Probing the Activity of O^6-alkylguanine-DNA Alkyltransferases on Alkylene Interstrand Cross-linked DNA

Francis Patrice McManus

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

March 2013

CONCORDIA UNIVERSITY
SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: Francis P. McManus

Entitled: Probing the Activity of O6-alkylguanine-DNA Alkyltransferases on Alkylene Interstrand Cross-linked DNA

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

Doctor of Philosophy in 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:

Dr. Andrew Chapman — Chair

Dr. Dindial Ramotar — External Examiner

Dr. Laszlo Kalman — External to Program

Dr. Joanne Turnbull — Examiner

Dr. Sébastien Robidoux — Examiner

Dr. Christopher Wilds — Thesis Supervisor

Approved by

Chair of Department or Graduate Program Director

Dean of Faculty

Probing the Activity of O^6-alkylguanine-DNA Alkyltransferases on Alkylene Interstrand Cross-linked DNA

Francis P. McManus, Ph.D.

Concordia University, 2013

O^6-Alkylguanine-DNA alkyltransferases (AGT) are responsible for the removal of numerous mutagenic O^6-alkyl 2'-deoxyguanosine (O^6-alkyl dG) and O^4-alkyl thymidine (O^4-alkyl dT) adducts. The function of AGT in the recognition and removal of both mono-adducts and DNA interstrand cross-links (ICL) were undertaken. To achieve this, oligonucleotide probes containing alkylene linkers varying in length that tethered the O^4 atoms of thymidine (dT) or 2'-deoxyuridine (dU) and O^6 atoms of 2'-deoxyguanosine (dG) were prepared. These ICL were generated using a combination of solution and solid phase synthesis. Various O^4-alkyl dT and O^4-alkyl dU mono-adducts were also prepared by similar methods and analyzed to further elucidate the limitation of AGT with respect to damage detection and repair. Three AGT homologues (human AGT/hAGT, and *E. Coli* OGT and Ada-C) and an engineered chimera were used in repair and binding studies with the O^4 and O^6 modified oligonucleotides to note variations amongst species. Studies with the AGTs and modified DNA probes revealed: 1) hAGT is capable of repairing O^6-2'-deoxyguanosine-butylene-O^6-2'-deoxyguanosine (O^6dG-butylene-O^6dG) and O^6-2'-deoxyguanosine-heptylene-O^6-2'-deoxyguanosine (O^6dG-heptylene-O^6dG) ICLs in a 5'-GNC sequence motif, designed to mimic the configuration of the ICL generated by hepsulfam (1,7-disulfamoyloxyheptane); 2) O^4-thymidine-alkylene-O^4-thymidine (O^4dT-alkylene-O^4dT) ICLs evade repair from all AGTs studied, whereas the O^4-butyl-4-ol and O^4-heptyl-7-ol dT mono-adducts undergoes repair by OGT; 3) Creating an hAGT:O^4-

alkyl dT covalent complex for crystallography by employing O^6-2'-deoxyguanosine-alkylene-O^4-thymidine (O^6dG-alkylene-O^4dT) ICL DNA shows promise since covalent complex formation is observed with O^6dG-heptylene-O^4dT cross-links; 4) Introducing the active site of OGT into the hAGT scaffold confers the chimera with enhanced repair capabilities of O^4-alkyl dT adducts with respect to hAGT while maintaining ICL repair activity; 5) AGT mediated repair of O^4-alkyl dT damage is hindered by the presence of the C5 methyl (up to 30-fold). This effect is most notable with hAGT and the chimera; 6) O^4-2'-deoxyuridine-alkylene-O^4-2'-deoxyuridine (O^4dU-alkylene-O^4dU) ICL DNAs are not repaired by AGT. This elusive property is attributed to the *E* conformation adopted by the linker about the C4-O^4 bond of the modified dU, as demonstrated from NMR studies. There are no working models that provide a feasible hAGT based repair mechanism for ICL damage. ICL DNA:hAGT crystal structures are needed to shed light on this poorly understood activity.

Acknowledgments

I would like to start off by thanking my family: my wife Christine and kids, Noah and Lucy. They are a tremendous source of motivation, but most importantly happiness. I am grateful for their support and continued love. Many thanks also go out to: my parents, Michelle Deslauriers and Gary McManus; my grandparents, Jeanine and Jean-Robert Deslauriers; my brother Casey; and my nephews Calvin, Owen and Dylan.

I am grateful to Dr. Christopher Wilds for giving me a chance to pursue my studies in his laboratory. He has been a great mentor, teacher and supervisor throughout my PhD. I would like to thank him for teaching me how to conduct chemistry in a lab, especially how to run FCC. My thanks go out to the current and past members of the lab: Dr. Anne Noronha, Derek O'Flaherty, Gang Sun, David de Bellefeuille, Jack Cheong, Vincent Grenier and Nadia Schoonhoven. An added thanks to Dr. Noronha for teaching me how to use the solid phase synthesizer, even though I still fear it.

I wish to extend my appreciation to my committee members, Dr. Sébastien Robidoux and Dr. Joanne Turnbull for their ideas and probing questions. Further thanks goes out to Dr. Robidoux for training me on the NMR and Dr. Turnbull for allocating a section of her lab for my radioactive work.

I would like to thank students from other labs: Michael Lowden, Daniel Foshag, Ian Jaworski, Vasken Parsekian, Peter Quashie, Natascha Hotz, Said Hassouna and Nicholas Sitaras.

My gratitude goes out to Alexey Denisov for solving the solution structures of the ICL DNAs. Without him CHAPTER VII would be incomplete.

Table of contents

List of figures

List of schemes

List of tables

List of abbreviations

ACN	- Acetonitrile
AGT	- O^6-alkylguanine-DNA alkyltransferase
Alloc-OBt	- Allyl 1-benzotriazolyl carbonate
AP	- Apurinic/Apyrimidinic
ATM	- Ataxia-telangiectasia-mutated protein kinase
BER	- Base excision repair
CD	- Circular dichroism
CHARMM	- Chemistry at HARvard Macromolecular Mechanics
CHO	- Chinese hamster ovary
CPD	- Cyclobutane-pyrimidine dimer
CPG	- Controlled pore glass
dA	- 2'-Deoxyadenosine
DBU	- 1,8-Diazabicycloundec-7-ene
dC	- 2'-Deoxycytidine
DCM	- Dichloromethane
dG	- 2'-Deoxyguanosine
DIAD	- Diisopropylazidodicarboxylate
DIPEA	- Diisopropylethylamine
DMAP	- *N*,*N*'-Dimethylaminopyridine
DMF	- *N*,*N*-Dimethylformamide
DMT	- 4,4'-Dimethoxytrityl
DNA	- 2'-deoxyribonucleic acid
DQF-COSY	- Double quantum filtered correlation spectroscopy
DR	- Direct Repair
dsDNA	- Double-stranded 2'-deoxyribonucleic acid
DSB	- Double strand break
DSS	- 4,4-Dimethyl-4-silapentane-1-sulfonic acid
dT	- Thymidine
dU	- 2'-Deoxyuridine
E. coli	- *Escherichia coli*
EDTA	- Ethylenediaminetetraacetic acid
EMSA	- Electro mobility shift assay
Fapy	- Formamido pyrimidine
hAGT	- Human O^6-alkylguanine-DNA alkyltransferase
HPLC	- High performance liquid chromatography
HR	- Homologous recombination
HSQC	- Heteronuclear single quantum correlation
ICL	- Interstrand cross-link
IPTG	- Isopropyl β-D-thiogalactopyranoside
K_d	- Dissociation constants
LB	- Luria broth
MLEV	- Malcolm Levitt's composite-pulse decoupling sequence
MMR	- Mismatch repair

MMS	- Methyl methanesulfonate
MNNG	- *N*-methyl-*N*'-nitro-*N*-nitrosoguanidine
MNU	- *N*-methyl-*N*'-nitrosourea
MS	- Mass spectrometry
NBS1	- Nijmegen breakage syndrome 1
NER	- Nucleotide excision repair
NHEJ	- Nonhomologous end joining
NMR	- Nuclear magnetic resonance
NOE	- Nuclear Overhauser effect
NOESY	- Nuclear Overhauser effect spectroscopy
O^4-BndU	- O^4-benzyl-2'-deoxyuridine
O^4-MedT	- O^4-methyl-thymidine
O^4dT-alkylene-O^4dT	- O^4-thymidine-alkylene-O^4-thymidine
O^4dU-alkylene-O^4dU	- O^4-2'-deoxyuridine-alkylene-O^4-2'-deoxyuridine
O^6-BndG	- O^6-benzyl-2'-deoxyguanosine
O^6dG-alkylene-O^4dT	- O^6-2'-deoxyguanosine-alkylene-O^4-thymidine
O^6dG-alkylene-O^6dG	- O^6-2'-deoxyguanosine-alkylene-O^6-2'-deoxyguanosine
O^6-MedG	- O^6-methyl-2'-deoxyguanosine
O^6-methyl-riboG	- O^6-methyl-guanosine
PAC	- Phenoxyacetyl
PAGE	- Polyacrylamide gel electrophoresis
PBS	- Phosphate buffered saline
PCR	- Polymerase chain reaction
PNK	- Polynucleotide kinase
RPA	- Replication protein A
SAX	- Strong anion exchange
SDS-PAGE	- Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SDS	- Sodium dodecyl sulphate
S_N1	- Nucleophilic substitution, first order
S_N2	- Nucleophilic substitution, second order
ssDNA	- Single-stranded 2'-deoxyribonucleic acid
TBAF	- 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
TOCSY	- Total correlated spectroscopy
UV	- Ultraviolet

Contribution of Authors

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

CHAPTER II:

"O^6-2'-deoxyguanosine-alkylene-O^6-2'-deoxyguanosine interstrand cross-link in a 5'-GNC motif and repair by human O^6-alkylguanine-DNA alkyltransferase"

Fang Q: Performed Colony Forming Assay in CHO Cells, which constituted Figure 4 of the article.

Booth JD: Performed the solution synthesis starting from 5′-*O*-dimethoxytrityl-N^2-phenoxyacetyl-2′-deoxyguanosine to generate the amidites 7a and 7b. These experiments constituted Scheme 1 of the article.

Noronha AM: Performed the solid-phase synthesis of the oligonucleotides including the later extension step after de-silylation. These experiments constitute part of Figure 2.

Pegg AE: Project supervisor for Fang Q.

Wilds CJ: Project supervisor for McManus FP, Booth JD and Noronha AM. Responsible for correcting the manuscript.

McManus FP: Wrote the first draft of the manuscript and corrected the manuscript in later versions. Performed remainder of experiments: solid support chemistry (except for extensions on the solid phase synthesizer); purified oligonucleotides; snake venom digestion analysis of oligonucleotides, circular dichroism and thermal denaturation experiments of oligonucleotides; purified AGT proteins; biophysical characterization of proteins; repair and binding assays.

CHAPTER III:

"O^4-alkyl thymidine cross-linked DNA to probe recognition and repair by O^6-alkylguanine DNA alkyltransferases"

O'Flaherty DK: Performed the solution synthesis starting from 3′-*O*-(*t*-butyldimethylsilyl)-5′-*O*-(4,4′-dimethoxytrityl)-C4-(1,2,4-triazol-1-yl)-thymidine to generate the pre-amidites 3a, 3b, 6a and 6b. These experiments constituted the majority of Schemes 1 and 2.

Noronha AM: Performed the solid-phase synthesis, purification and snake venom digestion of the oligonucleotides.

Wilds CJ: Project supervisor responsible for correcting the manuscript. Performed the phosphitylation reactions on 3a, 3b, 6a and 6b to generate 4a, 4b, 7a and 7b, respectively.

McManus FP: Wrote the first draft of the manuscript and corrected the manuscript in later versions. Performed remainder of experiments. These include: circular dichroism and thermal denaturation analysis of the oligonucleotides; thermodynamic analysis of the oligonucleotides; AGT purifications; repair assays; binding assays and molecular modelling.

CHAPTER IV:

"Preparation of Covalently Linked Complexes Between DNA and O^6-Alkylguanine-DNA Alkyltransferase Using Inter-strand Cross-linked DNA"

Khaira A: Performed the solution synthesis of compounds 1a and 2a. These experiments constitute half of Scheme 1.

Noronha AM: Performed the solid-phase synthesis of the oligonucleotides.

Wilds CJ: Project supervisor responsible for correcting the manuscript. Performed the phosphitylation reactions on 2a and 2b to generate 3a and 3b, respectively.

McManus FP: Wrote the first draft of the manuscript and corrected the manuscript in later versions. Performed remainder of experiments. These include: solution synthesis of compounds 1b and 2b; purification of oligonucleotides, snake venom analysis of the oligonucleotides; circular dichroism and thermal denaturation analysis of the oligonucleotides; AGT purifications; repair assays; binding assays; SDS-PAGE of cross-linking reaction and mass spectrometry analysis of the cross-linking reaction.

CHAPTER V:

"Engineering of a O^6-Alkylguanine-DNA alkyltransferase Chimera and Repair of O^4-Alkyl Thymidine Adducts and O^6-Alkylene-2'-Deoxyguanosine Cross-Linked DNA"

Wilds CJ: Project supervisor responsible for correcting the manuscript.

McManus FP: Wrote the first draft of the manuscript and corrected the manuscript in later versions. Performed all experiments.

CHAPTER VI:

"Effect of the C5 Methyl on O^6-Alkylguanine-DNA Alkyltransferase Mediated Repair of O^4-alkyl Thymidine Adducts"

Wilds CJ: Project supervisor responsible for correcting the manuscript. Performed the phosphitylation reactions on 2a and 5a to generate 3a and 6a, respectively.

McManus FP: Wrote the first draft of the manuscript and corrected the manuscript in later versions. Performed all other experiments.

CHAPTER VII:

"Structural Basis for hAGT Selective Mediated Repair of O^6-2'-Deoxyguanosine-heptylene-O^4-2'-deoxyguanosine and not O^4-2'-Deoxyuridine-heptylene-O^4-2'-deoxyuridine Interstrand Cross-Linked Containing DNA"

Denisov AY: Co-first author of the manuscript. Conducted NMR spectroscopy analysis of oligonucleotides and structural modelling. These results constitute Figure 5-9 and Table 2 of the document. Prepared the above mentioned figures and table and wrote the NMR related experimental and results sections of the manuscript.

O'Flaherty DK: Prepared the O^4dT-heptylene-O^4dT ICL DNA.

Noronha AM: Performed the solid-phase synthesis of the oligonucleotides.

Wilds CJ: Project supervisor responsible for correcting the manuscript.

McManus FP: Wrote the first draft of the manuscript (excluding the NMR related experimental and results sections) and corrected the manuscript. Performed remainder of the experiments. These include: solution synthesis of the dG and dU containing small molecules; purification of O^4dU-heptylene-O^4dU, O^6dG-heptylene-O^6dG ICL DNAs and respective controls; snake venom analysis of the aforementioned oligonucleotides; circular dichroism and thermal denaturation analysis of the oligonucleotides; AGT purifications; repair and binding assays.

CHAPTER I

General Introduction

1.1 Preamble

DNA encodes the information required for biosynthesis of RNA and proteins, it is therefore essential for life. DNA is the source of information in the cell, it dictates the amino acid sequence of each protein. It is clear that the integrity of the genomic DNA should not be compromised or else the ramifications of these modifications in the DNA will be felt on multiple levels. The DNA that constitutes our genome is constantly under attack by endogenous and exogenous agents. The damage induced by these agents can have a harmful effect on the host due to their ability to induce mutations and/or modifications in the DNA. Mutation are more prone to occur when the agent damages the heterocyclic base of the DNA, while alkylation of the phosphotriester moiety, for example, are virtually silent on the nucleotide context. ICLs are very cytotoxic since complementary DNA strands are covalently bound together, inhibiting DNA replication, RNA transcription and ultimately gene expression.

Cells have evolved repair mechanisms to deal with DNA damage that arise from exposure to modifying agents. The ultimate role of all DNA repair mechanisms are the same; revert the modified DNA back to its native state. Each DNA repair mechanism has its set of substrates, where some overlap amongst mechanisms is known. The repair mechanisms are not ubiquitous in all kingdoms of life. Different organisms have different/extra classes of repair systems with most of the repair pathways showing overlap amongst species. These vast arrays of repair mechanisms and variation amongst species constitute a fine balance between DNA damage formation and DNA damage removal.

1.2 DNA alkylation

1.2.1 Classes of DNA alkylating agents

Most alkylating agents can be broken down into two general classes: mono-functional and bi-functional. As the name infers, mono-functional alkylating agents cause one alkylation event in the DNA per molecule whereas the bi-functional alkylating agents can cause two alkylation events.

Mono-functional alkylating agents can be harmful to the cell by impeding the ability of certain DNA polymerases to continue DNA replication and hence stop cell growth. The ability of an alkylation event to impede DNA polymerase function depends on the atom in question undergoing modification. Modification of the N3 atom of 2'-deoxyadenosine (dA), the second most predominant point of alkylation, leads to replication inhibition.(*1*,*2*,*3*) Some sites have virtually no effect on polymerase activity, such as the N7 atom of dG, the most predominant point of alkylation.(*1*)

Mono-functional alkylating agents can also induce mutations in both genomic and mitochondrial DNA. For example, alkylation at the O^6 atom of dG is mutagenic since it participates in base-pairing, which will be elaborated upon later in the introduction.(*4*,*5*)

Bi-functional alkylating agents have shown an increased ability to suppress tumor cells with reports dating back 60 years.(*6*) Bi-functional alkylating agents are more harmful to the cell than their mono-functional counterparts due their ability to introduce mono-adducts and cross-links in DNA duplexes.(*7*) The cross-links induced by this class of alkylating agents on DNA can either be intrastrand cross-links, where both alkylation events occur on the same DNA strand, or interstrand cross-links (ICL), where the two

alkylation events occur on complementary DNA strands. Their ability to cause ICL lesions in DNA is thought to be the reason for the cytotoxicity of these compounds.(*8*) The introduction of a covalent linkage between two complementary strands of DNA, inhibits DNA strand segregation.(*7*,*9*) Many cellular processes rely on this segregation for function with DNA replication and RNA transcription as examples. This entails that all processes downstream of these two vital events will also be inhibited, ultimately leading to cell arrest and apoptosis.

1.2.2 Sites of DNA alkylation and their biological consequences

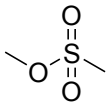
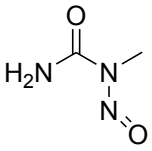
Not all atoms of the DNA duplex are acted upon by alkylating agents. Alkylating agents perform their function due to their electrophilic properties with the exception of psoralens and other agents which act by intercalating between DNA bases prior to photoactivation.(*7*) There exist no alkylating agents capable of exclusive selectivity in targeting a defined nucleophilic centre.

DNA alkylation occurs *via* a reaction of a nucleophilic center in the DNA on the electrophilic moiety of the alkylating agent and can be described to follow a mechanism that is either unimolecular (S_N1) or bimolecular in nature (S_N2).(*10*) In the S_N1 type mechanism the rate of the reaction is governed solely by the nature of the electrophile (alkylating agent) making compounds that fall under this category less specific. In the S_N2 type mechanism the rate of the reaction is not only governed by the electrophile but also by the nucleophile (nucleophilic centers of the DNA) making these agents more selective.

The "softer", more charged delocalized alkylating agents undergo S_N2 type mechanisms yielding predominantly N7-alkyl dG. The "harder", higher charge density

alkylating agents undergo S_N1 type mechanisms yielding less of N7-alkyl dG and more of the exocyclic oxygen adducts, such as O^6-alkyl dG and O^4-alkyl dT.(*1*,*10*,*11*) **Table 1.1** shows how the nature of the alkylating agent can alter which atoms in the DNA undergo modification.

Table 1.1: Percentage of total alkylation of each nucleophilic atom in DNA when treated with methyl methanesulfonate (MMS) or *N*-methyl-*N*'-nitrosourea (MNU) (adapted from *12*)

Atom Methylated		MMS (Soft) 		MNU (Hard) 	
		In vitro	*In vivo*	*In vitro*	*In vivo*
Adenine	N1	3.8		1.3	
	N3	10.4	8.7	9.0	4.2
	N7	1.8	2.4	1.7	3.1
Guanine	N3	0.6	1.3	0.8	0.5
	O^6	0.3	0.3	6.3	5.1
	N7	83.0	83.0	67.0	69.0
Thymine	O^2	nd		0.11	5.1
	N3	0.1		0.3	
	O^4	nd		0.4	
Cytosine	O^2	nd		0.1	
	N3	<1	<1	0.6	0.4
Diester		0.8		17.0	9.0

The carcinogenicity of the various alkylating agents are not directly correlated with the magnitude of alkylating events but rather with the mutagenicity of the lesions. Despite the relatively low proportion of O^6-alkyl dG formed with respect to the total DNA damage, this lesion along with O^4-alkyl dT, are used as biomarkers to rank the mutagenic/carcinogenic factor of the different alkylating agents.

O^6-alkyl dG is mutagenic due to the tendency of DNA polymerases to incorporate an erroneous dT nucleotide opposite to the modification during DNA replication and

repair by the mismatch repair (MMR) machinery.(*13*) In human O^6-alkylguanine-DNA alkyltransferase (hAGT) deficient tumor cells, one lethal event occurs for every 6 650 O^6-methyl-2'-deoxyguanosine (O^6-MedG) produced in the genome.(*14*)

Studies conducted with AGT deficient cells revealed that the MMR pathway was detrimental to the cell in the presence of O^6-MedG.(*15*,*16*) The MMR machinery perceives an O^6-MedG•dT pair as a substrate for repair as this base pairing does not constitute one of the two natural interactions found in DNA. In the case where the O^6-MedG is on the non-template strand the MMR machinery can repair the lesion correctly by removing the incorrectly incorporated modified nucleotide since it is on the newly synthesized strand. In the opposite case, where the lesion is present in the template strand, the MMR machinery cannot excise the modified nucleotide and thus it removes the dT erroneously incorporated opposite to the O^6-MedG. This is followed by the misincorporation of dT opposite to O^6-MedG. Multiple rounds of futile MMR eventually signal the cell to commit apoptosis in order to conserve genetic integrity.(*17*) This property has been exploited during chemotherapeutic regimens. Methylating agents capable of forming considerable amounts of O^6-MedG, such as temozolomide, are commonly used in the treatment of brain cancers where hAGT levels are at their lowest.(*18*) If the O^6-alkyl dG•dT pair evades all repair mechanisms and remains in the genomic DNA a dG•dC to dA•dT transition is ultimately produced.(*5*,*19*)

The thermal stability of various duplexes consisting of the self-complementary Dickerson dodecamers (CGXGAATTCYCG), where X is O^6-MedG and Y is any of the four natural 2'-deoxynucleotides revealed that the thermal stability of the duplex decreased according to the following trend: dG•dC > dG•T > O^6-MedG•dC > O^6-

MedG•dA > O^6-MedG•dG > O^6-MedG•T.(*20*) These results are counterintuitive considering that dT is generally incorporated opposite to O^6-MedG by DNA polymerases.(*21*) If H-bonding properties governed the incorporation of the nucleotide opposite to O^6-MedG then dT should be incorporated least. Rather, preferential incorporation of dT opposite to O^6-MedG is governed by Watson-Crick geometry. NMR studies revealed that O^6-MedG forms a non-wobble base pair only with dT, see **Figure 1.1**.(*22*,*23*)

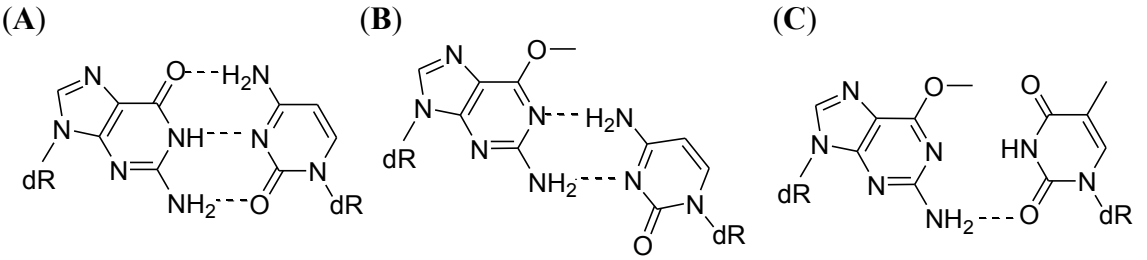


Figure 1.1: Watson-Crick pairing of (**A**) dG•dC, (**B**) O^6-MedG•dC and (**C**) O^6-MedG•dT

Much like O^6-alkyl dG, O^4-alkyl dT is highly mutagenic. O^4-alkyl dT is mutagenic due to its ability to cause DNA polymerases to incorporate dG instead of dA opposite to it during DNA replication.(*4*) O^4-methyl-thymidine (O^4-MedT) is a minor alkylation product with respect to its O^6-MedG counterpart. MNU treated DNA produces 126 times more O^6-MedG than O^4-MedT.(*24*) As with O^6-MedG, misincorporation of dG at O^4-MedT sites is governed by the geometry of the base pair. NMR studies demonstrated that O^4-MedT forms a non-wobble base pair only with dG, see **Figure 1.2**.(*25*,*26*)

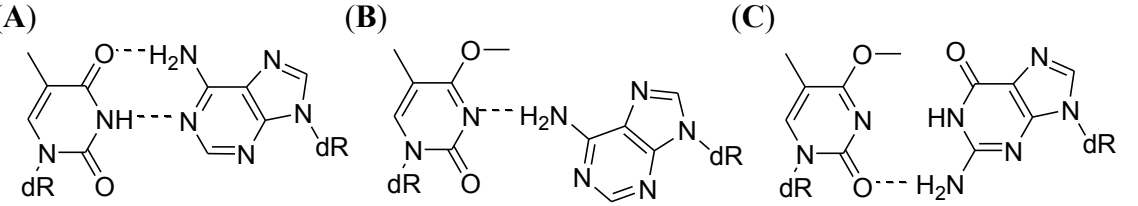


Figure 1.2: Watson-Crick pairing of (**A**) dT•dA, (**B**) O^4-MedT•dA and (**C**) O^4-MedT•dG

In repair proficient cells transformed with single-stranded DNA (ssDNA) containing specific lesions, mutations rose 12% opposite to the O^4-MedT while a < 2% mutation rate was observed opposite to O^6-MedG.(*27*) A similar trend was observed for double-stranded DNA (dsDNA) in *Escherichia coli* (*E. coli*), where O^4-MedT caused a 45% mutation rate as opposed to 6% by O^6-MedG.(*28*) In mammalian cells, O^4-MedT is more toxic than O^6-MedG both in normal and repair deficient systems.(*29*,*30*) Unlike O^6-MedG, O^4-MedT is not affected by the expression or suppression of hAGT activity nor by MMR, explaining its added toxicity.(*29*,*30*) Nucleotide excision repair (NER) deficient cells exhibit added sensitivity to O^4-MedT indicating that this lesion is eliminated *via* the NER pathway in mammalian cells.(*31*) *E. coli* has evolved two proteins from the direct repair pathway to process O^4-MedT lesions; OGT and the c-terminal of the adaptive response protein Ada-C.(*4*,*32*,*33*,*34*) As a result O^4-MedT is less toxic in *E. coli* cells.

ICL formation represents the most lethal event of all the possible reactions that can occur upon exposure to bi-functional alkylating agent (shown in **Figure 1.3**).(*8*) The presence of 40 ICL lesions in repair deficient mammalian cell is sufficient to cause cell death, while only one ICL is enough to kill repair deficient bacteria and yeast.(*35*,*36*) Due to their lethal properties ICL forming agents are employed in many instances during chemotherapy.(*37*,*38*,*39*) Some of the more commonly used and potent anti-cancer agents used today are cross-linking agents such as: nitrogen mustards, platinum drugs, alkyl methanesulfonates and mytomycin C.(*9*) In a study of 234 chemicals, which were tabulated in order of genotoxicity, bi-functional alkylating agents repeatedly appeared in

the most toxic chemicals.(*40*) Moreover, 12 out of the 15 most potent anti-cancer drugs were cross-linking agents.(*41*)

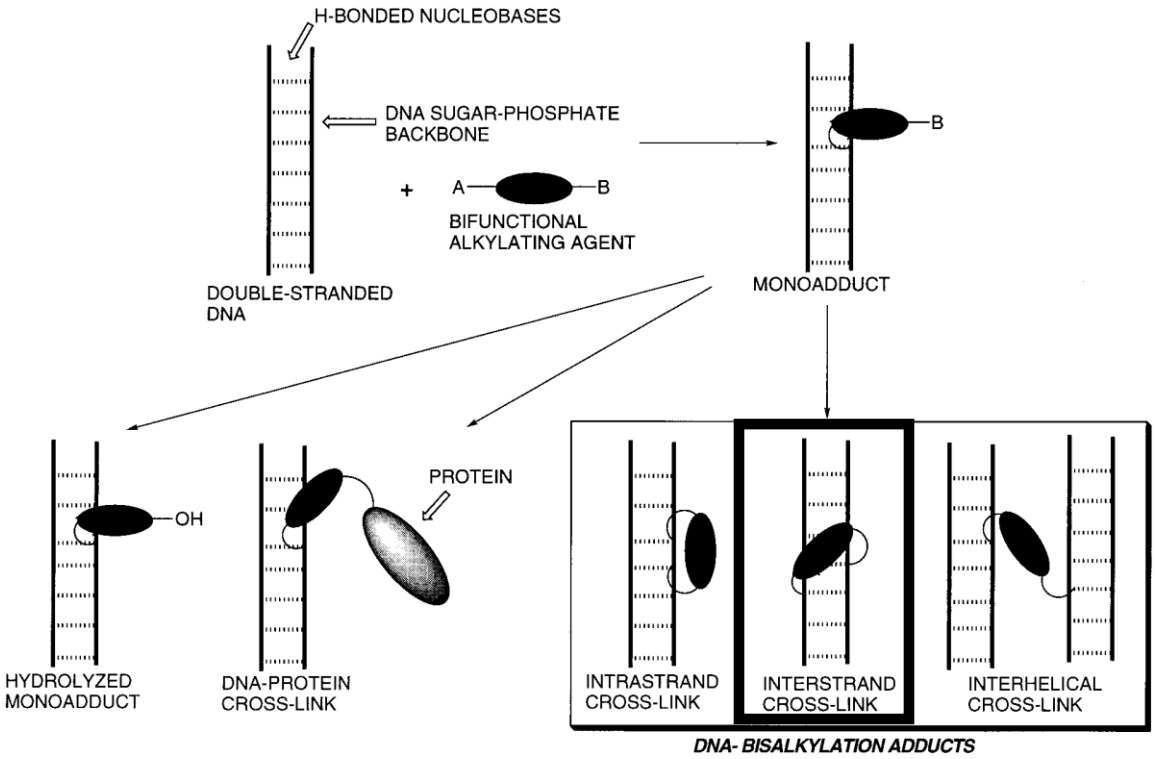


Figure 1.3: Possible fates of bi-functional alkylating agents (*42*)

In general, ICLs are challenging to eliminate causing their prolonged presence in the cell with respect to mono-alkylation.(*9*) This property allows ICL to target faster proliferating cells such as tumor and cancer cells.(*8*) As previously introduced, mono-functional alkylating agents can react with multiple atoms on the nucleobases. Bi-functional alkylating agents can react with those same atoms with an additional step of a second alkylation event increasing the number of different lesions induced by these agents. The various bi-functional alkylating agents display specific properties, which include which atoms are targeted and more importantly which nucleotide sequences are targeted for ICL formation.(*42*)

1.2.3 Sources of DNA alkylation

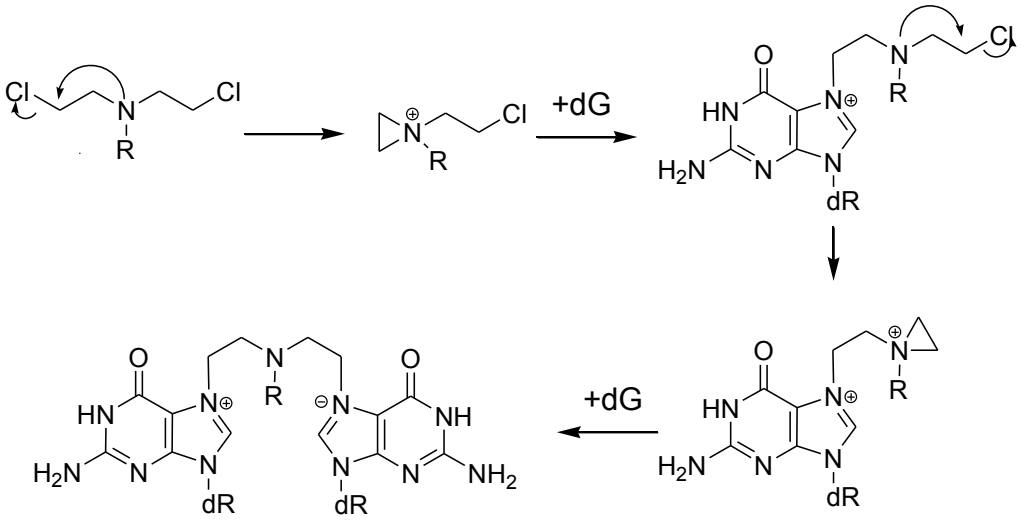
The most prevalent endogenous DNA damage is the N7-methyl dG adduct. It is thought that on a daily basis 4000 N7-methyl dG adducts form per cell and its steady-state level is around 3000 copies per cell.(*43*) S-adenosylmethionine, the major methyl donor in cells, is the main endogenous factor responsible for the formation of N7-methyl dG adducts.(*44*,*45*) S-adenosylmethionine is also responsible for the formation 600 N3-MedA and 10-30 O^6-MedG events, daily.(*46*) Other endogenous sources of DNA alkylation include: N-nitrosoglycocholic acid (N-nitroso bile acid conjugates) that form O^6-carboxymethyl dG and O^6-MedG;(*47*,*48*,*49*,*50*) and lipid peroxidation products, such as malondialdehyde that generate an array of adducts.(*51*,*52*,*53*)

DNA modification by exogenous factors have been extensively analyzed. The adverse effects of tobacco have been actively investigated for decades. The metabolic activation of tobacco leads to the formation of the methyldiazonium ion, a known DNA methylating agent, which is responsible for the production of N7-methyl dG, O^6-MedG and O^4-MedT.(*54*) Other tobacco derived adducts include: N7-[4-(3-pyridyl)-4-oxobutyl] dG, O^6-[4-(3-pyridyl)-4-oxobutyl] dG, N^2-[4-(3-pyridyl)-4-oxobutyl] dG, O^2-[4-(3-pyridyl)-4-oxobutyl] dT, O^2-[4-(3-pyridyl)-4-oxobutyl] dC, N7-[4-(3-pyridyl)-4-hydroxybut-1-yl] dG, O^6-[4-(3-pyridyl)-4-hydroxybut-1-yl] dG, N^2-[4-(3-pyridyl)-4-hydroxybut-1-yl] dG, O^2-[4-(3-pyridyl)-4-hydroxybut-1-yl] dT and O^2-[4-(3-pyridyl)-4-hydroxybut-1-yl] dC.(*55*,*56*,*57*,*58*)

If alkylation damage to DNA is not repaired, this may have various negative outcomes in the cell. On the other hand, the introduction of lesions in DNA by the action of alkylating drugs can be beneficial overall; such is the case with chemotherapeutic

agents. In this scenario, the lesions induced exert their toxic influence to the foreign body (such as a cancer) to a greater extent than to the host, ultimately resulting in elimination of the foreign body without being lethal to the host.

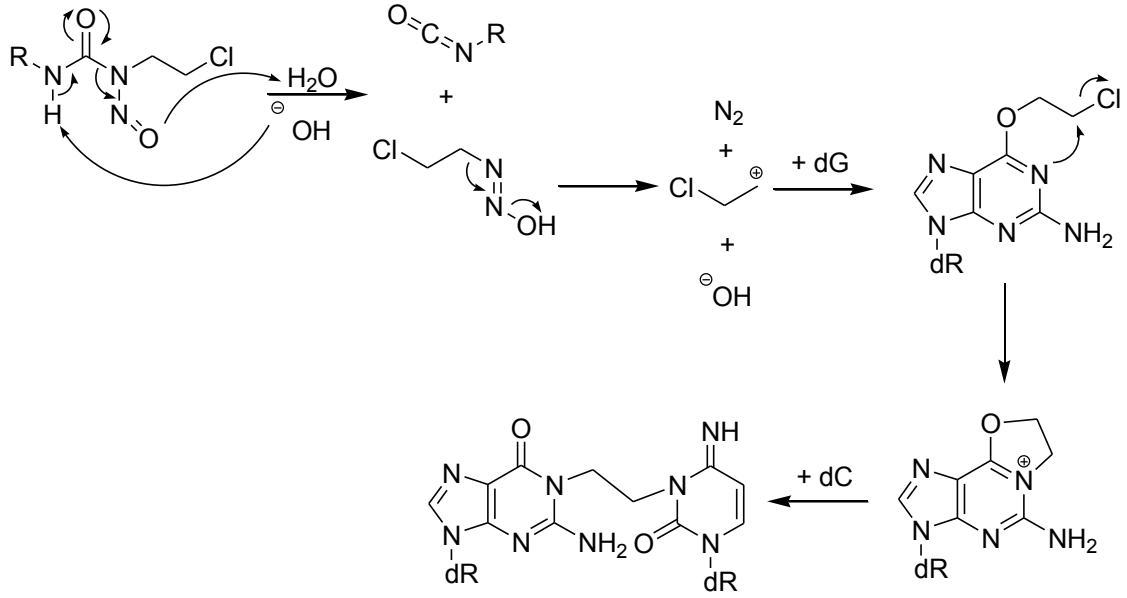
Despite its original purpose in chemical warfare during World War I, nitrogen mustard (mechlorethamine) was the first alkylating agent employed to combat cancer, dating back to 1946.(*59*) As shown in **Scheme 1.1**, the nitrogen mustards form DNA ICLs *via* reactions involving the displacement of the chlorine atom to form an aziridinium ion. This ion is involved in the first alkylation event on guanine and generated again to ultimately form the ICL.(*35*)



Scheme 1.1: ICL formation by nitrogen mustards, where R=Me for mechlorethamine (*7*)

The main ICL induced by nitrogen mustards tether the N7 atoms of dG in a 5'-GNC motif where the cross-link lies in the major groove of dsDNA.(*60*,*61*,*62*) Alkylation at the N7 atom of dG is unstable and for this reason the cross-links induced by nitrogen mustards are generally short lived with half-lives of one hour at room temperature under physiological conditions.(*63*)

Chemotherapeutic alkylnitrosoureas produce ICLs in DNA by a different mechanism, depicted in **Scheme 1.2.** ICL formation is catalyzed by base to initially yield a chloroethyl diazonium compound, which is responsible for the alkylation of DNA. The first alkylation event involves chloroethylation of the O^6 atom of dG. This is followed by cyclization to form an unstable five membered ring. To alleviate the formal positive charge, the cyclic structure is broken upon reaction with a nearby 2'-deoxycytidine (dC) at its N3 atom. Chloroethylating agents form ICLs between two complimentary nucleotides on opposite DNA strands (1,1 motif), which is unique to these agents.(*64*)



Scheme 1.2: ICL formation by nitrosourea compounds (*65*)

Alkyl sulfonates and their derivatives have also found use in the fight against cancer. The reactivity of alkyl sulfonates and their derivatives towards DNA nucleophilic centers stem from the stability of the leaving group as a result of the delocalization of the negative charge over three oxygen atoms. Interest in bi-functional alkyl sulfonates as chemotherapeutics originate from the discovery of busulfan (1,4-butanediol dimethanesulfonate) in the early 1950s.(*66*) Since its early stages busulfan was and is still

used to treat cases of leukemia, though its use is less common due to newer, less toxic agents.(*67*) Busulfan predominantly alkylates dG at the N7 atom. It is capable of forming cross-links in DNA though the nature of the cross-links are open for debate due to the length of the linker (6.0 Å) and the distance between the two nearest N7 atoms of dG on complementary strands (8.0 Å).(*68*,*69*,*70*) The general consensus is that busulfan predominantly generates intrastrand cross-links at 5'-GA-3' sites. To date, three dimethanesulfonates and one sulfomate have entered clinical trials, see **Figure 1.4** for structures.(*71*)

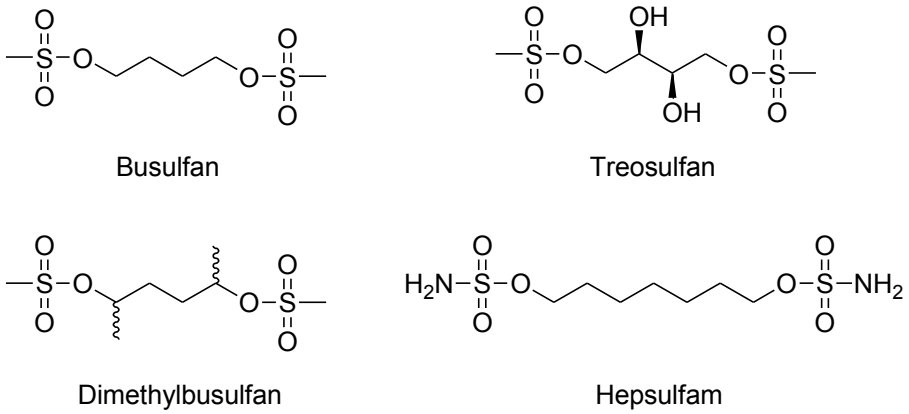
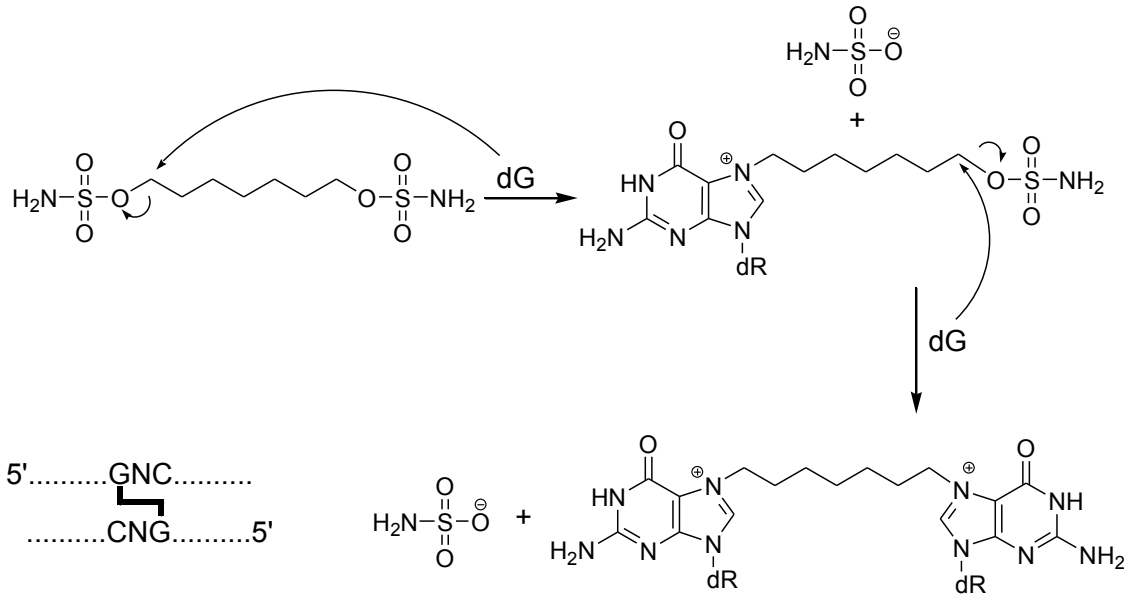


Figure 1.4: Structures of alkyl sulfonates and derivatives

Of the later generation alkyl sulfonates and derivatives generated, hepsulfam received the most attention due to its increased activity against cancers *in vitro*.(*72*,*73*,*74*) Hepsulfam alkylates DNA mostly at the N7 atom of dG to form mono-alkylation adducts involving the displacement of a sulfamate moiety by water. Moreover, ICL lesions were characterized, where the linker joined two N7 atoms of dG in a 5'-GNC motif (see **Scheme 1.3**), which was never isolated for the other alkyl sulfonates.(*75*)



Scheme 1.3: ICL formation by hepsulfam

1.3 DNA repair

The various sources of DNA damaging agents generally lead to a vast diversity of lesions, which require a range of repair mechanisms, each with their specific lesion detection and repair library. It is estimated that on a daily basis the average adult human requires 10^{16}–10^{18} repair events per day. The repair mechanisms are broken down into 6 general classes: base excision repair (BER), nucleotide excision repair (NER), mismatch repair (MMR), homologous recombination (HR), nonhomologous end joining (NHEJ) and direct repair (DR).(*76*,*77*)

1.3.1 Base excision repair

BER recognized and removes modification on the heterocyclic bases of DNA. Lesions include: O^6-MedG, thymine glycols, N3-MedA, 8-oxo-2'-deoxyguanosine, N7-alkyl dG derived formamidopyrimidines, deamination products and base mismatches.(*78,79*) BER requires multiple enzymes, as shown in **Figure 1.5**. The repair

mechanism is initiated by removal of the damaged base by a DNA glycosylase, which cleaves the glycosidic bond.(*80*) Various DNA glycosylases are present in the cell, each with specific substrate(s) they can repair. The glycosylase recruits an endonuclease to the DNA, where the endonuclease cleaves the damaged strand 5' of the newly formed apurinic/apyrimidinic (AP) site. Polymerase-β is subsequently recruited where it removes the 5'-deoxyribosephosphate with its AP lyase activity and incorporates the proper nucleoside where the damaged heterocycle was previously located. The repair cycle is concluded by the closing of the 5' nick by DNA ligase I.

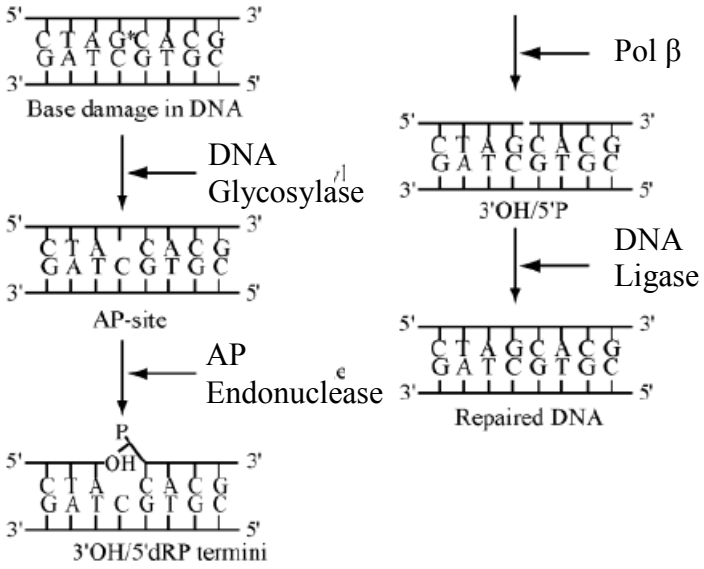


Figure 1.5: General short patch BER repair pathway (*81*)

1.3.2 Nucleotide excision repair

NER is responsible for the repair of most exogenous DNA adducts such as adducts produced when DNA is treated with cisplatin and UV light.(*82*,*83*,*84*) This repair pathway is versatile in its capability to eliminate a wide variety of damage ranging from the bulkier DNA adducts that cause distortion to the DNA duplex to small lesions such as O^6-MedG, O^4-alkyl dT and N^6-methyl dA.(*79*,*85*) NER recognizes a wide array of

distortions including: base mispairing, helix unwinding, helix bending and transcription blockage.(*86*,*87*)

As shown in **Figure 1.6**, the basic NER pathway begins with the recognition of the lesions. Incisions are then created on the damaged strand a few nucleotides 5' and 3' of the lesion where the incised DNA is excised out of the duplex. The generated gap is filled with new nucleotides by a DNA polymerase and the 3' nick is then sealed by a DNA ligase. The illustrated mechanism shows how NER can repair a wide array of lesions due to the incisions being far from the actual lesion. Indeed, very bulky lesions can easily be removed without inhibiting the repair process.

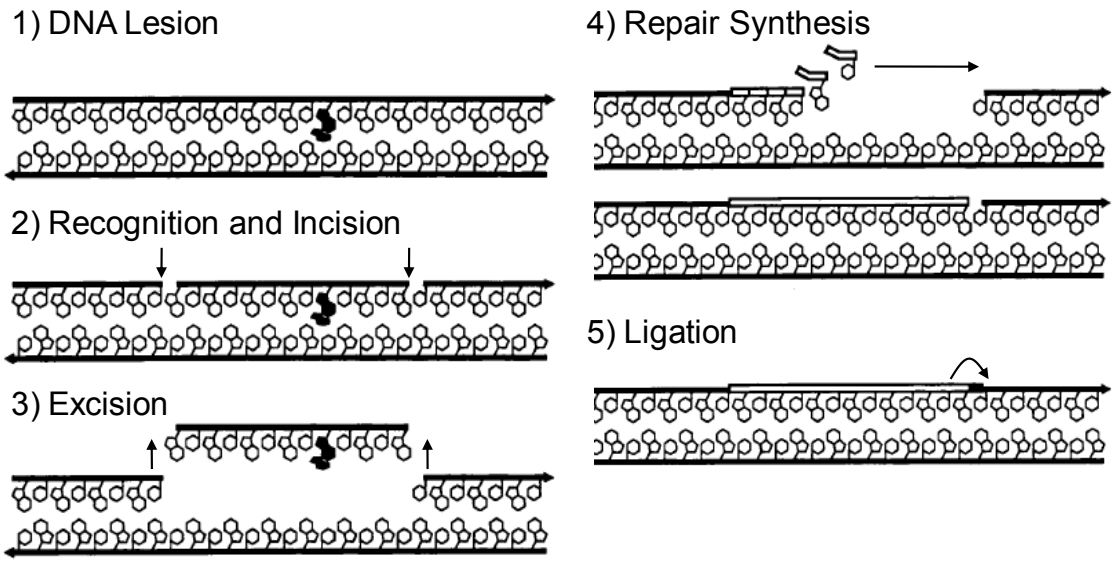


Figure 1.6: Overview of NER mechanism (adapted from *88*)

1.3.3 Mismatch repair

The main substrates for the MMR pathway are DNA base pair mismatches, which can arise from improper incorporation of nucleotides by polymerases, spontaneous mutations or by mutations brought about by alkylating agents.(*89*) Interestingly, the *E. coli* MMR machinery can repair any kind of base pair mismatch except for the dC•dC

mismatch.(*90*,*91*) *E. coli* DNA is methylated at the N^6 atom of dA in 5'-GATC sequences except for newly synthesized daughter DNA. Methylation of these 5'-GATC sequences allows MMR to distinguish between the template and the newly synthesized DNA. MMR requires three proteins specific for this pathway which are: MutS, MutL and MutH. Aside from these proteins, the pathway also requires a DNA helicase, an exonuclease, a single-stranded DNA binding protein, a polymerase and a ligase.(*92*) The repair process is initiated by a MutS homodimer, which recognizes the lesion as illustrated in **Figure 1.7**. This is followed by the recruitment of a MutL homodimer, which aids in the mismatch detection and is responsible for the subsequent binding of MutH to the complex. MutH, which recognized unmethylated DNA, is activated by MutS and MutL as a result of ATP hydrolysis. MutH cleaves the unmethylated strand at the 5' of the 5'-GATC sequence. A helicase subsequently unwinds the DNA and allows the exonuclease to cleave the newly nicked strand slightly past the mismatch at which point the single-stranded DNA binding protein caps the single strand to inhibit further DNA digestion. The newly generated gap is filled by a DNA polymerase and the 3' nick sealed by a ligase.

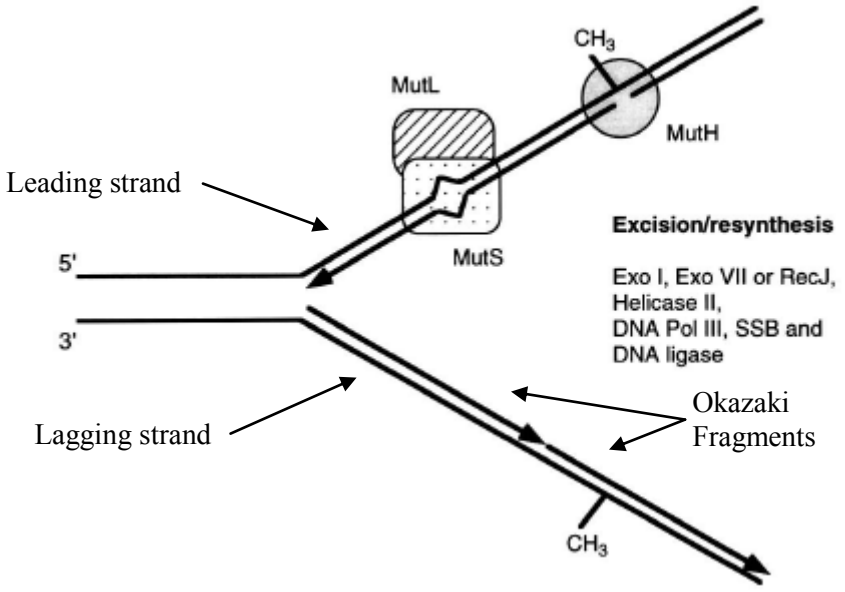


Figure 1.7: MMR pathway in *E. coli* (adapted from *89*)

1.3.4 Nonhomologous end joining

Ionizing radiation from X-rays or γ-rays, DNA damage from alkylating agents, replication errors and recombinational events are all sources of double strand breaks (DSBs).(*93*) These DSBs are repaired by the NHEJ repair pathway.(*94*) In mammalian cells, after formation of a DSB a Ku heterodimer, composed of Ku70 and Ku80 binds the ends of the dsDNA created. The Ku complex translocates away from the DNA ends which allows recruitment of DNA-PKcs, a serine/threonine protein kinase. This new complex, termed DNA-PK, promotes the formation of the 'synapsis', which occurs when the two DNA fragments are brought back together. This is followed by DNA processing by various enzymes to remove overhangs or DNA damage ensued in the DSB formation process. DNA ligase IV is recruited to the site of lesion and the DNA-PK complex released from the DNA. Ligation completes the repair cycle.(*95*,*96*)

1.3.5 Homologous recombination

HR repairs DSB in an error-free manner but unlike NHEJ requires a second copy of the genomic DNA entailing that the cell must be in the G2 or S phases where the sister chromatid is present (see **Figure 1.8)**.(*97*) Primary response to the DSB is initiated by activation of Ataxia-telangiectasia-mutated protein kinase (ATM) which phosphorylates Nijmegen breakage syndrome 1 (NBS1). Phosphorylated NBS1 in complex with RAD50 and MRE11 is thought to process the DSB using its 5'-3' exonuclease activity whereby 3' overhangs are created at the original DSB site.(*98*) RAD51 proteins polymerize as the site of the 3' overhang with the help of replication protein A (RPA) and RAD52, forming nucleoprotein filaments. The nucleoprotein filaments recruit the homologous sister DNA, which is used as a template by the polymerase. The nicks are later sealed by a ligase and resolvases restore the DNA to the duplex form resulting in the fully repaired DNA.(*99*)

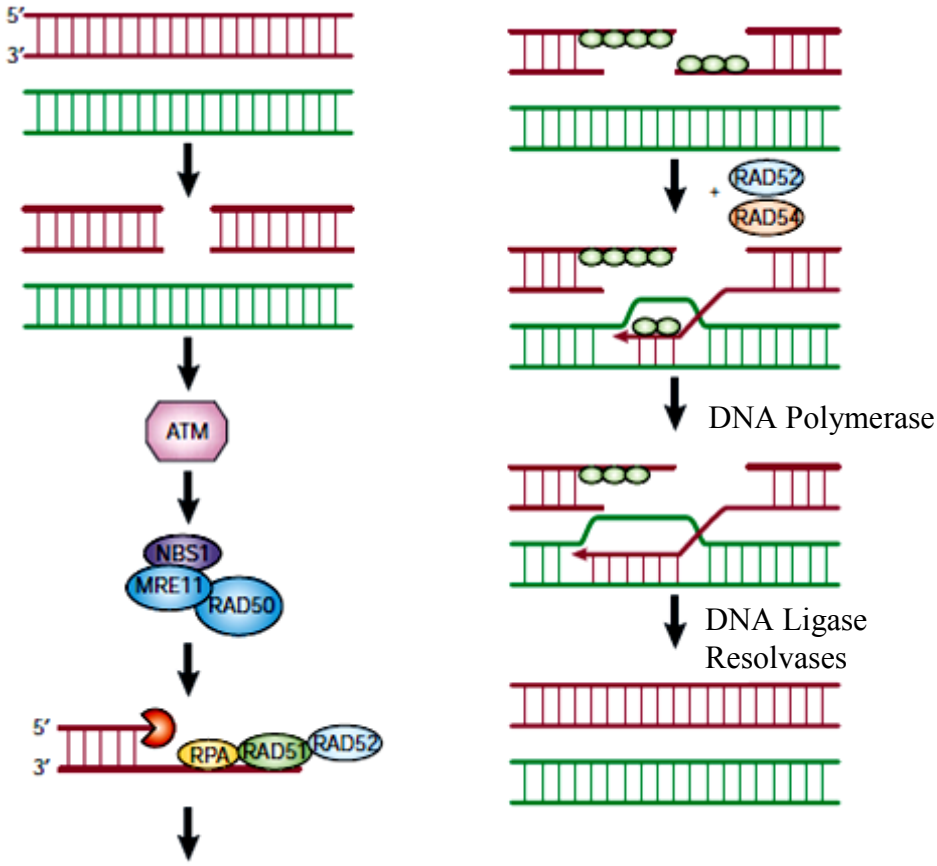


Figure 1.8: Homologous recombination pathway (*100*)

1.3.6 Direct repair

The direct repair pathway is unique since the original nucleotide remains after the repair event. Two direct repair pathway exist in mammals: oxidative demethylases/dioxygenases (ABH2 and ABH3) and AGT, which will be discussed later.

ABH2 and ABH3 repair methylation and ethylation at the N1 atom of dA and methylation at the N3 atom of dC, which are known to be cytotoxic due to DNA replication inhibition.(*101*,*102*) ABH proteins require O_2, Fe(II) and 2-oxoglutarate to perform their functions. In the process, 2-oxoglutarate is decarboxylated to form succinate generating a hydroxymethylated or hydroxyethylated nucleotide. The

hydroxylated alkyl group is spontaneously released as formaldehyde or acetaldehyde, producing the native nucleotide.

1.4 O^6-Alkylguanine-DNA alkyltransferases

Organisms have evolved AGT proteins to repair the highly mutagenic alkyl lesions at the O^6 atom of dG and O^4 atom of dT by a single step reaction without the need for cofactors or other proteins. The single step reaction irreversibly alkylates the protein making each protein able to accept only one alkyl group per active site.(*103*) Once alkylated, the protein is degraded by the ubiquitin pathway.(*104*) The importance of AGT for maintaining genomic integrity is evident considering its presence in Eubacteria, Archea and Eukaryotes.(*105*) Interestingly, the various AGT homologues display diverse substrate specificities.

All AGT have the totally conserved -PCHR- active site motif, an asparagine hinge and a DNA recognition helix, as shown in **Figure 1.9**. The alkyl group is transferred from the point of lesion located on an exocyclic oxygen to the central Cys residue of the active site motif by an S_N2 type reaction. The conserved asparagine hinge is responsible for protein degradation due to a rotation about this hinge that highly destabilizes the alkylated form of the protein. AGTs bind the minor groove of DNA rather than the major groove due to their DNA recognition helix.(*106*)

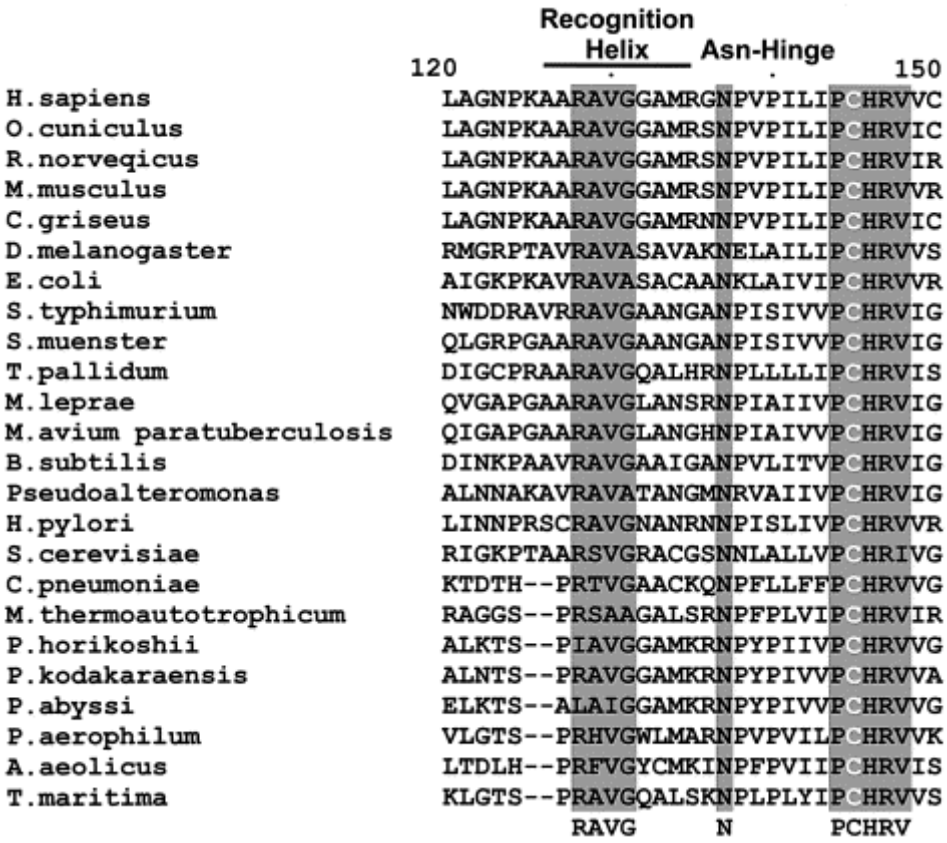


Figure 1.9: Sequence alignment of active site, asparagine hinge and recognition helix of AGT homologues (*106*)

1.4.1 AGT homologues

```
hAGT    1   --MDKDCEMKRTTLDSPLGKLELSGCEQGLHEIKLLGKGTSAADAVEVPAPAAVLGGPEP
Ada     181 RHGGENLAVRYALADCELGRCLVAESERGICAILLGDDDAT-----LISELQQMFPAAD-
OGT     1   -----MLRLLEEKIATPLGPLWVICDEQFRLRAVEWEEYSE-----RMVQLLDIHYRKEG

hAGT    59  L----MQCTAWLNAYFHQPEAIEEFPVPALHHPVFQQESFTRQVLWKLLKVVKFGEVISY
Ada     235 ----NAPADLMFQQHVREVIASLNQRDTPLTLPLDIRGTAFQQQVWQALRTIPCGETVSY
OGT     51  YERISATNPGGLSDKLREYFAGNLS--IIDTLPTATGGTPFQREVWKTLRTIPCGQVMHY

hAGT    115 QQLAALAGNPKAARAVGGAMRGNPVPILIPCHRVVCSSGAVGNYSGGLAVKEWLLAHEGH
Ada     291 QQLANAIGKPKAVRAVASACAANKLAIVIPCHRVVRGDGSLSGYRWGVSRKAQLLRREAE
OGT     109 GQLAEQLGRPGAARAVGAANGSNPISIVVPCHRVIGRNGTMTGYAGGVQRKEWLLRHEGY

hAGT    175 RLGKPGLGGSSGLAGAWLKGAGATSGSPPAGRN
Ada     351 NEER----------------------------
OGT     169 LLL----------------------------
```

Figure 1.10: hAGT, Ada and OGT sequence alignments

The first AGT protein discovered was isolated from *E. coli* cells and given the name Ada, short for "Adaptive Response to Alkylating Damage".(*107*) Its discovery

stemmed from *E. coli*'s ability to withstand greater levels of *N*-methyl-*N*'-nitro-*N*-nitrosoguanidine (MNNG) levels when the cells were pretreated with a lower dose of the drug.(*108*,*109*) This resistance was later attributed to Ada's ability to repair O^6-MedG.(*110*,*111*) Ada can repair O^6-MedG, O^6-ethyl dG, O^4-MedT, O^4-ethyl dT and the S_p diastereomer of the methylphosphotriester.(*33*,*34*,*112*,*113*,*114*)

This 354 amino acid protein has three discrete functions. The N-terminal domain of the protein, Ada-N, is responsible for two of these activities. This 20 kDa domain is responsible for the regulation of the adaptive response pathway and for the repair of the S_p diastereomer of the methylphosphotriester adduct.(*115*,*116*)

The third activity of the Ada protein, which involves the repair of O^6-MedG, O^6-ethyl dG, O^4-MedT and O^4-ethyl dT, is attributed to the C-terminal domain. This protein whose substrate range is limited, shows faster repair of O^6-MedG than it does O^4-MedT.(*117*) Moreover, the repair of O^6-ethyl dG is 1000 times slower than O^6-MedG but is still more than 10 times faster than the repair of O^4-MedT.(*34*) Clearly, DNA damage repair by Ada-C is not only affected by the nature of the alkylated atom but is also very sensitive to steric effects.

The crystal structure of Ada-C has been solved and resembles the true nature of AGT proteins with the telltale -PCHR- motif in its active site, unlike its N-terminal domain.(*118*) Ada Cys 321 (or Cys 146 for Ada-C domain) is the alkyl acceptor group. Normally, the pK_a of a Cys side chain is ~8.4 such that at physiological pH the thiolate anion should not be predominant. The pK_a of Cys 146 is reduced due to a nearby hydrogen bond network involving the His residue from the -PCHR- motif, a Glu and water molecule. Rough experimental data places the pK_a of Cys 146 at ~5.1.(*119*)

DNA damage recognition by Ada-C has been attributed, through cross-linking experiments, to weakened hydrogen bonding between complimentary nucleotides.(*120*) This mode of recognition is efficient since the lesions detected by Ada-C disrupt normal Watson-Crick base pairing between either T:A in the case of O^4-alkyl dT or G:C in the case of O^6-alkyl dG.

E. coli possess a second copy of an AGT protein called OGT.(*32*) The necessity for a second AGT arises from the low basal level of Ada prior to its adaptive response. Before the adaptive response Ada is present at a level of 1-2 molecules per cells. After the adaptive response, Ada is present at approximately 3000 molecules per cell. OGT is present at roughly 30 molecules per cell and is constitutively expressed.(*121*,*122*) OGT is therefore required for *E. coli* cells to survive mild initial alkylating damage.

OGT has shown the greatest activity at eliminating O^4-MedT from DNA. Its ability to repair this lesion is virtually identical to that of O^6-MedG.(*117*) Unfortunately, OGT has not been crystallized and hence its increased rate of repair of O^4-MedT cannot be concretely explained. The current belief is that the shape of the active site of OGT differs from most other AGT proteins, which would explain this increased activity.(*123*) To date, OGT's known substrates include: O^6-MedG, O^6-ethyl dG, O^6-benzyl dG (O^6-bndG), O^6-*n*-propyl dG, O^6-*n*-butyl dG, O^6-heptyl-7-ol dG, a dG nucleoside linked through its O^6 atom to an oligonucleotide through the O^6 atom of a dG by a heptamethylene linker and O^4-MedT.(*34*,*124*,*125*)

Like Ada-C, substrate repair by OGT is hindered by increasing the size of the lesion. The steric effect is less pronounced in OGT as observed by comparing repair rates of O^6-MedG and O^6-ethyl dG. For instance, OGT and Ada-C have virtually identical

repair rates for O^6-MedG while OGT is 173 times more efficient at repairing O^6-ethyl dG damage.(*34*)

hAGT is the most studied AGT. Like OGT, hAGT is constitutively expressed, though its levels vary depending on the tissue. The liver displays the highest levels and the brain the lowest.(*65*) Its known substrate range is more vast than the other AGTs due to the increased interest in this system. hAGT has the ability to repair large lesions at the O^6 atom of dG, which has not been observed in other homologues. Its known substrates are depicted in **Figure 1.11**.(*126*,*127*,*128*,*129*) Its inability to repair O^4-MedT efficiently distinguishes this protein from the *E. coli* homologues.(*117*,*130*,*131*)

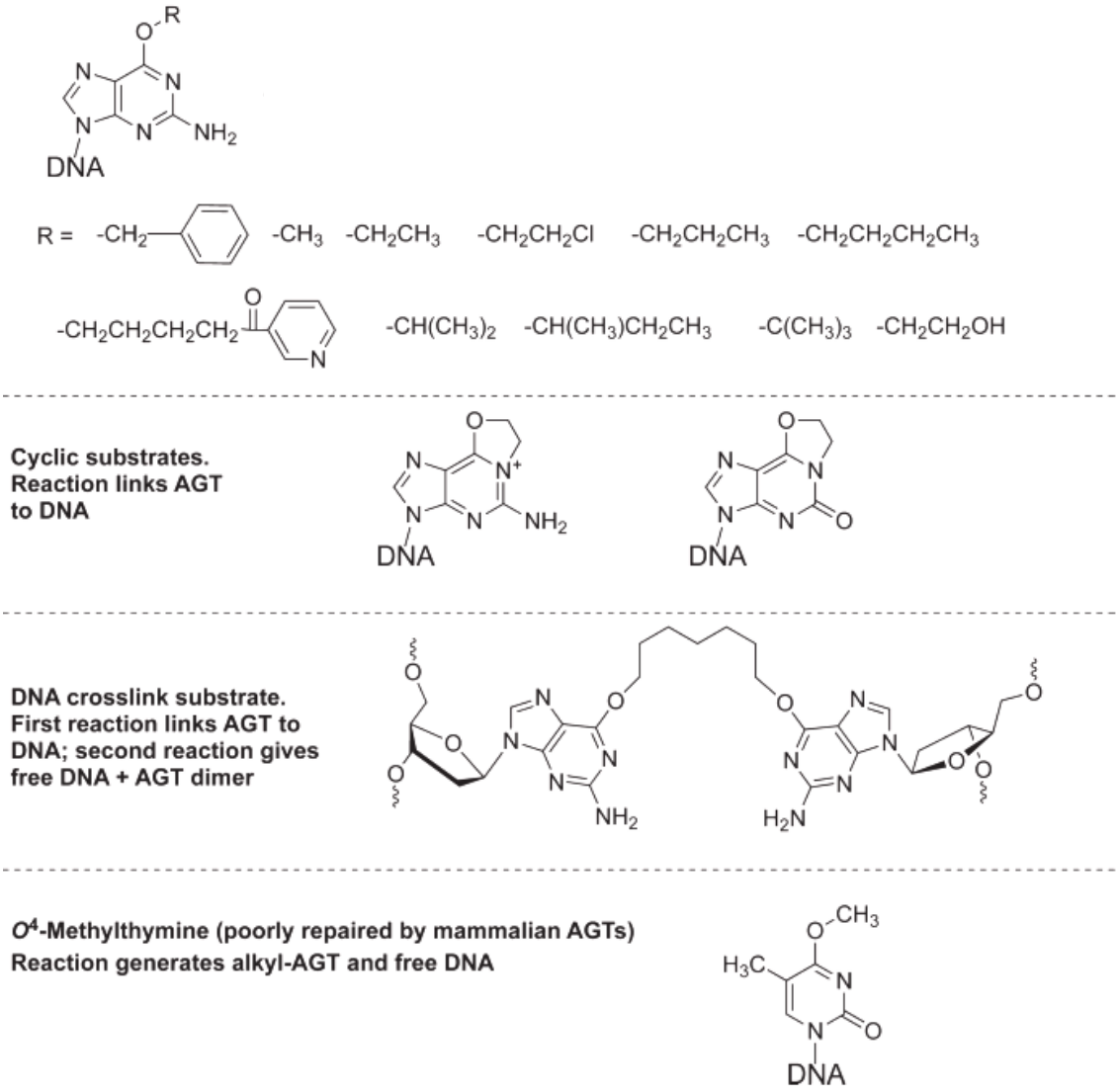


Figure 1.11: Known hAGT substrates (adapted from *132*)

In vivo, hAGT can recognize and bind O^4-MedT lesions but cannot remove them efficiently.(*133*) In effect, hAGT physically shields the lesion from the NER machinery rendering it counterproductive.(*123*) hAGT variants with increased repair capabilities of O^4-MedT have been investigated. The Loeb group employed a random sequence mutagenesis approach to generate multiple variants followed by a functional complementation assay to obtain a variant with 8 alterations (C150Y, S152R, A154S,

V155G, N157T, V164M, E166Q, and A170T). This variant displayed a rate of repair of O^4-MedT that was roughly 11.5 times that of native hAGT.(*134*,*135*) The Pegg group used a method that resembles a mix of rational design (based on information from PDB 1T38) and sexual PCR (horizontal gene transfer from OGT to hAGT) to generate various hAGT-OGT chimeras.(*123*) Their most promising construct harboured 8 alterations (V149I, C150G, S151R, S152N, A154T, V155M, G156T and N157G) and displayed a 47-fold increase in rate of repair of O^4-MedT.

O^6-benzylguanine (as a free base), a purely synthetic molecule, has shown great variation in repair rates between homologues. Ada-C has shown no repair of this substrate whether as a free base or in an oligonucleotide due to steric congestion in the active site. Site-directed mutagenesis studies revealed that the W336A and A316P double mutant of Ada-C is capable of repairing O^6-benzylguanine and O^6-BndG in DNA.(*127*,*136*) O^6-benzylguanine was repaired more rapidly by this variant in the presence of spectator DNA, indicating a conformational change upon DNA binding helps the Ada-C perform its repair function.

OGT showed great repair activity of O^6-BndG in an oligonucleotide while the repair of the free base was more modest. Finally, hAGT showed a high level of repair of O^6-benzylguanine and O^6-BndG in an oligonucleotide. hAGT repaired O^6-BndG in DNA with similar ease as OGT but was 400 times more efficient at repairing O^6-benzylguanine than OGT.(*137*) Interestingly, competition experiments revealed that hAGT repairs O^6-BndG more efficiently than O^6-MedG in short oligonucleotides.(*127*) Alternatively, O^4-benzylthymine was not repaired by hAGT, OGT nor Ada-C.(*138*)

hAGT repairs O^6-benzylguanine so readily that this molecule has undergone clinical trials for use during chemotherapeutic regimens. Indeed, increased levels of hAGT have been shown to cause cellular resistance to certain chemotherapeutic agents including the chloroethylating agents and nitrogen mustards.(*139*,*140*,*141*) Two hAGT inhibitors have undergone clinical trials to date, O^6-benzylguanine and O^6-(4-bromothenyl)guanine, see **Figure 1.12** for structures.(*142*,*143*) The later molecule has shown an increased hAGT inhibitory effect over O^6-benzylguanine, where its I_{50} is 0.003 as opposed to 0.04 μM.(*144*)

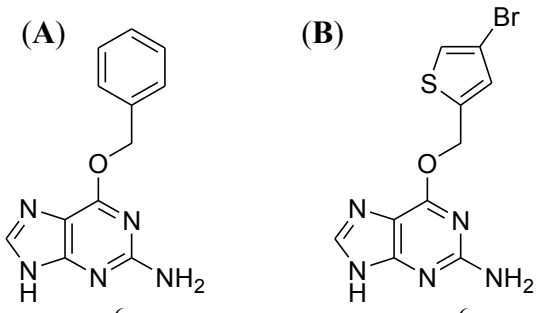


Figure 1.12: Structures of (**A**) O^6-benzylguanine and (**B**) O^6-(4-bromothenyl)guanine

AGT homologues also show dramatic differences in the processing of ICL lesions. AGT repair studies conducted with ICL DNA, where the cross-link tethered two O^6-atoms of dG in a directly opposed motif (1,1 orientation) by a 7 or 4 methylene chain, were investigated.(*145*,*146*) hAGT repaired both ICL DNAs, while Ada-C and OGT were unable to do so. It was proposed that the increased size of the hAGT active site with respect to the other homologues was responsible for this unique property. This hypothesis was further supported by the observed lack of repair with the smaller active site containing P140K hAGT variant.

1.4.2 hAGT structure

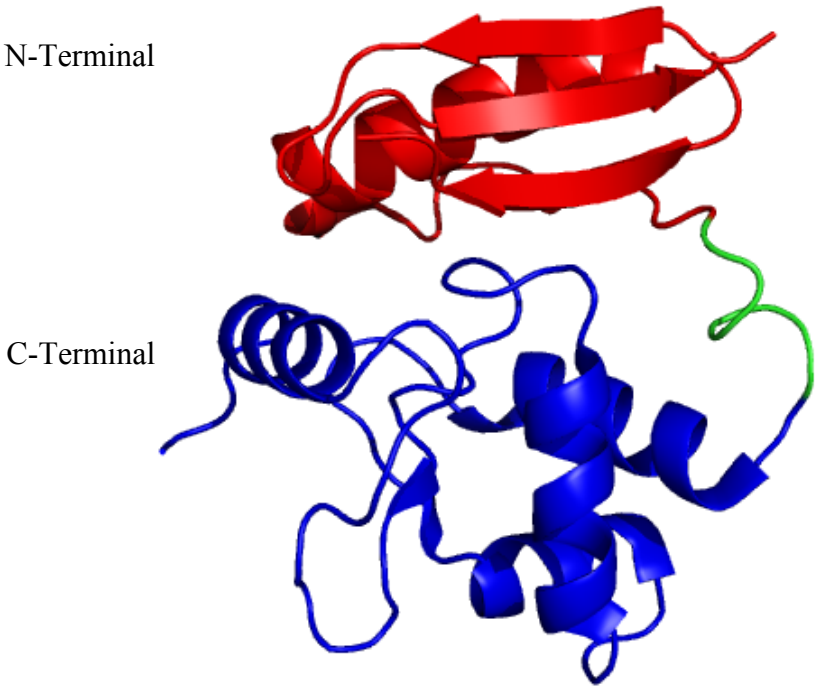


Figure 1.13: Crystal structure of hAGT (PDB 1T38)

hAGT is a 21 kDa monomeric protein.(*147*) It consists of two distinct domains, as depicted in **Figure 1.13**, that when expressed separately show no activity.(*105*) The N-terminal domain (residues 1-85) provides structural stability to the protein. This domain is responsible for the binding of Zn^{2+}, a known cofactor for higher eukaryote AGT. The binding of Zn^{2+} is believed to increase structural stability by causing better packing between the N- and C-domains due to the interaction of Helix 3 and the 3 β-strands of the N-terminal domain *via* the ion, as exemplified in **Figure 1.14**. Zn^{2+} is coordinated by the protein by Cys 5, Cys 24, His 29 and His 85 residues more than 20 Å away from the active site.(*148*) Zn^{2+} bound hAGT has shown an increase in repair of ~60-fold for both O^6-benzylguanine and methylated DNA. This increase in activity correlates with the observed decrease in pK_a of the active site Cys for the holoprotein versus its apo

form.(*119*) Zn^{2+} binding by hAGT stabilizes the protein, as assessed by urea denaturation experiments, supporting the notion that the role of Zn^{2+} is purely structural.(*149*)

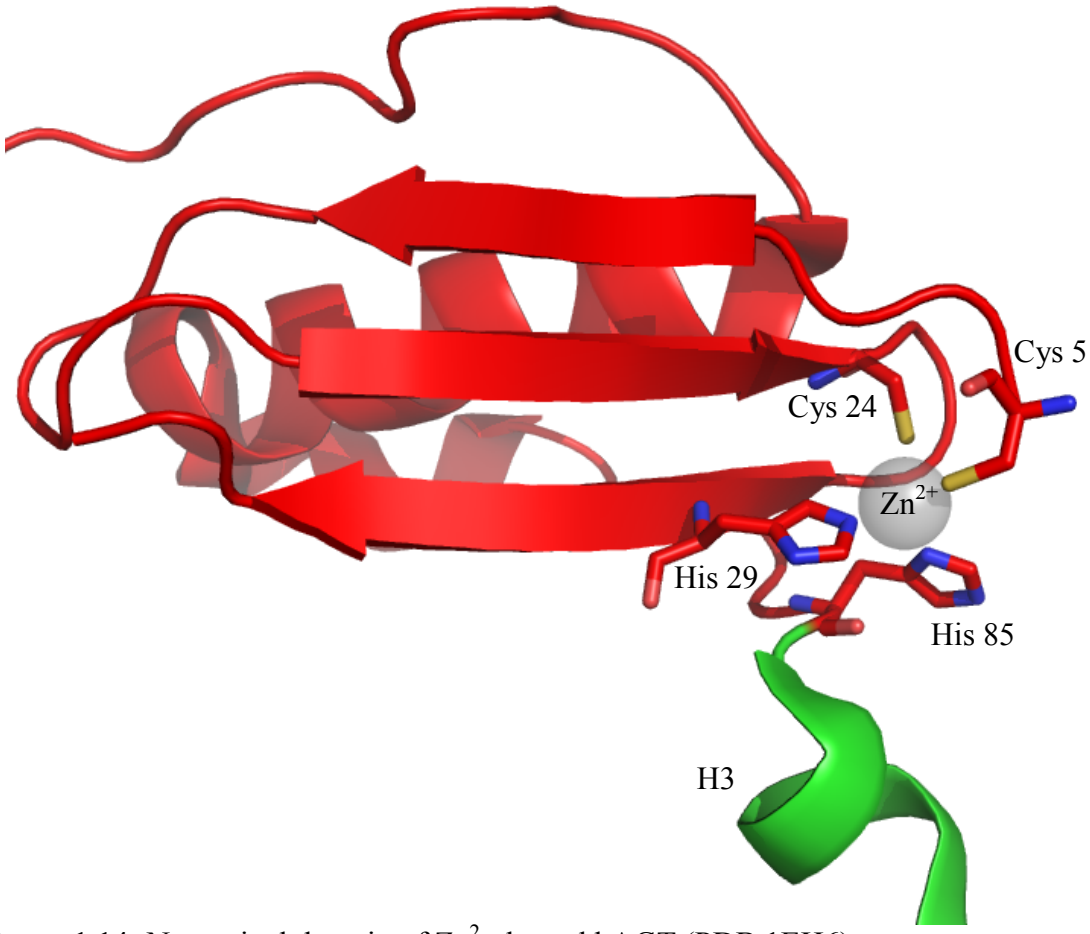
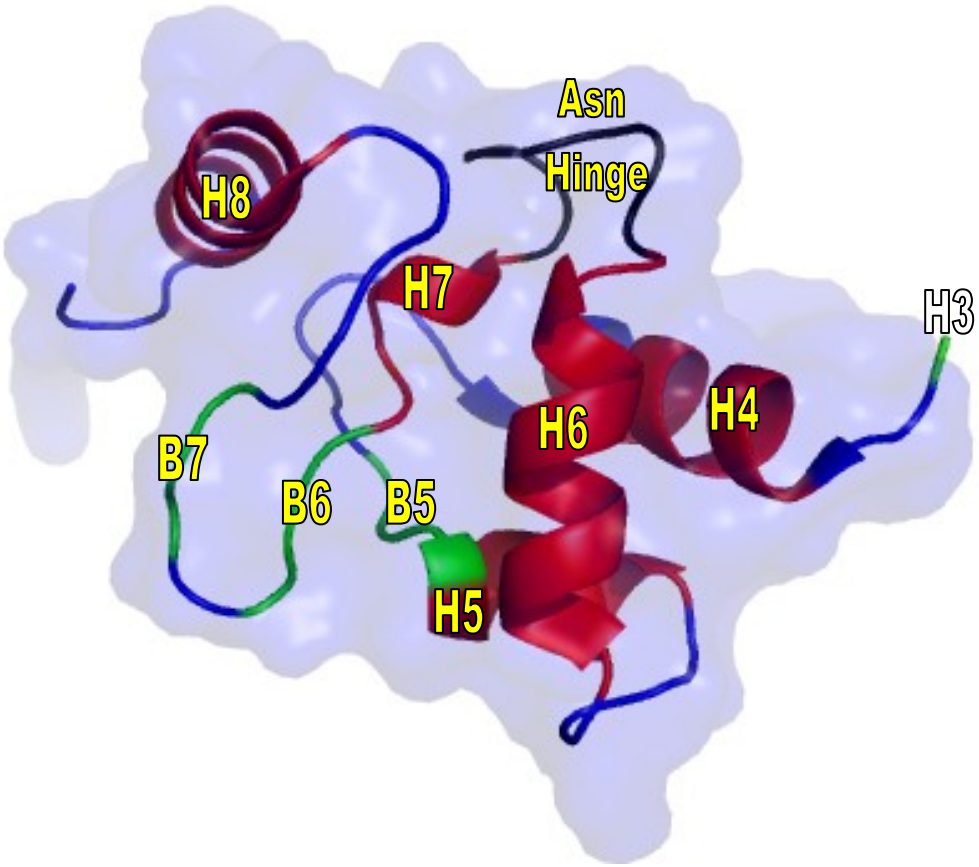


Figure 1.14: N-terminal domain of Zn^{2}+ bound hAGT (PDB 1EH6)

The C-terminal domain (residues 92-176), shown in **Figure 1.15**, contains the alkyltransferase activity and all the residues that make up the binding pocket as well as those conferring DNA-binding capabilities. This domain is absolutely conserved in all homologues as opposed to the highly variant N-terminal domain.

(**A**)



(**B**)

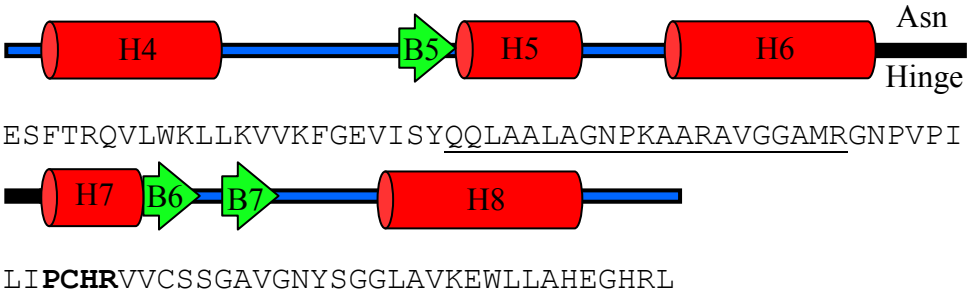


Figure 1.15: C-terminal domain of hAGT (PDB 1QNT). (**A**) Cartoon representation and (**B**) secondary structure representation

AGT binds DNA through the use of a helix-turn-helix motif (HTH). HTH motifs are usually reserved for major groove binders, whereas this motif in hAGT binds to the minor groove. The hAGT HTH motif is composed of Helix 5, a loop and Helix 6, underlined in **Figure 1.15B**.(*150*) Helix 6, also known as the "recognition helix", is

responsible for interacting with DNA. It partially interacts *via* its small hydrophobic residues that pack well into the minor groove. Also, Ala 127 and Ala 129 help widen the minor groove by clashing with the sugars on opposite strands. Finally, Arg 128, also called the "arginine finger", plays the critical role of flipping out the damaged nucleotide and maintaining the DNA conformation.(*151*) These 3 properties of Helix 6 allows for the minor groove to expand by 3 Å relative to B-form DNA making it ~15 Å wide while bending the duplex by 15°. Once the modified nucleotide is flipped out of the duplex an orphan dC is formed and is stabilized by a hydrogen bond by Arg 128.

Downstream of the HTH motif is the asparagine hinge. The Asn hinge is comprised of two tight turns which contains Asn 137. Asn 137 makes hydrogen bonds with the main chains of Val 139 and Ile 143 and the thiol group of Cys 145. The Asn hinge also interacts with the N-terminal domain and is responsible for roughly 40% of the surface buried area between the domains.(*150*) Upon alkylation of the protein a steric clash is produced between the Cys 145 thiol and Asn 137 causing the hydrogen bond network that make up the Asn hings to be disrupted.(*148*) This causes the hinge to open and expose hydrophobic areas of the protein. This conformational change in turn leads to ubiquitination of the protein and ultimatly degradation of alkylated protein by the proteosome.(*104*)

Opening of the Asn hinge exposes the 98-VLWKLLKVV-106 sequence and allows hAGT to bind to the estrogen receptor. This inhibits the binding of the estrogen receptor's natural partner, steroid receptor coactivator-1. The steric blocking of the estrogen receptor thus causes a reduction in cell proliferation rates.(*152*)

Lys 125, Ala 126, Ala 127 and Arg 128 (KAAR) make up a nuclear localization signal allowing hAGT to be targeted to the nucleus where RNA transcription occurs.(*153*) This motif is similar to that of histone H1 (KAAK). RNA polymerase does not stall at O^6-MedG sites and introduces mutations in the RNA. It is therefore essential that the DNA damage be repaired prior to transcription. As a result, hAGT must be concentrated in the area of transcription to remove the lesions as rapidly as possible.(*154*) The current hypothesis stipulates that hAGT binds vacated areas of DNA where histone H1 is absent such as the areas of DNA that undergo transcription. This partly explains how hAGT can detect lesions in the 3 billion bp human genome while having only a modest binding preference (~3-fold) for O^6-MedG over dG.

The highly conserved -PCHR- active site motif is present on Helix 7, a 3_{10} helix. Cys 145 of the -PCHR- motif, known to be the catalytic residue, resides in a ridge measuring 7 x 9 x 14 Å which corresponds to the binding pocket where the alkylated nucleotide is inserted for repair.(*106*) Tyr 114, Gly 131, Met 134, Asn 137, Pro 140, Ile 143, Pro 144, Cys 145, His 146, Arg 147, V148, Val 149, Val 155, Gly 156, Asn 157, Tyr 158, Ser 159 and Gly 160 make up the binding pocket.(*148*)

The benzylated form of hAGT has been crystallized. The structure shows how the adduct stacks in the active site between the side chain of Pro 140, the C_β of Ser 159 and contacts the side chain of Tyr 158 *via* a hydrophobic interaction as observed in **Figure 1.16**.

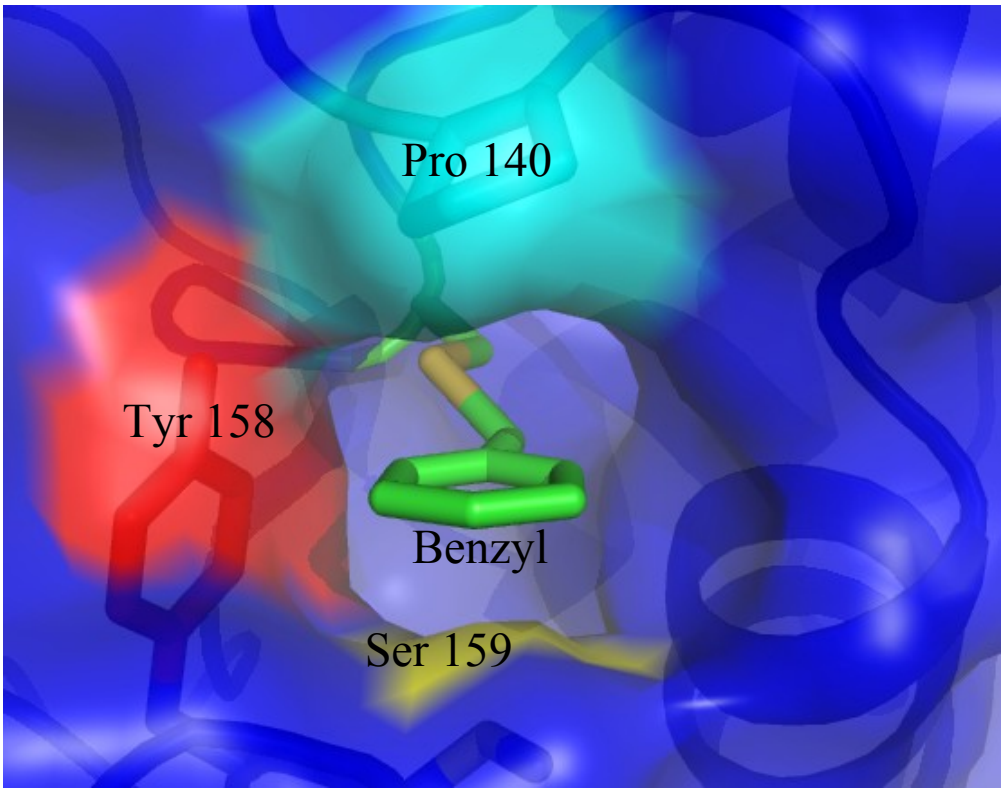


Figure 1.16: Interaction of benzyl group in hAGT binding pocket (PDB 1EH8)

hAGT recognizes dG and O^6-MedG over the other natural nucleotides due to steric factors and specific hydrogen bonds. O^6-alkyl dG fits specifically in the active site due to steric factors imposed by Met 134, Val 155, Gly 156, Asn 157, Tyr 158, Ser 159 and Gly 160 residues that form the proper hydrophobic ridge. The exocyclic amine of the nucleotide hydrogen bonds with the carbonyl groups of Cys 145 and Val 148. The hydroxyl group of Tyr 114 donates a hydrogen bond to the N3 atom of guanine while the amide nitrogen of Ser 159 donates a hydrogen bond to the O^6 atom. In summary, the steric factors allow a better fit for purine bases while the hydrogen bond network, as shown in **Figure 1.17**, allows hAGT to discriminate between guanine and adenine.(*150,155*)

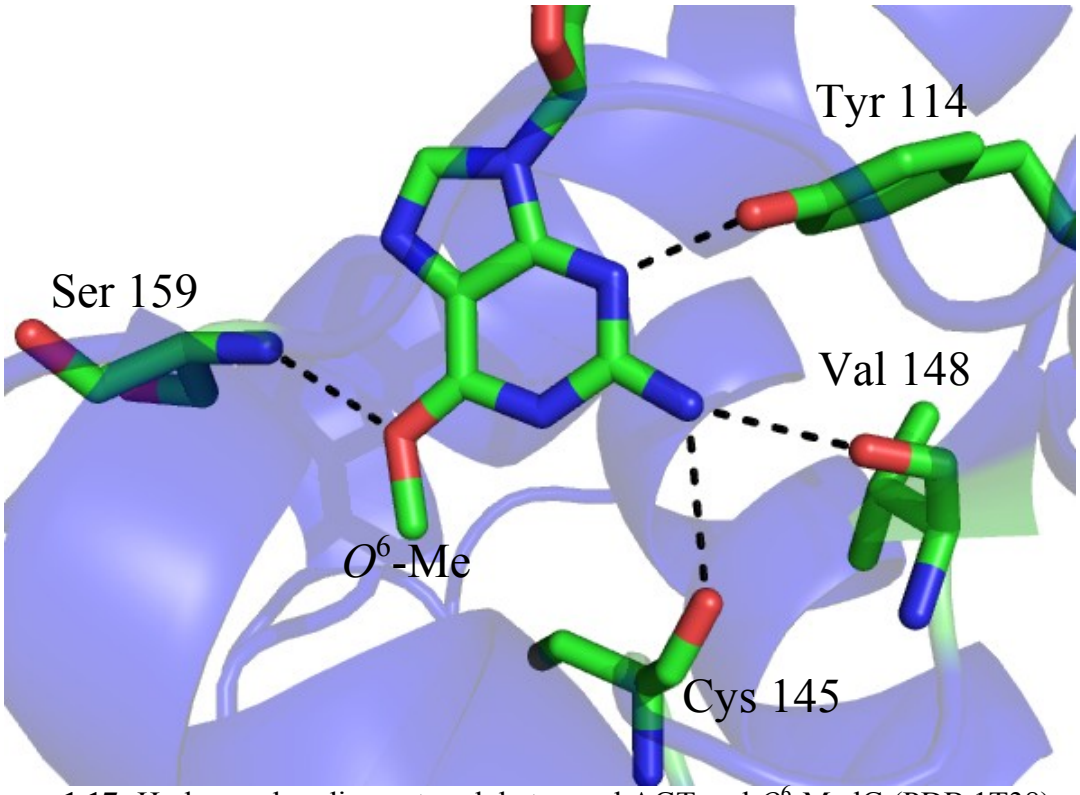


Figure 1.17: Hydrogen bonding network between hAGT and O^6-MedG (PDB 1T38)

The interaction of O^4-MedT and hAGT has not been studied as thoroughly as that of O^6-MedG resulting in a poor understanding of how hAGT can repair this adduct. The co-crystallization of hAGT with a modified oligonucleotide containing N^4-*p*-xylylenediamine-2'-deoxycytidine has given some insight as to how hAGT interacts with pyrimidine nucleotides. The only interaction observed with this probe was that of Tyr 114 donating a hydrogen bond to the O^2 atom of the modified dC as seen in **Figure 1.18**. It should be noted that Ser 159 was not built into the final structure indicating that flexibility of the sheet on which this residue resides was observed supporting the lack of a hydrogen bond between Ser 159 and the probe.(*151*) The O^2 atom of dC is also present in dT, indicating that the Tyr 114 interaction should be present with O^4-MedT.

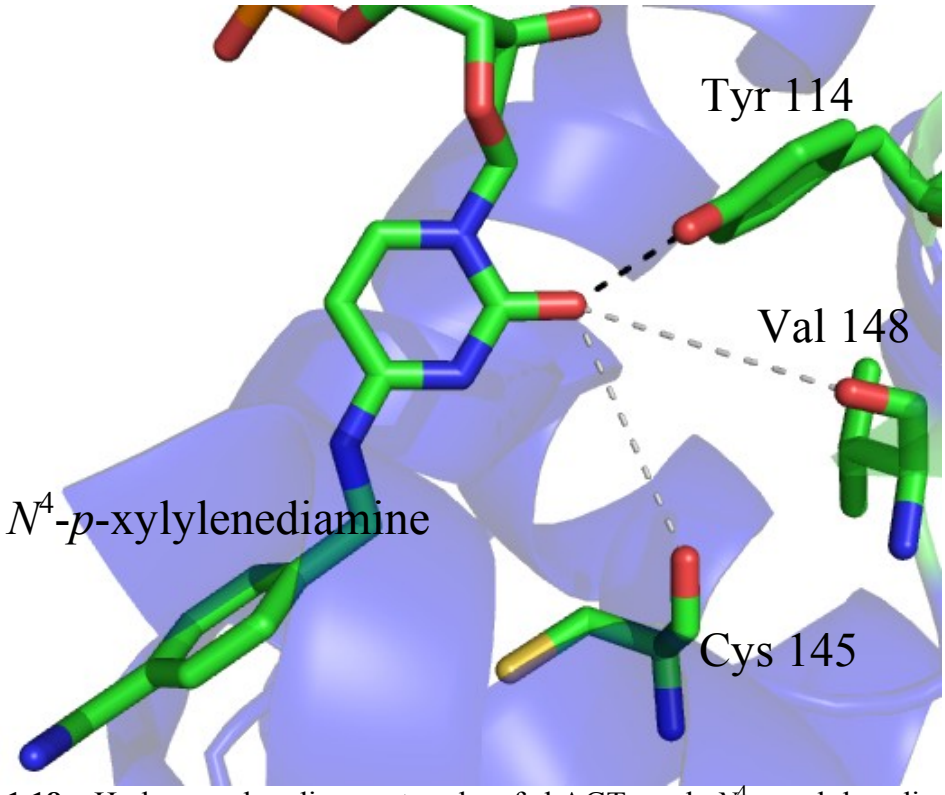


Figure 1.18: Hydrogen bonding network of hAGT and N^4-*p*-xylylenediamine-2'-deoxycytidine (PDB 1YFH)**.** Hydrogen bonds shown with black dashes represent those that are present with O^6-MedG and 1YFH, while grey dashes depict those found only with O^6-MedG.

Figure 1.19 displays an overlay of hAGT co-crystallized with O^6-MedG (PDB 1T38) and O^4-MedT (generated from N^4-*p*-xylylenediamine-2'-deoxycytidine from PDB 1YFH). The first apparent difference between both structures is the distance between the thiol receiving group from Cys 145 and the α-carbon of the alkyl group. For O^6-MedG, this distance is 2.68 Å while for O^4-MedT this distance is 4.00 Å. This marked difference explains why hAGT repairs O^6-MedG more rapidly than O^4-MedT. Upon substituting N^4-*p*-xylylenediamine-2'-deoxycytidine with O^4-MedT in PDB 1YFH it can be observed that the Ser 159 amide nitrogen can hydrogen bond with the O^4 atom with some movement of β-strands 6 and 7 to help coordinate the adduct. Tyr 114 hydrogen bonds with either the

N3 atom of dG or O^2 atom of dT with similar distances due to the virtually identical location of these atoms in the active site.

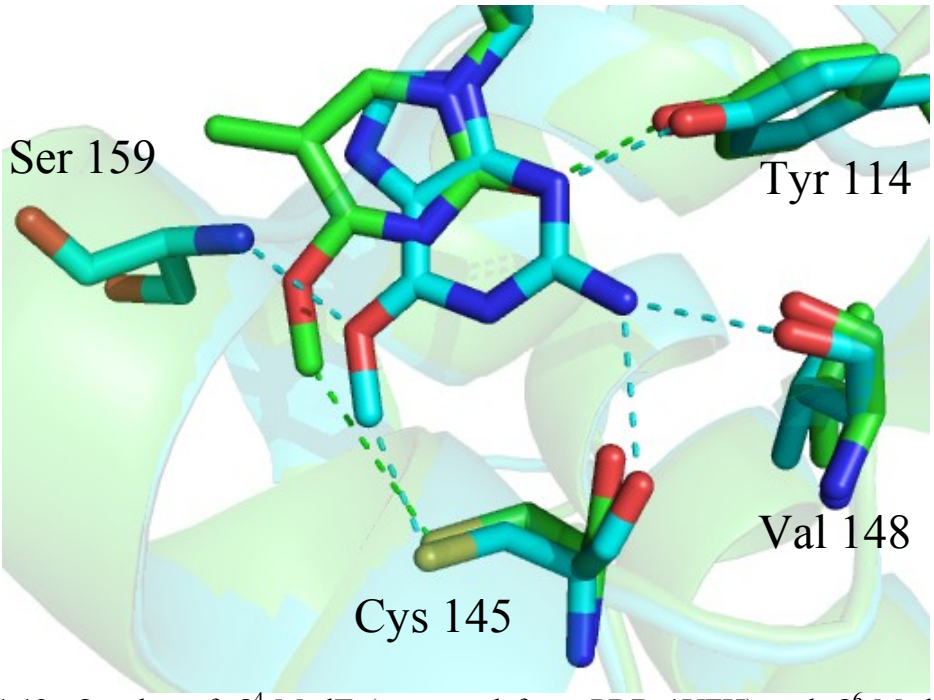
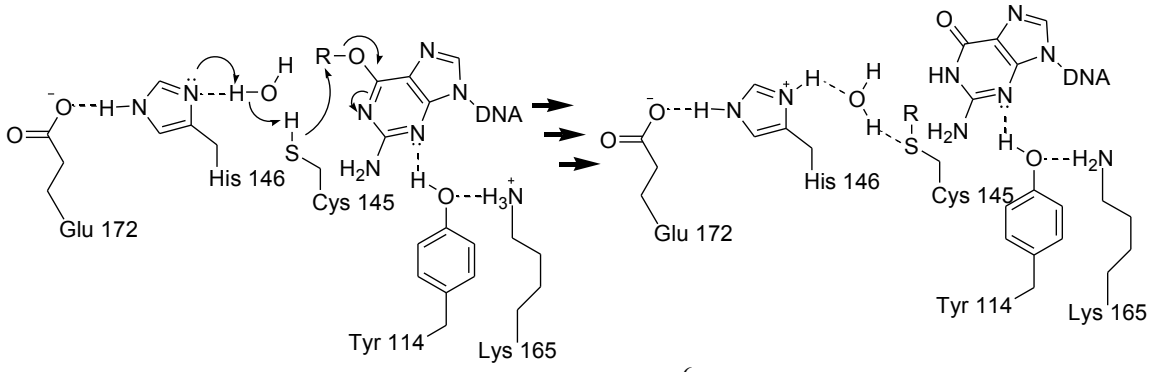


Figure 1.19: Overlay of O^4-MedT (generated from PDB 1YFH) and O^6-MedG (PDB 1T38) in hAGT active site. Interaction of O^6-MedG and hAGT shown in blue and O^4-MedT and hAGT shown in green.

1.4.3 hAGT repair mechanism

Prior to repair, the alkylated nucleotide must be flipped out of the DNA scaffold and placed into the active site of the AGT, where the catalytic triad composed of Glu-His-Cys can transfer the alkyl group to the Cys residue. This task is achieved partly by the binding of the AGT protein where an observed bend of 15° (for hAGT) is introduced into the dsDNA, widening the minor groove. The nucleotide is flipped out of the duplex by Tyr 114, due to steric and electrostatic interactions, as well as by Arg 128, which helps stabilize the irregular dsDNA structure as eluded to earlier.(*155*, *156*)

The Cys responsible for AGT activity is made highly nucleophilic as a result of the formation of its thiolate anion which is promoted by a Cys-His-Glu catalytic triad.(*148*) The flipped out alkylated nucleotide is oriented by the hydrophobic ridge of the AGT active site, thus promoting repair. In hAGT, Glu 172 activates His 146 *via* a hydrogen bond. His 146 abstracts a proton from a near by water molecule, which in turn abstracts a proton from Cys 145 to form the thiolate anion. The thiolate anion performs a nucleophilic attack on the α-carbon of the alkyl group through an S_N2 type mechanism, generating the alkylated protein and deprotonated form of the restored nucleotide.(*106*) Tyr 114 donates a proton to the nucleotide forming the fully repaired DNA and abstracts a proton from Lys 145, shown in **Scheme 1.4**.(*157*) The repaired DNA strand is then released by the protein.



Scheme 1.4: hAGT mediated repair mechanism of O^6-alkyl dG

According to quantum-mechanical/molecular-mechanical studies, hAGT mediated repair of O^6-MedG follows a distinctly different mechanism than the one usually accepted, noted above.(*158*) Unlike previously accepted, the proposed repair mechanism occurs in two distinct steps rather than one. The suggested model stipulates that the first step, being rate limiting, is the formation of the methyl carbocation, which is then followed by nucleophilic attack by the active site Cys residue.

1.4.4 DNA damage detection model

The mode by which AGT proteins are capable of locating damage in such a vast number of nucleotides as the genome is still a mystery. Three mechanisms have been proposed by Dr. He, which all assume that AGT binds the DNA at random sites prior to movement.(*1*) 1) AGT actively flips out every nucleotide into its active site as it migrates along the DNA until damage is detected. 2) The protein migrates along the DNA and flips out nucleotides that participate in weakened base pairs, which are intrahelical. 3) The protein migrates along the DNA and only captures transiently flipped out nucleotides (extrahelical) that arise from improper base pairing.

Cross-linking experiments were conducted with hAGT using a modified dC containing a disulfide moiety capable of forming a cross-link with the reactive Cys residue of the protein were undertaken to probe the proposed models.(*159*) hAGT could capture both damage that was transiently flipped out (hypothesis 3) and damage that was still intrahelical (hypothesis 2). Later experiments revealed that hAGT selectively flips out O^6-alkyl dG from the duplex, supporting the 2nd hypothesis.(*160*)

Findings inferred from the cross-linking experiments are in accordance with the structural data.(*151*,*155*) hAGT is documented by both crystallographic groups as imposing a bend in the DNA duplex upon binding. Such a stress on the DNA scaffold could hold a fine balance where only weakened base pairs will be disrupted, while the normal Watson-Crick base pairs remain unaffected. Such a "pinching" mechanism has been put forth for other DNA damage repair proteins, such as DNA glycosylases.(*161*)

The position of the modified nucleotide in the DNA sequence plays a role on how efficiently hAGT can perform its activity. Studies involving the repair of both O^6-MedG

and O^6-BndG at varying positions in short oligonucleotides revealed hAGTs poor capability to remove lesions that are 4 nucleotides or less from the 3' end in ssDNA. This effect is virtually eliminated when the damaged nucleotide is introduced into dsDNA, where the modified nucleotide must be 2 or more nucleotides from the 3' end for efficient repair. This positional effect is non-existent when the lesion is at the 5' end of the DNA (dsDNA or ssDNA).(*162*) hAGT binds the DNA in the same orientation when the lesion is found in ssDNA; while adding the complement strand allows hAGT to access lesions further down the 3' end, perhaps by interacting with the complement strand.

1.4.5 DNA binding by hAGT

Many binding experiments have been conducted with hAGT and DNA where dissociation constants (K_d) ranging from 300 nM to 3 μM are reported.(*163*,*164*,*165*) hAGT interacts with ssDNA and dsDNA with a similar affinities, yet hAGT repairs damage in dsDNA more efficiently than it does in ssDNA.(*162*,*166*,*167*)

The binding affinity of hAGT to unmodified DNA, whether in duplex or single stranded form, is length dependant.(*166*,*167*) Experimental evidence indicated that hAGT binds DNA with higher affinity when the nucleotide length is a multiple of 4 and when the DNA is longer than 11 nucleotides. A model was established by Dr. Fried to explain the DNA length dependence observed by hAGT, depicted in **Figure 1.20**.

The model proposes that hAGT occludes 4 nucleotides but spans 8. It also suggests that a strong contact exists between the *n* and *n*+3 protein on the DNA. Indeed, the model implies that few contacts are made between the *n* and *n*+1 or *n*+2 neighbours. This hypothesis was later validated with modelling, cross-linking and site directed mutagenesis experiments.(*168*,*169*) From the model, one can see that a DNA length >11

nucleotides would allow at least 4 proteins to bind to the DNA and thus allowing the *n* and *n*+3 contact to form. Lastly, hAGTs ability to occlude 4 nucleotides supports the notion that DNA length that are multiples of 4 nucleotides would promote tighter binding since all the DNA is occupied by the proteins favouring inter-protein contacts.

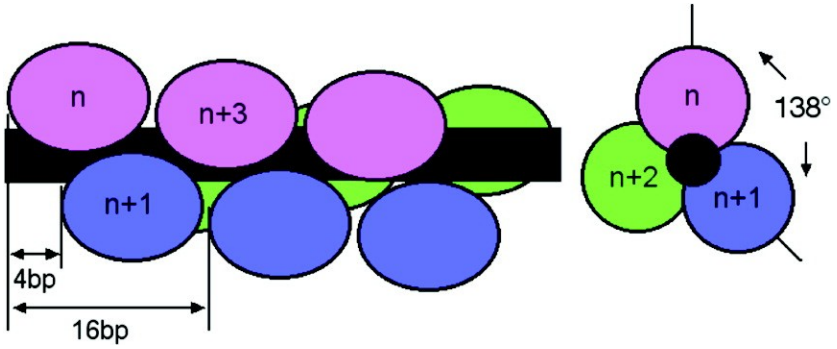


Figure 1.20: Model of hAGT binding to unmodified DNA. The proteins are depicted as filled ovals and DNA as a black rod. Left, side view of the complex, where the n and n+3 proteins are similarly coloured and in close contact. The model proposes a three-protein repeating motif where the n and n+3 proteins are on the same face of the DNA. Sequential proteins are rotated 138° with respect to their neighbours about the plane of the DNA. Right, cross-section of the DNA showing the protein arrangment about the axis of the duplex.(*167*)

The computational model produced by Dr. Fried, though rudimentary in nature, reveals other important properties of both hAGT-DNA and hAGT-hAGT interactions. According to the computational results, hAGT makes 3 ionic contacts with the phosphate backbone of dsDNA (Lys 125, Arg 128 and Arg 135) but only makes 2 with ssDNA (Arg 128 and Arg 135).(*168*) More importantly, the in *silico* data reveals a large protein-protein interface between the *n* and *n*+3 proteins, which is responsible for the high level of cooperativity observed upon binding of unmodified DNA by hAGT.(*163*,*166*,*167*) This interface spans ~1100 $Å^2$, which represents 70% of the protein-DNA interface (~1600 $Å^2$). The protein-protein interface is peppered with ionizable residues, which point to potential ionic interactions amongst proteins. Such ionic interactions would

promote the adoption of a head-to-tail pattern by hAGT monomers upon multimerization along the DNA scaffold.

Residues 1-7 and 163-169 of hAGT are important for the protein-protein interaction. Altering these two regions not only reduces the cooperativity of DNA binding but also reduces the binding affinity, highlighting the impact of cooperativity on binding affinity for this system.(*169*)

hAGT interacts with modified dsDNA differently than it does with unmodified DNA. The binding of hAGT to either a 24-mer or 26-mer harbouring a single O^6-MedG displays a binding stoichiometry of 2:1 hAGT:DNA.(*170*) This value is substantially less than those reported for unmodified 24-mer and 26-mer DNA (~6:1).(*167*) Binding experiments conducted with O^6-MedG containing DNA revealed the presence of 1:1 hAGT:DNA complex, which formed at lower hAGT concentration than the 2:1 complex. This lead to the finding that hAGT binds the O^6-MedG in the duplex first, directly or through migration from distal portions of the DNA. This is followed by a second hAGT molecule binding the DNA to form the cooperative protein-protein complex. This preliminary localization of hAGT to O^6-MedG prior to recruitment of another hAGT molecule was not observed with ssDNA, validating that modified dsDNA and ssDNA are processed differently by the protein.

Binding cooperativity takes on a new role when hAGT interacts with larger unmodified DNA fragments (1000 bp). Larger DNA substrates are not saturated with hAGT, unlike shorter oligonucleotides, but rather clusters of proteins are formed with up to 11 proteins per cluster.(*171*) The finite size of the hAGT clusters is believed to be a result of the deformation of the DNA upon hAGT binding. Every hAGT added to the

cluster adds a stress to the DNA structure until this stress outweighs the protein-protein interaction energy causing the cluster size to cease increasing. This hypothesis also supports hAGTs preference for DNA ends.

The observed cooperativity of binding by hAGT to DNA is not conserved amongst AGTs since Ada-C has been reported to lack this property.(*172*) Further dissimilarities exist amongst these AGTs. Two Ada-C proteins bind to dsDNA while only one binds to ssDNA, irrespective of DNA length. But like hAGT, Ada-C binds both ssDNA and dsDNA with similar affinities ($K_d \sim 2$ μM at 100 mM NaCl) and spans 7 ± 1 nucleotides, showing some parallels with regards to substrate binding.(*172*,*173*)

1.5 Objectives and arrangement of thesis

1.5.1 Objectives

The purpose of this research is to shed light on the role of AGT on ICL damage recognition and repair. The occurrence of ICL damage has been known for years but the exact mechanisms of repair are not fully elucidated and appear to be different depending on the nature of the ICL lesion. If AGT does play a role in ICL detection and repair, its inhibition may be beneficial during chemotherapeutic regimens. Indeed, increased levels of hAGT are correlated with resistance to certain chemotherapeutic agents such as the chloroethylating agent carmustine and the nitrogen mustards chlorambucil, mechlorethamine and cyclophosphamide, suggesting that hAGT may play a role in ICL damage processing.(*139*,*140*,*141*)

To achieve this goal, various AGT proteins were used in binding and repair assays with diverse ICL damage. These ICL containing either 4 or 7 methylene groups as cross-

links were designed to mimic the effects of busulfan or hepsulfam treated DNA, respectively. Based on the known AGT substrates, ICL involving the O^6 atom of dG and O^4 atom of dT were originally investigated, as these were the only potential ICL that could be repaired by the proteins. Repair results demonstrated hAGT's ability to eliminate O^6dG-alkylene-O^6dG ICLs but not the O^4dT-alkylene-O^4dT ICLs.

The next aim was then to determine the cause for this elusive behavior and to design an AGT protein capable of processing O^4dT-alkylene-O^4dT ICLs. The first means employed to gain an understanding of the elusive properties of the O^4dT-alkylene-O^4dT containing ICL was to generate AGT-ICL DNA covalent complexes to aid in crystallization studies, since there are no crystal or NMR structures reported for AGT in complex with O^4-alkyl dT.

A modelling approach was also invoked to determine the absence of repair of O^4dT-alkylene-O^4dT ICLs by modelling O^4-MedT into the hAGT active site. The modelling results spearheaded research that focussed on the design of a chimera and O^4-alkyl dU substrates, including O^4dU-alkylene-O^4dU ICLs. The AGTs failure to repair O^4dU-alkylene-O^4dU ICLs lead to the solution structures of O^4dU-heptylene-O^4dU and O^6dG-heptylene-O^6dG ICLs to establish structural causes for the selective repair by hAGT.

1.5.2 Arrangement of thesis

CHAPTER I

This chapter highlights the importance of DNA damage and repair in the cells. A brief overview of different classes of alkylating agents and the DNA damage they

produce is provided with a focus on the cytotoxic nature of ICL lesions. The mechanisms of ICL formation by nitrogen mustards, nitrosoureas and alkyl sulfonates and their derivatives are elaborated upon. The general mechanisms and substrate ranges of BER, NER, MMR, NHEJ, HR and direct repair are covered. The properties and roles of AGT homologues are discussed in great detail. An extra emphasis is placed on hAGT mediated DNA damage repair and binding.

CHAPTER II-CHAPTER VII Overview

CHAPTER II-CHAPTER V constitute published work, while CHAPTER VI and CHAPTER VII are manuscripts to be submitted. CHAPTER II ("O^6-2'-deoxyguanosine-alkylene-O^6-2'-deoxyguanosine interstrand cross-link in a 5'-GNC motif and repair by human O^6-alkylguanine-DNA alkyltransferase") and CHAPTER III ("O^4-alkyl thymidine cross-linked DNA to probe recognition and repair by O^6-alkylguanine DNA alkyltransferases") are published work in Organic & Biomolecular Chemistry (Royal Society of Chemistry), CHAPTER IV ("Preparation of Covalently Linked Complexes Between DNA and O^6-Alkylguanine-DNA Alkyltransferase Using Interstrand Cross-linked DNA") is published in Bioconjugate Chemistry (American Chemical Society) and CHAPTER V ("Engineering of a O^6-Alkylguanine-DNA alkyltransferase Chimera and Repair of O^4-Alkyl Thymidine Adducts and O^6-Alkylene-2'-Deoxyguanosine Cross-Linked DNA") is published material in Toxicology Research (Royal Society of Chemistry). CHAPTER VI ("Effect of the C5 Methyl on O^6-Alkylguanine-DNA Alkyltransferase Mediated Repair of O^4-alkyl Thymidine Adducts") and CHAPTER VII ("Structural Basis for hAGT Selective Mediated Repair of O^6-2'-Deoxyguanosine-heptylene-O^6-2'-deoxyguanosine and not O^4-2'-Deoxyuridine-heptylene-O^4-2'-

deoxyuridine Interstrand Cross-Linked Containing DNA") are to be submitted for publication in the near future.

The relationships between chapters will be made evident in this section rather than between chapters for clarity and brevity. This section provides a quick glance into each chapter; the major goals, including their experimental rational. Finally, the most note worthy results and drawn conclusions, which were used as a stepping stone for the subsequent chapter, are given.

CHAPTER II

This chapter presents the synthesis of cross-linked amidites linking the O^6 atoms of dG by a butylene and heptylene linkage and their incorporation into oligonucleotides in a 1,3 orientation for hAGT based studies. These ICL DNAs were prepared to supplement previously published data involving similar ICL DNAs where the cross-links were introduced in a directly opposed manner (1,1 motif). This 1,3 motif was employed to generate a more accurate mimic of the cross-links formed by hepsulfam (see **Scheme 1.3**). hAGT was capable of repairing both 1,3 ICL DNAs and did so at similar rates as was previously reported for the 1,1 motif. The orientation of the cross-link does not play a substantial role on how it is processed by hAGT.

CHAPTER III

Despite what its name infers, AGTs are capable of repairing O^4-alkyl dT adducts in addition to O^6-alkyl dG damage. This chapter presents the synthesis of cross-linked amidites linking the O^4 atoms of dT by butylene and heptylene tethers and their incorporation into oligonucleotides in a directly opposed motif to gain an understanding

of the AGT ICL substrate range. These ICL DNAs were prepared to compliment results obtained in CHAPTER II, which demonstrate AGTs ability to repair cross-links. These O^4dT-alkylene-O^4dT ICL DNAs evaded repair by hAGT, OGT and Ada-C. The respective mono-adducts, O^4-butyl-4-ol dT and O^4-Heptyl-7-ol dT were synthesized and used as probes to note the limitations of AGT with respect to O^4-alkyl dT repair. These mono-adducts were repaired solely by OGT. The cause for the poor repair of O^4-alkyl dT mono-adducts and cross-links is not understood considering AGTs have never been crystallized with O^4-alkyl dT containing DNA.

CHAPTER IV

The synthesis of O^6dG-alkylene-O^4dT containing ICL DNAs containing butylene and heptylene linkages and their use as substrates for AGT are reported in this chapter. Since hAGT is capable of repairing cross-links involving the O^6 atom of dG (CHAPTER II) but not those involving the O^4 atom of dT (CHAPTER III) the repair of such cross-links should, in theory, lead to the formation of a covalent complex composed of hAGT linked to the O^4 atom of the modified dT. Such covalent complexes could show use for crystallographic studies, which would shed light into hAGT mediated repair of O^4-alkyl dT damage. For this reason cross-links of different length were tested to obtain an idea of the optimal length for protein-DNA cross-link formation. The O^6dG-heptylene-O^4dT containing ICL DNA was repaired by hAGT but not the O^6dG-butylene-O^4dT one. The identity of the covalent complex formed with the O^6dG-heptylene-O^4dT containing ICL DNA was confirmed to be as hypothesized. The results indicated that use of O^6dG-alkylene-O^4dT containing ICL DNAs for the formation of covalent complexes to be used

for crystallographic purposes shows promise since the O^6dG-heptylene-O^4dT containing ICL DNA generated considerable levels of hAGT-DNA cross-links.

CHAPTER V

An attempt to generate an AGT protein capable of O^4dT-alkylene-O^4dT ICL repair is presented in this chapter. The engineered chimera protein was generated by replacing a portion of the active site of hAGT with the respective amino acids in OGT, with the omission of one amino acid for solubility purposes. The chimera was prepared in hopes that it would harbour properties of hAGT (capable of repairing ICL DNA, CHAPTER II and CHAPTER IV) and OGT (capable of repairing O^4-alkyl dT damage, CHAPTER III), therefore granting the protein with the ability to remove O^4dT-alkylene-O^4dT cross-links. The chimera displayed augmented repair of O^4-alkyl dT mono-adducts, with respect to hAGT, and was able to process O^6dG-alkylene-O^6dG and O^6dG-alkylene-O^4dT with similar rates as hAGT. The chimera was unable to eliminate O^4dT-alkylene-O^4dT cross-links. According to the modelling data, the chimera showed increased repair of O^4-alkyl dT mono-adducts due to the increased flexibility of its active site. The modelling results exposed a potential clash between the protein`s active site and the C5 methyl of O^4-alkyl dT, supporting why O^4-alkyl dT mono-adducts are not processed as readily by the chimera when compared to OGT.

CHAPTER VI

This chapter provides the synthesis O^4-alkyl dU mono-adduct containing DNA and their use as AGT substrates. These probes were generated to verify the adverse effect of the C5 methyl of dT on O^4-alkyl damage repair by AGT, as hypothesized from modelling

studies performed in CHAPTER V. Indeed, removal of the C5 methyl, as observed with the dU derivatives, leads to increased repair by AGTs when compared to similar adducts on the on the O^4 atom of dT (CHAPTER III). According to the modelling, removing the C5 methyl allows the modified nucleotides to penetrate further into the hAGT active site, placing the alkyl group closer to the reactive Cys residue.

CHAPTER VII

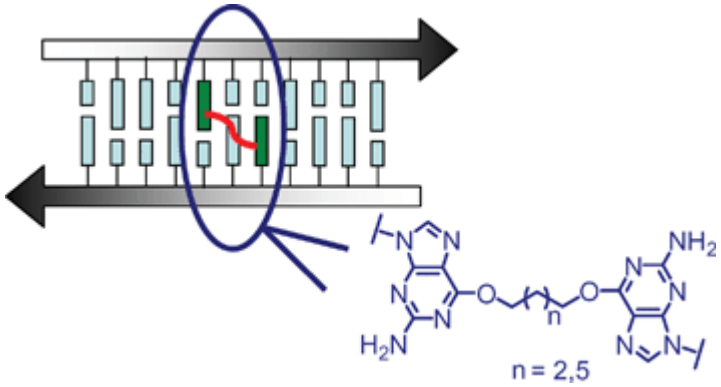
This chapter provides the synthesis and characterization of O^4dU-alkylene-O^4dU ICL DNAs and their use as AGT substrates. These probes were generated to verify the adverse effect of the C5 methyl of dT on ICL repair by AGT, as inferred from experimental and modelling results obtained in CHAPTER VI. Removal of the C5 methyl from the cross-linked nucleotides did not grant any AGT the ability to remove the damage. Biomolecular NMR studies conducted on O^4dU-heptylene-O^4dU and O^6dG-heptylene-O^6dG containing DNA revealed no dramatic differences in global structure between both ICL DNAs. On a more local level, the heptylene cross-link adopted the *E*-configuration about the C4-O^4 for the dU containing ICL DNA, whereas the *Z*-configuration about the C6-O^6 bond was adopted by the cross-link for the modified dG containing ICL DNA. Based on the co-crystal structures of hAGT, the protein orients the lesion in the *Z*-configuration about the C6-O^6 of dG (**Figure 1.17**) or C6-N^6 of pyrimidines (**Figure 1.18**) prior to repair, suggesting that the configuration may play a role in substrate recognition.

CHAPTER VIII

This chapter provides conclusions that summarize the results from CHAPTER II-CHAPTER VII, some future work and the significance of this thesis to the field of science.

CHAPTER II

O^6-2'-deoxyguanosine-alkylene-O^6-2'-deoxyguanosine interstrand cross-link in a 5'-GNC motif and repair by human O^6-alkylguanine-DNA alkyltransferase



Graphical Abstract: Representation of the ICL lesion in a 1,3 motif

Published as:

McManus FP, Fang Q, Booth JD, Noronha AM, Pegg AE, Wilds CJ.

Org Biomol Chem., **8** (2010) 4414-4426.

Abstract

O^6dG-alkylene-O^6dG ICL with a four and seven methylene linkage in a 5'-GNC- motif have been synthesized and their repair by hAGT investigated. Duplexes containing 11 base-pairs with the ICL in the center were prepared by automated DNA solid-phase synthesis using a cross-linked dG dimer phosphoramidite, prepared *via* a seven step synthesis which employed the Mitsunobu reaction to introduce the alkyl lesion at the O^6 atom of dG. Introduction of the four and seven carbon ICL resulted in no change in duplex stability based on UV thermal denaturation experiments compared to a non-cross-linked control. Circular dichroism (CD) spectra of these ICL duplexes exhibited features of a B-form duplex, similar to the control, suggesting that these lesions induce little overall change in structure. The efficiency of repair by hAGT was examined and it was shown that hAGT repairs both ICL containing duplexes, with the heptylene ICL repaired more efficiently relative to the butylene cross-link. These results were reproducible with various hAGT mutants including one that contains a novel V148L mutation. The ICL duplexes displayed similar binding affinities to a C145S hAGT variant compared to the unmodified control duplex. Experiments with Chinese hamster ovary (CHO) cells to investigate the sensitivity of these cells to busulfan and hepsulfam demonstrate that hAGT reduces the cytotoxicity of hepsulfam suggesting that the O^6dG-alkylene-O^6dG ICLs may account for at least part of the cytotoxicity of this agent.

2.1 Introduction

ICLs that are formed in DNA as a consequence of the action of bifunctional alkylating agents represent some of the most toxic lesions encountered by cells due to the obstruction of unwinding of the two strands, critical to the processes of DNA replication, transcription and recombination.(*8*) Interference with these key cellular processes by the presence of ICL is the basis of the mechanism of action of bifunctional alkylating agents, such as mechlorethamine, that are used as cancer therapeutics.(*42*) However, the potency of these agents is reduced by the ability of cancer cells to repair the lesions resulting in an overall resistance to therapy.

In eukaryotic cells the repair of ICL is a complex process with numerous repair pathways including nucleotide excision repair, homologous recombination and non-homologous end joining implicated in the removal of the damage.(*7*) Understanding the molecular basis of how ICL repair occurs will play an important role towards the development of new chemotherapeutic agents that may evade this process, thus increasing the efficacy of these drugs.

One approach employed in elucidating the roles various DNA repair pathways contribute in removing DNA ICL involves the use of chemically synthesized oligonucleotides that contain representative lesions formed by bifunctional alkylating chemotherapeutics.(*174*,*175*,*176*) These oligonucleotides are designed to contain lesions linking specific atoms in DNA in well defined orientations and can be incorporated into plasmids for DNA repair experiments. This approach uses solid-phase synthesis which

generates sufficient amounts of ICL DNA to enable structural studies and repair assays.(*177*,*178*,*179*,*180*,*181*)

Some of these ICL are challenging to prepare synthetically. For example, the bifunctional alkylating agent hepsulfam which has been investigated clinically (*182*), has been shown to form a cross-link between the N7 atoms of dG in 5'-GNC sequences (**Figure 2.1**) as demonstrated through the use of mass spectrometry and identification of 1,7-bis(guanyl)heptane.(*75*) N7-alkyl dG are chemically unstable, for example the presence of N7-methyl dG can result in an apurinic site in DNA or undergo ring opening to yield the ring opened formamido pyrimidine derivative.(*183*)

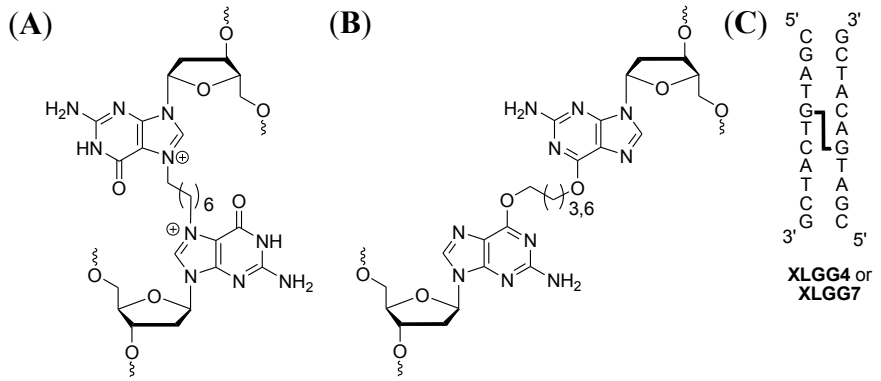


Figure 2.1: Structures of the (**A**) N7dG-heptylene-N7dG cross-link induced by hepsulfam, (**B**) the O^6dG-alkylene-O^6dG cross-link (where n = 3 or 6) and oligomers (**C**) **XLGG4** and **XLGG7**

In order to prepare oligonucleotides containing a cross-link that models the orientation of the lesion formed by hepsulfam, our group has explored methodologies to prepare ICL DNA that links the O^6 atoms of dG.(*184*) The O^6 atom of dG is a known site of alkylation by chemotherapeutic drugs such as the methylating agent temozolomide and the chloroethylating agent carmustine which proceeds to form a dG-dC ICL.(*185*)

Different chemical groups at the O^6 position of dG have been incorporated by organic synthesis for numerous purposes including protection during oligonucleotide

synthesis and as substrates for DNA repair studies.(*186*,*187*) Examples of synthetic methods that have been employed to introduce these groups include the post-synthetic modification of oligonucleotides containing the 2′-deoxyribonucleoside of 2-amino-6-methylsulfonylpurine and the Mitsunobu reaction.(*187*,*188*)

Using a combination of solution and solid-phase synthesis, a 5'-GNC ICL has been prepared (**Figure 2.1**) which required a O^6dG-alkylene-O^6dG phosphoramidite that contains various protecting groups to permit orthogonal removal, enabling a unit of asymmetry to be engineered in the final ICL DNA duplex to allow for a clinically relevant staggered 1,3 orientation. The synthetic scheme to produce this amidite involved the Mitsunobu reaction which was selected due to its mild conditions, compatibility with all protecting groups employed and direct route to introduce alkyl linkers of various lengths. This method allows for the production of ICL duplexes to explore the influence of linker length in substrates of defined structure and for the systematic investigation of susceptibility to DNA repair mechanisms.

The ability of wild hAGT and some mutants to repair these ICLs was also investigated. AGT are a class of repair proteins that are found in numerous organisms whose role is to repair alkylation at the O^6 position of the dG and thus they play an important role in maintaining genomic integrity.(*189*) Alkylation at the O^6 position of dG can be mutagenic, disrupting normal Watson-Crick base pairing leading to point mutations which may be fatal. The mechanism by which hAGT repairs alkylated DNA involves flipping the alkylated nucleoside out of the duplex and into the active site of the protein where transfer of the alkyl group from the O^6 position of dG to a C145 residue occurs.(*155*) The protein can only act once due to the irreversible alkylation of C145,

whereupon the protein is degraded by the ubiquitin pathway.(*104*) hAGT has been shown not only to repair small adducts at the O^6 position of dG such as a methyl group but also larger groups such as benzyl and 4-(3-pyridyl)-4-oxobutyl.(*129*) The study of these novel ICL substrates to undergo repair by hAGT will contribute to the range of DNA substrates that can undergo repair *via* this pathway.

2.2 Experimental

2.2.1 Materials

5'-*O*-Dimethoxytrityl-N^2-phenoxyacetyl-2'-deoxyguanosine, 3'-*O*-dimethoxytrityl-2'-deoxyribonucleoside-5'-*O*-(β-cyanoethyl-*N*,*N*'-diisopropyl) phosphoramidites and *N*,*N*-diisopropylamino cyanoethyl phosphonamidic chloride were purchased from ChemGenes Inc. (Wilmington, MA). 5'-*O*-Dimethoxytrityl-2'-deoxyribonucleoside-3'-*O*-(β-cyanoethyl-*N*,*N*'-diisopropyl)phosphoramidites and protected 2'-deoxyribonucleoside-polystyrene 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) obtained from Silicycle (Quebec City, QC). Thin layer chromatography (TLC) was performed 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 INOVA 300 MHz NMR spectrometer at room temperature. ^{1}H NMR spectra were recorded at a frequency of 300.0 MHz and chemical shifts were reported in parts per million downfield from tetramethylsilane. ^{31}P NMR spectra (^{1}H decoupled) were recorded at a frequency of

121.5 MHz with H_3PO_4 used as an external standard. 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. ESI mass spectra for small molecules were recorded at the McGill University Department of Chemistry Mass Spectrometry Facility with a Finnigan LCQ DUO mass spectrometer in methanol or acetone. 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, ethylenediaminetetraacetic acid-free (EDTA-free) Protease Inhibitor Cocktail Tablets were obtained from Roche (Laval, QC) Nitrocellulose filters (0.20 μm) were obtained from Millipore. XL-10 Gold *E. coli* cells were obtained from Stratagene (Cedar Creek, TX). *DpnI*, T4 PNK, Unstained Protein Molecular Weight Marker and restriction enzymes *EcoRI* and *KpnI* were obtained from Fermentas (Burlington, ON). [γ-^{32}P]ATP was purchased from Amersham Canada Ltd. (Oakville, ON).

2.2.2 Solution synthesis of small molecules

The synthesis of 3'-*O*-alloxycarbonyl-5'-*O*-dimethoxytrityl-N^2-phenoxyacetyl-2'-deoxyguanosine (**1**), 3'-*O*-alloxycarbonyl-5'-*O*-dimethoxytrityl-N^2-phenoxyacetyl-O^6-(4-*t*-butyldiphenylsiloxybutyl)-2'-deoxyguanosine (**2a**), 3'-*O*-alloxycarbonyl-5'-*O*-dimethoxytrityl-N^2-phenoxyacetyl-O^6-(7-*t*-butyldiphenylsiloxyheptyl)-2'-deoxyguanosine (**2b**), 3'-*O*-alloxycarbonyl-5'-*O*-dimethoxytrityl-N^2-phenoxyacetyl-O^6-

(4-hydroxybutyl)-2'-deoxyguanosine (**3a**), 3'-*O*-alloxycarbonyl-5'-*O*-dimethoxytrityl-N^2-phenoxyacetyl-O^6-(7-hydroxyheptyl)-2'-deoxyguanosine (**3b**), 3'-*O*-*t*-butyldimethylsilyl-5'-*O*-dimethoxytrityl-N^2-phenoxyacetyl-2'-deoxyguanosine (**4**), 1-{O^6-[3'-*O*-alloxycarbonyl-5'-*O*-dimethoxytrityl-N^2-phenoxyacetyl-2'-deoxyguanidyl]}-4-{O^6-[3'-*O*-*t*-butyldimethylsilyl-5'-*O*-dimethoxytrityl-N^2-phenoxyacetyl-2'-deoxyguanidyl]}-butane (**5a**), 1-{O^6-[3'-*O*-alloxycarbonyl-5'-*O*-dimethoxytrityl-N^2-phenoxyacetyl-2'-deoxyguanidyl]}-7-{O^6-[3'-*O*-*t*-butyldimethylsilyl-5'-*O*-dimethoxytrityl-N^2-phenoxyacetyl-2'-deoxyguanidyl]}-heptane (**5b**), 1-{O^6-[3'-*O*-*t*-butyldimethylsilyl-5'-*O*-dimethoxytrityl-N^2-phenoxyacetyl-2'-deoxyguanidyl]}-4-{O^6-[5'-*O*-dimethoxytrityl-N^2-phenoxyacetyl-2'-deoxyguanidyl]}-butane (**6a**), 1-{O^6-[3'-*O*-*t*-butyldimethylsilyl-5'-*O*-dimethoxytrityl-N^2-phenoxyacetyl-2'-deoxyguanidyl]}-7-{O^6-[5'-*O*-dimethoxytrityl-N^2-phenoxyacetyl-2'-deoxyguanidyl]}-heptane (**6b**), 1-{O^6-[3'-*O*-*t*-butyldimethylsilyl-5'-*O*-dimethoxytrityl-N^2-phenoxyacetyl-2'-deoxyguanidyl]}-4-{O^6-[5'-*O*-dimethoxytrityl-N^2-phenoxyacetyl-2'-deoxyguanidyl-3'-*O*-(β-2-cyanoethyl-*N*,*N*'-diisopropyl)phosphoramidite]}-butane (**7a**) and 1-{O^6-[3'-*O*-*t*-butyldimethylsilyl-5'-*O*-dimethoxytrityl-N^2-phenoxyacetyl-2'-deoxyguanidyl]}-7-{O^6-[5'-*O*-dimethoxytrityl-N^2-phenoxyacetyl-2'-deoxyguanidyl-3'-O-(β-2-cyanoethyl-*N*,*N*'-diisopropyl)phosphoramidite]}-heptane (**7b**) were performed as previously described.(*190*)

2.2.3 Oligonucleotide synthesis and purification

The cross-linked duplexes **XLGG4** and **XLGG7**, whose sequences are shown in **Figure 2.1**, were assembled using an Applied Biosystems Model 3400 synthesizer on a 1 μmole scale employing standard β-cyanoethylphosphoramidite cycles supplied by the

manufacturer with slight modifications to coupling times described below. The nucleoside phosphoramidites containing standard protecting groups were prepared in anhydrous acetonitrile (ACN) at a concentration of 0.1 M for the 3'-*O*-2'-deoxyphosphoramidites, 0.15 M for the cross-linked 3'-*O*-2'-deoxyphosphoramidites (**7a** and **7b**) and 0.2 M for the 5'-*O*-2'-deoxyphosphoramidites. Assembly of sequences first involved detritylation (3% trichloroacetic acid (TCA) in DCM), followed by nucleoside phosphoramidite coupling with commercial 3'-*O*-2'-deoxyphosphoramidites (2 min), 5'-*O*-2'-deoxyphosphoramidites (3 min) or cross-linked phosphoramidites **7a** or **7b** (10 min); Subsequent capping with phenoxyacetic anhydride/pyridine/tetrahydrofuran (1:1:8, v/v/v; solution A, and 1-methyl-1 H-imidazole/tetrahydrofuran 16:84 w/v; solution B) and oxidation (0.02 M iodine in tetrahydrofuran/water/pyridine 2.5:2:1) followed every coupling. To complete assembly of the cross-linked duplexes, the cyanoethyl groups were removed from the polystyrene-linked oligomers by treating the support with 1 mL of anhydrous triethylamine (TEA) for at least 12 h. The support was then washed with 30 mL of anhydrous ACN followed by anhydrous THF. The *t*-butyldimethylsilyl (TBS) group was removed from the partial duplex by treating the support with 2 x 1 mL triethylamine trihydrofluoride (TEA•3HF) for a total of 1 h. The support was then washed with 30 mL each of anhydrous THF and ACN followed by drying *via* high vacuum (20 min). The final extension of the cross-linked duplex was then achieved using 5'-*O*-2'-deoxyphosphoramidites with a total detritylation exposure of 130 seconds and removal of the 3'-terminal trityl group by the synthesizer to yield duplexes **XLGG4** and **XLGG7** on the solid support.

The oligomer-derivatized polystyrene beads were transferred from the reaction column to screw cap microfuge tubes fitted with teflon lined caps and the oligomer released from the support and protecting groups removed by treatment with a mixture of concentrated ammonium hydroxide/ethanol (0.3 mL:0.1 mL) for 4 h at 55 ºC. The cross-linked final products were separated from pre-terminated products by strong anion exchange (SAX) HPLC using a Dionex DNAPAC PA-100 column (0.4 cm x 25 cm) purchased from Dionex Corp, (Sunnyvale, CA) with a linear gradient of 0–50% buffer B over 30 min (buffer A: 100 mM Tris HCl, pH 7.5, 10% ACN and buffer B: 100 mM Tris HCl, pH 7.5, 10% ACN, 1 M NaCl) at 40 °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.(*175*)

2.2.4 Oligonucleotide characterization by ESI-MS and nuclease digestion

The cross-linked oligomers (0.1 A_{260} units) were characterized by enzymatic digestion (snake venom phosphodiesterase: 0.28 units and calf intestinal alkaline phosphatase: 5 units, in 10 mM Tris, pH 8.1 and 2 mM magnesium chloride) for a minimum of 36 h at 37 °C as previously described.(*175*) The resulting mixture of nucleosides was analyzed by reversed phase HPLC carried out using a Symmetry® C-18 5µm column (0.46 x 15 cm) purchased from Waters Inc, Milford, MA. The C-18 column was eluted with a linear gradient of 0-60% buffer B over 30 min (buffer A, 50 mM sodium phosphate, pH 5.8, 2% ACN and buffer B, 50 mM sodium phosphate, pH 5.8, 50% ACN). The resulting peaks were identified by co-injection with the corresponding

standards and eluted at the following times: dC (4.6 min), dG (7.9 min), dT (8.5 min), dA (9.6 min), cross-link dimers (18.1 for the four carbon and 26.2 min for the seven carbon cross-link), and the ratio of nucleosides was determined. The molecular weights of the cross-linked oligomers were determined by ESI-MS and these were in agreement with the calculated values.

2.2.5 UV thermal denaturation studies

Molar extinction coefficients for the oligonucleotides and cross-linked duplexes were calculated from those of the mononucleotides and dinucleotides according to nearest-neighbour approximations (M^{-1} cm^{-1}). Non-cross-linked 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 90 mM sodium chloride, 10 mM sodium phosphate, 1 mM EDTA buffer (pH 7.0) to give a final concentration of 2.8 μM duplex. The cross-linked duplexes were dissolved in the same buffer to give a final concentration of 2.8 μM. The solutions were then heated to 90 °C for 10 min, cooled slowly to room temperature, and stored at 4 °C overnight before measurements. Prior to the thermal run, samples were degassed by placing them in a speed-vac concentrator for 2 min. Denaturation curves were acquired at 260 nm at a rate of heating of 0.5 °C/min, using a Varian CARY Model 3E spectrophotometer fitted with a 6-sample thermostated cell block and a temperature controller. The data were analyzed in accordance with the convention of Puglisi and Tinoco (*191*) and transferred to Microsoft Excel™.

2.2.6 CD spectroscopy

CD 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 10 °C in 90 mM sodium chloride, 10 mM sodium phosphate, 1 mM EDTA (pH 7.0), at a final concentration of 2.8 μM for the cross-linked duplexes and 2.8 μM for control duplex. Each spectrum was an average of 5 scans. Spectra were collected at a rate of 100 nm/min, with a bandwidth of 1 nm and sampling wavelength of 0.2 nm using fused quartz cells. The CD spectra were recorded from 350 to 200 nm at 10 °C. The molar ellipticity was calculated from the equation $[\phi] = \varepsilon/Cl$, 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).

2.2.7 Molecular modelling of ICL duplexes

The DNA duplex (5'-dCGATGTCATCG)/(5'-CGATGACATCG) and cross-linked duplexes **XLGG4** and **XLGG7** were built using HyperChem™ molecular modelling software. All duplexes were geometry optimized using the AMBER forcefield.

2.2.8 Protein expression and purification

Site directed mutagenesis as well as transformation into XL-10 *E. coli* cells were performed as directed by the Stratagene manual. Cells containing either a plasmid coding for wild-type or mutant hAGT were grown in a 1 L cultures of LB (Luria broth) + 100 μg/mL ampicillin until an $OD_{600} = 0.6$ was reached. The cells were induced with 0.3 mM IPTG, incubated at 37 °C for 4 h with shaking at 225 rpm and harvested by centrifugation at 6000 x g at 4 °C for 30 min. Pellets were weighed and resuspended in 5 mL of

resuspension buffer (20 mM Tris HCl (pH 8.0), 250 mM NaCl, 20 mM β-mercaptoethanol supplemented with Complete, Mini, EDTA-free Protease Inhibitor Cocktail Tablets) for each gram of pellet. The cells were homogenized by using approximately 20 strokes in a dounce homogenizer, lysed using two rounds of French press, centrifuged at 17000 x g for 45 min at 4 °C. The lysate was applied to a pre-equilibrated Ni-NTA Superflow column containing 3.5 mL of resin. The column was washed with 200 mL resuspension buffer and 20 mM imidazole, until the OD_{280} of the eluant was constant. The protein eluted with 50 mL resuspension buffer supplemented with 200 mM imidazole. 1 mL fractions were collected and fractions that displayed protein content (generally fractions 4-11) were pooled followed by dialysis versus 4 L of dialysis buffer (50 mM Tris HCl (pH 7.6), 250 mM NaCl, 20 mM β-mercaptoethanol and 0.1 mM EDTA) using 8000 Da cutoff dialysis tubing. Typically a yield of 8-10 mg of purified protein was obtained per L of culture inoculated.

2.2.9 Protein characterization

All proteins were analyzed by ESI-MS and were prepared by precipitation using one volume of TCA for every four volumes of protein. Samples were incubated for 10 min on ice, centrifuged at 4 °C for 5 min at 14 krpm on a table-top centrifuge, the supernatant removed, the pellet washed with 200 µL of cold acetone and washing repeated. The protein pellet was dried at 95 °C for 5 min, resuspended in a 50% ACN and 1% formic acid solution in water at a protein concentration of 10 µM. The samples were analyzed with a Waters Micromass Q-ToF-2 mass spectrometer operating in positive-ion mode.

Far-UV CD spectra were obtained on a JASCO 815 spectropolarimeter using a 2 mm path length cell. The scans were performed at 20 °C, by averaging 5 wavelength scans from 260 to 200 nm (1 nm bandwidth) in 0.2 nm steps at a rate of 20 nm/min, and 1 sec response at a protein concentration of 5 μM in CD buffer (50 mM potassium phosphate (pH 7.5), 75 mM NaCl and 2 mM dithiothreitol (DTT)).

Thermal denaturation curves were obtained by monitoring the change in mdeg at 222 nm (corresponding to the α-helical content of the protein). The heating rate was set at 15 °C/h using a Peltier-type temperature controller, which ranged from 40-70 °C. The T_m was obtained by noting the temperature at which the 1st derivative of the denaturation plot was the highest.

Fluorescence spectra were obtained using a Varian Cary Eclipse Fluorescence Spectrophotometer in a 10 mm quartz cuvette with 3.5 μM protein in CD buffer at room temperature. Excitation and emission slits were set at 5 nm with a detector voltage of 650 and the spectra were the average of 10 accumulations. Tyrosine and tryptophan fluorescence were obtained using an excitation wavelength of 280 nm and recording the emission spectrum between 300 and 400 nm. To monitor the tryptophan fluorescence solely an excitation wavelength of 295 nm was used and the emission spectrum was monitored between 300 and 400 nm.

2.2.10 Cross-link repair assay

The cross-linked oligonucleotides **XLGG4** and **XLGG7** as well as the control DNA (non-cross-linked duplex) were labeled at the 5'-end using [γ-^{32}P]ATP. 10 μL reaction volumes were used for the labeling of the DNA. The reaction mixture was comprised of 1x T4 PNK buffer (Fermentas), 21 μM oligonucleotide, 10 μM cold-ATP, 1

μL [γ-^{32}P]ATP (10 μCi/μL) and 5 units of T4 PNK. The reaction mixture was incubated at 37 °C for 1 h and terminated by placing the reaction in boiling water for 10 min. 60 pmol of wild-type and mutant hAGT protein was incubated with 2 pmol of labeled DNA in a total reaction volume of 15 μL comprised of hAGT buffer (80 mM Tris HCl (pH 7.6), 0.1 mM EDTA and 5 mM DTT). The reactions were incubated at 37 °C for 14 to 16 h with similar results indicating that the reaction was complete by 14 h. Denaturing PAGE in 7 M urea was used to analyze the products. The reaction was terminated by adding 18.2 μL of stop reaction buffer (81 mM Tris HCl, 81 mM boric acid, 1.8 mM EDTA and 1% sodium dodecyl sulphate (SDS) (pH 8.0) in 80% formamide) to the reaction tube and placed in boiling water for 10 min. The samples were loaded onto 14 cm x 16 cm, 17% polyacrylamide gels (19:1) in the presence of 7 M urea. The gels were run using 1 X TBE (89 mM Tris HCl, 89 mM boric acid, 2 mM EDTA (pH 8.0)) for 1.5 h at 700 V, which heated the gels to 40 °C. After electrophoresis, the gels were covered in Saran wrap and exposed to a storage phosphor screen. The image was captured using a Typhoon 9400 (GE Healthcare, Piscataway, NJ) and the counts of the radiolabelled products quantified using ImageQuant™ (Amersham Biosciences). A control lane was used as a standard in order to quantitate the other bands on the gel. For the repair time course assay of the ICL a 150 μL master mix composed of 600 pmol of wild-type hAGT and 20 pmol of DNA substrate in hAGT buffer was prepared. The sample was placed at 37 °C. At every time point 7.5 μL (1 pmol of DNA) was removed from the master mix and placed in a tube with 9.1 μL of stop reaction buffer and terminated as above.

2.2.11 Binding studies of C145S hAGT

Reaction tubes consisting of 5 nM DNA and increasing C145S hAGT concentrations (ranging from 1 to 35.69 μM) were prepared in a total solution volume of 20 μl of binding buffer (10 mM Tris HCl (pH 7.6), 5 mM DTT and 0.1 mM EDTA). The samples were allowed to equilibrate on ice for 30 min. Triplicate samples were incubated for longer (45 min and 2 h) and yielded similar results indicating that equilibrium had been attained. Each sample (0.1 pmol) was loaded on 14 cm x 16 cm, 4 °C, preequilibrated native 6% polyacrylamide (75:1) gels. The gels were electrophoresed in 10 mM Tris acetate (pH 7.6) and 0.25 mM EDTA buffer at 125 V for 1 h. On completion, the gels were covered and processed as described above.

Estimates of the monomeric dissociation constant (K_d) and the cooperativity factor (n) were obtained from the electro mobility shift assay (EMSA) as explained previously.(*164*) The binding of n moles of hAGT protein [P] to 1 mole of cross-linked DNA [D] follows the equation: $n\mathrm{P} + \mathrm{D} \leftrightarrow \mathrm{P}_n\mathrm{D}$. Isolating the variables and taking the logarithm of the equation yields:

$$\log\left(\frac{[\mathrm{P}_n\mathrm{D}]}{[\mathrm{D}]}\right) = n \log [\mathrm{P}]_{free} + \log K_d$$

Plotting log $[\mathrm{P}_n\mathrm{D}]/[\mathrm{P}]$ as a function of log [P] gives a slope equal to n. At half-saturation: log $[\mathrm{P}_n\mathrm{D}]/[\mathrm{P}] = 0$ and observed K_d can be obtained from the relation observed $K_d = -n \ln [\mathrm{P}]$. The monomeric K_d can be obtained by taking the n^{th} root of the observed K_d.

2.2.12 Colony forming assay in CHO cells treated with busulfan and hepsulfam.

CHO cells transfected with either the empty pCMV vector or pCMV-hAGT were grown in α-MEM media (Gibco-Invitrogen, Carlsbad, CA) containing 26 mM $NaHCO_3$, 10% fetal bovine serum (Atlanta Biologicals, Lawrenceville, CA), 100 U/ml penicillin and 100 μg/ml streptomycin. The cells were maintained by seeding at $1x10^5$ cells/25 cm^2 flask at weekly intervals. For colony forming assays, cells were plated at a density of 10^6 per 25 cm^2 flask and 24 h later were treated with different concentrations of busulfan (Sigma, St Louis, MO) or hepsulfam (NCI, Frederick, MD) for 4 h. The cells were washed with phosphate buffered saline (PBS) twice, and then the medium was replaced with fresh medium and the cells allowed to grow for another 16-18 h. They were then replated at densities of 250-1000 cells per 25 cm^2 flask and grown for 6-8 days until discrete colonies had formed. The colonies were washed with 0.9% saline solution, stained with 0.5% crystal violet and counted.

2.3 Results and discussion

2.3.1 Syntheses and characterization of the ICL duplexes

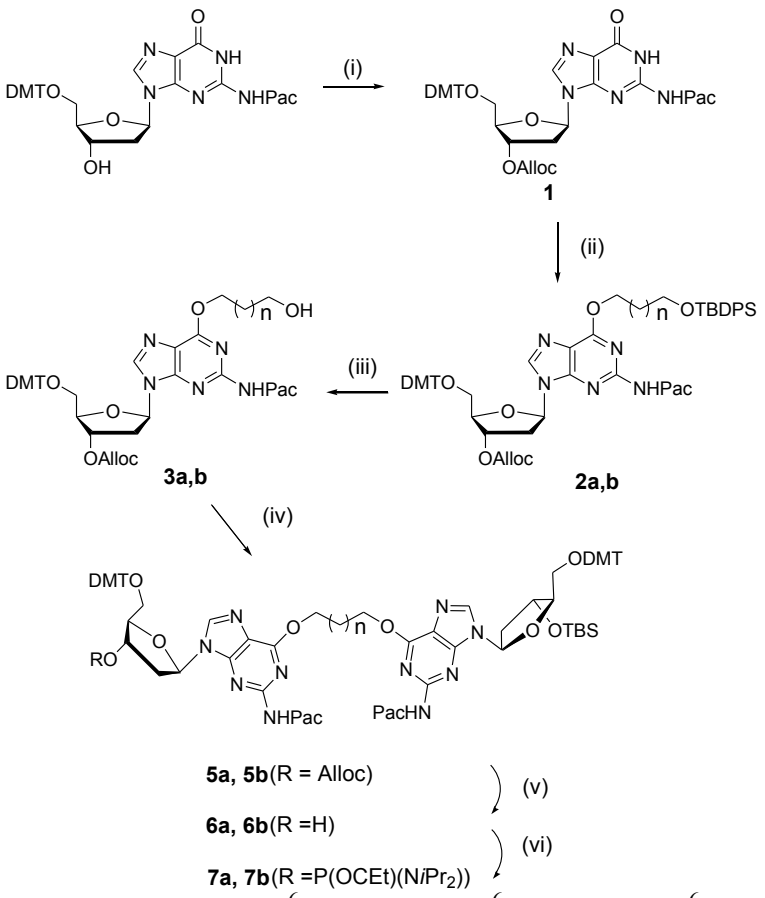
ICL DNA repair studies require access to substrates of well defined structure. The bifunctional alkylating agent hepsulfam has been shown to alkylate specifically at the N7 atom of dG in sequences containing a 5'-GNC motif placing a heptylene linkage between the two dG.(*75*) Because they are synthetically challenging to prepare, we have focused on the synthesis of a cross-link that joins the O^6 atoms in this particular sequence motif. It should be emphasized that this specific cross-link has not been identified as a product of

alkylation by hepsulfam. The methodology to produce this ICL was used to prepare duplexes containing alkyl linkages of various lengths to investigate the affect of linker length on ICL stability, structure and repair by hAGT. The chemotherapeutic agent busulfan is another example of a bifunctional alkylating agent that has been used for the treatment of chronic myelogenous leukemia.(*192*) This agent introduces a butylene linkage between the nucleobases with the demonstration that this agent reacts with dG to form 1,4-di(7-guanosinyl)butane.(*193*) Mass spectrometry of the reaction products of busulfan with oligonucleotides suggested that this agent forms intrastrand cross-links with 5′-GA-3′ sequences.(*68*)

The structure of the O^6dG-alkylene-O^6dG cross-link and the sequences containing the butylene and heptylene cross-links (**XLGG4** and **XLGG7**) are shown in **Figure 2.1**. The synthesis of the cross-linked amidites **7a** and **7b** were performed according to the synthetic pathway in **Scheme 2.1**. Starting from commercially available 5'-*O*-dimethoxytrityl-N^2-phenoxyacetyl-O^6-2'-deoxyguanosine the free 3'-alcohol was protected as an allyloxycarbonyl ester **1** using allyl 1-benzotriazoyl carbonate.(*194*) The adducts **2a** and **2b** were prepared by introduction of 1-*t*-butyldiphenylsilyloxy-butan-4-ol or 1-*t*-butyldiphenylsilyloxy-heptan-7-ol at the O^6 position of **1** *via* the Mitsunobu reaction.(*186*,*188*) Removal of the *t*-butyldiphenylsilyl protecting group on compounds **2a** and **2b** was accomplished with TBAF at room temperature for 30 min to yield the adducts **3a** and **3b**. The formation of dimers **5a** or **5b** were achieved *via* a second Mitsunobu reaction using 5'-*O*-dimethoxytrityl-N^2-phenoxyacetyl-3'-*O*-*t*-butyldimethyl-silyl-2'-deoxyguanosine (**4**) in yields of 63 and 71 %, respectively. Amidite precursors **6a** and **6b** required the removal of the 3'-*O*-alloxycarbonyl group from **5a** or **5b** with

palladium (0) tetrakistriphenylphosphine in THF at room temperature. These were then converted to phosphoramidites **7a** and **7b** using a slight excess of *N*,*N*-diisopropylamino cyanoethyl phosphonamidic chloride and isolated by hexane precipitation. The isolated phosphoramidites **7a** and **7b** were analyzed by mass spectrometry and were found to have the expected molecular masses. ^{31}P NMR analysis of these phosphoramidites revealed the presence of two signals for **7a** (143.92 and 144.14 ppm) and **7b** (143.91 and 144.11 ppm) in the region diagnostic for a phosphoramidite.

The synthesis of the 1,3-GNC engineered ICL using phosphoramidite dimers **7a** or **7b** required three different protecting groups around the cross-linked site enabling an orthogonal approach for their removal. This would allow for the synthesis of three different branches in terms of sequences around the site of the ICL. The Mitsunobu reaction was chosen for its specificity at the O^6 position of a protected dG (*186*,*188*) and the versatility of this method which allows for the synthesis of cross-links of various lengths, depending on the diol linker used. This reaction was employed twice in the synthetic pathway outlined in **Scheme 2.1**, first for the synthesis of the monomers **2a** and **2b** containing the four and seven carbon linkers at the O^6 atom and later in the scheme for dimers **5a** and **5b**. The dimerization reaction to give dimer **5b** proceeded in a slightly higher yield relative to **5a** (71 versus 63 %), presumably due to the reduced steric hindrance of the longer alkyl linker of **3b** that reacts with **4**.



Scheme 2.1: Synthesis of O^6dG-butylene-O^6dG and O^6dG-heptylene-O^6dG phosphoramidites

Reagents and conditions: (i) Alloc-OBt, DCM/pyridine (9:1), 12 h; (ii) 4-(*t*-butyldiphenylsiloxy) butan-1-ol or 7-(*t*-butyldiphenylsiloxy) heptan-1-ol (1.1 eqv.), PPh_3 (1.25 eqv.), DIAD (1.2 eqv.) (iii) TBAF (1 M in THF), 30 min; (iv) 4 (1.01 eqv.), PPh_3 (1.25 eqv.), DIAD (1.2 eqv.), 1 h; (v) $Pd(PPh_3)_4$ (5mol%), PPh_3 (0.25 eqv.), *n*-butylamine/formic acid (1:1, 2.5 eqv.), THF, 20 min; (vi) *N*,*N*-diisopropylamino cyanoethyl phosphonamidic chloride, DIPEA, THF, 30 min.

Solid phase synthesis of **XLGG4** and **XLGG7** was performed on a 1 μmol scale using phosphoramidites **7a** or **7b** according to **Figure 2.2**. The first arm of the duplex was synthesized on a polystyrene support using commercially available 3'-*O*-2'-deoxyphosphoramidites. Coupling of the cross-linked phosphoramidites **7a** or **7b** introduced the ICL at the 5'-ends to the linear oligomers. The dimethoxytrityl groups of the dimer moiety were removed by brief treatment with 3% TCA in DCM to expose the

5'-OH groups from which the second and third arms of the duplex were then synthesized with a noted increase in the trityl conductivity reading on the synthesizer. Repetitive coupling with protected 3'-*O*-2'-deoxyphosphoramidites gave, after detritylation and capping (phenoxyacetic anhydride), the branched oligomer **Y-TBS** (**Figure 2.2**). The 3'-*O*-TBS group was then removed from **Y-TBS** by treating the support with anhydrous TEA for 16 h followed by TEA•3HF twice for 30 min at room temperature. C-18 Reversed-phase HPLC analysis of the deprotected intermediate **Y-OH** revealed complete removal of the silyl group with the shift of the major peak from 11.2 to 8.2 min in the case of the butylene containing cross-link (data not shown). Continued synthesis with repetitive coupling of 5'-*O*-2'-deoxyphosphoramidites at the 3'-end of intermediate **Y-OH** gave full length ICL duplexes **XLGG4** and **XLGG7**.

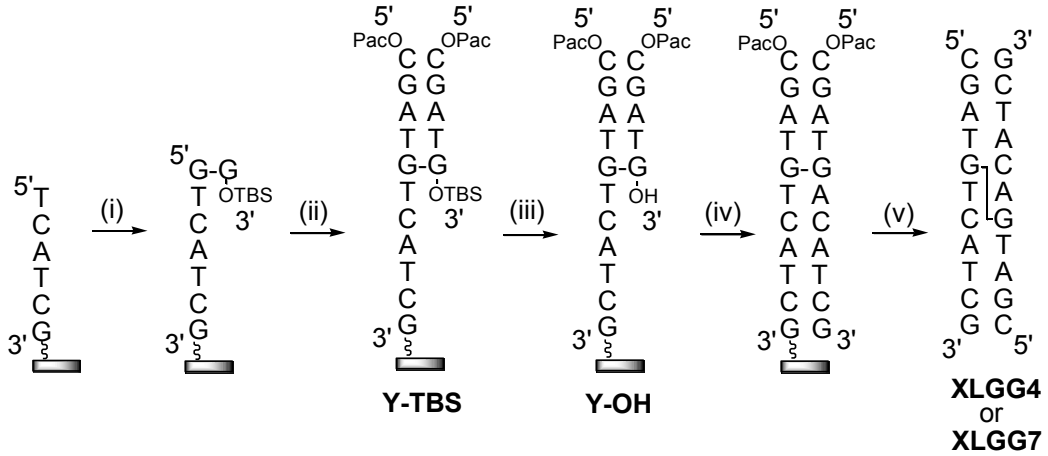


Figure 2.2: Methodology to construct the cross-linked duplexes **XLGG4** and **XLGG7** *via* solid phase synthesis.

(i) Coupling of amidite **7a** or **7b**; (ii) extension with 3'-*O*-2'-deoxyphosphoramidites followed by capping with phenoxyacetic anhydride; (iii) removal of the 3'-*O*-*t*-butyldimethylsilyl group with TEA·3HF; (iv) extension with 5'-*O*-2'-deoxyphosphoramidites and (v) cleavage from the solid support.

The solid-phase synthesis of the ICL containing oligonucleotides **XLGG4** and **XLGG7** required some changes from the standard synthesis cycle and reagents employed

during solid-phase oligonucleotide synthesis. First, synthesis was performed using a polystyrene rather than controlled-pore glass (CPG) solid-support due to the incompatibility of TEA•3HF with the latter.(*175*) As an added precaution, the cyanoethyl protecting groups were removed using TEA as it has been observed that prolonged fluoride treatment using TBAF could lead to chain cleavage.(*195*) Phenoxyacetic anhydride rather than acetic anhydride was used as capping reagent due to a undesirable N-acetylation reaction that has been observed with 'fast-deprotecting' amidites.(*196*) Additionally, the protection of the exocyclic amine functionalities of the cross-linked guanine bases with the phenoxyacetyl group would allow for milder deprotection conditions and ease of its removal at the final stage.

Chain assembly in the 3' to 5' direction proceeded smoothly using 3'-*O*-2'-deoxyphosphoramidites and the cross-linked dimers **7a** and **7b** with essentially quantitative coupling observed by conductivity measurements to assess the trityl values. A higher concentration (0.15 M) and longer coupling time (10 min) was used for phosphoramidites **7a** and **7b** to ensure a high coupling efficiency due to their larger size compared to the standard 3'-*O*-2'-deoxyphosphoramidites. At this point, capping with phenoxyacetyl anhydride gave intermediate **Y-TBS** (**Figure 2.2**) which was desilylated to give the free 3'OH functionality. The efficiency of this step was readily monitored by C-18 reversed phase HPLC (data not shown). A small amount (approximately 1 mg) of the solid support was deprotected using NH_4OH/ethanol (3:1), milder conditions that have been developed for the deprotection of RNA to ensure that the TBS groups are not removed.(*197*) The removal of the TBS group decreases the hydrophobicity of the oligomer and a change in the retention time for the oligomer from 11 to 8 min on the C-

18 reverse phase HPLC indicates removal to give the more hydrophilic **Y-OH** intermediate (**Figure 2.2**). The final step in the assembly of **XLGG4** and **XLGG7** is continued chain extension with 5'-*O*-2'-deoxyphosphoramidites. Cleavage and deprotection of **XLGG4** and **XLGG7** were performed with concentrated ammonium hydroxide/ethanol (3:1) at 55 °C for 4 h to yield 141.8 and 108.6 OD of material, respectively (**Table 2.1**). Analysis by SAX HPLC revealed that the major product was the desired ICL whose retention times are shown in **Table 2.1**. The combination of mild deprotection conditions and bulky nature of the adduct that the ICL represents at the O^6 position may account for the stability under these deprotection conditions. Purification by SAX HPLC, followed by desalting gave **XLGG4** and **XLGG7** in approximately 26 and 36% overall isolated yields.

Table 2.1: Amounts, retention times, nucleoside ratios and mass spectral data for cross-linked duplexes **XLGG4** and **XLGG7**.

Duplex	A_{260} Units[a]	Retention Time (min)[b]	Nucleoside	Ratios		Mass	
				exp	obs	exp	obs
XLGG4	141.8 (36.9)	23.7	dC	6.0	6.1	6727.5	6726.8
			dG	4.0	4.2		
			dT	5.0	5.0		
			dA	5.0	5.1		
			dG-dG	1.0	1.1		
XLGG7	108.6 (39.1)	25.5	dC	6.0	5.8	6769.6	6769.0
			dG	4.0	4.1		
			dT	5.0	5.0		
			dA	5.0	5.1		
			dG-dG	1.0	1.1		

[a] Amount of crude ICL duplex purified by SAX HPLC. The numbers in parentheses indicate the amount of pure duplex obtained.
[b] Retention times (min) of ICL duplexes on SAX HPLC using a 0.0-0.5 M linear gradient of sodium chloride.

These ICL duplexes were digested to the constituent nucleosides with a combination of snake venom phosphodiesterase and calf intestinal alkaline phosphatase

and analyzed by C-18 reversed phase HPLC. In addition to the four standard 2'-deoxynucleosides, one additional peak was observed with a retention time of 18.1 min for **XLGG4** and 25.5 min for **XLGG7**. As shown in **Table 2.1**, the ratios of the component 2'-deoxynucleosides and cross-linked nucleosides were consistent with the theoretical composition of ICL duplexes **XLGG4** and **XLGG7**. The ICL oligomers **XLGG4** and **XLGG7** required additional time for digestion compared to single stranded oligonucleotides (48 h as opposed to 30 min). ESI mass spectrometry of the duplexes **XLGG4** and **XLGG7** had masses of 6726.8 and 6769.0 Da, consistent with the theoretical values for the cross-linked duplexes.

2.3.2 UV thermal denaturation and CD of ICL duplexes

UV denaturation experiments were carried out to assess the effect of the alkyl linkers on duplex stability. The UV thermal denaturation (T_m) curves for the ICL duplexes **XLGG4**, **XLGG7**, and the corresponding non-cross-linked control duplex all exhibited a sigmoidal denaturation profile with similar hyperchromicities for the transition curves (see Appendix II). The melting temperature for **XLGG4**, **XLGG7** and the non-cross-linked control were all similar with a value of ~ 48 °C.

When this ICL was present as a directly opposed O^6dG-heptylene-O^6dG mismatch in an 11-bp duplex, an increase in T_m of 23 °C over the control duplex was observed.(*184*) UV thermal denaturation experiments with oligonucleotides containing an O^6-MedG / dC base pair have shown that a single O^6-methyl substitution reduced duplex stability by at least 18 °C per substitution in 1 M NaCl buffer and DNA duplexes containing two of these alkylated lesions were found to have a T_m 40 °C lower than that of the unmodified duplex.(*20*,*198*,*199*) It is believed that the destabilizing nature of the

O^6-alkyl lesions are counterbalanced by the preorganization of the covalently linked strands of the duplex resulting in a stability comparable to the non-cross-linked control.(*175*,*184*)

The CD spectra of cross-linked duplexes **XLGG4**, **XLGG7** and the non-cross-linked control were recorded at 10 °C (spectra are shown in Appendix II). In all cases, the CD spectra of the duplexes exhibited signatures characteristic of B-form DNA with a positive maximum peak centered around 275 nm, a negative peak at approximately 250 nm and a cross over around 260 nm.(*200*,*201*) The CD spectra of non-cross-linked controls revealed some minor differences, particularly a reduction of the signal at 275 nm.

These results suggests that the cross-links had minimal effect on the global B-form structure. NMR studies performed with dodecanucleotides containing a O^6-MedG •C base pair implicate the formation of a wobble base pair, with the methylated dG sliding towards the major groove and the dC oriented towards the minor groove. This base pair is stacked in the duplex between the flanking base pairs inducing only a small conformational change from the wild-type duplex.(*23*) In a standard B-form duplex, the distance between the O^6-O^6 atoms in a 5'-GNC sequence is approximately 6.4 Å. The O-O distance in the fully extended 1,4-butanediol and 1,7-heptanediol are 6.2 and 9.9 Å, respectively, which is sufficient to span the distance linking the two O^6 atoms in a 5'-GNC sequence motif without inducing a significant structural change. Molecular models for the oligomers **XLGG4**, **XLGG7** and the non-cross-linked control duplex that were geometry optimized using the AMBER forcefield suggest that the alkylene linkers are

oriented towards the major groove and do not greatly distort the duplex, consistent with our findings (see Appendix II).

2.3.3 Cross-link repair

Repair assays were performed on the ICL containing oligomers **XLGG4** and **XLGG7** with wild-type hAGT and the variants C145S, P140A and V148L. The C145S variant was selected as a negative control due to its lack of alkyltransferase activity and for the investigation of protein binding to the O^6dG-alkylene-O^6dG ICLs.(*163*) The P140A mutation has been investigated for the influence of reduced size of the active site pocket on hAGT activity.(*202*,*203*) In addition, some AGTs such as the *E.coli* Ada-C, have alanine in place of proline in the equivalent position of hAGT. The V148L variant was prepared to probe the effect of the side chain of Val 148 on the repairing ability of the protein. This variant was selected since little is known about this amino acid other than the carbonyl group of Val 148 has been shown in wild-type hAGT to accept a H-bond from the N^2 atom of a bound dG.(*148*,*155*) We suspected that substituting Val to a larger amino acid would reduce the size of the active site pocket yielding a protein with the same properties as P140A.

All protein (wild-type hAGT, C145S, P140A and V148L) masses were verified by ESI-MS and were in agreement with the calculated values (see Appendix II). The secondary structure of the proteins by far UV CD was examined to ensure that the alterations did not compromise the structure. CD scans of the P140A and V148L hAGT variants revealed similar secondary structures to the wild-type hAGT due the highly conservative nature of these alterations (see Appendix II). The C145S mutation had a greater impact on the global secondary structure of the protein. Although the size of the

cysteine and serine side groups are relatively similar, the H-bonding capabilities of these two amino acids differ.

The stability of the proteins (wild-type hAGT, C145S, P140A and V148L) were assessed *via* thermal denaturation studies to ensure that the proteins were properly folded at the assay temperature (37 °C) by monitoring the α-helical content of the protein at 222 nm as a function of increasing temperature. The thermal denaturation studies demonstrated that all hAGT proteins were still properly folded at 37 °C due to their T_m's being well above this temperature. Thermal denaturation studies of the wild-type hAGT correlated with values reported of complete inactivation of the protein at 60 °C.(*204*) All mutations destabilized the protein based on CD spectra where the stability of the secondary structure of the protein was compromised (see Appendix II).

The intrinsic fluorescence of the hAGT proteins were investigated by monitoring the local environments containing both tryptophan and tyrosine residues using an excitation wavelength of 280 nm while in the case of tryptophan alone the excitation wavelength of 295 nm was used. It is clear from the emission wavelengths and intensities that the mutations had no dramatic affect on the tertiary structure of the protein with V148L having a slight impact on tertiary structure of the protein relative to the other mutants (see Appendix II).

The 5'-^{32}P labeled ICL DNA duplexes **XLGG4** and **XLGG7** (2 pmol) were incubated with 60 pmol of either the wild-type hAGT or the variants (C145S, P140A and V148L) overnight at 37 °C to determine repair. Trials for either 14 or 16 h were performed with virtually identical results obtained, indicating that the reaction was complete as monitored by denaturing PAGE (**Figure 2.3**). Lanes 1 and 2 contain 2 pmol

of ssDNA corresponding to the sequence of one of the repaired strands with the only difference being the presence of hAGT in lane 2. Lanes 3 and 8 contain 2 pmol of **XLGG4** and **XLGG7**, respectively, representing the unrepaired ICL duplexes with a reduced mobility relative to that of completely repaired substrate. Lanes 4 and 9 contain 60 pmol of hAGT with 2 pmol of either **XLGG4** (lane 4) or **XLGG7** (lane 9). In both cases two additional bands are seen, one that migrates more quickly corresponding to the completely repaired single stranded product and the other at the top of the gel with significantly reduced mobility which is likely a single hAGT-bound covalently to a DNA species, a median repair product. For both **XLGG4** and **XLGG7**, no repair products are observed with 60 pmol of the variants C145S (lanes 5 and 10), while P140A (lanes 6 and 11) or V148L (lanes 7 and 12) show very minimal repair of **XLGG7** and no repair of **XLGG4**. The repair assays demonstrate that both **XLGG4** and **XLGG7** are repaired by hAGT with **XLGG7** undergoing a greater amount of repair (with 57.0% of ICL repaired) over **XLGG4** (31.4% of ICL repaired) based on quantitation of ICL remaining as measured by the relative counts of the labelled products present on the gel by ImageQuant™ (see Appendix II).

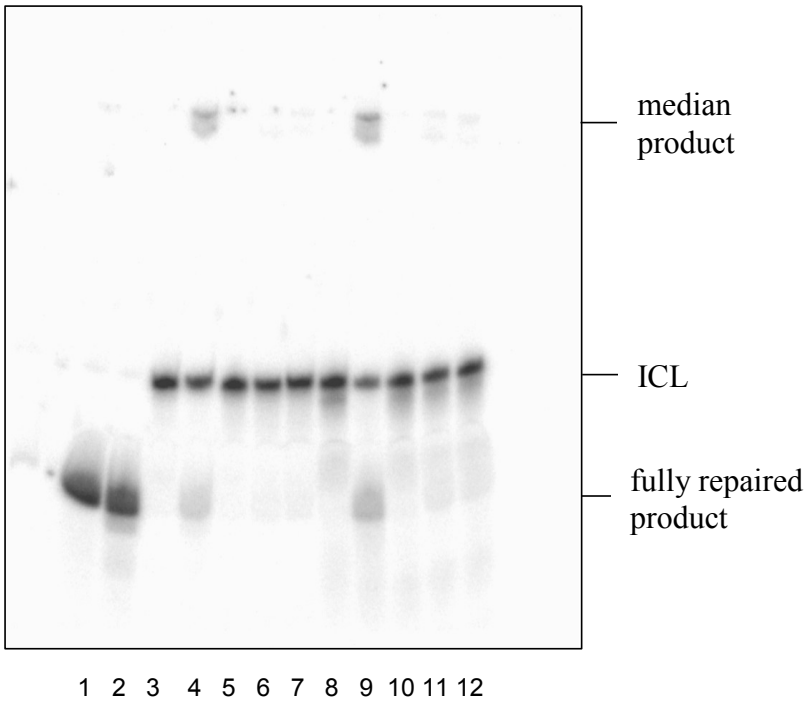
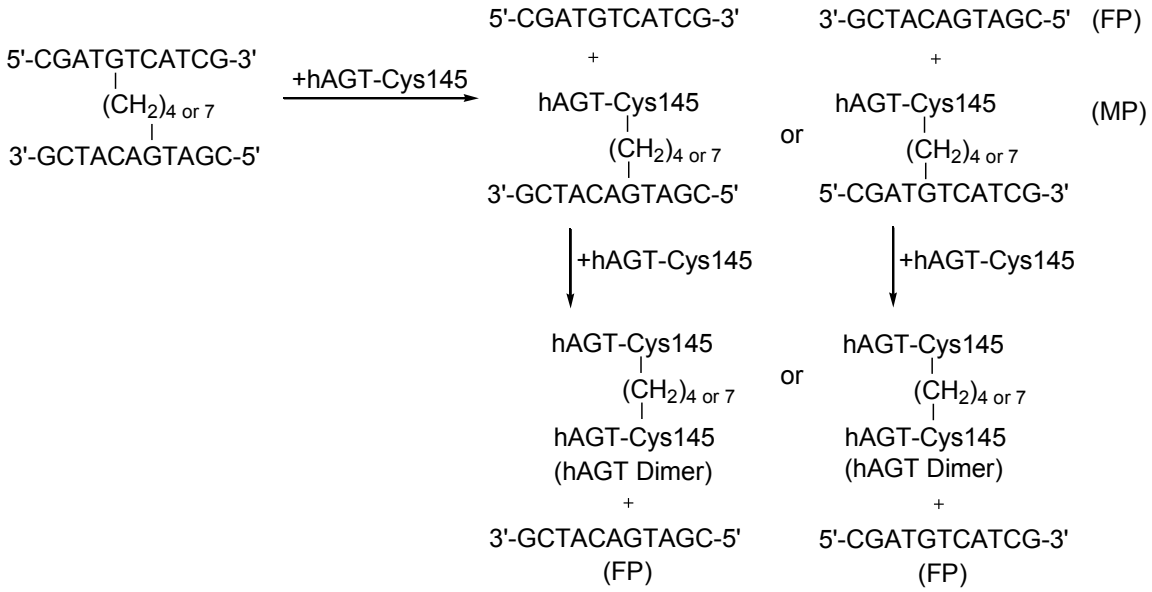


Figure 2.3: Repair of **XLGG4** and **XLGG7** by wild-type hAGT, C145S, P140A and V148L.

(**A**) Denaturing PAGE of repair reactions as described in the experimental section: Lane 1, 2 pmol Control; lane 2, 2 pmol Control + 60 pmol hAGT; lane 3, 2 pmol **XLGG4**; lane 4, 2 pmol **XLGG4** + 60 pmol hAGT; lane 5, 2 pmol **XLGG4** + 60 pmol C145S; lane 6, 2 pmol **XLGG4** + 60 pmol P140A; lane 7, 2 pmol **XLGG4** + 60 pmol V148L; lane 8, 2 pmol **XLGG7**; lane 9, 2 pmol **XLGG7** + 60 pmol hAGT; lane 10, 2 pmol **XLGG7** + 60 pmol C145S; lane 11, 2 pmol **XLGG7** + 60 pmol P140A; lane 12, 2 pmol **XLGG7** + 60 pmol V148L.

For the repair assays involving **XLGG4** and **XLGG7**, the samples were denatured in stop reaction buffer before being loaded on a 7 M urea denaturing gel to prevent self-complexation of the repaired DNA strands which could complicate determination of ratios of the repaired intermediate and fully repaired product versus intact (unrepaired) ICL (**Scheme 2.2**).



Scheme 2.2: Proposed repair pathway of cross-link species by wild-type hAGT, P140A and V148L (MP - median product; FP- final product).

The inactive C145S variant was unable to repair both ICL duplexes as expected due to the lack of the active site Cys residue which is known to be the alkyl group acceptor. The other P140A and V148L variants showed no repair of **XLGG4** and very poor repair for **XLGG7** (4.5% and 3.0%, respectively). The diminished capacity for repair by the P140A variant has been observed as it is much less active than wild-type hAGT in dealkylating O^6-benzylguanine, which has been explained by the reduced size of the binding pocket.(*205*)

Time course assays were performed using 60 pmol wild-type hAGT and 2 pmol of the **XLGG4** and **XLGG7** substrates by quantitating the amount of unrepaired, partially and fully repaired species (see Appendix II). **XLGG4** was repaired at a much slower rate with 25% repair of the ICL occurring in 8.5 h whereas the same level of repair for **XLGG7** required approximately 1 h.

The results obtained from the time course experiments for the 5'-GNC-3' ICL for **XLGG4** and **XLGG7** follow similar trends to repair results for the directly opposed O^6dG-alkylene-O^6dG ICL in a mismatch duplex.(*146*) It was shown that 50% final product was formed after 2 h for a directly opposed O^6dG-heptylene-O^6dG mismatch ICL, also observed in the repair of **XLGG7**. These results are surprising, as the ICL spanning a 5'-GNC-3' motif would be expected to be a less flexible substrate relative to directly opposed O^6dG-heptylene-O^6dG mismatch ICL. The mechanism by which hAGT repairs DNA alkylation involves rotation of the alkylated nucleoside from the DNA duplex from the minor groove into the active site allowing for transfer of the alkyl group to the residue C145.(*189*) The improved repair of **XLGG7** versus **XLGG4** can be rationalized by the difference in distances of the cross-link. For the four carbon linker of **XLGG4**, the O-O distance of the fully extended linker is just sufficient (6.2 Å) to span the distance necessary to link the O^6 atoms in a 5'-GNC sequence (6.4 Å). In a geometry optimized model of **XLGG4**, the O-O distance was found to be 4.5 Å with some tilting of the guanine bases towards each other (see Appendix II). The model of **XLGG7** has an O-O distance of 7.1 Å with the longer alkylene chain protruding into the major groove. This would suggest that rotation of one of the alkylated dG of **XLGG7** into the active site of hAGT would be less difficult relative to the more strained **XLGG4** ICL. These findings demonstrate that there could have been some distortion in the cross-linked duplexes inherent in the nature of the substrates prior to the repair process and that **XLGG4** is too short to fit optimally in the active site relative to the **XLGG7** counterpart.

The formation of the final product occurs at a faster rate than the formation of the median product for both **XLGG4** and **XLGG7**, as the % final product observed is always

greater than the % median product. Final and median product formation occur at different rates, suggesting that hAGT preferentially repairs the median product over the initial substrate as indicated by the lack of accumulation of median product at any time. If hAGT preferentially repaired the median product over the initial substrate the median product should not increase with time as the median product would be consumed as it would be formed.

One possible explanation as to why this reaction does not go to completion is that the reacted hAGT binds to another lesion containing ICL or oligonucleotide in our assay. It is known that alkylated hAGT can also bind to DNA.(*163*) *In vivo*, dissocation is most likely reinforced by ubiquitination resulting in degradation of the alkylated hAGT.(*206*) In addition, it is possible that the cross-link may be oriented in an alternate, less reactive conformation, as has been postulated for O^6-(2-hydroxyethyl) dG and O^6-[4-oxo-4-(3-pyridyl)but-1-yl] dG containing oligonucleotides.(*129*)

2.3.4 Binding studies of C145S hAGT

Binding cooperativity (Hill factor) and monomeric K_d values were determined for the non-cross-linked control duplex, **XLGG4** and **XLGG7** with the C145S hAGT mutant (see **Table 2.2** and Appendix II). Similar cooperativity values were obtained for the control duplex (1.41 ± 0.06) and **XLGG7** (1.32 ± 0.06) whereas this value was found to be slightly higher for **XLGG4** (1.75 ± 0.13). All three DNA duplexes showed a similar monomeric dissociation constant ranging from 8.36 for the control duplex to 9.87 μM for **XLGG4** (see Appendix II).

Table 2.2: Hill factor and monomeric K_d values of C145S hAGT-DNA complexes

DNA	Cooperativity Factor	K_d (μM)
Control	1.41 ± 0.06	8.36±1.11
XLGG4	1.75 ± 0.13	9.87±1.22
XLGG7	1.32 ± 0.06	9.67±1.12

Experiments to measure cooperativity and dissociation constants for **XLGG4** and **XLGG7** showed only two bands by EMSA corresponding to the free and bound ICL DNA. The absence of an intermediate band on the native gel indicates that there is cooperativity in the binding of hAGT for the ICL DNA. **XLGG7** and the control duplex showed similar cooperativity and dissociation constants indicating that the presence of the 7-methylene cross-link had little influence on hAGT interaction with the DNA. **XLGG4** had a slightly higher cooperativity of interaction with hAGT, however, all the measured Hill factors for ICL duplexes **XLGG4** and **XLGG7** were between 1 and 2 in accordance with previous work which showed that 1 hAGT protein binds every 4 nucleotides.(*164*) A Hill factor of 2 indicates perfect cooperativity, as previously shown in the case of hAGT for many unmodified oligonucleotide sequences under different assay conditions whereas a value of 1 indicates no cooperativity.(*164*)

In the absence of NMR or X-ray structural data for **XLGG4** or **XLGG7**, it can only be speculated that the butylene cross-link induces a structural change in the DNA that has an influence on the interaction of the ICL with hAGT contributing to the slight increase in cooperativity that is observed. It is known that binding of hAGT to DNA causes structural changes in the DNA, which may play a role in aiding the rotation of the damage base from the DNA double helix and subsequently promote hAGT cooperative binding.(*151*,*155*) However, the shorter alkylene linker of **XLGG4** versus **XLGG7** may

hinder rotation of the alkylated dG into the active site contributing to a reduction in repair by hAGT.

2.3.5 Colony forming assay in CHO cells treated with busulfan and hepsulfam and relevance with O^6dG-alkylene-O^6dG ICL

CHO cells that do not express AGT were sensitive to killing by hepsulfam (**Figure 2.4**). Expression of hAGT provided a significant but not complete protection from this agent. This is consistent with the concept that an O^6dG-alkylene-O^6dG cross-link accounts for at least part of the cytotoxicity of this agent and that the efficient repair of this 7 carbon adduct by hAGT prevents this killing. There was no protection from busulfan, which is in accord with the poor repair of the 4 carbon adduct produced by this agent.

Extensive study of the hAGT reaction has shown that DNA repair by this protein is specific for adducts on the O^6 position with a minor capability to repair adducts on the O^4 position of dT. The results shown here using *in vitro* assays show that a DNA interstrand cross-link containing 7 carbon atoms linking two O^6 atoms of dG residues is efficiently repaired by hAGT. Such a cross-link would be expected to be formed in treated cells by hepsulfam and, although it has not yet been characterized, the finding that hAGT reduces the cytotoxicity of hepsulfam is strong evidence that it does occur. Unrepaired DNA interstrand cross-links are very toxic and only a small number of such cross-links may be formed. The results also suggest that a high level of hAGT expression would render tumor cells resistant to therapy with hepsulfam.

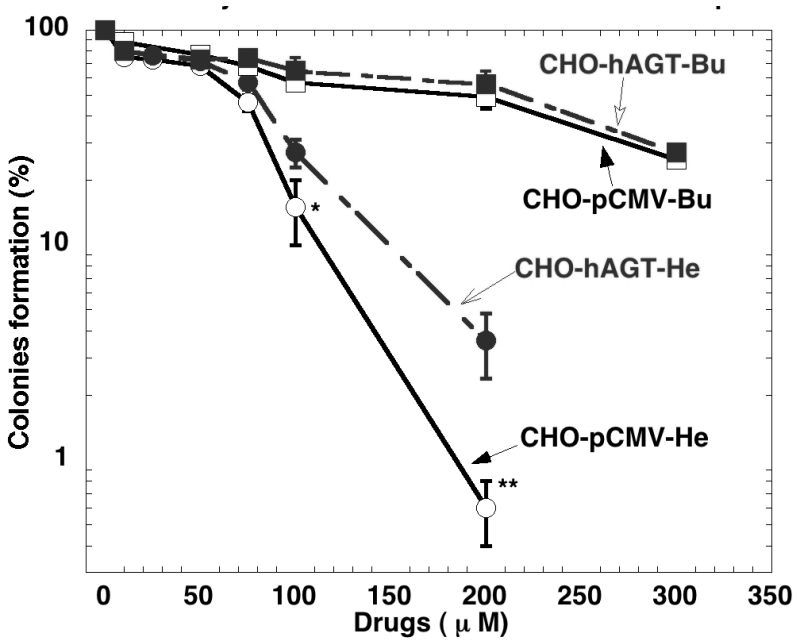


Figure 2.4: The effect of the expression of hAGT on exposure of CHO cells to busulfan and hepsulfam.

Cell survival was measured with a colony forming assay as described in the experimental section. * $p<0.05$, Compared to cells expressing hAGT; ** $p<0.01$, Compared to cells expressing hAGT.

2.4 Conclusions

The synthesis of nucleoside dimers containing a O^6dG-alkylene-O^6dG cross-link that enables the construction of a 1,3-staggered 5'-GNC motif by solid-phase synthesis using phosphoramidite chemistry and an orthogonal approach to removing protecting groups is described. These ICL were obtained in high yield and purity that is required for DNA repair studies. ICL duplexes containing a four and seven methylene linkage were found to have similar thermal stability and structure relative to a non-cross-linked control. Both were repaired by hAGT, where hAGT displayed a higher efficiency for the ICL containing the seven methylene linkage. Binding studies suggested similar affinity of

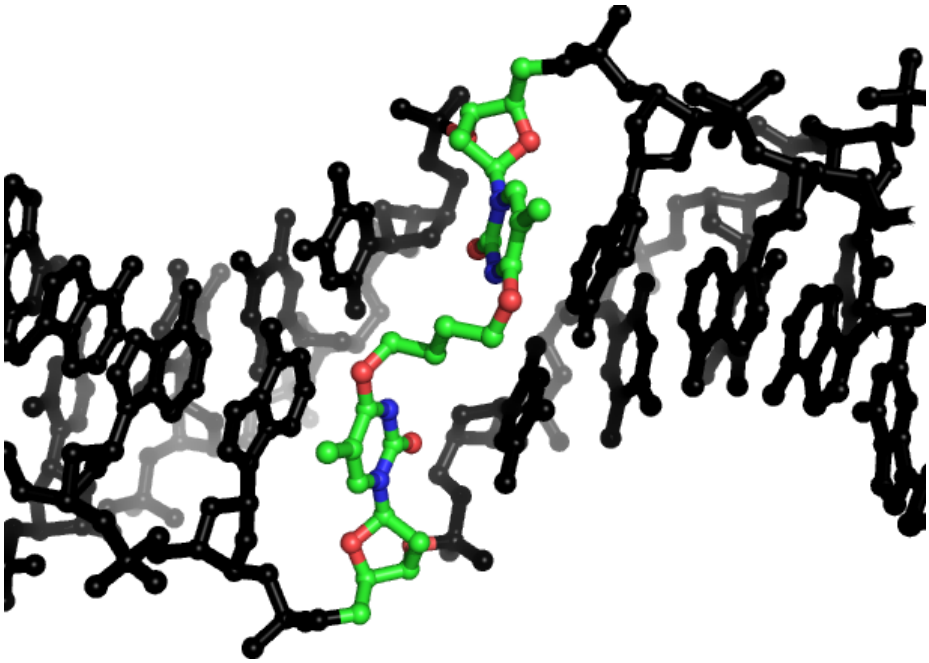
hAGT for both ICL. Studies with cells treated with hepsulfam demonstrate that hAGT reduces the cytotoxicity that this bifunctional alkylating agent induces and serves as evidence that the O^6dG-alkylene-O^6dG cross-link may account for at least part of the cytotoxicity of this agent.

2.5 Acknowledgements

We thank the Natural Sciences and Engineering Research Council of Canada (NSERC), the Canada Foundation for Innovation (CFI), the Canada Research Chair (CRC) program (to CJW) and grant CA-018137 from the NCI, US Public Health Service (to AEP) for financial support of this project. FPM is the recipient of graduate fellowships from NSERC, the Fonds Québécois de la Recherche sur la Nature et les Technologies (FQRNT) and Groupe de Recherche Axé sur la Structure des Protéines (GRASP). We also thank Mr. Nadim Saadeh and Dr. Bruce Lennox of McGill University for ESI-MS analysis.

CHAPTER III

O^4-alkyl thymidine cross-linked DNA to probe recognition and repair by O^6-alkylguanine DNA alkyltransferases



Graphical Abstract: Representation of O^4dT-butylene-O^4dT ICL lesion in a DNA duplex

Published as:

McManus FP, O'Flaherty DK, Noronha AM, Wilds CJ.

Org Biomol Chem., **10** (2012) 7078-7090.

Abstract

DNA duplexes containing a directly opposed O^4dT-alkylene-O^4dT interstrand cross-link 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. CD experiments on the ICL duplexes indicated minimal difference from the standard B-form DNA structure. These ICL were used for DNA repair studies with hAGT, Ada-C and OGT, whose purpose is to remove O^6-alkyl dG and in some cases O^4-alkyl dT lesions. It has been previously shown that hAGT can repair O^6dG-alkylene-O^6dG ICL. The O^4dT-alkylene-O^4dT ICL prepared in this study were found to evade repair by hAGT, OGT and Ada-C. 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 dT and O^4-heptyl-7-ol dT. Binding experiments conducted with hAGT demonstrated that the protein bound O^4-alkyl dT lesions with similar affinities as O^6-MedG, which hAGT repairs efficiently, suggesting the lack of O^4-alkyl dT repair by hAGT is not a function of substrate recognition.

3.1 Introduction

Interference by ICLs of critical cell events involving DNA unwinding is exploited in cancer chemotherapy regimens employing bis-alkylating agents.(*35*,*36*,*42*) 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 DR, BER, NER, HR, NHEJ and MMR remove various DNA lesions.(*8*,*9*,*78*,*207*) 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.(*7*,*208*,*209*)

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.(*62*) A number of elegant examples have been reported including the preparation of numerous phosphoramidites of nucleosides and other molecules to introduce site-specific lesions in DNA for various experiments.(*176*,*180*,*210*,*211*,*212*) Repair and binding experiments can be conducted by incorporating the oligonucleotides containing lesions into specific plasmids.(*213*)

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, the alkyl sulfamate hepsulfam and the nitrogen mustard mechlorethamine both form similar ICL between two N7 atoms of dG in DNA containing a 5'-GNC.(*60*,*61*,*75*) Alkylation at the N7 atom of dG destabilizes the nucleoside. For example, the half-life of the mechlorethamine induced ICL is on the order of hours which can limit experiments that may be conducted.(*63*) Our lab recently reported the preparation of a chemically stable O^6dG-alkylene-O^6dG ICL *via* a versatile synthetic strategy that enables introduction of an alkyl lesion of any desired length and orientation.(*184*) Repair studies of the O^6dG-alkylene-O^6dG ICL with AGT proteins were conducted. AGT proteins, which are found in all kingdoms of life, are responsible predominantly for the repair of O^6-MedG lesions which are renowned to cause point mutations by disrupting normal Watson–Crick base pairing proving that this protein is integral in maintaining genomic integrity.(*189*) hAGT, the most thoroughly characterized AGT protein, repairs alkylated DNA by flipping the damaged nucleotide out of the helix and into its active site where the alkyl group is transferred from the point of lesion to the active site Cys 145 residue.(*155*) Once alkylated, this protein is degraded by the ubiquitin pathway.(*104*)

It was found that hAGT can remove an O^6dG-alkylene-O^6dG ICL in mismatched (*146*) and 5'-GNC sequence motif DNA, the latter designed to mimic the lesion formed by hepsulfam.(*214*) We reported that hAGT reduces hepsulfam cytotoxicity in experiments with CHO cells, suggesting that this drug may form O^6dG-alkylene-O^6dG

ICL given that hAGT repairs only O^6-alkyl dG. Interestingly, neither AGT homologues from *E. coli* (Ada-C or OGT) were capable of repairing these ICL.

DNA mismatches can arise from heteroduplexes formed during homologous recombination, errors in DNA replication, deamination and base damage by alkylating agents.(*215*,*216*,*217,218*) The exocyclic atoms of mispaired nucleobases are more susceptible to alkylation as a result of altered hydrogen bonding patterns that expose these atoms to the external environment, as observed by the enhanced reactivity of mispaired dT residues to osmium tetroxide.(*219*) Moreover, formation of mutagenic alkylated nucleobases, such as O^4-MedT, can cause mispairs when encountered by DNA polymerases to form, for example, an O^4-MedT:dG mismatch.(*4*)

Formation of pyrimidine-pyrimidine ICL introduced by bifunctional alkylating agents have been reported. DNA containing mispaired dC residues treated with mechlorethamine have been shown to form ICL linking N3 atoms.(*220*) This recent discovery of ICL formation involving two mispaired pyrimidines is unique and unprecedented. Not much is known about the properties of these ICL because of their recent appearance and is an area of research that requires further investigation.

Based on a straightforward synthetic methodology and the observation that directly opposed O^6dG-alkylene-O^6dG ICL could be repaired by hAGT, we set out to prepare O^4dT-alkylene-O^4dT ICL in a directly opposed motif as well as their respective mono-adducts (**Figure 3.1**) by adapting methods described by Swann's group requiring the preparation of convertible nucleosides.(*221*) These ICL were prepared by solid-phase oligonucleotide synthesis and the influence of various alkyl linker lengths on duplex

stability and structure as well as recognition and repair by AGTs from human and *E. coli* were explored.

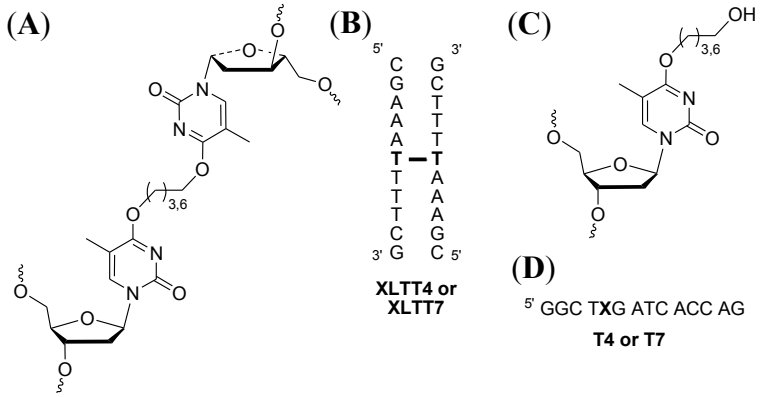


Figure 3.1: Structures of (**A**) O^4dT-alkylene-O^4dT interstrand cross-link, (**B**) duplex where T-T is the ICL, (**C**) O^4-alkyl dT monoadducts and (**D**) DNA sequence where X contains the monoadduct.

The formation of O^4-MedT in DNA is highly mutagenic due to the preferential incorporation of dG opposite this alkylated nucleotide during replication by DNA polymerases, as indicated previously.(*4*) O^4-MedT is a minor alkylation product with respect to its mutagenic O^6-MedG counterpart. DNA treated with MNU produces 126 times more O^6-MedG than O^4-MedT.(*24*) Though found in relatively low abundance this lesion is important as it has been observed in the DNA of smokers and is partly responsible for the adverse effects of tobacco.(*54*)

In mammalian cells, O^4-MedT is more toxic than O^6-MedG both in normal and repair deficient systems.(*29*,*30*) Unlike O^6-MedG, O^4-MedT is not affected by the expression or suppression of hAGT activity nor by MMR explaining its added toxicity. NER deficient cells exhibit increased sensitivity to O^4-MedT indicating that this lesion is possibly repaired *via* the NER pathway in mammalian cells.(*31*) *E. coli* has evolved two proteins from the direct repair pathway to eliminate O^4-MedT lesions, OGT and the C-

terminal of the adaptive response protein Ada, highlighting the importance of repairing this lesion *in vivo*, underscoring an interesting difference in substrate preference between the human and *E. coli* proteins.(*32*,*33*,*34*,*123*)

3.2 Experimental

3.2.1 Materials

5'-*O*-Dimethoxytrityl-thymidine, 3'-*O*-dimethoxytrityl-2'-deoxyribonucleoside-5'-*O*-(β-cyanoethyl-*N*,*N*'-diisopropyl) phosphoramidites and *N*,*N*-diisopropylaminocyanoethylphosphonamidic chloride were purchased from ChemGenes Inc. (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 Inc. (Gibbstown, NJ). Flash column chromatography was performed using silica gel 60 (230–400 mesh) obtained from Silicycle (Quebec City, QC). TLC was performed 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 MHz and chemical shifts were reported in parts per million downfield from tetramethylsilane. ^{13}C NMR spectra (^{1}H decoupled) were recorded at a frequency of 125.7 MHz with tetramethylsilane as a reference. ^{31}P NMR spectra (^{1}H decoupled) were recorded at a frequency of 202.3 MHz with H_3PO_4 used as an external standard. High resolution mass

spectrometry of modified nucleosides were obtained using a LTQ OrbitrapVelos - ETD (Thermo Scientific) at the Concordia University Centre for Biological Applications of Mass Spectrometry. The mass spectrometer was operated in full scan, positive ion detection mode. Ampicillin, 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) XL-10 Gold and BL21(DE3) *E. coli* cells were obtained from Stratagene (Cedar Creek, TX). *DpnI*, T4 PNK, Unstained Protein Molecular Weight Marker and restriction enzymes *EcoRI* and *KpnI* were 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). Oligonucleotides containing O^4-MedT required for repair studies were a kind gift from TriLink Biotechnologies (San Diego, CA).

3.2.2 Solution synthesis of small molecules

3'-*O*-(*t*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-O^4-(hydroxybutyl)-thymidine (1a)

1,4-butanediol (1.60 g, 17.8 mmol) was dissolved in THF (15 mL). Sodium metal (2.35 g, 102 mmol) was added. After 5 h, the salt was extracted to be immediately used in the subsequent reaction. 3'-*O*-(*t*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-C4-(1,2,4-triazol-1-yl)-thymidine (preperad as described in *221*) (2.5 g, 3.52 mmol) was dissolved in dioxane (10 mL). Sodium 4-hydroxybutan-1-olate was added dropwise.

After 15 h, the solvent was evaporated in vacuo, the crude product was taken up in DCM (50 mL) and the solution was washed with two portions of sodium bicarbonate (2 x 50 mL). The organic layer was dried over sodium sulphate and concentrated to produce a yellow gum. The crude product was purified by flash column chromatography using a hexanes : ethyl acetate solvent system (2 : 3) to afford 2.05 g (79.8 %) of product as a colorless foam.

R_f (SiO_2 TLC): 0.24 hexanes : ethyl acetate (2 : 3). $\lambda_{max(ACN)}$ = 282 nm. ^{1}H NMR (500 MHz, $CDCl_3$, ppm): 7.98 (s, 1H, H6), 7.30-7.45 (m, 10H, Ar), 6.85-6.88 (m, 3H, Ar), 6.36 (dd, 1H, H1', J = 6.0 Hz), 4.53 (m, 1H, H3'), 4.47 (m, 2H, $ArOCH_2$), 4.02 (m, 1H, H4'), 3.85 (s, 6H, OCH_3), 3.75 (t, 2H, CH_2), 3.56 (dd, 1H, H5'), 3.31 (dd, 1H, H5''), 2.56 (m, 1H, H2'), 2.27 (m, 1H, H2''), 1.90 (m, 2H, CH_2), 1.74 (m, 2H, CH_2), 1.64 (s, 1H, OH), 1.57 (s, 3H, $ArCH_3$), 0.87 (s, 9H, $SiC(CH_3)_3$), 0.08 (s, 6H, $Si(CH_3)_2$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 170.4, 158.7, 156.0 144.4, 139.5, 135.53, 135.51, 130.11, 130.10, 128.2, 127.9, 127.1, 113.24, 113.21, 104.6, 86.7, 86.5, 86.3, 71.0, 67.1, 62.4, 62.2, 42.2, 29.1, 25.7, 25.1, 17.9, 11.7, -4.62, -4.94. IR (thin film); ν_{max} (cm^{-1}) = 3414, 3061, 2953, 1669, 1509, 1252, 1178, 1109, 1035,835, 782. HRMS (ESI-MS) *m/z* calculated for $C_{41}H_{54}N_2O_8SiNa^+$ 753.3546: found 753.3555 $[M+Na]^+$

3'-*O*-(*t*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-*O*^{4}-(hydroxyheptyl)-thymidine (1b)

1,7-heptanediol (1.7 g, 11.35 mmol) was dissolved in THF (10 mL). Sodium metal (1.7 g, 65.2 mmol) was added. After 5 h, the salt was extracted to be immediately used in the subsequent reaction. 3'-*O*-(*t*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-C4-(1,2,4-triazol-1-yl)-thymidine (1.7 g, 1.13 mmol) was dissolved in dioxane (5 mL).(*221*)

Sodium 7-hydroxyheptan-1-olate was added dropwise. After 15 h, the crude product was taken up in ethyl acetate (50 mL) and the solution was washed with two portions of sodium bicarbonate (2 x 50 mL). The organic layer was dried over sodium sulphate and concentrated to produce a yellow gum. The crude product was purified by flash column chromatography using a hexanes : ethyl acetate solvent system (6 : 4). Further purification of the product was achieved by flash column chromatography using a ACN : DCM (1 : 9 → 3 : 25) solvent system to afford 0.594 g (42.1 %) of product as a colorless foam.

R_f (SiO_2 TLC): 0.43 hexanes : ethyl acetate (2 : 3). $\lambda_{max(ACN)}$ = 282 nm. ^{1}H NMR (500 MHz, $CDCl_3$, ppm): 7.96 (s, 1H, H6), 7.29-7.34 (m, 10H, Ar), 6.86-6.88 (m, 3H, Ar), 6.37 (dd, 1H, H1', J = 5.7 Hz), 4.53 (dd, 1H, H3'), 4.24 (t, 2H, $ArOCH_2$), 4.02 (m, 1H, H4'), 3.85 (s, 6H, OCH_3), 3.69 (m, 2H, CH_2OH), 3.56 (dd, 1H, H5'), 3.30 (dd, 1H, H5''), 2.55 (m, 1H, H2'), 2.25 (m, 1H, H2''), 1.80 (m, 2H, CH_2), 1.63 (m, 2H, CH_2), 1.57 (s, 3H, CH_3Ar), 1.47 (m, 6H, $(CH_2)_3$), 0.89 (s, 9H, $SiC(CH_3)_3$), 0.08 (s, 6H, $Si(CH_3)_2$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 170.5, 158.7, 156.0 144.4, 139.4, 135.56, 135.55, 130.12, 130.10, 128.2, 127.9, 127.1, 113.24, 113.21, 104.6, 86.7, 86.5, 86.3, 71.1, 67.3, 62.9, 62.4, 55.3, 42.2, 32.7, 29.8, 29.0, 28.4, 25.8, 25.7, 25.6, 17.9, 11.8, -4.62, -4.94. IR (thin film); ν_{max} (cm^{-1}) = 3420, 3060, 2930, 1668, 1509, 1252, 1177, 1109, 1035, 834, 782. HRMS (ESI-MS) *m/z* calculated for $C_{44}H_{60}N_2O_8SiNa^+$ 795.4017: found 795.4017 $[M+Na]^+$

1-{*O*^{4}-[3'-*O*-(*t*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-thymidinyl]}-4-{*O*^{4}-[3'-*O*-(*t*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-thymidinyl]}-butane (2a)

1a (0.154 g, 0.205mmol) and 3'-*O*-(*t*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-C4-(1,2,4-triazol-1-yl)-thymidine (0.117 g, 0.164 mmol) was dissolved in pyridine (3 mL). 1,8-Diazabicycloundec-7-ene (DBU) (0.062 g, 0.410 mmol) was added dropwise. After 3 days, the crude product was taken up in ethyl acetate (100 mL) and the solution was washed with four portions of sodium bicarbonate (4 x 50 mL). The organic layer was dried over sodium sulphate and concentrated to produce a yellow gum. The crude product was purified by flash column chromatography using a hexanes : ethyl acetate (1 : 1 →2 : 3) solvent system to afford 0.151 g (67.7 %) of product as a colorless foam.

R_f (SiO_2 TLC): 0.22 hexanes : ethyl acetate (2 : 3). $\lambda_{max(ACN)}$ = 282 nm. ^{1}H NMR (500 MHz, $CDCl_3$, ppm): 7.96 (s, 2H, H6), 7.30-7.46 (m, 20H, Ar), 6.86-6.88 (m, 6H, Ar), 6.37 (dd, 1H, H1', J = 5.7 Hz), 4.53 (m, 2H, H3'), 4.47 (m, 4H, CH_2), 4.01 (m, 2H, H4'), 3.84 (s, 12H, OCH_3), 3.57 (dd, 2H, H5'), 3.31 (dd, 2H, H5''), 2.55 (m, 2H, H2'), 2.26 (m, 2H, H2''), 1.93 (m, 4H, CH_2), 1.57 (s, 6H, $ArCH_3$), 0.87 (s, 18H, $SiC(CH_3)_3$), 0.07 (s, 12H, $Si(CH_3)_2$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 170.4, 158.7, 155.9 144.4, 139.6, 135.51, 135.50, 130.11, 130.10, 128.2, 127.9, 127.1, 113.24, 113.21, 104.5, 86.7, 86.5, 86.3, 71.1, 66.6, 62.4, 55.2, 42.2, 25.7, 25.3, 17.9, 11.7, -4.62, -4.94. IR (thin film); ν_{max} (cm^{-1}) = 3057, 2954, 1671, 1608, 1509, 1463, 1329, 1252, 1178, 1035, 835, 736. HRMS (ESI-MS) *m/z* calculated for $C_{79}H_{98}N_4O_{14}Si_2Na^+$ 1393.6515: found 1393.6526 $[M+Na]^+$

1-{*O*^{4}-[3'-*O*-(*t*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-thymidinyl]}-7-{*O*^{4}-[3'-*O*-(*t*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-thymidinyl]}-heptane (2b)

1b (0.500 g, 0.647 mmol) and 3'-*O*-(*t*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-C4-(1,2,4-triazol-1-yl)-thymidine (0.368 g, 0.518 mmol) was dissolved in pyridine (10 mL). DBU (0.197 g, 1.29 mmol) was added dropwise. After 4.5 days, the crude product was taken up in ethyl acetate (150 mL) and the solution was washed with four portions of sodium bicarbonate (4 x 50 mL). The organic layer was dried over sodium sulphate and concentrated to produce a yellow gum. The crude product was purified by flash column chromatography using a ACN : DCM (1 : 9 → 3 : 25) solvent system to afford 0.358 g (39.5 %) of product as a colorless foam.

R_f (SiO_2 TLC): 0.49 hexanes : ethyl acetate (2 : 3). $\lambda_{max(ACN)}$ = 282 nm. ^{1}H NMR (500 MHz, $CDCl_3$, ppm): 7.99 (s, 2H, H6), 7.33-7.49 (m, 20H, Ar), 6.89-6.91 (m, 6H, Ar), 6.40 (dd, 2H, H1', J = 12.0 Hz), 4.54 (dd, 2H, H3'), 4.44 (t, 4H, $ArOCH_2$), 4.03 (m, 2H, H4'), 3.86 (s, 12H, OCH_3), 3.58 (dd, 2H, H5'), 3.34 (dd, 2H, H5''), 2.58 (m, 2H, H2'), 2.26 (m, 2H, H2''), 1.81 (m, 4H, CH_2), 1.59 (s, 6H, CH_3Ar), 1.48 (m, 6H, $(CH_2)_3$), 0.88 (s, 18H, $SiC(CH_3)_3$), 0.07 (s, 12H, $Si(CH_3)_2$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 170.5, 158.7, 156.0 144.4, 139.4, 135.5, 130.13, 130.11, 128.2, 127.9, 127.1, 113.24, 113.21, 104.6, 86.7, 86.5, 86.3, 71.1, 67.3, 62.4, 55.3, 42.2, 29.0, 28.5, 25.9, 25.7, 17.9, 11.8, -4.61, -4.94. IR (thin film); ν_{max} (cm^{-1}) = 3058, 2952, 1671, 1533, 1508, 1327, 1252, 1177, 1035, 834. HRMS (ESI-MS) *m/z* calculated for $C_{81}H_{104}N_4O_{14}Si_2Na^+$ 1435.6985: found 1435.7037 $[M+Na]^+$

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

2a (0.658 g, 0.485 mmol) was dissolved in THF (4.8 mL). TBAF (1M in THF) (1.97 mL, 1.97 mmol) was added dropwise. After 30 min, the crude product was taken up in ethyl acetate (50 mL) and the solution was washed with two portions of sodium bicarbonate (2 x 50 mL). The organic layer was dried over sodium sulphate and concentrated to produce a yellow gum. The crude product was purified by flash column chromatography using a methanol : hexanes : ethyl acetate (1 : 5 : 4 → 1 : 4 : 5) solvent system to afford 0.514 g (93.0 %) of product as a colorless foam.

R_f (SiO_2 TLC) : 0.21 methanol : hexanes : ethyl acetate (1 : 5 : 4). $\lambda_{max(ACN)}$ = 282 nm. ^{1}H NMR (500 MHz, $CDCl_3$, ppm): 7.78 (s, 2H, H6), 7.23-7.37 (m, 20H, Ar), 6.78-6.81 (m, 6H, Ar), 6.34 (dd, 2H, H1', J = 6.0 Hz), 4.52 (m, 2H, H3'), 4.40 (m, 4H, CH_2), 4.05 (m, 2H, H4'), 3.77 (s, 12H, OCH_3), 3.46 (dd, 2H, H5'), 3.35 (dd, 2H, H5''), 2.58 (m, 2H, H2'), 2.25 (m, 2H, H2''), 2.11 (s, 2H, OH),1.85 (m, 4H, CH_2), 1.56 (s, 6H, $ArCH_3$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 170.4, 158.7, 158.6, 144.5,139.8, 135.6, 135.5, 130.1, 130.1, 128.1, 128.0, 127.0, 113.3, 105.0, 86.8, 86.5, 86.3, 66.8, 63.5, 60.4, 55.2 42.2, 25.2, 21.1, 14.2, 11.7. IR (thin film); ν_{max} (cm^{-1}) = 3353, 3057, 2953, 1661, 1509, 1431, 1329, 1251, 1177, 1095, 1035, 829, 735. HRMS (ESI-MS) *m/z* calculated for $C_{66}H_{70}N_4O_{14}Na^+$ 1165.4786: found 1165.4806 $[M+Na]^+$

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

2b (0.357 g, 0.263 mmol) was dissolved in THF (2 mL). TBAF (1M in THF) (0.63 mL, 0.63 mmol) was added dropwise. After 25 min, the crude product was taken up in

ethyl acetate (50 mL) and the solution was washed with three portions of sodium bicarbonate (3 x 50 mL). The organic layer was dried over sodium sulphate and concentrated to produce a yellow gum. The crude product was purified by flash column chromatography using a methanol : hexanes : ethyl acetate (1 : 5 : 4 → 1 : 4 : 5) solvent system to afford 0.256 g (82.0 %) of product as a colorless foam.

R_f (SiO_2 TLC): 0.22 methanol : hexanes : ethyl acetate (1:11:8). $\lambda_{max(ACN)}$ = 282 nm. ^{1}H NMR (500 MHz, $CDCl_3$, ppm): 7.77 (s, 2H, H6), 7.18-7.37 (m, 20H, Ar), 6.79-6.80 (m, 6H, Ar), 6.34 (dd, 2H, H1', J = 12.5 Hz), 4.50 (m, 2H, H3'), 4.29-4.39 (m, 4H, CH_2), 4.05 (m, 2H, H4'), 3.78 (s, 12H, OCH_3), 3.46 (dd, 2H, H5'), 3.36 (dd, 2H, H5''), 2.58 (m, 2H, H2'), 2.25 (m, 2H, H2''), 1.70-1.74 (m, 4H, CH_2), 1.56 (s, 6H, $ArCH_3$), 1.39-1.41 (m, 6H, CH_2). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 170.6, 158.7, 156.2, 144.5, 139.5, 135.6, 135.5, 130.1, 128.1, 128.0, 127.1, 113.3, 105.0, 86.8, 86.4, 86.1, 71.9, 67.4, 63.4, 55.3, 42.1, 28.5, 25.8, 11.8. IR (thin film); ν_{max} (cm^{-1}) = 3350, 3058, 2932, 1654, 1532, 1508, 1458, 1328, 1251, 1177, 1035, 829, 735. HRMS (ESI-MS) *m/z* calculated for $C_{69}H_{76}N_4O_{14}Na^+$ 1207.5255: found 1207.5270 $[M+Na]^+$

1-{*O*^{4}-[3'-*O*-(*β*-cyanoethyl-*N*,*N*'-diisopropyl)-5'-*O*-(4,4'-dimethoxytrityl)-thymidinyl]}-4-{*O*^{4}-[3'-*O*-(*β*-cyanoethyl-*N*,*N*'-diisopropyl)-5'-*O*-(4,4'-dimethoxytrityl)-thymidinyl]}-butane (4a)

3a (0.200 g, 0.175 mmol) and DIPEA (0.068 g, 0.525 mmol) were dissolved in THF (2 mL), followed by *N*,*N*-diisopropylaminocyanoethylphosphonamidic chloride (0.104 g, 0.438 mmol) and the reaction was allowed to stir at room temperature for 30 min. Upon completion, the reaction was quenched by the addition of ethyl acetate (50 mL), and the solution was washed with sodium bicarbonate (3%, 2 x 25 mL) and brine

(25 mL). The organic layer was dried over sodium sulphate, decanted, and evaporated. The product, a colorless powder, was precipitated from hexanes (0.152 g, 56.2 %).

R_f (SiO_2 TLC): 0.12, 0.18, 0.27 ethyl acetate. $\lambda_{max(ACN)}$ = 282 nm. ^{1}H NMR (500 MHz, d_6-acetone, ppm): 7.83-7.91 (s, 2H, H6), 7.47-7.50 (m, 4H, Ar), 7.21-7.39 (m, 14H, Ar), 6.87-6.91 (m, 8H, Ar), 6.29-6.34 (m, 2H, H1'), 4.70-4.74 (m, 2H, H3'), 4.37 (m, 4H, CH_2), 4.17-4.24 (m, 2H, H4'), 3.57-3.89 (m, 20H, CH_2OP, OCH_3), 3.34-3.51 (m, 4H, H5', H5''), 2.74 (t, 2H. CH_2CN), 2.52-2.63 (m, 4H, H2',CH_2CN), 2.34-2.41 (m, 2H, H2''), 1.89-1.91 (m, 4H, $(CH_2)_2$), 1.57-1.60 (m, 6H, $ArCH_3$) 1.15-1.20 (m, 18H, CH_3), 1.07-1.08 (m, 6H, CH_3). ^{31}P NMR (202.3 MHz, d_6-acetone, ppm): 148.28, 148.07. ^{13}C NMR (125.7 MHz, d_6-acetone, ppm): 170.0, 158.9, 158.8, 154.7, 144.97, 145.95, 140.03, 139.97, 135.7, 135.63, 135.60, 135.56, 130.17, 130.15, 128.2, 128.1, 127.87, 127.86, 126.91, 126.88, 118.1, 117.9, 113.1, 103.5, 103.4, 86.64, 86.61, 85.99, 85.9, 85.5, 85.3, 85.2, 73.6, 73.5, 73.1, 73.0, 66.2, 63.2, 63.0, 58.70, 58.69, 54.70, 54.68, 43.1, 43.0, 40.4, 40.3, 29.4, 25.1, 24.05, 24.03, 23.99, 23.93, 19.9, 19.8, 11.18, 11.17. IR (thin film); ν_{max} (cm^{-1}) = 3031, 2966, 1672, 1534, 1509, 1462, 1329, 1251, 1179, 1035. HRMS (ESI-MS) *m/z* calculated for $C_{84}H_{104}N_8O_{16}P_2Na^+$ 1565.6943: found 1565.6854 $[M+Na]^+$

1-{*O*^{4}-[3'-*O*-(*β*-cyanoethyl-*N*,*N*'-diisopropyl)-5'-*O*-(4,4'-dimethoxytrityl)thymidinyl]}-7-{*O*^{4}-[3'-*O*-(*β*-cyanoethyl-*N*,*N*'-diisopropyl)-5'-*O*-(4,4'-dimethoxytrityl)-thymidinyl]}-heptane (4b)

3b (0.170 g, 0.143 mmol) and DIPEA (0.056 g, 0.430 mmol) were dissolved in THF (2 mL), followed by *N*,*N*-diisopropylaminocyanoethylphosphonamidic chloride (0.085 g, 0.359 mmol) and the reaction was allowed to stir at room temperature for 30 min. Upon completion, the reaction was quenched by the addition of ethyl acetate (50

mL), and the solution was washed with sodium bicarbonate (3%, 2 x 25 mL) and brine (25 mL). The organic layer was dried over sodium sulphate, decanted, and evaporated. The product, a colorless powder, was precipitated from hexanes (0.146 g, 64.3 %).

R_f (SiO_2 TLC): 0.31, 0.46, 0.62 ethyl acetate. $\lambda_{max(ACN)}$ = 282 nm. ^{1}H NMR (500 MHz, d_6-acetone, ppm): 7.88-7.92 (s, 2H, H6), 7.48-7.51 (m, 4H, Ar), 7.23-7.40 (m, 14H, Ar), 6.87-6.93 (m, 8H, Ar), 6.30-6.35 (m, 2H, H1'), 4.70-4.76 (m, 2H, H3'), 4.30-4.32 (m, 4H, CH_2), 4.18-4.24 (m, 2H, H4'), 3.60-3.91 (m, 20H, CH_2OP, OCH_3), 3.39-3.52 (m, 4H, H5', H5''), 2.77 (t, 2H, CH_2CN), 2.53-2.64 (m, 4H, H2',CH_2CN), 2.37-2.42 (m, 2H, H2''), 1.74-1.78 (m, 4H, $(CH_2)_2$), 1.58-1.61 (m, 6H, $ArCH_3$) 1.44-1.49 (m, 6H, $(CH_2)_3$), 1.18-1.21 (m, 18H, CH_3), 1.09-1.11 (m, 6H, CH_3). ^{31}P NMR (202.3 MHz, d_6-acetone, ppm): 148.28, 148.07. ^{13}C NMR (125.7 MHz, d_6-acetone, ppm): 170.09, 170.08 158.88, 158.86, 154.7, 145.0, 140.0, 139.9, 135.7, 135.63, 135.61, 135.57, 130.2, 130.1, 128.2, 128.1, 127.9, 127.8, 126.89, 126.86, 118.1, 117.9, 113.1, 103.5, 103.4, 86.62, 86.59, 86.0, 85.9, 85.5, 85.3, 85.2, 73.6, 73.5, 73.1, 73.0, 66.6, 63.2, 63.0, 58.7, 58.5, 54.68, 54.66, 43.1, 43.0, 40.4, 40.3, 29.4, 25.7, 24.04, 24.02, 23.97, 23.96, 23.9, 19.94, 19.89, 19.83, 19.77, 13.6, 11.17, 11.16. IR (thin film); ν_{max} (cm^{-1}) = 3057, 2966, 1672, 1510, 1463, 1329, 1252, 1179, 1036. HRMS (ESI-MS) *m/z* calculated for $C_{87}H_{110}N_8O_{16}P_2Na^+$ 1607.7412: found 1607.7285 $[M+Na]^+$

3'-*O*-(*t*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-*O*^{4}-(phenoxyacetyloxybutyl)-thymidine (5a)

1a (0.300 g, 0.410 mmol) was dissolved in THF (1.2 mL). Triethylamine (0.145 g, 1.43 mmol) followed by phenoxyacetyl chloride (0.122 g, 0.715 mmol) was added dropwise. The reaction was set in an ice bath for 25 min. After 1.5 h, the crude product

was taken up in ethyl acetate (50 mL) and the solution was washed with three portions of sodium bicarbonate (3 x 50 mL). The organic layer was dried over sodium sulphate and concentrated to produce a yellow gum. The crude product was purified by flash column chromatography using a hexanes : ethyl acetate (3 : 7) solvent system to afford 0.266 g (75.1 %) of product as a colorless foam.

R_f (SiO_2 TLC): 0.39 hexanes : ethyl acetate (1 : 1). $\lambda_{max(ACN)}$ = 276 nm. ^{1}H NMR (500 MHz, $CDCl_3$, ppm): 7.98 (s, 1H, H6), 7.26-7.50 (m, 13H, Ar), 6.85-7.09 (m, 5H, Ar), 6.37 (dd, 1H, H1', J = 6.0 Hz), 4.69 (s, 2H, $PhOCH_2CO$), 4.53 (m, 1H, H3'), 4.42 (m, 2H, $ArOCH_2$), 4.30 (t, 2H, $COCH_2$),4.01 (m, 1H, H4'), 3.84 (s, 6H, OCH_3), 3.58 (dd, 1H, H5'), 3.31 (dd, 1H, H5''), 2.55 (m, 1H, H2'), 2.26 (m, 1H, H2''), 1.84 (m, 4H, CH_2), 1.57 (s, 3H, $ArCH_3$), 0.87 (s, 9H, $SiC(CH_3)_3$), 0.08 (s, 6H, $Si(CH_3)_2$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 170.3, 169.0, 158.7, 157.8, 155.6, 144.4, 139.7, 135.52, 135.50, 130.11, 130.09, 129.6, 128.2, 127.9, 127.1, 121.8, 114.6, 113.24, 113.21, 104.4, 86.7, 86.5, 86.3, 71.1, 66.5, 65.3, 64.8, 62.4, 55.2, 42.2, 25.7, 25.3, 25.1, 17.9, 11.7, -4.62, -4.94. IR (thin film); ν_{max} (cm^{-1}) = 3061, 2954, 1760, 1672, 1532, 1509, 1497, 1252, 1178, 835, 783. HRMS (ESI-MS) *m/z* calculated for $C_{49}H_{60}N_2O_{10}SiNa^+$ 887.3914: found 887.3921$[M+Na]^+$

3'-*O*-(*t*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-*O*^{4}-(phenoxyacetyloxy-heptyl)-thymidine (5b)

1b (0.527 g, 0.682 mmol) was dissolved in THF (0.5 mL). Triethylamine (0.110 g, 1.09 mmol) was added followed by cooling of the reaction in an ice bath for 10 min. Phenoxyacetyl chloride (0.175 g, 1.02 mmol) was added dropwise. After 1.5 h, the solvent was evaporated in vacuo, the crude product was taken up in DCM (50 mL) and

the solution was washed with two portions of sodium bicarbonate (2 x 50 mL). The organic layer was dried over sodium sulphate and concentrated to produce a yellow gum. The crude product was purified by flash column chromatography using a hexanes : ethyl acetate (4 : 1 → 2 : 3) solvent system to afford 0.491 g (79.4 %) of product as a colorless foam.

R_f (SiO_2 TLC): 0.49 hexanes : ethyl acetate (1 : 1). $\lambda_{max(ACN)}$ = 276 nm. ^{1}H NMR (500 MHz, $CDCl_3$, ppm): 8.00 (s, 1H, H6), 7.45 (dd, 2H, Ar), 7.25-7.34 (m, 9H, Ar), 6.85-7.04 (m, 7H, Ar), 6.37 (dd, 1H, H1', J = 6.0 Hz), 4.68 (s, 2H, $PhOCH_2CO$), 4.53 (dd, 1H, H3'), 4.41 (t, 2H, $ArOCH_2$), 4.25 (t, 2H, CH_2OCO), 4.02 (m, 1H, H4'), 3.84 (s, 6H, OCH_3), 3.58 (dd, 1H, H5'), 3.30 (dd, 1H, H5''), 2.55 (m, 1H, H2'), 2.26 (m, 1H, H2''), 1.77 (quintet, 2H, CH_2), 1.71 (quintet, 2H, CH_2), 1.58 (s, 3H, CH_3Ar), 1.37-1.46 (m, 6H, $(CH_2)_3$), 0.88 (s, 9H, $SiC(CH_3)_3$), 0.07 (s, 6H, $Si(CH_3)_2$. ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 170.5, 169.1, 158.7, 157.8, 156.0, 144.4, 139.4, 135.5, 130.11, 130.09, 129.5, 128.2, 127.9, 127.1, 121.7, 114.6, 113.23, 113.21, 104.5, 86.7, 86.5, 86.3, 71.1, 67.2, 65.4, 65.3, 62.4, 55.2, 42.2, 28.8, 28.5, 28.4, 25.8, 25.71, 25.69, 17.9, 11.8, -4.62, -4.95. IR (thin film); ν_{max} (cm^{-1}) = 3062, 2928, 1760, 1675, 1510, 1464, 154, 1065, 832, 734. HRMS (ESI-MS) *m/z* calculated for $C_{52}H_{66}N_2O_{10}SiNa^+$ 929.4382: found 929.4384 $[M+Na]^+$

5'-*O*-(4,4'-dimethoxytrityl)-*O*^{4}-(phenoxyacetyloxybutyl)-thymidine (6a)

5a (0.200 g, 0.231 mmol) was dissolved in THF (2.4 mL). TBAF (1M in THF) (0.277 mL, 0.277 mmol) was added dropwise. After 15 min, the crude product was taken up in ethyl acetate (50 mL) and the solution was washed with two portions of sodium bicarbonate (2 x 50 mL). The organic layer was dried over sodium sulphate and

concentrated to produce a yellow gum. The crude product was purified by flash column chromatography using a methanol : hexanes : ethyl acetate (2 : 11 : 7) solvent system to afford 0.164 g (94.1 %) of product as a colorless foam.

R_f (SiO_2 TLC): 0.29 methanol : hexane : ethyl acetate (2 : 11 : 7). $\lambda_{max(ACN)}$ = 276 nm. ^{1}H NMR (500 MHz, $CDCl_3$, ppm): 7.79 (s, 1H, H6), 7.18-7.37 (m, 13H, Ar), 6.79-6.96 (m, 5H, Ar), 6.34 (dd, 1H, H1', J = 6.5 Hz), 4.61 (s, 2H, $PhOCH_2CO$), 4.50 (m, 1H, H3'), 4.36 (m, 2H, $ArOCH_2$), 4.23 (t, 2H, $COCH_2$),4.05 (m, 1H, H4'), 3.77 (s, 6H, OCH_3), 3.47 (dd, 1H, H5'), 3.36 (dd, 1H, H5''), 2.58 (m, 1H, H2'), 2.23 (m, 1H, H2''), 1.75-1.78 (m, 4H, CH_2), 2.04 (s, 1H, OH), 1.55 (s, 3H, $ArCH_3$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 170.4, 169.0, 158.7, 158.6, 157.8, 156.1, 144.5, 139.9, 135.6, 135.5, 130.08, 130.07, 129.6, 128.1, 128.0, 127.0, 121.8, 114.6, 113.3, 104.9, 86.8, 86.5, 86.4, 72.0, 66.7, 65.3, 64.8, 63.5, 55.2, 42.2, 25.3, 25.1, 11.7. IR (thin film); ν_{max} (cm^{-1}) = 3342, 3061, 2956, 1758, 1663, 1533, 1509, 1251, 1177, 1090, 755. HRMS (ESI-MS) *m/z* calculated for $C_{43}H_{46}N_2O_{10}Na^+$ 773.3049: found 773.3055 $[M+Na]^+$

5'-*O*-(4,4'-dimethoxytrityl)-O^4-(phenoxyacetyloxyheptyl)-thymidine (6b)

5b (0.491 g, 0.541 mmol) was dissolved in THF (4.5 mL). TBAF (1M in THF) (0.65 mL, 0.65 mmol) was added dropwise. After 25 min, the solvent was evaporated in vacuo, the crude product was taken up in DCM (50 mL) and the solution was washed with two portions of sodium bicarbonate (2 x 50 mL). The organic layer was dried over sodium sulphate and concentrated to produce a yellow gum. The crude product was purified by flash column chromatography using a methanol : hexanes : ethyl acetate (0 : 1 : 1→ 1 : 19 : 30) solvent system to afford 0.373 g (86.8 %) of product as a colorless foam.

$\lambda_{max(ACN)}$ = 276 nm. ^{1}H NMR (500 MHz, $CDCl_3$, ppm): 8.00 (s, 1H, H6), 7.29-7.40 (m, 2H, Ar), 7.18-7.27 (m, 9H, Ar), 6.94-6.97 (m, 1H, Ar), 6.87-6.88 (m, 2H, Ar), 6.79-6.81 (m, 4H, Ar), 6.37 (dd, 1H, H1', J = 6.2 Hz), 4.61 (s, 2H, $PhOCH_2CO$), 4.52 (m, 1H, H3'), 4.34 (t, 2H, $ArOCH_2$), 4.18 (t, 2H, CH_2OCO), 4.07 (m, 1H, H4'), 3.78 (s, 6H, OCH_3), 3.47 (dd, 1H, H5'), 3.36 (dd, 1H, H5''), 2.60 (m, 1H, H2'), 2.33 (s, 1H, OH), 2.25 (m, 1H, H2''), 1.72 (quintet, 2H, CH_2), 1.64 (quintet, 2H, CH_2), 1.56 (s, 3H, CH_3Ar), 1.33-1.37 (m, 6H, $(CH_2)_3$). ^{13}C NMR (125.7 MHz, $CDCl_3$, ppm): 170.6, 169.1, 158.6, 157.8, 156.3, 144.5, 139.7, 135.6, 135.5, 130.09, 130.07, 129.5, 128.1, 127.9, 127.0, 121.7, 114.6, 113.2, 105.0, 86.8, 86.5, 86.3, 72.0, 67.3, 65.4, 65.3, 63.5, 55.2, 42.2, 28.8, 28.5, 28.4, 25.8, 25.7, 11.7. IR (thin film); ν_{max} (cm^{-1}) = 3346, 3060, 2934, 1757, 1663, 1532, 1509, 1251, 1177, 1091, 735. HRMS (ESI-MS) *m/z* calculated for $C_{46}H_{52}N_2O_{10}Na^+$ 815.3519: found 815.3517 $[M+Na]^+$

3'-*O*-(*β*-cyanoethyl-*N*,*N*'-diisopropyl)-5'-*O*-(4,4'-dimethoxytrityl)-O^4-(phenoxyacetyloxybutyl)-thymidine (7a)

6a (0.200 g, 0.266 mmol) and DIPEA (0.051 g, 0.398 mmol) were dissolved in THF (2.5 mL), followed by *N*,*N*-diisopropylaminocyanoethylphosphonamidic chloride (0.076 g, 0.319 mmol). The reaction was allowed to stir at room temperature for 30 min, where TLC revealed the presence of starting material. DIPEA (0.006 g, 0.049 mmol) followed by *N*,*N*-diisopropylaminocyanoethylphosphonamidic chloride (0.011 g, 0.045 mmol) were added and the reaction was allowed to proceed for another 30 min. Upon completion, the reaction was quenched by the addition of ethyl acetate (50 mL), and the solution was washed with sodium bicarbonate (3%, 2 x 25 mL) and brine (25 mL). The

organic layer was dried over sodium sulphate, decanted, and evaporated. The product, a colorless powder, was precipitated from hexanes (0.116 g, 46.2 %).

R_f (SiO_2 TLC): 0.49, 0.67 hexanes : ethyl acetate (2 : 8). $\lambda_{max(ACN)}$ = 276 nm. ^{1}H NMR (500 MHz, d_6-acetone, ppm): 7.90-7.94 (s, 1H, H6), 7.46-7.51 (m, 2H, Ar), 7.24-7.39 (m, 9H, Ar), 6.89-6.96 (m, 7H, Ar), 6.31-6.36 (m, 1H, H1'), 4.73-4.75 (m, 3H, H3', $PhOCH_2CO$), 4.33 (m, 2H, $ArOCH_2$), 4.18-4.24 (m, 3H, H4' & CH_2OCO), 3.60-3.89 (m, 10H, CH_2OP, OCH_3), 3.40-3.50 (m, 2H, H5', H5''), 2.76 (t, 2H. CH_2CN), 2.54-2.63 (m, 4H, H2',CH_2CN), 2.35-2.42 (m, 1H, H2''), 1.79-1.81 (m, 4H, $(CH_2)_2$), 1.57-1.61 (m, 3H, $ArCH_3$), 1.17-1.23 (m, 9H, CH_3), 1.09-1.10 (m, 3H, CH_3). ^{31}PNMR (202.3 MHz, d_6-acetone, ppm): 148.28, 148.08. ^{13}C NMR (125.7 MHz, d_6-acetone, ppm): 170.00, 169.99, 168.6, 158.9, 158.2, 154.7, 144.9, 140.1, 140.0, 135.7, 135.64, 135.62, 135.58, 130.18, 130.15, 129.4, 128.2, 128.1, 127.9, 127.0, 126.9, 121.1, 117.9, 114.5, 113.1, 103.5, 103.4, 86.64, 86.60, 86.0, 85.9, 85.5, 85.3, 85.2, 73.1, 66.1, 64.3, 63.2, 63.0, 58.7, 58.5, 54.70, 54.68, 43.1, 43.0, 40.4, 40.3, 25.1, 25.0, 24.05, 24.03, 23.99, 23.9, 19.9, 19.84, 19.79, 13.6, 11.17, 11.16. IR (thin film); ν_{max} (cm^{-1}) = 3060, 2966, 1758, 1671, 1607, 1508, 1330, 1251, 1180, 1034, 755. HRMS (ESI-MS) *m/z* calculated for $C_{52}H_{63}N_4O_{11}PNa^+$ 973.4128: found 973.4131, $[M+Na]^+$

3'-*O*-(*β*-cyanoethyl-*N*,*N*'-diisopropyl)-5'-*O*-(4,4'-dimethoxytrityl)-*O*^{4}-(phenoxyacetyloxyheptyl)-thymidine (7b)

6b (0.205 g, 0.242 mmol) and DIPEA (0.050 g, 0.390 mmol) were dissolved in THF (1 mL), followed by *N*,*N*-diisopropylaminocyanoethylphosphonamidic chloride (0.073 g, 0.310 mmol). The reaction was allowed to stir at room temperature for 30 min, where TLC revealed the presence of starting material. DIPEA (0.006 g, 0.049 mmol)

followed by *N*,*N*-diisopropylaminocyanoethylphosphonamidic chloride (0.011 g, 0.045 mmol) were added and the reaction was allowed to react another 30 min. Upon completion, the reaction was quenched by the addition of ethyl acetate (50 mL), and the solution was washed with sodium bicarbonate (3%, 2 x 25 mL) and brine (25 mL). The organic layer was dried over sodium sulphate, decanted, and evaporated. The product, a colorless powder, was precipitated from hexanes (0.079 g, 32.2 %).

R_f (SiO_2 TLC): 0.70, 0.83 hexanes : ethyl acetate (2 : 8). $\lambda_{max(ACN)}$ = 276 nm. ^{1}H NMR (500 MHz, d_6-acetone, ppm): 7.90-7.93 (s, 1H, H6), 7.46-7.51 (m, 2H, Ar), 7.24-7.39 (m, 9H, Ar), 6.89-6.97 (m, 7H, Ar), 6.31-6.36 (m, 1H, H1'), 4.72-4.75 (m, 3H, H3', $PhOCH_2CO$), 4.29 (m, 2H, $ArOCH_2$), 4.16-4.25 (m, 3H, H4' & CH_2OCO), 3.60-3.90 (m, 10H, CH_2OP, OCH_3), 3.40-3.52 (m, 2H, H5', H5''), 2.76 (t, 2H. CH_2CN), 2.53-2.63 (m, 4H, H2',CH_2CN), 2.34-2.42 (m, 1H, H2''), 1.71-1.77 (m, 2H, CH_2), 1.57-1.68 (m, 5H, CH_2 & $ArCH_3$), 1.37-1.44 (m, 6H, $(CH_2)_3$), 1.16-1.23 (m, 9H, CH_3), 1.09-1.10 (m, 3H, CH_3). ^{31}PNMR (202.3 MHz, d_6-acetone, ppm): 148.28, 148.08. ^{13}C NMR (125.7 MHz, d_6-acetone, ppm): 170.09, 170.08, 168.6, 158.9, 158.2, 154.7, 145.0, 140.0, 139.9, 135.7, 135.64, 135.62, 135.58, 130.18, 130.16, 129.4, 128.20, 128.15, 127.87, 127.86, 126.91, 126.88, 121.2, 118.1, 117.9, 114.5, 113.2, 103.54, 103.46, 86.64, 86.61, 86.0, 85.9, 85.51, 85.47, 85.3, 73.6, 73.5, 73.1, 66.6, 64.8, 64.6, 63.2, 63.0, 59.7, 58.7, 58.5, 54.71, 54.69, 43.1, 43.0, 40.4, 40.3, 25.7, 25.5, 24.1, 24.04, 24.01, 23.9, 19.91, 19.85, 19.8, 13.7, 11.20, 11.18, 11.14. IR (thin film); ν_{max} (cm^{-1}) = 3060, 2965, 1758, 1671, 1607, 1509, 1252, 1180, 1086, 1035, 755. HRMS (ESI-MS) *m/z* calculated for $C_{55}H_{69}N_4O_{11}PNa^+$ 1015.4597: found 1015.4620 $[M+Na]^+$

3.2.3 Oligonucleotide synthesis and purification

DNA duplexes (**XLTT4** and **XLTT7**, whose sequences are shown in **Figure 3.1**) and single stranded DNA (control strands, O^4-butyl-4-ol dT (**T4**) and O^4-heptyl-7-ol dT (**T7**)), were assembled on an Applied Biosystems Model 3400 synthesizer on a 1 μmole scale employing standard β-cyanoethylphosphoramidite cycles as described in **Section 2.2.3** with the modifications indicated below.

The modified 3'-*O*-2'-deoxyphosphoramidites (**7a** and **7b**) were dissolved in anhydrous ACN to a concentration of 0.15 M and the modified bis-3'-*O*-2'-deoxyphosphoramidites (**4a** and **4b**) to 0.05 M. The coupling times for the modified phosphoramidites **7a** and **7b** were 10 min. For the modified bis-phosphoramidite **4a** and **4b** the coupling times were 30 min.

The oligomer-derivatized CPG support was removed from the column and placed into screw cap microfuge tubes fitted with teflon lined caps. The oligonucleotides containing O^4dT modifications were removed from the support by incubation of the CPG in 500 μL of a 10% DBU in the respective alcohol (1,4-butanediol for O^4-butyl-4-ol dT and 1,7-heptanediol for O^4-heptyl-7-ol dT), or ethanol for cross-linked oligonucleotides, for 5 days in the dark at room temperature. The DBU was neutralized with acetic acid and the DNA solubilised in ACN prior to being transferred into clean vials to eliminate the CPG. The DNA was dried down in a speed vacuum and then desalted using C-18 SEP PAK cartridges (Waters Inc.) prior to purification.

The cross-link and mono-adduct final products were purified from pre-terminated products by 20% 7 M urea denaturing PAGE (19 : 1) using 1 X TBE (89 mM Tris HCl, 89 mM boric acid, 2 mM EDTA (pH 8.0)) as running buffer to eliminate any residual

DBU, which could not be removed by SAX-HPLC. Prior to gel separation, 25 O.D. of each sample were dried *via* a speed-vac concentrator and resuspended in 100 μL of formamide. The bands corresponding to the desired products were excised from the gels and placed into 15 mL falcon tubes and the gel slices submerged in 0.1 M sodium acetate. The tubes were allowed to shake overnight and the extracted oligomers desalted using C-18 SEP PAK cartridges.

3.2.4 Oligonucleotide characterization by ESI-MS and nuclease digestion

ESI mass spectra and exonuclease analysis for oligonucleotides were obtained as described in **Section 2.2.4**.

The retention times of the eluted peaks were compared to the standard nucleotides which eluted at the following times: dC (4.4 min), dG (7.5 min), dT (8.2 min), dA (9.2 min), O^6-MedG (12.0 min), O^4-MedT (13.4 min), O^4-Butyl-4-ol dT (15.3 min), O^4-Heptyl-7-ol dT (25.6 min) and the cross-link dimers (21.7 for the four carbon and 29.9 min for the seven carbon cross-link), and the ratio of nucleosides was determined. The molecular weights of the modified oligomers were determined by ESI-MS, which correlated with the calculated mass.

3.2.5 UV thermal denaturation studies

Thermal denaturation studies were performed and processed as described in **Section 2.2.5**.

Thermodynamic parameters of the non-crosslinked control DNA were obtained by performing thermal denaturation studies in triplicate at concentrations of 0.6, 2.7, 12.6

and 58.5 μM duplex DNA in 1, 0.5, 0.2 and 0.1 cm cuvettes, respectively. A plot of $\ln(C_{tot})$ as a function of $1/T_m$, typically used for a non-self-complimentary bimolecular systems, was employed to obtain ΔH° and ΔS° for the system as described by Puglisi and Tinoco.(*191*) A similar procedure was employed with the cross-linked DNA at concentrations of 0.6 and 2.7 μM. The T_m of the ICL were independent of oligonucleotide concentration due to the uni-molecular nature of the DNA and therefore the thermodynamic parameters were obtained by plotting ln(K) versus 1/T at a single concentration of 2.95 μM. The experiments were performed in triplicate. (see Appendix III).

3.2.6 CD spectroscopy

CD experiments were carried out as described as in **Section 2.2.6**.

3.2.7 Protein expression and purification

All proteins were purified under native conditions as described in **Section 2.2.8**.

3.2.8 Mono-adduct repair assay

A general oligonucleotide sequence was created with a *BclI* cut site to monitor the amount of repair of mono-adduct based on restriction digestion. The sequence of the damage containing strand was 5' GGC TXG ATC ACC AG where: X represents dT (control sequence), O^4-MedT, O^4-butyl-4-ol dT (**T4**) or O^4-heptyl-7-ol dT (**T7**). The sequence of the complement strand was 5' CTG GT/G ATC AAG CC where the "/" represent the *BclI* cut site.

The damage containing strand was ^{32}P labelled at its 5' end as previously described.(*14*) Briefly, a 20 μM solution of DNA was made in 1X PNK buffer along with

1 μL [γ-^{32}P]ATP (10 μCi/μL) 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 10 min.

40 pmol of labelled DNA was added to 50 pmol of the complement strand in a total volume of 20 μL of water making a 2 μM dsDNA solution with 20 % excess of the non-damaged strand. The solution was boiled for 5 min, cooled slowly to room temperature and placed at 4 °C overnight to ensure proper duplex formation.

The repair reaction mixtures were constituted of 2 pmol of the duplex DNA and 10 pmol of AGT in a total volume of 15 μL in Activity Buffer (10 mM Tris HCl (pH 7.6), 100 mM NaCl and 1 mM DTT) and allowed to react at 37 °C for 30 min, unless otherwise indicated. The reaction was terminated by boiling for 10 min. Prior to *BclI* digestion, $MgCl_2$ was added at a final concentration of 10 mM and the reaction mixture allowed to cool to room temperature. 7.5 units (0.5 μL) of *BclI* was added to the mixture and the solution incubated for 1 h at 37 °C. 24 μL of stop buffer (81 mM Tris HCl, 81 mM boric acid, 1.8 mM EDTA and 1% SDS (pH 8.0) in 80 % formamide) was added, boiled for 10 min and loaded on a 14 cm × 16 cm, 20 % 7 M urea denaturing polyacrylamide gel (19 : 1) for separation. The gels were run using 1 X TBE for 40 min 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 ImageQuant™ (Amersham Biosciences).

3.2.9 Binding studies of C145S hAGT variant with mono-adduct substrates

Binding reactions of the 14 bp mono-adduct DNA substrates consisted of 0.5 nM dsDNA (where an excess of 5% of undamaged strand was present to force duplex formation) and increasing C145S hAGT (ranging from 1 to 35.69 μM) in a total solution volume of 20 μl of binding buffer (10 mM Tris HCl (pH 7.6), 100 mM NaCl, 1 mM DTT, 10 μg mL^{-1} BSA and 2.5 % glycerol). The samples were equilibrated for 30 min at room temperature and loaded on a 10 % native polyacrylamide gel (75 : 1) and pre-ran with 10 mM Tris acetate (pH 7.6) and 100 mM NaCl. Electrophoresis was carried out at 21 °C at 100 V for 45 min.

3.2.10 Cross-link repair assay

Cross-link repair assays were performed as described in **Section 2.2.10**.

3.2.11 Binding studies of C145S hAGT variant with cross-linked substrates

Binding reactions were conducted as for the mono-adduct substrates with minor modifications due to the different substrate size. Binding reactions consisted of 0.5 nM dsDNA, increasing C145S AGT ranging from 1 to 45.5 μM for the control duplex and from 1 to 6.5 μM for the cross-linked DNA in a total solution volume of 20 μl of binding buffer. The samples were equilibrated for 1 h at room temperature and loaded on a 15 % native polyacrylamide gel (75:1) and ran with 0.5 X TBE, which allowed the electrophoresis voltage to be increased without heating the system. Electrophoresis was carried out at 21 °C at 250 V for 25 min.

The monomeric dissociation constant (K_d) and stoichiometry (n) of AGT binding to DNA were obtained from the EMSA as described previously.(*14*) The binding of n molecules of AGT protein [P] to 1 molecule of DNA [D] can be expressed by the equation: nP + D ↔ PnD. Taking the logarithm and rearranging the variables yields: log ([PnD]/[D]) = n log [P]free + log K_d.

The data can be expressed with a plot of log [PnD]/[P] as a function of log [P]. The slope of the graph represents the stoichiometry (n) and the observed K_d can be obtained by the relationship: $K_d = 10^{-(\text{y-intercept})}$. The monomeric K_d shown in the tables was obtained by taking the nth root of the observed K_d.

3.2.12 Modelling of O^4-MedT and O^4-butyl-4-ol dT in hAGT active site

The alkylated DNA was placed in the active site of hAGT based on PDB 1T38 by mutating O^6-MedG to O^4-MedT. The placement of O^4-MedT was verified with PDB 1YFH. PDB 1T38 was used as the template for hAGT where Ser 145 was mutated back to Cys and the proper rotamer used as based on PDBs 1T39, 1EH6 and 1YFH. This primary model was used as the template for the addition of the butyl-4-ol adduct, which was added manually by PyMol.

The preliminary models underwent conjugate gradient minimization, simulated annealing, and torsion angle dynamics using Crystallography & NMR System.(*222*) CHARMM (Chemistry at HARvard Macromolecular Mechanics) topology files were modified to use the altered nucleotides by adding restraints to keep the α-carbon of the

lesion in a *syn* conformation with respect to the N3 atom of the base, as observed in PBDs 1T38, 1T39, and 1YFH.(*223*)

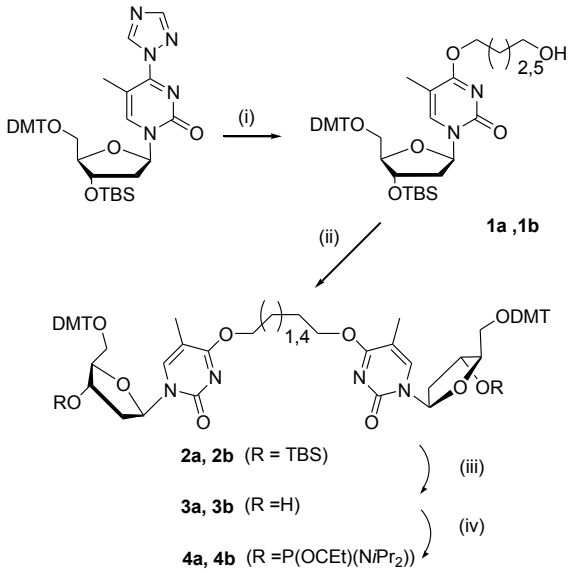
The models were subjected to 2 sets of 100 steps of gradient minimization and 200 steps of simulated annealing. The first cycle was carried out at 500 K with slow cooling at a rate of 4.5 K per cycle of dynamics. The second cycle was carried out at 350 K with slow cooling at a rate of 2.5 K per cycle of dynamics. The minimized models were displayed using PyMol with the absence of a few amino acids for clarity of the active site.

3.3 Results and discussion

3.3.1 Chemical synthesis of modified nucleosides

The structures of the O^4dT-alkylene-O^4dT cross-links and the mono-adducts are shown in **Figure 3.1** in addition to the sequences of the respective cross-links (**XLTT4** and **XLTT7**) and single stranded oligonucleotides (**T4** and **T7**). The synthesis of the O^4dT-alkylene-O^4dT cross-linked bis-phosphoramidites **4a** and **4b**, containing a four and seven carbon linker, is shown in **Scheme 3.1**. The convertible nucleoside 3'-*O*-(*t*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-C4-(1,2,4-triazol-1-yl)-thymidine was transformed into adducts **1a** or **1b** with the sodium salt of 1,4-butanediol or 1,7-heptanediol, respectively.(*221*) Reacting either **1a** or **1b** with the convertible nucleoside yielded the fully protected dimers **2a** and **2b**. The 3'-OH groups of these compounds, protected as silyl ethers, were liberated by treatment with TBAF to produce compounds **3a** and **3b** which were then phosphitylated with *N*,*N*-diisopropylaminocyanoethylphosphonamidic chloride in the presence of DIPEA to

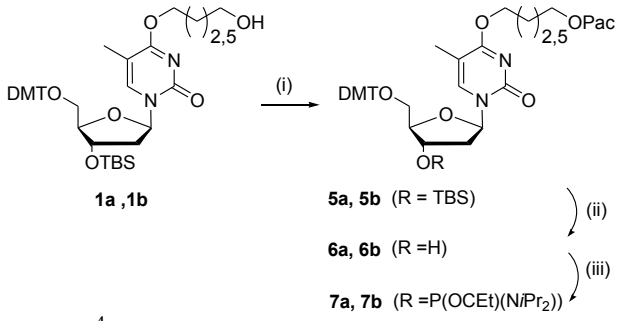
produce bis-phosphoramidites **4a** and **4b**. ^{31}P NMR signals at 148.07 and 148.28 ppm for **4a** and 148.07 and 148.28 ppm for **4b** confirmed the presence of the phosphoramidite functionality in these molecules.



Scheme 3.1: Synthesis of O^4dT-butylene-O^4dT and O^4dT-heptylene-O^4dT bis-phosphoramidites

Reagents and conditions: (i) sodium salt of 1,4-butanediol or 1,7-heptanediol, dioxane. (ii) 3'-*O*-(*t*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-C4-(1,2,4-triazol-1-yl)-thymidine, DBU, pyridine.(iii) TBAF (1M in THF), THF. (iv) *N*,*N*-diisopropylaminocyanoethylphosphonamidic chloride, DIPEA, THF.

The mono-adduct lesions were introduced at the O^4 atom of dT as shown in **Scheme 3.2**. Starting with compounds **1a** or **1b**, the free terminal alcohol functionality of these adducts were protected with phenoxyacetyl chloride to yield **5a** and **5b**. The 3'-OH group of these nucleosides were exposed prior to the phosphitylation step by treatment with TBAF to yield **6a** and **6b**. They were then converted to phosphoramidites as confirmed by the presence of two diagnostic signals in ^{31}P NMR at 148.08 and 148.28 ppm for **7a** and 148.08 and 148.28 ppm for **7b**.



Scheme 3.2: Synthesis of O^4-alkyl dT mono-adduct amidites

Reagents and conditions: (i) phenoxyacetyl chloride, triethylamine, THF. (ii) TBAF (1M in THF), THF. (iii) *N,N*-diisopropylaminocyanoethylphosphonamidic chloride, DIPEA, THF.

3.3.2 Synthesis of modified and ICL DNA

The solid-phase synthesis of cross-linked duplexes **XLTT4** and **XLTT7** were performed on a 1 μmol scale using either bis-phosphoramidite **4a** or **4b** to prepare the oligomers containing the four or seven carbon cross-link, respectively. Given the lability of the cross-link, 'fast-deprotecting' 3'-*O*-2'-deoxyphosphoramidites were employed (0.1 M in ACN) with phenoxyacetic anhydride as the capping agent to prevent an undesired *N*-acetylation reaction that could occur if acetic anhydride was used.(*196*) Moreover, the directly opposed motif allows for ICL duplex synthesis without the need to remove additional protecting groups around the cross-linked site which could compromise the ICL or the use of 5'-*O*-2'-deoxyphosphoramidites. The 3'-*O*-bis-phosphoramidite concentrations of **4a** and **4b** in ACN were 0.05 M and allowed to couple for an extended period (30 min) to ensure both phosphoramidite functionalities would couple to two oligonucleotide chains on the solid-support, to ensure ICL formation.(*175*,*224*) Solid-phase assembly of the duplex was completed by extension with 3'-*O*-2'-deoxyphosphoramidites at the cross-linked site.

A variation to the standard oligonucleotide deprotection protocol was required due to the labile nature of O^4-alkyl dT adducts. These milder deprotection conditions involved a solution of 10% DBU in anhydrous ethanol for 48 h at room temperature to reduce the likelihood of ICL cleavage linking the O^4 atoms upon deprotection.(*221*) The crude deprotection mixture containing duplexes of **XLTT4** and **XLTT7** were purified by PAGE, rather than SAX HPLC, due to difficulties in removing excess DBU used in the deprotection step. The purified oligonucleotides were eluted from the gels and desalted. ESI mass spectrometry analysis of **XLTT4** and **XLTT7** revealed they had molecular weights of 6715.6 and 6758.3 Da (expected 6716.5 and 6758.6 Da), consistent with the expected values. The composition of the cross-linked oligomers **XLTT4** and **XLTT7** were further confirmed by digestion to their constituent nucleosides with a mixture of snake venom phosphodiesterase and calf intestinal alkaline phosphatase followed by C-18 reversed phase HPLC analysis. The presence of cross-linked nucleosides was established by the appearance of one additional peak, other than the four standard nucleosides, that eluted from the HPLC with retention times of 21.7 for the four carbon and 29.9 min for the seven carbon O^4dT-alkylene-O^4dT cross-link. The later elution of the seven carbon cross-link can be attributed to the greater hydrophobicity of the longer alkyl linker. The ratios extracted from the chromatogram of the component 2'-deoxynucleosides and cross-linked nucleosides were in good agreement with the theoretical compositions.

For the mono-adduct lesions, O^4-butyl-4-ol dT and O^4-heptyl-7-ol dT introduced in sequences **T4** and **T7**, similar solid-phase synthesis conditions as described for the ICL substrates were employed. For these oligonucleotides, cleavage and deprotection

conditions involved the incubation of the CPG with a solution composed of 10% DBU in 1,4-butanediol (for **T4**) or 1,7-heptanediol (for **T7**) for 5 days at room temperature to ensure that the adducts were not compromised.(*221*) The DBU was neutralized with acetic acid the CPG rinsed with ACN (50 % in water) to solubilize the fully deprotected product as a result of the poor solubility of DNA in the diols. These oligonucleotides were also characterized by ESI mass spectrometry and nuclease digestion to ensure their composition.

3.3.3 UV thermal denaturation studies

The effect of the directly opposed O^4dT-alkylene-O^4dT ICL on duplex stability was assessed through UV thermal denaturation experiments. The melting profiles of **XLTT4** and **XLTT7**, shown in **Figure 3.2**, exhibit a sigmoidal denaturation and hyperchromicities comparable to the control duplex containing a T-A base pair instead of the cross-link. The melting temperatures observed for the cross-linked oligonucleotides **XLTT4** and **XLTT7** were 66 and 50 °C, respectively, higher than the control duplex (with a T_m of 44 °C).

This melting temperature trend has been observed earlier by our group for duplexes containing a directly opposed O^6dG-alkylene-O^6dG mismatch present in an 11-bp duplex with increases of 23 and 33 °C in T_m for the heptylene and butylene linkages, respectively, over the control duplex.(*145*) This same tendency has also been reported when ICL of various lengths are introduced between the N3 atoms of dT where an increase in T_m of 16 and 31 °C were observed for heptylene and butylene linkages over the non-alkyated control.(*225*)

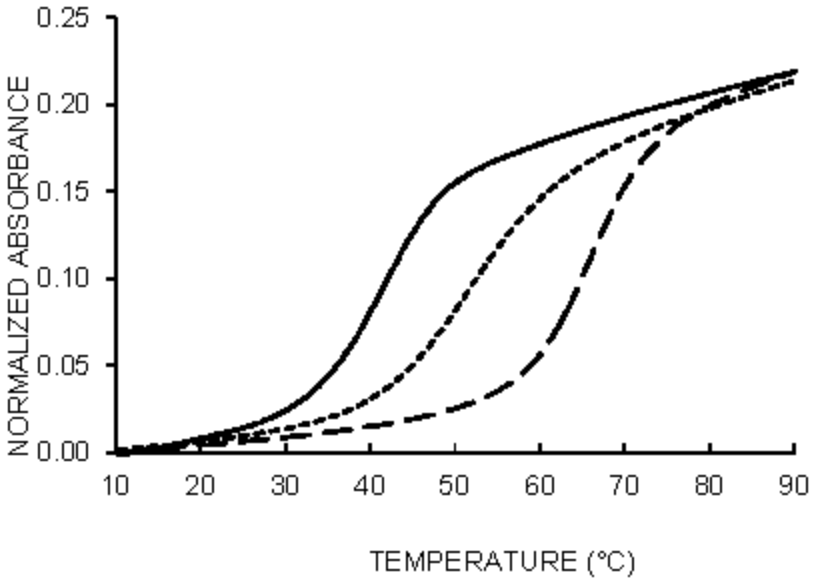


Figure 3.2: Absorbance (A_{260}) versus temperature profiles of cross-linked duplexes **XLTT4** (— —), **XLTT7** (- - -) and non-cross-linked DNA (—)

Thermodynamic parameters of the ICL and control duplexes were studied to investigate the origin of the increased thermal stability observed for the ICL DNA (**Table 3.1**). Varying duplex concentration did not affect the T_m observed for the ICL duplexes but did influence that of the control, as expected. The thermodynamic parameters indicate a reduction in the enthalpy for the ICL containing duplex, which was more pronounced for the heptyl versus butyl linker. Possible reasons include disruption of the base pairs and base stacking around the cross-linked site as a result of steric effects of the larger linker.

Table 3.1: Thermodynamic parameters of DNA Duplexes

DNA	ΔH° (KJ/mol)	ΔS° (J/(mol K)
Control (T-A)	-298 ± 23	-823 ± 71
XLTT4	-249 ± 7	-735 ± 20
XLTT7	-179 ± 4	-551 ± 12

The presence of the ICL caused an increase in the entropy of the duplex due to the preorganized nature of the unimolecular system thus favouring thermal stability of the

DNA. The contribution of the increased entropy outweighed the negative enthalpic effect in the ICL system, consistent with what has been previously observed with ICL formed by trans-diamminedichloroplatinum(II).(*226*)

The influence of alkyl chain length of the mono-adducts at the O^4 atom of dT on stability was also assessed. There was a 10 °C reduction in duplex stability for the duplexes containing O^4-MedT, O^4-butyl-4-ol dT (**T4**) or O^4-heptyl-7-ol dT (**T7**) (48 °C) compared to the non-alkylated control (58 °C) demonstrating that for these adducts the length of the alkyl chain attached at the O^4 atom of dT had no substantial effect on duplex stability (see Appendix III). These results are supported by previous studies demonstrating that O^4-ethyl dT/dA base pair within a duplex decreased the T_m by 14°C with respect to the native control.(*227*)

3.3.4 CD spectroscopy of DNA duplexes

CD spectra of the cross-linked duplexes **XLTT4** and **XLTT7** exhibited signatures characteristic of B-form DNA with a positive signal centered around 275 nm, a crossover around 260 nm and a negative signal around 250 nm as shown in **Figure 3.3**.(*200*,*201*) These results indicate that the O^4dT-alkylene-O^4dT ICL caused minimal distortion to the global B-form structure.

Similarly, duplexes containing O^4-MedT, O^4-butyl-4-ol dT (**T4**) or O^4-heptyl-7-ol dT (**T7**), showed little deviation in their CD signature relative to the unmodified control duplex demonstrating minimal distortion from the global B-form structure. These findings are in agreement with NMR reports on O^4-MedT containing duplexes that suggest such lesions do not cause major structural deformation of the duplex.(*26*)

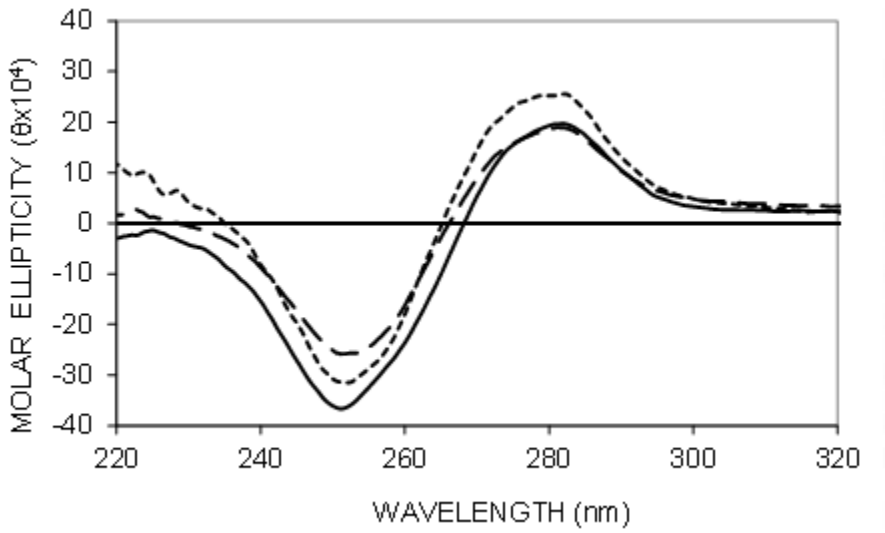


Figure 3.3: Far-UV CD spectra of cross-linked duplexes **XLTT4** (— —), **XLTT7** (- - -) and non-cross-linked DNA (—).

3.3.5 Repair and binding of AGTs with cross-linked and mono-alkylated substrates

Repair studies conducted on the **XLTT4** and **XLTT7** duplexes revealed that none of the AGT proteins from human or *E. coli* were able to completely eliminate the cross-links, as observed in **Figure 3.4**. Moreover none of the proteins were able to create the intermediate product, which consists of one repair step leading to an AGT-ssDNA species.(*146*,*214*)

EMSA were conducted with **XLTT4**, **XLTT7** and a non cross-linked control duplex, which has a T-A match where the T-T cross-link is located, to determine if the lack of activity of hAGT towards these lesions was inherent to the binding and detection of the lesion or if the lack of repair was a function of catalysis.

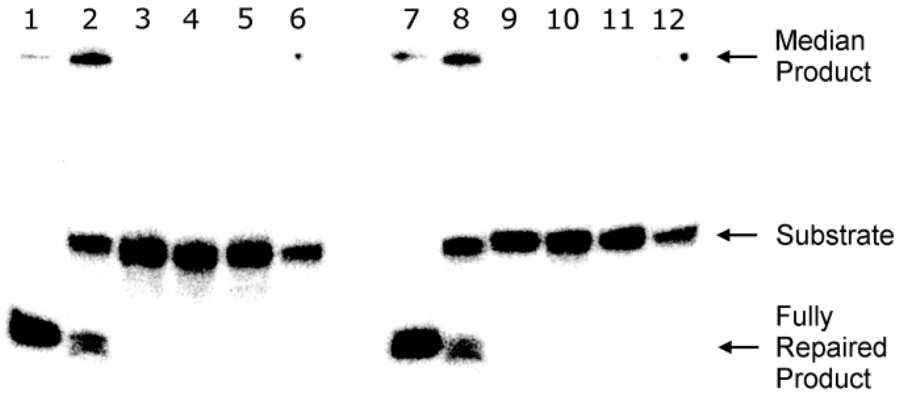


Figure 3.4: Repair gel of **XLTT4** and **XLTT7** by hAGT, Ada-C and OGT.

Denaturing gel of the repair of 2 pmol of DNA by 60 pmol of AGT. Lane 1, Control DNA; lane 2, **XLGG7** + hAGT; lane 3, **XLTT4**; lane 4, **XLTT4** + hAGT; lane 5, **XLTT4** + Ada-C; lane 6, **XLTT4** + OGT; lane 7, Control DNA; lane 8, **XLGG7** + hAGT; lane 9, **XLTT7**; lane 10, **XLTT7** + hAGT; lane 11, **XLTT7** + Ada-C; lane 12, **XLTT7** + OGT.

Results of the EMSA indicate that the hAGT stoichiometry for the various DNA was the same with two proteins for every 11-bp duplex (see Appendix III). These results are consistent with findings of others, which stipulates that hAGT binds every 4 nucleotides.(*167*,*168*) This stoichiometry was the same for the cross-linked DNA and for the control demonstrating that the presence of the cross-link did not affect how hAGT interact together when bound to the DNA. The dissociation constant of the control duplex was much higher than the cross-links, which was unanticipated.

To address the low binding affinity of hAGT for the control duplex a set a oligos were designed where the central dT in the control sequence (5' CGA AAT TTT CG) was replaced by the other 3 natural 2'-deoxynucleotides and assessed for binding by hAGT. Binding reactions were conducted over a large range of protein to obtain a rough estimate of the K_d for the hAGT-DNA complexes. Results indicate that hAGT preferentially binds to DNA with a dG nucleotide that is further than 2 nucleotides from the ends (see

Appendix III). This same hAGT binding trend has been observed for long A-T tracts.(*164*) The lack of a central dG explains the low binding affinity observed for the control sequence. The increased binding affinity observed for **XLTT4** and **XLTT7** over the control may be attributed to a hydrophobic interaction between alkyl adduct and hAGT, which has been proposed previously as the basis of substrate discrimination by the protein for O^6-MedG and O^6-BndG over dG.(*148*)

Experiments probing the repair of larger alkyl adducts at the O^4 atom of dT show a different trend than the ones observed for methylation (see Appendix III). For the larger lesions, O^4-butyl-4-ol dT and O^4-heptyl-7-ol dT, only OGT showed any ability to remove these adducts to a reasonable level, indicating only OGT out of all 3 proteins is capable of repairing large lesions at the O^4 atom of dT.

To assess the lack of repair of the adducts at the O^4 atom of dT by hAGT, binding assays were conducted. A general trend from the EMSA could be observed for the binding affinity of hAGT to the various mono-adduct DNA (see Appendix III). First, hAGT had very similar binding affinities to all mono-adduct DNA, which was roughly 1.8 times better than the non-alkylated control. Second, hAGT bound O^4-alkyl dT lesions with very similar affinities as it did O^6-MedG and the size of the lesion at the O^4 atom of dT had no substantial effect on binding for the various alkyl chains.

The observed increase in binding affinity for the adduct or ICL containing DNA and hAGT over their unalkylated controls suggests hAGT is capable of flipping out the alkylated dT and placing the lesion into its active site as required for repair, as demonstrated previously through the repair of O^6dG-alkylene-O^6dG ICL.(*146*,*214*)

3.3.6 Modelling of O^4-MedT and O^4-butyl-4-ol dT in hAGT active site

To understand the underlying substrate discrimination observed for AGT with respect to the mono-adducts, molecular modelling was performed. hAGT is the only AGT that has been co-crystallized with a DNA substrate providing valuable insight into its active conformation. The substrate differences observed between various AGT homologues suggest substantial variations in the active sites amongst proteins. Attempts to model methyl and butyl-4-ol adducts with a homology model of OGT based on hAGT (generated from PDB 1T38) were performed. However, the minimized structures did not correlate with the experimental evidence. Our modelling results suggests the methyl adduct fits in the active site of hAGT without steric clashes. The model of the butyl-4-ol adduct proposes a steric clash in the active site of the protein between the γ-carbon of the lesion and the protein due to the adoption of the gauche conformation at the β-carbon of the adduct. This interaction causes Cys 145 to adopt the incorrect rotamer for the alkyl transfer reaction to take place when the butyl-4-ol adduct is present. In the hAGT repair mechanism, His 146 is responsible for proton abstraction of a nearby water molecule from Cys 145 allowing the formation of the thiolate anion. In the case where Cys 145 is rotated away from His 146 (as observed for the butyl-4-ol lesion) the thiolate anion will not be produced and the protein will be inactive.

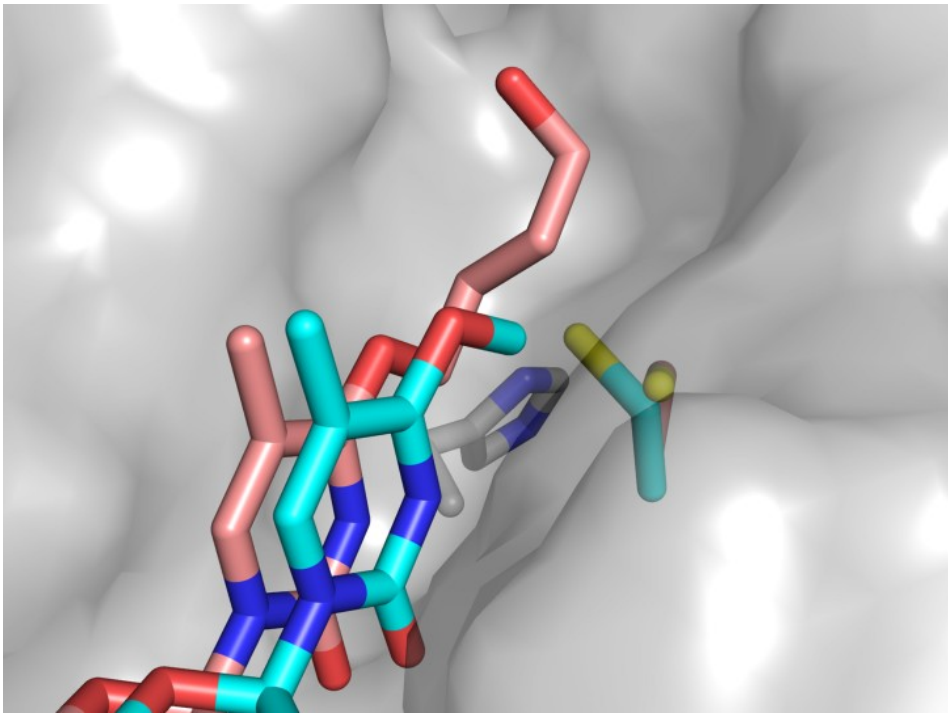


Figure 3.5: Overlay of O^4-MedT and O^4-butyl-4-ol dT in the active site of hAGT

The adopted rotamer of Cys 145 are shown for O^4-MedT (cyan) and O^4-butyl-4-ol dT (pink) and His 146 in grey.

3.4 Conclusions

DNA duplexes containing an opposed O^4dT-alkylene-O^4dT ICL were synthesized using a convertible nucleoside strategy to prepare a cross-linked nucleoside dimer that was converted into a bis-phosphoramidite. Solid phase synthesis was employed to incorporate the cross-link into a specific orientation and position in a DNA duplex. This method is versatile and allows for various linker lengths to be introduced for numerous studies including DNA repair. ICL duplexes containing a four and seven carbon linker evaded repair from various AGT proteins from human and *E. coli*.

Studies to investigate AGT proteins that can be engineered to repair these ICL and their repair by other systems are ongoing to enhance our understanding of how ICL

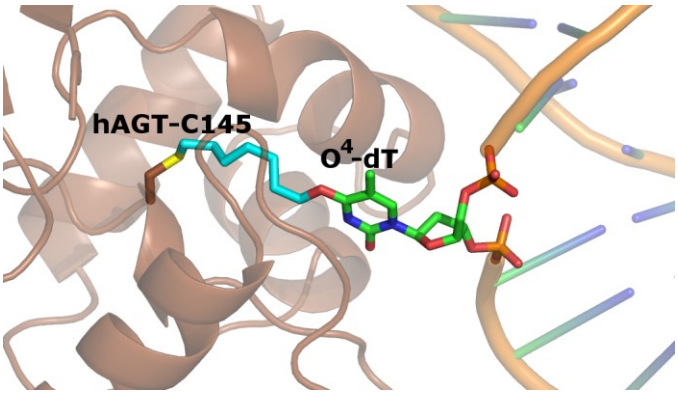
lesions may be processed. Furthermore this may result in increasing the potency of bifunctional chemotherapeutic agents which create ICL lesions and hopefully limit their mutagenicity and other side effects such as the onset of resistance to these drugs.

3.5 Acknowledgements

We are grateful to Dr. Anthony. E. Pegg (Penn. State University) for the plasmids encoding the wild-type Ada-C, OGT and hAGT genes. 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. Both FPM and DKO were recipients of postgraduate fellowships from NSERC.

CHAPTER IV

Preparation of Covalently Linked Complexes Between DNA and *O*^{6}-Alkylguanine-DNA Alkyltransferase Using Interstrand Cross-linked DNA



Graphical Abstract: Representation of hAGT-ICL covalent complex

Published as:

McManus FP, Khaira A, Noronha AM, Wilds CJ.

Bioconjugate Chem., DOI: 10.1021/bc300553u

Abstract

AGT are responsible for the removal of alkylation at both the O^6 atom of dG and O^4 atom of dT. AGT homologues show vast substrate differences with respect to the size of the adduct and which alkylated atoms they can restore. hAGT has poor capabilities for the removal of methylation at the O^4 atom of dT, which is not the case in the majority of homologues. No structural data is available to explain this poor hAGT repair. We prepared and characterized O^6dG-butylene-O^4dT (**XLGT4**) and O^6dG-heptylene-O^4dT (**XLGT7**) ICL DNA as probes for hAGT, OGT and Ada-C for the formation of DNA-AGT covalent complexes. **XLGT7** reacted only with hAGT and did so with a cross-linking efficiency of 25%, while **XLGT4** was inert to all AGT tested. The hAGT mediated repair of **XLGT7** occurred slowly, on the order of hours as opposed to the repair of O^6-MedG which requires seconds. Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis of the repair reaction revealed the formation of a covalent complex with an observed migration in accordance with a DNA-AGT complex. The identity of this covalent complex, as determined by mass spectrometry, was composed of a heptamethylene bridge between the O^4 atom of dT (in an 11-mer DNA strand) to residue Cys 145 of hAGT. This procedure can be applied to produce well defined covalent complexes between AGT with DNA.

4.1 Introduction

Modified nucleic acids with novel properties have had a long standing interest and application in the field of bioorganic chemistry.(*228*) One example of how these probes can be employed involves the formation of covalent complexes with proteins where weak binding was previously observed leading to a wide range of applications.(*229*,*230*,*231*)

Formation of covalent complexes for crystallographic purposes can be employed for various systems due to the control one can have on the nature of the chemical modification, which can be tailored to specific protein complexes.(*232*,233) For example, the structure of hAGT complexed with DNA has been determined with the aid of modified nucleic acids.(*151*,*155*) One of these structures, significant in enhancing our understanding of the mechanism of hAGT, contains a chemical cross-link formed between the hAGT active site and DNA containing the modified nucleoside N1,O^6-ethanoxanthosine. hAGT has also been employed in an elegant method to observe the movement of an hAGT fusion protein which reacts with O^6-benzylguanine derivatives.(*234*)

AGT proteins, which are found in all kingdoms of life, are responsible for the repair of the mutagenic O^6-MedG and O^4-MedT lesions.(*4*,*13*,*189*) The mutagenicity of these lesions rise from the wobble base pair that these modified nucleotides adopt with respect to their natural compliments. These wobble base pairs adversely affect the processivity of DNA polymerase during DNA replication. O^6-MedG adopts a non-wobble base pair with dT causing a GC to AT transition while O^4-MedT adopts a non-wobble base pair with dG causing a TA to CG transition. These two modifications

account for 8% of the total DNA damage caused by MNU.(*12*) Endogenously, O^6-MedG is formed at a rate of 10-30 events daily by S-adenosylmethionine.(*46*)

The formation of O^6-MedG is detrimental to cells where the bioaccumulation of 6 650 O^6-MedG modifications in hAGT deficient tumor cells is sufficient to cause cell lethality.(*14*) Unlike O^6-MedG, O^4-MedT is not affected by the levels of hAGT nor by MMR, which is why this lesion is more mutagenic in mammalian cells.(*29*,*30*) Processing of O^4-MedT in mammalian cells is believed to rely solely on the NER pathway.(*31*)

hAGT and Ada-C, the most thoroughly characterized AGT proteins, repair alkylated DNA by flipping the damaged nucleotide out of the DNA duplex and into the active site where the alkyl group is transferred from the point of lesion to the active site-Cys residue.(*173*) Once alkylated, this suicide protein is degraded by the ubiquitin pathway in mammalian cells.(*104*) There are interesting differences in the ability of AGTs to process alkylation damage. For example, hAGT is unable to process O^4-MedT efficiently, unlike OGT and Ada-C.(*117*,*131*)

Mimics of the ICL generated by hepsulfam have been studied by our group.(*146*,*214*,*235*) Results from our previous work indicate hAGT, but not the *E. coli* homologues, can repair ICL that involve two O^6 atoms of dG that are either directly opposed, in a 1,2 stagger or in the more clinically relevant 1,3 staggered orientation. On the other hand, ICL DNA containing a heptamethylene linkage bridging the O^4 atoms of dT are inert to hAGT and the *E. coli* homologues.

In the current study the preparation of ICL DNA containing an O^6dG-alkylene-O^4dT cross-links are described. The approach to prepare these ICL involves the method described by Swann's group to prepare a "convertible" dT to introduce an alkyl chain

with a terminal alcohol at the O^4 atom of dT followed by a Mitsunobu reaction at the O^6 atom of protected dG to form the alkylene linkage (**Figure 4.1A**).(*187*,*221*) hAGT, OGT and Ada-C were expressed and purified to homogeneity to investigate their role in the repair of O^6dG-alkylene-O^4dT ICL **XLGT4** and **XLGT7** (**Figure 4.1B**).

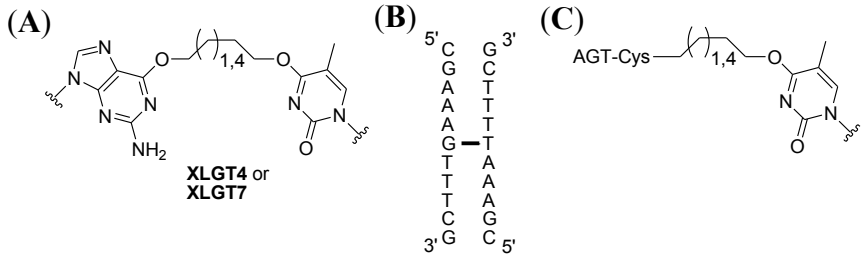


Figure 4.1: Structures of (**A**) the O^6dG-alkylene-O^4dT cross-links, (**B**) Oligonucleotides **XLGT4** and **XLGT7** and (**C**) AGT-DNA complex formed if **XLGT4** and **XLGT7** are repaired by AGT.

Aside from providing further insight into the substrate discrimination patterns of the various AGT, these ICL DNA have a potential use for the preparation of site specific covalently liked AGT-DNA adducts. The formation of AGT-oligonucleotide complexes between the various nucleobases have been shown to occur with 1,2-dibromoethane, attributed to the formation of an S-(2-bromoethyl) intermediate.(*236*) In our approach, repair by AGT at the O^6 atom of the modified dG by the active site Cys of AGT covalently links AGT to the O^4 atom of dT by an alkylene linker. The added facet of this approach is the ease with which the ICL linker can be varied allowing control over the length and chemical nature of the tether between the O^4 atom of dT and the active site of AGT (**Figure 4.1C**).

4.2 Experimental

4.2.1 Materials

5'-*O*-Dimethoxytrityl-N^2-phenoxyacetyl-2'-deoxyguanosine, 5'-*O*-dimethoxytrityl -thymidine and *N*,*N*-diisopropylamino cyanoethyl phosphonamidic chloride were purchased from ChemGenes Inc. (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 Inc. (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 performed using precoated TLC plates (Merck, Kieselgel 60 F_{254}, 0.25 mm) purchased from EMD Chemicals Inc. (Gibbstown, NJ). NMR spectra were recorded either on a Varian INOVA 300 MHz NMR spectrometer or on a Varian 500 MHz NMR spectrometer at room temperature. ^{1}H NMR spectra were recorded at a frequency of 300.0 MHz and chemical shifts were reported in parts per million downfield from tetramethylsilane. ^{31}P NMR spectra (^{1}H decoupled) were recorded at a frequency of 202.3 MHz with H_3PO_4 used as an external standard. Ampicillin, IPTG, DTT, EASY-BLUE, Protein Stain and most other biochemical reagents including 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) XL-10 Gold *E. coli* cells were obtained

from Stratagene (Cedar Creek, TX). T4 PNK and Unstained Protein Molecular Weight Marker were obtained from Fermentas (Burlington, ON). [γ-^{32}P]ATP was purchased from Amersham Canada Ltd. (Oakville, ON).

4.2.2 Solution synthesis of small molecules

1-{*O*^{4}-[3'-*O*-(*t*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-thymidinyl]}-4-{*O*^{6}-[3'-*O*-(*t*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-*N*^{2}-phenoxyacetyl-2'-deoxyguanosinyl]}-butane (1a)

To 5′-*O*-dimethoxytrityl-3′-*O*-*t*-butyldimethylsilyl-*N*^{2}-(phenoxyacetyl)-2'-deoxyguanosine (0.150 g, 0.183mmol) was added 5′-*O*-dimethoxytrityl-3′-*O*-*t*-butyldimethylsilyl-*O*^{4}-(hydroxybutyl)-thymidine (0.135 g, 0.185mmol) and triphenylphosphine (0.208 g, 7.70 mmol) in dioxane (2 mL) followed by drop wise addition of DIAD (1.521 g, 7.52 mmol) at room temperature. After 30 min the solvent was evaporated in vacuo, the crude product taken up in ethyl acetate (100 mL) and the solution washed with two portions of 3% sodium bicarbonate (2 x 100 mL). The organic layer was dried over sodium sulphate (4g) and concentrated to produce a yellow gum. The crude product was purified by flash column chromatography using a hexane / ethyl acetate (3:7) solvent system to afford 0.166 g (59 %) of product as a colorless foam.

R_f (SiO_2 TLC): 0.17 hexane / ethyl acetate (3:7). ^{1}H NMR (300 MHz, DMSO-d_6, ppm): 10.63 (s, 1H, NH), 8.45 (s, 1H, H8), 7.93 (s, 1H, H6), 7.46-6.79 (m, 31H, Ar), 6.45-6.40 (dd, 1H, H1'a, *J*=6.3Hz), 6.23-6.17 (dd, 1H, H1'b, *J*=6.6), 5.06 (s, 2H, $PhOCH_2CO$), 4.74-4.70 (m, 1H, H3'a), 4.68-4.63 (t, 2H, CH_2-O^6), 4.50-4.46 (m, 1H, H3'b), 4.43-4.39 (m, 2H, CH_2-O^4), 3.95-3.90 (m, 2H, H4'a & H4'b), 3.78 (s, 6H, O-CH_3), 3.76 (s, 6H, O-CH_3), 3.42-3.22 (m, 4H, H5'a, H5"a, H5'b & H5"b), 3.04-2.97 (m,

1H, H2'a), 2.43-2.28 (m, 3H, H2"a, H2'b & H2"b), 2.03-1.90 (m, 4H, O^4-CH_2-$C\underline{H_2}$-$C\underline{H_2}$-CH_2-O^6), 1.64 (s, 3H, C5-CH_3), 0.83 (s, 18H, Si-$C(CH_3)_3$), 0.06 (s, 6H, Si-CH_3), 0.01 (s, 3H, Si-CH_3), 0.00 (s, 3H, Si-CH_3). HRMS (ESI-MS) *m/z* calculated for $C_{86}H_{104}N_7O_{15}Si_2^+$ 1530.7129: found 1530.7087 $[M+H]^+$

1- {*O*^{4}-[3'-*O*-(*t*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-thymidinyl]}-7-{*O*^{6}-[3'-*O*-(*t*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-*N*^{2}-phenoxyacetyl-2'-deoxyguanosinyl]}-heptane (1b)

To 5′-*O*-dimethoxytrityl-3′-*O*-*t*-butyldimethylsilyl-*N*^{2}-(phenoxyacetyl)-2'-deoxyguanosine (0.700 g, 0.851 mmol) was added 5′-*O*-dimethoxytrityl-3′-*O*-*t*-butyldimethylsilyl-*O*^{4}-(hydroxyheptyl) thymidine (0.600 g, 0.774 mmol) and triphenylphosphine (0.850 g, 3.25 mmol) in dioxane (8.46 mL) followed by drop wise addition of DIAD (0.720 g, 3.56 mmol) at room temperature. After 22 hours the solvent was evaporated in vacuo, the crude product taken up in ethyl acetate (100 mL) and the solution washed with two portions of 3% sodium bicarbonate (2 x 100 mL). The organic layer was dried over sodium sulphate (4g) and concentrated to produce a yellow gum. The crude product was purified by flash column chromatography using a hexane / ethyl acetate (1:1 → 4:6) solvent system to afford 0.77 g (70 %) of product as a colorless foam.

R_f(SiO_2 TLC): 0.33 hexane / ethyl acetate (4:6). ^{1}H NMR (300 MHz, $CDCl_3$, ppm): 8.72 (s,1H, NH), 8.10 (s, 1H, H8), 8.00 (s, 1H, H6), 7.50-6.84 (m, 31H, Ar), 6.52-6.48 (dd, 1H, H1'a, *J*=6.6 Hz), 6.43-6.39 (dd, 1H, H1'b, *J*=6.0 Hz), 4.86 (s, 2H, $PhOCH_2CO$), 4.68-4.64(m, 2H, CH_2-O^6), 4.58-4.52 (m, 1H, H3'a), 4.46-4.41 (t, 2H, CH_2-O^4), 4.19-4.16 (m, 1H, H3'b), 4.05-4.01 (m, 2H, H4'a & H4'b), 3.86 (s, 6H, O-CH_3), 3.84 (s, 6H, O-CH_3), 3.60-3.31 (m, 4H, H5'a, H5''a, H5'b & H5''b), 2.82-2.73 (m, 1H, H2'a), 2.61-

2.48 (m, 2H, H2''a& H2'b), 2.32-2.24 (m, 1H, H2''b), 2.00-1.95 (m, 2H, R-$\underline{CH_2}$-CH_2-O^6), 1.85-1.80 (m, 2H, R-$\underline{CH_2}$-CH_2-O^4), 1.66 (s, 3H, C5-CH_3), 1.58 (m, 2H, R-$\underline{CH_2}$-CH_2-CH_2-O^6), 1.50 (m, 4H, R-$\underline{CH_2}$-$\underline{CH_2}$-CH_2-CH_2-O^4), 0.93 (s, 9H, Si-$C(CH_3)_3$), 0.88 (s, 9H, Si-$C(CH_3)_3$), 0.12 (s, 3H, Si-CH_3), 0.07 (s, 3H, Si-CH_3), 0.00 (s, 6H, Si-CH_3). HRMS (ESI-MS) *m/z* calculated for $C_{89}H_{110}N_7O_{15}Si_2^+$ 1572.7598: found 1572.7592 $[M+H]^+$

1- {O^4-[5'-*O*-(4,4'-dimethoxytrityl)-thymidinyl]}-4-{O^6-[5'-*O*-(4,4'-dimethoxytrityl)-N^2-phenoxyacetyl-2'-deoxyguanosinyl]}-butane (2a)

To compound **1a** (0.150 g, 0.100 mmol) was added TBAF (1M in THF) (0.059 g, 0.21 mmol) in THF (1 mL). After 30 min, the solvent was evaporated in vacuo and the crude product taken up in ethyl acetate (50 mL). The solution was washed with three portions of sodium bicarbonate (3 x 50 mL) followed by a two back extractions of the sodium bicarbonate washes with ethyl acetate (2 x 25 mL). The organic layer was dried over sodium sulphate, evaporated in vacuo and purified by flash column chromatography using an ethyl acetate / methanol (95:5) solvent system to afford 0.083 g (64 %) of product as a colorless foam.

R_f (SiO_2 TLC): 0.14 ethyl acetate / methanol (95:5). ^{1}H NMR (300 MHz, DMSO-d_6, ppm): 10.60 (s, 1H, NH), 8.36 (s, 1H, H8), 7.81 (s, 1H, H6), 7.42-6.71 (m, 31H, Ar), 6.43-6.38 (dd,1H,H1'a, *J*=6.3Hz), 6.22-6.17 (dd, 1H, H1'b *J*=6.3), 5.36-5.33 (m, 2H, OHa & OHb), 5.04 (s, 2H, $PhOCH_2CO$), 4.63-4.58 (t, 2H, CH_2-O^6), 4.53-4.49 (m, 1H, H3'a), 4.39-4.34(m, 3H, $C\underline{H_2}$-O^4& H3'b), 4.02-3.95 (m, 2H, H4'a & H4'b), 3.74 (s, 6H, O-CH_3), 3.72 (s, 3H, O-CH_3), 3.71 (s, 3H, O-CH_3), 3.32-3.08 (m, 4H, H5'a, H5'b,H5''a & H5''b), 2.94-2.84 (m, 1H, H2'a), 2.42-2.12 (m, 3H, H2''a, H2'b & H2''b), 1.96-1.89

(m, 4H, O^4-CH_2-C$\underline{H_2}$-$\underline{CH_2}$-CH_2-O^6). HRMS (ESI-MS) *m/z* calculated for $C_{74}H_{76}N_7O_{15}^+$ 1302.5399: found 1302.5392 $[M+H]^+$

1- {*O*^{4}-[5'-*O*-(4,4'-dimethoxytrityl)- thymidinyl]}-7-{*O*^{6}-[5'-*O*-(4,4'-dimethoxytrityl)-*N*^{2}-phenoxyacetyl-2'-deoxyguanosinyl]}-heptane (2b)

To compound **1b** (0.760g, 0.480mmol) was added TBAF (1M in THF) (0.282 g, 1.01 mmol) in THF (3 mL). After 30 min, the solvent was evaporated in vacuo and the crude product taken up in ethyl acetate (50 mL). The solution was washed with three portions of sodium bicarbonate (3 x 75 mL) followed by two back extractions of the sodium bicarbonate washes with ethyl acetate (2 x 25 mL). The organic layer was dried over sodium sulphate, evaporated in vacuo and purified by flash column chromatography using an ethyl acetate / methanol (199:1 → 19:1) solvent system to afford 0.620 g (95 %) of product as a colorless foam.

R_f (SiO_2 TLC): 0.20 ethyl acetate / methanol (49:1). ^{1}H NMR (300 MHz, $CDCl_3$, ppm): 8.79 (s,1H, NH), 8.02 (s, 1H, H8), 7.85 (s, 1H, H6), 7.43-6.78 (m, 31H, Ar), 6.59-6.55 (dd, 1H, H1'a, *J*=6.6 Hz), 6.44-6.37 (dd, 1H, H1'b, *J*=6.3 Hz), 4.83 (m, 1H, OHa), 4.72 (s, 2H, $PhOCH_2CO$), 4.61-4.57 (m, 3H, CH_2-O^6 & H3'a), 4.39-4.35 (t, 2H, CH_2-O^4), 4.22-4.18 (m, 1H, H3'b), 4.16-4.11 (m, 2H, H4'a & H4'b), 3.80 (s, 6H, O-CH_3), 3.77 (s, 6H, O-CH_3), 3.51-3.33 (m, 4H, H5'a, H5''a, H5'b & H5''b), 2.99 (m, 1H, OHb), 2.82-2.76 (m, 1H, H2'a), 2.67-2.59 (m, 2H, H2''a & H2'b), 2.32-2.23 (m, 1H, H2''b), 1.93-1.86 (m, 2H, R-$\underline{CH_2}$-CH_2-O^6), 1.76-1.70 (m, 5H, R-$\underline{CH_2}$-CH_2-O^4 & C5-CH_3),1.58-1.44 (m, 6H, O^4-CH_2-CH_2-$\underline{CH_2}$-$\underline{CH_2}$-$\underline{CH_2}$-CH_2-CH_2-O^6). HRMS (ESI-MS) *m/z* calculated for $C_{77}H_{82}N_7O_{15}^+$ 1344.5869: found 1344.5861 $[M+H]^+$

1-{O^4-[3′-*O*-(β-cyanoethyl-*N*,*N*′-diisopropyl)-5’-*O*-(4,4’-dimethoxytrityl)-thymidinyl]}-4-{O^6-[3′-*O*-(β-cyanoethyl-*N*,*N*′-diisopropyl)-5’-*O*-(4,4’-dimethoxytrityl)-N^2-phenoxyacetyl-2’-deoxyguanosinyl]}-butane (3a)

2a (0.200 g, 0.154 mmol) and DIPEA (0.060 g, 0.461 mmol) were dissolved in THF (1.5 mL). Then *N*,*N*-diisopropylaminocyanoethylphosphonamidic chloride (0.091 g, 0.383 mmol) was added dropwise and the reaction allowed to stir at room temperature for 30 min. This was followed by two subsequent additions of DIPEA (2 x 0.012 g, 0.091 mmol) and *N*,*N*-diisopropylaminocyanoethylphosphonamidic chloride (2 x 0.019 g, 0.081 mmol) with 1 hour intervals. Upon completion the solvent was evaporated in vacuo, the crude product taken up in ethyl acetate (25 mL), and the solution washed with 3% sodium bicarbonate (25 mL) and brine (50 mL). The organic layer was dried over sodium sulphate, decanted, and evaporated. The product, a colorless powder, was precipitated from hexanes (0.221 g, 85 %).

R_f (SiO_2 TLC): 0.56, 0.61, 0.66, 0.73 ethyl acetate. ^{31}P NMR (202.3 MHz, d_6-acetone, ppm): 153.61, 153.41, 153.40, 153.23. HRMS (ESI-MS) *m/z* calculated for $C_{92}H_{110}N_{11}O_{17}P_2^+$ 1702.7556: found 1702.7557 $[M+H]^+$

1-{O^4-[3′-*O*-(β-cyanoethyl-*N*,*N*′-diisopropyl)-5’-*O*-(4,4’-dimethoxytrityl)-thymidinyl]}-7-{O^6-[3′-*O*-(β-cyanoethyl-*N*,*N*′-diisopropyl)-5’-*O*-(4,4’-dimethoxytrityl)-N^2-phenoxyacetyl-2’-deoxyguanosinyl]}-heptane (3b)

2b (0.200 g, 0.149 mmol) and DIPEA (0.058 g, 0.446 mmol) were dissolved in THF (1.5 mL). Then *N*,*N*-diisopropylaminocyanoethylphosphonamidic chloride (0.088 g, 0.372 mmol) was added and the reaction allowed to stir at room temperature for 30 min. This was followed by two subsequent additions of DIPEA (2 x 0.012 g, 0.091 mmol) and

N,*N*-diisopropylaminocyanoethylphosphonamidic chloride (2 x 0.019 g, 0.081 mmol) with 1 hour intervals. Upon completion the solvent was evaporated in vacuo, the crude product taken up in ethyl acetate (25 mL), and the solution washed with 3% sodium bicarbonate (25 mL) and brine (50 mL). The organic layer was dried over sodium sulphate, decanted, and evaporated. The product, a colorless powder, was precipitated from hexanes (0.218 g, 84 %).

R_f (SiO_2 TLC): 0.50, 0.58, 0.67, 0.75 ethyl acetate. ^{31}P NMR (202.3 MHz, d_6-acetone, ppm): 153.60, 153.40, 153.39, 153.22. HRMS (ESI-MS) *m/z* calculated for $C_{95}H_{116}N_{11}O_{17}P_2^+$ 1745.8032: found 1745.8059 $[M+H]^+$

4.2.3 Oligonucleotide synthesis and purification

DNA duplexes **XLGT4** and **XLGT7** were assembled as described in **Section 3.2.3**, with the modifications indicated below.

Commercially available fast deprotecting 3'-*O*-2'deoxynucleoside phosphoramidites were dissolved in anhydrous ACN at a concentration of 0.1 M and the modified bis-3'-*O*-2'-deoxyphosphoramidites (**3a** and **3b**) at 0.05 M. Phosphoramidite coupling times for the commercial 3'-*O*-2'-deoxyphosphoramidites were 2 min and for the modified bis-phosphoramidites **3a** and **3b** 30 min.

Due to the presence of an O^4-alkyl dT adduct deprotection of the oligonucleotide was achieved by incubating the CPG in 500 μL of 10% DBU in ethanol for 3 days in the dark at room temperature.

XLGT4 and **XLGT7** were purified as described in **Section 3.2.3**.

4.2.4 ESI-MS of small molecules and DNA

ESI-MS of **XLGT4** and **XLGT7** were conducted at the Concordia University Centre for Biological Applications of Mass Spectrometry using a Micromass Q-tof2 mass spectrometer (Waters) equipped with a nanospray ion source. The mass spectrometer was operated in full scan, negative ion detection mode. ESI mass spectra for small molecules were recorded on a LTQ Orbitrap Velos-ETD (Thermo Scientific) in positive ion mode in ACN.

4.2.5 Exonuclease digestion

Exonuclease analysis of **XLGT4** and **XLGT7** were conducted as described in **Section 2.2.4**.

The retention times of the eluted peaks were compared to the standard nucleotides which eluted at the following times: dC (4.7 min), dG (7.5 min), dT (8.1 min), dA (8.9 min), O^6dG-butylene-O^4dT (18.4 min) and O^6dG-heptylene-O^4dT (25.3 min), and the ratio of nucleosides was determined based on known extinction coefficients, aside from the cross-linked nucleosides where an extinction coefficient of 6200 M^{-1} cm^{-1} was used corresponding to the sum of the extinction coefficients of O^4-MedT and O^6-MedG.

4.2.6 UV thermal denaturation studies

The molar extinction coefficients for **XLGT4** and **XLGT7** were determined using the nearest-neighbour method (M^{-1} cm^{-1}) assuming a central G-C match. Thermal denaturation profiles were obtained and processed as described in **Section 2.2.5**.

4.2.7 CD spectroscopy

CD signatures were obtained on a Jasco J-815 spectropolarimeter as described in **Section 2.2.6**.

4.2.8 Protein expression and purification

All AGT homologues were expressed under the promoter of the pQE30 vector in XL-10 Gold *E. coli* cells as described in **Section 2.2.8**.

4.2.9 Cross-link repair assay

Repair studies of **XLGT4** and **XLGT7** were performed as described in **Section 2.2.10**.

4.2.10 Identification of XLGT7 repair product by SDS-PAGE

600 pmol of hAGT was incubated with 625 pmol of **XLGT7** in 15 µL of Activity Buffer for 5 h at 37 °C. The reaction was terminated by adding 10 µL of SDS Buffer (125 mM Tris HCl (pH 7.6), 2% SDS, 30% glycerol and 7.5% 2-mercaptoethanol), boiling the sample for 5 min and separated on a 15% SDS-PAGE. The gel was stained with EASY-BLUE, Protein Stain as per the manufacturer's protocol.

4.2.11 Identification of XLGT7 repair product by ESI-MS

4000 pmol of hAGT was incubated with 1200 pmol of **XLGT7** in 100 µL of Activity Buffer for 3 h at 37 °C. 10 µL of the sample was loaded on a Spursil® C-18 column (3 µm, 2.1 mm × 150 mm) using an Agilent 1200 Series LC system at a flow rate of 0.2 mL/min with a gradient program of: 0-5 min, hold at 5% B; 5-30 min, linear gradient from 5% B to 50% B; 30-40 min, linear gradient from 50% B to 90% B; 40-42 min, hold at 90% B; 42-50 min, linear gradient from 90% B to 5% B (buffer A, 0.1%

formic acid in water and buffer B, 0.1 % formic acid in ACN). ESI-MS was conducted on a Micromass Q-Tof Ultima API with the following conditions: source voltage 3.5 kV, mass range of 700-1999 *m*/*z* in positive ion mode. The theoretical masses of the hAGT-DNA species were calculated to be either 25303.6 Da if the repair event occurred at the modified dG (21876.2 Da (hAGT) + 96.2 Da (Linker) + 3331.2 Da (CGAAATTTTCG)) or 25328.6 Da if the repair event occurred at the modified dT (21876.2 Da (hAGT) + 96.2 Da (Linker) + 3356.2 Da (CGAAAGTTTCG)).

4.2.12 Binding studies of C145S hAGT

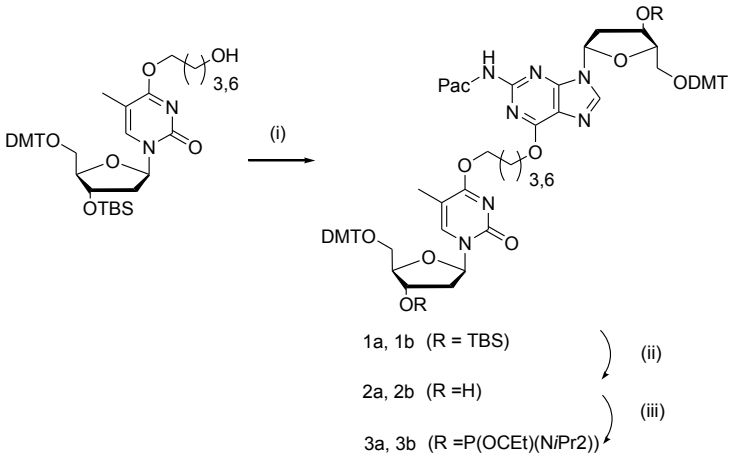
Cross-link binding assays were conducted as described in **Section 3.2.11**.

4.3 Results

4.3.1 Substrate preparation

The structures of **XLGT4** and **XLGT7** are shown in **Figure 4.1**. The O^6dG-butylene-O^4dT (**3a**) and O^6dG-heptylene-O^4dT (**3b**) cross-linked amidites were prepared as shown in **Scheme 4.1** which began with the nucleoside adduct containing a tetra- or heptamethylene linker with a terminal alcohol which was prepared by reacting the appropriate di-alcohol with 3'-*O*-(*t*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-C4-(1,2,4-triazol-1-yl)-thymidine.(*235*) The 3' and 5' bis-protected dT adducts were subsequently reacted with 5'-*O*-dimethoxytrityl-3'-*O*-*t*-butyldimethylsilyl-N^2-(phenoxyacetyl)-2'-deoxyguanosine *via* a Mitsunobu reaction to give the fully protected dimers **1a** and **1b**.(*188*) Desilylation of the dimers gave **2a** and **2b** and was followed by a phosphitylation reaction yielding the cross-linked bis-amidites **3a** and **3b**, which were employed for solid phase synthesis. ^{31}P NMR signals at 153.61, 153.41, 153.40 and

153.23 ppm for **3a** and 153.60, 153.40, 153.39 and 153.22 ppm for **3b** confirmed the presence of both phosphoramidite groups. As a result of the asymmetry for the phosphitylated cross-linked nucleosides four ^{31}P NMR signals were observed.



Scheme 4.1: Synthesis of O^6dG-butylene-O^4dT and O^6dG-heptylene-O^4dT bis-phosphoramidites

Reagents and conditions: (i) 5'-*O*-dimethoxytrityl-3'-*O*-*t*-butyldimethylsilyl-N^2-(phenoxyacetyl)-2'-deoxyguanosine, triphenylphosphine, DIAD, dioxane. (ii) TBAF (1M in THF), THF. (iii) *N*,*N*-diisopropylamino cyanoethyl phosphonamidic chloride, DIPEA, THF.

4.3.2 ESI-MS and exonuclease digestion of DNA

XLGT4 and **XLGT7** were analyzed by mass spectrometry to ensure that the base modifications were present after the deprotection and purification steps. The observed masses were in agreement with calculated values.

XLGT4 and **XLGT7** nucleoside compositions were confirmed by exonuclease digestion to their constituent nucleosides with a mixture of snake venom phosphodiesterase and calf intestinal alkaline phosphatase, then analyzed by C-18 reversed phase HPLC. The presence of modified nucleosides was established by the appearance of additional peaks, aside from the four standard nucleosides, on the HPLC chromatogram. The extracted ratios of the component 2'-deoxynucleosides were in good

agreement with the theoretical compositions except for the lower than anticipated proportion of cross-linked modifications due to their assumed extinction coefficients.

4.3.3 UV thermal denaturation studies

The stability of **XLGT4** and **XLGT7** containing duplexes were determined by UV thermal denaturing studies by monitoring A_{260} at increasing temperatures, as illustrated in **Figure 4.2**. Both modified duplexes displayed sigmoidal denaturation patterns comparable to the control duplex. The presence of the ICL stabilized the DNA duplex with the 4 methylene linker having a greater contribution than the 7 methylene linker. The melting temperatures for **XLGT4**, **XLGT7** and the control (G-C pair instead of a G-T cross-link) were 68, 56 and 46 °C, respectively.

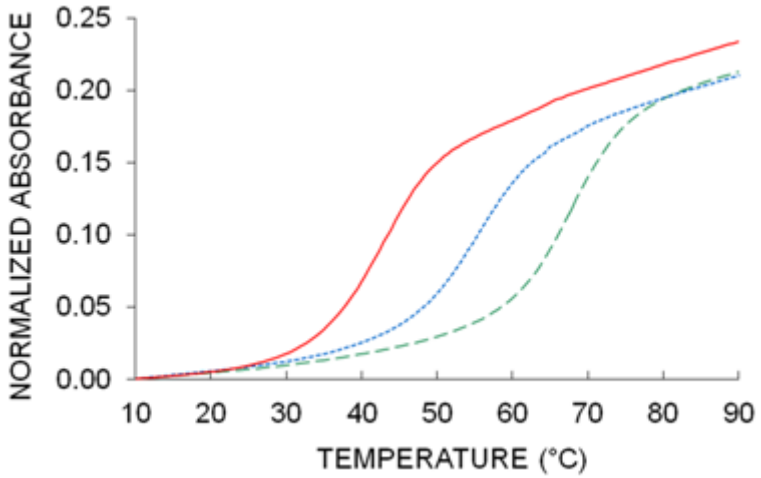


Figure 4.2: T_m curves of **XLGT4**, **XLGT7** and control DNA duplexes.

Control DNA (_____), **XLGT4** (_ _ _ _) and **XLGT7** (……..).

4.3.4 CD spectroscopy

The alkylene linkers present in **XLGT4** and **XLGT7** showed little effect on overall secondary structure as indicated by the characteristic B-form DNA signatures observed for all samples with positive signals centered around 275 nm, cross overs around 260 nm

and negative signals around 250 nm, as shown in **Figure 4.3**.(*200*,*201*) These results were similarly observed with ICL DNA having central O^4dT-alkylene-O^4dT and O^6dG-alkylene-O^6dG mismatches.(*145*,*235*)

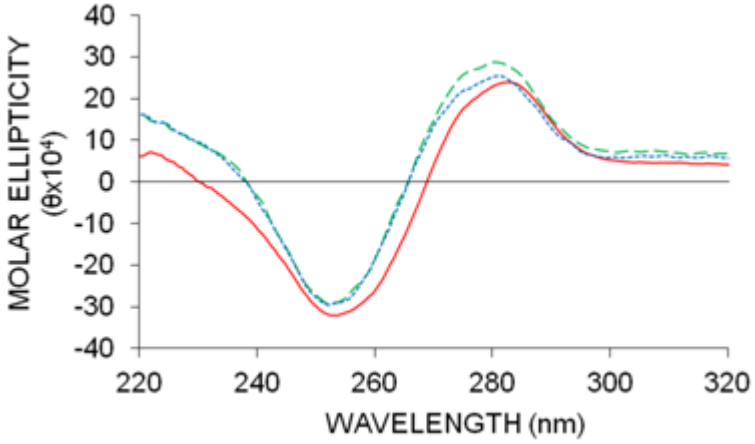


Figure 4.3: Far-UV CD spectra of **XLGT4**, **XLGT7** and control DNA duplexes. Control DNA (_____), **XLGT4** (_ _ _ _) and **XLGT7** (……..).

4.3.5 Cross-link repair

16 h repair reactions with the various AGT and **XLGT4** or **XLGT7** were conducted to detect the total amount of repair by the different AGT homologues. The positive control **XLGG7** (O^6dG-heptylene-O^6dG ICL) incubated with hAGT was used as a marker since the various species produced by this reaction are known and aid in the identification of the various products formed in our assay, including two slowly migrating bands corresponding to AGT-DNA complexes that are believed to be either fully reduced or partially reduced by 2-mercaptoethanol.

No appreciable repair was observed by any of the AGT tested with **XLGT4** (data not shown). Repair of **XLGT7**, shown in **Figure 4.4**, was achieved only by hAGT, where 25% of the starting material was converted to the AGT-DNA covalent complex. Moreover the ratio of the released DNA strand and AGT-DNA complex was 1:1

suggesting hAGT only repaired one of the alkylated atoms involved in the cross-link. This hAGT repair activity was abolished by introducing the C145S mutation (data not shown).

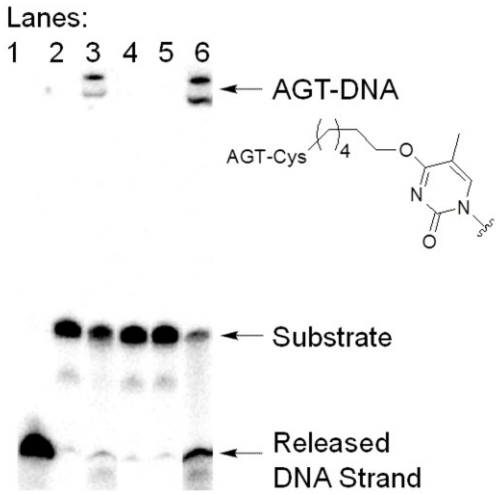


Figure 4.4: Denaturing PAGE gel illustrating the repair of **XLGT7** by hAGT, Ada-C and OGT

Repair reactions were composed of 2 pmol of DNA and 60 pmol AGT at 37 °C for 16 h. Lane 1, Control DNA + hAGT; lane 2, **XLGT7**; lane 3, **XLGT7** + hAGT; lane 4, **XLGT7** + OGT; lane 5, **XLGT7** + Ada-C; lane 6, **XLGG7** + hAGT.

The repair rate of **XLGT7** and formation of the hAGT-DNA complex was monitored over the course of three hours to obtain information on the kinetics of this reaction, as shown in **Figure 4.5**. The results indicate that the repair occurred slowly where only 13% of the AGT-DNA species is present after the three hour period. As was the case for the repair assay conducted overnight, the ratio of AGT-DNA complex to released DNA strand was 1:1.

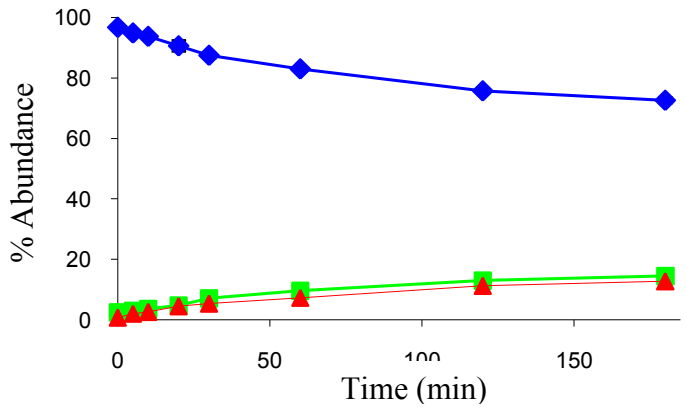


Figure 4.5: Time course repair of **XLGT7** by hAGT.

Substrate (♦), hAGT-DNA complex (▲) and released DNA strand (■).

4.3.6 Identification of XLGT7 repair product by SDS-PAGE and ESI-MS

Since the radioactivity based repair assays only show the DNA species, SDS-PAGE was conducted on the hAGT mediated repair of **XLGT7** to observe the various protein species formed. The SDS-PAGE shown in **Figure 4.6** reveals the formation of a protein complex with a greater molecular weight than unmodified hAGT. Based on the results from **Figure 4.4** and **Figure 4.6** we hypothesize that the new species corresponds to an hAGT-DNA covalent complex, which coincides with the calculated mass of (25.3 kDa).

The identity of the hAGT-DNA complex hypothesized from **Figure 4.6** was determined using ESI-MS. Since the cross-link involves two different alkylated nucleotides that could potentially be repaired by hAGT, it was necessary to identify the hAGT-DNA species in order to establish the site of repair. Repair at either of the alkylation sites will yield an hAGT-DNA species of a different mass aiding in identification.

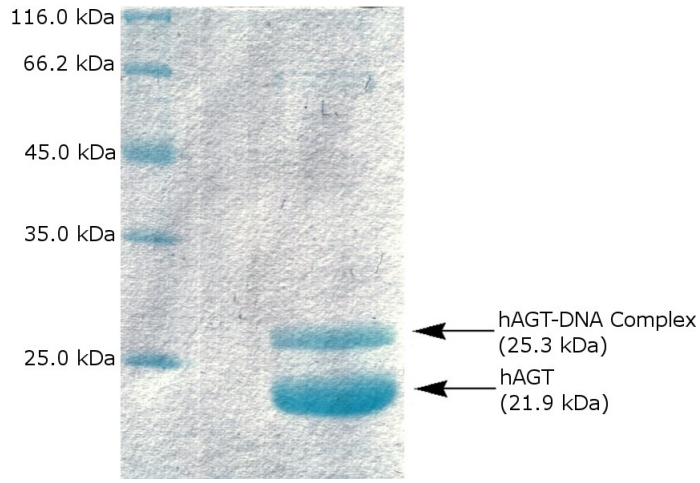


Figure 4.6: 15% SDS-PAGE of hAGT mediated repair of **XLGT7**.

Repair of 625 pmol **XLGT7** by 600 pmol hAGT for 5 h at 37 °C. Unstained Protein Molecular Weight Marker (left); **XLGT7** + hAGT reaction (right).

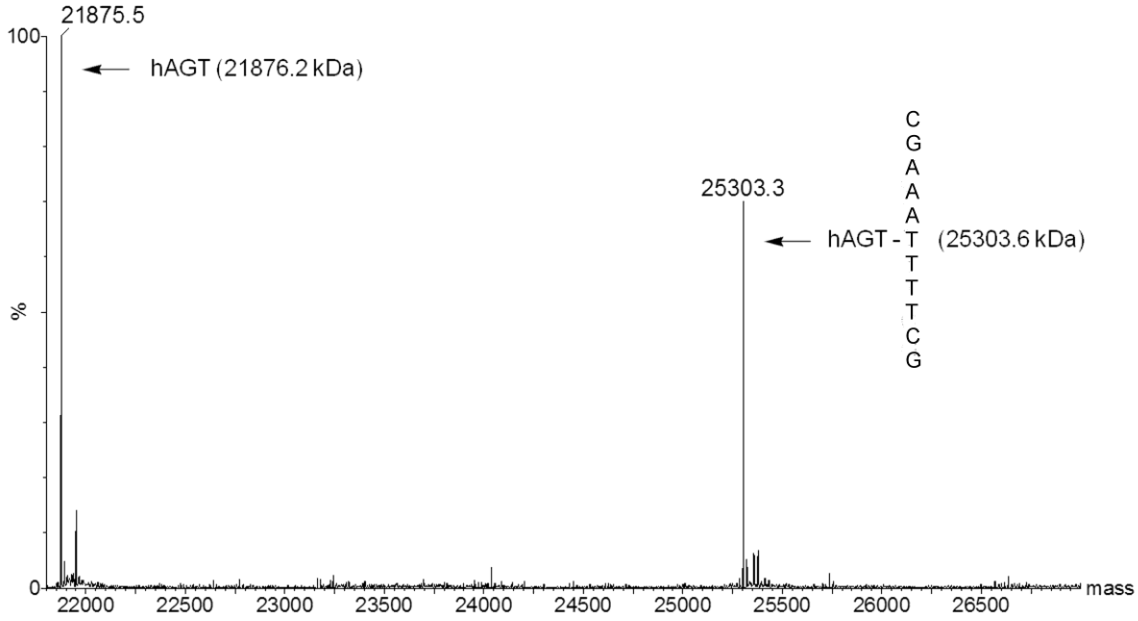


Figure 4.7: Identification of hAGT-DNA covalent complex by ESI-MS

The ESI-MS spectrum of the hAGT mediated repair of **XLGT7** is shown in **Figure 4.7**. Two major signals are observed in the deconvoluted spectrum. The first corresponds to the unalkylated form of hAGT and the second the hAGT-DNA covalent complex.

Based on the observed and calculated masses, the hAGT-DNA complex occurs between the protein and the DNA strand alkylated at the O^4 atom of dT, indicating hAGT preferentially repairs the alkylated dG (**Figure 4.8**).

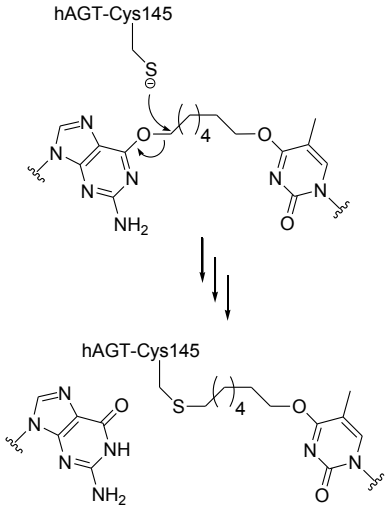


Figure 4.8: hAGT mediated repair of **XLGT7**

4.3.7 Binding studies of C145S hAGT

Binding assays were carried out to study the effect of the ICL on hAGT substrate discrimination in hopes of understanding the preferential repair observed for **XLGT7** over **XLGT4** (see Appendix IV). Results presented in **Table 4.1** indicate that hAGT binds both ICL DNA in a 2:1 (protein:DNA) stoichiometry, similar to the control DNA, suggesting the cross-links have no evident effect on the mode of hAGT binding. The K_d values extracted from the EMSA indicate that hAGT binds **XLGT7** 2-fold better than the control DNA and **XLGT4**, which are both bound in the low μM range.

Table 4.1: K_d and stoichiometry of binding by hAGT to **XLGT4**, **XLGT7** and control duplex.

DNA Duplex	K_d (μM)	Stoichiometry
Control (G-C)	3.55 ± 0.14	2.03 ± 0.08
XLGT4	3.64 ± 0.20	1.99 ± 0.17
XLGT7	2.05 ± 0.05	1.97 ± 0.02

4.4 Discussion

Directly opposed O^6dG-alkylene-O^4dT ICL were prepared using a combination of solution synthesis to create the cross-linked phosphoramidite and solid phase synthesis to incorporate the cross-link into an oligonucleotide. The identities of the ICL DNA **XLGT4** and **XLGT7** were confirmed by ESI-MS and exonuclease digestion. The nucleoside ratios extracted from the digestion were in accordance with the theoretical composition of the oligonucleotide except for a lower than anticipated ratio of the cross-linked nucleosides. These deviations in ratios are attributed to the unknown extinction coefficient for these novel nucleosides where the sum of the extinctions of O^4-MedT and O^6-MedG was used to get an estimate of the relative ratio of these modifications. The complimenting nature of the ESI-MS validates the presence of only one ICL modification per DNA molecule.

The presence of the cross-links increased the thermal stability of the DNA duplex, where the presence of a four methylene bridge had a greater stabilizing effect than the seven as observed by our UV thermal denaturation studies. The smaller increase in thermal stability brought upon by the larger ICL is believed to be due to a greater reduction in the enthalpy for the heptylene versus butylene linker containing DNA, which may be attributed to the disruption of the base pairs around the cross-linked site as a result of steric effects of the larger linker.(*235*) A comparable trend has been observed previously by our group for an O^4dT-alkylene-O^4dT mismatch, where an increase in T_m of 9 °C and 25 °C was observed over the control duplex for the heptylene and butylene linkages, respectively.(*235*) This also holds true for an O^6dG-alkylene-O^6dG mismatch in

a similar oligonucleotide sequence with 12 °C and 24 °C increases in T_m for the heptylene and butylene linkages over our control DNA, respectively.(*145*) Indeed, the presence of an ICL, irrespective of the length, causes a reduction in the change of entropy upon denaturation of the duplex due to the organized nature of the unimolecular system. As previously observed, the contribution of the entropy factor overrides the negative enthalpic effect on the ICL system affording a more stable species.(*226*)

The control DNA, **XLGT4** and **XLGT7** all assumed B-form structures as assessed by far-UV CD. These results are in accordance with previously studied ICL DNA having cross-links between the N3 atoms of dT and the N^4 atoms of dC.(*225*,*237*) Moreover, NMR studies on ICL DNA containing an N3T-butylene-N3T cross-link revealed no change in the overall B-form structure of the DNA but did show widening of the major groove without compromising the size of the minor groove.(*238*)

The conformation of the DNA probe might play a substantial role on hAGT repair. The repair of O^6-methyl-guanosine (O^6-methyl-riboG) in a DNA strand by hAGT is non-existent as opposed to its natural substrate, O^6-MedG which is efficiently processed by the protein.(*239*) Based on hAGT-DNA co-crystals, the lack of RNA repair by hAGT has been attributed to a potential steric clash between the 2' hydroxyl of the modified ribonucleoside and the alpha-carbon of Gly131.(*155*) This only partially addresses the discriminatory pattern. Other nucleic acid properties may be responsible for this discerning trend. Notably, differences in the ribose and 2'-deoxyribose sugar puckers, affecting the overall secondary structure of nucleic acids, could play a part in poor RNA repair by hAGT. For this reason the B-form structure of DNA may be primordial for AGT repair.

Our radioactivity based repair assay revealed no repair of **XLGT4** by any AGT tested. Lack of repair by OGT and Ada-C was anticipated based on previous work, which indicates that both proteins are inactive to ICL DNA.(*146*,*235*) The absence of any appreciable repair of **XLGT4** by hAGT was surprising considering its ability to repair O^6dG-butylene-O^6dG ICL (**XLGG4**) in a virtually identical sequence.(*146*) The difference in size between dT and dG may be responsible for the lack of **XLGT4** repair as opposed to **XLGG4**. Replacing one dG by a dT reduces the linear distance between the anomeric carbons of the modified nucleosides by roughly 1 Å.

A molecular model generated by Fang *et al.* of hAGT in complex with **XLGG7** suggests the preferential repair of longer cross-links by hAGT is brought upon by the shape of the hAGT active site in the DNA bound conformation which can be represented as a tunnel.(*146*) In effect, the larger the distance between the anomeric carbons of the modified nucleosides, the greater the ICL repair since the cross-linked nucleobases will sit at each end of the hAGT active site "tunnel". This allows the linker to lay in the "tunnel" while at the same time placing the alpha-carbon of the adduct in close proximity to the reactive Cys residue.

Local denaturation of the DNA duplex upon hAGT binding may be the cause for the extension of the cross-link, separating the tethered nucleotides.(*240*) As eluded to in earlier studies based on the thermal denaturation results, the larger cross-link causes the disruption of the base pairs around the lesion itself generating a substrate that is already denatured around the point of alkylation prior to protein binding.

Repair of **XLGT7** was achieved only by hAGT, as was anticipated based on the "tunnel" model. Increasing the length of the ICL allowed the nucleobases to span the

distance of the hAGT active site, permitting repair to occur. Repair of **XLGT7** by hAGT shown in lane 3 of **Figure 4.4** suggests the formation of a species that corresponds to a AGT-DNA complex, based on the migration pattern of the **XLGG7** control from lane 6, as well as a fully repaired DNA strand, which migrated similarly to the control DNA in lane 1. The formation of the AGT-DNA complex and the release of the repaired strand in a 1:1 ratio implies that repair by hAGT only occurred at one site per DNA molecule. Repair of **XLGT7** by hAGT required hours for completion as indicated by the time course assay shown in **Figure 4.5**. The C145S variant of hAGT was unable to repair **XLGT7** identifying Cys 145 as the reactive residue.

To overcome the limitations of the radioactivity based assay, which only monitors the DNA species, an SDS-PAGE analysis of the reaction was performed to analyze the various proteinaceous materials formed. The absence of a protein band at 43.8 kDa, which would indicate an hAGT dimer linked by the seven methylene linker, supports the finding that hAGT only reacts at one location on **XLGT7**. A band just above the 25 kDa marker correlates to a covalent species migrating with a molecular weight corresponding to an AGT-DNA complex in a 1:1 ratio (roughly 25.3 kDa), which is in agreement with our radioactivity based repair assay.

The identity of the covalent complex was determined by ESI-MS. The spectrum shown in **Figure 4.7**, displays some unreacted hAGT (21875.5 Da) and a peak with a mass of 25303.3 Da, corresponding to an hAGT-$(CH_2)_7$-CGAAA<u>T</u>TTTCG molecule. The absence of a peak at 25328.6 Da entails that hAGT reacts selectively at one location on the DNA duplex. To form this hAGT-$(CH_2)_7$-CGAAA<u>T</u>TTTCG product, where the underlined nucleotide depicts the cross-linked position, hAGT must react on the modified

CGAAAGTTTCG strand whereby releasing the repaired CGAAAGTTTCG oligonucleotide in the process. The absence of any formation of a covalent complex when reacting hAGT with the control DNA signifies the reactive moiety on the DNA as the O^6 atom of the cross-linked dG. The repair events of **XLGT7** by hAGT have been summarized in **Figure 4.8**. These results are consistent with our previous results that demonstrate that hAGT can repair O^6dG-alkylene-O^6dG cross-links but is inert towards O^4dT-alkylene-O^4dT cross-links.(*146*,*214*,*235*)

hAGTs inability to repair **XLGT4** was not attributed to poor substrate binding as established from our EMSA results, shown in **Table 4.1**. A two-fold decrease in K_d observed for **XLGT7** over **XLGT4** and the control DNA is not substantial enough to support the repair trend. Also the ICL had no consequence on the binding stoichiometry of hAGT to the DNA. This is in agreement with the minor groove binding property of hAGT and the solution structure of similar ICL DNA showing the cross-link spanning the major groove without influencing the size of the minor groove. (*155*,*238*,*241*)

4.5 Conclusions

Using our approach a well-defined hAGT-DNA covalent complex was generated covalently linking an O^4 alkylated dT *via* a seven methylene linker to hAGT. Unfortunately, the protein-DNA cross-linking reaction was limited to the human homologue where no complex was observed for Ada-C and OGT *E. coli* proteins. No complex was observed with the four methylene linker containing DNA **XLGT4**. The reason for this substrate discrimination by hAGT is not known but we hypothesize that the longer cross-link found in **XLGT7** allows for easier access to the point of lesion since

1) the two modified nucleosides can be separated further apart due to the increased length of the tether 2) the larger cross-link allows for substantial local destabilization of the DNA duplex at the point of repair granting easier access to hAGT.

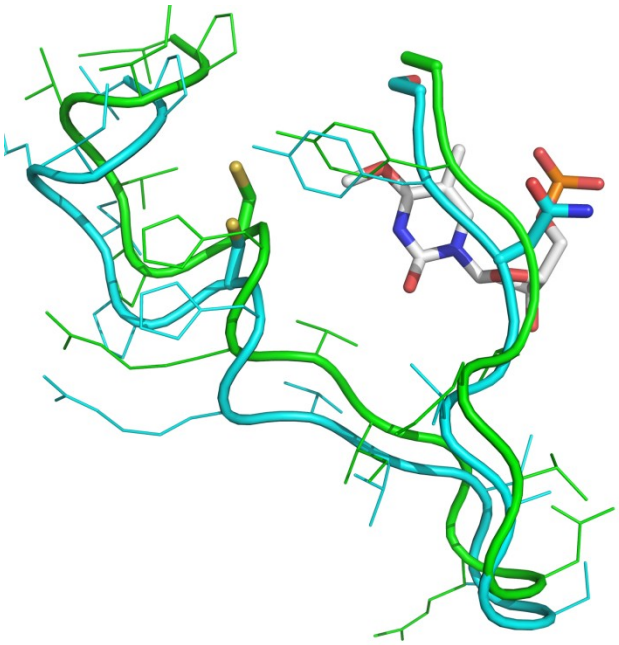
We conclude the use of ICL DNA as substrates for the formation of site specific AGT-DNA covalent complexes is promising. Synthesis of these modified oligonucleotides is straightforward and allows great control over the length of the desired tether. As demonstrated herein, hAGT specifically reacts with the O^6 alkyated dG of the cross-link allowing the formation of defined covalent complexes that may find applications in experiments involving the labeling of hAGT.

4.6 Acknowledgments

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. FPM is the recipient of postgraduate fellowships from the Groupe de Recherche Axé sur la Structure des Protéines (GRASP) and NSERC. Members of the group would like to thank Dr. Anthony. E. Pegg for the plasmids encoding the wild-type Ada-C, OGT and hAGT genes.

CHAPTER V

Engineering of a O^6-Alkylguanine-DNA alkyltransferase Chimera and Repair of O^4-Alkyl Thymidine Adducts and O^6-Alkylene-2'-Deoxyguanosine Cross-Linked DNA



Graphical Abstract: Modelled residues 139-159 of chimera (green) and hAGT (cyan) in the presence of O^4-MedT (white)

Published as:

McManus FP, Wilds CJ.

Toxicol. Res., DOI: 10.1039/c2tx20075a

Abstract

A soluble human hAGT chimera was engineered containing the active site of OGT (residues 139-159) and an additional S134P mutation. The resulting hAGT chimera not only retained hAGT's ability to repair bulky O^6dG-alkylene-O^6dG ICL DNA damage but also displayed enhanced repair of various O^4-alkyl dT adducts.

5.1 Introduction

Formation of O^4-MedT in the genome that results from exposure to tobacco derived *N*-nitrosamines or chemotherapeutic agents is highly mutagenic because of its ability to form altered hydrogen bonding patterns.(*25*,*54*,*242*,*243*) The stable O^4-MedT:dG pair resulting from a single round of replication, causes a dT to dC transition mutation, which is especially detrimental to the cell if this event occurs on proto-oncogenes or tumour suppressor genes.(*4*)

O^6-MedG, another well known mutagenic nucleobase modification, generates dG:dC to dA:dT transition.(*5*) Unlike its dT counter-part the O^6-MedG:dT intermediate (formed after a single round of replication) is a substrate for mismatch repair and direct repair in humans.(*126*,*244*,*245*) The O^4-MedT:dG mismatch persists in the cell since it is neither a good substrate for mismatch repair nor direct repair, contributing to the mutagenic potential of O^4-MedT.(*133*,*245*) O^4-MedT, which is believed to be a minor lesion, has been found at similar levels as O^6-MedG in the liver tissue of healthy volunteers, demonstrating its importance and prominence in the absence of exogenous alkylating agents.(*246*)

AGT are part of the direct repair pathway and are responsible for the removal of both O^6-alkyl dG and O^4-alkyl dT adducts.(*132*,*247*) The repair occurs in a single step process where the active site Cys thiolate anion performs a nucleophilic attack on the α-carbon of the adduct located on the exocyclic oxygen.(*155*) In the process the native DNA is restored and the AGT protein is irreversibly alkylated where it eventually undergoes degradation by the ubiquitin mediated pathway.(*104*,*248*) AGT homologues

show vast substrate differences. hAGT has been studied extensively and is believed to possess the greatest substrate range for O^6-alkyl dG adducts. Among some of the "bulkier" substrates that have been shown to undergo repair by this protein are O^6dG-alkylene-O^6dG ICL, which are DNA lesions that covalently link complementary DNA strands inhibiting DNA unwinding and segregations and in the process inhibits cell proliferation due to physical obstruction.(*146*,*214*,*235*,*249*)

E. Coli OGT is believed to be the most efficient AGT at eliminating O^4-alkyl dT.(*33*,*117*) This is contrary to hAGT, which is extremely poor at removing even O^4-MedT. hAGT recognizes and binds O^4-MedT but cannot remove them efficiently, physically shielding the lesion from the NER machinery and increasing the lesion's toxicity.(*133*)

The Loeb and Pegg groups have dedicated some effort to generate hAGT variants capable of increased repair of O^4-MedT. The Loeb laboratory employed a random sequence mutagenesis approach to generate multiple variants followed by a functional complementation assay to obtain a variant with 8 alterations (C150Y, S152R, A154S, V155G, N157T, V164M, E166Q, and A170T).(*134*) This variant showed a rate of repair of O^4-MedT that was roughly 11.5 times greater than that of hAGT.(*135*,*250*) A horizontal gene transfer from OGT to hAGT was invoked by the Pegg group to generate their hAGT-OGT chimera. Their most promising construct harboured 8 alterations (V149I, C150G, S151R, S152N, A154T, V155M, G156T and N157G) and was achieved by substituting the hAGT amino acids by their respective residues in OGT. Unfortunately, incorporating numerous successive mutations generated a chimeric protein that was not soluble, requiring refolding procedures to be adopted. Their chimera

displayed valuable properties when expressed in alkyltransferase and NER deficient *E. Coli* cells, where reduced levels of mutations at dT:dA sites were observed upon cell exposure to methylating, ethylating and propylating agents when compared to similar cells expressing wild-type hAGT.(*123*)

In the present work, we extended the portion of OGT residues in the chimera towards the C-terminal but kept Pro 140 from hAGT to generate a soluble construct.(*126*) The chimeras ability to repair alkylation adducts and ICL at both the O^4 atom of dT and O^6 atom of dG was investigated using the substrates shown in **Figure 5.1**, which were synthesized and characterized as previously described.(see **Sections 2.2.2**, **3.2.2** and **4.2.2**)

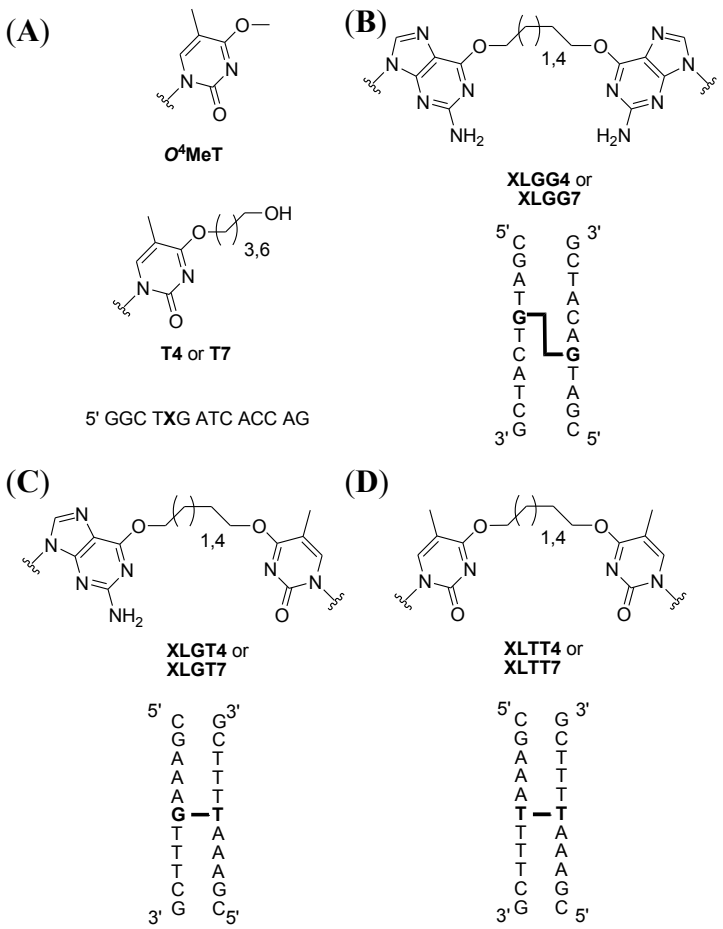


Figure 5.1: Structures of (**A**) O^4-alkyl dT monoadducts and DNA sequence where X contains the modified residue, (**B**) O^6dG-alkylene-O^6dG ICL and duplex where G-G is the ICL, (**C**) O^6dG-alkylene-O^4dT ICL and duplex where G-T is the ICL and (**D**) O^4dT-alkylene-O^4dT ICL and duplex where T-T is the ICL.

5.2 Experimental

Note: This manuscript was published as a communication and as such lacked a detailed experimental section. All novel methods are described here and previously explained methods are refered to.

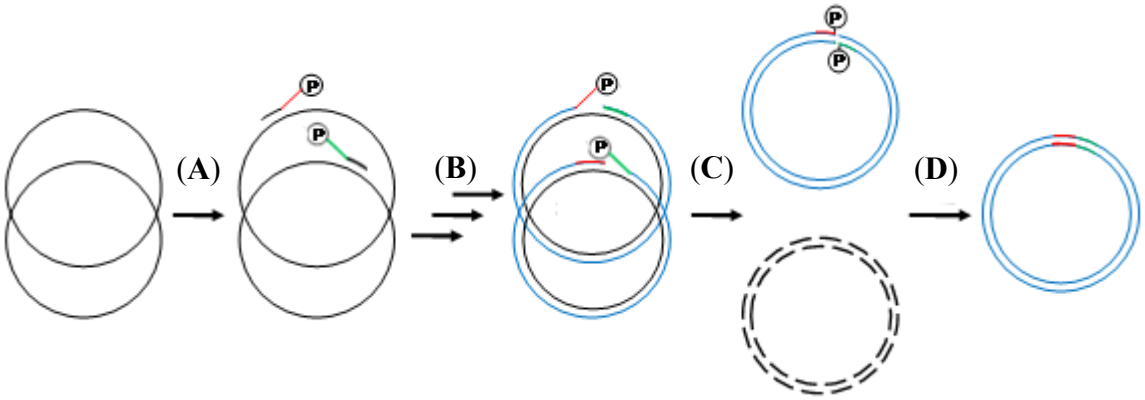
5.2.1 Materials

XL-10 Gold and BL21(DE3) *E. coli* cells were obtained from Stratagene (Cedar Creek, TX). [γ-^{32}P]ATP was purchased from Amersham Canada Ltd. (Oakville, ON).

Phusion polymerase (New England BioLabs, Ipswich, MA, USA), Pfu, T4 DNA ligase, *KpnI*, *EcoRI*, *DpnI*, Unstained Protein Molecular Weight Marker, O'GeneRuler™ 1 kb Plus DNA Ladder (Fermentas, Burlington, ON, CAN), 5' Phosphorylated Primers (Biocorp, Montreal, QC, CAN), 10 mM stock dNTPs , Agarose, Tris base, Borate, EDTA, Imidazole, Ampicillin, IPTG, and most other biochemical reagents as well as polyacrylamide gel materials (Bioshop Canada Inc, Burlington, ON, CAN), Ni-NTA Superflow Resin (Qiagen, Mississauga, ON, CAN) and Complete, Mini, EDTA-free Protease Inhibitor Cocktail Tablets (Roche, Laval, QC, CAN).

5.2.2 Creating pQE30-Chimera plasmid

An overview of the procedures is shown in **Scheme 5.1**. The primers, shown in **Table 5.1**, used in this study were designed such that the 3' portion of the primers that annealed to the template had a T_m > 50 °C, while having either a G or a C at the 3' end of the primer to ensure proper annealing. The 5' end of the reverse primer coded for the first 33 mutations while the 5' end of the forward primers coded for the remaining 30 mutations. 5' phosphate groups were incorporated into the primer *via* solid phase synthesis as these phosphates were needed to carry out the blunt-end ligation. These primers allowed for a GTC CCC ATC CTC ATC CCG TGC CAC AGA GTG GTC TGC AGC AGC GGA GCC GTG GGC AAC TAC TCC to *ATC CCC ATC GTC GTA CCT TGC CAT CGG GTT ATC* **GGC CGA AAC GGT ACC ATG ACC GGA TAT GCA** mutation. The italisized codons were coded for by the reverse primer while the bolded ones were coded for by the forward primer.



Scheme 5.1: Modified inverted site-directed mutagenesis.

(**A**) Annealing of the two 5' phosphorylated contiguous primers, (**B**) Extension by Phusion polymerase to give the linear mutated DNA, (**C**) Digestion of parental DNA by *DpnI* and (**D**) Blunt-end ligation of cleaned-up PCR product to yield the circularized mutated plasmid.

Table 5.1: Primers used to design the chimera gene from hAGT

Primer	Sequence	Additional Information
Reverse	P-GAT AAC CCG ATG GCA AGG TAC GAC GAT GGG GAT AGG ATT GCC TCT CAT GGC	Mutation of first 33 nucleotides
Forward	P-GGC CGA AAC **GGT ACC** ATG ACC GGA TAT GCA GGA GGG CTA GCC GTG AAG G	Mutation of remaining 30 nucleotides

Underlined nucleotides represent the non-annealing 5' moiety of the primers that are mutagenic
Introduced *KpnI* cut-site in forward primer shown in bold
P- Phosphate group

The PCR reaction mixtures were constituted of 10 ng of hAGT template DNA, 30 pmol of each primer, 200 µM of dNTPs, 10 µL of 5 X Phusion buffer, 1 µL of dimethyl sulfoxide and water to a total volume of 49.5 µL. Once the solution heated to its denaturation temperature (95 ^{o}C), 0.5 µL (1 unit) of Phusion polymerase was added.

The conditions used to perform the PCR reactions on the Applied Biosystems GeneAmp PCR System 9700 thermocycler were: a hot start of 1 min at 95 °C; 25 cycles of denaturation at 95 °C for 1 min followed by annealing at 60 °C for 1 min followed by

elongation at 72 °C for 12 min; a final cycle at 72 °C for 15 min was performed in order to establish proper termination of the PCR product strands for blunt end ligations.

The Phusion PCR reaction mixture was digested for 2 h at 37 °C using 10 units of *DpnI* to digest the template DNA. The resulting mixture was loaded on a 0.8 % agarose gel and electrophoresed at 80 V for 1 h. The band of the appropriate size (4 kb) was excised and the DNA extracted from the gel as specified by the Illustra GFX PCR DNA and Gel Band Purification Kit (GE Health, Baie d'Urfe, QC, CAN) using 30 μL of elution buffer type 4.

A blunt-end ligation reaction was performed for 2 h at 22 °C using 450 ng of the purified DNA (obtained above), 5 μL of 10 X T4 DNA ligase buffer, 34 μL of water and 1 μL (5 units) T4 DNA ligase. The ligase was subsequently inactivated by incubating the mixture at 70 °C for 5 min.

45 ng of DNA from the ligation mixture was transformed into 100 μL of competent XL-10 Gold cells. Briefly, the cells were thawed on ice and then transferred into a pre-chilled sterile 15 mL polypropylene tube. 5 μL of the ligation mixture was added to the tube and allowed to sit on ice for 5 min. The tube was subsequently placed in a 42 °C water bath for 40 sec, placed on ice for 2 min and the resulting mixture spread on LB + 100 μg/mL ampicillin agarose plates.

5.2.3 Creating the pET15-Chimera plasmid

Overlap extension PCR cloning was used to place the chimera gene into pET15b.(*251*) The tube for the 1st PCR step was made up of 10 ng of pQE30-Chimera template created in **Section 5.2.2**, 30 ng of the two primers, whose sequences are shown in **Table 5.2**, 200 μM of dNTPs, 10 μL of 5 X Phusion buffer and water to a total volume

of 49.5 µL. Once the solution heated to its denaturation temperature (98 ^{o}C) 0.5 µL (1 unit) of Phusion polymerase was added.

Table 5.2: Primers used to create pET15-Chimera

Primer	Sequence
pET-f	CCT CTA GAA ATA ATT TTG TTT AAC TTT AAG AAG GAG ATA TAC CAT GGA CAA GGA TTG TGA AAT GAA ACG CAC C
pET-r	CAA CTC AGC TTC CTT TCG GGC TTT GTT AGC AGC CGG ATC CTC TCA GTG ATG GTG ATG GTG GTG GGA GCC

Underlined nucleotides represent the portion of the primer which anneal to pET15b
Non-underlined nucleotides represent the nucleotides that anneal to the Chimera gene

The PCR conditions were as follows: a hot start of 2 min at 98 °C; 32 cycles of denaturation at 98 °C for 1 min followed by annealing at 60 °C for 1 min followed by elongation at 72 °C for 1.5 min; a final cycle at 72 °C for 10 min was performed. The PCR reaction was analyzed on a 1 % agarose gel and the desired band excised as indicated previously.

The tube for the 2nd PCR step was made up of 5 ng of pET15b, 160 ng of PCR product from step 1, 200 µM of dNTPs, 10 µL of 5 X Phusion buffer and water to a total volume of 49.5 µL. Once the solution heated to its denaturation temperature (98 ^{o}C) 0.5 µL (1 unit) of Phusion polymerase was added.

The PCR conditions were as follows: a hot start of 2 min at 98 °C; 20 cycles of denaturation at 98 °C for 1 min followed by annealing at 55 °C for 1.5 min followed by elongation at 72 °C for 12 min; a final cycle at 72 °C for 10 min was performed. The PCR reaction was then digested with *DpnI*, and transformed into XL-10 Gold cells and plated as indicated above.

A colony was picked from the plate and inoculated into 10 mL of LB + 100 μg/mL ampicillin overnight. The plasmids were extracted from the cells using GeneJET™ Plasmid Miniprep Kit. The purified plasmids, whose identities was confirmed by sequencing, were subsequently transformed into BL21(DE3) cells for protein expression under the T7 promoter.

5.2.4 Protein expression and purification under pQE30 vector

All proteins were purified under native conditions as described in **Section 2.2.8**.

5.2.5 Protein expression and purification under pET vector

Purification of the insoluble AGT protein was attempted using previously published protocols without success.(*123*) The only insoluble protein, the P140S chimera variant, was expressed in high copy under the pET15 system and purified under denaturing conditions. Purification of the P140S chimera variant was similar to the protocol used to purify the proteins under native condition with a few changes. The supernatant from the spin at 17000 x g after lyses of the cells using French press was discarded to remove all soluble proteins. The inclusion body pellet was taken up in resuspension buffer supplemented with 8 M urea. This was followed by a spin at 17000 x g for 45 min at room temperature. The supernatant was applied to a Ni-NTA column pre-equilibrated with equilibration buffer supplemented with 8 M urea. The column was extensively washed with equilibration buffer supplemented with 8 M urea and 10 mM imidazole. The protein was eluted with equilibration buffer supplemented with 8 M urea and 200 mM imidazole and immediately dialyzed versus refolding buffer (1 M Tris HCl (pH 8.5), 1 M NaCl, 0.3 M L-Arginine, 1 mM DTT, 10% glycerol, 0.1 mM EDTA) supplemented with

5 M urea using 8000 Da cutoff dialysis tubing for 16 hours. The protein was subsequently dialyzed against refolding buffer without urea and finally against storage buffer (50 mM Tris HCl (pH 7.6), 250 mM NaCl, 20 mM β-mercaptoethanol and 0.1 mM EDTA). This procedure yielded 30 mg of protein per L of culture, which was limited as a result of using only 3 mL of Ni-NTA resin during purification. It is noteworthy to include that on column refolding was attempted with no success as observed by the formation of a white precipitate on column when the urea concentration approached 2 M.

5.2.6 Protein characterization

CD, ESI-MS and intrinsic fluorescence analysis of the proteins were conducted as described in **Section 2.2.9**.

5.2.7 Mono-adduct repair assay

The mono-adduct repair assays were conducted as described in **Section 3.2.8**.

5.2.8 Modelling of chimera bound to O^4-alkyl dT mono-adducts

The initial model of the chimera was generated using Modeller 9.10 using PDB 1T38 as template. The modified DNA was placed in the chimera active site based on PDB 1T38 by substituting O^6-MedG to O^4-MedT, O^4-butyl-4-ol dT or O^4-heptyl-7-ol dT. The placement of the modified nucleotides were verified with PDB 1YFH.

The preliminary models underwent conjugate gradient minimization, simulated annealing, and torsion angle dynamics using Crystallography & NMR System.(*222*) CHARMM topology files were modified to use the altered nucleotides by adding restraints to keep the α-carbon of the lesion in a *syn* conformation with respect to the N3 atom of the heterocycle, as observed in PBDs 1T38, 1T39, and 1YFH.(*223*)

The models underwent 2 sets of 100 steps of gradient minimization and 200 steps of simulated annealing. The first cycle was carried out at 500 K with slow cooling at a rate of 4.5 K per cycle of dynamics. The second cycle was carried out at 350 K with slow cooling at a rate of 2.5 K per cycle of dynamics and the minimized models displayed using PyMol.

5.2.9 Cross-link repair assay

Cross-link repair assays were conducted as described in **Section 2.2.10**.

5.2.10 C145S chimera variant binding studies with ICL DNA

Binding reactions consisted of 0.5 nM dsDNA, increasing the C145S chimera variant ranging from 1 to 45.5 μM for the control (T-A) and from 1 to 6.5 μM for the ICL DNA and control (G-C) in a total solution volume of 20 μl of binding buffer (10 mM Tris HCl (pH 7.6), 100 mM NaCl, 1 mM DTT, 10 μg mL^{-1} and 2.5% glycerol). The samples were equilibrated for 1 hour at room temperature and loaded on a 15% native polyacrylamide gel (75 : 1) and run with 0.5 X TBE, which allowed the electrophoresis voltage to be increased without heating the system. Electrophoresis was carried out under liquid cooled gels at 21 °C at 250 V for 25 minutes.

The monomeric K_d and stoichiometry (n) of AGT binding to DNA were obtained from the EMSA as described in **Section 2.2.11**.

5.3 Results and discussion

5.3.1 Protein characterization

A slight variation of inverted PCR was employed to introduce residues 139-159 from OGT into the hAGT scaffold, while retaining Pro 140 (see **Table 5.3** for

sequences), which is known to confer hAGT with the ability to repair larger adducts such as ICL and O^6-benzyl guanine.(*146*,*202*,*203*,*214*,*249*) The protein was expressed and purified as previously reported for hAGT.(*214*) The protein was obtained in >90% purity as assessed by SDS-PAGE and its mass was in accordance with the calculated value, indicating no post-translational modification of the foreign protein (see Appendix V). The successive mutations introduced into the chimera did not substantially affect the secondary structure of the overall protein as assessed by far-UV CD, which revealed a similar signature to hAGT. The alterations destabilized the protein with respect to hAGT reducing the T_m by 7.5 °C. Characterization of the chimera by intrinsic fluorescence showed a blue shift in tryptophan emission indicating a slight change in tertiary structure. This change in fluorescence is most likely a result of a variation in the local environment of Trp 65 to a less polar position. Trp 65 is located in the N-terminal domain of the protein but packs next to the altered region and is perhaps affected by the V139I alteration.(*147*) C145S and R135G variants of the chimera protein were created by site-directed mutagenesis, purified and characterized (see Appendix V).

Table 5.3: Amino acid sequence of hAGT, OGT and chimera active sites

Protein	Amino Acid Sequence (139-159)[a]
hAGT	V***P***ILIPCHRVVCSSGAVGNYS
OGT	ISIVVPCHRVIGRNGTMTGYA
hAGT-03[b]	VPILIP**C**HRVIGRNGTMTGYS
Chimera	I***P***IVVPCHRVIGRNGTMTGYA

[a] Amino acid numbering based on hAGT sequence.
[b] Most active chimera designed in reference *123*.

The P140S variant (introducing Ser 140 found in OGT) of the chimera created an insoluble construct when expressed under the pQE30 vector. This protein was also expressed using pET15b and purified under denaturing conditions in order to obtain a

higher protein yield to be used in refolding attempts. Previously published refolding protocols for hAGT did not yield an appreciable amount of properly folded protein.(*105*) Dialysis mediated refolding of the P140S chimera variant was greatly aided by high salt, Tris and L-arginine. Unfortunately, the refolded protein had an improper conformation as assessed by far-UV CD and intrinsic fluorescence and was unable to repair O^6-MedG and O^4-MedT readily (data not shown).

5.3.2 Mono-adduct repair

The chimeras ability to repair O^6-MedG in a 14 bp oligonucleotide (5' GGC TT**X** ATC ACC AG, where **X** is O^6-MedG) appeared to be unaffected given the repair of O^6-MedG occurred in less than 15 sec at room temperature, as was the case with hAGT (data not shown). The alterations improved the AGTs ability to remove O^4-MedT in a similar oligonucleotide (5' GGC T**X**G ATC ACC AG, where **X** is O^4-MedT). As observed in **Figure 5.2**, the chimera repaired O^4-MedT 30 times faster than hAGT, which is comparable to results reported by the Pegg group for their insoluble constructs, where their most proficient variant had a 47-fold increase over hAGT.(*123*) Repair of O^4-MedT by OGT required less than 15 sec making it at least 100-fold more proficient than the chimera (data not shown).

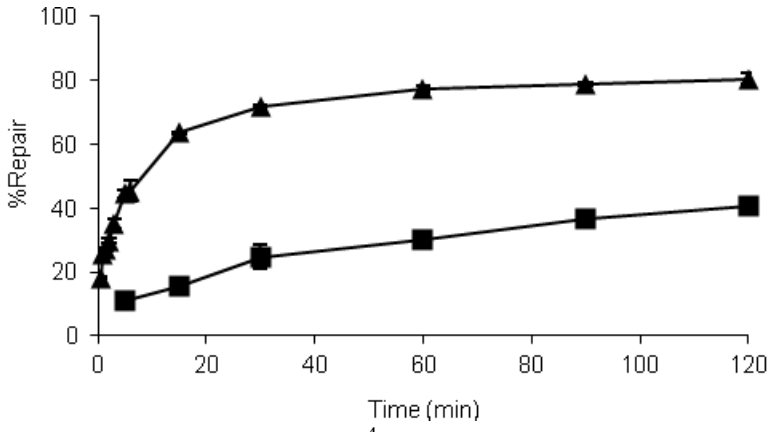


Figure 5.2: Time course dealkylation of O^4-MedT

Repair assays was conducted at room temperature with hAGT (■) and the chimera (▲).

5.3.3 Modelling of chimera bound to O^4-alkyl dT mono-adducts

Molecular modelling and molecular dynamics were employed to comprehend the reason for the chimeras increased ability to remove O^4-MedT over its human equivalent. Modelling of the chimera was performed using the hAGT-O^6MeG co-crystal (PDB 1T38) as template and placement of the Cys 145 side chain was guided by the hAGT crystal structures, PDBs 1YFH, 1T39 and 1EH6. O^6-MedG was mutated to O^4-MedT and the placement of O^4-MedT verified with PDB 1YFH, as reference. Conjugate gradient minimization, simulated annealing and torsion angle dynamics were conducted using Crystallography & NMR System as done previously by our group.(*222*,*235*)

The modelling data suggest the increased repair of O^4-MedT by the chimera may be due to changes in the shape of the loop that better accommodates the modified nucleotide as observed in **Figure 5.3**. The N157G and S159A alterations in the chimera also make the loop more flexible resulting in a reduction in distance between the Cys 145 thiolate anion in the chimera to the methyl adduct of O^4-MedT by 0.65 Å relative to the native hAGT, which may be responsible for the enhanced repair observed.

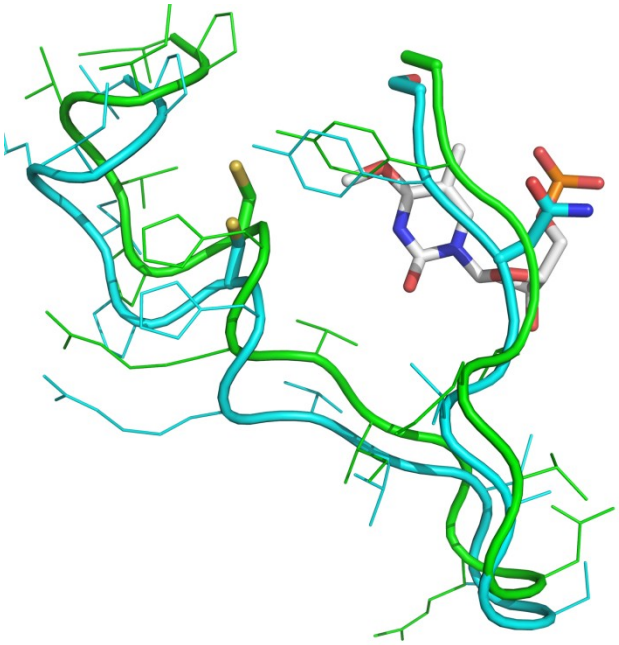


Figure 5.3: O^4-MedT (white) in the active site of hAGT (cyan/grey) and the chimera (green/black).

The side chains of Cys 145, Asn 157 and Ser/Ala 159 are represented as sticks.

hAGT has two unfavourable interactions; Asn 157 which sterically clashes with the O^4-MedT sugar moiety and Ser 159 that is in close proximity with the C5 methyl of O^4-MedT hindering the flexibility of the loop and consequently inhibiting any penetration of O^4-MedT into the protein's active site.

The pre-minimized model of the chimera in complex with O^4-MedT revealed a potential clash between Arg 135 and the C5 methyl of the modified nucleotide. This residue is replaced by a Gly in OGT and Ala in Ada-C, suggesting that small amino acids are usually located at this position in AGTs capable of repairing O^4-MedT with appreciable ease. For this reason the R135G variant of the chimera was generated in hope of creating a more active variant. The repair capabilities of the chimera were compared to its R135G variant using a small library of O^4-alkyl dT adducts. Overnight incubation at 37 °C of the AGT with the various O^4-alkylated DNA revealed the R135G alteration was

detrimental to the chimera. The R135G variant maintained the ability to repair O^4-MedT, but lost its capability to repair substrates **T4** and **T7** (see Appendix V). Repair of **T4** and **T7** by the original chimera was moderate, showing only 40% repair after an overnight incubation. These results are still promising since its human counterpart is unable to react with these substrates to a detectable extent.(*235*)

5.3.4 Cross-link repair

hAGT is the only AGT shown to date to repair ICL DNA and several independent alterations in its amino acid sequence are known to abolish this ability. These alterations (aside from those involved in the proteins activity) include: P140A, P140K, V148L and Y158H, all of which are located in the portion of hAGT that is altered to form the chimera.(*146*,*214*) To verify the effect of the alterations on the proteins ability to process ICL DNA time course repair assays with various ICL containing oligonucleotides were conducted. The time course assays revealed that the chimera is similar to hAGT and can repair ICL DNA involving the O^6 atom of dG. The repair of cross-linked substrates **XLGG4** and **XLGG7**, where the O^6 atoms of dG are attached by an alkylene linker, were carried out at virtually identical rates by hAGT and the chimera. The chimera, much like hAGT, repaired **XLGG7** more rapidly and extensively than it did **XLGG4**, 50% repair in 3 h for **XLGG7** as opposed to 30% repair in 8 h for **XLGG4** (see Appendix V).

Similar repair trends were observed for the **XLGT** series, where the O^6 atom of dG and O^4 atom of dT are attached by an alkylene linker. **XLGT7** was repaired with comparable rates by both AGTs whereas the repair of **XLGT4** was virtually non-existent (see Appendix V). ESI-MS analysis of the repair reaction revealed the chimera selectively repairs the cross-link at the O^6 atom of dG. The repair event produces a

complex comprised of the chimera covalently attached through its Cys 145 to the O^4 atom of the cross-linked dT *via* the seven methylene linkage (see Appendix V). The same repair specificity was reported by our group for hAGT.(*249*)

In light of the positive results with mono-adducts **T4** and **T7** repair assays with cross-linked DNA, where the O^4-atoms of dT were attached by an alkylene linker (**XLTT4** and **XLTT7**), were undertaken with the chimera. Unfortunately, the chimera was unable to repair **XLTT4** nor **XLTT7** (see Appendix V). As was the case with hAGT and OGT, **XLTT4** and **XLTT7** ICL DNA evaded repair by the chimera.(*235*)

5.3.5 Binding studies

The chimera bound **XLTT4** and **XLTT7** with low μM K_d as determined by EMSA indicating good binding of the ICL DNA by the protein. The chimera bound the **XLTT** DNA series more tightly than it did the **XLGT**, which the protein is able to partially repair, suggesting the absence of **XLTT** repair is not due to recognition of the substrate by the protein (see **Table 5.4**).

Table 5.4: K_d and stoichiometry of the C145S chimera variant binding to ICL DNA

DNA	K_d (μM)	Protein:DNA Stoichiometry
Control (T-A)[a]	37.57±6.48	1.91 ± 0.09
XLTT4	2.56±0.42	1.86 ± 0.16
XLTT7	1.99±0.12	1.80 ± 0.16
Control (G-C)[b]	3.96±0.82	1.89 ± 0.18
XLGT4	6.10±0.10	2.12 ± 0.20
XLGT7	3.88±0.47	1.99 ± 0.13

[a] DNA sequence is 5'-CGAAA<u>T</u>TTTCG/3'-GCTTT<u>A</u>AAAGC
[b] DNA sequence is 5'-CGAAA<u>G</u>TTTCG/3'-GCTTT<u>C</u>AAAGC

The chimera interacts with dsDNA in a virtually identical fashion to hAGT, as indicated by our binding experiments. Like hAGT, the chimera displays a preference for DNA with a dG that is more than 2 nucleotides from the 5' ends, as observed between the

T-A and G-C controls. The chimera bound all DNA species with comparable dissociation constants and stoichiometry as those reported for hAGT using identical substrates and conditions.(*235*,*249*) hAGT is believed to bind every 4 nucleotides where an observed stoichiometry of 2:1 hAGT:DNA has been documented for both 11-mer dsDNA and ssDNA, consistent with our findings with the chimera.(*166*,*167*)

Binding experiments with the R135G chimera variant showed no association of DNA in the tested range (0 to 30 μM) establishing the role of Arg 135 as a binding residue. The R135G alteration in the chimera appears to cause a decrease in binding affinity between the protein and DNA, limiting the proteins ability to repair bulkier adducts (such as **T4** and **T7**) since they require a longer time to react with the proteins thiolate anion as a result of the distance between the reacting groups. The role of Arg 135 in the chimera is equivalent to its function in hAGT, where it is speculated that this residue is 1 of only 3 basic amino acids involved in the ionic interaction between the protein and dsDNA as well as being 1 of 2 residues involved in the interaction with ssDNA.(*168*)

5.4 Conclusions

We successfully generated a hAGT-OGT chimera that can be purified under native conditions. The chimera, containing a portion of the hAGT active site changed to the respective amino acids in OGT, displays some properties from OGT since it can repair O^4-MedT 30 times faster than hAGT. In addition, the chimera can remove larger adducts at the O^4 atom of dT, such as those found in **T4** and **T7**, but does so poorly. The chimera maintained the ability to repair cross-links involving the O^6 atom of dG, a property that

was previously thought to be unique to hAGT (see **Table 5.5** for repair summary). The alterations in the primary construct were shown to have no effect on DNA binding affinity or stoichiometry, suggesting the modified residues are not involved in substrate recognition

Such a chimera may play a role in combinational chemotherapeutic regimens due to its added ability to repair undesired, highly mutagenic, O^4-alkyl dT adducts that are generated during cancer treatment and which are known to persist in mammalian cells.(*252*) Since the chimera is based mostly on the hAGT scaffold many of the biological properties of the protein are maintained. For example, the presence of the 125-KAAR sequence responsible for targeting and maintaining the protein in the nucleus and the presence of Met 134, which is required for the steric clash that allows hAGT to be ubiquitinated and degraded are retained; the presence of the terminal regions of hAGT that impart the protein with the ability to interact with DNA with a high inter-protein cooperativity, which is required for concentrating the protein to segments of DNA that are undergoing transcription are also preserved.(*153*,*154*,*155*,*171*).

Table 5.5: Summary of AGT mediated DNA repair

Lesion	hAGT	OGT	Chimera
O^4MeT[a (235)]	+	+++	++
T4[a (235)]	-	++	+
T7[a (235)]	-	+	+
XLGG4[b (146)]	+	-	+
XLGG7[b (146)]	+	-	+
XLGT4[b (249)]	-	-	-
XLGT7[b (249)]	+	-	+
XLTT4[b (235)]	-	-	-
XLTT7[b (235)]	-	-	-

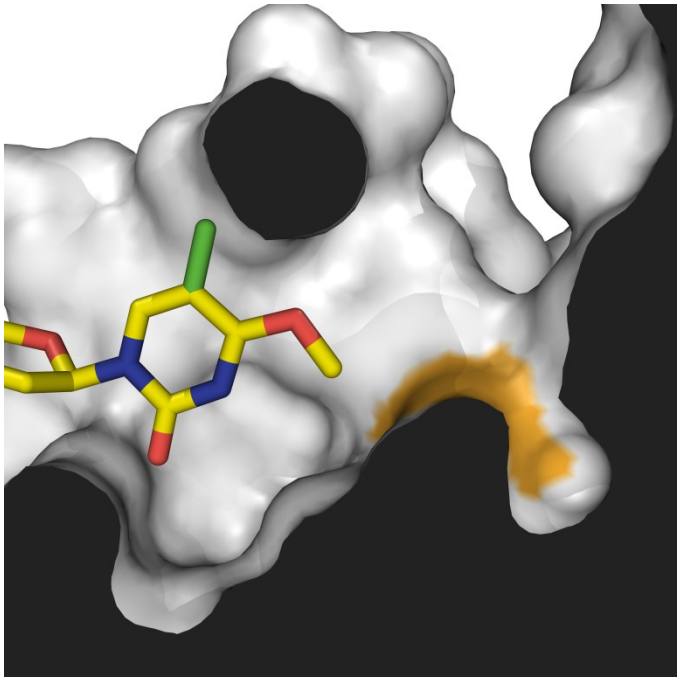
[a]10 pmol AGT with 2 pmol DNA at 37 °C. [b]60 pmol AGT with 2 pmol DNA at 37 °C. (+++) >80% repair within 1 minute, (++) >80% repair within 1 h, (+) >10% repair within 16 h and (-) <10% repair within 16 h.

5.5 Acknowledgements

We are grateful to Dr. Anthony. E. Pegg (Penn. State University) for the plasmid encoding the wild-type hAGT gene. 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. FPM is the recipient of a postgraduate fellowship from NSERC.

CHAPTER VI

Effect of the C5 Methyl on O^6-Alkylguanine-DNA Alkyltransferase Mediated Repair of O^4-alkyl Thymidine Adducts



Graphical Abstract: O^4-MedU (yellow) modelled into hAGT active site, where the C5 methyl that would be present if it were O^4-MedT is shown in green

McManus FP, Wilds CJ.

To be submitted to:

Biochemistry, February 2013.

Abstract

AGT are responsible for the removal of O^6-alkyl dG and O^4-alkyl dT adducts from the genome. Unlike the *E. coli* OGT protein, which can repair a vast range of O^4-alkyl dT lesions, hAGT is only able to remove O^4-MedT and does so poorly. We set out to uncover the cause of this substrate discrimination pattern by determining the effect of the C5 methyl of dT on AGT mediated repair of O^4-alkyl dT. The rate at which the *E. coli* AGTs, hAGT and an OGT/hAGT chimera removed methyl, butyl-4-ol and heptyl-7-ol adducts from the O^4 atom of dU were determined and compared to those of O^4-alkyl dT, which we previously reported under identical conditions. The presence of the C5 methyl hindered the repair of the methyl adducts by hAGT and the chimera by 25 and 30-fold, respectively. The absence of the C5 methyl allowed the chimera to repair the butyl-4-ol and heptyl-7-ol adducts with appreciable rates. The presence of the C5 methyl did not greatly affect the rate of repair of the butyl-4-ol adduct by OGT but had a 30-fold inhibitory affect on the removal of the heptyl-7-ol lesion. The ability of the various AGT to repair O^4-benzyl dU in an oligonucleotide was also investigated. This molecule was repaired efficiently by all AGTs except for Ada-C. No correlation was observed between the binding affinity of the AGT homologous to the adduct containing oligonucleotides and their rates of repair.

6.1 Introduction

The formation of O^6-alkyl dG and O^4-alkyl dT lesions in DNA through exposure to chemotherapeutic agents and environmental carcinogens such as tobacco derived N-nitrosamines is well documented.(*54*,*242*) Such modifications also occur endogenously where the formation of O^6-MedG is observed at a rate of 10-30 events daily by *S*-adenosylmethionine.(*46*) These lesions are often used as benchmarks to determine the mutagenic potential of DNA damaging agents due to their potency.(*243*) The source of the mutagenic potential of these modifications come from the wobble base pairs adopted by these altered heterocycles with their natural counterparts. As a result O^6-MedG and O^4-MedT lead to G to A and T to C transitions, respectively.(*21*,*25*)

In the genome, O^4-MedT is formed less readily than O^6-MedG. DNA treated with MNU produces 126 times more O^6-MedG than O^4-MedT.(*24*) Though formed to a lesser extent, O^4-MedT is observed at similar steady state levels as O^6-MedG in the liver tissue of healthy humans, proving its incidence, even in the absence of exogenous alkylating agents.(*246*) Moreover, O^4-MedT is more toxic than O^6-MedG in both normal and repair deficient mammalian cells.(*29*,*30*)

As the source of genetic information the nucleotide sequence found in DNA must be maintained and therefore the prevalence of O^6-alkyl dG and O^4-alkyl dT lesions limited. Organisms have evolved AGT proteins to remove these lesions and limit their effects on the cell. AGTs are virtually omnipresent and are found in all kingdoms of life apart from plants.(*189*) hAGT is the most thoroughly characterized AGT and has been crystallized in the absence and presence of different substrates.(*148*,*155*,*151*) AGTs

repair DNA damage by flipping the damaged nucleotide out of the duplex and into the active site of the protein allowing the alkyl group to be transferred from the modified nucleotide to the active site Cys residue.(*173*) The alkylated protein is subsequently ubiquitinated and targeted for degradation by the proteosome causing the protein to act only once.(*104*)

The augmented persistence of O^4-MedT in the genome and it's mutagenic potential is not based on its chemical nature but is rather attributed to its ability to evade repair by hAGT and MMR, unlike O^6-MedG.(*31*,*132*,*133*) The preferential repair of O^6-alkyl dG observed by hAGT is not ubiquitous amongst AGTs. For example, the *E. Coli* OGT and Ada-C proteins are capable of repairing O^4-MedT readily, where OGT is documented as having the greatest activity and repairs both O^4-MedT and O^6-MedG with comparable rates.(*33*,*117*)

Unfortunately, the origin of the substrate discrimination patterns by the AGT homologues is not known. Moreover the reason for hAGTs inability to process O^4-MedT is unknown and few hypothesis have been proposed. To date, there is no crystal structure of hAGT with a DNA duplex containing O^4-MedT. hAGT has been crystallized with the modified pyrimidine N^4-*p*-xylylenediamine, a cytosine analogue, but the nature of the modification may cause some potential interactions to be lost.(*155*) Notably, the absence of an O^4 atom in dC removes the Ser 159 backbone interaction with the O^4 atom of dT (observed with O^6 of dG in PDB 1T38).

Our group has generated a molecular model of hAGT in complex with O^4-MedT by substituting O^6-MedG in PDB 1T38.(*235*) The resulting model placed the methyl group on the O^4 atom of dT 2.39 Å further away from the thiolate anion of Cys 145 with respect

to the methyl group of O^6-MedG. These results are in agreement with experimental evidence but they did not explain the root cause of hAGTs reduced capacity to repair O^4-MedT. A closer look at the pre-minimized model revealed a potential clash between the C5 methyl of dT and Arg 135 of hAGT. Unfortunately, altering Arg 135 in hAGT would create a protein that has reduced DNA binding affinity because this residue is believed to be 1 of only 3 basic amino acids involved in the ionic interaction between the protein and phosphate backbone of dsDNA.(*168*)

We previously engineered a soluble chimera of OGT-hAGT where residues 139-159 of hAGT were replaced with their equivalents in OGT with the addition of an S140P alteration for solubility purposes.(*253*) This chimera possesses increased repair capabilities of O^4-alkyl dT damage with respect to its human homologue, while maintaining the ability to repair large adducts at the O^6-atom of dG. The R135G alteration in this chimera lead to a protein with reduced repair capabilities of O^4-alkyl dT, which was attributed to reduced DNA binding affinity, supporting the proposed role of Arg 135 in hAGT. Since only the active site is altered in the chimera and the HTH motif is essentially unchanged from that of hAGT, this protein will allow to determine the role of the active site residues on AGT mediated repair of various O^4-alkyl pyrimidine damage.

Since the residue involved in the potential clash cannot be altered due to its role in substrate binding, a synthetic approach was undertaken to synthesize similar oligonucleotides as used previously by our group for repair studies of O^4-alkyl dT. For this study, the O^4-alkyl dT nucleotide is substituted to its C5 lacking equivalent, O^4-alkyl dU. The synthesized substrates, shown in **Figure 6.1**, were prepared using methods

described by Swann's group for the modification of dT using a convertible nucleoside approach without the need for alterations.(*221*)

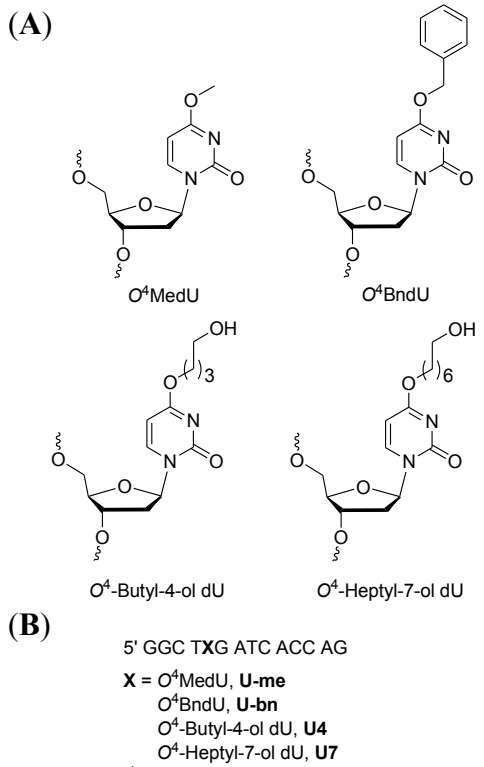


Figure 6.1: Structures of (**A**) O^4-alkyl dU adducts and (**B**) DNA sequence where X contains the adduct.

6.2 Experminental

6.2.1 Materials

2'-Deoxyuridine (dU) and *N*,*N*-diisopropylamino cyanoethyl phosphonamidic chloride were purchased from ChemGenes Inc. (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 Inc. (Gibbstown, NJ). Flash column chromatography was performed using silica gel 60 (230–400 mesh) obtained from Silicycle (Quebec City, QC). TLC was performed using precoated TLC plates (Merck, Kieselgel 60 F_{254}, 0.25 mm) purchased from EMD Chemicals Inc. (Gibbstown, NJ). NMR spectra were either recorded on a Varian INOVA 300 MHz NMR spectrometer or on a Varian 500 MHz NMR spectrometer at room temperature. ^{1}H NMR spectra were recorded at a frequency of 300 or 500 MHz and chemical shifts were reported in parts per million downfield from tetramethylsilane. ^{31}P NMR spectra (^{1}H decoupled) were recorded at a frequency of 202.3 MHz with H_3PO_4 used as an external standard. Ampicillin, IPTG, DTT 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). XL-10 Gold and B834(DE3)pLysS *E. coli* cells were obtained from Stratagene (Cedar Creek, TX). T4 PNK, Unstained Protein Molecular Weight Marker, *NdeI* and *EcoRI* restriction enzymes were obtained from Fermentas (Burlington, ON). [γ-^{32}P]ATP was purchased from Amersham Canada Ltd. (Oakville, ON).

6.2.2 Solution synthesis of small molecules

3'-*O*-(*t*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-C4-(1,2,4-triazol-1-yl)-2'-deoxyuridine (**1**) was prepared by a three step synthesis as previously described.(*254*)

5'-*O*-(4,4'-dimethoxytrityl)-O^4-(methyl)-2'-deoxyuridine (**2a**)

To an ice cold solution of (**1**) (0.600 g, 0.86 mmol) in ACN (4.7 mL) was added methanol (0.355 mL, 8.77 mmol) followed by DBU (0.295 mL, 1.97 mmol) and the

reaction allowed to progress on ice for 30 min then placed at room temperature. After 16 h, the solvent was evaporated in vacuo, the crude product taken up in DCM (50 mL) and the solution washed with two portions of brine (2 x 100 mL) followed by a back extraction of the brine washes with 5 mL of DCM. The organic layer was dried over sodium sulphate, concentrated and placed on the high vacuum for 2 hours to produce a brown foam.

To a solution of the brown foam in THF (10.0 mL) was added TBAF (1M) in THF (1.03 mL, 1.03 mmol). After 30 min, the solvent was evaporated in vacuo, the crude product taken up in DCM (25 mL), the solution was washed with two portions of sodium bicarbonate (2 x 125 mL) followed by a back extraction of sodium bicarbonate washes with DCM (5 mL). The organic layer was dried over sodium sulphate and concentrated to produce a brown gum. The crude product was purified by flash column chromatography using a DCM / methanol (49 : 1 → 97 : 3) solvent system to afford 0.460 g (98.2 %) of product as a colorless foam.

R_f (SiO_2 TLC): 0.20 DCM / methanol (49:1). ^1H NMR (500MHz, $CDCl_3$, ppm): 8.03 (d, 1H, H6, J= 7.6 Hz), 7.19-7.36 (m, 9H, Ar), 6.79-6.82 (m, 4H, Ar), 6.25 (dd, 1H, H1', J= 5.5 Hz), 5.60 (d, 1H, H5, J= 7.6 Hz), 4.47 (m, 1H, H3'), 4.02 (m, 1H, H4'), 3.92 (s, 3H, $ArOCH_3$), 3.78 (s, 6H, OCH_3), 3.50 (dd, 1H, H5', J= 10.8, 3.4 Hz), 3.41 (dd, 1H, H5'', J= 10.8, 3.4 Hz), 2.60 (m, 1H, H2'), 2.25 (m, 1H, H2''), 1.90 (d, 1H, OH). HRMS (ESI-MS) *m/z*: Observed 567.2100, calculated for $C_{31}H_3N_2O_7Na^+$ 567.2101 $[M+Na]^+$

5'-*O*-(4,4'-dimethoxytrityl)-O^4-(benzyl)-2'-deoxyuridine (2b)

To an ice cold solution of (**1**) (0.600 g, 0.86 mmol) in ACN (4.7 mL) was added benzyl alcohol (0.909 mL, 8.77 mmol) followed by DBU (0.295 mL, 1.97 mmol). The

reaction was allowed to progress on ice for 30 min then placed at room temperature. After 16 h, the solvent was evaporated in vacuo, the crude product taken up in DCM (100 mL) and the solution washed with two portions of brine (2 x 250 mL) followed by a back extraction of the brine washes with DCM (5 mL). The organic layer was dried over sodium sulphate and concentrated and placed on the high vacuum for 2 hours to produce a brown foam.

To a solution of the brown foam in THF (10.0 mL) was added TBAF (1M) in THF (1.58 mL, 1.58 mmol). After 30 min, the solvent was evaporated in vacuo, the crude product taken up in DCM (25 mL), the solution washed with two portions of sodium bicarbonate (2 x 125 mL) followed by a back extraction of the sodium bicarbonate washes with DCM (5 mL). The organic layer was dried over sodium sulphate and concentrated to produce a brown gum. The crude product was purified by flash column chromatography using a DCM / methanol (99 : 1 → 49 : 1) solvent system to afford 0.443 g (83.0 %) of product as a colorless foam.

R_f (SiO_2 TLC): 0.225 DCM / methanol (49:1). ^{1}H NMR (500MHz, $CDCl_3$, ppm): 8.05 (d, 1H, H6, J= 7.4 Hz), 7.18-7.36 (m, 14H, Ar), 6.79-6.81 (m, 4H, Ar), 6.28 (dd, 1H, H1', J= 5.5 Hz), 5.65 (d, 1H, H5, J= 7.4 Hz), 5.38 (s, 2H, $ArOCH_2Ar$), 4.49 (m, 1H, H3'), 4.04 (m, 1H, H4'), 3.77 (m, 6H, OCH_3), 3.50 (dd, 1H, H5', J= 10.6, 3.3 Hz), 3.41 (dd, 1H, H5'', J= 10.6, 3.3 Hz), 2.62 (m, 1H, H2'), 2.26 (m, 1H, H2''), 2.09 (d, 1H, OH). HRMS (ESI-MS) *m/z*: Observed 643.2414, calculated for $C_{37}H_{36}N_2O_7Na^+$ 643.2415 $[M+Na]^+$

3′-*O*-(*β*-cyanoethyl-*N*,*N*′-diisopropyl)-5'-*O*-(4,4'-dimethoxytrityl)-O^4-(methyl)-2'-deoxyuridine (**3a**)

To a solution of (**2a**) (0.350 g, 0.643 mmol) in THF (3.0 mL) was added DIPEA (168 μL, 0.964 mmol) followed by the dropwise addition of *N*,*N*-diisopropylamino cyanoethyl phosphonamidic chloride (172 μL, 0.770 mmol). After 30 min, the solvent was evaporated in vacuo, the crude product taken up in ethyl acetate (50 mL), the solution washed with two portions of sodium bicarbonate (2 x 100 mL) and once with brine (100 mL). The organic layer was dried over sodium sulphate and concentrated. The product, a colorless powder, was resuspended in 3 mL of ethyl acetate and precipitated from 1L of hexanes (0.177 g, 37.0 %).

R_f (SiO_2 TLC): 0.36, 0.55 hexanes / ethyl acetate (2:8). ^{31}P NMR (202.3 MHz, d_6-acetone, ppm): 148.36, 148.23. HRMS (ESI-MS) *m/z*: Observed 745.3369, calculated for $C_{40}H_{50}N_4O_8P^+$ 745.3366 $[M+H]^+$

3′-*O*-(*β*-cyanoethyl-*N*,*N*′-diisopropyl)-5'-*O*-(4,4'-dimethoxytrityl)-O^4-(benzyl)-2'-deoxyuridine (3b)

To a solution of (**2b**) (0.413 g, 0.665 mmol) in THF (3.0 mL) was added DIPEA (191 μL, 1.089 mmol) followed by the dropwise addition of *N*,*N*-diisopropylamino cyanoethyl phosphonamidic chloride (196 μL, 0.871 mmol). After 30 min, the solvent was evaporated in vacuo, the crude product taken up in ethyl acetate (50 mL), the solution washed with two portions of sodium bicarbonate (2 x 100 mL) and once with brine (100 mL). The organic layer was dried over sodium sulphate and concentrated. The product, a colorless powder, was resuspended in 3 mL of ethyl acetate and precipitated from 1L of hexanes (0.265 g, 48.8 %).

R_f (SiO_2 TLC): 0.72, 0.84 hexanes / ethyl acetate (2:8). ^{31}P NMR (202.3 MHz, d_6-acetone, ppm): 148.37, 148.24. HRMS (ESI-MS) *m/z*: Observed 821.3679, calculated for $C_{46}H_{54}N_4O_8P^+$ 821.3679 $[M+H]^+$

3'-*O*-(*t*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-O^4-(hydroxybutyl)-2'-deoxyuridine (4a)

To an ice cold solution of (**1**) (1.50 g, 2.16 mmol) in ACN (21.6 mL) was added 1,4-butanediol (0.960 mL, 10.80 mmol) followed by DBU (0.743 mL, 4.97 mmol), the reaction allowed to progress on ice for 30 min and placed at room temperature. After 1.5 h, the reaction was quenched with 7 mL of saturated sodium bicarbonate, the solvent was evaporated in vacuo and the crude product taken up in DCM (100 mL); then the solution was washed with two portions of sodium bicarbonate (2 x 150 mL) followed by a back extraction of the sodium bicarbonate washes with DCM (5 mL). The organic layer was dried over sodium sulphate and concentrated to produce a brown gum. The crude product was purified by flash column chromatography using a hexane / ethyl acetate solvent system (1 : 1 → 7 : 13) to afford 1.30 g (74.2 %) of product as a colorless foam.

R_f (SiO_2 TLC): 0.24 hexane / ethyl acetate (2:3). ^1H NMR (300MHz, $CDCl_3$, ppm): 8.28 (d, 1H, H6, J= 7.3 Hz), 7.29-7.45 (m, 10H, Ar), 6.87-6.90 (m, 3H, Ar), 6.31 (dd, 1H, H1', J= 4.2 Hz), 5.59 (d, 1H, H5, J= 7.3 Hz), 4.52 (m, 1H, H3'), 4.44 (t, 2H, $ArOCH_2$, J= 6.7 Hz), 4.00 (m, 1H, H4'), 3.85 (s, 6H, OCH_3), 3.74 (dt, 2H, CH_2OH, J= 6.6, 5.3 Hz), 3.58 (dd, 1H, H5', J= 10.8, 3.0 Hz), 3.36 (dd, 1H, H5'', J= 10.8, 3.0 Hz), 2.56 (m, 1H, H2'), 2.25 (m, 1H, H2''), 1.88 (m, 2H, CH_2), 1.72 (m, 2H, CH_2), 1.63 (s, 1H, OH), 0.84 (s, 9H, $SiC(CH_3)_3$), 0.04 (s, 6H, $Si(CH_3)_2$). HRMS (ESI-MS) *m/z*: Observed 739.3388, calculated for $C_{40}H_{52}N_2O_8SiNa^+$ 739.3391 $[M+Na]^+$

3'-*O*-(*t*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-O^4-(hydroxyheptyl)-2'-deoxyuridine(4b)

To an ice cold solution of (**1**) (1.70 g, 2.44 mmol) in ACN (21.6 mL) was added 1,7-heptanediol (1.70 mL, 12.2 mmol) followed by DBU (1.11 mL, 7.46 mmol), the reaction allowed to progress on ice for 30 min and placed at room temperature. After 7 h, the reaction was quenched with 7 mL of saturated sodium bicarbonate, the solvent was evaporated in vacuo, the crude product taken up in DCM (100 mL). The solution washed with two portions of sodium bicarbonate (2 x 150 mL) followed by a back extraction of the sodium bicarbonate washes with DCM (5 mL). The solution washed with brine (150 mL) followed by a back extraction with DCM (5 mL). The organic layer was dried over sodium sulphate and concentrated to produce a brown gum. The crude product was purified by flash column chromatography using a DCM / methanol solvent system (49 : 1 → 97 : 3) to afford 1.16 g (68.2%) of product as a colorless foam.

R_f (SiO_2 TLC): 0.5 DCM / methanol (19:1). ^{1}H NMR (300MHz, $CDCl_3$, ppm): 8.25 (d, 1H, H6, J= 7.4 Hz), 7.27-7.43 (m, 10H, Ar), 6.85-6.88 (m, 3H, Ar), 6.30 (dd, 1H, H1', J= 4.5 Hz), 5.58 (d, 1H, H5, J= 7.4 Hz), 4.49 (m, 1H, H3'), 4.37 (t, 2H, $ArOCH_2$, J= 6.7 Hz), 3.97 (m, 1H, H4'), 3.83 (s, 6H, OCH_3), 3.67 (dt, 2H, CH_2OH, J= 6.6, 5.2 Hz), 3.56 (dd, 1H, H5', J= 10.7, 2.7 Hz), 3.33 (dd, 1H, H5'', J= 10.7, 2.7 Hz), 2.52 (m, 1H, H2'), 2.23 (m, 1H, H2''), 1.75 (m, 2H, CH_2), 1.59 (m, 2H, CH_2), 1.40 (s, 6H, $(CH_2)_3$), 0.83 (s, 9H, $SiC(CH_3)_3$), 0.03 (s, 6H, $Si(CH_3)_2$). HRMS (ESI-MS) *m/z*: Observed 781.3855, calculated for $C_{43}H_{58}N_2O_8SiNa^+$ 781.3855 $[M+Na]^+$

5'-*O*-(4,4'-dimethoxytrityl)-O^4-(phenoxyacetyloxybutyl)-2'-deoxyuridine (5a)

To an ice cold solution of (**4a**) (0.500 g, 0.697 mmol) in THF (7.0 mL) was added triethylamine (0.157 μL, 1.12 mmol) followed by the dropwise addition of phenoxyacetyl chloride (0.122 μL, 0.715 mmol). The reaction was allowed to progress on ice for 5 min and then allowed to come up to room temperature. After 10 min, the solvent was evaporated in vacuo, the crude product taken up in DCM (50 mL), washed with sodium bicarbonate (2 x 100 mL) followed by a back extraction of the sodium bicarbonate washes with DCM (5 mL). The organic layer was dried over sodium sulphate and concentrated to produce a yellow gum. The crude product was placed on high vacuum for 1 hour.

To a solution of the yellow gum in THF (7.0 mL) was added dropwise TBAF (1M) in THF (1.68 mL, 1.68 mmol). After 30 min, the solvent was evaporated in vacuo, the crude product taken up in DCM (50 mL), the solution washed with sodium bicarbonate (2 x 150 mL) followed by a back extraction of the sodium bicarbonate washes with DCM (5 mL). The organic layer was dried over sodium sulphate and concentrated to produce a yellow gum. The crude product was purified by flash column chromatography using a DCM / methanol solvent system (49 : 1 → 193 : 7) to afford 0.462 g (77.7 %) of product as a colorless foam.

R_f (SiO_2 TLC): 0.26 DCM / methanol (19 : 1). ^{1}H NMR (500MHz, $CDCl_3$, ppm): 8.02 (d, 1H, H6, J= 7.3 Hz), 7.19-7.37 (m, 11H, Ar), 6.80-6.97 (m, 7H, Ar), 6.24 (dd, 1H, H1', J = 6.0 Hz), 5.59 (d, 1H, H5, J= 7.3 Hz), 4.62 (s, 2H, $PhOCH_2CO$), 4.46 (m, 1H, H3'), 4.34 (m, 2H, $ArOCH_2$), 4.23 (m, 2H, CH_2OCO), 4.01 (m, 1H, H4'), 3.78 (s, 6H, OCH_3), 3.50 (dd, 1H, H5', J= 10.7, 3.4 Hz), 3.41 (dd, 1H, H5'', J= 10.7, 3.4 Hz), 2.58

(m, 1H, H2'), 2.24 (m, 1H, H2''), 1.75 (m, 4H, $(CH_2)_2$). HRMS (ESI-MS) *m/z*: Observed 759.2891, calculated for $C_{42}H_{44}N_2O_{10}Na^+$ 759.2894 $[M+Na]^+$

5'-*O*-(4,4'-dimethoxytrityl)-O^4-(phenoxyacetyloxyheptyl)-2'-deoxyuridine (5b)

To an ice cold solution of (**4b**) (0.500 g, 0.660 mmol) in THF (6.6 mL) was added triethylamine (0.194 µL, 1.39 mmol) followed by the dropwise addition of phenoxyacetyl chloride (0.183 µL, 1.32 mmol). The reaction was allowed to progress on ice for 5 min and allowed to come up to room temperature. After 10 min, the solvent was evaporated in vacuo, the crude product taken up in DCM (50 mL), the solution washed with sodium bicarbonate (2 x 150 mL) followed by a back extraction of the sodium bicarbonate washes with DCM (5 mL). The organic layer was dried over sodium sulphate and concentrated to produce a yellow gum. The crude product was placed on high vacuum for 1 hour.

To a solution of the yellow gum in THF (6.6 mL) was added TBAF (1M) in THF (1.58 mL, 1.58 mmol) dropwise. After 30 min, the solvent was evaporated in vacuo, the crude product taken up in DCM (50 mL), the solution washed with sodium bicarbonate (2 x 150 mL) followed by a back extraction of the sodium bicarbonate washes with DCM (5 mL). The organic layer was dried over sodium sulphate and concentrated to produce a yellow gum. The crude product was purified by flash column chromatography using a hexanes / ethyl acetate solvent system (4 : 6) to afford 0.790 g (84.3 %) of product as a colorless foam.

R_f (SiO_2 TLC): 0.29 hexanes / ethyl acetate (4 : 6).^{1}H NMR (500MHz, $CDCl_3$, ppm): 8.02 (d, 1H, H6, J= 7.5 Hz), 7.20-7.37 (m, 12H, Ar), 6.80-6.97 (m, 6H, Ar), 6.26 (dd, 1H, H1', J = 6.0 Hz), 5.60 (d, 1H, H5, J= 7.5 Hz), 4.61 (s, 2H, $PhOCH_2CO$), 4.47

(m, 1H, H3'), 4.32 (t, 2H, $ArOCH_2$, J= 6.6 Hz), 4.18 (t, 2H, CH_2OCO, J= 6.6 Hz), 4.02 (m, 1H, H4'), 3.78 (s, 6H, OCH_3), 3.50 (dd, 1H, H5', J= 10.7, 3.4 Hz), 3.41 (dd, 1H, H5'', J= 10.7, 3.4 Hz), 2.59 (m, 1H, H2'), 2.25 (m, 1H, H2''), 1.80 (s, 1H, OH), 1.63-1.72 (m, 4H, $(CH_2)_2$), 1.33 (m, 6H, $(CH_2)_3$). HRMS (ESI-MS) *m/z*: Observed 801.3359, calculated for $C_{45}H_{50}N_2O_{10}Na^+$ 801.3363 $[M+Na]^+$

3′-*O*-(*β*-cyanoethyl-*N*,*N*′-diisopropyl)-5'-*O*-(4,4'-dimethoxytrityl)-*O*^{4}-(phenoxyacetyloxybutyl)-2'-deoxyuridine (6a)

To a solution of (**5a**) (0.440 g, 0.597 mmol) in THF (3.0 mL) was added DIPEA (172 µL, 0.986 mmol) followed by the dropwise addition of *N*,*N*-diisopropylamino cyanoethyl phosphonamidic chloride (176 µL, 0.789 mmol). After 30 min, the solvent was evaporated in vacuo, the crude product was taken up in ethyl acetate (50 mL), the solution washed with two portions of sodium bicarbonate (2 x 100 mL) and once with brine (100 mL). The organic layer was dried over sodium sulphate and concentrated. The product, a colorless gum, was resuspended in 3 mL of ethyl acetate and precipitated from 1L of hexanes (0.452 g, 80.4 %).

R_f (SiO_2 TLC): 0.53, 0.69 hexanes / ethyl acetate (2:8). ^{31}P NMR (202.3 MHz, d_6-acetone, ppm): 148.35, 148.22. HRMS (ESI-MS) *m/z*: Observed 937.4156, calculated for $C_{51}H_{62}N_4O_{11}P^+$ 937.4153 $[M+H]^+$

3′-*O*-(*β*-cyanoethyl-*N*,*N*′-diisopropyl)-5'-*O*-(4,4'-dimethoxytrityl)-*O*^{4}-(phenoxyacetyloxyheptyl)-2'-deoxyuridine (6b)

To a solution of (**5b**) (0.400 g, 0.514 mmol) in THF (3.0 mL) was added DIPEA (147 µL, 0.848 mmol) followed by the dropwise addition of *N*,*N*-diisopropylamino cyanoethyl phosphonamidic chloride (151 µL, 0.678 mmol). After 30 min, the solvent

was evaporated in vacuo, the crude product taken up in ethyl acetate (50 mL), the solution was washed with two portions of sodium bicarbonate (2 x 100 mL) and once with brine (100 mL). The organic layer was dried over sodium sulphate and concentrated. The product, a colorless gum, was resuspended in 3 mL of ethyl acetate and precipitated from 1L of hexanes (0.348 g, 69.0 %).

R_f (SiO_2 TLC): 0.69, 0.80 hexanes / ethyl acetate (2:8). ^{31}P NMR (202.3 MHz, d_6-acetone, ppm): 148.35, 148.22. HRMS (ESI-MS) *m/z*: Observed 979.4620, calculated for $C_{54}H_{68}N_4O_{11}P^+$ 979.4622 $[M+H]^+$

6.2.3 Oligonucleotide synthesis and purification

The solid-phase synthesis of O^4-alkyl dU containing oligonucleotides **U-Me, U-Bn, U4** and **U7** (**Figure 6.1**) were performed on an Applied Biosystems Model 3400 synthesizer on a 1 μmole scale employing standard β-cyanoethylphosphoramidite cycles as indicated by the manufacturer with certain modifications to coupling times as indicated below. 'Fast-deprotecting' 3'-*O*-2'-deoxyphosphoramidites were utilized at 0.1 M in ACN and modified phosphoramidites **3a**, **3b**, **6a** and **6b** (**Scheme 6.1**) at 0.15 M in ACN with phenoxyacetic anhydride as the capping agent due to the documented labile nature of adducts at the O^4-atom of dT upon extended deprotection times, which also applies to dU.(*196*)

Assembly of sequences was performed as described for **T4** and **T7** in **Section 3.2.3** with a phosphoramidite coupling time for modified phoshoramidites **3a**, **3b**, **6a** and **6b** of 10 min.

The support bound oligonucleotides were deprotected in a solution of 10% DBU in methanol (for **U-Me**), benzyl alcohol (for **U-Bn**), 1,4-butanediol (for **U4**) or 1,7-

heptanediol (for **U7**) for 5 days at room temperature in the dark to ensure that the adducts were not compromised.(*221*) For **U7**, which was deprotected with 1,7-heptanediol, an extraction of the aqueous layer with DCM was required prior to applying to the solution to a C-18 SEP PAK cartridge.

6.2.4 Oligonucleotide characterization by ESI-MS and nuclease digestion

ESI mass spectra and exonuclease analysis for oligonucleotides were obtained as described in **Section 2.2.4**.

The retention times of the eluted peaks were compared to the standard nucleotides which eluted at the following times: dC (4.6 min), dU (5.5 min), dG (6.7 min), dT (7.6 min), dA (8.9 min), O^4-MedU (9.6 min), O^4-BndU (27.8 min), O^4-butyl-4-ol dU (12.3 min) and O^4-heptyl-7-ol dU (22.7 min), and the ratio of nucleosides determined.

6.2.5 UV thermal denaturation studies

Thermal denaturation studies were performed and processed as described in **Section 2.2.5**.

6.2.6 CD spectroscopy

CD experiments were carried out as described as in **Section 2.2.6**.

6.2.7 Protein expression and purification

All AGT proteins except for the C146S Ada-C variant were expressed under the T5 promoter of the pQE30 vector as described in **Section 2.2.8**.

Due to the highly toxic nature of the inactive C146S Ada-C variant the gene was introduced into a tightly regulated system that would inhibit leaky expression of the

protein.(*173*,*255*) The wild-type Ada-C gene was introduced into the pET24B vector which harbours a lacI gene. The Ada-C gene was amplified from pQE30-Ada-C (*127*) by PCR using these primers: forward, 5'-GA GCA CAT ATG AGA GGA TCG CAT CAC CAT CAC CAT; reverse, 5'-CG GAT GAA TTC TTA CCT CTC CTC ATT TTC AGC TTC GCG, where the underlined nucleotides depict the *NdeI* and *EcoRI* cut-sites in the forward and reverse primers, respectively. PCR was conducted with Phusion polymerase with the following conditions: Hot start, 30 s, 98 °C; 35 cycles of denaturation at 98 °C for 10 s followed by annealing at 65 °C for 20 s followed by extension at 72 °C for 15 s; final extension at 72 °C for 10 s.

The amplified product was cleaned up using E. Z. N. A. Cycle Pure Kit (Omega bio-tek) digested with *NdeI* and *EcoRI*, separated on a 1% agarose gel, the band corresponding to 586 bp DNA fragment excised and extracted from the gel slice using E. Z. N. A. Gel Extraction Kit (Omega bio-tek). The digested product was ligated into pET24B, which was digested with *NdeI*, *EcoRI* and treated with calf intestinal alkaline phosphatase, and the ligation mixture transformed into XL-10 gold cells. The resulting mixture was plated on LB + 50 μg/mL kanamycin media. The plasmids were extracted from the colonies that grew and transformed into B834(pLysS) cells after the identity of the plasmids were verified by restriction digestion and sequencing.

Cells containing a plasmid coding for the C146S Ada-C variant were grown in 1L of LB broth supplemented with kanamycin (50 μg/mL) and chloramphenicol (100 μg/mL) until an $OD_{600} = 0.8$. IPTG was added at a concentration of 0.5 mM and the cells incubated for 4 h, 225 rpm at 37 °C. The cells were harvested by centrifugation at 6000 x g at 4 °C for 20 min, the supernatant discarded and the cell pellets taken up in

resuspension buffer composed of 20 mM Tris HCl (pH 8.0), 250 mM NaCl, 20 mM β-mercaptoethanol supplemented with Complete, Mini, EDTA-free Protease Inhibitor Cocktail Tablets at 5 mL buffer per gram of wet pellet. The cells were lysed using two passes on the French press followed by centrifugation at 17000 x g for 45 min at 4 °C. The cell lysate was introduced on a Ni-NTA column pre-equilibrated with equilibration buffer (resuspension buffer lacking the Complete, Mini, EDTA-free Protease Inhibitor). The column washed with equilibration buffer supplemented with 20 mM imidazole and the protein eluted with equilibration buffer supplemented with 200 mM imidazole. The pooled fractions were dialyzed against a buffer made up of 50 mM Tris HCl (pH 7.6), 250 mM NaCl, 20 mM β-mercaptoethanol and 0.1 mM EDTA using 8000 Da cutoff tubing.

6.2.8 Mono-adduct repair assay

An oligonucleotide sequence used by our group to monitor the repair of O^4-alkyl dT was used where an intrinsic diagnostic *BclI* cut site allows for the monitoring of dealkylation at the O^4 atom of dT or O^6 atom dG.(*235*) The sequence of the damage containing strand was 5' GGC TXG ATC ACC AG where: X represents dU (**U**), O^4-MedU (**U-Me**), O^4-BndU (**U-Bn**), O^4-butyl-4-ol dU (**U4**) and O^4-heptyl-7-ol dU (**U7**). The sequence of the complement strand was 5' CTG GT/G ATC AAG CC where the "/" represent the *BclI* cut site.

Radiolabelling and repair reactions with the substrates were conducted as described in **Section 3.2.8**.

6.2.9 Binding studies

Binding studies of the mono-adduct containing oligonucleotides were conducted as described in **Section 3.2.9**.

6.2.10 Modelling of O^4-alkyl dU into AGT active sites

The initial model of the chimera was generated using Modeller 9.10 using PDB 1T38 as template as previously described by our group.(*253*) The modified DNA was placed in the hAGT or chimera active site based on PDB 1T38 by substituting O^6-MedG to O^4-MedU, O^4-butyl-4-ol dU or O^4-heptyl-7-ol dU. The placement of the modified nucleoside was verified with PDB 1YFH.

The preliminary models underwent conjugate gradient minimization, simulated annealing, and torsion angle dynamics using Crystallography & NMR System.(*222*) CHARMM topology files were modified to use the altered nucleotides by adding restraints to keep the α-carbon of the lesion in a *syn* conformation with respect to the N3 atom of the nucleotide, as observed in PBDs 1T38, 1T39, and 1YFH.(*223*)

The models underwent 2 sets of 100 steps of gradient minimization and 200 steps of simulated annealing. The first cycle was carried out at 500 K with slow cooling at a rate of 4.5 K per cycle of dynamics. The second cycle was carried out at 350 K with slow cooling at a rate of 2.5 K per cycle of dynamics and the minimized models displayed using PyMol.

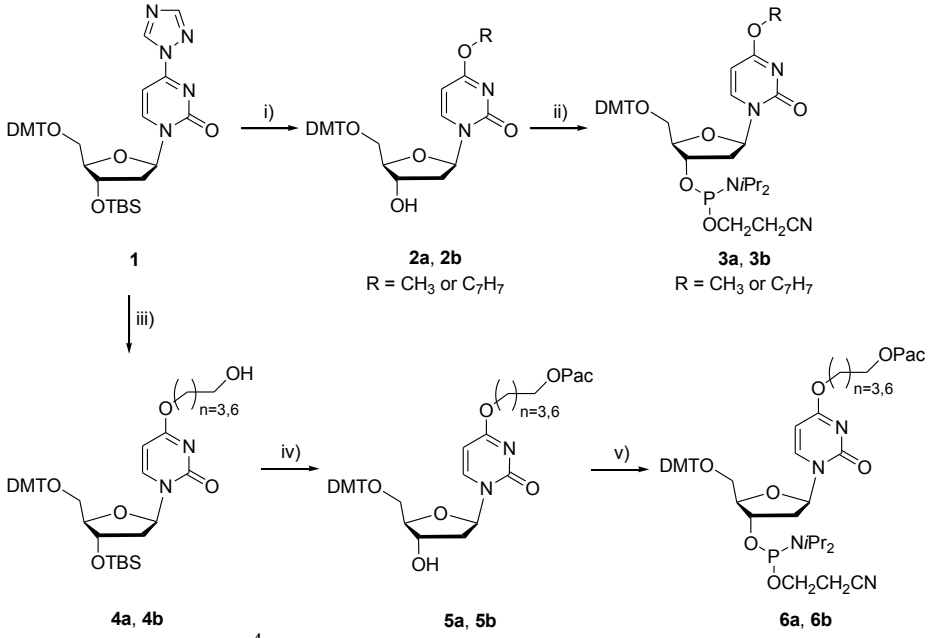
6.3 Results and discussion

6.3.1 Substrate preparation

The structures of the O^4-alkyl dU substrates generated and the oligonucleotide sequence into which they were incorporated for use in this study are shown in **Figure 6.1**. The adducts varied in size and shape to obtain an understanding on how different classes of lesions are acted upon by AGT. **U-Me**, **U4** and **U7** were synthesized to make the parallel with similar adducts studied by our group at the O^4 atom of dT in the same DNA context. **U-Bn** on the other hand was prepared to gain an understanding of : 1) substrate placement of adducts at the O^4 atom of dU in AGT active sites and 2) if this could be an O^6-BndG analogue with potent hAGT inhibiting properties.(*127*,*137*,*142*)

The synthesis of the phosphoramidite precursors required for solid phase synthesis are shown in **Scheme 6.1**. The synthesis began with a previously reported protected convertible dU derivative **1**.(*254*) The convertible nucleoside was transformed into various O^4 alkyl adducts using the appropriate alcohol (methanol for O^4-MedU, benzyl alcohol for O^4-BndU, 1,4-butanediol for O^4-butyl-4-ol dU and 1,7-heptanediol for O^4-heptyl-7-ol dU) and DBU. For the adducts that did not contain a terminal alcohol (O^4-MedU and O^4-BndU precursors), the 3'-OH were directly exposed *via* a fluoride treatment to afford compounds **2a** and **2b**, which were phosphitylated to generate **3a** and **3b,** respectively. For the terminal alcohol containing adducts **4a** and **4b**, the hydroxyl groups required introduction of protective groups that would be compatible with solid phase synthesis and easy to remove during oligonucleotide deprotection. This was achieved with a phenoxyacetyl group, which was the capping functionality to be used

during solid phase synthesis. For these compounds the 3'-OH were exposed *via* a fluoride treatment to afford **5a** and **5b** and then phosphitylated to yield **6a** and **6b**, respectively.



Scheme 6.1: Synthesis of O^4-alkyl dU amidites **3a**, **3b**, **6a** (n=3) and **6b** (n=6).

Reagents and conditions: i) MeOH (a) or BnOH (b), DBU, ACN, 16 h; TBAF (1 M in THF), 30 min; ii) *N*,*N*-diisopropylaminocyanoethylphosphonamidic chloride, DIPEA, THF, 30 min; iii) 1,4-butanediol (a) or 1,7-heptanediol (b), DBU, ACN, 1.5 h (a) or 7 h (b); iv) phenoxyacetyl chloride, triethylamine, THF, 10 min; TBAF (1 M in THF), 30 min; v) *N*,*N*-diisopropylaminocyanoethylphosphonamidic chloride, DIPEA, THF, 30 min.

The solid phase synthesis of the oligonucleotides was conducted using standard phosphoramidite chemistry, but some prospective hurdles involving the use of O^4-alkyl dU and deprotection conditions were circumvented by using a mixture of alcohol and DBU instead of the conventional ammonia ethanol treatment which is known to remove O^4-alkyl dT adducts.(*256*,*257*) Another difficulty, which was not foreseen, was the incompatibility between 1,7-heptanediol, used for the deprotection of **U7**, and the C-18 SEP PAK cartridges. The non-polar property of the diol obstructed binding of the DNA to the resin and hence desalting could not be achieved. To circumvent this issue a DCM

extraction was required prior to applying the aqueous layer to the C-18 SEP PAK cartridges.

6.3.2 Oligonucleotide characterization by ESI-MS and nuclease digestion

The molecular weights of the modified oligomers obtained from HR-MS were in agreement with the calculated masses indicating that the lesions were not compromised by the solid phase synthesis procedure or deprotection steps. The ESI-MS data was supported by the nucleoside ratios extracted from the exonuclease digestion. Exonuclease digestion of the oligonucleotides each revealed the presence of an additional peak aside from the 4 natural nucleosides, supporting the presence of the O^4-alkyl dU modifications or dU for DNA **U**.

6.3.3 UV thermal denaturation studies

The effect of the adducts at the O^4 atom of dU on thermal stability of the DNA duplex were examined. The melting temperatures of the different duplexes are shown in **Table 6.1** and their curves in Appendix VI. The non-alkylated control (**U**) displayed the highest thermal stability with a T_m of 58 °C, identical to the one observed for the same DNA having dT at **X**.(*235*) Introduction of an alkyl group at the O^4 atom of dU lead to a drop of ~10 °C in T_m, irrespective of the size or functionality of the lesion for the modifications investigated. This is also in agreement with previously published work, where methylation, hydroxybutylation and hydroxyheptylation of the O^4 atom of dT at **X** decreased the T_m of the duplex by 9-11 °C with respect to the non-alkylated control.

Clearly the absence of the C5 methyl of dU or of O^4-alkyl dU at **X** has no effect on DNA stability.

Table 6.1: Effect of O^4-alkyl dU modifications on thermal stability of DNA

DNA[1]	T_m (°C)
U	58
U-Me	48
U-Bn	48
U4	47
U7	46

[1] 5'-GGC T**X**G ATC ACC AG
CCG AAC TAG TGG TC-5'

6.3.4 CD spectroscopy

CD analysis of dsDNA containing **U**, **U-Me**, **U-Bn**, **U4** and **U7** exhibited virtually identical spectroscopic signatures characteristic of B-form DNA having a positive signal at 275 nm, a crossover at 260 nm and a negative signal at 250 nm, see **Figure 6.2**.(*200*) The adducts did not cause a dramatic effect on the overall structure of the DNA duplex.

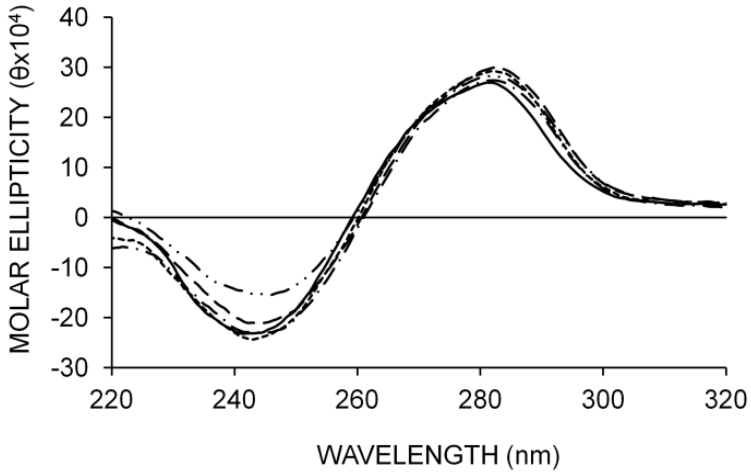


Figure 6.2: Far-UV CD spectra of duplex containing oligonucleotides **U** (——), **U-Me** (……), **U-Bn** (– • • –), **U4** (– • –) and **U7** (– – –) and the complement strand 5' CTGGTGATCCAGCC.

6.3.5 Mono-adduct repair

The repair of our O^4-alkyl dU substrate library was studied with 3 natural AGT homologues (and their variants) and a chimera (and its variants). Initially, the different substrates were incubated overnight with the various AGT to determine which proteins could repair our substrates (See Appendix VI). Time course repair studies with the "wild-type" proteins were conducted, with the omission of Ada-C since it only repaired **U-Me** and did so rapidly.

Figure 6.3A shows the repair of **U-Me** by hAGT at room temperature. The reaction required 45 min to reach completion. This represents a ~25-fold increase in rate over the repair of the alkylated dT counterpart. Complete repair of **U-Me** by the chimera occurred in less than 3 min, as observed in **Figure 6.3B**, making this protein ~20 times more efficient at repairing this lesion than its human equivalent. Repair of **U-Me** by the chimera occurred ~30 times faster than it did for O^4-MedT. The chimera was still less active than OGT since the later protein achieved total repair of the substrate in less than 15 sec, as was observed with O^4-MedT.(*235*)

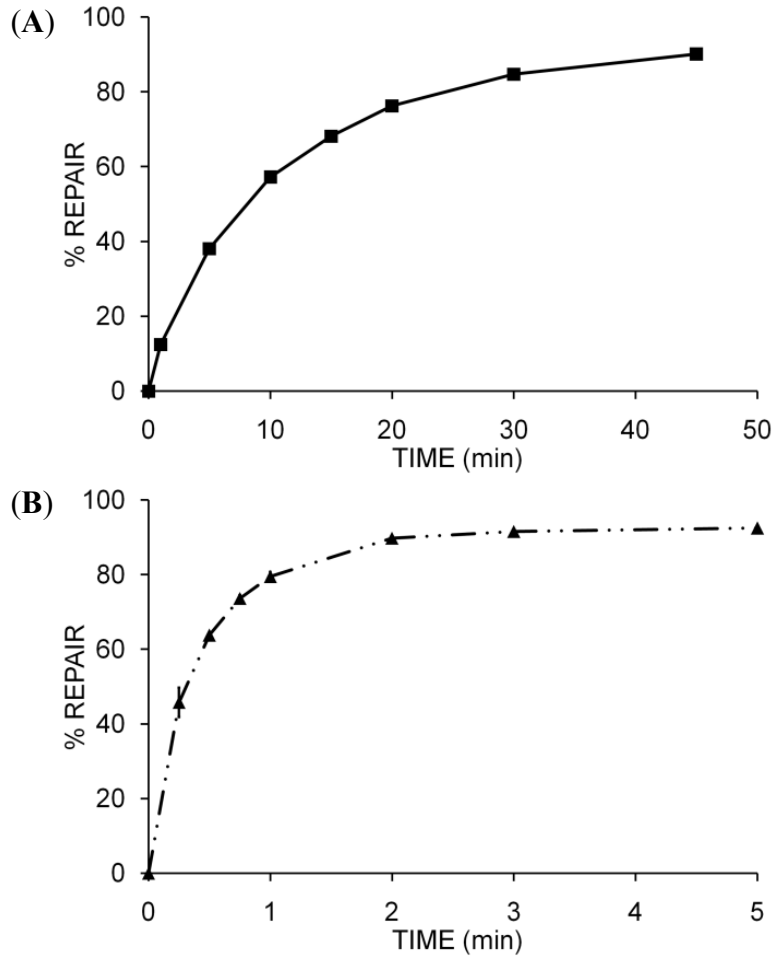


Figure 6.3: Time course repair of **U-Me** at room temperature by (**A**) hAGT and (**B**) chimera.

Repair studies of **U-Bn** (**Figure 6.4**) displayed hAGTs preference for the benzyl adduct over the methyl at the O^4 atom of dU. Once more, hAGT fell short of OGT and the chimera, where both proteins completely repaired the lesion in less than 15 sec. **U-Bn** was repaired ~100 times faster than **U-Me** by hAGT. These data correlates well with the reported repair rates of O^6-MedG and O^6-BndG, where the latter lesion is repaired ~80 faster by hAGT.(*160*) These similar results suggest that the benzyl adduct, whether at the O^6 atom of dG or O^4 atom of dU orient in similar fashions in the hAGT active site. Despite these promising data, O^4-benzyluracil would not make a potent inhibitor of

hAGT since the repair of O^6-MedG in an identical sequence and under the same conditions was achieved in less than 15 sec.(*235*) The effect of the added phenyl ring appeared to have the same consequence on the chimera where an increase in rate of repair was observed for **U-Bn** over **U-Me**.

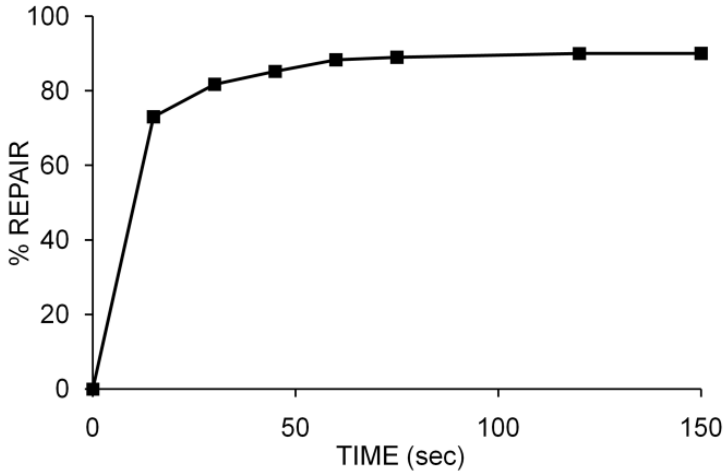


Figure 6.4: Time course repair of **U-Bn** at room temperature by hAGT

hAGT was unable to repair **U4** or **U7**, which follows the trend we observed previously for the repair of these same adducts on the O^4 atom of dT.(*235*) Unlike what was discovered for dT, the chimera was able to repair both **U4** and **U7** at reasonable rates. **U4** was fully repaired in ~2 hours (**Figure 6.5A**) representing a 60-fold decrease in activity with respect to OGT which repaired the damage in ~2 minutes (**Figure 6.5B**). The effect of the C5 methyl had a reduced influence on OGT, with respect to both the chimera and hAGT, where a 5-fold activation was observed upon the removal of the C5 appendage.

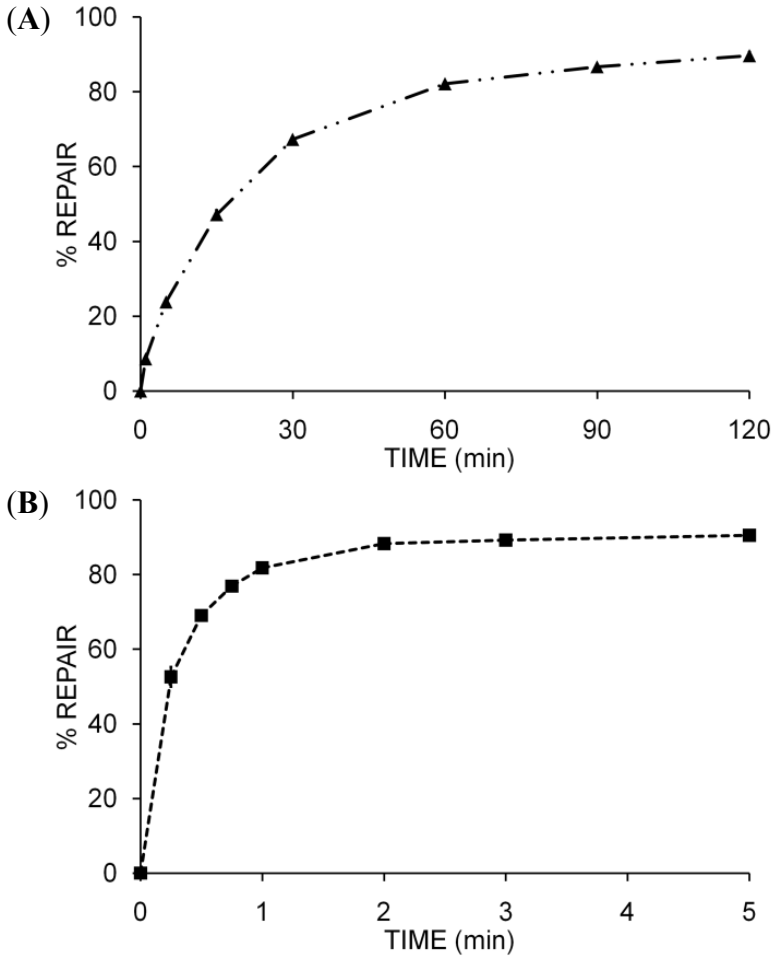


Figure 6.5: Time course repair of **U4** at room temperature by (**A**) the chimera and (**B**) OGT.

Repair of **U7** by OGT and the chimera occurred at virtually identical rates as shown in **Figure 6.6**. The time course repair of this adduct highlights large substrate discrimination patterns amongst AGTs. 1) Removing the C5 methyl of O^4-heptyl-7-ol dT not only permitted the chimera to repair the lesion but allowed the repair to occur at a rate similar to OGT, further demonstrating the more accrued effect of the C5 methyl on the chimera (and therefore hAGT) than on OGT. 2) The absence of the C5 methyl in this case increased OGTs ability to remove the adduct by 30-fold, suggesting a steric congestion in the OGT active site for **U7** as opposed to **U4**, a trend that is not observed

for the chimera. 3) Repair of the longer **U7** lesion occurred 10-fold slower by OGT than **U4**, while the chimera repaired **U7** 8 times faster than **U4**.

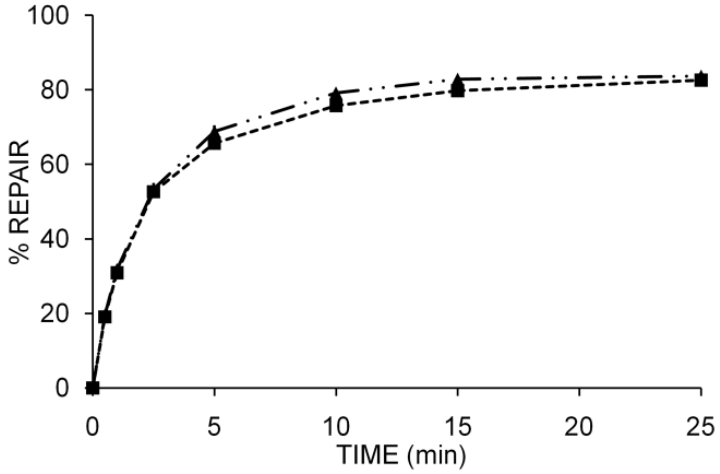


Figure 6.6: Time course repair of **U7** at room temperature by the chimera (– • • –) and OGT (- -)

The kinetic data support the notion that the active site cavities of OGT and hAGT differ considerably. By substituting a number of amino acids of the hAGT active site with the respective amino acids from OGT, the chimera was capable of removing all adducts in our substrate library. Indeed, residues 139-159 of hAGT, which are altered in the chimera, play a substantial role in the mediation of O^4-alkyl damage repair. Since part of the active site is still hAGT in nature some properties belonging to the human homologue persist in the chimera. The active site of hAGT can be viewed as a tunnel more than a pocket, allowing it, and the chimera, to process longer lesions. This was best exemplified when comparing the **U4** versus **U7** substrates. In OGT, **U7** seemed to spatially saturate the active-site, perhaps due to the more pocket-like shape active site of this protein, therefore reducing the processing rate of the protein. For the chimera, the longer alkyl chain may have allowed the terminal hydroxyl moiety to be occluded from the active site permitting repair to occur more efficiently. Since the effect of the C5 methyl of dT

hindered hAGT and the chimera to a similar degree, it can be infered that the amino acids that are substituted in the chimera (139-159) do not participate in the steric clash.

6.3.6 Binding studies

Other factors aside from the shape of the active site may play a role on substrate discrimination by AGTs. The substrate recognition amongst AGTs is of importance considering that of 3 basic amino acids that are believed to govern the hAGT-DNA ionic interaction (Lys 125, Arg 128 and Arg 135), only Arg 128 is present at a similar location in OGT.(*168*) Different interaction patterns amongst AGTs with DNA could dictate the orientation of the modified nucleotide into the active-site and contribute to the distance between the lesion's α-carbon and the reactive active site Cys residue. We conducted EMSAs with the inactive AGT variants to help determine if the various AGTs studied interact differently with our substrates to aid in the understanding of our kinetic data.

The binding data shown in **Table 6.2** reveal no correlation between repair kinetics and binding affinity or stoichiometry (all stoichiometries indicated 2 AGT per dsDNA). Indeed, OGT showed the fastest repair kinetics but interacts with the DNA with the lowest affinity with K_d values in the low μM range while the other AGTs display high nM K_d values. Perhaps binding affinity does not govern the reaction kinetics but the difference in affinities does support the concept that OGT binds DNA in a different fashion than the other homologues.

Table 6.2: K_d of AGTs binding to dsDNA containing **U**, **U-Me**, **U-Bn**, **U4** and **U7**

DNA[1]	OGT (μM)	Ada-C (μM)	hAGT (μM)	Chimera (μM)
U	3.77 ± 0.08	0.330 ± 0.014	0.381 ± 0.019	0.258 ± 0.013
U-Me	2.69 ± 0.26	0.660 ± 0.041	0.269 ± 0.010	0.149 ± 0.003
U-Bn	3.44 ± 0.22	0.603 ± 0.044	0.298 ± 0.010	0.179 ± 0.003
U4	3.72 ± 0.13	0.887 ± 0.049	0.593 ± 0.045	0.426 ± 0.004
U7	3.81 ± 0.08	0.974 ± 0.052	0.619 ± 0.021	0.414 ± 0.003

[1] 5'-GGC T**X**G ATC ACC AG
CCG AAC TAG TGG TC-5'

The chimera bound the DNA with similar affinities as hAGT but displayed different affinities than OGT. This is logical considering that the altered amino acids in the chimera are around the active site and not in the helix-turn-helix motif which makes up the AGT-DNA binding interface. The chimera's conferred ability to repair our substrates is therefore not caused by alteration in the binding interface and must therefore come from the shape of the active site that permits these substrates to fit such that the substrate and Cys side chain adopt a spatial arrangement that allow for the alkyl transfer to occur.

6.3.7 Modelling of O^4-alkyl dU into AGT active sites

To gain more insight into the interaction of the chimera with the O^4-alkyl dU adducts, relative to hAGT, molecular modelling was performed. In view of the chimeras similar binding properties to those of hAGT it is reasonable to prepare a molecular model based on the hAGT-DNA co-crystal 1T38 to model the adducts into both the hAGT and chimera active sites. Attempts to generate models of OGT based on the same template proved unsuccessful as the repair data and OGT models did not correlate due to the different binding properties of OGT and hAGT to DNA.

Table 6.3: Extracted distance between α-carbon of lesion and AGT reactive Cys 145 sulphur atom as determined by molecular modelling

Lesion	hAGT (Å)	Chimera (Å)
O^4-MedT	5.1 [235]	4.4 [253]
O^4-MedU	3.9	3.8
O^4-butyl-4-ol dU	ROTO[1]	5.3
O^4-heptyl-7-ol dU	ROTO[1]	5.0

[1] ROTO signifies Cys 145 side chain is not in correct rotamer for thiolate anion formation

In general, the modelling data agrees with our experimental results where the shorter distance between reactive groups is manifested by quicker reaction kinetics. hAGT is limited to O^4-MedU because of the rigidity of the active site producing steric clashes with the longer adducts, which was observed in our models as improper placement of the Cys 145 side chain for thiolate anion formation, see **Figure 6.7A**. Adoption of this rotamer by Cys 145 is not unreasonable considering the crystal structure of the C145S hAGT in complex with O^6-MedG has its Ser 145 in a similar rotameric state as those produced in our modelling.(*155*)

Amongst the many alterations in the chimera the N157G and S159A variations can be partially attributed as the basis for the chimera's ability to repair a vast array of lesions as opposed to hAGT. These point substitutions grant flexibility to the active site helping to orient the modified nucleoside for repair, see **Figure 6.7B**. Studies with similar chimera proteins by Dr. Anthony Pegg lead to comparable finding where the increased activity of their chimera were attributed to the flexibility of the proteins active site.(*123*)

According to the modelling, the increased rate of repair of **U-Me** over its dT counterpart by hAGT arises from the alleviation of a steric clash between the C5 group of dT and α-carbon of Ser 159 and the side chain of Arg 135. In the chimera, Arg 135 still

remains and the S159A alteration does not alleviate the former clash explaining why the chimera is affected in a similar fashion as hAGT by the C5 appendage.

U7 repair occurred more rapidly by the chimera than **U4** due to steric effects as noted in **Figure 6.7C**. The longer alkyl chain of O^4-heptyl-7-ol dU spans a distance vast enough to permit the terminal hydroxyl group to be solvent exposed. This manifested itself in a reduction in distance between the reactive moieties for O^4-heptyl-7-ol dU over O^4-butyl-4-ol dU by 0.32 Å.

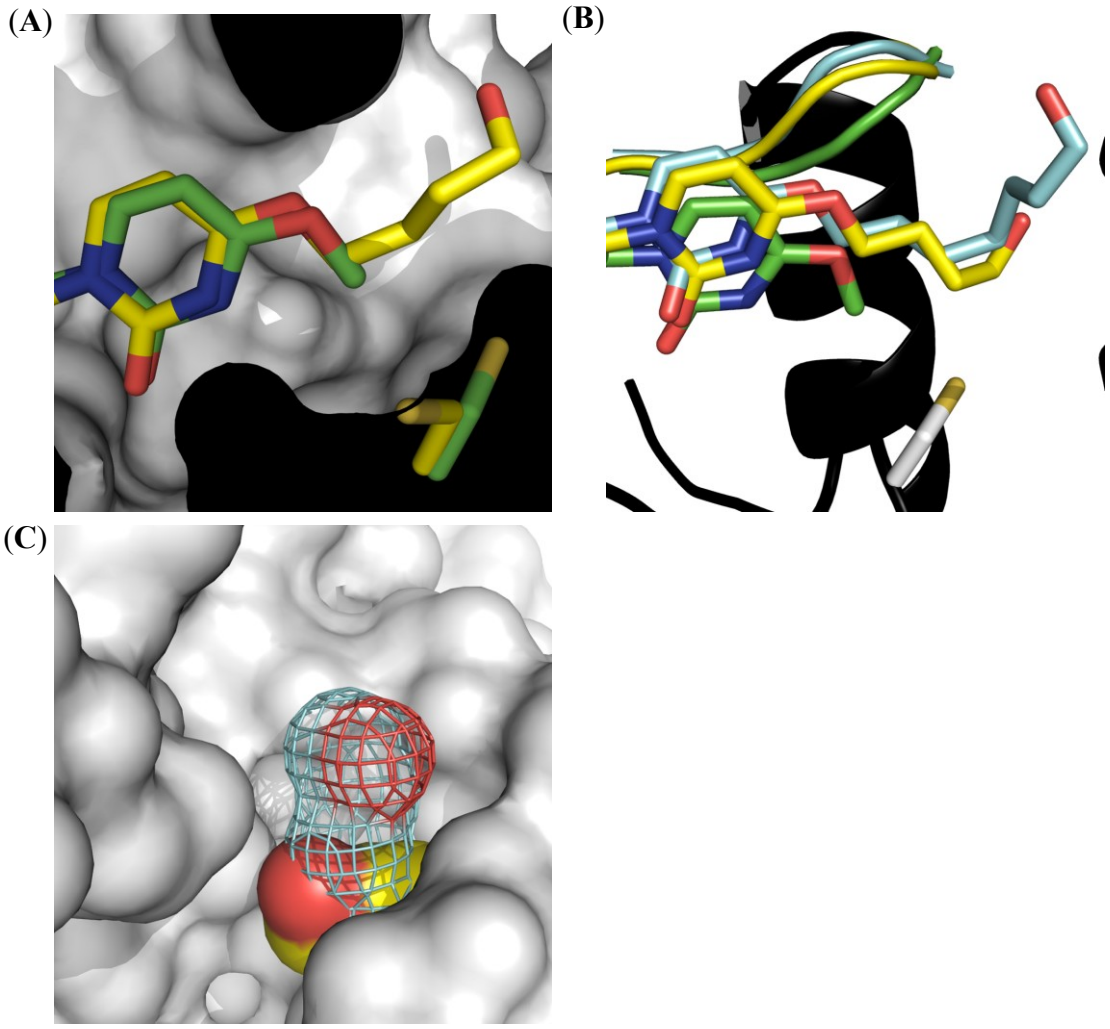


Figure 6.7: Molecular modelling results of O^4-alkyl dU adducts in hAGT and chimera active sites.

(**A**) Rotamer adoption of Cys 145 side chain of hAGT when bound to DNA containing O^4-MedU (green sticks) or O^4-butyl-4-ol dU (yellow sticks). (**B**) Mobility of the loop made up of amino acids 157- GYAGG (represented as ribbons) when the chimera is bound to DNA containing O^4-MedU (green), O^4-butyl-4-ol dU (yellow) and O^4-heptyl-7-ol dU (cyan), where Cys 145 side chain is shown as white sticks. (**C**) Active site of the chimera showing the space filling of O^4-butyl-4-ol dU (yellow) and wireframe of O^4-heptyl-7-ol dU (cyan), where the oxygen atoms are shown in red.

6.4 Conclusions

The C5 appendage of dT displays a negative effect on processing of O^4-alkyl pyrimidine lesions by AGTs. Its inhibitory effect is most observed by hAGT and the chimera protein. As the lesion at the O^4 position augments in size from a butyl-4-ol to a heptyl-7-ol, the role of the C5 methyl is greatly amplified on OGT mediated repair. These disparities between homologues signify substantial differences in active site shapes and/or sizes as well as DNA binding interfaces, the latter supported by our binding experiments. Modelling suggests substrate orientation as the determining factor for O^4-alkyl pyrimidine repair. The C5 methyl group, present in dT, clashes with Ser 159 and Arg 135 and limits the inclination of the nucleotide into the hAGT active (**Figure 6.8**). This clash also holds true for the chimera but the added mobility of the loop made up of 157- GYAGG allows the nucleotide to penetrate further into the active site and contributes to the promiscuity of this protein over its human equivalent.

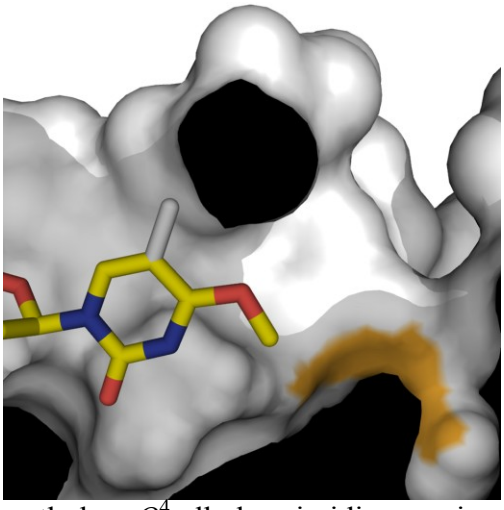


Figure 6.8: Effect of C5 methyl on O^4-alkyl pyrimidine repair.

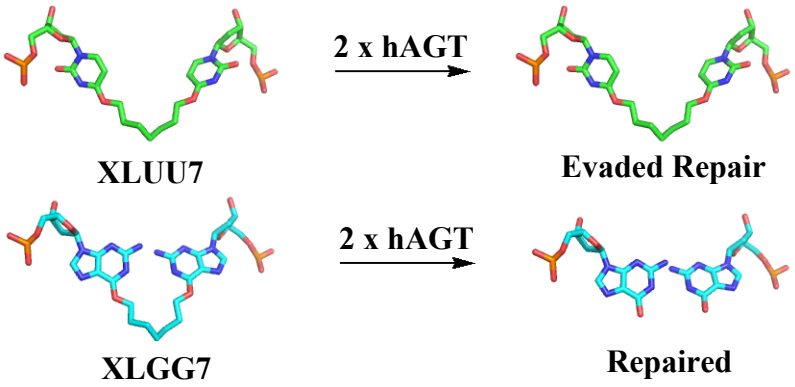
C5 methyl (white stick) inhibits inclination of the modified nucleoside (yellow sticks) into the hAGT active site, where the Cys 145 reactive sulphur is shown in orange on the surface representation.

6.5 Acknowledgments

We are grateful to Dr. Anthony. E. Pegg (Penn. State University) for the plasmid encoding the wild-type hAGT, OGT and Ada-C genes. 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. FPM is the recipient of a postgraduate fellowship from NSERC.

CHAPTER VII

Structural Basis for hAGT Selective Mediated Repair of O^6-2'-Deoxyguanosine-heptylene-O^6-2'-deoxyguanosine and not O^4-2'-Deoxyuridine-heptylene-O^4-2'-deoxyuridine Interstrand Cross-Linked Containing DNA



Graphical Abstract: XLUU7 evades hAGT repair (top), while **XLGG7** is repaired by the protein (bottom)

Denisov AY, McManus FP, O'Flaherty DK, Noronha, AM, Wilds CJ.

To be submitted to:

JACS, March 2013.

Abstract

AGT are responsible for the removal of alkylation damage from the O^6 atom of dG and the O^4 atom of dT. hAGT is capable of repairing ICL lesions between two O^6 atoms of dG when tethered by a butylene (**XLGG4**) or heptylene (**XLGG7**) chain but is inactive when the ICL involves the O^4 atoms of dT. We present the synthesis and biophysical characterization of ICL DNA containing either a dU O^4dU-butylene-O^4dU (**XLUU4**) or O^4dU-heptylene-O^4dU (**XLUU7**) ICL to study the effect of the C5 methyl of dT on AGT mediated repair of ICLs. **XLUU4** and **XLUU7** evaded repair by all AGTs analyzed. The AGT homologous interact with the ICL DNA with comparable low μM dissociation constants and 2:1 or 4:1 protein:DNA stoichiometries. The solution structures of **XLUU7** and **XLGG7** were solved and compared. The global structures are virtually identical with a RMSD of 1.22 Å. Moreover, the cross-link resides in the major groove for both DNA species. At a local level, the heptylene linkage adopts an *E* conformation about the C4-O^4 bond in **XLUU7** whereas a *Z* conformation about the C6-O^6 bond is observed for **XLGG7**. This *E* versus *Z* conformation may be the underlying cause for hAGTs substrate discrimination since substrates proficient at undergoing repair by hAGT have been shown to adopt the *Z* conformation.

7.1 Introduction

It was once believed that DNA, the source of the genetic code, was an inert molecule. This conjecture has been disproven and research now focuses on studying both the causes and consequences of modifying events. It is well accepted that many sites in DNA are susceptible to modifications, including the exocyclic atoms of the nitrogenous bases.(*12*) The sources of DNA damage are vast and can stem from both exogenous and endogenous agents.(*258*)

One class of chemicals that modify DNA are bi-functional alkylating agents. These agents are extremely cytotoxic due to the nature of the lesion they induce. Unlike their mono-adduct forming counterparts, bi-functional alkylating agents have two reactive moieties, which allow them to form cross-links between reactive residues. ICLs formed by bi-functional alkylating agents tether two complimentary strands of DNA together.(*7*) The potency and cytotoxic property of bi-functional alkylating agents are attributed to their ability to form these ICL since they inhibit many cellular processes such as DNA replication and RNA transcription. In a study of 234 DNA alkylating chemicals, bi-functional alkylating agents classify themselves as the most toxic where half of the top 20 agents were bi-functional in nature.(*40*) Moreover, 12 out of the 20 most potent anti-cancer drugs are cross-linking agents, with all of these ranking in the top 15.(*41*) Indeed, ICLs are cytotoxic, where 40 ICLs in repair deficient mammalian cell is enough to cause cell death, while a single ICL is sufficient to kill repair deficient bacteria and yeast.(*35*,*36*)

This cytotoxic nature has been exploited in chemotherapy to target cancer cells. Since both transcription and replication are hindered, the faster proliferating cancerous cells are affected to a greater level by these lesions. Due to the success of ICL forming agents many classes of bi-functional alkylating agents have emerged and find use in chemotherapy. Examples include: aziridines, nitrosoureas, nitrogen mustards and alkyl sulphonates and their derivatives.(*42*)

The chemistry of bifunctional alkylating agents is complex and many products can arise from their reaction within the cell. Specific lesions in DNA have been isolated and characterized such as the major ICL product formed when treating DNA with hepsulfam.(*75*) This product, composed of two dG crosslinked through their N7 atoms by a heptylene linkage, is unstable and leads to depurination and the imidazole ring opened formamidopyrimidine derivative over time. For this reason we synthesized stable mimics of the ICL formed by hepsulfam by linking the O^6 atoms of dG and O^4 atoms of dT by heptylene tethers to conduct biophysical and repair assays.(*146*,*214*,*235*) We have expanded on these mimics by varying the length of the linker by introducing a butylene linker within identical DNA sequences. Tetramethylene intrastrand cross-links have been shown to be produced by the agent busulfan.

Repair of ICL is complex and the mechanisms involved appear to be lesion specific.(*208*,*209*) No single repair mechanism can be attributed to ICL repair on its own but rather many mechanisms act together to achieve total removal of the damage. ICL removal is difficult since the genetic material on both strands of DNA are affected thus counteracting nature's failsafe template assisted resynthesis. Removal of the damage from one strand by an excision repair mechanism generates an intermediate that cannot be

resynthesized using the compliment strand since itself contains the remnant of the damage. NER, HR and TLS have shown activity in ICL repair.(reviewed in *259*)

AGT proteins are part of the direct repair pathway. Unlike most repair mechanisms in the cell this protein acts autonomously to achieve removal of DNA damage. The substrate range is quite limited where AGT is documented to repair O^6-alkyl dG, and in some homologues, O^4-alkyl dT lesions. O^6-alkyl dG and O^4-alkyl dT are mutagenic and if left unrepaired cause G to A and T to C transitions, respectively.(*21*,*25*) The human homologue has been extensively characterized and repairs DNA by flipping the modified nucleotide out of the DNA duplex and into its active site. This rotation positions the reactive Cys residue (Cys 145 in hAGT) in close proximity to the adduct promoting the nucleophilic attack by the Cys thiolate anion through an S_N2-type mechanism, restoring the natural nucleotide in the process.(*151*,*155*) The resulting alkylated AGT is subsequently targeted for proteolysis through the ubiquitin mediated pathway.(*104*) Substrate discrimination patterns amongst AGT homologues exist, where the *E. coli* homologues, Ada-C and OGT, are capable of removing O^4-MedT efficiently while hAGT shows virtually no repair of this adduct.(*33*,*117*) In contrast, hAGT repairs larger adducts at the O^6 atom of dG as opposed to its *E. coli* equivalents.(*127*,*128*)

Repair studies with hAGT revealed that the protein can repair both O^6dG-alkylene-O^6dG cross-links containing a tetra- or heptamethylene lesion in a mismatched, 1,2 staggered and 5'-GNC sequence context, where the O^6dG-heptylene-O^6dG cross-link is repaired more efficiently in all sequence contexts.(*146*,*214*) Ada-C and OGT lacked this function, further demonstrating disparities amongst homologues. Repair assays with O^4dT-alkylene-O^4dT cross-links containing a tetra- or heptamethylene lesion in a

mismatched sequence revealed hAGT and the *E. coli* homologues could not process the ICL.(*235*) These findings are in accordance with hAGTs poor capability to repair O^4-MedT. hAGT bound the O^4dT-alkylene-O^4dT cross-link containing DNA with reasonable affinity supporting the notion that the protein can recognize the lesion. Unfortunately, the substrate discrimination patterns observed between AGTs is not yet understood.

An AGT chimera composed of mainly hAGT residues and part of the active site of OGT has shown promise in conferring hAGT the ability to process O^4-MedT while maintaining its ability to process ICLs.(*253*) The chimera showed poor repair of larger O^4-butyl-4-ol dT and O^4-heptyl-7-ol dT adducts and was unable to repair the O^4dT-alkylene-O^4dT cross-links. Removal of the C5 methyl, such as is the case for O^4-butyl-4-ol dU and O^4-heptyl-7-ol dU allowed the chimera to remove the adducts.(See CHAPTER VI)

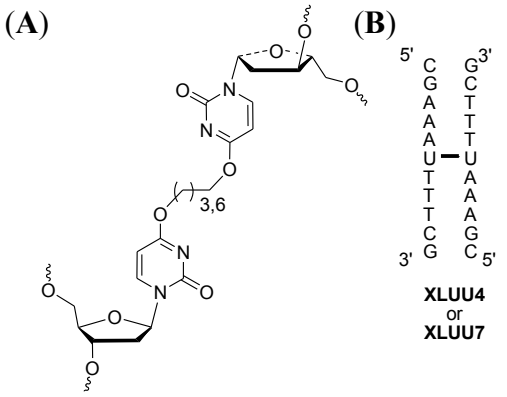


Figure 7.1: Structures of (**A**) the O^4dU-alkylene-O^4dU cross-links and (**B**) oligonucleotides **XLUU4** and **XLUU7**

We present the synthesis of O^4dU-butylene-O^4dU (**XLUU4**) and O^4dU-heptylene-O^4dU (**XLUU7**) as shown in **Figure 7.1**, their characterization and interaction with AGT to elucidate the role of the C5 methyl of the cross-linked dT of the **XLTT** ICL series on AGT mediated binding and repair. The solution structures of **XLGG7** and **XLUU7** were

solved and compared to aid in the understanding of the selective ICL repair pattern exhibited by hAGT and its chimera.

7.2 Experimental

7.2.1 Materials

See **Section 6.2.1** for materials.

7.2.2 Solution synthesis of small molecules

The solution and solid-phase synthesis of **XLGG7** have been described previously by our group.(*146*) 3'-*O*-(*t*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-C4-(1,2,4-triazol-1-yl)-2'-deoxyuridine was prepared by a three step synthesis.(*254*) 3'-*O*-(*t*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-O^4-(hydroxybutyl)-2'-deoxyuridine and 3'-*O*-(*t*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-O^4-(hydroxyheptyl)-2'-deoxyuridine were synthesized as described in CHAPTER VI.

1-{O^4-[3'-*O*-(*t*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-2'-deoxyuradinyl]}-4-{O^4-[3'-*O*-(*t*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-2'-deoxyuradinyl]}-butane (1a)

3'-*O*-(*t*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-O^4-(hydroxybutyl)-2'-deoxyuridine (0.500 g, 0.697 mmol) was dissolved in ACN (6 mL) and stirred on an ice bath. To this was added 3'-*O*-(*t*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-C4-(1,2,4-triazol-1-yl)-2'-deoxyuridine (0.582 g, 0.837 mmol) followed by the dropwise addition of DBU (0.313 µL, 2.09 mmol). The reaction allowed to proceed at 0 °C for 5 min and then allowed to warm up to room temperature. After 4 days, the solvent was evaporated in vacuo, the crude product was taken up in DCM (100 mL), the solution

washed with two portions of sodium bicarbonate (2 x 100 mL) and one portion of brine (100 mL). The organic layer was dried over sodium sulphate and concentrated to produce a yellow gum. The crude product was purified by flash column chromatography using a hexanes : ethyl acetate (2 : 3) solvent system to afford 0.790 g (84.3%) of product as a colorless foam.

R_f (SiO_2 TLC): 0.29 hexane / ethyl acetate (2 : 3). ^1H NMR (300MHz, $CDCl_3$, ppm): 8.27 (d, 2H, H6), 7.28-7.44 (m, 20H, Ar), 6.87-6.90 (m, 6H, Ar), 6.31 (dd, 2H, H1', J= 4.2 Hz), 5.60 (d, 2H, H5), 4.52 (m, 2H, H3'), 4.44 (m, 4H, $ArOCH_2$), 3.99 (m, 2H, H4'), 3.84 (s, 12H, OCH_3), 3.57 (dd, 2H, H5'), 3.36 (m, 2H, H5''), 2.55 (m, 2H, H2'), 2.25 (m, 2H, H2''), 1.88 (m, 4H, $(CH_2)_2$), 0.85 (s, 18H, $SiC(CH_3)_3$), 0.04 (s, 12H, $Si(CH_3)_2$). HRMS (ESI-MS) *m/z*: Observed 1343.6376, calculated for $C_{76}H_{95}N_4O_{14}Si_2^+$ 1343.6383 $[M+H]^+$

1- {*O*^{4}-[3'-*O*-(*t*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-2'-deoxyuradinyl]}-7-{*O*^{4}-[3'-*O*-(*t*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-2'-deoxyuradinyl]}-heptane (1b)

3'-*O*-(*t*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-*O*^{4}-(hydroxyheptyl)-2'-deoxyuridine (0.510 g, 0.672 mmol) was dissolved in ACN (6 mL) and stirred on an ice bath. To this was added 3'-*O*-(*t*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-C4-(1,2,4-triazol-1-yl)-2'-deoxyuridine (0.582 g, 0.837 mmol) followed by the dropwise addition of DBU (0.313 µL, 2.09 mmol). The reaction allowed to proceed at 0 °C for 5 min and then allowed to warm up to room temperature. After 4 days, the solvent was evaporated in vacuo, the crude product was taken up in DCM (100 mL), the solution washed with two portions of sodium bicarbonate (2 x 100 mL) and one portion of brine

(100 mL). The organic layer was dried over sodium sulphate and concentrated to produce a yellow gum. The crude product was purified by flash column chromatography using a hexanes : ethyl acetate (1 : 1) solvent system to afford 0.776 g (83.3%) of pure product as a colorless foam.

R_f (SiO_2 TLC): 0.47 hexane / ethyl acetate (2 : 3). ^1H NMR (300MHz, $CDCl_3$, ppm): 8.22 (d, 2H, H6), 7.22-7.41 (m, 20H, Ar), 6.83-6.86 (m, 6H, Ar), 6.28 (dd, 2H, H1', J= 4.0 Hz), 5.57 (d, 2H, H5), 4.48 (m, 2H, H3'), 4.35 (m, 4H, $ArOCH_2$), 3.96 (m, 2H, H4'), 3.80 (s, 12H, OCH_3), 3.54 (dd, 2H, H5'), 3.32 (dd, 2H, H5''), 2.51 (m, 2H, H2'), 2.23 (m, 2H, H2''), 1.73 (m, 4H, $(CH_2)_2$), 1.39 (m, 6H, $(CH_2)_3$), 0.81 (s, 18H, $SiC(CH_3)_3$), 0.01 (s, 12H, $Si(CH_3)_2$). HRMS (ESI-MS) *m/z*: Observed 1385.6847, calculated for $C_{79}H_{101}N_4O_{14}Si_2^+$ 1385.6853 $[M+H]^+$

1-{*O*^{4}-[5'-*O*-(4,4'-dimethoxytrityl)-2'-deoxyuradinyl]}-4-{*O*^{4}-[5'-*O*-(4,4'-dimethoxytrityl)-2'-deoxyuradinyl]}-butane (2a)

To a solution of (**1a**) (0.767 g, 0.570 mmol) in THF (7.0 mL) was added TBAF (1 M in THF) (2.57 mL, 2.57 mmol) dropwise. After 10 min, the solvent was evaporated in vacuo, the crude product taken up in DCM (50 mL), the solution washed with sodium bicarbonate (2 x 150 mL) followed by a two back extraction of the sodium bicarbonate washes with DCM (2 x 5 mL). The organic layer was dried over sodium sulphate and concentrated to produce a yellow gum. The crude product was purified by flash column chromatography using a DCM / methanol solvent system (49 : 1 → 9 : 1) to afford 0.634 g (99.6 %) of product as a colorless foam.

R_f (SiO_2 TLC): 0.24 dichloromethane (DCM) / methanol (19 : 1). ^1H NMR (500MHz, $CDCl_3$, ppm): 8.06 (d, 2H, H6), 7.19-7.37 (m, 20H, Ar), 6.80-6.82 (m, 6H,

Ar), 6.24 (dd, 2H, H1', J= 5.6 Hz), 5.60 (d, 2H, H5), 4.47 (m, 2H, H3'), 4.37 (m, 4H, $ArOCH_2$), 4.03 (m, 2H, H4'), 3.78 (s, 12H, OCH_3), 3.50 (dd, 2H, H5'), 3.41 (dd, 2H, H5''), 2.59 (m, 2H, H2'), 2.23 (m, 2H, H2''), 2.00 (m, 2H, OH), 1.83 (m, 4H, $(CH_2)_2$). HRMS (ESI-MS) *m/z*: Observed 1115.4658, calculated for $C_{64}H_{67}N_4O_{14}^+$ 1115.4654 $[M+H]^+$

1-{*O*^{4}-[5'-*O*-(4,4'-dimethoxytrityl)-2'-deoxyuradinyl]}-7-{*O*^{4}-[5'-*O*-(4,4'-dimethoxytrityl)-2'-deoxyuradinyl]}-heptane (2b)

To a solution of (**1b**) (0.575 g, 0.415 mmol) in THF (4.0 mL) was added TBAF (1 M) in THF (1.87 mL, 1.87 mmol) dropwise. After 10 min, the solvent was evaporated in vacuo, the crude product taken up in DCM (50 mL), the solution washed with sodium bicarbonate (150 mL), brine (150 mL) followed by a two back extraction of washes with DCM (2 x 5 mL). The organic layer was dried over sodium sulphate and concentrated to produce a yellow gum. The crude product was purified by flash column chromatography using a DCM / methanol solvent system (97 : 3 → 19 : 1) to afford 0.476 g (99.2 %) of product as a colorless foam.

R_f (SiO_2 TLC): 0.29 DCM / methanol (19 : 1). ^{1}H NMR (500MHz, $CDCl_3$, ppm): 8.06 (d, 2H, H6), 7.20-7.37 (m, 20H, Ar), 6.79-6.82 (m, 6H, Ar), 6.25 (dd, 2H, H1', J= 5.5 Hz), 5.60 (d, 2H, H5), 4.48 (m, 2H, H3'), 4.32 (m, 4H, $ArOCH_2$), 4.03 (m, 2H, H4'), 3.77 (s, 12H, OCH_3), 3.48 (dd, 2H, H5'), 3.40 (dd, 2H, H5''), 2.60 (m, 2H, H2'), 2.29 (d, 2H, OH), 2.25 (m, 2H, H2''), 1.71 (m, 4H, $(CH_2)_2$), 1.38 (m, 6H, $(CH_2)_3$). HRMS (ESI-MS) *m/z*: Observed 1157.5120, calculated for $C_{67}H_{73}N_4O_{14}^+$ 1157.5123 $[M+H]^+$

1- {O^4-[3′-*O*-(*β*-cyanoethyl-*N*,*N*′-diisopropyl)-5’-*O*-(4,4’-dimethoxytrityl)-2’-deoxyuradinyl]}-4-{O^4-[3′-*O*-(*β*-cyanoethyl-*N*,*N*′-diisopropyl)-5’-*O*-(4,4’-dimethoxytrityl)-2’-deoxyuradinyl]}-butane (3a)

To a solution of (**2a**) (0.614 g, 0.551 mmol) in THF (3.0 mL) was added DIPEA (288 μL, 1.65 mmol) followed by the dropwise addition of *N*,*N*-diisopropylamino cyanoethyl phosphonamidic chloride (308 μL, 1.38 mmol). After 30 min, the solvent was evaporated in vacuo, the crude product taken up in ethyl acetate (50 mL), the solution washed with two portions of sodium bicarbonate (2 x 100 mL) and once with brine (100 mL). The organic layer was dried over sodium sulphate and concentrated. The product, a colorless gum, was resuspended in 3 mL of ethyl acetate and precipitated from 1L of hexanes at -78 °C (0.757 g, 90.8 %).

R_f (SiO_2 TLC): 0.24, 0.44, 0.58 ethyl acetate. ^{31}P NMR (202.3 MHz, d_6-acetone, ppm): 148.3 and 148.2. HRMS (ESI-MS) *m/z*: Observed 1515.6805, calculated for $C_{82}H_{101}N_8O_{16}P_2^+$ 1515.6811 $[M+H]^+$

1-{O^4-[3′-*O*-(*β*-cyanoethyl-*N*,*N*′-diisopropyl)-5’-*O*-(4,4’-dimethoxytrityl)-2’-deoxyuradinyl]}-7-{O^4-[3′-*O*-(*β*-cyanoethyl-*N*,*N*′-diisopropyl)-5’-*O*-(4,4’-dimethoxytrityl)-2’-deoxyuradinyl]}-heptane (3b)

To a solution of (**2b**) (0.456 g, 0.394 mmol) in THF (3.0 mL) was added DIPEA (203 μL, 1.17 mmol) followed by the dropwise addition of *N*,*N*-diisopropylamino cyanoethyl phosphonamidic chloride (218 μL, 0.98 mmol). After 30 min, the solvent was evaporated in vacuo, the crude product taken up in ethyl acetate (50 mL), the solution washed with two portions of sodium bicarbonate (2 x 100 mL) and once with brine (100 mL). The organic layer was dried over sodium sulphate and concentrated. The product, a

colorless gum, was resuspended in 3 mL of ethyl acetate and precipitated from 1L of hexanes at -78 °C (0.488 g, 80.3 %).

R_f (SiO_2 TLC): 0.30, 0.44, 0.55 hexanes / ethyl acetate (2 : 8). ^{31}P NMR (202.3 MHz, d_6-acetone, ppm): 148.3 and 148.2. HRMS (ESI-MS) *m/z*: Observed 1557.7260, calculated for $C_{85}H_{107}N_8O_{16}P_2^+$ 1557.7280 $[M+H]^+$

7.2.3 Oligonucleotide synthesis and purification

Solid-phase synthesis of **XLUU4** and **XLUU7** was conducted on an Applied Biosystems Model 3400 synthesizer on a 1 μmole scale (**XLUU4**) or 5 x 2 μmole scale (**XLUU7**) as described in **Section 3.2.3**. The modified bis-phosphoramidites **3a** and **3b** were dissolved in anhydrous ACN at a concentration of 0.05 M. The coupling time for the modified bis-phosphoramidites **3a** and **3b** were 30 min.

The CPG was removed from the column, placed into screw cap microfuge tubes and 500 μL of a 10% DBU in *n*-propanol solution added to each μmole of material. Deprotection and cleavage of the oligonucleotide was conducted for 5 days in the dark at room temperature.

XLUU4 and **XLUU7** were purified as described in **Section 3.2.3**.

7.2.4 Oligonucleotide characterization by ESI-MS and nuclease digestion

ESI mass spectra and exonuclease analysis for oligonucleotides were obtained as described in **Section 2.2.4**.

The retention times of the eluted peaks were compared to the standard nucleotides obtained which eluted at the following times: dC (4.3 min), dU (5.6 min), dG (6.8 min),

dT (7.7 min), dA (9.0 min), O^4dU-butylene-O^4dU (17.8 min) and O^4dU-heptylene-O^4dU (26.5 min) and the ratio of nucleosides determined.

7.2.5 UV thermal denaturation studies

Thermal denaturation studies were performed and processed as described in **Section 2.2.5**.

7.2.6 CD spectroscopy

CD experiments were carried out as described as in **Section 2.2.6**.

7.2.7 Protein expression and purification

All AGT proteins except for the C146S Ada-C variant were expressed under the T5 promoter of the pQE30 vector as described in **Section 2.2.8.** The C146S Ada-C variant was expressed under the T7 promoter of the pET vector as described in **Section 6.2.7**.

7.2.8 Cross-link repair assay

Cross-link repair assays were conducted as described in **Section 2.2.10**.

7.2.9 Binding assay

Cross-link binding assays were conducted as described in **Section 3.2.11**. Binding reactions consisted of 0.5 nM radio-labelled dsDNA, increasing AGT (C139S OGT, 1.5-10.6 μM; C146S Ada-C, 0.3-1.2 μM; C145S hAGT; 0.5-1.8 μM or C145S chimera, 1.5-10.6 μM) in a total solution volume of 20 μL composed of 10 mM Tris HCl (pH 7.6), 100 mM NaCl, 1 mM DTT, 10 μg mL^{-1} BSA and 2.5% glycerol.

7.2.10 NMR spectroscopy of ICL DNA

The DNA samples were dissolved in 0.3 (**XLUU7**) or 0.4 mL (**XLGG7**) of 100% D_2O or 9:1 H_2O/D_2O buffer containing 15 mM sodium phosphate (pH=7.5) and 100 mM NaCl to a final strand concentration of 0.5 (**XLUU7**) or 1.0 mM (**XLGG7**).

The proton NMR spectra were recorded on Varian Unity Inova 500 and 800 MHz spectrometers equipped with cryoprobes. Duplex melting temperatures of approximately 50 °C for **XLUU7** and 60 °C for **XLGG7** were confirmed by recording proton 1D NMR spectra at different temperatures. NOESY (nuclear Overhauser effect spectroscopy) experiments were performed in D_2O at 25 °C using mixing times t_m of 70, 250 and 400 ms. All spectra were processed by NMRPipe and the volumes of crosspeaks of NOESY spectra were calculated using NMRView.(*260*,*261*) NOESY experiments for imino-protons in 9:1 H_2O/D_2O were performed at 0 °C with a mixing time of 200 ms. DQF-COSY (double quantum filtered correlation spectroscopy) spectra were collected without phosphorus decoupling. Proton TOCSY MLEV-16 (total correlated spectroscopy using Malcolm Levitt's composite-pulse decoupling sequence) experiments were performed with a mixing time of 76 ms. Phosphorus spectra and H,P-correlation HSQC (heteronuclear single quantum correlation) spectra (with J_{PH} = 15 Hz) were recorded on Varian VNMRS (^{31}P NMR 202.3 MHz and proton NMR 500 MHz). Proton chemical shifts were measured relative to internal DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid), and phosphorus resonances were indirectly referenced to 85% H_3PO_4.(*262*) Correlation times (τ_c=3-3.5 ns) were determined using the *H5-H6* NOESY cross-peaks of dC assuming isotropic motion.(*263*)

7.2.11 Structural modelling

The starting coordinates were generated from PDB 1N4B.(*241*) Crystallography & NMR System 1.2 software with nucleic acid all-hydrogen force field was used for molecular modelling.(*222*) The number of NOE-distance (nuclear Overhauser effect), torsion angle and other constraints are reported in the results section in **Table 7.2**. A 2-fold symmetry, noncrystallographic symmetry restraints were imposed on all heavy atoms. The gentle refinement was accomplished by molecular dynamics simulation (15 ps at 2000 K with slow-cooling annealing to 0 K) including a cartesian cooling stage. The final ensemble of the 10 best structures were collected, and the atomic coordinates of average and energy minimized structures have been deposited in the Protein Data Bank under ID code 2LZV for **XLUU7** and 2LZW for **XLGG7**.

Distance restraints were derived from NOESY spectra at different mixing times by crosspeak volume integration, using the d^{-6} distance relationship and average crosspeak volume values for *H5-H6* in dC and dU (d = 2.45 Å) and *Me-H6* in dT (d = 2.70 Å) as calibration. The distance constraints were given with a 10% lower and 15% upper limits. Sugar puckering was determined from vicinal J-couplings. The five sugar torsion angles were constrained corresponding to the range of *S*-type conformations (*C1'-exo, C2'-endo, C3'-exo*), which were determined from the strong *H1'-H2'* and very weak *H2''-H3'* crosspeaks in the DQF-COSY spectra. The β torsion angles were constrained using the information from the *H5'/H5''-P* and *H4'-P* crosspeaks in the H,P-HSQC spectra. The β angles were found to be in the *trans* conformation (180 ± 30 °) as determined by the relative low intensity of the *H5'/H5''-P* crosspeaks as well as strong ^{4}JH4'-P W-pathway couplings. The γ angles were constrained (60 ± 20 °) using the NOE *H4'-H1'* crosspeak

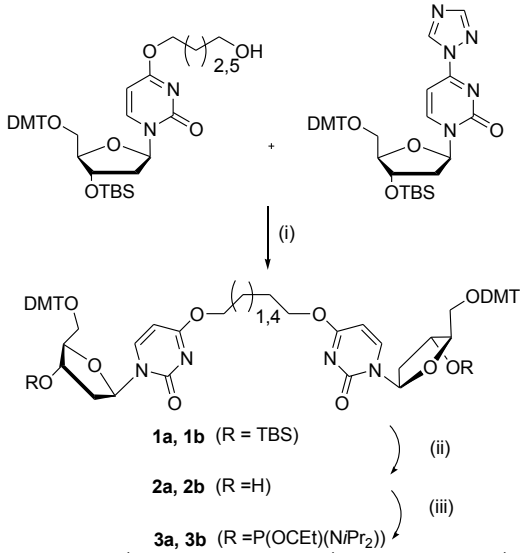
linewidths (9-11 Hz). Due to the larger values of the NOE *H4'* crosspeak linewidths for U6 in **XLUU7** and T7 in **XLGG7** residue (~14 Hz) and absence of ^{4}JH4'-P W-pathway coupling crosspeak (see Appendix VII), the γ angles for U6 in **XLUU7** and T7 in **XLGG7** were constrained to be out of the 60 ± 40 ° range.(*264*) Phosphorus chemical shifts spanned a very narrow region that is typically found for undistorted DNA, and the ε angles were constrained based on the moderate intensity of the *H3'-P* crosspeaks, in the range of B-DNA (170 ± 50 °). Finally, the glycosidic angles (χ) were not constrained in structure calculations but were fixed indirectly by intranucleotide aromatic-sugar distances. Global helical parameters were calculated using the CURVES+ program.(*265*,*266*)

7.3 Results

7.3.1 Substrate preparation

The structures of **XLUU4** and **XLUU7** are shown in **Figure 7.1**. The O^4dU-butylene-O^4dU (**3a**) and O^4dU-heptylene-O^4dU (**3b**) bis-amidites were synthesized as shown in **Scheme 7.1** starting from the mono-adduct containing nucleoside containing a tetra- or heptamethylene linker with a terminal alcohol which was prepared as previously reported by reacting the appropriate di-alcohol with 3'-*O*-(*t*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-C4-(1,2,4-triazol-1-yl)-2'-deoxyuridine (see CHAPTER VI). The protected monomers was reacted with 3'-*O*-(*t*-butyldimethylsilyl)-5'-*O*-(4,4'-dimethoxytrityl)-C4-(1,2,4-triazol-1-yl)-2'-deoxyuridine to achieve the protected dimers **1a** and **1b**. Desilylation of the dimers followed by phosphitylation produced the cross-linked bis-amidites **3a** and **3b**, required for solid phase synthesis. ^{31}P NMR signals at

148.35 and 148.22 ppm for **3a** and 148.35 and 148.21 ppm for **3b** confirmed the presence of the phosphoramidite groups.



Scheme 7.1: Synthesis of O^4dU-butylene-O^4dU and O^4dU-heptylene-O^4dU bis-phosphoramidites

Reagents and conditions: (i) DBU, ACN. (ii) TBAF (1M in THF), THF. (iii) *N*,*N*-diisopropylaminocyanoethylphosphonamidic chloride, DIPEA, THF.

7.3.2 Oligonucleotide characterization by ESI-MS and nuclease digestion

Mass spectrometry analysis of **XLUU4** and **XLUU7** confirmed the presence of the cross-links and absence of other modifications as the observed and calculated masses were in agreement.

The nucleoside compositions of **XLUU4**, **XLUU7** and the control sequence **U**, were confirmed by exonuclease digestion. All elution chromatograms showed the four standard nucleosides while **XLUU4** and **XLUU7** displayed an additional peak, each eluting later than the standards corresponding to the cross-linked nucleosides. The single stranded control sequence **U** diplayed an additional peak that eluted between dC and dG,

corresponding to dU. The extracted nucleotide ratios are in correlation with the theoretical compositions. Moreover, the elution chromatograms of **XLUU4** and **XLUU7** show no species that elutes at the same time as dU (from control oligonucleotide **U**) indicating that the cross-links were not compromised during solid-phase synthesis and oligonucleotide deprotection.

7.3.3 UV thermal denaturation studies

The effect of the cross-link in **XLUU4** and **XLUU7** on duplex stability was analyzed by thermal denaturation studies. Both modified duplexes displayed sigmoidal denaturation curves comparable to the control duplex as observed in **Figure 7.2**. The presence of the ICL stabilized the DNA duplex with the 4 methylene linker having a greater stabilizing effect than the 7 methylene linker. The melting temperatures for **XLUU4**, **XLUU7** and the control (U-A pair instead of U-U cross-link) were 61, 47 and 41 °C, respectively. These values follow the same trend as our previously prepared **XLTT** series, where **XLTT4**, **XLTT7** and the control (T-A pair instead of the T-T cross-link) displayed melting temperatures of 66, 50 and 41 °C, respectively.(*235*)

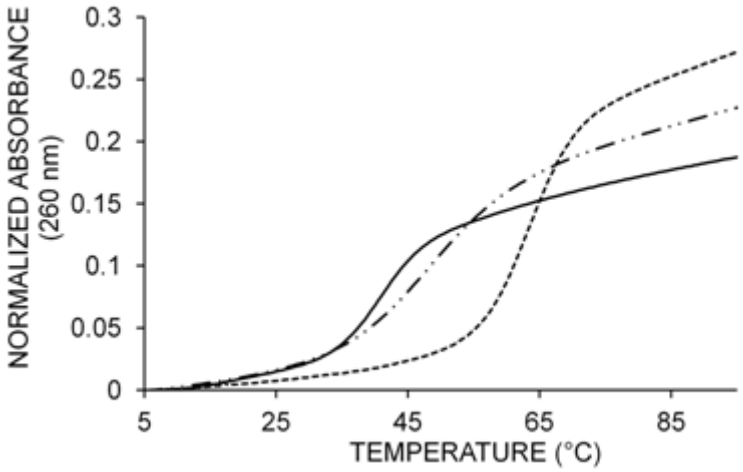


Figure 7.2: T_m curves of **XLUU4**, **XLUU7** and control DNA duplex containing **U**. Control DNA (——), **XLUU4** (……) and **XLUU7** (– • • –).

7.3.4 CD spectroscopy

The overall secondary structures of the control DNA, **XLUU4** and **XLUU7** appeared virtually identical demonstrating that the alkylene linker had no substantial effect on the DNA conformation. All DNA showed characteristic B-form DNA signatures with positive signals centered around 275 nm, cross overs around 260 nm and negative signals around 250 nm, as shown in **Figure 7.3**.(*200*,*201*) Similar Far-UV CD signatures were reported by our group for **XLTT4** and **XLTT7**.(*235*)

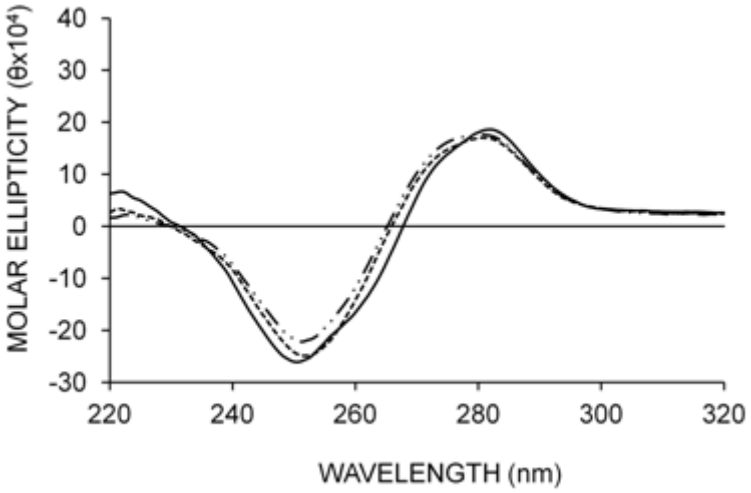


Figure 7.3: Far-UV CD spectra of **XLUU4**, **XLUU7** and control DNA duplex containing **U**. Control DNA (——), **XLUU4** (……) and **XLUU7** (– • • –).

7.3.5 Cross-link repair

Repair of **XLUU4** and **XLUU7** were performed with hAGT, OGT, Ada-C and the chimeric hAGT/OGT protein. The positive control, **XLGG7** incubated with hAGT (lanes 2 and 13), was used as a marker since the various species produced by this reaction are known and aid in the identification of the products formed in our assay. Sixteen hour reactions between AGT and ICL DNA lead to no detectable levels of fully repaired DNA species (bottom band) nor an AGT-DNA covalent complex (top band) representing a

repair intermediate (**Figure 7.4**). Clearly, **XLUU4** and **XLUU7** evaded repair by all AGTs tested.

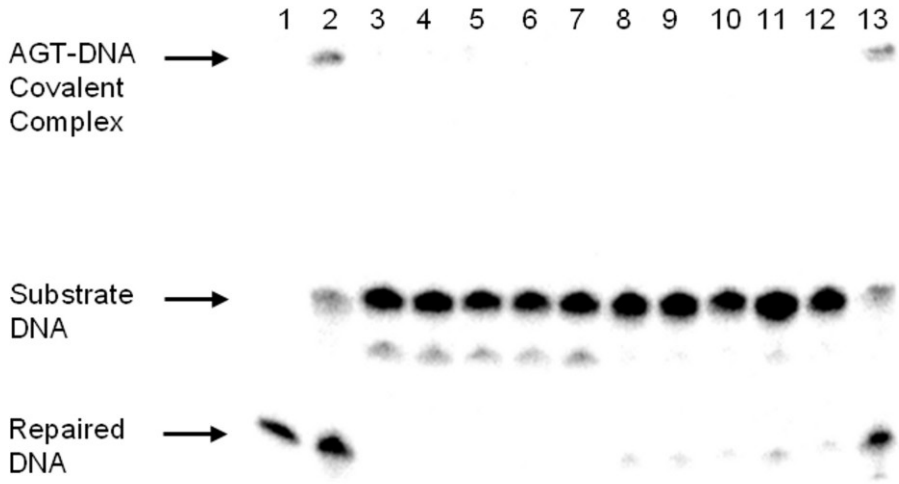


Figure 7.4: Repair gel of **XLUU4** and **XLUU7** by hAGT, OGT, Ada-C and the chimera.

Denaturing gel of the repair of 2 pmol of DNA by 60 pmol AGT. Lane 1, Control DNA **U**; lane 2, **XLGG7** + hAGT; lane 3, **XLUU7**; lane 4, **XLUU7** + hAGT; lane 5, **XLUU7** + OGT; lane 6, **XLUU7** + Ada-C; lane 7, **XLUU7** + chimera; lane 8, **XLUU4**; lane 9, **XLUU4** + hAGT; lane 10, **XLUU4** + OGT; lane 11, **XLUU4** + Ada-C; lane 12, **XLUU4** + chimera; lane 13, **XLGG7** + hAGT.

7.3.6 Binding studies

EMSA were invoked to gain an understanding on how the various AGT interact with **XLUU4** and **XLUU7**. All AGTs bound the ICL DNA in the low μM, while they showed no binding to the control DNA at these and higher concentrations (**Table 7.1**). **XLUU7** was bound with slightly higher affinity than its **XLUU4** equivalent by all AGTs representing commonality amongst AGTs. OGT and Ada-C, did not bind the ICL DNA as tightly as their mammalian homologue or the engineered chimera. Ada-C, hAGT and the chimera bound the ICL DNA with a 2:1 protein:DNA stoichiometry, while OGT showed a 4:1 stoichiometry under identical binding conditions (data not shown).

Table 7.1: K_d of AGTs binding to control and ICL DNA **XLUU4** and **XLUU7**.

DNA	OGT (μM)	Ada-C (μM)	hAGT (μM)	Chimera (μM)
Control[a]	>20	>100	>20	>20
XLUU4[b]	8.40 ± 0.18	8.73 ± 0.42	2.78 ± 0.03	4.02 ± 0.18
XLUU7[b]	7.61 ± 0.54	8.22 ± 0.60	1.74 ± 0.22	2.33 ± 0.17

[a] 5' CGAAAUTTTCG / 3' GCTTTAAAAGC

[b] 5' CGAAAXTTTCG / 3' GCTTTXAAAGC, where X are cross-linked by a butylene (**XLUU4**) or heptylene (**XLUU7**) linkers.

7.3.7 NMR Spectroscopy of ICL DNA

The assignment of the oligonucleotide non-exchangable proton resonances in the NOESY spectra (**Figure 7.5**) was performed using the *H1'(i)-H6/H8(i)-H1'(i-1)* pathway in the standard manner employed for right-handed DNA duplexes.(*267*) The DQF-COSY, TOCSY and H,P-HSQC spectra (see Appendix VII) were also used to assign the sugar protons and phosphorus signals. The dA H2 resonances were confirmed by interstrand *H2*(A3)-*H1'*(C10), *H2*(A4)-*H1'*(T9), *H2*(A5)-*H1'*(T8) NOE crosspeaks. The proton and phosphorous chemical shifts obtained for **XLUU7** and **XLGG7** (see Appendix VII) were in agreement with those published for a DNA duplex containing a N^4dC-ethylene-N^4dC cross-link where the cross-link was harbored in an identical DNA sequence apart from the cross-linked nucleotides.(*36*) The strong intraresidue NOE crosspeak observed for *H6-H1'*(U6) in **XLUU7** corresponds to a distance of ~2.5 Å (**Figure 7.5A**). This was coupled with a relatively weak *H6-H2'*(U6) crosspeak and weak sequential *H6*(U6)-*H1'/H2'/H2''*(A5) crosspeaks indicative that U6 adopted the *syn*-conformation about the glycosidic bond in **XLUU7**.(*268*) The same phenomenon was observed for T6 of **XLTT7** (data not shown). The heptylene linker protons of both **XLUU7** and **XLGG7** displayed various crosspeaks amongst each other but produced

very few crosspeaks with the oligonucleotide protons except for a strong *H5*(U6) to the neighboring CH_2(1) group crosspeak in **XLUU7**.

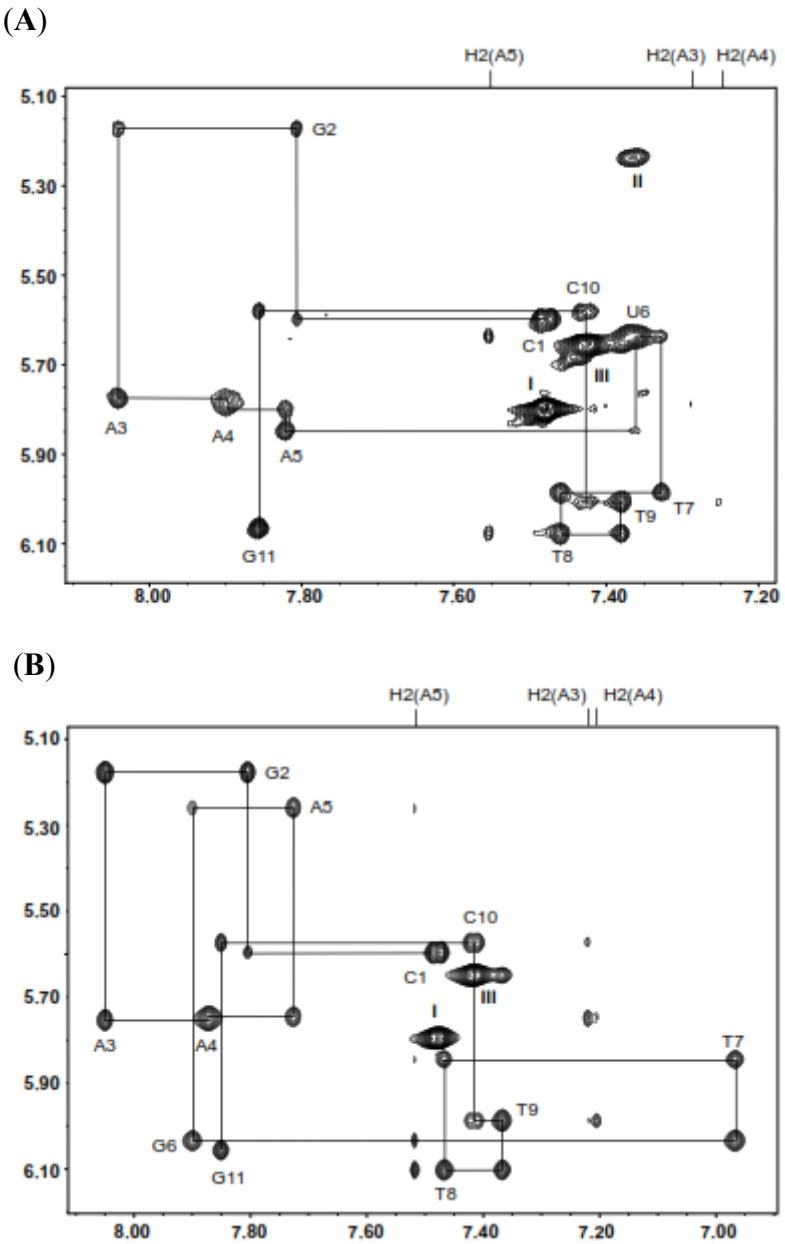


Figure 7.5: Expanded plots of NOESY spectra of (**A**) **XLUU7** and (**B**) **XLGG7**

NOESY spectra were recordedat 500 MHz at 25 °C (t_m=250 ms), the assignments of oligonucleotide protons are shown by solid line and nucleotide name with number; I-III *H5-H6*(C1,U6,C10) crosspeaks, respectively.

The most noteworthy difference amongst the ICL DNA analyzed stems from the NOE data linking the orientation of the cross-linked alkylene chain with respect to the N3 atom of U6 in **XLUU7** and the N1 atom of G6 in **XLGG7**. The adoption of the *syn*-conformation about the glycosidic bond of U6 in **XLUU7** causes the CH_2(1) (methylene attached to O^4) to adopt the *E*-configuration about the C4-O^4 bond of the cross-linked nucleotides as extrapolated from the strong NOE cross-peaks between *H5*(U6) and CH_2(1) protons in **XLUU7** (corresponding to a distance of ~2.5 Å). On the other hand, the CH_2(1) in **XLGG7** adopts the *Z*-configuration about the C6-O^6 bond as a result of the canonical *anti*-conformation adopted by G6 about the glycosidic bond as validated by the absence of cross-peaks between *H8*(G6) and CH_2(1) protons in **XLGG7**.

The presence of strong *H1'-H2'* crosspeaks and weak *H2''-H3'* crosspeaks for all residues except for the terminal G11 in the DQF-COSY spectra indicates that the most populated conformations are of the *South* type, corresponding to the *B*-DNA form (see Appendix VII).(*268*)

The imino-proton assignments were conducted based on NOESY spectra in H_2O by inter-imino crosspeaks and their crosspeaks with *H2*(A3,A4,A5) as well as with NH_2 proton signals (**Figure 7.6A**). The 1D NMR spectra of the imino protons for **XLUU7** at different temperatures are shown in **Figure 7.6B**. A strong inverse dependence between the intensity of the signal at 13.95 ppm with increasing temperature from 0 to 10 °C was detected, but the temperature dependence was greatly reduced for the signals of the T8 and G2 imino protons. This could be a result of signal overlaps between the T7 and T9 imino proton. Indeed, great fluctuations in the positions of the T7 imino protons have been reported for different ICL DNA within virtually identical sequences, ranging from

13.7-14.1 ppm.(*238*,*241*,*269*) The T7 imino proton in **XLUU7** did not show strong or medium intensity NOE crosspeaks with other protons suggesting a faster exchange with water protons (as in case of imino proton of terminal G11, see **Figure 7.6A**). This was not observed with **XLGG7** as it appeared to exist in a single conformation, where all imino protons were well defined at 0 °C (**Figure 7.6C**).

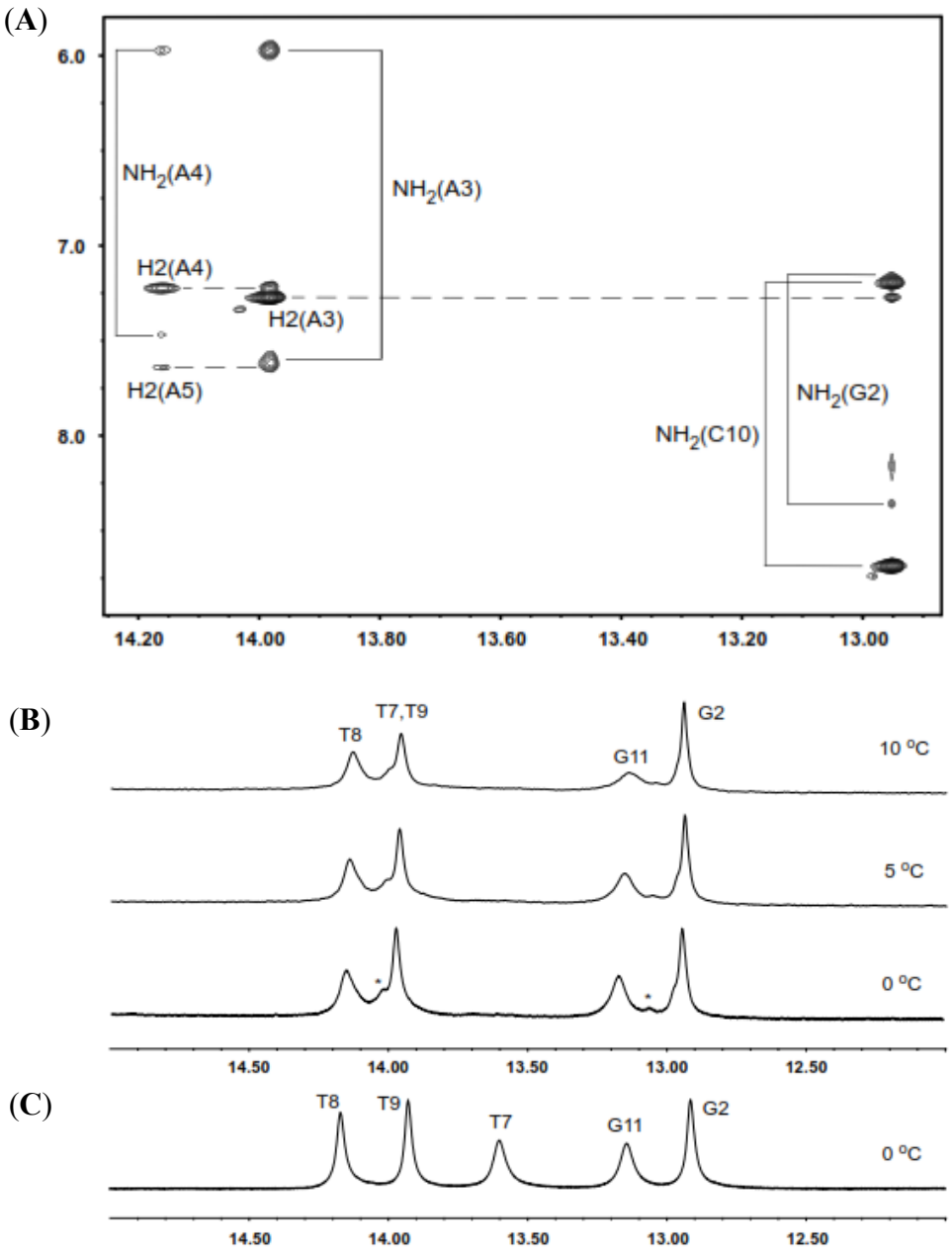


Figure 7.6: NMR spectroscopy of amine and imino protons of **XLUU7** and **XLGG7**

(**A**) Selected region of 800 MHz NOESY (t_m=200 ms) spectrum of **XLUU7** in H_2O at 0 °C, correlations between amino and imino protons are shown. (**B**) ^{1}H NMR spectra of imino-protons of **XLUU7** at low temperatures, the minor form signals are indicated with stars. (**C**) ^{1}H NMR spectrum of imino-protons of **XLGG7** at 0 °C.

7.3.8 Structure modelling and analysis

NMR-restrained molecular dynamic calculations were performed to determine the structures of **XLUU7** and **XLGG7**. The structural statistics are presented in **Table 7.2**. The superimposition of residues 1-5 of strand A and 7-11 or strand B (corresponding to half the ICL duplex) of 10 minimized structures of **XLUU7** are shown in **Figure 7.7A**. The DNA sugars adopted the *South* conformation with pseudorotation angles between 124 ° and 180 °. The cross-linked U6 residues in **XLUU7** adopted the *syn*-conformation with a χ-glycosidic angle of 77 °. The heptylene cross-links reside in the major groove for both ICL duplexes. The ICL DNA melting temperature profiles, as observed by NMR, suggest a greater flexibility for the central nucleotides in **XLUU7** with respect to **XLGG7**. This notion is supported by the larger atomic RMSD for **XLUU7**, as noted in **Table 7.2**, and also the broadened and lower intensity *H5-H6* NOESY crosspeaks for the U6 residues compared to C1 and C10 residues (**Figure 7.5A**).

The atomic RMSD for residues 1:5 of strand A and 7:11 of strand B for **XLUU7** and **XLGG7** was calculated to be 0.65 Å (**Figure 7.7B**). The atomic RMSD for the whole duplex, omitting the cross-linked nucleotides, was 1.22 Å for all heavy atoms.

Table 7.2: Structural statistics for 10 final individual structures of **XLUU7** and **XLGG7**

Duplex	**XLUU7**	**XLGG7**
Number NOE distance restraints	131	150
Intranucleotide	77	84
Internucleotide	54	66
Torsion angle restraints	87	87
Hydrogen bond restraints	34	34
NOE violation (>0.2 Å)	0	0
Torsion angle violation (>5 °)	0	0
Heavy atoms RMSD relative to average structure:		
Full duplex (Å)	0.58	0.41
Residues 1:5 of strand A and 7:11 of strand B (Å)	0.24	0.29
Average RMSD from covalent geometry:		
Bond lengths (Å)	0.0026 ± 0.0001	0.0026 ± 0.0001
Angles (°)	0.380 ± 0.005	0.411 ± 0.011
Impropers (°)	0.315 ± 0.023	0.355 ± 0.022

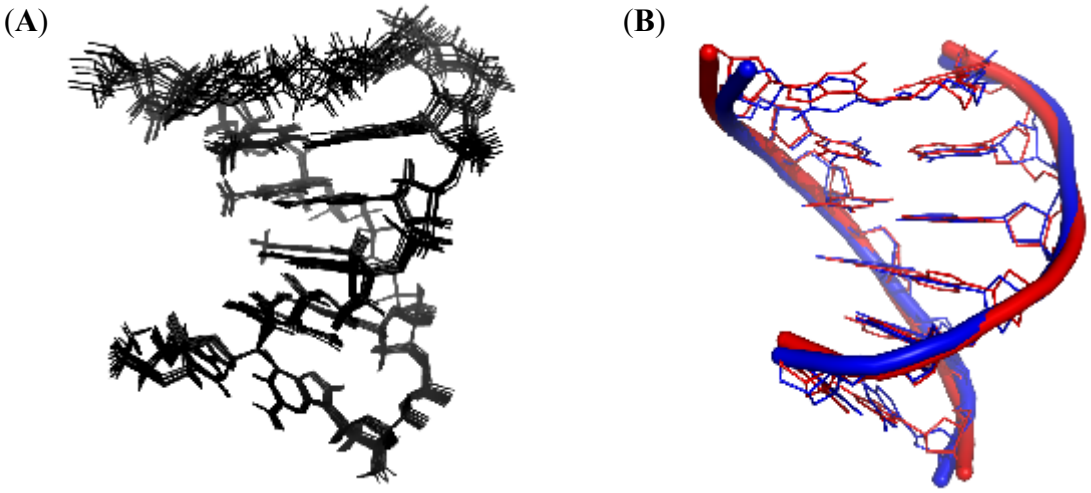


Figure 7.7: Superimposition of half of the ICL duplex by residues 1:5 of strand A and 7:11 of strand B.

(**A**) Superimposition of 10 final individual structures of **XLUU7**. (**B**) Superimposition of **XLUU7** (*in blue*) and **XLGG7** (*in red*) average minimized structures.

Upon comparing the overal minimized structures of **XLUU7** and **XLGG7**, as shown in **Figure 7.8**, it is obvious that the stacking interactions between the A5:T7 base pair and the neighboring U6 residues in **XLUU7** (**Figure 7.8C**) are poor compared to those in **XLGG7** (**Figure 7.8D**). This results in an increased mobility for the central

nucleotides of **XLUU7** where the imino-protons of the A5:T7 base pairs undergo a faster exchange with water, as eluded to earlier.

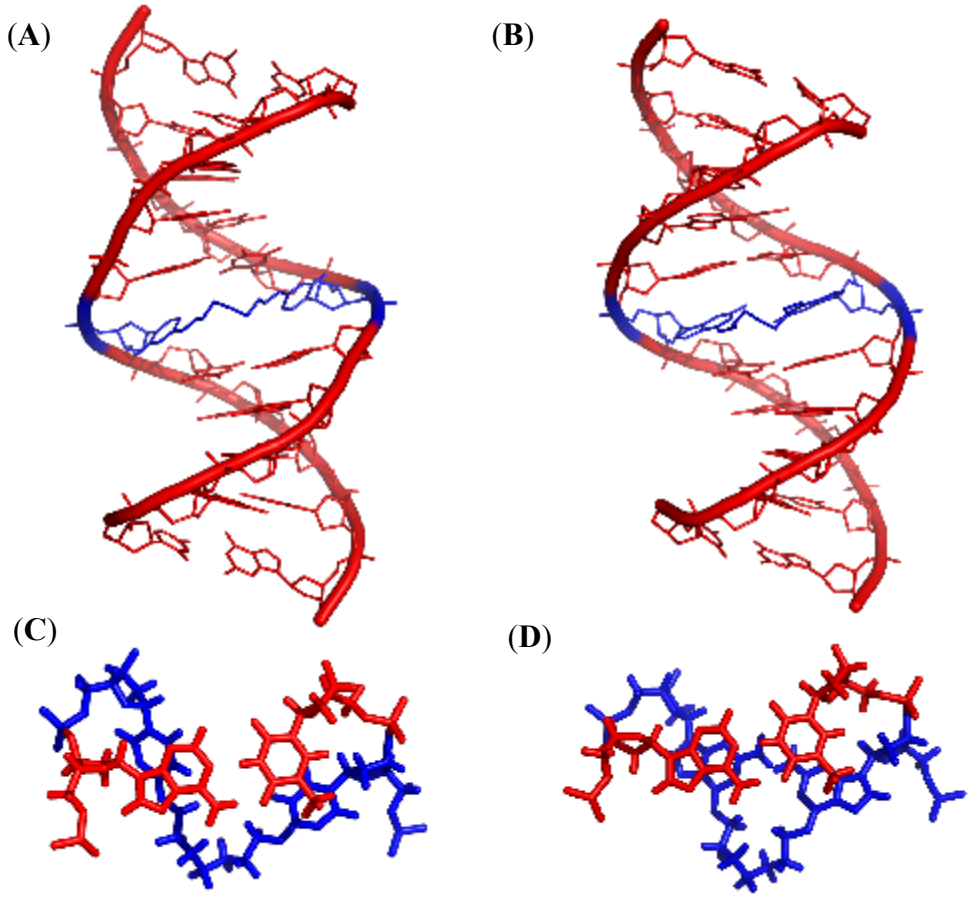


Figure 7.8: Comparison of the average minimized solution structures of **XLUU7** and **XLGG7**.

The view into the major groove of (**A**) **XLUU7** and (**B**) **XLGG7**. The stacking interaction between A5:T7 (*in red*) and cross-linked (*in blue*) (**C**) U6 and (**D**) G6 residues.

The helical and groove parameters extrapolated for **XLUU7** and **XLGG7** are shown in **Figure 7.9**. The minor groove is slightly larger around the G6 residues in **XLGG7** compared to the cross-linked U6 residues in **XLUU7**. The major groove in **XLGG7** is ~2 Å wider than in **XLUU7**. The rise between the A5:T7 base pair and

neighboring G6 residues in **XLGG7** is ~1 Å bigger than the remaining nucleotides in the DNA.

A noted increase of 10-12 Å in the major groove width around the site of the cross-links was observed when comparing the parameters of **XLUU7** and **XLGG7** to those of a previously solved structure harboring a short N^4dC-ethylene-N^4dC linker within an identical DNA sequence (except for the cross-linked nucleotides).(*241*) Rather, the widening of the major groove as observed in **XLUU7** and **XLGG7** is analogous to the structure of an N3T-butylene-N3T cross-link containing duplex.(*238*)

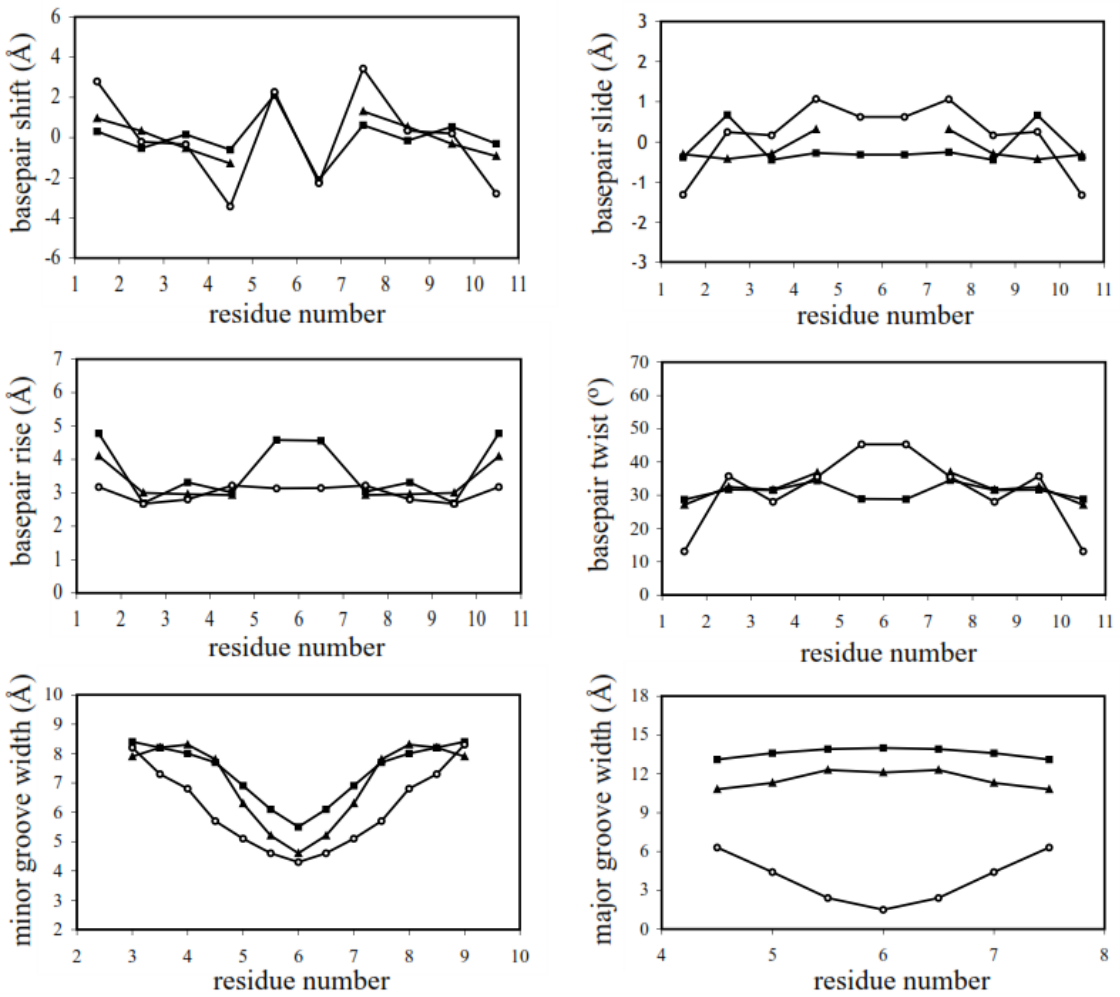


Figure 7.9: Helical parameters and groove widths of average minimized structures

XLUU7 (filled triangles), **XLGG7** (filled squares) and the previously characterized **XLCC2** (PDB 1N4B) (open circles). The helical parameters involving the central nucleotides in **XLUU7** could not be used in the comparison due to the *syn*-conformation of the U6-bases and were therefore omitted.

7.4 Discussion

The solution synthesis of the amidites required for the ICL DNA containing O^4dU-alkylene-O^4dU cross-links were prepared with high yield reactions such as those used by Swann for O^4-alkylated dT.(*221*) The solid-phase synthesis of the **XLUU** and **XLTT** ICL DNA generated little desired product. On average, 10 A_{260} (60 nmoles) of pure material

was obtained per 1 μmol column, representing only a 6 % yield. The basis for the low overall yield can be attributed in part to the method of deprotection of the oligonucleotide. Since O^4 alkylated dT and dU are susceptible to nucleophilic attack at C4, the standard ammonia/ethanol deprotection could not be utilized.(*256*,*257*) Moreover, oligonucleotide deprotection in methanol and DBU lead to the O^4 methylated products, removing our cross-links in the process as observed by denaturing PAGE and confirmed by mass spectrometry (data not shown). To circumvent this issue a mixture of 10 % DBU in *n*-propanol was adopted for deprotection. It is our belief that the larger size of the *n*-propanol alkoxide, in comparison to the methanol alkoxide, prevented breakdown of the cross-links at the expense of recovery from the CPG during the 5 day incubation.

The presence of the ICL in **XLUU4** and **XLUU7** increased the T_m of the duplex from 41 °C to 61 °C and 47 °C, respectively. This is comparable to results obtained with the **XLTT** series where the control, **XLTT4** and **XLTT7** had a T_m of 44 °C, 66 °C and 50 °C, respectively.(*235*) These ICL DNA are not as stable as the **XLGG** series where **XLGG4** and **XLGG7** had T_m of 70 °C to 60 °C, respectively.(*145*) The reduction in thermal stability of the **XLUU** series with respect to the **XLGG** may be attributed to the greater flexibility of the duplex at the cross-link site brought upon by disturbances to the base stacking interactions as observed in the NMR experiments.

Much like its **XLTT** counterparts, **XLUU4** and **XLUU7** evaded repair by all AGT tested. It is known that hAGT and the chimera are capable of repairing ICL DNA, whereas the *E. coli* homologues are not.(*146*,*214*,*235*) It has been recently documented that hAGT and the chimera have increased repair abilities of O^4-alkyl dU over the C5

methyl containing O^4-alkyl dT adducts (CHAPTER VI). This inhibitory effect of the C5 methyl on AGT mediated repair of O^4-alkylated dT was initially believed to be the cause for hAGT and chimeras inability to repair **XLTT** ICLs. Removing the C5 methyl conferred the chimera with the ability to repair larger lesions at the O^4 atom of the modified dU, where butyl-4-ol and heptyl-7-ol adducts were efficiently removed from dU by the chimera as opposed to the same lesions on dT which were left virtually untouched by the protein. Unfortunately, the **XLUU** series, lacking the C5 methyl groups, did not confer hAGT or the chimera with the ability to remove this ICL damage.

All AGT recognized the cross-links in **XLUU4** and **XLUU7** as observed by our binding results. The AGTs displayed poor binding to the control DNA. Such poor binding of similar DNA by hAGT was previously observed by our group. This was ultimately attributed to the lack of a central dG in the DNA duplex.(*235*) Indeed, a similar parallel can be made with Ada-C where it binds unmodified DNA with a K_d of 2 μm when dG is present in the central region of the duplex.(*173*) It appears that all AGT have evolved to recognize unmodified dG to aid in initial DNA binding prior to migration to sites of damage supporting the notion that the primary role of AGT is to repair O^6-alkyl dG.

Both AGT capable of ICL damage repair (hAGT and the chimera) bound **XLUU4** and **XLUU7** with the highest affinity. This higher binding affinity partially explain the source for hAGT and the chimeras ability to process ICLs over their *E. coli* counterparts. More interestingly, OGT bound the ICL DNA with a different stoichiometry than the other AGTs. Ada-C, hAGT and chimera bound the ICL DNA with a 2:1 protein:DNA ratio while OGT bound the DNA with a 4:1 ratio. OGT binds non-ICL DNA with a

similar stoichiometry as Ada-C, hAGT and the chimera, which indicates that with non-ICL DNA OGT binds DNA in a similar fashion as the other AGTs (CHAPTER VI). The increase in stoichiometry observed with OGT in the presence of the ICL could be a result of OGTs inability to migrate across the cross-link. This would cause OGT to recognize the ICL DNA as 4 distinct strands of DNA as opposed to 2, validating the increase from 2 to 4 proteins per ICL duplex. Irrespective of stoichiometry, all AGT bound the ICL DNAs in the low μM range, similar to the affinity reported with hAGT for **XLGG4** or **XLGG7**, which are repaired by the protein.(*214*) Therefore ICL DNA recognition by AGT is not the basis for the lack of activity observed.

An overlay of **XLUU7** and **XLGG7** exposes their similar overall structure. Both ICL DNA adopted the standard B-form fold as was expected from our and previous CD analysis of the DNA. **XLUU7** and **XLGG7** show similar bends in the DNA structure suggesting that preferential repair of ICL DNA by AGT is not brought upon by deviations of the DNA structure from its canonical form, such as is stipulated for other DNA binding proteins including minor groove binders.(*270*,*271*) As was previously modelled, the cross-link resides in the major groove for **XLGG7**.(*145*) The same pattern was observed for **XLUU7**. This entails that the location of the cross-link in **XLUU7** is not responsible for DNA repair evasion by this potential substrate.

Some substantial differences in structure, namely at the site of cross-link were observed between structures. These dissimilarities have the potential to play a role in substrate recognition and repair by AGT homologues. The most important finding is the adoption of the orientation of the cross-linked nucleotides about their glycosidic bonds. The modified heterocycles in **XLUU7** adopted the *syn*-configuration about the glycosidic

bond, which caused the cross-link to adopt the *E*-configuration about the C4-O^4 bond. The modified heterocycles in **XLGG7** adopted the usual *anti*-configuration about the glycosidic bond, placing the linker in a *Z*-configuration about the C6-O^6 bond. In view that hAGT has been crystallized repeatedly with mono-alkylated dG in the *Z*-configuration about the C6-O^6, it appears that the orientation of the linker may play a role in substrate repair by the protein.(*155*) Similar crystallographic data stemming from hAGT co-crystallized with DNA harboring a modified dC displayed the adduct in a *Z*-configuration about the C4-O^4 bond, supporting the notion that purine and pyrimidine orientations are conserved when present in the hAGT active site prior to repair.(*151*)

XLGG7 showed a slight increase in minor groove width (~1.5 Å) relative to **XLUU7**. Since hAGT is known to bind in the minor groove of DNA, this added space may be required for the protein to interact correctly with the substrate to place the adduct in the proper orientation for repair to occur.(*155*)

The greater level of flexibility observed in the core of **XLUU7** could aid in evading repair by AGT. Unlike nucleotide excision repair based DNA damage reversal, AGT mediated repair is not based on the flexibilty/bendability of the DNA scaffold.(*272*) Rather, flexibility may aid in evading repair since DNA binding by hAGT is transient.(*149*,*163*,*166*,*167*) Due to this nature, a preorganized substrate, such as **XLGG7**, would undergo repair more efficiently than a disorganized molecule since the protein must be associated with the substrate long enough for the proper conformation to be adopted by the substrate to promote repair.

Our solution structures shed some light on the selectivity of hAGT for different ICL substrates. Our findings would be greatly aided by AGT-ICL DNA crystal structures.

Due to the inherint complicated nature of DNA-Protein co-crystals, the preparation of such material do not appear readily accessible. Also, the cooperativity of AGT binding to DNA could be required for ICL repair. Since hAGT has only been crystallized in its truncate form, missing part of its C-terminal, its cooperativity has been abolished.(*168*,*169*) Moreover, biomolecular NMR of such complexes would be considerably complicated to solve considering the size of the DNA-AGT complex but more importantly the 2:1 stoichiometry observed for AGT binding to these ICL DNA.

7.5 Conclusions

The solution structures of two ICL DNA (**XLUU7** and **XLGG7**) have been determined by NMR which show that the α-methylene groups of the linkers must adopt the proper *Z*-configuration about the C6-O^6 bond in purines or C4-O^4 bond in pyrimidines to enable their removal by hAGT. In the former ICL, this properly preorganized system allows hAGT and the substrate to interact in an efficient manner, placing the reactive thiolate anion of the protein in close proximity to the α-carbon of the adduct undergoing nucleophilic attack leading to ICL repair. This was not the case for the latter DNA, which adopted the *E*-configuration and hence evaded hAGT based repair.

7.6 Acknowledgments

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. FPM and DKO are the recipients of a postgraduate fellowship from NSERC. The authors are grateful to Sameer Al-Abdul-

Wahid for recording proton NMR spectra at the Quebec-Eastern Canada High Field NMR Facility.

CHAPTER VIII

Conclusions, future works and significance

8.1 General conclusion

The methodology to generate ICL DNA using a combination of solution and solid phase synthesis proved versatile and effective during the course of this research. Using solution synthesis modified nucleosides containing various protecting groups were produced to either generate mono-adducts or cross-linked DNA in the desired motif by automated DNA synthesis. The methodology is versatile enabling the ICL to be engineered in the 1,3 motif. These 1,3 ICL motif (see CHAPTER II) were attained due to the added asymmetry in the ICL duplex, which was achieved by using the Alloc protecting group which was selectively removed prior to phosphitylation creating an amidite with two DMT and a TBS protective group.

Introduction of an alkyl group at the O^6 atom of dG was performed using the Mitsunobu reaction which is selective and can be performed under mild conditions.(*188*) Alkyl groups were introduced at the O^4 atoms of dT and dU using a method described by Swann's group, similar to approaches reported using an amine to generate N^4dC-alkylene-N^4dC cross-links.(*221*,*254*)

The results presented in this thesis provides evidence that hAGT may very well play a role in ICL damage detection and removal. *In vivo* assays conducted in **Section 2.2.12** provide convincing evidence that overexpression of hAGT limits the cytotoxic effects of hepsulfam. Whether this effect is due to removal of ICLs (such as those generated in **XLGG7**), mono-adducts or reacting directly with the agent itself is still unclear. Indeed, highly reactive Cys containing proteins are known to react with alkylating agents. Such an example is the reactivity of glutathione S-transferase and

hepsulfam, where increased levels of glutathione S-transferase are correlated with a reduced toxicity by the agent.(*273*) Certain steps still need to be taken to address this concern. For example, hAGT could be expressed after cell exposure to the agent. This would allow time for the DNA lesions to be induced prior to hAGT repair. If similar correlations are observed when hAGT expression is induced post alkylation with respect to those reported by our group, then it would suggest that hAGT reacts with the DNA adducts formed by hepsulfam rather than with the agent directly. hAGT is documented to increase the toxicity of certain bifunctional alkylating agents, such as 1,2-dibromoethane.(*236*) As eluded to in the introduction, hAGT is believed to directly react with 1,2-dibromoethane to create a bromoethylated hAGT, after which the alkylated protein interacts with DNA to form a protein-DNA covalent complex containing an ethylene linkage. According to the 1,2-dibromoethane mechanism, if hAGT reacts directly with hepsulfam we should, in theory, observe a positive correlation between hAGT and toxicity levels by the agent. Further *in vivo* studies are clearly needed to understand the exact role of hAGT in hepsulfam chemistry.

Irrespective of the *in vivo* studies, the *in vitro* repair assays demonstrate hAGTs ability to eliminate O^6dG-alkylene-O^6dG ICL in the clinically relevant 1,3 motif. The phenomenon occurred with butylene and heptylene cross-links, demonstrating substrate promiscuity by the protein. hAGT is limited to ICL that involve to the O^6 atom of dG as observed by the absence of any repair of O^4dT-alkylene-O^4dT and O^4dU-alkylene-O^4dU ICL DNA, irrespective of the tether length. The mode by which hAGT can eliminate these ICL remains to be uncovered and needs to be studied. Moreover, the substrate

selectivity presented by hAGT for the **XLGG** ICL DNA series is not understood nor documented in the literature.

Repair of the corresponding mono-adducts highlighted substantial difference amongst homologues. Repair of O^4-butyl-4-ol dT and O^4-heptyl-7-ol dT was achieved by OGT and to a minimal extent, by the engineered chimera, which was created by replacing a portion of the hAGT active site (residues 139-159) with the respective amino acids in OGT. hAGT, unlike the *E. Coli* homologues, did not repair O^4-MedT with an appreciable rate. These findings follow the same correlation as observed for the ICL DNA, where hAGT shows limited capability to eliminate O^4-alkyl damage.

The poor repair of O^4-alkyl dT damage by hAGT represent a significant evolutionary oversight, considering that many organisms have evolved AGT capable of eliminating such damage. We deemed it important to understand the basis for this discriminatory pattern amongst AGTs.

The first approach involved generating covalent complexes between hAGT and O^4-alkyl dT for crystallographic purposes. O^6dG-alkylene-O^4dT containing ICL DNA were employed to create hAGT-DNA covalent complexes where Cys 145 of hAGT would be tethered to the O^4 atom of dT by the alkylene chain due to the previously observed selectivity of hAGT for cross-links involving the O^6 atom of dG. The hAGT-alkylene-O^4dT cross-linking reaction was not efficient where the reaction was virtually absent when the tether was composed of 4 methylene groups (**XLGT4**). The cross-linking reaction conducted with **XLGT7** produced a considerable amount of hAGT-heptylene-O^4dT cross-linking.

The second approach, which stemmed from molecular modelling, entailed analyzing the effect of the C5 methyl of dT on O^4-alkyl damage repair by AGT. This was achieved by comparing the rates of repair of O^4-alkyl dU versus their O^4-alkyl dT analogues by the various AGT. Removal of the C5 methyl conferred hAGT with the ability to eliminate O^4-Me dT damage with a 30-fold increase in rate of repair. A similar trend was observed with the chimera, indicating that the C5 methyl of dT clashes somewhere in the hAGT active site (shown in red in **Figure 8.1**) in a location other than the altered residues found the chimera. This adverse effect of the C5 methyl on AGT processivity was less significant in OGT. To the best of our knowledge, these are the first results that provide a rational for the differential substrate repair trends documented for AGT homologues. The C5 methyl of dT is not the only source for the observed difference in rates between OGT and hAGT since its removal did not enable hAGT to repair the adducts with similar rates as OGT.

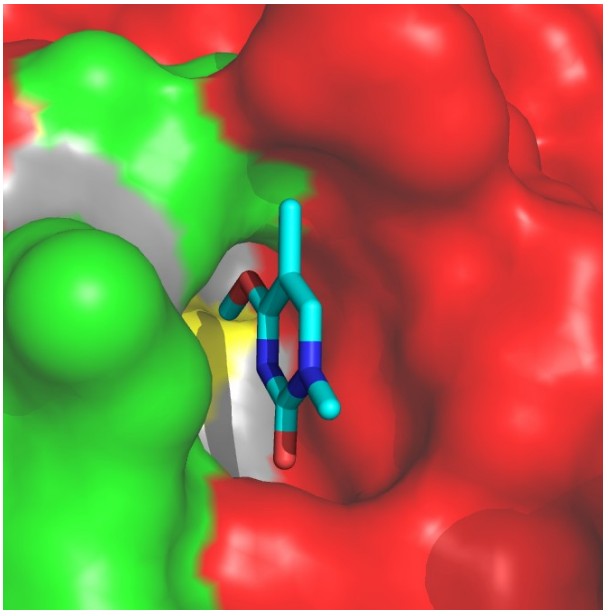


Figure 8.1: O^4-MedT in hAGT active site

Residues altered in chimera with respect to hAGT are shown in green, those conserved in white, Cys 145 in yellow and residues not altered in chimera in red.

Having had some success in determining the cause of poor O^4-alkyl mono-adduct damage repair by hAGT, efforts were subsequently focussed on determining the source of hAGTs ability to selectively repair **XLGG** and not **XLTT** nor **XLUU**. Insight into hAGT mediated repair of ICL DNA was provided by the solution structures of **XLUU7** and **XLGG7**. An overlay of the two structures (**Figure 7.7B**), solved by Dr. Denisov, highlight their virtually superimposable overall scaffold. The major differences between these two ICL DNA originate at the cross-link site. The modified heterocycles in **XLUU7** (and **XLTT7**) adopt the *syn*-configuration about the glycosidic bond, which places the cross-link in the *E*-configuration about the C4-O^4 bond. The modified heterocycles in **XLGG7** adopt the canonical *anti*-configuration about the glycosidic bond, placing the linker in a *Z*-configuration about the C6-O^6 bond. This disparity between ICL DNA duplexes is believed to be one of the causes for the way they are processed by hAGT.

Based on the hAGT repair mechanism for O^6-alkyl dG (**Scheme 1.4**), one of the events required to achieve repair of the adduct is the donation of a proton from Tyr 114 to the N3 of dG to stabilize the transition state and also aid in final product formation. An overlay of O^4-MedT and O^6-MedG as, shown in **Figure 1.19**, suggests that the O^2 atom of dT resides where the N3 atom of dG is situated when bound by hAGT. This indicates that the O^2 atom of dT should accept the proton from Tyr 114. For **XLUU7** (and **XLTT7**), the adoption of the *syn* orientation by the modified heterocycle rotates the O^2 atom of dT away from Tyr 114, inhibiting repair by hAGT.

This *syn* versus *anti* orientation adopted by the cross-linked dT and dU may be translatable to the mono-adducts. Since the C5 methyl of dT is presumed to clash in the hAGT active site, perhaps this clash is accentuated through adoption of the *syn* orientation about the glycosidic bond by the modified dT, see **Figure 8.2**. This rotation would relieve the clash produced by the C5 methyl of dT while still placing the adduct in a similar location in the hAGT active site, explaining the binding trends reported. In the case of dU, the absence of the C5 methyl would favour the *anti* orientation of the heterocycle about the glycosidic bond with respect to dT, explaining its preferential repair. The purine nature of dG would not allow the protein to bind to the *syn* orientation of the nucleoside due to the added aromatic ring that would clash considerable in the hAGT active site.

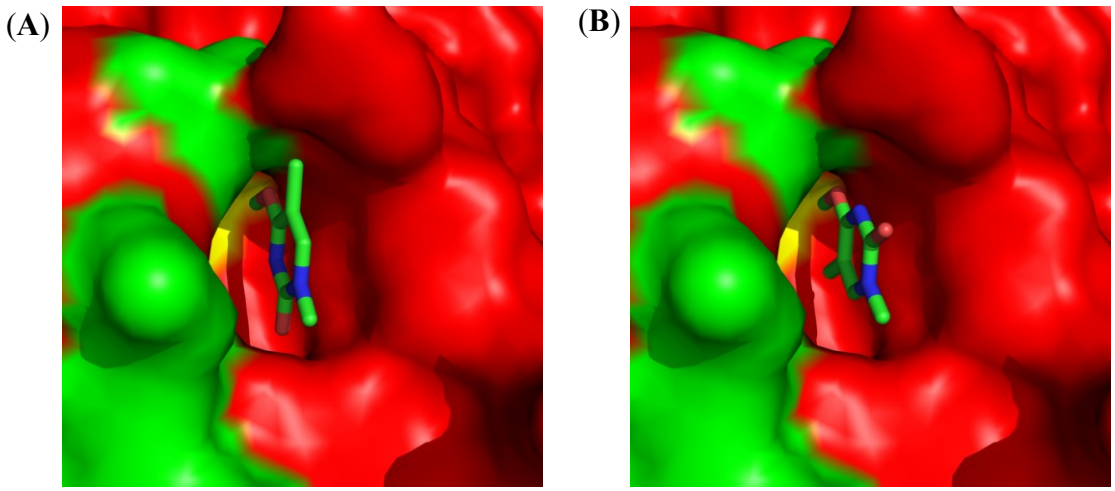


Figure 8.2: Modelling of *anti* and *syn* configuration of O^4-MedT in hAGT active site (**A**) in the classic *anti* orientation and (**B**) in the *syn* orientation.

The solution structure of **XLGG7** revealed that hydrogen bonding of the neighbouring nucleosides was unaffected by the presence of the cross-link. However there are disruptions to the base stacking interactions of the modified nucleotides which are pushed away from each other. This may account for the reduction in the T_m observed in the heptylene versus butylene containing ICL DNA.

8.2 Future works

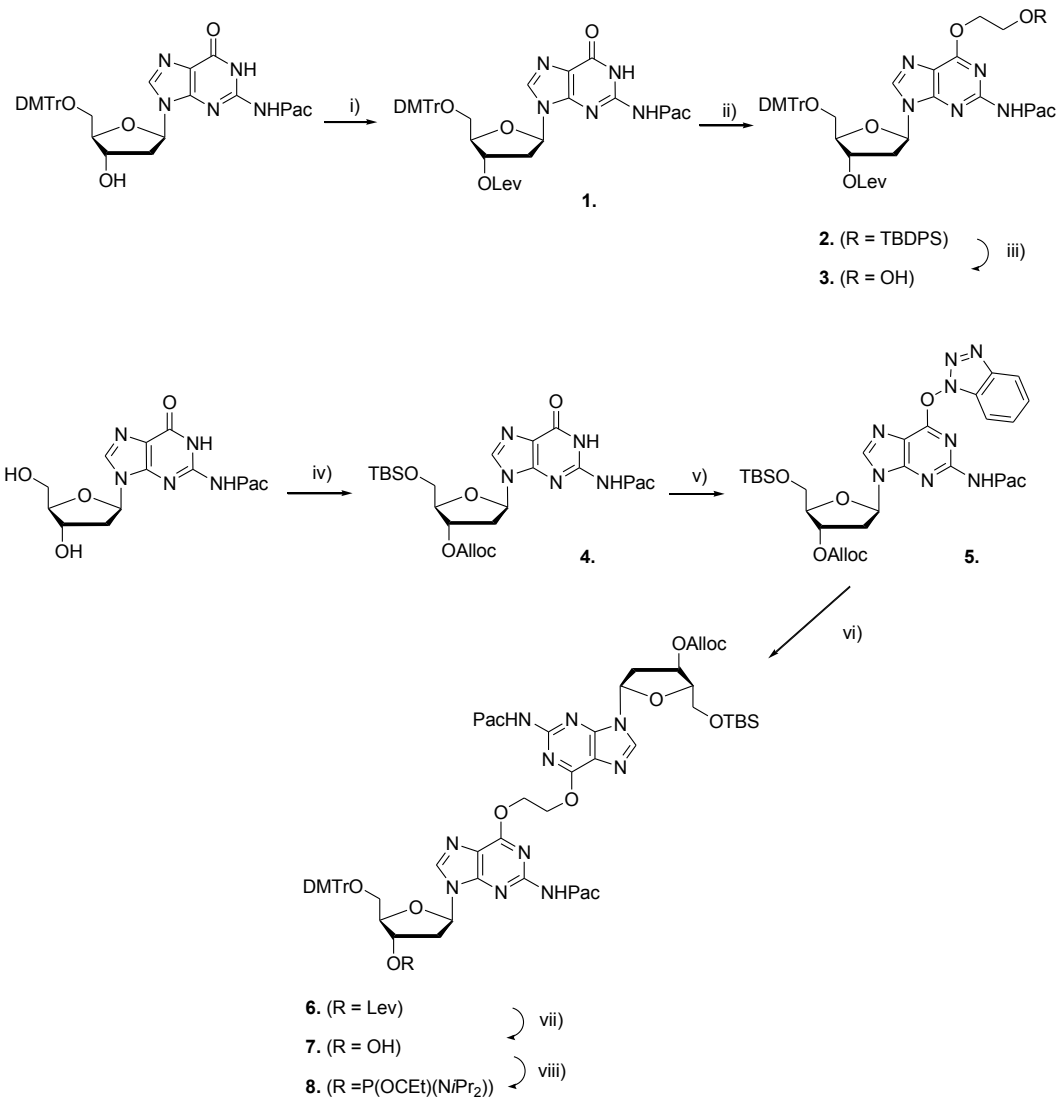
8.2.1 Preparation of O^6dG-ethylene-O^6dG ICL DNA

ICLs formed by chlorethylating agents (**Scheme 1.2**) have been of interest for decades. Recent advances have allowed the first chemical synthesis of their ICL using photocaging methodology.(*212*) Other agents such as 1,2-dichloroethane and 1,2-dibromoethane have the capability to make ethylene cross-links though they have yet to be observed between DNA strands but have been identified between proteins and

DNA.(*274*) It would therefore be of interest to note if ethylene cross-links bridging the O^6 atoms of dG are repaired by hAGT since such a cross-link could form in DNA.

Our findings indicate that reducing the cross-link length adversely affects repair by hAGT. An O^6dG-ethylene-O^6dG cross-link containing DNA (**XLGG2**) could provide further insight into the limitation of hAGT mediated ICL repair. The synthesis of the cross-linked nucleosides required for the solid phase synthesis of **XLGG2** is proposed in **Scheme 8.1**. A totally asymmetrical approach is explored so that the ICL can be incorporated into any motif within DNA. To achieve total asymmetry, the levulinyl protecting group is invoked in addition to the aforementioned Alloc, TBS and DMT groups due to its selective removal by hydrazine at the proper pH.

Previous attempts by our group to generate an O^6dG-ethylene-O^6dG cross-link have failed using the Mitsunobu reaction, where homo-dimerization of O^6-ethyl-2-ol dG is believed to occur (unpublished data). For this reason a "convertible" dG approach using benzotriazole-1-yl-oxy-tris-pyrrolidino-phosphonium hexafluorophosphate will be investigated. This methodology has been employed multiple times by the Lakshman group and has allowed the synthesis of O^6-2'-deoxyinosine-ethylene-O^6-2'-deoxyinosine cross-links.(*275*,*276*)



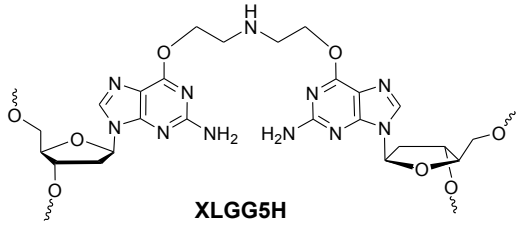
Scheme 8.1: Proposed synthesis of O^6dG-ethylene-O^6dG amidite

Reagents: (i) Levulinic acid, 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide, DMAP, dioxane. (ii) PPh_3, DIAD, 1-*t*-butyldiphenylsilylethanol, dioxane. iii) TBAF, THF. iv) a) TBS-Cl, pyridine. b) Alloc-OBT, DMAP, THF/pyridine (v) Benzotriazole-1-yl-oxy-tris-pyrrolidino-phosphonium hexafluorophosphate, DBU, THF (vi) Compound **3**, DBU, ACN. (vii) hydrazine, pyridine/acetic acid. viii) *N*,*N*-diisopropylaminocyanoethylphosphonamidic chloride, DIPEA, THF.

8.2.2 Mimic of mechlorethamine induced ICL

The tunnel model that is proposed which grants hAGT the ability to repair ICLs can be probed by introducing appendages to the ICL tether. Mimics of cross-links generated

by various nitrogen mustards could achieve this goal, see **Figure 8.3** for the structures of the mimics **XLGG5H** and **XLGG5Me**. A cross-link with a central methyl appendage (**XLGG5Me**) should cause a clash in the hAGT active site if it does indeed assume a tunnel shape. This clash should cause the ICL to evade repair while its methyl lacking control (**XLGG5H**) should be processed by the protein. A proposed synthesis of the linkers is presented in **Scheme 8.2**. Their use during solution synthesis are as described in CHAPTER II without the need for modifications since both compounds **6** and **8** are already mono-protected with TBDPS.



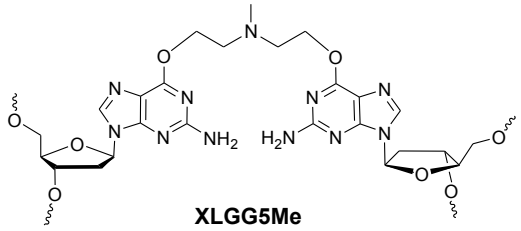
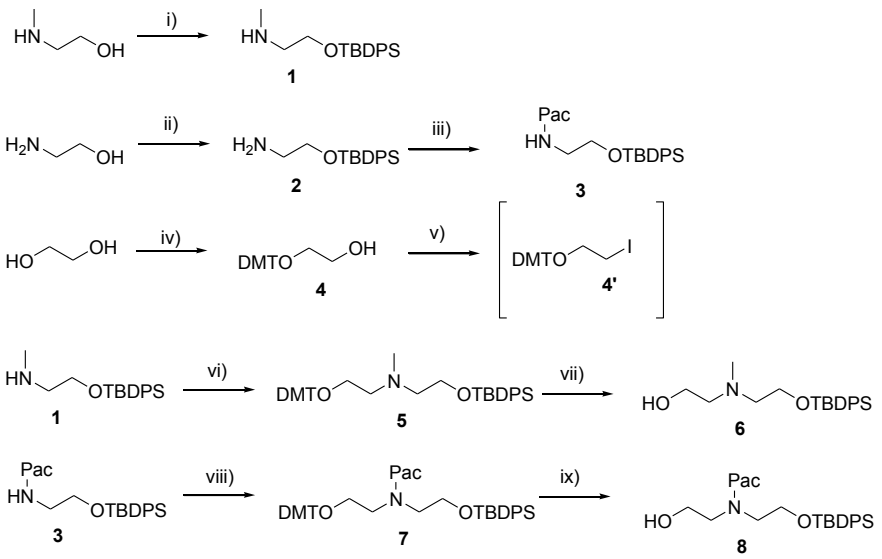


Figure 8.3: Structure of **XLGG5H** and **XLGG5Me**



Scheme 8.2: Proposed synthesis of cross-links for **XLGG5H** and **XLGG5Me**

Reagents: (i) TBDPS-Cl, imidazole, DMF. (ii) TBDPS-Cl, imidazole, DMF. (iii) PAC-Cl, TEA, THF. (iv) DMT-Cl, pyridine. v) PPh_3, imidazole, I_2, dioxane. vi) Compound **4'**, DBU, ACN. (vii) *p*-Toluenesulfonic acid, DCM/methanol. (viii) Compound **4'**, DBU, ACN. (ix) *p*-Toluenesulfonic acid, DCM/methanol.

8.2.3 Improvements on XLGT cross-linking

The short comings observed with the **XLGT** ICL series could be overcome by modifying the substrate. Eliminating one of the DNA strands from the **XLGT** ICL DNA to generate a O^6dG-alkylene-O^4-dT* (where dT* represents a dangling nucleoside) may achieve this effect, see **Figure 8.4A** for structure. Removal of one of the DNA strands from the cross-linked substrate could also serve as a substrate for OGT since it has been shown to repair O^6dG-heptylene-O^6-dG* cross-links, where dG* is a dangling nucleoside. This would promote the formation of OGT-O^4-alkyl dT* covalent complexes in addition to the hAGT-O^4-alkyl dT* covalent complexes. The proposed synthesis of the amidites to be used for solution synthesis are shown in **Scheme 8.3**. In this scenario the complex formed should have only a modified dT covalently tethered in the AGT active

site, which should be simpler to crystallize than a protein-DNA complex, see **Figure 8.4B**.

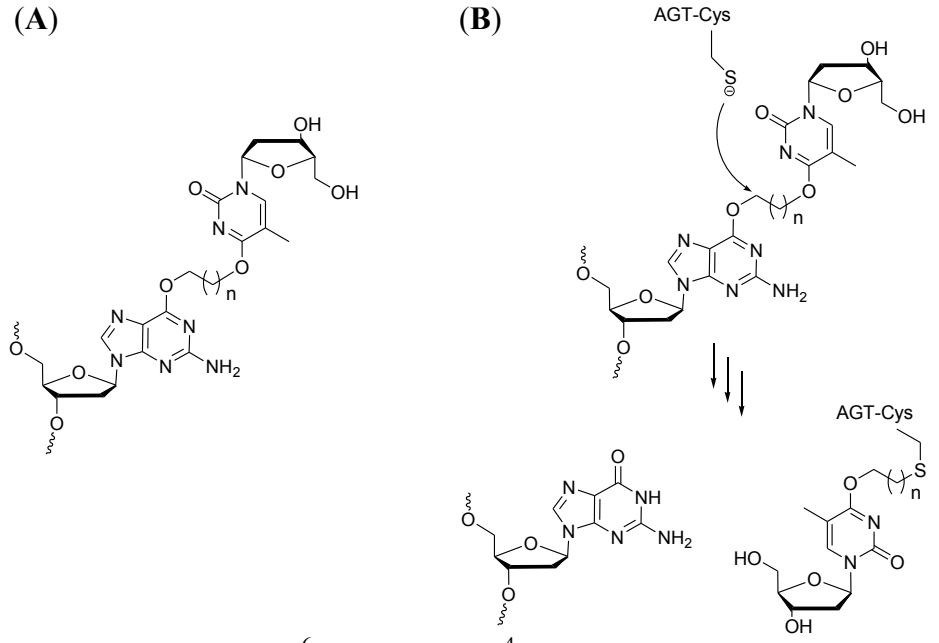
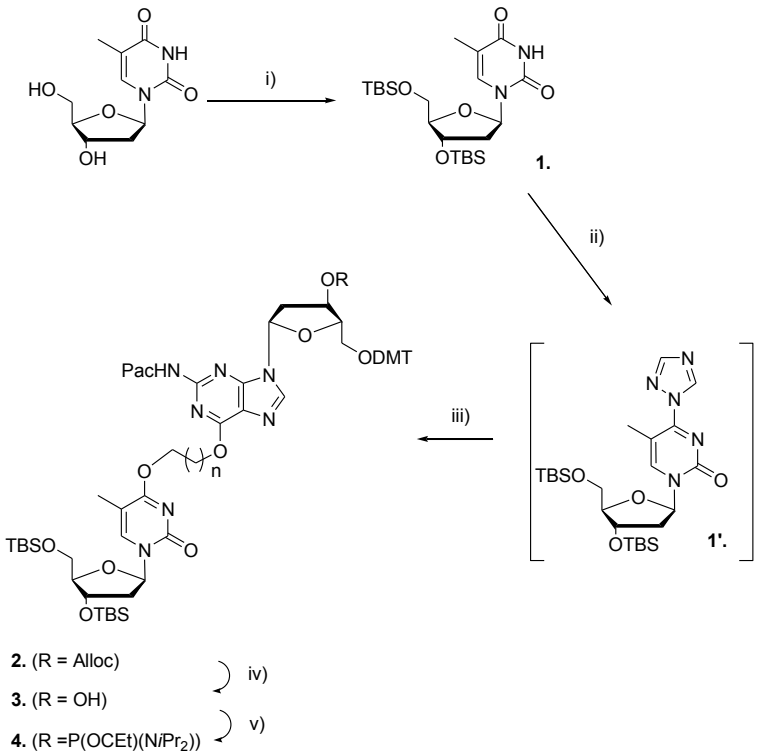


Figure 8.4: Structure of (**A**) O^6dG-alkylene-O^4-dT* cross-links and (**B**) AGT mediated repair of O^6dG-alkylene-O^4-dT* cross-links, where dT* represents a dangling nucleoside.



Scheme 8.3: Proposed synthesis of O^6dG-alkylene-O^4-dT cross-links

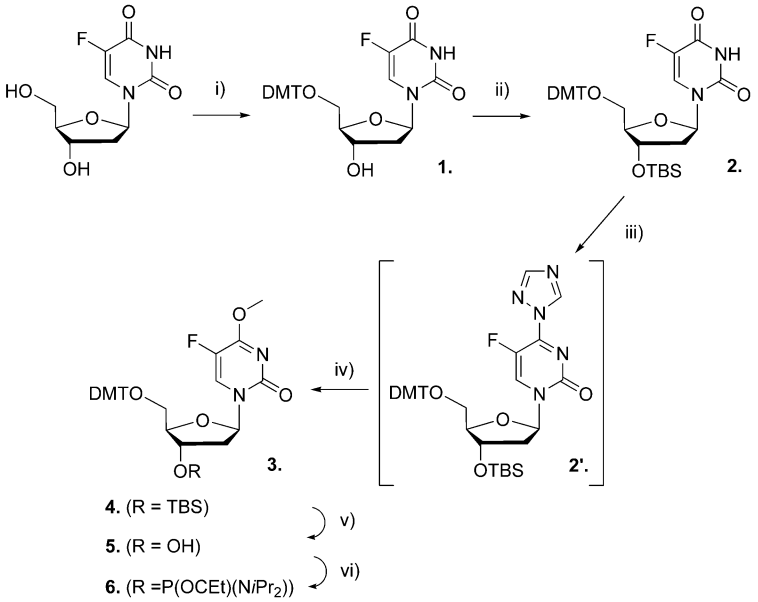
Reagents: (i) TBS-Cl, imidazole, DMF. (ii) 1,2,4-triazole, $POCl_3$, TEA, ACN. (iii) 3'-*O*-alloxycarbonyl-5'-*O*-dimethoxytrityl-N^2-phenoxyacetyl-O^6-(hydroxyalkyl)-2'-deoxyguanosine, DBU, ACN. (iv) $Pd(PPh_3)_4$, PPh_3, *n*-butylamine/formic acid, THF. (v) *N*,*N*-diisopropylaminocyanoethylphosphonamidic chloride, DIPEA, THF.

8.2.4 O^4-methyl-5-fluoro-2'-deoxyuridine

The negative impact of the C5 methyl of dT on O^4 alkyl damage repair was predicated based on steric clashes. Other properties of the C5 methyl might come into play. On a purely chemical level, the C5 methyl of dT donates electron density by induction into the heterocyclic ring, which would increase the transition state energy of the anionic intermediate formed during AGT repair (**Scheme 1.4**). To validate our sterics assumption, the 5-fluoro derivative of dU could be utilized. The preparation of the O^4-methyl-5-fluoro-dU amidite is presented in **Scheme 8.4** starting from commercially available 5-Fluoro-2′-deoxyuridine. The scheme shows the pathway to generate the

methyl adduct of 5-fluoro dU but using the appropriate alcohol in step iv) can vary the adduct with ease, as demonstrated in CHAPTER III and CHAPTER VI.

If the adverse effect of the C5 methyl of dT is based solely on sterics the rates of repair of the 5-fluoro analogues should be virtually identical to those of O^4-alkyl dU. If on the other hand electronics come into play, the electron withdrawing properties of the fluorine atom should cause the reaction rate of the 5-fluoro analogues to be greater than those of the O^4-alkyl dU series.



Scheme 8.4: Proposed synthesis of O^4-methyl-5-fluoro-dU amidite.

Reagents: i) DMT-Cl, pyridine. ii) TBS-Cl, imidazole, DMF. iii) 1,2,4-triazole, $POCl_3$, TEA, ACN. iv) methanol, DBU, ACN. v) TBAF, THF. vi) *N*,*N*-diisopropylaminocyanoethylphosphonamidic chloride, DIPEA, THF.

8.3 Significance of thesis

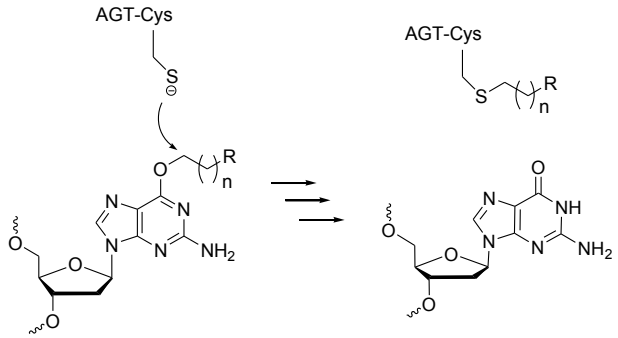
8.3.1 ICL DNA and derivatives for hAGT crystallography

The formation of covalent complexes for crystallographic purposes have been employed for various protein systems, including hAGT. Currently, there exists no

structural data, from NMR or crystallography, that explains hAGTs inability to remove O^4-alkyl dT damage proficiently. The correct ICL substrate with the proper nucleotide composition should generate hAGT-O^4-alkyl dT covalent complexes that could be of use during crystallography. These complexes, once crystallized and solved, would provide much needed insight into hAGTs selectivity. These crystal structures would also be a great starting point for rational design based engineering of hAGT to enhance this proteins ability to remove O^4-alkyl dT damage.

The advantage of these ICL substrates for use in the formation of covalent complexes is the amount of control one can have over the nature of the complex formed. Indeed, the length of the ICL, which eventually becomes the covalent cross-link between protein and DNA, can be varied with ease during solution synthesis. Such diversity cannot be introduced with poly-cyclic substrates such as N1,O^6-ethanoxanthosine generated by Noll due to ring strain.(*155*)

Moreover, the functionality of the cross-link itself can be varied, such as substituting methylene groups for phenyl rings, for example. Lastly, since hAGT selectively repairs the O^6-alkyl dG moiety of the cross-link, any other nucleobase, nucleoside, nucleotide or even molecule can be attached at the other extremity of the tether, see **Scheme 8.5**. This means that hAGT could, in theory, be crystallized with any molecule, as long as its tethered atom is not repaired. Being able to form these covalent complexes between hAGT and other molecules could grant some insight into the development of non-covalent hAGT inhibitors.



Scheme 8.5: Repair of ICL DNA derivatives by hAGT to form covalent complexes

8.3.2 hAGT variants in chemotherapy

Alkylating agents used during chemotherapy are known to alkylate various atoms of the DNA non-specifically. The O^4 atom of dT is no exception. For this reason an hAGT variant capable of eliminating O^4-alkyl dT damage, such as the chimera presented in this document, could find a home in gene therapy during cancer treatments with chemotherapeutic agents. The formation of O^4-alkyl dT is highly mutagenic and if left unrepaired in cancer cells can cause advanced cancer development.

An hAGT variant capable of O^4-alkyl dT damage removal would find best use during chemotherapeutic regimens involving bi-functional alkylators. Since the cytotoxic lesions they induce are ICLs that generally involve the N7 atom of dG, eliminating O^4-alkyl dT damage should not reduce the potency of the agent but would reduce some of the side effects.

For methylating chemotherapeutic agents such as temozolamide, whose main mode of action is through formation of O^6-MedG and cell apoptosis due to multiple rounds of futile MMR, further engineering of our chimera would show benefits. In this scenario the chimera displays half of the desired properties, but it can still eliminate O^6-MedG readily. To achieve total inversion of substrate specificity other alteration must be made to the

chimera. The chimera's substrate specificity could be altered by reshaping the active site since O^6-MedG is larger than O^4-MedT. Lys 165 and Val 148 alteration appear to be good candidates since it would remodel the active site where the exocyclic amine of dG lies, shown in blue in **Figure 8.5**. These amino acid alteration could induce a clash between the AGT active site and the N^2 atom of O^6-MedG hindering its repair. Since O^4-MedT lacks this functionality it should not affect how the AGT interacts with this lesion.

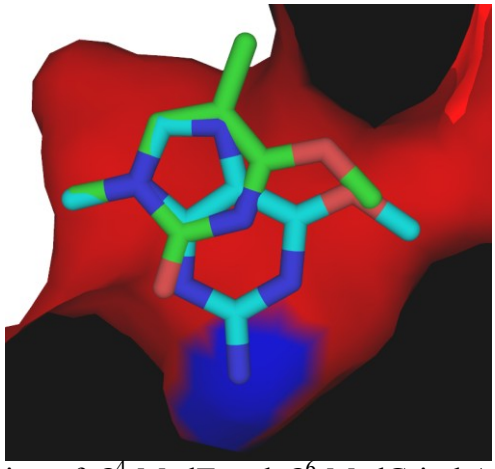


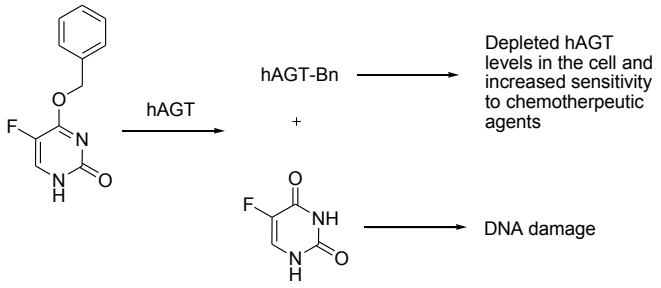
Figure 8.5: Superimposition of O^4-MedT and O^6-MedG in hAGT active site where Lys 165 and Val 148 surfaces are shown in blue.

8.3.3 Uracil derived hAGT inhibitors

The work shown in this thesis leads to the potential development of 5-fluoro-O^4-benzyl-uracil as a hAGT inhibitor with an added anti-cancer property. Since hAGT is often found at higher levels in cancer cells than in healthy cells this compound should target malignant tumours.(*277*,*278*,*279*) Based on the O^4-alkyl dU repair trends observed with hAGT it can be foreseen that 5-fluoro-O^4-benzyl-uracil should be acted upon by hAGT. The two products of this reaction in addition to chemotherapeutic agents should aid in the treatment of cancer, see **Scheme 8.6**. The first product of this reaction is alkylated hAGT, which leads to depletion of hAGT levels in the cell. These depleted

hAGT levels increase the potency of many chemotherapeutic agent. (reviewed in *132*) The novelty of this compound over classic hAGT inhibitors is the formation of 5-fluoro-uracil as the second product. 5-fluoro-uracil has been used as an anti-cancer agent for over 50 years due to its ability to inhibit thymidylate synthase. Thymidylate synthase, the protein responsible for methylating 2'-deoxyuridine monophosphate to generate thymidine monophosphate is required for proper DNA synthesis. Inhibition of this protein disturbs the general pool of nucleotides creating a biological imbalance and DNA damage due to replication mistakes. This compound's required bio activation by hAGT signifies that in tissues with low hAGT levels, such as the brain, the adverse off-target effects of 5-fluoro-uracil are alleviated.

Scheme 8.6: Products of hAGT mediated repair of O^4-benzyl-5-fluoro-uracil



References

(*1*) Mishina, Y., Duguid, E. M., & He, C. (2006). Direct reversal of DNA alkylation damage. *Chemical Reviews, 106*(2), 215-232.
(*2*) Labahn, J., Scharer, O. D., Long, A., Ezaz-Nikpay, K., Verdine, G. L., & Ellenberger, T. E. (1996). Structural basis for the excision repair of alkylation-damaged DNA. *Cell, 86*(2), 321-329.
(*3*) Wyatt, M. D., Allan, J. M., Lau, A. Y., Ellenberger, T. E., & Samson, L. D. (1999). 3-methyladenine DNA glycosylases: Structure, function, and biological importance. *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology, 21*(8), 668-676.
(*4*) Preston, B. D., Singer, B., & Loeb, L. A. (1986). Mutagenic potential of O4-methylthymine in vivo determined by an enzymatic approach to site-specific mutagenesis. *Proceedings of the National Academy of Sciences of the United States of America, 83*(22), 8501-8505.
(*5*) Loechler, E. L., Green, C. L., & Essigmann, J. M. (1984). In vivo mutagenesis by O6-methylguanine built into a unique site in a viral genome. *Proceedings of the National Academy of Sciences of the United States of America, 81*(20), 6271-6275.
(*6*) Biesele, J. J., Philips, F. S., Thiersch, J. B., Burchenal, J. H., Buckley, S. M., Stock, C. C., et al. (1950). Chromosome alteration and tumour inhibition by nitrogen mustards; the hypothesis of crosslinking alkylation. *Nature, 166*(4235), 1112-1114.
(*7*) Noll, D. M., Mason, T. M., & Miller, P. S. (2006). Formation and repair of interstrand cross-links in DNA. *Chemical Reviews, 106*(2), 277-301.
(*8*) Dronkert, M. L., & Kanaar, R. (2001). Repair of DNA interstrand cross-links. *Mutation Research, 486*(4), 217-247.
(*9*) McHugh, P. J., Spanswick, V. J., & Hartley, J. A. (2001). Repair of DNA interstrand crosslinks: Molecular mechanisms and clinical relevance. *The Lancet Oncology, 2*(8), 483-490.
(*10*) Hemminki, K. (1983). Nucleic acid adducts of chemical carcinogens and mutagens. Archives of *Toxicology, 52*(4), 249-285.
(*11*) Blans, P., & Fishbein, J. C. (2004). Determinants of selectivity in alkylation of nucleosides and DNA by secondary diazonium ions: Evidence for, and consequences of, a preassociation mechanism. *Chemical Research in Toxicology, 17*(11), 1531-1539.
(*12*) Kleibl, K. (2002). Molecular mechanisms of adaptive response to alkylating agents in escherichia coli and some remarks on O(6)-methylguanine DNA-methyltransferase in other organisms. *Mutation Research, 512*(1), 67-84.
(*13*) Spratt, T. E., & Levy, D. E. (1997). Structure of the hydrogen bonding complex of O6-methylguanine with cytosine and thymine during DNA replication. *Nucleic Acids Research, 25*(16), 3354-3361.
(*14*) Rasouli-Nia, A., Sibghat-Ullah, Mirzayans, R., Paterson, M. C., & Day, R. S.,3rd. (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. *Mutation Research, 314*(2), 99-113.
(*15*) Kat, A., Thilly, W. G., Fang, W. H., Longley, M. J., Li, G. M., & Modrich, P. (1993). An alkylation-tolerant, mutator human cell line is deficient in strand-specific mismatch repair. *Proceedings of the National Academy of Sciences of the United States of America, 90*(14), 6424-6428.
(*16*) Branch, P., Aquilina, G., Bignami, M., & Karran, P. (1993). Defective mismatch binding and a mutator phenotype in cells tolerant to DNA damage. *Nature, 362*(6421), 652-654.
(*17*) Sun, G., Jin, S., & Baskaran, R. (2009). MMR/c-abl-dependent activation of ING2/p73alpha signaling regulates the cell death response to N-methyl-N'-nitro-N-nitrosoguanidine. *Experimental Cell Research, 315*(18), 3163-3175.
(*18*) Jiang, G., Li, L. T., Xin, Y., Zhang, L., Liu, Y. Q., & Zheng, J. N. (2012). Strategies to improve the killing of tumors using temozolomide: Targeting the DNA repair protein MGMT. *Current Medicinal Chemistry, 19*(23), 3886-3892.

(*19*) Mackay, W. J., Han, S., & Samson, L. D. (1994). DNA alkylation repair limits spontaneous base substitution mutations in escherichia coli. *Journal of Bacteriology, 176*(11), 3224-3230.
(*20*) Gaffney, B. L., Marky, L. A., & 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*(24), 5686-5691.
(*21*) Basu, A. K., & Essigmann, J. M. (1988). Site-specifically modified oligodeoxynucleotides as probes for the structural and biological effects of DNA-damaging agents. *Chemical Research in Toxicology, 1*(1), 1-18.
(*22*) Patel, D. J., Shapiro, L., Kozlowski, S. A., Gaffney, B. L., & 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*(5), 1036-1042.
(*23*) Patel, D. J., Shapiro, L., Kozlowski, S. A., Gaffney, B. L., & 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*(5), 1027-1036.
(*24*) Dolan, M. E., & Pegg, A. E. (1985). Extent of formation of O4-methylthymidine in calf thymus DNA methylated by N-methyl-N-nitrosourea and lack of repair of this product by rat liver O6-alkylguanine-DNA-alkyltransferase. *Carcinogenesis, 6*(11), 1611-1614.
(*25*) Kalnik, M. W., Kouchakdjian, M., Li, B. F., Swann, P. F., & Patel, D. J. (1988). Base pair mismatches and carcinogen-modified bases in DNA: An NMR study of G.T and G.O4meT pairing in dodecanucleotide duplexes. *Biochemistry, 27*(1), 108-115.
(*26*) Kalnik, M. W., Kouchakdjian, M., Li, B. F., Swann, P. F., & 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*(1), 100-108.
(*27*) Dosanjh, M. K., Singer, B., & Essigmann, J. M. (1991). Comparative mutagenesis of O6-methylguanine and O4-methylthymine in escherichia coli. *Biochemistry, 30*(28), 7027-7033.
(*28*) Pauly, G. T., Hughes, S. H., & 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*(3), 457-461.
(*29*) Pauly, G. T., & Moschel, R. C. (2001). Mutagenesis by O(6)-methyl-, O(6)-ethyl-, and O(6)-benzylguanine and O(4)-methylthymine in human cells: Effects of O(6)-alkylguanine-DNA alkyltransferase and mismatch repair. *Chemical Research in Toxicology, 14*(7), 894-900.
(*30*) Altshuler, K. B., Hodes, C. S., & 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. *Chemical Research in Toxicology, 9*(6), 980-987.
(*31*) Klein, J. C., Bleeker, M. J., Roelen, H. C., Rafferty, J. A., Margison, G. P., Brugghe, H. F., et al. (1994). Role of nucleotide excision repair in processing of O4-alkylthymines in human cells. *The Journal of Biological Chemistry, 269*(41), 25521-25528.
(*32*) Potter, P. M., Wilkinson, M. C., Fitton, J., Carr, F. J., Brennand, J., Cooper, D. P., et al. (1987). Characterisation and nucleotide sequence of ogt, the O6-alkylguanine-DNA-alkyltransferase gene of E. coli. *Nucleic Acids Research, 15*(22), 9177-9193.
(*33*) Paalman, S. R., Sung, C., & 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*(37), 11118-11124.
(*34*) Wilkinson, M. C., Potter, P. M., Cawkwell, L., Georgiadis, P., Patel, D., Swann, P. F., et al. (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, 17*(21), 8475-8484.

(*35*) Lawley, P. D., & Phillips, D. H. (1996). DNA adducts from chemotherapeutic agents. *Mutation Research, 355*(1-2), 13-40.
(*36*) Magana-Schwencke, N., Henriques, J. A., Chanet, R., & Moustacchi, E. (1982). The fate of 8-methoxypsoralen photoinduced crosslinks in nuclear and mitochondrial yeast DNA: Comparison of wild-type and repair-deficient strains. *Proceedings of the National Academy of Sciences of the United States of America, 79*(6), 1722-1726.
(*37*) Haddad, R., Colevas, A. D., Tishler, R., Busse, P., Goguen, L., Sullivan, C., et al. (2003). Docetaxel, cisplatin, and 5-fluorouracil-based induction chemotherapy in patients with locally advanced squamous cell carcinoma of the head and neck: The dana farber cancer institute experience. *Cancer, 97*(2), 412-418.
(*38*) Balcome, S., Park, S., Quirk Dorr, D. R., Hafner, L., Phillips, L., & Tretyakova, N. (2004). Adenine-containing DNA-DNA cross-links of antitumor nitrogen mustards. *Chemical Research in Toxicology, 17*(7), 950-962.
(*39*) Erickson, L. C., Bradley, M. O., Ducore, J. M., Ewig, R. A., & Kohn, K. W. (1980). DNA crosslinking and cytotoxicity in normal and transformed human cells treated with antitumor nitrosoureas. *Proceedings of the National Academy of Sciences of the United States of America, 77*(1), 467-471.
(*40*) Lohman, P. H. (1999). Qualitative and quantitative procedures for health risk assessment. *Mutation Research, 428*(1-2), 237-254.
(*41*) Vogel, E. W., Barbin, A., Nivard, M. J., Stack, H. F., Waters, M. D., & Lohman, P. H. (1998). Heritable and cancer risks of exposures to anticancer drugs: Inter-species comparisons of covalent deoxyribonucleic acid-binding agents. *Mutation Research, 400*(1-2), 509-540.
(*42*) Rajski, S. R., & Williams, R. M. (1998). DNA cross-linking agents as antitumor drugs. *Chemical Reviews, 98*(8), 2723-2796.
(*43*) Kunkel, T. A. (1999). The high cost of living. american association for cancer research special conference: Endogenous sources of mutations, fort myers, florida, USA, 11-15 november 1998. *Trends in Genetics : TIG, 15*(3), 93-94.
(*44*) Avila, M. A., Garcia-Trevijano, E. R., Martinez-Chantar, M. L., Latasa, M. U., Perez-Mato, I., Martinez-Cruz, L. A., et al. (2002). S-adenosylmethionine revisited: Its essential role in the regulation of liver function. *Alcohol (Fayetteville, N.Y.), 27*(3), 163-167.
(*45*) Polevoda, B., & Sherman, F. (2007). Methylation of proteins involved in translation. *Molecular Microbiology, 65*(3), 590-606.
(46) Rydberg, B., & Lindahl, T. (1982). Nonenzymatic methylation of DNA by the intracellular methyl group donor S-adenosyl-L-methionine is a potentially mutagenic reaction. *The EMBO Journal, 1*(2), 211-216.
(*47*) Terasaki, M., Totsuka, Y., Nishimura, K., Mukaisho, K., Chen, K. H., Hattori, T., et al. (2008). Detection of endogenous DNA adducts, O-carboxymethyl-2'-deoxyguanosine and 3-ethanesulfonic acid-2'-deoxycytidine, in the rat stomach after duodenal reflux. *Cancer Science, 99*(9), 1741-1746.
(*48*) Shuker, D. E., Howell, J. R., & Street, B. W. (1987). Formation and fate of nucleic acid and protein adducts derived from N-nitroso-bile acid conjugates. *IARC Scientific Publications, 84*(84), 187-190.
(*49*) Shuker, D. E., & Margison, G. P. (1997). Nitrosated glycine derivatives as a potential source of O6-methylguanine in DNA. *Cancer Research, 57*(3), 366-369.
(*50*) Cupid, B. C., Zeng, Z., Singh, R., & Shuker, D. E. (2004). Detection of O6-carboxymethyl-2'-deoxyguanosine in DNA following reaction of nitric oxide with glycine and in human blood DNA using a quantitative immunoslot blot assay. *Chemical Research in Toxicology, 17*(3), 294-300.
(*51*) Marnett, L. J. (2000). Oxyradicals and DNA damage. *Carcinogenesis, 21*(3), 361-370.
(*52*) Burcham, P. C. (1998). Genotoxic lipid peroxidation products: Their DNA damaging properties and role in formation of endogenous DNA adducts. *Mutagenesis, 13*(3), 287-305.
(*53*) Nair, U., Bartsch, H., & Nair, J. (2007). Lipid peroxidation-induced DNA damage in cancer-prone inflammatory diseases: A review of published adduct types and levels in humans. *Free Radical Biology & Medicine, 43*(8), 1109-1120.

(*54*) Hecht, S. S. (1998). Biochemistry, biology, and carcinogenicity of tobacco-specific N-nitrosamines. *Chemical Research in Toxicology, 11*(6), 559-603.
(*55*) Wang, M., Cheng, G., Sturla, S. J., Shi, Y., McIntee, E. J., Villalta, P. W., et al. (2003). Identification of adducts formed by pyridyloxobutylation of deoxyguanosine and DNA by 4-(acetoxymethylnitrosamino)-1-(3-pyridyl)-1-butanone, a chemically activated form of tobacco specific carcinogens. *Chemical Research in Toxicology, 16*(5), 616-626.
(*56*) Hecht, S. S., Villalta, P. W., Sturla, S. J., Cheng, G., Yu, N., Upadhyaya, P., et al. (2004). Identification of O2-substituted pyrimidine adducts formed in reactions of 4-(acetoxymethylnitrosamino)-1-(3-pyridyl)-1-butanone and 4-(acetoxymethylnitros- amino)-1-(3-pyridyl)-1-butanol with DNA. *Chemical Research in Toxicology, 17*(5), 588-597.
(*57*) Wang, L., Spratt, T. E., Liu, X. K., Hecht, S. S., Pegg, A. E., & Peterson, L. A. (1997). Pyridyloxobutyl adduct O6-[4-oxo-4-(3-pyridyl)butyl]guanine is present in 4-(acetoxymethylnitrosamino)-1-(3-pyridyl)-1-butanone-treated DNA and is a substrate for O6-alkylguanine-DNA alkyltransferase. *Chemical Research in Toxicology, 10*(5), 562-567.
(*58*) Upadhyaya, P., Sturla, S. J., Tretyakova, N., Ziegel, R., Villalta, P. W., Wang, M., et al. (2003). Identification of adducts produced by the reaction of 4-(acetoxymethylnitrosamino)-1-(3-pyridyl)-1-butanol with deoxyguanosine and DNA. *Chemical Research in Toxicology, 16*(2), 180-190.
(*59*) Gilman, A., & Philips, F. S. (1946). The biological actions and therapeutic applications of the B-chloroethyl amines and sulfides. *Science (New York, N.Y.), 103*, 409-415.
(*60*) Ojwang, J. O., Grueneberg, D. A., & Loechler, E. L. (1989). Synthesis of a duplex oligonucleotide containing a nitrogen mustard interstrand DNA-DNA cross-link. *Cancer Research, 49*(23), 6529-6537.
(*61*) Dong, Q., Barsky, D., Colvin, M. E., Melius, C. F., Ludeman, S. M., Moravek, J. F., et al. (1995). A structural basis for a phosphoramide mustard-induced DNA interstrand cross-link at 5'-d(GAC). *Proceedings of the National Academy of Sciences of the United States of America, 92*(26), 12170-12174.
(*62*) Guainazzi, A., & Scharer, O. D. (2010). Using synthetic DNA interstrand crosslinks to elucidate repair pathways and identify new therapeutic targets for cancer chemotherapy. *Cellular and Molecular Life Sciences : CMLS, 67*(21), 3683-3697.
(*63*) Kallama, S., & Hemminki, K. (1986). Stabilities of 7-alkylguanosines and 7-deoxyguanosines formed by phosphoramide mustard and nitrogen mustard. *Chemico-Biological Interactions, 57*(1), 85-96.
(*64*) Kohn, K. W. (1977). Interstrand cross-linking of DNA by 1,3-bis(2-chloroethyl)-1-nitrosourea and other 1-(2-haloethyl)-1-nitrosoureas. *Cancer Research, 37*(5), 1450-1454.
(*65*) Kaina, B., Christmann, M., Naumann, S., & Roos, W. P. (2007). MGMT: Key node in the battle against genotoxicity, carcinogenicity and apoptosis induced by alkylating agents. *DNA Repair, 6*(8), 1079-1099.
(*66*) Boland, I., Vassal, G., Morizet, J., Terrier-Lacombe, M. J., Valteau-Couanet, D., Kalifa, C., et al. (1999). Busulphan is active against neuroblastoma and medulloblastoma xenografts in athymic mice at clinically achievable plasma drug concentrations. *British Journal of Cancer, 79*(5-6), 787-792.
(*67*) Avendaño, C., and Menéndez, J. C. (2008) Medicinal Chemistry of Anticancer Drugs, Elsevier, Oxford, UK.
(*68*) Iwamoto, T., Hiraku, Y., Oikawa, S., Mizutani, H., Kojima, M., & Kawanishi, S. (2004). DNA intrastrand cross-link at the 5'-GA-3' sequence formed by busulfan and its role in the cytotoxic effect. *Cancer Science, 95*(5), 454-458.
(*69*) Bedford, P., & Fox, B. W. (1982). Repair of DNA interstrand crosslinks after busulphan. A possible mode of resistance. *Cancer Chemotherapy and Pharmacology, 8*(1), 3-7.
(*70*) Ponti, M., Souhami, R. L., Fox, B. W., & Hartley, J. A. (1991). DNA interstrand crosslinking and sequence selectivity of dimethanesulphonates. *British Journal of Cancer, 63*(5), 743-747.
(*71*) Westerhof, G. R., Ploemacher, R. E., Boudewijn, A., Blokland, I., Dillingh, J. H., McGown, A. T., et al. (2000). Comparison of different busulfan analogues for depletion of hematopoietic stem cells and

promotion of donor-type chimerism in murine bone marrow transplant recipients. *Cancer Research, 60*(19), 5470-5478.
(*72*) Hincks, J. R., Adlakha, A., Cook, C. A., Johnson, C. S., Furmanski, P., & Gibson, N. W. (1990). In vitro studies on the mechanism of action of hepsulfam in chronic myelogenous leukemia patients. *Cancer Research, 50*(23), 7559-7563.
(*73*) Pacheco, D. Y., Cook, C., Hincks, J. R., & Gibson, N. W. (1990). Mechanisms of toxicity of hepsulfam in human tumor cell lines. *Cancer Research, 50*(23), 7555-7558.
(*74*) Pacheco, D. Y., Stratton, N. K., & Gibson, N. W. (1989). Comparison of the mechanism of action of busulfan with hepsulfam, a new antileukemic agent, in the L1210 cell line. *Cancer Research, 49*(18), 5108-5110.
(*75*) Streeper, R. T., Cotter, R. J., Colvin, M. E., Hilton, J., & Colvin, O. M. (1995). Molecular pharmacology of hepsulfam, NSC 3296801: Identification of alkylated nucleosides, alkylation site, and site of DNA cross-linking. *Cancer Research, 55*(7), 1491-1498.
(*76*) Chaney, S. G., & Sancar, A. (1996). DNA repair: Enzymatic mechanisms and relevance to drug response. *Journal of the National Cancer Institute, 88*(19), 1346-1360.
(*77*) Hoeijmakers, J. H. (2001). Genome maintenance mechanisms for preventing cancer. *Nature, 411*(6835), 366-374.
(*78*) Sancar, A., Lindsey-Boltz, L. A., Unsal-Kacmaz, K., & Linn, S. (2004). Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annual Review of Biochemistry*, *73*, 39-85.
(*79*) Sancar, A. (1996). DNA excision repair. *Annual Review of Biochemistry, 65*, 43-81.
(*80*) Lindahl, T., & Wood, R. D. (1999). Quality control by DNA repair. *Science (New York, N.Y.)*, *286*(5446), 1897-1905.
(*81*) Mitra, S., Boldogh, I., Izumi, T., & Hazra, T. K. (2001). Complexities of the DNA base excision repair pathway for repair of oxidative DNA damage. *Environmental and Molecular Mutagenesis, 38*(2-3), 180-190.
(*82*) Reardon, J. T., Vaisman, A., Chaney, S. G., & Sancar, A. (1999). Efficient nucleotide excision repair of cisplatin, oxaliplatin, and bis-aceto-ammine-dichloro-cyclohexylamine-platinum(IV) (JM216) platinum intrastrand DNA diadducts. *Cancer Research, 59*(16), 3968-3971.
(*83*) Reed, E. (1998). Platinum-DNA adduct, nucleotide excision repair and platinum based anti-cancer chemotherapy. *Cancer Treatment Reviews, 24*(5), 331-344.
(*84*) Sinha, R. P., & Hader, D. P. (2002). UV-induced DNA damage and repair: A review. *Photochemical & Photobiological Sciences : Official Journal of the European Photochemistry Association and the European Society for Photobiology, 1*(4), 225-236.
(*85*) Theis, K., Skorvaga, M., Machius, M., Nakagawa, N., Van Houten, B., & Kisker, C. (2000). The nucleotide excision repair protein UvrB, a helicase-like enzyme with a catch. *Mutation Research, 460*(3-4), 277-300.
(*86*) Gunz, D., Hess, M. T., & Naegeli, H. (1996). Recognition of DNA adducts by human nucleotide excision repair. evidence for a thermodynamic probing mechanism. *The Journal of Biological Chemistry, 271*(41), 25089-25098.
(*87*) Gillet, L. C., & Scharer, O. D. (2006). Molecular mechanisms of mammalian global genome nucleotide excision repair. *Chemical Reviews, 106*(2), 253-276.
(*88*) Van Houten, B., & McCullough, A. (1994). Nucleotide excision repair in E. coli. *Annals of the New York Academy of Sciences, 726*, 236-251.
(*89*) Kolodner, R. (1996). Biochemistry and genetics of eukaryotic mismatch repair. *Genes & Development, 10*(12), 1433-1442.
(*90*) Modrich, P. (1991). Mechanisms and biological effects of mismatch repair. *Annual Review of Genetics, 25*, 229-253.
(*91*) Kunkel, T. A., & Erie, D. A. (2005). DNA mismatch repair. *Annual Review of Biochemistry, 74*, 681-710.

(*92*) Li, G. M. (2008). Mechanisms and functions of DNA mismatch repair. *Cell Research, 18*(1), 85-98.
(*93*) Pfeiffer, P., Goedecke, W., & Obe, G. (2000). Mechanisms of DNA double-strand break repair and their potential to induce chromosomal aberrations. *Mutagenesis, 15*(4), 289-302.
(*94*) Jackson, S. P. (2002). Sensing and repairing DNA double-strand breaks. *Carcinogenesis, 23*(5), 687-696.
(*95*) Mahaney, B. L., Meek, K., & Lees-Miller, S. P. (2009). Repair of ionizing radiation-induced DNA double-strand breaks by non-homologous end-joining. *The Biochemical Journal, 417*(3), 639-650.
(*96*) Lees-Miller, S. P., & Meek, K. (2003). Repair of DNA double strand breaks by non-homologous end joining. *Biochimie, 85*(11), 1161-1173.
(*97*) Johnson, R. D., & Jasin, M. (2000). Sister chromatid gene conversion is a prominent double-strand break repair pathway in mammalian cells. *The EMBO Journal, 19*(13), 3398-3407.
(*98*) Lavin, M. F. (2004). The Mre11 complex and ATM: A two-way functional interaction in recognising and signaling DNA double strand breaks. *DNA Repair, 3*(11), 1515-1520.
(*99*) Ohnishi, T., Mori, E., & Takahashi, A. (2009). DNA double-strand breaks: Their production, recognition, and repair in eukaryotes. *Mutation Research, 669*(1-2), 8-12.
(*100*) van Gent, D. C., Hoeijmakers, J. H., & Kanaar, R. (2001). Chromosomal stability and the DNA double-stranded break connection. *Nature Reviews.Genetics, 2*(3), 196-206.
(*101*) Sedgwick, B. (2004). Repairing DNA-methylation damage. *Nature Reviews.Molecular Cell Biology, 5*(2), 148-157.
(*102*) Duncan, T., Trewick, S. C., Koivisto, P., Bates, P. A., Lindahl, T., & Sedgwick, B. (2002). Reversal of DNA alkylation damage by two human dioxygenases. *Proceedings of the National Academy of Sciences of the United States of America, 99*(26), 16660-16665.
(*103*) Liu, L., & Gerson, S. L. (2006). Targeted modulation of MGMT: Clinical implications. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research, 12*(2), 328-331.
(*104*) Srivenugopal, K. S., Yuan, X. H., Friedman, H. S., & Ali-Osman, F. (1996). Ubiquitination-dependent proteolysis of O6-methylguanine-DNA methyltransferase in human and murine tumor cells following inactivation with O6-benzylguanine or 1,3-bis(2-chloroethyl)-1-nitrosourea. *Biochemistry, 35*(4), 1328-1334.
(*105*) Fang, Q., Kanugula, S., & Pegg, A. E. (2005). Function of domains of human O6-alkylguanine-DNA alkyltransferase. *Biochemistry, 44*(46), 15396-15405.
(*106*) Daniels, D. S., & Tainer, J. A. (2000). Conserved structural motifs governing the stoichiometric repair of alkylated DNA by O(6)-alkylguanine-DNA alkyltransferase. *Mutation Research, 460*(3-4), 151-163.
(*107*) Jeggo, P. (1979). Isolation and characterization of escherichia coli K-12 mutants unable to induce the adaptive response to simple alkylating agents. *Journal of Bacteriology, 139*(3), 783-791.
(*108*) Jeggo, P., Defais, T. M., Samson, L., & Schendel, P. (1977). An adaptive response of E. coli to low levels of alkylating agent: Comparison with previously characterised DNA repair pathways. *Molecular & General Genetics : MGG, 157*(1), 1-9.
(*109*) Samson, L., & Cairns, J. (1977). A new pathway for DNA repair in escherichia coli. *Nature, 267*(5608), 281-283.
(*110*) Olsson, M., & Lindahl, T. (1980). Repair of alkylated DNA in escherichia coli. methyl group transfer from O6-methylguanine to a protein cysteine residue. *The Journal of Biological Chemistry, 255*(22), 10569-10571.
(*111*) Lindahl, T., Demple, B., & Robins, P. (1982). Suicide inactivation of the E. coli O6-methylguanine-DNA methyltransferase. *The EMBO Journal, 1*(11), 1359-1363.
(*112*) Preston, B. D., Singer, B., & Loeb, L. A. (1987). Comparison of the relative mutagenicities of O-alkylthymines site-specifically incorporated into phi X174 DNA. *The Journal of Biological Chemistry, 262*(28), 13821-13827.

(*113*) Sedgwick, B., Robins, P., Totty, N., & Lindahl, T. (1988). Functional domains and methyl acceptor sites of the escherichia coli ada protein. *The Journal of Biological Chemistry, 263*(9), 4430-4433.
(*114*) Graves, R. J., Li, B. F., & 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*(4), 661-666.
(*115*) Teo, I., Sedgwick, B., Kilpatrick, M. W., McCarthy, T. V., & Lindahl, T. (1986). The intracellular signal for induction of resistance to alkylating agents in E. coli. *Cell, 45*(2), 315-324.
(*116*) Sakumi, K., & 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. *Journal of Molecular Biology, 205*(2), 373-385.
(*117*) Sassanfar, M., Dosanjh, M. K., Essigmann, J. M., & 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. *The Journal of Biological Chemistry, 266*(5), 2767-2771.
(*118*) Moore, M. H., Gulbis, J. M., Dodson, E. J., Demple, B., & Moody, P. C. (1994). Crystal structure of a suicidal DNA repair protein: The ada O6-methylguanine-DNA methyltransferase from E. coli. *The EMBO Journal, 13*(7), 1495-1501.
(*119*) Guengerich, F. P., Fang, Q., Liu, L., Hachey, D. L., & Pegg, A. E. (2003). O6-alkylguanine-DNA alkyltransferase: Low pKa and high reactivity of cysteine 145. *Biochemistry, 42*(37), 10965-10970.
(*120*) Yang, C. G., Garcia, K., & He, C. (2009). Damage detection and base flipping in direct DNA alkylation repair. *Chembiochem : A European Journal of Chemical Biology, 10*(3), 417-423.
(*121*) Mitra, S., Pal, B. C., & Foote, R. S. (1982). O6-methylguanine-DNA methyltransferase in wild-type and ada mutants of escherichia coli. *Journal of Bacteriology, 152*(1), 534-537.
(*122*) Rebeck, G. W., Smith, C. M., Goad, D. L., & Samson, L. (1989). Characterization of the major DNA repair methyltransferase activity in unadapted escherichia coli and identification of a similar activity in salmonella typhimurium. *Journal of Bacteriology, 171*(9), 4563-4568.
(*123*) Fang, Q., Kanugula, S., Tubbs, J. L., Tainer, J. A., & Pegg, A. E. (2010). Repair of O4-alkylthymine by O6-alkylguanine-DNA alkyltransferases. *The Journal of Biological Chemistry, 285*(11), 8185-8195.
(*124*) Pegg, A. E., Dolan, M. E., Scicchitano, D., & Morimoto, K. (1985). Studies of the repair of O6-alkylguanine and O4-alkylthymine in DNA by alkyltransferases from mammalian cells and bacteria. *Environmental Health Perspectives, 62*, 109-114.
(*125*) Schoonhoven, N. (2010) The role of residue -5 on O^6-alkylated DNA duplex repair by O^6-alkylguanine-DNA-alkyltransferase in *Chemistry and Biochemistry*, Concordia University, Montreal.
(*126*) Koike, G., Maki, H., Takeya, H., Hayakawa, H., & Sekiguchi, M. (1990). Purification, structure, and biochemical properties of human O6-methylguanine-DNA methyltransferase. *The Journal of Biological Chemistry, 265*(25), 14754-14762.
(*127*) Goodtzova, K., Kanugula, S., Edara, S., Pauly, G. T., Moschel, R. C., & Pegg, A. E. (1997). Repair of O6-benzylguanine by the escherichia coli ada and ogt and the human O6-alkylguanine-DNA alkyltransferases. *The Journal of Biological Chemistry, 272*(13), 8332-8339.
(*128*) Mijal, R. S., Thomson, N. M., Fleischer, N. L., Pauly, G. T., Moschel, R. C., Kanugula, S., et al. (2004). The repair of the tobacco specific nitrosamine derived adduct O6-[4-oxo-4-(3-pyridyl)butyl]guanine by O6-alkylguanine-DNA alkyltransferase variants. *Chemical Research in Toxicology, 17*(3), 424-434.
(*129*) Coulter, R., Blandino, M., Tomlinson, J. M., Pauly, G. T., Krajewska, M., Moschel, R. C., et al. (2007). Differences in the rate of repair of O6-alkylguanines in different sequence contexts by O6-alkylguanine-DNA alkyltransferase. *Chemical Research in Toxicology, 20*(12), 1966-1971.
(*130*) O'Toole, S. M., Pegg, A. E., & Swenberg, J. A. (1993). Repair of O6-methylguanine and O4-methylthymidine in F344 rat liver following treatment with 1,2-dimethylhydrazine and O6-benzylguanine. *Cancer Research, 53*(17), 3895-3898.

(*131*) Zak, P., Kleibl, K., & Laval, F. (1994). Repair of O6-methylguanine and O4-methylthymine by the human and rat O6-methylguanine-DNA methyltransferases. *The Journal of Biological Chemistry, 269*(1), 730-733.
(*132*) Pegg, A. E. (2011). Multifaceted roles of alkyltransferase and related proteins in DNA repair, DNA damage, resistance to chemotherapy, and research tools. *Chemical Research in Toxicology, 24*(5), 618-639.
(*133*) Samson, L., Han, S., Marquis, J. C., & Rasmussen, L. J. (1997). Mammalian DNA repair methyltransferases shield O4MeT from nucleotide excision repair. *Carcinogenesis, 18*(5), 919-924.
(*134*) Encell, L. P., Coates, M. M., & Loeb, L. A. (1998). Engineering human DNA alkyltransferases for gene therapy using random sequence mutagenesis. *Cancer Research, 58*(5), 1013-1020.
(*135*) Encell, L. P., & Loeb, L. A. (1999). Redesigning the substrate specificity of human O(6)-alkylguanine-DNA alkyltransferase. mutants with enhanced repair of O(4)-methylthymine. *Biochemistry, 38*(37), 12097-12103.
(*136*) Crone, T. M., Kanugula, S., & Pegg, A. E. (1995). Mutations in the ada O6-alkylguanine-DNA alkyltransferase conferring sensitivity to inactivation by O6-benzylguanine and 2,4-diamino-6-benzyloxy-5-nitrosopyrimidine. *Carcinogenesis, 16*(8), 1687-1692.
(*137*) Elder, R. H., Margison, G. P., & Rafferty, J. A. (1994). Differential inactivation of mammalian and escherichia coli O6-alkylguanine-DNA alkyltransferases by O6-benzylguanine. *The Biochemical Journal, 298 (Pt 1)*(Pt 1), 231-235.
(*138*) Pegg, A. E., Boosalis, M., Samson, L., Moschel, R. C., Byers, T. L., Swenn, K., et al. (1993). Mechanism of inactivation of human O6-alkylguanine-DNA alkyltransferase by O6-benzylguanine. *Biochemistry, 32*(45), 11998-12006.
(*139*) Bodell, W. J., Aida, T., Berger, M. S., & Rosenblum, M. L. (1985). Repair of O6-(2-chloroethyl)guanine mediates the biological effects of chloroethylnitrosoureas. *Environmental Health Perspectives, 62*, 119-126.
(*140*) Loeber, R., Michaelson, E., Fang, Q., Campbell, C., Pegg, A. E., & Tretyakova, N. (2008). Cross-linking of the DNA repair protein Omicron6-alkylguanine DNA alkyltransferase to DNA in the presence of antitumor nitrogen mustards. *Chemical Research in Toxicology, 21*(4), 787-795.
(*141*) Cai, Y., Wu, M. H., Ludeman, S. M., Grdina, D. J., & Dolan, M. E. (1999). Role of O6-alkylguanine-DNA alkyltransferase in protecting against cyclophosphamide-induced toxicity and mutagenicity. *Cancer Research, 59*(13), 3059-3063.
(*142*) McElhinney, R. S., Donnelly, D. J., McCormick, J. E., Kelly, J., Watson, A. J., Rafferty, J. A., et al. (1998). Inactivation of O6-alkylguanine-DNA alkyltransferase. 1. novel O6-(hetarylmethyl)guanines having basic rings in the side chain. *Journal of Medicinal Chemistry, 41*(26), 5265-5271.
(*143*) Middleton, M. R., Kelly, J., Thatcher, N., Donnelly, D. J., McElhinney, R. S., McMurry, T. B., et al. (2000). O(6)-(4-bromothenyl)guanine improves the therapeutic index of temozolomide against A375M melanoma xenografts. *International Journal of Cancer.Journal International Du Cancer, 85*(2), 248-252.
(*144*) Margison, G. P., & Santibanez-Koref, M. F. (2002). O6-alkylguanine-DNA alkyltransferase: Role in carcinogenesis and chemotherapy. *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology, 24*(3), 255-266.
(*145*) Booth, J. D., Murphy, S. P., Noronha, A. M., & Wilds, C. J. (2008). Effect of linker length on DNA duplexes containing a mismatched O6-2'-deoxyguanosine-alkyl interstrand cross-link. *Nucleic Acids Symposium Series (2004), (52)*(52), 431-432.
(*146*) Fang, Q., Noronha, A. M., Murphy, S. P., Wilds, C. J., Tubbs, J. L., Tainer, J. A., et al. (2008). Repair of O6-G-alkyl-O6-G interstrand cross-links by human O6-alkylguanine-DNA alkyltransferase. *Biochemistry, 47*(41), 10892-10903.
(*147*) Wibley, J. E., Pegg, A. E., & Moody, P. C. (2000). Crystal structure of the human O(6)-alkylguanine-DNA alkyltransferase. *Nucleic Acids Research, 28*(2), 393-401.

(*148*) Daniels, D. S., Mol, C. D., Arvai, A. S., Kanugula, S., Pegg, A. E., & Tainer, J. A. (2000). Active and alkylated human AGT structures: A novel zinc site, inhibitor and extrahelical base binding. *The EMBO Journal, 19*(7), 1719-1730.
(*149*) Rasimas, J. J., Kanugula, S., Dalessio, P. M., Ropson, I. J., Fried, M. G., & Pegg, A. E. (2003). Effects of zinc occupancy on human O6-alkylguanine-DNA alkyltransferase. *Biochemistry, 42*(4), 980-990.
(*150*) Tubbs, J. L., Pegg, A. E., & Tainer, J. A. (2007). DNA binding, nucleotide flipping, and the helix-turn-helix motif in base repair by O6-alkylguanine-DNA alkyltransferase and its implications for cancer chemotherapy. *DNA Repair, 6*(8), 1100-1115.
(*151*) Duguid, E. M., Rice, P. A., & He, C. (2005). The structure of the human AGT protein bound to DNA and its implications for damage detection. *Journal of Molecular Biology, 350*(4), 657-666.
(*152*) Teo, A. K., Oh, H. K., Ali, R. B., & Li, B. F. (2001). The modified human DNA repair enzyme O(6)-methylguanine-DNA methyltransferase is a negative regulator of estrogen receptor-mediated transcription upon alkylation DNA damage. *Molecular and Cellular Biology, 21*(20), 7105-7114.
(*153*) Lim, A., & Li, B. F. (1996). The nuclear targeting and nuclear retention properties of a human DNA repair protein O6-methylguanine-DNA methyltransferase are both required for its nuclear localization: The possible implications. *The EMBO Journal, 15*(15), 4050-4060.
(*154*) Ali, R. B., Teo, A. K., Oh, H. K., Chuang, L. S., Ayi, T. C., & Li, B. F. (1998). Implication of localization of human DNA repair enzyme O6-methylguanine-DNA methyltransferase at active transcription sites in transcription-repair coupling of the mutagenic O6-methylguanine lesion. *Molecular and Cellular Biology, 18*(3), 1660-1669.
(*155*) Daniels, D. S., Woo, T. T., Luu, K. X., Noll, D. M., Clarke, N. D., Pegg, A. E., et al. (2004). DNA binding and nucleotide flipping by the human DNA repair protein AGT. *Nature Structural & Molecular Biology, 11*(8), 714-720.
(*156*) Hu, J., Ma, A., & Dinner, A. R. (2008). A two-step nucleotide-flipping mechanism enables kinetic discrimination of DNA lesions by AGT. *Proceedings of the National Academy of Sciences of the United States of America, 105*(12), 4615-4620.
(*157*) Jena, N. R., Shukla, P. K., Jena, H. S., Mishra, P. C., & Suhai, S. (2009). O6-methylguanine repair by O6-alkylguanine-DNA alkyltransferase. *The Journal of Physical Chemistry.B, 113*(51), 16285-16290.
(*158*) Hou, Q., Du, L., Gao, J., Liu, Y., & Liu, C. (2010). QM/MM study on the reaction mechanism of O6-alkylguanine-DNA alkyltransferase. *The Journal of Physical Chemistry.B, 114*(46), 15296-15300.
(*159*) Duguid, E. M., Mishina, Y., & He, C. (2003). How do DNA repair proteins locate potential base lesions? a chemical crosslinking method to investigate O6-alkylguanine-DNA alkyltransferases. *Chemistry & Biology, 10*(9), 827-835.
(*160*) Zang, H., Fang, Q., Pegg, A. E., & Guengerich, F. P. (2005). Kinetic analysis of steps in the repair of damaged DNA by human O6-alkylguanine-DNA alkyltransferase. *The Journal of Biological Chemistry, 280*(35), 30873-30881.
(*161*) Parikh, S. S., Mol, C. D., Hosfield, D. J., & Tainer, J. A. (1999). Envisioning the molecular choreography of DNA base excision repair. *Current Opinion in Structural Biology, 9*(1), 37-47.
(*162*) Luu, K. X., Kanugula, S., Pegg, A. E., Pauly, G. T., & Moschel, R. C. (2002). Repair of oligodeoxyribonucleotides by O(6)-alkylguanine-DNA alkyltransferase. *Biochemistry, 41*(27), 8689-8697.
(*163*) Rasimas, J. J., Pegg, A. E., & Fried, M. G. (2003). DNA-binding mechanism of O6-alkylguanine-DNA alkyltransferase. effects of protein and DNA alkylation on complex stability. *The Journal of Biological Chemistry, 278*(10), 7973-7980.
(*164*) Fried, M. G., Kanugula, S., Bromberg, J. L., & Pegg, A. E. (1996). DNA binding mechanism of O6-alkylguanine-DNA alkyltransferase: Stoichiometry and effects of DNA base composition and secondary structure on complex stability. *Biochemistry, 35*(48), 15295-15301.
(*165*) Bender, K., Federwisch, M., Loggen, U., Nehls, P., & Rajewsky, M. F. (1996). Binding and repair of O6-ethylguanine in double-stranded oligodeoxynucleotides by recombinant human O6-alkylguanine-DNA

alkyltransferase do not exhibit significant dependence on sequence context. *Nucleic Acids Research, 24*(11), 2087-2094.
(*166*) Rasimas, J. J., Kar, S. R., Pegg, A. E., & Fried, M. G. (2007). Interactions of human O6-alkylguanine-DNA alkyltransferase (AGT) with short single-stranded DNAs. *The Journal of Biological Chemistry, 282*(5), 3357-3366.
(*167*) Melikishvili, M., Rasimas, J. J., Pegg, A. E., & Fried, M. G. (2008). Interactions of human O(6)-alkylguanine-DNA alkyltransferase (AGT) with short double-stranded DNAs. *Biochemistry, 47*(52), 13754-13763.
(*168*) Adams, C. A., Melikishvili, M., Rodgers, D. W., Rasimas, J. J., Pegg, A. E., & Fried, M. G. (2009). Topologies of complexes containing O6-alkylguanine-DNA alkyltransferase and DNA. *Journal of Molecular Biology, 389*(2), 248-263.
(*169*) Adams, C. A., & Fried, M. G. (2011). Mutations that probe the cooperative assembly of O(6)-alkylguanine-DNA alkyltransferase complexes. *Biochemistry, 50*(10), 1590-1598.
(*170*) Melikishvili, M., & Fried, M. G. (2012). Lesion-specific DNA-binding and repair activities of human O(6)-alkylguanine DNA alkyltransferase. *Nucleic Acids Research, 40*(18), 9060-9072.
(*171*) Tessmer, I., Melikishvili, M., & Fried, M. G. (2012). Cooperative cluster formation, DNA bending and base-flipping by O6-alkylguanine-DNA alkyltransferase. *Nucleic Acids Research, 40*(17), 8296-8308.
(*172*) Takahashi, M., Sakumi, K., & Sekiguchi, M. (1990). Interaction of ada protein with DNA examined by fluorescence anisotropy of the protein. *Biochemistry, 29*(14), 3431-3436.
(*173*) Verdemato, P. E., Brannigan, J. A., Damblon, C., Zuccotto, F., Moody, P. C., & Lian, L. Y. (2000). DNA-binding mechanism of the escherichia coli ada O(6)-alkylguanine-DNA alkyltransferase. *Nucleic Acids Research, 28*(19), 3710-3718.
(*174*) Alzeer, J., & Scharer, O. D. (2006). A modified thymine for the synthesis of site-specific thymine-guanine DNA interstrand crosslinks. *Nucleic Acids Research, 34*(16), 4458-4466.
(*175*) Wilds, C. J., Xu, F., & 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. *Chemical Research in Toxicology, 21*(3), 686-695.
(*176*) Angelov, T., Guainazzi, A., & Scharer, O. D. (2009). Generation of DNA interstrand cross-links by post-synthetic reductive amination. *Organic Letters, 11*(3), 661-664.
(*177*) Stevens, K., & Madder, A. (2009). Furan-modified oligonucleotides for fast, high-yielding and site-selective DNA inter-strand cross-linking with non-modified complements. *Nucleic Acids Research, 37*(5), 1555-1565.
(*178*) Ali, M. M., Imoto, S., Li, Y., Sasaki, S., & Nagatsugi, F. (2009). Incorporation of an inducible nucleotide analog into DNA by DNA polymerases. *Bioorganic & Medicinal Chemistry, 17*(7), 2859-2863.
(*179*) Taniguchi, Y., Kurose, Y., Nishioka, T., Nagatsugi, F., & Sasaki, S. (2010). The alkyl-connected 2-amino-6-vinylpurine (AVP) crosslinking agent for improved selectivity to the cytosine base in RNA. *Bioorganic & Medicinal Chemistry, 18*(8), 2894-2901.
(*180*) Huang, H., Kim, H. Y., Kozekov, I. D., Cho, Y. J., Wang, H., Kozekova, A., et al. (2009). Stereospecific formation of the (R)-gamma-hydroxytrimethylene interstrand N2-dG:N2-dG cross-link arising from the gamma-OH-1,N2-propano-2'-deoxyguanosine adduct in the 5'-CpG-3' DNA sequence. *Journal of the American Chemical Society, 131*(24), 8416-8424.
(*181*) Noll, D. M., Webba da Silva, M., Noronha, A. M., Wilds, C. J., Colvin, O. M., Gamcsik, M. P., et al. (2005). Structure, flexibility, and repair of two different orientations of the same alkyl interstrand DNA cross-link. *Biochemistry, 44*(18), 6764-6775.
(*182*) Ravdin, P. M., Havlin, K. A., Marshall, M. V., Brown, T. D., Koeller, J. M., Kuhn, J. G., et al. (1991). A Phase I Clinical and Pharmacokinetic Trial of Hepsulfam. *Cancer Res. 51*(23), 6268-6272.
(*183*) Haines, J. A., Reese, C. B., & Lord Todd. (1962). The methylation of guanosine and related compounds with diazomethane. *J. Chem. Soc*. 5281-5288.

(*184*) Wilds, C.J., Booth. J. D., & Noronha, A. M. (2006). Synthesis of oligonucleotides containing an O^6-G-alkyl-O^6-G interstrand cross-link. *Tetrahedron Lett*. *47*(51), 9125-9128.
(*185*) Khan, O., & Middleton, M. R. (2007). The therapeutic potential of O6-alkylguanine DNA alkyltransferase inhibitors. *Expert Opinion on Investigational Drugs, 16*(10), 1573-1584.
(*186*) Himmelsbach, F., Schulz, B. S., Trichtinger, T., Charubala, R., & Pfleiderer, W. (1984) The *p*-nitrophenylethyl (NPE) group: A versatile new blocking group for phosphate and aglycone protection in nucleosides and nucleotides. *Tetrahedron*, *40*(1), 59-72.
(*187*) Shibata, T., Glynn, N., McMurry, T. B., McElhinney, R. S., Margison, G. P., & Williams, D. M. (2006). Novel synthesis of O6-alkylguanine containing oligodeoxyribonucleotides as substrates for the human DNA repair protein, O6-methylguanine DNA methyltransferase (MGMT). *Nucleic Acids Research, 34*(6), 1884-1891.
(*188*) Mitsunobu, O. (1981). The Use of Diethyl Azodicarboxylate and Triphenylphosphine in Synthesis and Transformation of Natural Products. Synthesis **1**, 1–28
(*189*) Pegg, A. E. (2000). Repair of O(6)-alkylguanine by alkyltransferases. *Mutation Research, 462*(2-3), 83-100.
(*190*) Booth, J. (2008) Synthesis and characterization of oligonucleotides containing an O^6-deoxyguanosine-alkyl-O^6-deoxyguanosine interstrand cross-link in *Chemistry and Biochemistry*, Concordia University, Montreal.
(*191*) Puglisi, J. D., & Tinoco, I.,Jr. (1989). Absorbance melting curves of RNA. *Methods in Enzymology, 180*, 304-325.
(*192*) Buggia, I., Locatelli, F., Regazzi, M. B., & Zecca, M. (1994). Busulfan. *The Annals of Pharmacotherapy, 28*(9), 1055-1062.
(*193*) Tong, W. P., & Ludlum, D. B. (1980). Crosslinking of DNA by busulfan. formation of diguanyl derivatives. *Biochimica Et Biophysica Acta, 608*(1), 174-181.
(*194*) Hayakawa, Y., Kato, H., Uchiyama, M., Kajino H., & Noyori, R. (1986). Allyloxycarbonyl group: a versatile blocking group for nucleotide synthesis. *J. Org. Chem., 51*(12), 2400-2402.
(*195*) Braich, R. S., & Damha, M. J. (1997). Regiospecific solid-phase synthesis of branched oligonucleotides. effect of vicinal 2',5'- (or 2',3'-) and 3',5'-phosphodiester linkages on the formation of hairpin DNA. *Bioconjugate Chemistry, 8*(3), 370-377.
(*196*) Zhu, Q., Delaney, M. O., & Greenberg, M. M. (2001). Observation and elimination of N-acetylation of oligonucleotides prepared using fast-deprotecting phosphoramidites and ultra-mild deprotection. *Bioorganic & Medicinal Chemistry Letters, 11*(9), 1105-1107.
(*197*) Damha, M. J., & Ogilvie, K K. (1993) Oligoribonucleotide synthesis - the silyl-phosphoramidite method in "Protocols for Oligonucleotide and Analogs: Synthesis and Properties" S. Agrawal (ed.), *Methods in Molecular Biology,* pp.81-114, The Humana Press Inc., Totowa, New Jersey.
(*198*) Gaffney, B. L., & 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*(14), 5881-5889.
(*199*) Kuzmich, S., Marky, L. A., & Jones, R. A. (1983). Specifically alkylated DNA fragments. synthesis and physical characterization of d[CGC(O6Me)GCG] and d[CGT(O6Me)GCG]. *Nucleic Acids Research, 11*(10), 3393-3403.
(*200*) Johnson, W. C. Jr., (2006) Circular Dichroism and Its Empirical Application to Biopolymers, *Methods of Biochemical Analysis*, pp.61-163, John Wiley & Sons, New York.
(*201*) Johnson, W. C. Jr., (1996) Determination of the Conformation of Nucleic Acids by Electronic CD, *Circular Dichroism and the Conformational Analysis of Biomolecules*, Plenum Press, New York, 1996, pp. 433-468.
(*202*) Crone, T. M., & Pegg, A. E. (1993). A single amino acid change in human O6-alkylguanine-DNA alkyltransferase decreasing sensitivity to inactivation by O6-benzylguanine. *Cancer Research, 53*(20), 4750-4753.

(*203*) Crone, T. M., Goodtzova, K., Edara, S., & Pegg, A. E. (1994). Mutations in human O6-alkylguanine-DNA alkyltransferase imparting resistance to O6-benzylguanine. *Cancer Research, 54*(23), 6221-6227.
(*204*) Kanugula, S., & Pegg, A. E. (2003). Alkylation damage repair protein O6-alkylguanine-DNA alkyltransferase from the hyperthermophiles aquifex aeolicus and archaeoglobus fulgidus. *The Biochemical Journal, 375*(Pt 2), 449-455.
(*205*) Goodtzova, K., Crone, T. M., & Pegg, A. E. (1994). Activation of human O6-alkylguanine-DNA alkyltransferase by DNA. *Biochemistry, 33*(28), 8385-8390.
(*206*) Rasimas, J. J., Dalessio, P. A., Ropson, I. J., Pegg, A. E., & Fried, M. G. (2004). Active-site alkylation destabilizes human O6-alkylguanine DNA alkyltransferase. *Protein Science : A Publication of the Protein Society, 13*(1), 301-305.
(*207*) Christmann, M., Tomicic, M. T., Roos, W. P., & Kaina, B. (2003). Mechanisms of human DNA repair: An update. *Toxicology, 193*(1-2), 3-34.
(*208*) McCabe, K. M., Olson, S. B., & Moses, R. E. (2009). DNA interstrand crosslink repair in mammalian cells. *Journal of Cellular Physiology, 220*(3), 569-573.
(*209*) Ho, T. V., & Scharer, O. D. (2010). Translesion DNA synthesis polymerases in DNA interstrand crosslink repair. *Environmental and Molecular Mutagenesis, 51*(6), 552-566.
(*210*) Stevens, K., Claeys, D. D., Catak, S., Figaroli, S., Hocek, M., Tromp, J. M., et al. (2011). Furan-oxidation-triggered inducible DNA cross-linking: Acyclic versus cyclic furan-containing building blocks--on the benefit of restoring the cyclic sugar backbone. *Chemistry (Weinheim an Der Bergstrasse, Germany), 17*(25), 6940-6953.
(*211*) Hattori, K., Hirohama, T., Imoto, S., Kusano, S., & Nagatsugi, F. (2009). Formation of highly selective and efficient interstrand cross-linking to thymine without photo-irradiation. *Chemical Communications (Cambridge, England), (42):6463-5. doi*(42), 6463-6465.
(*212*) Hentschel, S., Alzeer, J., Angelov, T., Scharer, O. D., & Luedtke, N. W. (2012). Synthesis of DNA interstrand cross-links using a photocaged nucleobase. *Angewandte Chemie (International Ed.in English), 51*(14), 3466-3469.
(*213*) Hlavin, E. M., Smeaton, M. B., Noronha, A. M., Wilds, C. J., & Miller, P. S. (2010). Cross-link structure affects replication-independent DNA interstrand cross-link repair in mammalian cells. *Biochemistry, 49*(18), 3977-3988.
(*214*) McManus, F. P., Fang, Q., Booth, J. D., Noronha, A. M., Pegg, A. E., & Wilds, C. J. (2010). Synthesis and characterization of an O(6)-2'-deoxyguanosine-alkyl-O(6)-2'-deoxyguanosine interstrand cross-link in a 5'-GNC motif and repair by human O(6)-alkylguanine-DNA alkyltransferase. *Organic & Biomolecular Chemistry, 8*(19), 4414-4426.
(*215*) Waldman, A. S. (2008). Ensuring the fidelity of recombination in mammalian chromosomes. *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology, 30*(11-12), 1163-1171.
(*216*) Goodman, M. F., Creighton, S., Bloom, L. B., & Petruska, J. (1993). Biochemical basis of DNA replication fidelity. *Critical Reviews in Biochemistry and Molecular Biology, 28*(2), 83-126.
(*217*) Kunkel, T. A., & Bebenek, K. (2000). DNA replication fidelity. *Annual Review of Biochemistry, 69*, 497-529.
(*218*) Lindahl, T. (1993). Instability and decay of the primary structure of DNA. *Nature, 362*(6422), 709-715.
(*219*) Bhattacharyya, A., & Lilley, D. M. (1989). Single base mismatches in DNA. long- and short-range structure probed by analysis of axis trajectory and local chemical reactivity. *Journal of Molecular Biology, 209*(4), 583-597.
(*220*) Rojsitthisak, P., Jongaroonngamsang, N., Romero, R. M., & 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*(6), e20745.
(*221*) Xu, Y. Z., & Swann, P. F. (1990). A simple method for the solid phase synthesis of oligodeoxynucleotides containing O4-alkylthymine. *Nucleic Acids Research, 18*(14), 4061-4065.

(*222*) Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., et al. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallographica.Section D, Biological Crystallography, 54*(Pt 5), 905-921.
(*223*) Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., & Karplus, M. (1983). CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J Comp Chem. 4*(2), 187–217.
(*224*) Wilds, C. J., Palus, E., & Noronha, A. M. (2007). An approach for the synthesis of duplexes containing N^3T-butyl-N^3T interstrand cross-links via a bisphosphoramidite strategy. *Can. J. Chem.* 85, 249-256.
(*225*) Wilds, C. J., Noronha, A. M., Robidoux, S., & 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. *Journal of the American Chemical Society, 126*(30), 9257-9265.
(*226*) Hofr, C., & Brabec, V. (2005). Thermal stability and energetics of 15-mer DNA duplex interstrand crosslinked by trans-diamminedichloroplatinum(II). *Biopolymers, 77*(4), 222-229.
(*227*) Fabrega, C., Eritja, R., Sinha, N. D., Dosanjh, M. K., & Singer, B. (1995). Synthesis and properties of oligonucleotides containing the mutagenic base O4-benzylthymidine. *Bioorganic & Medicinal Chemistry, 3*(1), 101-108.
(*228*) Egli, M., & Pallan, P. S. (2010). The many twists and turns of DNA: Template, telomere, tool, and target. *Current Opinion in Structural Biology, 20*(3), 262-275.
(*229*) Corneillie, T. M., Whetstone, P. A., Lee, K. C., Wong, J. P., & Meares, C. F. (2004). Converting weak binders into infinite binders. *Bioconjugate Chemistry, 15*(6), 1389-1391.
(*230*) Niemeyer, C. M. (2002). The developments of semisynthetic DNA-protein conjugates. *Trends in Biotechnology, 20*(9), 395-401.
(*231*) Niemeyer, C. M. (2010). Semisynthetic DNA-protein conjugates for biosensing and nanofabrication. *Angewandte Chemie (International Ed.in English), 49*(7), 1200-1216.
(*232*) Staker, B. L., Hjerrild, K., Feese, M. D., Behnke, C. A., Burgin, A. B., Jr, and Stewart, L. (2002) The Mechanism of Topoisomerase I Poisoning by a Camptothecin Analog. *Proc. Natl. Acad. Sci. U. S. A. 99*, 15387-15392.
(*233*) Golan, G., Zharkov, D. O., Grollman, A. P., Dodson, M. L., McCullough, A. K., Lloyd, R. S., et al. (2006). Structure of T4 pyrimidine dimer glycosylase in a reduced imine covalent complex with abasic site-containing DNA. *Journal of Molecular Biology, 362*(2), 241-258.
(*234*) Keppler, A., Gendreizig, S., Gronemeyer, T., Pick, H., Vogel, H., & Johnsson, K. (2003). A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nature Biotechnology, 21*(1), 86-89.
(*235*) McManus, F. P., O'Flaherty, D. K., Noronha, A. M., & Wilds, C. J. (2012). O4-alkyl-2'-deoxythymidine cross-linked DNA to probe recognition and repair by O6-alkylguanine DNA alkyltransferases. *Organic & Biomolecular Chemistry, 10*(35), 7078-7090.
(*236*) Liu, L., Pegg, A. E., Williams, K. M., & Guengerich, F. P. (2002). Paradoxical enhancement of the toxicity of 1,2-dibromoethane by O6-alkylguanine-DNA alkyltransferase. *The Journal of Biological Chemistry, 277*(40), 37920-37928.
(*237*) Noronha, A. M., Noll, D. M., Wilds, C. J., & Miller, P. S. (2002). N(4)C-ethyl-N(4)C cross-linked DNA: Synthesis and characterization of duplexes with interstrand cross-links of different orientations. *Biochemistry, 41*(3), 760-771.
(*238*) Webba da Silva, M. W., Bierbryer, R. G., Wilds, C. J., Noronha, A. M., Colvin, O. M., Miller, P. S., et al. (2005). Intrastrand base-stacking buttresses widening of major groove in interstrand cross-linked B-DNA. *Bioorganic & Medicinal Chemistry, 13*(14), 4580-4587.
(*239*) Pegg, A. E., Morimoto, K., & Dolan, M. E. (1988). Investigation of the specificity of O6-alkylguanine-DNA-alkyltransferase. *Chemico-Biological Interactions, 65*(3), 275-281.

(*240*) Federwisch, M., Hassiepen, U., Bender, K., Rajewsky, M. F., & Wollmer, A. (1997). Recombinant human O6-alkylguanine-DNA alkyltransferase induces conformational change in bound DNA. *FEBS Letters, 407*(3), 333-336.
(*241*) Webba da Silva, M., Noronha, A. M., Noll, D. M., Miller, P. S., Colvin, O. M., & Gamcsik, M. P. (2002). Solution structure of a DNA duplex containing mispair-aligned N4C-ethyl-N4C interstrand cross-linked cytosines. *Biochemistry, 41*(51), 15181-15188.
(*242*) Lawley, P. D., Orr, D. J., Shah, S. A., Farmer, P. B., & Jarman, M. (1973). Reaction products from N-methyl-N-nitrosourea and deoxyribonucleic acid containing thymidine residues. synthesis and identification of a new methylation product, O4-methylthymidine. *The Biochemical Journal, 135*(1), 193-201.
(*243*) Singer, B. (1986). O-alkyl pyrimidines in mutagenesis and carcinogenesis: Occurrence and significance. *Cancer Research, 46*(10), 4879-4885.
(*244*) Toorchen, D., & Topal, M. D. (1983). Mechanisms of chemical mutagenesis and carcinogenesis: Effects on DNA replication of methylation at the O6-guanine position of dGTP. *Carcinogenesis, 4*(12), 1591-1597.
(*245*) Duckett, D. R., Drummond, J. T., Murchie, A. I., Reardon, J. T., Sancar, A., Lilley, D. M., et al. (1996). Human MutSalpha recognizes damaged DNA base pairs containing O6-methylguanine, O4-methylthymine, or the cisplatin-d(GpG) adduct. *Proceedings of the National Academy of Sciences of the United States of America, 93*(13), 6443-6447.
(*246*) Kang, H., Konishi, C., Kuroki, T., & Huh, N. (1995). Detection of O6-methylguanine, O4-methylthymine and O4-ethylthymine in human liver and peripheral blood leukocyte DNA. *Carcinogenesis, 16*(6), 1277-1280.
(*247*) Gerson, S. L. (2004). MGMT: Its role in cancer aetiology and cancer therapeutics. *Nature Reviews.Cancer, 4*(4), 296-307.
(*248*) Xu-Welliver, M., & Pegg, A. E. (2002). Degradation of the alkylated form of the DNA repair protein, O(6)-alkylguanine-DNA alkyltransferase. *Carcinogenesis, 23*(5), 823-830.
(*249*) McManus, F. P., Khaira, A., Noronha, A. M., & Wilds, C. J. (2013). Preparation of covalently linked complexes between DNA and O(6)-alkylguanine-DNA alkyltransferase using interstrand cross-linked DNA. *Bioconjugate Chemistry,* DOI. 10.1021/bc300553u
(*250*) Encell, L. P., & Loeb, L. A. (2000). Enhanced in vivo repair of O(4)-methylthymine by a mutant human DNA alkyltransferase. *Carcinogenesis, 21*(7), 1397-1402.
(*251*) Bryksin, A., & Matsumura, I. (2010). Overlap extension PCR cloning: A simple and reliable way to create recombinant plasmids. *BioTechniques, 48*(6), 463-465.
(*252*) Huh, N., & Rajewsky, M. F. (1986). Enzymatic elimination of O6-ethylguanine and stability of O4-ethylthymine in the DNA of malignant neural cell lines exposed to N-ethyl-N-nitrosourea in culture. *Carcinogenesis, 7*(3), 435-439.
(*253*) McManus, F. P., & Wilds, C. J., (2013) Engineering of a O(6)-alkylguanine-DNA alkyltransferase chimera and repair of O(4)-alkyl thymidine adducts and O(6)-alkylene-2′-deoxyguanosine cross-linked DNA. *Toxicology Research*, DOI: 10.1039/c2tx20075a
(*254*) Noll, D. M., Noronha, A. M., & Miller, P. S. (2001). Synthesis and characterization of DNA duplexes containing an N(4)C-ethyl-N(4)C interstrand cross-link. *Journal of the American Chemical Society, 123*(15), 3405-3411.
(*255*) Doherty, A. J., Ashford, S. R., Brannigan, J. A., & Wigley, D. B. (1995). A superior host strain for the over-expression of cloned genes using the T7 promoter based vectors. *Nucleic Acids Research, 23*(11), 2074-2075.
(*256*) Li, B. F., Reese, C. B., & Swann, P. F. (1987). Synthesis and characterization of oligodeoxynucleotides containing 4-O-methylthymine. *Biochemistry, 26*(4), 1086-1093.
(*257*) Borowy-Borowski, H., & Chambers, R. W. (1989). Solid-phase synthesis and side reactions of oligonucleotides containing O-alkylthymine residues. *Biochemistry, 28*(4), 1471-1477.

(*258*) Middleton, M. R., & Margison, G. P. (2003). Improvement of chemotherapy efficacy by inactivation of a DNA-repair pathway. *The Lancet Oncology, 4*(1), 37-44.
(*259*) Hlavin, E. M., Smeaton, M. B., & Miller, P. S. (2010). Initiation of DNA interstrand cross-link repair in mammalian cells. *Environmental and Molecular Mutagenesis, 51*(6), 604-624.
(*260*)Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., & Bax, A. (1995). NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR, 6*(3), 277-293.
(*261*) Johnson, B. A., & Blevins, R. A. (1994). NMR view: A computer program for the visualization and analysis of NMR data. *Journal of Biomolecular NMR, 4*(5), 603-614.
(*262*) Denisov, A. Y., Noronha, A. M., Wilds, C. J., Trempe, J. F., Pon, R. T., Gehring, K., et al. (2001). Solution structure of an arabinonucleic acid (ANA)/RNA duplex in a chimeric hairpin: Comparison with 2'-fluoro-ANA/RNA and DNA/RNA hybrids. *Nucleic Acids Research, 29*(21), 4284-4293.
(*263*) Denisov, A. Y., Zamaratski, E. V., Maltseva, T. V., Sandstrom, A., Bekiroglu, S., Altmann, K. H., et al. (1998). The solution conformation of a carbocyclic analog of the dickerson-drew dodecamer: Comparison with its own X-ray structure and that of the NMR structure of the native counterpart. *Journal of Biomolecular Structure & Dynamics, 16*(3), 547-568.
(*264*) Kim, S. G., Lin, L. J., & Reid, B. R. (1992). Determination of nucleic acid backbone conformation by 1H NMR. *Biochemistry, 31*(14), 3564-3574.
(*265*) Lavery, R., Moakher, M., Maddocks, J. H., Petkeviciute, D., & Zakrzewska, K. (2009). Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Research, 37*(17), 5917-5929.
(*266*) Blanchet, C., Pasi, M., Zakrzewska, K., & Lavery, R. (2011). CURVES+ web server for analyzing and visualizing the helical, backbone and groove parameters of nucleic acid structures. *Nucleic Acids Research, 39*(Web Server issue), W68-73.
(*267*) Wuthrich, K. (1986) NMR of Proteins and Nucleic Acids, John Wiley & Sons, New York.
(*268*) Saenger, W. (1984) Principles of Nucleic Acids Structure, Springer-Verlag, New York.
(*269*) Webba da Silva, M., Wilds, C. J., Noronha, A. M., Colvin, O. M., Miller, P. S., & Gamcsik, M. P. (2004). Accommodation of mispair aligned N3T-ethyl-N3T DNA interstrand cross link. *Biochemistry, 43*(39), 12549-12554.
(*270*) Dickerson, R. E., & Chiu, T. K. (1997). Helix bending as a factor in protein/DNA recognition. *Biopolymers, 44*(4), 361-403.
(*271*) Cohen, S. M., Jamieson, E. R., & Lippard, S. J. (2000). Enhanced binding of the TATA-binding protein to TATA boxes containing flanking cisplatin 1,2-cross-links. *Biochemistry, 39*(28), 8259-8265.
(*272*) Camenisch, U., Dip, R., Schumacher, S. B., Schuler, B., & Naegeli, H. (2006). Recognition of helical kinks by xeroderma pigmentosum group A protein triggers DNA excision repair. *Nature Structural & Molecular Biology, 13*(3), 278-284.
(*273*) Armstrong, D. K., Gordon, G. B., Hilton, J., Streeper, R. T., Colvin, O. M., & Davidson, N. E. (1992). Hepsulfam sensitivity in human breast cancer cell lines: The role of glutathione and glutathione S-transferase in resistance. *Cancer Research, 52*(6), 1416-1421.
(*274*) Rannug, U., Sundvall, A., & Ramel, C. (1978). The mutagenic effect of 1,2-dichloroethane on salmonella typhimurium I. activation through conjugation with glutathion in vitro. *Chemico-Biological Interactions, 20*(1), 1-16.
(*275*) Kokatla, H. P., & Lakshman, M. K. (2010). One-pot etherification of purine nucleosides and pyrimidines. *Organic Letters, 12*(20), 4478-4481.
(*276*) Lakshman, M. K., & Frank, J. (2009). A simple method for C-6 modification of guanine nucleosides. *Organic & Biomolecular Chemistry, 7*(14), 2933-2940.
(*277*) Citron, M., Decker, R., Chen, S., Schneider, S., Graver, M., Kleynerman, L., et al. (1991). O6-methylguanine-DNA methyltransferase in human normal and tumor tissue from brain, lung, and ovary. *Cancer Research, 51*(16), 4131-4134.

(*278*) Povey, A. C., Hall, C. N., Cooper, D. P., O'Connor, P. J., & Margison, G. P. (2000). Determinants of O(6)-alkylguanine-DNA alkyltransferase activity in normal and tumour tissue from human colon and rectum. *International Journal of Cancer.Journal International Du Cancer, 85*(1), 68-72.
(*279*) Belanich, M., Randall, T., Pastor, M. A., Kibitel, J. T., Alas, L. G., Dolan, M. E., et al. (1996). Intracellular localization and intercellular heterogeneity of the human DNA repair protein O(6)-methylguanine-DNA methyltransferase. *Cancer Chemotherapy and Pharmacology, 37*(6), 547-555.

APPENDIX I: First page to published manuscripts

CHAPTER II

PAPER www.rsc.org/obc | Organic & Biomolecular Chemistry

Synthesis and characterization of an O^6-2′-deoxyguanosine-alkyl-O^6-2′-deoxyguanosine interstrand cross-link in a 5′-GNC motif and repair by human O^6-alkylguanine-DNA alkyltransferase†

Francis P. McManus,[a] Qingming Fang,[b] Jason D. M. Booth,[a] Anne M. Noronha,[a] Anthony E. Pegg[b] and Christopher J. Wilds*[a]

Received 7th May 2010, Accepted 28th June 2010
DOI: 10.1039/c0ob00093k

O^6-2′-Deoxyguanosine-alkyl-O^6-2′-deoxyguanosine interstrand DNA cross-links (ICLs) with a four and seven methylene linkage in a 5′-GNC- motif have been synthesized and their repair by human O6-alkylguanine-DNA alkyltransferase (hAGT) investigated. Duplexes containing 11 base-pairs with the ICLs in the center were assembled by automated DNA solid-phase synthesis using a cross-linked 2′-deoxyguanosine dimer phosphoramidite, prepared *via* a seven step synthesis which employed the Mitsunobu reaction to introduce the alkyl lesion at the O^6 atom of guanine. Introduction of the four and seven carbon ICLs resulted in no change in duplex stability based on UV thermal denaturation experiments compared to a non-cross-linked control. Circular dichroism spectra of these ICL duplexes exhibited features of a B-form duplex, similar to the control, suggesting that these lesions induce little overall change in structure. The efficiency of repair by hAGT was examined and it was shown that hAGT repairs both ICL containing duplexes, with the heptyl ICL repaired more efficiently relative to the butyl cross-link. These results were reproducible with various hAGT mutants including one that contains a novel V148L mutation. The ICL duplexes displayed similar binding affinities to a C145S hAGT mutant compared to the unmodified duplex with the seven carbon containing ICLs displaying slightly higher binding. Experiments with CHO cells to investigate the sensitivity of these cells to busulfan and hepsulfam demonstrate that hAGT reduces the cytotoxicity of hepsulfam suggesting that the O^6-2′-deoxyguanosine-alkyl-O^6-2′-deoxyguanosine interstrand DNA cross-link may account for at least part of the cytotoxicity of this agent.

Introduction

Interstrand cross-links (ICLs) that are formed in DNA as a consequence of the action of bifunctional alkylating agents represent some of the most toxic lesions encountered by cells due to the obstruction of unwinding of the two strands, critical to the processes of DNA replication, transcription and recombination.[1] Interference with these key cellular processes by the presence of ICLs is the basis of the mechanism of action of bifunctional alkylating agents, such as mechlorethamine, that are used as cancer therapeutics.[2] However, the potency of these agents is reduced by the ability of cancer cells to repair the lesions resulting in an overall resistance to therapy.

In eukaryotic cells the repair of ICLs is a complex process with numerous repair pathways including nucleotide excision repair, homologous recombination and non-homologous end joining implicated in the removal of the damage.[3] Understanding the molecular basis of how ICL repair occurs will play an important role towards the development of new chemotherapeutic agents that may evade this process, thus increasing the efficacy of these drugs.

One approach employed in elucidating the roles various DNA repair pathways contribute in removing DNA ICLs involves the use of chemically synthesized oligonucleotides that contain representative lesions formed by bifunctional alkylating chemotherapeutics.[4,5,6] These oligonucleotides are designed to contain lesions linking specific atoms in DNA in well defined orientations and can be incorporated into plasmids for DNA repair experiments. This approach uses solid-phase synthesis which generates sufficient amounts of ICL DNA to enable structural studies and repair assays.[7,8,9,10,11]

Some of these ICLs are challenging to prepare synthetically. For example, the bifunctional alkylating agent hepsulfam (1,7-heptanediol disulfamate) which has been investigated clinically,[12] has been shown to form a cross-link between the N^7 atoms of guanines in 5′-GNC sequences (Fig. 1) as demonstrated through the use of mass spectrometry and identification of 1,7-bis(guanyl)heptane.[13] N^7-alkylated guanines are chemically unstable, for example the presence of N^7-methylguanine can result in an apurinic site in DNA or undergo ring opening to yield the ring opened formamido pyrimidine (Fapy) derivative.[14,15]

[a]*Department of Chemistry and Biochemistry, Concordia University, 7141 Sherbrooke St. West, Montréal, QC, Canada, H4B 1R6. E-mail: cwilds@alcor.concordia.ca; Fax: 514-848-2868; Tel: 514-848-2424 ext. 5798*
[b]*Departments of Cellular and Molecular Physiology and Pharmacology, The Pennsylvania State University College of Medicine, PO Box 850, Hershey, PA, USA, 17033*

† Electronic supplementary information (ESI) available: ^{1}H NMR and ^{31}P NMR of compounds **6a**, **6b**, **7a** and **7b**, HPLC chromatographs and MS spectra for oligonucleotide characterization and repair and binding data of hAGT to the ICL. See DOI: 10.1039/c0ob00093k

CHAPTER III

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

CHAPTER IV

Bioconjugate Chemistry

Article

pubs.acs.org/bc

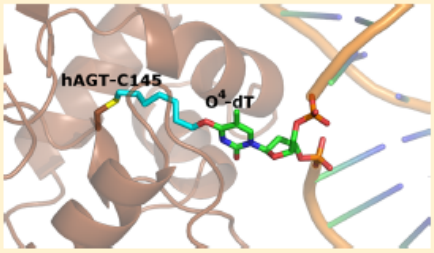
Preparation of Covalently Linked Complexes Between DNA and O^6-Alkylguanine-DNA Alkyltransferase Using Interstrand Cross-Linked DNA

Francis P. McManus,[‡] Amardeep Khaira,[‡] Anne M. Noronha,[‡] and Christopher J. Wilds*[,‡]

[‡]Department of Chemistry and Biochemistry, Concordia University, 7141 Sherbrooke St. West, Montréal, QC, Canada H4B 1R6

Supporting Information

ABSTRACT: O^6-alkylguanine-DNA alkyltransferases (AGT) are responsible for the removal of alkylation at both the O^6 atom of guanine and O^4 atom of thymine. AGT homologues show vast substrate differences with respect to the size of the adduct and which alkylated atoms they can restore. The human AGT (hAGT) has poor capabilities for removal of methylation at the O^4 atom of thymidine, which is not the case in most homologues. No structural data are available to explain this poor hAGT repair. We prepared and characterized O^6G-butylene-O^4T (XLGT4) and O^6G-heptylene-O^4T (XLGT7) interstrand cross-linked (ICL) DNA as probes for hAGT and the *Escherichia coli* homologues, OGT and Ada-C, for the formation of DNA-AGT covalent complexes. XLGT7 reacted only with hAGT and did so with a cross-linking efficiency of 25%, while XLGT4 was inert to all AGT tested. The hAGT mediated repair of XLGT7 occurred slowly, on the order of hours as opposed to the repair of O^6-methyl-2′-deoxyguanosine which requires seconds. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis of the repair reaction revealed the formation of a covalent complex with an observed migration in accordance with a DNA-AGT complex. The identity of this covalent complex, as determined by mass spectrometry, was composed of a heptamethylene bridge between the O^4 atom of thymidine (in an 11-mer DNA strand) to residue Cys145 of hAGT. This procedure can be applied to produce well-defined covalent complexes between AGT with DNA.



INTRODUCTION

Modified nucleic acids with novel properties have had a long-standing interest and application in the field of bioorganic chemistry.[1] One example of how these probes can be employed involves the formation of covalent complexes with proteins where weak binding was previously observed leading to a wide range of applications.[2−4]

Formation of covalent complexes for crystallographic purposes can be employed for various systems due to the control one can have on the nature of the chemical modification, which can be tailored to specific protein complexes.[5,6] For example, the structure of hAGT complexed with DNA has been determined with the aid of modified nucleic acids.[7,8] One of these structures, significant in enhancing our understanding of the mechanism of hAGT, contains a chemical cross-link formed between the hAGT active site and DNA containing the modified nucleoside N1,O^6-ethanoxanthosine.

AGT proteins, which are found in all kingdoms of life, are responsible for the repair of the mutagenic O^6-methylguanosine (O^6 MedG) and O^4-methylthymidine (O^4 MedT) lesions.[9−11] The mutagenicity of these lesions rise from the wobble base pair that these modified bases adopt with respect to their natural compliments. These wobble base pairs adversely affect the processivity of DNA polymerase during DNA replication. O^6 MedG adopts a nonwobble base pair with dT causing a GC to AT transition, while O^4 MedT adopts a nonwobble base pair with dG causing a TA to CG transition. These two modifications account for 8% of the total DNA damage caused by *N*-methyl-*N*-nitrosourea (MNU).[12] Endogenously, O^6 MedG is formed at a rate of 10−30 events daily by *S*-adenosylmethionine.[13]

The formation of O^6 MedG is detrimental to cells where the bioaccumulation of 6 650 O^6 MedG modifications in hAGT deficient tumor cells is sufficient to cause cell lethality.[14] Unlike O^6 MedG, O^4 MedT is affected by neither the levels of hAGT nor mismatch repair (MMR), which is why this lesion is more mutagenic in mammalian cells.[15,16] Processing of O^4 MedT in mammalian cells is believed to rely solely on the nucleotide excision repair (NER) pathway.[17]

hAGT and Ada-C, the most thoroughly characterized AGT proteins, repair alkylated DNA by flipping the damaged base out of the DNA duplex and into the active site where the alkyl group is transferred from the point of lesion to the active site-

Received: October 9, 2012
Revised: December 19, 2012

CHAPTER V

Toxicology Research

RSC Publishing

COMMUNICATION

View Article Online
View Journal

Engineering of a O^6-alkylguanine-DNA alkyltransferase chimera and repair of O^4-alkyl thymidine adducts and O^6-alkylene-2′-deoxyguanosine cross-linked DNA†

Cite this: DOI: 10.1039/c2tx20075a

Received 12th October 2012,
Accepted 28th December 2012

DOI: 10.1039/c2tx20075a

www.rsc.org/toxicology

Francis P. McManus and Christopher J. Wilds*

A soluble human O^6-alkylguanine-DNA alkyltransferase (hAGT) chimera was engineered containing the active site of OGT (residues 139–159) and an additional S134P mutation. The resulting hAGT chimera not only retained hAGT's ability to repair bulky O^6-alkylene-2′-deoxyguanosine interstrand cross-linked DNA damage but also displayed enhanced repair of various O^4-alkyl thymidine adducts.

O^4-Methyl-thymidine (O^4MeT) in the genome that results from exposure to tobacco derived *N*-nitrosamines or chemotherapeutic agents is highly mutagenic because of its ability to form altered hydrogen bonding patterns.[1] The stable O^4MeT : dG pair resulting from a single round of replication, causes a dT to dC transition mutation, which is especially detrimental to the cell if this event occurs on proto-oncogenes or tumour suppressor genes.[2]

O^6-Methyl-2′-deoxyguanosine (O^6MeG), another well known mutagenic nucleobase modification, generates dG : dC to dA : dT transition.[3] Unlike its thymidine counter-part the O^6MeG : dT intermediate (formed after a single round of replication) is a substrate for mismatch repair and direct repair in humans.[4] The O^4MeT : dG mismatch persists in the cell since it is neither a good substrate for mismatch repair nor direct repair, contributing to the mutagenic potential of O^4MeT.[5] O^4-Alkyl thymidine, which is believed to be a minor lesion, has been found at similar levels as O^6MeG in the liver tissue of healthy volunteers, demonstrating its importance and prominence in the absence of exogenous alkylating agents.[6]

O^6-Alkylguanine-DNA alkyltransferases (AGT) are part of the direct repair pathway and are responsible for the removal of both O^6-alkyl dG and O^4-alkyl dT adducts.[7] The repair occurs in a single step process where the active site Cys thiolate anion performs a nucleophilic attack on the α-carbon of the adduct located on the exocyclic oxygen.[8] In the process the native DNA is restored and the AGT protein is irreversibly alkylated where it eventually undergoes degradation by the ubiquitin mediated pathway.[9] AGT homologues show vast substrate differences. hAGT has been studied extensively and is believed to possess the greatest substrate range for O^6-alkyl dG adducts. Among some of the "bulkier" substrates that have been shown to undergo repair by this protein are O^6-2′-deoxyguanosine-alkylene-O^6-2′-deoxyguanosine interstrand cross-links (ICL), which are DNA lesions that covalently link complementary DNA strands obstructing DNA unwinding and segregation, in the process inhibiting cell proliferation due to physical obstruction.[10]

Escherichia coli OGT is believed to be the most efficient AGT at eliminating O^4-alkyl dT.[11] This is contrary to hAGT, which is extremely poor at removing even O^4MeT. hAGT recognizes and binds O^4MeT but cannot remove them efficiently, physically shielding the lesion from the nucleotide excision repair (NER) machinery and increasing the lesion's toxicity.[12]

The Loeb and Pegg groups have dedicated some effort to generate hAGT variants capable of increased repair of O^4MeT. The Loeb laboratory employed a random sequence mutagenesis approach to generate multiple variants followed by a functional complementation assay to obtain a variant with 8 point mutations (C150Y, S152R, A154S, V155G, N157T, V164M, E166Q, and A170T).[13] This variant showed a rate of repair of O^4MeT that was roughly 11.5 times greater than that of hAGT.[14] A horizontal gene transfer from OGT to hAGT was invoked by the Pegg group to generate their hAGT–OGT chimera. Their most promising construct harboured 8 mutations (V149I, C150G, S151R, S152N, A154T, V155M, G156T and N157G), which was achieved by substituting the hAGT amino acids by their respective residues in OGT. Unfortunately, incorporating numerous successive mutations generated a chimeric protein that was not soluble, requiring refolding procedures to be adopted. Their chimera displayed valuable properties when expressed in alkyltransferase and NER deficient *E. coli* cells, where reduced levels of mutations

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

†Electronic supplementary information (ESI) available: Chimera engineering and biophysical characterization, ICL repair and binding data. See DOI: 10.1039/c2tx20075a

APPENDIX II: Supporting information to CHAPTER II

O^6-2'-deoxyguanosine-alkylene-O^6-2'-deoxyguanosine interstrand cross-link in a 5'-GNC motif and repair by human O^6-alkylguanine-DNA alkyltransferase

Francis P. McManus, Qingming Fang, Jason D. M. Booth, Anne M. Noronha, Anthony E. Pegg and Christopher J. Wilds*

Figure A2.1: Absorbance (A_{260}) versus temperature profiles of non cross-linked duplex (_____) 5'-dCGATGTCATCG-3'/5'-dCGATGACATCG-3', cross-linked duplex **XLGG4** (_ _ _ _), and cross-linked duplex **XLGG7** (……..). Solutions containing a total strand concentration of 2.8 μM for the cross-linked (**XLGG4** and **XLGG7**) and non-cross-linked control duplexes in 90 mM sodium chloride, pH = 7.0, 10 mM sodium phosphate, and 1 mM EDTA buffer, were heated at 0.5°C/min.

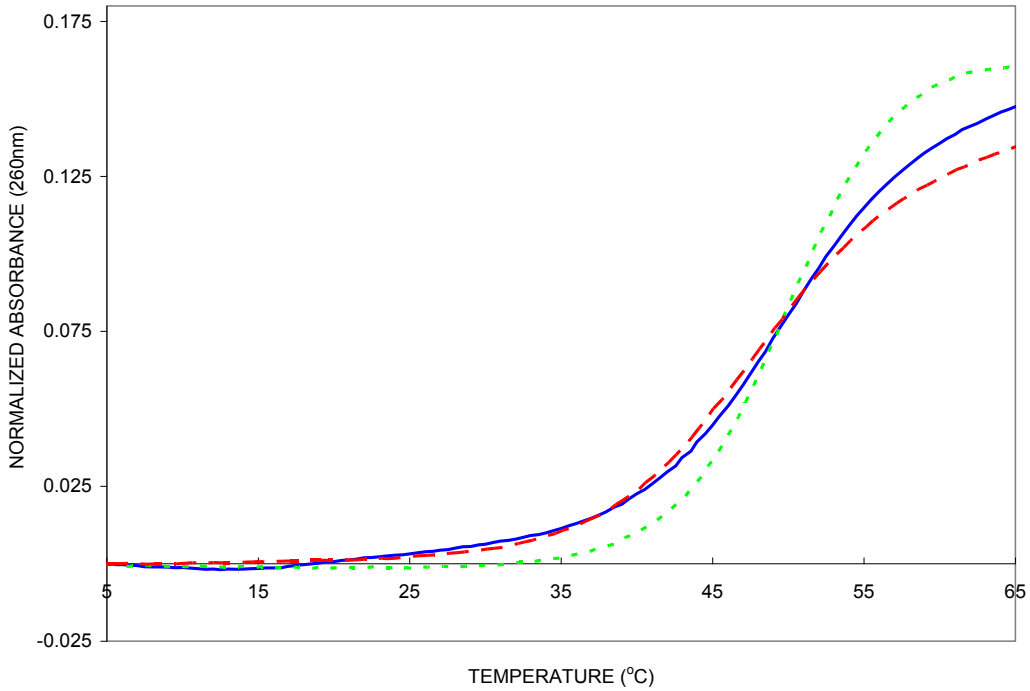


Figure A2.2: CD spectra of non cross-linked duplex (_____) 5'- dCGATGTCATCG-3'/5'-dCGATGACATCG3', cross-linked duplex **XLGG4** (_ _ _ _), and cross-linked duplex **XLGG7** (……..). Solutions containing a total strand concentration of 2.8 μM for the cross-linked duplex **XLGG 4**,**7** and 2.8 μM of the non-crosslinked control duplexes in 10 mM sodium phosphate, pH 7.0, 90 mM sodium chloride, and 1 mM EDTA. Spectra are the average of 5 scans and were recorded at 10 ºC.

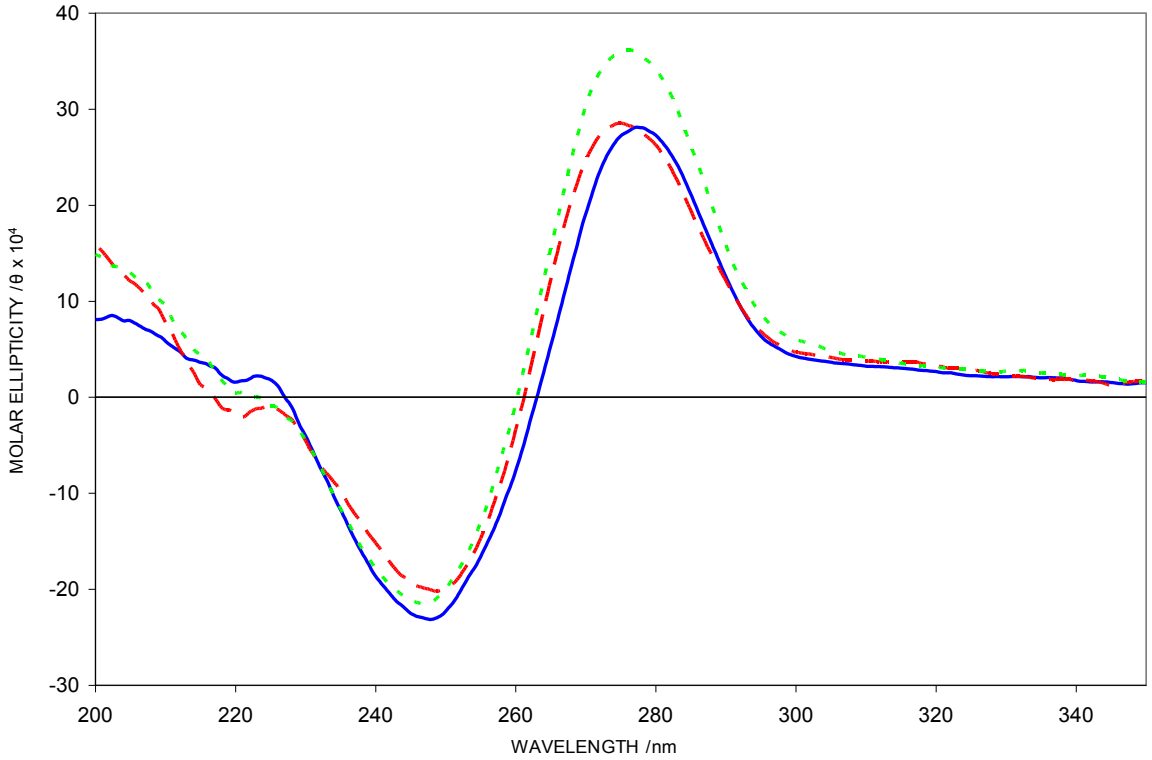
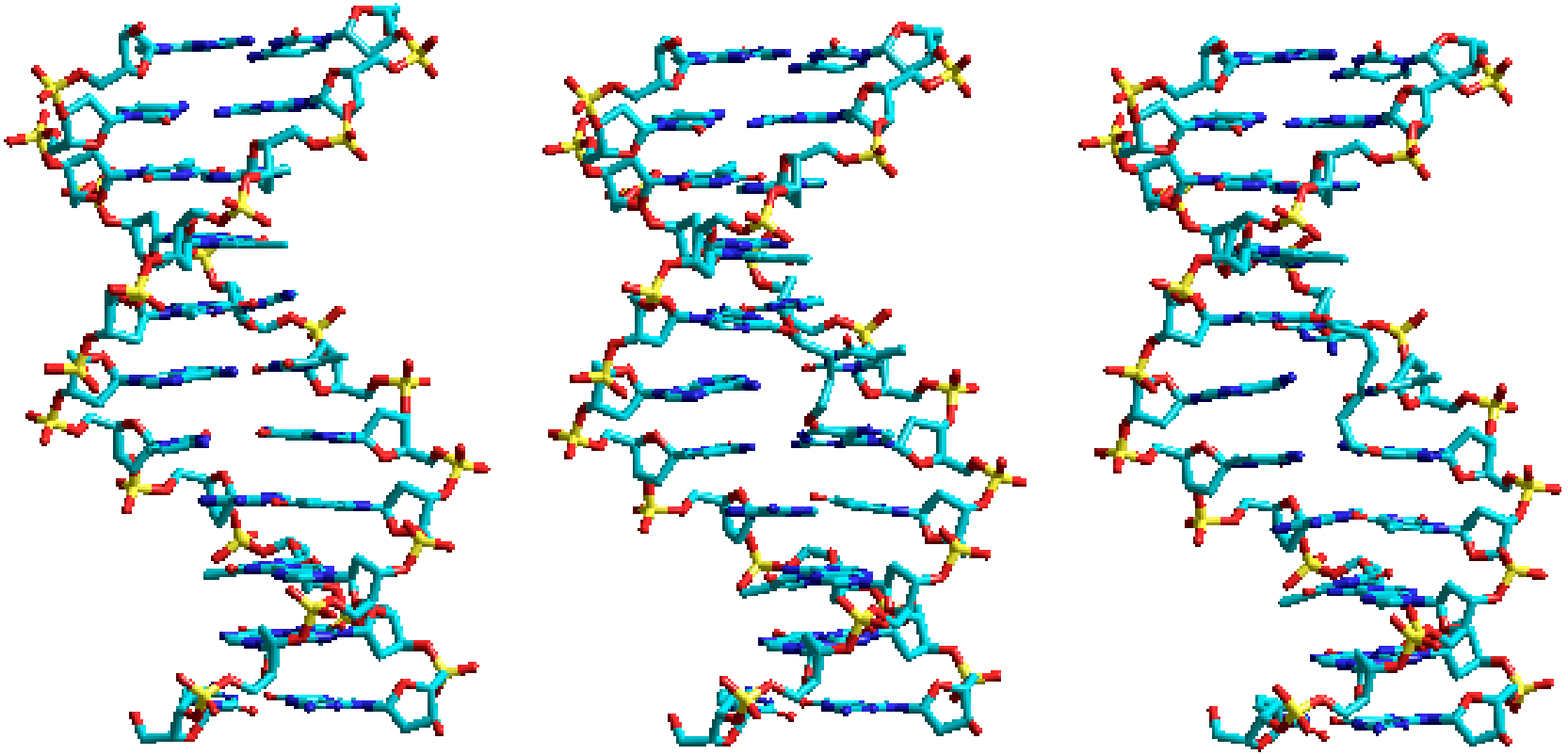


Figure A2.3: Molecular models of non cross-linked control duplex and cross-linked duplexes **XLGG4** and **XLGG7** that were geometry optimized using the AMBER forcefield.



non cross-linked control **XLGG4** **XLGG7**

Figure A2.4: Effects of mutations on secondary structure of hAGT by CD. Scans were taken with 5 µM wild-type hAGT (_____), C145S (__ _ _ __), P140A (__ _ __) and V148L (__ __) between 260 and 200 nm in CD buffer (260-203 nm shown due to high voltage below 203nm).

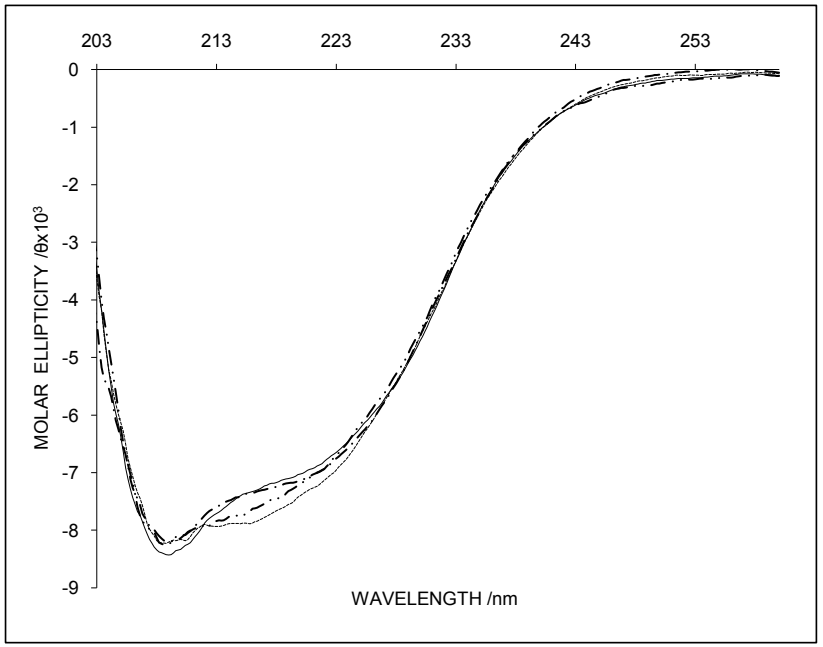
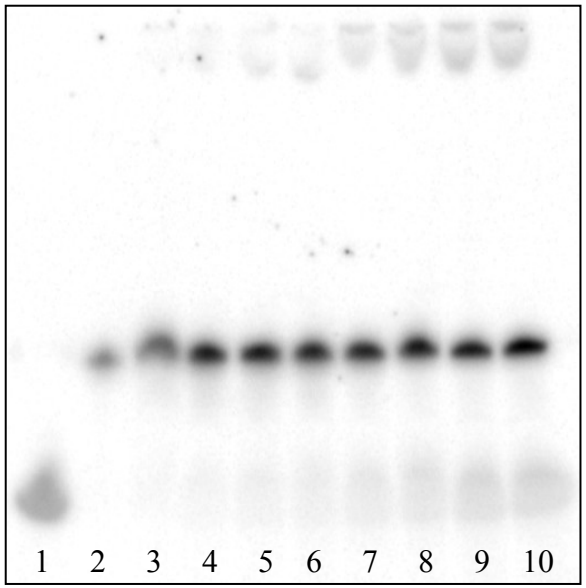


Figure A2.5: Time course repair of **XLGG4** and **XLGG7** by hAGT. (**A**) Denaturing gel of the repair of 2 pmol of **XLGG4** by 60 pmol hAGT as a function of time: lane 1, 2 pmol Control; lanes 2-10, 2 pmol **XLGG4** + 60 pmol hAGT incubated for 0, 5, 15, 30, 60, 120, 240, 510 and 540 min, respectively (**B**) Denaturing gel of the repair of 2 pmol of **XLGG7** by 60 pmol hAGT as a function of time: lane 1, 2 pmol Control; lanes 2-10, 2 pmol **XLGG7** + 60 pmol hAGT incubated for 0, 1, 2, 5, 10, 15, 30, 60, 120 and 180 min, respectively.

(**A**)



(**B**)

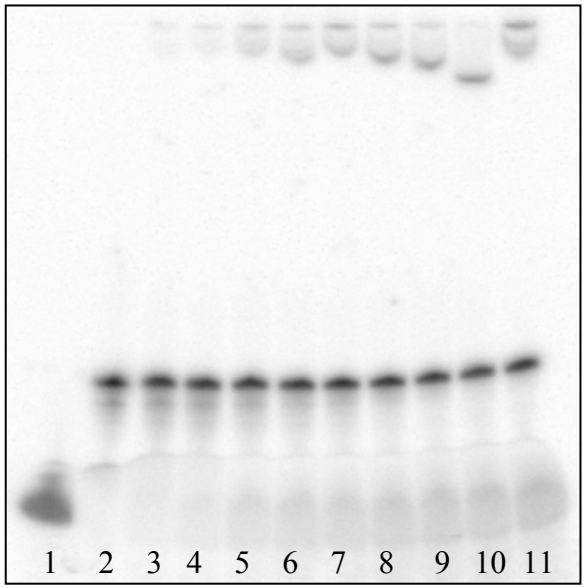
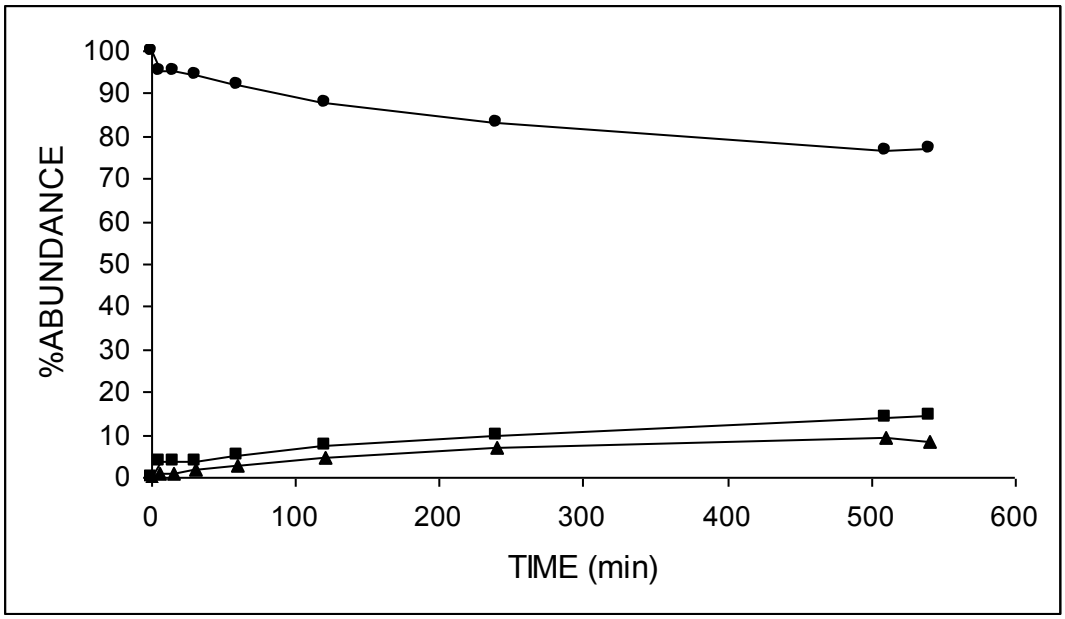


Figure A2.6: Graphic representation of the % abundance of each species in the time course repair by hAGT of (**A**) **XLGG4** and (**B**) **XLGG7**. hAGT-DNA complex/partially repaired product (▲); free DNA/ fully repaired product (■); ICL/ unrepaired substrate (●).

(**A**)



(**B**)

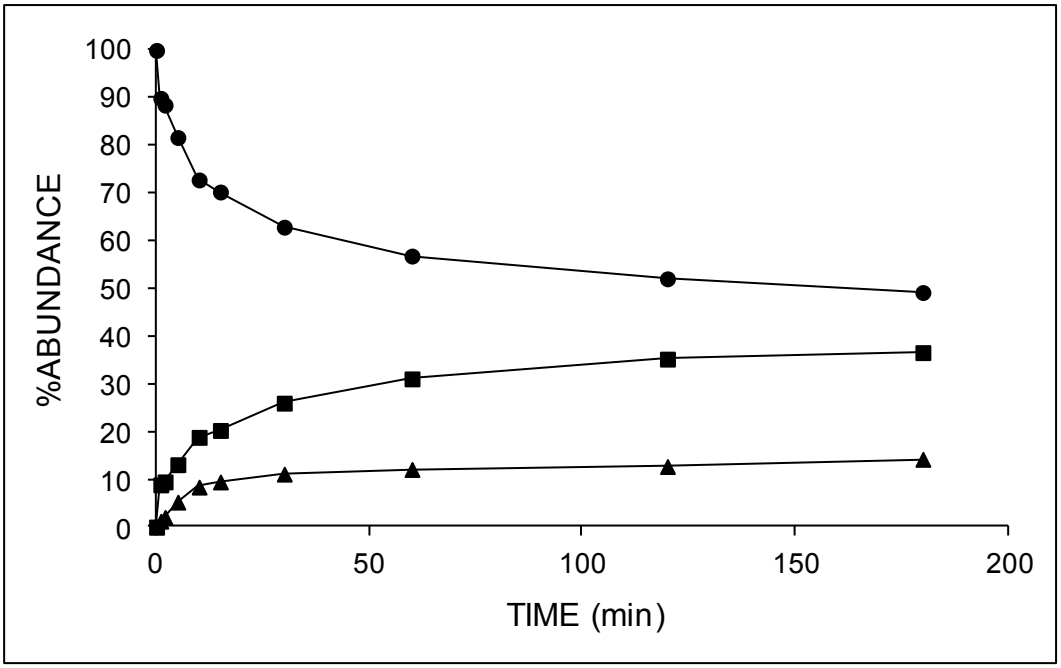


Figure A2.7: EMSA Gel of C145S hAGT and the control 11-mer DNA duplex. 5 nM control DNA and 0 - 35.69 μM C145S hAGT.

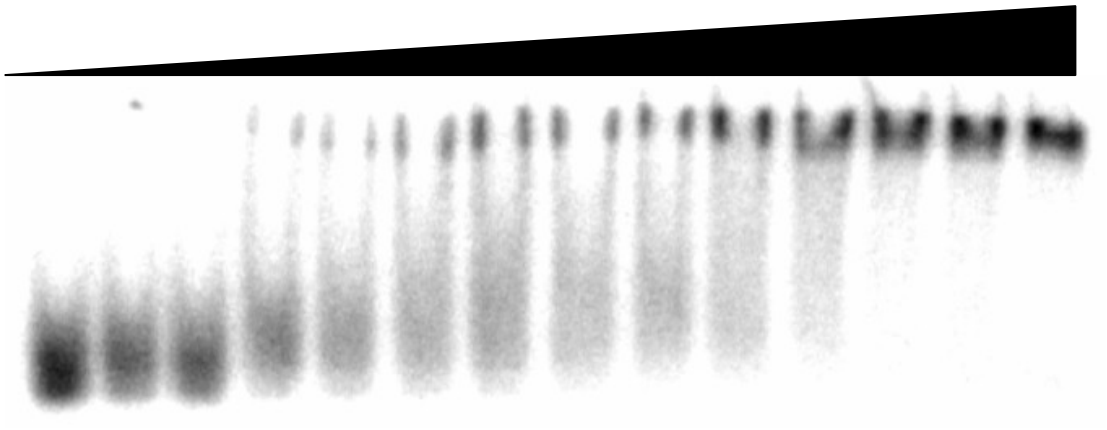


Figure A2.8: Hill plot representation of log[PD]/[D] versus log[P] for the control duplex (■), **XLGG4** (♦) and **XLGG7** (▲).

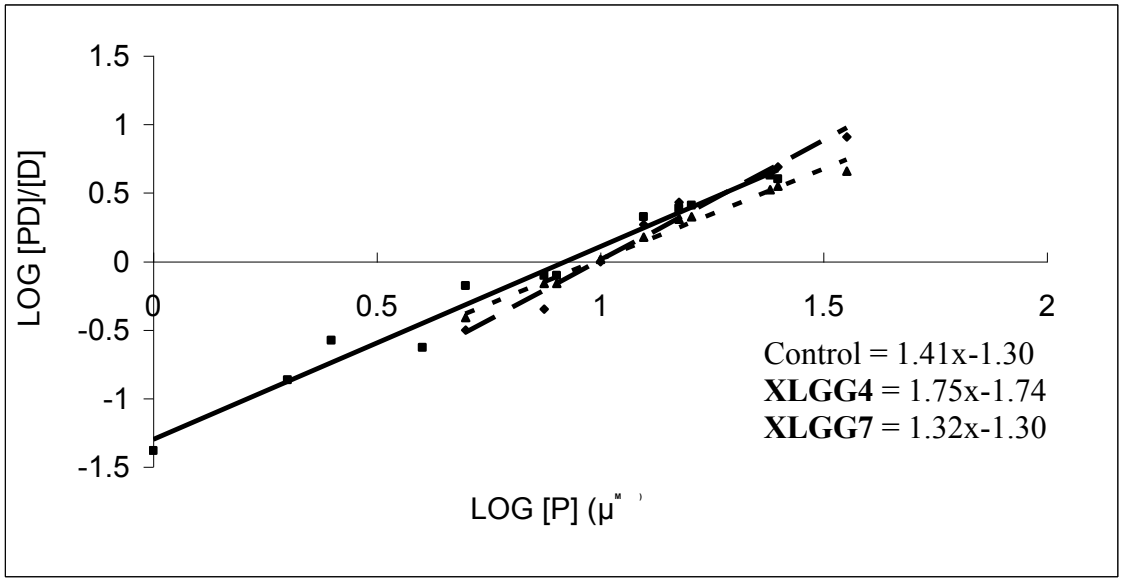


Table A2.1: ESI-MS results of wild-type hAGT and variants

	Molecular Weight (Da)	
Protein	Calculated	Observed
Wild-Type hAGT	21876.2	21875.0
C145S	21860.1	21860.0
P140A	21850.1	21850.5
V148L	21890.2	21889.5

Table A2.2: Effect of mutations on fluorescence emission signals of intrinsic hAGT on Tryptophan and Tyrosine (λex 280nm) and Tryptophan only (λex 295nm).

	λex 280nm		λex 295nm	
Protein	λem (nm)	Fluorescence Intensity	λem (nm)	Fluorescence Intensity
Wild-Type hAGT	346	438	350	184
C145S	350	409	350	176
P140A	346	440	348	169
V148L	350	428	350	173

Table A2.3: Difference in T_m between mutants and wild-type hAGT.

Protein	Melting Temperature (^oC)		
	Observed	Wild-Type	Difference
Wild-Type hAGT	56.5	56.5	0.0
C145S	50.7	56.5	-5.8
P140A	47.0	56.5	-9.5
V148L	48.0	56.5	-8.5

APPENDIX III: Supporting information to CHAPTER III

O^4-alkyl thymidine 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*

Figure A3.1: Thermodynamic data for the non-crosslinked control, **XLTT4** and **XLTT7** DNA

	Control (T-A)				**XLTT4**		**XLTT7**	
Trial	0.585 uM	2.719 uM	12.57 uM	58.46 uM	0.585 uM	2.95 uM	0.585 uM	2.95 uM
1	38.0°C	42.0°C	47.0°C	52.0°C	66.5°C	66.0°C	52.0°C	52.0°C
2	38.0°C	43.0°C	48.0°C	52.0°C	65.0°C	67.0°C	53.5°C	51.0°C
3	39.0°C	43.0°C	49.0°C	50.0°C	66.0°C	66.0°C	51.0°C	53.0°C
Avg	38.3°C	42.7°C	48.0°C	51.3°C	65.8°C	66.3°C	52.2°C	52.0°C
StdDev	0.5°C	0.5°C	0.8°C	0.9°C	0.6°C	0.5°C	1.0°C	0.8°C

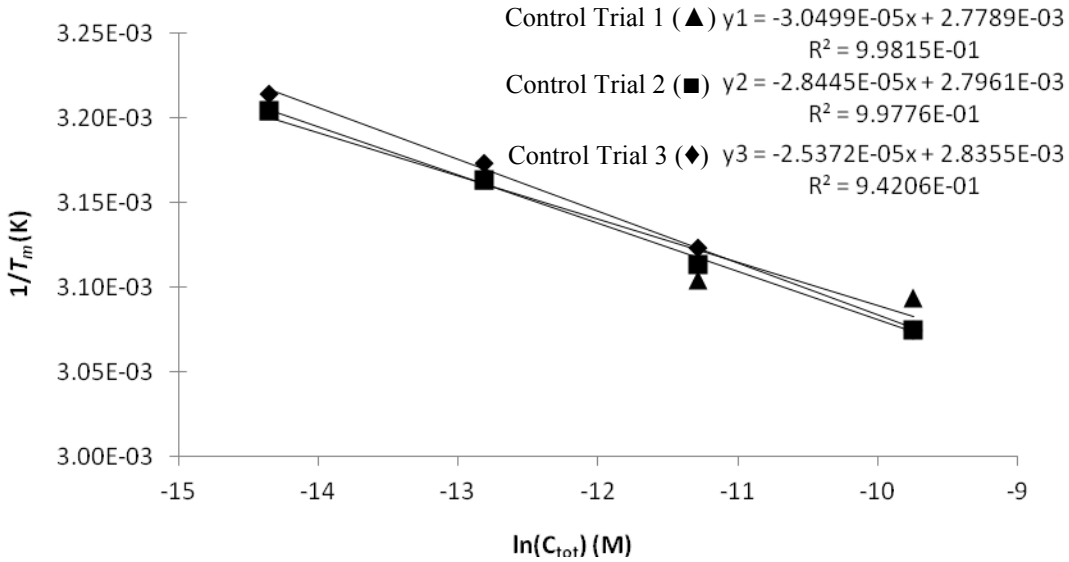


Figure A3.2: T_m curves of various O^4-alkyl dT mono-adducts in DNA duplexes 5' GGCTXGATCACCAG 3' / 5' CTGGTGATCCAGCC 3' where X is dT (——), O^4-MedT (……), O^4-butyl-4-ol dT for **T4** (– • –) and O^4-heptyl-7-ol dT for **T7** (– – –). Solutions containing a total strand concentration of 2.8 μM for the duplexes in 90 mM sodium chloride, pH = 7.0, 10 mM sodium phosphate, and 1 mM EDTA buffer, were heated at 0.5 °C/min.

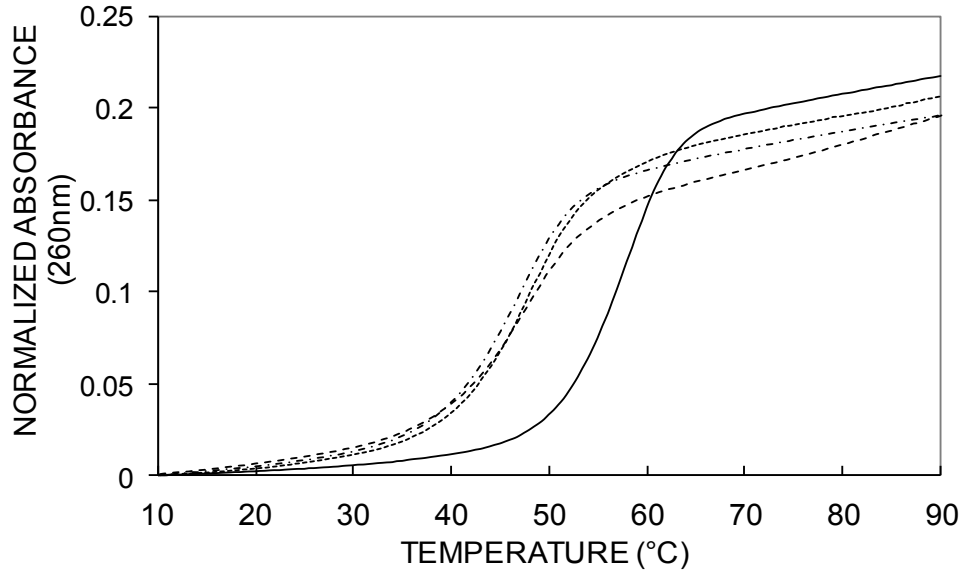


Figure A3.3: CD profiles of various O^4-alkyl dT mono-adducts in DNA duplexes 5' GGCTXGATCACCAG 3' / 5' CTGGTGATCCAGCC 3' where X is dT (——), O^4-MedT (……), O^4-butyl-4-ol dT for **T4** (– • –) and O^4-heptyl-7-ol dT for **T7** (– – –).

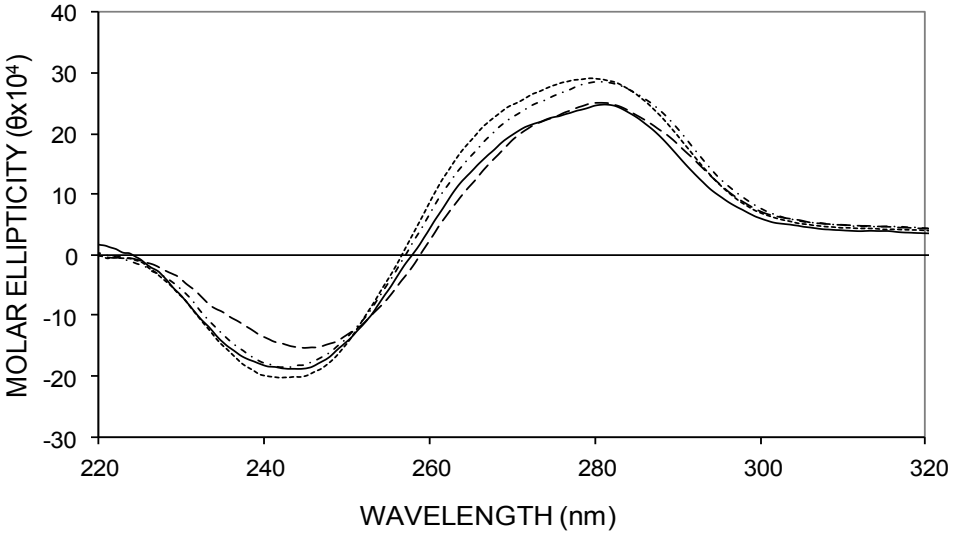


Figure A3.4: Repair gel of O^4-MedT, **T4** and **T7** DNA by hAGT, Ada-C and OGT. Denaturing gel of the repair of 2 pmol of DNA by 10 pmol AGT. Lane 1, Control; lane 2, O^4-MedT DNA; lane 3, O^4-MedT DNA + hAGT; lane 4, O^4-MedT DNA + Ada-C, lane 5, O^4-MedT DNA + OGT; lane 6, **T4** DNA; lane 7, **T4** DNA + hAGT; lane 8, **T4** DNA + Ada-C, lane 9, **T4** DNA + OGT; lane 10, **T7** DNA; lane 11, **T7** DNA + hAGT; lane 12, **T7** DNA + Ada-C, lane 13, **T7** DNA + OGT.

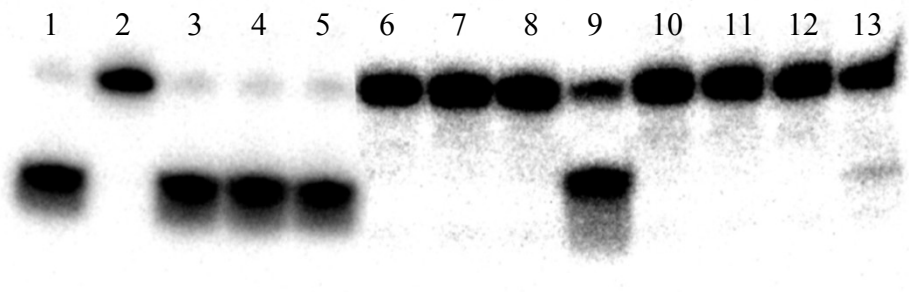


Figure A3.5: Repair of O^4-MedT, **T4** and **T7** by hAGT (red), OGT (green) and Ada-C (blue).

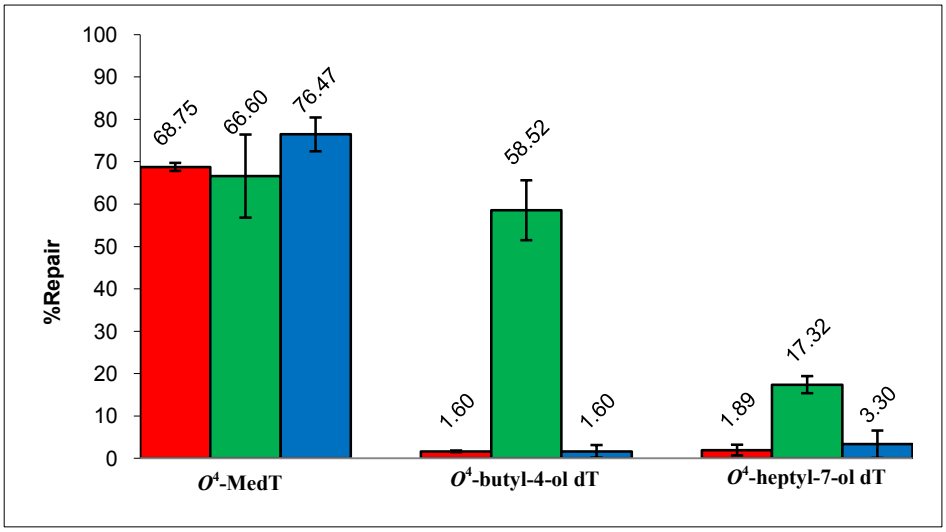


Table A3.1: K_d and stoichiometry of hAGT binding with **XLTT4**, **XLTT7** and control duplex

DNA Duplex	K_d (μM)	Stoichiometry
Control (T-A)	30.57±3.00	1.94 ± 0.03
XLTT4	2.45±0.05	1.90 ± 0.11
XLTT7	2.58±0.67	2.09 ± 0.02

Table A3.2: K_d of hAGT binding to single stranded 5' CGAAAXTTTCG

X	K_d approximate (μM)
T	>30
G	1-10
C	30
A	30

Table A3.3: K_d of hAGT binding to duplexes containing various O^4-alkyl dT mono adducts and O^6-MedG

DNA Duplex	K_d (μM)
Control	6.90 ± 0.04
O^4-MedT	2.87 ± 0.19
T4 (O^4-butyl-4-ol dT)	3.41 ± 0.28
T7 (O^4-heptyl-7-ol dT)	3.68 ± 0.15
O^6-MedG	2.86 ± 0.16

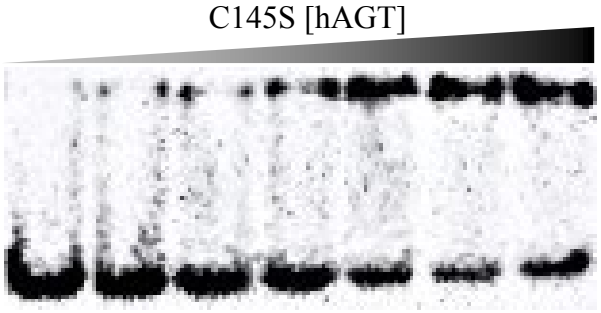
APPENDIX IV: Supporting information to CHAPTER IV

Preparation of Covalently Linked Complexes Between DNA and O^6-Alkylguanine-DNA Alkyltransferase Using Interstrand Cross-linked DNA

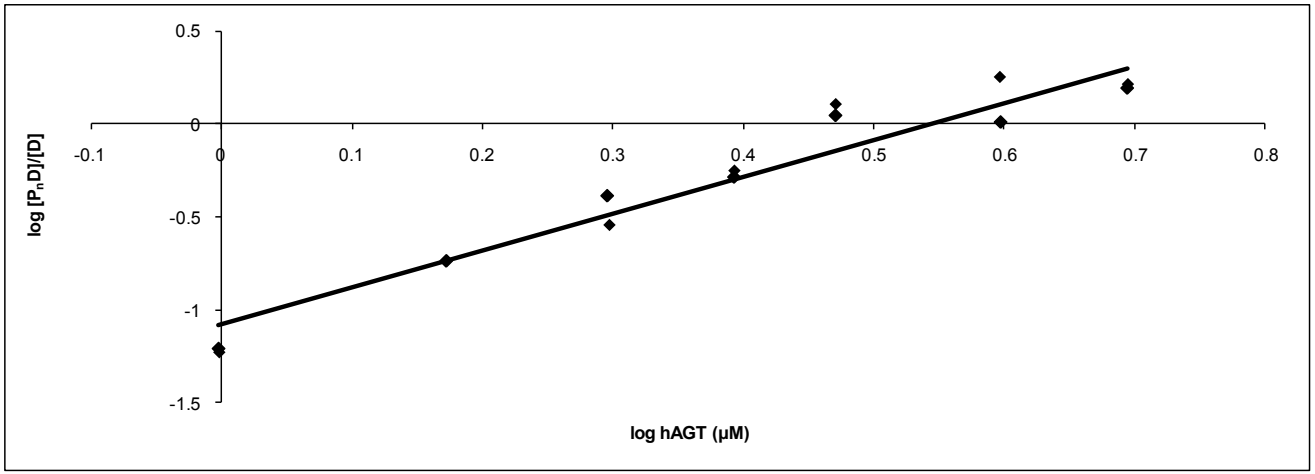
Francis P. McManus, Amardeep Khaira, Anne M. Noronha and Christopher J. Wilds*

Figure A4.1: Electro mobility shift assay of C145S hAGT binding to **XLGT4**. (**A**) EMSA Gel of C145S hAGT and **XLGT4**. 0.5 nM DNA and 1 - 5 μM C145S hAGT. (**B**) Hill plot representation of log[P_nD]/[D] versus log[P].

(**A**)



(**B**)



APPENDIX V: Supporting information to CHAPTER V

Engineering of a O^6-Alkylguanine-DNA alkyltransferase Chimera and Repair of O^4-Alkyl Thymidine Adducts and O^6-Alkylene-2'-Deoxyguanosine Cross-Linked DNA

Francis P. McManus and Christopher J. Wilds*

Figure A5.1: 12% SDS-PAGE of purified chimera proteins. Loaded: lane 1, 10 µL of Unstained Protein Molecular Weight Marker (Fermentas); lane 2, 5 µg chimera protein; lane 3, 5 µg C145S chimera variant protein; lane 4, 5 µg R135G chimera variant protein.

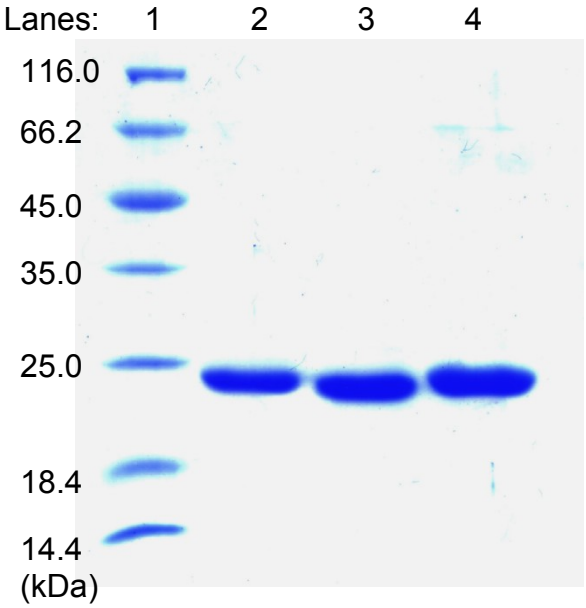
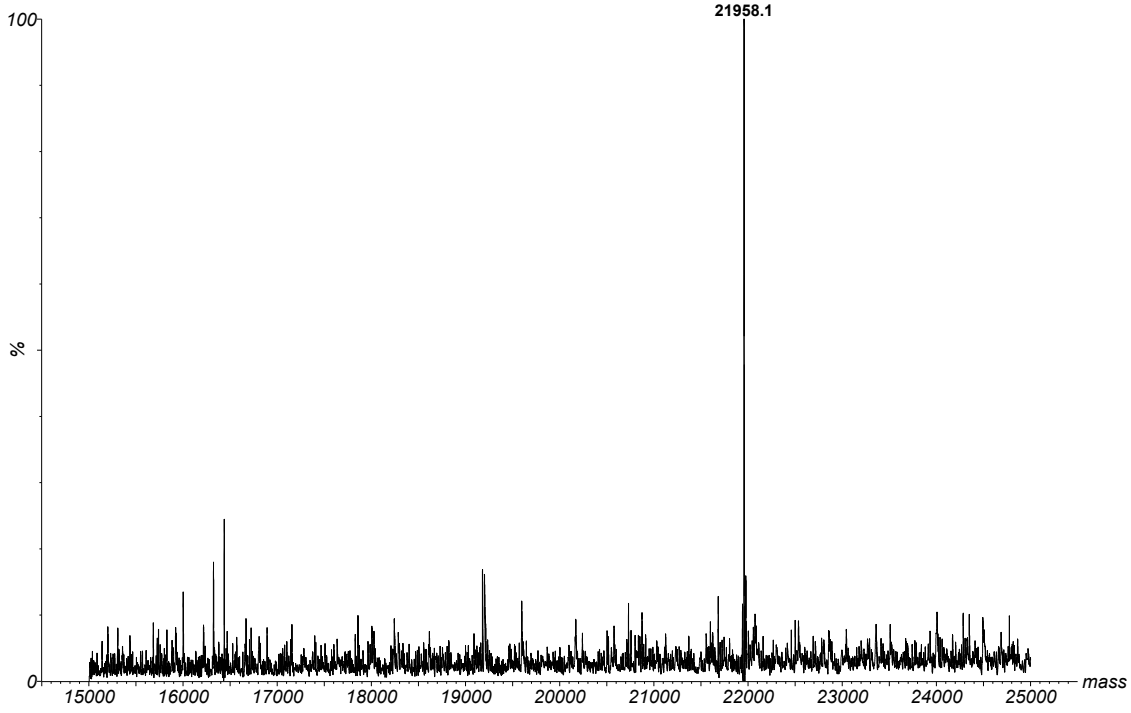
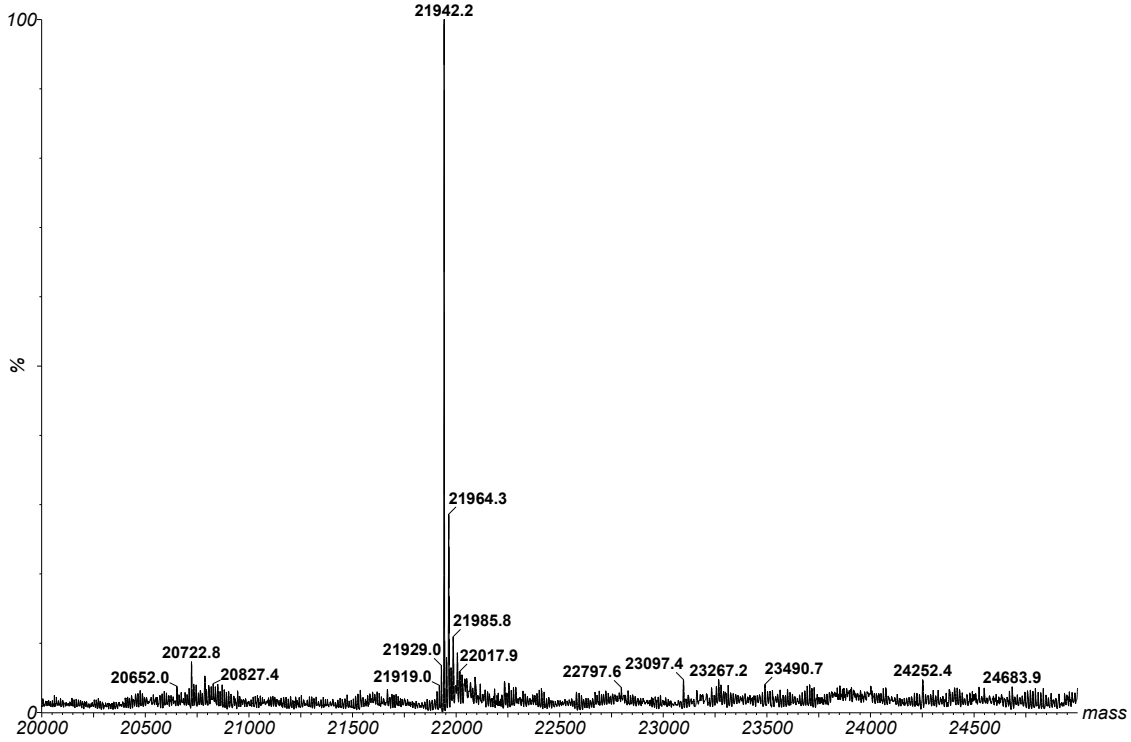


Figure A5.2: ESI-MS of purified chimera protein and variants.
(**A**) Chimera



(**B**) C145S Chimera Variant



(**C**) R135G Chimera Variant

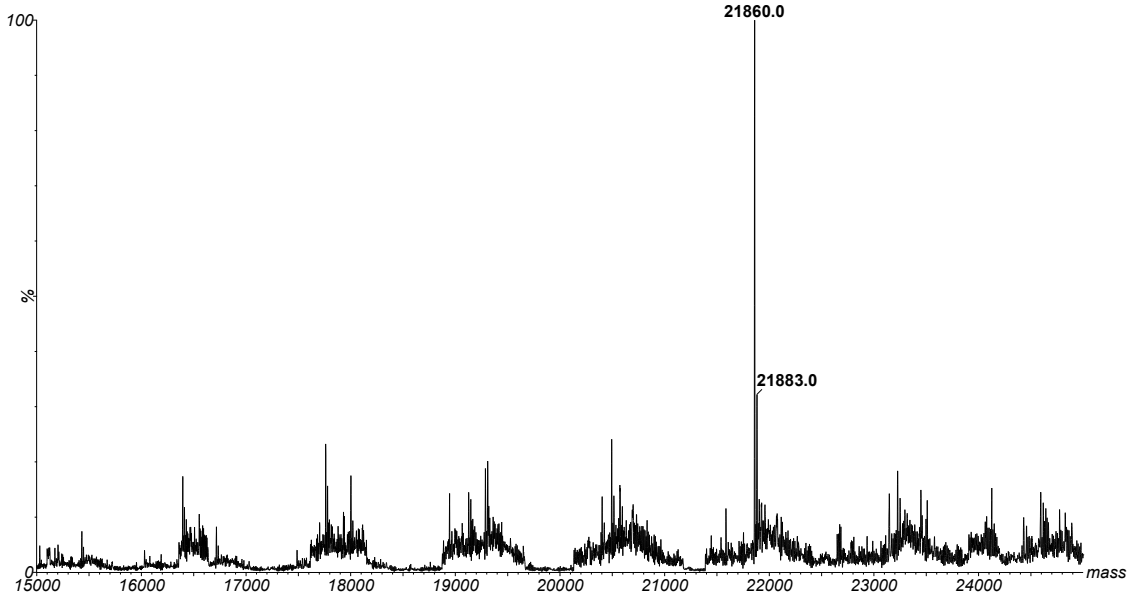


Figure A5.3: Effects of mutations on secondary structure of chimera by far-UV CD. Scans were taken with 5 µM hAGT (_____), chimera (__ _ _ __), C145S chimera variant (__ _ __) and R135G chimera variant (__ __) between 260 and 200 nm.

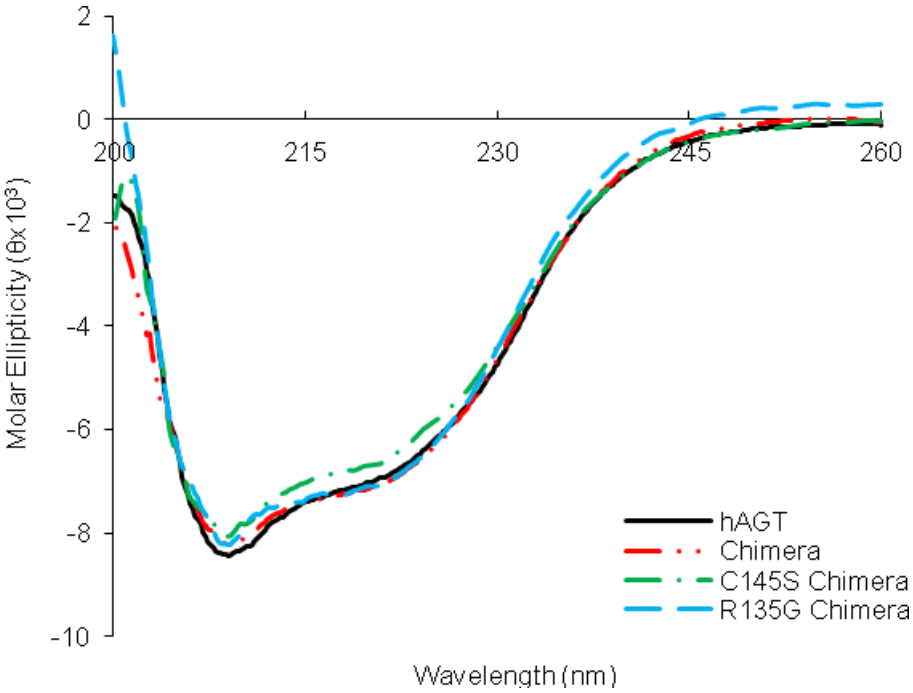


Figure A5.4: Effects of mutations on thermal stability of chimera. Scans were taken with 5 µM hAGT (_____), chimera (__ _ _ __), C145S chimera variant (__ _ __) and R135G chimera variant (__ __) by monitoring the change in molar ellipticity at 222 nm.

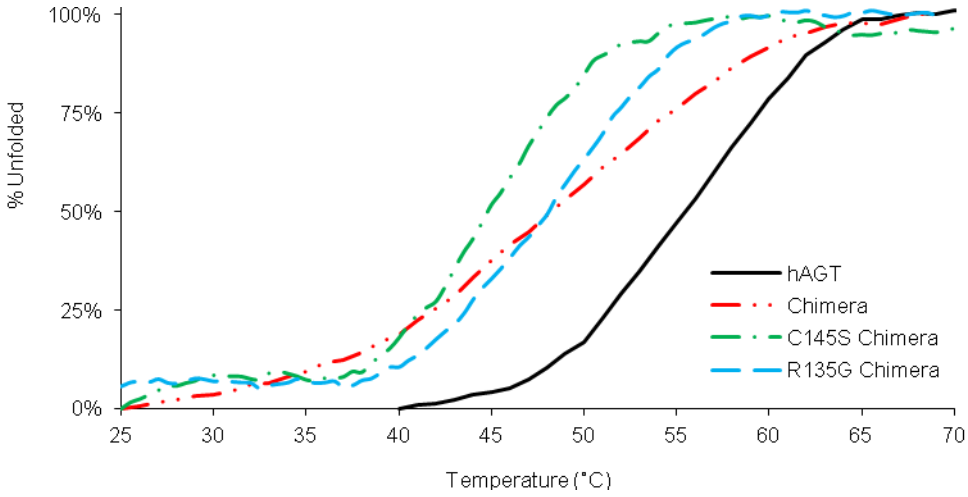
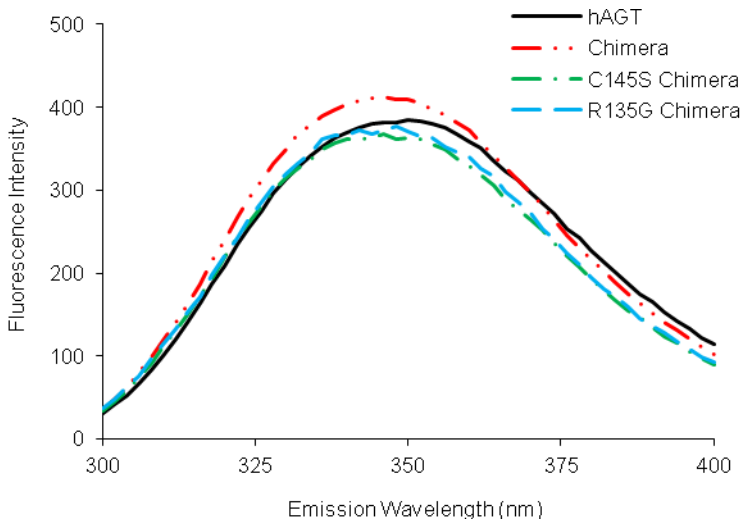


Figure A5.5: Effects of mutations on tertiary structure of chimera by studying the intrinsic fluorescence signals of 1 µM hAGT (_____), chimera (__ _ _ __), C145S chimera variant (__ _ __) and R135G chimera variant (__ __). (**A**) Monitoring of intrinsic tryptophan and tyrosine fluorescence. (**B**) Monitoring of intrinsic tryptophan fluorescence.

(**A**)



(**B**)

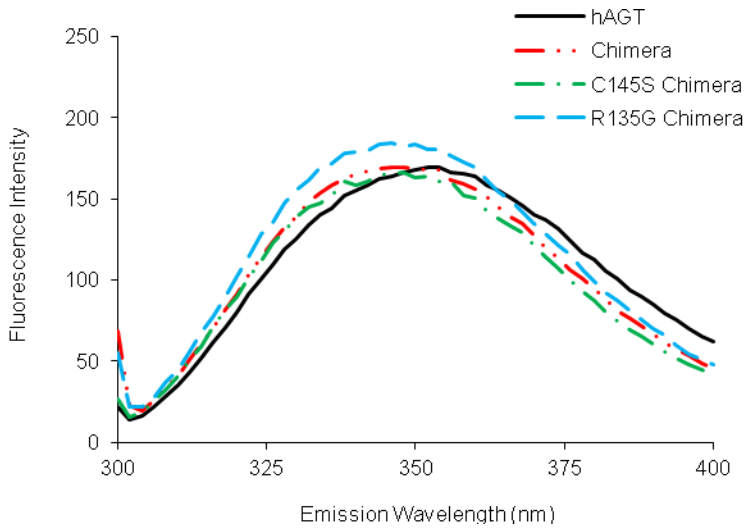


Figure A5.6: Repair gel of O^4-MedT, **T4** and **T7** by chimera and its R135G variant. Denaturing gel of the repair of 2 pmol of DNA by 10 pmol of AGT. Lane 1, O^4-MedT + R135G chimera; lane 2, **T4** + R135G chimera; lane 3, **T7** + R135G chimera; lane 4, O^4-MedT + chimera; lane 5, **T4** + chimera; lane 6, **T7** + chimera; lane 7, Positive Control (dT); lane 8, O^4-MedT; lane 9, **T4**; lane 10, **T7**.

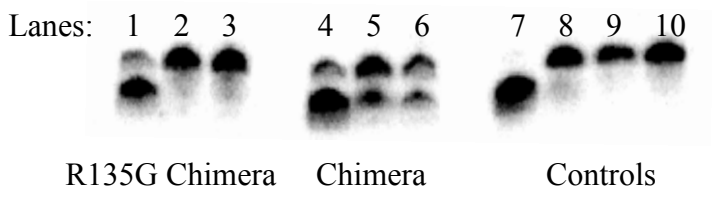
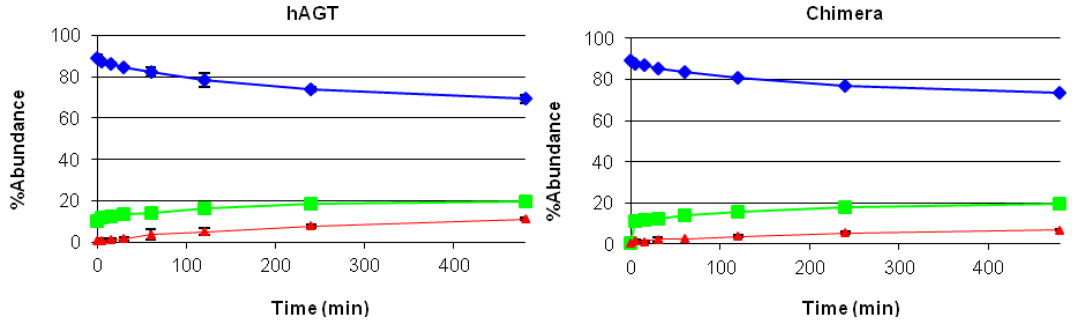


Figure A5.7: Time course repair of **XLGG4** (**A**) and **XLGG7** (**B**) by hAGT and chimera. AGT-DNA complex/partially repaired product (▲); free DNA/ fully repaired product (■); ICL/unrepaired substrate (♦).

(A) XLGG4



(B) XLGG7

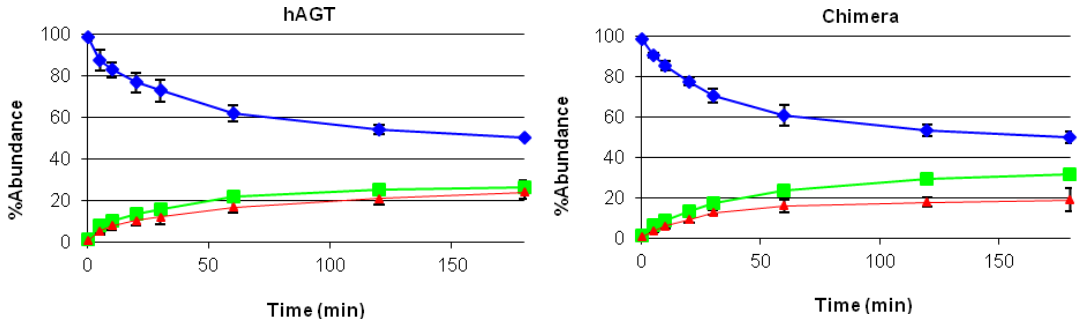
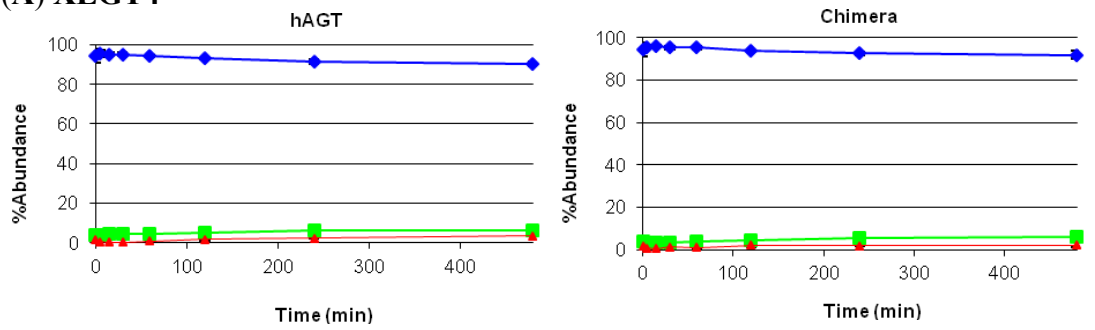


Figure A5.8: Time course repair of **XLGT4** (**A**) and **XLGT7** (**B**) by hAGT and chimera. AGT-DNA complex/partially repaired product (▲); free DNA/ fully repaired product (■); ICL/unrepaired substrate (♦).

(A) XLGT4



(B) XLGT7

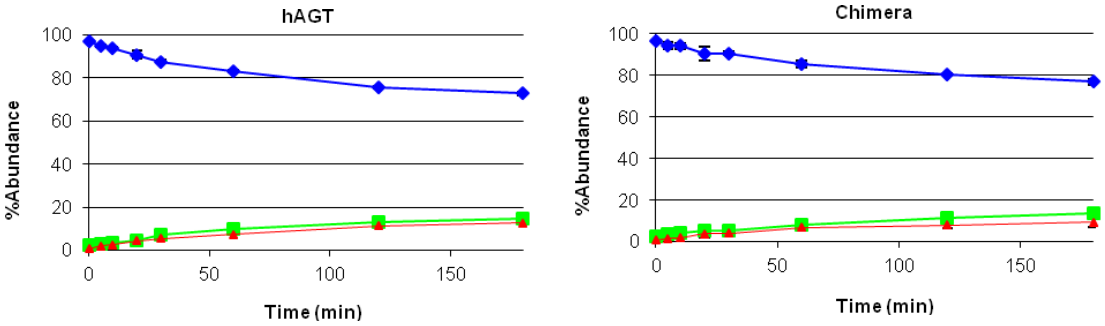


Figure A5.9: 15% SDS-PAGE of chimera mediated repair of **XLGT7**. Repair of 625 pmol **XLGT7** by 600 pmol chimera for 5 h at 37 °C. **XLGT7** + chimera reaction (left); Unstained Protein Molecular Weight Marker (right).

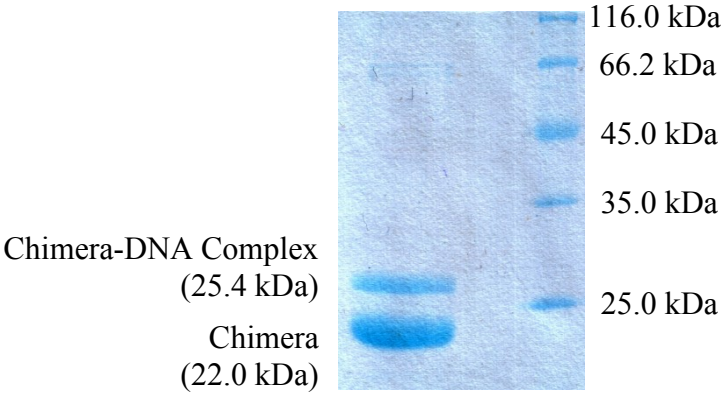


Figure A5.10: Identification of chimera-DNA covalent complex by ESI-MS.

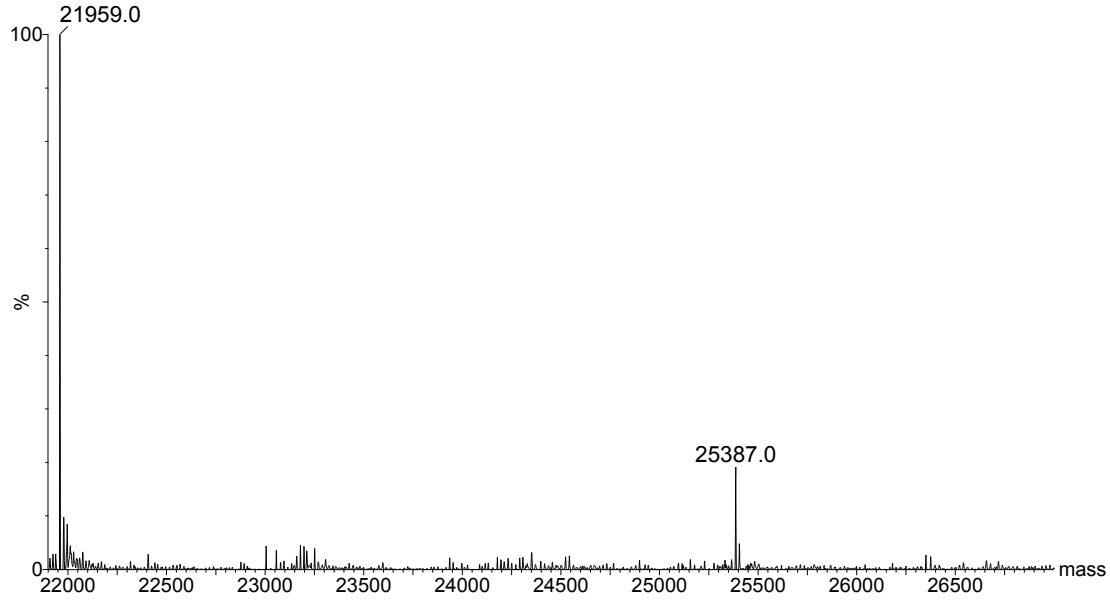
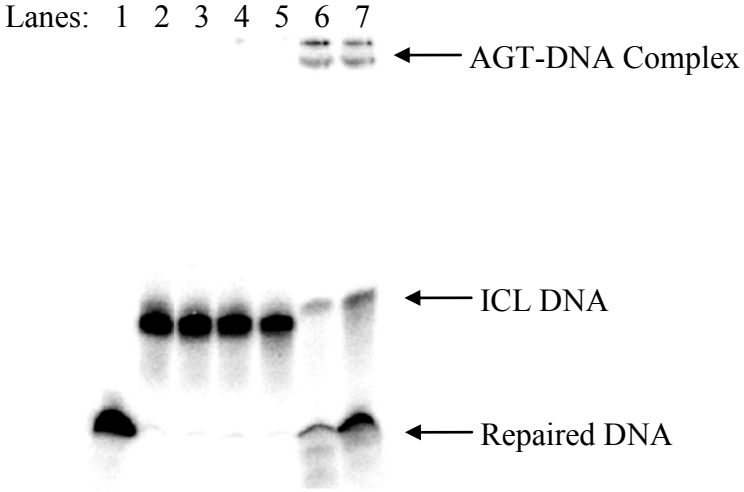


Figure A5.11: Repair gel of **XLTT4** (**A**) and **XLTT7** (**B**) by chimera.
(**A**) Denaturing gel of the repair of 2 pmol of DNA by 60 pmol of AGT. Lane 1, Positive Control; lane 2, **XLTT4**; lane 3, **XLTT4** + chimera; lane 4, **XLTT4** + C145S chimera; lane 5, **XLTT4** + R135G chimera; lane 6, **XLGG7** + hAGT; lane 7, **XLGG7** + chimera.



(**B**) Denaturing gel of the repair of 2 pmol of DNA by 60 pmol of AGT. Lane 1, Positive Control; lane 2, **XLTT7**; lane 3, **XLTT7** + chimera; lane 4, **XLTT7** + C145S chimera; lane 5, **XLTT7** + R135G chimera; lane 6, **XLGG7** + hAGT; lane 7, **XLGG7** + chimera.

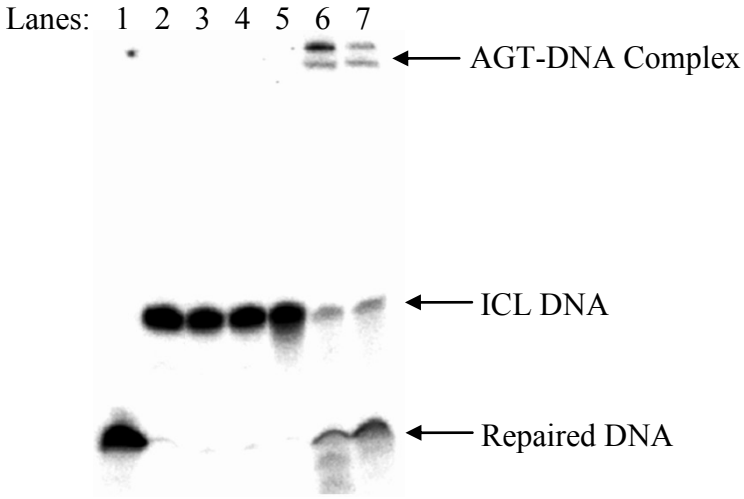
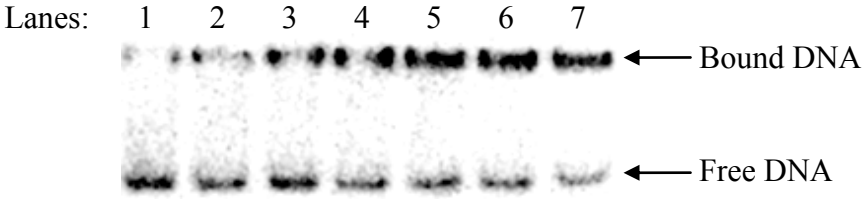


Figure A5.12: Electro mobility shift assay of C145S chimera binding to **XLGT7**. (**A**) EMSA Gel of C145S chimera and **XLGT7**. 5 nM **XLGT7** and 1.11 - 5.55 μM C145S chimera increasing from lanes 1 to 7. (**B**) Hill plot representation of log[PD]/[D] versus log[P] for **XLGT7** (■).

(**A**)



(**B**)

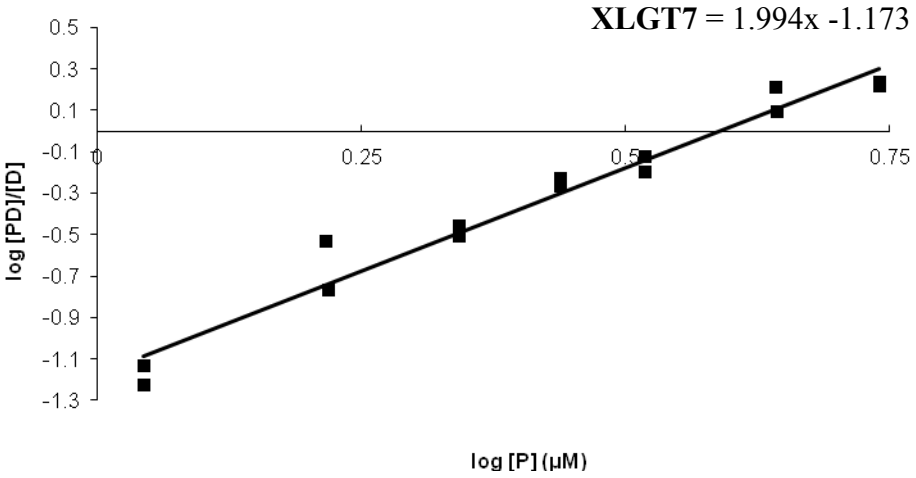


Table A5.1: ESI-MS results of chimera and variants.

	Molecular Weight (Da)	
Protein	Calculated	Observed
Chimera	21959.3	21958.1
C145S variant	21943.2	21942.2
R135G variant	21860.2	21860.0

Table A5.2: Difference in T_m between the chimera, its variants and hAGT.

Protein	Melting Temperature (°C)		
	Observed	Wild-Type	Difference
hAGT	56.5	56.5	0.0
Chimera	49.0	56.5	-7.5
C145S Chimera	45.5	56.5	-11.0
R135G Chimera	49.0	56.5	-7.5

Table A5.3: Effect of mutations on chimera intrinsic fluorescence on tryptophan and tyrosine (λ_{ex} 280 nm) and tryptophan only (λ_{ex} 295 nm).

	λ_{ex} 280 nm		λ_{ex} 295 nm	
Protein	λ_{em} (nm)	Fluorescence Intensity	λ_{em} (nm)	Fluorescence Intensity
hAGT	350	385	353	169
Chimera	346	413	347	169
C145S Chimera	346	368	348	166
R135G Chimera	342	372	346	184

APPENDIX VI: Supporting information to CHAPTER VI

Effect of the C5 Methyl on O^6-Alkylguanine-DNA Alkyltransferase Mediated Repair of O^4-alkyl Thymidine Adducts

Francis P. McManus and Christopher J. Wilds*

Figure A6.1: T_m curves of the various O^4-alkyl dU adducts in DNA duplexes. 5' GGCT**X**GATCACCAG 3' / 5' CTGGTGATCCAGCC 3' where **X** is dU for **U** (——), O^4-MedU for **U-Me** (……), O^4-BenzyldU for **U-Bn** (– • • –), O^4-butyl-4-ol dU for **U4** (– • –) and O^4-heptyl-7-ol dU for **U7** (– – –). Solutions containing 2 μM for the duplexes in 90 mM sodium chloride, pH = 7.0, 10 mM sodium phosphate, and 1 mM EDTA buffer, were heated at 0.5°C/min.

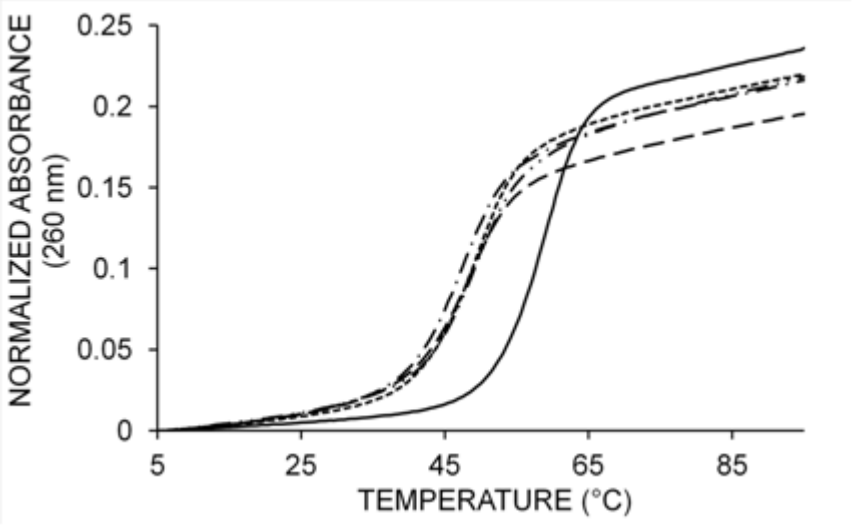
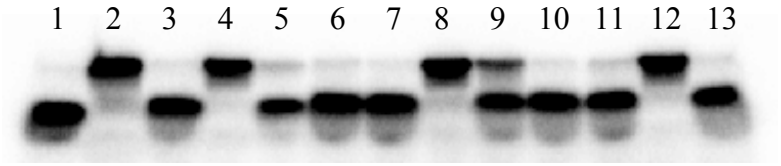
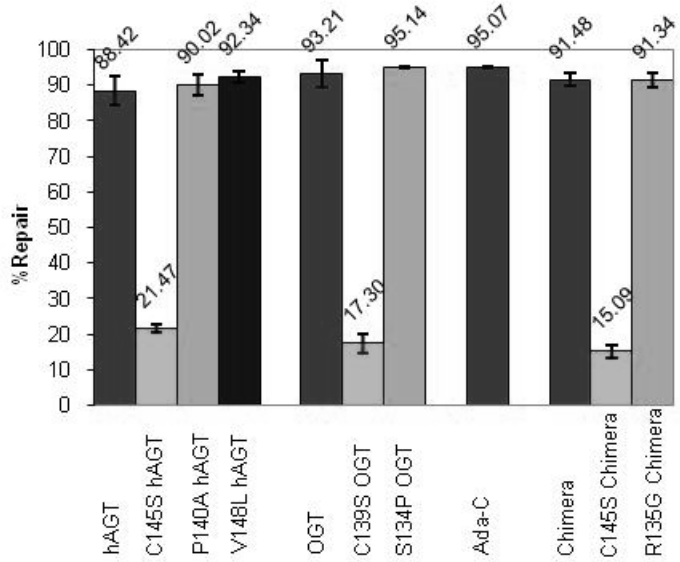


Figure A6.2: Overnight repair of O^4-alkyl dU by AGTs. (**A**) Denaturing gel of the repair of 2 pmol of DNA by 10 pmol AGT: lane 1, **U** (control); lane 2, **U-Me**; lane 3, **U-Me** + hAGT; lane 4, **U-Me** + C145S hAGT; lane 5, **U-Me** + P140V hAGT; lane 6, **U-Me** + V148L hAGT; lane 7, **U-Me** + OGT; lane 8, **U-Me** + C139S OGT; lane 9, **U-Me** + S134P OGT; lane 10, **U-Me** + Ada-C; lane 11, **U-Me** + Chimera; lane 12, **U-Me** + C145S Chimera variant; lane 13, **U-Me** + R135G Chimera variant. Bar graph representation of the overnight repair or **U-Me** (**B**), **U-Bn** (**C**), **U4** (**D**) and **U7** (**E**) by the AGTs.

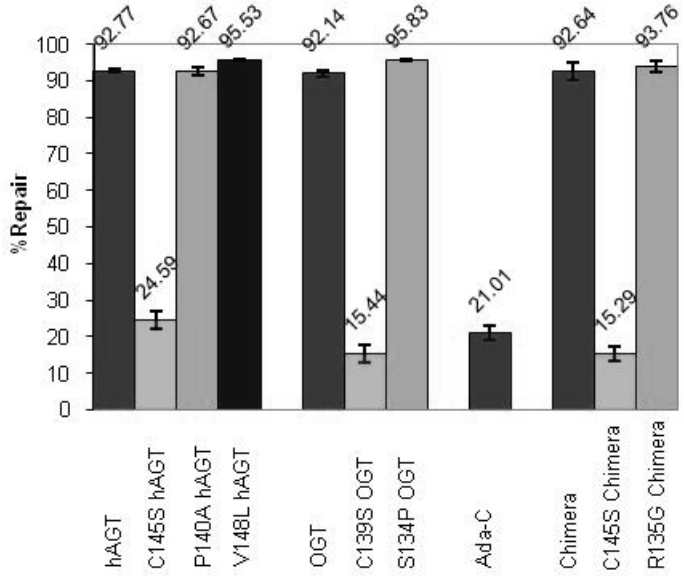
(**A**)



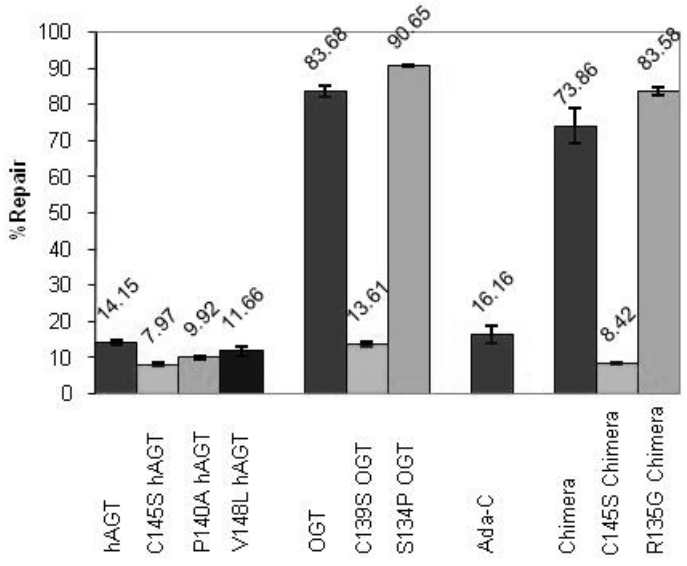
(**B**)



(**C**)



(D)


% Repair
100
90
80
70
60
50
40
30
20
10
0
14.15
7.97
9.92
11.66
83.68
13.61
90.65
16.16
73.86
8.42
83.58
hAGT
C145S hAGT
P140A hAGT
V148L hAGT
OGT
C139S OGT
S134P OGT
Ada-C
Chimera
C145S Chimera
R135G Chimera

(E)

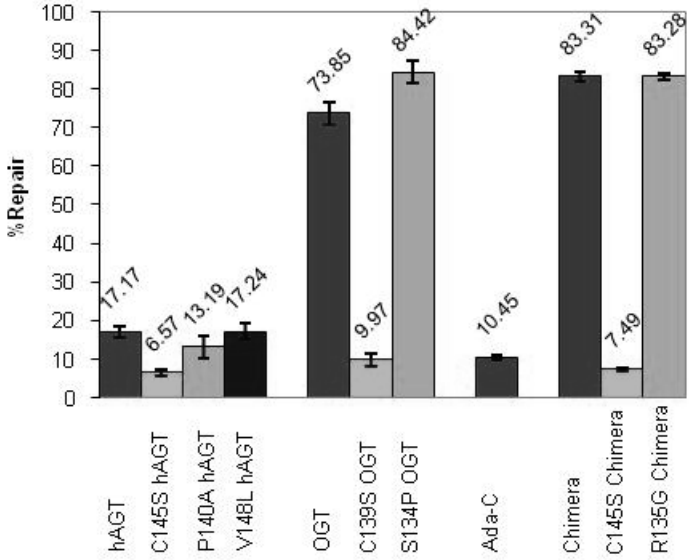
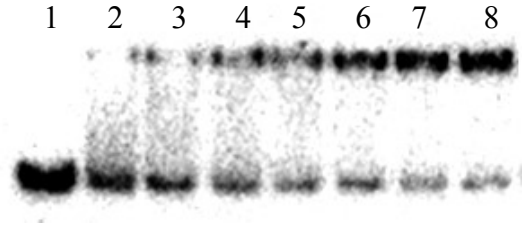
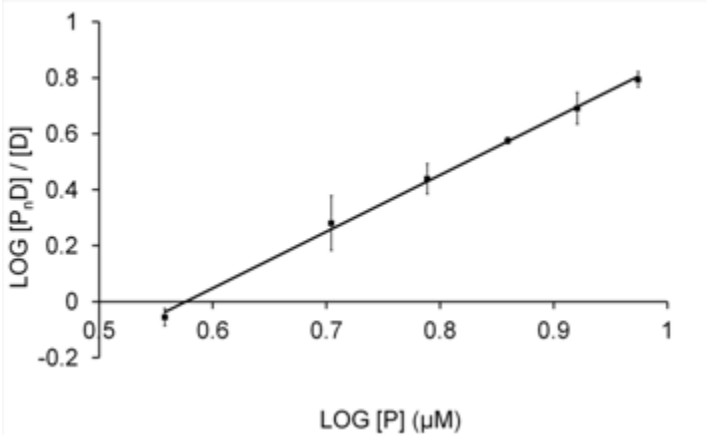

% Repair
100
90
80
70
60
50
40
30
20
10
0
17.17
6.57
13.19
17.24
73.85
9.97
84.42
10.45
83.31
7.49
83.28
hAGT
C145S hAGT
P140A hAGT
V148L hAGT
OGT
C139S OGT
S134P OGT
Ada-C
Chimera
C145S Chimera
R135G Chimera

Figure A6.3: EMSA of C139S OGT binding to **U**. (**A**) EMSA gel of C139S OGT and **U**. 5 nM dsDNA and 0, 3.7, 5.1, 6.2, 7.3, 8.4, 9.5 and 10.6 μM C139S OGT (lanes 1-7, respectively). (**B**) Hill plot representation of log[P_nD]/[D] versus log[P] for **U** (■).

(**A**)



(**B**)



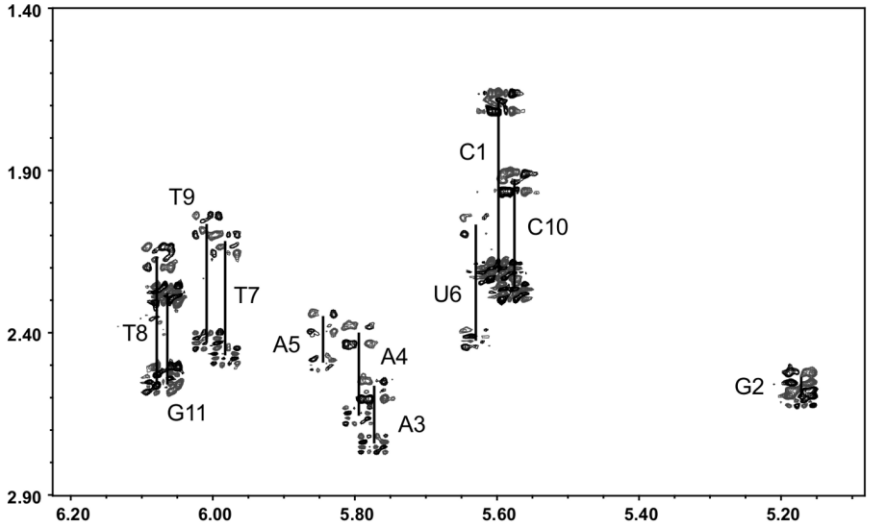
APPENDIX VII: Supporting information to CHAPTER VII

Structural Basis for hAGT Selective Mediated Repair of O^6-2'-Deoxyguanosine-heptylene-O^6-2'-deoxyguanosine and not O^4-2'-Deoxyuridine-heptylene-O^4-2'-deoxyuridine Interstrand Cross-Linked Containing DNA

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

Figure A7.1: Spectral regions of (**A**) DQF-COSY and (**B**) H,P-HSQC spectra at 25 °C. *H1'-H2'* and *H1'-H2"* cross-peaks are connected as well as ^{31}*P-H4'* and ^{31}*P(i)-H3'(i-1)* cross-peaks.

(**A**)



(**B**)

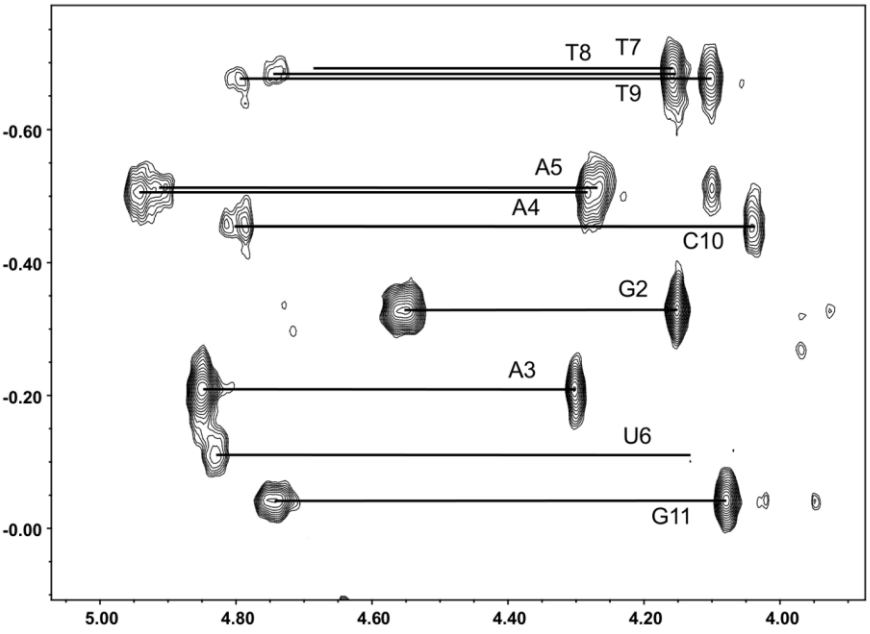


Table A7.1: Proton and phosphorus chemical shifts (ppm) of **XLUU7** (*1*) and **XLGG7** (*2*) duplexes at 25 °C[a].

Duplex	Residue	*H6/H8*	*H5/H2/Me*	*H1'*	*H2'*	*H2''*	*H3'*	*H4'*	*^{31}P(5')*
1	**C1**	7.48	5.80	5.60	1.68	2.20	4.55	3.93	-
2		7.48	5.79	5.59	1.69	2.20	4.55	3.93	-
1	**G2**	7.81	-	5.17	2.53	2.57	4.85	4.14	-0.33
2		7.81	-	5.17	2.54	2.57	4.85	4.15	-0.34
1	**A3**	8.04	7.29	5.77	2.57	2.73	4.95	4.30	-0.21
2		8.05	7.22	5.75	2.58	2.74	4.94	4.30	-0.24
1	**A4**	7.90	7.25	5.80	2.40	2.64	4.91	4.28	-0.50
2		7.87	7.21	5.74	2.35	2.59	4.89	4.27	-0.57
1	**A5**	7.82	7.55	5.85	2.37	2.48	4.83	4.26	-0.51
2		7.72	7.52	5.26	2.23	2.13	4.74	4.20	-0.72
1	**U6**	7.36	5.24	5.64	2.06	2.40	4.68	4.13	-0.11
2	**G6**	7.90	-	6.03	2.65	2.74	4.88	4.33	0.44
1	**T7**	7.33	1.40	5.99	2.11	2.46	4.74	4.16	-0.69
2		6.97	1.07	5.84	1.98	2.48	4.61	4.16	-0.73
1	**T8**	7.46	1.56	6.08	2.15	2.54	4.80	4.15	-0.69
2		7.47	1.51	6.10	2.15	2.55	4.82	4.19	-0.61
1	**T9**	7.38	1.60	6.01	2.06	2.43	4.80	4.10	-0.68
2		7.37	1.60	5.98	2.04	2.40	4.79	4.09	-0.66
1	**C10**	7.43	5.65	5.58	1.92	2.24	4.74	4.04	-0.45
2		7.42	5.65	5.57	1.91	2.25	4.74	4.03	-0.49
1	**G11**	7.86	-	6.06	2.51	2.27	4.58	4.07	-0.05
2		7.85	-	6.06	2.52	2.26	4.58	4.07	-0.06

[a] Proton signals for linker - CH_2(1) 3.47/3.62, CH_2(2) 1.18, CH_2(3) 0.86, CH_2(4) 0.95 ppm for **XLUU7**, and CH_2(1) 3.55/3.88, CH_2(2) 1.23/1.42, CH_2(3) 1.09, CH_2(4) 1.01 ppm for **XLGG7**.