COMPUTATIONAL INVESTIGATIONS OF ALKYLATION DAMAGE TO THE DNA NUCLEOBASE THYMINE

PRIYA BHUTANI
Bachelor of Science, University of New Brunswick, 2016

A thesis submitted
in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in

CHEMISTRY

Department of Chemistry and Biochemistry
University of Lethbridge
LETHBRIDGE, ALBERTA, CANADA

COMPUTATIONAL INVESTIGATIONS OF ALKYLATION DAMAGE TO THE DNA NUCLEOBASE THYMINE

PRIYA BHUTANI

Date of Defence: September 20, 2019

Dr. Stacey Wetmore Thesis Supervisor	Professor	Ph.D.
Dr. Peter Dibble Thesis Examination Committee Member	Associate Professor	Ph.D.
Dr. Paul Hayes Thesis Examination Committee Member	Professor	Ph.D.
Dr. Gino DiLabio External Examiner University of British Columbia Okanagan Kelowna, British Columbia	Professor	Ph.D.
Dr. Michael Gerken Chair, Thesis Examination Committee	Professor	Ph.D.

Abstract

DNA alkylation damage is caused by numerous sources in the environment. Alkylation damage can stall standard DNA polymerases and replication instead occurs through a process called translesion synthesis (TLS). This thesis uses a combination of density functional theory calculations and molecular dynamics simulations to rationalize the mutagenic patterns of O-alkylthymine adducts and elucidate the role of human TLS polymerase η in their replication. Specifically, this thesis unveils the effect of alkyl chain position by considering O2-methylthymine and O4-methylthymine. Furthermore, the impact of alkyl chain size is revealed by studying the O4-[4-(3-pyridyl)-4-oxobut-1-yl]-thymine lesion, which is derived from tobacco smoke. Finally, insight into the impact of alkyl chain shape is obtained by considering the synthetic O2-n-butylthymine and O2-iso-butylthymine adducts. Overall, this thesis addresses some of the unanswered questions surrounding the replication of thymine alkylation damage and provides a better understanding of human polymerase η that has been implicated in disease.

Acknowledgements

To start off, I would like to thank my supervisor, Dr. Stacey Wetmore. The support you have provided over the past two years is unmatched, and I will always be grateful for the opportunities you have given me. I would like to thank my supervisory committee members, Dr. Peter Dibble and Dr. Paul Hayes, for their helpful comments and support throughout my degree. I would also like to thank Dr. Michael Gerken for chairing my defense. Finally, thank you to Dr. Gino DiLabio (University of British Columbia Okanogan) for taking an interest in this thesis and agreeing to be my external examiner.

I have had the privilege of working with many other Wetmore Lab members over the past two years. A special thanks to Dr. Katie Wilson, Makay Murray and Dylan Nikkel, whose research has directly contributed to this thesis. Additionally, thank you to past and present graduate students and post-doctoral fellows (Dr. Mohamed Aboelnga, Dr. Preethi Seelam Prabhakar, Dr. Stefan Lenz, Dr. Shahin Sowlati, Ryan Kung, Rajwinder Kaur, Hanzala Hussain, Briana Boychuk) for your support and friendship along the way. Thank you to past and present undergraduate students (Lindey Felske, Cynthia Fonderson, Rebecca Jeong, Nathania Takyi, Chalalai Phawaphootanon, Bhadra Pandya, Michaela Reitsma, Kohl Yee and Trinity Deak) for always lightening the mood in stressful times.

Finally, thank you to my family and friends from coast to coast. To those from Fredericton, especially those who have ventured out west in the past two years for a visit, I appreciate the support and understanding you have provided throughout my degree. To the friends I have made in Lethbridge, whether it was on the running trails, through the graduate students' association, or from teaching Chem 2000, thank you for making my short time here memorable. I couldn't imagine my life without you all.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

A	adenine
C	cytosine
DFT	density functional theory
DNA	deoxyribose nucleic acid
dNTP	deoxyribonucleotide triphosphate
dR	deoxyribose
G	guanine
GAFF	general AMBER force field
HG	Hoogsteen
MD	molecular dynamics
NAB	nucleic acid builder
NNK	nicotine-derived nitrosamine ketone
O2-iBu-T	O2-iso-butylthymine
O2-Me-T	O2-methylthymine
O2-nBu-T	O2-n-butylthymine
O4-Me-T	O4-methylthymine
O4-POB-T	O4-[4-(3-pyridyl)-4-oxobut-1-yl]-thymine
PDB	protein data bank
PES	potential energy surface
POB	pyridyloxobutyl
pol	polymerase
QM	quantum mechanics
RESP	restrained electrostatic potential
rmsd	root mean squared deviation
SAM	S-Adenosyl methionine
T	thymine
TLS	translesion synthesis
UV	ultraviolet
WC	Watson-Crick

Chapter 1: Introduction[a]

1.1 Overview

Genetic information is encoded within deoxyribose nucleic acid (DNA), which must be replicated in order for most living organisms to grow and survive. However, DNA can be damaged by a variety of endogenous and exogenous sources, such as ultraviolet (UV) radiation, tobacco smoke, pollution, chemotherapeutic drugs and S-adenosyl methionine (SAM).[1-3] When DNA is damaged, standard DNA replication can be stalled, leading to a process called translesion synthesis (TLS) in which a TLS polymerase attempts to replicate past the damaged site. TLS polymerases have very flexible active sites to accommodate different types of damaged DNA, ultimately resulting in these enzymes being more error-prone than standard DNA polymerases. For example, TLS polymerases often lead to base mispairs, as well as insertion or deletion mutations.[4] Experimental work has provided insight into the mutagenic patterns of a number of damaged DNA bases and the role of TLS in processing these lesions.[4-6] Atomic level details of the TLS replication process are difficult and time consuming to obtain with current experimental methods. Computational chemistry is an excellent tool to gain structural insight about DNA damage and the associated replication processes.[7-10] This thesis uses a combination of density functional theory (DFT) calculations and molecular dynamics (MD) simulations to understand how DNA damage, particularly alkylation damage to the thymine (T) nucleobase, is replicated by human TLS polymerase η (pol η). Pol η has been experimentally shown to be the primary TLS polymerase that replicates past many O-alkyl-T lesions.[11-16] Interestingly, experimental studies in different environments indicate that

[a]*The Journal of American Chemical Society* reference style is used in this chapter.

methylation at the O4 position of T results in a range in the frequency of guanine (G) mispairing, whereas methylation at the O2 position is primarily non-mutagenic (i.e. adenine (A) is predominantly inserted opposite these lesions), emphasizing the impact of the damage position on mutagenicity.[11, 16] Furthermore, the size of the alkyl chain and the degree of branching have been shown to impact the mutagenic patterns of T lesions.[11-16] Together, these results raise questions regarding how different T alkylation adducts are replicated by pol η. The objective of this thesis is to provide structural insight into the replication of alkylation damage at the O2 and O4 positions of T in order to understand how the location, size and shape of the damage affect the replication outcomes. Overall, this chapter provides the necessary background on DNA structure, damage and replication to understand the scope of this thesis.

1.2 DNA Structure and Damage

DNA is comprised of four canonical nitrogenous nucleobases, specifically two purines, namely A and G, and two pyrimidines, namely T and cytosine (C, Figure 1.1). In canonical DNA, these nucleobases form complementary base pairs that are stabilized through Watson-Crick (WC) hydrogen bonding, where A pairs with T and G pairs with C (Figure 1.2).

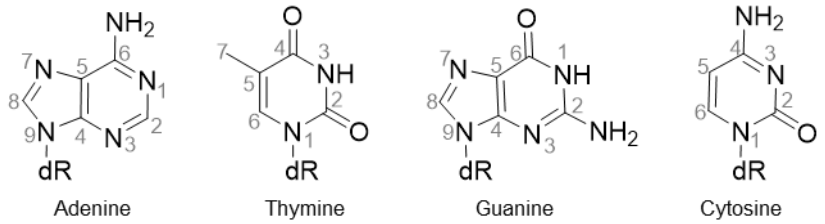


Figure 1.1: Chemical structure and atomic numbering of the four canonical DNA nucleosides (dR = deoxyribose).

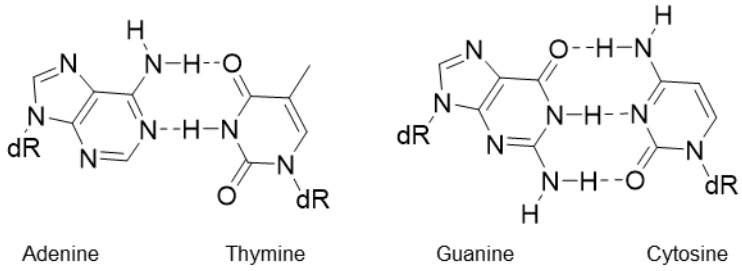


Figure 1.2: Chemical structure and hydrogen-bonding patterns of the canonical A:T (left) and G:C (right) pairs (dR = deoxyribose).

The glycosidic bond connects the nucleobase to the deoxyribose, which is also connected to the phosphodiester portion of the backbone, forming a nucleotide (Figure 1.3a). Although base pairs primarily interact through WC hydrogen bonding, the nucleobase can rotate about the glycosidic bond, (i.e. around the dihedral angle (χ) that is defined as $\angle$(O4′–C1′–N1–C2) for pyrimidines and as $\angle$(O4′–C1′–N9–C4) for purines, Figure 1.4). Specifically, WC base pairs are more common and form when nucleosides adopt *anti* orientations ($\chi = 180 \pm 90°$), while Hoogsteen (HG) hydrogen-bonded base pairs form when nucleobases rotate about χ to adopt *syn* conformations ($\chi = 0 \pm 90°$). Two complementary DNA strands come together to form a DNA duplex, most commonly in the B-DNA form, although other conformations also exist (such as A-DNA and Z-DNA). Key structural features of B-DNA include a twist of 36° that occurs approximately every 10 base pairs, and the presence of deep major and minor grooves (Figure 1.3b).[17] Overall, the integrity of this helical structure is important for maintaining our genetic information.

Unfortunately, DNA can be exposed to a variety of damaging agents, including UV radiation, tobacco smoke, pollution, and many other sources.[2] Common types of DNA damage that directly affect the canonical nucleobases include alkylation, oxidation, and deamination.[1] These damaging agents can modify the nucleobase or DNA backbone.[2-3]

When nucleobase damage occurs, the structure of DNA may be compromised, which can ultimately affect important cellular processes such as DNA replication.

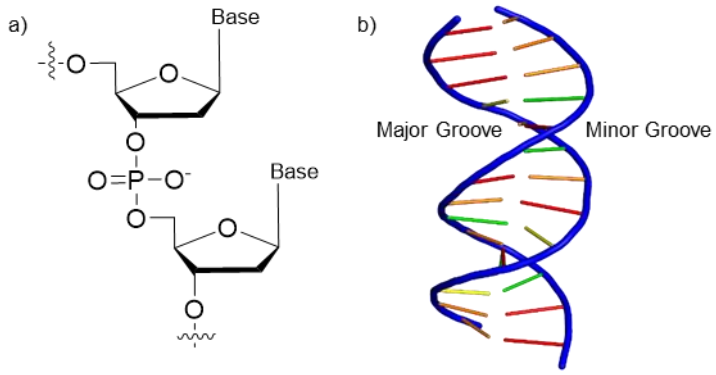


Figure 1.3: a) Chemical structure of a single DNA strand, highlighting the backbone. b) Global DNA helical structure, with the major and minor grooves labelled.

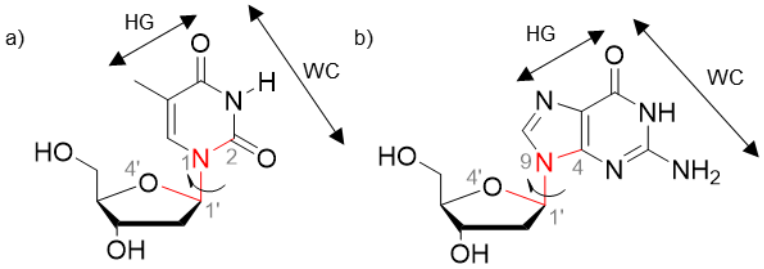


Figure 1.4: Definition of the dihedral angle (at 180°) about the glycosidic bond (χ) highlighted in red for a representative a) pyrimidine and b) purine, as well as the Watson-Crick (WC) and Hoogsteen (HG) faces labelled.

1.3 DNA Replication

DNA replication is a key process in living organisms required to grow and survive. During replication, the DNA double helix is unwound by a DNA helicase, which breaks the WC hydrogen bonding within the canonical base pairs. The two resulting DNA single strands are then processed by DNA polymerases, which generates two genetically identical DNA duplexes by inserting the complementary base opposite each template base in both single strands. Due to standard DNA replication polymerases being tightly bound to the

growing DNA strand, DNA damage may block standard DNA replication. In order for replication to continue, the TLS pathway is used. TLS polymerases have more flexible active sites compared to standard replicative polymerases, which allows these polymerases to accommodate various types of DNA damage.[4-5] These polymerases are more error-prone than standard DNA polymerases and can potentially result in base mispairs (transition or transversion mutations), as well as insertion or deletion mutations upon replication.[4] Ultimately, these replication outcomes can lead to cell death or various diseases, such as cancer and Alzheimer's disease.[18-19]

In humans, there are four known TLS polymerases, namely pol η, pol κ, pol ι and Rev1.[4] These TLS polymerases have different features that dictate their role in the bypass of DNA damage. For example, pol η can replicate past a number of lesions due to an open pocket on the major groove side of the bound DNA, and therefore this polymerase typically replicates damage in the major groove (e.g. the O4 position of T).[4, 20-22] Similarly, pol κ has an open pocket on the minor groove side of the growing DNA strand, making minor groove DNA lesions common substrates (e.g. damage at O2 of T).[4-5] Although pol η and pol κ have broad substrate specificities, pol ι exhibits a preference for bypassing less frequently observed Hoogsteen base pairs, while Rev1 acts as an insertase by inserting C opposite lesions.[4-5] Overall, these TLS polymerases play important roles in replicating damaged DNA and their ability to replicate past DNA alkylation damage is the focus of this thesis.

1.4 DNA Alkylation Damage

DNA alkylation damage commonly occurs from exposure to tobacco smoke, chemotherapeutic drugs, SAM, and atmospheric halocarbons.[23] These sources can act as methylating and ethylating agents, directly adding alkyl chains to the four canonical DNA bases.[24] In particular, methyl methane sulfonate (MMS) is an example of a chemotherapeutic drug that methylates DNA in cancer and non-cancer cells.[18] Alternatively, tobacco smoke contains the 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) carcinogen, which is activated by cytochrome P450 enzymes to yield a variety of reactive intermediates upon metabolism.[25-27] These reactive intermediates lead to methyl, ethyl, and bulkier 4-(3-pyridyl)-4-oxobut-1-yl (POB) adducts, among others, which can form at various positions in any of the four DNA nucleobases.[27-29] Positions at which the associated alkyl damage is most commonly detected include N7 and O6 of G, and O2 of T and C.[28] However, other positions that can be alkylated have been more recently identified including O4 of T, N4 of C, N6 of A, and the phosphate backbone.[27] This thesis focuses on alkylation damage to the O2 and O4 positions of T, as many questions remain unanswered about the replication outcomes of these lesions.

1.4.1 Thymine Alkylation Damage

DNA alkylation damage to T most commonly occurs at the O2 and O4 positions (Figure 1.5). Alkylation damage to T at either position changes N3 from a hydrogen-bonding donor to a hydrogen-bonding acceptor, altering the canonical WC binding face (Figure 1.5). The smallest alkylated T lesions, O2-methylthymine (O2-Me-T) and O4-methylthymine (O4-Me-T) can form from exogenous methylating agents (e.g. MMS and

tobacco smoke) and endogenous sources (e.g. SAM).[3] Larger O-alkyl-T adducts, such as O-POB-T adducts, are known to form from exposure to tobacco smoke.[25, 27] These O-alkyl-T adducts can be resistant to repair processes, indicating that they will persist during DNA replication.[24, 30-35] Additionally, other O-alkyl-T lesions of varying lengths and shapes have been synthesized to determine how these structural features affect replication.[11-16] Furthermore, O-alkyl-T lesions stall standard DNA polymerases,[3] and thus undergo replication through the TLS process.[11-16] Together, these results suggest that O-alkyl-T lesions exhibit different types and frequencies of misinsertion depending on the alkyl chain position, length and shape, the details of which are discussed in the following section.

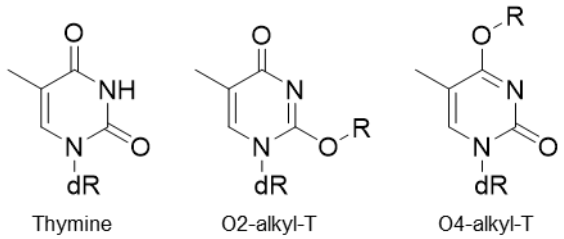


Figure 1.5: Chemical structure of T (left) and the alteration of the WC face upon alkylation at the O2 (middle) or O4 (right) positions, dR = deoxyribose and R = alkyl chain.

1.4.2 Replication Outcomes of O-alkyl-T Adducts

Alkylation damage to the O2 and O4 positions of T has been studied experimentally *in vitro* and *in vivo* (in both human and *E. coli* cells).[11-16] Although O2-T and O4-T adducts similarly alter the WC face, they have vastly different experimental mutagenic outcomes. Specifically, methylation at the O2 position (O2-Me-T, Figure 1.6a) is primarily non-mutagenic, with A being correctly inserted opposite the lesion 98%, 95% and 64% of the time in human cells, *in vitro*, and in *E. coli*, respectively.[11, 14-15] In contrast, methylation at the O4 position (O4-Me-T, Figure 1.6b) results in 60% and 70% G misinsertion opposite

this lesion during replication in human and *E. coli* cells, respectively, while misinsertion of G occurs twice as often as A *in vitro*.[12-13, 16] Additionally, larger alkyl chains at both positions (namely POB (O4-POB-T),[36] ethyl, n-propyl, iso-propyl, n-butyl (O2-nBu-T), iso-butyl (O2-iBu-T, Figure 1.6) result in similar mispairing patterns as their methyl counterpart, but with varying frequencies of misinsertion.[11-16] For example, increasing the length of the alkyl chain at the O2 position primarily results in an increase in the misinsertion of C and T (and sometimes G), making alkylation damage more mutagenic with increasing size of the alkyl chain.[11, 14-15] In contrast, increasing the length of the alkyl chain at the O4 position, results in a decrease in mutagenicity (e.g. dGTP insertion decreases from 60% for methyl to 35% for POB in human cells).[16] In addition, branched O2-alkyl-T lesions exhibit lower frequencies of mutation than their straight-chained counterparts (e.g. the addition of an iso-butyl group to the O2 position of T is less mutagenic that the addition of an n-butyl group).[12, 14, 16] The subtle differences in branched and straight chained O2-alkyl-T lesions likely arise from how these alkyl chains interact with the polymerase. To my knowledge, structural insight to rationalize the mutagenic patterns for O2-alkyl-T and O4-alkyl-T adducts has not appeared in the literature to date.

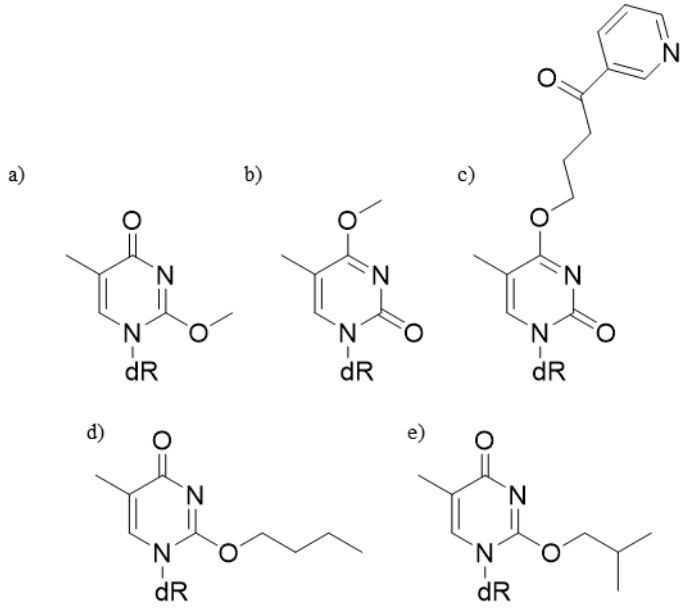


Figure 1.6: Chemical structure of various alkylated T adducts considered in this thesis including a) O2-Me-T, b) O4-Me-T, c) O4-POB-T, d) O2-nBu-T and e) O2-iBu-T, dR = deoxyribose.

The activity of three human TLS polymerases (pol η, pol κ and pol ι) has been carefully studied to determine the polymerase(s) responsible for replicating the O2 and O4-T alkylation adducts.[11, 16] Overall, knockout studies indicate that pol η has the largest effect on the mutagenicity for O4-T lesions. This is likely due to the alkyl chains pointing into the open pocket in the active site on the major groove side of the growing DNA strand.[12-13, 16] Similarly, knockout of pol η results in a significant decrease in the mutagenicity of O2-alkyl-T lesions.[11, 14-15] Due to the fact that pol η has an open pocket on the major groove side of the growing DNA strand, the presence of damage on the minor groove side may disrupt key active site residues. Current experimental results are unable to explain how pol η can accommodate these lesions, and therefore structural insights from computational chemistry may help rationalize how both O4 and O2 adducts are replicated by pol η.

1.5 Using Computational Chemistry to Study Damage to T

To help explain the observed experimental mutagenic patterns, computational chemistry has been used to gain structural insight into T lesions.[8, 37-40] Previous computational studies have used QM methods, specifically semi-empirical or MP2 calculations, to look at potential base-pairing properties of smaller alkylation T adducts (i.e. methylation damage).[38-40] Discrepancies between the theoretical calculations and experimental results arise because O-Me-T adducts have only been examined in isolated nucleobase models and have not considered interactions within the pol η active site or within the context of a post-extension DNA strand. Nevertheless, these isolated models provide insight into G misinsertion opposite O4-Me-T, with the predicted base pair containing two strong hydrogen bonds and key base pair structural features that are comparable to canonical T:A.[38, 41] Notably, O4-Me-T is not always mutagenic,[12-13, 16] which contrasts a predicted unfavourable O4-Me-T:A interaction in the isolated base-pair model.[38-40] Similarly, although O2-Me-T is primarily non-mutagenic,[11, 14-15] the WC face is altered analogously to that of O4-Me-T and previous work using isolated models suggest that this lesion may also pair favourably with G.[41] Thus, no structural evidence is currently available to explain why these two similar lesions produce different mutagenic results. By considering the dynamics of these base pairs within the pol η active site and post-extension DNA duplexes, the replication outcomes of O2 and O4 alkylated T lesions can be better understood.

In contrast to computational work on O-Me-T lesions, modelling on other T lesions has provided key structural insight into their mutagenic patterns and the role of the TLS pathway in their replication. Specifically, recent work on the tobacco-derived O2-POB-T

adduct has used a combination of DFT and MD simulations to understand how this lesion is accommodated within the active sites of pol η and pol κ.[8] Experimental work indicates that pol η preferentially inserts T and A opposite this lesion, whereas pol κ inserts G.[36, 42] Computational modelling unveiled that the minor groove O2-POB-T lesion must undergo a conformational change about the glycosidic bond prior to dNTP (deoxyribonucleotide triphosphate) insertion by pol η and thereby provided the first example of *syn* pyrimidine replication.[8] However, due to the open pocket of pol κ being on the minor groove side of the growing DNA strand, the bulky POB moiety adopts the standard *anti* conformation during pol κ replication, revealing a key difference between how these TLS polymerases replicate past this bulky damage and rationalizing the different mutagenic outcomes for these human polymerases. Furthermore, computational work on TT-dimers, which are formed from exposure to UV radiation, provided structural insight into the mutagenic patterns upon replication by pol η[9-10, 37] and elucidated the roles of key pol η active site residues.[9] Taken together, these works validate that computational chemistry is a useful tool for gaining structural insight into the TLS process and the replication of T lesions.

1.6 Thesis Overview

This thesis uses a combination of DFT methods and MD simulations to probe how the position, size and shape of O-alkyl-T adducts affect the replication outcomes associated with human TLS pol η. Initially, DFT calculations will be used to look at the flexibility and preferred orientations of the alkyl chains on T. Subsequently, MD simulations on the pol η pre-insertion complexes (i.e. in the absence of an incoming dNTP) containing O-alkyl-T adducts will be performed to determine how these lesions are positioned in the

active site of pol η and which active site residues, if any, are displaced. Next, DFT methods will be used to understand the hydrogen-bonding potential between O-alkyl-T adducts and select bases in order to identify favourable base pairs. Finally, MD simulations will be used to investigate the ternary complex of pol η bound to a damaged DNA strand, as well as post-extension DNA duplexes, which will uncover the structure and dynamics of these complexes. In combination, these approaches will provide insight into the different mutagenic patterns observed for O2 and O4 alkylation adducts of T. In the process, this thesis will solidify the computational methodology that can be reliably used to study other forms of DNA damage and the role that TLS polymerases play in human DNA replication.

Using the general approach described above, Chapter 2 uncovers the reason for the significant differences in the observed mutagenic patterns of the two smallest T alkylation products, namely O2-Me-T and O4-Me-T (Figure 1.6a,b), and why the position of the methyl group has such a large effect on the mutagenic outcomes.[11-16] Specifically, although these lesions similarly alter the WC face, O2-Me-T is primarily non-mutagenic upon replication (98% dATP insertion),[11] whereas O4-Me-T results in a high percentage of base mispairing (> 60% dGTP insertion in human cells).[16] Next, Chapter 3 explains the observed mutagenic pattern of the recently detected O4-POB-T adduct (Figure 1.6c) and clarifies the differences between the replication outcomes of this lesion (dGTP mispairs) and the O2-POB-T counterpart (dTTP mispairs).[36, 42] Additionally, although both O4-POB-T and O4-Me-T can be accommodated in the open pocket of pol η, experimental results indicate that dGTP insertion significantly decreases when the bulky POB moiety is attached (35% insertion)[36] compared to a methyl group (> 60% insertion).[12-13, 16] As a result, this thesis clarifies how varying the alkyl chain size can lead to different replication outcomes.

Chapter 4 then explains the subtle differences in the pol η replication of straight-chained and branched alkyl lesions by considering O2-nBu-T (Figure 1.6d) and O2-iBu-T (Figure 1.6e) as prototypical examples. Upon replication, O2-nBu-T is more mutagenic (14% dTTP insertion) than O2-iBu-T (4% dTTP insertion), and therefore, my work answers questions about how the shape of the alkyl chain impacts lesion accommodation by pol η.[11] Finally, Chapter 5 summarizes the major conclusions of this thesis and provides directions for future work in this area. Taken together, this thesis fills current holes in the literature on the replication of thymine alkylation damage and provides a greater understanding of the TLS process facilitated by polymerase η.

1.7 References

(1) Bauer, N. C.; Corbett, A. H.; Doetsch, P. W. The current state of eukaryotic DNA base damage and repair. *Nucleic Acids Res.* **2015,** *43,* 10083-10101.

(2) Friedberg, E. C. DNA damage and repair. *Nature* **2003,** *421,* 436-440.

(3) Yu, Y.; Wang, P.; Cui, Y.; Wang, Y. Chemical analysis of DNA damage. *Anal. Chem.* **2018,** *90,* 556-576.

(4) Quinet, A.; Lerner, L. K.; Martins, D. J.; Menck, C. F. M. Filling gaps in translesion DNA synthesis in human cells. *Mutat. Res. Genet. Toxicol. Environ. Mutagen.* **2018,** *836,* 127-142.

(5) Vaisman, A.; Woodgate, R. Translesion DNA polymerases in eukaryotes: What makes them tick? *Crit. Rev. Biochem. Mol. Biol.* **2017,** *52,* 274-303.

(6) Nair, D. T.; Kottur, J.; Sharma, R. A rescue act: Translesion DNA synthesis past N2-deoxyguanosine adducts. *IUBMB Life* **2015,** *67,* 564-574.

(7) Wilson, K. A.; Wetmore, S. D. Molecular insights into the translesion synthesis of benzyl-guanine from molecular dynamics simulations: Structural evidence of mutagenic and nonmutagenic replication. *Biochemistry* **2017,** *56,* 1841-1853.

(8) Wilson, K. A.; Holland, C. D.; Wetmore, S. D. Uncovering a unique approach for damaged DNA replication: A computational investigation of a mutagenic tobacco-derived thymine lesion. *Nucleic Acids Res.* **2019,** *47,* 1871-1879.

(9) Ucisik, M. N.; Hammes-Schiffer, S. Effects of active site mutations on specificity of nucleobase binding in human DNA polymerase η. *J. Phys. Chem. B* **2017,** *121,* 3667-3675.

(10) Ucisik, M. N.; Hammes-Schiffer, S. Relative binding free energies of adenine and guanine to damaged and undamaged DNA in human DNA polymerase η: Clues for fidelity and overall efficiency. *J. Am. Chem. Soc.* **2015,** *137,* 13240-13243.

(11) Wu, J.; Li, L.; You, C.; Wang, Y.; Wang, P.; Wang, Y. Cytotoxic and mutagenic properties of minor-groove O(2)-alkylthymidine lesions in human cells. *J. Biol. Chem.* **2018,** *293,* 8638-8644.

(12) Wang, P.; Amato, N. J.; Zhai, Q.; Wang, Y. Cytotoxic and mutagenic properties of O4-alkylthymidine lesions in *Escherichia coli* cells. *Nucleic Acids Res.* **2015,** *43,* 10795-10803.

(13) Williams, N. L.; Wang, P.; Wu, J.; Wang, Y. *In vitro* lesion bypass studies of O4-alkylthymidines with human DNA polymerase η. *Chem. Res. Toxicol.* **2016,** *29,* 669-675.

(14) Williams, N. L.; Wang, P.; Wang, Y. Replicative bypass of O2-alkylthymidine lesions *in vitro*. *Chem. Res. Toxicol.* **2016,** *29,* 1755-1761.

(15) Zhai, Q.; Wang, P.; Cai, Q.; Wang, Y. Syntheses and characterizations of the *in vivo* replicative bypass and mutagenic properties of the minor-groove O2-alkylthymidine lesions. *Nucleic Acids Res.* **2014,** *42,* 10529-10537.

(16) Wu, J.; Li, L.; Wang, P.; You, C.; Williams, N. L.; Wang, Y. Translesion synthesis of O4-alkylthymidine lesions in human cells. *Nucleic Acids Res.* **2016,** *44,* 9256-9265.

(17) Drew, H. R.; Wing, R. M.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E. Structure of a B-DNA dodecamer: Conformation and dynamics. *Proc. Natl. Acad. Sci. U. S. A.* **1981,** *78,* 2179-83.

(18) Cheung-Ong, K.; Giaever, G.; Nislow, C. DNA-damaging agents in cancer chemotherapy. *Chem. Biol.* **2013,** *20,* 648-659.

(19) Yamanaka, K.; Chatterjee, N.; Hemann, M. T.; Walker, G. C. Inhibition of mutagenic translesion synthesis: A possible strategy for improving chemotherapy? *PLoS Genet.* **2017,** *13,* e1006842/1-e1006842/16.

(20) Yang, W. An overview of Y-family DNA polymerases and a case study of human DNA polymerase η. *Biochemistry* **2014,** *53,* 2793-2803.

(21) Silverstein, T. D.; Johnson, R. E.; Jain, R.; Prakash, L.; Prakash, S.; Aggarwal, A. K. Structural basis for the suppression of skin cancers by DNA polymerase η. *Nature* **2010,** *465,* 1039-1043.

(22) Biertuempfel, C.; Zhao, Y.; Kondo, Y.; Ramon-Maiques, S.; Gregory, M.; Lee, J. Y.; Masutani, C.; Lehmann, A. R.; Hanaoka, F.; Yang, W. Structure and mechanism of human DNA polymerase η. *Nature* **2010,** *465,* 1044-1048.

(23) Drabløs, F.; Feyzi, E.; Aas, P. A.; Vaagbø, C. B.; Kavli, B.; Bratlie, M. S.; Peña-Diaz, J.; Otterlei, M.; Slupphaug, G.; Krokan, H. E. Alkylation damage in DNA and RNA—repair mechanisms and medical significance. *DNA Repair* **2004,** *3,* 1389-1407.

(24) Shrivastav, N.; Li, D.; Essigmann, J. M. Chemical biology of mutagenesis and DNA repair: Cellular responses to DNA alkylation. *Carcinogenesis* **2010,** *31,* 59-70.

(25) Hecht, S. S. DNA adduct formation from tobacco-specific N-nitrosamines. *Mutat. Res. Fundam. Mol. Mech. Mutagen.* **1999,** *424,* 127-142.

(26) Sturla, S. J.; Scott, J.; Lao, Y.; Hecht, S. S.; Villalta, P. W. Mass spectrometric analysis of relative levels of pyridyloxobutylation adducts formed in the reaction of DNA with a chemically activated form of the tobacco-specific carcinogen 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone. *Chem. Res. Toxicol.* **2005,** *18,* 1048-1055.

(27) Ma, B.; Stepanov, I.; Hecht, S. S. Recent studies on DNA adducts resulting from human exposure to tobacco smoke. *Toxics* **2019,** *7*.

(28) Hecht, S. S. Biochemistry, biology, and carcinogenicity of tobacco-specific N-nitrosamines. *Chem. Res. Toxicol.* **1998,** *11,* 559-603.

(29) Hecht, S. S. Progress and challenges in selected areas of tobacco carcinogenesis. *Chem. Res. Toxicol.* **2008,** *21,* 160-171.

(30) Riazuddin, S.; Athar, A.; Saffhill, R. Chemical adaptation of *M. Luteus* induces repair functions for O-alkylated DNA pyrimidines. *Nucleic Acids Res.* **1985,** *13,* 7153-66.

(31) Bronstein, S. M.; Skopek, T. R.; Swenberg, J. A. Efficient repair of O6-ethylguanine, but not O4-ethylthymine or O2-ethylthymine, is dependent upon O6-alkylguanine-DNA alkyltransferase and nucleotide excision repair activities in human-cells. *Cancer Res.* **1992,** *52,* 2008-2011.

(32) McCarthy, T. V.; Karran, P.; Lindahl, T. Inducible repair of O-alkylated DNA pyrimidines in *Escherichia coli*. *The EMBO Journal* **1984,** *3,* 545-550.

(33) McManus, F. P.; O'Flaherty, D. K.; Noronha, A. M.; Wilds, C. J. O4-alkyl-2'-deoxythymidine cross-linked DNA to probe recognition and repair by O6-alkylguanine DNA alkyltransferases. *Org. Biomol. Chem.* **2012,** *10,* 7078-7090.

(34) Brent, T. P.; Dolan, M. E.; Fraenkel-Conrat, H.; Hall, J.; Karran, P.; Laval, L.; Margison, G. P.; Montesano, R.; Pegg, A. E.; Potter, P. M. Repair of O-alkylpyrimidines in mammalian cells: A present consensus. *Proc. Natl. Acad. Sci. U.S.A.* **1988,** *85,* 1759-1762.

(35) Leng, J.; Wang, Y. Liquid chromatography-tandem mass spectrometry for the quantification of tobacco-specific nitrosamine-induced DNA adducts in mammalian cells. *Anal. Chem.* **2017,** *89,* 9124-9130.

(36) Du, H.; Leng, J.; Wang, P.; Li, L.; Wang, Y. Impact of tobacco-specific nitrosamine-derived DNA adducts on the efficiency and fidelity of DNA replication in human cells. *J. Biol. Chem.* **2018,** *293,* 11100-11108.

(37) Ucisik, M. N.; Hammes-Schiffer, S. Comparative molecular dynamics studies of human DNA polymerase η. *J. Chem. Inf. Model.* **2015,** *55,* 2672-81.

(38) Qiu, Z. M.; Wang, H. L.; Liu, Y. Z.; Hou, D. N. MP2 study on the hydrogen-bonding interaction between O4-methylthymine and DNA bases: A, C, G, and T. *J. Struc. Chem.* **2014,** *25,* 767-774.

(39) Pohorille, A.; Loew, G. H. Base-pairing properties of O-methylated bases of nucleic acids. Energetic and steric considerations. *Biophys. Chem.* **1985,** *22,* 37-51.

(40) Flood, A.; Hubbard, C.; Forde, G.; Hill, G.; Gorb, L.; Leszczynski, J. Theoretical *ab initio* study of the effects of methylation on the nature of hydrogen bonding in A:T base pair. *J. Biomol. Struct. Dyn.* **2003,** *21*, 297-302.

(41) Venkateswarlu, D.; Lyngdoh, R. H. D. Structural, steric and energetic requirements for induction of base substitutional mutations by methylated guanines and thymines. *J. Chem. Soc. Perkin Trans. 2* **1995,** 839-846.

(42) Gowda, A. S. P.; Spratt, T. E. DNA polymerases η and ζ combine to bypass O2-[4-(3-pyridyl)-4-oxobutyl]thymine, a DNA adduct formed from tobacco carcinogens. *Chem. Res. Toxicol.* **2016,** *29*, 303-316.

Chapter 2: Computational Insight into the Differential Mutagenic Patterns of O-methylthymine Lesions[a,b,c]

2.1 Introduction

Human DNA is regularly exposed to damaging agents, such as sunlight, tobacco smoke, automobile exhaust, red food dye, and charred red meat.[1] Although there are many repair processes in our bodies that target DNA lesions,[1] some forms of damage are not repaired efficiently, and therefore persist in cells and impact key biological processes. For example, when damaged DNA is replicated, various biological outcomes are possible, including non-mutagenic replication and the formation of base mispairs, as well as insertion and deletion mutations.[2-3] The mutagenic events that occur upon replication of damaged DNA can change our genetic makeup and lead to ailments such as Alzheimer's disease and cancer.[4-5] Characterizing the replication outcomes for different types of DNA modifications is an important first step to understanding the implications of damage formation and the function of key enzymes that process modified DNA. This information can ultimately be used to develop new therapeutic approaches for related diseases.

One common type of DNA damage is nucleobase alkylation. The smallest alkylation products are methylated lesions, which are commonly caused by industrial

---

[a] *The Journal of American Chemical Society* reference style is used in this chapter.

[b] Adapted from Bhutani, P. Nikkel, D.J. Wilson, K.A. Wetmore, S.D. (2019). Computational Insight into the Differential Mutagenic Patterns of O-methylthymine Lesions. *Chemical Research in Toxicology*, (accepted, tx-2019-00291y.R1).

[c] P.B. performed calculations and data analysis on the O4-Me-T base-pair models, as well as nucleoside, polymerase and DNA models for both lesions, and wrote the first draft of the manuscript. D.J.N. performed calculations on the O2-Me-T base-pair models. K.A.W. generated parameters for the O2-Me-T lesion and contributed to manuscript writing. S.D.W. contributed to project and calculation design, data interpretation and manuscript writing.

pollutants, halocarbons in the atmosphere, tobacco smoke, and fruits and vegetables containing myosmine.[6-7] Although alkylation can occur at any of the four canonical DNA nucleobases, O2-methylthymine (O2-Me-T) and O4-methylthymine (O4-Me-T, Figure 2.1) methylation adducts persist in cells due to low repair efficiencies.[8-13] O2-Me-T and O4-Me-T are potentially detrimental since the Watson-Crick (WC) binding face of canonical T has changed (i.e. N3 is a hydrogen-bond acceptor rather than a hydrogen-bond donor, Figure 2.1), which may lead to base mispairing upon replication. To add to this problem, O2 and O4 thymine alkylation lesions typically stall standard (replicative) DNA polymerases,[14-19] leading to replication through the translesion synthesis (TLS) pathway.[20-22] Although TLS polymerases can accommodate a vast array of damaged nucleobases due to their large and flexible active sites, the same properties make these polymerases more error prone.[23-24]

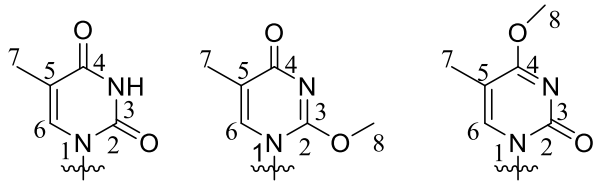


Figure 2.1: Chemical structure and numbering of undamaged T (left), O2-Me-T (middle) and O4-Me-T (right).

Despite similar WC hydrogen-bonding faces, the TLS process results in significantly different mutagenic outcomes for O2-Me-T and O4-Me-T, with the mutagenicity trends prevailing regardless of whether *in vitro* or *in vivo* data are considered.[14-19] Specifically, *in vitro* studies suggest that dATP is primarily inserted opposite O2-Me-T, with low frequencies of mispair formation (< 5% total), while O4-Me-T leads to dGTP insertion almost twice as often as dATP insertion.[16-17] Similarly, studies using *E. coli* cells reveal that O2-Me-T mispairs with dTTP, dCTP and dGTP with

relatively low mutagenicity rates (~ 20%, 10% and 6%, respectively), but O4-Me-T results in a high occurrence of dGTP insertion (~ 70% mutational frequency).[15] Finally, in human cells, O2-Me-T is non-mutagenic (dATP insertion occurs ~ 98% of the time),[19] while O4-Me-T replication causes T → C mutations (dGTP insertion) with ~ 60% frequency.[14] The roles of three human polymerases (pol η, pol κ, and pol ι) and one yeast polymerase (pol ζ) in the replication of O2-Me-T and O4-Me-T have been investigated.[14-19] Pol η has been determined to bypass both lesions, with knockout of pol η significantly decreasing the damage bypass efficiency of O2-Me-T and showing a modest effect on the bypass efficiency of O4-Me-T.[14, 19] Despite this work identifying that pol η is involved in replication, there is little structural information available to explain the differences in the mutagenic outcomes associated with O2-Me-T and O4-Me-T. Additionally, although crystal structures are available for insertion of dATP and dGTP opposite O4-Me-T by pol η, O4-Me-T is not fully bound within the active site, which prevents hydrogen bonding with either dATP or dGTP and generates more questions about how O4-Me-T is replicated.[25]

Computational chemistry has proven to be a useful tool for obtaining valuable insight into the replication of DNA adducts by TLS polymerases.[23, 26-32] Although similar investigations have not yet been conducted to understand the replication of O2-Me-T and O4-Me-T, computational studies have considered select hydrogen-bonded base pairs between the isolated adducts and canonical nucleobases.[33-36] Specifically, semi-empirical methods (PM3 SCF-MO) were initially used to investigate how O4-Me-T interacts with the four canonical nucleobases,[36] although the structural features and strengths of the hydrogen-bonded pairs do not explain the experimental mutagenic outcomes.[14-19, 36]

Additionally, *ab initio* (MP2) calculations have been used to study the O4-Me-T and O2-Me-T hydrogen-bonded pairs with canonical A, but structural parameters of the canonical T:A pair are not maintained and therefore the calculations do not explain the non-mutagenic patterns.[35] Most recently, an *ab initio* (MP2) study on O4-Me-T characterized the structures and stabilities of base pairs involving each of the four canonical bases.[33] Despite uncovering that O4-Me-T can form two hydrogen bonds with G, no structural support for dATP insertion opposite the lesion was obtained.[33] Although these studies provide some insight into O4-Me-T and O2-Me-T hydrogen-bonding patterns, many questions remain about the lesion replication outcomes, including why two similar lesions produce vastly different mutagenic patterns. Thus, further computational studies are required to provide a thorough assessment of the ability of these adducts to form mispairs, and the impact of the DNA and polymerase environments.

To afford the atomic level details of the base mispairing potential of the O2-Me-T and O4-Me-T lesions necessary to understand their observed differential mutagenicity, the present work uses a combination of quantum mechanical (QM) calculations and molecular dynamics (MD) simulations. Specifically, MD simulations are initially used to determine how these adducts are accommodated in the active site of pol η, the human TLS polymerase that most efficiently bypasses these lesions.[14, 19] Subsequently, density functional theory (DFT) calculations are used to understand the hydrogen-bonding patterns between the lesions and select canonical nucleobases. MD simulations are then employed to probe the pol η ternary complexes for insertion of these dNTPs opposite each adduct. Finally, MD simulations reveal the structure, dynamics and stability of post-replication duplexes containing the lesions paired opposite these canonical bases. Together, this work provides

a structural rationalization for the impact of the aberrant methyl location on damaged thymine replication, and the greater mutagenicity of O4-Me-T compared to O2-Me-T. In addition to clarifying the biological impact of pyrimidine methylation damage, this work sets the foundation for studying larger alkylation adducts and enhances our understanding of the function of a key human TLS polymerase (pol η), the absence of which in humans has been directly linked to cancer.[37]

2.2 Methods

2.2.1 DFT Calculations: The potential energy surface (PES) of each lesion was searched with respect to the ∠(N3C4O4C8) or ∠(N3C2O2C8) dihedral angle to determine the favored orientation(s) of the methyl group with respect to the nucleoside. Additionally, the PES of the damaged nucleoside was computed with respect to ∠(O4′C1′N1C2) to determine the rotational barrier about the glycosidic bond. Specifically, B3LYP/6-31G(d,p) optimizations were performed with ∠(N3CXOXC8) or ∠(O4′C1′N1C2) constrained in 10° increments for a total of 36 points and the relative energies were determined using B3LYP-D3(BJ)/6-311++G(2df,2p) single–point calculations. The lowest energy conformation for each lesion with respect to ∠(N3CXOXC8) was subsequently fully optimized and paired opposite select canonical nucleobases. In both the monomer and base-pair models, the deoxyribose moiety was replaced with a methyl group since the aberrant methyl group is not positioned to interact with the DNA backbone, and therefore the sugar moiety is anticipated to have a negligible impact on the lesion structure and binding energies.[38] Optimization and frequency calculations of the monomers and base pairs were performed using B3LYP/6-31G(d,p), while single–point calculations were

completed with B3LYP-D3(BJ)/6-311++G(2df,2p). All binding energies include counterpoise corrections, deformation energies, and scaled (0.9857) zero–point vibrational energy (ZPVE) corrections.[39] The interstrand distance (R(C1′···C1′)) and base pair opening angle (∠(N9C1′C1′)) for each hydrogen-bonded pair were measured using GaussView 5,[40] while the interplanar angle (i.e. angle between nucleobase ring planes) was measured using Mercury.[41] All DFT calculations were performed using Gaussian 09 (revision D.01).[42]

2.2.2 MD Simulations: Polymerase η models were generated from X-ray crystal structures of dATP and dGTP insertion opposite O4-ethylthymine (O4-Et-T, PDB ID: 5DQG and 5DQH, respectively),[25] which are the highest-resolution structures available of an alkylated T lesion correctly poised for replication. The experimental structures were modified in GaussView 5 to insert unresolved residues (T155–E159) and change the damaged site (O4-Et-T) to the methylated T lesions of interest in the present work. For the pol η pre-insertion complexes (i.e, the DNA–polymerase complex prior to dNTP insertion), the active site Mg(II) ions were removed because of evidence that shows these bind with the incoming dNTP and do not play a role until dNTP binding.[43-44] Glycerol used in the crystallization process was removed for all complexes. Protein residues with multiple orientations were visually inspected to determine the most likely conformation based on possible interactions with neighboring residues. The DNA sequence was edited using GaussView 5 to the sequence used in experimental work on the O-methylated-T lesions (namely 5′-GXGCTATGC-3′, where X is the damage site).[14-19, 45] Partial charges for the O2-Me-T and O4-Me-T lesions were generated using the RED.vIII.4 script,[46] and parm10 and GAFF parameters were assigned using Antechamber 1.4 (Table A.1 and A.2).[47] The AMBER

OL15 force field was used to describe DNA and AMBERff14SB parameters were used for the enzyme.[48] dNTP parameters were adopted from previous work,[49] while the Allner *et al*. parameters were used for active site Mg(II) ions.[50] Na^+ ions were added to neutralize the system, while Na^+ and Cl^- ions were added to achieve a physiological salt concentration of 150 mM. All systems were solvated using an octahedral TIP3P water box with edges 10 Å from the solute in each direction.[51]

The pol η pre-insertion and ternary complexes were subjected to seven minimization steps, each using 5000 steps of steepest descent and 5000 steps of conjugant gradient minimization, while placing a restraint of 50 kcal mol^{-1} $Å^{-2}$ on regions of the system for the first six steps. Specifically, the solvent was initially minimized, followed by the hydrogen atoms. Subsequently, the side chains of the enzyme were minimized, followed by both the enzyme side chains and DNA helix (excluding the lesion base pair). The lesion base pair was then included in the minimization step, followed by the solute. Finally, the entire system was minimized. The system was then heated using the Langevin thermostat ($\gamma = 1$) to 310 K in 50 K increments. The solute then underwent 5 phases of equilibration using a 2 fs time step, while the force restraint constant was lowered from 20 kcal mol^{-1} $Å^{-2}$ to 1.5 kcal mol^{-1} $Å^{-2}$. A non-bonded cut-off of 10 Å was implemented, as well as the periodic boundary condition and particle-mesh Ewald algorithms to describe electrostatic interactions. Each pre-insertion and ternary complex was then simulated at a constant temperature (310 K) and pressure (1 atm) for 1 μs. Interaction energies and interplanar angles were computed using an in-house script to extract lesion nucleobase pairings every 5 ns. The interaction energies were evaluated using B3LYP-D3(BJ)/6-311++G(2df,2p) counterpoise-corrected calculations, as implemented in G09 (revision

D.01), on models capped with hydrogen atoms at the glyosidic bond truncation points (N1 for pyrimidines and N9 for purines). Hydrogen bonding and structural analyses, including the Mg(II) coordination and reaction parameters, were conducted every 0.05 ns using CPPTRAJ in Amber14.[52]

Select lesion base pairs were inserted into a 16-mer DNA strand 5′-GCATGGCGXGCTATGC-3′ at position X, which parallels the sequence used in experimental work on these adducts,[14, 19] with two additional G:C pairs added to both strand ends to prevent duplex unraveling. The DNA duplex was built using the nucleic acid builder (NAB) in Amber14[45] and the damaged base inserted using GaussView 5. Na^+ counterions were added to neutralize the system using the tleap program in Amber14.[45] The damaged duplexes were described using the AMBER OL15 force field. The resulting structures were minimized in four phases, using 2500 steps of steepest descent and 2500 steps of conjugate gradient, and a 50 kcal mol^{-1} $Å^{-2}$ force constant restraint for a portion of the system. Specifically, the first phase minimized the solvent only, the second phase minimized the hydrogen atoms, the third phase minimized the solute, and the final phase minimized the entire system. The system was then heated and equilibrated, and production simulations performed using the same protocol discussed for the polymerase models. Interaction energies, interplanar angles, hydrogen-bonding patterns and other structural parameters were analyzed as described for the polymerase models.

2.3 Results and Discussion

2.3.1 The position of O4-Me-T is stable in the pol η pre-insertion complex: In a crystal structure representing dATP insertion opposite O4-Me-T by pol η (PDB ID: 5DLF),[25] the

lesion is not positioned in the active site in a manner that permits hydrogen bonding with an opposing dNTP.[25] Furthermore, there is currently little information about the orientation and flexibility of the aberrant methyl group in the polymerase active site or the impact of the change in the WC binding face of T upon damage on nucleoside accommodation by pol η.[25] Therefore, the pol η pre-insertion complex corresponding to O4-Me-T replication was initially considered to assess how the lesion is accommodated within the polymerase active site prior to dNTP insertion. Throughout the MD simulation, the aberrant methyl group of O4-Me-T is well accommodated within the open pocket of pol η on the major groove side of the DNA (Figure A.1a). As a result, the O4 methyl group of the lesion is void of interactions with active site residues and is flexible (∠(N3C4O4C8) = 13.8 ± 39.1°, Figure A.1a). Since the damage does not affect the minor groove of canonical T, key protein interactions with the lesion are preserved. Specifically, Q38 has been suggested to play a role in stabilizing the position of the template base through hydrogen bonding with O2 of pyrimidines or N3 of purines, which maintains alignment of the template base for the incoming dNTP.[27, 32, 53-55] Indeed, this Q38 interaction occurs in the crystal structure corresponding to dATP insertion opposite natural T by pol η,[56] and Q38 interacts with O2 of the 3′-T in the TT-dimer in MD simulations (29% occupancy).[27, 32] In the O4-Me-T pre-insertion complex, Q38 periodically interacts with O2 (24%) and O4′ (45%) of the lesion throughout the simulation (Figure 2.2a). These contacts help O4-Me-T maintain stacking interactions with the 3′-intrastrand flanking base (average nucleobase–nucleobase stacking interaction = –35.3 kJ/mol; Figure 2.2b). As a result, unlike observed in crystal structures,[25] the lesion is positioned in the pol η active site to permit replication, with an open void occurring opposite the lesion for an incoming dNTP to bind. Nevertheless, the aberrant

methyl group is primarily directed towards the lesion WC face, which may disrupt base pairing with a future incoming dNTP (Figure A.1a).

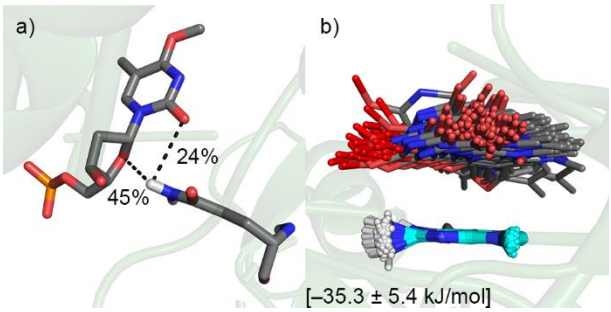


Figure 2.2: a) Hydrogen bonding between O4-Me-T and Q38 in the pol η pre-insertion complex. b) MD snapshots taken at 5 ns intervals overlaid according to the ring atoms of 3′-G with respect to O4-Me-T and the associated stacking energy.

2.3.2 O4-Me-T intrinsically prefers to pair with G: Since the aberrant methyl group location may impact canonical WC hydrogen-bonding patterns, the base-pairing properties of O4-Me-T must be considered. Focus will be placed on dATP and dGTP insertion since these are the primary replication outcomes under several experimental conditions.[15-16, 19] However, prior to considering O4-Me-T:dNTP pairing within the pol η active site, potential hydrogen-bonding patterns between the O4-Me-T and A or G nucleobases were characterized using DFT. To provide a reference point to understand the structure and stability of the lesion pairs, the natural T:A Watson-Crick pair was characterized (Figure A.2), which contains two strong O4(T)···H–N6(A) and N3–H(T)···N1(A) hydrogen bonds. The two canonical bases adopt a planar orientation (interplanar angle = 0.0°), while the interstrand distance (R(C1′···C1′)) and base pair opening angle (∠(N9C1′C1′)) are 10.576 Å and 55.2°, respectively. Overall, these structural features result in an interaction energy of –69.8 kJ/mol, which is similar to a previously reported CCSD(T) value (–68.6 kJ/mol).[57]

Although one conformation of the O4-Me-T methyl group dominates during MD simulations on the pol η pre-insertion complex directing the methyl group toward the lesion WC face (Figure A.1a), multiple orientations can be envisioned. DFT calculations reveal two shallow minima in which the methyl group is directed out of the O4-Me-T nucleobase plane (± 130°, Figure A.3). Nevertheless, a small (~ 5 kJ/mol) barrier connects these minima to the ~ 50 kJ/mol more stable conformation that points the methyl group toward the lesion WC face (Figure A.3). Since this structural preference aligns with the dominant conformation predicted by MD simulations on the pol η pre-insertion complex, the most stable O4-Me-T conformation was paired with A and G to determine the hydrogen-bonding potential of the lesion.

As mentioned in the Introduction, the WC hydrogen-bond face of T is altered upon methylation at O4 such that N3 becomes a hydrogen-bond acceptor. As a result, the O4-Me-T:A pair is highly distorted, with the N1–C1′ bond of O4-Me-T and the N9–C1′ bond of A directed in a parallel orientation (Figure 2.3a), which is consistent with previous computational work.[33, 35-36] This base-pair structure cannot be accommodated within the context of B-DNA. Furthermore, the corresponding interaction energy (–40.2 kJ/mol) is significantly less than that of natural T:A (–69.8 kJ/mol) due to the presence of only one strong hydrogen bond upon damage (N3(O4-Me-T)···H–N6(A)). In contrast, the O4-Me-T:G mispair contains two strong hydrogen bonds (O2(O4-Me-T)···H–N2(G) and N3(O4-Me-T)···H–N1(G), Figure 2.3a). Although most structural parameters of the G mispair are comparable to undamaged DNA, the bases adopt a nonplanar arrangement (interplanar angle = 44.1°), which affords additional interactions between O6 of G and the O4 methyl substituent. As a result, the O4-Me-T:G base pair has an interaction energy

(–78.0 kJ/mol) that is slightly stronger than canonical T:A (–69.8 kJ/mol) and significantly stronger than O4-Me-T:A (–40.2 kJ/mol).

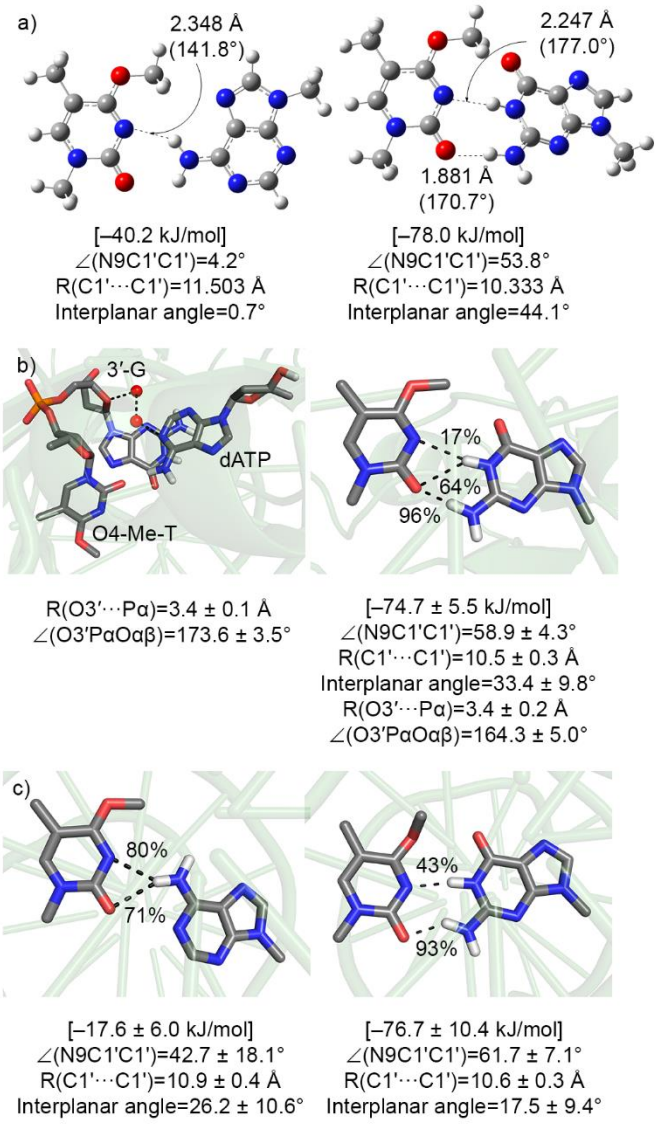


Figure 2.3: Key structural features (Å and °), binding strengths [kJ/mol], hydrogen-bond occupancies (%) and reaction parameters (Å and °) where applicable for O4-Me-T pairs with A (right) or G (left) from a) isolated nucleobase models, b) pol η ternary complexes, and c) damaged DNA duplexes.

Overall, the inherently distorted O4-Me-T:A pair correlates with the reported high mutagenicity of O4-Me-T.[15-16, 19] Furthermore, the DFT models indicate that isolated O4-Me-T forms a viable base mispair with G, which is the most commonly observed mutation

upon O4-Me-T replication.[15-16, 19] Nevertheless, based on the structure and stability of O4-Me-T:A, it is unclear how dATP is inserted by pol η under a variety of experimental conditions,[15-16, 19] including ~ 40% insertion in human cells.[19] To better understand these replication outcomes, the pol η ternary complexes corresponding to dATP and dGTP insertion opposite O4-Me-T are considered in the following section.

2.3.3 Ternary complexes and post-extension duplexes support both A and G insertion opposite O4-Me-T by pol η: There are several structural features known to be important for a polymerase–DNA complex to be conducive for dNTP insertion.[44, 49, 58] First, the incoming dNTP must be properly bound within the active site. In a crystal structure corresponding to dATP insertion opposite canonical T (PDB ID: 4ECS),[56] dATP is positioned in the pol η active site through direct interactions with C16, F18, Y52, R55, R61 and K231. Over the course of the MD simulation, there is little movement in these key active site residues that surround the incoming dATP or dGTP relative to the crystal structure corresponding to natural T replication by pol η (rmsd=0.7 Å, Figure A.4). Second, for the insertion complex to be conducive for the reaction, the Mg(II) ions must maintain octahedral coordination (Figure A.5a). In the O4-Me-T ternary insertion complexes, all residues maintain a Mg(II) binding distance < 2.5 Å for > 87% of the simulation (Tables A.3 and A.4). Third, the dNTP must be properly aligned for the nucleotidyl transfer reaction, with ideal reaction parameters including a O3′–Pα distance (R(O3′···Pα)) < 3.5 Å and a ∠(O3′–Pα–Oαβ) angle close to 180° (Figure A.5b). Average reaction parameters over the MD simulations for both dATP and dGTP incorporation are favourable, with an average reaction distance (angle) of 3.4 Å (173.8°) and 3.2 Å (166.4°), respectively. Thus,

both dATP and dGTP can be accommodated in a position in the pol η active site that is conducive for insertion. However, in addition to the dNTP being correctly bound, the lesion must be properly oriented within the pol η active site.

For dATP insertion, the methyl group of O4-Me-T primarily adopts one conformation throughout the MD simulation on the pol η insertion complex (∠(N3C4O4C8) = –1.6 ± 40.1°), which directs the methyl group toward the incoming dATP (Figure A.6a). However, interactions between Q38 and O2 of the lesion template base occur for only 4% of the simulation and Q38 interactions with the lesion deoxyribose (< 18%, Figure A.7) are less frequent than in the pre-insertion complex. In fact, the O4-Me-T lesion primarily resides in the open pocket of pol η on the major groove side of DNA (Figure 2.3b). As a result, no persistent hydrogen bond occurs between the damaged template base and the incoming dATP.

Despite no direct hydrogen bonding between O4-Me-T and dATP, a water chain occurs between N1 of dATP and O4′ of the 3′-interstrand nucleotide sugar on the template strand throughout the simulation (Figure 2.3b). This is similar to interactions in a crystal structure corresponding to dATP insertion opposite an abasic site by pol η,[59] where a water chain forms between N3 of dATP and the 3′-dA sugar of the template strand, and additional water chains form between N1 and N6 of dATP and the 5′-phosphate backbone of the template strand (Figure A.8a). The subtle differences in the position of the water chain in the crystal structure and these MD simulations could arise because of numerous factors including inherent differences between a template abasic site and the O4-Me-T lesion, and the impact of structural dynamics in these simulations compared to the static X-ray structure. Regardless, the mechanism for the preferential insertion of A opposite O4-Me-T

is anticipated to be similar to the preferential insertion of A opposite an abasic site (the so-called "A-rule").[60] Furthermore, previous computational work[26] on pol η replication of O2-4-(3-pyridyl)-4-oxobut-1-yl-T (O2-POB-T), a lesion that arises from exposure to tobacco smoke, has used an analogous predicted active site configuration to explain the observed replication outcomes,[61] with a water chain forming between N1 of dATP and the 3′-flanking base as reported herein for O4-Me-T. Thus, despite O4-Me-T not being aligned for dNTP insertion in the pol η pre-insertion complex, these MD simulations suggests that dATP can be accommodated opposite the lesion through a mechanism similar to that for dATP insertion opposite an abasic site[59] and other T alkylation lesions (O2-POB-T).[26] Further support for the O4-Me-T:A pairing comes from MD simulations on post-extension damaged duplexes, which reveal that the lesion does not impact the overall structure of B-DNA (Figure A.9a), and interactions between N3 or O2 of O4-Me-T and N6–H of A result in a stable base pair (average interaction energy = –17.6 kJ/mol, Figure 2.3c).

For dGTP incorporation, the O4-Me-T methyl group is directed toward the incoming dGTP (∠(N3C4O4C8) = –0.6 ± 37.6°, Figure A.6b) and the lesion flexibility is comparable to that discussed for dATP insertion. Q38 interacts with O2 of the damaged T for 25% and O3′ for 14% of the simulation trajectory (Figure A.7). These interactions help stabilize the lesion position in the active site and maintain a suitable template base alignment for hydrogen bonding with the incoming dGTP. Specifically, a O2(O4-Me-T)···(N2–H)G hydrogen bond is maintained for 96% of the simulation, while two additional hydrogen bonds between O2 (64%) or N3 (17%) of O4-Me-T and N1–H of G also occur (Table A.5 and Figure 2.3b). This hydrogen-bonding pattern is similar to what is observed in the crystal structure corresponding to dGTP insertion opposite O4-Et-T.[25]

dGTP maintains an interstrand distance with respect to the template base comparable to natural DNA (Figure 2.3b), and the resulting lesion base pair is very stable (average interaction energy = –74.7 kJ/mol). The propensity of O4-Me-T to base mispair is underscored by the persistence of this hydrogen-bonding pattern and stability in post-extension damaged duplexes (Figure 2.3c).

Overall, these MD simulations suggest that the pol η ternary complexes for dATP and dGTP insertion opposite O4-Me-T are catalytically conducive. Although O4-Me-T:A maintains interstrand hydrogen bonding and does not distort the canonical duplex, stable WC hydrogen bonding does not occur in the insertion complex. Instead, the damaged T is located in the major groove of the growing DNA strand and dATP binding is stabilized by a water chain to the template strand, as previously reported for abasic site and O2-POB-T replication.[26] In contrast, dGTP forms two hydrogen bonds with the damaged T, which is held in a catalytically conducive position in the pol η active site by key residues (Q38). Additionally, the G mispair is much more stable than the A pair within the duplex environment. Thus, this data rationalizes the experimentally observed mutagenicity of O4-Me-T.[15-16, 19]

2.3.4 The position of O2-Me-T in the pol η pre-insertion complex is dynamic: Since O2 and O4-Me-T impact the WC binding face of T in a similar fashion, it is unclear why O2-Me-T is non-mutagenic,[14, 17-18] while O4-Me-T commonly mispairs with G.[15-16, 19] As discussed for O4-Me-T replication, the O2-Me-T aberrant methyl group is directed toward the WC binding face of the lesion throughout MD simulations on the pol η pre-insertion complex (average ∠(N3C2O2C8) = –6.7 ± 33.4°, Figure A.1b). Although pol η has an open

pocket on the major groove side of DNA, the damaged moiety of O2-Me-T falls in the DNA minor groove and therefore disrupts Q38 in the polymerase active site. Specifically, Q38 hydrogen bonds with O2 of O2-Me-T for 26% of the simulation, and interacts with O4′ of the lesion for only 9% and O3′ for only 5% of the simulation (Figure 2.4a). These interactions are significantly less persistent than those in the O4-Me-T pre-insertion complex. Additionally, the O2-Me-T orientation in the pol η active site is highly variable (Figure 2.4b), which reduces the 3′-interstrand stacking interaction energy (–13.5 kJ/mol) compared to the O4-Me-T pre-insertion complex (–35.3 kJ/mol). As a result, although O2-Me-T is aligned to interact with an incoming dNTP in the pol η active site, the lesion position is dynamic.

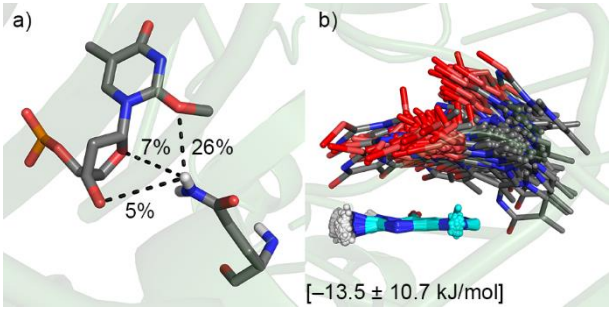


Figure 2.4: a) Hydrogen bonding between O2-Me-T and Q38 in the pol η pre-insertion complex. b) MD snapshots taken at 5 ns intervals overlaid according to the ring atoms of 3′-G with respect to O2-Me-T and the associated stacking energy.

2.3.5 Despite inherent distortion in the isolated O2-Me-T:A pair, ternary complexes and post-replication duplexes support dATP insertion by pol η: As discussed for O4-Me-T, the hydrogen-bonding pattern in the non-mutagenic O2-Me-T:A pair was first investigated using DFT nucleobase models. In contrast to O4-Me-T, only one minimum exists for isolated O2-Me-T, which directs the methyl group toward the lesion WC face (∠(N3C2O2C8) = 0°), and the rotational barrier about the nucleobase–methyl linker is relatively high (~ 78 kJ/mol, Figure A.10). These features underscore the low inherent flexibility of O2-Me-T, as observed in the MD simulations on the pol η pre-insertion complex. The alteration of the WC hydrogen-bonding face of T upon methylation at O2 results in only one strong N3(O2-Me-T)···H–N6(A) hydrogen bond with complementary A (Figure 2.5a). Nevertheless, interplanar distortion in the base pair (by 37.5°) affords weak C–H···N interactions between the adducted methyl group and N1 of A. However, the interstrand distance is ~ 1 Å greater than canonical DNA, and the stability of O2-Me-T:A (–40.9 kJ/mol) is reduced compared to canonical T:A (–69.8 kJ/mol).

To understand the impact of the pol η active site on the distorted O2-Me-T:A pair, the pol η ternary complex corresponding to dATP insertion opposite O2-Me-T is considered. Similar to O4-Me-T, there is little movement in key active site residues that surround the incoming dATP in the O2-Me-T insertion complex, including C16, F18, Y52, R55, R61 and K231 (rmsd = 0.6 Å, Figure A.11a). Furthermore, all Mg(II) binding residues adopt a coordination distance of < 2.5 Å for > 95% of the simulation (Table A.6), and the average reaction parameters are favourable (distance = 3.4 Å; angle = 173.5°). Thus, the incoming dATP is well aligned for the nucleotidyl transfer reaction throughout the MD trajectory.

As discussed for the pre-insertion complex, the methyl group of O2-Me-T primarily adopts one conformation throughout the simulation, which directs the damaged group toward the lesion WC face (∠(N3C2O2C8) = –14.5 ± 25.2°, Figure A.12a). However, as observed for O4-Me-T, O2-Me-T is positioned toward the open pocket of pol η (Figure 2.5b). This occurs since Q38 interacts with each of O2 and deoxyribose of the lesion template base for less than 10% of the simulation (Figure A.7). Since O2-Me-T is not bound within the pol η active site, no persistent hydrogen bonds occur between the damaged template base and the incoming dATP (Figure 2.5b). Nevertheless, a water chain forms between N3 of dATP and O4′ of the 3′ sugar on the template strand. This suggests that the experimentally observed[15-16, 19] O2-Me-T:A pairing likely arises due to a similar water-mediated pathway as dATP insertion opposite O4-Me-T, as well as an abasic site and other alkylation T lesions (O2-POB-T).[26] However, although the overall B-DNA structure is maintained in post-extension duplexes containing O2-Me-T:A (Figure A.13), hydrogen bonding at the lesion site only occurs intermittently, which results in a less stable damaged pair than canonical T:A (Figure 2.5c). Thus, despite these calculations clarifying a potential pathway for dATP insertion opposite O2-Me-T, the significant flexibility of O2-Me-T in the pol η pre-insertion complex, the reduced stability of O2-Me-T:A in the pol η active site and DNA duplexes, and the similar insertion pathway proposed for O2 and O4-Me-T raise questions about why mispairs do not occur for O2-Me-T, while O4-Me-T exhibits a high frequency of dGTP insertion.[15-16, 19]

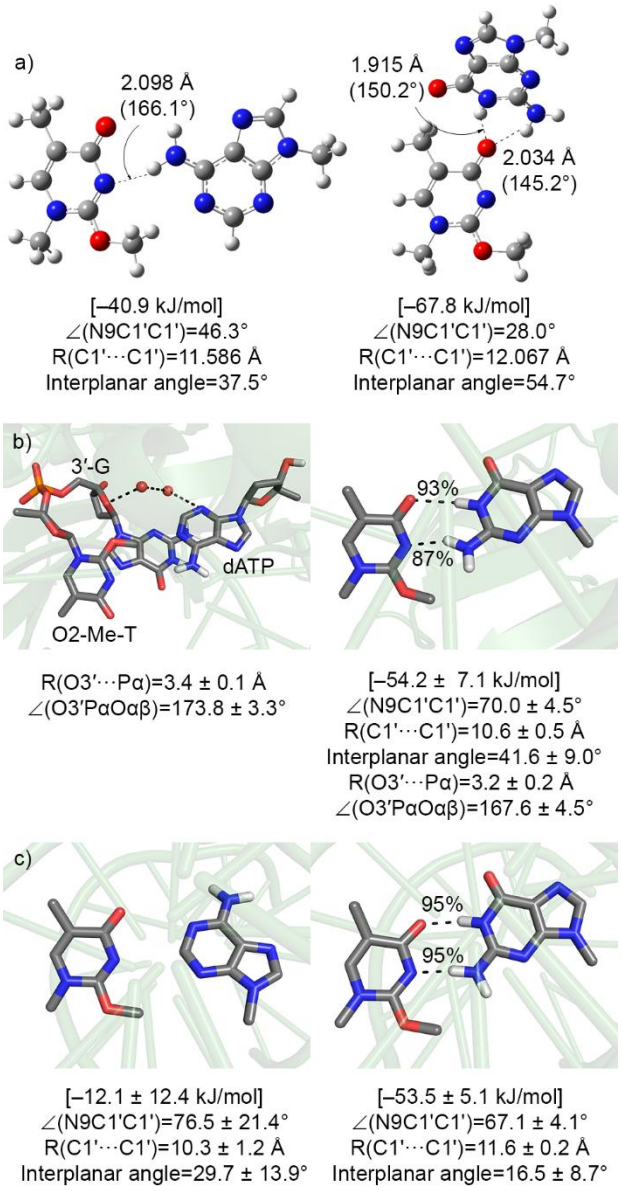


Figure 2.5: Key structural features (Å and °), binding strengths [kJ/mol], hydrogen-bond occupancies (%) and reaction parameters (Å and °) where applicable for O2-Me-T pairs with A (right) or G (left) from a) isolated nucleobase models, b) pol η ternary complexes, and c) damaged DNA duplexes.

2.3.6 MD simulations on the ternary complex support dGTP insertion opposite O2-Me-T, contrasting experimental evidence: To understand why O2-Me-T replication does not result in base mispairing, the insertion of dGTP opposite the lesion is considered since this is the major replication outcome associated with O4-Me-T.[15-16, 19] Unlike O4-Me-T,

DFT models suggest that O2-Me-T:G is highly distorted due to steric repulsion between the O2-Me-T methyl group and the N2 amino group of G (Figure 2.5a). Although this nicely correlates with O2-Me-T primarily leading to non-mutagenic replication,[14, 17-18] two strong hydrogen bonds form between O2-Me-T and the incoming dGTP in the pol η ternary complex (O4(O2-Me-T)···H–N1(dGTP) and N3(O2-Me-T)···H–N2(dGTP), Figure 2.5b and Table A.5). Indeed, a stable base pair (–54.2 kJ/mol) occurs within the pol η active site that is the width of canonical DNA (10.6 Å), albeit slightly nonplanar (41.6°). In contrast to dATP insertion, the lesion is correctly positioned for hydrogen bonding with dGTP since persistent Q38 interactions occur with O2 (16%), O3′ (34%) and O4′ (17%) of O2-Me-T (Figure A.7). Furthermore, stabilizing contacts exist between dGTP and key active site residues (C16, F18, Y52, R55, R61 and K231; rmsd=0.5 Å, Figure A.11b), the Mg(II) coordination is maintained (coordination distances ≤ 2.5 Å for 99% of the simulation, Table A.7), and the reaction parameters remain favorable (R(O3′···Pα) =3.2 Å and ∠(O3′–Pα–Oαβ) = 168.1°). If extension occurs past the lesion site, the structure and stability of the O2-Me-T:G base pair is generally maintained in the resulting DNA duplex (two strong hydrogen bonds with 95% occupancy and average interaction energy = –53.5 kJ/mol). Additionally, there is no major distortion to the helical structure (Figure A.13b), even though the interstrand distance is elongated at the damaged site (11.6 Å) compared to canonical DNA (Figure 2.5c). Overall, the G mispair is less stable for O2-Me-T compared to O4-Me-T due to differences in the hydrogen-bonding patterns arising from the location of the aberrant methyl group. Nevertheless, this data suggests O2-Me-T can be paired opposite dGTP within the context of the pol η active site and DNA duplex environment, which contrasts the experimental mutagenic patterns.[14, 17-18] However, previous

computational work on another O2-alkylation T lesion, namely O2-POB-T, rationalized observed mutagenic outcomes[61-62] based on the lesion undergoing a conformational change to adopt the *syn* glycosidic orientation prior to replication (Figure A.14a).[26] Although O2-Me-T is a much smaller lesion than O2-POB-T, the significant dynamics of bound O2-Me-T, and the associated disruption of lesion contacts with Q38 and the 3′-flanking base suggest even methylation damage at O2 of T can prevent the template base from being sufficiently stabilized within the pol η active site to permit hydrogen bonding with an incoming dNTP. Therefore, to further explore the apparent discrepancy between experimental replication outcomes and the simulation data for O2-Me-T, replication of the *syn* lesion orientation must be considered.

2.3.7 *syn*-O2-Me-T is better accommodated in the pol η pre-insertion complex than *anti*-O2-Me-T: The ability of O2-Me-T to adopt the *syn* conformation is supported by nearly isoenergetic *syn* and *anti* nucleoside minima (within ~ 3 k/mol) connected by a modest rotational barrier (~ 43 kJ/mol, Figure A.14b). When the *syn* glycosidic orientation of O2-Me-T is situated in the pol η active site, the aberrant methyl group is directed into the open pocket of pol η (Figure A.1c), and therefore would not interfere with an incoming dNTP. Although the change in glycosidic orientation prohibits interactions between Q38 and O2 of the lesion, Q38 forms hydrogen bonds with the O2-Me-T deoxyribose (78% with O4' and 7% O3', Figure 2.6a), which are significantly more persistent than those discussed for *anti*-O2-Me-T. Furthermore, although *anti*-O2-Me-T exhibits significant flexibility in the active site and limited stacking with the 3'-flanking base (Figure 2.4b), the *syn* orientation is more restricted and maintains strong stacking with 3'-G (–40.8 kJ/mol,

Figure 2.6b). Thus, in contrast to the *anti* lesion conformation, *syn*-O2-Me-T is well accommodated in the pol η active site and securely situated to interact with an incoming dNTP.

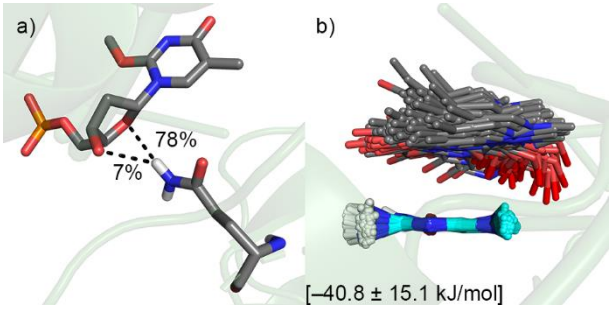


Figure 2.6: a) Hydrogen bonding between *syn*-O2-Me-T and Q38 in the pol η pre-insertion complex. b) MD structures taken at 5 ns intervals overlaid according to the ring atoms of 3′-G with respect to *syn*-O2-Me-T and the associated stacking energy.

2.3.8 Predicted pol η ternary structure is not conducive for dGTP insertion opposite *syn*-O2-Me-T, which correlates with the reported non-mutagenic replication: Despite the energetic accessibility of the *syn*-O2-Me-T conformation and accommodation of this orientation in the pol η active site in the absence of an incoming dNTP, the reduced hydrogen-bonding potential of the O2-Me-T Hoogsteen binding face causes G to reposition and interact with the WC binding face in isolated nucleobase models (Figure 2.7a). This orientation aligns the sugars in a parallel fashion that is not feasible in a DNA duplex. Furthermore, there is a stark lack of persistent hydrogen bonding at the lesion site in the corresponding ternary complex (Figure 2.7b). Specifically, in the pol η ternary complex corresponding to dGTP insertion opposite *syn*-O2-Me-T, dGTP is nicely bound within the active site through interactions with key residues (rmsd = 0.6 Å, Figure A.15), the Mg(II) coordination is appropriate (99% < 2.5 Å, Table A.8), and the reaction parameters are aligned for insertion (R(O3′···Pα) = 3.4 Å and ∠(O3′–Pα–Oαβ) = 149.0°). However, the

lesion is positioned in the open pocket of pol η on the major groove side of DNA due to restricted contacts with Q38, which solely interacts with O3′ of the lesion (45%, Figure A.7). As a result, there are no direct noncovalent contacts with the incoming dGTP (Figure 2.7b), which would impede dGTP insertion.

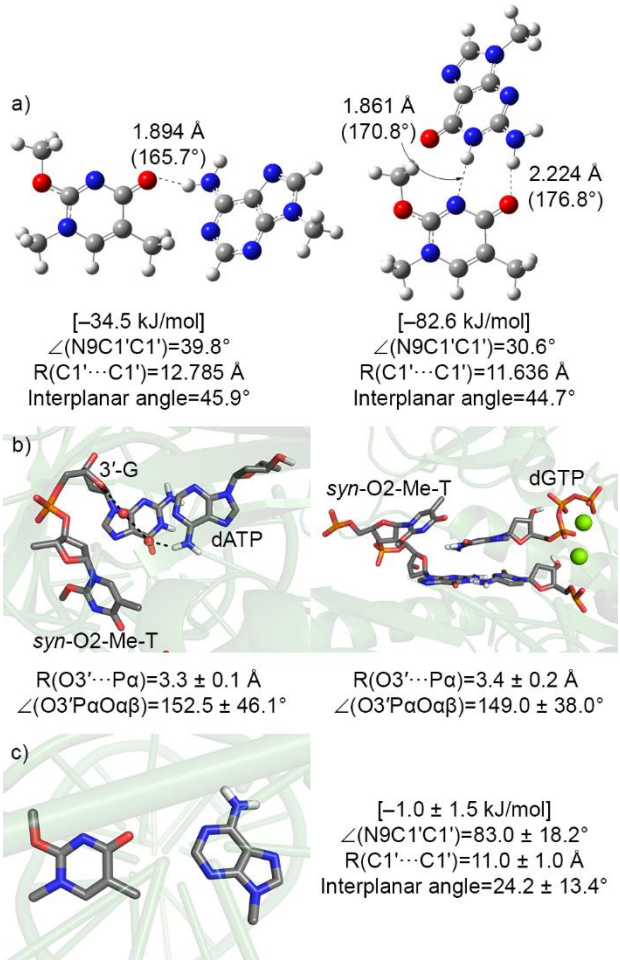


Figure 2.7: Key structural features (Å and °), binding strengths [kJ/mol], hydrogen-bond occupancies (%) and reaction parameters (Å and °) where applicable for *syn*-O2-Me-T pairs with A (right) or G (left) from a) isolated nucleobase models, b) pol η ternary complexes, and c) damaged DNA duplex containing the *syn*-O2-Me-T:A pair.

Nevertheless, analogous to dATP insertion opposite O4-Me-T and *anti*-O2-Me-T, it is possible that dGTP insertion opposite *syn*-O2-Me-T could occur through a mechanism similar to dGTP insertion opposite an abasic site. Indeed, experimental evidence supports

dGTP insertion opposite an abasic site by pol η,[59] which was proposed to be facilitated by direct interactions with Q38 that help align the incoming dGTP, and a water chain between O6 of dGTP and the phosphate backbone of the lesion (Figure A.8b). However, in MD simulations on the DNA–polymerase complex corresponding to dGTP insertion opposite *syn*-O2-Me-T, Q38 only interacts with deoxyribose of the lesion (45%, Figure A.7), and the *syn*-O2-Me-T nucleotide orientation prevents a water chain between O6 of dGTP and the template backbone. Therefore, it is unlikely that dGTP will be inserted opposite *syn*-O2-Me-T based on a similar mechanism to abasic site replication.[59]

Overall, based on the distorted isolated pairs and ternary pol η complex, there is no driving force for dGTP insertion opposite *syn*-O2-Me-T. As a result, the post-extension duplex containing the *syn*-O2-Me-T:G pair was not considered. Nevertheless, the adoption of an alternative glycosidic orientation provides a structural explanation for experimental work that indicates dGTP is not inserted opposite this lesion.[14] This suggestion parallels the lesion orientation previously proposed based on computational work for the larger O2-POB-T lesion,[26] which also rationalized the experimentally-observed mutagenic outcomes.[61-62]

2.3.9 dATP can be inserted opposite *syn*-O2-Me-T by pol η, providing an alternate explanation for non-mutagenic lesion bypass: Despite the *syn* lesion conformation providing an explanation for the lack of dGTP insertion opposite O2-Me-T by pol η, it is important to understand whether dATP insertion can readily occur for this lesion glycosidic orientation, which could account for the observed non-mutagenic bypass.[14, 17-18] Similar to *anti*-O2-Me-T, DFT models suggest the *syn*-O2-Me-T:A pair contains a large interstrand

distance (12.785 Å) and only one strong (O4(O2-Me-T)···H–N6(A)) hydrogen bond (Figure 2.7a), which leads to a less stable pair (–34.5 kJ/mol) than canonical T:A (–69.8 kJ/mol). Nevertheless, the incoming dATP remains well positioned throughout MD simulations on the pol η ternary complex for *syn*-O2-Me-T replication. Specifically, there is little movement in key active site residues surrounding dATP (rmsd = 0.7 Å, Figure A.15), the Mg(II) coordination is maintained (91% < 2.5 Å, Table A.9), and the reaction parameters are comparable to ideal values (R(O3′···Pα) =3.3 Å and ∠(O3′–Pα–Oαβ) = 152.5°). In contrast, *syn*-O2-Me-T is positioned in the open pocket of pol η with the aberrant methyl group directed toward the WC binding face (∠(N3C2O2C8) = –14.1 ± 23.7°, Figure A.16) and Q38 interactions are restricted to deoxyribose of the lesion (69%, Figure A.7). However, despite no direct interactions with the incoming dATP (Figure 2.7b), a water chain forms between N6 of dATP and O4′ of the 3′-base with respect to the lesion. This is similar to the water chain discussed for dATP insertion opposite *anti*-O2-Me-T and O4-Me-T. Although the resulting base pair lacks strong lesion site interactions in the fully-extended DNA duplex (Figure 2.7c), the helix is not distorted outside the lesion site upon incorporation of *syn*-O2-Me-T (Figure A.17). Thus, pol η likely bypasses *syn*-O2-Me-T in a manner similar to an abasic site[59] and other T alkylation products.[26] Interestingly, regardless of the lesion glycosidic orientation, this data supports the non-mutagenic replication of O2-Me-T. The fact that dATP can be accommodated opposite both the *anti* and *syn* conformations of O2-Me-T likely contributes to the strong preference for dATP insertion opposite the lesion.[14, 17-18] Nevertheless, only replication in the *syn* orientation rationalizes the lack of dGTP insertion opposite O2-Me-T.[14, 17-18] As this is the second report of the replication of an alkylated pyrimidine lesion in the *syn* orientation,[26]

both of which form at the O2 position of T, this data hints at a possible unified mechanism for the replication of such DNA damage products.

2.4 Conclusion

Although both O2-Me-T and O4-Me-T are resistant to repair[8-13] and similarly modify the WC binding face of canonical T, these lesions exhibit vastly different mutagenic outcomes upon replication, with O4-Me-T preferentially resulting in T $\rightarrow$ C mutations (dGTP insertion)[15-16, 19] and O2-Me-T being non-mutagenic.[14, 17-18] To understand this differential mutagenicity, the present study employs a multiscale computational approach to investigate lesion replication by the human TLS polymerase η, the enzyme responsible for bypass of these DNA adducts.[14-19] Although O4-Me-T does not intrinsically form a stable base pair with an opposing A, dATP insertion opposite the lesion is stabilized in the pol η active site by water-mediated hydrogen bonding to O4′ of the 3′ flanking base and the corresponding damaged duplex maintains the structure of canonical B-DNA. Therefore, dATP insertion opposite O4-Me-T likely occurs through a mechanism similar to the "A-rule" that has been previously established for the replication of an abasic site by pol η.[59] Nevertheless, the suitable structure and high stability of the O4-Me-T:G pair in isolation, within the polymerase active site and in DNA duplexes rationalize the preferential insertion of dGTP opposite the O4-Me-T lesion.[15-16, 19] Despite reported differences in lesion mutagenicity, this data for *anti*-O2-Me-T parallels that for O4-Me-T, with dATP insertion opposite the lesion being stabilized by water hydrogen-bonding networks and the *anti*-O2-Me-T:G pair being a viable replication outcome. In contrast, these calculations suggest that the *syn*-O2-Me-T conformation is better accommodated in

the pol η active site than the *anti* orientation. Furthermore, the predicted replication outcomes for O2-Me-T in the *syn* glycosidic orientation nicely align with experimental work that suggests this lesion is non-mutagenic. Specifically, when dGTP insertion is considered, *syn*-O2-Me-T is poorly positioned in the pol η open pocket and the complex is void of direct or water mediated interactions that would promote dGTP incorporation. Nevertheless, dATP can be accommodated opposite *syn*-O2-Me-T via a water chain similar to that observed for the *anti* lesion orientation. In addition to agreeing with the experimental mutagenic outcomes, the proposed replication of O2-Me-T in the *syn* glycosidic orientation directly aligns with the previously reported replication pathway for the O2-POB-T lesion.[26] Overall, this chapter rationalizes the differential mutagenic patterns of O4-Me-T and O2-Me-T, and highlights an emerging theme of the pol η bypass of minor groove O2-T lesions in a unique *syn* glycosidic orientation. Future work must continue to probe the replication of other thymine alkylation adducts to further understand the effects of the thymine damage site and composition on replication outcomes.

2.5 References

(1) Friedberg, E. C. DNA damage and repair. *Nature* **2003,** *421,* 436-440.

(2) Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **1993,** *362,* 709-15.

(3) Sander, M.; Cadet, J.; Casciano, D. A.; Galloway, S. M.; Marnett, L. J.; Novak, R. F.; Pettit, S. D.; Preston, R. J.; Skare, J. A.; Williams, G. M.; Van Houten, B.; Gollapudi, B. B. Proceedings of a workshop on DNA adducts: Biological significance and applications to risk assessment. *Toxicol. Appl. Pharmacol.* **2005,** *208,* 1-20.

(4) Abolhassani, N.; Leon, J.; Sheng, Z.; Oka, S.; Hamasaki, H.; Iwaki, T.; Nakabeppu, Y. Molecular pathophysiology of impaired glucose metabolism, mitochondrial dysfunction, and oxidative DNA damage in alzheimer's disease brain. *Mech. Ageing Dev.* **2017,** *161,* 95-104.

(5) Gasser, S.; Raulet, D. The DNA damage response, immunity and cancer. *Semin. Cancer Biol.* **2006,** *16,* 344-347.

(6) Drabløs, F.; Feyzi, E.; Aas, P. A.; Vaagbø, C. B.; Kavli, B.; Bratlie, M. S.; Peña-Diaz, J.; Otterlei, M.; Slupphaug, G.; Krokan, H. E. Alkylation damage in DNA and RNA—repair mechanisms and medical significance. *DNA Repair* **2004,** *3,* 1389-1407.

(7) Yu, Y.; Wang, P.; Cui, Y.; Wang, Y. Chemical analysis of DNA damage. *Anal. Chem.* **2018,** *90,* 556-576.

(8) Riazuddin, S.; Athar, A.; Saffhill, R. Chemical adaptation of *M. Luteus* induces repair functions for O-alkylated DNA pyrimidines. *Nucleic Acids Res.* **1985,** *13,* 7153-66.

(9) Brent, T. P.; Dolan, M. E.; Fraenkel-Conrat, H.; Hall, J.; Karran, P.; Laval, L.; Margison, G. P.; Montesano, R.; Pegg, A. E.; Potter, P. M. Repair of O-alkylpyrimidines in mammalian cells: A present consensus. *Proc. Natl. Acad. Sci. U.S.A.* **1988,** *85,* 1759-1762.

(10) Bronstein, S. M.; Skopek, T. R.; Swenberg, J. A. Efficient repair of O6-ethylguanine, but not O4-ethylthymine or O2-ethylthymine, is dependent upon O6-alkylguanine-DNA alkyltransferase and nucleotide excision repair activities in human-cells. *Cancer Res.* **1992,** *52,* 2008-2011.

(11) Wani, A. A.; Wani, A. A.; Wani, G.; D'Ambrosio, S. M. A human DNA repair activity specific for O4-ethylthymine: Identification and partial characterization. *Carcinogenesis* **1991,** *12,* 153-153.

(12) Ahmmed, Z.; Laval, J. Enzymatic repair of O-alkylated thymidine residues in DNA: Involvement of a O4-methylthymine-DNA methyltransferase and a O2-methylthymine DNA glycosylase. *Biochem. Biophys. Res. Commun.* **1984,** *120,* 1-8.

(13) McCarthy, T. V.; Karran, P.; Lindahl, T. Inducible repair of O-alkylated DNA pyrimidines in *Escherichia coli*. *The EMBO Journal* **1984,** *3,* 545-550.

(14) Wu, J.; Li, L.; You, C.; Wang, Y.; Wang, P.; Wang, Y. Cytotoxic and mutagenic properties of minor-groove O(2)-alkylthymidine lesions in human cells. *J. Biol. Chem.* **2018,** *293,* 8638-8644.

(15) Wang, P.; Amato, N. J.; Zhai, Q.; Wang, Y. Cytotoxic and mutagenic properties of O4-alkylthymidine lesions in *Escherichia coli* cells. *Nucleic Acids Res.* **2015,** *43,* 10795-10803.

(16) Williams, N. L.; Wang, P.; Wu, J.; Wang, Y. *In vitro* lesion bypass studies of O4-alkylthymidines with human DNA polymerase η. *Chem. Res. Toxicol.* **2016,** *29,* 669-675.

(17) Williams, N. L.; Wang, P.; Wang, Y. Replicative bypass of O2-alkylthymidine lesions *in vitro*. *Chem. Res. Toxicol.* **2016,** *29,* 1755-1761.

(18) Zhai, Q.; Wang, P.; Cai, Q.; Wang, Y. Syntheses and characterizations of the *in vivo* replicative bypass and mutagenic properties of the minor-groove O2-alkylthymidine lesions. *Nucleic Acids Res.* **2014,** *42,* 10529-10537.

(19) Wu, J.; Li, L.; Wang, P.; You, C.; Williams, N. L.; Wang, Y. Translesion synthesis of O4-alkylthymidine lesions in human cells. *Nucleic Acids Res.* **2016,** *44,* 9256-9265.

(20) Vaisman, A.; Woodgate, R. Translesion DNA polymerases in eukaryotes: What makes them tick? *Crit. Rev. Biochem. Mol. Biol.* **2017,** *52,* 274-303.

(21) Zhao, L.; Washington, M. T. Translesion synthesis: Insights into the selection and switching of DNA polymerases. *Genes* **2017,** *8,* 24/1-24/25.

(22) Yang, W. An overview of Y-family DNA polymerases and a case study of human DNA polymerase η. *Biochemistry* **2014,** *53,* 2793-2803.

(23) Walker, A. R.; Cisneros, G. A. Computational simulations of DNA polymerases: Detailed insights on structure/function/mechanism from native proteins to cancer variants. *Chem. Res. Toxicol.* **2017,** *30,* 1922-1935.

(24) Nair, D. T.; Kottur, J.; Sharma, R. A rescue act: Translesion DNA synthesis past N2-deoxyguanosine adducts. *IUBMB Life* **2015,** *67,* 564-574.

(25) O'Flaherty, D. K.; Patra, A.; Su, Y.; Guengerich, F. P.; Egli, M.; Wilds, C. J. Lesion orientation of O4-alkylthymidine influences replication by human DNA polymerase η. *Chem. Sci.* **2016,** *7,* 4896-4904.

(26) Wilson, K. A.; Holland, C. D.; Wetmore, S. D. Uncovering a unique approach for damaged DNA replication: A computational investigation of a mutagenic tobacco-derived thymine lesion. *Nucleic Acids Res.* **2019,** *47,* 1871-1879.

(27) Ucisik, M. N.; Hammes-Schiffer, S. Comparative molecular dynamics studies of human DNA polymerase η. *J. Chem. Inf. Model.* **2015,** *55,* 2672-81.

(28) Wilson, K. A.; Garden, J. L.; Wetmore, N. T.; Wetmore, S. D. Computational insights into the mutagenicity of two tobacco-derived carcinogenic DNA lesions. *Nucleic Acids Res.* **2018,** *46,* 11858-11868.

(29) Ucisik, M. N.; Hammes-Schiffer, S. Relative binding free energies of adenine and guanine to damaged and undamaged DNA in human DNA polymerase η: Clues for fidelity and overall efficiency. *J. Am. Chem. Soc.* **2015,** *137,* 13240-13243.

(30) Wilson, K. A.; Wetmore, S. D. Molecular insights into the translesion synthesis of benzyl-guanine from molecular dynamics simulations: Structural evidence of mutagenic and nonmutagenic replication. *Biochemistry* **2017,** *56,* 1841-1853.

(31) Wilson, K. A.; Wetmore, S. D. Conformational flexibility of the benzyl-guanine adduct in a bypass polymerase active site permits replication: Insights from molecular dynamics simulations. *Chem. Res. Toxicol.* **2017,** *30,* 2013-2022.

(32) Ucisik, M. N.; Hammes-Schiffer, S. Effects of active site mutations on specificity of nucleobase binding in human DNA polymerase η. *J. Phys. Chem. B* **2017,** *121,* 3667-3675.

(33) Qiu, Z. M.; Wang, H. L.; Liu, Y. Z.; Hou, D. N. MP2 study on the hydrogen-bonding interaction between O4-methylthymine and DNA bases: A, C, G, and T. *J. Struc. Chem.* **2014,** *25,* 767-774.

(34) Pohorille, A.; Loew, G. H. Base-pairing properties of O-methylated bases of nucleic acids. Energetic and steric considerations. *Biophys. Chem.* **1985,** *22,* 37-51.

(35) Flood, A.; Hubbard, C.; Forde, G.; Hill, G.; Gorb, L.; Leszczynski, J. Theoretical *ab initio* study of the effects of methylation on the nature of hydrogen bonding in A:T base pair. *J. Biomol. Struct. Dyn.* **2003,** *21,* 297-302.

(36) Venkateswarlu, D.; Lyngdoh, R. H. D. Structural, steric and energetic requirements for induction of base substitutional mutations by methylated guanines and thymines. *J. Chem. Soc. Perkin Trans. 2* **1995,** 839-846.

(37) Silverstein, T. D.; Johnson, R. E.; Jain, R.; Prakash, L.; Prakash, S.; Aggarwal, A. K. Structural basis for the suppression of skin cancers by DNA polymerase η. *Nature* **2010,** *465,* 1039-1043.

(38) Millen, A. L.; Wetmore, S. D. Glycosidic bond cleavage in deoxynucleotides - a density functional study. *Can. J. Chem.* **2009,** *87,* 850-863.

(39) Merrick, J. P.; Moran, D.; Radom, L. An evaluation of harmonic vibrational frequency scale factors. *J. Phys. Chem. A* **2007,** *111,* 11683-11700.

(40) Frisch, *Æ.* H. H. P.; Dennington, R. D. II; Keith, T.; A.; Millam, J. N. A. B.; Holder, A. J.; Hiscocks, J. *GaussView 5*; Wallingford, CT, **2009**.

(41) Macrae, C. F.; Bruno, I. J.; Chisholm, J. A.; Edgington, P. R.; McCabe, P.; Pidcock, E.; Rodriguez-Monge, L.; Taylor, R.; van de Streek, J.; Wood, P. A. Mercury CSD 2.0 - new features for the visualization and investigation of crystal structures. *J. Appl. Crystallogr.* **2008,** *41,* 466-470.

(42) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Jr. J. A. M.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Revisions D.01 and E.01; Gaussian, Inc.: Wallingford CT, **2009**.

(43) Mendieta, J.; Cases-Gonzalez, C. E.; Matamoros, T.; Ramirez, G.; Menendez-Arias, L. A Mg2+-induced conformational switch rendering a competent DNA polymerase catalytic complex. *Proteins: Struct. Funct. Bioinf.* **2008,** *71,* 565-574.

(44) Ummat, A.; Silverstein, T. D.; Jain, R.; Buku, A.; Johnson, R. E.; Prakash, L.; Prakash, S.; Aggarwal, A. K. Human DNA polymerase η is pre-aligned for dNTP binding and catalysis. *J. Mol. Biol.* **2012,** *415,* 627-634.

(45) Case, D. A.; Darden, T. A.; Cheatham, T. E. III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossvary, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A. *Amber tools*, Version 1.0; University of California: San Francisco, **2008**.

(46) Dupradeau, F.-Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P. The R.E.D. Tools: Advances in RESP and ESP

charge derivation and force field library building. *Phys. Chem. Chem. Phys.* **2010,** *12,* 7821–7839.

(47) Wang, J.; Wang, W.; Kollman, P. A. Antechamber: An accessory software package for molecular mechanical calculations. *J. Am. Chem. Soc.* **2001,** *222,* U403.

(48) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **2015,** *11,* 3696-3713.

(49) Perlow, R. A.; Broyde, S. Toward understanding the mutagenicity of an environmental carcinogen: Structural insights into nucleotide incorporation preferences. *J. Mol. Biol.* **2002,** *322,* 291-309.

(50) Allner, O.; Nilsson, L.; Villa, A. Magnesium ion-water coordination and exchange in biomolecular simulations. *J. Chem. Theory Comput.* **2012,** *8,* 1493-1502.

(51) Price, D. J.; Brooks, C. L. III, A modified TIP3P water potential for simulation with ewald summation. *J. Chem. Phys.* **2004,** *121,* 10096-10103.

(52) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **2013,** *9,* 3084-3095.

(53) Su, Y.; Patra, A.; Harp, J. M.; Egli, M.; Guengerich, F. P. Roles of residues arg-61 and gln-38 of human DNA polymerase η in bypass of deoxyguanosine and 7,8-dihydro-8-oxo-2'-deoxyguanosine. *J. Biol. Chem.* **2015,** *290,* 15921-15933.

(54) Biertuempfel, C.; Zhao, Y.; Kondo, Y.; Ramon-Maiques, S.; Gregory, M.; Lee, J. Y.; Masutani, C.; Lehmann, A. R.; Hanaoka, F.; Yang, W. Structure and mechanism of human DNA polymerase η. *Nature* **2010,** *465,* 1044-1048.

(55) Suarez, S. C.; Beardslee, R. A.; Toffton, S. M.; McCulloch, S. D. Biochemical analysis of active site mutations of human polymerase η. *Mutat. Res. Fundam. Mol. Mech. Mutagen.* **2013,** *745-746,* 46-54.

(56) Nakamura, T.; Yamagata, Y.; Yang, W. Watching DNA polymerase η make a phosphodiester bond. *Nature* **2013,** *53,* 254-257.

(57) Dabkowska, I.; Gonzalez, H. V.; Jurecka, P.; Hobza, P. Stabilization energies of the hydrogen-bonded and stacked structures of nucleic acid base pairs in the crystal geometries of CG, AT, and AC DNA steps and in the NMR geometry of the 5'-d(GCGAAGC)-3' hairpin: Complete basis set calculations at the MP2 and CCSD(T) levels. *J. Phys. Chem. A* **2005,** *109,* 1131-1136.

(58) Batra, V. K.; Beard, W. A.; Shock, D. D.; Krahn, J. M.; Pedersen, L. C.; Wilson, S. H. Magnesium-induced assembly of a complete DNA polymerase catalytic complex. *Structure* **2006,** *14*, 757-766.

(59) Patra, A.; Zhang, Q.; Lei, L.; Su, Y.; Egli, M.; Guengerich, F. P. Structural and kinetic analysis of nucleoside triphosphate incorporation opposite an abasic site by human translesion DNA polymerase η. *J. Biol. Chem.* **2015,** *290*, 8028-8038.

(60) Choi, J.-Y.; Lim, S.; Kim, E.-J.; Jo, A.; Guengerich, F. P. Translesion synthesis across abasic lesions by human B-family and Y-family DNA polymerases α, δ, η, ι, κ, and rev1. *J. Mol. Biol.* **2010,** *404*, 34-44.

(61) Gowda, A. S. P.; Spratt, T. E. DNA polymerases η and ζ combine to bypass O2-[4-(3-pyridyl)-4-oxobutyl]thymine, a DNA adduct formed from tobacco carcinogens. *Chem. Res. Toxicol.* **2016,** *29*, 303-316.

(62) Du, H.; Leng, J.; Wang, P.; Li, L.; Wang, Y. Impact of tobacco-specific nitrosamine-derived DNA adducts on the efficiency and fidelity of DNA replication in human cells. *J. Biol. Chem.* **2018,** *293*, 11100-11108.

Chapter 3: Computational Study on the Replicative Outcomes of the O4-4-(3-pyridyl)-4-oxobut-1-yl-thymine Lesion[a,b,c]

3.1 Introduction

DNA damage occurs from a variety of endogenous and exogenous sources. A common cause of DNA damage is tobacco smoke since approximately one billion people world-wide smoke tobacco products.[1-2] Indeed, tobacco contains ~ 7000 chemicals, 70 of which are known carcinogens.[3] One important carcinogen, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK), is activated in cells by cytochrome P450 enzymes and has been shown to cause cancer in humans, monkeys, pigs, rabbits and certain rodents.[4] Although NNK is known to primarily cause tumors in the lungs of these species, exposure of rats to NNK indicates that tumors can also grow in the pancreas, liver, and nasal mucosa.[4]

Metabolism of NNK yields a number of reactive intermediates that lead to DNA methylation and ethylation, as well as the formation of bulkier 4-(3-pyridyl)-4-oxobut-1-yl (POB) adducts.[4] POB adducts most commonly occur at the O6 and N7 positions of G, and the O2 positions of C and T (Figure 3.1a).[5] However, POB adducts have also been detected at other positions, including O4 of T, N2 of G, N3 and N4 of C, and N6 of A, as well as the phosphate backbone.[4] Although numerous POB lesions have been shown to

---

[a] *The Journal of American Chemical Society* reference style is used in this chapter.
[b] Adapted from Bhutani, P. Murray, M.T. Wilson, K.A. Wetmore, S.D. (2019). Computational Study on the Replicative Outcomes of the O4-4-(3-pyridyl)-4-oxobut-1-yl-thymine Lesion (in preparation).
[c] M.T.M. performed preliminary calculations and analysis on the nucleobase, base pair, polymerase and DNA models. P.B. generated O4-POB-T parameters and performed final calculations and data analysis on all models, as well as wrote the first draft of the manuscript. K.A.W. contributed to the calculation design. S.D.W. contributed to project and calculation design, data interpretation and manuscript writing.

form experimentally,[3] many questions remain about their impact on DNA structure and their interactions with DNA polymerases.

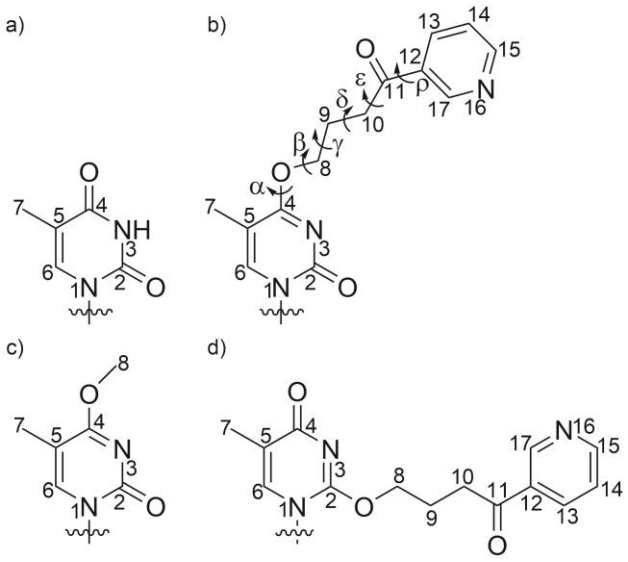


Figure 3.1: Chemical structures and numbering of a) canonical T; b) O4-POB-T (with definitions of key dihedrals angles about the bulky POB moiety including α=∠(N3C4O4C8), β=∠(C4O4C8C9), γ=∠(O4C8C9C10), δ=∠(C8C9C10C11), ε=∠(C9C10C11C12), and ρ=∠(C10C11C12C13)); c) O4-Me-T; d) O2-POB-T.

Recent experimental work examined the repair and replication of O4-[4-(3-pyridyl)-4-oxobut-1-yl]-thymine (O4-POB-T, Figure 3.1b).[6-7] Since O2-[4-(3-pyridyl)-4-oxobut-1-yl]-thymine (O2-POB-T) can be repaired by the nucleotide excision repair (NER) process,[6] the NER of O4-POB-T was studied in mammalian cells.[6] Results indicate that O4-POB-T is resistant to the NER pathway, which means this lesion will likely persist for replication. Similar to other T alkylation adducts,[8-9] O4-POB-T stalls standard DNA polymerases[7] and instead undergoes replication via the translesion synthesis (TLS) process.[10-11] The Y-family TLS polymerase responsible for the bypass and mutagenicity of O4-POB-T in human cells is polymerase η (pol η), although knockout of both pol η and B-family pol ζ most significantly affects replication bypass.[7] This aligns with previous work

showing that pol η possesses an open pocket on the major groove side of the growing DNA strand and can readily accommodate major groove DNA lesions.[12]

When alkylation damage occurs at the O4 position of T, N3 is transformed from a hydrogen-bonding donor to a hydrogen-bonding acceptor, which alters the base-pairing properties of canonical T. Therefore, it is not surprising that O4-POB-T leads to base mispairs upon replication. Specifically, experimental work in human cells show that dGTP is incorrectly inserted opposite O4-POB-T 35% of the time, with no other mispairs being detected.[7] This mutagenic pattern aligns with other O4-T alkylation adducts, such as O4-methylthymine (O4-Me-T, Chapter 2).[9, 13-14] However, this 35% misinsertion of dGTP is significantly less than observed for smaller O4-T alkylation adducts derived from tobacco smoke. For example, as mentioned in Chapter 2, replication of O4-Me-T (Figure 3.1c) by pol η results in dGTP insertion > 60% of the time.[9] This significant difference is interesting because although the O4-POB-T alkyl chain is larger and more flexible than the methyl group of O4-Me-T, both lesions have similarly altered Watson-Crick (WC) faces, and both damaged groups are anticipated to point into the open pocket of pol η to facilitate replication. Additionally, a more commonly formed POB-T lesion (O2-POB-T, Figure 3.1d) leads to a different mutagenic pattern upon replication, specifically dATP (85%) and dTTP insertion (15%).[7, 15] Interestingly, despite being a minor groove lesion, O2-POB-T is most efficiently bypassed by pol η and pol ζ[15] as found for O4-POB-T.[7] Although experimental work indicates that O2-POB-T and O4-POB-T result in different base mispairs, previous computational work has only provided insight into the mutagenicity of O2-POB-T,[16] leaving questions regarding how the position of the POB moiety affects pol η replication outcomes.

Systematic modelling has been previously conducted to rationalize experimental replication outcomes associated with a variety of DNA adducts.[16-21] Specifically, the effects of the shape and size of bulky G adducts have been studied using density functional theory (DFT) calculations and molecular dynamics (MD) simulations, which provided structural insight into the observed differential mutagenic patterns of G lesions.[17] Additionally, using similar computational techniques, the replication of the O6-POB-G and O2-POB-T adducts has been studied to understand how these lesions fit within the active site of TLS polymerases.[16, 18] These latter works have developed an accurate protocol for systematically studying flexible DNA adducts.[16] Furthermore, MD simulations have been used to understand the mutagenicity associated with a common form of DNA damage caused by UV radiation, namely TT-dimers,[19-20] as well as the role of pol η in the bypass of these lesions.[20] Taken together, these studies illustrate that computational modelling is a useful tool for rationalizing the mutagenic outcomes of large, flexible DNA adducts.

In this chapter, the computational protocol successfully employed to understand the replication of other flexible DNA adducts is used to provide structural insight into the bypass of O4-POB-T. Initially, a conformational search is conducted to explore the inherent flexibility of O4-POB-T. Next, MD simulations on the pre-insertion complex are used to understand how the lesion can be accommodated by pol η. To give insight into how O4-POB-T base pairs, DFT models yield information on the intrinsic hydrogen-bonding patterns between the isolated lesion and select canonical nucleobases. Subsequently, MD simulations on pol η ternary complexes for the insertion of select dNTPs opposite O4-POB-T, as well as the corresponding post-extension DNA duplexes, are used to understand the stability and inherent dynamics of the lesion base pairs during and after replication.

Together, these results rationalize the previously reported mutagenic outcomes of O4-POB-T.[7] Comparison of the data for O4-POB-T presented in this chapter to that for O4-Me-T presented in the previous chapter provides insight into the significant decrease in the amount of dGTP insertion (i.e. from 60% to 35%) with an increase in the size of the alkyl group. Furthermore, a comparison of O4-POB-T and O2-POB-T provides insight into the position of the flexible bulky moiety, and how it impacts mutagenic outcomes,[16] which is the second example of the effect of T alkylation position on the replication outcomes. This chapter also further solidifies the computational protocol required to accurately model flexible DNA adducts, which can be used in a predictive way to study a wide variety of lesions in future work to further understand the role of the TLS process in the replication of damaged DNA.

3.2 Methods

The methodology implemented in this chapter is based on previous computational work on O2-POB-T,[16] which will permit an accurate comparison of the two lesions to reveal the effect of bulky moiety position on replication outcomes.

3.2.1 DFT Calculations: To explore the conformational flexibility of the O4-POB-T lesion, a HyperChem (version 8.0)[22] conformational search was initially performed with respect to the six dihedrals angles within the bulky moiety (α′, β′, γ′, δ′, ε′ and ρ′, Figure 3.1b) using the AMBER96 force field (Table B.1). This isolated model was capped with a hydrogen atom in the place of deoxyribose as the bulky POB moiety is not positioned to interact with the DNA backbone and this substitution is not anticipated to affect the accessible bulky moiety conformations. Partial charges for O4-POB-T were generated

using the RED.v.III program on the optimized (B3LYP/6-31(d,p)) nucleoside structures (*anti* and *syn*) (Table B.2).[23] The resulting structures were subsequently optimized using M06-2X/6-31G(d) and single-point energies were calculated using B3LYP-D3(BJ)/6-311++G(2df,2p). Each unique structure was visually inspected and categorized into either an extended, T-shaped, or stacked conformation depending on how the bulky POB moiety interacts with the T ring. All conformations were subsequently manually docked into the active site of pol η to identify feasible bound conformations.

The lowest energy orientation of O4-POB-T from the conformational search that fits within the pol η active site was chosen to pair with select canonical nucleobases. In these models, all nucleobases were capped with a methyl group in the place of deoxyribose to save computational time as this will unlikely affect the computed hydrogen-bonding patterns or interaction energies.[24] Base pairs were optimized using M06-2X/6-31G(d), while interaction energies were calculated using B3LYP-D3(BJ)/6-311++G(2df,2p). Reported interaction energies include counterpoise corrections and scaled (0.9813) zero-point vibrational energy (ZPVE), but do not account for the deformation energy. The interstrand distance (R(C1′···C1′)) and base pair opening angle (∠(N9C1′C1′)) were measured using GaussView 6,[25] while the interplanar angle (i.e. the angle between the nucleobase ring planes) was measured using Mercury.[26] These parameters were also evaluated at the same level of theory for the natural T:A pair. DFT calculations were performed using Gaussian 09 (revision E.01).[27]

3.2.2 MD Simulations: Polymerase η pre-insertion and ternary complexes were generated from an X-ray crystal structure corresponding to dATP insertion opposite canonical T

(PDB ID: 4ECS), which was chosen because this well resolved structure corresponds to a reactant complex.[28] Ca(II) ions in the crystal structure that were used to inhibit the reaction were replaced with Mg(II) ions. This structure was modified using GaussView 6[25] to insert missing residues (T155–E159) and replace canonical T with O4-POB-T. For the pre-insertion complex (i.e. the DNA–polymerase complex without an incoming dNTP), Mg(II) ions were removed due to evidence these do not come into the active site until dNTP binding.[29-30] However, two Mg(II) ions were included in ternary complexes since these have been proposed to be key for dNTP binding and facilitating the nucleotidyl transfer reaction.[30-31] The Mg(II) parameters developed by Allner *et al*. were used to describe the metal ions,[32] while dNTP parameters were taken from previous work.[33] The DNA sequence was modified to ensure the sequence context around O4-POB-T corresponds to the experimental work investigating O4-POB-T replication by pol η.[7, 34] Protein residues with multiple possible orientations were visually inspected to ensure the simulated conformation permitted favourable interactions with neighbouring residues. O4-POB-T parameters were assigned using Antechamber 1.4[35] as implemented in Amber14,[36] with missing parameters adopted from the General Amber Force Field (GAFF).[37] The AMBER14SB force field was used to describe the unmodified portions of the enzyme and DNA strand.[38] For all DNA–polymerase models, tleap was used to add Na^+ to neutralize the system, while Na^+ and Cl^- ions were added to bring the systems to a physiological salt concentration of ~ 150 mM. The DNA duplex models with the 5′-GCATGGCGXGCTATGC-3′ sequence were built using the Nucleic Acid Builder (NAB) and the damaged lesion was added using GaussView 6.[25] The sequence was chosen to match previous experimental work investigating the replication of O4-POB-T by pol η, with two additional G:C pairs added

to both ends to stabilize the duplex. Na^+ counterions were added to neutralize the system and a 25 kcal mol^{-1} $Å^{-1}$ force restraint constant was applied to the hydrogen bonds within the terminal base pairs to prevent helical unwinding. The AMBER OL15 force field was used to describe unmodified DNA in the duplex model.[39]

For all simulations, a non-bonded cut-off of 10 Å and the periodic boundary condition were implemented. The particle-mesh Ewald algorithm was used to describe electrostatic interactions throughout the equilibration and production simulations. All systems were solvated using a TIP3P[40] octahedral water box in which the solute was 10.0 Å from the edge of the box in each direction. All complexes were minimized using 2500 steps of steepest descent and 2500 steps of conjugant gradient stepwise minimization. First, the solvent was minimized, while the solute position was held in place using a force restraint constant of 50 kcal mol^{-1} $Å^{-1}$. Next, the hydrogen atoms were minimized using the same force restraint constant on the solvent and remaining solute. The solute was then fully minimized while restraining the solvent. Finally, the entire system was minimized. The system was then heated using a Langevin thermostat ($\gamma = 1$) from 10 K to 310 K in 50 K increments. The system was subsequently equilibrated using a 2 fs time step over 5 phases that sequentially lowered the force restraint constant from 20 kcal mol^{-1} $Å^{-1}$ to 1.5 kcal mol^{-1} $Å^{-1}$. MD production simulations were then performed on each the system for 1 μs.

Average structural features similar to those discussed for the base-pair models were analyzed over the MD production simulations, including the average interaction energy and interplanar angle every 5 ns (structures were extracted using in-house scripts), as well as the interstrand distance (R(C1′⋯C1′)) and base pair opening angle (∠(N9C1′C1′)) every 0.05 ns using the CPPTRAJ[41] module in AMBER. Interaction energies include

counterpoise corrections, but do not include the deformation energy, and were computed at the same level of theory as reported for the base-pair model. Additionally, the hydrogen-bonding occupancies, average reaction parameters (R(O3′···Pα) and ∠(O3′–Pα–Pβ)) and the Mg(II) coordination were analyzed over the MD production simulation every 0.05 ns using CPPTRAJ.[41]

3.3 Results

3.3.1 The sequestered O4-POB-T lesion is inherently flexible: To explore the inherent flexibility of the isolated O4-POB-T adduct, the energetically accessible lesion conformations with respect to six key dihedrals angles within the POB moiety (Figure 3.1b) were characterized. In total, 197 lesion conformations were isolated from the AMBER conformational search, which collapsed to 169 unique structures upon optimization with DFT that span over ~ 85 kJ/mol. The most stable conformation adopted a stacked arrangement between the T ring and the bulky moiety. However, the stacked conformations ranged in energy by ~ 62 kJ/mol and only account for 33 (20%) of the 169 structures. The most common conformations were classified as extended since there are no direct interactions between the T ring and POB moiety. In fact, 130 (77%) of the total conformers were classified as extended, with the most stable conformation being ~ 12 kJ/mol higher in energy and the least stable conformation being ~ 85 kJ/mol higher in energy than the overall most stable stacked conformation. Finally, 6 (4%) structures involved a T-shaped interaction between the POB moiety and the T ring, where the most stable conformation is ~ 15 kJ/mol higher in energy and the least stable conformation is ~ 27 kJ/mol higher in energy than the global minimum (Figure 3.2 and B.1a). Overall, although the POB moiety

is very flexible, it is unclear how the bulky group will be accommodated in the pol η active site. Specifically, even though stacked conformations are the most stable orientations, such nucleobase–POB interactions may interfere with base stacking or key contacts with active site. Furthermore, although most conformations are extended, questions remain about how such structures may interfere with lesion base pairing. Therefore, the pol η pre-insertion complexes, and lesion base pairing in isolation and ternary complexes will be considered in the following sections.

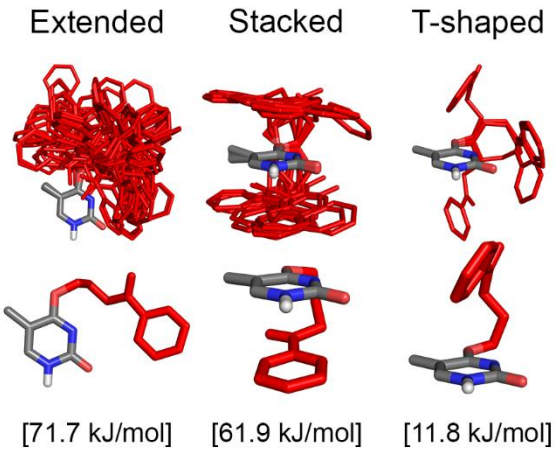


Figure 3.2: Overlay of O4-POB-T nucleobase conformations with respect to the T ring atoms (top) and the most stable conformation (bottom), as well as the energetic difference between the lowest and highest energy conformation [kJ/mol] within each structural category.

3.3.2 O4-POB-T can be suitably positioned in the pol η active site prior to dNTP insertion despite the flexibility of the POB moiety: Prior to investigating the DNA–polymerase pre-insertion complex, all 169 unique structures of isolated O4-POB-T were manually docked within the pol η active site. Specifically, the conformations were overlaid onto the template T in an X-ray structure corresponding to dATP insertion during the replication of canonical DNA according to T ring atoms (PDB ID: 4ECS).[28] Due to the size of the bulky moiety, many conformations clash with the flanking DNA bases in the

template and primer strands, as well as active site residues. Among the original 169 conformations, only 25 maintain a heavy atom distance between the lesion and the nearest active site residue or flanking base > 2.0 Å, or do not take the space of the incoming dNTP (Figure 3.3a). All 25 conformations that satisfy these criteria were from the extended structural category. To further assess lesion accommodation by pol η, an MD simulation was initiated from an O4-POB-T conformation that pointed the bulky moiety into the open pocket. Throughout the simulation, O4-POB-T remains in the active site (Figure B.2) and the POB moiety remains very flexible (Figure 3.3b). An overlay of MD snapshots taken every 5 ns shows that most lesion orientations from the conformational search that fit within the active site (Figure 3.3b) were sampled over the simulation. In the pre-insertion complex, O4-POB-T primarily adopts extended conformations (Figure B.2a), which avoid steric repulsion with surrounding DNA and enzyme residues. Nevertheless, some lesion conformations involve stacking between the POB moiety and flanking bases (< 10%, Figure B.2b), which could disrupt intrastrand stacking upon dNTP insertion, while other conformations place the bulky moiety in the location the incoming dNTP must bind, which could impede dNTP insertion (Figure 3.3b). Such conformations correlate with experimental evidence that O4-POB-T moderately blocks replication by pol η (< 30% bypass).[7] In the X-ray crystal structure corresponding to dATP insertion opposite canonical T by pol η,[28] only Q38 directly interacts with the template nucleobase. Indeed, Q38 has been proposed to stabilize the template base in the active site through hydrogen bonding with the O2 position of pyrimidines.[42-43] Throughout the simulation, Q38 forms direct interactions with O2 (26%) and O4′ (19%) of the lesion site (Figure B.3a). Overall, the MD simulation on the pre-insertion complex shows that the bulky POB moiety is very flexible

within the active site of pol η, which correlates with the conformational search on the isolated lesion. Nevertheless, O4-POB-T, as well as the flanking bases, are generally well positioned in the active site to permit the accommodation of a dNTP opposite the lesion (Figure B.2a). However, it is unclear how the incoming dNTP will be accommodated opposite the lesion because of the change in the WC hydrogen-bonding face compared to canonical T, as well as the possibility that the POB moiety favorably interacts or sterically clashes with the pairing base. Therefore, potential hydrogen-bonding patterns between O4-POB-T and select canonical bases are closely examined in the following section.

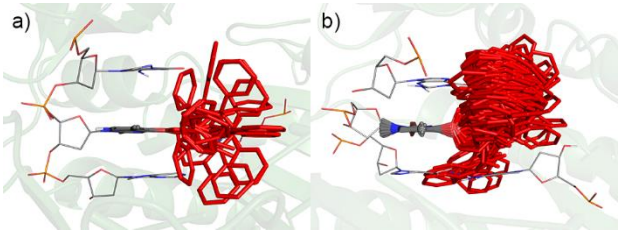


Figure 3.3: a) DFT conformational search orientations that fit within the pol η active site (PDB ID: 4ECS) and b) MD snapshots taken every 5 ns overlaid based on T-ring atoms in the pol η pre-insertion complex (nucleobase in grey, POB in red).

3.3.3 O4-POB-T intrinsically prefers to pair with G over A: Prior to considering the base-pairing potential of O4-POB-T within the pol η ternary complex, DFT calculations were used to understand how the isolated O4-POB-T lesion hydrogen bonds with A and G since these pairs correspond to experimental replication outcomes.[7] Key base pair structural parameters, as well as the interaction energies, will be compared to canonical T:A computed at the same level of theory (Figure B.4). The canonical T:A pair contains two strong hydrogen bonds (O4(T)···H–N6(A) and N3(T)–H···N1(A)). The resulting interstrand distance and base pair opening angle are 10.481 Å and 56.4°, respectively, while the base pair is almost planar (interplanar angle = 0.1°). Overall, these structural features result in an interaction energy of –70.8 kJ/mol.

The lowest energy conformation of O4-POB-T from the conformational search that fit within the pol η active site was selected to pair opposite A and G. As previously mentioned, unlike canonical T, the N3 position of O4-POB-T is a hydrogen-bonding acceptor. As a result, the O4-POB-T:A pair contains only one strong N3(O4-POB-T)···H–N6(A) hydrogen bond (Figure 3.4a). However, the methyl groups that replace deoxyribose moieties in this model are directed such that the sugars would be directed parallel to each other instead of the anti-parallel orientation necessary for B-DNA. This arrangement affords an additional weak hydrogen bond between N7 of A and the ring of the POB moiety. Nevertheless, the O4-POB-T:A pair has a weaker interaction energy (–42.6 kJ/mol) compared to T:A (–70.8 kJ/mol), which supports the mutagenic potential of this lesion.

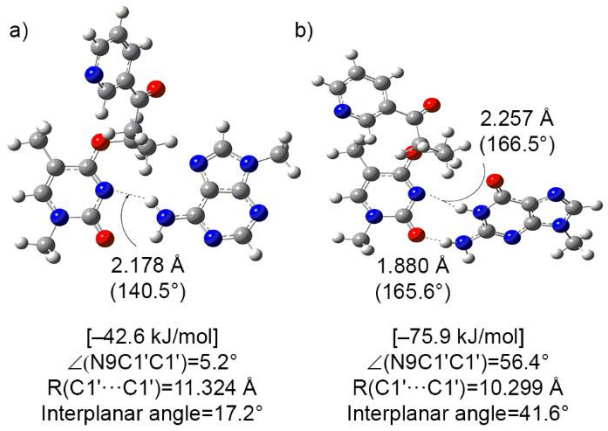


Figure 3.4: Key DFT structural features (Å and deg.) and interaction energies [kJ/mol] for the O4-POB-T lesion hydrogen bonded with a) A or b) G.

The O4-POB-T:G mispair contains two strong hydrogen bonds, namely N3(O4-POB-T)···H–N1(G) and O2(O4-POB-T)···H–N2(G). An additional weak interaction forms between the alkyl chain of the POB moiety and O6 of the opposing G, which is accompanied by distortion of the native interplanar angle to 41.6°. The interstrand distance

and base pair opening angle in the O4-POB-T:G mispair (R(C1′···C1′)=10.299 Å and ∠(N9C1′C1′)=56.4°, Figure 3.4b) are comparable to those of canonical T:A (Figure B.4). In fact, the O4-POB-T:G mispair has a slightly stronger interaction energy (–75.9 kJ/mol) than canonical T:A (–70.8 kJ/mol). Nevertheless, it is unclear how the distorted pair will be accommodated in the pol η active site.

Overall, the intrinsic base-pairing preference of the O4-POB-T lesion suggests that O4-POB-T:G is more favourable than the O4-POB-T:A pair both in terms of the structural parameters and stability. Although this at least in part correlates with the experimentally-observed 35% dGTP incorporation by pol η,[7] the unfavourable base pairing with A makes it unclear why this lesion is non-mutagenic 65% of the time. Therefore, to further understand the replication of the O4-POB-T lesion, the pol η ternary complexes corresponding to dGTP and dATP insertion opposite the damaged site are considered in the following sections.

3.3.4 The bulky POB moiety disrupts hydrogen bonding between O4-POB-T and the incoming dGTP in the pol η active site: Although the isolated base pairs suggest that only the O4-POB-T:G mispair has intrinsic structural features comparable to canonical T:A, experimental results indicate dGTP insertion opposite O4-POB-T occurs only 35% of the time.[7] To explain the discrepancy between DFT models and the experimental mutagenic patterns, the O4-POB-T:G mispair was considered within the active site of pol η. There are many structural features that are required for the pol η ternary complex to be conducive for dNTP insertion. First, there are several residues that stabilize the position of the incoming dNTP, including C16, F18, Y52, R55, R61 and K231. For dGTP insertion, there are no

significant deviations in the orientations of these residues compared to the pol η crystal structure corresponding to dATP insertion opposite T (PDB ID: 4ECS,[28] rmsd=0.5 ± 0.1, Figure B.5a). Second, the coordination of the two active site Mg(II) ions must be maintained (Figure B.6a), which occurs in this ternary complex for a minimum of 64% (< 2.5 Å) over the simulation (Table B.3). Finally, the incoming dNTP must be aligned for the nucleotidyl transfer reaction, which can be quantified as the reaction distance (R(O3′···Pα)) being < 3.5 Å and the reaction angle (∠(O3′–Pα–Pβ)) being ~ 180° (Figure B.6b). Although the reaction distance (R(O3′···Pα) = 3.2 ± 0.2 Å) is conducive for the nucleotidyl reaction, the average reaction angle (117.3 ± 58.9°) significantly deviates from the ideal value. Nevertheless, as reflected in the large standard deviation, the reaction angle periodically adopts two primary values throughout the simulation. Specifically, the reaction angle is near the ideal value of ~ 180° for 59% of the simulation, and is ~ 50° for 41% of the simulation (Figure B.7). This suggests that the reaction angle is frequently not conducive for dGTP insertion opposite O4-POB-T by pol η.

In addition to the position of the incoming dGTP, the flexibility and location of the O4-POB-T lesion in the ternary complex must be considered. For dGTP insertion, the bulky POB moiety adopts many conformations over the simulation. Most of the lesion conformations extend into the open pocket of pol η, while others are in close proximity to 5′-G (< 13%, Figure 3.5a and B.8). Nevertheless, as discussed for the pol η pre-insertion complex, Q38 interacts with O2 (21%) and O3′ (20%) of the lesion throughout the simulation, which helps align O4-POB-T opposite the incoming dGTP (Figure B.3b). As a result, three hydrogen bonds between the lesion and dGTP periodically occur throughout the simulation, namely O2(O4-POB-T)···H–N2(dGTP) (68%), N3(O4-POB-T)···H–

N2(dGTP) (19%) and N3(O4-POB-T)···H–N1(dGTP) (19%, Table B.4). The interstrand distance is comparable to that of canonical DNA (R(C1′···C1′)=10.6 Å). However, the base pair opening angle is extremely flexible (∠(N9C1′C1′)=80.6 ± 30.3°) and the nucleobases are non-planar (by 31.6°), suggesting some distortion within the base pair in the pol η active site (Figure 3.5b). Interestingly, the flexibility in the base pair opening angle directly correlates with the previously discussed fluctuations in the reaction angle. Specifically, when the base pair opening angle is > 70°, the reaction angle is also less ideal (< 100°, Figure B.7). Regardless of these fluctuations in the structural parameters, O4-POB-T:dGTP has a strong average base pair interaction energy (–80.2 kJ/mol).

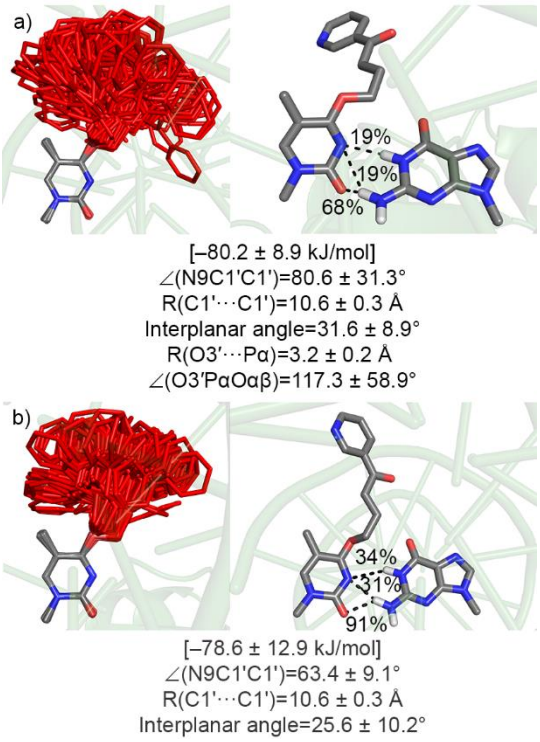


Figure 3.5: MD snapshots of the O4-POB-T lesion taken every 5 ns overlaid based on T-ring atoms (left), and base-pair interaction (right), as well as key structural features (Å and °), binding strengths [kJ/mol], hydrogen-bond occupancies (%) and reaction parameters (Å and °) where applicable for the O4-POB-T:G mispair in the a) pol η ternary complex or b) damaged DNA duplex.

When G is paired opposite O4-POB-T in the post-extension DNA duplex, the POB moiety exhibits high flexibility, but primarily points toward the major groove of the DNA helix (Figure 3.5c). This is due to the fact that the lesion can extend into the solvent surrounding DNA and does not clash with flanking DNA bases. The average interstrand distance is comparable to that observed in the isolated model (R(C1′···C1′)=10.6 Å), and the base pair opening angle and interplanar angle are 63.4° and 25.6°, respectively. These structural parameters help maintain one O2(O4-POB-T)···H–N2(G) hydrogen bond for the majority of the simulation (92% occupancy), and two periodic hydrogen bonds between N3 of O4-POB-T and N2 (31%) or N1 (34%) of G (Figure 3.5d and Table B.4). This leads to an average interaction energy of –78.6 kJ/mol. Furthermore, there are no significant distortions to the DNA duplex containing the O4-POB-T:G base pair (Figure B.9a).

Overall, these results indicate that the flexible POB moiety disrupts the O4-POB-T:G base mispair in the pol η active site. Indeed, the resulting active site position of dGTP may impede the nucleotidyl transfer reaction. Despite this, favourable structural and reaction parameters are periodically adopted throughout the simulation, which likely plays a contributing role to the mutagenicity of this lesion (35%).[7] Furthermore, there are no distortions to the overall helical structure in the post-extension DNA duplex containing O4-POB-T:G. Regardless, due to the instability in the active site upon dGTP insertion, dATP insertion may still be feasible. In fact, experimental data suggests dATP insertion (65%) occurs far more often than dGTP insertion (35%),[7] which contrasts DFT models that reveal an unfavourable O4-POB-T:A base-pair interaction. Therefore, dATP insertion opposite the lesion by pol η is carefully considered in the following section.

3.3.5 dATP insertion opposite O4-POB-T is stabilized through the formation of a water chain, supporting the experimentally-observed non-mutagenic replication: In the ternary complex corresponding to dATP insertion opposite O4-POB-T, the orientations of residues that interact with the incoming dATP (i.e. C16, F18, Y52, R55, R61 and K231) do not significantly deviate from the crystal structure orientation for dATP insertion opposite undamaged T by pol η (rmsd=0.6 ± 0.2, Figure B.5b). Additionally, the coordinating residues interact with the two Mg(II) for a minimum of 86% occupancy (maintaining a distance < 2.5 Å, Table B.5). Furthermore, throughout the simulation, the reaction distance and angle (3.4 ± 0.1 Å and 165.4 ± 30.7°) are near ideal values. Overall, this indicates that dATP is well positioned in the active site to promote insertion.

When dATP is inserted opposite O4-POB-T, Q38 only interacts with deoxyribose at O3′ (31%), while direct interactions between Q38 and the O2 of T are minimal (3%, Figure B.3c). Although the bulky POB moiety is very flexible in this ternary complex, the entire lesion is positioned in the pol η open pocket, which prevents direct WC hydrogen bonding between the damaged T and the incoming dATP. However, a water chain forms between N1 of dATP and O4′ of the 3′-G (Figure 3.6a). This interaction is observed in the crystal structure corresponding to dATP insertion by pol η opposite an abasic site,[44-45] explaining what is known as "the A rule" (Figure B.10).[45] This result was also observed in computational studies of dATP insertion by pol η opposite O-Me-T lesions (Chapter 2) and O2-POB-T.[16] Therefore, MD simulations suggest that the experimentally-observed 65% insertion of dATP likely occurs through a similar mechanism as proposed for the replication of an abasic site,[44] as well as other O-alkyl-T lesions.[16]

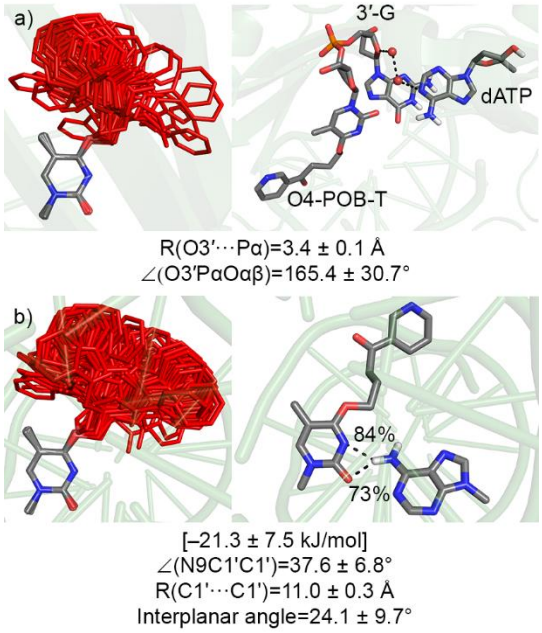


Figure 3.6: MD snapshots of the O4-POB-T lesion taken every 5 ns overlaid based on T-ring atoms (left) and key structural features (Å and °), binding strengths [kJ/mol], hydrogen-bond occupancies (%) and reaction parameters (Å and °) where applicable for the O4-POB-T:A pair (right) in the a) pol η ternary complex or b) damaged DNA duplex.

In the DNA duplex, the POB moiety exhibits high flexibility, and similar to the O4-POB-T:G pair in this environment, nearly all lesion conformations are extended (Figure 3.6b). The O4-POB-T:A base pair maintains two hydrogen bonds over the simulation, namely N3(O4-POB-T)···H–N6(A) (84%) and O2(O4-POB-T)···H–N6(A) (73%, Figure 3.6d and Table B.4). The interstrand distance and base pair opening angle (R(C1′···C1′)=11.0 Å) and ∠(N9C1′C1′)=37.6°) are slightly deviated from canonical B-DNA. To afford highly occupied hydrogen bonds and an interstrand distance similar to B-DNA, the base pair is non-planar by 24.1°. These structural parameters lead to a weak interaction energy (–21.3 kJ/mol, Figure 3.6d). Despite the reduced stability of the O4-POB-T:A pair compared to T:A, there are no major distortions to the overall DNA helical structure upon incorporation of O4-POB-T (Figure B.9b).

Overall, in correlation with DFT models that suggest an unfavourable interaction in the O4-POB-T:A pair, O4-POB-T is positioned in the pol η open pocket and does not directly interact with the incoming dATP. Nevertheless, dATP is stabilized by a water chain to the 3′-G with respect to O4-POB-T. When coupled with the integrity of the post-extension DNA duplex, this indicates that bypass may occur through a similar mechanism proposed for the replication of an abasic site[44] and other O-alkyl-T lesions,[16] and thus, this data rationalizes the observed 65% dATP insertion.[7]

3.4 Discussion

3.4.1 Mutagenic and non-mutagenic O4-POB-T bypass by pol η likely arise through different pathways: The results of this chapter indicate that the experimentally-observed pol η replication outcomes for O4-POB-T, namely dATP (65%) and dGTP (35%) insertion,[7] are likely due to different insertion pathways. DFT models suggest that base-pair interactions favourable for B-DNA only occur between O4-POB-T and G. Indeed, the O4-POB-T:G mispair contains two strong hydrogen bonds and structural parameters comparable to those of canonical T:A, and this mispair is almost twice as strong as O4-POB-T:A. However, MD simulations on the pol η ternary complex corresponding to dGTP insertion opposite O4-POB-T suggest the size and flexibility of the alkyl chain disrupt the base-pair interaction, as well as the position of dGTP in the polymerase active site. Specifically, the hydrogen-bonding occupancies, the base pair opening angle and the reaction angle only periodically adopt optimal values, meaning that bypass is sometimes feasible. Nevertheless, the overall structure of the DNA helix is maintained upon incorporation of the O4-POB-T:G mispair. Together, this data correlates with the

experimentally-observed 35% insertion frequency.[7] In contrast, DFT models indicate that the O4-POB-T:A pair is unstable, resulting in the methyl groups that replace the deoxyribose moiety to be in an arrangement that is not compatible with B-DNA. As a result of the poor hydrogen-bonding potential, the lesion is displaced into the open pocket of pol η in the ternary complex corresponding to dATP insertion opposite O4-POB-T and does not interact with the incoming dATP. However, a water chain forms between dATP and 3′-G of the template strand, which helps hold the dATP in a catalytically conducive position for insertion. This water network is comparable to interactions present in the crystal structure for the replication of an abasic site by pol η,[44] as well as the proposed replication patterns for other alkylated T lesions, including O2-POB-T[16] and O4-Me-T (Chapter 2). Furthermore, the O4-POB-T:A pair does not disrupt the structural integrity of the DNA duplex. Overall, despite O4-POB-T forming favourable mispairs with G, the size and flexibility of the bulky moiety can affect lesion interactions in the pol η active site, leading to a low frequency of dGTP insertion and favoring dATP insertion through a pathway similar to an abasic site.[44]

3.4.2 The size and flexibility of the POB moiety reduces dGTP insertion opposite O4-POB-T compared to O4-Me-T: Despite both being O4-alkyl-T lesions, O4-POB-T and O4-Me-T lead to different mutational frequencies upon replication by pol η. Specifically, although dGTP is the only mispair detected for both lesions in human cells,[7, 9] replication of O4-Me-T results in 60% dGTP insertion compared to only 35% insertion for O4-POB-T. This significant decrease in the mutational frequency is particularly surprising due to the fact that pol η has an open pocket on the major groove side of the growing DNA strand

and should be able to readily accommodate many forms of damage to the major groove side of DNA (i.e. the O4 position of T). Since both lesions similarly alter the WC binding face of T (i.e. N3 becomes a hydrogen-bond acceptor), these results indicate that both O4-Me-T and O4-POB-T do not directly hydrogen bond with an incoming dATP. Instead, it is proposed that these O4-alkyl-T lesions will promote dATP insertion through a water chain between the 3′-base with respect to the template base, with the lesion positioned in the open pocket of pol η and void of direct interactions with the incoming dATP. In contrast, both O4-Me-T and O4-POB-T can form stable hydrogen bonds with an incoming dGTP. However, dGTP is not optimally positioned in the pol η ternary complex for O4-POB-T replication due to interference from the large and flexible bulky moiety. Specifically, the O4-Me-T:dGTP mispair results in favourable reaction (∠(O3′–Pα–Pβ) = 164.3 ± 5.0°) and structural (∠(N9C1′C1′) = 58.9°) parameters throughout the MD simulation, while the O4-POB-T:dGTP mispair frequently has less optimal values (∠(O3′–Pα–Pβ) = 117.3 ± 58.9° and ∠(N9C1′C1′) = 80.6°). Additionally, O4-Me-T results in a persistent O2(O-Me-T)···H–N2(dGTP) (96% hydrogen-bond occupancy), while the corresponding interaction between dGTP and O4-POB-T is greatly reduced (68% occupancy). Thus, when compared to Chapter 2, this chapter emphasizes the impact of the size of the bulky POB moiety on the base-pair interactions within the pol η active site. These differences rationalize the experimentally-observed decreased dGTP (or increased dATP) insertion with an increase in the size of the alkyl chain.[7, 9]

3.4.3 O4-POB-T adopts an *anti* conformation upon replication by pol η, which contrasts the preferred O2-POB-T conformation and rationalizes the reported differential mutagenicity: Previously, the Wetmore group has used a similar computational methodology to that applied in this chapter to obtain structural insight into the mutagenic patterns of the minor groove O2-POB-T lesion,[16] which results in dATP (85%) and dTTP (15%) insertion upon replication by pol η.[7, 15] Interestingly, both lesions similarly alter the WC face, and only differ in the position of the bulky moiety. Therefore, the reasons for the contrasting mutagenic patterns are not immediately apparent.[7, 15] However, there are some key differences between the O2-POB-T and the O4-POB-T adducts predicted by both the isolated models and pol η complexes. Specifically, some O2-POB-T adduct conformations from the conformational search hydrogen bond with the N1 position of T (10% of the total number of structures),[16] which would interfere with deoxyribose in the full duplex context. Furthermore, interactions between the POB moiety and T ring occur more often for O2-POB-T than O4-POB-T, leading to a higher percentage of extended conformations for the O4 lesion (77% ranging over ~ 70 kJ/mol, Figure B.2) than the O2 adduct (49% ranging over ~ 50 kJ/mol).[16] Although stacked conformations occur ~ 20% of the time regardless of bulky moiety position, the T-shaped interactions are more preferred for O2-POB-T (21%) over O4-POB-T (4%). These differences in the intrinsic conformational preferences of the lesions may impact the base pairing potential. Nevertheless, when these conformations are manually docked into the pol η active site, only extended conformations do not interfere with active site residues or flanking bases regardless of the bulky moiety position.

In both the pol η pre-insertion and ternary complexes for O2-POB-T replication, the POB moiety clashes with key active site residues that would impede replication when this lesion adopts an *anti* orientation (Figure B.11). As a result, O2-POB-T must adopt a *syn* orientation (Figure B.11), which directs the bulky moiety into the pol η open pocket and avoids such disruptive contacts. This rotation about the glycosidic bond changes the hydrogen-bonding face prior to binding of the incoming dNTP, ultimately leading to a preference for dTTP insertion opposite the O2 lesion. In contrast, because the *anti* orientation of O4-POB-T points the bulky moiety into the open pocket of pol η, this lesion is well accommodated within the active site without a conformational change. In this instance, the lesion WC face is directed toward the incoming dNTP and therefore O4-POB-T mispairs with dGTP are preferred. Nevertheless, dATP insertion opposite both the O2 and O4-POB-T lesions likely proceeds through a similar mechanism, where a water chain is formed between the incoming dATP and the 3′ base of the template strand as proposed for an abasic site, as well as O-Me-T lesions (Chapter 2).[44]

Thus, this data reveals that O2-POB-T and O4-POB-T lead to very different mutagenic outcomes since different hydrogen-bonding faces are available to interact with an incoming dNTP. This conclusion is similar to that made in Chapter 2. Specifically, *anti*-O2-Me-T was proposed to promote dATP and dGTP insertion, which contrasts experimental results indicating this lesion is non-mutagenic.[7] However, rotation about the glycosidic bond such that the lesion interacts with the Hoogsteen face (*syn*-O2-Me-T) only promotes dATP insertion. Therefore, the current chapter provides the second example of the effect of alkyl group position on the replication outcomes for O-alkyl-T lesions. Overall, although both O-POB-T lesions have similarly modified WC faces, there is a

strong dependence of the mutagenicity on the location of the POB moiety due to differences in how the lesions are accommodated within the pol η active site.

3.5 Conclusion

The experimental mutagenicity pattern of O4-POB-T indicates that this lesion is primarily non-mutagenic upon replication, although mispairs with G occur with a frequency of 35%.[7] Interestingly, compared to smaller O4-alkyl-T lesions (i.e. O4-Me-T, 60%), the insertion of dGTP is significantly reduced for O4-POB-T.[9] In contrast, O2-POB-T prefers to form mispairs with dTTP (15%).[7] Together, this data suggests that there is an effect of the size and position of the O-alkyl T group on the replication outcomes. Since structural studies to date have only considered the mutagenic patterns of O4-Me-T (Chapter 2) and O2-POB-T,[16] the present chapter investigates the replication of O4-POB-T to clarify the impact of the bulky moiety size and shape on the mutagenic outcomes.

All of these O-alkyl-T lesions are proposed to insert dATP when the lesion positions itself in the pol η open pocket, and the damaged group is on the major groove of the growing DNA strand. This leaves space for a water chain to form between dATP and the 3′-base on the template strand, similar to observed in a crystal structure corresponding to replication of an abasic site by pol η.[44] Thus, a common theme is emerging for the non-mutagenic replication pathway of thymine alkylated lesions by pol η, regardless of the bulky moiety size or damage position.

The results from this chapter suggest that the low percentage of dGTP insertion opposite O4-POB-T can be rationalized by the large bulky moiety compromising the lesion site base-pair interaction. In contrast, dGTP insertion is the preferred replication outcome

for O4-Me-T as the smaller methyl group does not significantly interfere with the base pairing. These structural insights support the observed relative mutagenicities of these two lesions with respect to the size of the alkyl group.[7, 9] In terms of bulky moiety position, recent computational work rationalized the mutagenic outcomes upon replication of O2-POB-T by pol η by the lesion adopting the *syn* conformation.[7, 15-16] Unlike O2-POB-T, O4-POB-T does not require a conformational change prior to replication by pol η. As a result, O2-POB-T is predicted to interact with the opposing base through the lesion Hoogsteen face, while O4-POB-T uses the WC face for dNTP insertion. This rationalizes why these lesions form different mispairs upon replication and highlights that the position of the bulky moiety greatly impacts how the lesion is accommodated within the pol η active site. Thus, this chapter highlights the importance of both the size and position of the bulky moiety for the replication of O-alkylated-T lesions, and sets the stage to examine how the shape of the alkyl chain affects the replication outcomes.

3.6 References

(1) Bilano, V.; Gilmour, S.; Moffiet, T.; Tuyl, F.; Hudson, I.; d'Espaignet, E. T.; Commar, A.; Stevens, G. A.; Shibuya, K. Global trends and projections for tobacco use, 1990-2025: An analysis of smoking indicators from the who comprehensive information systems for tobacco control. *Lancet* **2015,** *385,* 966-76.

(2) Yu, Y.; Wang, P.; Cui, Y.; Wang, Y. Chemical analysis of DNA damage. *Anal. Chem.* **2018,** *90,* 556-576.

(3) Ma, B.; Stepanov, I.; Hecht, S. S. Recent studies on DNA adducts resulting from human exposure to tobacco smoke. *Toxics* **2019,** *7*.

(4) Hecht, S. S. Biochemistry, biology, and carcinogenicity of tobacco-specific N-nitrosamines. *Chem. Res. Toxicol.* **1998,** *11,* 559-603.

(5) Brown, P. J.; Bedard, L. L.; Massey, T. E. Repair of 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone-induced DNA pyridyloxobutylation by nucleotide excision repair. *Cancer Lett.* **2008,** *260,* 48-55.

(6) Leng, J.; Wang, Y. Liquid chromatography-tandem mass spectrometry for the quantification of tobacco-specific nitrosamine-induced DNA adducts in mammalian cells. *Anal. Chem.* **2017,** *89,* 9124-9130.

(7) Du, H.; Leng, J.; Wang, P.; Li, L.; Wang, Y. Impact of tobacco-specific nitrosamine-derived DNA adducts on the efficiency and fidelity of DNA replication in human cells. *J. Biol. Chem.* **2018,** *293,* 11100-11108.

(8) Wu, J.; Li, L.; You, C.; Wang, Y.; Wang, P.; Wang, Y. Cytotoxic and mutagenic properties of minor-groove O(2)-alkylthymidine lesions in human cells. *J. Biol. Chem.* **2018,** *293,* 8638-8644.

(9) Wu, J.; Li, L.; Wang, P.; You, C.; Williams, N. L.; Wang, Y. Translesion synthesis of O4-alkylthymidine lesions in human cells. *Nucleic Acids Res.* **2016,** *44,* 9256-9265.

(10) Quinet, A.; Lerner, L. K.; Martins, D. J.; Menck, C. F. M. Filling gaps in translesion DNA synthesis in human cells. *Mutat. Res. Genet. Toxicol. Environ. Mutagen.* **2018,** *836,* 127-142.

(11) Vaisman, A.; Woodgate, R. Translesion DNA polymerases in eukaryotes: What makes them tick? *Crit. Rev. Biochem. Mol. Biol.* **2017,** *52,* 274-303.

(12) Yang, W. An overview of Y-family DNA polymerases and a case study of human DNA polymerase η. *Biochemistry* **2014,** *53,* 2793-2803.

(13) Wang, P.; Amato, N. J.; Zhai, Q.; Wang, Y. Cytotoxic and mutagenic properties of O4-alkylthymidine lesions in *Escherichia coli* cells. *Nucleic Acids Res.* **2015,** *43,* 10795-10803.

(14) Williams, N. L.; Wang, P.; Wu, J.; Wang, Y. *In vitro* lesion bypass studies of O4-alkylthymidines with human DNA polymerase η. *Chem. Res. Toxicol.* **2016,** *29,* 669-675.

(15) Gowda, A. S. P.; Spratt, T. E. DNA polymerases η and ζ combine to bypass O2-[4-(3-pyridyl)-4-oxobutyl]thymine, a DNA adduct formed from tobacco carcinogens. *Chem. Res. Toxicol.* **2016,** *29,* 303-316.

(16) Wilson, K. A.; Holland, C. D.; Wetmore, S. D. Uncovering a unique approach for damaged DNA replication: A computational investigation of a mutagenic tobacco-derived thymine lesion. *Nucleic Acids Res.* **2019,** *47,* 1871-1879.

(17) Kung, R. W.; Sharma, P.; Wetmore, S. D. Effect of size and shape of nitrogen-containing aromatics on conformational preferences of DNA containing damaged guanine. *J. Chem. Inf. Model.* **2018,** *58,* 1415-1425.

(18) Wilson, K. A.; Garden, J. L.; Wetmore, N. T.; Wetmore, S. D. Computational insights into the mutagenicity of two tobacco-derived carcinogenic DNA lesions. *Nucleic Acids Res.* **2018,** *46,* 11858-11868.

(19) Ucisik, M. N.; Hammes-Schiffer, S. Comparative molecular dynamics studies of human DNA polymerase η. *J. Chem. Inf. Model.* **2015,** *55,* 2672-81.

(20) Ucisik, M. N.; Hammes-Schiffer, S. Effects of active site mutations on specificity of nucleobase binding in human DNA polymerase η. *J. Phys. Chem. B* **2017,** *121,* 3667-3675.

(21) Wilson, K. A.; Wetmore, S. D. Conformational flexibility of the benzyl-guanine adduct in a bypass polymerase active site permits replication: Insights from molecular dynamics simulations. *Chem. Res. Toxicol.* **2017,** *30,* 2013-2022.

(22) *Hyperchem*^{tm}, Professional 8.0; Hypercube, Inc.: Gainesville, Florida 32601, USA.

(23) Dupradeau, F.-Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P. The R.E.D. Tools: Advances in RESP and ESP charge derivation and force field library building. *Phys. Chem. Chem. Phys.* **2010,** *12,* 7821–7839.

(24) Millen, A. L.; Wetmore, S. D. Glycosidic bond cleavage in deoxynucleotides - a density functional study. *Can. J. Chem.* **2009,** *87,* 850-863.

(25) Roy Dennington, T. A. K. and John M. Millam, Semichem Inc. Shawnee Mission, GaussView, Version 6. **2016**.

(26) Macrae, C. F.; Bruno, I. J.; Chisholm, J. A.; Edgington, P. R.; McCabe, P.; Pidcock, E.; Rodriguez-Monge, L.; Taylor, R.; van de Streek, J.; Wood, P. A. Mercury CSD 2.0 - new features for the visualization and investigation of crystal structures. *J. Appl. Crystallogr.* **2008,** *41,* 466-470.

(27) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Jr. J. A. M.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Revision E.01; Gaussian, Inc.: Wallingford CT, **2009**.

(28) Nakamura, T.; Yamagata, Y.; Yang, W. Watching DNA polymerase η make a phosphodiester bond. *Nature* **2013,** *53,* 254-257.

(29) Batra, V. K.; Beard, W. A.; Shock, D. D.; Krahn, J. M.; Pedersen, L. C.; Wilson, S. H. Magnesium-induced assembly of a complete DNA polymerase catalytic complex. *Structure* **2006,** *14,* 757-766.

(30) Mendieta, J.; Cases-Gonzalez, C. E.; Matamoros, T.; Ramirez, G.; Menendez-Arias, L. A Mg2+-induced conformational switch rendering a competent DNA polymerase catalytic complex. *Proteins: Struct. Funct. Bioinf.* **2008,** *71,* 565-574.

(31) Ummat, A.; Silverstein, T. D.; Jain, R.; Buku, A.; Johnson, R. E.; Prakash, L.; Prakash, S.; Aggarwal, A. K. Human DNA polymerase η is pre-aligned for dNTP binding and catalysis. *J. Mol. Biol.* **2012,** *415,* 627-634.

(32) Allner, O.; Nilsson, L.; Villa, A. Magnesium ion-water coordination and exchange in biomolecular simulations. *J. Chem. Theory Comput.* **2012,** *8,* 1493-1502.

(33) Perlow, R. A.; Broyde, S. Toward understanding the mutagenicity of an environmental carcinogen: Structural insights into nucleotide incorporation preferences. *J. Mol. Biol.* **2002,** *322,* 291-309.

(34) Guo, S.; Leng, J.; Tan, Y.; Price, N. E.; Wang, Y. Quantification of DNA lesions induced by 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol in mammalian cells. *Chem. Res. Toxicol.* **2019,** *32,* 708-717.

(35) Wang, J.; Wang, W.; Kollman, P. A. Antechamber: An accessory software package for molecular mechanical calculations. *J. Am. Chem. Soc.* **2001,** *222,* U403.

(36) Case, D.; Babin, V.; Berryman, J.; M. Betz, R.; Cai, Q.; S. Cerutti, D.; Cheatham, T.; Darden, T.; Duke, R.; Gohlke, H.; Götz, A.; Gusarov, S.; Homeyer, N.; Janowski, P.; Kaus, J.; Kolossváry, I.; Kovalenko, A.; Lee, T.-S.; Legrand, S.; A. Kollman, P. *Amber 2014*. 2014.

(37) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004,** *25,* 1157-1174.

(38) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **2015,** *11,* 3696-3713.

(39) Galindo-Murillo, R.; Robertson, J. C.; Zgarbova, M.; Sponer, J.; Otyepka, M.; Jurecka, P.; Cheatham, T. E. Assessing the current state of amber force field modifications for DNA. *J. Chem. Theory Comput.* **2016,** *12,* 4114-4127.

(40) Price, D. J.; Brooks, C. L. III, A modified TIP3P water potential for simulation with ewald summation. *J. Chem. Phys.* **2004,** *121,* 10096-10103.

(41) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **2013,** *9,* 3084-3095.

(42) Su, Y.; Patra, A.; Harp, J. M.; Egli, M.; Guengerich, F. P. Roles of residues arg-61 and gln-38 of human DNA polymerase η in bypass of deoxyguanosine and 7,8-dihydro-8-oxo-2'-deoxyguanosine. *J. Biol. Chem.* **2015,** *290,* 15921-15933.

(43) Ucisik, M. N.; Hammes-Schiffer, S. Relative binding free energies of adenine and guanine to damaged and undamaged DNA in human DNA polymerase η: Clues for fidelity and overall efficiency. *J. Am. Chem. Soc.* **2015,** *137,* 13240-13243.

(44) Patra, A.; Zhang, Q.; Lei, L.; Su, Y.; Egli, M.; Guengerich, F. P. Structural and kinetic analysis of nucleoside triphosphate incorporation opposite an abasic site by human translesion DNA polymerase η. *J. Biol. Chem.* **2015,** *290,* 8028-8038.

(45) Choi, J.-Y.; Lim, S.; Kim, E.-J.; Jo, A.; Guengerich, F. P. Translesion synthesis across abasic lesions by human B-family and Y-family DNA polymerases α, δ, η, ι, κ, and rev1. *J. Mol. Biol.* **2010,** *404,* 34-44.

Chapter 4: Impact of Alkyl Chain Shape on the Replication Outcomes of O2-alkyl-T Lesions[a,b,c]

4.1 Introduction

Experimental work has considered a number of synthetic O-alkyl-T lesions to uncover the mutagenic patterns of these adducts and the role of TLS in their replication *in vivo* (human and *E. coli* cells) and *in vitro*.[1-6] These studies have considered how increasing the length, as well as varying the shape, of the alkyl chain affect the replication outcomes. In human cells and *in vitro*, straight-chained O2-alkyl-T lesions were found to be more mutagenic than their branched counterparts.[1-2] In contrast, straight-chained O4-alkyl-T lesions are less mutagenic than their branched counterparts in similar environments.[4-5] When coupled with the work discussed in Chapters 2 and 3, this data further reveals the effect of alkyl chain position on the mutagenicity. More interestingly, these outcomes highlight that subtle changes in the alkyl chain shape lead to differences in the lesion mutagenicities. Unfortunately, there is a lack of structural evidence to understand the underlying driving force for these results.

Previous chapters in this thesis and recent structural work[7] have shown that computational chemistry can be a key tool for rationalizing the mutagenic patterns of O-alkyl-T lesions,[1, 4, 7, 8] specifically O-Me-T and O-POB-T adducts. In particular, previous

---

[a] *The Journal of American Chemical Society* reference style is used in this chapter.
[b] Adapted from Bhutani, P. Nikkel, D.J. Wilson, K.A. Wetmore, S.D. (2019). Impact of Alkyl Chain Shape on the Replication Outcomes of O2-alkyl-T Lesions (in preparation).
[c] P.B. performed calculations and data analysis on the base pair and polymerase models, and wrote the first draft of the manuscript. D.J.N. performed calculations on the O2-alkyl-T nucleobase and base-pair models. K.A.W. generated parameters for O2-alkyl-T lesions. S.D.W. contributed to project and calculation design, data interpretation and manuscript writing.

computational work in this thesis on alkylated T lesions focused on the impact of position (i.e. O2 versus O4) and size (i.e. the addition of a methyl group versus a POB group) of the alkyl chain. Despite differences in the position or composition of the associated alkyl chains, O2-alkyl-T and O4-alkyl-T lesions similarly alter the WC face (i.e. N3 becomes a hydrogen-bonding acceptor). The data presented in this thesis emphasize that the experimental mutagenic patterns in human cells associated with O2-Me-T (non-mutagenic)[1] and O2-POB-T (primarily non-mutagenic, 15% dTTP insertion)[8] can only be rationalized if the lesions adopt a *syn* conformation about the glycosidic bond.[7] This orientation points the damaged group into the pol η open pocket and aligns the adduct in the active site such that the lesion Hoogsteen (HG) face interacts with an incoming dNTP. In contrast, the alkyl chains of O4-Me-T and O4-POB-T are preferentially pointed into the pol η open pocket when the lesion adopts the *anti* glycosidic conformation native to B-DNA. Therefore, unlike their O2 counterparts, the lesion Watson-Crick (WC) binding face of the O4 adducts is available to interact with an incoming dNTP. This difference in glycosidic orientation explains why both O4 lesions preferentially form mispairs with G.[4, 8] The computational results also rationalize the higher percentage of dGTP insertion for O4-Me-T (60%)[4] than O4-POB-T (35%)[8] based on the bulky POB moiety periodically disrupting the base-pair interaction. Regardless of differences in the frequency of mispair formation, dATP insertion opposite all of the previously mentioned O-alkyl-T lesions is suggested to be promoted through a water-mediated interaction similar to that previously proposed for the replication of an abasic site by pol η based on crystallographic data.[9]

Despite the previous computational work on O-alkyl-T lesions, no studies to date have addressed questions surrounding the effect of alkyl chain shape on the mutagenic

outcomes.[1-2] Therefore, to begin to answer these questions, this chapter focuses on understanding the impact of alkyl chain shape for O2-alkyl-T lesions. Both *in vitro* and *in vivo* (human cells) studies on O2-alkyl-T lesions indicate that straight-chained alkyl lesions (such as n-propyl and n-butyl) are more mutagenic compared to their branched alkyl chains counterparts (such as iso-propyl and iso-butyl), although dATP insertion dominates for both lesion types.[1-2] Specifically, in human cells, O2-n-propyl-T and O2-n-butyl-T (O2-nBu-T, Figure 4.1, left) result in 13% and 12% dTTP insertion, respectively, whereas O2-iso-propyl-T and O2-iso-butyl-T (O2-iBu-T, Figure 4.1, right) result in < 5% dTTP insertion.[1] Similarly, *in vitro* studies show that mispairs can occur opposite straight-chained lesions as often as dATP insertion, but branched alkyl lesions preferentially insert dATP upon replication (i.e. mispairs occur about one-fifth as often as dATP insertion).[2] Comparable to other O-alkyl-T lesions discussed in this thesis, knockout studies indicate that human TLS pol η plays an important role in the mutagenic outcomes of these O2-alkyl-T lesions in human cells.[1-2] This is an interesting finding since pol η has an open pocket on the major groove side of the growing DNA strand, while O2-alkyl-T adducts are minor groove lesions. Therefore, it is unclear how an alkyl chain at the O2 position may disrupt the active site and how pol η can accommodate a range of O2-alkyl-T lesions with different alkyl chain shapes.[7]

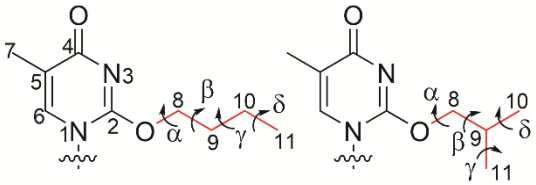


Figure 4.1: Chemical structure and numbering of O2-nBu-T (left) and O2-iBu-T (right), with definitions of key dihedral angles within the alkyl chain including α=∠(N3C4O4C8), β=∠(C4O4C8C9), γ=∠(O4C8C9C10), and δ=∠(C8C9C10C11).

In order to provide the structural insight necessary to better understand how the shape of the alkyl chain affects the replication outcomes,[1-2] the present chapter uses a combination of DFT calculations and MD simulations to study O2-nBu-T and O2-iBu-T (Figure 4.1). The O2-nBu-T and O2-iBu-T lesions were chosen as prototypical examples in this chapter since the general trend for straight-chained and branched alkyl lesions is similar regardless of lesion size.[1] Initially, a conformational search on the isolated lesions was performed to consider the alkyl chain flexibility. Next, MD simulations on the pol η pre-insertion complexes were conducted to understand how the lesions fit within the active site and how the alkyl chain is accommodated relative to the surrounding residues. Following this, DFT calculations on the lesions paired opposite select canonical nucleobases provide insight into the potential lesion hydrogen-bonding patterns. Finally, MD simulations were performed to probe the structure and dynamics of the pol η ternary complex corresponding to select dNTP insertion opposite O2-nBu-T or O2-iBu-T in order to determine which pairs are best stabilized in the active site. Taken together, this chapter provides a better understanding of how the shape of the alkyl chain affects O2-alkyl-T lesion accommodation by pol η, and uncovers key interactions that may play a role in facilitating replication of O2-alkyl-T lesions by a polymerase that has historically been thought to process major groove lesions.[10-11]

4.2 Methods

The same general computational approach as described in Chapters 2 and 3 was used in the present work. Specifically, conformational searches were initially performed with respect to the dihedral angles within the lesion alkyl chain (α, β, γ and δ, Figure 4.1) for both O2-nBu-T and O2-iBu-T using AMBER96 (Tables C.1 and C.2) as implemented

in Hyperchem (version 8.0).[12] In these models, a methyl group replaced the deoxyribose moiety as the alkyl chains are not anticipated to interact with the DNA backbone. The resulting conformations were optimized with B3LYP/6-31G(d,p) and single-point energies were computed using B3LYP-D3(BJ)/6-311++G(2df,2p). Unique conformations isolated from the nucleobase model were manually docked into the pol η active site with respect to T ring atoms to identify the conformations that can be readily accommodated by the enzyme. Potential energy surfaces (PES) for nucleoside models with the lowest energy conformation of the bulky moiety were computed to determine the rotational barrier about the glycosidic bond for these lesions. Specifically, the PES was searched with respect to ∠(O4′C1′N1C2) by constraining the dihedral angle in 10° increments (for a total of 36 points) using the same methods as described above. DFT was used to analyze base-pair models between the lowest energy conformation of each lesion that fit within the pol η active site and select canonical nucleobases. Optimizations and single-point energies for the base-pair models were computed at the same level of theory as in the conformational search. All binding energies include counterpoise corrections, deformation energies and zero-point energy corrections (0.9857). The interstrand distances (R(C1′···C1′)) and base pair opening angles (∠(N9/N1C1′C1′)) were measured using GaussView 5,[13] and the interplanar angles were calculated using Mercury.[14] All structural parameters were compared to the canonical T:A pair calculated at the same level of theory. All monomer and base-pair calculations were completed with Gaussian 09 (revision D.01 or E.01).[15]

Pre-insertion and ternary complexes were generated from the crystal structure corresponding to dATP insertion opposite O4-ethylthymine by pol η (PDB ID: 5DQG).[16] The structure was modified using Gaussview 5 as described in Chapter 2. In the pre-

insertion complexes, the incoming dNTP and Mg(II) ions were removed since these have been suggested to not play a role until dNTP binding.[17-18] In the ternary complexes, two Mg(II) ions were included as these have been shown to correctly align the incoming dNTP for the reaction.[18] Partial charges of the O2-nBu-T and O2-iBu-T lesions were derived from the RED.vIII.4 program[19] and atom types were assumed using the Antechamber 1.4 script.[20] Parameters were generated using the parm10 and General Amber Force Field (GAFF)[21] files as implemented in Amber14 (Tables C.3 and C.4).[22] In both the pre-insertion and ternary complexes, the AMBER OL15 force field was used to describe DNA,[23] while the AMBER14SB force field was used to describe the protein.[24] Systems were neutralized, solvated, and brought to a biologically relevant temperature as previously described for the DNA–polymerase models in Chapter 2. Pre-insertion and ternary complexes were simulated for 1 μs. For the ternary complexes, the average hydrogen-bonding parameters, interstrand distance (R(C1′···C1′)) and base pair opening angle (∠(N9/N1C1′C1′)) were analyzed, as well as the R(O3′···Pα) and ∠(O3′–Pα–Pβ) reaction parameters, key active site residue interactions and the Mg(II) coordination, similar to Chapters 2 and 3. Data was recorded every 0.05 ns using the CPPTRAJ module of Amber 18.[25] Average interaction energies and interplanar angles were computed every 5 ns using in-house scripts. Interaction energies were computed at the same level of theory as DFT models and included counterpoise corrections.

4.3 Results and Discussion

4.3.1 Although O2-nBu-T is more flexible than O2-iBu-T, both lesions disrupt the active site in the pol η pre-insertion complex: The conformational search for the isolated

O2-nBu-T nucleobase resulted in 27 conformations, which collapsed into 25 unique conformations upon optimization with DFT. All conformations fall within an energy range of 4.5 kJ/mol (Figure C.1). Prior to considering the pol η pre-insertion complex, the unique conformations were overlaid onto an X-ray crystal structure of pol η corresponding to dATP insertion opposite canonical T (PDB ID: 4ECS).[26] For O2-nBu-T, 14 of 25 conformations are in close proximity to key active site residues (< 2.0 Å distance), including Q38 and I48, or the 3′-G with respect to the template base (Figure 4.2a, top). One structure that does not clash with surrounding residues in the active site was selected as the starting structure for the MD simulation on the pol η pre-insertion complex corresponding to O2-nBu-T replication. Throughout the simulation, the alkyl chain maintains a high degree of flexibility (rmsd = 0.3 ± 0.2 Å, Figure C.2a), which disrupts the active site of pol η by pushing away the 5′-G into the open pocket (Figure 4.2a, middle). Additionally, an overlay of snapshots taken over the simulation with respect to the 3′-G ring atoms shows significant movement of Q38 and I48 over the course of the simulation (rmsd of Q38 and I48 = 1.6 ± 0.1 Å and 1.4 ± 0.6 Å, respectively, Figure C.2a). Since Q38 directly interacts with O2 of the template base in an X-ray crystal structure corresponding to the replication of canonical DNA by pol η (PDB ID: 4ECS),[26] this residue has been proposed to play a role in stabilizing the template base.[27-29] However, the flexible O2 alkyl chain disrupts interactions with Q38 such that this residue only interacts with the lesion deoxyribose (30% with O3′, Figure C.3a). Nevertheless, O2-nBu-T is still positioned in a manner that would permit hydrogen bonding with an incoming dNTP and further stabilization of the active site (Figure 4.2a, bottom).

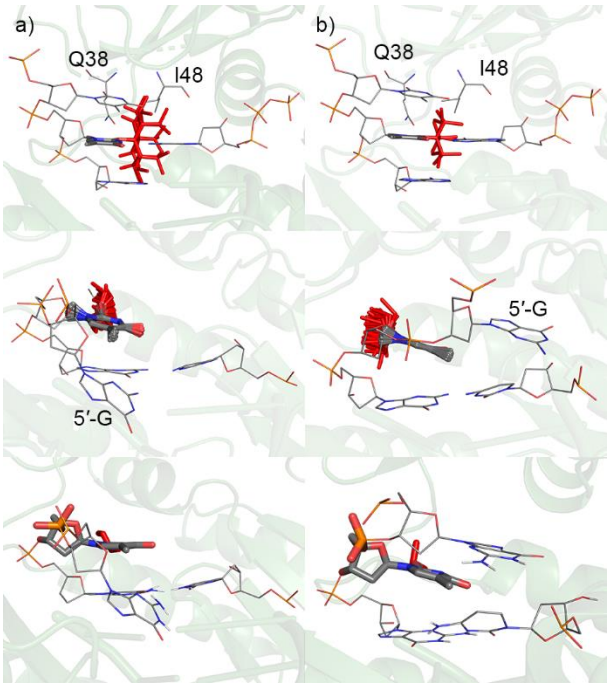


Figure 4.2: Overlay of lesion conformations from the conformational search onto pol η crystal structure (PDB ID: 4ECS, top) according to the T ring atoms. MD snapshots taken every 5 ns (middle) and MD representative structure of the lesion (bottom) in the pol η pre-insertion complex for a) anti-O2-nBu-T and b) anti-O2-iBu-T.

Both the AMBER and subsequent DFT optimizations from the conformational search for isolated O2-iBu-T resulted in only 7 conformations, which ranged over 3.8 kJ/mol (Figure C.1). When the conformations were overlaid onto the same crystal structure considered for the O2-nBu-T lesion, 4 of 7 O2-iBu-T conformers are in close proximity < 2.0 Å distance) to active site residues (Q38 and I48) or the 3′-G (Figure 4.2b, top). In the pre-insertion complex corresponding to O2-iBu-T replication, the alkyl chain is much more restricted over the MD simulation (rmsd = 0.1 ± 0.0 Å, Figure 4.2b, middle and C.2b) compared to O2-nBu-T. However, the 5′-G is disrupted and takes the space of the incoming dNTP (Figure 4.2b, middle). Additionally, both Q38 and I48 remain very dynamic throughout the simulation (rmsd of Q38 and I48 = 1.4 ± 0.1 Å and 1.1 ± 0.4 Å, respectively, Figure C.2), with Q38 only interacting with the lesion deoxyribose (15% with O3′ and 7%

with O4′, Figure C.3b). Thus, although both the O2-nBu-T and O2-iBu-T lesions are positioned to hydrogen bond with an incoming dNTP (Figure 4.2b, bottom), it is unclear how the active site disruptions observed in the pre-insertion complexes will affect the alignment of the incoming dNTP. Therefore, the hydrogen-bonding properties of both lesions in the pol η active site will be considered in the following sections.

4.3.2 dTTP insertion opposite O2-alkyl-T lesions is unfavourable, contrasting the experimental mutagenic patterns: Experimental mutagenic data collected in human cells indicate that straight-chained O2-alkyl-T lesions exhibit higher frequencies of dTTP insertion (13%) than their branched counterparts (4%).[1] To provide insight into the impact of alkyl chain shape on these outcomes, the base-pairing properties of O2-nBu-T and O2-iBu-T are compared to canonical T:A (Figure C.4). Prior to considering these lesion pairs within the active site of pol η, the isolated base pairs were computed to better understand the lesion hydrogen-bonding potential. The O2-nBu-T:T pair only contains one strong O4(O2-nBu-T)···H–N3(T) hydrogen bond and the interstrand distance is ~ 1 Å shorter (9.423 Å) than canonical T:A (Figure C.4, top). Although the alkyl chain slightly disrupts the interplanar angle (18.2°) of canonical DNA, which affords an additional weak hydrogen bond between the alkyl chain and the O2 of T, these structural features result in a weaker interaction energy (–39.2 kJ/mol) compared to canonical T:A (–69.8 kJ/mol). In the pol η ternary complex corresponding to dTTP insertion opposite O2-nBu-T, dTTP is properly positioned by active site residues (C16, F18, Y52, R55, R61 and K231) and there are no major differences in the orientations of these residues compared to the crystal structure of dATP insertion opposite canonical T by pol η (rmsd = 0.7 ± 0.1 Å, Figure C.5a). The

reaction parameters (3.5 Å and 147.7°, Figure 4.3a, bottom) do not significantly deviate from their ideal values (< 3.5 Å and ~ 180°), which makes the complex conducive for the nucleotidyl transfer reaction. Additionally, the coordinating residues around the two Mg(II) ions, which are important for dNTP binding and enzyme catalysis,[17-18] are tightly bound for the majority of the simulation (82% < 2.5 Å, Table C.5). However, interactions between O2-nBu-T and Q38 occur for < 1% of the simulation (Figure C.7a). Furthermore, the alkyl chain is highly flexible (rmsd = 0.7 ± 0.1 Å, Figure C.6a). As a result of reduced Q38 interactions and disruption from the alkyl chain, there are no base-pair interactions between the lesion and the incoming dTTP (Figure 4.3a, bottom), which would greatly impact dTTP insertion.

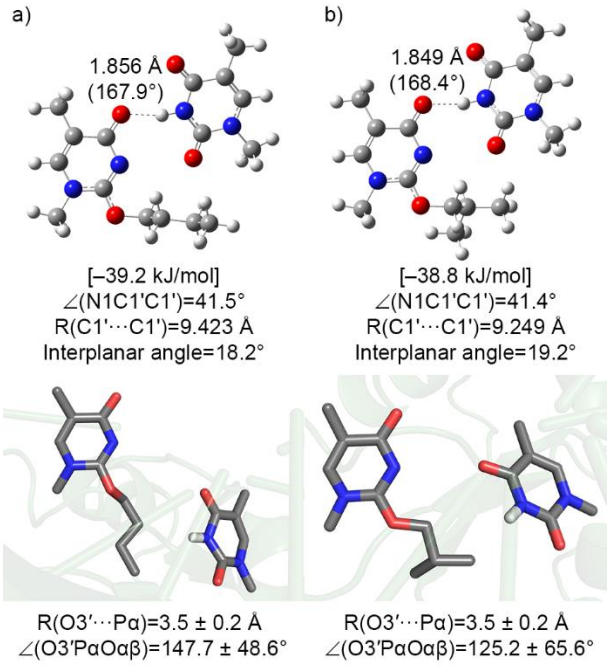


Figure 4.3: Key DFT structural features (Å and deg.) and interaction energies [kJ/mol] hydrogen bonded with T (top) and MD representative structure of the ternary complex with reaction parameters (Å and deg.) for dTTP insertion (bottom) opposite a) anti-O2-nBu-T and b) anti-O2-iBu-T.

Similar to O2-nBu-T, one O4(O2-iBu-T)···H–N3(T) hydrogen bond forms in the isolated base-pair model and the interstrand distance is ~ 1 Å shorter (9.429 Å, Figure 4.3b, top) than that for canonical T:A. Taken together, the structural features lead to a weak interaction energy (–38.8 kJ/mol). Overall, the DFT models do not show any significant differences in the lesion mispairs with T based on alkyl chain shape. Furthermore, in the pol η ternary complex corresponding to dTTP insertion opposite O2-iBu-T, the Mg(II) coordination (74%, Table C.6) and the position of key active site residues (C16, F18, Y52, R55, R61 and K231, rmsd = 0.6 ± 0.2 Å, Figure C.5b) are maintained. However, the reaction angle deviates from ideal value (125.2°), despite a favourable reaction distance (3.5 Å). Additionally, although the alkyl chain is restricted (rmsd = 0.1 ± 0.0 Å, Figure C.6) compared to its straight-chained counterpart discussed above, Q38 interactions are still disrupted (O4′ for 23% and O5′ for 21%, Figure C.7) and the lesion is positioned in the open pocket of pol η, resulting in no direct interactions with the incoming dTTP (Figure 4.3b, bottom).

Overall, the poorly aligned ternary complexes corresponding to O2-nBu-T and O2-iBu-T replication cannot explain the experimentally observed dTTP insertion opposite these lesions.[1] However, disruptions occur to the pol η active site because the alkyl chain is not situated in the open pocket. Furthermore, other O2-alkyl-T lesions (i.e. O2-Me-T and O2-POB-T) have been proposed to undergo rotation about the glycosidic bond to facilitate pol η replication by directing the damaged group into the open pocket. This rotation is also possible for the O2-alkyl-T lesions of interest in this chapter (Figure C.8 and C.9). Therefore, the accommodation of the *syn* conformation of these lesions by pol η will be considered in the following sections.

4.3.3 Pol η pre-insertion complexes containing *syn*-O2-nBu-T and *syn*-O2-iBu-T preserve the active site conformation and are aligned for dNTP insertion: As discussed for O2-Me-T in Chapter 2, the *anti* and *syn* orientations about the glycosidic bond of the O2-nBu-T nucleoside are almost isoenergetic, with an energy difference of ~ 4 kJ/mol and a barrier of ~ 45 kJ/mol (Figure C.8). An overlay of the O2-nBu-T conformations isolated from the AMBER/DFT conformational search into the pol η active site according to the template base ring atoms such that the lesion adopts the *syn* orientation indicates that 15 out of 25 conformations of the alkyl chain fit (< 2.0 Å distance, Figure 4.4a, top). Conformations that do not fit either directly clash with or fall in close proximity to deoxyribose of the 5′-G. However, unlike *anti*-O2-nBu-T, the alkyl chain points into the open pocket of pol η and therefore no active site residues are disrupted by the damaged group. Nevertheless, in the pol η pre-insertion complex, the *syn*-O2-nBu-T alkyl chain remains very flexible (rmsd = 0.3 ± 0.2, Figure C.2c) and displaces the 5′-G (Figure 4.4a, middle), but the lesion does not sterically clash with Q38 or I48. Although rotation about the glycosidic bond prevents interactions between Q38 and O2 of the lesion, a hydrogen bond exists between Q38 and O3′ on the lesion deoxyribose moiety (13%, Figure C.3c). Regardless, *syn*-O2-nBu-T maintains a position in the pol η active site such that an incoming dNTP can be situated in an open void in order to interact with the lesion Hoogsteen face (Figure 4.4a, bottom).

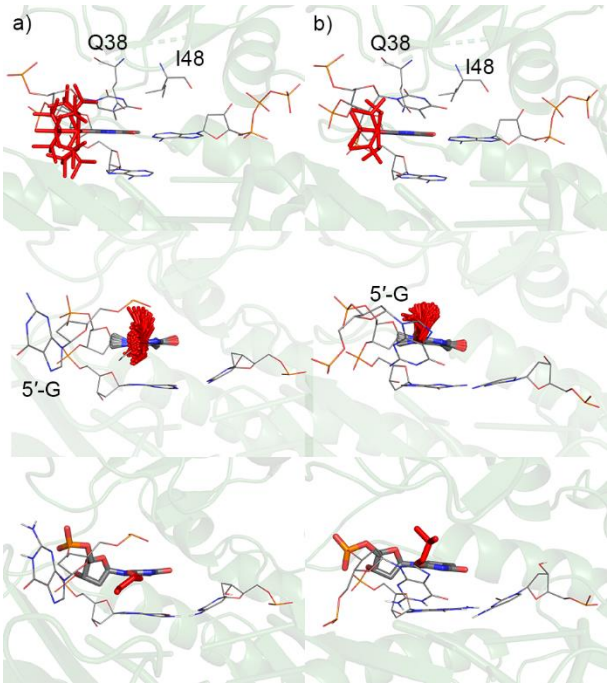


Figure 4.4: Overlay of lesion conformations from the conformational search onto pol η crystal structure (PDB ID: 4ECS, top) according to the T ring atoms. MD snapshots taken every 5 ns (middle) and MD representative structure of the lesion (bottom) in the pol η pre-insertion complex for a) *syn*-O2-nBu-T and b) *syn*-O2-iBu-T.

As discussed for O2-nBu-T, rotation about the glycosidic bond of O2-iBu-T indicates the *anti* and *syn* orientations are nearly isoenergetic (~ 1 kJ/mol) and are connected by a barrier of ~ 41 kJ/mol (Figure C.9). When the AMBER/DFT conformational search results were manually docked into the pol η active site such that the lesion adopts the *syn* orientation, there is no disruption to active site residues or the incoming dNTP as the alkyl chain points into the pol η open pocket (Figure 4.4b, top). However, 5 of 7 adduct conformations clash with the deoxyribose of the 5′-G (< 2.0 Å distance). In the pol η pre-insertion complex corresponding to *syn*-O2-iBu-T replication, the alkyl chain exhibits low flexibility (rmsd = 0.1 ± 0.0 Å, Figure C.2d) despite being in the open pocket, which highlights the inherently conformationally restricted nature of this branched lesion. However, the 5′-G is pushed toward the open pocket in the simulation

(Figure 4.4b, bottom). Like *syn*-O2-nBu-T, rotation about the glycosidic bond prevents interactions between Q38 and the lesion nucleobase, but Q38 forms a hydrogen bond with the deoxyribose moiety for a small portion of the simulation (7% with O4′, Figure C.3). Additionally, active site residues surrounding the lesion (i.e. Q38 and I48) are not as dynamic over the simulation (Figure C.2d) compared to the corresponding *anti*-O2-iBu-T pre-insertion complex. Perhaps most importantly, the Hoogsteen face of *syn*-O2-iBu-T is aligned to form hydrogen bonds with the incoming dNTP (Figure 4.4b, bottom).

Thus, the pre-insertion complexes corresponding to *syn*-O2-nBu-T and *syn*-O2-iBu-T replication appear to be more stable and the lesion is better aligned with an opposing void in the active site to permit dNTP insertion. To determine whether this orientation about the glycosidic bond is conducive for the reaction and thereby can rationalize the experimental mutagenic patterns,[1] the pol η ternary complexes corresponding to dTTP insertion opposite *syn*-O2-nBu-T and *syn*-O2-iBu-T are considered in the following section.

4.3.4 dTTP forms a stable base pair with O2-nBu-T but not O2-iBu-T in the pol η insertion complex, which aligns with the experimental mutagenic patterns: In contrast to *anti*-O2-nBu-T, DFT models suggest the interstrand distance of the *syn*-O2-nBu-T:T mispair (10.293 Å, Figure 4.5a, top) is comparable to that of canonical T:A. Furthermore, although only one strong O4(O2-nBu-T)···H–N3(T) hydrogen bond forms in the lesion mispair, the pair distorts to form an additional weak hydrogen bond between O2 of T and the C5 methyl group of O2-nBu-T (Figure 4.5a, top). Nevertheless, these structural parameters lead to a weaker interaction energy (–34.0 kJ/mol) than canonical T:A (–69.8

kJ/mol). In the pol η ternary complex corresponding to dTTP insertion opposite *syn*-O2-nBu-T, the incoming dTTP is well positioned due to the interactions between key active site residues, with the exception of R61, which points toward the lesion base pair (rmsd = 0.7 ± 0.2 Å, Figure C.5c). The Mg(II) exhibits tight coordination (82% < 2.5 Å, Table C.7) and, although the reaction distance is near the ideal value (3.5 Å), the reaction angle fluctuates throughout the simulation (113.6 ± 58.9°, Figure 4.5a, bottom). Similar to the situation for the O4-POB-T lesion discussed in Chapter 3, the deviation in the reaction parameter results in one conformation that maintains a near-ideal reaction angle (48%) and two conformations in which the reaction angle is < 100° (52% total occupancy, Figure C.10). Regardless, this suggests that dTTP is at least sometimes correctly aligned for the nucleotidyl transfer reaction.

Despite the *syn*-O2-nBu-T alkyl chain pointing into the open pocket of pol η and void of interactions with the active site residues, the alkyl group maintains similar flexibility throughout the simulation as observed for dTTP insertion opposite *anti*-O2-nBu-T (rmsd = 0.5 ± 0.2 Å, Figure C.6c). Lesion interactions with Q38 are maintained between both O3' (39%) and O4' (13%, Figure C.7c), which helps position the adduct in the active site opposite the incoming dTTP. However, the high flexibility of the alkyl chain disrupts the stacking interaction between *syn*-O2-nBu-T and the 5'-G, forcing the flanking base into the pol η open pocket (Figure C.11), which is similar to the discussion for the pre-insertion complex (Figure 4.4a, middle). Despite the loss of this stacking interaction, one consistent O4(*syn*-O2-nBu-T)···H–N3(dTTP) hydrogen bond is formed (57% occupancy, Table C.8), and the base pair contains reasonable average structural parameters (R(C1′···C1′) = 10.9 Å, ∠(N9C1′C1′) = 83.0° and interplanar angle = 26°). These structural features lead to an

interaction energy of –27 kJ/mol (Figure 4.5a). As discussed for the O4-POB-T lesion in Chapter 3, the base pair opening angle has a high standard deviation, with an unfavourable opening angle (> 70°) directly correlating with a non-optimal reaction angle (< 100°) (Figure C.10). Despite these deviations, both the structural and reaction parameters periodically maintain values that would permit dTTP insertion. Most notably, the position of the flanking 5'-G allows for hydrogen bonding between R61 and O4 of dTTP (68% occupancy), as well as periodic interactions between R61 and O4 of *syn*-O2-nBu-T (15%, Figure C.11a). Previous work has proposed that R61 is very flexible and is a catalytically important residue.[28-30] Therefore, the R61 interactions observed in the simulation for dTTP insertion opposite *syn*-O2-nBu-T likely play a role in stabilizing the lesion base pair in the pol η active site.

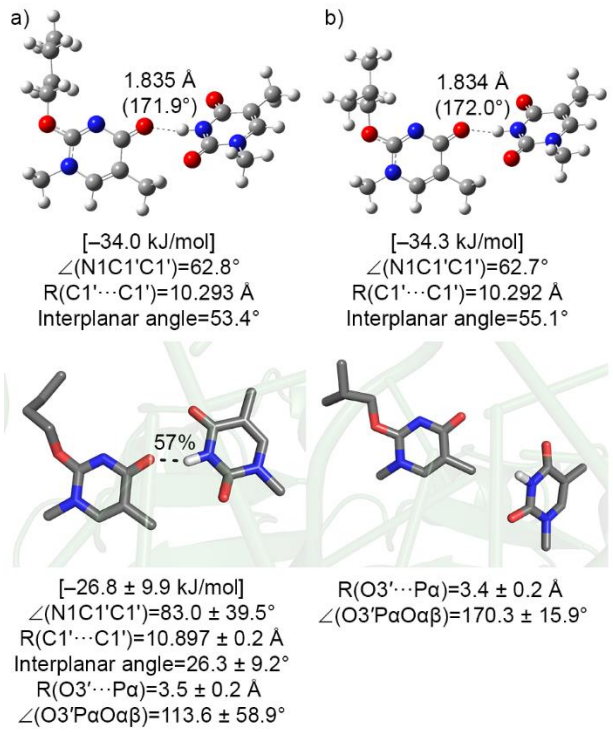


Figure 4.5: Key DFT structural features (Å and deg.) and interaction energies [kJ/mol] hydrogen bonded with T (top) and MD representative structure of the ternary complex with applicable structural and reaction parameters (Å and deg.) for dTTP insertion (bottom) opposite a) *syn*-O2-nBu-T and b) *syn*-O2-iBu-T.

Due to the fact that the alkyl chain does not interfere with WC hydrogen bonding, *syn*-O2-iBu-T:T exhibits almost identical structural features (Figure 4.5b, top) compared to the *syn*-O2-nBu-T:T pair (Figure 4.5a, top). In the corresponding pol η ternary complex containing *syn*-O2-iBu-T, dTTP is well positioned by active site residues (rmsd = 0.7 ± 0.2 Å, Figure C.5d), Mg(II) is coordinated for the majority of the simulation (72% < 2.5 Å, Table C.9), and the average reaction parameters are conducive for the nucleotidyl transfer reaction (3.4 Å and 170.3°, Figure 4.5d, bottom). However, unlike *syn*-O2-nBu-T, there are no major fluctuations in the reaction angle, which remains near the ideal value throughout the simulation. The conformationally restricted alkyl chain of the lesion (rmsd = 0.1 ± 0.0 Å, Figure C.6d) points into the open pocket of pol η. Interactions between Q38 and both O3′ and O4′ are maintained for ~ 30% of the simulation (Figure C.7d). Most importantly, although the incoming dTTP is positioned to interact with *syn*-O2-iBu-T, no persistent hydrogen bond occurs between the bases (Figure 4.5b, bottom). In fact, in contrast to *syn*-O2-nBu-T, the conformationally restricted branched alkyl chain does not disrupt the stacking interaction between the lesion and 5′-G, which prevents R61 from stabilizing the hydrogen-bonded pair between the lesion and incoming dTTP (Figure C.11b).

Overall, the present work suggests that both the O2-nBu-T and O2-iBu-T lesions likely adopt a *syn* orientation when replicated by pol η. *syn*-O2-nBu-T can maintain consistent hydrogen bonding with an opposing dTTP, which is stabilized by R61. Nevertheless, the repositioning of R61 to fulfill this role requires displacement of 5'-G by the flexible alkyl chain, which may impact extension past the lesion site. When coupled with the non-optimal reaction angle and base pair opening angle, these results correlate

with the reported low (13%) mutagenicity.[1] In contrast, the conformationally restricted alkyl chain of *syn*-O2-iBu-T does not disrupt the pol η active site to stabilize the lesion base pair. This data directly correlate with the reduced frequency of dTTP insertion (4%) opposite O2-iBu-T compared to its straight-chained counterpart (13%).[1] Interestingly, O2-POB-T leads to a similar frequency of dTTP insertion by pol η compared to O2-nBu-T in human cells,[8] and computational work suggests that the POB moiety is very flexible.[7] Comparison to the present chapter indicates that there are similarities in how O2-nBu-T and O2-POB-T are accommodated by pol η. Regardless, both O2-nBu-T and O2-iBu-T are primarily non-mutagenic,[1] and additional calculations are required to ensure that dATP insertion is promoted opposite these synthetic lesions in the *syn* orientation in order to rationalize the observed mispair frequencies. Therefore, the corresponding base pairs and pol η ternary complexes will be considered in the following section.

4.3.5 dATP insertion is stabilized by the formation of a water chain opposite both O2-nBu-T and O2-iBu-T, supporting non-mutagenic replication past these lesions: DFT models indicate that although the *syn*-O2-nBu-T:A pair maintains one strong O4(O2-nBu-T)···H–N6(A) hydrogen bond, the interstrand distance is elongated (R(C1′···C1′) = 12.805 Å, Figure 4.6a, top) compared to canonical T:A. Furthermore, this unfavourable pairing results in a weaker interaction energy (–34.9 kJ/mol) than canonical T:A (–69.8 kJ/mol). Nevertheless, in the corresponding pol η ternary complex, the reaction distance (3.4 Å) and Mg(II) coordination (89% < 2.5 Å, Table C.10), along with the orientation of stabilizing active site residues (C16, F18, Y52, R55, R61 and K231, rmsd = 0.6 ± 0.1, Figure C.5e), favourably position the incoming dATP opposite *syn*-O2-nBu-T. However, the reaction

angle fluctuates throughout the simulation (117.2 ± 64.8°). Similar to dTTP insertion opposite *syn*-O2-nBu-T, deviations in the reaction angle result in an ideal conformation being adopted periodically (55% of the time, Figure C.12), which makes this complex sometimes conducive for the nucleotidyl transfer reaction. The alkyl chain remains flexible (rmsd = 0.5 ± 0.2, Figure C.6e) and Q38 interactions are limited to O3′ (54%, Figure C.7e). Most importantly, the lesion does not form direct interactions with the incoming dATP. Nevertheless, a water chain forms between N3 of dATP and the 3′-G (Figure 4.6a, bottom), which is similar to the configuration observed for other O-alkyl-T lesions in this thesis (Chapters 2 and 3) and for the previously studied O2-POB-T,[7] as well as in the crystal structure corresponding to dATP insertion opposite an abasic site by pol η.[9]

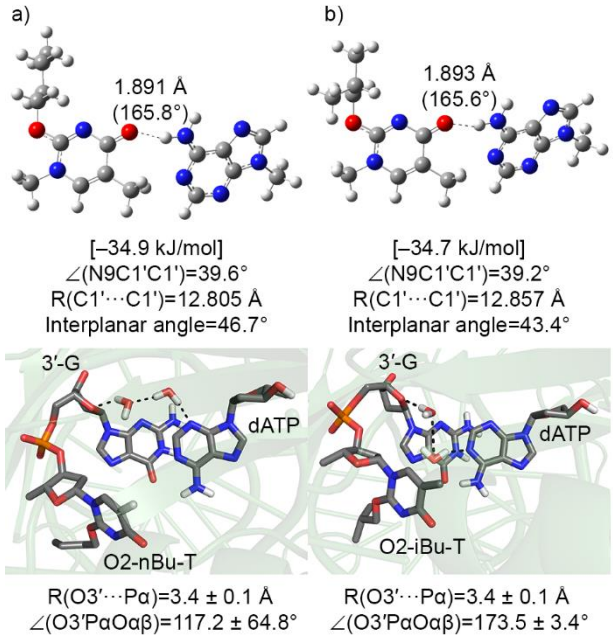


Figure 4.6: Key DFT structural features (Å and deg.) and interaction energies [kJ/mol] hydrogen bonded with A (top) and MD representative structure of the ternary complex with reaction parameters (Å and deg.) for dATP insertion (bottom) opposite a) *syn*-O2-nBu-T and b) *syn*-O2-iBu-T.

DFT calculations suggest that the *syn*-O2-iBu-T:A base pair has very similar structural parameters and interaction energy to the *syn*-O2-nBu-T:A pair, as the alkyl chain does not interfere with the hydrogen-bonding face in this orientation (Figure 4.6b, top). In the corresponding pol η ternary complex, dATP is stabilized by active site residues (rmsd = 0.8 ± 0.2 Å, Figure C.5f), and the reaction parameters (3.4 Å and 173.5°, Figure 4.6b, bottom) and Mg(II) coordination (87% < 2.5 Å, Table C.11) is maintained. In this binding orientation, Q38 hydrogen bonds with O3′ of the lesion for most of the simulation (82% occupancy, Figure C.7f). The restricted (rmsd = 0.1 ± 0.0 Å, Figure C.6f) alkyl chain of the *syn* conformer points toward the major groove side of the growing DNA strand, which prevents interference with active site residues. However, despite favourable dATP positioning, the lesion is located in the open pocket of pol η and void of direct interactions with the incoming base. As a result, a water chain forms between dATP and 3′-G with respect to *syn*-O2-iBu-T (Figure 4.6b, bottom). This interaction has been previously proposed for other O2-alkyl-T lesions,[7] including O2-nBu-T as discussed above, as well as the O-alkyl-T lesion studied in Chapters 2 and 3, and previous work on O2-POB-T,[7] indicating that the replication of *syn*-O2-iBu-T likely proceeds through a similar mechanism to the replication of an abasic site by pol η.[9]

Overall, regardless of the shape of the alkyl chain, dATP insertion is proposed to occur opposite O2-nBu-T and O2-iBu-T through a water chain to the 3′-G of the template strand. In fact, this proposed mechanism is becoming a common theme for O-alkyl-T lesions, regardless of the alkyl chain position, shape and size. Interestingly, dATP insertion opposite *syn*-O2-nBu-T is associated with fluctuations in the reaction angle, which contrasts the near-ideal values for *syn*-O2-iBu-T. Therefore, dATP alignment may be an

additional factor leading to the observed greater preference for base mispairing for the straight-chained lesion.[1]

4.4 Conclusion

Overall, although O2-nBu-T and O2-iBu-T have similar alkyl chains, previous experimental data highlights that the shape and flexibility of the alkyl chain impacts the lesion mutagenic outcomes.[1-2] Specifically, there is a greater percentage of dTTP insertion opposite O2-nBu-T (13%) compared to O2-iBu-T (4%) in human cells.[1] In the present chapter, it was revealed that a stable base pair between an incoming dTTP and the lesion cannot form when the adduct adopts the *anti* glycosidic conformation. This contrasts the reported experimental mutagenic patterns.[1] However, ternary complexes corresponding to dTTP insertion opposite *syn*-O2-nBu-T maintain lesion site hydrogen bonding and are overall conducive for the nucleotidyl transfer reaction. Furthermore, the flexible n-butyl alkyl chain disrupts lesion stacking interactions with 5′-G, leaving room for R61 to hydrogen bond and stabilize the *syn*-O2-nBu-T:dTTP base mispair. Nevertheless, the reaction angle and base pair opening angle only periodically adopt catalytically conducive values. Together, this data correlates with the overall low mutagenicity of this lesion (13% dTTP insertion).[1] Unlike the straight-chained lesion, the restricted movement of the *syn*-O2-iBu-T alkyl chain does not lead to the same active site configuration as discussed for *syn*-O2-nBu-T replication. Specifically, R61 no longer stabilizes the lesion base mispair. The differential alkyl chain conformational heterogeneity rationalizes the reduced mutagenicity of O2-iBu-T (4% dTTP insertion) compared to its straight-chained counterpart.[1] Regardless, both lesions are primarily non-mutagenic[1-2] and dATP insertion

is proposed to be promoted via the formation of a water chain opposite the *syn* lesion glycosidic orientation, similar to other O-alkyl-T adducts studied in this thesis and in previous work.[7] Due to similarities with a crystal structure corresponding to dATP insertion opposite an abasic site by pol η,[9] this work emphasizes a recurring theme in the literature for the replication of O2-alkyl-T lesions by pol η, namely a mechanism similar to abasic site replication.[9] Taken together, these results highlight the impact of alkyl chain shape on the mutagenic outcomes of O2-alkyl-T lesions,[1] afford a greater understanding of the replication of O2-alkyl-T lesions by pol η, and emphasize that computational chemistry is a key tool for understanding the role TLS plays in replicating damaged DNA.

4.5 References

(1) Wu, J.; Li, L.; You, C.; Wang, Y.; Wang, P.; Wang, Y. Cytotoxic and mutagenic properties of minor-groove O(2)-alkylthymidine lesions in human cells. *J. Biol. Chem.* **2018,** *293,* 8638-8644.

(2) Williams, N. L.; Wang, P.; Wang, Y. Replicative bypass of O2-alkylthymidine lesions *in vitro*. *Chem. Res. Toxicol.* **2016,** *29,* 1755-1761.

(3) Zhai, Q.; Wang, P.; Cai, Q.; Wang, Y. Syntheses and characterizations of the *in vivo* replicative bypass and mutagenic properties of the minor-groove O2-alkylthymidine lesions. *Nucleic Acids Res.* **2014,** *42,* 10529-10537.

(4) Wu, J.; Li, L.; Wang, P.; You, C.; Williams, N. L.; Wang, Y. Translesion synthesis of O4-alkylthymidine lesions in human cells. *Nucleic Acids Res.* **2016,** *44,* 9256-9265.

(5) Williams, N. L.; Wang, P.; Wu, J.; Wang, Y. *In vitro* lesion bypass studies of O4-alkylthymidines with human DNA polymerase η. *Chem. Res. Toxicol.* **2016,** *29,* 669-675.

(6) Wang, P.; Amato, N. J.; Zhai, Q.; Wang, Y. Cytotoxic and mutagenic properties of O4-alkylthymidine lesions in *Escherichia coli* cells. *Nucleic Acids Res.* **2015,** *43,* 10795-10803.

(7) Wilson, K. A.; Holland, C. D.; Wetmore, S. D. Uncovering a unique approach for damaged DNA replication: A computational investigation of a mutagenic tobacco-derived thymine lesion. *Nucleic Acids Res.* **2019,** *47,* 1871-1879.

(8) Du, H.; Leng, J.; Wang, P.; Li, L.; Wang, Y. Impact of tobacco-specific nitrosamine-derived DNA adducts on the efficiency and fidelity of DNA replication in human cells. *J. Biol. Chem.* **2018,** *293,* 11100-11108.

(9) Patra, A.; Zhang, Q.; Lei, L.; Su, Y.; Egli, M.; Guengerich, F. P. Structural and kinetic analysis of nucleoside triphosphate incorporation opposite an abasic site by human translesion DNA polymerase η. *J. Biol. Chem.* **2015,** *290,* 8028-8038.

(10) Quinet, A.; Lerner, L. K.; Martins, D. J.; Menck, C. F. M. Filling gaps in translesion DNA synthesis in human cells. *Mutat. Res. Genet. Toxicol. Environ. Mutagen.* **2018,** *836,* 127-142.

(11) Vaisman, A.; Woodgate, R. Translesion DNA polymerases in eukaryotes: What makes them tick? *Crit. Rev. Biochem. Mol. Biol.* **2017,** *52,* 274-303.

(12) *Hyperchem*^{tm}, Professional 8.0; Hypercube, Inc.: Gainesville, Florida 32601, USA.

(13) Frisch, Æ. H. H. P.; Dennington, R. D. II; Keith, T.; A.; Millam, J. N. A. B.; Holder, A. J.; Hiscocks, J. *GaussView 5*; Wallingford, CT, **2009**.

(14) Macrae, C. F.; Bruno, I. J.; Chisholm, J. A.; Edgington, P. R.; McCabe, P.; Pidcock, E.; Rodriguez-Monge, L.; Taylor, R.; van de Streek, J.; Wood, P. A. Mercury CSD 2.0 - new features for the visualization and investigation of crystal structures. *J. Appl. Crystallogr.* **2008,** *41,* 466-470.

(15) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Jr. J. A. M.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09,* Revisions D.01 and E.01; Gaussian, Inc.: Wallingford CT, **2009**.

(16) O'Flaherty, D. K.; Patra, A.; Su, Y.; Guengerich, F. P.; Egli, M.; Wilds, C. J. Lesion orientation of O4-alkylthymidine influences replication by human DNA polymerase η. *Chem. Sci.* **2016,** *7,* 4896-4904.

(17) Batra, V. K.; Beard, W. A.; Shock, D. D.; Krahn, J. M.; Pedersen, L. C.; Wilson, S. H. Magnesium-induced assembly of a complete DNA polymerase catalytic complex. *Structure* **2006,** *14,* 757-766.

(18) Mendieta, J.; Cases-Gonzalez, C. E.; Matamoros, T.; Ramirez, G.; Menendez-Arias, L. A Mg2+-induced conformational switch rendering a competent DNA polymerase catalytic complex. *Proteins: Struct. Funct. Bioinf.* **2008,** *71,* 565-574.

(19) Dupradeau, F.-Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P. The R.E.D. Tools: Advances in RESP and ESP charge derivation and force field library building. *Phys. Chem. Chem. Phys.* **2010,** *12,* 7821–7839.

(20) Wang, J.; Wang, W.; Kollman, P. A. Antechamber: An accessory software package for molecular mechanical calculations. *J. Am. Chem. Soc.* **2001,** *222,* U403.

(21) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004,** *25,* 1157-1174.

(22) Case, D.; Babin, V.; Berryman, J.; M. Betz, R.; Cai, Q.; S. Cerutti, D.; Cheatham, T.; Darden, T.; Duke, R.; Gohlke, H.; Götz, A.; Gusarov, S.; Homeyer, N.; Janowski, P.; Kaus, J.; Kolossváry, I.; Kovalenko, A.; Lee, T.-S.; Legrand, S.; A. Kollman, P. *Amber 2014*. 2014.

(23) Galindo-Murillo, R.; Robertson, J. C.; Zgarbova, M.; Sponer, J.; Otyepka, M.; Jurecka, P.; Cheatham, T. E. Assessing the current state of amber force field modifications for DNA. *J. Chem. Theory Comput.* **2016,** *12,* 4114-4127.

(24) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **2015,** *11,* 3696-3713.

(25) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **2013,** *9,* 3084-3095.

(26) Nakamura, T.; Yamagata, Y.; Yang, W. Watching DNA polymerase η make a phosphodiester bond. *Nature* **2013,** *53,* 254-257.

(27) Suarez, S. C.; Beardslee, R. A.; Toffton, S. M.; McCulloch, S. D. Biochemical analysis of active site mutations of human polymerase η. *Mutat. Res. Fundam. Mol. Mech. Mutagen.* **2013,** *745-746,* 46-54.

(28) Ucisik, M. N.; Hammes-Schiffer, S. Effects of active site mutations on specificity of nucleobase binding in human DNA polymerase η. *J. Phys. Chem. B* **2017,** *121,* 3667-3675.

(29) Su, Y.; Patra, A.; Harp, J. M.; Egli, M.; Guengerich, F. P. Roles of residues arg-61 and gln-38 of human DNA polymerase η in bypass of deoxyguanosine and 7,8-dihydro-8-oxo-2'-deoxyguanosine. *J. Biol. Chem.* **2015,** *290,* 15921-15933.

(30) Genna, V.; Gaspari, R.; Dal, P. M.; De, V. M. Cooperative motion of a key positively charged residue and metal ions for DNA replication catalyzed by human DNA polymerase-η. *Nucleic Acids Res.* **2016,** *44,* 2827-36.

Chapter 5: Conclusions[a]

5.1 Thesis Overview

The overall focus of this thesis was to provide structural insight into the replication outcomes of O-alkylthymine lesions associated with TLS polymerase η. Until this thesis, very little was understood about the effects of alkyl chain position, shape and size on the mutagenic patterns of O-alkyl-T lesions. Differential mutagenic patterns were recently reported for O-methylated thymine lesions, which indicate that O4-Me-T frequently leads to G misinsertions (60% in human cells), whereas O2-Me-T is primarily non-mutagenic (98% in human cells).[1-6] The reasons for these differences are unclear since both lesions similarly alter the Watson-Crick binding face of T. Although previous computational work focused on the isolated base-pair models of these lesions,[7-9] these models could not rationalize the non-mutagenic patterns of O-Me-T adducts, indicating that more complete models are required to understand the replication outcomes for the lesions of interest. To rationalize these replication outcomes at a molecular level, Chapter 2 uses DFT calculations and MD simulations to probe the lesion base-pairing properties, as well as lesion accommodation by human pol η and post-extension DNA duplexes. O4-Me-T forms two strong hydrogen bonds with an opposing G in the active site of pol η, which rationalizes the observed lesion mutagenicity.[6] Nevertheless, dATP insertion opposite O4-Me-T can proceed through water-mediated hydrogen bonding, which is similar to the pathway previously proposed for pol η bypass of abasic sites[10] and other T alkylation lesions.[11] In contrast, the position of O2-Me-T in the pol η active site is dynamic due to the presence of the aberrant methyl group on the minor groove side of DNA. In fact, the experimental

[a]*The Journal of American Chemical Society* reference style is used in this chapter.

replication outcomes[1] can only be rationalized when the syn glycosidic orientation of O2-Me-T is considered, which stabilizes the pre-insertion complex by placing the damage in the polymerase open pocket on the major groove side of DNA. Although dATP insertion can occur opposite *syn*-O2-Me-T through a water-mediated pathway similar to O4-Me-T replication, rotation about the glycosidic bond precludes a stable pol η ternary complex corresponding to dGTP insertion, which correlates with the reported non-mutagenic bypass of O2-Me-T.[1] Taken together, Chapter 2 provided key insights into how pol η accommodates two common DNA methylation adducts and highlights the effect of the methyl group position on the replication outcomes.

Chapter 3 examined the effect of size and position of the alkyl chain by considering O4-POB-T compared to both O4-Me-T (Chapter 2) and previously studied O2-POB-T.[11] This work uncovered that the bulky POB moiety disrupts the base-pair interaction between the lesion and the incoming dGTP in the pol η ternary complex. The recurrent non-optimal structural and reaction parameters associated with the addition of a larger group supports the low mutagenicity of O4-POB-T (35% dGTP insertion).[12] This contrasts the pol η ternary complex corresponding to dGTP insertion opposite O4-Me-T discussed in Chapter 2 in which favourable structural and reaction parameters were maintained throughout the simulation, providing insight into the greater mutagenicity of O4-Me-T. Thus, Chapter 3 underscores the importance of alkyl chain size on the replication outcomes. Regardless of the size, dATP insertion is found to be stabilized through similar interactions proposed in the crystal structure corresponding to the replication of an abasic site by pol η.[10] In Chapter 3, O4-POB-T is proposed to forms dGTP mispairs by adopting an *anti* conformation about the glycosidic bond in the pol η active site, making the WC face available to interact with

the incoming dNTP. This contrasts previous work on O2-POB-T (forms dTTP mispairs),[12-13] which proposed the O2 lesion adopts a *syn* orientation and interacts with the incoming dNTP through the HG face.[11] The different glycosidic bond orientations of these lesions orients the bulky POB moiety into the pol η open pocket regardless of the T damage position. This ensures the bulky moiety is void of most steric clashes with the surrounding DNA and active site residues in the ternary complex. The different hydrogen-bonding faces being available to the incoming dNTP upon replication rationalize the differential mutagenicity of O2-POB-T and O4-POB-T.[12] Thus, Chapter 3 builds upon the conclusions presented in Chapter 2 and further emphasizes the effect of alkyl chain position on O-alkyl-T lesions mutagenicity.

Chapter 4 investigated the impact of subtle differences in the alkyl chain shape by using a prototypical example of straight-chained (i.e. O2-nBu-T) and branched (i.e, O2-iBu-T) O2-alkyl-T lesions. Experimental studies illustrated that straight-chained and branched alkyl lesions exhibit different mutational frequencies for O2-alkyl-T lesions.[1, 3-4, 6] For example, in human cells, O2-alkyl-T lesions result in small frequencies of dTTP insertion,[1] with straight-chained lesions leading to higher frequencies of mispairs than their branched counterparts. In contrast to the experimental mutagenicities,[1, 4] MD simulations reveal that dTTP insertion opposite both O2-alkyl-T lesions in the *anti* glycosidic orientation do not result in suitable base pairs. When the lesion adopts the *syn* orientation, the highly flexible n-butyl group of O2-nBu-T disrupts stacking interactions with the 5′-base, allowing R61 to reposition and stabilize the lesion base mispair. Together, the low mutagenicity of this lesion (13% dTTP insertion)[1] is supported by the fact that R61 has been proposed to be catalytically important for positioning the incoming dNTP[14-16] and

may not always be available to stabilize this base pair. Furthermore, the loss of the 5′ stacking interaction with respect to the template base may impede replication. More importantly, unlike O2-nBu-T, the reduced flexibility of the O2-iBu-T alkyl chain does not similarly disrupt the active site. As a result, there is no space for R61 to move to stabilize the corresponding mispair, which directly correlates with the reduced mutagenicity of this lesion in human cells (4% dTTP insertion) compared to O2-nBu-T.[1] Regardless, dATP insertion is proposed to be promoted via water-mediated interactions when both lesions are in the *syn* orientation, which supports the observed primarily non-mutagenic patterns.[1] Taken together, Chapter 4 builds on Chapters 2 and 3 by highlighting the effect of the alkyl chain shape and flexibility on the replication outcomes. Finally, Chapter 4 supports the proposal of a required rotation about the glycosidic bond for O2-alkyl-T lesions to permit replication by pol η, which was initially proposed for O2-POB-T.[11]

Overall, although pol η has been shown to have a flexible active site to accommodate a wide spectrum of DNA lesions, the replication outcomes of O-alkyl-T lesions depend on the size, shape and flexibility of the damage, as well as the damage position. Thus, this thesis broadens our understanding of TLS pol η and highlights a practical computational methodology that can be used to study DNA lesions of varying shapes and sizes in order to successfully rationalize the experimental mutagenic frequencies.

5.2 Future Work: Other DNA Lesions

Although this thesis contributes to the overall understanding of the replication of T alkylation damage by pol η, questions still remain about the role of TLS in the replication

of other DNA lesions. Since large and flexible lesions may be hard to study using current experimental methods, this thesis emphasizes that computational chemistry can rationalize the experimental mutagenic outcomes and may be used in a predictive way to provide structural insight into the replication of other forms of DNA damage. For example, these methods could extend our understanding of flexible lesions formed upon exposure to NNK. In particular, the major groove N6-POB-A and N6-4-(3-pyridyl)-4-hydroxybut-1-yl-A (N6-PHB-A, Figure 5.1) lesions were recently detected in rats, but their mutagenic patterns have not yet been revealed.[17] Furthermore, the addition of a methyl group or POB/PHB moiety to the phosphate backbone occurs upon exposure to NNK (Figure 5.2).[18-20] Unfortunately, very little is known about the impact of DNA phosphate lesions on the duplex environment, as well as how this form of damage may interfere with replicative or TLS DNA polymerases. By using a combination of DFT calculations and MD simulations on these PHB or POB lesions, insight may be obtained into the hydrogen-bonding properties, as well as the structure and dynamics of these lesions, as well as the potential role of pol η in their replication. Together, this could provide the first atomic level insight about the potential mutagenicity of these newly identified lesions.

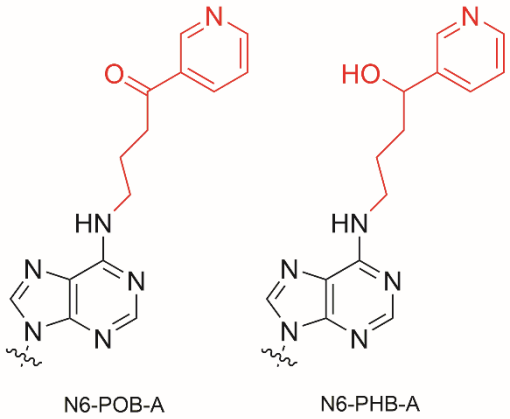


Figure 5.1: Chemical structures of recently detected tobacco-derived adenine lesions.[17]

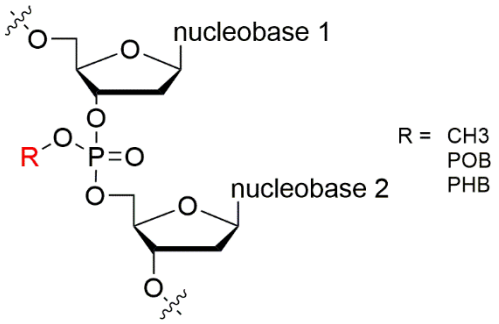


Figure 5.2: Chemical structure of DNA backbone lesions (R = alkyl group).[18-20]

Aside from the methyl and POB adducts studied in this thesis, there are still questions that remain about other O-alkyl-T lesions that could be addressed in future work to understand the role of pol η in replication. For example, the replication outcomes of synthetic O4-alkyl-T adducts studied in the experimental work highlighted in Chapter 4 indicates that branched adducts are more mutagenic than straight-chained alkyl lesions in human cells (Figure 5.3),[6] which contrasts the O2-alkyl-T lesions[1, 4] studied in this thesis. Since only G mispairs are detected in this environment,[6] it is unclear how the shape of the alkyl chain affects the base-pair interactions. Specifically, compared to O4-Me-T (Chapter 2), the frequency of insertion is reduced with the increase in alkyl chain size, whereas compared to O4-POB-T (Chapter 3), there is an increase in the mutagenicity. Using a similar approach as applied in this thesis, future work on these lesions can be compared to the O4-alkyl-T adducts studied in Chapters 2 and 3, as well as the O2-alkyl-T adducts studied in Chapter 4. Overall, this would contribute to the understanding of T alkylation adducts and the role of pol η in their replication.

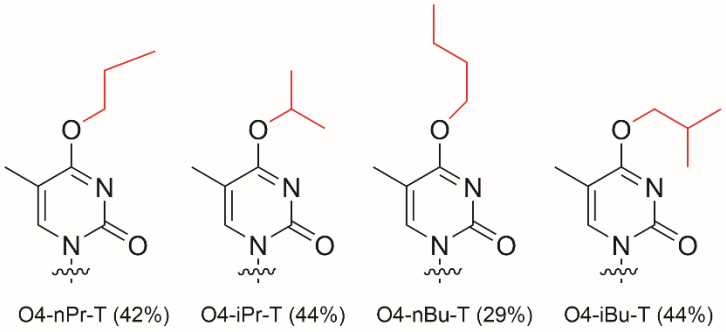


Figure 5.3: Chemical structures of synthetic O4-alkyl-T lesions and the frequency of dGTP insertion observed in human cells (%).[6]

Throughout this thesis, dATP insertion is proposed to occur via a similar mechanism to abasic site replication by pol η, where dATP is stabilized in the pol η active site by the formation of a water chain to the 3′-base with respect to the template base.[10] Despite this comparison, subtle differences in water chains were noted between the static crystal structure (Figure 5.1a) and the dynamic simulation data presented in this thesis. Specifically, in the crystal structure, a water chain forms between N3 of the incoming dATP and O4' of the 3'-base with respect to the lesion. Notably, MD simulations on some O-alkyl-T lesions indicate that the interaction instead stems from N1 or N6 of dATP. To better understand the impact of the water chain location, computational work should be conducted. Furthermore, previous crystallographic and kinetic data of dNTP insertion opposite an abasic site suggested that pol η follows a "purine rule" (i.e. insertion of either dATP or dGTP, Figure 5.4)[10] rather than simply the "A rule" reported for many other polymerases.[21] Overall, by considering the structure and dynamics of the corresponding DNA–polymerase complexes, a better understanding of how pol η replicates abasic sites can be obtained.

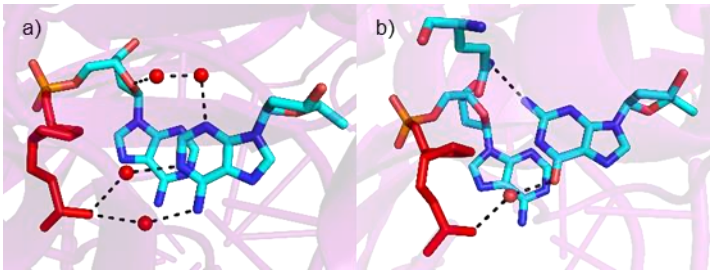


Figure 5.4: X-ray crystal structure corresponding to pol η insertion of a) dATP (PDB ID: 4RNM) or b) dGTP (PDB ID: 4RNN) opposite an abasic site (red).[10]

Finally, future work could consider other DNA lesions and the role of TLS in their replication. For example, the mutagenic patterns and bypass efficiencies of the C1' and C3'-epimeric lesions involving all DNA bases were recently revealed.[22] These lesions are derived from the improper repair of a hydrogen atom abstraction from either C1' or C3' of deoxyribose. The replication and bypass outcomes are dependent on the location of the initial abstraction and the corresponding nucleobase being replicated. Interestingly, the products from abstraction of a hydrogen at C1' (α-dN lesions, Figure 5.5) stall standard replicative polymerases, and therefore their replication occurs via the TLS pathway, which results in mutagenic bypass of all lesions except α-dA.[22] Furthermore, although pol η is primarily responsible for the mutagenicity of α-dG, bypass is most affected by knockout of pol κ. In contrast, C3'-epimeric lesions (x-dN, Figure 5.5) are non-mutagenic and do not stall standard replication polymerases or induce mutations. This is suggested to occur because these lesions do not interfere with the WC face, although structural insight has yet to be provided to support this proposal. Overall, computational modelling of these lesions could provide the insight required to understand how these changes to deoxyribose affect the replication outcomes. Furthermore, since different TLS polymerases play a role in the mutagenicity and bypass of some of these lesions, future computational work may clarify how these lesions can be accommodated by other TLS polymerases (pol κ) or standard

replicative polymerases and, as a result, provide a better understanding of the broader TLS process.

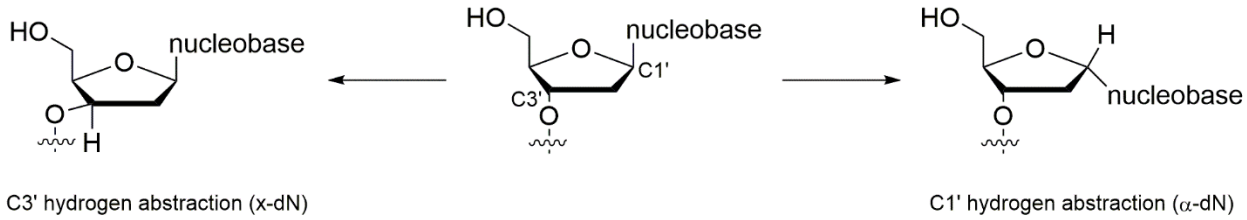


Figure 5.5: Chemical structures of canonical deoxyribose (middle); and deoxyribose after improper repair of C3′ hydrogen abstraction (left) and C1′ hydrogen abstraction (right).

5.3 Final Remarks

This thesis provides key structural insight into the experimental mutagenic outcomes of O-alkylthymine lesions.[1-6] Specifically, systematic modelling that involved a combination of DFT calculations and MD simulations was used to examine the effects of the position, shape and size of the O-alkyl-T adducts on the replication outcomes from human TLS polymerase η. Furthermore, this thesis highlights the importance of understanding key DNA–polymerase interactions in the replication of these DNA adducts. Overall, this work extensively studied T alkylation damage, and addressed some of the key unanswered questions that arose from recent experimental mutagenic studies on O-alkyl-T lesions.

5.4 References

(1) Wu, J.; Li, L.; You, C.; Wang, Y.; Wang, P.; Wang, Y. Cytotoxic and mutagenic properties of minor-groove O(2)-alkylthymidine lesions in human cells. *J. Biol. Chem.* **2018,** *293*, 8638-8644.

(2) Wang, P.; Amato, N. J.; Zhai, Q.; Wang, Y. Cytotoxic and mutagenic properties of O4-alkylthymidine lesions in *Escherichia coli* cells. *Nucleic Acids Res.* **2015,** *43*, 10795-10803.

(3) Williams, N. L.; Wang, P.; Wu, J.; Wang, Y. *In vitro* lesion bypass studies of O4-alkylthymidines with human DNA polymerase η. *Chem. Res. Toxicol.* **2016,** *29*, 669-675.

(4) Williams, N. L.; Wang, P.; Wang, Y. Replicative bypass of O2-alkylthymidine lesions *in vitro*. *Chem. Res. Toxicol.* **2016,** *29*, 1755-1761.

(5) Zhai, Q.; Wang, P.; Cai, Q.; Wang, Y. Syntheses and characterizations of the *in vivo* replicative bypass and mutagenic properties of the minor-groove O2-alkylthymidine lesions. *Nucleic Acids Res.* **2014,** *42*, 10529-10537.

(6) Wu, J.; Li, L.; Wang, P.; You, C.; Williams, N. L.; Wang, Y. Translesion synthesis of O4-alkylthymidine lesions in human cells. *Nucleic Acids Res.* **2016,** *44*, 9256-9265.

(7) Qiu, Z. M.; Wang, H. L.; Liu, Y. Z.; Hou, D. N. MP2 study on the hydrogen-bonding interaction between O4-methylthymine and DNA bases: A, C, G, and T. *J. Struc. Chem.* **2014,** *25*, 767-774.

(8) Flood, A.; Hubbard, C.; Forde, G.; Hill, G.; Gorb, L.; Leszczynski, J. Theoretical *ab initio* study of the effects of methylation on the nature of hydrogen bonding in A:T base pair. *J. Biomol. Struct. Dyn.* **2003,** *21*, 297-302.

(9) Venkateswarlu, D.; Lyngdoh, R. H. D. Structural, steric and energetic requirements for induction of base substitutional mutations by methylated guanines and thymines. *J. Chem. Soc. Perkin Trans. 2* **1995,** 839-846.

(10) Patra, A.; Zhang, Q.; Lei, L.; Su, Y.; Egli, M.; Guengerich, F. P. Structural and kinetic analysis of nucleoside triphosphate incorporation opposite an abasic site by human translesion DNA polymerase η. *J. Biol. Chem.* **2015,** *290*, 8028-8038.

(11) Wilson, K. A.; Holland, C. D.; Wetmore, S. D. Uncovering a unique approach for damaged DNA replication: A computational investigation of a mutagenic tobacco-derived thymine lesion. *Nucleic Acids Res.* **2019,** *47*, 1871-1879.

(12) Du, H.; Leng, J.; Wang, P.; Li, L.; Wang, Y. Impact of tobacco-specific nitrosamine-derived DNA adducts on the efficiency and fidelity of DNA replication in human cells. *J. Biol. Chem.* **2018,** *293*, 11100-11108.

(13) Gowda, A. S. P.; Spratt, T. E. DNA polymerases η and ζ combine to bypass O2-[4-(3-pyridyl)-4-oxobutyl]thymine, a DNA adduct formed from tobacco carcinogens. *Chem. Res. Toxicol.* **2016,** *29*, 303-316.

(14) Genna, V.; Gaspari, R.; Dal, P. M.; De, V. M. Cooperative motion of a key positively charged residue and metal ions for DNA replication catalyzed by human DNA polymerase-η. *Nucleic Acids Res.* **2016,** *44*, 2827-36.

(15) Ucisik, M. N.; Hammes-Schiffer, S. Effects of active site mutations on specificity of nucleobase binding in human DNA polymerase η. *J. Phys. Chem. B* **2017,** *121*, 3667-3675.

(16) Su, Y.; Patra, A.; Harp, J. M.; Egli, M.; Guengerich, F. P. Roles of residues arg-61 and gln-38 of human DNA polymerase η in bypass of deoxyguanosine and 7,8-dihydro-8-oxo-2'-deoxyguanosine. *J. Biol. Chem.* **2015,** *290*, 15921-15933.

(17) Carlson, E. S.; Upadhyaya, P.; Villalta, P. W.; Ma, B.; Hecht, S. S. Analysis and identification of 2'-deoxyadenosine-derived adducts in lung and liver DNA of f-344 rats treated with the tobacco-specific carcinogen 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone and enantiomers of its metabolite 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol. *Chem. Res. Toxicol.* **2018,** *31*, 358-370.

(18) Ma, B.; Zarth, A. T.; Carlson, E. S.; Villalta, P. W.; Upadhyaya, P.; Stepanov, I.; Hecht, S. S. Methyl DNA phosphate adduct formation in rats treated chronically with 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone and enantiomers of its metabolite 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol. *Chem. Res. Toxicol.* **2018,** *31*, 48-57.

(19) Haglund, J.; Henderson, A. P.; Golding, B. T.; Toernqvist, M. Evidence for phosphate adducts in DNA from mice treated with 4-(n-methyl-n-nitrosamino)-1-(3-pyridyl)-1-butanone (NNK). *Chem. Res. Toxicol.* **2002,** *15*, 773-779.

(20) Ma, B.; Zarth, A. T.; Carlson, E. S.; Villalta, P. W.; Upadhyaya, P.; Stepanov, I.; Hecht, S. S. Identification of more than 100 structurally unique DNA-phosphate adducts formed during rat lung carcinogenesis by the tobacco-specific nitrosamine 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone. *Carcinogenesis* **2018,** *39*, 232-241.

(21) Choi, J.-Y.; Lim, S.; Kim, E.-J.; Jo, A.; Guengerich, F. P. Translesion synthesis across abasic lesions by human B-family and Y-family DNA polymerases α, δ, η, ι, κ, and rev1. *J. Mol. Biol.* **2010,** *404*, 34-44.

(22) Du, H.; Wang, P.; Li, L.; Amato, N. J.; Wang, Y. Cytotoxic and mutagenic properties of C1' and C3'-epimeric lesions of 2'-deoxyribonucleosides in human cells. *ACS Chem. Biol.* **2019,** *14*, 478-485.

Appendix A: Supplementary Information for Chapter 2:
Computational Insight into the Differential Mutagenic Patterns of O-methylthymine Lesions

Figures A.1 – A.17 and Tables A.1 – A.9

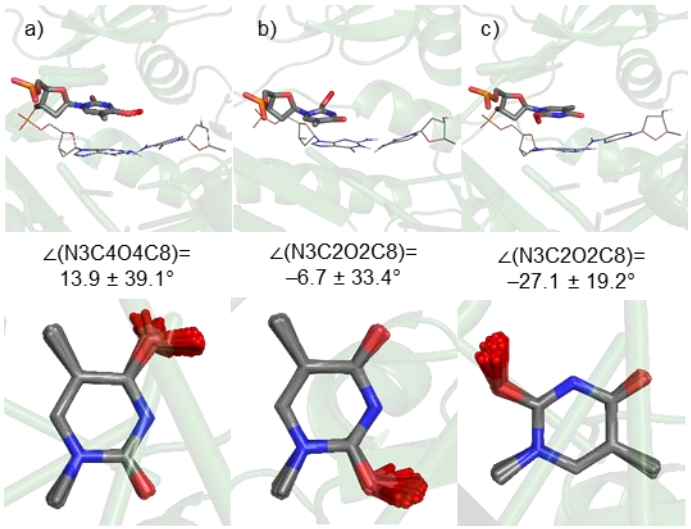


Figure A.1: MD representative structure of the lesion with in the active side of pol η (top) and MD snapshots taken every 5 ns overlaid with respect to T ring atoms, highlighting the orientation and flexibility of the aberrant methyl group in a pol η pre-insertion complex (bottom) for a) O4-Me-T, b) *anti*-O2-Me-T, and c) *syn*-O2-Me-T, as well as the ∠(N3CXOXC8) dihedral angle (deg.) over the simulation.

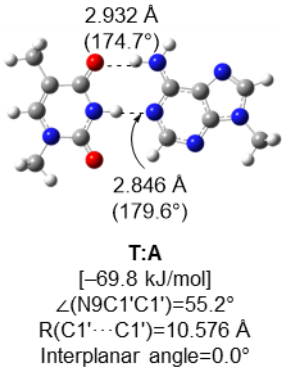


Figure A.2: Key structural features (Å and deg.) and interaction energy [kJ/mol] of the natural T:A pair.

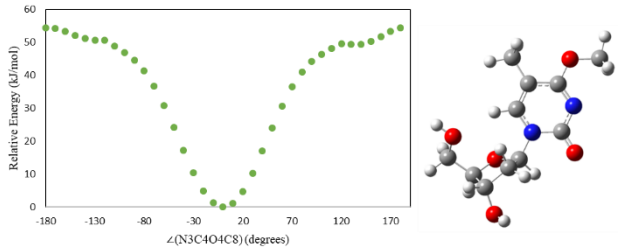


Figure A.3: B3LYP-D3(BJ)/6-311++G(2df,2p)//B3LYP/6-31G(d,p) relative energy (kJ/mol) as a function of the ∠(N3C4O4C8) dihedral angle that dictates the orientation of the O4-Me-T aberrant methyl group (left). Most stable ∠(N3C4O4C8) = 0° orientation (right).

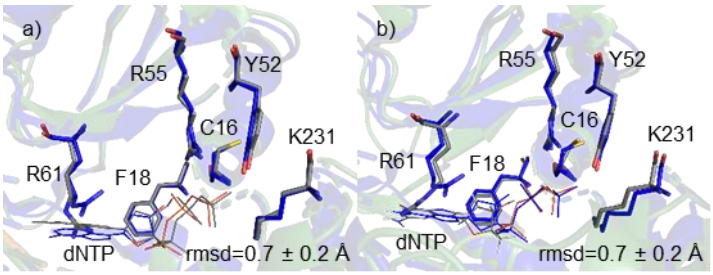


Figure A.4: Overlay of MD representative structure (grey) and X-ray crystal structure of dATP insertion opposite T (PDB ID: 4ECS, blue) for pol η insertion of (a) dATP and (b) dGTP opposite O4-Me-T, highlighting alignment of key active site residues that interact with the dNTP.

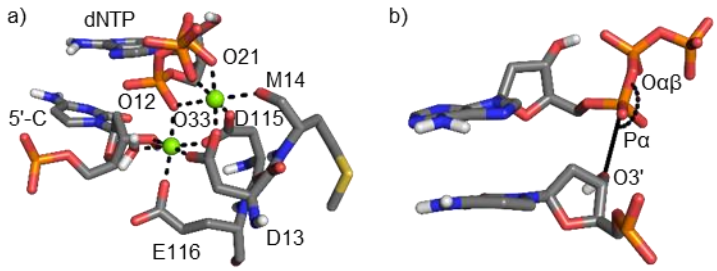


Figure A.5: a) Mg(II) coordination required to promote catalytic activity of pol η and b) definition of reaction parameters for the incoming dNTP with respect to the corresponding 5′ base.

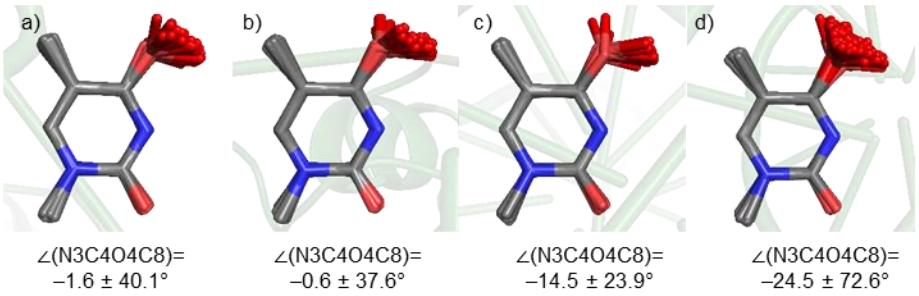


Figure A.6: MD snapshots taken every 5 ns overlaid with respect to T ring atoms, highlighting the orientation and flexibility of the aberrant methyl group in a pol η ternary complex for a) dATP or b) dGTP insertion opposite O4-Me-T, and the post-extension duplexes containing the lesion paired opposite c) A or d) G, as well as average ∠(N3C4O4C8) dihedral angle (°) over the simulation.

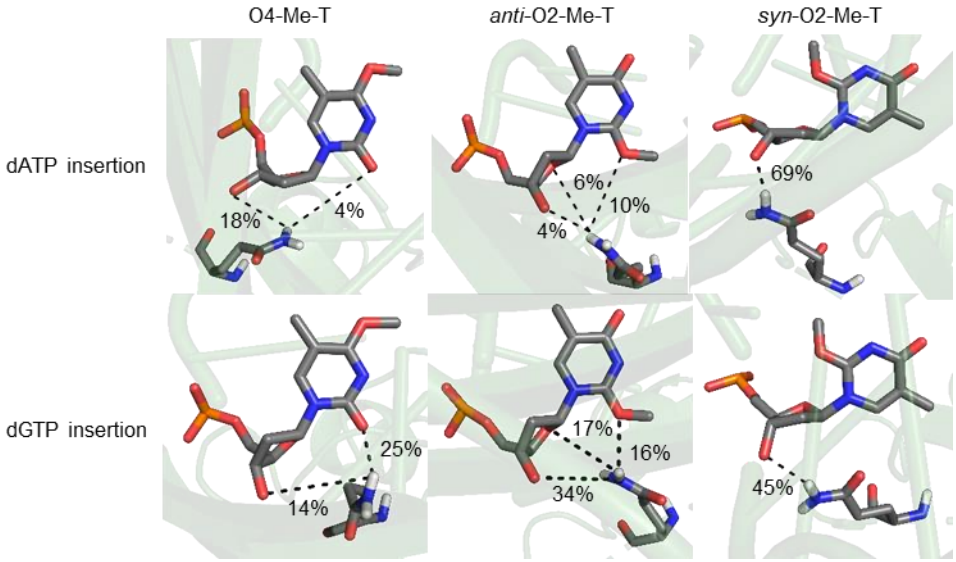


Figure A.7: MD representative structure, and interactions between Q38 and the T lesion in the pol η ternary complex corresponding to dATP (top) or dGTP (bottom) insertion opposite O4-Me-T (left), *anti*-O2-Me-T (middle) and *syn*-O2-Me-T (right).

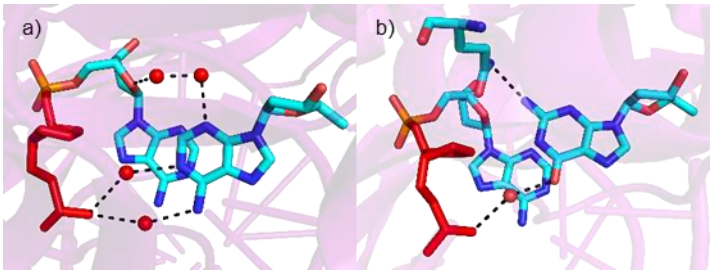


Figure A.8: X-ray crystal structure corresponding to pol η insertion of a) dATP (PDB ID: 4RNM) or b) dGTP (PDB ID: 4RNN) opposite an abasic site (red).

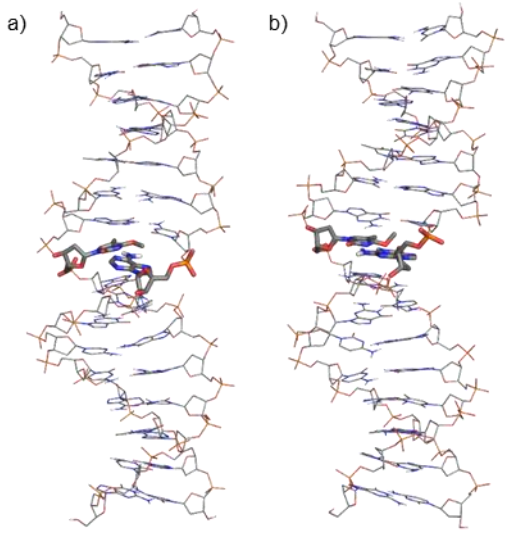


Figure A.9: MD representative structure of a 16-mer DNA duplex containing O4-Me-T opposite a) A and b) G.

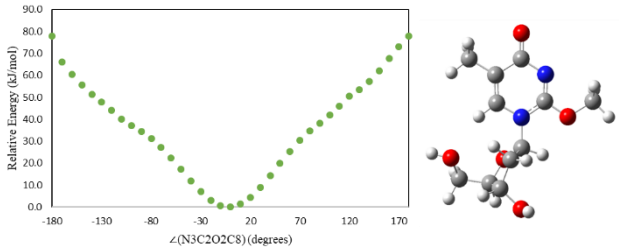


Figure A.10: B3LYP-D3(BJ)/6-311++G(2df,2p)//B3LYP/6-31G(d,p) relative energy (kJ/mol) as a function of the ∠(N3C2O2C8) dihedral angle that dictates the orientation of the O2-Me-T aberrant methyl group (left). Most stable ∠(N3C2O2C8) = 0° orientation (right).

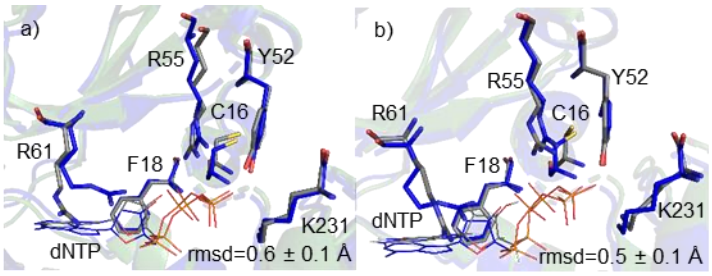


Figure A.11: Overlay of MD representative structure (grey) and X-ray crystal structure of dATP insertion opposite T (PDB ID: 4ECS, blue) for pol η insertion of (a) dATP and (b) dGTP opposite *anti*-O2-Me-T, highlighting alignment of key active site residues that interact with the dNTP.

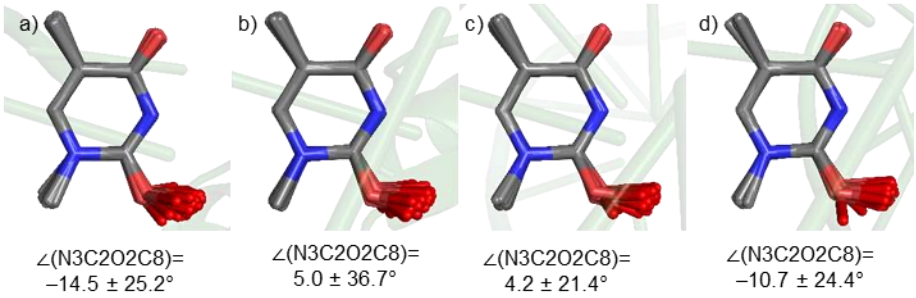


Figure A.12: MD snapshots taken every 5 ns overlaid with respect to T ring atoms, highlighting the orientation and flexibility of the aberrant methyl group in a pol η ternary complex for a) dATP or b) dGTP insertion opposite *anti*-O2-Me-T, and the post-extension duplexes containing the lesion paired opposite c) A or d) G, as well as average ∠(N3C2O2C8) dihedral angle (°) over the simulation.

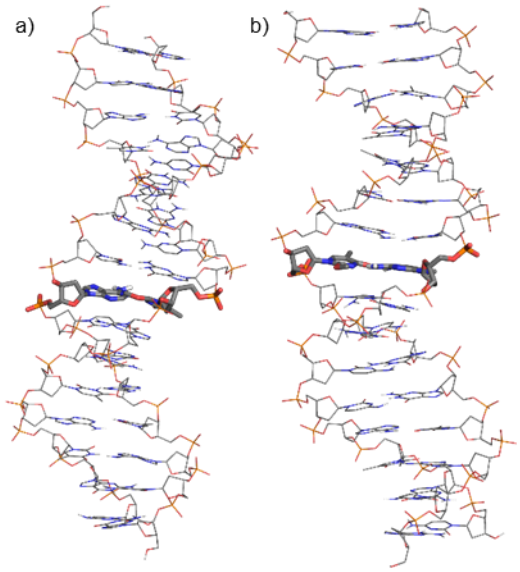


Figure A.13: MD representative structure of a 16-mer DNA duplex containing anti-O2-Me-T opposite a) A or b) G.

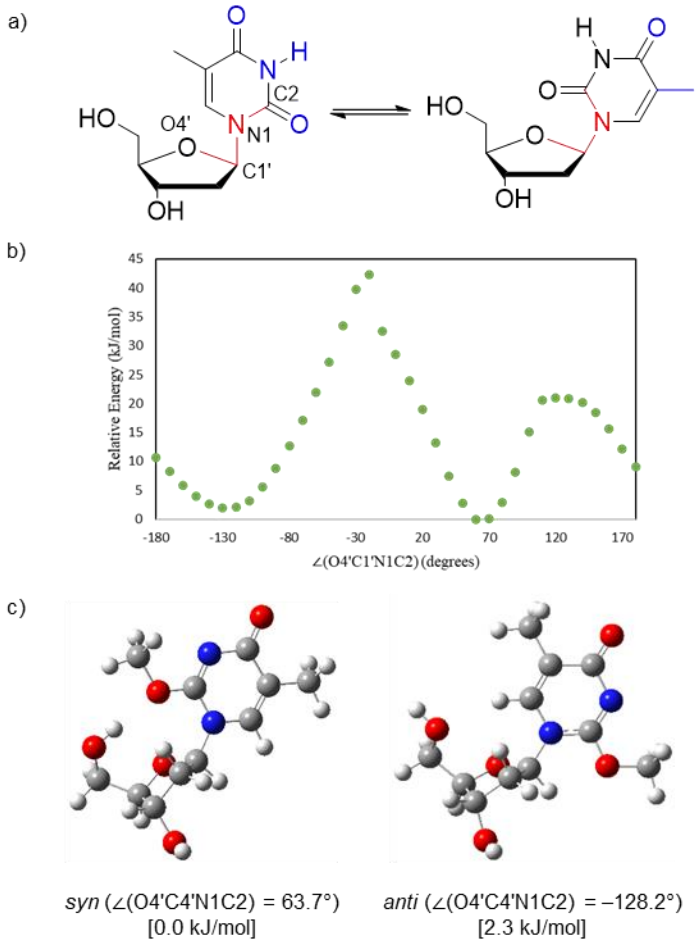


Figure A.14: a) *anti* (left) and *syn* (right) glycosidic orientations of canonical T, with the χ dihedral angle highlighted in red (∠(O4′C4′N1C2)), and Watson-Crick (left) and Hoogsteen (right) hydrogen-bonding faces highlighted in blue. b) B3LYP-D3(BJ)/6-311++G(2df,2p)//B3LYP/6-31G(d,p) relative energy (kJ/mol) as a function of the ∠(O4′C4′N1C2) dihedral angle that dictates the orientation of the glycosidic bond. c) Fully optimized structures of the most stable syn and anti orientations, with associated ∠(O4′C4′N1C2) dihedral angle and relative energies [kJ/mol].

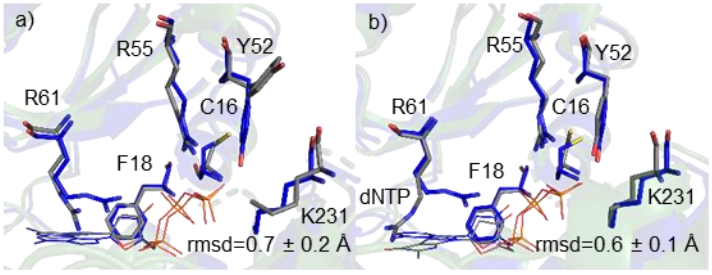


Figure A.15: Overlay of MD representative structure (grey) and X-ray crystal structure of dATP insertion opposite T (PDB ID: 4ECS, blue) for pol η insertion of (a) dATP and (b) dGTP opposite *syn*-O2-Me-T, highlighting alignment of key active site residues that interact with the dNTP.

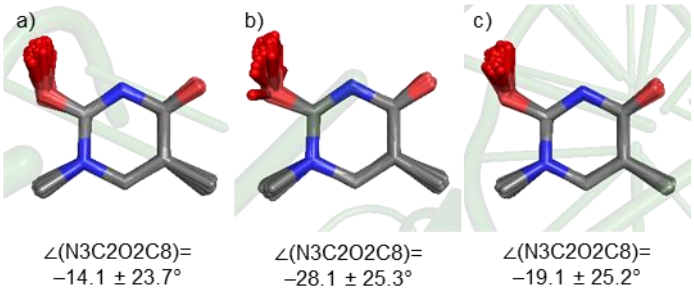


Figure A.16: MD snapshots taken every 5 ns overlaid with respect to T ring atoms, highlighting the orientation and flexibility of the aberrant methyl group in a pol η ternary complex for a) dATP or b) dGTP insertion opposite *syn*-O2-Me-T, and c) the post-extension duplex containing the lesion paired opposite A, as well as average ∠(N3C2O2C8) dihedral angle (°) over the simulation.

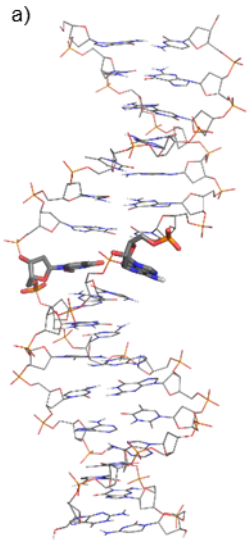


Figure A.17: MD representative structure of the post-extension DNA duplex containing *syn*-O2-Me-T opposite A.

Table A.1: O4-Me-T atom types and partial charges.

Atom Name	Atom Type	Charge
P	P	1.1624
O5'	OS	-0.4938
OP1	O2	-0.7624
OP2	O2	-0.7701
O3'	OS	-0.548
C5'	CT	0.0115
H5'	H1	0.0701
H5"	H1	0.0701
C4'	CT	0.1706
H4'	H1	0.0998
O4'	OS	-0.4091
C1'	CT	0.1755
H1'	H2	0.0879
C3'	CT	0.1586
H3'	H1	0.0675
C2'	CT	-0.0809
H2'	HC	0.0489
H2"	HC	0.0489
N1	N*	-0.138
C6	CM	-0.1066
H6	H4	0.218
C5	CM	-0.1835
C7	CT	-0.1891
H71	HC	0.0706
H72	HC	0.0706
H73	HC	0.0706
C4	CA	0.7583
O4	OS	-0.3659
N3	NC	-0.7523
C2	C	0.7962
O2	O	-0.5911
C8	CT	-0.0369
H81	H1	0.0906
H82	H2	0.0906
H83	H3	0.0906

Table A.2: O2-Me-T atom types and partial charges.

Atom Name	Atom Type	Charge
P	P	1.1623
O5'	OS	-0.4986
OP1	O2	-0.7636
OP2	O2	-0.7691
O3'	OS	-0.5409
C5'	CT	-0.0276
H5'	H1	0.0823
H5''	H1	0.0823
C4'	CT	0.1783
H4'	H1	0.1084
O4'	OS	-0.4078
C1'	CT	0.1608
H1'	H2	0.1258
C3'	CT	0.1281
H3'	H1	0.0831
C2'	CT	-0.0630
H2'	HC	0.0488
H2''	HC	0.0488
N1	N*	-0.1961
C6	CM	-0.1909
H6	H4	0.2716
C5	CM	-0.0807
C7	CT	-0.2700
H71	HC	0.0874
H72	HC	0.0874
H73	HC	0.0874
C4	C	0.7724
O4	O	-0.6005
N3	NC	-0.7361
C2	CA	0.7734
O2	OS	-0.3821
C8	CT	-0.0585
H81	H1	0.0990
H82	H1	0.0990
H83	H1	0.0990

Table A.3: Occupancy of Mg(II) coordinating residues throughout 1 μs MD simulations on the pol η ternary complex corresponding to dATP insertion opposite O4-Me-T.[a]

Interacting Residue (Coordinating Atom)	Mg(II) Distance (Å)	Standard Deviation (Å)	Occupancy (%)
dATP (O12)	2.1	0.1	100
D13 (OD1)	2.0	0.0	100
E116 (OE2)	1.9	0.0	100
D115 (OD1)	2.0	0.1	100
5'-C (O3')	2.2	0.1	100
Water (O)	2.1	0.1	100
D13 (OD2)	2.1	0.1	100
M17 (O)	2.0	0.0	100
D115 (OD2)	2.0	0.0	100
dATP (O33)	1.9	0.0	100
dATP (O22)	1.9	0.0	100
dATP (O12)	2.3	0.2	87

[a]The Mg(II) coordination occupancies are based on a heavy atom distance cut-off of 2.5 Å

Table A.4: Occupancy of Mg(II) coordinating residues throughout 1 μs MD simulations on the pol η ternary complex corresponding to dGTP insertion opposite O4-Me-T.[a]

Interacting Residue (Coordinating Atom)	Mg(II) Distance (Å)	Standard Deviation (Å)	Occupancy (%)
dGTP (O12)	2.2	0.1	98
D13 (OD1)	2.0	0.0	100
E116 (OE2)	1.9	0.0	100
D115 (OD1)	2.0	0.0	100
5'-C (O3')	2.2	0.1	100
Water (O)	2.1	0.1	100
D13 (OD2)	2.1	0.1	100
M17 (O)	2.0	0.1	100
D115 (OD2)	2.0	0.0	100
dGTP (O33)	1.9	0.0	100
dGTP (O22)	2.0	0.1	100
dGTP (O12)	2.2	0.1	99

[a]The Mg(II) coordination occupancies are based on a heavy atom distance cut-off of 2.5 Å

Table A.5: Average base pair hydrogen-bonding distances (Å) and angles (deg.) between O4-Me-T or O2-Me-T and an opposing A or G over 1 μs MD simulations on ternary complexes and post-extension duplexes.[a]

Hydrogen Bond	Distance	Angle
O4-Me-T:dGTP (ternary complex)		
O2(O4-Me-T)···(N2–H)G	2.9 Å	151.3°
N3(O4-Me-T)···(N1–H)G	3.3 Å	162.9°
O2(O4-Me-T)···(N1–H)G	3.1 Å	142.5°
O4-Me-T:A (post-extension)		
N3(O4-Me-T)···(N6–H)A	3.1 Å	140.1°
O2(O4-Me-T)···(N6–H)A	3.1 Å	155.1°
O4-Me-T:G (post-extension)		
O2(O4-Me-T)···(N2–H)G	2.9 Å	153.8°
N3(O4-Me-T)···(N1–H)G	3.2 Å	155.6°
O2-Me-T:dGTP (ternary complex)		
N3(O2-Me-T)···(N2–H)G	3.0 Å	145.6°
O4(O2-Me-T)···(N1–H)G	3.0 Å	159.0°
O4(O4-Me-T)···(N2–H)G	3.2 Å	141.7°
O2-Me-T:G (post-extension)		
N3(O2-Me-T)···(N2–H)G	3.0 Å	158.8°
O4(O2-Me-T)···(N1–H)G	2.9 Å	158.7°

[a]The hydrogen-bonding occupancies are based on a heavy atom distance and angle cut-offs of 3.4 Å and 120°

Table A.6: Occupancy of Mg(II) coordinating residues throughout 1 μs MD simulations on the pol η ternary complex corresponding to dATP insertion opposite O2-Me-T.[a]

Interacting Residue (Coordinating Atom)	Mg(II) Distance (Å)	Standard Deviation (Å)	Occupancy (%)
dATP (O12)	2.1	0.1	100
D13 (OD1)	2.0	0.0	100
E116 (OE2)	1.9	0.0	100
D115 (OD1)	2.0	0.1	100
5'-C (O3')	2.2	0.1	100
Water (O)	2.1	0.1	100
D13 (OD2)	2.1	0.1	100
M17 (O)	2.0	0.0	100
D115 (OD2)	2.0	0.0	100
dATP (O33)	1.9	0.0	100
dATP (O22)	1.9	0.0	100
dATP (O12)	2.3	0.2	89

[a]The Mg(II) coordination occupancies are based on a heavy atom distance cut-off of 2.5 Å

Table A.7: Occupancy of Mg(II) coordinating residues throughout 1 μs MD simulations on the pol η ternary complex corresponding to dGTP insertion opposite O2-Me-T.[a]

Interacting Residue (Coordinating Atom)	Mg(II) Distance (Å)	Standard Deviation (Å)	Occupancy (%)
dGTP (O12)	2.2	0.1	98
D13 (OD1)	2.0	0.0	100
E116 (OE2)	1.9	0.0	100
D115 (OD1)	2.0	0.0	100
5'-C (O3')	2.2	0.1	100
Water (O)	2.1	0.1	100
D13 (OD2)	2.1	0.1	100
M17 (O)	2.0	0.1	100
D115 (OD2)	2.0	0.1	100
dGTP (O33)	1.9	0.0	100
dGTP (O22)	2.0	0.1	100
dGTP (O12)	2.2	0.1	100

[a]The Mg(II) coordination occupancies are based on a heavy atom distance cut-off of 2.5 Å

Table A.8: Occupancy of Mg(II) coordinating residues throughout 1 μs MD simulations on the pol η ternary complex corresponding to dGTP insertion opposite *syn*-O2-Me-T.[a]

Interacting Residue (Coordinating Atom)	Mg(II) Distance (Å)	Standard Deviation (Å)	Occupancy (%)
dGTP (O12)	2.2	0.1	99
D13 (OD1)	2.0	0.0	100
E116 (OE2)	1.9	0.0	100
D115 (OD1)	2.0	0.0	100
5'-C (O3')	2.2	0.1	100
Water (O)	2.1	0.1	100
D13 (OD2)	2.0	0.1	100
M17 (O)	2.0	0.0	100
D115 (OD2)	2.0	0.0	100
dGTP (O33)	1.9	0.0	100
dGTP (O22)	2.0	0.1	100
dGTP (O12)	2.2	0.1	99

[a]The Mg(II) coordination occupancies are based on a heavy atom distance cut-off of 2.5 Å

Table A.9: Occupancy of Mg(II) coordinating residues throughout 1 μs MD simulations on the pol η ternary complex corresponding to dATP insertion opposite *syn*-O2-Me-T.[a]

Interacting Residue (Coordinating Atom)	Mg(II) Distance (Å)	Standard Deviation (Å)	Occupancy (%)
dATP (O12)	2.1	0.1	100
D13 (OD1)	2.0	0.0	100
E116 (OE2)	1.9	0.0	100
D115 (OD1)	2.0	0.0	100
5'-C (O3')	2.2	0.1	100
Water (O)	2.1	0.1	100
D13 (OD2)	2.1	0.1	100
M17 (O)	1.9	0.0	100
D115 (OD2)	2.0	0.0	100
dATP (O33)	1.9	0.0	100
dATP (O22)	1.9	0.0	100
dATP (O12)	2.3	0.2	91

[a]The Mg(II) coordination occupancies are based on a heavy atom distance cut-off of 2.5 Å

Appendix B: Supplementary Information for Chapter 3:
Computational Study on the Replicative Outcomes of the O4-4-(3-pyridyl)-4-oxobut-1-yl-thymine Lesion

Figures B.1 – B.11 and Tables B.1 – B.5

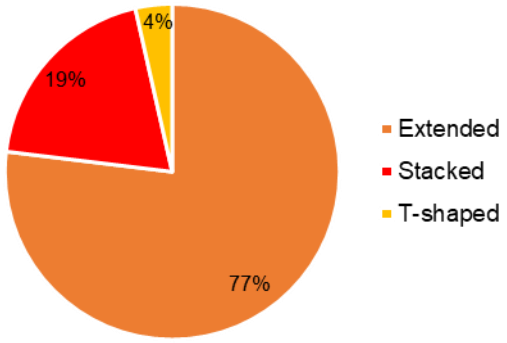


Figure B.1: Percentage of extended (orange), stacked (red) and T-shaped (yellow) O4-POB-T conformations identified from the DFT conformational search.

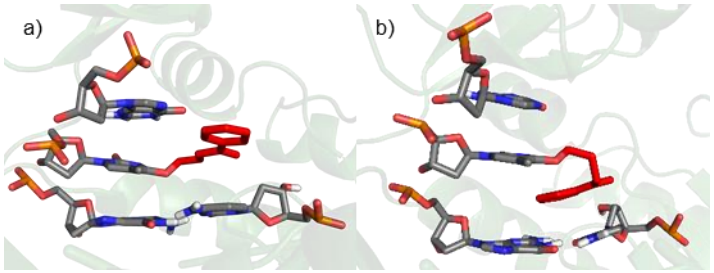


Figure B.2: MD representative structure of a) the lesion extended into the open pocket (> 90% occupancy) and b) a stacked conformation with the 3′ base pair (< 10% occupancy) in the pol η pre-insertion complex containing O4-POB-T (nucleobase in grey, POB in red).

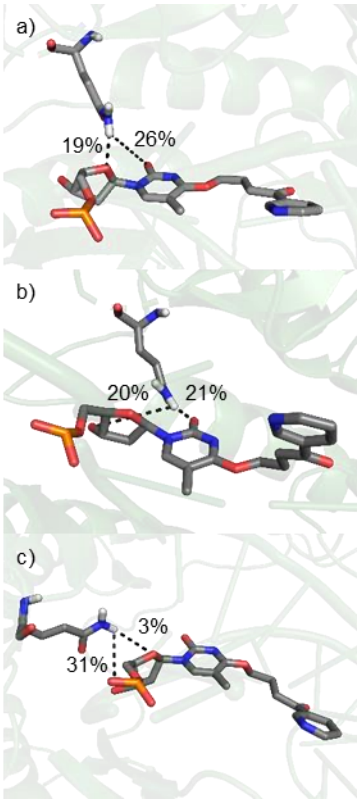
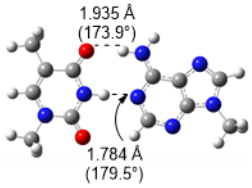


Figure B.3: Hydrogen-bonding occupancies between O4-POB-T and key active site residue Q38 in the pol η a) pre-insertion complex, and ternary complex corresponding to b) dGTP or c) dATP insertion.



T:A
[–70.8 kJ/mol]
∠(N9C1'C1')=56.4°
R(C1'···C1')=10.481 Å
Interplanar angle=0.1°

Figure B.4: Key DFT structural parameters of the canonical T:A base pair (Å and °) and associated interaction energy [kJ/mol].

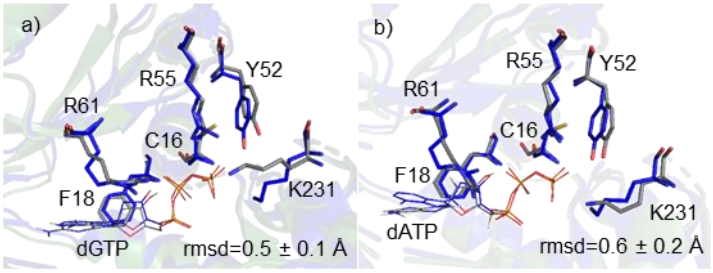


Figure B.5: Overlay of X-ray crystal structure (PDB ID: 4ECS, blue) (grey) and MD representative structure for pol η insertion of a) dGTP and b) dATP opposite O4-POB-T, highlighting alignment of key active site residues that interact with the dNTP.

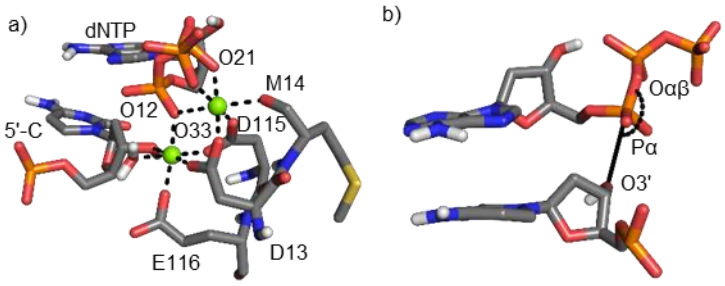


Figure B.6: a) Mg(II) coordination required to promote catalytic activity of pol η and b) definition of reaction parameters for the incoming dNTP with respect to the corresponding 5′ base.

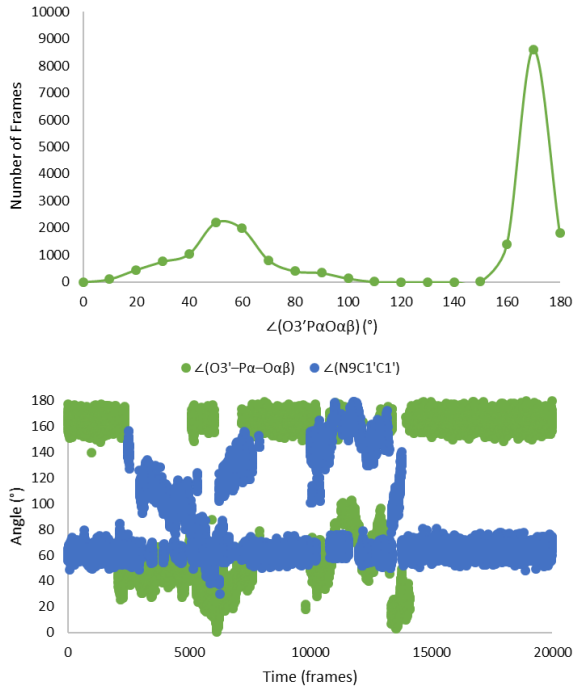


Figure B.7: Histogram of the reaction angle adopted over the 1 μs MD simulation of the pol η ternary complex corresponding to dGTP insertion (top). The reaction angle (∠(O3′–Pα–Oαβ), ideally 180°, green) and base pair opening angle (∠(N9C1′C1′), ideally 55°, blue) over the 1 μs MD simulation, highlighting the overlap between unfavourable angles over time (bottom).

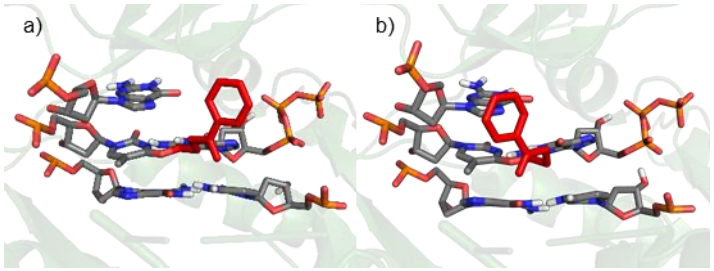


Figure B.8: a) MD representative structure of the pol η ternary complex corresponding to dGTP insertion opposite O4-POB-T with the lesion extended into the open pocket; b) MD representative structure of the pol η ternary complex corresponding to dGTP insertion opposite O4-POB-T with the lesion interacting with 5′-G, (nucleobase in grey, POB in red).

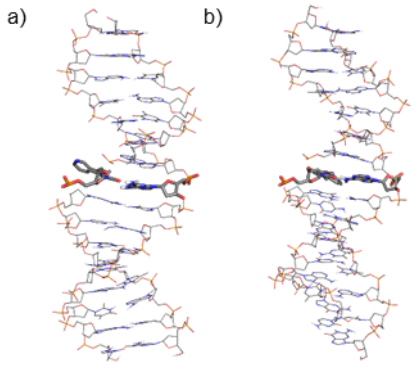


Figure B.9: MD representative structure of a DNA duplex containing the a) O4-POB-T:G mispair or b) O4-POB-T:A pair.

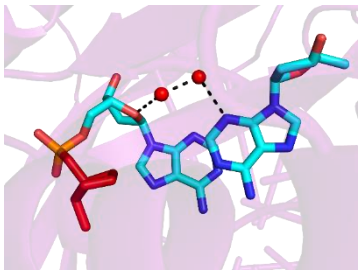


Figure B.10: X-ray crystal structure (PDB ID: 4RNM) corresponding to pol η insertion of dATP opposite an abasic site (red).

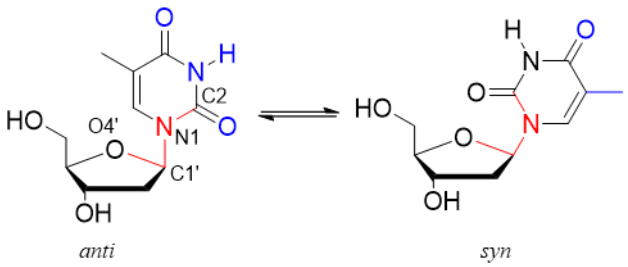


Figure B.11: Rotation about the glycosidic bond (∠(O4′C1′N1C2), red) in the canonical T nucleoside between the *anti* and *syn* conformations, as well as identification of the hydrogen-bonding face (blue) for each conformation.

Table B.1: Atom names and types, as well as associated partial charges for O4-POB-T in the isolated conformational search.

Atom Name	Atom Type	Partial Charge
N1	N*	-0.1311
C6	CM	-0.0997
H6	H4	0.201
C5	CM	-0.1846
C7	CT	-0.2316
H71	HC	0.0838
H72	HC	0.0838
H73	HC	0.0838
C4	CA	0.7792
O4	OS	-0.358
N3	NC	-0.7797
C2	C	0.7887
O2	O	-0.5808
C8	CT	-0.0221
H81	H1	0.0966
H82	H1	0.0966
C9	CT	-0.0053
H91	HC	0.0353
H92	HC	0.0353
C10	CT	0.0044
H101	HC	0.0252
H102	HC	0.0252
C11	C	0.3549
O11	O	-0.486
C12	CA	-0.0915
C13	CA	-0.0546
H13	HA	0.18
C14	CA	-0.3497
H14	HA	0.1587
C15	CA	0.3553
H15	H4	0.0699
N16	NC	-0.5937
C17	CA	0.3099
H17	H4	0.0899
C1'	CT	0.1272
H1'1	H1	-0.0054

H1'2	H1	-0.0054
H1'3	H1	-0.0054

Table B.2: Atom names and types, as well as associated partial charges for O4-POB-T in the pol η complexes and post-extension DNA duplexes.

Atom Name	Atom Type	Charge
P	P	1.1624
O5'	OS	-0.4935
O1P	O2	-0.7628
O2P	O2	-0.77
O3'	O	-0.5447
C5'	CT	-0.0071
H5'	H1	0.0822
H5"	H1	0.0822
C4'	CT	0.1368
H4'	H1	0.1077
O4'	OS	-0.4001
C1'	CT	0.1772
H1'	H2	0.0842
C3'	CT	0.1521
H3'	H1	0.0753
C2'	CT	-0.0596
H2'	HC	0.0433
H2"	HC	0.0433
N1	N*	-0.1353
C6	CM	-0.1013
H6	H4	0.2053
C5	CM	-0.1881
C7	CT	-0.2016
H71	HC	0.0759
H72	HC	0.0759
H73	HC	0.0759
C4	CA	0.78
O4	OS	-0.3578
N3	NC	-0.7838
C2	C	0.7944
O2	O	-0.5824
C8	CT	-0.0282
H81	H1	0.0999
H82	H1	0.0999
C9	CT	-0.0094
H91	HC	0.0365
H92	HC	0.0365

C10	CT	0.0048
H10	HC	0.0252
H10	HC	0.0252
C11	C	0.3552
O11	O	-0.486
C12	CA	-0.0913
C13	CA	-0.0542
H13	HA	0.1799
C14	CA	0.3096
H14	H4	0.09
C15	CA	-0.35
H15	HA	0.1589
C17	CA	0.355
H17	H4	0.0701
N16	NC	-0.5935

Table B.3: Occupancy of Mg(II) coordinating residues throughout 1 μs MD simulation on the pol η ternary complex corresponding to dGTP insertion opposite O4-POB-T.[a]

Interacting Residue (Coordinating Atom)	**Mg(II) Distance (Å)**	**Standard Deviation (Å)**	**Occupancy (%)**
dGTP (O12)	2.4	0.2	64
D13 (OD1)	2.0	0.0	100
E116 (OE2)	1.9	0.0	100
D115 (OD1)	2.0	0.0	100
5'-C (O3')	2.1	0.1	100
Water (O)	2.1	0.1	100
D13 (OD2)	2.0	0.1	100
M17 (O)	1.9	0.0	100
D115 (OD2)	2.0	0.0	100
dGTP (O31)	1.9	0.0	100
dGTP (O22)	2.4	0.1	70
dGTP (O12)	2.1	0.1	100

[a]The occupancies are based on a heavy atom distance cut-off of 2.5 Å

Table B.4: Average base-pair hydrogen-bonding distances (Å) and angles (deg.) between O4-POB-T and an opposing A or G over 1 μs MD simulations on ternary complexes and post-extension duplexes.[a]

Hydrogen Bond	Distance	Angle
O4-POB-T:dGTP (ternary complex)		
O2(O4-POB-T)···(N2–H)G	2.9 Å	149.1°
N3(O4-POB-T)···(N1–H)G	3.3 Å	155.8°
N3(O4-POB-T)···(N2–H)G	3.2 Å	154.7°
O4-POB-T:G (post-extension)		
O2(O4-POB-T)···(N2–H)G	2.8 Å	153.4°
N3(O4-POB-T)···(N1–H)G	3.2 Å	155.6°
N3(O4-POB-T)···(N2–H)G	3.1 Å	154.4°
O4-POB-T:A (post-extension)		
N3(O4-POB-T)···(N6–H)A	3.1 Å	139.7°
O2(O4-POB-T)···(N6–H)A	3.1 Å	156.5°

[a]The hydrogen-bonding occupancies are based on a heavy atom distance and angle cut-off of 3.4 Å and 120°

Table B.5: Occupancy of Mg(II) coordinating residues throughout 1 μs MD simulation on the pol η ternary complex corresponding to dATP insertion opposite O4-POB-T.[a]

Interacting Residue (Coordinating Atom)	Mg(II) Distance (Å)	Standard Deviation (Å)	Occupancy (%)
dATP (O12)	2.1	0.1	100
D13 (OD1)	2.0	0.0	100
E116 (OE2)	1.9	0.0	100
D115 (OD1)	2.0	0.1	100
5'-C (O3')	2.2	0.1	100
Water (O)	2.1	0.1	100
D13 (OD2)	2.1	0.1	100
M17 (O)	2.0	0.0	100
D115 (OD2)	2.0	0.0	100
dATP (O31)	1.9	0.0	100
dATP (O22)	1.9	0.0	100
dATP (O12)	2.3	0.2	86

[a]The occupancies are based on a heavy atom distance cut-off of 2.5 Å

Appendix C: Supplementary Information for Chapter 4:
Impact of Alkyl Chain Shape on the Replication Outcomes of O2-alkyl-T Lesions

Figures C.1 – C.12 and Tables C.1 – C.11

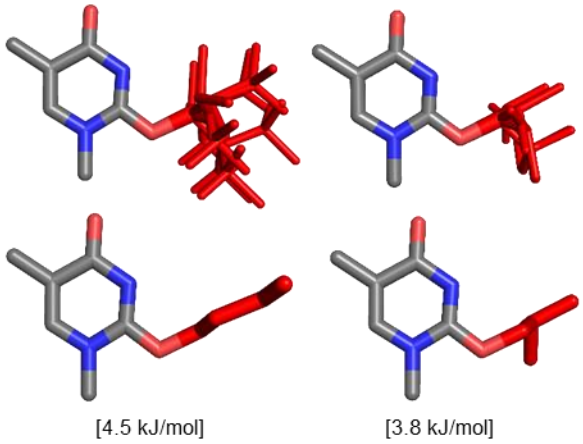


Figure C.1: Overlay of conformations based on the T ring atoms (top) and lowest energy conformation that fit within the pol η active site (bottom), with associated relative energy span over the conformations (kJ/mol).

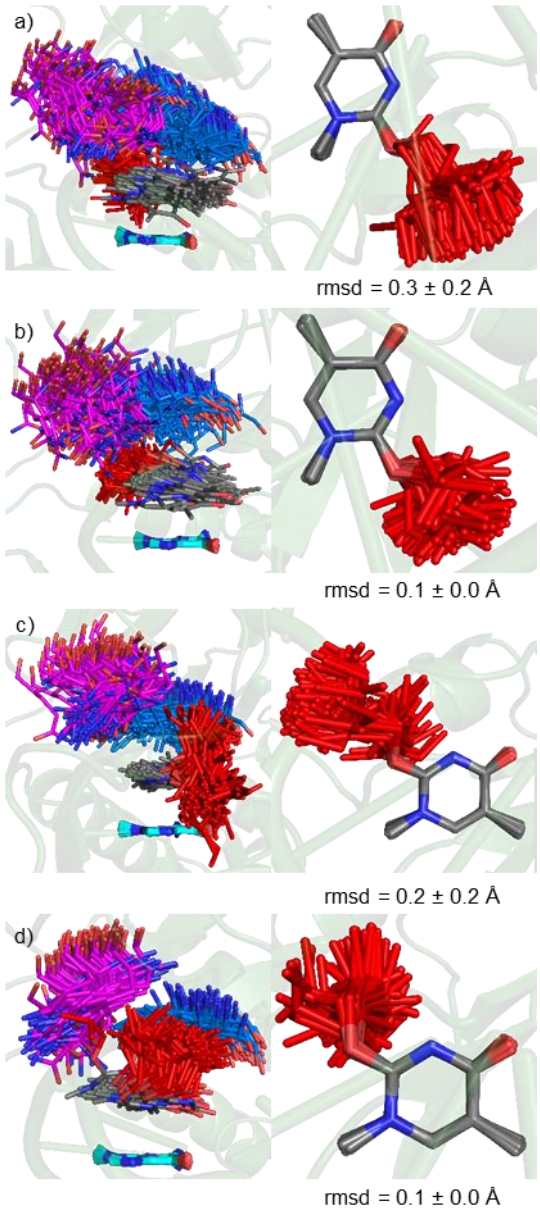


Figure C.2: MD snapshots taken every 5 ns overlaid with respect to 3′-G ring (cyan) atoms, highlighting the dynamics of the lesion (grey) and alkyl chain (red), as well as active site residues Q38 (pink) and I48 (blue) (left) and overlaid with respect to T-ring atoms highlighting flexibility of the alkyl chain (right) in the pol η pre-insertion complex corresponding to a) *anti*-O2-nBu-T, b) *anti*-O2-iBu-T, c) *syn*-O2-nBu-T and d) *syn*-O2-iBu-T.

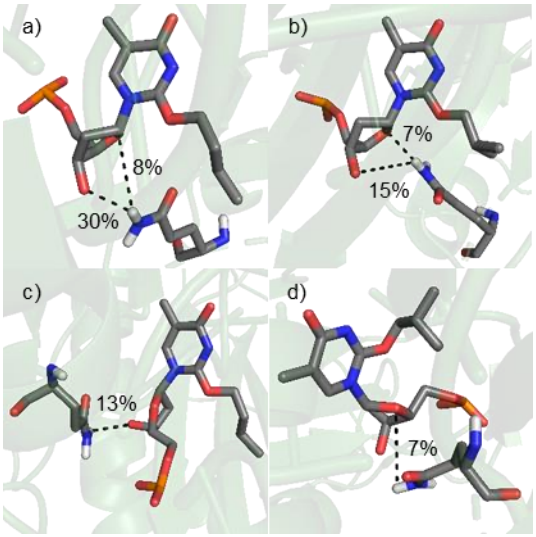


Figure C.3: MD representative structure and interactions between Q38 and the T lesion (%) in the pol η pre-insertion complex corresponding to a) *anti*-O2-nBu-T, b) *anti*-O2-iBu-T, c) *syn*-O2-nBu-T and d) *syn*-O2-iBu-T.

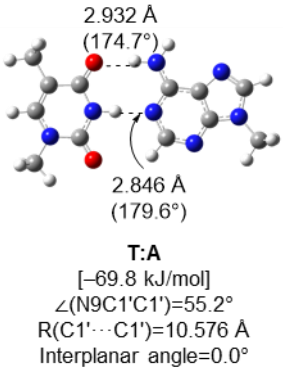


Figure C.4: Key structural features (Å and deg.) and interaction energy [kJ/mol] of the canonical T:A pair.

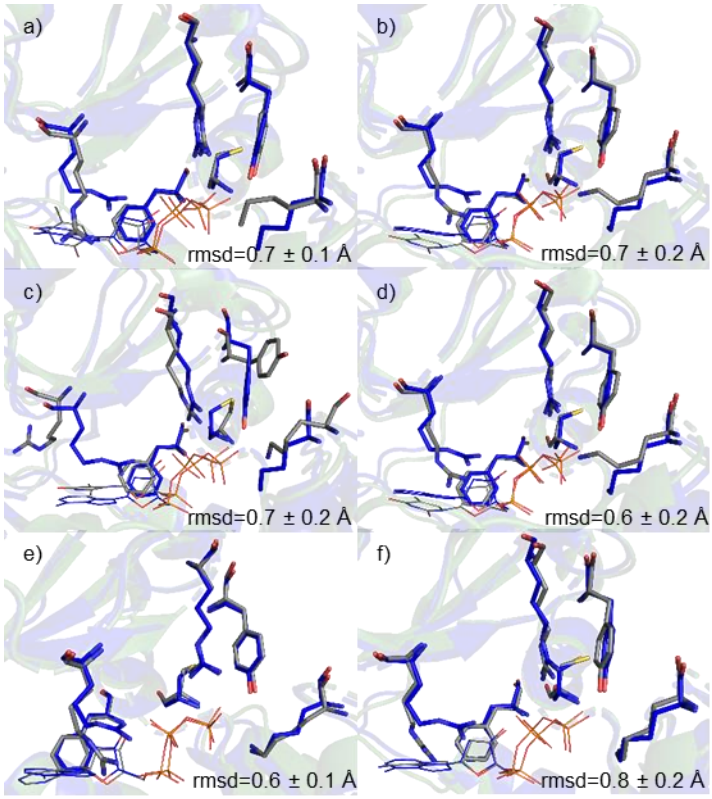


Figure C.5: Overlay of X-ray crystal structure (PDB ID: 4ECS, blue) and MD representative structure (grey) for key active sites in the pol η ternary complexes corresponding to dTTP insertion opposite a) *anti*-O2-nBu-T and b) *anti*-O2-iBu-T, c) *syn*-O2-nBu-T and d) *syn*-O2-iBu-T, and dATP insertion opposite e) *syn*-O2-nBu-T and f) *syn*-O2-iBu-T.

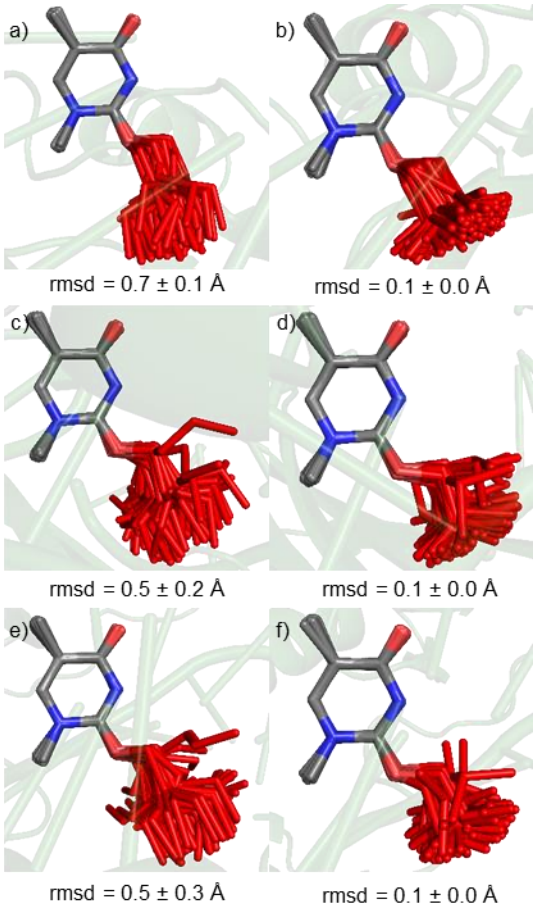


Figure C.6: MD snapshots taken every 5 ns overlaid with respect to T ring atoms, highlighting the orientation and flexibility of the alkyl chain in the pol η ternary complexes corresponding to dTTP insertion with resepective rmsds opposite a) *anti*-O2-nBu-T and b) *anti*-O2-iBu-T, c) *syn*-O2-nBu-T and d) *syn*-O2-inBu-T, and dATP insertion opposite e) *syn*-O2-nBu-T and f) *syn*-O2-iBu-T.

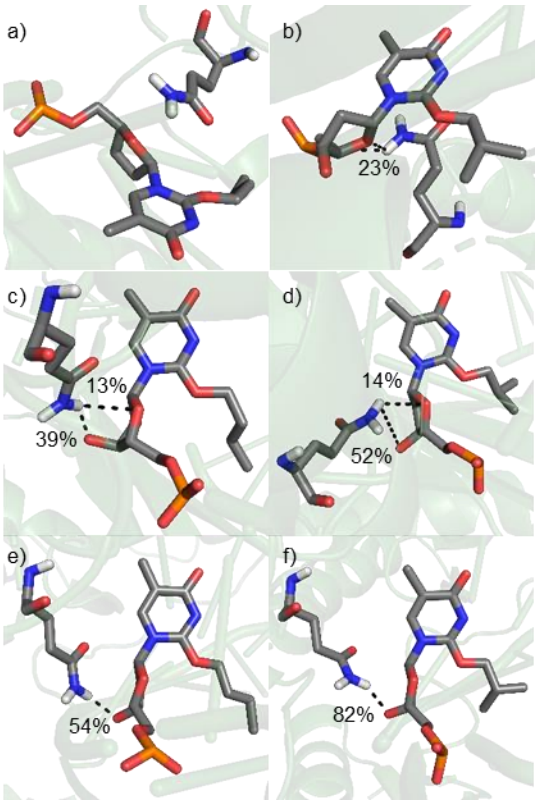


Figure C.7: MD representative structure and interactions between Q38 and the T lesion (%) in the pol η ternary complex corresponding to dTTP insertion opposite a) *anti*-O2-nBu-T and b) *anti*-O2-iBu-T, c) *syn*-O2-nBu-T and d) *syn*-O2-inBu-T, and dATP insertion opposite e) *syn*-O2-nBu-T and f) *syn*-O2-iBu-T.

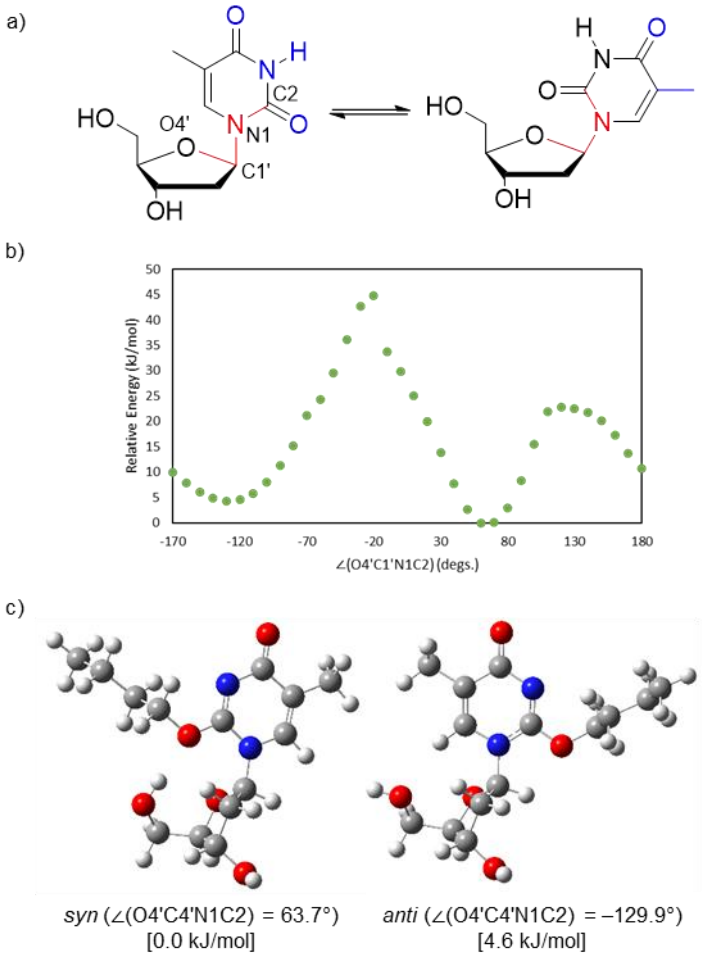


Figure C.8: a) Chemical structure for the rotation about χ (∠(O4′C4′N1C2, red) with the hydrogen-bonding faces highlighted in blue. B) B3LYP-D3(BJ)/6-311++G(2df,2p)//B3LYP/6-31G(d,p) relative energy (kJ/mol) as a function of the ∠(O4′C4′N1C2) dihedral angle that dictates the orientation of the glycosidic bond for O2-nBu-T. c) Fully optimized structures of the most stable *syn* and *anti* orientations, with associated ∠(O4′C4′N1C2) dihedral angle and relative energies [kJ/mol].

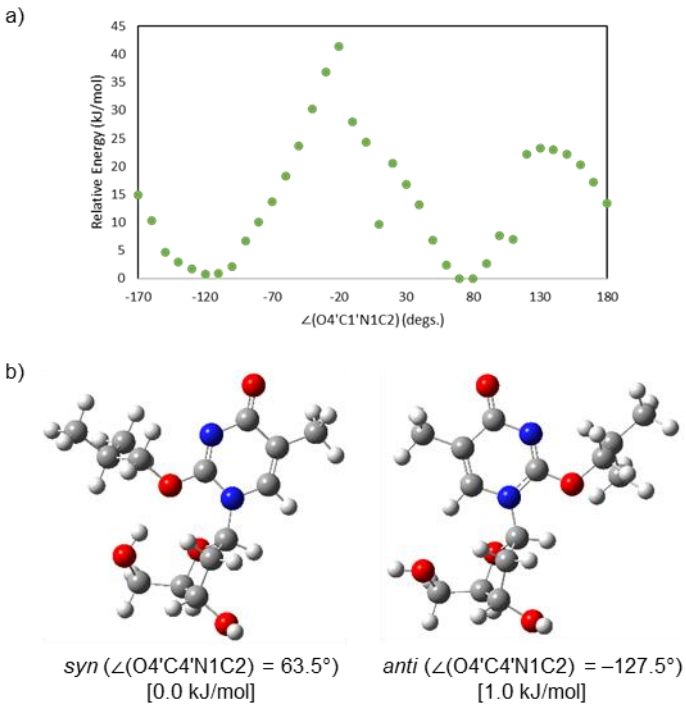


Figure C.9: a) B3LYP-D3(BJ)/6-311++G(2df,2p)//B3LYP/6-31G(d,p) relative energy (kJ/mol) as a function of the ∠(O4′C4′N1C2) dihedral angle that dictates the orientation of the glycosidic bond for O2-iBu-T. b) Fully optimized structures of the most stable *syn* and *anti* orientations, with associated ∠(O4′C4′N1C2) dihedral angle and relative energies [kJ/mol].

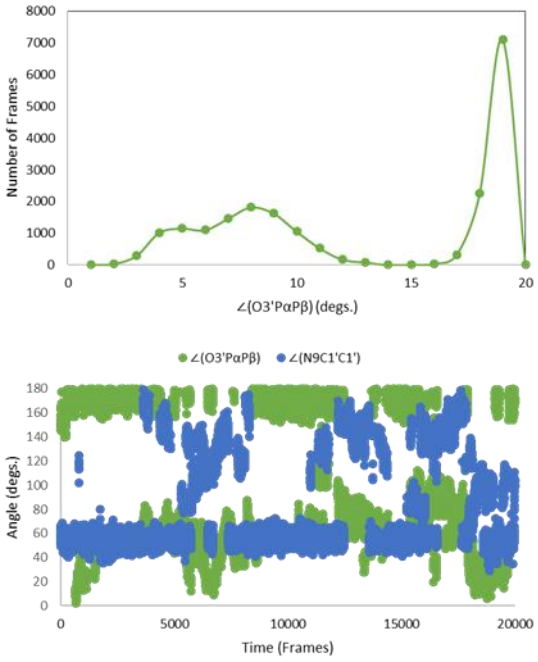


Figure C.10: Histogram of the reaction angle (deg.) conformations adopted from the pol η ternary complex corresponding to dTTP insertion opposite *syn*-O2-nBu-T (top). The reaction angle (∠(O3′–Pα–Oαβ), ideally 180°, green) and base pair opening angle (∠(N9C1′C1′), ideally 55°, blue) highlighting the overlap between unfavourable angles over time (bottom).

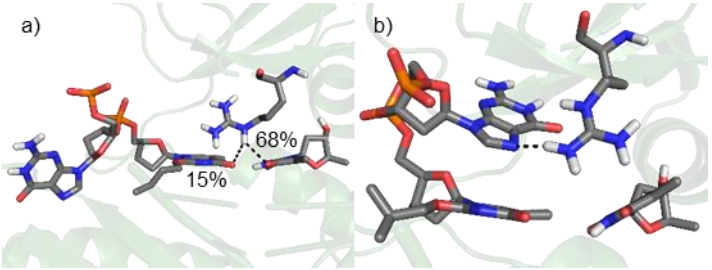


Figure C.11: Interactions between R61 and the lesion base pair in the pol η ternary complexes corresponding to dTTP insertion opposite) *syn*-O2-nBu-T and b) *syn*-O2-iBu-T.

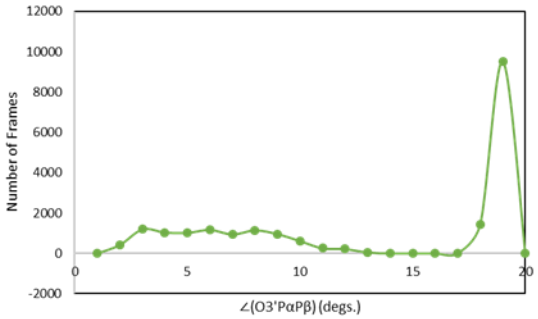


Figure C.12: Histogram of the reaction angle conformations adopted from the pol η ternary complex corresponding to dATP insertion opposite *syn*-O2-nBu-T.

Table C.1: Atom types and partial charges for O2-nBu-T in the isolated conformational search.

Atom Name	Atom Type	Partial Charge
N1	N*	-0.1308
C6	CM	-0.2142
H6	H4	0.2466
C5	CM	-0.0425
C7	CT	-0.2911
H71	HC	0.0909
H72	HC	0.0909
H73	HC	0.0909
C4	C	0.7659
O4	O	-0.5944
N3	NC	-0.735
C2	CA	0.7656
O2	OS	-0.4187
C8	CT	0.0277
H81	H1	0.0791
H82	H1	0.0791
C9	CT	0.0036
H91	HC	0.0262
H92	HC	0.0262
C10	CT	0.0409
H101	HC	0.0033
H102	HC	0.0033
C11	CT	-0.0807
H111	HC	0.0179
H112	HC	0.0179
H113	HC	0.0179
C1'	CT	0.1222
H1'1	H1	0.0029
H1'2	H1	0.0029
H1'3	H1	0.0029

Table C.2: Atom types and partial charges for O2-iBu-T in the isolated conformational search.

Atom Name	Atom Type	Partial Charge
N1	N*	-0.1153
C6	CM	-0.2413
H6	H4	0.2572
C5	CM	-0.0394
C7	CT	-0.276
H71	HC	0.087
H72	HC	0.087
H73	HC	0.087
C4	C	0.7616
O4	O	-0.5917
N3	NC	-0.7529
C2	CA	0.8137
O2	OS	-0.4651
C8	CT	-0.0266
H81	H1	0.0979
H82	H1	0.0979
C9	CT	0.2154
H91	HC	0.0094
C10	CT	-0.2452
H101	HC	0.0579
H102	HC	0.0579
H103	HC	0.0579
C11	CT	-0.1722
H111	HC	0.0415
H112	HC	0.0415
H113	HC	0.0415
C1'	CT	0.1042
H1'1	H1	0.0031
H1'2	H1	0.0031
H1'3	H1	0.0031

Table C.3: Atom types and partial charges for O2-nBu-T in the pol η pre-insertion and ternary complexes.

Atom Name	Atom Type	Partial Charge
P	P	1.1623
O5'	OS	-0.4983
OP1	O2	-0.7635
OP2	O2	-0.7692
O3'	OS	-0.5413
C5'	CT	-0.0211
H5'	H1	0.08
H5"	H1	0.08
C4'	CT	0.1781
H4'	H1	0.1081
O4'	OS	-0.4055
C1'	CT	0.1222
H1'	H2	0.1386
C3'	CT	0.1286
H3'	H1	0.0837
C2'	CT	-0.073
H2'	HC	0.0519
H2"	HC	0.0519
N1	N*	-0.1308
C6	CM	-0.2142
H6	H4	0.2466
C5	CM	-0.0425
C7	CT	-0.2911
H71	HC	0.0909
H72	HC	0.0909
H73	HC	0.0909
C4	C	0.7659
O4	O	-0.5944
N3	NC	-0.735
C2	CA	0.7656
O2	OS	-0.4187
C8	CT	0.0277
H81	H1	0.0791
H82	H1	0.0791
C9	CT	0.0036
H91	HC	0.0262
H92	HC	0.0262
C10	CT	0.0409
H101	HC	0.0033
H102	HC	0.0033
C11	CT	-0.0807
H111	HC	0.0179

H112	HC	0.0179
H113	HC	0.0179

Table C.4: Atom types and partial charges for O2-iBu-T in the pol η pre-insertion and ternary complexes.

Atom Name	Atom Type	Partial Charge
P	P	1.1623
O5'	OS	-0.4986
OP1	O2	-0.7634
OP2	O2	-0.7691
O3'	OS	-0.5418
C5'	CT	-0.0211
H5'	H1	0.0801
H5"	H1	0.0801
C4'	CT	0.1778
H4'	H1	0.1084
O4'	OS	-0.407
C1'	CT	0.1042
H1'	H2	0.1593
C3'	CT	0.1278
H3'	H1	0.0846
C2'	CT	-0.0737
H2'	HC	0.0518
H2"	HC	0.0518
N1	N*	-0.1153
C6	CM	-0.2413
H6	H4	0.2572
C5	CM	-0.0394
C7	CT	-0.276
H71	HC	0.087
H72	HC	0.087
H73	HC	0.087
C4	C	0.7616
O4	O	-0.5917
N3	NC	-0.7529
C2	CA	0.8137
O2	OS	-0.4651
C8	CT	-0.0266
H81	H1	0.0979
H82	H1	0.0979
C9	CT	0.2154
H91	HC	0.0094
C10	CT	-0.2452
H101	HC	0.0579
H102	HC	0.0579
H103	HC	0.0579
C11	CT	-0.1722

H111	HC	0.0415
H112	HC	0.0415
H113	HC	0.0415

Table C.5: Occupancy of Mg(II) coordinating residues throughout 1 μs MD simulation on the pol η ternary complex corresponding to dTTP insertion opposite *anti*-O2-nBu-T.[a]

Interacting Residue (Coordinating Atom)	**Mg(II) Coordination (Å)**	**Standard Deviation (Å)**	**Occupancy (%)**
dTTP (O12)	2.2	0.1	98
D13 (OD1)	2.0	0.0	100
E116 (OE2)	1.9	0.0	100
D115 (OD1)	2.0	0.1	100
5'-C (O3')	2.2	0.1	100
Water (O)	2.1	0.1	100
D13 (OD2)	2.0	0.1	100
M17 (O)	1.9	0.0	100
D115 (OD2)	1.9	0.0	100
dTTP (O31)	1.9	0.0	100
dTTP (O22)	2.0	0.1	100
dTTP (O12)	2.3	0.2	82.4

[a]The occupancies are based on a heavy atom distance cut-off of 2.5 Å

Table C.6: Occupancy of Mg(II) coordinating residues throughout 1 μs MD simulation on the pol η ternary complex corresponding to dTTP insertion opposite *anti*-O2-iBu-T.[a]

Interacting Residue (Coordinating Atom)	Mg(II) Coordination (Å)	Standard Deviation (Å)	Occupancy (%)
dTTP (O12)	2.2	0.1	96
D13 (OD1)	2.0	0.0	100
E116 (OE2)	1.9	0.0	100
D115 (OD1)	2.0	0.1	100
5'-C (O3')	2.2	0.1	99
Water (O)	2.1	0.1	100
D13 (OD2)	2.0	0.1	100
M17 (O)	1.9	0.0	100
D115 (OD2)	1.9	0.0	100
dTTP (O31)	1.9	0.0	100
dTTP (O22)	2.0	0.1	100
dTTP (O12)	2.4	0.2	74

[a]The occupancies are based on a heavy atom distance cut-off of 2.5 Å

Table C.7: Occupancy of Mg(II) coordinating residues throughout 1 μs MD simulation on the pol η ternary complex corresponding to dTTP insertion opposite *syn*-O2-nBu-T.[a]

Interacting Residue (Coordinating Atom)	Mg(II) Coordination (Å)	Standard Deviation (Å)	Occupancy (%)
dTTP (O12)	2.2	0.1	97
D13 (OD1)	2.0	0.0	100
E116 (OE2)	1.9	0.0	100
D115 (OD1)	2.0	0.1	100
5'-C (O3')	2.2	0.1	99
Water (O)	2.1	0.1	100
D13 (OD2)	2.0	0.1	100
M17 (O)	1.9	0.0	100
D115 (OD2)	1.9	0.0	100
dTTP (O31)	1.9	0.0	100
dTTP (O22)	2.0	0.1	100
dTTP (O12)	2.4	0.3	69

[a]The occupancies are based on a heavy atom distance cut-off of 2.5 Å

Table C.8: Average base-pair hydrogen-bonding distances (Å) and angles (deg.) between *syn*-O2-nBu-T and dTTP in the pol η ternary complex.[a]

Hydrogen Bond	Distance	Angle
O4(O2-nBu-T)···(N3–H)T	2.9 Å	157.2°

[a]The hydrogen-bonding occupancies are based on a heavy atom distance and angle cut-off of 3.4 Å and 120°

Table C.9: Occupancy of Mg(II) coordinating residues throughout 1 μs MD simulation on the pol η ternary complex corresponding to dTTP insertion opposite *syn*-O2-iBu-T.[a]

Interacting Residue (Coordinating Atom)	Mg(II) Coordination (Å)	Standard Deviation (Å)	Occupancy (%)
dTTP (O12)	2.1	0.1	98
D13 (OD1)	2.0	0.0	100
E116 (OE2)	1.9	0.0	100
D115 (OD1)	2.0	0.1	100
5'-C (O3')	2.2	0.1	99
Water (O)	2.1	0.1	100
D13 (OD2)	2.1	0.1	100
M17 (O)	1.9	0.0	100
D115 (OD2)	1.9	0.0	100
dTTP (O31)	1.9	0.0	100
dTTP (O22)	2.0	0.1	100
dTTP (O12)	2.4	0.3	72

[a]The occupancies are based on a heavy atom distance cut-off of 2.5 Å

Table C.10: Occupancy of Mg(II) coordinating residues throughout 1 μs MD simulation on the pol η ternary complex corresponding to dATP insertion opposite *syn*-O2-nBu-T.[a]

Interacting Residue (Coordinating Atom)	Mg(II) Coordination (Å)	Standard Deviation (Å)	Occupancy (%) < 2.5 Å
dATP (O12)	2.1	0.1	100
D13 (OD1)	2.0	0.0	100
E116 (OE2)	1.9	0.0	100
D115 (OD1)	2.0	0.1	100
5'-C (O3')	2.2	0.1	100
Water (O)	2.1	0.1	100
D13 (OD2)	2.1	0.1	100
M17 (O)	2.0	0.0	100
D115 (OD2)	1.9	0.0	100
dATP (O31)	1.9	0.0	100
dATP (O22)	1.9	0.0	100
dATP (O12)	2.2	0.2	89

[a]The occupancies are based on a heavy atom distance cut-off of 2.5 Å

Table C.11: Occupancy of Mg(II) coordinating residues throughout 1 μs MD simulation on the pol η ternary complex corresponding to dATP insertion opposite *syn*-O2-iBu-T.[a]

Interacting Residue (Coordinating Atom)	Mg(II) Coordination (Å)	Standard Deviation (Å)	Occupancy (%)
dATP (O12)	2.1	0.1	100
D13 (OD1)	2.0	0.0	100
E116 (OE2)	1.9	0.0	100
D115 (OD1)	2.0	0.1	100
5'-C (O3')	2.2	0.1	100
Water (O)	2.1	0.1	100
D13 (OD2)	2.1	0.1	100
M17 (O)	2.0	0.0	100
D115 (OD2)	2.0	0.0	100
dATP (O31)	1.9	0.0	100
dATP (O22)	1.9	0.0	100
dATP (O12)	2.3	0.2	87

[a]The occupancies are based on a heavy atom distance cut-off of 2.5 Å