmol), EDC (3.1 g, 16 mmol) and DMAP (cat) in dioxane (64 mL) was added levulinic acid (1.9 g, 16 mmol) and allowed to stir. After 4h, the solvent was removed in vacuo and the crude was taken up in DCM (50 mL), washed with aqueous $NaHCO_3$ (3% w/v, 2 x 50 mL) and once with brine (50 mL), and dried over anhydrous Na_2SO_4 (~ 4 g). The solvent was removed in vacuo and the product was purified *via* flash column chromatography using MeOH : DCM (1% to 2.5%) as eluent to afford 3.2 g (85% over two steps) of a slightly beige powder. R_f (SiO_2 TLC): 0.53 MeOH : DCM (5%). $\lambda_{max(MeCN)}$ = 264 nm. ^{1}H NMR (500MHz, $CDCl_3$, ppm): 9.38 (s, 1H, NH), 7.54 (d, 1H, H6, J = 1 Hz), 6.33 (dd, 1H, H1', J = 9, 6 Hz), 5.24 (m, 1H, H3'), 4.08 (m, 1H, H4'), 2.75 (m, 2H, CH_2CO), 2.58 (m, 2H, $OOCCH_2$), 2.39 (m, 1H, H2'), 2.18 (s, 3H, $COCH_3$),

2.08 (m, 1H, H2''), 1.90 (d, 3H, $ArCH_3$, J = 1 Hz), 0.91 (s, 9H, $SiC(CH_3)_3$), 0.109 (s, 3H, $SiCH_3$), 0.105 (s, 3H, $SiCH_3$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 206.3, 205.7, 176.8, 172.4, 168.4, 164.0, 150.5, 135.2, 111.1, 85.4, 84.7, 75.6, 63.6, 38.0, 37.7, 29.7, 27.9, 25.9, 18.3, 12.4, , -5.42, -5.52. IR (thin film); ν_{max} (cm^{-1}) = 3066, 2955, 2929, 2856, 1715, 1471, 1276, 1257, 1073, 1005, 837. HRMS (ESI-MS) *m/z* calculated for $C_{21}H_{35}N_2O_7Si^+$ 455.2208: found 455.2211 $[M+H]^+$

1-{O^4-[3'-O-(tert-Butyldimethylsilyl)-5'-O-(4,4'-dimethoxytrityl)-thymidinyl]}-4-{O^4-[5'-O-(tert-butyldimethylsilyl)-3'-O- (levulinyl)-thymidinyl]}-butane (**6.5a**)

To a solution of compound **6.2** (1.2 g, 2.6 mmol) , triazole (4.0 g, 58 mmol), NEt_3 (8.3 mL, 59 mmol) in MeCN : DCM (32mL, 1/1 v/v) stirring at 0^{o}C for 10 min was added $POCl_3$ dropwise (0.36 mL, 3.9 mmol). After 5 min, another portion of $POCl_3$ (0.27 mL, 2.9 mmol) was added which resulted in the formation of a precipitate. After 20 min, the solvent was removed in vacuo and the crude was taken up in DCM (50 mL). The content was washed with aqueous $NaHCO_3$ (3% w/ v, 2 x 50 mL) and once with brine (50 mL), followed by drying over anhydrous Na_2SO_4 (~ 4 g). The solvent was removed in vacuo and crude intermediate **6.3** was used directly for the subsequent dimerization reaction. To a solution of **6.3** (~ 1.0 g, ~2.1 mmol) and **6.4a** (1.7 g, 2.3 mmol) in MeCN (15 mL) was added DBU (0.92 mL, 6.2 mmol) and allowed to stir at 21 ^{o}C. After 20h, the solvent was removed in vacuo and the crude was taken up in DCM (50 mL). The content was washed with aqueous $NaHCO_3$ (3% w/ v, 2 x 50 mL), dried over anhydrous Na_2SO_4 (~ 4 g), and concentrated in vacuo. The crude product was purified *via* flash column chromatography using EtOAc : hexanes (7 :3 to 100:0) as eluent to afford 1.2 g (52 %) of a colorless foam. R_f (SiO_2 TLC): 0.35 EtOAc . $\lambda_{max(MeCN)}$ =283 nm. ^1H NMR (500MHz, $CDCl_3$, ppm): 7.94 (s, 1H, H6), 7.77 (s, 1H, H6), 7.42-7.40 (m, 2H Ar), 7.31-7.23 (m, 7H, Ar), 6.84-6.81

(m, 4H, Ar), 6.41 (dd, 1H, H1'a, J = 9, 5.5 Hz), 6.32 (dd, 1H, H1'b, J = 6 Hz), 5.24 (m, 1H, H3'a), 4.48-4.43 (m, 5H, H3'b & 2 x $ArOCH_2$), 4.13 (m, 1H, H4'a), 3.96 (m, 1H, H4'b), 3.91-3.90 (m, 2H, H5'a & H5''a), 3.79 (s, 6H, 2 x OCH_3), 3.51 (m, 1H, H5'b), 3.25 (m, 1H, H5''b), 2.75 (m, 2H, CH_2CO), 2.64-2.47 (m, 4H, H2'a, H2'b & $OOCCH_2$), 2.23-2.18 (m, 4H, H2''b & $COCH_3$), 2.04-1.92 (m, 4H, H2''a & $ArCH_3$), 1.88 (m, 4H, CH_2CH_2), 1.51 (s, 3H, $ArCH_3$), 0.90 (s, 9H, $SiC(CH_3)_3$), 0.80 (s, 9H, $SiC(CH_3)_3$), 0.11- 0.10 (m, 9H, 3 x $SiCH_3$), - 0.08 (s, 3H, $SiCH_3$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 206.3, 205.7, 176.8, 172.4, 168.4, 164.0, 150.5, 135.2, 111.1, 85.3, 84.8, 84.7, 63.6, 38.0, 37.8, 37.7, 37.3, 29.8, 29.7, 29.0, 27.9, 27.7, 25.9, 18.3, 12.4, -5.42, -5.52. IR (thin film); ν_{max} (cm^{-1}) = 3063, 2954, 2928, 2855, 1739, 1720, 1673, 1537, 1509, 1330, 1253, 1229, 1178, 1033, 835. HRMS (ESI-MS) *m/z* calculated for $C_{62}H_{87}N_4O_{14}Si_2^+$ 1167.5752: found 1167.5740 $[M+H]^+$

1-{O^4-[3'-O-(tert-Butyldimethylsilyl)-5'-O-(4,4'-dimethoxytrityl)-thymidinyl]}-7-{O^4-[5'-O-(tert-butyldimethylsilyl)-3'-O-(levulinyl)-thymidinyl]}-heptane (**6.5b**)

Intermediate **6.3** was prepared according to the procedure for product **6.5a** as described above. To a solution of **6.3** (~ 0.48 g, ~0.95 mmol) and **6.4b** (0.77 g, 1.0 mmol) in MeCN (7.9 mL) was added DBU (0.81 mL, 5.4 mmol) and allowed to stir at 21 ^oC. After 20h, the solvent was removed in vacuo and the crude was taken up in DCM (50 mL). The content was washed with aqueous $NaHCO_3$ (3% w/ v, 2 x 50 mL) and once with brine (50 mL), then dried over anhydrous Na_2SO_4 (~ 4 g), and concentrated in vacuo. The crude product was purified *via* flash column chromatography using MeOH : DCM (1% to 3%) as eluent to afford 0.78 g (68 %) of a colorless foam. R_f (SiO_2 TLC): 0.54 MeOH : DCM (5%). $\lambda_{max(MeCN)}$ = 283 nm. ^1H NMR (500MHz, $CDCl_3$, ppm): 7.92 (d, 1H, H6, J = 1 Hz), 7.75 (d, 1H, H6, J = 1 Hz), 7.42-7.40 (m, 2H Ar), 7.31-7.21 (m, 7H, Ar), 6.84-6.81 (m, 4H, Ar), 6.41 (dd, 1H, H1'a, J = 9.0, 5.5 Hz), 6.33

(dd, 1H, H1'b, J = 6 Hz), 5.24 (m, 1H, H3'a), 4.47 (m, 1H, H3'b), 4.38-4.35 (m, 4H, 2 x $ArOCH_2$), 4.13 (m, 1H, H4'a), 3.96 (m, 1H, H4'b), 3.91-3.90 (m, 2H, H5'a & H5''a), 3.78 (s, 6H, 2 x OCH_3), 3.51 (m, 1H, H5'b), 3.25 (m, 1H, H5''b), 2.74 (m, 2H, CH_2CO), 2.64-2.46 (m, 4H, H2'a, H2'b & $OOCCH_2$), 2.22-2.17 (m, 4H, H2''b & $COCH_3$), 2.00 (m, 1H, H2''a), 1.93 (d, 3H, $ArCH_3$, J = 1 Hz), 1.76 (m, 4H, CH_2CH_2), 1.51 (d, 3H, $ArCH_3$, J = 1 Hz), 1.38 (m, 6H, $CH_2CH_2CH_2$), 0.89 (s, 9H, $SiC(CH_3)_3$), 0.80 (s, 9H, $SiC(CH_3)_3$),0.11 (s, 3H, $SiCH_3$), 0.10 (s, 3H, $SiCH_3$),-0.01 (s, 3H, $SiCH_3$), - 0.08 (s, 3H, $SiCH_3$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 206.3, 172.4, 170.5, 170.5, 158.7, 156.0, 155.9, 144.4, 139.4, 139.0, 135.5, 130.1, 130.1, 128.2, 127.90, 127.0, 113.2, 113.2, 104.7, 104.5, 86.6, 86.5, 86.4, 86.2, 85.6, 75.9, 71.1, 67.4, 67.3, 63.6, 62.4, 55.2, 42.2, 39.0, 37.8, 29.8, 29.0, 28.5, 28.5, 28.0, 25.9, 25.9, 25.7, 18.2, 17.9, 12.3, 11.8, -4.64, -4.97, -5.45, -5.54. IR (thin film); ν_{max} (cm^{-1}) = 3063, 2953, 2928, 2855, 2739, 1720, 1673, 1608, 1536, 1509, 1329, 1253, 1178, 1122, 1033, 835. HRMS (ESI-MS) *m/z* calculated for $C_{65}H_{93}N_4O_{14}Si_2^+$ 1209.6221: found 1209.6213 $[M+H]^+$

1-{O^4-[3'-O-(tert-Butyldimethylsilyl)-5'-O-(4,4'-dimethoxytrityl)-thymidinyl]}-4-{O^4-[5'-O-(tert-butyldimethylsilyl)- thymidinyl]}-butane (**6.6a**)

A solution (1.5 mL) of 0.5 M N_2H_4•H_2O in pyridinium acetate (1/1, v/v) was added to **6.5a** (0.39 g, 0.33 mmol) and allowed to stir. After 5 min, the reaction was diluted with DCM (100 mL) and washed several times with saturated aqueous $NaHCO_3$ (6 x 50 mL), followed by water (50 mL) and brine (50 mL). The organic layer was dried over anhydrous Na_2SO_4 (~ 4 g), concentrated in vacuo and purified *via* flash column chromatography using MeOH : DCM (1% to 5%) as eluent to afford 0.3 g (83 %) of a colorless foam. R_f (SiO_2 TLC): 0.20 MeOH : DCM (3%). $\lambda_{max(MeCN)}$ =283 nm. ^1H NMR (500MHz, $CDCl_3$, ppm): 7.94 (d, 1H, H6, J = 1 Hz), 7.79 (d, 1H, H6, J = 1 Hz), 7.41-7.40 (m, 2H Ar), 7.29-7.21 (m, 7H, Ar), 6.83-6.80 (m, 4H, Ar), 6.38

(dd, 1H, H1'a, J = 7.5, 5.5 Hz), 6.32 (dd, 1H, H1'b, J = 6 Hz), 4.49-4.42 (m, 6H, H3'a, H3'b & 2 x $ArOCH_2$), 4.10 (m, 1H, H4'a), 3.96 (m, 1H, H4'b), 3.91 (m, 1H, H5'a), 3.83 (m, 1H, H5''a), 3.78 (s, 6H, 2 x OCH_3), 3.51 (m, 1H, H5'b), 3.25 (m, 1H, H5''b), 3.04 (d, 1H, OH, J = 3.5), 2.60 (m, 1H, H2'a), 2.47 (m, 1H, H2'b), 2.19 (m, 1H, H2''b), 2.03 (m, 1H, H2''a), 1.93 (d, 3H, $ArCH_3$, J = 1 Hz), 1.87 (m, 4H, CH_2CH_2), 1.50 (d, 3H, $ArCH_3$, J = 1Hz), 0.90 (s, 9H, $SiC(CH_3)_3$), 0.80 (s, 9H, $SiC(CH_3)_3$), 0.10 (s, 3H, $SiCH_3$), 0.09 (s, 3H, $SiCH_3$), -0.01 (s, 3H, $SiCH_3$), - 0.08 (s, 3H, $SiCH_3$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 170.34, 170.32, 158.7, 156.0, 155.9, 144.4, 139.5, 135.49, 135.48, 130.08, 130.06, 128.2, 127.9, 127.0, 113.2, 113.2, 104.50, 104.45, 87.4, 86.6, 86.5, 86.2, 72.3, 71.0, 66.73, 66.65, 63.5, 62.4, 55.2, 42.23, 42.19, 25.9, 25.7, 25.3, 25.2, 18.3, 17.9, 12.3, 11.7, -4.64, -4.97, -5.42, -5.48. IR (thin film); ν_{max} (cm^{-1}) =3384, 3061, 2954, 2929, 2855, 1669, 1580, 1509, 1472, 1329, 1121, 835. HRMS (ESI-MS) *m/z* calculated for $C_{57}H_{81}N_4O_{12}Si_2^+$ 1069.5384: found 1069.5378 $[M+H]^+$

1-{O^4-[3'-O-(tert-Butyldimethylsilyl)-5'-O-(4,4'-dimethoxytrityl)-thymidinyl]}-7-{O^4-[5'-O-(tert-butyldimethylsilyl)- thymidinyl]}-heptane (**6.6b**)

A solution (2.9 mL) of 0.5 M N_2H_4•H_2O in pyridinium acetate (1/1, v/v) was added to **6.5b** (1.2 g, 0.98 mmol) and allowed to stir. After 5 min, the reaction was diluted with EtOAc (75 mL) and washed several times with saturated aqueous $NaHCO_3$ (6 x 80 mL), followed by brine (50 mL). The organic layer was dried over anhydrous Na_2SO_4 (~ 4 g), concentrated in vacuo and purified *via* flash column chromatography using MeOH : DCM (1.5% to 3.5%) as eluent to afford 0.74 g (69 %) of a colorless foam. R_f (SiO_2 TLC): 0.10 EtOAc. $\lambda_{max(MeCN)}$ = 283 nm. ^1H NMR (500MHz, $CDCl_3$, ppm): 7.92 (d, 1H, H6, J = 1 Hz), 7.78 (d, 1H, H6, J = 1 Hz), 7.41-7.40 (m, 2H Ar), 7.29-7.21 (m, 7H, Ar), 6.83-6.81 (m, 4H, Ar), 6.38 (dd, 1H, H1'a, J = 7.5, 6.0 Hz), 6.32 (dd, 1H, H1'b, J = 6 Hz), 4.48-4.44 (m, 2H, H3'a & H3'b), 4.37-4.34 (m, 4H, 2 x

$ArOCH_2$), 4.09 (m, 1H, H4'a), 3.96 (m, 1H, H4'b), 3.92 (m, 1H, H5'a), 3.84 (m, 2H, H5''a), 3.78 (s, 6H, 2 x OCH_3), 3.51 (m, 1H, H5'b), 3.25 (m, 1H, H5''b), 2.99 (d, 1H, OH, J = 3.5 Hz), 2.60 (m, 1H, H2'a), 2.48 (m, 1H, H2'b), 2.20 (m, 1H, H2''b), 2.04 (m, 1H, H2''a), 1.93 (d, 3H, $ArCH_3$, J = 1 Hz), 1.73 (m, 4H, CH_2CH_2), 1.51 (d, 3H, $ArCH_3$, J = 1 Hz), 1.39 (m, 6H, $CH_2CH_2CH_2$), 0.90 (s, 9H, $SiC(CH_3)_3$), 0.80 (s, 9H, $SiC(CH_3)_3$),0.11 (s, 3H, $SiCH_3$), 0.10 (s, 3H, $SiCH_3$),-0.01 (s, 3H, $SiCH_3$), - 0.08 (s, 3H, $SiCH_3$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 170.47, 170.46, 158.7, 156.1, 156.0, 144.4, 139.4, 139.3, 135.5, 130.08, 130.06, 128.2, 127.9, 127.0, 113.20, 113.18, 104.6, 104.5, 87.3, 86.6, 86.6, 86.5, 86.2, 72.3, 71.1, 67.3, 67.3, 63.5, 62.4, 55.2, 42.22, 42.19, 28.9, 28.5, 28.4, 25.9, 25.83, 25.81, 25.7, 18.3, 17.9, 12.3, 11.8, -4.64, -4.97, -5.42, -5.48. IR (thin film); ν_{max} (cm^{-1}) = 3365, 3061, 2953, 2930, 2856, 1671, 1608, 1536, 1509, 1472, 1430, 1328, 1253, 1178, 1121, 835. HRMS (ESI-MS) *m/z* calculated for$C_{60}H_{87}N_4O_{12}Si_2^+$ 1111.5854: found 1111.5820 $[M+H]^+$

1-{O^4-[3'-O-(tert-Butyldimethylsilyl)-5'-O-(4,4'-dimethoxytrityl)-thymidinyl]}-4-{O^4-[5'-O-(tert-butyldimethylsilyl)-3'-O-(2-cyanoethoxy(diisopropylamino)-phosphino)-thymidinyl]}-butane

(**6.7a**)

To a solution of **6.6a** (0.30 g, 0.28 mmol) and DIPEA (92 µL, 0.53 mmol) in THF (2.8 mL) was added Cl-P(OCEt)(N*i*Pr_2) (93 µL, 0.42 mmol) and the reaction was allowed to stir at room temperature. After 20 min, the solvent was removed in vacuo and the crude was taken up in EtOAc (50 mL), washed with aqueous $NaHCO_3$ (3% w/v, 2 x 40mL), followed with brine (40 mL). The organic layer was dried over anhydrous Na_2SO_4 (~ 1 g), decanted and then evaporated in vacuo. The crude product was purified *via* flash column chromatography using EtOAc (containing 0.1% v/v NEt_3) as eluent to afford 0.27 g (74 %) of a colorless foam. R_f (SiO_2 TLC): 0.53, 0.41 EtOAc. $\lambda_{max(MeCN)}$ = 283 nm. ^1H NMR (500MHz, d_6-acetone, ppm): 7.94 (m, 1H, H6),

7.82-7.79 (m, 1H, H6), 7.51-7.48 (m, 2H Ar), 7.39-7.24 (m, 7H, Ar), 6.92-6.89 (m, 4H, Ar), 6.32-6.25 (m, 2H, H1'a & H1'b), 4.63-4.59 (m, 2H, H3'a & H3'b), 4.41-4.36 (m, 4H, 2 x $ArCH_2$), 4.25, 4.16 (m, 1H, H4'a), 4.03-3.82 (m, 5H, $POCH_2$, H4'b, H5'a & H5''a), 3.79 (s, 6H, 2 x OCH_3), 3.73-3.65 (m, 2H, 2 x NCH), 3.49 (m, 1H, H5'b), 3.35 (m, 1H, H5''b), 2.81-2.77 (m, 2H, CH_2CN), 2.62-2.51 (m, 1H, H2'a), 2.42-2.37 (m, 1H, H2'b), 2.33-2.27 (m, 1H, H2''b), 2.16-2.11 (m, 1H, H2''a), 1.94-1.92 (m, 7H, $ArCH_3$ & CH_2CH_2), 1.62-1.61 (m, 3H, $ArCH_3$), 1.23-1.21 (m, 12H, 4 x CH_3), 0.94-0.93 (m, 9H, $SiC(CH_3)_3$), 0.86-0.85 (m, 9H, $SiC(CH_3)_3$), 0.16-0.14 (m, 6H, 2 x $SiCH_3$), 0.07 (s, 3H, $SiCH_3$), 0.01 (s, 3H, $SiCH_3$). ^{13}C NMR (125.7 MHz, d_6-acetone, ppm): 170.88, 170.87, 159.77, 155.58, 155.57, 155.56, 145.82, 140.87, 140.69, 140.62, 136.54, 136.50, 131.01, 131.00, 129.01, 128.72, 127.77, 118.93, 114.02, 114.01, 104.12, 104.10, 104.01, 87.96, 87.92, 87.57, 87.53, 87.46, 87.34, 87.24, 87.04, 86.87, 75.11, 75.06, 74.98, 74.93, 72.76, 67.10, 64.23, 64.08, 63.80, 59.55, 59.41, 55.57, 44.01, 43.98, 43.91, 43.88, 42.54, 41.49, 41.45, 41.34, 41.31, 26.35, 26.15, 26.04, 24.92, 24.91, 24.89, 24.87, 24.85, 24.83, 24.81, 20.79, 20.77, 20.73, 20.72, 18.93, 18.91, 18.50, 12.48, 12.46, 12.12, -4.44, -4.69, -5.17, -5.18, -5.22.^{31}P NMR (202.3 MHz, d_6-acetone, ppm): 148.05, 147.68. IR (thin film); ν_{max} (cm^{-1}) = 3064, 2957, 2929, 2856, 1714, 1673, 1608, 1536, 1509, 1471, 1363, 1329, 1253, 1124, 1035, 1002, 835, 782. HRMS (ESI-MS) *m/z* calculated for$C_{66}H_{98}N_6O_{13}PSi_2^+$ 1269.6463: found 1269.6499 $[M+H]^+$

1-{O^4-[3'-O-(tert-Butyldimethylsilyl)-5'-O-(4,4'-dimethoxytrityl)-thymidinyl]}-7-{O^4-[5'-O-(tert-butyldimethylsilyl)-3'-O-(2-cyanoethoxy(diisopropylamino)-phosphino)-thymidinyl]}-heptane

(**6.7b**)

To a solution of **6.6b** (0.31 g, 0.28 mmol) and DIPEA (91 µL, 0.52 mmol) in THF (2.8 mL) was added Cl-P(OCEt)(N*i*Pr$_2$) (93 µL, 0.42 mmol) and the reaction was allowed to stir at room temperature. After 20 min, the solvent was removed in vacuo and the crude was taken up

in EtOAc (50 mL), washed with aqueous $NaHCO_3$ (3% w/v, 2 x 40mL), followed with brine (40 mL). The organic layer dried over anhydrous Na_2SO_4 (~ 1 g), decanted and the solvent evaporated in vacuo. The crude product was purified *via* flash column chromatography using EtOAc (containing 0.1% v/v NEt_3) as eluent to afford 0.32 g (87 %) of a colorless foam. R_f (SiO_2 TLC): 0.81, 0.70 EtOAc. $\lambda_{max(MeCN)}$ = 283 nm. ^{1}H NMR (500MHz, d_6-acetone, ppm): 7.94-7.93 (m, 1H, H6), 7.81-7.78 (m, 1H, H6), 7.51-7.48 (m, 2H Ar), 7.39-7.24 (m, 7H, Ar), 6.92-6.89 (m, 4H, Ar), 6.32-6.26 (m, 2H, H1'a & H1'b), 4.63-4.60 (m, 2H, H3'a & H3'b), 4.33-4.30 (m, 4H, 2 x $ArCH_2$), 4.25, 4.16 (m, 1H, H4'a), 4.02-3.82 (m, 5H, $POCH_2$, H4'b, H5'a & H5''a), 3.79 (s, 6H, 2 x OCH_3), 3.72-3.65 (m, 2H, 2 x NCH), 3.49 (m, 1H, H5'b), 3.35 (m, 1H, H5''b), 2.80-2.78 (m, 2H, CH_2CN), 2.63-2.51 (m, 1H, H2'a), 2.42-2.37 (m, 1H, H2'b), 2.33-2.28 (m, 1H, H2''b), 2.18-2.11 (m, 1H, H2''a), 1.94-1.93 (m, 3H, $ArCH_3$), 1.79-1.75 (m, 4H, 2 x CH_2), 1.60 (m, 3H, $ArCH_3$), 1.50-1.47 (m, 6H, $CH_2CH_2CH_2$), 1.23-1.21 (m, 12H, 4 x CH_3), 0.94-0.93 (m, 9H, $SiC(CH_3)_3$), 0.86-0.84 (m, 9H, $SiC(CH_3)_3$), 0.16-0.14 (m, 6H, 2 x $SiCH_3$), 0.06 (s, 3H, $SiCH_3$), 0.01 (s, 3H, $SiCH_3$). ^{13}C NMR (125.7 MHz, d_6-acetone, ppm): 170.04, 170.03, 158.87, 154.70, 144.91, 139.92, 139.73, 139.66, 135.64, 135.61, 130.11, 130.10, 128.13, 127.81, 126.87, 118.05, 113.12, 113.10, 103.20, 103.19, 103.09, 87.05, 87.02, 86.66, 86.61, 86.56, 86.41, 86.33, 86.13, 85.93, 74.21, 74.16, 74.08, 74.03, 71.83, 66.53, 66.51, 63.33, 63.18, 62.88, 58.66, 58.51, 54.67, 43.11, 43.08, 43.02, 42.98, 41.63, 40.57, 40.54, 40.42, 40.39, 28.77, 28.32, 25.71, 25.46, 25.26, 24.03, 24.01, 23.99, 23.97, 23.95, 23.94, 23.92, 19.89, 19.88, 19.84, 19.82, 18.03, 18.01, 17.61, 11.60, 11.57, 11.23, -5.33, -5.59, -6.07, -6.08, -6.12. ^{31}P NMR (202.3 MHz, d_6-acetone, ppm): 148.04, 147.67. IR (thin film); ν_{max} (cm^{-1}) = 3064, 2957, 2931, 2857, 1714, 1673, 1608, 1536, 1430, 1328, 1254, 1223, 1180, 1126, 1034, 1003, 835, 782. HRMS (ESI-MS) *m/z* calculated for $C_{69}H_{104}N_6O_{13}PSi_2^+$ 1311.6932: found 1311.7026 $[M+H]^+$

The synthesis of phosphoramidites **6.11a** and **6.11b** are shown in **Scheme 6.2**

5'-O-(tert-butyldimethylsilyl)-3'-O-levulinyl-O^4-hydroxybutyl-thymidine (**6.8a**)

To a solution of **6.2** (1.0 g, 2.2 mmol) and triazole (3.4 g, 50 mmol) in MeCN : DCM (14 mL : 14 mL) stirring at 0 ^{o}C was added NEt_3 (7.1 mL, 51 mmol) followed by the dropwise addition of $POCl_3$ (0.31 mL, 3.3 mmol). Additional $POCl_3$ (0.20 mL, 2.1 mmol) was added dropwise after 10 min. After 30 min, the solvent was removed in vacuo and the content taken up in DCM (50 mL), then washed with a 3 % (w/v) aqueous solution of $NaHCO_3$ (2 x 50mL). The solvent was dried over anhydrous Na_2SO_4, decanted and evaporated to produce the intermediate as a yellow gum. To a solution of this intermediate and 1,4-butanediol (0.22 g, 2.4 mmol) in MeCN (13 mL) was added DBU (0.40 mL, 2.6 mmol). After 16h, the solvent was removed in vacuo and the content taken up in EtOAc (50 mL), then washed with a 3 % (w/v) aqueous solution of $NaHCO_3$ (2 x 50mL) followed with brine (50 mL). The solvent was dried over anhydrous Na_2SO_4 (≈ 4 g), decanted and then evaporated to afford a yellow gum. The product was purified *via* flash column chromatography using MeOH : DCM (1 % → 5 %) as eluent to afford 0.50 g (43 % over two steps) of a slightly yellow oil. R_f (SiO_2 TLC): 0.10 EtOAc. $\lambda_{max(MeCN)}$ = 287 nm. ^{1}H NMR (500MHz, $CDCl_3$, ppm): 7.74 (d, 1H, H6, J = 1 Hz), 6.37 (dd, 1H, H1', J = 9 Hz & 5 Hz), 5.22 (m, 1H, H3'), 4.40 (t, 2H, $ArOCH_2$, J = 6.0 Hz), 4.11 (m, 1H, H4'), 3.89-3.87 (m, 2H, H5' & H5''), 3.68 (m, 2H, $C\underline{H_2}OH$), 2.75 (m, 2H, $COCH_2$), 2.61-2.55 (m, 3H, H2' & $OOCCH_2$), 2.17 (s, 3H, $COCH_3$), 1.99 (m, 1H, H2''), 1.91-1.81 (m, 6H, $ArCH_3$, OH, CH_2), 1.67 (m, 2H, CH_2), 0.87 (s, 9H, $SiC(CH_3)_3$), 0.084 (s, 3H, $SiCH_3$), 0.079 (s, 3H, $SiCH_3$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 206.3, 172.4, 170.4, 155.9, 139.1, 104.7, 86.4, 85.6, 75.9, 67.1, 63.5, 62.2, 39.0, 37.7, 29.8, 29.1, 27.9, 25.8, 25.0, 18.2, 12.2, -5.47, -5.56. IR (thin film); ν_{max} (cm^{-1}) = 3419, 2958, 2929, 2856, 2362, 2335, 1739, 1719, 1669, 1534, 1475,

1432, 1331, 1259, 1231, 1126, 836, 783. HRMS (ESI-MS) *m/z* calculated for $C_{25}H_{43}N_2O_8Si^+$ 527.2783: found 527.2779 $[M+H]^+$.

5'-O-(tert-butyldimethylsilyl)-3'-O-levulinyl-O^4-hydroxyheptyl-thymidine (**6.8b**)

To a solution of **6.2** (2.3 g, 5.1 mmol) and triazole (10.3 g, 148 mmol) in MeCN : DCM (41 mL : 41 mL) stirring at 0 ^oC was added NEt_3 (21 mL, 151 mmol) followed by the dropwise addition of $POCl_3$ (0.92 mL, 9.9 mmol). Additional $POCl_3$ (0.68 mL, 7.3 mmol) was added dropwise after 15 min. After 30 min, the solvent was removed in vacuo and the content taken up in DCM (50 mL), then washed with a 3 % (w/v) aqueous solution of $NaHCO_3$ (2 x 75mL). The solvent was dried over anhydrous Na_2SO_4 (≈ 4 g), decanted and then evaporated to produce the intermediate as a yellow gum. To a solution of this intermediate and 1,7-butanediol (0.84 g, 6.3 mmol) in MeCN (37 mL) was added DBU (1.14 mL, 7.6 mmol). After 16h, the solvent was removed in vacuo and the content taken up in EtOAc (70 mL) then washed with a 3 % (w/v) aqueous solution of $NaHCO_3$ (2 x 50mL) and followed with brine (50 mL). The solvent was dried over anhydrous Na_2SO_4 (≈ 4 g), decanted and evaporated to afford a yellow gum. The product was purified *via* flash column chromatography using MeOH : DCM (0.5 % → 2 %) as eluent to afford 1.2 g (43 % over two steps) of a slightly yellow oil. R_f (SiO_2 TLC): 0.18 EtOAc:Hex (8 : 2). $\lambda_{max(MeCN)}$ = 288 nm. ^1H NMR (500MHz, $CDCl_3$, ppm): 7.72 (d, 1H, H6, *J* = 1 Hz), 6.36 (dd, 1H, H1', *J* = 9 Hz & 5 Hz), 5.21 (m, 1H, H3'), 4.34 (t, 2H, $ArOCH_2$, *J* = 6.5 Hz), 4.10 (m, 1H, H4'), 3.90-3.84 (m, 2H, H5' & H5''), 3.60 (m, 2H, $C\underline{H_2}OH$), 2.73 (m, 2H, $COCH_2$), 2.60-2.53 (m, 3H, H2' & $OOCCH_2$), 2.16 (s, 3H, $COCH_3$), 1.97 (m, 1H, H2''), 1.89 (d, 3H, $ArCH_3$, *J* = 1Hz), 1.74-1.68 (m, 3H, OH & CH_2), 1.53 (m, 2H, CH_2), 1.40-1.33 (m, 6H, $CH_2CH_2CH_2$) 0.86 (s, 9H, $SiC(CH_3)_3$), 0.072 (s, 3H, $SiCH_3$), 0.066 (s, 3H, $SiCH_3$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 206.3, 172.4, 170.5, 156.0, 138.9, 104.7, 86.4, 85.6, 75.8, 67.3, 63.5,

62.8, 38.9, 37.7, 32.6, 29.7, 28.9, 28.4, 27.9, 25.8, 25.8, 25.8, 25.6, 18.2, 12.2, -5.49, -5.57. IR (thin film); ν_{max} (cm^{-1}) = 3426, 2961, 2931, 2856, 2362, 2336, 1739, 1721, 1669, 1535, 1473, 1432, 1331, 1259, 1230, 1126, 837, 783. HRMS (ESI-MS) *m/z* calculated for $C_{28}H_{49}N_2O_8Si^+$ 569.3253: found 569.3247 $[M+H]^+$.

1-{O⁴-[3'-O-(tert-Butyldimethylsilyl)-5'-O-(4,4'-dimethoxytrityl)-N2-phenoxyacetyl-2'-deoxyguanidyl]}-4-{O⁴-[5'-O-(tert-butyldimethylsilyl)-3'-O-levulinyl-thymidinyl]}-butane (**6.9a**)

To a solution of **6.8a** (0.50 g, 0.95 mmol), 3'-*O*-(*tert*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-*N*2-phenoxyacetyl-2'-deoxyguanosine (0.78 g, 0.95 mmol) and Ph_3P (0.50, 1.9 mmol) in dioxane (2.8 mL) was added DIAD (0.37 mL, 1.9 mmol) dropwise over 5 min while stirring. After 16h, additional DIAD (0.1 mL, 0.51 mmol) was added and the reaction was allowed to stir for another 2h. The solvent was removed in vacuo and the content taken up in DCM (50 mL) and washed with a 3 % (w/v) aqueous solution of $NaHCO_3$ (2 x 50 mL). The solvent was dried over anhydrous Na_2SO_4 (≈ 4 g), decanted and evaporated to produce a yellow gum. The product was purified *via* flash column chromatography using MeOH : DCM (1 % → 2 %) as eluent to afford 1.2 g of crude material (verified by ^1H NMR, data not shown). The product was repurified *via* flash column chromatography using EtOAc : hexanes (3 : 2 → 4: 1) as eluent to afford 0.59 g (46 %) of a colorless foam. R_f (SiO_2 TLC): 0.22 MeOH : DCM (2 %). $\lambda_{max(MeCN)}$ = 271 nm. ^1H NMR (500MHz, $CDCl_3$, ppm): 8.67 (broad s, 1H, NH), 8.03 (s, 1H, H8), 7.76 (d, 1H, H6, J = 1 Hz), 7.39-7.17 (m, 11H, Ar), 7.04-6.99 (m, 3H, Ar), 6.78-6.76 (m, 4H, Ar), 6.42-6.38 (m, 2H, H1'a & H1'b), 5.24 (m, 1H, H3'b), 4.76 (broad s, 2H, $PhOCH_2CO$), 4.65 (m, 2H, $ArOCH_2$), 4.57 (m, 1H, H3'a), 4.45 (m, 2H, $ArOCH_2$), 4.13 (m, 1H, H4'b), 4.06 (m, 1H, H4'a), 3.92-3.86 (m, 2H, H5'b & H5''b), 3.75 (s, 6H, 2 x OCH_3), 3.32 (m, 2H, H5'a & H5''a), 2.78-2.54 (m, 6H, $COCH_2$, $OOCCH_2$, H2'a & H2'b), 2.43 (m, 1H, H2''a), 2.19 (s, 3H, $COCH_3$),

2.02-1.97 (m, 5H, H2''b & CH_2CH_2), 1.92 (s, 3H, $ArCH_3$), 0.89 (s, 9H, $SiC(CH_3)_3$), 0.85 (s, 9H, $SiC(CH_3)_3$), 0.10 (m, 6H, 2 x $SiCH_3$), 0.03 (s, 3H, $SiCH_3$), - 0.01 (s, 3H, $SiCH_3$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 206.29, 172.40, 170.38, 161.02, 158.48, 157.22, 155.89, 152.39, 151.08, 144.51, 140.12, 139.12, 135.72, 135.68, 129.98, 129.97, 129.77, 128.09, 127.79, 126.83, 122.25, 118.89, 114.91, 113.10, 104.66, 86.95, 86.45, 86.41, 85.64, 84.28, 75.91, 72.50, 68.01, 67.16, 66.84, 63.56, 63.44, 55.17, 40.93, 38.96, 37.75, 29.77, 27.96, 25.85, 25.72, 25.47, 25.21, 18.23, 17.93, 12.27, -4.70, -4.84, -5.45, -5.54. IR (thin film); ν_{max} (cm^{-1}) = 3412, 2956, 2928, 2856, 2361, 2336, 1735, 1719, 1670, 1607, 1539, 1509, 1472, 1437, 1332, 1251, 1177, 1124, 835, 782. HRMS (ESI-MS) *m/z* calculated for $C_{70}H_{92}N_7O_{15}Si_2^+$ 1326.6184: found 1326.6177$[M+H]^+$.

1-{O^4-[3'-O-(tert-Butyldimethylsilyl)-5'-O-(4,4'-dimethoxytrityl)-N2-phenoxyacetyl-2'-deoxyguanidyl]}-7-{O^4-[5'-O-(tert-butyldimethylsilyl)-3'-O-levulinyl-thymidinyl]}-heptane

(**6.9b**)

To a solution of **6.8b** (0.75 g, 1.3 mmol), 3'-*O*-(*tert*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-*N*2-phenoxyacetyl-2'-deoxyguanosine (0.98 g, 1.2 mmol) and Ph_3P (0.66, 2.5 mmol) in dioxane (4.0 mL) was added DIAD (0.50 mL, 2.5 mmol) dropwise over 5 min while stirring. After 16h, the solvent was removed in vacuo and the content taken up in EtOAc (50 mL), then washed with a 3 % (w/v) aqueous solution of $NaHCO_3$ (2 x 50 mL) and followed with brine (50 mL). The solvent was dried over anhydrous Na_2SO_4 (≈ 4 g), decanted and then evaporated to produce a yellow gum. The product was purified *via* flash column chromatography using EtOAc : hexanes (3 : 2 → 4 : 1) as eluent to afford 0.98 g (71 %) of a colorless foam. R_f (SiO_2 TLC): 0.70 EtOAc. $\lambda_{max(MeCN)}$ = 271 nm. ^1H NMR (500MHz, $CDCl_3$, ppm): 8.64 (broad s, 1H, NH), 8.02 (s, 1H, H8), 7.75 (d, 1H, H6, J = 1 Hz), 7.40-7.17 (m, 11H, Ar), 7.05-7.00 (m,

3H, Ar), 6.78-6.76 (m, 4H, Ar), 6.42-6.39 (m, 2H, H1'a & H1'b), 5.24 (m, 1H, H3'b), 4.78 (broad s, 2H, $PhOCH_2CO$), 4.60-4.57 (m, 3H, H3'a & $ArOCH_2$), 4.37 (t, 2H, $ArOCH_2$, J = 7 Hz), 4.13 (m, 1H, H4'b), 4.07 (m, 1H, H4'a), 3.93-3.87 (m, 2H, H5'b & H5''b), 3.76 (s, 6H, 2 x OCH_3), 3.33 (m, 2H, H5'a & H5''a), 2.79-2.55 (m, 6H, $COCH_2$, $OOCCH_2$, H2'a & H2'b), 2.44 (H2''a), 2.19 (s, 3H, $COCH_3$), 2.00 (m, 1H, H2''b), 1.92-1.86 (m, 5H, $ArCH_3$ & CH_2), 1.75 (m, 2H, CH_2), 1.52 (m, 2H, CH_2), 1.43-1.42 (m, 4H, CH_2CH_2), 0.89 (s, 9H, $SiC(CH_3)_3$), 0.85 (s, 9H, $SiC(CH_3)_3$) 0.104 (s, 3H, $SiCH_3$), 0.098(s, 3H, $SiCH_3$), 0.04 (s, 3H, $SiCH_3$), - 0.01 (s, 3H, $SiCH_3$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 206.30, 172.41, 170.50, 161.21, 158.47, 155.96, 152.33, 151.11, 144.51, 140.02, 138.93, 135.72, 135.69, 129.97, 129.76, 128.09, 127.79, 126.83, 122.23, 118.93, 114.91, 113.10, 113.09, 104.76, 86.94, 86.44, 86.36, 85.60, 84.27, 75.91, 72.52, 68.04, 67.82, 67.43, 63.57, 63.45, 55.17, 40.90, 38.94, 37.76, 29.77, 29.07, 28.78, 28.50, 27.96, 25.90, 25.85, 25.72, 18.23, 17.94, 12.29, -4.70, -4.84, -5.45, -5.54. IR (thin film); ν_{max} (cm^{-1}) = 3412, 2956, 2928, 2856, 2362, 2336, 1735, 1719, 1670, 1607, 1533, 1509, 1464, 1437, 1331, 1250, 1177, 1124, 835, 781. HRMS (ESI-MS) *m/z* calculated for $C_{73}H_{98}N_7O_{15}Si_2^+$ 1368.6654: found 1368.6658 $[M+H]^+$.

1-{O^4-[3'-O-(tert-Butyldimethylsilyl)-5'-O-(4,4'-dimethoxytrityl)-N2-phenoxyacetyl-2'-deoxyguanidyl]}-4-{O^4-[5'-O-(tert-butyldimethylsilyl)-thymidinyl]}-butane (**6.10a**)

A solution (2.5 mL) of 0.5 M $N_2H_4 \bullet H_2O$ in pyridinium acetate (1/1, v/v) was added to **6.9a** (0.75 g, 0.57 mmol). After 15 min of stirring, the reaction was diluted with EtOAc (75 mL added), washed several times with a 3 % (w/v) aqueous solution of $NaHCO_3$ (6 x 75 mL) and then with brine (75 mL). The solvent was dried over anhydrous Na_2SO_4 (≈ 4 g), decanted and evaporated to produce a yellow gum. The product was purified *via* flash column chromatography using MeOH : DCM (1.5 % → 3.5 %) as eluent to afford 0.41 g (58 %) of a colorless foam. R_f

(SiO_2 TLC): 0.23 MeOH : DCM (3 %). $\lambda_{max(MeCN)}$ = 271 nm. ^{1}H NMR (500MHz, $CDCl_3$, ppm): 8.68 (broad s, 1H, NH), 8.03 (s, 1H, H8), 7.77 (d, 1H, H6, *J* = 1 Hz), 7.40-7.16 (m, 11H, Ar), 7.06-6.99 (m, 3H, Ar), 6.78-6.76 (m, 4H, Ar), 6.42 (dd, 1H, H1'a, *J* = 6.5 Hz), 6.38 (dd, 1H, H1'b, *J* = 7.5 , 6 Hz), 4.76 (broad s, 2H, $PhOCH_2CO$), 4.65 (m, 2H, $ArOCH_2$), 4.57 (m, 1H, H3'a), 4.46-4.43 (m, 3H, H3'b & $ArOCH_2$), 4.08-406 (m, 2H, H4'a & H4'b), 3.91 (dd, 1H, H5'b, *J* = 11, 3 Hz), 3.83 (dd, 1H, H5''b, *J* = 11, 3 Hz), 3.76 (s, 6H, 2 x OCH_3), 3.33-3.32 (m, 2H, H5'a & H5''a), 2.83 (d, 1H, OH, *J* = 4.5 Hz), 2.69 (m, 1H, H2'a), 2.59 (m, 1H, H2'b), 2.44 (m, 1H, H2''a), 2.02-1.97 (m, 5H, H2''b & CH_2CH_2), 1.92 (d, 3H, $ArCH_3$, *J* = 1 Hz), 0.89 (s, 9H, $SiC(CH_3)_3$), 0.85 (s, 9H, $SiC(CH_3)_3$), 0.09 (s, 3H, $SiCH_3$), 0.08 (s, 3H, $SiCH_3$), 0.04 (s, 3H, $SiCH_3$), - 0.01 (s, 3H, $SiCH_3$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 170.38, 161.03, 158.48, 157.23, 156.01, 152.39, 151.08, 144.51, 140.11, 139.45, 135.73, 135.69, 129.99, 129.97, 129.77, 128.09, 127.80, 126.83, 122.25, 118.89, 114.92, 113.11, 104.48, 87.31, 86.95, 86.63, 86.45, 84.28, 72.50, 72.39, 68.01, 67.18, 66.80, 63.47, 55.18, 42.22, 40.94, 25.87, 25.72, 25.48, 25.22, 18.30, 17.94, 12.29, -4.69, -4.84, -5.42, -5.49. IR (thin film); ν_{max} (cm^{-1}) = 3409, 3057, 2954, 2929, 2857, 2362, 1724, 1663, 1608, 1532, 1509, 1496, 1462, 1441, 1380, 1362, 1251, 1177, 1067, 940, 835, 782.HRMS (ESI-MS) *m/z* calculated for $C_{65}H_{86}N_7O_{13}Si_2^+$ 1228.5817: found 1228.5822 $[M+H]^+$.

1-{O^4-[3'-O-(tert-Butyldimethylsilyl)-5'-O-(4,4'-dimethoxytrityl)-N2-phenoxyacetyl-2'-deoxyguanidyl]}-7-{O^4-[5'-O-(tert-butyldimethylsilyl)-thymidinyl]}-heptane (**6.10b**)

A solution (3.1 mL) of 0.5 M N_2H_4•H_2O in pyridinium acetate (1/1, v/v) was added to **6.9b** (0.95 g, 0.70 mmol). After 20 min of stirring, the reaction was diluted with EtOAc (70 mL added), then washed several times with a 3 % (w/v) aqueous solution of $NaHCO_3$ (6 x 75 mL) followed with brine (75 mL). The solvent was dried over anhydrous Na_2SO_4 (≈ 4 g), decanted

and evaporated to produce a yellow gum. The product was purified *via* flash column chromatography using MeOH : DCM (2 % → 3.5 %) as eluent to afford 0.64 g (73 %) of a colorless foam. R_f (SiO_2 TLC): 0.68 EtOAc. $\lambda_{max(MeCN)}$ = 271 nm. ^{1}H NMR (500MHz, $CDCl_3$, ppm): 8.66 (broad s, 1H, NH), 8.02 (s, 1H, H8), 7.78 (d, 1H, H6, *J* = 1 Hz), 7.39-7.17 (m, 11H, Ar), 7.05-6.99 (m, 3H, Ar), 6.78-6.76 (m, 4H, Ar), 6.43-6.37 (m, 2H, H1'a & H1'b), 4.78 (broad s, 2H, $PhOCH_2CO$), 4.58-4.56 (m, 3H, H3'a & $ArOCH_2$), 4.44 (m, 1H, H3'b), 4.36 (t, 2H, $ArOCH_2$, *J* = 6.5 Hz), 4.10-406 (m, 2H, H4'a & H4'b), 3.91 (dd, 1H, H5'b, *J* = 11, 3 Hz), 3.82 (dd, 1H, H5''b, *J* = 11, 3 Hz), 3.76 (s, 6H, 2 x OCH_3), 3.33-3.32 (m, 2H, H5'a & H5''a), 3.26 (d, 1H, OH, *J* = 4 Hz), 2.69 (m, 1H, H2'a), 2.61 (m, 1H, H2'b), 2.43 (m, 1H, H2''a), 2.02 (m, 1H, H2''b), 1.94-1.86 (m, 5H, $ArCH_3$ & CH_2), 1.75 (m, 2H, CH_2), 1.52 (m, 2H, CH_2), 1.43-1.42 (m, 4H, 2 x CH_2), 0.89 (s, 9H, $SiC(CH_3)_3$), 0.85 (s, 9H, $SiC(CH_3)_3$), 0.09 (s, 3H, $SiCH_3$), 0.08 (s, 3H, $SiCH_3$), 0.04 (s, 3H, $SiCH_3$), - 0.01 (s, 3H, $SiCH_3$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 170.5, 161.2, 158.5, 156.1, 152.3, 151.1, 144.5, 140.0, 139.3, 135.73, 135.70, 130.0, 129.8, 128.1, 127.8, 126.8, 122.2, 118.9, 114.9, 113.1, 113.1, 104.6, 87.4, 86.9, 86.7, 86.4, 84.3, 72.5, 72.3, 68.1, 67.8, 67.4, 63.50, 63.45, 55.2, 42.2, 40.9, 29.1, 28.8, 28.5, 25.9, 25.7, 18.3, 17.9, 12.3, -4.69, -4.84, -5.42, -5.48. IR (thin film); ν_{max} (cm^{-1}) = 3409, 3060, 2953, 2930, 2857, 2361, 2337, 1723, 1663, 1608, 1533, 1508, 1496, 1464, 1437, 1375, 1329, 1250, 1177, 1065, 1034, 940, 834, 781. HRMS (ESI-MS) *m/z* calculated for $C_{68}H_{92}N_7O_{13}Si_2^+$ 1270.6286: found 1270.6294 $[M+H]^+$.

1-{O^4-[3'-O-(tert-Butyldimethylsilyl)-5'-O-(4,4'-dimethoxytrityl)-N2-phenoxyacetyl-2'-deoxyguanidyl]}-4-{O^4-[5'-O-(tert-butyldimethylsilyl)-3'-O-(2-cyanoethoxy(diisopropylamino)-phosphino)-thymidinyl]}-butane (**6.11a**)

To a solution of compound **6.10a** (0.20 g, 0.16 mmol) and DIPEA (53 µL, 0.31 mmol) in THF (2 mL) was added Cl-POCEN*i*Pr$_2$ (54 µL, 0.25 mmol) dropwise while stirring. After 30 min, the solvent was evaporated in vacuo and the content was diluted with EtOAc (50 mL), then washed with $NaHCO_3$ (aq, 3%) (2 x 35 mL) and followed by brine (35 mL). The solvent was dried over anhydrous Na_2SO_4, decanted and then evaporated to produce a yellow gum. Purification was achieved *via* short flash column chromatography using EtOAc (containing 0.1 % NEt_3) as eluent to afford 0.16 g (69 %) of a colorless foam. R_f (SiO_2 TLC): 0.86, 0.81 EtOAc. $\lambda_{max(MeCN)}$ = 271 nm. ^{1}H NMR (500MHz, d_6-acetone, ppm): 9.33 (s, 1H, NH), 8.22 (s, 1H, H8), 7.82-7.79 (m, 1H, H6), 7.42-7.40 (m, 2H, Ar), 7.35-7.17 (m, 6H, Ar), 7.24-7.16 (m, 3H, Ar), 7.05-6.98 (m, 3H, Ar), 6.91-6.75 (m, 4H, Ar), 6.48-6.45 (m, 1H, H1'a), 6.33-6.29 (m, 1H, H1'b), 5.04 (bs, 2H, $PhOCH_2CO$), 4.90 (m, 1H, H3'a), 4.68-4.60 (m, 3H, H3'b & $ArOCH_2$), 4.45-4.42 (m, 2H, $ArOCH_2$), 4.26 & 4.16 (m, 1H, H4'b), 4.06-4.04 (m, 1H, H4'a), 3.99-3.83 (m, 4H, H5'b, H5''b & CH_2OP), 3.77-3.66 (m, 8H, 2x (CH)N & 2 x OCH_3), 3.48-3.44 (m, 1H, H5'a), 3.38-3.35 (m, 1H, H5''a), 3.11-3.08 (m, 1H, H2'a), 2.82-2.79 (m, 2H, CH_2CN), 2.64-2.59 (m, 1H, H2'b), 2.50-2.45 (m, 1H, H2''a), 2.19-2.12 (m, 1H, H2''b), 2.08-1.97 (m, 4H, CH_2CH_2, overlapping with acetone solvent residual peak), 1.94 (m, 3H, $ArCH_3$), 1.24-1.21 (m, 12H, 4 x CH_3), 0.94 (m, 9H, $SiC(CH_3)_3$) , 0.87 (m, 9H, $SiC(CH_3)_3$), 0.16-0.15 (m, 6H, 2 x $SiCH_3$), 0.09 (m, 3H, $SiCH_3$), 0.04 (m, 3H, $SiCH_3$) . ^{13}C NMR (125.7 MHz, d_6-acetone, ppm): 170.0, 160.9, 158.6, 158.6, 158.2, 154.8, 154.8, 152.7, 151.5, 145.3, 141.2, 139.8, 139.7, 136.0, 135.9, 130.1, 130.0, 129.5, 128.1, 127.5, 126.5, 121.3, 118.8, 118.1, 114.8, 112.9, 112.8, 103.3, 103.2, 87.1, 87.0, 86.8, 86.7, 86.6, 86.4, 86.2, 86.1, 84.5, 74.21, 74.16, 74.1, 74.0, 72.5, 67.9, 66.7, 66.1, 64.1, 63.3, 63.2, 58.7, 58.5, 54.6, 43.1, 43.1, 43.03, 42.99, 40.6, 40.4, 39.5, 25.5, 25.4, 25.3, 25.1, 24.03, 24.02, 23.98, 23.96, 23.9, 19.90, 19.88, 19.84, 19.83, 18.04, 18.02, 17.6, 11.60, 11.58, -

5.36, -5.50, -6.07, -6.12. ^{31}P NMR (202.3 MHz, d_6-acetone, ppm): 148.03, 147.68. HRMS (ESI-MS) *m/z* calculated for $C_{74}H_{103}N_9O_{14}PSi_2^+$ 1428.6895: found 1428.6890 $[M+H]^+$.

1-{O^4-[3'-O-(tert-Butyldimethylsilyl)-5'-O-(4,4'-dimethoxytrityl)-N2-phenoxyacetyl-2'-deoxyguanidyl]}-7-{O^4-[5'-O-(tert-butyldimethylsilyl) -3'-O-(2-cyanoethoxy(diisopropylamino)-phosphino)-thymidinyl]}-heptane (**6.11b**)

To a solution of compound **6.10b** (0.20 g, 0.16 mmol) and DIPEA (52 µL, 0.31 mmol) in THF (2 mL) was added Cl-POCEN*i*Pr$_2$ (52 µL, 0.25 mmol) dropwise while stirring. After 30 min, the solvent was evaporated in vacuo and the content was diluted with EtOAc (50 mL), then washed with $NaHCO_3$ (aq, 3%) (2 x 35 mL) followed with brine (35 mL). The solvent was dried over anhydrous Na_2SO_4, decanted and then evaporated to produce a yellow gum. Purification was achieved *via* short flash column chromatography using EtOAc (containing 0.1 % NEt_3) as eluent to afford 0.21 g (89 %) of a colorless foam. R_f (SiO_2 TLC): 0.92, 0.90 EtOAc. $\lambda_{max(MeCN)}$ = 272 nm. ^{1}H NMR (500MHz, d_6-acetone, ppm): 9.30 (s, 1H, NH), 8.23 (s, 1H, H8), 7.80 & 7.80 (m, 1H, H6), 7.43-7.40 (m, 2H, Ar), 7.34-7.16 (m, 6H, Ar), 7.24-7.16 (m, 3H, Ar), 7.05-6.98 (m, 3H, Ar), 6.81-6.75 (m, 4H, Ar), 6.47-6.45 (m, 1H, H1'a), 6.33-6.29 (m, 1H, H1'b), 5.05 (bs, 2H, $PhOCH_2CO$), 4.89 (m, 1H, H3'a), 4.64-4.59 (m, 3H, H3'b & $ArOCH_2$), 4.33-4.31 (m, 2H, $ArOCH_2$), 4.26 & 4.16 (m, 1H, H4'b), 4.07-4.04 (m, 1H, H4'a), 3.99-3.81 (m, 4H, H5'b, H5''b & CH_2OP), 3.75-3.65 (m, 8H, 2x (CH)N & 2 x OCH_3), 3.47-3.44 (m, 1H, H5'a), 3.38-3.35 (m, 1H, H5''a), 3.11-3.07 (m, 1H, H2'a), 2.82-2.78 (m, 2H, CH_2CN), 2.64-2.53 (m, 1H, H2'b), 2.50-2.45 (m, 1H, H2''a), 2.17-2.12 (m, 1H, H2''b), 1.94 (m, 3H, $ArCH_3$), 1.92-1.89 (m, 2H, CH_2), 1.79-1.77 (m, 2H, CH_2), 1.54-1.48 (m, 6H, $(CH_2)_3$), 1.24-1.22 (m, 12H, 4 x CH_3), 0.95-0.94 (m, 9H, $SiC(CH_3)_3$) , 0.87 (m, 9H, $SiC(CH_3)_3$), 0.17-0.15 (m, 6H, 2 x $SiCH_3$), 0.09 (m, 3H, $SiCH_3$), 0.04 (m, 3H, $SiCH_3$). ^{13}C NMR (125.7 MHz, d_6-acetone, ppm): 170.1, 161.0, 158.6, 158.6,

158.2, 154.8, 152.6, 151.6, 145.3, 141.2, 139.7, 139.7, 136.0, 135.9, 130.1, 130.0, 129.5, 128.1, 127.5, 126.5, 121.3, 118.7, 114.8, 112.9, 112.8, 103.3, 103.2, 86.8, 86.3, 86.14, 86.06, 84.5, 74.0, 72.5, 67.9, 67.1, 66.5, 64.1, 63.3, 63.2, 59.6, 58.7, 58.5, 54.6, 43.1, 43.1, 43.02, 42.99, 39.5, 28.3, 25.7, 25.7, 25.5, 25.3, 24.01, 23.97, 23.95, 19.88, 19.83, 18.0, 17.6, 13.6, 11.6, 11.6, -5.37, -5.51, -6.09, -6.13. ^{31}P NMR (202.3 MHz, d_6-acetone, ppm): 148.03, 147.67. HRMS (ESI-MS) *m/z* calculated for $C_{77}H_{109}N_9O_{14}PSi_2^+$ 1470.7365: found 1470.7380 $[M+H]^+$.

The synthesis of phosphoramidites **6.16a** and **6.16b** are shown in **Scheme 6.3**

5'-O-(tert-butyldimethylsilyl)-3'-O-levulinyl-N2-phenoxyacetyl-2'-deoxyguanosine (**6.13**)

*N*2-Phenoxyacetyl-2'-deoxyguanosine (**6.12**) (3.0 g, 7.5 mmol) was co-evaporated thrice with anhydrous pyridine (3 x 10 mL). To a solution of the dried residue in pyridine (45 mL) stirring at 0 ^{o}C was added TBS-Cl (1.2 g, 7.8 mmol) which was then allowed to warm slowly to 22 ^{o}C. After 16h, the solvent was removed in vacuo and the content taken up in DCM (100 mL), washed with distilled water (50 mL) followed with a saturated solution of $NaHCO_{3(aq)}$ (75 mL). A gel-like material formed in the organic layer after the second wash, which was then solubilized by the addition of MeOH (5 mL). The solvent was then dried over anhydrous Na_2SO_4 (≈ 4 g), decanted and evaporated to produce a yellow gum. The residue was placed on high vacuum for 2h before being used for the subsequent step. To a solution of the dried residue, EDAC • HCl (2.9 g, 14.9 mmol), and DMAP (cat) in dioxane (54 mL) was added levulinic acid (1.73 g, 14.9 mmol) and allowed to stir at 22 ^{o}C. After 16h, the solvent was removed in vacuo and the content taken up in DCM (50 mL), washed with a saturated solution of $NaHCO_{3(aq)}$ (2 x 50 mL) followed with brine (50 mL). The solvent was dried over anhydrous Na_2SO_4 (≈ 4 g), decanted and evaporated to produce a yellow gum. The product was purified *via* flash column chromatography

using MeOH : DCM (1.0 % → 1.5 %) as eluent to afford 3.1 g (67 % over two steps) of a slightly yellow solid. R_f (SiO_2 TLC):0.46 MeOH : DCM (5 %). $\lambda_{max(MeCN)}$ = 285, 259 nm. ^{1}H NMR (500MHz, $CDCl_3$, ppm): 11.80 (br, 1H, NH), 9.40 (br, 1H, NH), 8.06 (s, 1H, H8), 7.34-7.31 (m, 2H, Ar), 7.05 (m, 1H, Ar), 6.99-6.97 (m, 2H, Ar), 6.26 (dd, 1H, H1', J = 8, 6.5 Hz), 5.38 (m, 1H, H3'), 4.70 (s, 2H, $PhOCH_2CO$), 4.16 (m, 1H, H4'), 3.85-3.78 (m, 2H, H5' & H5''), 2.79-2.77 (m, 2H, $COCH_2$), 2.60-2.55 (m, 2H, $OOCCH_2$), 2.10 (s, 3H, $COCH_3$), 0.86 (s, 9H, $SiC(CH_3)_3$), 0.059 (s, 3H, $SiCH_3$), 0.057 (s, 3H, $SiCH_3$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 206.5, 172.3, 169.7, 156.5, 155.3, 147.6, 146.3, 137.0, 129.9, 122.9, 121.6, 114.9, 85.7, 83.8, 75.5, 67.03, 66.98, 63.6, 38.9, 37.8, 29.8, 27.9, 25.9, 25.8, 18.3, -5.45, -5.61. IR (thin film); ν_{max} (cm^{-1}) = 3417, 2956, 2928, 2856, 2363, 2335, 1717, 1700, 1616, 1559, 1506, 1496, 1473, 1457, 1405, 1360, 1258, 1208, 1155, 1075, 1004, 936, 838, 782. HRMS (ESI-MS) *m/z* calculated for $C_{29}H_{40}N_5O_8Si^+$ 614.2641: found 614.2640 $[M+H]^+$.

1-{O^4-[3'-O-(tert-Butyldimethylsilyl)-5'-O-(4,4'-dimethoxytrityl)- thymidinyl]}-4-{O^4-[5'-O-(tert-butyldimethylsilyl)-3'-O-levulinyl-N2-phenoxyacetyl-2'-deoxyguanidyl]}-butane (**6.14a**)

To a solution of **6.13** (0.76 g, 1.2 mmol), **6.4a** (1.0 g, 1.4 mmol) and Ph_3P (0.68, 2.6 mmol) in dioxane (4.0 mL) was added DIAD (0.51 mL, 2.6 mmol) dropwise over 5 min while stirring. After 16h, additional DIAD (0.11 mL, 0.54 mmol) was added and the reaction was allowed to stir for another 2h. The solvent was removed in vacuo and the content taken up in EtOAc (75 mL) then washed with a 3 % (w/v) aqueous solution of $NaHCO_3$ (2 x 75 mL) followed with brine (75 mL). The solvent was dried over anhydrous Na_2SO_4 (≈ 4 g), decanted and evaporated to produce a yellow gum. The product was purified *via* flash column chromatography using EtOAc : hexanes (3 : 2 → 100 % EtOAc) as eluent to afford 0.85 g (52 %) of a colorless foam. R_f (SiO_2 TLC): 0.45 EtOAc. $\lambda_{max(MeCN)}$ = 270 nm. ^{1}H NMR (500MHz,

$CDCl_3$, ppm): 8.77 (broad s, 1H, NH), 8.23 (s, 1H, H8), 7.93 (d, 1H, H6, *J* = 1 Hz), 7.41-7.21 (m, 11H, Ar), 7.05-7.00 (m, 3H, Ar), 6.84-6.80 (m, 4H, Ar), 6.51 (dd, 1H, H1'a, *J* = 6 & 8 Hz), 6.32 (dd, 1H, H1'b, *J* = 6 Hz), 5.43 (m, 1H, H3'a), 4.77 (broad s, 2H, $PhOCH_2CO$), 4.64 (t, 2H, $ArOCH_2$, *J* = 6 Hz), 4.49-4.43 (m, 3H, $ArOCH_2$ & H3'b), 4.19 (m, 1H, H4'a), 3.96 (m, 1H, H4'b), 3.90-3.88 (m, 2H, H5'a & H5''a), 3.78 (s, 6H, 2 x OCH_3), 3.51 (dd, 1H, H5'b, J= 3 & 11 Hz), 3.25 (d, 1H, H5''b, *J* = 3 & 11 Hz), 2.78 (m, 2H, $OOCCH_2$), 2.67-2.61 (m, 4H, CH_2CO, H2'a & H2''a), 2.48 (m, 1H, H2'b), 2.22-2.17 (m, 4H, H2''b & $COCH_3$), 2.04-1.96 (m, 4H, CH_2CH_2), 1.50 (d, 3H, $ArCH_3$, *J* = 1 Hz), 0.90 (s, 9H, $SiC(CH_3)_3$), 0.79 (s, 9H, $SiC(CH_3)_3$), 0.10 (s, 3H, $SiCH_3$), 0.09 (s, 3H, $SiCH_3$), -0.01 (s, 3H, $SiCH_3$), - 0.08 (s, 3H, $SiCH_3$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 206.3, 172.4, 170.3, 161.0, 158.7, 157.2, 155.9, 152.5, 151.2, 144.4, 139.8, 139.5, 135.49, 135.47, 130.08, 130.06, 129.8, 128.2, 127.9, 127.0, 122.3, 118.6, 114.9, 113.2, 113.2, 104.5, 86.6, 86.5, 86.2, 85.7, 84.0, 75.7, 71.0, 68.0, 67.2, 66.7, 63.7, 62.4, 55.2, 42.2, 39.1, 37.8, 29.8, 28.0, 26.0, 25.7, 25.4, 25.2, 18.4, 17.9, 11.7, -4.64, -4.97, -5.40, -5.56. IR (thin film); ν_{max} (cm^{-1}) = 3411, 3060, 2956, 2929, 2856, 2362, 2336, 1734, 1718, 1669, 1607, 1532, 1509, 1472, 1437, 1329, 1301, 1251, 1177, 1154, 1070, 836, 781. HRMS (ESI-MS) *m/z* calculated for $C_{70}H_{92}N_7O_{15}Si_2^+$ 1326.6184: found 1326.6191 $[M+H]^+$.

1-{O^4-[3'-O-(tert-Butyldimethylsilyl)-5'-O-(4,4'-dimethoxytrityl)- thymidinyl]}-7-{O^4-[5'-O-(tert-butyldimethylsilyl)-3'-O-levulinyl-N2-phenoxyacetyl-2'-deoxyguanidyl]}-heptane (**6.14b**)

To a solution of **6.13** (0.71 g, 1.2 mmol), **6.4b** (1.0 g, 1.3 mmol) and Ph_3P (0.64, 2.5 mmol) in dioxane (3.8 mL) was added DIAD (0.48 mL, 2.5 mmol) dropwise over 5 min while stirring. After 16h, the solvent was removed in vacuo and the content taken up in EtOAc (75 mL) then washed with a 3 % (w/v) aqueous solution of $NaHCO_3$ (2 x 75 mL) followed with brine (75 mL). The solvent was dried over anhydrous Na_2SO_4 (≈ 4 g), decanted and evaporated to produce

a yellow gum. The product was purified *via* flash column chromatography using EtOAc : hexanes (1 : 1 → 4 : 1) as eluent to afford 0.76 g (48 %) of a colorless foam. R_f (SiO_2 TLC): 0.70 EtOAc. $\lambda_{max(MeCN)}$ = 270 nm. ^{1}H NMR (500MHz, $CDCl_3$, ppm): 8.75 (broad s, 1H, NH), 8.22 (s, 1H, H8), 7.91 (d, 1H, H6, *J* = 1 Hz), 7.68-7.20 (m, 11H, Ar), 7.04-7.01 (m, 3H, Ar), 6.84-6.80 (m, 4H, Ar), 6.51 (dd, 1H, H1'a, *J* = 6 & 9 Hz), 6.32 (dd, 1H, H1'b, *J* = 6 Hz), 5.42 (m, 1H, H3'a), 4.79 (broad s, 2H, $PhOCH_2CO$), 4.58 (t, 2H, $ArOCH_2$, *J* = 6.5 Hz), 4.47 (m, 1H, H3'b), 4.35 (t, 2H, $ArOCH_2$, *J* = 7 Hz), 4.19 (m, 1H, H4'a), 3.95 (m, 1H, H4'b), 3.90-3.88 (m, 2H, H5'a & H5''a), 3.78 (s, 6H, 2 x OCH_3), 3.51 (dd, 1H, H5'b, J= 3 & 11 Hz), 3.24 (dd, 1H, H5''b, *J* = 3 & 11 Hz), 2.78 (m, 2H, $OOCCH_2$), 2.67-2.61 (m, 4H, CH_2CO, H2'a & H2''a), 2.48 (m, 1H, H2'b), 2.22-2.17 (m, 4H, H2''b & $COCH_3$), 1.89 (m, 2H, CH_2), 1.75-1.71 (m, 4H, 2 x CH_2), 1.50 (d, 3H, $ArCH_3$, *J* = 1 Hz), 1.43-1.40 (m, 4H, 2 x CH_2), 0.90 (s, 9H, $SiC(CH_3)_3$), 0.79 (s, 9H, $SiC(CH_3)_3$), 0.10 (s, 3H, $SiCH_3$), 0.09 (s, 3H, $SiCH_3$), -0.02 (s, 3H, $SiCH_3$), - 0.08 (s, 3H, $SiCH_3$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 206.3, 172.4, 170.5, 161.2, 158.6, 155.9, 152.5, 151.2, 144.4, 139.7, 139.3, 135.5, 132.1, 132.0, 130.08, 130.06, 129.8, 128.5, 128.4, 128.2, 127.9, 127.0, 122.3, 118.6, 114.9, 113.2, 113.2, 104.6, 86.6, 86.5, 86.2, 85.7, 84.0, 75.7, 71.1, 68.0, 67.8, 67.3, 63.7, 62.4, 55.2, 42.2, 39.1, 37.8, 29.8, 29.1, 28.9, 28.5, 28.0, 26.0, 25.9, 25.7, 18.4, 17.9, 14.2, 11.8, -4.64, -4.97, -5.40, -5.56. IR (thin film); ν_{max} (cm^{-1}) = 3411, 3059, 2954, 2930, 2856, 2362, 2336, 1734, 1718, 1669, 1607, 1529, 1509, 1472, 1437, 1378, 1328, 1301, 1252, 1177, 1155, 1071, 836, 782, 734. HRMS (ESI-MS) *m/z* calculated for $C_{73}H_{98}N_7O_{15}Si_2^+$ 1368.6654: found 1368.6661 $[M+H]^+$.

1-{O^4-[3'-O-(tert-Butyldimethylsilyl)-5'-O-(4,4'-dimethoxytrityl)-thymidinyl]}-4-{O^4-[5'-O-(tert-butyldimethylsilyl)-N2-phenoxyacetyl-2'-deoxyguanidyl]}-butane (**6.15a**)

A solution (2.8 mL) of 0.5 M N_2H_4•H_2O in pyridinium acetate (1/1, v/v) was added to **6.14a** (0.85 g, 0.64 mmol). After 30 min of stirring, the reaction was diluted with EtOAc (70 mL added), washed several times with a 3 % (w/v) aqueous solution of $NaHCO_3$ (6 x 75 mL) followed with brine (75 mL). The solvent was dried over anhydrous Na_2SO_4 (≈ 4 g), decanted and evaporated to produce a yellow gum. The product was purified *via* flash column chromatography using MeOH : DCM (1.5 % → 2.5 %) as eluent to afford 0.57 g (73 %) of a colorless foam. R_f (SiO_2 TLC):0.17 MeOH : DCM (4 %). $\lambda_{max(MeCN)}$ = 270 nm.. ^{1}H NMR (500MHz, $CDCl_3$, ppm): 8.85 (broad s, 1H, NH), 8.23 (s, 1H, H8), 7.95 (d, 1H, H6, *J* = 1 Hz), 7.42-7.40 (m, 2H, Ar), 7.36-7.23 (m, 9H, Ar), 7.07-7.01 (m, 3H, Ar), 6.84-6.82 (m, 4H, Ar), 6.70 (dd, 1H, H1'a, *J* = 6.8 Hz), 6.33 (dd, 1H, H1'b, *J* = 6 Hz), 4.77 (m, 3H, H3'a & $PhOCH_2CO$), 4.62 (t, 2H, $ArOCH_2$, *J* = 6 Hz), 4.50-4.43 (m, 3H, $ArOCH_2$ & H3'b), 4.17 (m, 1H, H4'a), 3.97 (m, 1H, H4'b), 3.89-3.88 (m, 2H, H5'a & H5''a), 3.79 (s, 6H, 2 x OCH_3), 3.52 (dd, 1H, H5'b, J= 3 & 11 Hz), 3.26 (dd, 1H, H5''b, *J* = 3 & 11 Hz), 2.64-2.61 (m, 2H, H2'a & H2''a), 2.49 (m, 1H, H2'b), 2.20 (m, 1H, H2''b), 2.03-1.96 (m, 4H, CH_2CH_2), 1.51 (d, 3H, $ArCH_3$, *J* = 1 Hz), 0.91 (s, 9H, $SiC(CH_3)_3$), 0.80 (s, 9H, $SiC(CH_3)_3$), 0.10 (s, 3H, $SiCH_3$), 0.09 (s, 3H, $SiCH_3$), 0.00 (s, 3H, $SiCH_3$), - 0.07 (s, 3H, $SiCH_3$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 170.4, 161.0, 158.7, 157.1, 155.9, 152.4, 150.9, 144.4, 140.4, 139.5, 135.49, 135.47, 130.09, 130.06, 129.8, 128.2, 127.9, 127.1, 122.4, 118.8, 114.9, 113.21, 113.18, 104.6, 87.8, 86.6, 86.5, 86.2, 84.3, 72.5, 71.0, 67.9, 67.1, 66.7, 63.9, 62.3, 58.5, 55.2, 42.2, 41.6, 29.7, 26.0, 25.7, 25.4, 25.2, 18.4, 17.9, 11.7, 8.26, -4.63, -4.96, -5.36, -5.50. IR (thin film); ν_{max} (cm^{-1}) = 3411, 3059, 2957, 2928, 2856, 2362, 2335, 1717, 1663, 1608, 1532, 1509, 1472, 1437, 1381, 1329, 1301, 1251, 1176, 1150, 1064, 1034, 835, 781, 734. HRMS (ESI-MS) *m/z* calculated for $C_{65}H_{86}N_7O_{13}Si_2^+$ 1228.5817: found 1228.5821 $[M+H]^+$.

1-{O^4-[3'-O-(tert-Butyldimethylsilyl)-5'-O-(4,4'-dimethoxytrityl)-thymidinyl]}-7-{O^4-[5'-O-(tert-butyldimethylsilyl)-N2-phenoxyacetyl-2'-deoxyguanidyl]}-heptane (**6.15b**)

A solution (2.4 mL) of 0.5 M N_2H_4•H_2O in pyridinium acetate (1/1, v/v) was added to **6.14b** (0.74 g, 0.54 mmol). After 30 min of stirring, the reaction was diluted with EtOAc (70 mL added), then washed several times with a 3 % (w/v) aqueous solution of $NaHCO_3$ (6 x 75 mL) followed with brine (75 mL). The solvent was dried over anhydrous Na_2SO_4 (≈ 4 g), decanted and evaporated to produce a yellow gum. The product was purified *via* flash column chromatography using MeOH : DCM (1.5 % → 2.5 %) as eluent to afford 0.55 g (80 %) of a colorless foam. R_f (SiO_2 TLC):0.23 MeOH : DCM (4 %). $\lambda_{max(MeCN)}$ = 270 nm. ^{1}H NMR (500MHz, $CDCl_3$, ppm): 8.80 (broad s, 1H, NH), 8.21 (s, 1H, H8), 7.92 (d, 1H, H6, *J* = 1 Hz), 7.42-7.40 (m, 2H, Ar), 7.36-7.21 (m, 9H, Ar), 7.06-7.01 (m, 3H, Ar), 6.84-6.81 (m, 4H, Ar), 6.66 (dd, 1H, H1'a, *J* = 6.5 Hz), 6.33 (dd, 1H, H1'b, *J* = 6 Hz), 4.75 (m, 3H, H3'a & $PhOCH_2CO$), 4.57 (t, 2H, $ArOCH_2$, *J* = 6.5 Hz), 4.47 (m, 1H, H3'b), 4.35 (t, 2H, $ArOCH_2$, *J* = 7 Hz), 4.15 (m, 1H, H4'a), 3.96 (m, 1H, H4'b), 3.89-3.88 (m, 2H, H5'a & H5''a), 3.79 (s, 6H, 2 x OCH_3), 3.52 (dd, 1H, H5'b, J= 3 & 11.5 Hz), 3.26 (dd, 1H, H5''b, *J* = 3 & 11.5 Hz), 2.64-2.61 (m, 2H, H2'a & H2''a), 2.49 (m, 1H, H2'b), 2.20 (m, 1H, H2''b), 1.88 (m, 2H, CH_2), 1.74 (m, 2H, CH_2), 1.51 (m, 5H, CH_2 & $ArCH_3$), 1.43-1.41 (m, 4H, CH_2CH_2) 0.91 (s, 9H, $SiC(CH_3)_3$), 0.80 (s, 9H, $SiC(CH_3)_3$), 0.10 (s, 3H, $SiCH_3$), 0.09 (s, 3H, $SiCH_3$), -0.01 (s, 3H, $SiCH_3$), - 0.07 (s, 3H, $SiCH_3$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 170.5, 161.2, 158.7, 156.0, 152.3, 151.0, 144.4, 140.3, 139.3, 135.5, 132.1, 132.0, 130.09, 130.07, 129.8, 128.5, 128.4, 128.2, 127.9, 127.0, 122.4, 118.8, 114.9, 113.20, 113.18, 104.6, 87.6, 86.6, 86.5, 86.2, 84.2, 72.5, 71.1, 68.0, 67.8, 67.3, 67.1, 63.8, 62.4, 55.2, 42.2, 41.5, 29.1, 28.8, 28.5, 26.0, 25.9, 25.7, 18.4, 17.9, 11.8, -4.64, -4.96, -5.36, -5.50. IR (thin film); ν_{max} (cm^{-1}) = 3411, 3059, 2953, 2930, 2856, 2362, 2336, 1717,

1668, 1607, 1531, 1509, 1471, 1437, 1379, 1328, 1302, 1251, 1176, 1110, 1066, 1034, 835, 781, 733. HRMS (ESI-MS) *m/z* calculated for $C_{68}H_{91}N_7O_{13}Si_2Na^+$ 1292.6106: found 1292.6110 $[M+Na]^+$.

1-{O⁴-[3'-O-(tert-Butyldimethylsilyl)-5'-O-(4,4'-dimethoxytrityl)-thymidinyl]}-4-{O⁴-[5'-O-(tert-butyldimethylsilyl)-3'-O-(2-cyanoethoxy(diisopropylamino)-phosphino)-N2-phenoxyacetyl-2'-deoxyguanidyl]}-butane (**6.16a**)

To a solution of compound **6.15a** (0.20 g, 0.16 mmol) and DIPEA (53 µL, 0.31 mmol) in THF (2 mL) was added Cl-POCEN*i*Pr$_2$ (54 µL, 0.25 mmol) dropwise while stirring. After 35 min, the solvent was evaporated in vacuo and the content was diluted with EtOAc (50 mL), then washed with 3% (w/v) $NaHCO_{3(aq)}$ (2 x 35 mL) followed by brine (35 mL). The solvent was dried over anhydrous Na_2SO_4, decanted and evaporated to produce a yellow gum. Purification was achieved *via* short flash column chromatography using an EtOAc (containing 0.1 % NEt_3) as eluent to afford 0.19 g (83 %) of a colorless foam. R_f (SiO_2 TLC): 0.84, 0.76 EtOAc. $\lambda_{max(MeCN)}$ = 271 nm. ^{1}H NMR (500MHz, d_6-acetone, ppm)): 9.43 (s, 1H, NH), 8.28-8.27 (s, 1H, H8), 7.94 (m, 1H, H6), 7.51-7.49 (m, 2H, Ar), 7.40-7.25 (m, 9H, Ar), 7.06-6.97 (m, 3H, Ar), 6.93-6.90 (m, 4H, Ar), 6.51-6.47 (m, 1H, H1'a), 6.29-6.27 (m, 1H, H1'b), 5.10 (s, 2H, $PhOCH_2CO$), 4.89-4.83 (m, 1H, H3'a), 4.69-4.61 (m, 3H, H3'b & $ArOCH_2$), 4.44-4.41 (m, 2H, $ArOCH_2$), 4.26 & 4.17 (m, 1H, H4'a), 4.03-3.62 (m, 13H, H4'b, H5'a, H5''a, CH_2OP, 2x (CH)N & 2 x OCH_3), 3.51-3.48 (m, 1H, H5'b), 3.37-3.34 (m, 1H, H5''b), 3.07-2.98 (m, 1H, H2'a), 2.82-2.61 (m, 3H, H2"a & CH_2CN), 2.43-2.38 (m, 1H, H2'b), 2.34-2.29 (m, 1H, H2''b) , 2.05-2.02 (m, 2H, CH_2, overlapping with acetone solvent residual peak), 2.00-1.96 (m, 2H, CH_2), 1.61 (m, 3H, $ArCH_3$), 1.25-1.20 (m, 12H, 4 x CH_3), 0.93-0.91 (m, 9H, $SiC(CH_3)_3$) , 0.87-0.85 (m, 9H, $SiC(CH_3)_3$), 0.10 (s, 3H, $SiCH_3$), 0.09 (m, 3H, $SiCH_3$), 0.07 (m, 3H, $SiCH_3$), 0.02 (s, 3H, $SiCH_3$). ^{13}C NMR (125.7

MHz, d_6-acetone, ppm): 170.0, 166.92, 160.90, 158.9, 158.2, 154.7, 152.80, 151.79, 144.9, 140.6, 140.5, 140.0, 135.7, 135.6, 130.1, 129.4, 128.1, 127.8, 126.9, 121.2, 118.4, 118.1, 114.7, 113.1, 103.3, 87.1, 86.9, 86.8, 86.6, 86.4, 86.0, 84.4, 74.1, 73.9, 73.8, 73.7, 71.8, 67.9, 67.4, 66.7, 66.1, 65.7, 63.4, 63.2, 62.9, 61.3, 59.6, 58.8, 58.7, 58.5, 54.7, 43.2, 43.12, 43.06, 43.02, 41.6, 39.1, 38.8, 25.5, 25.4, 25.3, 25.1, 24.1, 24.03, 24.00, 23.98, 19.92, 19.87, 18.1, 17.6, 13.6, 11.2, -5.33, -5.59, -6.06, -6.09. ^{31}P NMR (202.3 MHz, d_6-acetone, ppm): 148.24, 148.13. HRMS (ESI-MS) *m/z* calculated for $C_{74}H_{103}N_9O_{14}PSi_2^+$ 1428.6895: found 1428.6898 $[M+H]^+$.

1-{O^4-[3'-O-(tert-Butyldimethylsilyl)-5'-O-(4,4'-dimethoxytrityl)-thymidinyl]}-7-{O^4-[5'-O-(tert-butyldimethylsilyl)-3'-O-(2-cyanoethoxy(diisopropylamino)-phosphino)-N2-phenoxyacetyl-2'-deoxyguanidyl]}-heptane (**6.16b**)

To a solution of compound **6.15b** (0.20 g, 0.16 mmol) and DIPEA (52 µL, 0.31 mmol) in THF (2 mL) was added Cl-POCEN*i*Pr$_2$ (52 µL, 0.25 mmol) dropwise while stirring. After 30 min, the solvent was evaporated in vacuo and the content was diluted with EtOAc (50 mL), then washed with $NaHCO_3$ (aq, 3%) (2 x 35 mL) followed with brine (35 mL). The solvent was dried over anhydrous Na_2SO_4, decanted and then evaporated to produce a yellow gum. Purification was achieved *via* short flash column chromatography using an EtOAc (containing 0.1 % NEt_3) as eluent to afford 0.19 g (82 %) of a colorless foam. R_f (SiO_2 TLC): 0.87, 0.86 EtOAc. $\lambda_{max(MeCN)}$ = 271 nm. ^{1}H NMR (500MHz, d_6-acetone, ppm)): 9.42 (s, 1H, NH), 8.28-8.27 (s, 1H, H8), 7.94 (m, 1H, H6), 7.51-7.50 (m, 2H, Ar), 7.36-7.25 (m, 9H, Ar), 7.05-7.03 (m, 2H, Ar), 7.00-6.97 (m, 1H, Ar), 6.92-6.91 (m, 4H, Ar), 6.50-6.47 (m, 1H, H1'a), 6.30-6.27 (m, 1H, H1'b), 5.11 (s, 2H, PhOCH_2CO), 4.88-4.84 (m, 1H, H3'a), 4.64-4.59 (m, 3H, H3'b & ArOCH_2), 4.33-4.30 (m, 2H, ArOCH_2), 4.24 & 4.17 (m, 1H, H4'a), 4.09-3.80 (m, 11H, H4'b, H5'a, H5''a, CH_2OP & 2 x OCH_3), 3.72-3.65 (m, 2H, 2x (CH)N), 3.51-3.48 (m, 1H, H5'b), 3.37-3.34 (m, 1H, H5''b), 3.07-

2.98 (m, 1H, H2'a), 2.82-2.76 (m, 2H, CH_2CN), 2.74-2.61 (m, 1H, H2"a), 2.43-2.38 (m, 1H, H2'b), 2.34-2.29 (m, 1H, H2''b) , 1.92-1.87 (m, 2H, CH_2) 1.78-1.74 (m, 2H, CH_2), 1.61 (m, 3H, $ArCH_3$), 1.55-1.48 (m, 6H, $(CH_2)_3$), 1.26-1.21 (m, 12H, 4 x CH_3), 0.92-0.91 (m, 9H, $SiC(CH_3)_3$) , 0.86-0.85 (m, 9H, $SiC(CH_3)_3$), 0.10 (s, 3H, $SiCH_3$), 0.09 (m, 3H, $SiCH_3$), 0.07 (m, 3H, $SiCH_3$), 0.02 (s, 3H, $SiCH_3$).^{13}C NMR (125.7 MHz, d_6-acetone, ppm): 170.1, 161.0, 158.9, 158.3, 154.8, 152.8, 151.8, 144.9, 140.5, 140.4, 139.9, 135.7, 135.6, 130.12, 130.11, 129.4, 128.1, 127.8, 126.9, 121.2, 118.4, 118.1, 114.7, 113.13, 113.11, 103.3, 87.1, 86.8, 86.6, 86.4, 86.0, 84.4, 84.3, 74.1, 73.9, 73.8, 73.7, 71.8, 67.9, 67.1, 66.5, 63.4, 63.2, 62.9, 59.6, 58.8, 58.7, 58.5, 54.7, 43.2, 43.12, 43.06, 43.0, 41.6, 39.1, 38.8, 28.3, 25.72, 25.71, 25.49, 25.48, 25.3, 24.1, 24.04, 24.01, 23.98, 23.95, 19.93, 19.90, 19.87, 19.85, 18.1, 17.6, 13.6, 11.2, -5.33, -5.58, -6.04, -6.06, -6.08, -6.10. ^{31}P NMR (202.3 MHz, d_6-acetone, ppm): 148.23, 148.13. HRMS (ESI-MS) *m/z* calculated for $C_{77}H_{109}N_9O_{14}PSi_2^+$ 1470.7365: found 1470.7375 $[M+H]^+$.

6.5.1.3 Preparation, Purification and Characterization of the TT, GT, and TG IaCL Oligonucleotide

The cross-linked duplexes, the sequences for which are shown in **Figure 6.1**, were assembled with an Applied Biosystems Model 3400 synthesizer on a 1.5 mmol scale using β-cyanoethylphosphoramidite chemistry supplied by the manufacturer with slight modifications to coupling times. The nucleoside phosphoramidites protected with "fast-deprotecting" groups were prepared in anhydrous MeCN at a concentration of 0.1 M for the 3'-*O*-deoxyphosphoramidites, and 0.13-0.15 M for the cross-linked 3'-*O*-deoxyphosphoramidite. Oligomer sequence assembly was carried out as previously described.[323] The capping step of the assembly was carried out using phenoxyacetic anhydride/pyridine/tetrahydrofuran 1:1:8 (v/v/v; solution A) and 1-methyl-imidazole/ tetrahydrofuran 16:84 (w/v; solution B). Coupling wait times for phosphoramidites

6.7a, **6.7b**, **6.11a**, **6.11b**, **6.16a**, and **6.16b** were extended to 10 min (compared to 2 min for the commercially available phosphoramidites). Protecting group removal and cleavage from the solid support was carried out by treatment with 0.05 M K_2CO_4 in MeOH for 4 h at room temperature with mild rocking in 2 mL screw-cap microfuge tubes fitted with Teflon lined caps. The base was neutralized with an equimolar amount of AcOH and crude oligomers were transferred and lyophilized in a Speedvac concentrator. Silyl protecting groups were removed from oligomers by treatment with NEt_3•3HF (200 μL, pellet initially sonicated (2 x 15 s)) for 16h at room temperature under gentle rocking. Oligomers were precipitated using cool *n*-butanol (400 μL) and the resulting mixture was allowed to cool to -20^{o}C for 10 min, followed by spinning the samples down. The supernatant was removed and the pellet was washed with another aliquot of *n*-butanol (400 μL). Purification was achieved by strong anion exchange HPLC using a Dionex DNAPAC PA-100 column (0.4 cm x 25 cm) purchased from Dionex Corp, Sunnyvale, CA using a linear gradient of 0–52% buffer B over 24 min (buffer A: 100 mm Tris HCl, pH 7.5, 10% MeCN and buffer B: 100 mm Tris HCl, pH 7.5, 10% MeCN, 1M NaCl) at 55 ^{o}C. The columns were monitored at 260 nm for analytical runs or 280 nm for preparative runs. The purified oligomers were desalted using C-18 SEP PAK cartridges (Waters Inc.) as previously described.[317] The molecular mass of the modified oligomers were identified by ESI-MS and the measured values were in agreement with the expected masses (**Supporting Information Figure S6.2 – S6.7** for MS spectra).

6.5.1.4 UV thermal denaturation

Molar extinction coefficients for the unmodified and cross-linked oligonucleotides were calculated from those of the mononucleotides and dinucleotides using the nearest-neighbor approximations (M^{-1} cm^{-1}). All duplexes were prepared by mixing equimolar amounts of the

interacting strands and lyophilizing the mixture to dryness. The resulting pellet was then dissolved in 90mM sodium chloride, 10 mM sodium phosphate, 1 mM EDTA buffer (pH 7.0) to give a final concentration of 3.5 μM duplex. Prior to the thermal run, samples were degassed by placing them in a speed- vac concentrator for 2 min. Annealing curves were acquired at 260 nm starting at 95 ^{o}C and decreasing temperature at a rate of cooling of 0.5 ^{o}C min^{-1} until 15 ^{o}C, using a Varian CARY Model 3E spectrophotometer fitted with a 6-sample thermostated cell block and a temperature controller. The samples were then denatured by heating from 15 ^{o}C to 95 ^{o}C at an increasing temperature rate of 0.5 ^{o}C min^{-1} to show reversibility. Denaturing data processing was carried out as described by Puglisi and Tinoco[254] and transferred to Microsoft Excel™ for viewing.

6.5.1.5 Circular dichroism (CD) spectroscopy

Circular dichroism spectra were obtained on a Jasco J-815 spectropolarimeter equipped with a Julaba F25 circulating bath. Samples were allowed to equilibrate for 5 – 10 min at 15 ^{o}C in 90 mM sodium chloride, 10 mM sodium phosphate, 1 mM EDTA (pH 7.0), at a final concentration of 3.4 μM. Each spectrum was an average of 3 scans, collecting at a rate of 50 nm min^{-1} , with a bandwidth of 1 nm and sampling wavelength of 0.2 nm using fused quartz cells (Starna 29-Q-10). The CD spectra were recorded from 350 to 220 nm at 15 ^{o}C. The molar ellipticity (φ) was calculated from the equation $\varphi = \varepsilon/Cl$, where ε is the relative ellipticity (mdeg), C is the molar concentration of oligonucleotides (moles/L), and l is the path length of the cell (cm). The data were processed using software supplied by the manufacturer (JASCO, Inc.) and transferred into Microsoft Excel™ for viewing.

6.5.1.6 Molecular modelling

Molecular modeling was performed by using the Hyperchem 7.5 software package from Hypercube utilizing the AMBER force field. Hybridized oligomers containing a dGG·dCC, dGpG4·dCC, dGpG4·dCC base pair were constructed from the nucleic acid template option using a B-form duplex. Sequence contexts were truncated to the following 5'-GGC TXX GAT C and 3'-CCG AYY CTA G for proper solvation. Duplexes were solvated with water using a periodic box occupying at least 3 times the volume of the duplex alone. Standard Amber99 parameters were used with the dielectric set to constant. "One to four scale factors" non-bonded interactions were set to 0.5 (both electrostatic and van der Waals). Cutoffs were applied to "switched" to an outer and inner radius of 14.5 and 10.5 Å. All structures were geometry optimized using Polak-Ribiere conjugate gradient until the RMS gradient was less than 0.1 kcal/(Å mol) using the periodic boundary condition option.

6.5.1.7 Protein expression and purification

Ampicillin, isopropyl β-D-thiogalactopyranoside (IPTG), and most other biochemical reagents as well as polyacrylamide gel materials were purchased from Bioshop Canada Inc (Burlington, ON). Ni-NTA Superflow Resin was purchased from Qiagen (Mississauga, ON). Complete, Mini, EDTA-free Protease Inhibitor Cocktail Tablets were obtained from Roche (Laval, QC). Nitro-cellulose filters (0.20 μm) were obtained from Millipore. XL-10 Gold and Bl21(DE3) *E. coli* cells were obtained from Stratagene (Cedar Creek, TX). T4 polynucleotide kinase (PNK) was obtained from Fermentas (Burlington, ON). [γ-^{32}P]ATP was purchased from PerkinElmer (Woodbridge, ON). Phusion Polymerase was obtained from New England Biolabs (Ipswich, MA). DNA primers for site directed mutagenesis and cloning were purchased from Biocorp (Montreal, QC). All AGT homologues were expressed under the promoter of the pQE30 vector in XL-10 Gold *E. coli* cells, as previously described.[65,149–151]

6.5.1.8 AGT repair assay of IaCL DNA duplexes

DNA substrates were labeled at the using γ-[^{32}P]-ATP as previously described.[65] Briefly, a 20 μM solution of DNA was made in 1X PNK buffer along with 1 μL γ-[^{32}P]-ATP (10 μCi μL^{-1}) and 5 units of T4 PNK. The labelling reaction was conducted for 1 h at 37 °C after which the reaction was terminated by boiling the sample for 5 min. 100 pmol of labeled DNA was added to 110 pmol of the complement strand in a total volume of 50 μL of water making a 2 μM dsDNA solution with 10% excess of the non-damaged strand. The solution was boiled for 2 min, cooled slowly to room temperature and kept in a refrigerator at 4 °C overnight to ensure proper annealing of the duplex. The repair reaction mixtures were constituted of 2 pmol of the DNA duplex and 10 pmol or 60 pmol of AGT in a total volume of 15 μL of Activity Buffer [10 mM Tris–HCl (pH 7.6), 100 mM NaCl and 1 mM DTT] and allowed to react at 37 °C overnight. The reaction was terminated by the addition of 18.2 μL of stop buffer [81 mM Tris–HCl, 81 mM boric acid, 1.8 mM EDTA and 1% SDS (sodium dodecyl sulfate) (pH 8.0) in 80% formamide] followed by boiling for 2 min. Samples (10 μL) were loaded on a 14 cm × 16 cm, 20% 7 M urea denaturing polyacrylamide gel (19 : 1) for separation. The gels were run using 1X TBE for 1h-1.5h at 400 V and the gels exposed to a storage phosphor screen. The image was captured on a Typhoon 9400 (GE Healthcare, Piscataway, NJ) and the autoradiography counts obtained by Image-Quant™ (Amersham Biosciences). For the repair time course assays, master mixes of 150 μL composed of 100 pmol (5-fold) or 600 pmol (30-fold) of AGT and 20 pmol of DNA substrate were prepared. Each sample was placed at 37 ^{o}C and at each time point, 7.5 μL was removed from the master mix and placed in a tube with 9.1 μL of stop reaction buffer and analysis was achieved as described above. All samples were analyzed on polyacrylamide gel electrophoresis (using 1X TBE for 1 h-1.5 h at 400 V) and visualized as described above.

6.5.1.9 Identification of IaCL Repair Product by SDS-PAGE

600 pmol of hAGT or Ada-C was incubated with 600 pmol of IaCL (**GG4**, **GG7**, **TG4** or **TG7** in 25 μL of Activity Buffer [10 mM Tris–HCl (pH 7.6), 100 mM NaCl and 1 mM DTT] overnight at 37 °C. The reaction was terminated by adding 6 μL of SDS Buffer [200 mM Tris-HCl (pH 7.6), 3.2% SDS, 48% glycerol and 0.4 M DTT], boiling the sample for 2 min, and separated on a 15% SDS- PAGE. The gel was stained with Coomassie Blue Stain as per the manufacturer's protocol. Detection and visualization of DPC products was conducted with Typhoon Trio Variable Mode Imager in fluorescence mode with the green (532 nm) laser in conjunction with the 526-nm short-pass filter.

6.5.1.10 Identification of Repair Product (DPC) from hAGT-mediated Repair of IaCL DNA by LC-MS

600 pmol of hAGT was incubated with 600 pmol of **GG4**, **GG7**, **TG4** or **TG7** in 25 μL of Activity Buffer [10 mM Tris–HCl (pH 7.6), 100 mM NaCl and 1 mM DTT] for 30 min at 37 °C. A reaction aliquot (5 μL) was diluted in aqueous 0.1 % (v/v) formic acid subject to HPLC (Agilent 1200 Series system) using a Grace^{TM} Vydac^{TM} C4 (214MS) column (Fisher Scientific) (100mm x 2.1 mm) operated at a flow rate of 0.25 mL/min at room temperature (22 ^{o}C) with the following gradient method: 0 - 0.1 min, linear gradient from 10 - 20 % B, 0.1 - 6 min, linear gradient from 20 - 95 % B, 6 - 8 min, hold at 95 % B, 8 - 10 min, linear gradient from 95 – 10 % B, 10 - 19 min, hold at 10 % B (buffer A, 0.1% formic acid in water and buffer B, 0.1% formic acid in acetonitrile). The LC system was interfaced to a Micromass Q-Tof Ultima API equipped with an electrospray source set with the following conditions: source voltage 3.5 kV, mass range of 700−1999 m/z in positive ion mode. The theoretical mass of the hAGT-DNA species were calculated by summing the mass of hAGT (21876 Da), either a butylene linker (54 Da) or

heptylene linker (96 Da), and the corresponding fragments GGCTT (1494 Da) or GGCTG (1519 Da). The expected masses are reported in **Figure 6.7** of the main text. If the repair event had occurred at 5'-residue of the IaCL adduct, DPC masses would have been significantly different since the DNA fragment differed in size and sequence context (GGA TCA CCAG (3037 Da)).

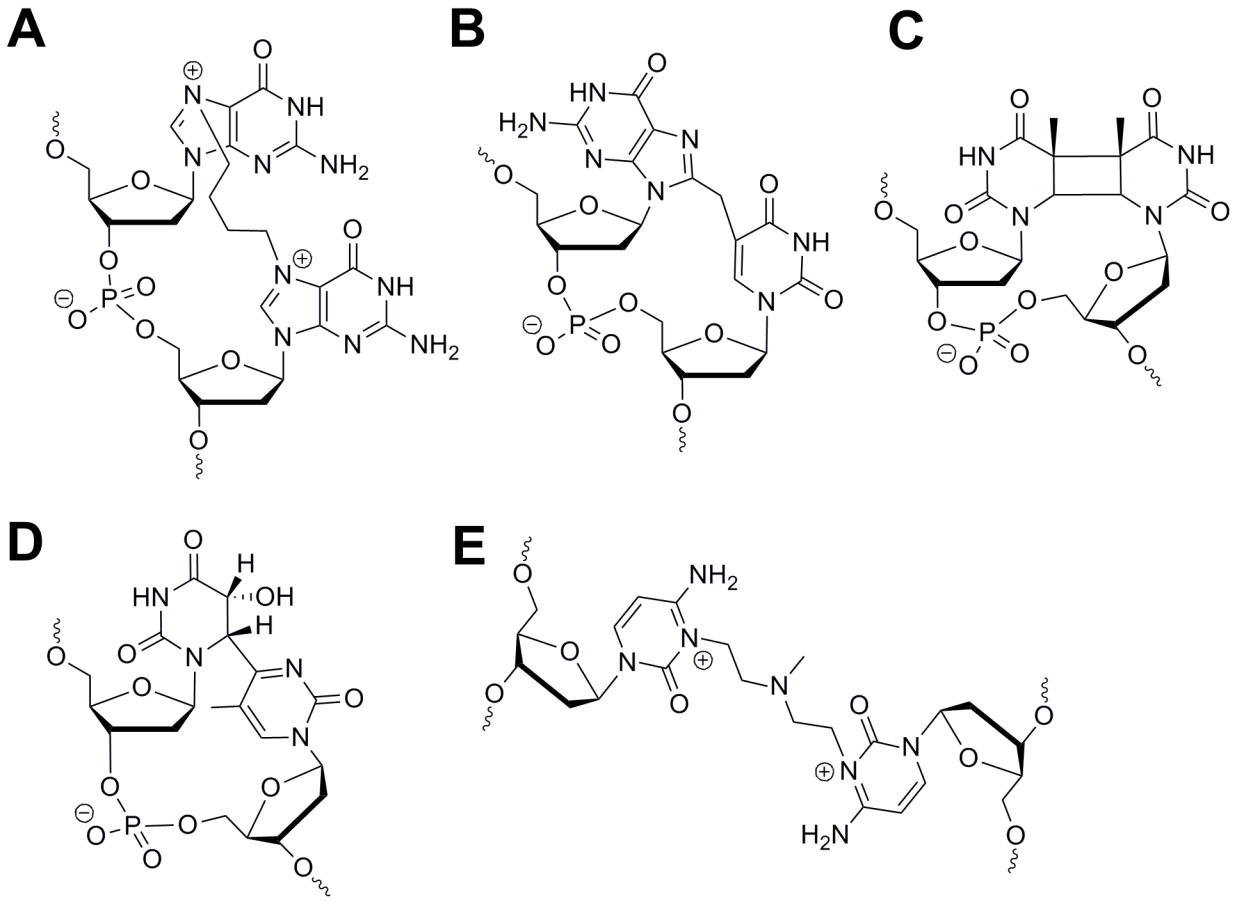


Figure S6.1: Examples of different DNA adducts mentioned in the main text. **A** An example of an IaCL formed from exposure of DNA to the chemotherapeutic drug busulfan.[69] **B** An example of γ-radiation induced IaCL lesion (dG[8,5-Me]dT).[3,223] **C** An example of a UV-induced cyclubutane pyrimidine dimer. [207] **D** An example of a UV-induced (6-4) pyrimidone photoproduct lesion.[82] **E** An example of an ICL lesion generated by mechlorethamine between two 2'-deoxycytosine residues.[364]

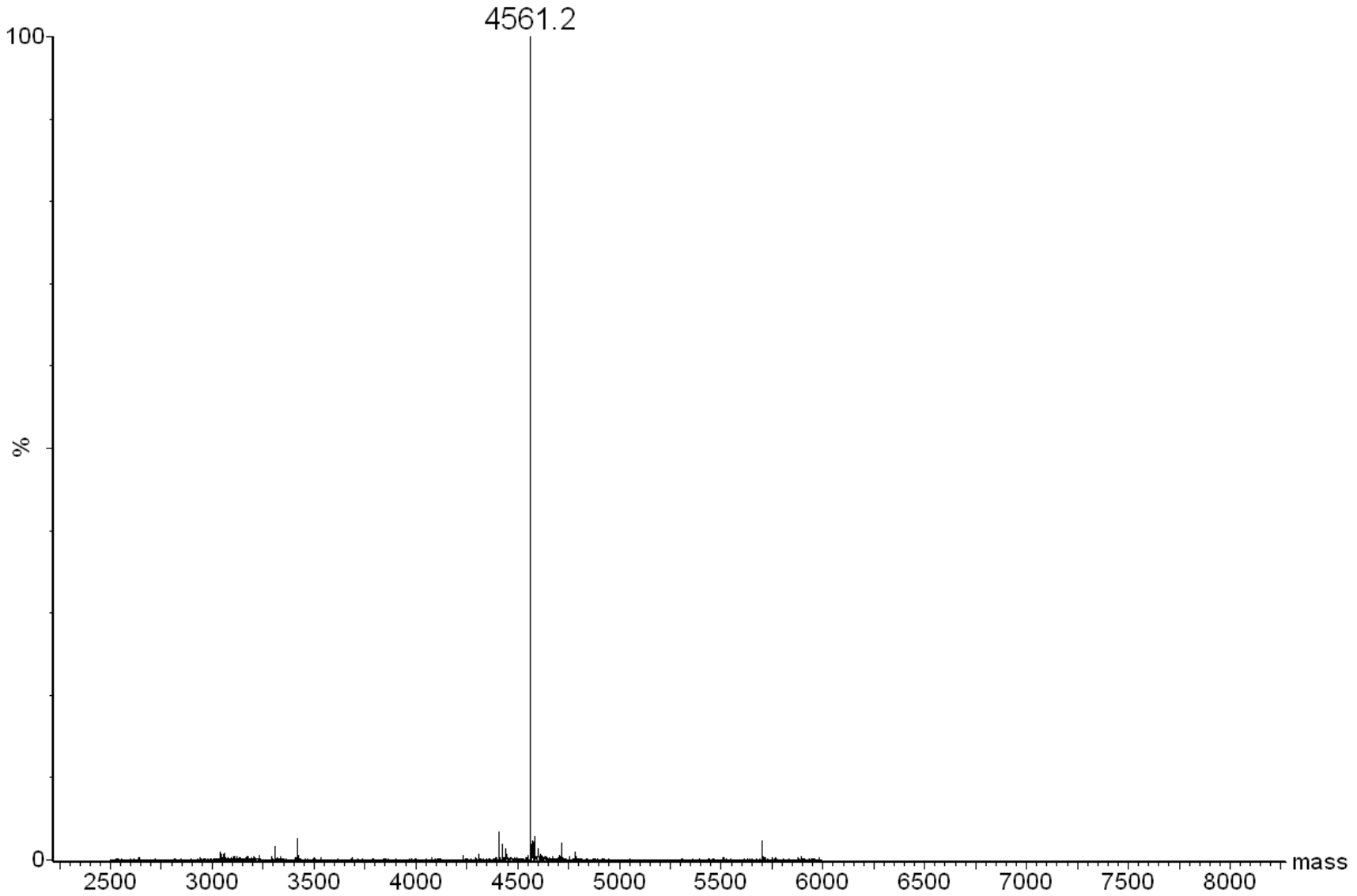


Figure S6.2 - Deconvoluted ESI MS spectrum of oligonucleotide **TT4** (expected mass of 4560.1)

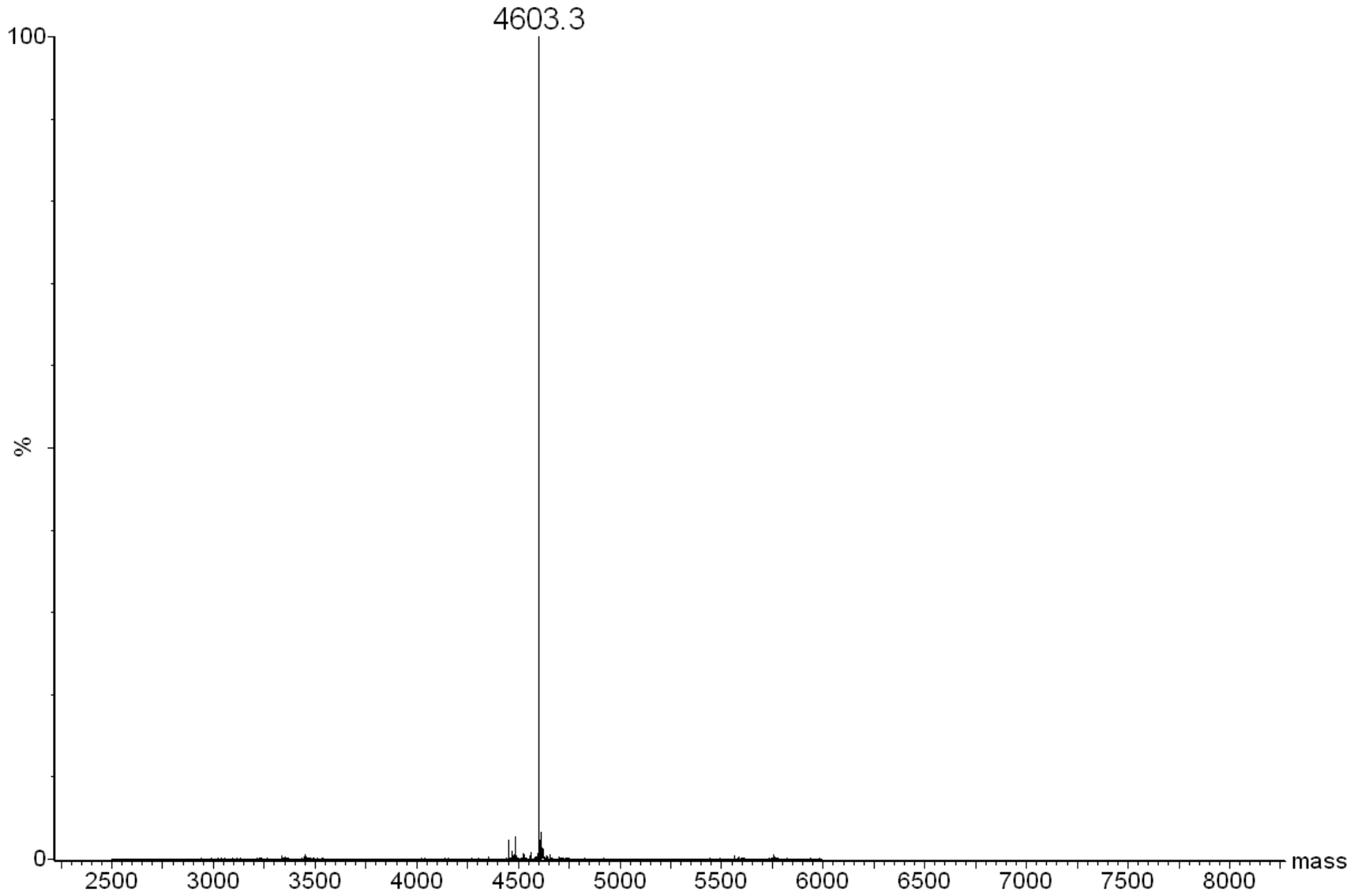


Figure S6.3 - Deconvoluted ESI MS spectrum of oligonucleotide **TT7** (expected mass of 4562.2)

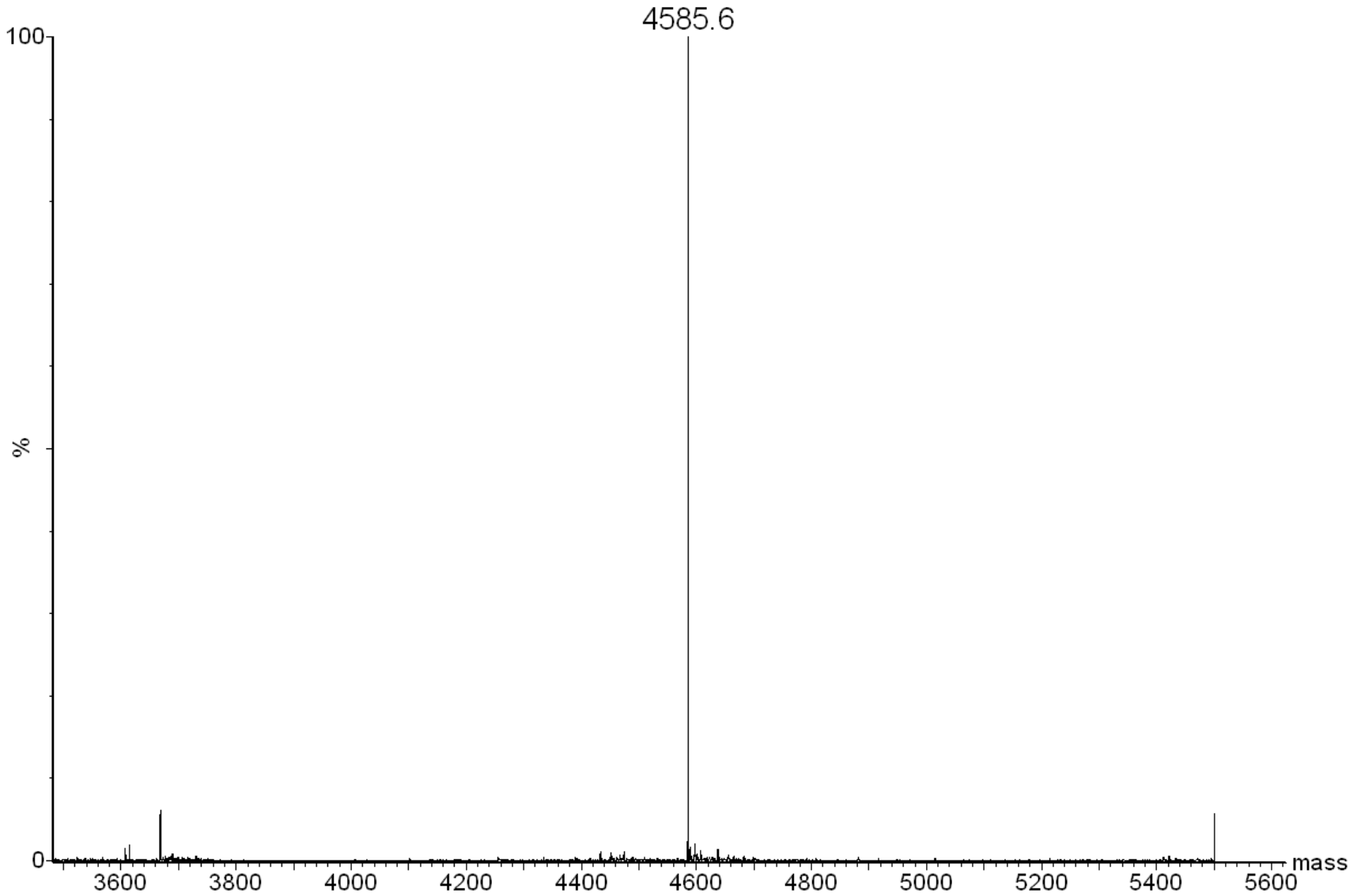


Figure S6.4 - Deconvoluted ESI MS spectrum of oligonucleotide **GT4** (expected mass of 4585.1)

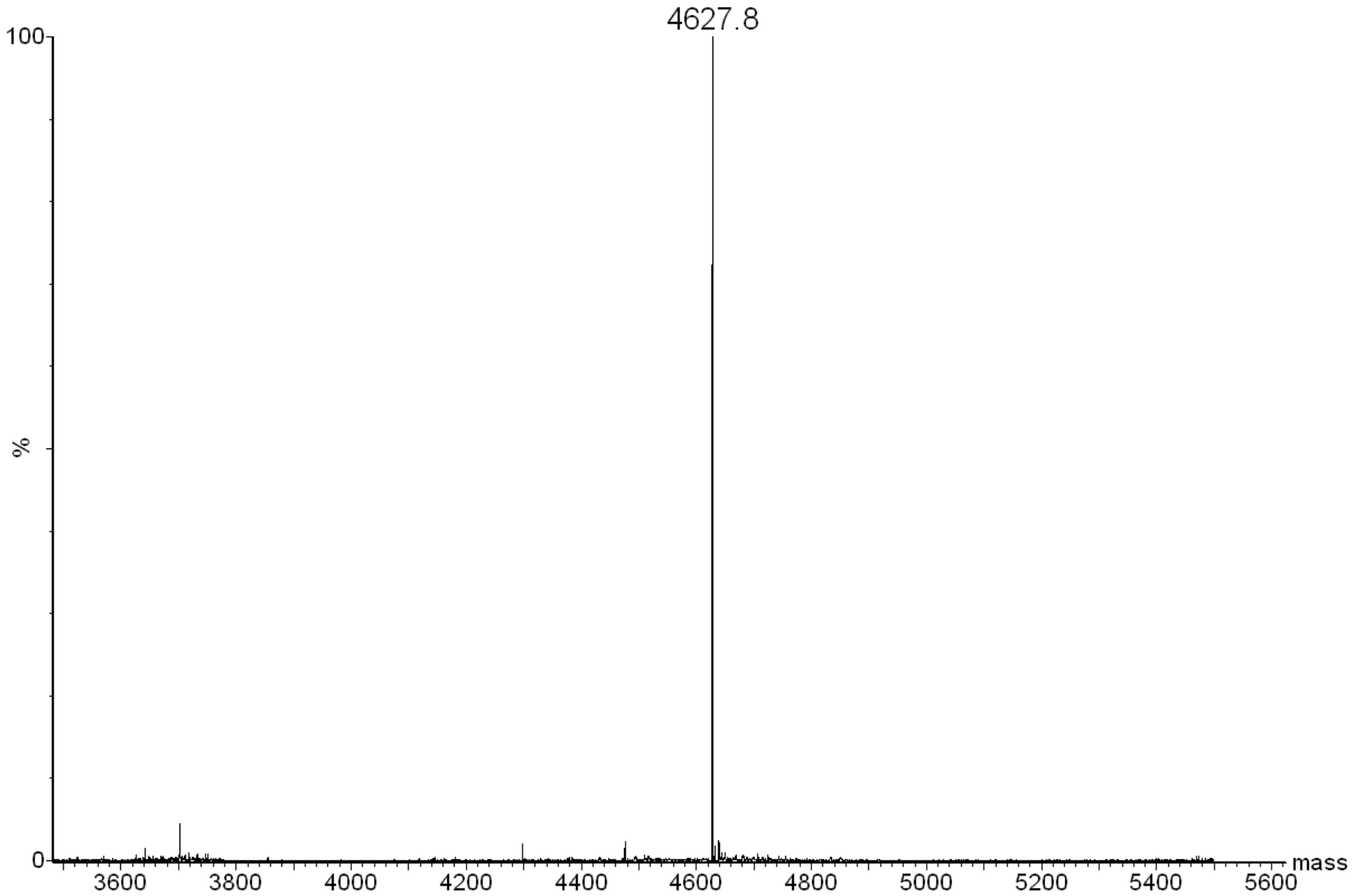


Figure S6.5 - Deconvoluted ESI MS spectrum of oligonucleotide **GT7** (expected mass of 4627.2)

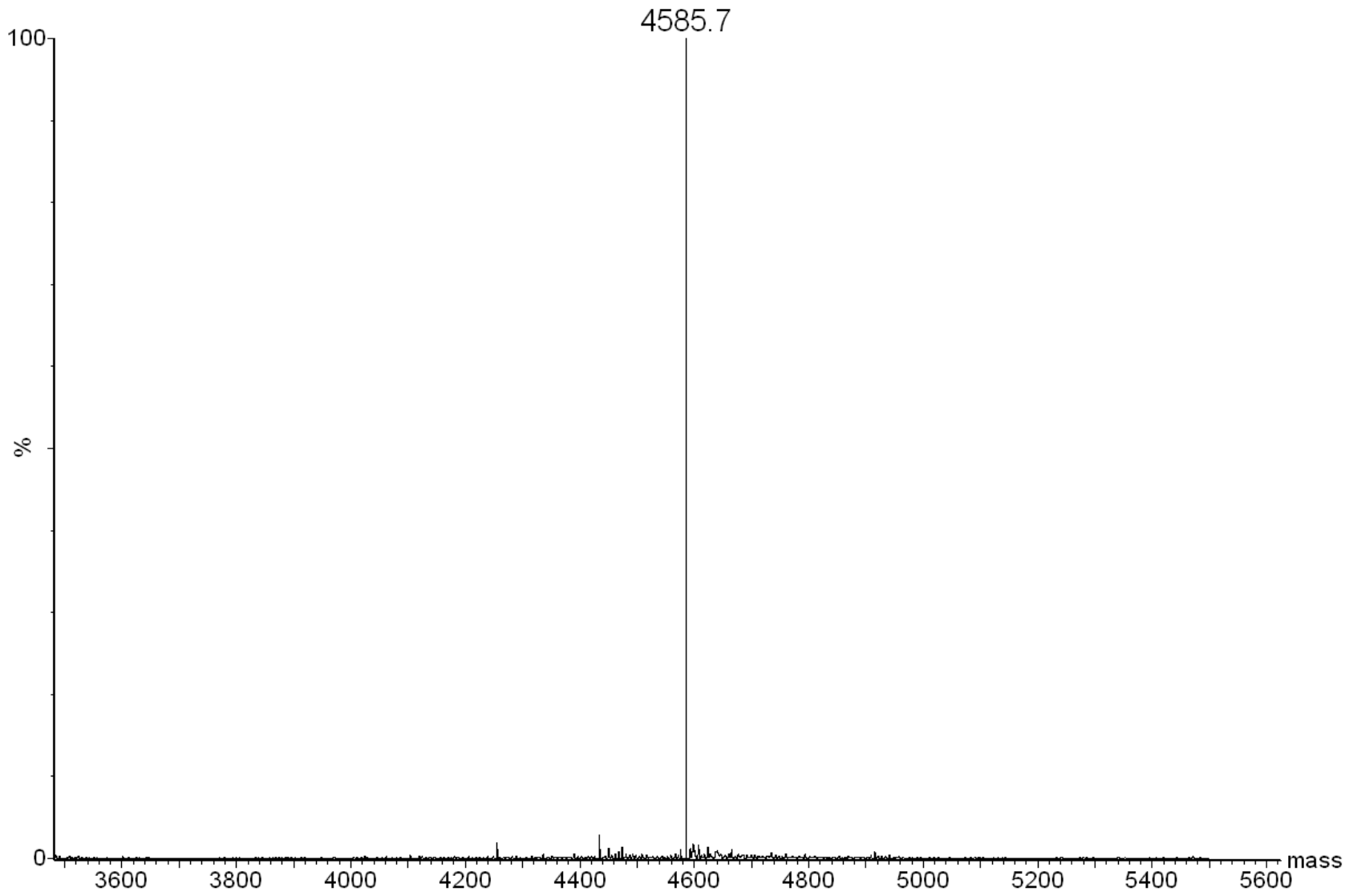


Figure S6.6 - Deconvoluted ESI MS spectrum of oligonucleotide **TG4** (expected mass of 4585.1)

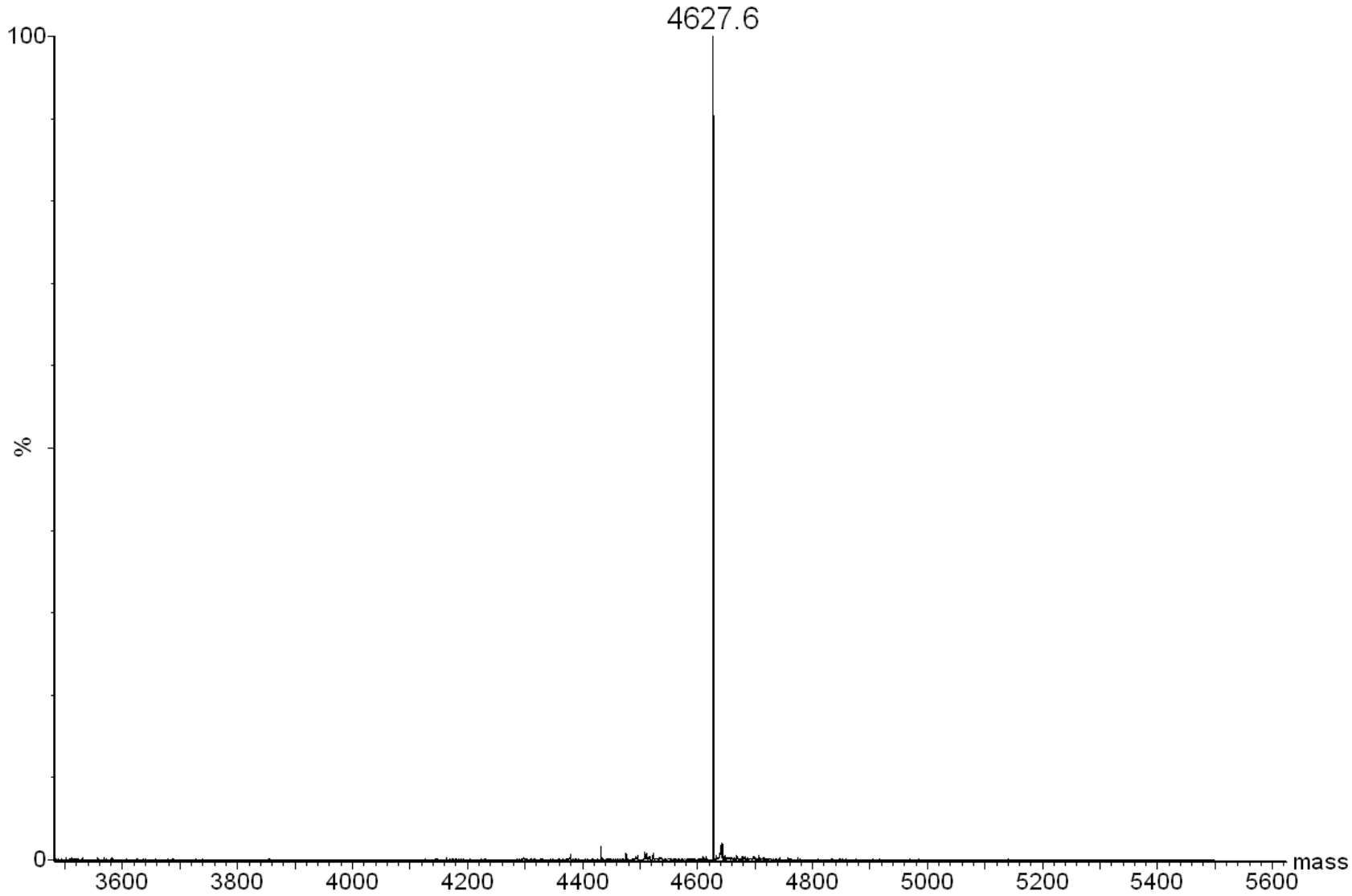


Figure S6.7 - Deconvoluted ESI MS spectrum of oligonucleotide **TG7** (expected mass of 4627.2)

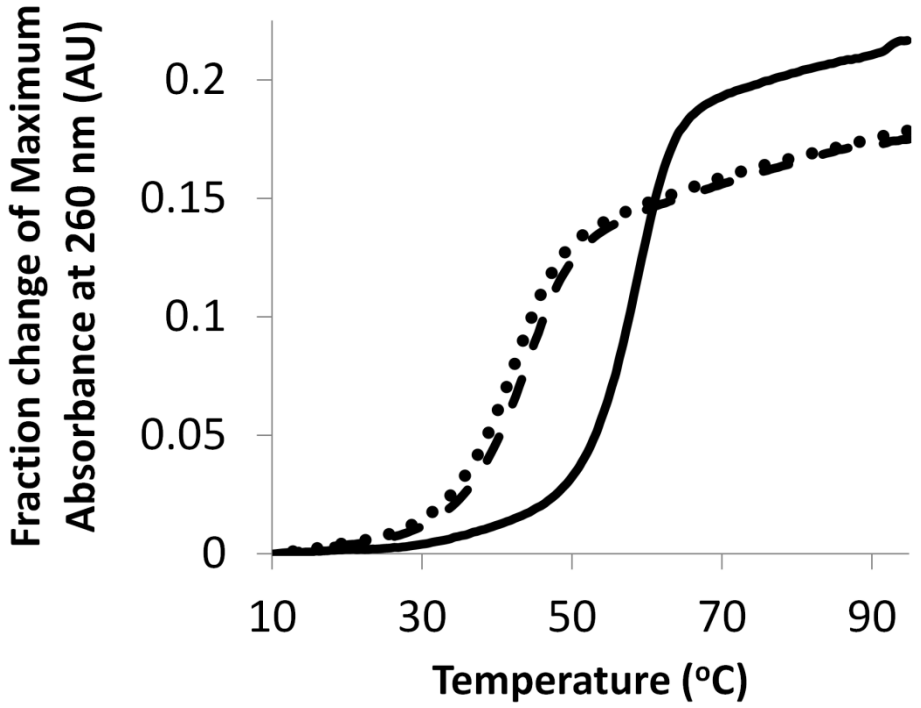


Figure S6.8 - Hyperchromicity change (A_{260}) *versus* temperature (°C) profiles of duplexes containing **TT4** (•••), **TT7** (— — —) and unmodified **TT** control DNA (—).

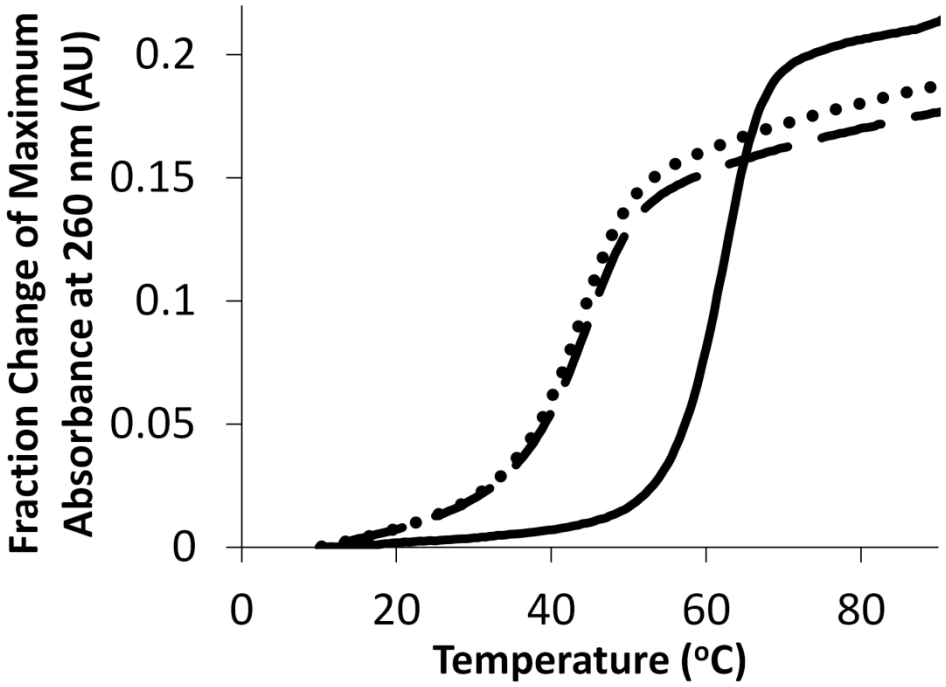


Figure S6.9 - Hyperchromicity change (A_{260}) *versus* temperature (°C) profiles of duplexes containing **GT4** (•••), **GT7** (— — —) and unmodified **GT** control DNA (—).

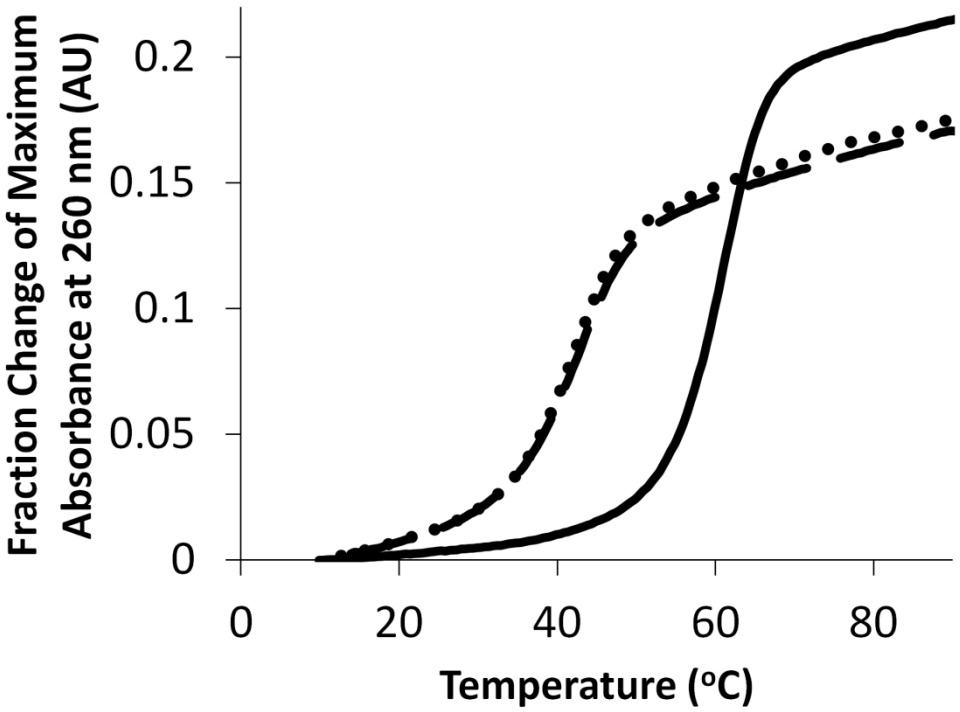


Figure S6.10 - Hyperchromicity change (A_{260}) *versus* temperature (^{o}C) profiles of duplexes containing **TG4** (•••), **TG7** (— — —) and unmodified **TG** control DNA (—).

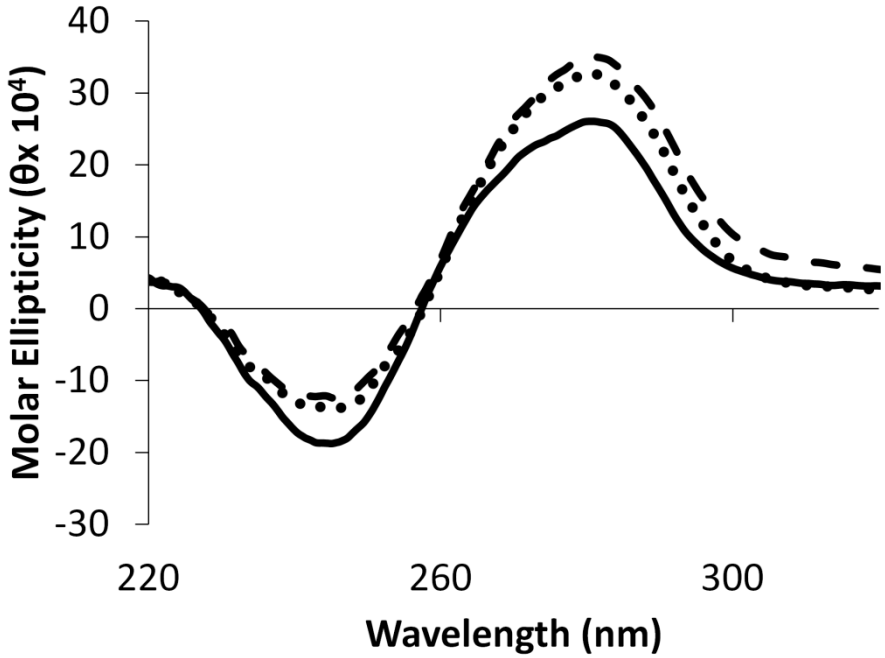


Figure S6.11 - Circular dichroism spectra of IaCL duplexes **TT4** (•••), **TT7** (— — —) and unmodified control **TT** DNA (—).

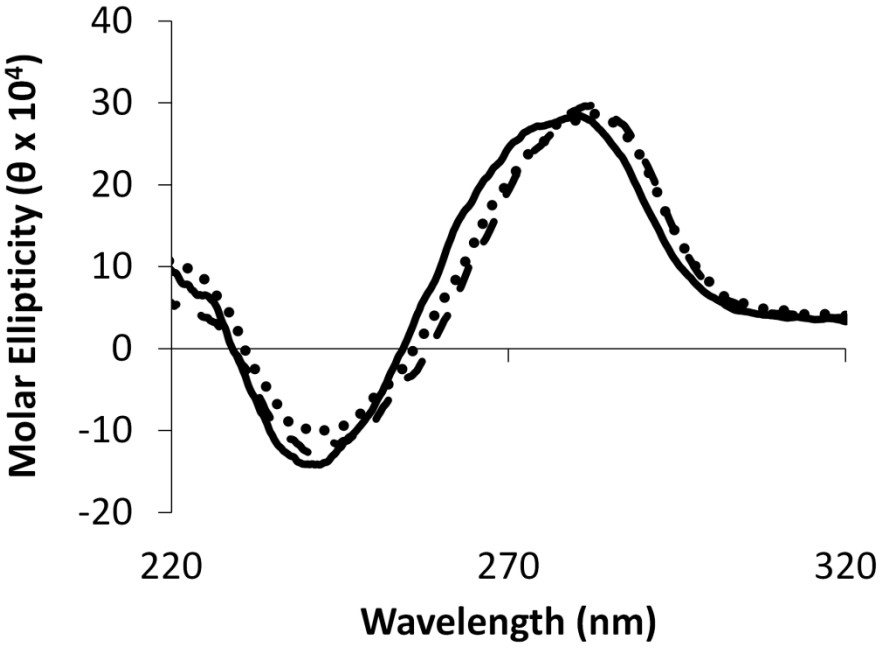


Figure S6.12 - Circular dichroism spectra of IaCL duplexes **GT4** (•••), **GT7** (— — —) and unmodified control **GT** DNA (—).

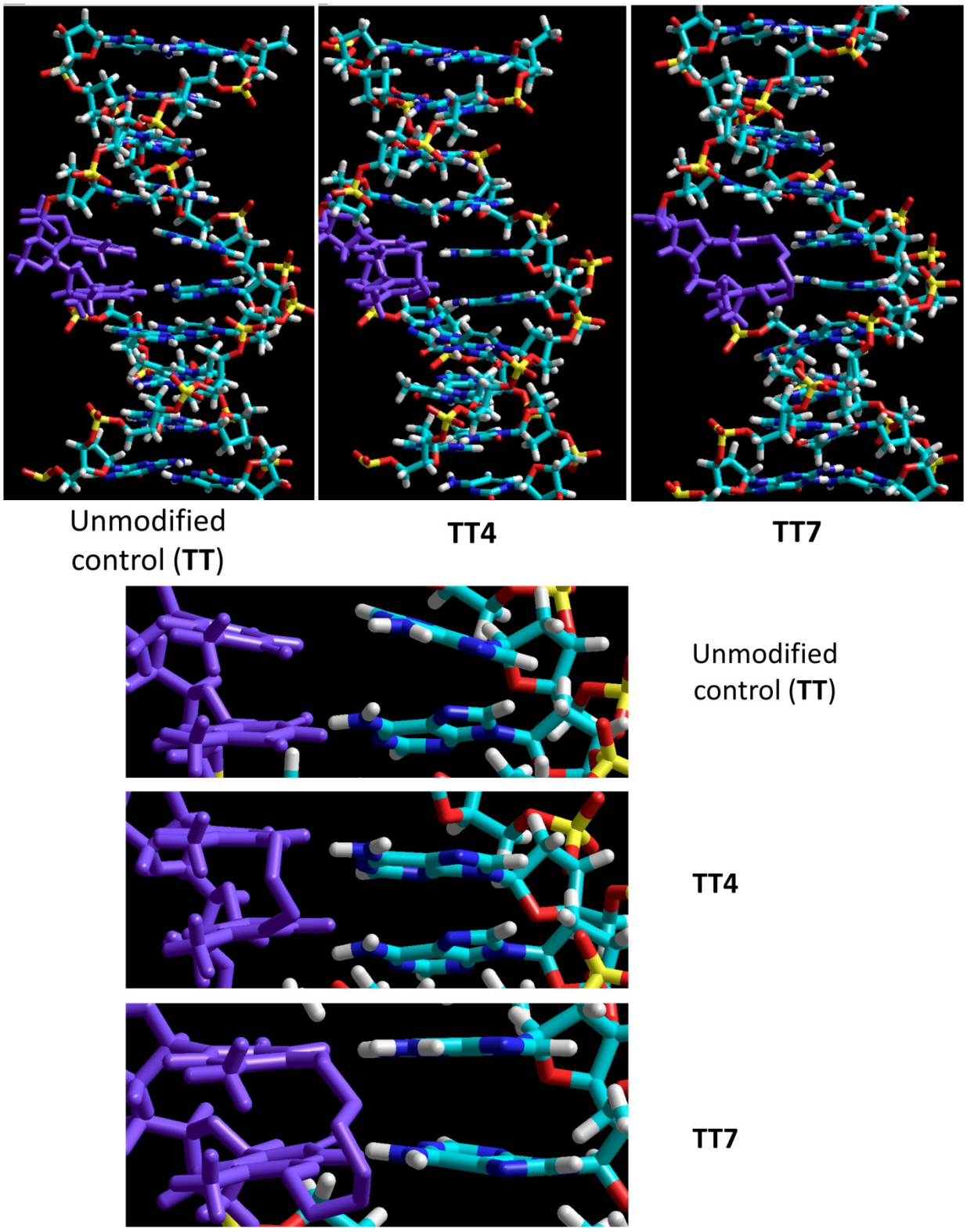


Figure S6.13 - Molecular models of unmodified control (**TT**) duplex and duplexes containing **TT4** and **TT7** that were geometry optimized using the AMBER forcefield.

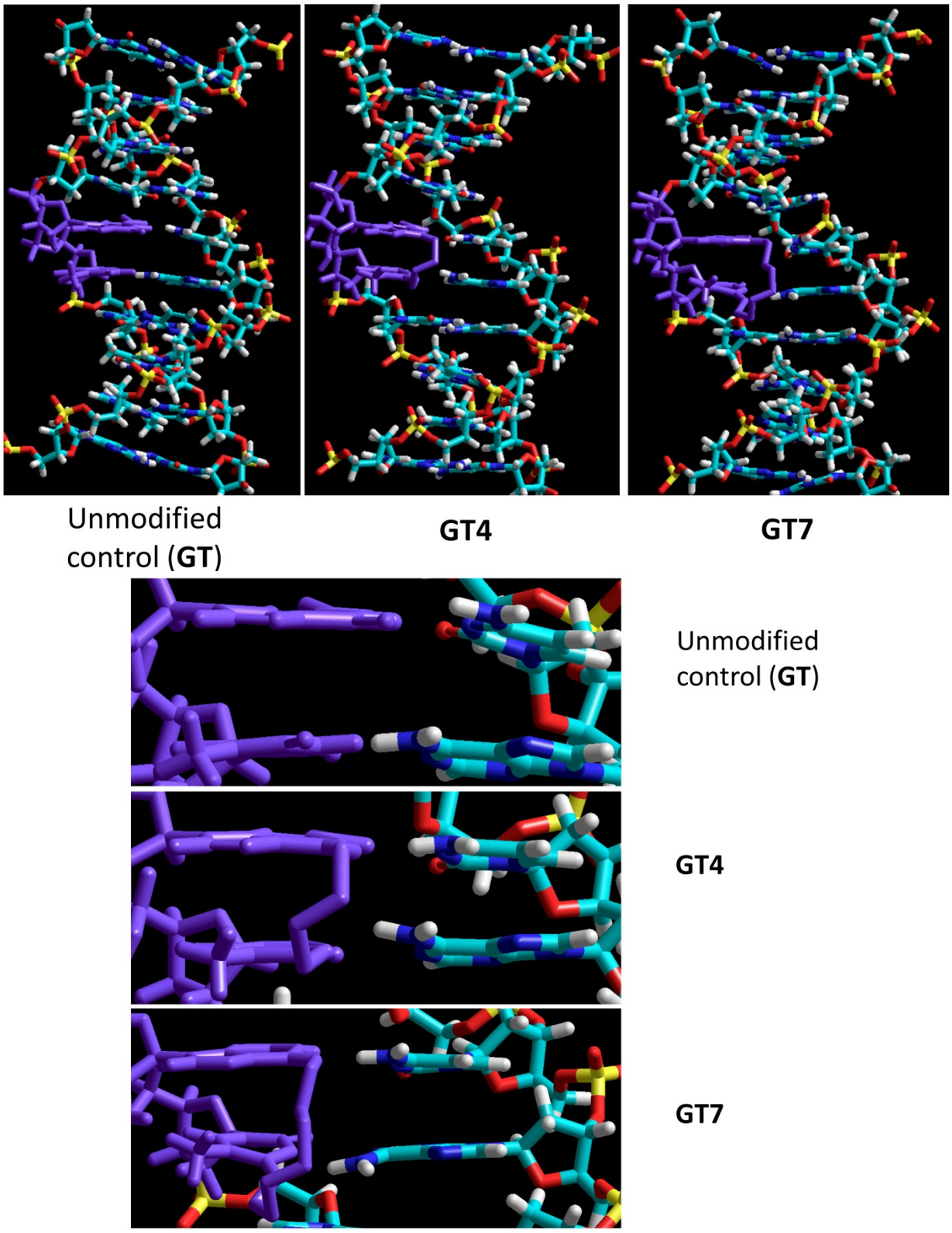


Figure S6.14 - Molecular models of unmodified control (**GT**) duplex and duplexes containing **GT4** and **GT7** that were geometry optimized using the AMBER forcefield.

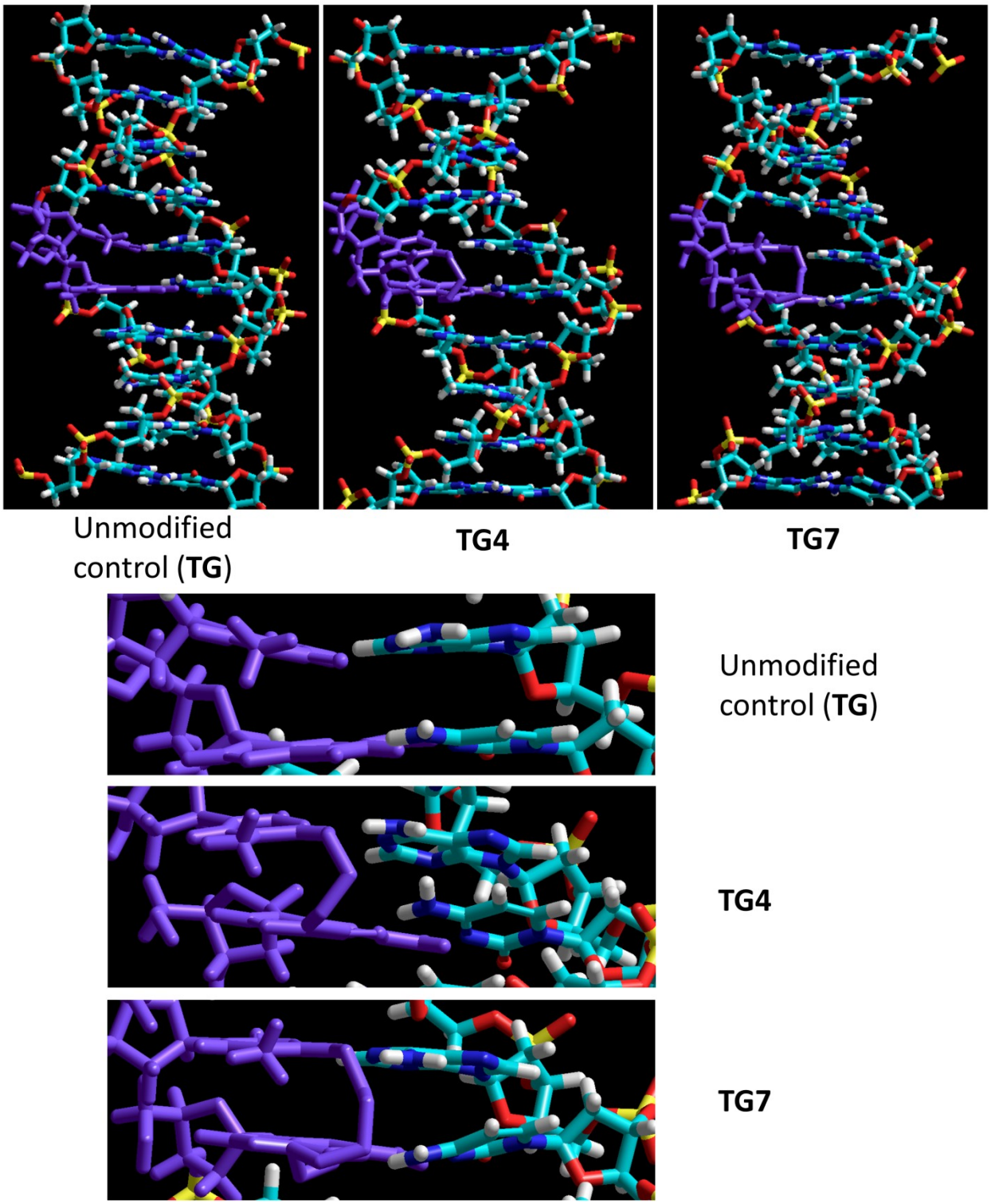


Figure S6.15 - Molecular models of unmodified control (**TG**) duplex and duplexes containing **TG4** and **TG7** that were geometry optimized using the AMBER forcefield.

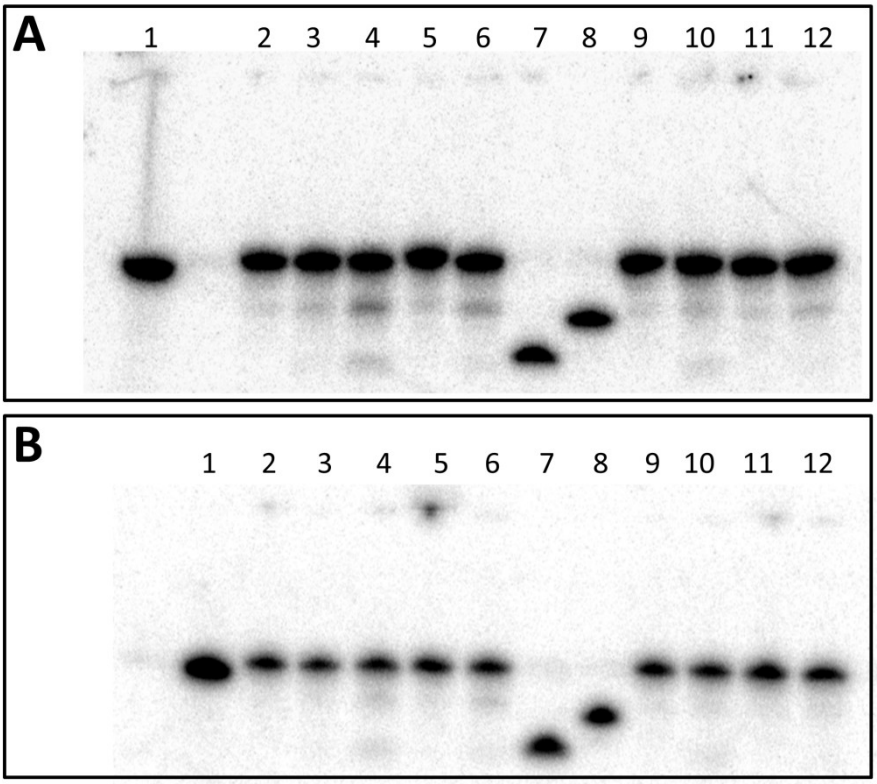


Figure S6.16 - Repair of (**A**) **TT4** and (**B**) **TT7** by hAGT, OGT, Ada-C, and OGT S134P for 16h at 37°C. Denaturing PAGE of repair reactions as described in the experimental section with varying protein concentration (either 5- or 30-fold protein). Panel **A**; Lane 1, 2 pmol control unmodified **TT** DNA; lane 2, 2 pmol **TT4**; lane 3, **TT4** + 10 pmol hAGT; lane 4 **TT4** + 60 pmol hAGT; lane 5 **TT4** + 10 pmol OGT; lane 6 **TT4** + 60 pmol OGT; lane 7 unmodified 5-mer control DNA 5'-pTGCTT; lane 8 unmodified 10-mer control 5'-pTGC ATC ACC AG; lane 9 **TT4** + 10 pmol Ada-C; lane 10 **TT4** + 60 pmol Ada-C; lane 11 **TT4** + 10 pmol OGT S134P; lane 12 **TT4** + 60 pmol OGT S134P. Panel **B**; identical to panel **A** except with **TT7** instead of **TT4**.

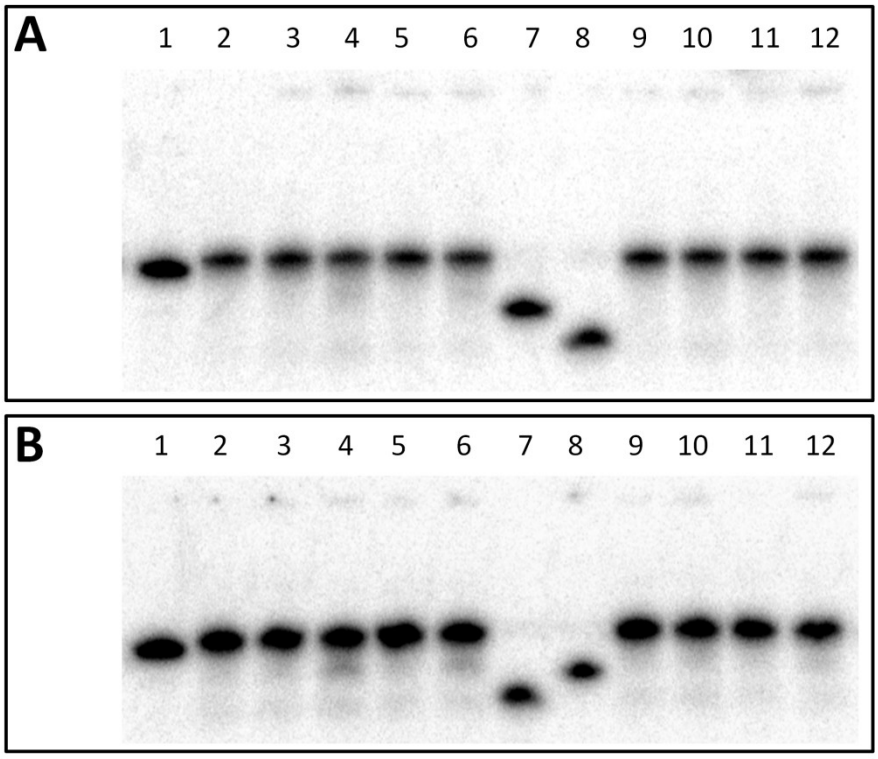


Figure S6.17 - Repair of (**A**) **GT4** and (**B**) **GT7** by hAGT, OGT, Ada-C, and OGT S134P for 16h at 37^{o}C. Denaturing PAGE of repair reactions as described in the experimental section with varying protein concentration (either 5- or 30-fold protein). Panel **A**; Lane 1, 2 pmol control unmodified **GT** DNA; lane 2, 2 pmol **GT4**; lane 3, **GT4** + 10 pmol hAGT; lane 4 **GT4** + 60 pmol hAGT; lane 5 **GT4** + 10 pmol OGT; lane 6 **GT4** + 60 pmol OGT; lane 7 unmodified 5-mer control DNA 5'-pTGCTT; lane 8 unmodified 10-mer control 5'-pTGC ATC ACC AG; lane 9 **GT4** + 10 pmol Ada-C; lane 10 **GT4** + 60 pmol Ada-C; lane 11 **GT4** + 10 pmol OGT S134P; lane 12 **GT4** + 60 pmol OGT S134P. Panel **B**; identical to panel **A** except with **GT7** instead of **GT4**.

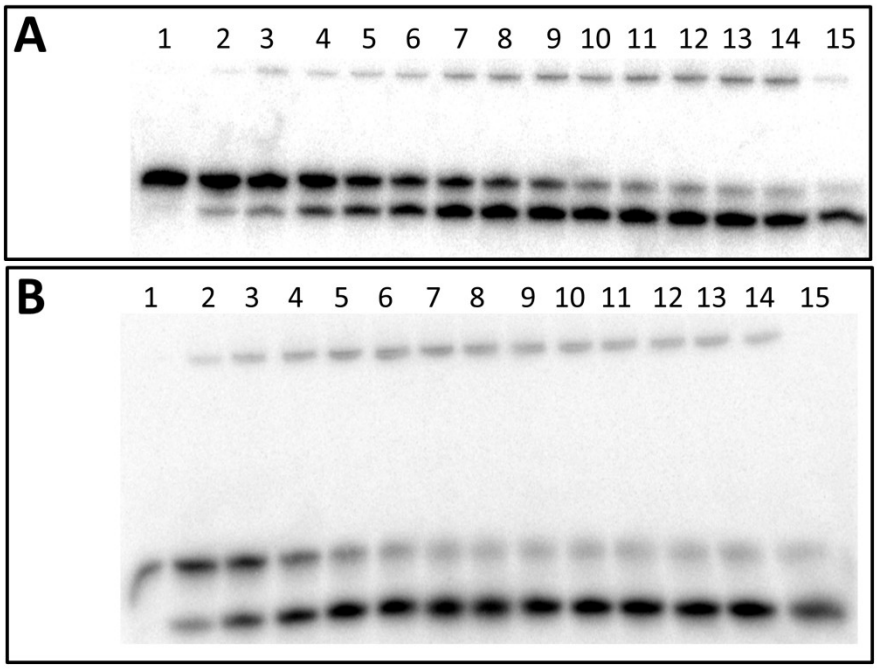


Figure S6.18 - Time course repair gel of duplexes containing (**A**) **TG4** or (**B**) **TG7** by hAGT. (**A**) Denaturing gel of the repair of 2 pmol of **TG4** by 10 pmol hAGT as a function of time: lane 1, 2 pmol Control; lanes 2-15, 2 pmol **TG4** + 10 pmol hAGT incubated for 1, 2.5, 5, 10, 15, 25, 35, 45, 60, 90, 120, 180, 240, 1020 min, respectively. (**B**) Denaturing gel of the repair of 2 pmol of **TG7** by 10 pmol hAGT as a function of time: lane 1, 2 pmol Control; lanes 2-15, 2 pmol **TG7** + 10 pmol hAGT incubated for 1, 2.5, 5, 10, 15, 25, 35, 45, 60, 90, 120, 180, 240 and 1020 min, respectively.

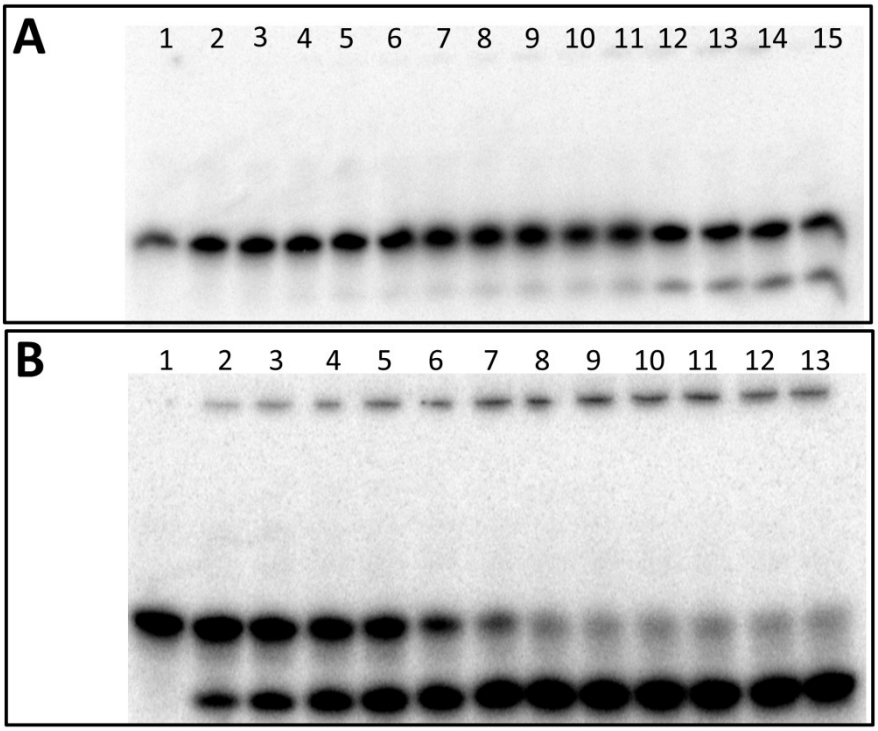


Figure S6.19 - Time course repair gel of duplexes containing **TG7** by 30-fold protein (**A**) Ada-C and (**B**) hAGT. (**A**) Denaturing gel of the repair of 2 pmol of **TG4** by 60 pmol Ada-C as a function of time: lane 1, 2 pmol Control; lanes 2-15, 2 pmol **TG4** + 60 pmol hAGT incubated for 1, 2.5, 5, 10, 15, 25, 35, 45, 60, 90, 120, 180, 240, 1020 min, respectively. (**B**) Denaturing gel of the repair of 2 pmol of **TG7** by 60 pmol hAGT as a function of time: lane 1, 2 pmol Control; lanes 2-13, 2 pmol **TG7** + 60 pmol hAGT incubated for 0.25, 0.5, 0.75, 1, 1.5, 2.5, 5, 10, 15, 25, 60, 120 min, respectively.

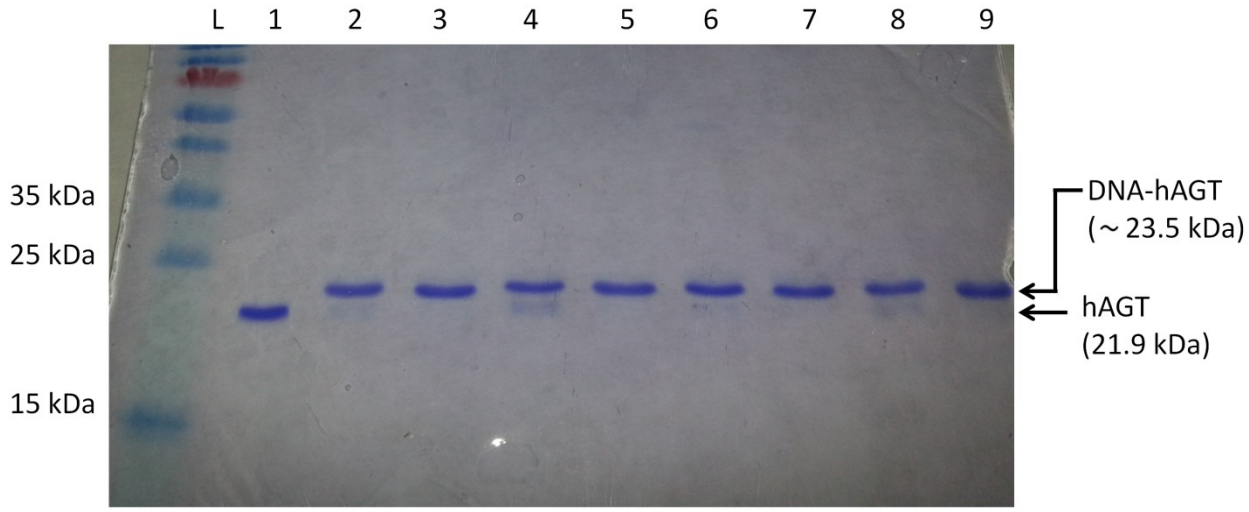


Figure S6.20 - 15% SDS-PAGE (Coomassie stained) of repair reaction between hAGT (300 pmol) and IaCL DNA (300 pmol) for either 1h (lanes 2-5) or 3.5 h (lanes 6-9) at 37°C. Lane L, molecular weight marker (ladder); lane 1, hAGT control; lane 2, hAGT + **GG4**; lane 3, hAGT + **GG7**; lane 4, hAGT + **TG4**; lane 5, hAGT + **TG7**; lane 6, hAGT + **GG4**; lane 7, hAGT + **GG7**; lane 8, hAGT + **TG4**; lane 9, hAGT + **TG7**. Results demonstrate virtually quantitative conversion to the DPC after only 1h of reaction time.

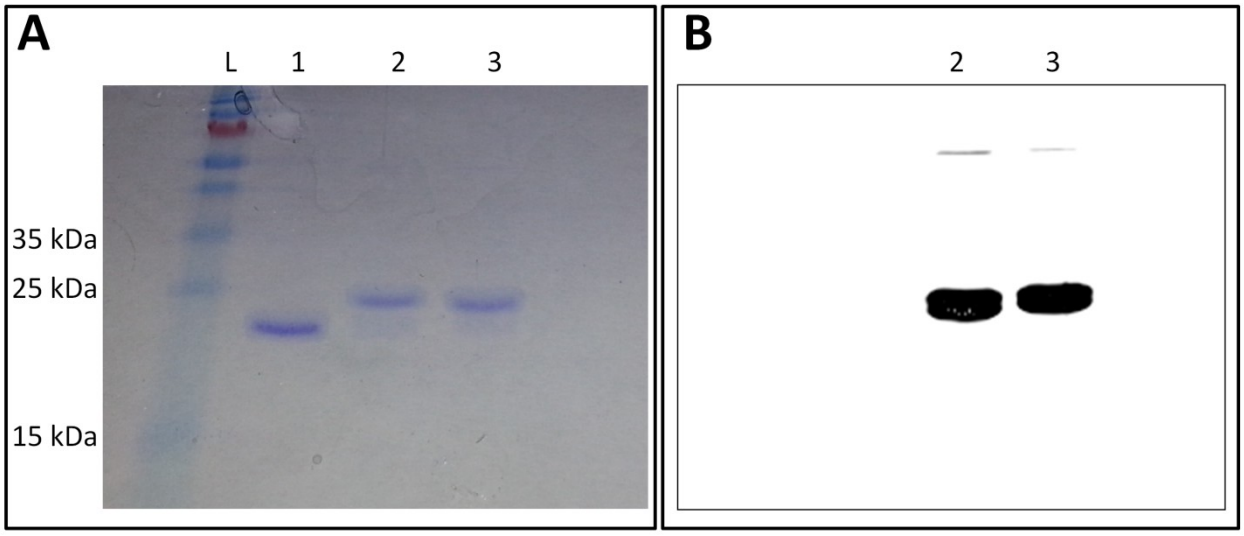


Figure S6.21 - 15% SDS-PAGE of repair reaction between hAGT (300 pmol) and fluorescently labelled **GG4** or **GG7** (300 pmol) for 1h at 37°C showing the feasibility of the hAGT protein-tagging with a functionalized DNA fragment. (**A**) Coomassie Blue stained. (**B**) Fluorescence monitoring of the fluorescein tag. Lane L, molecular weight marker (ladder); lane 1, hAGT control; lane 2, hAGT + **GG4**; lane 3, hAGT + **GG7**.

CHAPTER VII
Conclusions and Future Work

7.1 General Conclusions

The work presented in this thesis has enabled the introduction of a number of site-specific chemical modifications in DNA. This was accomplished using small molecule synthesis to generate the nucleoside(tide) phosphoramidite already containing the chemical alterations. These building blocks were incorporated into DNA oligomer sequences *via* automated solid-phase synthesis. The developed approach enabled a versatile methodology to incorporate different chemical modifications that were subsequently utilized for biochemical purposes.

First, DNA containing bicyclic pyrimidine analogues (DFP and TPP) were prepared as mimics of the biologically relevant O^4-MedT and O^4-EtdT DNA damage. The synthesis of *C*5-modified pyrimidines was accomplished by the use of different size lactone starting materials. Once inserted in DNA oligomers, the conformationally restricted analogues behaved similarly to the open-chain modifications in terms of biophysical characterization (thermal stability, circular dichroism and basic molecular modeling). Although these probes remained inert towards AGT-mediated repair (hAGT, OGT and Ada-C), they provided critical insights in the bypass mechanism of the model TLS polymerase, hPol η. The bypass profiles by hPol η of DNA containing O^4-MedT and O^4-EtdT were significantly different in comparison to their respective conformationally locked mimics. All adducts causes blockage of hPol η, with stronger blockage observed for the open-chain analogues relative to the mimics, respectively. All adducts caused error-prone hPol η activity, with the major product containing the dGMP error. The incorporation frequency of dGMP decrease according to the following series TPP > DFP ≈ O^4EtdT > O^4MedT. Full extension assays revealed a noteworthy stall at the n+1 primer product for templates containing O^4-MedT and O^4-EtdT, but not for DFP and TPP, with a larger accumulation observed for in the case of O^4-MedT. Also, a significant increase in frameshift

adduct formation (-1) was observed across O^4-MedT, relative to other adducts and the unmodified control. Crystal structures of ternary complexes revealed that the O^4-MedT nucleobase adopted a unique (and undocumented) conformation, not engaging with the incoming nucleotide, but rather, engaged in a hydrophobic pocket located at the top of the hPol η active site. Interestingly, a similar conformation was not shared for the larger O^4-EtdT nucleobase in its ternary complex crystal structure. The investigation provided mechanistic insights about the bypass mechanism of hPol η. That is, perhaps hPol η makes use of the conformational liberty of the O^4-MedT to aid in its nucleotide discrimination (*e.g.* slowing the extension step by allowing different conformational populations inside the active site). This may explain why the fidelity decreases with the conformationally locked analogue.

Previous investigation from our group has demonstrated the potential involvement of direct repair by AGT in the mending of ICL DNA. This was shown using O^6-dG-alkylene-O^6-dG ICL DNA with butylene or heptylene linkers, for which hAGT demonstrated repair.[64,65] In light of these results, we questioned whether AGTs were capable of acting on IaCL substrates. As an initial study, IaCL mimics lacking the phosphodiester linkage were prepared since these were less synthetically challenging and could provide insights on the processing of such substrates by AGTs. The key intermediate step in the synthesis of the target dG-dG dimer was the Mitsunobu reaction, known for its regioselectivity towards the O^6-atom of (5', 3'-protected) dG, which has previously been used for the introduction of ICL in DNA.[311,312,368] Our initial findings unveiled the potential participation of hAGT in the restoration of IaCL DNA damage. The *in vitro* assays demonstrated efficient repair of the model IaCL (**GG4** and **GG7**). These were processed in superior levels compared to their respective ICL DNA versions (even at lower protein : DNA

ratios). Furthermore, the heptylene adducted DNA (**GG7**) was consumed at a faster rate compared to the butylene analogue (**GG4**).

The latter study prompted us to develop the next generation of IaCL DNA, which contained the phosphodiester linkage (**GpG4** and **GpG7**). Albeit more synthetically challenging than the previous mimics, these DNA probes were prepared by first synthesizing the dG-dG dimer phosphoramidite containing the alkylene cross-link and intradimer phosphodiester backbone. Ring closing of the medium size ring was achieved using methodology developed by Begley's group, by means of the MSNT condensing reagent.[325] Interestingly, UV thermal denaturation experiments of DNA duplexes containing the IaCL were in agreement regardless of the presence, or lack, of the internucleotide linkage. Surprisingly, duplexes containing **GpG4** and **GpG7** were only modestly repaired by hAGT, with greater efficiency towards **GpG7**. The decrease in repair efficiency, relative to **GG4** and **GG7**, was potentially attributed to: (i) the lesion restriction prevented optimal α-carbon orientation necessary for Cys145 to attack, (ii) linkage to the neighboring nucleotide prevents efficient flipping of the nucleotide undergoing repair, (iii) other vital interactions such as Tyr114 cannot be accomplished correctly. Nonetheless, repair of the IaCL DNA were in agreement with previous *in vivo* experiments, which showed that hAGT contributes to cell survival upon exposure to hepsulfam, but not busulfan.[65] Whether this resistance effect was due to removal of IaCL DNA, or adducts other than mono-adducts, remains to be determined.

Results from **CHAPTER II** to **IV** highlighted a few aspects that would warrant investigating bypass experiments of **GG4**, **GG7**, **GpG4**, and **GpG7** using hPol η. Conformationally restricted analogues tested in **CHAPTER II** displayed enhanced rates of bypass compared to the freely rotating lesions, although erroneous. Perhaps reduced flexibility

allows for more efficient bypass by hPol η. Adducts such as CPD and Pt-GG lesions are proficiently bypassed by hPol η while being relatively rigid in nature. It was also noted that the increased conformational rigidity seemed to have a negative impact on hAGT-mediated repair. Our hPol η bypass investigation supported the rigidity hypothesis. As expected, all alkylene IaCL blocked hPol η bypass relative to the unmodified control. This was observed across both residues part of the IaCL. Steady-state experiments revealed that although hPol η increases promiscuity across the first O^6-alkylene-dG residue (3'-dG), the preferred nucleotide was the correct dCMP. The same was not said for the second O^6-alkylene-dG (5'-dG), where an error-prone behavior prevailed in all cases. This was, however, more pronounced for **GG4** and **GG7**, relative to **GpG4** and **GpG7**, respectively. Full extension assays revealed an accumulation of unextended primers in the case of **GG4** and **GG7** (n+2 products), but not in the case of **GpG4** and **GpG7**. IaCL containing the intradimer phosphodiester were extended to the full length, with the majority containing nucleotides dCMP and dTMP across the 3'-dG and 5'-dG, respectively. This brought about the conception that lesion orientation may play a critical role in bypass mechanism since the linker of the 3'-O^6-alkylene-dG is permanently pointing towards the 5'-end of the DNA template, whereas the 5'-O^6-alkylene-dG is oriented in the opposite 3'-direction.

Longer extension times of **GG4** and **GG7** caused uncontrollable extension by hPol η with no defined "fully" extended product. This inaccurate behavior has been observed for extension across and past abasic sites, regardless of 5'-nucleotide next to the damaged site.[202,236] Moreover, frameshift adduct formation accounted for about 65-80 % of the total products, and the remaining was largely attributed to dAMP or dGMP. The study on abasic sites supported the so-called purine-rule, where purines are most-frequently inserted across the abasic site. Structural investigation suggested that a network of water molecules mediate the necessary interactions for

the purine rule (previously described as the A-rule).[236] In the case of **GG4** and **GG7**, the major product across the 5'-dG residue is a frameshift adduct which may be mediated by water as well. However, structural investigations are required to support this hypothesis.

Each nucleotide unit can be divided into a nucleobase, a sugar and a phosphodiester group. As expected, the abasic site can mislead DNA Pols given that the identity of the nucleotide is defined by its nucleobase moiety (and abasic sites lack the nucleobase). It was interesting to observe similar profiles for **GG4** / **GG7** where the intradimer phosphodiester linkage was *deleted*, and revealed the direct importance of the phosphodiester group in the bypass of IaCL DNA (those linking the O^6-atoms of dG, at least). The effect may be an artifact of the alkylene tether acting as the sugar-phosphate backbone, given that the alkylene moiety does not have electrostatic character and is inherently more flexible, whereas the latter has anionic character. As a result, it would be of interest to test whether altering the tether size (*e.g.* ethylene) or its character (*e.g.* diethylamine) would have an effect on hPol η bypass.

DNA repair and TLS mechanisms must be orchestrated in the cell.[369,370] The results acquired throughout this research potentially suggest a synergy between hAGT-repair and TLS, particularly for IaCL DNA. That is, hAGT repair is efficient for the flexible IaCL analogues, whereas TLS is relatively efficient for the more-rigid IaCL DNA. To further probe AGT direct repair pathway, we developed a series of three novel sets of IaCL DNA lacking the intradimer phosphodiester. The IaCLs linking two O^4-atoms of dT nucleotides were prepared using the convertible nucleoside methodology developed by Swann's group, and procedures developed by our group.[150,242,371] The mixed IaCL dimers between the O^4-position of dT and the O^6-position of dG were prepared from the dT mono-adducts and dG monomers *via* the Mitsunobu reaction. Using different protecting groups on the 5'-*O* and 3'-*O* sugar functionalities of either dT or dG

allowed us to generate IaCL DNA with different orientations (5'-TG and 5'-GT). Ultra-mild deprotection strategies were employed for the DNA containing the novel IaCL. Biophysical studies of duplexes containing the various IaCLs displayed similar behavior relative to **GG4** and **GG7**. Repair assay profiles of the different IaCLs (**TT4**, **TT7**, **GT4**, **GT7**, **TG4**, and **TG7**) showcased only efficient repair of **TG4** and **TG7** by hAGT, and low-moderate repair of **TG7** by Ada-C, with repair occurring only at the dG moiety. Interestingly, repair of **TG4** and **TG7** by hAGT occurred at a faster rate compared to **GG4** and **GG7**. An NMR structure of a O^4-dU-heptylene-O^4-dU ICL revealed higher mobility of the cross-linked site compared to identical duplexes containing an O^6-dG-heptylene-O^6-dG ICL adduct (unpublished data). The **TG** IaCL hybrid may have higher mobility (flexibility) compared to the **GG** series, which could account for the enhanced repair rates observed. The lack of repair of **GT** IaCL DNA by hAGT suggests that the 3'-phosphodiester may be necessary for efficient repair of O^6-alkyl-dG adducts, which was in agreement with minimal repair of the 5'-dG residue of **GG4** and **GG7** by hAGT at elevated protein : DNA ratio. In addition, the free nucleoside O^6-MedG is repaired very slowly by hAGT, which may be partly due to the lack of the 3'-phosphodiester linkage. Alternatively, it may simply be that the linker orientation is not favourable for repair. The preference of the 3'-dG residue of **GG4**, **GG7**, **TG4**, and **TG7** was important in supporting the necessity of the 3'-phosphodiester group in order for efficient repair by hAGT. Our investigation revealed an insight pertinent to the mechanism of repair, whether the interaction occurs with Tyr114 (or not), *via* sterics or electronics.[137,168] Deleting this group prevents proficient repair of the 5'-O^6-alkylene-dG residue part of two IaCL DNA types (**GG4**, **GG7**, **GT4**, and **GT7**). More research must be conducted to elucidate reasons as to why the 5'-dG remains inert against hAGT activity.

The substrate discrimination based on the phosphodiester linkage allows for the production of site-specific DPC with excellent conversions in relatively short times (approximately 1h). The procedure can be applied to single stranded IaCL DNA to generate hAGT-DNA covalent complexes, making this approach "greener" compared to ICL DNA in terms of atom economy. The DPC is tethered between the Cys145 of hAGT and the 3'-nucleoside of a DNA fragment. The IaCL approach is versatile in producing DPC linked to either the O^6-atom of dG or the O^4-atom of dT, by either a butylene or heptylene linker. A feature to be noted of this DPC is the linear connectivity between the DNA and the protein (**Figure 7.1**). **CHAPTER VI** showcases an example of functionalizing hAGT with fluorescently-labelled DNA (5'-fluorescein). The DPC product was, as expected, fluorescent. This new technology can be envisioned as having a DNA spacer between the hAGT molecule and any label compatible with DNA and protein (bio)chemistry. The programmability of DNA allows for added versatility as well. For instance, the distance between the label and the hAGT molecule can easily be tailored by synthesizing longer initial IaCL DNA substrates. Many DNA sequences are known to fold into specific secondary structures, which may be of interest to position the tag in a particular orientation with respect to the protein. Lastly, the fact that hAGT cleaves the single stranded DNA also brings about potential applications in nanotechnology switch mechanisms.

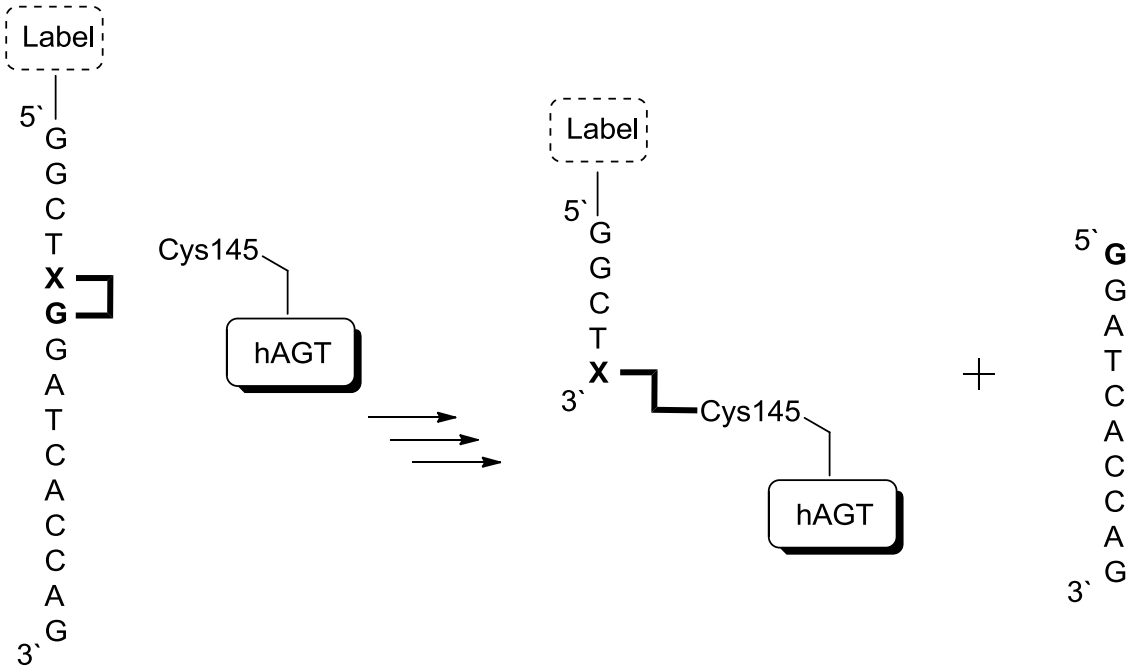


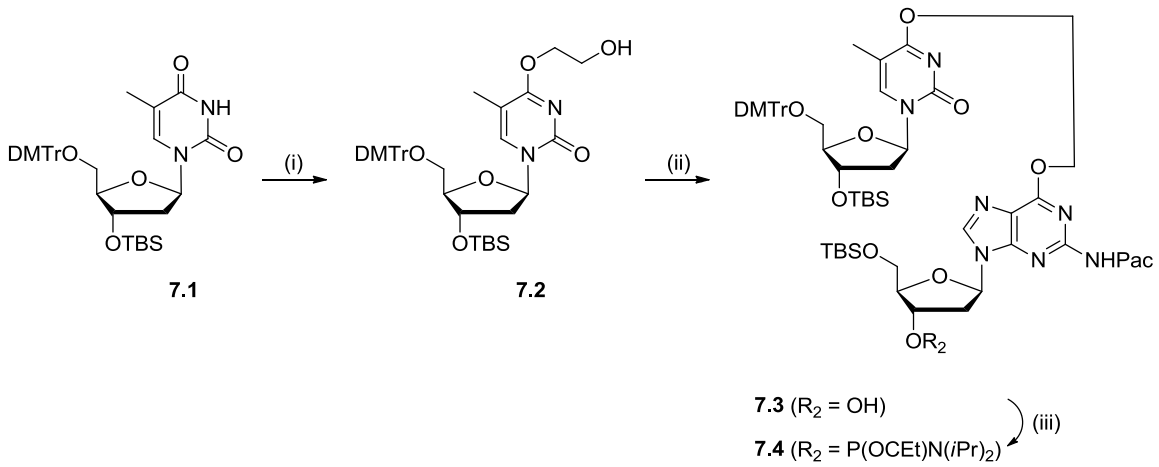
Figure 7.1: General hAGT-mediated repair of IaCL lacking the intradimer phosphodiester linkage. The presence of a 5'-tag is optional but may find use in hAGT tagging experiments. Experiments have been conducted with DNA containing O^4-alkylene-dT or O^6-alkylene dG as **X**.

7.2 Future Work

7.2.1 The Ethylene IaCL DNA Probe for AGT-Repair

The work presented herein opens a new area of bifunctional intrastrand adduct repair by AGTs. Substrate discrimination of AGT homologues is an ongoing investigation. For instance, why hAGT is capable of proficient repair of many O^6-dG adducts, but limited solely to O^4-MedT adducts. McManus and co-workers demonstrated a method to generate DPC with hAGT covalently bound to a DNA strand *via* a heptylene linker to the thymidinyl residue at the O^4-atom. The tether may be too long to verify the interaction of the hAGT active site with the thymidine nucleotide, especially because highly mobile ("disordered") regions may not be visible in crystallographic structures. Although the new procedure described herein allows for a DPC with a shorter, butylene-linked thymidinyl adduct, an even shorter length such as an ethylene

bridge may be more suitable. The proposed synthesis the cross-linked dimer is shown in **Scheme 7.1** to prepare **TG2**. This probe would also be relevant to various chloroethylating agents used in the clinic, which are known to generate hydroxyethylene mono-adducts, IaCL, and ICL DNA. In addition, exposure to 1,2-dibromoethane generates ethylene DPCs with all four nucleotides.[372,373] An assortment of ethylene-linked hAGT-DNA complexes could be devised by modifying the synthesis shown in **Scheme 7.1** (*e.g.* changing the mono-adduct from a thymidinyl derivative to a 2'-deoxycytidinyl or 2'-deoxyadenosyl). It should be noted, however, that the synthesis of O^6-dG-ethylene- dG has been previously attempted by our group (unpublished data) with no success to date.

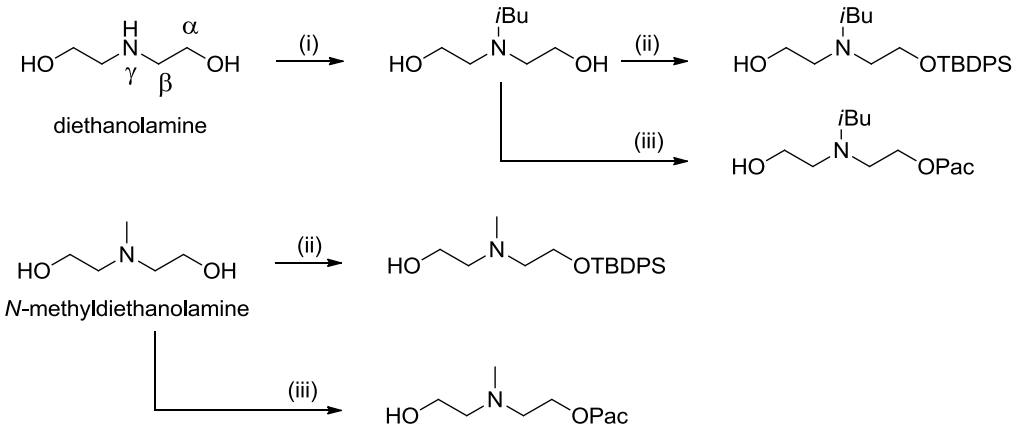


Scheme 7.1: Reagents and conditions: (i) ethylene glycol, DBU, MeCN, 21°C, 4 h. (ii) 1) 5'-*O*-(*tert*-butyldimethylsilyl)-3'-*O*-levulinyl-*N*2-phenoxyacetyl-2'-deoxyguanosine, Ph_3P, DIAD, dioxane, 21°C, 16 h. 2) $N_2H_4 \bullet H_2O$, pyridinium acetate (1:1, v/v), 21°C, 30 min. (iii) Cl-P(OCEt)N(*i*Pr)$_2$, NEt_3, THF, 21°C, 30 min.

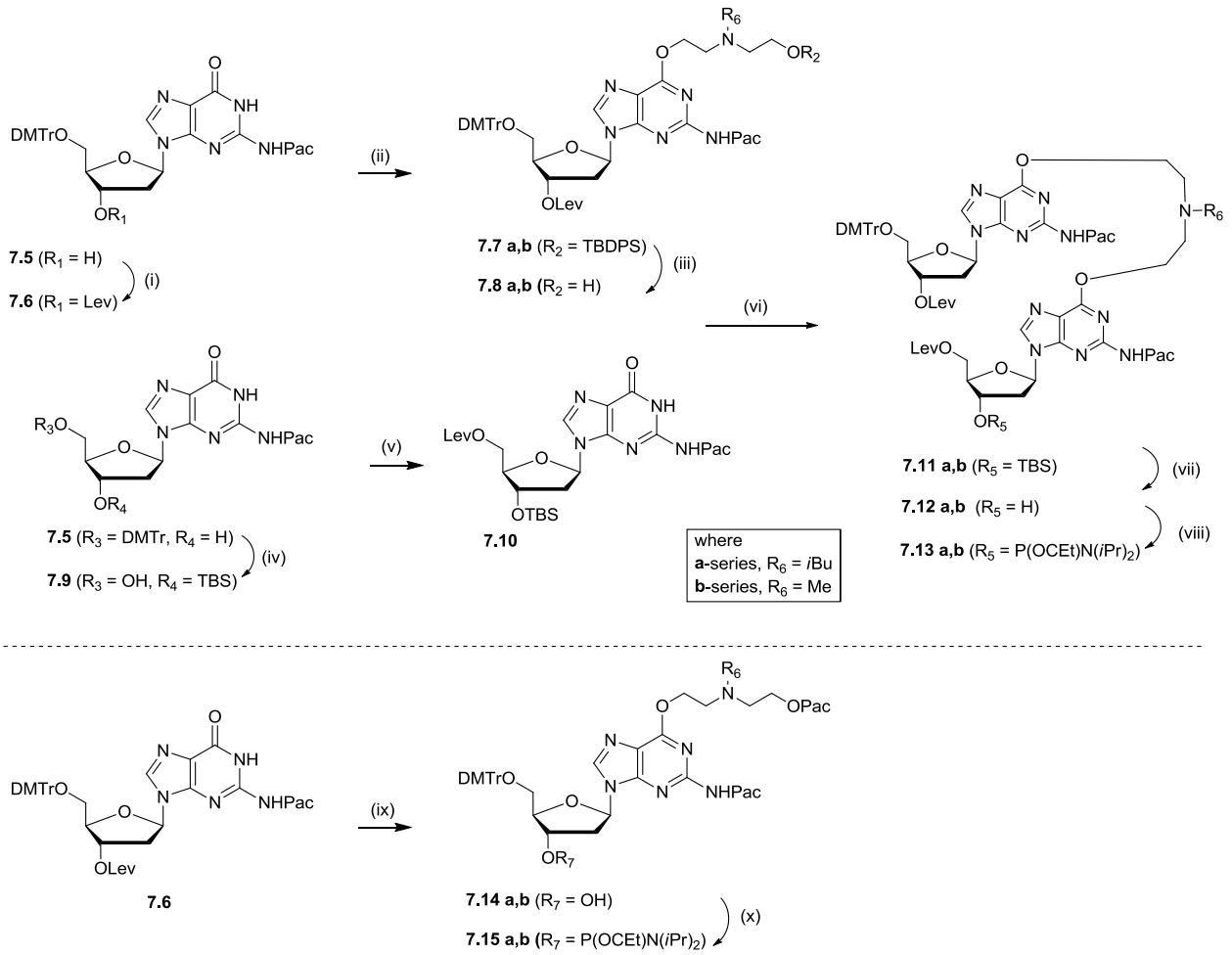
7.2.2 Synthesis of IaCL Adducts with Diverse Tethers for AGT-Repair

Our group has mainly focused on generating IaCL and ICL DNA containing alkylene tethers. However, certain chemotherapeutic agents such as mechlorethamine generate IaCL and ICL with tethers containing a central amino functionality (at the γ-position of the linker **Scheme**

7.2). This central amino functionality would aid in probing the sterics of the linker (by using the tertiary amine *N*-methyldiethanolamine) as well as probing for the electronics as this functionality would most-likely be charged at physiological pH. Investigating the O^6-alkylene-dG mono-adduct and IaCL lacking the intradimer phosphodiester group (**Scheme 7.3**) may serve as a valuable starting point, since these generally undergo better repair by AGTs compared to ICL DNA and the O^4-alkylene-dT adducts. Depending on the results, expansion to the other types of cross-links would definitely be warranted.



Scheme 7.2: Diethanolamine and *N*-methyldiethanolamine are commercially available (Sigma Aldrich, for instance). Note the central nitrogen in the linker (at the γ-position).
Reagents and conditions: (i) 1) HMDS, TMS-Cl, Py, reflux, 1h. 2) isobutyryl chloride (Cl-*i*Bu), NEt_3, THF, 21°C, 4 h. 3) saturated aqueous solution of NH_3, 0°C, 30 min. (ii) TBDPS-Cl, Im, CH_2Cl_2, 21°C, 4 h. (iii) Pac-Cl, NEt_3, THF, 0°C, 1 h.



Scheme 7.3: Reagents and conditions for **a**-series: (i) Lev-OH, EDAC•HCl, dioxane, 21°C, 4 h. (ii) *N*-(isobutyryl)-*N*-(2-((tert-butyldiphenylsilyl)oxy)ethyl)-ethanolamine, Ph_3P, DIAD, dioxane, 21°C, 16h. (iii) TBAF, THF, 21°C, 30 min. (iv) 1. TBS-Cl (2.2 eq), Im (4.4 eq), DMF, 21°C, 30 min, 2. *p*-TsOH•H_2O (1.2 eq), MeOH/DCM (1:4), 21°C, 30 min. (v) Lev-OH, EDAC•HCl, dioxane, 21°C, 4 h. (vi) **6**, Ph_3P, DIAD, dioxane, 21°C, 16h. (vii) TBAF (1M in THF), THF, 21°C, 30 min. (viii) Cl-P(OCEt)N(*i*Pr)$_2$, DIPEA, THF, 21°C, 30 min. (ix) *N*-(isobutyryl)-*N*-(2-((phenoxyacetyl)oxy)ethyl)-ethanolamine, Ph_3P, DIAD, dioxane, 21°C, 16h. (x) Cl-P(OCEt)N(iPr)$_2$, DIPEA, THF, 21°C, 30 min.

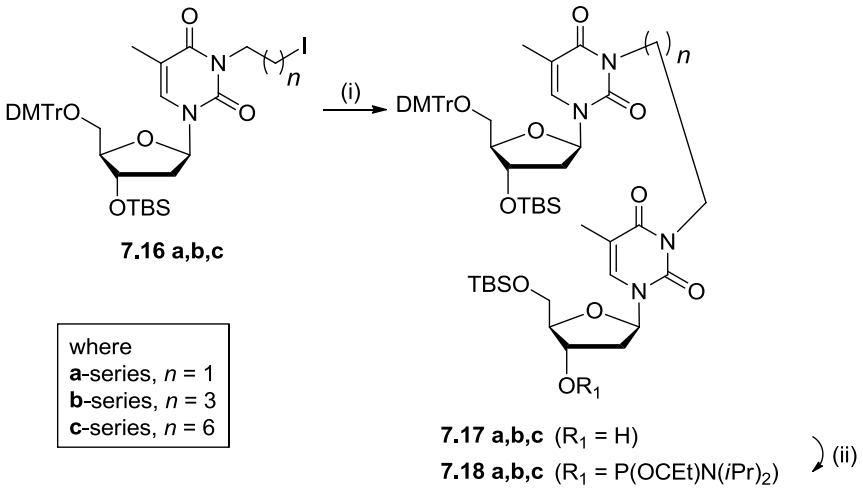
Reagents and conditions for **b**-series: Identical to **a**-series except steps (ii) *N*-(methyl)-*N*-(2-((tert-butyldiphenylsilyl)oxy)ethyl)-ethanolamine, Ph_3P, DIAD, dioxane, 21°C, 16h. (ix) *N*-(methyl)-*N*-(2-((phenoxyacetyl)oxy)ethyl)-ethanolamine, Ph_3P, DIAD, dioxane, 21°C, 16h.

7.2.3 Probing the TLS Pathway Further

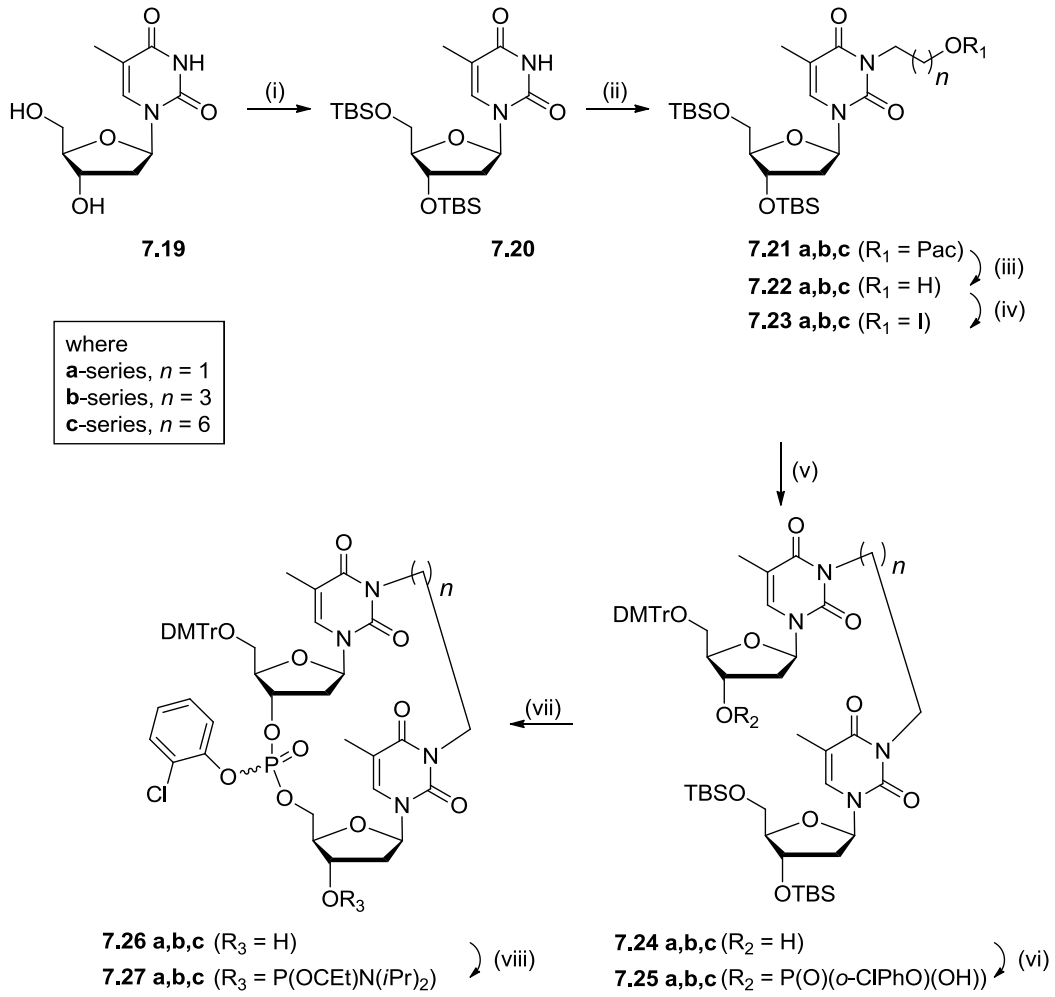
Our study on TLS revealed a potential dependence on lesion rigidity for efficient processivity from hPol η. This hypothesis is drawn from a limited amount of data. To make such claims gain more credibility, other Y-family DNA polymerases would have to be investigated such as hPol κ, hPol ι, Rev1, yPol η, yRev1, Dpo4 (from *Sulfolobus Solfataricus*). Moreover, DNA Pol bypass seems to be dependent on the DNA sequence identity, and thus, it would be of interest to incorporate the adducts described in **CHAPTER II** and **CHAPTER V** into other DNA sequences to identify if hPol η (and other Pols) behave similarly or differently. For **CHAPTER II**, crystal structures are required for the bicyclicpyrimidine (**DFP** and **TPP**) inserts to acquire additional insights on the bypass by hPol η. Similarly, structural investigations for the IaCL DNA described in **CHAPTER V** may illuminate why the lack of a phosphodiester linkage promotes error-prone behavior by hPol η.

The d(TT) CPD adducts is a naturally occurring lesion and is one of the substrates bypassed efficiently by hPol η. On the other hand the pyrimidine (6-4) pyrimidone adducts is poorly bypassed by hPol η. To further probe this substrate discrimination, a flexible *N*3-dT-alkylene-*N*3-dT IaCL crosslink is proposed (**Scheme 7.4 - 7.5**). The methodology to generate *N*3-adducted thymidine has been developed previously.[299,374,375] The presence or lack of the intradimer phosphate group allows for a similar comparison as those drawn in **CHAPTER V** for hPol η (or any other Pol tested for that matter). Moreover, the linker protrudes the Watson-Crick face and would disrupt H-bonding with the incoming nucleotide. The *N*3-atom of dT was selected on the basis of its chemical stability, in comparison to the more labile O^4-alkylene-dT adducts.[150,371] The *N*3-alkylene-dT adducts have, also, previously served as mimics of the *N*3-alkylene-dC adducts,[298,374,376] which may aid in deciphering whether hPol η is proficient in

bypassing *N*3-alkylene-pyrimidines in general. Finally, comparison of this study with the investigation presented in **CHAPTER V** would supply insights on the substrate discrimination of hPol η towards flexible pyrimidine-pyrimidine and purine-purine IaCL DNA.



Scheme 7.4: Mono-adducts **7.16 a**,**b**,**c** have been prepared previously.[374] Reagents and conditions: (i) 5'-*O*-(*tert*-butyldimethylsilyl)-3'-*O*-levulinyl-thymidine, DBU, MeCN, 21°C, 16 h. (ii) Cl-P(OCEt)N(*i*Pr)$_2$, DIPEA, THF, 21°C, 30 min.



Scheme 7.5: Reagents and conditions for **a**-series: (i) TBS-Cl, Im, CH_2Cl_2, 21°C, 4 h. (ii) 2-iodoethyl-1-phenoxyacetate, DBU, MeCN, 21°C, 4 h. (iii) *n*-butylamine, THF, 45°C, 4h. (iv) I_2, Ph_3P, Im, THF, 21°C, 1 h. (v) 1) 5'-*O*-(4,4'-dimethoxytrityl)-3'-*O*-levulinyl-thymidine, DBU, MeCN, 21°C, 16 h. 2) $N_2H_4 \bullet H_2O$, pyridinium acetate (1:1, v/v), 21°C, 30 min. (vi) 2-chlorophenyl dichlorophosphate, 1,2,4-triazole, NEt_3, THF, Py, 21°C, 2 h. (vii) 1. TBAF, THF, 21°C, 1 h. 2. MSNT, Py, 21°C, 4 h. (viii) Cl-P(OCEt)N(*i*Pr)$_2$, DIPEA, THF, 21°C, 30 min.

Reagents and conditions for **b**-series: Same as **a**-series except step (ii) 4-iodobutyl-1-phenoxyacetate, DBU, MeCN, 21°C, 4 h.

Reagents and conditions for **c**-series: Same as **a**-series except step (ii) 7-iodoheptyl-1-phenoxyacetate, DBU, MeCN, 21°C, 4 h.

References

(1) Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D., and Darnell, J. (2004) Molecular Cell Biology 5th ed. W. H. Freeman, New York.
(2) Box, H. C., Budzinski, E. E., Dawidzik, J. B., Gobey, J. S., and Freund, H. G. (1997) Free radical-induced tandem base damage in DNA oligomers. *Free Radic. Biol. Med. 23*, 1021–1030.
(3) Box, H. C., Patrzyc, H. B., Dawidzik, J. B., Wallace, J. C., Freund, H. G., Iijima, H., and Budzinski, E. E. (2000) Double base lesions in DNA X-irradiated in the presence or absence of oxygen. *Radiat Res. 153*, 442–446.
(4) Box, H. C., Dawidzik, J. B., and Budzinski, E. E. (2001) Free radical-induced double lesions in DNA. *Free Radic. Biol. Med. 31*, 856–868.
(5) Hussain, S. P., Hofseth, L. J., and Harris, C. C. (2003) Radical causes of cancer. *Nat. Rev. Cancer 3*, 276–285.
(6) Hoeijmakers, J. H. J. (2001) Genome maintenance mechanisms for preventing cancer. *Nature 411*, 366–374.
(7) Lindahl, T. (1993) Instability and decay of the primary structure of DNA. *Nature 362*, 709–715.
(8) Gates, K. S., Nooner, T., and Dutta, S. (2004) Biologically relevant chemical reactions of N7-alkylguanine residues in DNA. *Chem. Res. Toxicol. 17*, 840–852.
(9) Lindahl, T., and Nyberg, B. (1972) Rate of depurination of native deoxyribonucleic acid. *Biochemistry 11*, 3610–3618.
(10) Lindahl, T., and Karlström, O. (1973) Heat-induced depyrimidination of deoxyribonucleic acid in neutral solution. *Biochemistry 12*, 5151–5154.
(11) Haddow, A. (1973) On the Biological Alkylating Agents. *Perspect. Biol. Med. 16*, 503–524.
(12) Goldacre, R. J., Loveless, A., and Ross, W. C. J. (1949) Mode of production of chromosome abnormalities by the nitrogen mustards: The possible role of cross-linking. *Nature 163*, 667–669.
(13) Lawley, P. D., and Brookes, P. (1967) Interstrand cross-linking of DNA by difunctional alkylating agents. *J. Mol. Biol. 25*, 143–160.
(14) Rydberg, B., and Lindahl, T. (1982) Nonenzymatic methylation of DNA by the intracellular methyl group donor S-adenosyl-L-methionine is a potentially mutagenic reaction. *EMBO J. 1*, 211–216.
(15) Nair, U., Bartsch, H., and Nair, J. (2007) Lipid peroxidation-induced DNA damage in cancer-prone inflammatory diseases: A review of published adduct types and levels in humans. *Free Radic. Biol. Med. 43*, 1109–1120.
(16) Marnett, L. J. (1999) Lipid peroxidation-DNA damage by malondialdehyde. *Mutat. Res. 424*, 83–95.
(17) Beranek, D. T. (1990) Distribution of methyl and ethyl adducts following alkylation with monofunctional alkylating agents. *Mutat. Res-Fund. Mol. M. 231*, 11–30.
(18) Stone, M. P., Huang, H., Brown, K. L., and Shanmugam, G. (2011) Chemistry and structural biology of DNA damage and biological consequences. *Chem. Biodivers. 8*, 1571–1615.
(19) Warren, J. J., Forsberg, L. J., and Beese, L. S. (2006) The structural basis for the mutagenicity of O6-methyl-guanine lesions. *PNSA 103*, 19701–19706.
(20) Singer, B. (1986) O-Alkyl pyrimidines in mutagenesis and carcinogenesis: Occurrence and significance. *Cancer Res. 46*, 4879–4885.
(21) Altshuler, K. B., Hodes, C. S., and Essigmann, J. M. (1996) Intrachromosomal probes for mutagenesis by alkylated DNA bases replicated in mammalian cells : A comparison of the

mutagenicities of O4 -methylthymine and O6 -Methylguanine in cells with different DNA repair backgrounds. *Chem. Res. Toxicol. 9*, 980–987.
(22) Singer, B., Spengler, S. J., Fraenkel-Conrat, H., and Kuśmierek, J. T. (1986) O4-Methyl, -ethyl, or -isopropyl substituents on thymidine in poly(dA-dT) all lead to transitions upon replication. *Proc. Natl. Acad. Sci. USA 83*, 28–32.
(23) Patel, D. J., Shapiro, L., Kozlowski, S. A., Gaffney, B. L., and Jones, R. A. (1986) Structural studies of the O6meG•C interaction in the d(C-G-C-G-A-A-T-T-C-O6meG-C-G) duplex. *Biochemistry 25*, 1027–1036.
(24) Patel, D. J., Shapiro, L., Kozlowski, S. A., Gaffney, B. L., and Jones, R. A. (1986) Structural studies of the O6meG•T interaction in the d(C-G-T-G-A-A-T-T-C-O6meG-C-G) duplex. *Biochemistry 25*, 1036–1042.
(25) Kalnik, M. W., Kouchakdjian, M., Li, B. F. L., Swann, P. F., and Patel, D. J. (1988) Base pair mismatches and carcinogen-modified bases in DNA: An NMR study of A•C and A•O4meT pairing in dodecanucleotide duplexes. *Biochemistry 27*, 100–108.
(26) Kalnik, M. W., Kouchakdjian, M., Li, B. F. L., Swann, P. F., and Patel, D. J. (1988) Base pair mismatches and carcinogen-modified bases in DNA: An NMR study. *Biochemistry 27*, 108–115.
(27) Kaina, B., Christmann, M., Naumann, S., and Roos, W. P. (2007) MGMT: key node in the battle against genotoxicity, carcinogenicity and apoptosis induced by alkylating agents. *DNA Repair (Amst). 6*, 1079–1099.
(28) Kondo, N., Takahashi, A., Ono, K., and Ohnishi, T. (2010) DNA damage induced by alkylating agents and repair pathways. *J. Nucleic Acids* 543531.
(29) Wheeler, G. P. (1962) Studies related to the mechanisms of action of cytotoxic alkylating agents: A review. *Cancer Res.*
(30) Rajski, S. R., and Williams, R. M. (1998) DNA cross-linking agents as antitumor drugs. *Chem. Rev. 98*, 2723–2796.
(31) Shrivastav, N., Li, D., and Essigmann, J. M. (2010) Chemical biology of mutagenesis and DNA repair: cellular responses to DNA alkylation. *Carcinogenesis 31*, 59–70.
(32) Marnett, L. J., and Burcham, P. C. (1993) Endogenous DNA adducts: potential and paradox. *Chem. Res. Toxicol. 6*, 771–785.
(33) Kleibl, K. (2002) Molecular mechanisms of adaptive response to alkylating agents in Escherichia coli and some remarks on O6-methylguanine DNA-methyltransferase in other organisms. *Mutat. Res. - Rev. Mutat. Res. 512*, 67–84.
(34) B. Singer, D. G. (1983) Molecular biology of mutagens and carcinogens. Plenum Press, New York.
(35) A, P., and B, P. (1981) Molecular electrostatic potential of the nucleic acids. *Q Rev Biophys. 3*, 289–380.
(36) Singer, B. (1975) The chemical effects of nucleic acid alkylation and their relationship to mutagenesis and carcinogenesis. *Prog. Nucleic Acid Res. Mol. Biol. 15*, 219–284.
(37) Rasouli-Nia, A., Sibghat-Ullah, Mirzayans, R., Paterson, M., and Day, R. I. (1994) On the quantitative relationship between O6-methylguanine residues in genomic DNA and production of sister-chromatid exchanges, mutations and lethal events in a Mer- human tumor cell line. *Mutat Res. 314*, 99–113.
(38) Tubbs, J. L., Pegg, A. E., and Tainer, J. A. (2008) DNA binding, nucleotide flipping, and the helix-turn-helix in base repair by O6-alkylguanine-DNA alkyltransferase and its implications for cancer chemotherapy. *DNA Repair 6*, 1100–1115.

(39) Peterson, L. A., and Hecht, S. S. (1991) O6-Methylguanine is a critical determinant of 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone tumorigenesis in A/J mouse lung. *Cancer Res. 51*, 5557–5564.
(40) Swenberg, J. A., Dyroff, M. C., Bedell, M. A., Popp, J. A., Huht, N., Kirsteint, U. W. E., and Rajewskyf, M. F. (1984) O4-Ethyldeoxythymidine, but not O6-ethyldeoxyguanosine, accumulates in hepatocyte DNA of rats exposed continuously to diethylnitrosamine. *Proc. Natl. Acad. Sci. USA 81*, 1692–1695.
(41) Zak, P., Kleibl, K., and Laval, F. (1994) O6-Methylguanine and O4-methylthymine by human the and rat O6-methylguanine-DNA methyltransferases. *J Biol. Chem. 269*, 730–733.
(42) Voigt, J. M., and Topal, M. D. (1995) O6-Methylguanine-induced replication blocks. *Carcinogenesis 16*, 1775–1782.
(43) Singh, J., Su, L., and Snow, E. T. (1996) Replication across O6-methylguanine by human DNA polymerase β in vitro. *Biochemistry 271*, 28391–28398.
(44) Haracska, L., Prakash, S., and Prakash, L. (2000) Replication past O6 -Methylguanine by Yeast and Human DNA Polymerase eta. *Mol. Cell. Biol. 20*, 8001–07.
(45) Choi, J.-Y., Chowdhury, G., Zang, H., Angel, K. C., Vu, C. C., Peterson, L., and Guengerich, F. P. (2006) Translesion synthesis across O6-alkylguanine DNA adducts by recombinant human DNA polymerases. *J Biol. Chem. 281*, 38244–38256.
(46) Dosanjh, M. K., Galeros, G., Goodman, M. F., and Singer, B. (1991) Kinetics of extension of O6-methylguanine paired with cytosine or thymine in defined oligonucleotide sequences. *Biochemistry 30*, 11595–11599.
(47) Andersen, N., Wang, J., Wang, P., Jiang, Y., and Wang, Y. (2012) In-vitro replication studies on O2-methylthymidine and O4-methylthymidine. *Chem. Res. Toxicol. 25*, 2523–2531.
(48) Swann, P. F. (1990) Why do O6-alkylguanine and O4-alkylthymine miscode? The relationship between the structure of DNA containing O6-alkylguanine and O4-alkylthymine and the mutagenic properties of these bases. *Mutat. Res. 233*, 81–94.
(49) Duckett, D. R., Drummond, J. T., Murchiet, A. I. H., Reardont, J. T., Sancart, A., Lilleyt, D. M. J., and Modrich, P. (1996) Human MutScv recognizes damaged DNA base pairs containing O6-methylguanine, O4-methylthymidine, or the cisplatin-d(GpG) adduct. *Proc. Natl. Acad. Sci. USA 93*, 6443–6447.
(50) Kat, A., Thilly, W. G., Fang, W. H., Longley, M. J., Li, G. M., and Modrich, P. (1993) An alkylation-tolerant, mutator human cell line is deficient in strand-specific mismatch repair. *Proc. Natl. Acad. Sci. U. S. A. 90*, 6424–6428.
(51) Branch, P., Aquilina, G., Bignami, M., and Karran, P. (1993) Defective mismatch binding and a mutator phenotype in cells tolerant to DNA damage. *Nature 362*, 652–654.
(52) Sun, G., Jin, S., and Baskaran, R. (2009) MMR/c-Abl-dependent activation of ING2/p73alpha signaling regulates the cell death response to N-methyl-N'-nitro-N-nitrosoguanidine. *Exp. Cell Res. 315*, 3163–3175.
(53) Jiricny, J. (2006) The multifaceted mismatch-repair system. *Nat. Rev. Mol. Cell Biol. 7*, 335–346.
(54) Noll, D. M., Mason, T. M., and Miller, P. S. (2006) Formation and repair of interstrand cross-links in DNA. *Chem. Rev. 106*, 277–301.
(55) Lawley, P. D., and Phillips, D. H. (1996) DNA adducts from chemotherapeutic. *Mutat. Res. 355*, 13–40.
(56) Magaňa-Schwencke, N., Henriques, J.-A. P., Chanet, R., and Moustacchi, E. (1982) The fate of 8-methoxypsoralen photoinduced crosslinks in nuclear and mitochondrial yeast DNA :

Comparison of wild-type and repair-deficient strains. *Proc. Natl. Acad. Sci. USA 79*, 1722–1726.
(57) Dronkert, M. L., and Kanaar, R. (2001) Repair of DNA interstrand cross-links. *Mutat. Res. 486*, 217–247.
(58) Noll, D. M., Webba, M., Noronha, A. M., Wilds, C. J., Colvin, O. M., Gamcsik, M. P., and Miller, P. S. (2005) Structure , flexibility, and repair of two different orientations of the same alkyl interstrand DNA cross-link. *Biochemistry 44*, 6764–6775.
(59) Perry, J. J. P., Cotner-Gohara, E., Ellenberger, T., and Tainer, J. a. (2010) Structural dynamics in DNA damage signaling and repair. *Curr. Opin. Struct. Biol. 20*, 283–294.
(60) Scott, B. R., Pathak, M. A., and Mohn, G. R. (1976) Molecular and genetic basis of furocoumarin reactions. *Mutat. Res. 39*, 29–74.
(61) Price, N. E., Johnson, K. M., Wang, J., Fekry, M. I., Wang, Y., and Gates, K. S. (2014) Interstrand DNA-DNA cross-link formation between adenine residues and abasic sites in duplex dna. *J. Am. Chem. Soc. 136*, 3483–3490.
(62) Yang, Z., Price, N. E., Johnson, K. M., and Gates, K. S. (2015) Characterization of interstrand DNA–DNA cross-links derived from abasic sites using bacteriophage ϕ29 DNA polymerase. *Biochemistry 54*, 4259–4266.
(63) Price, N. E., Catalano, M. J., Liu, S., Wang, Y., and Gates, K. S. (2015) Chemical and structural characterization of interstrand cross-links formed between abasic sites and adenine residues in duplex DNA. *Nucleic Acids Res. 43*, 3434–3441.
(64) Fang, Q., Noronha, A. M., Murphy, S. P., Wilds, C. J., Tubbs, J. L., Tainer, J. A., Chowdhury, G., Guengerich, F. P., and Pegg, A. E. (2008) Repair of O6-G-alkyl-O6-G interstrand cross-links by human O6-alkylguanine-DNA alkyltransferase. *Biochemistry 47*, 10892–10903.
(65) McManus, F. P., Fang, Q., Booth, J. D. M., Noronha, A. M., Pegg, A. E., and Wilds, C. J. (2010) Synthesis and characterization of an O6-2’-deoxyguanosine-alkyl-O6-2'-deoxyguanosine interstrand cross-link in a 5'-GNC motif and repair by human O6-alkylguanine-DNA alkyltransferase. *Org. Biomol. Chem. 8*, 4414–4426.
(66) O’Donovan, A., Davies, a a, Moggs, J. G., West, S. C., and Wood, R. D. (1994) XPG endonuclease makes the 3’ incision in human DNA nucleotide excision repair. *Nature 371*, 432–435.
(67) Huang, Y., and Li, L. (2013) DNA crosslinking damage and cancer - a tale of friend and foe. *Transl. Cancer Res. 2*, 144–154.
(68) Kozmin, S. G., and Jinks-Robertson, S. (2013) The mechanism of nucleotide excision repair-mediated UV-induced mutagenesis in nonproliferating cells. *Genetics 193*, 803–817.
(69) Iwamoto, T., Hiraku, Y., Olkawa, S., Mizutani, H., Kojima, M., and Kawanishi, S. (2004) DNA intrastrand cross-link at the 5′-GA-3′ sequence formed by busulfan and its role in the cytotoxic effect. *Cancer Sci. 95*, 454–458.
(70) Rink, S. M., Lipman, R., Alley, S. C., Hopkins, P. B., and Tomasz, M. (1996) Bending of DNA by the mitomycin C-induced, GpG intrastrand cross-link. *Chem. Res. Toxicol. 9*, 382–389.
(71) Chaney, S. G., Campbell, S. L., Bassett, E., and Wu, Y. (2005) Recognition and processing of cisplatin- and oxaliplatin-DNA adducts. *Crit. Rev. Oncol. Hematol. 53*, 3–11.
(72) Rink, S. M., and Hopkins, P. B. (1995) Direct evidence for DNA intrastrand cross-linking by the nitrogen mustard mechlorethamine in synthetic oligonucleotides. *Bioorg. Med. Chem. Lett. 5*, 2845–2850.
(73) Kallama, S., and Hemminki, K. (1986) Stabilities of 7-alkylguanosines and 7-deoxyguanosines formed by phosphoramide mustard and nitrogen mustard. *Chem. Biol. Interact.*

57, 85–96.
(74) Baik, M.-H., Friesner, R. A., and Lippard, S. J. (2002) Theoretical study on the stability of N-glycosyl bonds: why does N7-platination not promote depurination? *J. Am. Chem. Soc. 124*, 4495–4503.
(75) You, Y. H., Szabó, P. E., and Pfeifer, G. P. (2000) Cyclobutane pyrimidine dimers form preferentially at the major p53 mutational hotspot in UVB-induced mouse skin tumors. *Carcinogenesis 21*, 2113–2117.
(76) Sinha, R. P., and Häder, D.-P. (2002) UV-induced DNA damage and repair: a review. *Photochem. Photobiol. Sci. 1*, 225–236.
(77) Bellon, S., Ravanat, J. L., Gasparutto, D., and Cadet, J. (2002) Cross-linked thymine-purine base tandem lesions: synthesis, characterization, and measurement in γ-irradiated isolated DNA. *Chem. Res. Toxicol. 15*, 598–606.
(78) Gu, C., and Wang, Y. (2004) LC-MS/MS identification and yeast polymerase η bypass of a novel γ-irradiation-induced intrastrand cross-link lesion G[8-5]C. *Biochemistry 43*, 6745–6750.
(79) Zhang, Q. Q., and Wang, Y. (2005) Generation of 5-(2'-deoxycytidyl)methyl radical and the formation of intrastrand cross-link lesions in oligodeoxyribonucleotides. *Nucleic Acids Res. 33*, 1593–1603.
(80) Esteban-Fernández, D., Moreno-Gordaliza, E., Cañas, B., Palacios, M. A., and Gómez-Gómez, M. M. (2010) Analytical methodologies for metallomics studies of antitumor Pt-containing drugs. *Metallomics 2*, 19–38.
(81) Hong, H., Cao, H., and Wang, Y. (2007) Formation and genotoxicity of a guanine-cytosine intrastrand cross-link lesion in vivo. *Nucleic Acids Res. 35*, 7118–7127.
(82) Yao, J., Dixon, K., and Carty, M. P. (2001) A single (6-4) photoproduct inhibits plasmid DNA replication in xeroderma pigmentosum variant cell extracts. *Environ. Mol. Mutagen. 38*, 19–29.
(83) Zamble, D. B., Mu, D., Reardon, J. T., Sancar, A., and Lippard, S. J. (1996) Repair of cisplatin-DNA adducts by the mammalian excision nuclease. *Biochemistry 35*, 10004–10013.
(84) McA'Nulty, M. M., Whitehead, J. P., and Lippard, S. J. (1996) Binding of Ixr1, a yeast HMG-domain protein, to cisplatin-DNA adducts in vitro and in vivo. *Biochemistry 35*, 6089–6099.
(85) Huang, J. C., Zamble, D. B., Reardon, J. T., Lippard, S. J., and Sancar, a. (1994) HMG-domain proteins specifically inhibit the repair of the major DNA adduct of the anticancer drug cisplatin by human excision nuclease. *Proc. Natl. Acad. Sci. U. S. A. 91*, 10394–10398.
(86) Trimmer, E. E., Zamble, D. B., Lippard, S. J., and Essigmann, J. M. (1998) Human testis-determining factor SRY binds to the major DNA adduct of cisplatin and a putative target sequence with comparable affinities. *Biochemistry 37*, 352–362.
(87) S. Chow, C., M. Barnes, C., and J. Lippard, S. (1995) A single HMG domain in high-mobility group 1 protein binds to DNAs as small as 20 base pairs containing the major cisplatin adduct. *Biochemistry 34*, 2956–2964.
(88) Ohndorf, U. M., Whitehead, J. P., Raju, N. L., and Lippard, S. J. (1997) Binding of tsHMG, a mouse testis-specific HMG-domain protein, to cisplatin-DNA adducts. *Biochemistry 36*, 14807–14815.
(89) Park, H., Zhang, K., Ren, Y., Nadji, S., Sinha, N., Taylor, J., and Kang, C. (2002) Crystal structure of a DNA decamer containing a cis-syn thymine dimer. *Proc. Natl. Acad. Sci. U. S. A. 99*, 15965–15970.
(90) Yokoyama, H., Mizutani, R., Satow, Y., Komatsu, Y., Ohtsuka, E., and Nikaido, O. (2000)

Crystal structure of the 64M-2 antibody Fab fragment in complex with a DNA dT(6-4)T photoproduct formed by ultraviolet radiation. *J. Mol. Biol. 299*, 711–723.
(91) Yokoyama, H., and Mizutani, R. (2014) Structural biology of DNA (6-4) photoproducts formed by ultraviolet radiation and interactions with their binding proteins. *Int. J. Mol. Sci. 15*, 20321–20338.
(92) Gelasco, a, and Lippard, S. J. (1998) NMR solution structure of a DNA dodecamer duplex containing a cis-diammineplatinum(II) d(GpG) intrastrand cross-link, the major adduct of the anticancer drug cisplatin. *Biochemistry 37*, 9230–9239.
(93) Garrec, J., Patel, C., Rothlisberger, U., and Dumont, E. (2012) Insights into intrastrand cross-link lesions of DNA from QM/MM molecular dynamics simulations. *J. Am. Chem. Soc. 134*, 2111–2119.
(94) Dumont, E., Dršata, T., Guerra, C. F., and Lankaš, F. (2015) Insights into the structure of intrastrand cross-Link DNA lesion-containing oligonucleotides: G[8–5m]T and G[8–5]C from molecular dynamics simulations. *Biochemistry 54*, 1259–1267.
(95) Travers, A. A. (1990) Why bend DNA ? *Cell 60*, 177–180.
(96) Nagaich, a K., Appella, E., and Harrington, R. E. (1997) DNA bending is essential for the site-specific recognition of DNA response elements by the DNA binding domain of the tumor suppressor protein p53. *J. Biol. Chem. 272*, 14842–14849.
(97) Wheate, N. J., Walker, S., Craig, G. E., and Oun, R. (2010) The status of platinum anticancer drugs in the clinic and in clinical trials. *Dalt. Trans. 39*, 8097–8340.
(98) Brabec, V., Kašpárková, J., Vrána, O., Nováková, O., Cox, J. W., Qu, Y., and Farrell, N. (1999) DNA modifications by a novel bifunctional trinuclear platinum phase I anticancer agent. *Biochemistry 38*, 6781–6790.
(99) Cox, J. W., Berners-price, S. J., Davies, M. S., Qu, Y., Farrell, N., Commonwealth, V., and Uni, V. (2006) Kinetic analysis of the stepwise formation of a long-range DNA interstrand cross-link by a dinuclear platinum antitumor complex: Evidence for aquated intermediates and formation of both kinetically and thermodynamically controlled Conformers. *J. Am. Chem. Soc. 4*, 1316–1326.
(100) Stingele, J., and Jentsch, S. (2015) DNA–protein crosslink repair. *Nat. Rev. Mol. Cell Biol. 16*, 455–460.
(101) Chen, S. H., Chan, N.-L., and Hsieh, T. (2013) New mechanistic and functional insights into DNA topoisomerases. *Annu. Rev. Biochem. 82*, 139–170.
(102) Pourquier, P., Ueng, L. M., Kohlhagen, G., Mazumder, A., Gupta, M., Kohn, K. W., and Pommier, Y. (1997) Effects of uracil incorporation, DNA mismatches, and abasic sites on cleavage and religation activities of mammalian topoisomerase I. *J. Biol. Chem. 272*, 7792–7796.
(103) Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J. R., Cole, P. A., Casero, R. A., and Shi, Y. (2004) Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell 119*, 941–953.
(104) Solomon, M. J., Larsen, P. L., and Varshavsky, A. (1988) Mapping protein-DNA interactions in vivo with formaldehyde: evidence that histone H4 is retained on a highly transcribed gene. *Cell 53*, 937–947.
(105) Lu, K., Ye, W., Zhou, L., Collins, L. B., Chen, X., Gold, A., Ball, L. M., and Swenberg, J. a. (2010) Structural characterization of formaldehyde-induced cross-links between amino acids and deoxynucleosides and their oligomers. *J. Am. Chem. Soc. 132*, 3388–3399.
(106) Loeber, R. L., Michaelson-Richie, E. D., Codreanu, S. G., Liebler, D. C., Campbell, C., and Tretyakova, N. Y. (2009) Proteomic analysis of DNA-protein cross-linking by antitumor

nitrogen mustards. *Chem. Res. Toxicol. 22*, 1151–1162.
(107) K. Wozniak, Z. W. (2000) Induction of DNA-protein cross-links by platinum compounds. *Z. Naturforsch. C. 55*, 731–736.
(108) Zhang, H., Koch, C. J., Wallen, C. A., and Wheeler, K. T. (1995) Radiation-induced DNA damage in tumors and normal tissues. III. Oxygen dependence of the formation of strand breaks and DNA-protein crosslinks. *Radiat. Res. 142*, 163–168.
(109) Nakano, T., Ouchi, R., Kawazoe, J., Pack, S. P., Makino, K., and Ide, H. (2012) T7 RNA polymerases backed up by covalently trapped proteins catalyze highly error prone transcription. *J. Biol. Chem. 287*, 6562–6572.
(110) Nakano, T., Miyamoto-Matsubara, M., Shoulkamy, M. I., Salem, A. M. H., Pack, S. P., Ishimi, Y., and Ide, H. (2013) Translocation and stability of replicative DNA helicases upon encountering DNA-protein cross-links. *J. Biol. Chem. 288*, 4649–4658.
(111) Fu, Y. V, Yardimci, H., Long, D. T., Ho, T. V., Guainazzi, A., Bermudez, V. P., Hurwitz, J., van Oijen, A., Schärer, O. D., and Walter, J. C. (2011) Selective bypass of a lagging strand roadblock by the eukaryotic replicative DNA helicase. *Cell 146*, 931–941.
(112) Yeo, J. E., Wickramaratne, S., Khatwani, S., Wang, Y., Distefano, M. D., and Tretyakova, N. Y. (2014) Synthesis of site-specific DNA−protein conjugates and their effects on DNA replication. *ACS Chem. Biol. 9*, 1860–1868.
(113) Novakova, O., Kasparkova, J., Malina, J., Natile, G., and Brabec, V. (2003) DNA-protein cross-linking by trans-[PtCl2(E-iminoether)(2)]. A concept for activation of the trans geometry in platinum antitumor complexes. *Nucleic Acids Res. 31*, 6450–6460.
(114) Chválová, K., Brabec, V., and Kašpárková, J. (2007) Mechanism of the formation of DNA-protein cross-links by antitumor cisplatin. *Nucleic Acids Res. 35*, 1812–1821.
(115) Andreassen, P. R., and Ren, K. (2009) Fanconi anemia proteins, DNA interstrand crosslink repair pathways, and cancer therapy. *Curr. Cancer Drug Targets 9*, 101–117.
(116) Marteijn, J. a, Lans, H., Vermeulen, W., and Hoeijmakers, J. H. J. (2014) Understanding nucleotide excision repair and its roles in cancer and ageing. *Nat Rev Mol Cell Biol 15*, 465–481.
(117) Kazak, L., Reyes, A., and Holt, I. J. (2012) Minimizing the damage: repair pathways keep mitochondrial DNA intact. *Nat. Rev. Mol. Cell Biol. 13*, 659–671.
(118) Panier, S., and Boulton, S. J. (2013) Double-strand break repair: 53BP1 comes into focus. *Nat. Rev. Mol. Cell Biol. 15*, 7–18.
(119) Renkawitz, J., Lademann, C. A., and Jentsch, S. (2014) Mechanisms and principles of homology search during recombination. *Nat. Rev. Mol. Cell Biol. 15*, 369–383.
(120) Robertson, A. B., Klungland, A., Rognes, T., and Leiros, I. (2009) Base excision repair: the long and short of it. *Cell. Mol. Life Sci. 66*, 981–993.
(121) Dianov, G. L., and Hubscher, U. (2013) Mammalian base excision repair: the forgotten archangel. *Nucleic Acids Res. 41*, 3483–3490.
(122) San Filippo, J., Sung, P., and Klein, H. (2008) Mechanism of eukaryotic homologous recombination. *Annu. Rev. Biochem. 77*, 229–257.
(123) Weterings, E., and Chen, D. J. (2008) The endless tale of non-homologous end-joining. *Cell Res. 18*, 114–124.
(124) Deriano, L., and Roth, D. B. (2013) Modernizing the nonhomologous end-joining repertoire: alternative and classical NHEJ share the stage. *Annu. Rev. Genet. 47*, 433–455.
(125) Mishina, Y., Duguid, E. M., and He, C. (2006) Direct reversal of DNA alkylation damage. *Chem. Rev. 106*, 215–232.
(126) Li, Y. F., Kim, S. T., and Sancar, A. (1993) Evidence for lack of DNA photoreactivating

enzyme in humans. *Proc. Natl. Acad. Sci. U. S. A. 90*, 4389–4393.
(127) Carell, T., Burgdorf, L. T., Kundu, L. M., and Cichon, M. (2001) The mechanism of action of DNA photolyases. *Curr. Opin. Chem. Biol. 5*, 491–498.
(128) Sancar, A. (2003) Structure and function of DNA photolyase and cryptochrome blue-light photoreceptors. *Chem. Rev. 103*, 2203–2237.
(129) Mishina, Y., and He, C. (2006) Oxidative dealkylation DNA repair mediated by the mononuclear non-heme iron AlkB proteins. *J. Inorg. Biochem. 100*, 670–678.
(130) Fedeles, B. I., Singh, V., Delaney, J. C., Li, D., and Essigmann, J. M. (2015) The AlkB family of Fe(II)/α-ketoglutarate-dependent dioxygenases: Repairing nucleic acid alkylation damage and beyond. *J. Biol. Chem. 290*, 20734–20742.
(131) Sedgwick, B. (2004) Repairing DNA-methylation damage. *Nat. Rev. Mol. Cell Biol. 5*, 148–157.
(132) Pegg, A. E. (2011) Multifaceted roles of alkyltransferase and related proteins in DNA repair, DNA damage, resistance to chemotherapy, and research tools. *Chem. Res. Toxicol. 24*, 618–639.
(133) Gerson, S. L. (2004) MGMT: its role in cancer aetiology and cancer therapeutics. *Nat. Rev. Cancer 4*, 296–307.
(134) Liu, L., and Gerson, S. L. (2006) Targeted modulation of MGMT: Clinical implications. *Clin. Cancer Res. 12*, 328–331.
(135) Srivenugopal, K. S., Yuan, X., Friedman, H. S., and Ali-osman, F. (1996) Ubiquitination-dependent proteolysis of O6-methylguanine-DNA methyltransferase in human and murine tumor cells following inactivation with O6-benzylguanine. *Biochemistry 35*, 1328–1334.
(136) Daniels, D. S., and Tainer, J. a. (2000) Conserved structural motifs governing the stoichiometric repair of alkylated DNA by O6-alkylguanine-DNA alkyltransferase. *Mutat. Res. 460*, 151–163.
(137) Daniels, D. S., Woo, T. T., Luu, K. X., Noll, D. M., Clarke, N. D., Pegg, A. E., and Tainer, J. A. (2004) DNA binding and nucleotide flipping by the human DNA repair protein AGT. *Nat. Struct. Mol. Biol. 11*, 714–720.
(138) Dolan, M. E., and Pegg, A. E. (1997) O6-benzylguanine and its role in chemotherapy. *Clin. Cancer Res. 3*, 837–847.
(139) Wedge, S. R., Porteous, J. K., and Newlands, E. S. (1997) Effect of single and multiple administration of an O6-benzylguanine/temozolomide combination: an evaluation in a human melanoma xenograft model. *Cancer Chemoth. Pharm. 40*, 266–272.
(140) Olsson, M., and Lindahl, T. (1980) Repair of alkylated DNA in Escherichia coli. *J. Biol. Chem. 255*, 10569–10571.
(141) Lindahl, T., Demple, B., and Robins, P. (1982) Suicide inactivation of the E. coli O6-methylguanine-DNA methyltransferase. *EMBO J. 1*, 1359–1363.
(142) Samson, L., and Cairns, J. (1977) A new pathway for DNA repair in Escherichia coli. *Nature 267*, 281–283.
(143) Wilkinson, M. C., Potter, P. M., Cawkwell, L., Georgiadis, P., Patel, D., Swann, P. F., and Margison, G. P. (1989) Purification of the E. coli ogt gene product to homogeneity and its rate of action on O6-methylguanine, O6-ethylguanine and O4-methylthymine in dodecadeoxyribonucleotides. *Nucleic Acids Res. 21*, 8475–8484.
(144) Teo, I., Sedgwick, B., Kilpatrick, M. W., McCarthy, T. V, and Lindahl, T. (1986) The intracellular signal for induction of resistance to alkylating agents in E. coli. *Cell 45*, 315–324.
(145) Sakumi, K., and Sekiguchi, M. (1989) Regulation of expression of the ada gene controlling

the adaptive response. Interactions with the ada promoter of the Ada protein and RNA polymerase. *J Mol Biol 205*, 373–385.
(146) Graves, R. J., Li, B. F., and Swann, P. F. (1989) Repair of O6-methylguanine, O6-ethylguanine, O6-isopropylguanine and O4-methylthymine in synthetic oligodeoxynucleotides by Escherichia coli ada gene O6-alkylguanine-DNA-alkyltransferase. *Carcinogenesis 10*, 661–666.
(147) Paalman, S. R., Sung, C., and Clarke, N. D. (1997) Specificity of DNA repair methyltransferases determined by competitive inactivation with oligonucleotide substrates: evidence that Escherichia coli Ada repairs O6-methylguanine and O4-methylthymine with similar efficiency. *Biochemistry 36*, 11118–11124.
(148) Hospital, C., Radium, H., and Hospital, M. (1989) Alkyltransferase Assay Extracts of E. coli harbouring various plasmids, fractions collected during chromatography. *Nucleic Acids Res. 17*, 8475–8484.
(149) McManus, F. P., and Wilds, C. J. (2014) O6-Alkylguanine-DNA alkyltransferase-mediated repair of O4-alkylated 2′-Deoxyuridines. *ChemBioChem 15*, 1966–1977.
(150) McManus, F. P., O'Flaherty, D. K., Noronha, A. M., and Wilds, C. J. (2012) O4-alkyl-2'-deoxythymidine cross-linked DNA to probe recognition and repair by O6-alkylguanine DNA alkyltransferases. *Org. Biomol. Chem. 10*, 7078–7090.
(151) McManus, F. P., Khaira, A., Noronha, A. M., and Wilds, C. J. (2013) Preparation of covalently linked complexes between DNA and O6-alkylguanine-DNA alkyltransferase using interstrand cross-linked DNA. *Bioconjugate Chem. 24*, 224–233.
(152) Rebeck, G. W., Smith, C. M., Goad, D. L., and Samson, L. (1989) Characterization of the major DNA repair methyltransferase activity in unadapted Escherichia coli and identification of a similar activity in Salmonella typhimurium. *J. Bacteriol. 171*, 4563–4568.
(153) Sassanfar, M., Dosanjh, M. K., Essigmann, J. M., and Samson, L. (1991) Relative efficiencies of the bacterial, yeast, and human DNA methyltransferases for the repair of O6-methylguanine and O4-methylthymine: Suggestive evidence for O4-methylthymine repair by eukaryotic methyltransferases. *J. Biol. Chem. 266*, 2767–2771.
(154) Pegg, A. E., Dolan, M. E., Scicchitano, D., and Morimoto, K. (1985) Studies of the repair of O6-alkylguanine and O4-alkylthymine in DNA by alkyltransferases from mammalian cells and bacteria. *Environ. Health Perspect. 62*, 109–114.
(155) Schoonhoven, N. M., Murphy, S. P., O'Flaherty, D. K., Noronha, A. M., Kornblatt, M. J., and Wilds, C. J. (2008) Synthesis, biophysical and repair studies of O6-2'-deoxyguanosine adducts by Escherichia coli OGT. *Nucleic Acids Symp. Ser. No. 52* 449–50.
(156) Schoonhoven, N. M. (2010) The role of residue-5 on O6-alkylated DNA duplex repair by O6-alkylguanine-DNA-alkyltransferase. Concordia University.
(157) Newlands, E. E., Stevens, M. F. G., Wedge, S. R., Wheelhouse, R. T., and Brock, C. (1997) Temozolomide: a review of its discovery, chemical properties, pre-clinical development and clinical trials. *Cancer Treat. Rev. 23*, 35–61.
(158) Brent, T. P. (1984) Suppression of cross-link formation in chloroethylnitrosourea-treated DNA by an activity in extracts of human leukemic lymphoblasts. *Cancer Res. 44*, 1887–1892.
(159) Erickson, L. C., Laurent, G., Sharkey, N. A., and Kohn, K. W. (1980) DNA cross-linking and monoadduct repair in nitrosoureatreated human tumor cells. *Nature 288*, 727–29.
(160) Noll, D. M., and Clarke, N. D. (2001) Covalent capture of a human O6-alkylguanine alkyltransferase-DNA complex using N1,O6-ethanoxanthosine, a mechanism-based crosslinker. *Nucleic Acids Res. 29*, 4025–4034.

(161) Lindahl, T., Sedgwick, B., Sekiguchi, M., and Nakabeppu, Y. (1988) Regulation and expression of the adaptive response to alkylating agents. *Ann. Rev. Biochem. 57*, 133–157.
(162) Moore, M. H., Gulbis, J. M., Dodson, E. J., Demple, B., and Moody, P. C. (1994) Crystal structure of a suicidal DNA repair protein: the Ada O6-methylguanine-DNA methyltransferase from E. coli. *EMBO J. 13*, 1495–1501.
(163) Wibley, J. E., Pegg, a E., and Moody, P. C. (2000) Crystal structure of the human O6-alkylguanine-DNA alkyltransferase. *Nucleic Acids Res. 28*, 393–401.
(164) Daniels, D. S., Mol, C. D., Arvai, A. S., Kanugula, S., Pegg, A. E., and Tainer, J. A. (2000) Active and alkylated human AGT structures: a novel zinc site, inhibitor and extrahelical base binding. *EMBO J. 19*, 1719–1730.
(165) Fang, Q., Kanugula, S., and Pegg, A. E. (2005) Function of domains of human O6-alkylguanine-DNA alkyltransferase. *Biochemistry 44*, 15396–15405.
(166) Rasimas, J. J., Kanugula, S., Dalessio, P. M., Ropson, I. J., Fried, M. G., and Pegg, A. E. (2003) Effects of zinc occupancy on human O6-alkylguanine-DNA alkyltransferase. *Biochemistry 42*, 980–990.
(167) Goodtzova, K., Kanugula, S., Edara, S., and Pegg, A. E. (1998) Investigation of the role of tyrosine-114 in the activity of human O6-alkylguanine-DNA alkyltranferase. *Biochemistry 37*, 12489–12495.
(168) Hu, J., Ma, A., and Dinner, A. R. (2008) A two-step nucleotide-flipping mechanism enables kinetic discrimination of DNA lesions by AGT. *Proc. Natl. Acad. Sci. USA 105*, 4615–4620.
(169) Rasimas, J. J., Pegg, A. E., and Fried, M. G. (2003) DNA-binding mechanism of O6-alkylguanine-DNA alkyltransferase: Effects of protein and DNA alkylation on complex stability. *J Biol. Chem. 278*, 7973–7980.
(170) Liu, L., Pegg, a E., Williams, K. M., and Guengerich, F. P. (2002) Paradoxical enhancement of the toxicity of 1,2-Dibromoethane by O6-alkylguanine-DNA alkyltransferase. *J. Biol. Chem. 277*, 37920–37928.
(171) Zang, H., Fang, Q., Pegg, A. E., and Guengerich, F. P. (2005) Kinetic analysis of steps in the repair of damaged DNA by human O6-alkylguanine-DNA alkyltransferase. *J. Biol. Chem. 280*, 30873–30881.
(172) Moschel, R. C., McDougall, M. G., Dolan, M. E., Stine, L., and Pegg, A. E. (1992) Structural features of substituted purine derivatives compatible with depletion of human O6-alkylguanine-DNA alkyltransferase. *J. Med. Chem. 35*, 4486–4491.
(173) Kindermann, M., George, N., Johnsson, N., and Johnsson, K. (2003) Covalent and selective immobilization of fusion proteins. *J. Am. Chem. Soc. 125*, 7810–7811.
(174) Keppler, A., Kindermann, M., Gendreizig, S., Pick, H., Vogel, H., and Johnsson, K. (2004) Labeling of fusion proteins of O6-alkylguanine-DNA alkyltransferase with small molecules in vivo and in vitro. *Methods 32*, 437–444.
(175) Kamiya, M., and Johnsson, K. (2010) Localizable and highly sensitive calcium indicator based on a BODIPY fluorophore. *Anal. Chem. 82*, 6472–6479.
(176) Cole, N. B., and Donaldson, J. G. (2012) Releasable SNAP-tag probes for studying endocytosis and recycling. *ACS Chem. Biol. 7*, 464–469.
(177) Prifti, E., Reymond, L., Umebayashi, M., Hovius, R., Riezman, H., and Johnsson, K. (2014) A fluorogenic probe for snap-tagged plasma membrane proteins based on the solvatochromic molecule nile red. *ACS Chem. Biol. 9*, 606–612.
(178) Keppler, A., Gendreizig, S., Gronemeyer, T., Pick, H., Vogel, H., and Johnsson, K. (2003)

A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nat. Biotechnol. 21*, 86–89.
(179) Yang, W. (2014) An overview of Y-Family DNA polymerases and a case study of human DNA polymerase η. *Biochemistry 53*, 2793–2803.
(180) Pringle, M. J., and Loeb, L. A. (2012) DNA polymerase delta in DNA replication and genome maintenance. *Environ. Mol. Mutagen. 53*, 666–682.
(181) McCulloch, S. D., and Kunkel, T. A. (2008) The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Res. 18*, 148–161.
(182) Garg, P., and Burgers, P. M. J. (2005) DNA polymerases that propagate the eukaryotic DNA replication fork. *Crit. Rev. Biochem. Mol. Biol. 40*, 115–128.
(183) Ohmori, H., Friedberg, E. C., Fuchs, R. P. P., Goodman, M. F., Hanaoka, F., Hinkle, D., Kunkel, T. a., Lawrence, C. W., Livneh, Z., Nohmi, T., Prakash, L., Prakash, S., Todo, T., Walker, G. C., Wang, Z., and Woodgate, R. (2001) The Y-family of DNA polymerases. *Mol. Cell 8*, 7–8.
(184) Yang, W., and Woodgate, R. (2007) What a difference a decade makes: insights into translesion DNA synthesis. *Proc. Natl. Acad. Sci. U. S. A. 104*, 15591–15598.
(185) Waters, L. S., Minesinger, B. K., Wiltrout, M. E., D'Souza, S., Woodruff, R. V, and Walker, G. C. (2009) Eukaryotic translesion polymerases and their roles and regulation in DNA damage tolerance. *Microbiol. Mol. Biol. Rev. 73*, 134–154.
(186) Prakash, S., Johnson, R. E., and Prakash, L. (2005) Eukaryotic translesion synthesis DNA polymerases: Specificity of structure and function. *Annu. Rev. Biochem. 74*, 317–353.
(187) Biertümpfel, C., Zhao, Y., Kondo, Y., Ramón-Maiques, S., Gregory, M., Lee, J. Y., Masutani, C., Lehmann, A. R., Hanaoka, F., and Yang, W. (2010) Structure and mechanism of human DNA polymerase η. *Nature 465*, 1044–1048.
(188) Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G., and Steitz, T. A. (1985) Structure of large fragment of Escherichia coli DNA polymerase I complexed with dTMP. *Nature 313*, 762–766.
(189) Rothwell, P. J., and Waksman, G. (2005) Structure and mechanism of DNA polymerases. *Adv. Protein Chem. 71*, 401–440.
(190) Garcia-Diaz, M., and Bebenek, K. (2007) Multiple functions of DNA polymerases. *CRC. Crit. Rev. Plant Sci. 26*, 105–122.
(191) Steitz, T. a. (1999) DNA polymerases: Structural diversity and common mechanisms. *J. Biol. Chem. 274*, 17395–17398.
(192) Johnson, S. J., and Beese, L. S. (2004) Structures of mismatch replication errors observed in a DNA polymerase. *Cell 116*, 803–816.
(193) Joyce, C. M., and Benkovic, S. J. (2004) DNA polymerase fidelity: Kinetics, structure, and checkpoints. *Biochemistry 43*, 14317–14324.
(194) Joyce, C. M., and Steitz, T. A. (1994) Function and structure relationships in DNA Polymerases. *Annu. Rev. Biochem. 63*, 777–822.
(195) Brody, R. S., and Frey, P. a. (1981) Unambiguous determination of the stereochemistry of nucleotidyl transfer catalyzed by DNA polymerase I from Escherichia coli. *Biochemistry 20*, 1245–1252.
(196) Yang, W., Lee, J. Y., and Nowotny, M. (2006) Making and breaking nucleic acids: Two-Mg2+-ion catalysis and substrate specificity. *Mol. Cell 22*, 5–13.
(197) Rothwell, P. J., Mitaksov, V., and Waksman, G. (2005) Motions of the fingers subdomain of klentaq1 are fast and not rate limiting: Implications for the molecular basis of fidelity in DNA

polymerases. *Mol. Cell 19*, 345–355.
(198) Nakamura, T., Zhao, Y., Yamagata, Y., Hua, Y., and Yang, W. (2012) Watching DNA polymerase η make a phosphodiester bond. *Nature 487*, 196–201.
(199) Guengerich, F. P. (2006) Interactions of carcinogen-bound DNA with individual DNA polymerases. *Chem. Rev. 106*, 420–452.
(200) Sale, J. E., Lehmann, A. R., and Woodgate, R. (2012) Y-family DNA polymerases and their role in tolerance of cellular DNA damage. *Nat. Rev. Mol. Cell Biol. 13*, 141–152.
(201) Zhang, Y., Wu, X., Rechkoblit, O., Geacintov, N. E., Taylor, J.-S., and Wang, Z. (2002) Response of human REV1 to different DNA damage: preferential dCMP insertion opposite the lesion. *Nucleic Acids Res. 30*, 1630–1638.
(202) Choi, J.-Y., Lim, S., Kim, E.-J., Jo, A., and Guengerich, F. P. (2010) Translesion synthesis across abasic lesions by human B-family and Y-family DNA polymerases α, δ, η, ι, κ, and REV1. *J. Mol. Biol. 404*, 34–44.
(203) Washington, M. T., Minko, I. G., Johnson, R. E., Haracska, L., Harris, T. M., Lloyd, R. S., Prakash, L., and Prakash, S. (2004) Efficient and error-free replication past a minor-groove N2 -guanine adduct by the sequential action of yeast Rev1 and DNA polymerase ζ. *Mol. Cell. Biol. 24*, 6900–6906.
(204) Nair, D. T., Johnson, R. E., Prakash, L., Prakash, S., and Aggarwal, A. K. (2005) Rev1 employs a novel mechanism of DNA synthesis using a protein template. *Science (80-.). 309*, 2219–2222.
(205) Swan, M. K., Johnson, R. E., Prakash, L., Prakash, S., and Aggarwal, A. K. (2009) Structure of the human Rev1-DNA-dNTP ternary complex. *J. Mol. Biol. 390*, 699–709.
(206) Nair, D. T., Johnson, R. E., Prakash, S., Prakash, L., and Aggarwal, A. K. (2004) Replication by human DNA polymerase-iota occurs by Hoogsteen base-pairing. *Nature 430*, 377–380.
(207) Masutani, C., Kusumoto, R., and Yamada, A. (1999) The XPV (xeroderma pigmentosum variant) gene encodes human DNA polymerase η. *Nature 399*, 700–704.
(208) Masutani, C., Araki, M., Yamada, A., Kusumoto, R., Nogimori, T., Maekawa, T., Iwai, S., and Hanaoka, F. (1999) Xeroderma pigmentosum variant (XP-V) correcting protein from HeLa cells has a thymine dimer bypass DNA polymerase activity. *EMBO J. 18*, 3491–3501.
(209) McCulloch, S. D., Kokoska, R. J., Masutani, C., Iwai, S., Hanaoka, F., and Kunkel, T. A. (2004) Preferential cis – syn thymine dimer bypass by DNA polymerase η occurs with biased fidelity. *Nature 428*, 97–100.
(210) Vassylyev, D. G., Kashiwagi, T., Mikami, Y., Ariyoshi, M., Iwai, S., Ohtsuka, E., and Morikawa, K. (1995) Atomic model of a pyrimidine dimer excision repair enzyme complexed with a DNA substrate: structural basis for damaged DNA recognition. *Cell 83*, 773–782.
(211) Min, J.-H., and Pavletich, N. P. (2007) Recognition of DNA damage by the Rad4 nucleotide excision repair protein. *Nature 449*, 570–575.
(212) Li, Y., Dutta, S., Doublié, S., Bdour, H. M., Taylor, J.-S., and Ellenberger, T. (2004) Nucleotide insertion opposite a cis-syn thymine dimer by a replicative DNA polymerase from bacteriophage T7. *Nat. Struct. Mol. Biol. 11*, 784–790.
(213) Woodside, A. M., and Guengerich, F. P. (2002) Effect of the O6 substituent on misincorporation kinetics catalyzed by DNA polymerases at O6-methylguanine and O6-benzylguanine. *Biochemistry 41*, 1027–1038.
(214) Thomson, N. M., Kenney, P. M., and Peterson, L. A. (2003) The pyridyloxobutyl DNA adduct, O6-[4-Oxo-4-(3-pyridyl)butyl]guanine, is detected in tissues from 4-

(Methylnitrosamino)- 1-(3-pyridyl)-1-butanone-treated A/J mice *16*, 1–6.
(215) Pauly, G. T., Peterson, L. A., and Moschel, R. C. (2002) Mutagenesis by O6-[4-oxo-4-(3-pyridyl)butyl]guanine in Escherichia coli and human cells. *Chem. Res. Toxicol. 15*, 165–169.
(216) Mijal, R. S., Loktionova, N. a, Vu, C. C., Pegg, A. E., and Peterson, L. a. (2005) O6-pyridyloxobutylguanine adducts contribute to the mutagenic properties of pyridyloxobutylating agents. *Chem. Res. Toxicol. 18*, 1619–1625.
(217) Andersen, N., Wang, P., and Wang, Y. (2013) Replication across regioisomeric ethylated thymidine lesions by purified DNA polymerases. *Chem. Res. Toxicol. 26*, 1730–1738.
(218) Takasawa, K., Masutani, C., Hanaoka, F., and Iwai, S. (2004) Chemical synthesis and translesion replication of a cis-syn cyclobutane thymine-uracil dimer. *Nucleic Acids Res. 32*, 1738–1745.
(219) Song, Q., Sherrer, S. M., Suo, Z., and Taylor, J. S. (2012) Preparation of site-specific T = mCG cis-syn cyclobutane dimer-containing template and its error-free bypass by yeast and human polymerase η. *J. Biol. Chem. 287*, 8021–8028.
(220) Lemaire, D. G. E., and Ruzsicska, B. (1993) Kinetic analysis of the deamination reactions of cyclobutane dimers of thymidylyl-3',5'-2'-deoxycytidine and 2'-deoxycytidylyl-3',5'-thymidine. *Biochemistry 32*, 2525–2533.
(221) Yamamoto, J., Oyama, T., Kunishi, T., Masutani, C., Hanaoka, F., and Iwai, S. (2014) A cyclobutane thymine-N4-methylcytosine dimer is resistant to hydrolysis but strongly blocks DNA synthesis. *Nucleic Acids Res. 42*, 2075–2084.
(222) Colis, L. C., Raychaudhury, P., and Basu, A. K. (2008) Mutational specificity of γ - radiation-induced guanine - thymine and thymine - guanine intrastrand cross-links in mammalian cells and translesion synthesis past the guanine - thymine lesion by human DNA polymerase η. *Biochemistry 47*, 8070–8079.
(223) Raychaudhury, P., and Basu, A. K. (2010) Replication past the γ-radiation-induced guanine-thymine cross-link G[8,5-Me]T by human and yeast DNA polymerase η. *J. Nucleic Acids 2010*, doi:10.4061/2010/101495.
(224) Vaisman, A., Masutani, C., Hanaoka, F., and Chaney, S. G. (2000) Efficient translesion replication past oxaliplatin and cisplatin GpG adducts by human DNA polymerase η. *Biochemistry 39*, 4575–4580.
(225) Masutani, C., Kusumoto, R., Iwai, S., and Hanaoka, F. (2000) Mechanisms of accurate translesion synthesis by human DNA polymerase eta. *EMBO J. 19*, 3100–3109.
(226) Bassett, E., King, N. M., Bryant, M. F., Hector, S., Pendyala, L., Chaney, S. G., and Cordeiro-Stone, M. (2004) The role of DNA polymerase η in translesion synthesis past platinum − DNA adducts in human fibroblasts. *Cancer Res. 64*, 6469–6475.
(227) Zhao, Y., Biertumpfel, C., Gregory, M. T., Hua, Y.-J., Hanaoka, F., and Yang, W. (2012) Structural basis of human DNA polymerase η-mediated chemoresistance to cisplatin. *Proc. Natl. Acad. Sci. USA 109*, 7269–7274.
(228) Abdu, K., Aiertza, M. K., Wilkinson, O. J., Grasby, J. a., Senthong, P., Povey, A. C., Margison, G. P., and Williams, D. M. (2012) Synthesis of oligodeoxyribonucleotides containing a conformationally-locked anti analogue of O6-methyl-2′-deoxyguanosine and their recognition by MGMT and Atl1. *Chem. Commun. 48*, 11214–11216.
(229) Patra, A., Nagy, L. D., Zhang, Q., Su, Y., Müller, L., Guengerich, F. P., and Egli, M. (2014) Kinetics, structure, and mechanism of 8-oxo-7,8-dihydro-2'-deoxyguanosine bypass by human DNA polymerase η. *J Biol. Chem. 289*, 16867–16882.
(230) Albertella, M. R., Green, C. M., Lehmann, A. R., and O'Connor, M. J. (2005) A role for

polymerase η in the cellular tolerance to cisplatin-induced damage. *Cancer Res. 65*, 9799–9806.
(231) Singer, B. (1976) All oxygens in nucleic acids react with carcinogenic ethylating agents. *Nature 264*, 333–339.
(232) Tannenbaum, S. R. (1983) N-nitroso compounds: A perspective on human exposure. *Lancet 1*, 6296–32.
(233) Bronstein, S. M., Skopek, T. R., and Swenberg, J. A. (2011) Efficient repair of O6-ethylguanine, but not O4 -ethylthymine or O2-ethylthymine, is dependent upon O6-alkylguanine-DNA alkyltransferase and nucleotide excision repair Activities in Human Cells. *Cancer Res. 52*, 2008–2011.
(234) Preston, B. D., Singert, B., and Loeb, L. A. (1986) Mutagenic potential of O4-methylthymine in vivo determined by an enzymatic approach to site-specific mutagenesis. *Proc. Natl. Acad. Sci. USA 83*, 8501–8505.
(235) Brennan, R. G., Pyżaksja, D., Blonski, W. J. P., Hruska, F. E., and Sundaralingam, M. (1986) Crystal structure of the promutagen O4-methylthymidine: Importance of the anti conformation of the O4 methoxy group and possible mispairing of O4-methylthymidine with guanine. *Biochemistry 25*, 1181–1185.
(236) Patra, A., Zhang, Q., Lei, L., Su, Y., Egli, M., and Guengerich, F. P. (2015) Structural and kinetic analysis of nucleoside triphosphate incorporation opposite an abasic site by human translesion DNA polymerase η. *J. Biol. Chem. 290*, 8028–8038.
(237) Madril, A. C., Johnson, R. E., Washington, M. T., Prakash, L., and Prakash, S. (2001) Fidelity and damage bypass ability of Schizosaccharomyces pombe Eso1 protein, comprised of DNA polymerase η and sister chromatid cohesion protein Ctf7. *J. Biol. Chem. 276*, 42857–42862.
(238) Swanson, A. L., Wang, J., and Wang, Y. (2011) In vitro replication studies of carboxymethylated DNA lesions with Saccharomyces cerevisiae polymerase η. *Biochemistry 50*, 7666–7673.
(239) Rao, S., Chenna, A., Slupska, M. I., and Singer, B. (1996) Replication of O4-methylthymine-containing oligonucleotides : Effect of 3 ' and 5 ' flanking bases on formation and extension of O4-methylthymine•guanine basepairs. *Mutat. Res. 356*, 179–185.
(240) Haracska, L., Yu, S. L., Johnson, R. E., Prakash, L., and Prakash, S. (2000) Efficient and accurate replication in the presence of 7,8-dihydro-8-oxoguanine by DNA polymerase η. *Nat. Genet. 25*, 458–461.
(241) Haracska, L., Prakash, S., and Prakash, L. (2000) Replication past O6-methylguanine by yeast and human DNA polymerase η. *Mol. Cell. Biol. 20*, 8001–8007.
(242) Xu, Y. Z., and Swann, P. F. (1990) A simple method for the solid phase synthesis of oligodeoxynucleotides containing O4-alkylthymine. *Nucleic Acids Res. 18*, 4061–4065.
(243) Fissekis, J. D., Myles, A., and Bosworth Brown, G. (1964) Synthesis of 5-hydroxyalkylpyrimidines from lactones. *J. Org. Chem. 29*, 2670–2673.
(244) Gazidova, T., Raić-Malić, S., Hegold-Brundić, A., and Cetina, M. (2008) Synthesis and structural characterization of the 5-(2-haloethyl)pyrimidines – hydrogen-bonded chains in α-(1-carbamyliminomethylene)-γ-butyrolactone. *Molecules 13*, 2786–2795.
(245) Denny, G. H., and Ryder, M. A. (1974) α-(ureidomethylene)lactones and derived 5-(hydroxyalkyl)uracils. *J. Med. Chem. 17*, 1230–1231.
(246) Leonard, N. J., and Cundall, R. L. (1974) Stereochemically controlled photoreactions between two thymine rings. *J. Am. Chem. Soc. 96*, 5904–5910.
(247) Fissekis, J. D., and Sweet, F. (1973) The chemistry of some 5-(2-hydroxyalkyl)uracil

derivatives and a synthesis of 5-vinyluracil. *J. Org. Chem. 28*, 264–269.
(248) Eger, K., Mohammad, J., and Schmidt, M. (1995) Steric fixation bromovinyluracil: synthesis of furo[2,3-d]pyrimidine nucleosides. *J. Heterocycl. Chem. 31*, 211–218.
(249) Zhu, Q., Delaney, M. O., and Greenberg, M. M. (2001) Observation and elimination of N-acetylation of oligonucleotides prepared using fast-deprotecting phosphoramidites and ultra-mild deprotection. *Bioorg. Med. Chem. Lett 11*, 1105–1107.
(250) Fabrega, C., Eritja, R., Sinha, N. D., Dosanjh, M. K., and Singer, B. (1995) Synthesis and properties of oligonucleotides containing the mutagenic base O4-benzylthymidine. *Bioorg. Med. Chem. 3*, 101–108.
(251) Xu, Y.-Z., and Swann, P. F. (1994) Oligodeoxynucleotides containing O2-alkylthymine: synthesis and characterization. *Tetrahedron Lett. 35*, 303–306.
(252) Cruzeiro-Hansson, L., and Goodfellow, J. M. (1994) Flexibility and curvature of DNA duplexes containing O4-methylthymine: implications for DNA repair. *Carcinogenesis 15*, 1525–1533.
(253) Li, B. F. L., Reese, C. B., and Swann, P. F. (1987) Synthesis and characterization or oligodeoxnucleotides containing 4-O-methylthymine. *Biochemistry 26*, 1086–1093.
(254) Puglisi, J. D., and Tinoco, I. J. (1989) Absorbance melting curves of RNA. *Methods Enzym. 180*, 304–325.
(255) Lowe, L. G., and Guengerich, F. P. (1996) Steady-state and pre-steady-state kinetic analysis of dNTP insertion opposite 8-oxo-7,8-dihydroguanine by Escherichia coli polymerases I exo- and II exo-. *Biochemistry 35*, 9840–9849.
(256) Chowdhury, G., and Guengerich, F. P. (2011) Liquid chromatography-mass spectrometry analysis of DNA polymerase reaction products. *Curr. Protoc. Nucleic Acid Chem. 7*, 7.16.1–7.16.11.
(257) Christov, P. P., Angel, K. C., Guengerich, F. P., and Rizzo, C. J. (2009) Replication past the N5-methyl-formamidopyrimidine lesion of deoxyguanosine by DNA polymerases and an improved procedure for sequence analysis of in vitro bypass products by mass spectrometry. *Chem. Res. Toxicol. 22*, 1086–1095.
(258) Zang, H., Goodenough, A. K., Choi, J.-Y., Irimia, A., Loukachevitch, L. V, Kozekov, I. D., Angel, K. C., Rizzo, C. J., Egli, M., and Guengerich, F. P. (2005) DNA adduct bypass polymerization by sulfolobus solfataricus DNA polymerase Dpo4: Analysis and crystal structures of multiple base pair substitution and frameshift products with the adduct 1,N2-ethenoguanine. *J Biol Chem. 280*, 29750–29764.
(259) Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*
(260) (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr. 50*, 760–763.
(261) McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr. 40*, 658–674.
(262) Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr. 66*, 213–221.
(263) Murshudov, G. N., Skubák, P., Lebedev, A. a., Pannu, N. S., Steiner, R. a., Nicholls, R. a., Winn, M. D., Long, F., and Vagin, A. a. (2011) REFMAC5 for the refinement of

macromolecular crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr. 67*, 355–367.
(264) Emsley, P., and Cowtan, K. (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr. 60*, 2126–2132.
(265) Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem. 25*, 1605–1612.
(266) Houtgraaf, J. H., Versmissen, J., and van der Giessen, W. J. (2006) A concise review of DNA damage checkpoints and repair in mammalian cells. *Cardiovasc. revascularization Med. 7*, 165–172.
(267) Fu, D., Calvo, J. a, and Samson, L. D. (2012) Balancing repair and tolerance of DNA damage caused by alkylating agents. *Nat. Rev. Cancer 12*, 104–120.
(268) Eastman, A. (1986) Reevaluation of interaction of cis-dichloro(ethylenediamine)platinum(II) with DNA. *Biochemistry 25*, 3912–3915.
(269) Kowalczyk, A., Carmical, J. R., Zou, Y., Van Houten, B., Lloyd, R. S., Harris, C. M., and Harris, T. M. (2002) Intrastrand DNA cross-links as tools for studying DNA replication and repair: two-, three-, and four-carbon tethers between the N(2) positions of adjacent guanines. *Biochemistry 41*, 3109–3118.
(270) C., M. M., Powell, K. L., and Tran, N. (1995) Binding of the transcription factor, Sp1, to non-target sites in DNA modified by benzo[a]pyrene diol epoxide. *Carcinogenesis 16*, 975–983.
(271) Novakova, O., Kasparkova, J., Bursova, V., Hofr, C., Vojtiskova, M., Chen, H., Sadler, P. J., and Brabec, V. (2005) Conformation of DNA modified by monofunctional Ru(II) arene complexes: recognition by DNA binding proteins and repair. Relationship to cytotoxicity. *Chem. Biol. 12*, 121–129.
(272) Treiber, D. K., Zhai, X., Jantzen, H. M., and Essigmann, J. M. (1994) Cisplatin-DNA adducts are molecular decoys for the ribosomal RNA transcription factor hUBF (human upstream binding factor). *Proc. Natl. Acad. Sci. 91*, 5672–5676.
(273) Bellon, S., Gasparutto, D., Saint-Pierre, C., and Cadet, J. (2006) Guanine-thymine intrastrand cross-linked lesion containing oligonucleotides: from chemical synthesis to in vitro enzymatic replication. *Org. Biomol. Chem. 4*, 3831–7.
(274) Brash, D. E., Rudolph, J. a, Simon, J. a, Lin, A., McKenna, G. J., Baden, H. P., Halperin, a J., and Pontén, J. (1991) A role for sunlight in skin cancer: UV-induced p53 mutations in squamous cell carcinoma. *Proc. Natl. Acad. Sci. U. S. A. 88*, 10124–10128.
(275) Ziegler, a, Leffell, D. J., Kunala, S., Sharma, H. W., Gailani, M., Simon, J. a, Halperin, a J., Baden, H. P., Shapiro, P. E., and Bale, a E. (1993) Mutation hotspots due to sunlight in the p53 gene of nonmelanoma skin cancers. *Proc Natl Acad Sci 90*, 4216–4220.
(276) Dumaz, N., Drougard, C., Sarasin, A., and Daya-Grosjean, L. (1993) Specific UV-induced mutation spectrum in the p53 gene of skin tumors from DNA-repair-deficient xeroderma pigmentosum patients. *Proc. Natl. Acad. Sci. U. S. A. 90*, 10529–10533.
(277) Cleaver, J. E., and Crowley, E. (2002) UV damage, DNA repair and skin carcinogenesis. *Front. Biosci. 7*, 1024–1043.
(278) Yoon, J. H., Lee, C. S., O'Connor, T. R., Yasui, A., and Pfeifer, G. P. (2000) The DNA damage spectrum produced by simulated sunlight. *J. Mol. Biol. 299*, 681–693.
(279) Clingen, P. H., Arlett, C. F., Roza, L., Mori, T., Nikaido, O., and Green, M. H. (1995) Induction of cyclobutane pyrimidine dimers, pyrimidine(6-4)pyrimidone photoproducts, and Dewar valence isomers by natural sunlight in normal human mononuclear cells. *Cancer Res. 55*, 2245–2248.

(280) Rastogi, R. P., Richa, Kumar, A., Tyagi, M. B., and Sinha, R. P. (2010) Molecular mechanisms of ultraviolet radiation-induced DNA damage and repair. *J. Nucleic Acids 2010*, 592980.
(281) Eastman, A. (1987) The formation, isolation and characterization of DNA adducts produced by anticancer platinum complexes. *Pharmacol. Ther. 34*, 155–166.
(282) Zwelling, L. a, Anderson, T., and Kohn, K. W. (1979) DNA-protein and DNA interstrand cross-linking by cis- and trans- platinum (II) diamminedichloride in L1210 mouse leukemia cells and relation to cytotoxicity Platinum (ll) Diamminedichloride In L1210 Mouse Leukemia Cells and Relation to Cytotoxicity. *Cancer Res. 39*, 365–369.
(283) Olinski, R., Wedrychowski, A., Schmidt, W. N., Briggs, R. C., and Hnilica, L. S. (1987) In vivo DNA-protein cross-linking by cis- and trans-diamminedichloroplatinum(II). *Cancer Res. 47*, 201–205.
(284) Bizanek, R., McGuinness, B. F., Nakanishi, K., and Tomasz, M. (1992) Isolation and structure of an intrastrand cross-link adduct of mitomycin C and DNA. *Biochemistry 31*, 3084–3091.
(285) Rink, S. M., and Hopkins, P. B. (1995) A mechlorethamine-induced DNA interstrand cross-link bends duplex DNA. *Biochemistry 34*, 1439–1445.
(286) Balcome, S., Park, S., Quirk Dorr, D. R., Hafner, L., Phillips, L., and Tretyakova, N. (2004) Adenine-containing DNA-DNA cross-links of antitumor nitrogen mustards. *Chem. Res. Toxicol. 17*, 950–962.
(287) Gillet, L. C. J., and Schärer, O. D. (2006) Molecular mechanisms of mammalian global genome nucleotide excision repair. *Chem. Rev. 106*, 253–276.
(288) Xie, K., Doles, J., Hemann, M. T., and Walker, G. C. (2010) Error-prone translesion synthesis mediates acquired chemoresistance. *Proc. Natl. Acad. Sci. U. S. A. 107*, 20792–20797.
(289) Yang, Z., Colis, L. C., Basu, A. K., and Zou, Y. (2005) Recognition and incision of γ-radiation-induced cross-linked guanine−thymine tandem lesion G[8,5-Me]T by UvrABC nuclease. *Chem. Res. Toxicol. 18*, 1339–1346.
(290) Szymkowski, D. E., Yarema, K., Essigmann, J. M., Lippard, S. J., and Wood, R. D. (1992) An intrastrand d(GpG) platinum crosslink in duplex M13 DNA is refractory to repair by human cell extracts. *Proc. Natl. Acad. Sci. U. S. A. 89*, 10772–10776.
(291) Wood, R. (1994) Studying nucleotide excision repair of mammalian DNA in a cell-free system. *Ann. N. Y. Acad. Sci. 726*, 274–79.
(292) Lehmann, A. R. (2002) Replication of damaged DNA in mammalian cells: new solutions to an old problem. *Mutat. Res. 509*, 23–34.
(293) Chen, Y., Cleaver, J. E., Hanaoka, F., Chang, C., and Chou, K. (2006) A novel role of DNA polymerase eta in modulating cellular sensitivity to chemotherapeutic agents. *Mol. Cancer Res. 4*, 257–265.
(294) Ho, T. V., and Schärer, O. D. (2010) Translesion DNA synthesis polymerases in DNA interstrand crosslink repair. *Environ. Mol. Mutagen. 566*, 552–566.
(295) Li, H., Qiu, Y., Moyroud, E., and Kishi, Y. (2001) Synthesis of DNA oligomers possessing a covalently cross-linked watson-crick base pair model. *Angew. Chem. Int. Ed. 40*, 1471–1475.
(296) Noronha, A. M., Noll, D. M., Wilds, C. J., and Miller, P. S. (2002) N4C-ethyl-N4C cross-linked DNA: synthesis and characterization of duplexes with interstrand cross-links of different orientations. *Biochemistry 41*, 760–771.
(297) Murata, S., Mizumura, Y., Hino, K., Ueno, Y., Ichikawa, S., and Matsuda, A. (2007) Modular bent DNAs: A new class of artificial DNAs with a protein binding ability. *J. Am. Chem.*

Soc. 129, 10300–10301.
(298) Wilds, C. J., Xu, F., and Noronha, A. M. (2008) Synthesis and characterization of DNA containing an N1-2'-deoxyinosine-ethyl-N3-thymidine interstrand cross-link: A structural mimic of the cross-link formed by 1,3-bis-(2-chloroethyl)-1-nitrosourea. *Chem. Res. Toxicol. 21*, 686–695.
(299) Sun, G., Noronha, A., and Wilds, C. (2012) Preparation of N3-thymidine–butylene–N3-thymidine interstrand cross-linked DNA via an orthogonal deprotection strategy. *Tetrahedron 68*, 7787–7793.
(300) Haines, J. A., Reese, C. B., and Todd, L. (1961) The methylation of guanosine and related compounds with diazomethane. *J. Chem. Soc.* 5281–5288.
(301) Townsend, L. B., and Robins, R. K. (1963) Ring cleavage of purine nucleosides to yield possible biogenetic precursors of pteridines and riboflavin. *J. Am. Chem. Soc. 85*, 242–243.
(302) Bodell, W. J., and Pongracz, K. (1993) Chemical synthesis and detection of the cross-link 1-[N3-(2'-deoxycytidyl)]-2-[N1-(2'-deoxyguanosinyl)]ethane in DNA reacted with 1-(2-chloroethyl)-1-nitrosourea. *Chem. Res. Toxicol. 6*, 434–438.
(303) Iwakuma, T., Sakumi, K., Nakatsuru, Y., Kawate, H., Igarashi, H., Shiraishi, A., Tsuzuki, T., Ishikawa, T., and Sekiguchi, M. (1997) High incidence of nitrosamine-induced tumorigenesis in mice lacking DNA repair methyltransferase. *Carcinogenesis 51*, 1631–1635.
(304) Bugni, J. M., Meira, L. B., and Samson, L. D. (2009) Alkylation-induced colon tumorigenesis in mice deficient in the Mgmt and Msh6 proteins. *Oncogene 28*, 734–741.
(305) Sakumi, K., Shiraishi, A., Shimizu, S., Tsuzuki, T., Ishikawa, T., and Sekiguchi, M. (1997) Methylnitrosourea-induced tumorigenesis in MGMT gene knockout mice. *Cancer Res. 57*, 2415–2418.
(306) Pegg, A. E. (2000) Repair of O6-alkylguanine by alkyltransferases. *Mutat. Res. 462*, 83–100.
(307) Tintoré, M., Gállego, I., Manning, B., Eritja, R., and Fàbrega, C. (2013) DNA origami as a DNA repair nanosensor at the single-molecule level. *Angew. Chemie - Int. Ed. 52*, 7747–7750.
(308) Hellman, L. M., Spear, T. J., Koontz, C. J., Melikishvili, M., and Fried, M. G. (2014) Repair of O6-methylguanine adducts in human telomeric G-quadruplex DNA by O6-alkylguanine-DNA alkyltransferase. *Nucleic Acids Res. 42*, 9781–9791.
(309) Borowy-Borowski, H., and Chambers, R. W. (1987) A study of side reactions occurring during synthesis of oligodeoxynucleotides containing O6-alkyldeoxyguanosine residues at preselected sites. *Biochemistry 26*, 2465–2471.
(310) Spratt, T. E., and Campbell, C. R. (1994) Synthesis of oligodeoxynucleotides containing analogs of O6-methylguanine and reaction with O6-alkylguanine-DNA alkyltransferases. *Biochemistry 33*, 11364–11371.
(311) Wilds, C. J., Booth, J. D. M., and Noronha, A. M. (2011) Synthesis of building blocks and oligonucleotides with {G}O6-alkyl-O6{G} cross-links. *Curr. Protoc. Nucleic Acid Chem 44*, 5.9.1–5.9.19.
(312) Wilds, C. J., Booth, J. D., and Noronha, A. M. (2006) Synthesis of oligonucleotides containing an O6-G-alkyl-O6-G interstrand cross-link. *Tetrahedron Lett. 47*, 9125–9128.
(313) Gaffney, B. L., Marky, L. A., and Jones, R. A. (1984) Synthesis and characterization of a set of four dodecadeoxyribonucleoside undecaphosphates containing O6-methylguanine opposite adenine, cytosine, guanine, and thymine. *Biochemistry 23*, 5686–5691.
(314) Gaffney, B. L., and Jones, R. A. (1989) Thermodynamic comparison of the base pairs formed by the carcinogenic lesion O6-methylguanine with reference both to watson-crick pairs

and to mismatched pairs. *Biochemistry 28*, 5881–5889.
(315) Kuzmich, S., Marky, L. A., and Jones, R. A. (1983) Specifically alkylated DNA fragments. Synthesis and physical characterization of d[CGC(O6Me)GCG] and d[CGT(O6Me) GCG]. *Nucleic Acids Res. 11*, 3393–3404.
(316) Coulter, R., Blandino, M., Tomlinson, J. M., Pauly, G. T., Krajewska, M., Moschel, R. C., Peterson, L. a, Pegg, A. E., and Spratt, T. E. (2007) Differences in the rate of repair of O6-alkylguanines in different sequence contexts by O6-alkylguanine-DNA alkyltransferase. *Chem. Res. Toxicol. 20*, 1966–1971.
(317) Noll, D. M., Noronha, a M., and Miller, P. S. (2001) Synthesis and characterization of DNA duplexes containing an N4C-ethyl-N4C interstrand cross-link. *J. Am. Chem. Soc. 123*, 3405–3411.
(318) Marini, V., Christofis, P., Novakova, O., Kasparkova, J., Farrell, N., and Brabec, V. (2005) Conformation, protein recognition and repair of DNA interstrand and intrastrand cross-links of antitumor trans-[PtCl2(NH3)(thiazole)]. *Nucleic Acids Res. 33*, 5819–5828.
(319) Jung, Y., and Lippard, S. J. (2007) Direct cellular responses to platinum-induced DNA damage direct cellular responses to platinum-induced DNA damage. *Chem. Rev. 107*, 1387–1407.
(320) Klapacz, J., Meira, L. B., Luchetti, D. G., Calvo, J. a, Bronson, R. T., Edelmann, W., and Samson, L. D. (2009) O6-methylguanine-induced cell death involves exonuclease 1 as well as DNA mismatch recognition in vivo. *Proc. Natl. Acad. Sci. U. S. A. 106*, 576–581.
(321) McManus, F. P., and Wilds, C. J. (2013) Engineering of a O6-alkylguanine-DNA alkyltransferase chimera and repair of O4-alkyl thymidine adducts and O6-alkylene-2′-deoxyguanosine cross-linked DNA. *Toxicol. Res. (Camb). 2*, 158–162.
(322) Xu-welliver, M., and Pegg, A. E. (2002) Degradation of the alkylated form of the DNA repair protein, O6-alkylguanine-DNA alkyltransferase. *Carcinogenesis 23*, 823–830.
(323) O'Flaherty, D. K., and Wilds, C. J. (2015) Synthesis, characterization, and repair of a flexible O6-2'-deoxyguanosine-alkylene-O6 -2'-deoxyguanosine intrastrand cross-link. *Chem. Eur. J. 21*, 10522–10529.
(324) Streeper, R. T., Cotter, R. J., Colvin, M. E., Hilton, J., and Colvin, O. M. (1995) Molecular pharmacology of hepsulfam , NSC 3296801 : Identification of alkylated nucleoside, alkylation site, and site of DNA cross-linking. *Cancer Res. 55*, 1491–1498.
(325) Kim, S. J., Lester, C., and Begley, T. P. (1995) Synthesis of the dinucleotide spore photoproduct. *J. Org. Chem. 60*, 6256–6257.
(326) Jian, Y., and Li, L. (2013) Chemical syntheses of oligodeoxyribonucleotides containing spore photoproduct. *J Org Chem. 78*, 3021–3029.
(327) Bowman, K. J., Pla, R. L., Guichard, Y., Farmer, P. B., and Jones, G. D. (2001) Evaluation of phosphodiesterase I-based protocols for the detection of multiply damaged sites in DNA: the detection of abasic, oxidative and alkylative tandem damage in DNA oligonucleotides. *Nucleic Acids Res. 29*, E101.
(328) Park, S., Seetharaman, M., Ogdie, A., Ferguson, D., and Tretyakova, N. (2003) 3'-Exonuclease resistance of DNA oligodeoxynucleotides containing O6-[4-oxo-4-(3-pyridyl)butyl]guanine. *Nucleic Acids Res. 31*, 1984–1994.
(329) Verdemato, P. E., Brannigan, J. A., Damblon, C., Zuccotto, F., Moody, P. C. E., and Lian, L. (2000) DNA-binding mechanism of the Escherichia coli Ada O6-alkylguanine – DNA alkyltransferase. *Nucleic Acids Res. 28*, 3710–3718.
(330) Butenandt, J., Eker, A. P. M., and Carell, T. (1998) Synthesis, crystal structure, and

enzymatic evaluation of a DNA-photolesion isostere. *Chem. Eur. J. 4*, 642–654.
(331) Mcmanus, F. P., Fang, Q., Booth, J. D. M., Noronha, A. M., Anthony, E., and Wilds, C. J. (2010) DON'T Supplemental-Synthesis and Characterization of Oligonucleotides Containing an O 6 -2 ' - GNC Motif and Repair by Human O 6 -Alkylguanine-DNA Alkyltransferase. *Org. Biomol. Chem. 8*, 4414–26.
(332) Eastman, A. (1983) Characterization of the adducts produced in DNA by cis-diamminedichloroplatinum (II) and cis-dichloro(ethylenediamine)platinum(II). *Biochemistry 22*, 3927–3933.
(333) Eastman, A. (1985) Interstrand cross-links and sequence specificity in the reaction of cis-dichloro(ethylenediamine)platinum(II) with DNA. *Biochemistry 24*, 5027–5032.
(334) Fichtinger-Schepman, A. M. J., Oosterom, A. T. Van, Lohman, P. H. M., and Berends, F. (1987) cis-Diamminedichloroplatinum (II)-induced DNA adducts in peripheral leukocytes from seven cancer patients: Quantitative immunochemical detection of the adduct induction and removal after a single dose of cis-diamminedichloroplatinum (II). *Cancer Res. 47*, 3000–3004.
(335) Fichtinger-Schepman, A. M. J., Veer, J. L. Van Der, Hartog, J. H. J. Den, and Lohman, P. H. M. (1985) Adducts of the antitumor drug cis-diamminedichloroplatinum (II) with DNA: Formation, identification, and quantitation. *Biochemistry 24*, 707–713.
(336) Fichtinger-Schepman, A. M. J., Velde-Visser, S. D. Van Der, Dijk-Knijnenburg, H. C. M. Van, Oosterom, A. T. Van, Baan, R. A., and Berends, F. (1990) Kinetics of the formation and removal of cisplatin-DNA adducts in blood cells and tumor tissue of cancer patients receiving chemotherapy: Comparison with in vitro adduct formation. *Cancer Res. 50*, 7887–7894.
(337) Plooy, A. C. M., Fichtinger-Schepman, A. M. J., Schutte, H. H., van Dijk, M., and Lohman, P. H. M. (1985) The quantitative detection of various Pt-DNA-adducts in Chinese hamster ovary cells treated with cisplatin: application of immunochemical techniques. *Carcinogenesis 6*, 561–566.
(338) Zou, Y., Van Houten, B., and Farrell, N. (2006) Sequence specificity of DNA-DNA interstrand cross-link formation by cisplatin and dinuclear platinum complexes. *Biochemistry 33*, 5404–5410.
(339) Lin, Q., Clark, A. B., McCulloch, S. D., Yuan, T., Bronson, R. T., Kunkel, T. A., and Kucherlapati, R. (2006) Increased susceptibility to UV-induced skin carcinogenesis in polymerase η-deficient mice. *Cancer Res. 66*, 87–94.
(340) Yamada, K., Takezawa, J., and Ezaki, O. (2003) Translesion replication in cisplatin-treated xeroderma pigmentosum variant cells is also caffeine-sensitive: Features of the error-prone DNA polymerase(s) involved in UV-mutagenesis. *DNA Repair (Amst). 2*, 909–924.
(341) Jiang, Y., Hong, H., Cao, H., and Wang, Y. (2007) In vivo formation and in vitro replication of a guanine-thymine intrastrand cross-link lesion. *Biochemistry 46*, 12757–12763.
(342) Pauly, G. T., Hughes, S. H., and Moschel, R. C. (1998) Comparison of mutagenesis by O6-methyl- and O6-ethylguanine and O4-methylthymine in Escherichia coli using double-stranded and gapped plasmids. *Carcinogenesis 19*, 457–461.
(343) Pauly, G. T., and Moschel, R. C. (2001) Mutagenesis by O6-Methyl-, O6-Ethyl-, and O6-Benzylguanine and O4-Methylthymine in Human Cells: Effects of O6-Alkylguanine-DNA Alkyltransferase and Mismatch Repair. *Chem. Res. Toxicol. 14*, 894–900.
(344) Pauly, G. T., Hughes, S. H., and Moschel, R. C. (1995) Mutagenesis in Escherichia coli by three O6-substituted guanines in double-stranded or gapped plasmids. *Biochemistry 34*, 8924–8930.
(345) Pauly, G. T., Hughes, S. H., and Moschel, R. C. (1994) Response of repair-competent and

repair-deficient Escherichia coli to three O6-substituted guanines and involvement of methyl-directed mismatch repair in the processing of O6-methylguanine residues. *Biochemistry 33*, 9169–9177.
(346) Bishop, R. E., Pauly, G. T., and Moschel, R. C. (1996) O6-Ethylguanine and O6-benzylguanine incorporated site-specifically in codon 12 of the rat H-ras gene induce semi-targeted as well as targeted mutations in Rat4 cells *17*, 849–856.
(347) Boosalis, M. S., Petruska, J., and Goodman, M. F. (1987) DNA polymerase insertion fidelity: Gel assay for site-specific kinetics. *J Biol. Chem. 262*, 14689–14696.
(348) O'Flaherty, D. K., and Guengerich, F. P. (2014) Steady-state kinetic analysis of DNA polymerase single-nucleotide incorporation products. *Curr. Protoc. Nucleic Acid Chem. 59*, 7.21.1–7.21.13.
(349) Choi, J.-Y., and Guengerich, F. P. (2006) Kinetic evidence for inefficient and error-prone bypass cross bulky N2-guanine DNA adducts by human DNA polymerase ι. *J. Biol. Chem. 281*, 12315–12324.
(350) Vaisman, A., and Chaney, S. G. (2000) The efficiency and fidelity of translesion synthesis past cisplatin and oxaliplatin GpG adducts by human DNA polymerase β. *J Biol. Chem. 275*, 13017–13025.
(351) Boiteux, S., and Laval, J. (1982) Coding properties of poly(deoxycytidylic acid) templates containing uracil or apyrimidinic sites: In vitro modulation of mutagenesis by deoxyribonucleic acid repair enzyme. *Biochemistry 21*, 6746–6751.
(352) Sagher, D., and Strauss, B. (1983) Insertion of nucleotides opposite apurinic / apyrimidinic sites in deoxyribonucleic acid during in vitro synthesis: Uniqueness of adenine nucleotides. *Biochemistry 22*, 4518–4526.
(353) Schaaper, R. M., Glickman, B. W., and Loeb, L. A. (1982) Mutagenesis resulting from depurination is an SOS process. *Mut. Res. 106*, 1–9.
(354) Pfeifer, G. P., and Besaratinia, A. (2012) UV wavelength-dependent DNA damage and human non-melanoma and melanoma skin cancer. *Photochem. Photobiol. Sci. 11*, 90–7.
(355) Singer, B., and Essigmann, J. M. (1991) Site-specific mutagenesis: Retrospective and prospective. *Carcinogenesis 12*, 949–955.
(356) Samson, L., Han, S., and Marquis, J. C. (1997) Mammalian DNA repair methyltransferases shield O4MeT from nucleotide excision repair. *Carcinogenesis 18*, 919–924.
(357) Brent, T. P., and Remack, J. S. (1988) Formation of covalent complexes between human O6-alkylguanine-DNA alkyltransferase and BCNU-treated defined length synthetic oligodeoxynucleotides. *Nucleic Acids Res. 16*, 6779–6788.
(358) Gonzaga, E., Margison, P., Potter, P. M., Brent, T. P., Yu, D., Ludlum, D. B., and Rafferty, A. (1992) Identification of the cross-link between human O6-methylguanine-DNA methyltransferase and chloroethylnitrosourea-treated DNA. *Cancer Res. 52*, 6052–6058.
(359) Darkes, M. J. M., Plosker, G. L., and Jarvis, B. (2002) Temozolomide. *Am J. Cancer 1*, 55–80.
(360) Loeber, R., Rajesh, M., Fang, Q., Pegg, A. E., and Tretyakova, N. (2006) Cross-linking of the human DNA repair protein O6-alkylguanine DNA alkyltransferase to DNA in the presence of 1,2,3,4-diepoxybutane. *Chem. Res. Toxicol. 19*, 645–654.
(361) Liu, H., Xu-Welliver, M., and Pegg, a E. (2000) The role of human O(6)-alkylguanine-DNA alkyltransferase in promoting 1,2-dibromoethane-induced genotoxicity in Escherichia coli. *Mutat. Res. 452*, 1–10.
(362) Abril, N., LuqueRomero, F. L., PrietoAlamo, M. J., Rafferty, J. a, Margison, G. P., and

Pueyo, C. (1997) Bacterial and mammalian DNA alkyltransferases sensitize Escherichia coli to the lethal and mutagenic effects of dibromoalkanes. *Carcinogenesis 18*, 1883–1888.
(363) Gentil, A., Page, F. Le, Margot, A., Lawrence, C. W., Borden, A., and Sarasin, A. (1996) Mutagenicity of a unique thymine – thymine dimer or thymine – thymine pyrimidine pyrimidone (6-4) photoproduct in mammalian cells. *Nucleic Acids Res. 24*, 1837–1840.
(364) Rojsitthisak, P., Jongaroonngamsang, N., Romero, R. M., and Haworth, I. S. (2011) HPLC-UV, MALDI-TOF-MS and ESI-MS/MS analysis of the mechlorethamine DNA crosslink at a cytosine-cytosine mismatch pair. *PLoS One 6*, e20745.
(365) Samson, L., Thomale, J., and Rajewsky, M. F. (1988) Alternative pathways for the in vivo repair of O6-alkylguanine and O4-alkylthymine in Escherichia coli: the adaptive response and nucleotide excision repair. *EMBO J. 7*, 2261–2267.
(366) Wiechelman, K., and Taylor, E. R. (1998) Anti-syn conformational range of pyrimidines with deoxyribofuranose. *J. Biomol. Struct. Dyn. 15*, 1181–1194.
(367) Pegg, A. E., Morimoto, K., and Dolan, M. E. (1988) Investigation of the specificity of O6-alkylguanine-DNA-alkyltransferase. *Chem. Biol. Interact. 65*, 275–281.
(368) Mitsunobu, O. (1981) The use of diethyl azodicarboxylate and triphenylphosphine in synthesis and transformation of natural products. *Synthesis (Stuttg).* 1–28.
(369) Zannis-Hadjopoulos, M., and Rampakakis, E. (2011) Synergy between DNA replication and repair mechanisms, in *DNA repair - On the pathways to fixing DNA damage and errors* (Storici, F., Ed.), pp 25–42. InTech.
(370) Ghosal, G., and Chen, J. (2013) DNA damage tolerance: a double-edged sword guarding the genome. *Transl. Cancer Res. 2*, 107–129.
(371) O'Flaherty, D. K., Mcmanus, F. P., Noronha, A. M., and Wilds, C. J. (2013) Synthesis of building blocks and oligonucleotides containing {T}O4-alkylene-O4{T} interstrand cross-links. *Curr. Protoc. Nucleic Acid Chem. 55*, 5.13.1–5.13.19.
(372) Liu, L., Hachey, D. L., Valadez, G., Williams, K. M., Guengerich, F. P., Loktionova, N. A., Kanugula, S., and Pegg, A. E. (2003) Characterization of a mutagenic DNA adduct formed from 1,2-dibromoethane by O6-alkylguanine-DNA alkyltransferase. *J. Biol. Chem. 279*, 4250–4259.
(373) Guengerich, F. P. (2005) Principles of covalent binding of reactive metabolites and examples of activation of bis-electrophiles by conjugation. *Arch. Biochem. Biophys. 433*, 369–378.
(374) Wilds, C. J., Noronha, A. M., Robidoux, S., and Miller, P. S. (2004) Mispair-aligned N3T-alkyl-N3T interstrand cross-linked DNA: synthesis and characterization of duplexes with interstrand cross-links of variable lengths. *J. Am. Chem. Soc. 126*, 9257–9265.
(375) Sun, G., Noronha, A. M., Miller, P. S., and Wilds, C. J. (2012) Synthesis of building blocks and oligonucleotides with {T}N3-alkylene-N3{T} cross-links. *Curr. Protoc. nucleic acid Chem. 51*, 5.11.1–5.11–17.
(376) O'Flaherty, D. K., Denisov, A. Y., Noronha, A. M., and Wilds, C. J. (2014) NMR structure of an ethylene interstrand cross-linked DNA which mimics the lesion formed by 1,3-bis(2-chloroethyl)-1-nitrosourea. *ChemMedChem 9*, 2099–103.

Appendix I

First Page to Published Manuscripts

Organic & Biomolecular Chemistry

Dynamic Article Links

Cite this: *Org. Biomol. Chem.*, 2012, **10**, 7078

www.rsc.org/obc

PAPER

O^4-Alkyl-2′-deoxythymidine cross-linked DNA to probe recognition and repair by O^6-alkylguanine DNA alkyltransferases†‡

Francis P. McManus, Derek K. O'Flaherty, Anne M. Noronha and Christopher J. Wilds*

Received 10th April 2012, Accepted 3rd July 2012
DOI: 10.1039/c2ob25705j

DNA duplexes containing a directly opposed O^4-2′-deoxythymidine-alkyl-O^4-2′-deoxythymidine (O^4-dT-alkyl-O^4-dT) interstrand cross-link (ICL) have been prepared by the synthesis of cross-linked nucleoside dimers which were converted to phosphoramidites to produce site specific ICL. ICL duplexes containing alkyl chains of four and seven methylene groups were prepared and characterized by mass spectrometry and nuclease digests. Thermal denaturation experiments revealed four and seven methylene containing ICL increased the T_m of the duplex with respect to the non-cross-linked control with an observed decrease in enthalpy based on thermodynamic analysis of the denaturation curves. Circular dichroism experiments on the ICL duplexes indicated minimal difference from B-form DNA structure. These ICL were used for DNA repair studies with O^6-alkylguanine DNA alkyltransferase (AGT) proteins from human (hAGT) and *E. coli* (Ada-C and OGT), whose purpose is to remove O^6-alkylguanine and in some cases O^4-alkylthymine lesions. It has been previously shown that hAGT can repair O^6-2′-deoxyguanosine-alkyl-O^6-2′-deoxyguanosine ICL. The O^4-dT-alkyl-O^4-dT ICL prepared in this study were found to evade repair by hAGT, OGT and Ada-C. Electromobility shift assay (EMSA) results indicated that the absence of any repair by hAGT was not a result of binding. OGT was the only AGT to show activity in the repair of oligonucleotides containing the mono-adducts O^4-butyl-4-ol-2′-deoxythymidine and O^4-heptyl-7-ol-2′-deoxythymidine. Binding experiments conducted with hAGT demonstrated that the protein bound O^4-alkylthymine lesions with similar affinities to O^6-methylguanine, which hAGT repairs efficiently, suggesting the lack of O^4-alkylthymine repair by hAGT is not a function of recognition.

Introduction

Interference by ICLs of critical cell events involving DNA unwinding is exploited in cancer chemotherapy regimens employing bis-alkylating agents.[1] The potency of these agents may be diminished by the ability of cancer cells to repair the very lesions they induce. Numerous DNA repair pathways including direct, base- and nucleotide-excision repair (NER), homologous recombination (HR), non-homologous end joining and DNA-mismatch repair remove various DNA lesions.[2] Some pathways, such as NER, are complex with broad substrate specificities whereas direct repair by AGT, which involves one protein, has a narrow range of damage that it repairs. ICL damage is challenging to repair as information on both DNA strands is affected. Removal of the ICL on one strand of DNA leaves behind damage on the opposing strand, complicating error-free repair from a template strand. Repair pathways such as NER, HR and translesion DNA synthesis (TLS) have all been implicated in ICL repair in mammalian cells, however, there is an increasing realization that ICL processing may depend on the nature of the lesion making it challenging to generalize how specific ICLs are repaired.[3]

An approach pursued by a number of groups to enhance our understanding of these processes involves the preparation of ICL DNA substrates by solution and solid-phase synthesis for repair studies.[4] A number of elegant examples have been reported including the preparation of various phosphoramidites of nucleosides and other molecules to introduce site-specific lesions in DNA for various experiments.[5] Repair and binding experiments can be conducted by incorporating the oligonucleotides containing lesions into specific plasmids.[6]

Due to the instability of certain ICL formed in DNA as a consequence of DNA treatment with bifunctional alkylating agents it is at times necessary to modify the structure of the ICL by producing a mimic similar in structure but exhibiting improved stability to enable biophysical and repair studies. For example,

Department of Chemistry and Biochemistry Concordia University, 7141 Sherbrooke Street West, Montreal, Canada.
E-mail: cwilds@alcor.concordia.ca; Fax: +1-514-848-2424 ext.5798

† In memory of Professor Har Gobind Khorana (1922–2011), acknowledging his legacy to the scientific community.

‡ Electronic supplementary information (ESI) available: ^{1}H and ^{31}P NMR Spectra, ESI-MS Spectra of DNA, C-18 RP-HPLC traces of phosphodiesterase digests of DNA, T_m and CD data of mono-adduct DNA, DNA Repair gels and results by AGTs and EMSA binding data. See DOI: 10.1039/c2ob25705j

Synthesis of Building Blocks and Oligonucleotides Containing {T}O^4-Alkylene-O^4{T} Interstrand Cross-Links

Derek K. O'Flaherty,[1] Francis P. McManus,[1] Anne M. Noronha,[1] and Christopher J. Wilds[1]

[1]Department of Chemistry and Biochemistry, Concordia University, Montréal, Québec, Canada

ABSTRACT

This protocol describes the preparation of O^4-thymidine-alkylene-O^4-thymidine dimer bis-phosphoramidites and precursors for incorporation into DNA sequences to produce site-specific DNA interstrand cross-links. Linkers are introduced at the 4-position of thymidine by reacting the sodium salt of a diol with a pyrimidinyl-convertible nucleoside to produce mono-adducts, which then undergo reaction with a stoichiometric equivalent of a pyrimidinyl-convertible nucleoside under basic conditions to form O^4-thymidine-alkylene-O^4-thymidine dimers. Bis-phosphoramidites are incorporated into oligonucleotides by solid-phase synthesis, and mild conditions for deprotection and cleavage from the solid support are employed to prevent degradation of the thymidine modifications. Purification of these cross-linked oligonucleotides is performed by denaturing polyacrylamide gel electrophoresis. This approach allows for the preparation of cross-linked DNA substrates in quantities and purity sufficient for a wide range of biophysical experiments and biochemical studies as substrates to investigate DNA repair pathways. *Curr. Protoc. Nucleic Acid Chem.* 55:5.13.1-5.13.19. © 2013 by John Wiley & Sons, Inc.

Keywords: DNA interstrand cross-link • chemically modified oligonucleotide • oligonucleotide synthesis • solid-phase synthesis • DNA damage • DNA repair

INTRODUCTION

This unit describes methods to prepare O^4-hydroxyalkyl-thymidine phosphoramidite mono-adducts and O^4-thymidine-alkylene-O^4-thymidine cross-linked phosphoramidite dimers (Fig. 5.13.1A,B, respectively) and their incorporation into oligonucleotides in a site-specific fashion. The cross-linked oligonucleotides can be synthesized with moderate yields to produce substrates for DNA repair studies. Furthermore, their biophysical properties can be investigated using UV thermal denaturation and circular dichroism to study the role that modifications have on the stability of the duplex and global structure, respectively.

The bis-phosphoramidite approach is used for solid-phase assembly of interstrand cross-linked duplexes. This strategy requires protection of the thymidine at the free alcohol functionalities. The introduction of the linker is performed by generating a convertible thymidine nucleoside containing a triazolyl moiety at the C4 position (Xu and Swann, 1990). The triazolyl group is displaced by a nucleophile, in this case the sodium salt of a bifunctional alkane containing two terminal hydroxyl functionalities, to produce a mono-adduct, which is then coupled with another convertible thymidine nucleoside under basic conditions to produce O^4-thymidine-alkylene-O^4-thymidine (O^4-dT-alkylene-O^4-dT) cross-linked dimers. Orthogonally-protected dimers are desilylated at the

Current Protocols in Nucleic Acid Chemistry 5.13.1-5.13.19, December 2013
Published online December 2013 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/0471142700.nc0513s55

CHEMMEDCHEM
FULL PAPERS

ChemPubSoc Europe

DOI: 10.1002/cmdc.201402121

NMR Structure of an Ethylene Interstrand Cross-Linked DNA which Mimics the Lesion Formed by 1,3-Bis(2-chloroethyl)-1-nitrosourea

Derek K. O'Flaherty, Alexey Y. Denisov, Anne M. Noronha, and Christopher J. Wilds*[a]

Dedicated to Professor Martin Egli on the occasion of his 50th birthday

The bisalkylating agent 1,3-bis(2-chloroethyl)-1-nitrosourea (BCNU), used in cancer chemotherapy to hinder cellular proliferation, forms lethal interstrand cross-links (ICLs) in DNA. BCNU generates an ethylene linkage connecting the two DNA strands at the N1 atom of 2′-deoxyguanosine and N3 atom of 2′-deoxycytidine, which is a synthetically challenging probe to prepare. To this end, an ICL duplex linking the N1 atom of 2′-deoxyinosine to the N3 atom of thymidine via an ethylene linker was devised as a mimic. We have solved the structure of this ICL duplex by a combination of molecular dynamics and high-field NMR experiments. The ethylene linker is well-accommodated in the duplex with minimal global and local perturbations relative to the unmodified duplex. These results may account for the substantial stabilization of the ICL duplex observed by UV thermal denaturation experiments and provides structural insights of a probe that may be useful for DNA repair studies.

Introduction

Interstrand cross-links (ICLs) in DNA represent one of the most lethal forms of damage for the cell as the genetic information on both strands is compromised. These lesions interfere with unwinding of the DNA helix thereby disrupting vital cellular processes such as replication and transcription. More specifically, ICL DNA has shown to trigger DNA polymerase infidelity, slippage, and blockage, which have been correlated with critical events such as mutations, chromosomal instability, and cellular apoptosis stimulation.[1] Albeit detrimental to healthy cells, certain cancer regimens have exploited the introduction of ICL in DNA as a strategy to hinder cellular proliferation. ICL can be generated in DNA by bifunctional alkylating agents, a category of molecules that have been employed as chemotherapeutic drugs.[2] The lesions formed on DNA are dependent both on the agent's chemical nature and the nucleobase. For example, 1,3-bis(2-chloroethyl)-1-nitrosourea (BCNU), used in the treatment of certain brain cancers,[3,4] generates an ICL in DNA, linking the N1 atom of 2′-deoxyguanosine with the N3 atom of 2′-deoxycytidine via an ethylene linker (Figure 1).[5–7]

Cellular repair machinery plays a pivotal role in the survival of cells after suffering ICL damage events. In contrast, repair has also been associated with efficacy reduction of chemotherapy regimens by reverting the damage.[8–10] Mammalian ICL

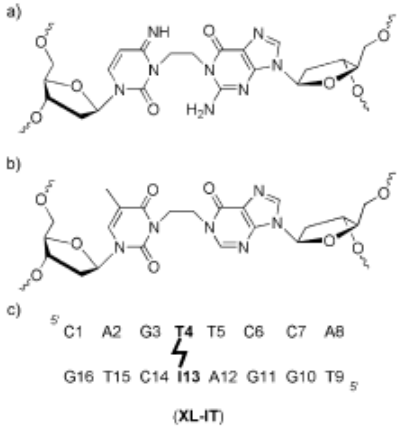


Figure 1. Structures of the a) N3-2′-deoxycytidine-ethylene-N1-2′-deoxyguanosine ICL formed by BCNU, b) N3-thymidine-ethylene-N1-2′-deoxyinosine ICL and c) the sequence of the complementary duplex studied by NMR where T4⁄I13 is the ICL.

DNA repair remains an active area of research in understanding cellular survival and chemotherapy resistance.[11–14] The current understanding of ICL DNA repair is not limited to any one repair pathway, but rather, a combination of several repair pathways. Nucleotide excision repair,[15] trans-lesion synthesis (TLS), homologous recombination, mismatch repair (MMR), and Fanconi anemia (FA)-associated proteins have been implicated in the processing of ICL DNA.[2] For these reasons, it is important to identify key structural features of ICL DNA that could provide insights as to why these lesions are recognized or even processed by repair proteins from different repair path-

[a] D. K. O'Flaherty,[+] Dr. A. Y. Denisov,[+] Dr. A. M. Noronha, Prof. C. J. Wilds
Department of Chemistry & Biochemistry
Concordia University
Montréal, QC H4B 1R6 (Canada)
E-mail: Chris.Wilds@concordia.ca

[+] These two authors made an equal contribution on this paper.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/cmdc.201402121.

Steady-State Kinetic Analysis of DNA Polymerase Single-Nucleotide Incorporation Products

Derek K. O'Flaherty[1] and F. Peter Guengerich[2]

[1]Department of Chemistry and Biochemistry, Concordia University, Montréal, Québec, Canada
[2]Department of Biochemistry and Center in Molecular Toxicology, Vanderbilt University, School of Medicine, Nashville, Tennessee

This unit describes the experimental procedures for the steady-state kinetic analysis of DNA synthesis across DNA nucleotides (native or modified) by DNA polymerases. In vitro primer extension experiments with a single nucleoside triphosphate species followed by denaturing polyacrylamide gel electrophoresis of the extended products is described. Data analysis procedures and fitting to steady-state kinetic models is presented to highlight the kinetic differences involved in the bypass of damaged versus undamaged DNA. Moreover, explanations concerning problems encountered in these experiments are addressed. This approach provides useful quantitative parameters for the processing of damaged DNA by DNA polymerases. © 2014 by John Wiley & Sons, Inc.

Keywords: DNA polymerase • translesion synthesis • steady-state kinetics

How to cite this article:
O'Flaherty, D. K. and Guengerich, F. P. 2014. Steady-State Kinetic Analysis of DNA Polymerase Single-Nucleotide Incorporation Products. *Curr. Protoc. Nucleic Acid Chem.* 59:7.21.1-7.21.13.
doi: 10.1002/0471142700.nc0721s59

BASIC PROTOCOL

SINGLE-NUCLEOTIDE INCORPORATION REACTION OF 2′-DEOXYNUCLEOSIDE TRIPHOSPHATES OPPOSITE A TEMPLATE CONTAINING AN O^4-METHYLTHYMIDINE INSERT

The protocol describes the steady-state primer incorporation of a single nucleoside triphosphate across, or past, a DNA lesion by DNA polymerases, specifically translesion DNA polymerases. As a model system, the damaged DNA contained an O^4-methylthymidine (O^4MeT; Fig. 7.21.1) insert and human DNA polymerase η (hpol η) was the Y-Family translesion polymerase of choice. The complete procedure for the single-nucleotide primer extension assay across the DNA lesion, followed by primer extension product and substrate resolution by polyacrylamide gel electrophoresis (PAGE), are described here. Lastly, data fitting to the Michaelis-Menten kinetic model is presented to enable a quantitative comparison of kinetic parameters for the damaged and wild-type DNA.

STRATEGIC PLANNING

Lesion-containing DNA template

A modified oligomers template can be acquired via commercial sources or can be synthesized using automated solid-phase synthesis employing a phosphoramidite strategy. The lability of the particular modification is paramount and should be taken into consideration when planning these experiments.

Current Protocols in Nucleic Acid Chemistry 7.21.1-7.21.13, December 2014
Published online December 2014 in Wiley Online Library (wileyonlinelibrary.com).
doi: 10.1002/0471142700.nc0721s59

ChemPubSoc Europe

DOI: 10.1002/chem.201501103

CHEMISTRY
A European Journal
Full Paper

Chemical Biology

Synthesis, Characterization, and Repair of a Flexible O^6-2′-Deoxyguanosine-alkylene-O^6-2′-deoxyguanosine Intrastrand Cross-Link

Derek K. O'Flaherty and Christopher J. Wilds*[a]

Abstract: Oligonucleotides tethered by an alkylene linkage between the O^6-atoms of two consecutive 2′-deoxyguanosines, which lack a phosphodiester linkage between these residues, have been synthesized as a model system of intrastrand cross-linked (laCL) DNA. UV thermal denaturation studies of duplexes formed between these butylene- and heptylene-linked oligonucleotides with their complementary DNA sequences revealed about 20 °C reduction in stability relative to the unmodified duplex. Circular dichroism spectra of the model laCL duplexes displayed a signature characteristic of B-form DNA, suggesting minimal global perturbations are induced by the lesion. The model laCL containing duplexes were investigated as substrates of O^6-alkylguanine DNA alkyltransferase (AGT) proteins from human and *E. coli* (Ada-C and OGT). Human AGT was found to repair both model laCL duplexes with greater efficiency towards the heptylene versus butylene analog adding to our knowledge of substrates this protein can repair.

Introduction

DNA is susceptible to modification by various endogenous metabolites and environmental agents.[1,2] Any chemical or structural change to this important biomolecule may have a profound effect on an organism since genetic instructions may be compromised. Lesions that can form in DNA include interstrand cross-links (ICL), intrastrand cross-links (laCL), DNA–protein cross-links (DPCs) and various monoalkylated adducts. The presence of ICL affects both strands of DNA and can block vital cellular processes that require unwinding such as replication and transcription.[3–5] The introduction of laCL in DNA modifies only one of the strands but can lead to a distortion of DNA structure which can result in the interference of critical cellular processes.[6–9] These modifications can have an impact on DNA structural features such as helical bending that can affect the affinity of DNA-binding proteins which may consequently impair cellular functions that involve DNA–protein complex formation.[9–12]

DNA laCL can arise from numerous sources. For example, γ-radiation can induce a G[8,5-Me]T laCL lesion inducing structural perturbations in the DNA that has been shown to lead to blockage of Pol IV (dinB) of *E. coli* and Taq polymerase in vitro.[13] UV irradiation is the main cause of photocarcinogenesis inducing high levels of p53 mutations at bipyrimidine sites in skin tumors.[14–18] Cyclobutane–pyrimidine dimers (CPDs) and pyrimidine (6-4) pyrimidone photoproducts (in addition to their Dewar isomers) are two major classes of laCL DNA lesions introduced by UV irradiation.[17, 19,20] Structural investigation of oligomers containing different UV induced laCL lesions revealed variation in global and local structures, all of which impair synthesis by DNA polymerases.[21]

Some chemical agents can introduce both DNA ICL and laCL lesions and have applications in various chemotherapeutic regimens. One of the most widely studied and administered chemotherapeutic cancer drug, cisplatin, has an estimated lesion distribution of 90–95 % laCL and the remaining 5–10% is attributed to ICL, monoadducts and DPCs.[7,22–25] Bifunctional electrophilic drugs such as busulfan,[26] mitomycin C,[26,27] and mechlorethamine[28] have all been shown to introduce laCL lesions in DNA. A common cross-linking site on DNA occurs between the N^7-atoms of adjacent purines.[26,29] For example, busulfan's main laCL lesion occurs at 5′-d(GA) sites linking the N^7-atoms.[26]

A major mechanism of chemotherapeutic resistance involves DNA damage recognition and repair. laCL DNA is thought to be repaired primarily by nucleotide excision repair (NER)[30,31] and translesion synthesis (TLS).[32,33] Studies with an oxidatively generated laCL between guanine and pyrimidine bases have revealed recognition and, in some cases, incision by the NER machinery in *E. coli*.[34,35] Interestingly, proteins containing a high mobility group (HMG) domain DNA-binding motif showed inhibition of mammalian NER factors from acting on laCL DNA.[36] Certain platinum laCL in DNA, 1,2 d(GG) and 1,3 d(GNG) in particular, are readily recognized and excised by the NER pathway.[37,38] Y-family DNA polymerases, such as Pol η, are efficient and accurate in bypassing the platinum-based laCL between guanine residues.[39–41] Cell lines with mutations in the POLH gene, responsible for encoding Pol η, are hypersensitive

[a] D. K. O'Flaherty, Prof. C. J. Wilds
Department of Chemistry and Biochemistry, Concordia University
7141 Sherbrooke St. West, Montréal, Québec (Canada)
E-mail: Chris.Wilds@concordia.ca

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/chem.201501103.

BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

Backbone Flexibility Influences Nucleotide Incorporation by Human Translesion DNA Polymerase η opposite Intrastrand Cross-Linked DNA

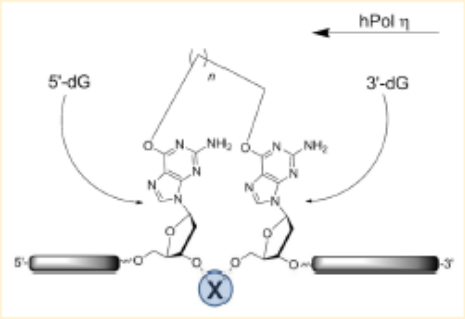
Derek K. O'Flaherty,[†] F. Peter Guengerich,[‡] Martin Egli,[*,‡] and Christopher J. Wilds[*,†]

[†]Department of Chemistry and Biochemistry, Concordia University, 7141 Sherbrooke Street West, Montréal, Québec, Canada H4B 1R6

[‡]Department of Biochemistry and Center in Molecular Toxicology, Vanderbilt University School of Medicine, Nashville, Tennessee 37232-0146, United States

Supporting Information

ABSTRACT: Intrastrand cross-links (IaCL) connecting two purine nucleobases in DNA pose a challenge to high-fidelity replication in the cell. Various repair pathways or polymerase bypass can cope with these lesions. The influence of the phosphodiester linkage between two neighboring 2′-deoxyguanosine (dG) residues attached through the O^6 atoms by an alkylene linker on bypass with human DNA polymerase η (hPol η) was explored *in vitro*. Steady-state kinetics and mass spectrometric analysis of products from nucleotide incorporation revealed that although hPol η is capable of bypassing the 3′-dG in a mostly error-free fashion, significant misinsertion was observed for the 5′-dG of the IaCL containing a butylene or heptylene linker. The lack of the phosphodiester linkage triggered an important increase in frameshift adduct formation across the 5′-dG by hPol η, in comparison to the 5′-dG of IaCL DNA containing the phosphodiester group.



Chemotherapeutic agents such as the platinum-containing drugs used in the treatment of cancer exert their therapeutic effect mainly via the formation of cytotoxic DNA damage. The lesions that these agents produce have been identified primarily as intrastrand cross-links (IaCL) between the N7 atoms of purines with the distribution of these IaCL determined to be 65% 1,2(GpG), 25% 1,2(ApG), and 5−10% 1,3(GpTpG).[1−9] In addition, formation of minor products, including interstrand cross-links (ICL), monoadducts, and DNA−protein cross-links, occur.[9] The presence of these adducts on the DNA scaffold impedes vital cellular processes such as DNA replication and transcription, ultimately leading to cell death. Drugs used in cancer regimens, other than platinum-containing agents, such as mechlorethamine,[10,11] mitomycin C,[12,13] and busulfan,[14] have also been shown to introduce IaCL into DNA, in particular between adjacent purine nucleobases. Using drugs that act directly on DNA to treat cancer has intrinsic and acquired drug resistance as a major limitation, which is mediated by the cellular response processes of DNA repair and translesion DNA synthesis (TLS).

The four TLS DNA polymerases identified in humans are Pol η, Pol κ, Pol ι, and Rev1. The most widely studied of these is Pol η given its crucial involvement in bypassing UV-induced intrastrand cross-linked DNA lesions. Disruption of the proper function of the *POLH* gene leads to xeroderma pigmentosum variant (XPV), a disease characterized by hypersensitivity to UV irradiation and an increased incidence of skin cancer.[15] As suspected, *POLH* knockout mice demonstrated heightened incidences of skin cancer compared to the control group upon being exposed to UV irradiation.[16] XPV cell extracts displayed replication inhibition of plasmid DNA containing a single (6−4) pyrimidone photoproduct lesion.[17] Moreover, human cells deficient in Pol η revealed a greater number of cell death events upon being treated with platinum-based chemotherapeutic agents.[18−21] Exposure of DNA to γ-irradiation leads to the formation of a mixture of the IaCL lesions G[8,5]C and G[8,5]T, among others, formed via a radical mechanism.[22] Their bypass by yeast and/or human Pol η demonstrated reduced fidelity and processivity, in particular across the dG portion of the lesion.[23−25] Accounts of Pol η bypass are numerous, and the search for other biologically relevant DNA damage, or mimics thereof, is ongoing.

DNA alkylating agents such as *N*-nitroso-*N*-methylurea readily modify the N7 atom of dG and, to a lesser extent, the O^6 position. Lesions at the O^6 atom of dG have also been detected after exposure to the methylating and chloroethylating chemotherapeutic drugs temozolimide and carmustine. Endogenous formation of O^6-MedG by *S*-adenosylmethionine is estimated at 10−30 damage events per cell per day.[26] The

Received: October 1, 2015
Revised: November 26, 2015
Published: December 1, 2015

DOI: 10.1021/acs.biochem.5b01078
Biochemistry 2015, 54, 7449−7456

ACES Asian Chemical Editorial Society

DOI: 10.1002/asia.201501253

CHEMISTRY
AN ASIAN JOURNAL
Full Paper

DNA Repair

O^6-Alkylguanine DNA Alkyltransferase Repair Activity Towards Intrastrand Cross-Linked DNA is Influenced by the Internucleotide Linkage

Derek K. O'Flaherty and Christopher J. Wilds*[a]

Abstract: Oligonucleotides containing an alkylene intrastrand cross-link (IaCL) between the O^6-atoms of two consecutive 2'-deoxyguanosines (dG) were prepared by solid-phase synthesis. UV thermal denaturation studies of duplexes containing butylene and heptylene IaCL revealed a 20 °C reduction in stability compared to the unmodified duplexes. Circular dichroism profiles of these IaCL DNA duplexes exhibited signatures consistent with B-form DNA. Human O^6-alkylguanine DNA alkyltransferase (hAGT) was capable of repairing both IaCL containing duplexes with slightly greater efficiency towards the heptylene analog. Interestingly, repair efficiencies of hAGT towards these IaCL were lower compared to O^6-alkylene linked IaCL lacking the 5'-3'-phosphodiester linkage between the connected 2'-deoxyguanosine residues. These results demonstrate that the proficiency of hAGT activity towards IaCL at the O^6-atom of dG is influenced by the backbone phosphodiester linkage between the cross-linked residues.

Introduction

DNA insults may be incurred by environmental agents, endogenous metabolic processes and chemotherapeutic drugs.[1,2] Some chemotherapy regimens exploit the use of bifunctional electrophilic agents, which act on DNA to produce lesions and interfere with vital processes such as DNA replication.[3] Stalling these processes or modification of the information encoded in DNA can have severe effects on the cell. Bifunctional electrophilic agents can form adducts in DNA which link the atoms of two nucleotides on the same (intra) or opposing (inter) strands. Interstrand cross-links (ICL) in DNA are particularly cytotoxic given that the unwinding of the individual strands is prevented, which impedes cellular proliferation.[3–5] Although damage occurs on one strand for intrastrand cross-linked (IaCL) DNA, information content and structure can be affected. Cisplatin, a platinum-based drug used for the treatment of some forms of cancer, predominantly introduces lesions that are IaCL in nature.[6–10] Busulfan[11] and mitomycin C[12,13] are examples of bis-electrophilic agents which can introduce IaCL lesions in DNA, and can lead to the blockage of DNA replication and activation of apoptosis.[14,15]

O^6-Alkyl-2'-deoxyguanosine adducts can be formed by exposure to carcinogenic agents such as nitrosamines and chemotherapeutic alkylating agents such as temozolomide.[16] This class of lesions can induce transitional mutations via proficient base pairing with thymidine during DNA replication,[17] and the activation of mismatch repair can result in futile excision-resynthesis, ultimately leading to apoptosis.[18,19] Our group has shown that human O^6-alkylguanine DNA alkyltransferase (hAGT) is capable of repairing certain ICL.[20–24] The primary role of AGTs is to remove alkyl lesions found at the O^6-position of dG and, to a lesser extent, at the O^4-position of dT.[25] In the active site of the protein, the alkyl lesion becomes transferred to an activated cysteine, rendering the AGT protein irreversibly inactivated and ultimately degraded in vivo by the ubiquitin pathway.[26,27]

The human AGT (hAGT) variant is capable of repairing butylene and heptylene ICL linking the O^6-positions of dG residues in a directly opposed fashion, as well as in a 5'-GNC motif.[20,21] Recently, we demonstrated that hAGT could also act on a DNA duplex containing an IaCL on one of the strands consisting of an alkylene functionality attaching the O^6-atoms of two dG residues that lack a phosphodiester linkage in the backbone between them (Figure 1A).[28] hAGT was proficient towards removing the lesion, with almost complete consumption observed for a heptylene-linked substrate within 60 min. These results prompted us to explore AGT activity towards the IaCL containing a phosphodiester linkage at this site (Figure 1B).

In the current investigation, a methodology to introduce this O^6-2'-deoxyguanosine-alkylene-O^6-2'-deoxyguanosine IaCL (O^6-dG-alkylene-O^6-dG, Figure 1B) was developed. The influence of this IaCL modification in a DNA duplex was assessed by UV thermal denaturation and circular dichroism. Repair of this IaCL DNA by a variety of AGTs (human and *E. coli*) was evaluated. Given the efficient action of hAGT on the IaCL DNA probes lacking the phosphodiester linkage between the 3'-

[a] D. K. O'Flaherty, Prof. C. J. Wilds
Department of Chemistry&Biochemistry
Concordia University
7141 Sherbrooke St. West, Montréal, Québec (Canada)
E-mail: Chris.Wilds@concordia.ca

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/asia.201501253.