The Dissertation Committee for Yi Kou Certifies that this is the approved version of the following dissertation:

Structural and Kinetic Study of N7-methyl, N7-benzyl and C8-chloro Guanine Lesions Using Human DNA Polymerase β

Committee:

Seongmin Lee, Supervisor

Brent L. Iverson

Hung-Wen Liu

Tanya T. Paull

Christian P Whitman

Structural and Kinetic Study of N7-methyl, N7-benzyl and C8-chloro Guanine Lesions Using Human DNA polymerase β

by

Yi Kou, B.Med

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2015

Dedication

To Mingxian Kou and Shufang Li

And

My TATA

Acknowledgements

This work would not have been possible without the assistance of countless people who have played a part in shaping me as both a person and a scientist. I must first thank my parents, Mingxian Kou and Shufang Li, for their loving support and encouragement, which helps me overcoming many difficulties. I also would like to thank my advisor, Dr. Seongmin Lee. For the last five years, Dr. Lee has been the main force molding the scientist I am today. I thank him for exemplifying the excellence of scientific persistence and creativity.

I thank my doctoral committee members: Dr. Brent Iverson, Dr. Hung-wen Liu, Dr. Tanya Paull and Dr. Christian Whitman for their constructive suggestions and kind criticism.

I would also like to thank many others, especially my lab mates in Lee lab: Dr. Myong Koag, Young and Hala, and Dr. Lin. Science is with much fun to work and talk with friends.

Finally, I would like to thank my loving wife, Fang. I cannot thank her enough for her support during this graduate research life. She has sacrificed so many precious moments of her life to help me.

All I can say for my beloved ones, thank you, thank you for all the time and patience tolerating my ignorance, prejudice and arrogance. Thank you for all the time spent selflessly.

I love you all.

Structural and Kinetic Study of N7-methyl, N7-benzyl and C8-chloro Guanine Lesions Using Human DNA polymerase β

Yi Kou, PhD.

The University of Texas at Austin, 2015

Supervisor: Seongmin Lee

DNA bases are constantly under the damages from both outside and inside, bringing possible mutagenic changes. To elucidate the detailed mechanisms, structural method of X-ray crystallography and steady state kinetics together with other necessary means are used in human DNA polymerase β (polβ) model to study the N7-methyl (N7meG), N7-benzyl (N7bnG) and C8-chloro (8ClG) guanine lesions. The results show that polβ will maintain the DNA fidelity by deterring dTTP insertion across N7meG lesion. But dCTP insertion is also severely slowed down. For the N7meG post-insertion base pairs, surprisingly unlike others, N7meG:dT base pair shows stabilizing effect on the DNA duplex, as indicated by the structure and thermodynamic parameters. When N7 substitution is replaced by bulky and flexible group to form N7bnG, correct insertion of dCTP shows metal ion dependence. And the previously minimal insertion of dTTP can be seen with increased efficiency. Both of which are also contributed by the flexibility and hydrophobic interactions made by benzyl group. For 8ClG lesion, it shows Watson-Crick base pair for incoming dCTP, Hoogsteen base pair for incoming dGTP, similar to 8BrG in the polβ model. However, from the structural comparison, 8ClG seems to be more mutagenic than 8BrG. Additionally, it can have some dATP and dTTP insertions, although they are in less extent both structurally and kinetically.

Table of Contents

Chapter I: Human DNA polymerase β proceeds across the chemically unstable lesion N7-methylguanine[1]

1.1 BACKGROUND INTRODUCTION

Deoxyribonucleic acid (DNA), the polymer made of repeating nucleotides, is the molecule that carries the most of the genetic information and instructions, which are essential to the development and normal functioning of all known living organisms. Important as it is, DNA, as all other macrobiomolecules, is also very fragile and susceptible to various kinds of damage. And by which, its sequence may change, leading to many potential diseases including cancer.

The genetic information stored in DNA exists preciously in majority of the cells: a quite risky situation considering that it accumulates the damage over the whole time without aid from remanufacturing and depends merely on the repair machinery. Although not all lesions will cause inheritable deleterious mutation, once the mutation is created, it is permanent even in the descendant cells. Additionally, the compiling effects could be resulted from erroneous operations during the replication, recombination, lesion recognition and even repair process. Furthermore, not only the sequence instability could lead to cancer, studies have been confirmed that cancers could also be stemmed from the epigenetic change[1], and which could possibly exist even as a permanent cause[2].

From the general view, the damage made to DNA can be categorized by sources into two kinds: endogenous and exogenous. Endogenous damage can be further divided into two categories: i) The unstable nature of the DNA molecule itself. And ii) various metabolites as the damaging reagents. Unlike the weakness of phosphodiester bond

[1] Chapter I is based on the publication :Transition-state destabilization reveals how human DNA polymerase β proceeds across the chemically unstable lesion N7-methylguanine. Koag MC*, Kou Y*, Ouzon-Shubeita H, Lee S. Nucleic Acids Res. 2014 Jul;42(13):8755-66. * These authors contributed equally to the work.

contributed by 2'-OH in the ribonucleic acid (RNA), in DNA, where 2' position is in deoxy state, the structural and chemical weakness now moves to N-glycosidic bond. This bond is very labile under various physical, chemical or biological conditions, such as heating, base ring modifications and certain enzymes, leading to the production of abasic sites (apurinic or apyrimidinic, AP). Once not repaired, AP sites can lead to backbone break and mutation because of preferred insertion of adenine (A). Another aspect of unstable nature of DNA molecule is hydrolytic deamination. The A can be converted into hypoxanthine, which will then be base paired with cytosine (C), instead of thymine (T)[3]. Deamination of guanine (G), changes it into xanthine, which will remain base pair with C, but it is also easy to be hydrolyzed to generate abasic site[4]. Deamination of C changes it into uracil (U) which can be efficiently repaired, while deaminated 5-methyl cytosine can result in G:T mismatch which could be mutagenic[4].

Other endogenous DNA damages include oxidative DNA damage. The reactive oxygen species (ROS) can lead to the formation of 8-oxo-guanine (8oxoG), which will preferentially base pair with A rather than C, creating a transversion mutation[5]. Other bases can also be damaged oxidatively as well. Endogenous DNA damage can come from the adductions that are stemmed from lipid peroxidations. Lipid hydroperoxides can react with metals to generate epoxides, acrolein, crotonaldehyde, malondialdehyde (MDA) and 4-hydroxynonenal (HNE). These chemicals can modify the bases to make different adducts like N2-(3-hydroxybutylidene)-G, N4-etheno-C, and N6-(3-oxopropenyl)-A[6]. These lesions can either block the replication or generate mutations through deletion and base replacement. Estrogen induced DNA damage can also be an endogenous damage. It can generate bulky adducts, strand break or base hydroxylation modifications[7].

DNA can also be damaged from outside, such as various chemical and physical factors. Various chemicals can modify bases. As will be discussed in the chapter IV,

some of these base modifications, like 8ClG can lead to mismatch mutations. Even without chemical modification, there exist the intercalators like ethidium bromide that can be fitted between the bases, distorting duplex structure and causing trouble to replication and transcription. Physical factors such as high energy level radiation like ultra violet (UV) and X rays can also damage DNA. A well known example is the TT dimer formed by UV irradiation[8].

One important damage on DNA is alkylation damage, which is the topic for most of the chapters in this thesis. This type of damage can be done both endogenously and exogenously. The major alkylation type is methylation. The most important endogenous methylating reagent is *S*-adenosylmethionine (SAM). SAM is produced from adenosine triphosphate (ATP) and methionine by methionine adenosyltransferase, and is also involved in trans-sulfuration and aminopropylation besides transmethylation pathway. It has a chemically active methyl group (-CH_3, Me) attached to the methionine sulfur atom as shown below (Figure 1.1):

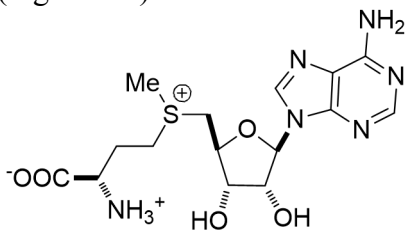


Figure 1.1 *S*-Adenosyl methionine

A variety of substrates, including nucleic acids, proteins and other metabolites, can accept the methyl group donated by SAM via transmethylation. Additional to its endogenous production, SAM is also approved to be on the market in north America as the dietary supplement for treating depression, osteoarthritis, and other diseases[9][10]. Other endogenous sources of methylating reagents include choline and its derivative betaine, besides their respective role as osmolyte and cell membrane components.

There are many other exogenous sources of methylation reagents. They are mainly in two categories: the methylating (alkylating) anti-cancer drugs and the tobacco related products. Dacarbazine, a triazenoimidazole compound used in treating malignant melanoma, is one of the examples for DNA methylating anti-cancer drugs[11]. Nitrosamines such as 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) is a famous tobacco related DNA methylating reagent. After activation *in vivo*, such as the α-methylene hydroxylation, methane diazohydroxide or the methyldiazonium ion is generated from NNK, both of which can methylate DNA[12], as shown below (Figure 1.2):

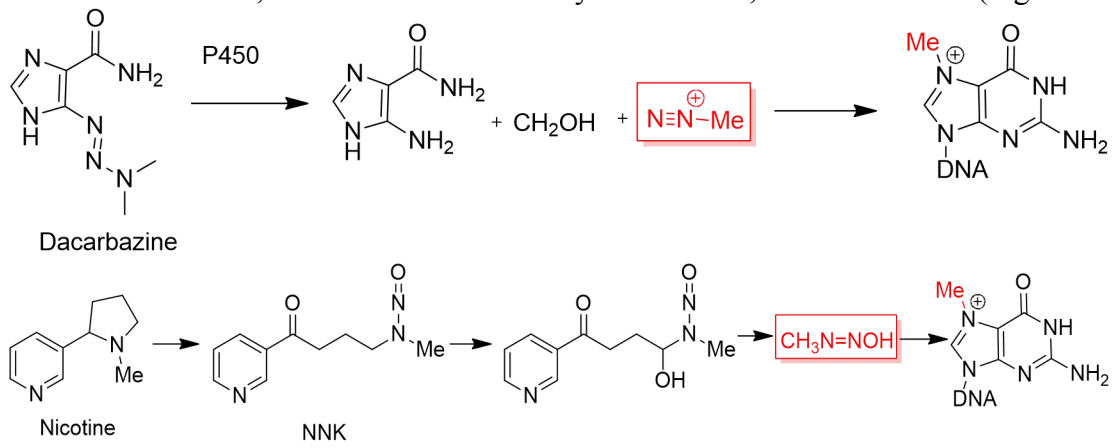


Figure 1.2 Methylation mechanisms of Dacarbazine and Nicotine derivative

On the DNA side, N or O atoms on the base ring can act as the nucleophiles to react with electrophilic methylating reagents to become methylated. There are many such sites on the nucleobases, according to their nucleophilicity, they can be arranged from the most active to the least active in such a sequence: N-7 of G > N-3 of A > N-7 of A > N-3 of G> N-1 of A > N-1 of C[13]. In addition to these hot spots, N-3 of C, O-6 of G, and the phosphate groups can also be methylated. As can be seen from the reactivity, N7 of G is the most reactive site, plus its easy accessibility towards the major groove side which is not involved in base pair H bonding (hydrogen bonds), N7 methylation accounts for ~70-80% of the total methylated lesions[14].

N7 of G is not only the site for methylation, it is also a well known site for almost all other DNA alkylations as well. As shown below (Figure 1.3):

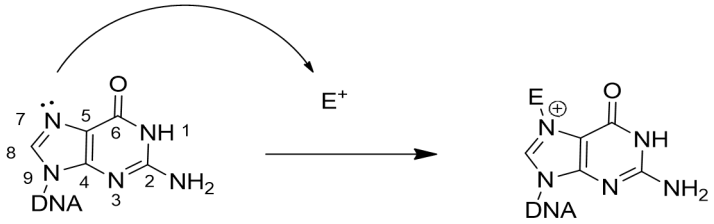


Figure 1.3 N7G alkylation

The adduction (covalent bond formation) at the N7 position can be seen from many other electrophiles: bulky or flexible carcinogens (as will be discussed in detail in chapter III), anti-cancer drugs, or DNA intra- or inter-strand cross-linkers. Examples of these adducts can be seen from aflatoxins to cisplatin. In fact, the methyl or ethyl adduction on N7 has been considered as not so toxic ones that are tolerated by the cells[15]. An indirect reflection of this can be seen from the naturally occurred structure of N7 quarternized methyl group in mRNA cap and in many secondary metabolites such as herbipoline, as shown below (Figure 1.4):

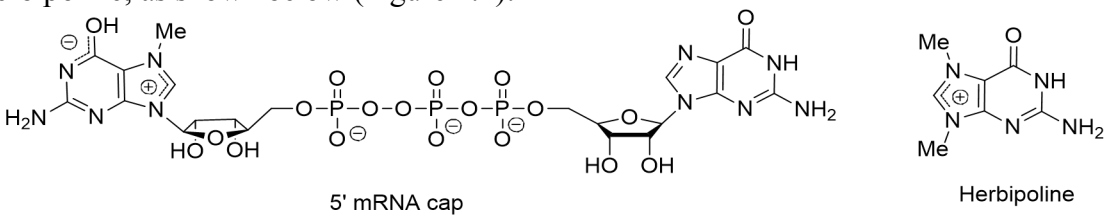


Figure 1.4 Naturally occurred N7 methylations

After the N7 alkylation, there are many types of subsequent reactions and even more concerning only the direct derivatives produced out of these reactions. This is so because the N7 alkylation, even to be considered in a limited range of local guanine base (without the steric or intercalating effects into neighboring environment), brings the positive charge onto the saturated N atom on a previous stabilized and electron-rich ring system. Thus by alkylation, the electrons from such a conjugated system are to some

extent depleted, and as a result, the whole balance of base ring system is destroyed. All the subsequent reactions featured for N7 alkylation can be regarded and understood in such a way that these reactions are there to salvage or revert the base ring system.

There are three main categories of subsequent reactions after the N7 alkylation of guanine (N7-alkyl G): i) acidity changes of C8 and N1 of N7-alkyl G; ii) the formation of formamidopyrimidine guanine (FAPy G); and most importantly, iii) spontaneous depurination of N7-alkyl G.

One of the characteristic and quantification assays of N7 alkylation is to measure the quantity of labeled C8-H from the N7-alkyl G, after it is exchanged from the water. The alkylation on the N7 is known to facilitate the deprotonation of C8-H to an extent that even at neutral pH and mild temperature, it would be exchanged with solvent proton in minutes[16]. The mechanisms of such process has been suggested to include a neutral dipolar transition state shown below ():

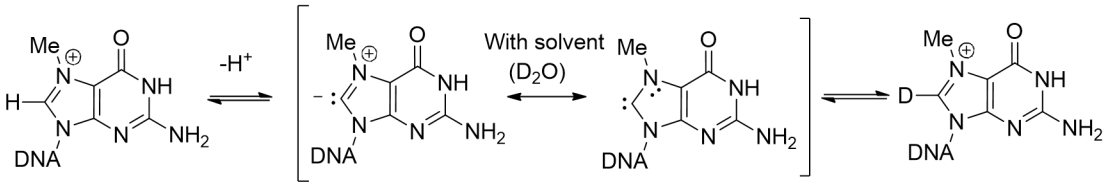


Figure 1.5 C8-H exchange with solvent

Another important acidity increase brought by N7 alkylation is the p*K*a change of N1H. It could drop from around 9 to about 7 after the N7 is alkylated on the guanine[17][18]. This p*K*a change makes the N1-H easily to be deprotonated and after which it can then be considered as H bond acceptor in forming the potential N7meG:T mismatch. This pseudo-Watson-Crick mismatch has two H bonds, in which one is formed on the basis of N1H deprotonation, as shown below (Figure 1.6). Upon which, the N7-alkyl G (shown as N7meG) can now take the zwitterionic form. As will be shown in the

chapter II of host-guest system of N7meG, this type N7meG:T mismatch can stay stably in the DNA duplex.

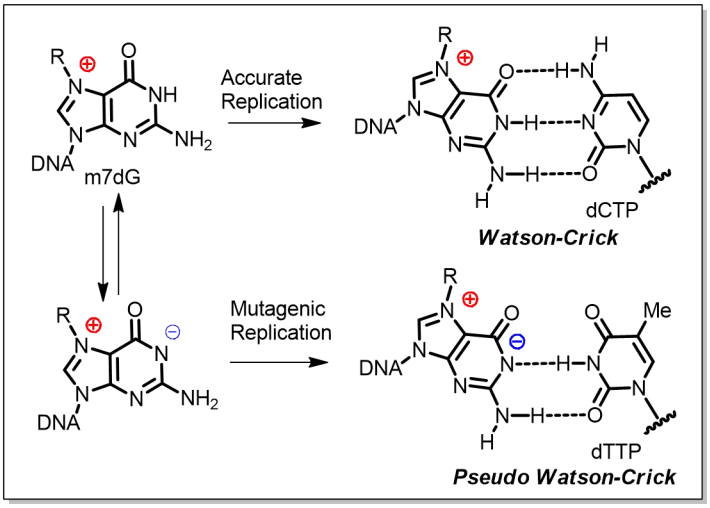


Figure 1.6 N7meG:T mismatch hypothesis

Because of the p*K*a decreasing tendency of both C8 and N1 upon N7 alkylation, a special type of C8-N7 cyclic adduct[19] can be formed (as shown below, Figure 1.7):

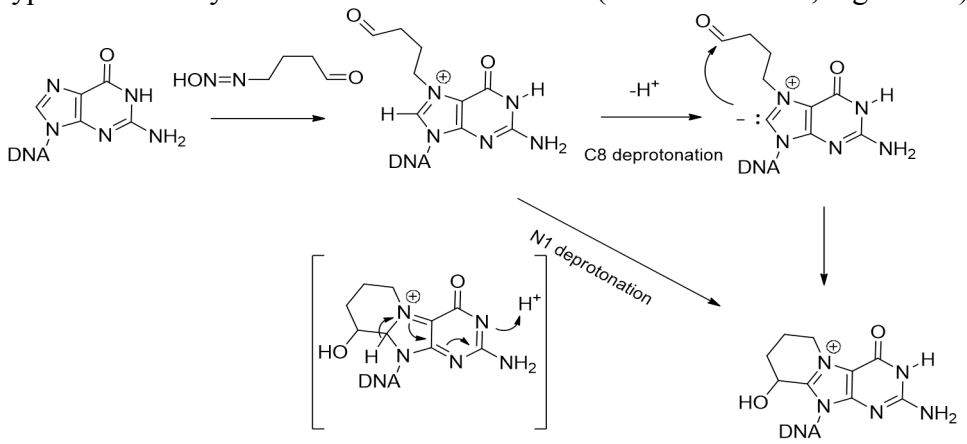


Figure 1.7 C8-N7 cyclic adduct

Another well known product for N7-alkyl G is the formation of formamidopyrimidine guanine (FAPy G, Figure 1.8):

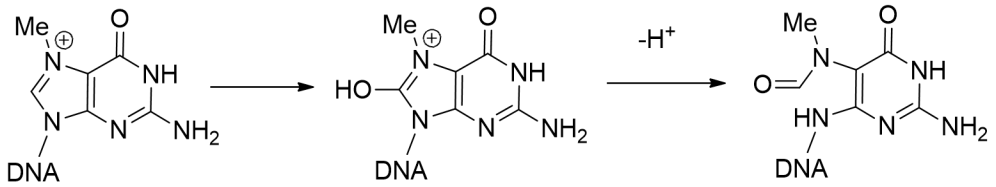


Figure 1.8 FAPy G formation

As shown in the figure above, FAPy G usually forms when the C8 of N7-alkyl G is further covalently bonded to hydroxyl group. Then the 5 member imidazolium ring would be opened to form the formamidopyridinic product. The formyl group is confirmed to be associated with N7[20], probably because that the pyrimidine ring with unsaturated carbonyl group can serve to delocalize and stablize the transition state charges developed during the ring opening. The bond between N7-C5, N9-C1' and N7-C8 (used as the same numbering system before ring opening) from the FAPy G can easily rotate according to the size of N7 alkylation, sugar and specific conformations in the DNA duplex, further creating the problems of mutagenicity and toxicity[21][22]. The formation of FAPy G can be facilitated by the presence of base[23].

One of the important reactions occurring after N7 alkylation of guanine in DNA duplex is the spontaneous depurination and possible backbone break, as shown below(Figure 1.9):

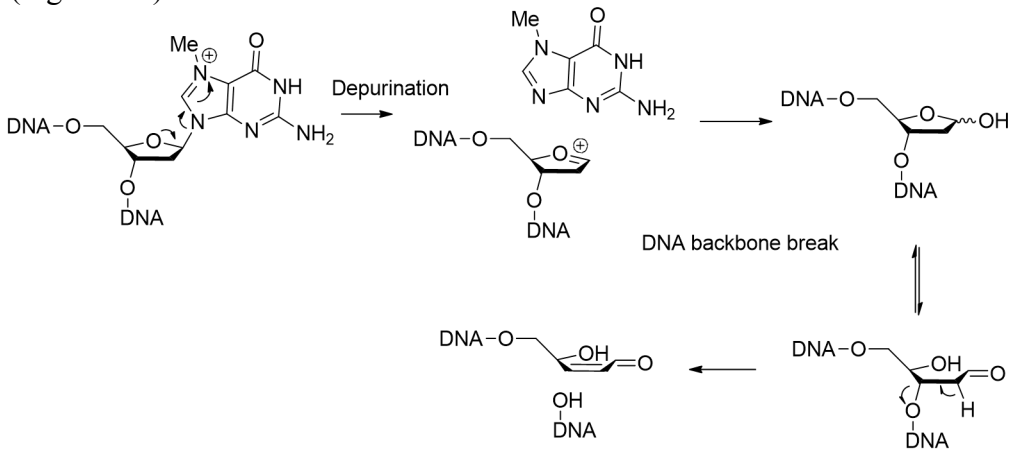


Figure 1.9 N7 methylation leading to auto-depurination and backbone break

The positive charge gained upon the N7 alkylation makes the total guanine base not a stable but a good leaving group for a S_N1 typed A_N+D_N (attack by nucleophile followed by departure of the nucleophile) reaction. The oxocarbenium ion can be converted easily by hydroxylation into cyclic acetal and deprotonated ring-opened aldehyde form, which could then undergo the elimination breaking the downstream linkage of the DNA phosphate backbone[24]. Acid can facilitate the depurination while mild base can facilitate the strand break from the abasic site[25][26]. Either abasic site or backbone break can easily serve as the potential spots for future mutagenicity, further adding the carcinogenesis risk to the N7 alkylation.

Though this is a spontaneous process for N7-alkyl G, half life studies suggest that the rate only speeds up to a significant level when the N7 substitution group is very bulky and electron rich. In other words, specifically for N7 methylation, the auto-depurination process is slow: half life remains as long as 10 days at room temperature[27] and more than a week upon physiological conditions[28] in DNA duplex. This has created a large time window for N7 methyl guanine (N7meG) to impact on DNA or related bio-activity *in vivo*.

However, to explore those unknown impacts remains hard, for the simple reason that N7meG is not stable enough to withstand the acidic and basic conditions to be put sequence specifically into the DNA during the oligo synthesis. Thus it had been imposing the difficulty for the related mechanism studies of N7meG, especially for those relying on the structural method, such as X-ray crystallography. To overcome this, there have been several methods[29], and the principle of which is to add the electron withdrawing group onto the 2' position of the sugar to strengthen the glycosidic bond. It has been confirmed that one of these methods which utilizes the 2'-fluorine (2'-F) group can produce reliable structural result[30], since the introduction of 2'-F mimics that of the hydrogen and it does

not change the sugar pucker conformation. The addition of 2'-F on the sugar destabilizes the oxocarbenium ion transition state, as shown below (Figure 1.10), and thus prevents the N7meG from spontaneous depurination, even upon harsh conditions during DNA oligo synthesis[14]:

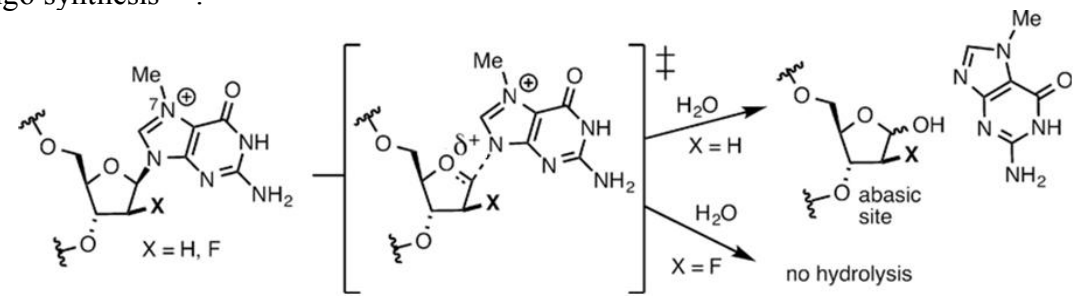


Figure 1.10 Transition state destabilization of N7meG by fluorination

Summarily, N7meG predominates in the nonenzymatic methylation of DNA. However, the effect of intact N7meG on DNA replication has remained elusive. As mentioned above, although N7meG undergoes spontaneous depurination, the lesion in duplex DNA is quite stable that it can enjoy the half life of around a week and persistently exist in genomic DNA at the level of several adducts per 10 million bases[31]. Additionally, as shown above, the N7-methylation decreases the p*K*a of N1 to a value of around 7[17][32], which might facilitate the formation of pseudo-Watson-Crick N7meG:T mismatch base pair during DNA replication by taking the zwitterionic form [17][33]. If this occurs *in vivo*, then it could promote G to A transition mutations, which will make N7meG a promutagenic lesion. Additionally, there have been only limited reports that describe the effects of intact N7meG on biological processes: i) the presence of N7meG in methylated CpG has been shown to preclude the binding of methyl-CpG-binding domain 1[34], and ii) N7meG in DNA inhibits binding of major-groove-interacting proteins[35]. Despite these, a detailed systematic study on the effect of N7meG on DNA replication has not been performed so far. It would be of great significance because such

study will also provide insights into the effects of many other N7 guanine alkylation adducts on DNA replication and mutagenesis.

To obtain this goal, the initial and crucial step is to see the situation of this N7meG lesion during the DNA replication. The essential enzyme in the DNA replication is the polymerase. During each cell division, DNA polymerase is required to duplicate the DNA, so that a copy of the original genetic information can be passed to each of the daughter cells. Based on sequence homology, DNA polymerases can be divided into 7 different families: A, B, C, D, X, Y, and RT. In this study, human DNA polymerase β (pol β) from X family is chosen as a model polymerase. The X family of DNA polymerases in eukaryotic cells consists of terminal deoxynucleotidyl transferase, and DNA polymerases β, λ, and μ. All these enzymes lack proofreading function, so their intrinsic fidelity is much lower than a normal polymerase that is in charge of replication. They all share 8 kDa domain, with the 31 kDa polymerase domain further divided into thumb, palm, and fingers subdomains. They also share two Helix-hairpin-Helix motifs that function in DNA-polymerase interactions. One resides in the 8 kDa subdomain, interacting with the downstream DNA template. The other is located in the thumb subdomain, interacting usually with the primer strand. The palm subdomain has 3 highly conserved aspartic acid residues, crucial for the nucleotidyl transfer chemistry in the active site. The fingers subdomain has residues crucial for nucleotide interaction. Thus, as a member from X family, DNA polymerase β is also an error-prone polymerase lacking an intrinsic proofreading 3' to 5' exonuclease activity. Polβ participates actively in base-excision DNA repair (BER) pathway. This is a DNA repair pathway that is responsible for lesions caused by single base modifications usually from reactive oxygen species or alkylators. It is responsible for repairing at least 20,000 endogenous lesions per cell per day[36]. Polβ does this by filling a short nucleotide gap generated by DNA

glycosylase and AP endonuclease. After the gap is filled, the nicked DNA duplex is then sealed by DNA ligase. Additionally and importantly in this study, pol β has been implicated in the translesion synthesis (TLS) of various types of DNA damage such as cisplatin induced intra-strand cross-link and bulky UV-induced cyclobutane TT dimers[37]. This, plus that a great amount of kinetic and structural data already obtained for pol β [38][39], and its smallest molecular weight among all DNA polymerases (39 kDa), makes it an ideal human DNA polymerase model for the structural and kinetic exploration of nucleotidyl transfer chemistry and substrate discrimination mechanisms concerning the N7meG lesion.

1.2 EXPERIMENTAL PROCEDURE

Synthesis of the N7meG phosphoramidite

The N7meG phosphoramidite is prepared starting from a ribose derivative according to the synthetic procedures[30]. Briefly, as shown below (Figure S2.1), the sugar was first fluorinated by diethylamino sulfur trifluoride (DAST) followed by bromination to give the bromo compound 2. N-Glycosidation of the bromide 2 with the 2-amino-6-chloropurine with NaH in acetonitrile generates mainly the desired β -anomer 3, which was then installed with the 6-benzyloxy group and with the deprotection of 3' and 5' hydroxyl groups. Phenoxyacetylation ($(Pac)_2O$) of 4 under conditions of transient TMS (trimethylsaline) protection protects the exocyclic amine, which is then followed by reductive removal of the benzyl moiety to afford 5. Regioselective alkylation of 5 with iodomethane (MeI) provided 2'-deoxy-2'-fluoro-N7-methyl guanosine 6 in satisfying yield. Standard tritylation of the 5'-OH followed by phosphitylation of the 3'-OH produced the desired N7meG phosphoramidite.

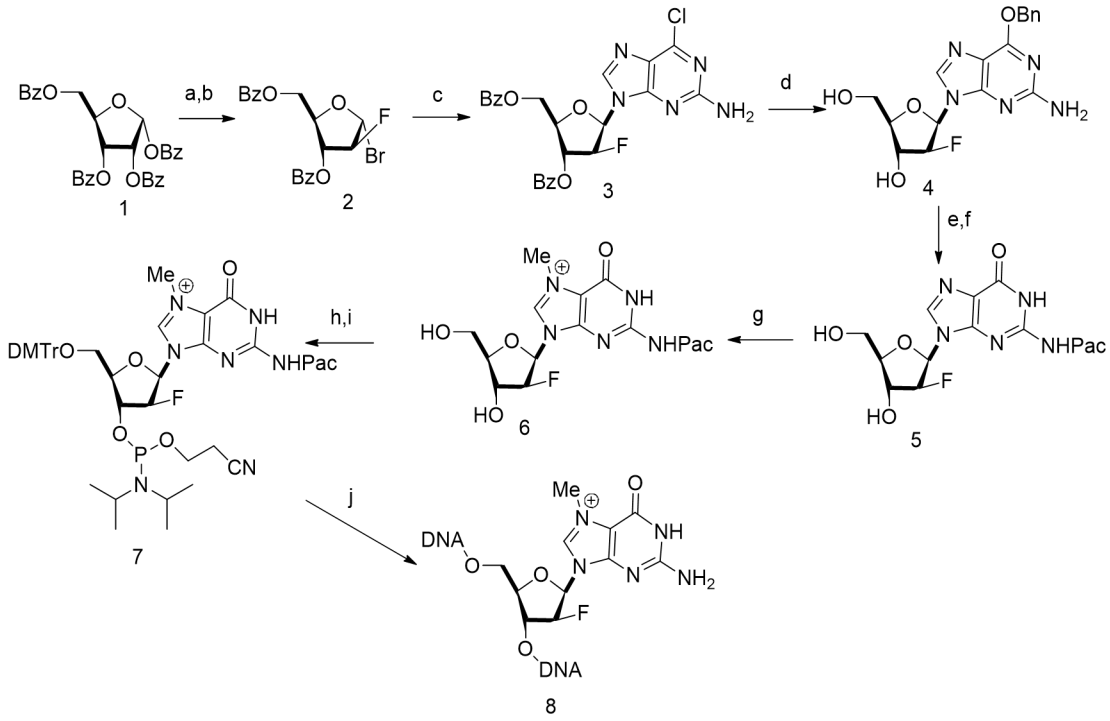


Figure 1.11 Synthesis of N7meG amidite

Reagents and conditions: (a) DAST, DCM (Dichloromethane), reflux, 15 h, 90%; (b) HBr (hydrobromide), AcOH (Acetic acid), 25˚C 24 h, 98%; (c) 2-amino-6-chloropurine, NaH, MeCN (Acetonitrile), 50˚C 6 h, 65%; (d) BnOH (Benzyl alcohol), NaH, DMF (Dimethylformamide), 25˚C, 75%; (e) (i) TMSCl, pyridine; (ii) $(Pac)_2O$; (iii) $NaHCO_3$ 90%; (f) Pd/C, H_2 , EtOH/THF, 25˚C, 98%; (g) MeI, DMF, 25˚C, 95%; (h) DMTrCl (Dimethoxytrityl chloride), pyridine, 70%; (i) 4,5-dicyanoimidazole, 2-cyanoethyltetraisopropyl phosphordiamidite, DCM, 25˚C, 1 h, 70%; (j) solid phase DNA synthesis.

DNA sequences used for kinetic and X-ray crystallographic studies

The oligonucleotides for kinetic assays (upstream primer, 5'-FAM /GGGGGCTCGTAAGGATTC-3', downstream primer, 5'-phosphate /AGTCGG-3' and

template, 5'-CCGACT(X)GAATCCTTACGAGCCCCC-3', X = dG, FdG or N7meG) were synthesized by Midland Certified Reagent Co. (Midland, TX, USA) and Integrated DNA Technologies (Coralville, IA, USA). The template DNA was prepared via solid-phase DNA synthesis using ultramild deprotection conditions (50 mM K_2CO_3 in MeOH, 25˚C, 24h). The DNA sequences used for co-crystallization were 5'-CCGAC(N7meG)TCGCATCAGC-3' for template, 5'-GCTGATGCGA-3' for the upstream primer and 5'-phosphate/GTCGG-3' for the downstream primer. The oligonucleotides were annealed to give a single-nucleotide gapped DNA, which was used for the crystallization of both binary and ternary polβ complex structures[38].

Expression and purification of polβ

Polβ used for kinetic and crystallographic studies was expressed and purified from *Escherichia coli* as described previously[39].

Steady state kinetics

Steady-state kinetic parameters for dCTP insertion opposite templating dG, 2′-fluoro-2′-deoxyguanosine (FdG) and N7meG by polβ were determined using the conditions described previously[39]. DNA substrates containing a single-nucleotide gap opposite the templating dG, FdG and N7meG were prepared by annealing the template oligonucleotide (5′-CCGACT(X)GAATCCTTACGAGCCCCC-3′; X = dG, FdG and N7meG) with the upstream (5′-FAM/GGGGGCTCGTAAGGATTC-3′) and the downstream primers (phosphate/AGTCGG-3′) at 95°C for 3 min followed by slow cooling to 4°C. Recessed DNA substrates that do not have the downstream primer were prepared similarly. Polβ activities were determined using the reaction mixture containing 50 mM Tris-HCl pH 7.5, 5 mM $MgCl_2$, 100 mM KCl, 80 nM substrate DNA and varying

concentrations of the incoming nucleotides. The nucleotidyl transfer reactions were initiated by adding polβ and incubated at 37°C for 2 min, and quenched by adding a stop solution containing 95% formamide, 45 mM Tris-borate, 20 mM ethylenediaminetetraacetic acid, 0.1% bromophenol blue and 0.1% xylene cyanol. The nucleotidyl transfer reaction products were resolved on 18–20% denaturing urea polyacrylamide gels, and the product formation was analyzed using a PhosphorImager. The efficiency of nucleotide insertion was calculated as k_{cat}/K_m. The relative efficiency of dCTP incorporation opposite N7meG was determined as $f = (k_{cat}/K_m)_{[dCTP:N7meG]}/(k_{cat}/K_m)_{[dCTP:dG]}$.

Cocrystallization and structure determination

The binary and ternary polβ complexes with templating N7meG were crystallized using the similar conditions described previously[38]. Diffraction data were collected at 100 K at the beamline 5.0.3 at the Advanced Light Source, Lawrence Berkeley National Laboratory. All diffraction data were processed using HKL2000. Structures were solved by molecular replacement using a gapped binary complex structure with an open conformation (PDB ID: 1BPX) and a ternary complex structure with a closed conformation (PDB ID: 1BPY) as the search models. The model was built using COOT[156] and refined using Phenix[157].

1.3 EXPERIMENTAL RESULTS AND DISCUSSIONS

Steady state kinetics results

To assess whether N7meG at the templating position affects the polymerase activity of polβ, the kinetic parameters for the polβ-catalyzed insertion of dCTP and

dTTP opposite the templating dG, FdG and N7meG (Table1, shown below) are determined. The efficiencies for dCTP insertion opposite the templating FdG and dG in the single-nucleotide gapped DNA are almost the same, indicating that 2′-fluorination has a minimal effect on polβ catalysis. In contrast, the presence of the templating N7meG increases K_m ~10-fold and decreases k_{cat} ~30-fold compared to those of dG, thereby reducing the relative efficiency for dCTP insertion opposite N7meG by ~300-fold (Table1). The sharp decrease in the relative efficiency is surprising, because the N7-methyl moiety perturbs neither the Watson–Crick H-bonding nor the minor-groove edges of dG. To evaluate whether the p*K*a of N7meG contributes to the decrease in the relative efficiency, we determined kinetic parameters for dCTP insertion opposite N7meG at different pH conditions (pH 6.5 and 8.5). The relative efficiency increases only ~2-fold when pH changes from 6.5 to 8.5, indicating that the p*K*a of N7meG does not significantly contribute to the observed reduction in the relative efficiency. In addition to the N7meG:dCTP kinetics, the effect of the N7-methylation on the formation of G:T mismatches was evaluated, which comprise about 60% of polβ-induced spontaneous mutations[40], by determining kinetic parameters for dTTP insertion opposite N7meG. Surprisingly, the dTTP incorporation was not observed to the level of significant, suggesting that the templating N7meG greatly deters the dTTP insertion opposite the lesion in the active site of polβ.

Table1.1: steady state kinetic parameters

Template:dNTP	pH	K_m (μM)	k_{cat} (10^{-3} s^{-1})	k_{cat}/K_m (10^{-3} s^{-1}μM^{-1})	f [a]
dG:dCTP[b]	7.5	269.55 ± 18.87	7.56 ± 0.38	0.03	
dG:dCTP[c]	7.5	0.59 ± 0.03	20.38 ± 0.50	34.54	1
FdG:dCTP[c]	7.5	0.34 ± 0.07	12.52 ± 0.53	36.82	1.1
N7meG:dCTP[b,d]	7.5	-	-	-	-
N7meG:dCTP[c]	6.5	3.86 ± 0.09	0.40 ± 0.06	0.1	2.9×10^{-3}
N7meG:dCTP[c]	7.5	6.01 ± 0.75	0.69 ± 0.03	0.11	3.2×10^{-3}
N7meG:dCTP[c]	8.5	3.87 ± 0.07	0.75 ± 0.02	0.19	5.5×10^{-3}
N7meG:dTTP[c,d]	7.5	-	-	-	-

[a]Relative efficiency:$(k_{cat}/K_m)_{[dNTP:N7meG]}/(k_{cat}/K_m)_{[dCTP:dG]}$.

[b]A recessed DNA (i.e. DNA without the downstream primer) was used as a substrate.

[c]Single-nucleotide gapped DNA was used as a substrate.

[d]Nucleotide incorporation was not observed or minimal.

Data collection and refinement statistics

Table 1.2: data collection and refinement statistics

PDB code	N7meG binary (4O5C)	N7meG:C ternary (4O5K)	N7meG:T–Mg^{2+} ternary (4O5E)	N7meG:T–Mn^{2+} ternary (4P2H)
Space group	***P*2$_1$**	***P*2$_1$**	***P*2$_1$**	***P*2$_1$**
Cell constants				
a (Å)	54.396	50.87	54.425	54.767
b	79.758	79.568	78.92	79.167
c	54.83	55.67	54.819	54.982
α (°)	90	90	90	90
β	105.54	107.71	105.97	106.18

Table 1.2: (continued)

γ	90	90	90	90
Resolution (Å)[a]	20–2.37	20–2.06	20–2.54	20–1.99
	(2.41–2.37)	(2.10–2.06)	(2.58–2.54)	(2.02–1.99)
$\langle I/\sigma \rangle$	15.4 (2.66)	15.8 (1.86)	9.7 (1.52)	22.2 (2.52)
Completeness (%)	99.2 (97.1)	98.8 (96.6)	99.7 (97.6)	100 (100)
R_{merge}[b] (%)	10.6 (47.6)	8.2 (52.0)	13.5 (61.1)	8.6 (44.5)
Redundancy	5.5 (5.0)	3.7 (3.5)	3.4 (3.0)	5.6 (5.6)
R_{work}[c]/R_{free}[d] (%)	20.02/24.54	20.03/25.16	20.54/27.28	20.40/24.95
Unique reflections	17819	24650	13821	30969
Mean *B* factor (Å^2)				
Protein	26.2	23.9	34.7	29.1
Ligand	25.2	32.8	34.8	34.4
Solvent	24.2	28.3	31.5	28.9
Ramachandran plot				
Most favored (%)	97.2	98.5	96	97.5
Add. allowed (%)	2.8	1.5	4	2.5
r.m.s.d.				
Bond lengths (Å)	0.003	0.004	0.005	0.005
Bond angles (º)	0.822	1.124	1.089	1.219

[a] Values in parentheses are for the highest resolution shell.

[b] $R_{merge} = \Sigma|I-\langle I \rangle|/ \Sigma I$ where I is the integrated intensity of a given reflection.

[c] $R_{work} = \Sigma|F(obs)-F(calc)|/\Sigma F(obs)$.

[d] $R_{free} = \Sigma|F(obs)-F(calc)|/\Sigma F(obs)$, calculated using 5% of the data.

Binary structure of polβ in complex with DNA bearing a single-nucleotide gap opposite N7meG

To evaluate whether N7meG at the templating position perturbs the conformation of DNA and protein, a binary structure of polβ bound to DNA containing a single-nucleotide gap opposite the templating N7meG is solved (Figure 1.12A; see Table 1.1 for refinement statistics). The gapped binary structure, refined to 2.4 Å resolution, indicates that whereas N7meG does not induce any substantial distortion of the polβ conformation, it induces a slight conformational change of DNA near the lesion (Figure 1.12B). The overall structure of the N7meG gapped binary complex is very similar to that of a published dG gapped structure (PDB ID: 1BPX, r.m.s.d.= 0.619 Å, Figure 1.12E)[38]. Polβ assumes an open protein conformation. The α-helix N bearing the minor-groove recognition amino acids, Asn279 and Arg283, does not engage in H-bonding interactions with DNA. The templating N7meG adopts an *anti*-base conformation and C2′-endo sugar pucker, as does templating dG (Figure1.12C and D). Although the 2′-F is evident on the electron density map, the N7-methyl group is not, indicating the flexibility of templating N7meG when there is no incoming nucleotide. Although similar to the normal dG binary structure, N7meG at the templating position still induces a minor conformational change at the 5′ and the 3′ sides of the lesion (Figure 1.12E). The 3′-OH of the primer terminus moves 1.7 Å toward the templating base relative to the position observed in published gapped structure (Figure 1.12E). The base moiety of templating N7meG slightly rotates toward the minor groove and the 5′ side of it shifts ~1 Å from the position observed in the templating dG-containing structure (Figure 1.12F). Overall, N7meG gapped binary complex structure indicates that templating N7meG induces a slight conformational change of DNA, especially around the lesion site.

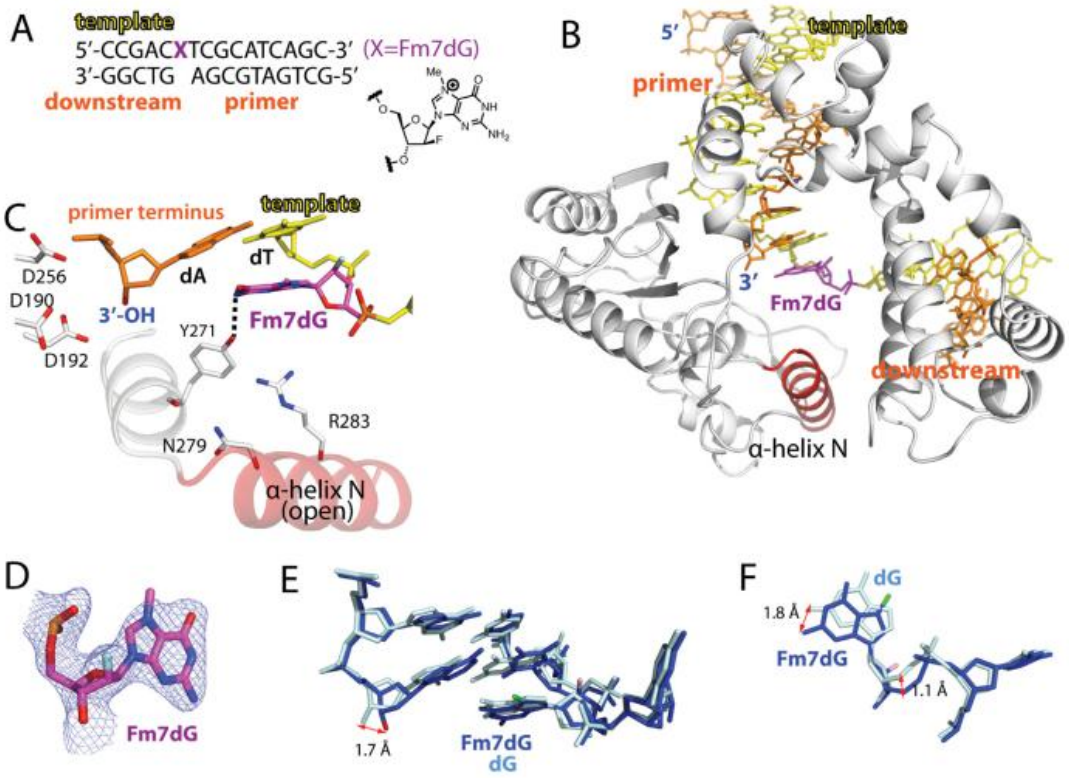


Figure 1.12[30]. Single-nucleotide gapped binary structure of polβ in complex with DNA containing templating N7meG (PDB ID: 4O5C). (A) The DNA sequence used for crystallography and chemical structure of N7meG are shown. The downstream primer contains 5'-phosphate. (B) Overall crystal structure of the N7meG gapped binary complex. Polβ is shown in white. The α-helix N bearing the minor groove recognition amino acid residues Asn279 and Arg283 is shown in red. The template strand is shown in yellow and the upstream and the downstream primers are shown in orange with templating N7meG indicated. (C) Active site structure. The α-helix N is in an open conformation. Y271 side chain is H-bonded to amine of N7meG (D) A 2Fo-Fc electron density map is contoured at 1σ around N7meG. Templating N7meG adopts the 2'-endo sugar pucker and the *anti*-base conformation. (E) Comparison of the N7meG gapped structure (blue) with published dG gapped structure (cyan, PDB ID: 1BPX). Distance

Figure 1.12[30]. Single-nucleotide gapped binary structure of polβ in complex with DNA containing templating N7meG (PDB ID: 4O5C).

between the 3'-OHs of the primer terminus is indicated. (F) Comparison of templating N7meG (blue) with dG (cyan) in (E).

Ternary structure of polβ incorporating dCTP analog opposite templating N7meG

To gain insight into the observed slow insertion of dCTP opposite N7meG by the enzyme and to elucidate the base-pairing properties of N7meG:dCTP in the enzyme active site, a ternary structure of polβ with templating N7meG base-paired with the incoming nonhydrolyzable dCTP analog dCMPNPP (dCTP* hereafter) is solved. The nonhydrolyzable dCTP analog contains a bridging N–H group, which makes the molecule resistant to nucleotidyl transfer and hydrolysis. The use of dNMPNPP enables a structural determination of the pre-chemistry state of the polymerase ternary complex with the catalytic metal ion coordinated to the primer terminus 3′-OH. The active site coordination of dNMPNPP is indistinguishable from that of dNTP and has been used in structural studies of various DNA polymerases.

The polβ-N7meG:dCTP* ternary complex structure, refined to 2.1 Å resolution, shows that N7meG:dCTP* base pair is well accommodated in the polβ active site (Figure 1.13). The overall structure of the N7meG:dCTP* ternary complex (Figure 1.13A) is essentially identical to that of published ternary structure with correct insertion (PDB ID: 2FMS[41], r.m.s.d. = 0.399 Å, Figure 1.13D). In the N7meG:dCTP* complex, polβ undergoes an open-to-closed conformational change, with α-helix N moving ~10 Å from the position in the N7meG gapped binary complex to sandwich the N7meG:dCTP* base pair between the primer terminus base pair and α-helix N (Figure 1.13B). Similar to the

observed in published polβ ternary structure with the correct insertion, Tyr271, Asn279 and Arg283 engage in H-bonding interactions with the minor groove edges of the primer terminus base, the incoming dCTP* and the templating base, respectively.

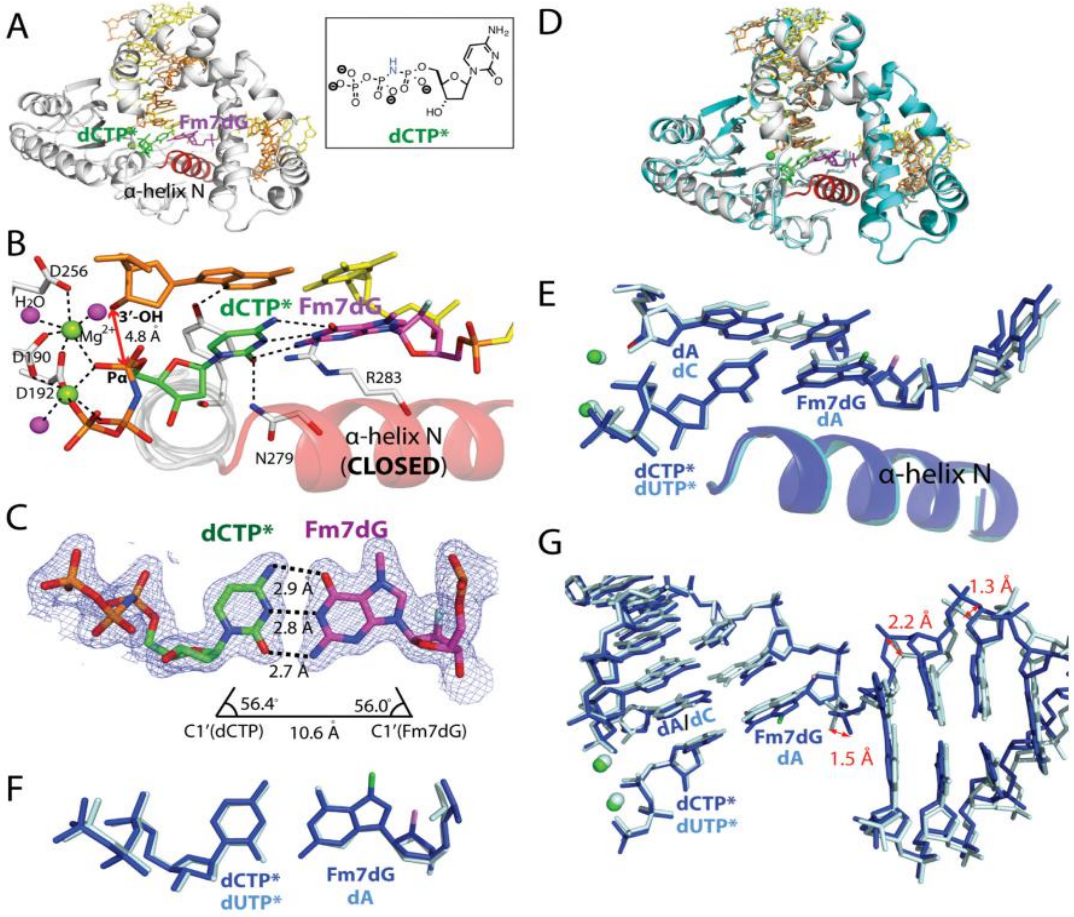


Figure 1.13.[30] Ternary structure of polβ in complex with templating N7meG paired with an incoming nonhydrolyzable dCTP analog (PDB ID: 4O5K). (A) Overall structure of the N7meG:dCTP* complex. The protein adopts the closed conformation and the nascent base pair adopts a coplanar conformation. (B) Close-up view of the active-site structure. The α-helix N is in a closed conformation. Key interactions are indicated as dotted lines. Active-site Mg^{2+} ions are shown as green spheres, and water molecules are shown as magenta spheres. 3′-OH of the primer terminus is not coordinated to the catalytic metal

Figure 1.13.[30] Ternary structure of polβ in complex with templating N7meG paired with an incoming nonhydrolyzable dCTP analog (PDB ID: 4O5K).

ion. The distance between 3′-OH of the primer terminus and Pα of the incoming nucleotide is indicated. (C) Base-pairing mode of N7meG:dCTP*. A 2Fo − Fc electron density map is contoured at 1σ around N7meG and dCTP*. The H-bonding distances, the C1′ − C1′ distance and λ angles are very similar to those observed in a polβ structure with the correct insertion (PDB ID: 2FMS). (D) Overlay of the N7meG:dTTP*–Mg^{2+} ternary structure with published polβ ternary structure (PDB ID: 2FMS) with nascent dA:dUTP* base pair (r.m.s.d.= 0.400Å) (E) Overlay of the active site of the N7meG:dTTP*–Mg^{2+} structure (blue) and published dA:dUTP* ternary structure (cyan). (F) Top view of the N7meG:dTTP* base pair (blue) and the dA:dUTP* base pair (cyan) in (E). (G) Overlay of DNA in the N7meG:dTTP*–Mg^{2+} structure (blue) and published dA:dUTP* ternary structure (cyan).

The polβ-N7meG:dCTP* structure reveals the base-pairing features of N7meG:dCTP* in the confines of the enzyme's active site. The base-pairing nature of N7meG:dCTP* is quite different from that of N7meG:dC observed in published AlkA-N7meG:dC structure. Unlike a relaxed Watson–Crick N7meG:dC base pair with an average H-bond distance of 3.4 Å observed in the AlkA-N7meG:dC structure[30], the N7meG:dCTP* base pair adopts canonical Watson–Crick conformation with an average H-bond distance of 2.8 Å (Figure 1.13C). The base pair geometry such as the C1′-C1′ distance and the λ angles of N7meG:dCTP* is essentially identical to that of canonical Watson–Crick base pair (Figure 1.13C, E and F). The difference between the base-paring properties of N7meG:dCTP* and N7meG:dC (AlkA) appears to result from the absence and the presence of constraints by protein. Whereas the N7meG:dC base pair in the AlkA-N7meG:dC complex is devoid of any interaction with protein, the N7meG:dCTP*

base pair in the polβ-N7meG:dCTP* complex is now confined in the polβ active site that discriminates between correct and wrong nucleotides using the rigid geometric selection mechanism. The nascent base pair binding pocket with its rigid geometric constraints would prevent the formation of the relaxed base pair conformation as observed in the AlkA-N7meG:dC structure and instead, promote the formation of canonical Watson–Crick N7meG:dCTP* base pair conformation.

Although the N7meG:dCTP* ternary complex shares the characteristics of the polβ ternary complex with the correct insertion, the structure indicates that the active-site conformation of this complex is not very optimal for nucleotidyl transfer reaction (Figure 1.13B). Specifically, the distance between Pα of dCTP* and the 3′-OH of the primer terminus is ~1.4 Å longer than that for correct insertion[41] (4.8 Å versus 3.4 Å, Figure 1.13B). Additionally, the primer terminus 3′-OH is not positioned for an ideal in-line nucleophilic attack on the Pα of the incoming nucleotide. Furthermore, the critical coordination of the 3′-OH of the primer terminus to the catalytic metal ion[42] is missing in the N7meG:C structure. Instead, an ordered water molecule is liganded to the catalytic metal ion. The coordination of the 3′-OH to the catalytic metal ion has been suggested to lower the p*K*a of the 3′-OH[43] and facilitate proton transfer from the 3′-OH to a catalytic carboxylate residue Asp256[44]. Previous molecular dynamics studies have indicated that nucleophilic attack of the non-coordinated 3′-OH toward the Pα of an incoming nucleotide is energetically unfavorable for polβ catalysis[42]. The N7meG:dCTP* ternary structure thus indicates that, despite its ability to form the canonical Watson–Crick base pairing with the incoming dCTP, the presence of N7meG at the templating position may slow polβ catalysis across the lesion by perturbing the coordination of the catalytic metal ion, as indicated in Table 1.0. Also, the lack of coordination of the primer terminus 3′-OH

to the catalytic metal ion likely contributes to the observed reduced binding of dCTP opposite the templating N7meG (Table 1.0). The N7meG:dCTP* ternary complex will require a subtle conformational adjustment of the active site to reach the ideal catalytical state.

There is some distortion of the 3′ end of the primer terminus, which is ~12 Å away from the N7-methyl moiety. dCTP* is unlikely to cause the perturbation at the primer 3′ end since the conformation of the incoming dCTP* is nearly indistinguishable from that of an incoming nucleotide in the published structure with the correct insertion (PDB ID: 2FMS[41], Figure 1.13F). A comparison of the polβ-N7meG:dCTP* complex and the published polβ-dA:dUTP* complex (PDB ID: 2FMS) reveals a significant difference in the conformation of the downstream primer base pairs (Figure 1.13G), suggesting that the presence of N7meG at the templating position induces a relatively considerable conformational change of DNA. Polβ may probably react to such conformational difference and thus takes a catalytically suboptimal conformation for dCTP incorporation opposite N7meG. The perturbation of the 3′ end of the primer terminus has been also observed with published polβ structures with incorrect insertions, showing a longer distance between primer terminus 3′-OH and the catalytic metal ion (e.g. dC:dATP (PDB ID: 3C2L) and dA:dGTP (PDB ID: 3C2M), 4.4 and 4.5 Å, respectively[45]). Mismatched structures also show a staggered base pair conformation and an upstream shift (~3 Å) of template strand. Overall, the N7meG:dCTP* and published mismatched structures illustrate that polβ is highly sensitive to the presence of the structural change within the enzyme active site and it would compensate by taking an alternate and suboptimal conformation for chemistry in the presence of active site distortion.

Ternary structure of pol β incorporating dTTP analog opposite N7meG in the presence of Mg^{2+}

As mentioned above, previous studies suggested that N7meG may be a promutagenic lesion that can base pair with dTTP through an ionized form, which will lead to the formation of pseudo-Watson–Crick N7meG:dTTP base pair[17][33]. To investigate whether a templating N7meG promotes mutagenic replication by forming such pseudo-Watson–Crick base pairing with dTTP in the active site of a DNA polymerase, the X-ray structure of polβ in complex with templating N7meG base-paired with incoming nonhydrolyzable dTTP analog dTMPNPP (dTTP* hereafter, Figure 3) in the presence of the active-site Mg^{2+} is solved.

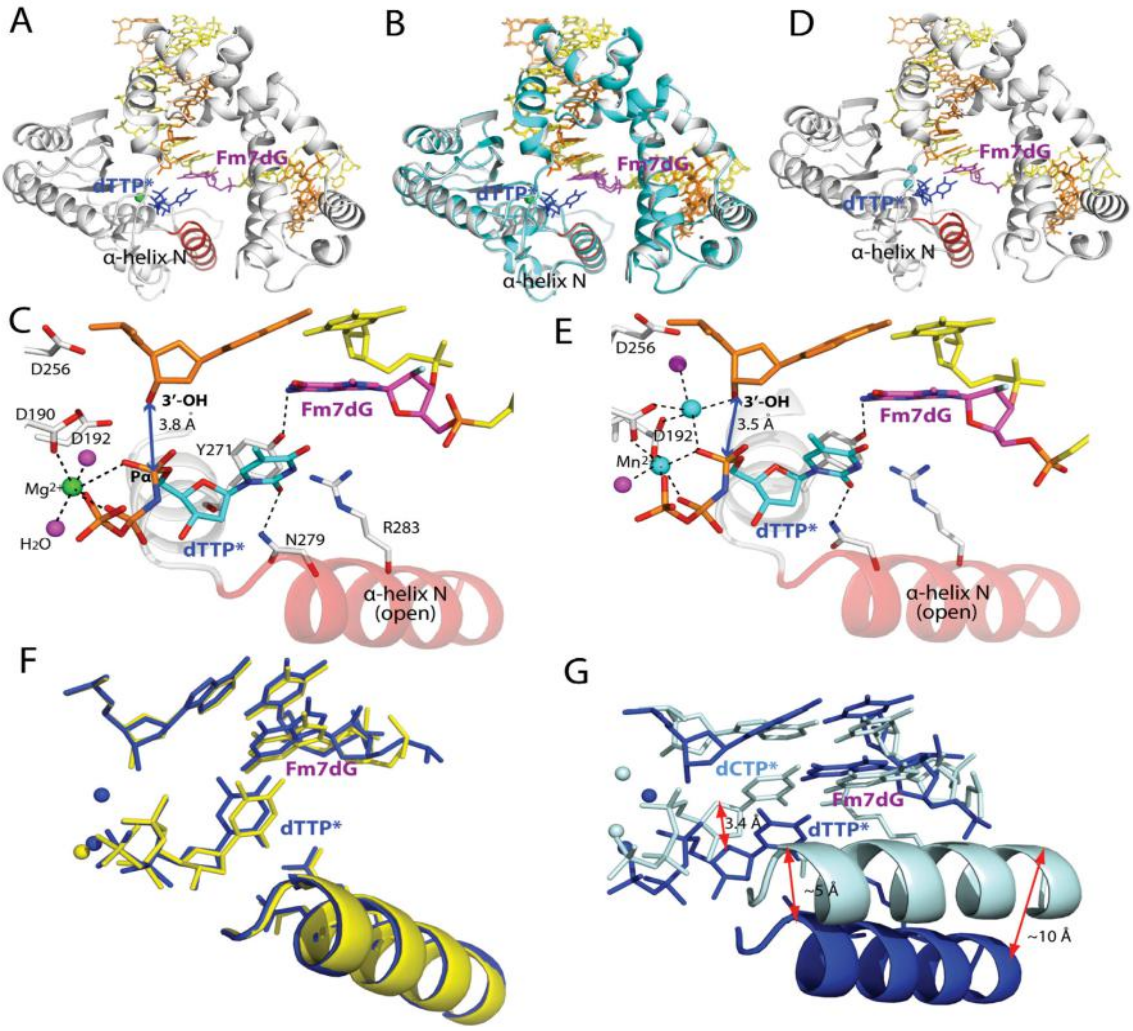


Figure 1.14[30]. Ternary structure of pol β with templating N7meG paired with an incoming nonhydrolyzable dTTP analog (PDB ID: 4O5E and 4P2H). (A) Overall structure of the N7meG:dTTP*-Mg^{2+} complex. (B) Overlay of the N7meG:dTTP*-Mg^{2+} structure (white) with the N7meG gapped binary structure (cyan) (r.m.s.d.= 0.316 Å). (C) Close-up view of the active-site structure of the N7meG:dTTP*-Mg^{2+} complex. N7meG does not H-bond with the incoming dTTP*. Only the nucleotide-binding metal ion is observed. (D) Overall structure of the N7meG:dTTP*-Mn^{2+} complex. (E) Active-site structure of the N7meG:dTTP*-Mn^{2+} complex. Both the nucleotide-binding and the

Figure 1.14[30]. Ternary structure of pol β with templating N7meG paired with an incoming nonhydrolyzable dTTP analog (PDB ID: 4O5E and 4P2H).

catalytic metal ions are present. (F) Comparison of the active site structure of the N7meG:dTTP*-Mn^{2+} complex (blue) with that of the N7meG:dTTP*-Mg^{2+} complex (yellow). (G) Comparison of the active site structure of the N7meG:dTTP*-Mn^{2+} complex (blue) with that of the N7meG:dCTP*-Mg^{2+} complex (cyan).

The N7meG:dTTP*-Mg^{2+} ternary structure, solved to 2.5 Å resolution, is significantly different from the N7meG:dCTP*–Mg^{2+} ternary structure (r.m.s.d. = 1.367Å). The N7meG:dTTP*–Mg^{2+} ternary complex adopts a catalytically incompetent state with an open protein conformation, a staggered base pair, and one active-site metal ion (Figure 1.14A), these are all in contrast to the N7meG:dCTP* ternary complex with its closed protein conformation, coplanar base pair and two active-site metal ions. The protein and DNA conformations observed in the N7meG:dTTP*–Mg^{2+} ternary structure are essentially identical to those of the N7meG gapped binary structure (Figure 1.14B, r.m.s.d. = 0.316 Å), indicating that binding of dTTP opposite templating N7meG does not induce any significant conformational change of protein and DNA. More specifically, in the mismatched N7meG:dTTP*–Mg^{2+} structure, pol β does not undergo an open-to-closed conformational activation, and the minor-groove recognition amino acid residues (Tyr271, Asn279 and Arg283) are with the same positions as observed in the N7meG gapped binary structure. In addition, the incoming dTTP* neither engages in H-bonding interactions with the templating N7meG nor it stacks with the primer terminus base. Furthermore, the active-site structure shows the binding only of the nucleotide-binding metal ion (Figure 1.14C) and the lack of coordination of Asp192 to the nucleotide-binding metal ion, which implies that even the binding of dTTP opposite N7meG in polβ active site is weak.

The N7meG:dTTP*–Mg^{2+} structure, which represents a rare pol β ternary structure with an open protein conformation and a single active-site metal ion, strongly suggests that polβ deters the misincorporation of dTTP opposite N7meG by adpoting an alternate conformation that is incompetent for catalysis. This catalytically unfavorable structure suggests that the N7meG:dTTP*–Mg^{2+} complex would have minimal accessibility to the catalytically competent state than the N7meG:dCTP complex and that, in the pol β active site, the formation of the promutagenic pseudo-Watson–Crick N7meG:T base pair is not favored.

Ternary structure of pol β incorporating dTTP analog opposite N7meG in the presence of Mn^{2+}

The N7meG:dTTP*–Mg^{2+} ternary structure most likely represents a ground-state structure that is unfavorable for nucleotdiyl transfer chemistry[42]. To gain insight into the pre-catalytic state of pol β inserting dTTP opposite the templating N7meG, a ternary structure of polβ-N7meG:dTTP* is determined in the presence of active-site Mn^{2+}, which is known to increase the misincorporation rate of DNA polymerase and facilitate the formation of a closed polβ conformation during misincorporation. The N7meG:dTTP*–Mn^{2+} ternary complex was refined to 2.0 Å resolution.

Even upon using the Mn^{2+} ion, the N7meG:dTTP*–Mn^{2+} ternary and the N7meG:dTTP*–Mg^{2+} ternary complex structurres are very similar, with both forming an open protein conformation and a staggered nascent base pair conformation (Figure 1.14D, r.m.s.d. = 0.274 Å). Even so, there is some structural difference concerning the metal ion-binding site. Unlike the N7meG:dTTP*–Mg^{2+} complex with only the nucleotide-binding metal ion, the N7meG:dTTP*–Mn^{2+} complex contains both the catalytic and the nucleotide-binding metal ions (Figure 1.14E). However, the catalytic metal ion is not

coordinated to Asp256, which is critical for pol β catalysis as aforementioned[44], but is instead coordinated to a water molecule (Figure 1.14E).

The structural similarity between the N7meG:dTTP*–Mg^{2+}/Mn^{2+} complexes indicates that the replacement of Mn^{2+} for Mg^{2+} does not facilitate an open-to-closed conformational transition of the N7meG:dTTP* complex (Figure 1.14F). The N7meG:dTTP*–Mn^{2+} structure adopts an open protein conformation that is observed in the N7meG gapped binary structure (r.m.s.d. = 0.316 Å), indicating the open-to-closed conformational activation is deterred in the presence of dTTP opposite templating N7meG in the active site of pol β. The huge structural differences between the N7meG:dTTP*–Mn^{2+} and the N7meG:dCTP*–Mg^{2+} complexes (Figure 1.14G) indicate that pol β strongly discriminates between N7meG:dTTP and N7meG:dCTP in the enzyme active site and that the templating N7meG does not form the promutagenic pseudo-Watson–Crick base pair with dTTP in the nascent base pair binding pocket of pol β. Consistent with the catalytically incompetent conformation observed in the N7meG:dTTP*–Mg^{2+}/Mn^{2+} complexes, the incorporation of dTTP opposite N7meG by pol β is minimal under various catalytic conditions, indicating that the presence of N7meG in DNA is unlikely to form the mutagenic replication by polβ.

Base pairing features of alkylated dG in the confines of the polβ active site

The structures from this study and recently published O6-methyl-dG (O6MedG) structures[46] provide insight into the base-pairing nature of modified dG in the pol β active site. If the zwitterionic form of N7meG exists as a major conformation in the enzyme active site, N7meG is likely to facilitate the formation of a wobble base pairing with dCTP* and a pseudo-Watson–Crick base pairing with dTTP* as shown below[46]. However, this N7meG structural study reveals that neither the wobble N7meG:dCTP*

nor the pseudo-Watson–Crick N7meG:dTTP* base pair forms in the enzyme's active site. Instead, the Watson–Crick N7meG:dCTP* and the staggered N7meG:dTTP* base pair conformations are observed, which indicates that the cationic form of N7meG rather than the zwitterionic form of N7meG exists as a major isomer in the active site of polβ. Compared to the bulkier and electron rich nucleobase, it is possible that this microenvironment around may reduce the effect of the N7-methyl moiety on the ionization of N1 of G, thereby disable the formation of the zwitterionic N7meG.

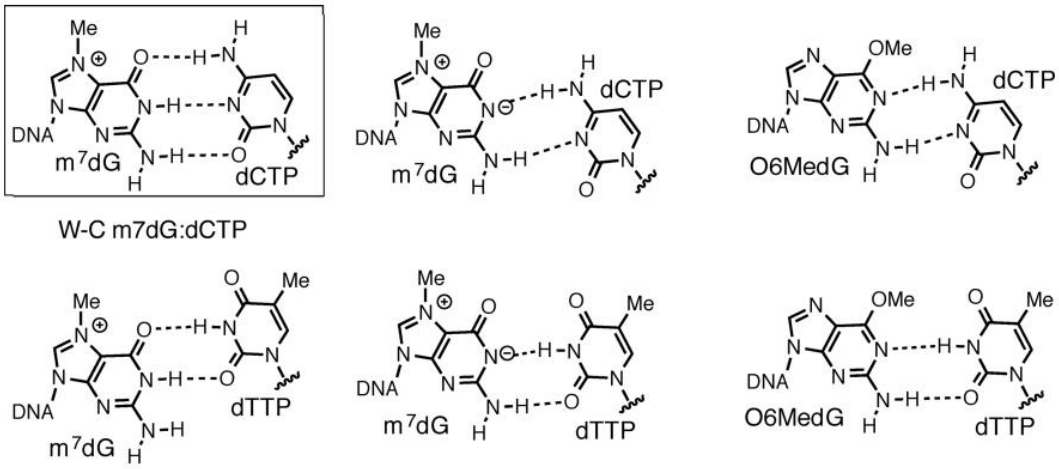


Figure 1.15 Possible N7meG base pairs with dCTP and dTTP

Published polβ-O6MedG:dCTP*/dTTP* structures show that the pseudo-Watson–Crick O6MedG:dTTP* base pair, but not the wobble O6MedG:dCTP* base pair, is accommodated in the enzyme active site. Instead, O6MedG:dCTP* forms a staggered base pair conformation (as shown above, Figure 1.15). The failure to observe the wobble base pair conformations of N7meG:dCTP*, N7meG:dTTP* and O6MedG:dCTP* in the enzyme's active site suggests that polβ strongly discriminates between base pairs with a canonical Watson–Crick geometry and a wobble geometry. Various DNA polymerases such as the A-family DNA polymerase *Bacillus stearothermophilus* DNA polymerase I large fragment[47], the B-family DNA polymerase RB69[48] and the Y-family DNA

polymerase polη[49] have been shown to accommodate wobble G:T or A:C base pair in their active sites. Unlike those polymerases, polβ does not allow the formation of the wobble conformations of A:C, N7meG:dCTP/dTTP and O6MedG:dCTP in the active site, highlighting a stringent geometric constraints of polβ. The observation of the pseudo-Watson–Crick base pair conformation for O6MedG:dTTP*, but not for N7meG:dTTP* also implies that the zwitterionic form of N7meG does not exist as a major conformation in the active site of polβ.

Insight into the replication fidelity of polβ

Comparison of the N7meG:dTTP*–Mn^{2+} structure with published polβ structures with base pair mismatch and active-site Mn^{2+} provides insight into the replication fidelity mechanism of polβ (Figure 1.16). Previous polβ structures with dC:dATP and dG:dATP mismatches show that it deters nucleotide misincorporation by inducing an alternate conformation that is suboptimal for catalysis (Figure 1.16A) [39][45]. Interestingly, the overall structure of our N7meG:dTTP*–Mn^{2+} mismatch complex is very different from those of published polβ ternary complexes with dC:dATP and dG:dATP mismatches (r.m.s.d. = 1.610 and 1.515 Å, respectively). Reported mismatched structures show a staggered base pair conformation, an upstream shift (~3 Å) of template strand near the active site, and a nearly closed protein conformation. Whereas the N7meG:dTTP*–Mn^{2+} complex also adopts a staggered base pair conformation, the template DNA strand does not shift significantly, and the protein remains in an open conformation (Figure 1.16A). So it seems that the binding of the incoming nucleotide in the polβ active site induces significant conformational change of protein and DNA for the dC:dATP and the dG:dATP mismatched ternary complexes, but not for the N7meG:dTTP* complex. The huge conformational differences observed among these

structures indicate that the insertion of dTTP opposite N7meG would be less favorable than the insertion of dATP opposite the templating dC or dG. Previous kinetic studies show that dATP incorporation opposite dC and dG by polβ is slow in the presence of active-site Mg^{2+}, but increases ~50-fold in the presence of Mn^{2+}[45]. In contrast to those studies, incorporation of dTTP opposite templating N7meG is not observed in the presence of either Mg^{2+} or Mn^{2+}, which is consistent with the catalytically incompetent conformation observed in the N7meG:dTTP*–Mg^{2+}/Mn^{2+} structures. The significant variation in the active-site conformations of the published mismatched structures and this N7meG structures suggests that polβ, which lacks an intrinsic proofreading exonuclease activity, remains highly sensitive to the presence of base pair mismatch. It also suggests that polβ increases its replication fidelity by varying the conformations of protein and substrate DNA (Figure 1.16B, below). Apparently, polβ deters the possible formation of the pseudo-Watson–Crick N7meG:dTTP base pair in the enzyme active site while accommodating the Watson–Crick N7meG:dCTP base pair. This indicates that polβ increases its replication fidelity by permitting the closed protein conformation and coplanar base pair conformation with near ideal coordination around nucleotidyl transfer chemistry site, only when a nascent base pair can adopt a canonical Watson–Crick base pair conformation.

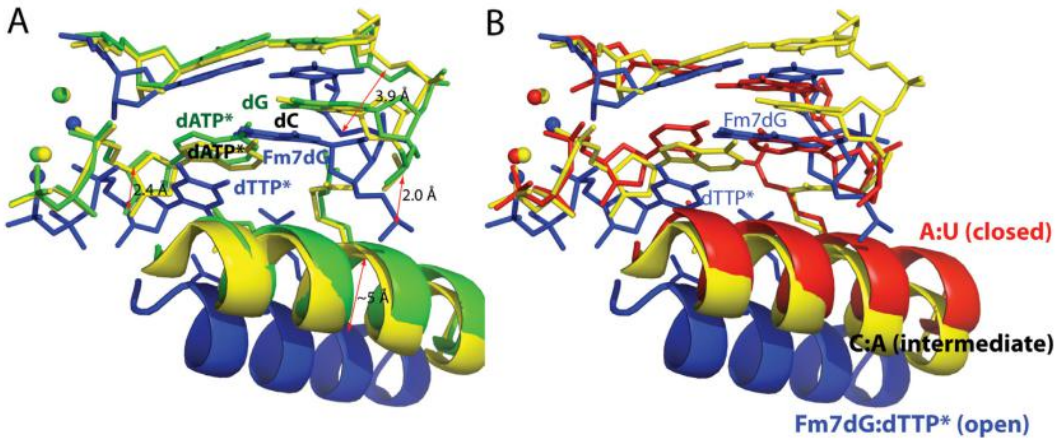


Figure 1.16[30]. Comparison of the N7meG:dTTP*-Mn^{2+} structure with published mismatched structures. Overlay of the N7meG:dTTP*-Mn^{2+} structure (blue) with published: (A) dC:dATP-Mn^{2+} structure (yellow, PDB ID: 3C2L) and dG:dATP*-Mn^{2+} structure (green, PDB ID: 3C2M), (B) dC:dATP*-Mn^{2+} structure (yellow, PDB ID: 3C2L) and dA:dUTP*-Mg^{2+} structure (red, PDB ID: 2FMS).

Effects of N7-alkyl-dG on DNA replication and mutagenesis

This kinetic and structural study, which represent the first systematic investigation of N7medG's mutagenic potentials and its effect on polymerase activity, provide insight into the effect of the predominant N7-alkyl-dG on DNA replication and mutagenesis. Recently, Stone et al. reported kinetic and structural studies on the Y-family DNA polymerase Dpo4 replicating across the N7-aflatoxin B1-dG adduct[50]. In the aflatoxin B1-dG adduct structure, aflatoxin B1 moiety is intrahelical and stacks with the templating dG and an incoming nucleotide, thereby promoting an insertion of dATP opposite the templating N7-aflatoxin B1-dG adduct. In addition, the bulky N7-dG acridine half-mustard adduct has been shown to preferentially induce G to A transition mutations[51].

As similarly seen in the aflatoxin B1-dG adduct structure, the acridine moiety that stacks with an adjacent nucleobase may contribute to the observed transition mutations. Unlike bulky N7-arylalkyl-dG adducts that can form stacking interactions with their aromatic moieties, N7-alkyl-dG lesions with small alkyl groups (e.g. ethyl, propyl) may not be mutagenic, as supported by our kinetic and structural studies on m7dG. The small N7-aklyl-dG lesions may form the canonical Watson-Crick base pair with dCTP in the active site of a DNA polymerase. Our studies also indicate that the presence of small N7-alkyl-dG adducts at the templating position may block replication by some DNA polymerases that are sensitive to an abnormality of the templating base.

Misincorporation of dTTP opposite N7meG may be less favored than that opposite dG in the polβ active site

Polβ is an error-prone DNA polymerase that makes a mistake every several thousands of the nucleotides transferred. The predominant base substitution errors made by polβ are G:T mismatches, comprising ~60% of the total base substitution errors[52]. The failure to incorporate dTTP opposite the templating N7meG by polβ thus implies that the presence of the intact N7meG lesion at the templating position may significantly decrease the dTTP misincorporation rates of some DNA polymerases. The crystal structure of the X-family DNA polymerase pol λ , whose active-site structure is essentially identical to that of polβ, shows the formation of a pseudo-Watson-Crick dT:dGTP base pair in the enzyme's active site presumably via ionization[53], implying that a pseudo-Watson-Crick G:T base pair may also form in the polβ active site. Indeed, our polβ ternary structure with dG:dTTP* and active-site Mn^{2+} shows that the templating dG forms a coplanar pseudo-Watson-Crick base pairing with the incoming dTTP* with an average H-bonding distance of 2.9 Å (PDB ID: 4PGX). In addition, this polβ-dG:dTTP*-

Mn^{2+} structure shows the closed protein conformation and the completion of coordination of the catalytic metal ion, while the polβ-N7meG:dTTP*-Mn^{2+} structure shows an open protein conformation, the incomplete coordination of the catalytic metal ion and the staggered N7meG:dTPP* base pair as shown above. The large structural differences between the catalytically unfavorable polβ-N7meG:dTTP*-Mn^{2+} complex and the catalytically competent polβ-dG:dTTP*-Mn^{2+} complex, together with the lack of dTTP incorporation opposite N7meG by polβ, thus suggest that the templating N7meG in the polβ active site deters dTTP misincorporation to a greater extent than that of the templating dG.

1.4 SUMMARY OF CHAPTER I

Although polβ is an error-prone DNA polymerase that makes a mistake every several thousands of the nucleotides transferred. The failure to incorporate dTTP opposite the templating N7meG by polβ indicates that the intact N7meG lesion at the templating position does not change the sequence and thus retains the integrity. The reasons behind this might be that the methyl group is small and also fixed to the N7 position, which neither participates the Watson-Crick base pair nor does it affect the sugar ring. The inefficient insertion indicates several possibilities: i) *In vivo*, it could imply that the translesion synthesis across the N7meG site is probably performed by some other polymerases, such as pol η , or other more error-tolerating polymerases. ii) It can also imply that this low efficiency creates a time window to bring in the repair machinery for the N7meG site. iii) As for its gap-filling role in the BER, the low polymerizing efficiency across N7meG by polβ probably implies a mechanism that retains the sequential performance of the enzymes in the BER process. If polβ is somewhere

polymerizing the lesion while the lesion is not detected and cut off by the glycosylase, the low insertion efficiency would hamper the translesion synthesis of polβ, and restart the whole BER process.

From the structural perspective, the results imply that the coordination of 3'-OH of the primer terminus is crucial in affecting the insertion efficiency, though it might not be the determining factor for insertion success. From the N7meG:dTTP* ternary structure, it can be seen that the nucleotidyl transfer reaction distance is not the crucial factor in the insertion efficiency. This distance is actually shortened in the dTTP* structure compared to the dCTP* structure. It looks like that the coplanarity, minor groove recognition and crucial catalytic ion coordination value more in the insertion efficiency than this distance.

Chapter II: Novel base pair patterns of N7-methylguanine lesion in the DNA duplex revealed by a new host-guest system using polβ[2]

2.1 BACKGROUND INTRODUCTION

Previously, the insertion across the templating N7meG has been studied. As confirmed by both kinetic and structural results, N7meG can pair with the incoming dCTP through canonical Watson-Crick base pair. And for the incoming dTTP, there is no significant binding[14]. These conclusions are obtained in the polβ model, which has a relatively constrictive active site that often requires multiple minor groove interactions to perform the nucleotidyl transfer reaction. It is highly possible with lesion tolerating polymerase such as pol η , N7meG lesion can form mismatched base pair with other dNTPs. Additionally, considering that even for the normal dCTP insertion, the insertion efficiency is quite slow (decreases ~300 fold)[14], and that incoming dTTP*, though before the precatalytic step, can still stay relatively stably in the active site of polβ to be observed with clear electron density, these mismatches of dNTPs (other than dCTP): N7meG can probably take place *in vivo*.

Furthermore, DNA duplex can adopt a variety of multi-stranded structures that go beyond the standard Watson-Crick paired antiparallel double helix, and even in the traditional B-DNA form there are a great deal of mismatches that can stably exist[54]. One major part of the mismatch is the base pair that contains guanine, which is usually the stablest mismatch among all other mismatches. Mismatch of the guanine can be seen frequently from base pair such as G:A, G:T and even G:G, and they all have very important features concerning biological significance: i) G:T is the most common mismatch. It can be easily formed through cytosine methylation and continuous

[2] Chapter II is based on the publication: N7 methylation alters hydrogen-bonding patterns of guanine in duplex DNA. Kou Y*, Koag MC*, Lee S. J Am Chem Soc. 2015 Nov 11;137(44):14067-70. * These authors contributed equally to the work.

spontaneous deamination[55]. ii) G:A mismatch has the highest escaping rate from the mismatch repair enzymes[56], and they can also be formed very easily as G:G mismatch from the triplet repeating sequence[57]. iii) Besides its easy formation, G:G mismatch can be efficiently repaired. However, its unique structural feature is often indispensable in stabilizing the telomere structures[58], observed in DNA triplexes[59] and remains as the well known structure study model of many RNA stem bulges[57]. Since the N7 methylation on guanine is the most prevalent alkylation, it is natural to question what are the potential effects of N7 methylguanine to the G:X (X=A,T,G, C) base pairs in the DNA duplex structure.

Apparently, all these N7meG lesion base pairs are biologically important, since they could bring mutagenicity to cells from either sequence change, or DNA duplex structure distortion. Either way, it is very valuable if the structures of these lesions and surrounding environment alike are known. Besides the straightforward advantage of observing the effects brought by these lesions, knowing the detailed structures is also beneficial in another two ways: i) Exploration of the potential DNA repair machinery, since the lesions *in vivo* are often detected and repaired by DNA repair pathways. The structure of the intact DNA lesions provides at least the basic mode for repair recognition and initial contacts from the lesion repair machinery. ii) Study of the exogenous DNA modifications, which is the hot spot of carcinogenesis. Studies in this area would be much advanced if the DNA lesion structures are known and easily obtainable. Exogenous mono-alkylating reagents, inter- or intra-strand cross-linkers, and potential mechanisms of the drug leads, all of which can be explored on the molecular level with vivid structural information that requires less or even no further confirmation.

However, obtaining the detailed structures of DNA duplex, especially the lesion containing structures, which are usually thermodynamically unstable, remains hard. NMR

(Nuclear magnetic resonance) has long been used as method to confirm the chemical structure. With 2D NOESY (2-dimension Nuclear Overhauser Effect Spectroscopy) at hand, solution structures of macromolecules such as proteins have been solved. These structures are useful and complementary to single crystal X-ray structures, since they can to some extent reveal the dynamic nature of proteins *in vivo*[60]. Unlike the protein, DNA consists of the deoxyribose and the base with phosphodiester linkers connecting in between. While amino acids side chains are made mainly of carbon chains branching or forming different shapes with small percentage of hetero-atoms such as nitrogen or sulfur, the DNA bases are made of conjugating ring system composed of large portions of nitrogen or oxygen atoms. This makes it possess less protons in a way that intra-molecular base distances or specific structure details are hard to obtain. Although some of the structures are available[61], their accuracy often depends largely on the model choice and mechanic calculations, which may bring some subjective bias[62]. Additionally, the flexibility of the solution structure may severely affect the interpretation accuracy.

To address this issue, the methodology of solving X-ray structures of single crystals of DNA duplex to explain the particular structural problems have been developed, especially for the B-DNA[63]. B-form nucleic acid is the classic and the most common form of right-handed double duplex at neutral pH and physiological salt concentrations. In addition to this canonical form, two other uncommon forms also exist: i) A thicker and more compact right-handed duplex has been observed for RNA-DNA and RNA-RNA duplexes which is called A-form. ii) A left handed duplex consists of alternating purines and pyrimidines that may be involved in transcription and regulations is called the Z-form. DDD (Drew Dickerson Dodecamer)[63] is the major X-ray crystallography solved B DNA structure that enjoys extensive studies. However there are several potential problems with DDD structure: i) High concentration of the negative

charges inherently possessed by the DNA duplex often acts as the repulsive force pushing against each other, making them difficult to be packed in the crystalline lattice. To overcome this, high salt concentration is used in mere DNA duplex crystallization (not complex structure with protein), which often brings additional duplex bending and sometimes it brings other unnatural forms[64]. ii) Because of its innate ionic repulsiveness, the introduction of certain lesions in DNA duplex like DDD would make the base pairs more unstable, which would further destruct the repeating order of the lattice. iii) Studies using the DDD shows that innate structural properties are sequence dependent. So they are often changeable, which may make the structural details questionable. iv) The sequence of DDD is a 12 mer palindromic sequence. This sequence feature facilitates the overlapping and interlocking of the duplex ends with the neighboring duplex in the unit and thus creating a 19° bending over the middle 11 base pairs[64]. v) Additionally, both the flanking sequences of DDD are alternating GC sequence, which has been shown to possess a Z-form tendency[65]. Although it has been proposed to be counter-effected by the AT rich region in the middle, the actual extent of bias or even the compensated AT region 'kink' tendency is not fully understood or quantified.

Therefore, it is a natural idea of using protein-nucleic acid complex to overcome the crystallization difficulty, and trying to get such structures as many as possible to bring in various sequences to get the general idea. This principle is implemented and named 'Host Guest Complex' system (HGC). The protein serves as the host to hold the guest nucleic acid duplex. And by the complex stability, DNA duplex structure with various lesions can be studied in detail. Initially, HGC was used to get several high-resolution RNA structures[66]. Later, the idea is practiced by some groups to develop HGC using DNA as the 'guest'[67][68]. However, the total number of HGC is still quite low, especially considering the quantity of various lesions that can occur on the DNA molecules.

Herein, a new high-resolution HGC system using human DNA polymerase β is reported. And it shows for the first time of the N7 methylguanine (2'-F analog as in Chapter I) lesions base pairing with all 4 dNTPs in the duplex with some novel patterns. Taken together with the thermodynamic parameters determined from the melting assays, it reveals the distinctive features of N7 methylguanine lesions upon different base pairing conditions.

2.2 EXPERIMENTAL PROCEDURES

Synthesis of the N7me/bnG phosphoramidite

The N7meG phosphoramidite was prepared starting from a ribose derivative according to the synthetic procedures described previously[30] in Chapter I. N7bnG phosphoramidite is synthesized as described in Chapert III.

DNA Sequences

Oligonucleotides were purchased from Integrated DNA Technologies (IDT) or synthesized from the above amidites by Midland Certified Reagent Co. (Midland, TX). They were purified by the manufacturer, and their sequences were confirmed by MALDI-TOF mass spectrometry. DNA substrates used for crystallographic studies consisted of a 16-mer template, a complementary 10-mer primer, and 5-mer downstream oligonucleotides. The template DNA sequence used for crystallization was 5'-CCGACGTCGCATYAGC-3' (Y=G,A,T,C). The upstream primer sequence was 5'-GCTXATGCGA-3' (X=N7meG, N7bnG). The downstream oligonucleotide sequence was 5'-pGTCGG-3'. The DNA sequence almost identical to a published binary

structure[38] (PDB ID: 1BPX) was used to minimize sequence-dependent structural differences. The oligonucleotides were mixed and annealed to give a 1 mM mixture of gapped DNA as decribed[14].

DNA duplex melting assay

DNA melting curves were acquired and fitted on LightCycler 480 (Roche Applied Science). For acquisition of fluorescence signal, dye SYBR Green I was used (10,000X concentrate in DMSO, Life Technologies). The 20 μl reaction mixture was obtained by mixing the 10 μl of DNA duplex (CCGACXTCCGCATCAGC, X=N7meG/N7bnG/dG/2FG, 7.5μM), with 1 μl (500X) SYBR green I dye in 10mM potassium phosphate (pH 7.4) buffer containing 200mM NaCl. The reaction mixtures were then loaded into 96 well PCR Microplate (white, Axygen Scientific). The melting experiment was set up by choosing SYBR Green I filter from LightCycler, and after stabilizing at 20 ˚C for 15 seconds, temperature was linearly increased with a step of 0.2 ˚C from 20 ˚C to the highest 99 ˚C and then returned to 20 ˚C. 10 Data points was collected and integrated over each temperature step. 3 repeats were performed with standard deviation calculated.

Protein Expression and Purification

The whole process of polβ expression and purification is the same as aforementioned in Chapter I.

Protein-DNA Co-crystallization

The whole process of co-crystallization is the same as aforementioned in Chapter I.

Data collection and refinement

Diffraction data is collected and processed in the same way as previously mentioned in Chapter I. The structures of the HCG complex containing the N7 methyl guanine lesions with all 4 base pairs were solved by molecular replacement with polβ with a single-nucleotide gapped DNA (PDB code 1BPX) as the search model[38]. The model was built using COOT[156] and refined using Phenix[157] software. MolProbity[158] was used to make Ramachandran plots.

Table 2.1. Data collection and refinement statistics.

PDB code	N7mG:dC	N7mG:dT	N7mG:dA	N7mG:dG	dG:dT	dG:dA	dG:dG
Data Collection							
Space Group	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Cell constants							
a (Å)	54.138	54.396	54.149	54.201	54.344	54.389	54.458
b	78.828	79.480	78.904	79.278	79.345	79.372	79.352
c	54.764	54.912	54.751	54.720	54.584	54.938	54.864
α (°)	90.00	90.00	90.00	90.00	90.00	90.00	90.00
β	105.92	105.36	105.39	105.63	106.14	105.45	105.45
γ	90.00	90.00	90.00	90.00	90.00	90.00	90.00
Resolution (Å)[a]	20-2.83 (2.88-2.83)	20-2.22 (2.26-2.22)	20-2.54 (2.58-2.54)	20-2.45 (2.49-2.45)	20-1.97 (2.00-1.97)	20-2.25 (2.29-2.25)	20-2.40 (2.44-2.40)
R_{merge} [b] (%)	0.143 (0.395)	0.084 (0.584)	0.137 (0.505)	0.117 (0.446)	0.072 (0.210)	0.130 (0.378)	0.109 (0.349)
$\langle I/\sigma \rangle$	9.4 (2.10)	20.4 (3.54)	11.2 (2.00)	13.8 (2.00)	12.9 (3.07)	17.9 (2.89)	16.0 (2.60)
Completeness (%)	100 (99.8)	99.9 (99.7)	100 (99.7)	99.4 (96.0)	99.6 (99.1)	100 (99.5)	94.7 (94.6)
Redundancy	4.1 (3.7)	4.8 (4.5)	5.4 (4.5)	4.8 (4.1)	4.1 (3.7)	4.2 (3.5)	3.2 (3.0)
Refinement							
R_{work} [c]/R_{free} [d] (%)	19.9/26.6	22.1/27.3	20.0/26.2	20.2/25.3	20.2/24.5	20.5/26.5	21.4/27.9
Unique reflections	10562	22542	14558	16411	31502	21462	16613
Mean B factor ($Å^2$)							
Protein	23.5	35.4	29.7	33.3	37.6	34.3	33.1
Ligand	27.2	38.5	31.1	36.5	36.6	35.1	32.0
Solvent	19.3	34.2	29.5	33.3	35.2	36.6	30.5
Ramachandran plot							
Most favored (%)	96.0	96.6	97.2	97.5	96.0	96.6	96.0
Additional allowed (%)	3.7	3.1	2.5	2.2	3.7	3.4	4.0
RMSD							
Bond lengths (Å)	0.003	0.003	0.003	0.003	0.008	0.003	0.003
Bond angles (°)	0.834	0.855	0.876	0.820	1.143	0.642	0.739

[a] Values in parentheses are for the highest resolution shell.

[b] $R_{merge} = \Sigma|I-\langle I \rangle|/ \Sigma I$ where I is the integrated intensity of a given reflection.

[c] $R_{work} = \Sigma|F(obs)-F(calc)|/\Sigma F(obs)$.

[d] $R_{free} = \Sigma|F(obs)-F(calc)|/\Sigma F(obs)$, calculated using 5% of the data.

2.3 EXPERIMENTAL RESULTS AND DISCUSSIONS:

Table 2.2. Melting temperature and thermodynamic parameters determination

	ΔH	ΔS	ΔG	SD: ΔH	SD: ΔS	SD: ΔG	Tm	SD:Tm	C
N7MeG	-107.6	-0.292	-20.6	0.7	0.002	0.7	64.32	0.06	
N7BnG	-112.3	-0.288	-26.5	1.2	0.003	1.2	66.57	0.07	
2FG	-104.3	-0.277	-21.8	0.5	0.003	0.5	65.3	0.03	
dG	-105.6	-0.279	-22.5	0.6	0.002	0.6	65.36	0.04	
									G
N7MeG	-90.2	-0.242	-18.1	1.3	0.003	1.3	58.82	0.04	
2FG	-98.9	-0.256	-21.3	0.5	0.002	0.5	61.2	0.03	
dG	-100.6	-0.269	-20.4	0.5	0.003	0.5	61.3	0.03	
									A
N7MeG	-85.8	-0.232	-16.7	0.9	0.005	0.9	55.72	0.04	
2FG	-92.7	-0.244	-20	0.4	0.002	0.4	60.2	0.03	
dG	-93.6	-0.247	-20	0.6	0.004	0.6	60.8	0.05	
									T
N7MeG	-99.8	-0.257	-23.2	0.7	0.004	0.7	61.1	0.03	
2FG	-94.1	-0.248	-20.2	0.4	0.003	0.4	60.7	0.04	
dG	-93.9	-0.246	-20.6	0.4	0.002	0.4	60.6	0.06	

ΔG, ΔH, and ΔS° are in units of kcal mol^{-1}. T = 298 K. Tm in Celsius degree.

Crystal structure of the human DNA polymerase β-lesion DNA complex

Our lab has previously solved many complex structures of polβ regarding the translesion synthesis across many DNA lesions to an overall good resolution around 2Å[14]. As can be seen from the Figure 2.0A, the templating strand in the structure near active site of the polβ is bent to almost 90˚. However the upstream primer and the counterpart from the templating strand remain largely (~70%) loose without any contacts to polymerase. In the unit cell (Figure 2.0C), the free ends of both templating (3'-) and upstream primer (5'-) strands are only stabilized by the weak hydrophobic interactions with side chains of Phe143 and Phe99 of the second polβ as shown in the amplified

figure. Thus, between these two polymerases in the unit cell, there creates a DNA duplex section whose structure is not affected by the protein-DNA interactions, but remains intact and maintained only by the internal interactions from the duplex itself. Therefore, this part of the DNA duplex can act as the 'Guest' in this complex with polβ (Host).

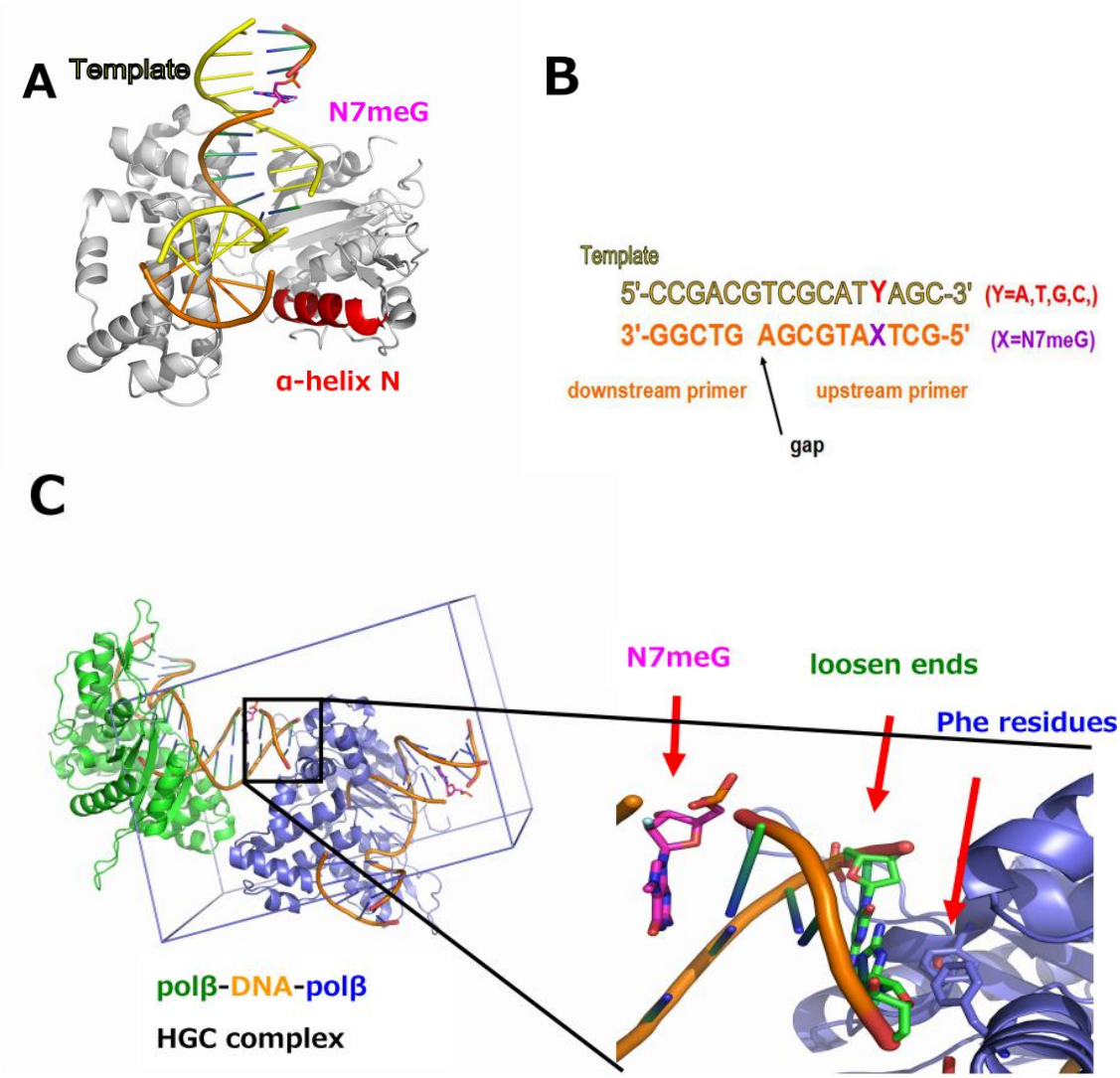


Figure 2.1. HGC system overview. (A) The overall structure of polβ-HGC complex containing N7meG (purple), the polymerase is shown in white, template strand in yellow and primers in brown. (B) The 'guest' DNA duplex paradigm, the lesion is on the upstream primer. (C) The unit cell with HGC system inside. Note that the DNA duplex is positioned in the middle and the free ends are hydrophobically stabilized by the side chains of Phe residues from the second polβ.

The guest part, as shown in the Figure 2.0C, contains 7 nucleotides on each single strand of the duplex, builds almost 3/4 of a full duplex turn. It contains both the minor

and major groove without any bending or distortion of the structure. The sequence is equally and evenly distributed with pyrimidines and purines, with a ratio of 3 to 4. And in the first 4 nucleotides near the free ends, it already has the positions for two purine and two pyrimidine base pairs. The DNA helix differentiating parameters of the 'guest' part, such as rise per bp (3.2Å), helix diameters of each strand (19.5Å/20.8Å), and major/minor groove width (16.4Å/11.0Å) are all conform to the canonical B form DNA. Additionally, the main chain and its torsion angles(α, β, γ, δ, ε, ζ, χ), sugar conformation parameters (v1, v2, v3, v4, p, tm) and base pair parameters (λ 1, λ 2,C1'-C1', RN9-YN1, RC8-YC6) are all indicating the standard B form DNA. Therefore, 'guest' DNA in this polβ-DNA host-guest system can be used as a good candidate for studying lesions or other related influence on the B form DNA.

In this study, N7meG is the lesion to be incorporated. And the 4th position from the 5' end of the upstream primer is chosen (Figure 2.0B), since this position is originally occupied by a dG. Additionally, it is in the middle of 'guest' section, which is relatively distant from the both ends, where some minor influence might exist because of the polymerase contact. As can be seen from Figure 2.0B, the 4th purine position is joined by another purine (5th) and 2 pyrimidines (2nd and 3rd), which can serve as the suitable positions for potential intra-stand cross-link lesions of purines or pyrimidines in the future. The advantage of this HGC system also lies in the easiness of preparation. Following the aforementioned procedures, it has been testified in our hands that the polβ can be expressed and purified in a relatively high yield (1L culture of several milligrams is enough for several trays), and can be readily crystallized in short time and good quality (generally requires only 1 to 3 days, with near half showing good hits). And from the following, it can be seen that the electron density of the lesion site N7meG and its base pairing partner can be clearly observed and modeled. Those base pairing partners

includes the mismatched ones, which are quite unstable and thus the clear density of which should be difficult to obtain. The clear observation of these from this HGC system further adds to its potential of wide application in exploring various DNA lesion structures.

Structure of the N7meG base pair with the cytosine

As aforementioned, our lab has solved the precatalytic structure of the incoming dCTP* with the templating N7meG base[14]. The two in the polβ active site have shown the classic Watson-Crick base pair. However, this is only the precatalytic step with the incoming dCTP* still as a nonhydrolyzable analog. The post extension status of the N7meG:C base pair has been studied in an AlkA host-guest complex[30]. However, the precatalytic N7meG:dCTP* base pair in polβ is different to the N7meG:dC base pair in AlkA complex structure. Unlike the relaxed base pair in the AlkA study where H bonds average distance is 3.4Å, N7meG:dCTP* base pair in polβ is much more similar to the canonical conformation. This is probably because of the active site confinement from the polβ. And also, although the DNA duplex in the AlkA host-guest complex remains as the B-form DNA, the axis is bent about 16˚ through the contacts of the guest DNA duplex and the host AlkA protein, which puts the N7meG:dC in the middle of the bent duplex. This might also explain the relaxation of Watson-Crick base pair of N7meG:C in AlkA host-guest complex. Additionally, the resolution of AlkA-N7meG host-guest complex is 2.9Å. And the electron density of the N7-methyl and N2-amine groups of the N7meG is not clear, with both models remaining unassigned by the the electron density at 1 σ . All these further adds up the need of reexamining the post-extension status of N7meG:C base pair.

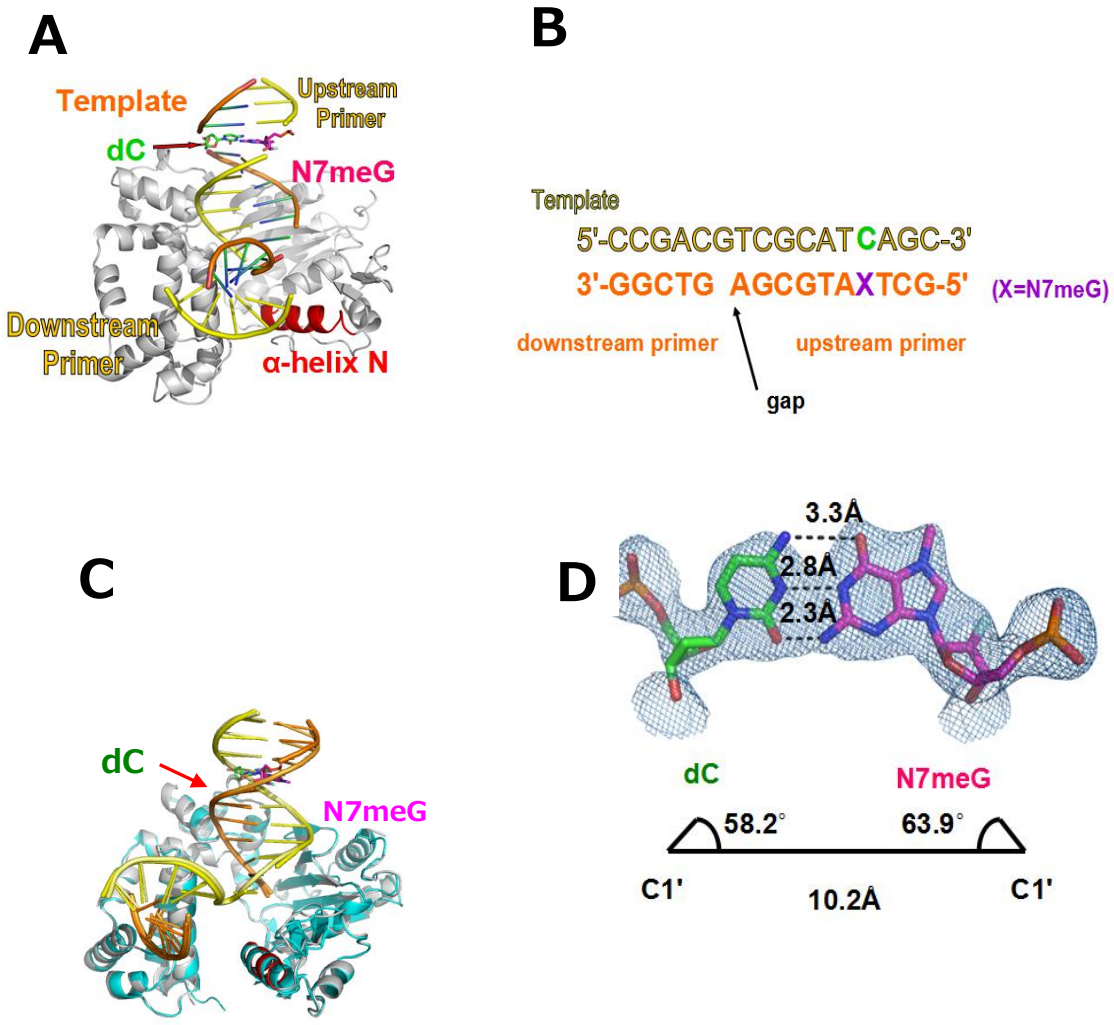


Figure 2.2. N7meG:C in HGC system. (A) Overall structure of polβ-HGC-N7meG:C. N7meG is shown in purple with pairing dC shown in green. (B) Paradigm of 'guest' DNA duplex sequence. (C) polβ-HGC-N7meG:C complex structure alignment with known polβ gapped binary structure (PDB ID = 1BPX, in cyan, r.m.s.d.=0.699Å). (D) Electron density contoured at 1 σ level with base pair parameters shown.

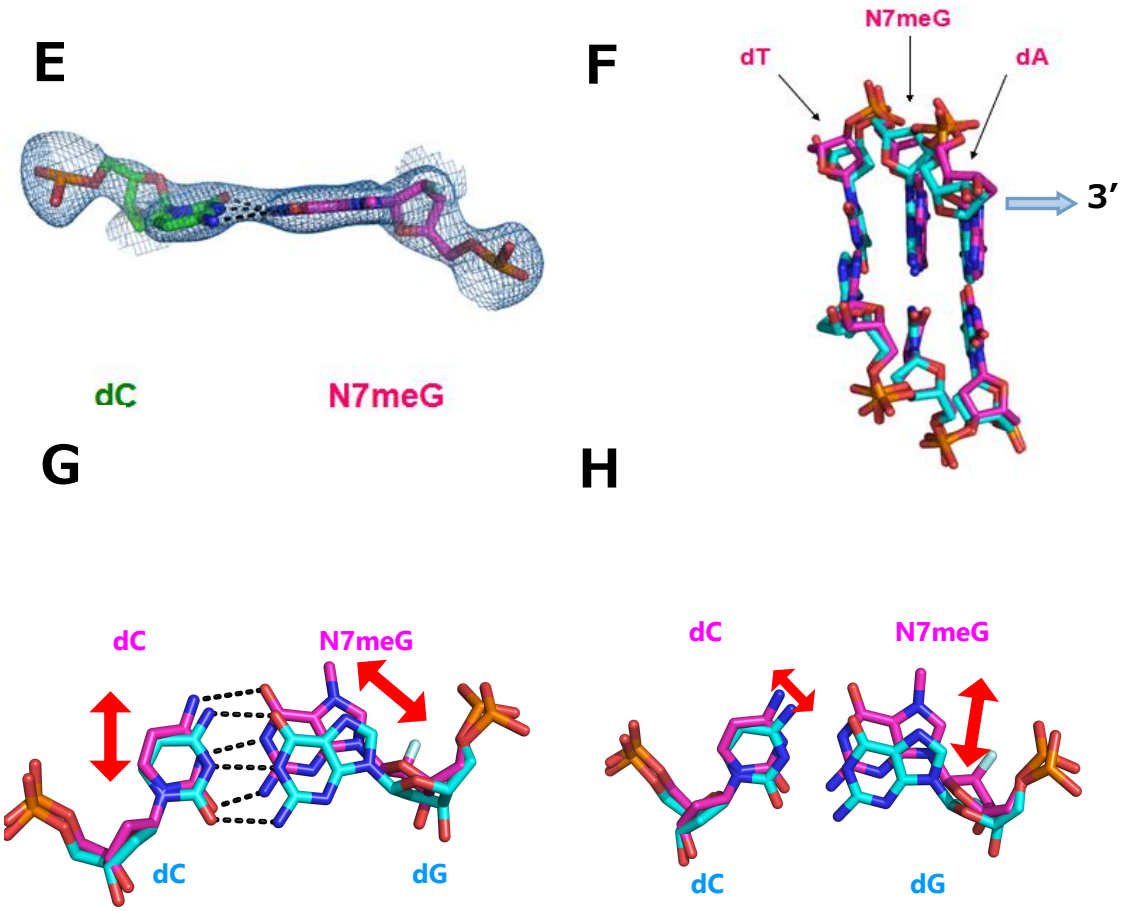


Figure 2.2. N7meG:C in HGC system. (E) View from the minor groove side, showing propeller distortion of the base pair. (F) Distortions brought by N7meG:C in the surrounding duplex, as shown in purple, normal G:C in surrounding duplex is shown in cyan (PDB ID = 1BPX). (G) H bonds distance comparison of N7meG:C (in purple) to normal G:C (PDB ID = 1BPX, in cyan). (H) Shift of N7meG (in purple) compared to the normal G (PDB ID = 1BPX, in cyan).

As can be seen, the overall structure of the N7meG:C is shown in Figure 2.1A, and refined to 2.83Å (Table 2.0). As aforementioned, there is no incoming dNTP, so the

α-helix N of the gapped polβ is taking the open conformation (shown in red). By alignment, the overall structure of polβ-N7meG:C HCG is very similar to that of the previously published normal binary structure of polβ (Figure 2.1C, PDB ID=1BPX, r.m.s.d.=0.699Å). The base pairing of N7meG:C in the HCG is taking the Watson Crick base pair with 3 H bonds, the same pattern as normal G:C base pair in the published normal binary structure. However, there are some differences. As shown in Figure 2.1D and G, whereas the H bonds between the normal G:C are approximately equal with average distance of 2.86Å, the H bonds between the N7meG:C is wide open from the major groove side of 3.34Å between N4(C) and O6(N7meG), but gradually narrowed via 2.76Å between N3(C) and N1(N7meG) to 2.31Å between O2(C) and N2(N7meG). And whereas the normal G:C pair is making ideal coplanarity, the N7meG:C is obviously in a propeller configuration, with the π angle of -1.52˚ (G:C) in contrast to -15.59˚(N7meG:C), as shown in the Figure 2.1E. To better illustrate the structural conformation, the propeller angle and other configurations are described in the following schematic diagrams, in which the the x-axis points towards the major groove edge of a base or base pair and the y-axis points toward the 5' of the templating strand (strand I)[figure from 69].

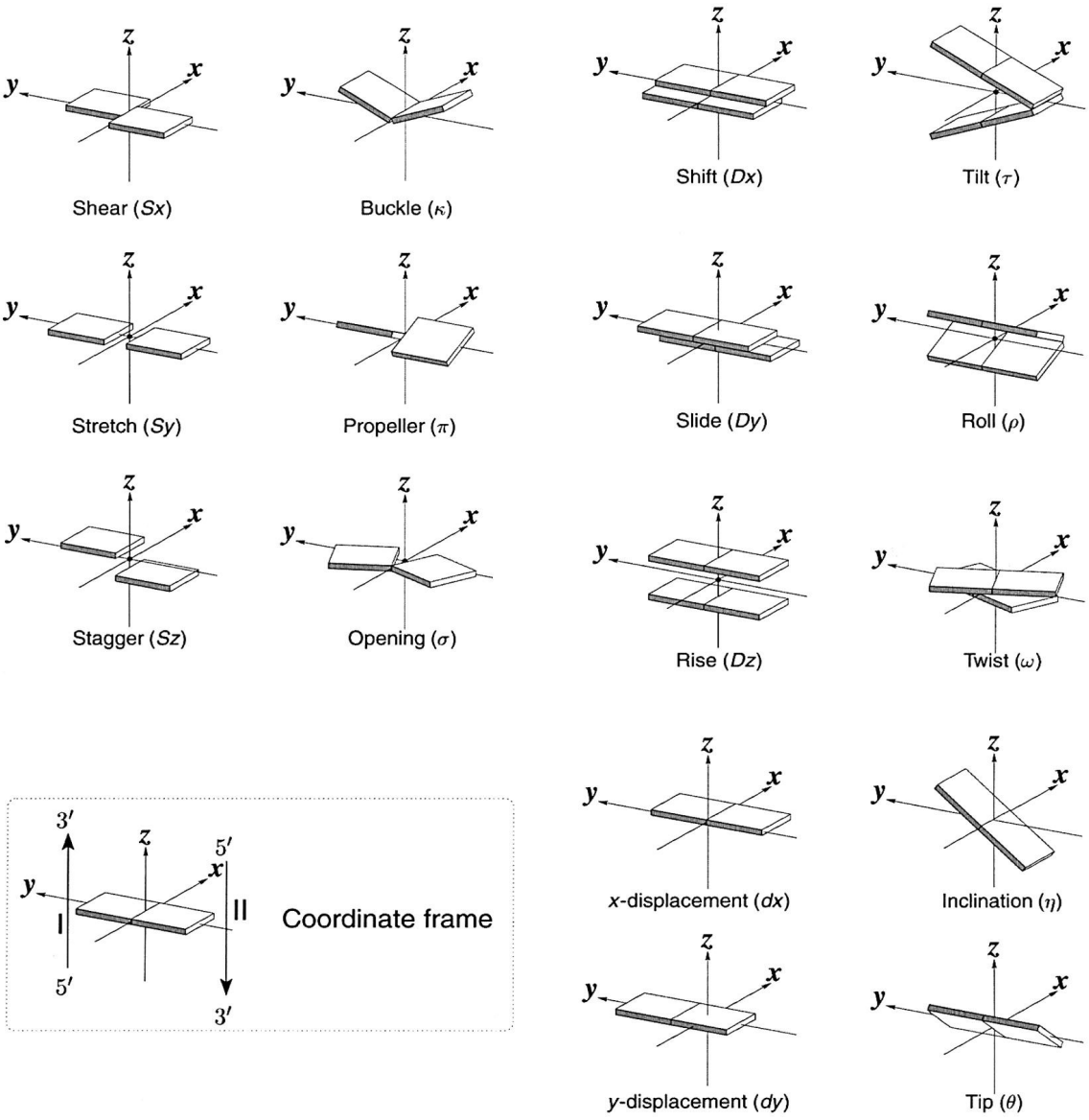


Figure 2.3 Schematic diagram of DNA base pair configurations[figure from 69]

Comparing the local base pair parameters[70] of N7meG:C with those of the normal G:C, as shown in the table below, the N7meG:C base pair shows angle change not only from propeller distortion, but also from others like buckle (change from -0.28˚ to 10.68˚) and opening (change from -0.47˚ to 14.08˚) distortions.

Table 2.3 Local base pair parameters of N7meG:C

Local base pair parameters	Shear	Stretch	Stagger	Buckle	Propeller	Opening
N7meG:C	-0.14	-0.02	-0.33	10.68	-15.59	14.08
G:C	0.11	-0.07	0.17	-4.60	-13.41	-1.19

Parameters for Shear, Stretch and Stagger are distances (Å). Parameters for Buckle, Propeller and Opening are angles (˚). The same as follows.

The overall effect of these angle changes is to deform the ideal coplanarity between the N7meG:C base pair. The λ angles of C:N7meG also changes from C:G (λ 1: 60.0˚, λ 2:54.6˚) to λ 1: 58.2˚, λ 2:63.9˚, with the C1'-C1' distance almost the same 10.2Å (N7meG:C) to 10.4Å (G:C). Additionally, the deformed coplanarity between the N7meG:C base pair affects the neighboring base pairs as well, as shown in the Figure 2F and the table below:

Table 2.4 Neighboring base pair parameters of N7meG:C

Local base pair parameters	Shear	Stretch	Stagger	Buckle	Propeller	Opening
A5:T12 (N7meG)	0.21	-0.16	0.28	2.46	-17.39	-1.01
T3:A14 (N7meG)	-0.17	-0.12	0.23	7.49	-16.72	1.78
A5:T12 (G)	0.15	-0.16	0.03	5.75	-11.47	-4.13
T3:A14 (G)	-0.08	-0.06	0.09	-4.84	-8.30	3.05

Numbers are given in the order from 5'-3' on either templating or primer strand.

As can be seen from table the main influence on the surrounding base pairs brought by N7meG:C base pair is the propeller distortion. Both the A5:T12 and T3:A14 base pairs increase their propeller distortion angle approximately 10 times. This propeller

distortion change is also reflected by the tilt and roll angles between the base pairs, as can be seen from the base pair step parameters:

Table 2.5 Local base pair step parameters of N7meG

Local base pair step parameters	Shift	Slide	Rise	Tilt	Roll	Twist
T12C13/N7meG4A5	1.61	0.67	2.91	10.05	0.08	39.60
C13A14/T3N7meG4	-1.63	0.36	3.64	-8.23	4.03	32.99
T12C13/G4A5	0.19	-0.24	3.36	1.31	0.31	41.71
C13A14/T3G4	-0.31	-0.030	3.24	0.10	5.76	34.13

Parameters for Shift, Slide and Rise are distances (Å). Parameters for Tilt, Roll and Twist are angles (°). The same as follows.

The increased tilt angle between the base pairs around the N7meG shifts the phosphate back bone of both strands and induces the widening of the major groove width (P-P distance) from 16.7Å (C13A14/T3G4) to 18.3Å (C13A14/T3N7meG4).

As can be seen from the alignment (Figure 2H), the driving force of all these structural differences originates from the methylation on N7. This induces the shift and rotation of the guanine base towards the minor groove. Comparing the base pair parameters with that of the N7meG:C base pair in the aforementioned AlkA complex (PDB ID: 3D4V), which also shows the gradual H bonds shortening from major groove side to the minor groove side, we can see that a common feature of N7meG:C base pair is deformed coplanarity from ideal status by increasing propeller, opening and buckle distortion angles. And among which, the propeller distortion is the most prominent.

Table 2.6 Local base pair parameters comparison

Local base pair parameters	Shear	Stretch	Stagger	Buckle	Propeller	Opening
N7meG:C (Polβ)	-0.14	-0.02	-0.33	10.68	-15.59	14.08
N7meG:C (AlkA)	0.01	0.10	0.21	-3.95	-12.81	5.60
G:C	0.11	-0.07	0.17	-4.60	-13.41	-1.19

Our thermodynamic results shows that the Tm is lowered by the presence of N7meG. However, the ΔG, ΔH and ΔS indicates that the destabilization is limited and small. Compare to the normal G:C, N7meG:C in the duplex has a favorable change of ΔH, which indicates N7meG:C is actually more stable. And compared to the normal G:C, N7meG:C in the duplex has a less favorable change of ΔS which indicates that the N7meG presence makes the duplex stay in a more ordered state. From the structure and its parameters, it is clear that N7meG:C base pair itself is not the stabilizing factor. So the stabilizing effect must come from the surrounding environment.

Table 2.7 Neighbor bases overlapping area of N7meG:C

Neighbor bases overlapping	i1-i2	i1-j2	j1-i2	j1-j2	sum
T12C13/N7meG4A5	5.48 (1.17)	0.00 (0.00)	0.00 (0.00)	1.65(0.00)	7.13 (1.17)
C13A14/T3N7meG4	0.00 (0.00)	0.00 (0.00)	0.12 (0.00)	6.96(2.04)	7.08 (2.04)
T12C13/G4A5	2.69 (0.05)	0.00 (0.00)	0.00 (0.00)	3.58 (0.87)	6.26 (0.92)
C13A14/T3G4	0.64 (0.00)	0.00 (0.00)	0.11 (0.00)	1.92 (0.17)	2.67 (0.17)

All units: $Å^2$, Values in parentheses measure the overlap of base ring atoms only. Those outside parentheses include exocyclic atoms on the ring. Directions: i1-i2 (templating strand 5'-3'), j1-j2(primer strand 5'-3')

From the base overlapping area calculated above, we can see that the presence of N7meG makes more stacking with the surrounding base pairs than that of the normal G. This probably is the reason of ΔH and ΔS changes while N7meG's presence. The probable reason of lowered Tm of N7meG containing duplex compared to normal G duplex is that more repulsiveness is shown by N7meG to the surround base pair during the despiralization, as it contains the methyl group and with less favored coplanarity.

Structure of the N7meG base paired with the adenine

As aforementioned, base pair mismatches occurs frequently during the DNA replication, so it must be recognized and repaired by mismatch repair machinery to lower down the potential mutation hazards. This is particularly important for the G:A mismatch, since, as mentioned above, it has the highest escaping rate from the mismatch-detection and subsequent correction process[56]. At the time, the reason for this escape had been suggested of its possible base pair stability that would not disrupt the local conformation so much. However, through the independent validation of a large set of oligonucleotides, the G:A mismatch itself has been shown to be the least stable one among all the mismatches involving guanine[71]. The seemingly conflicting results indicate that G:A induced stability or instability may be varying and depends on different conditions or repair recognition for this base pair is simply ineffective. Additionally, G:A mismatch is of great interesting is because that the base pairing pattern does not follow the Watson-Crick type all the way. As can be seen from below, there have been discovered of 4 types of G:A mismatches (Figure 2.3) (figure remade from [72]).

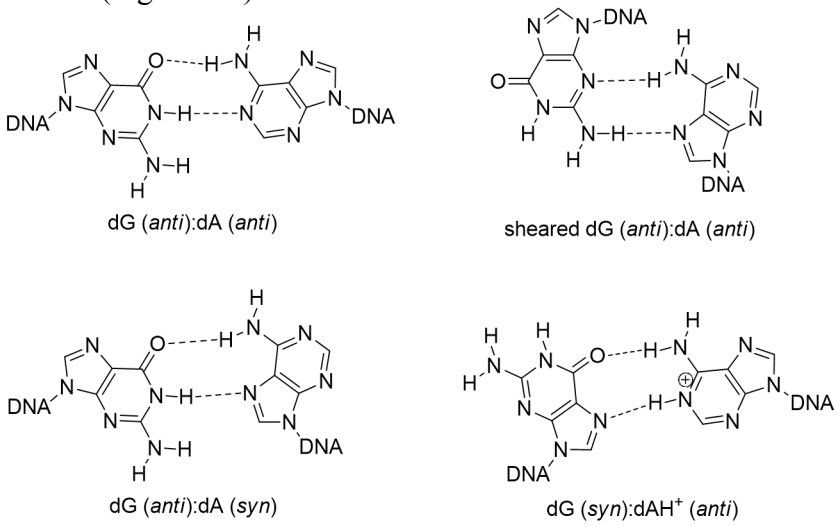


Figure 2.4 dG:dA base pair patterns

Both of which can either take *anti* or *syn* conformation, except there is no G(*syn*):A(*syn*) pairing pattern. For both in *anti* conformations, there exists a special base pairing pattern, which is called the sheared type. This conformation is frequently seen in the unusual secondary structures such as hairpins[73]. The sheared motif of G:A mismatch can be quite stable especially in the triplet sequence such as GNA (N = nucleoside), and especially when it is in a single strand while N serves as the loop end[74]. In such a conformation as shown below (Figure 2.4)(figure from[72]), the sheared arrangement of the G:A base pairs allows the single residue N in the loop to span both helical strands, thus expanding the base stacking of the loop, stabilizing this structure conformation. As can be seen, this kind of secondary structure conformation could be easily generated in lengthy single DNA strand, which will probably bring some steric hindrance for destined biological activity.

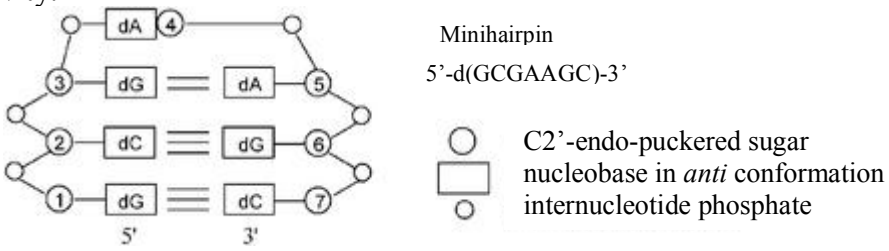


Figure 2.5 hairpin formation of G:A mismatch(figure from[72])

The base pair pattern of the other G(*anti*):A(*anti*) and G(*anti*):A(*syn*) have been found both exist in the neutral pH environment[72]. Though it has been suggested that the both have the similar stability in the duplex[75], it is obvious the G(*anti*):A(*anti*) causes greater distortion of the sugar phosphate backbone. This is probably also the reason that its rate of adenine cleavage is faster[76]. G(*syn*): AH^+(*anti*) conformation exists at pH ranges of 4.0 to 5.5[72], which is reasonable because one of its H bonds requires the protonation of the N1 atom from adenine.

The stability and the variety of G:A mismatches together with its high rate of repair-escaping and adaptability in different environments makes it also a probable target for methylation, especially for N7 methylation on G. Additionally as aforementioned, the correct insertion towards N7meG is quite inefficient, for some lesion tolerating polymerase there is a chance of dATP insertion across the N7meG lesion. Whatever the source, there is no study of N7meG:A mismatch. Herein, we present for the first time the N7meG:A mismatch structure and thermodynamic data in the DNA duplex from the polβ-HGC system.

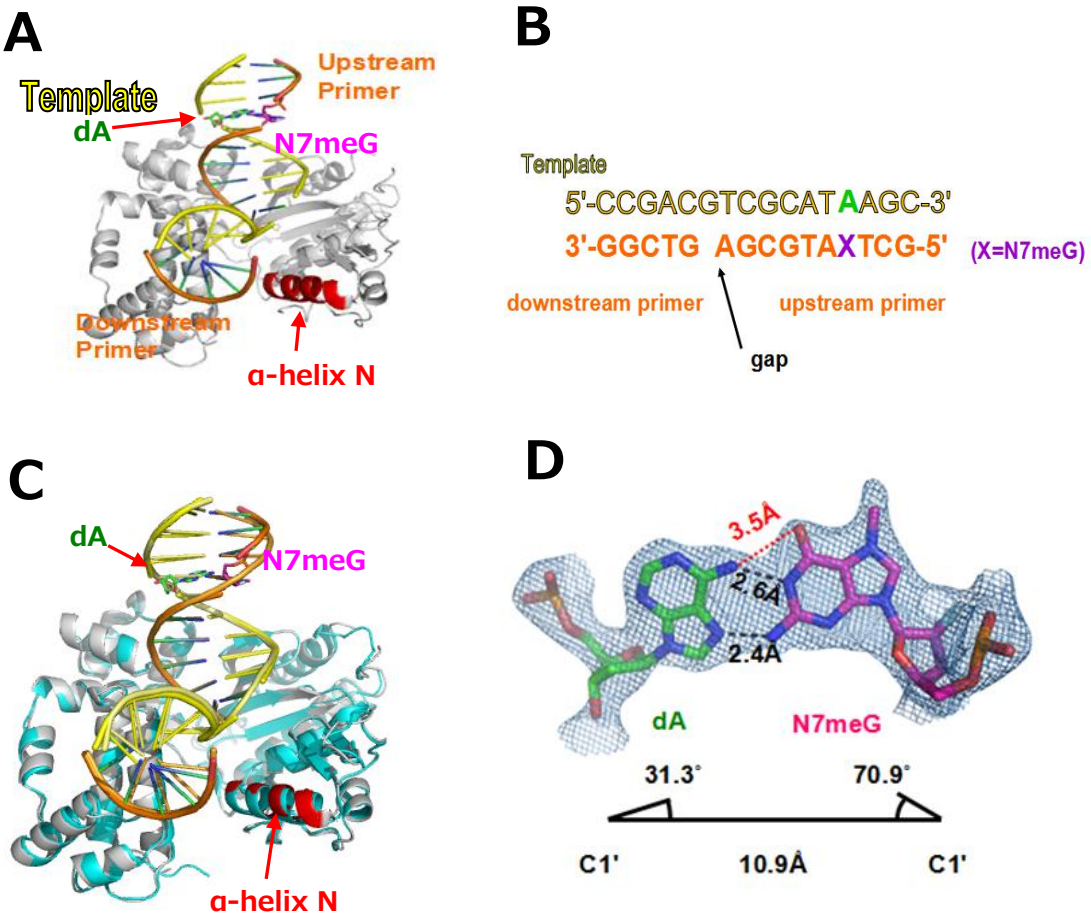


Figure 2.6. N7meG:A in HGC system. (A) Overall structure of polβ-HGC-N7meG:A. N7meG is shown in purple with pairing dA shown in green. (B) Paradigm of 'guest' DNA duplex sequence. (C) polβ-HGC-N7meG:A complex structure alignment with known polβ gapped binary structure (PDB ID = 1BPX, in cyan, r.m.s.d.=0.700Å). (D) Electron density contoured at 1 σ level with base pair parameters shown.

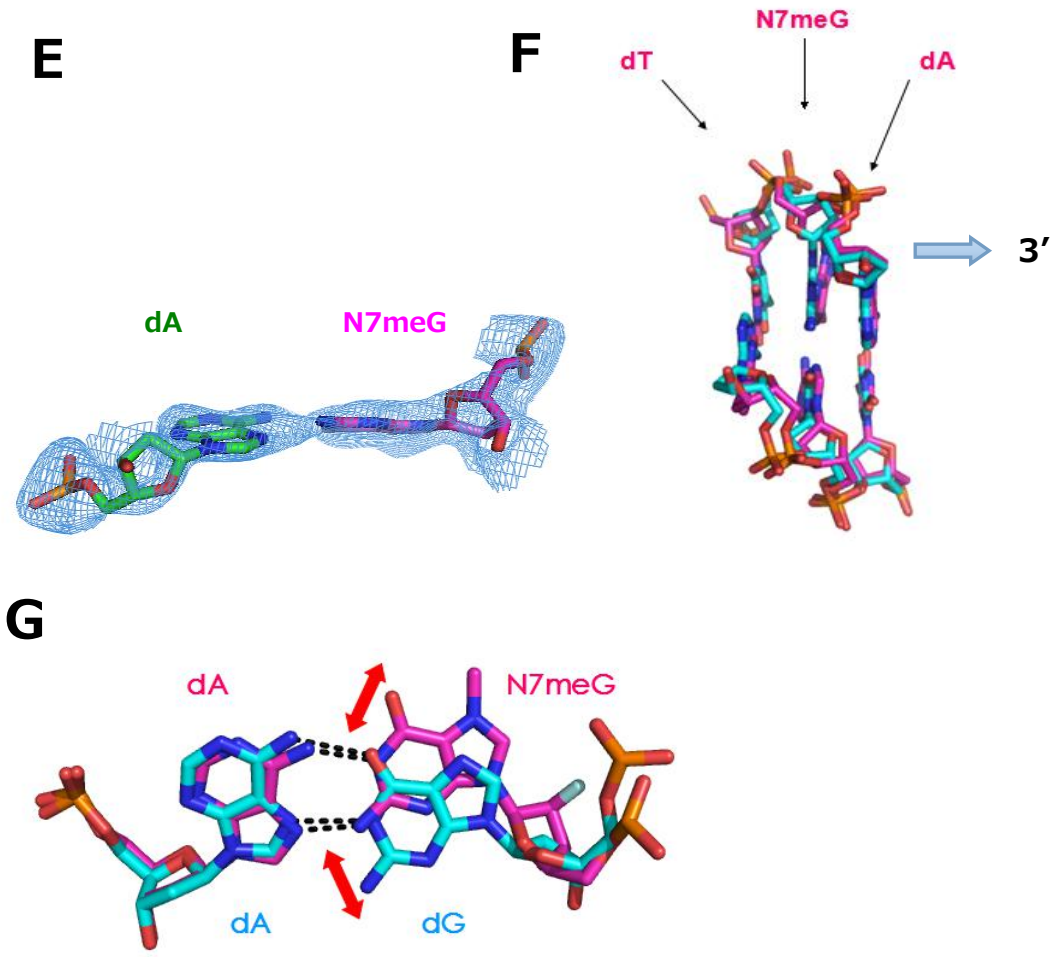
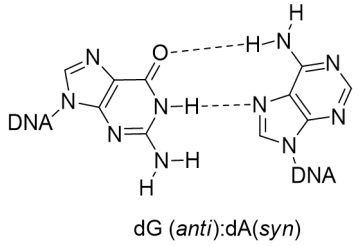


Figure 2.6. N7meG:A in HGC system. (E) View from the minor groove side, showing propeller distortion of the base pair. N7meG is shown in purple. (F) Distortions brought by N7meG:A in the surrounding duplex, as shown in purple, normal G:C in surrounding duplex is shown in cyan (PDB ID = 1BPX). (G) Shift of N7meG (in purple):A compared to the normal G:A in HGC complex.

As can be seen from the Figure 2.5A, the overall HGC of polβ with N7meG:A base pair in the DNA duplex has been solved to 2.54Å, indicating that N7meG:A base pair can stay relatively stable in the DNA duplex. The polβ is taking the open conformation as shown by the α-helix N since there is no incoming dNTP. The overall

complex structure is very similar to the normal binary structure (G:C) previously compared (PDB ID= 1BPX, r.m.s.d.=0.700Å), as shown in Figure 2.5C.

As can be seen from the figure, at 1 σ level, the N7meG is shown (Figure 2.5D) in clear electron density with methyl group and the 2'-F on the sugar. The base pairing partner adenine is also modeled in the clear electron density. The N7meG base is taking the *anti* conformation with the C2'-endo sugar pucker while the adenine is taking the *syn* conformation with the C2'-endo sugar pucker. However, unlike the traditional G(*anti*):A(*syn*) conformation, the N7meG(*anti*):A(*syn*) is taking a shifted conformation with totally different H bonds between the two. In the traditional G(*anti*):A(*syn*) pattern, as shown below (Figure 2.6)[77]:



dG (*anti*):dA(*syn*)

Figure 2.7 normal dG(*anti*): dA(*syn*) pattern

O6 of the guanine is H bond acceptor, while the N6 of the adenine acts as the donor. The N1 of the guanine is donating the H to the N7 of the adenine. Through these two hydrogen bonds, guanine makes the base pair from its Watson-Crick face to the Hoogsteen face of the adenine via 2 H bonds. Here in the polβ-HGC complex, as shown above in Figure 2.5D, N7meG is still making 2 H bonds to adenine, but with different acceptor and donor. In the N7meG(*anti*):A(*syn*) base pair, N1 of the guanine is acting as the H bond acceptor, accepting the H from the N6 amine of adenine. And the N2 exocyclic amine of the guanine is acting as the H bond donor to donate the H to the N7 saturated nitrogen of the adenine. Thus compared to the normal G(*anti*):A(*syn*) mismatch pattern, the N7meG(*anti*):A(*syn*) is making a novel shifted base pair. This new shifted

Hoogsteen base pair also largely changes the base pair values: the λ angles and C1'-C1' distances now changes to λ1: 30.0°, λ2:71.7°, with the C1'-C1' distance increased with to 11.0Å.

The primary reason of this difference comes from the N7 methylation. As aforementioned, it has been suggested that the N7 methylation could decrease the N1H p*K*a from 9 to 7[17][32] and thus the guanine base would be in a zwitterionic form, as shown below (Figure 2.7)[14]. In this way, at the pH 7.5, the N1H of the N7meG can stay largely in the ionic form, which could act as a good candidate for H bond acceptor instead of the donor and form the aforementioned shifted N7meG(*anti*):A(*syn*) base pair.

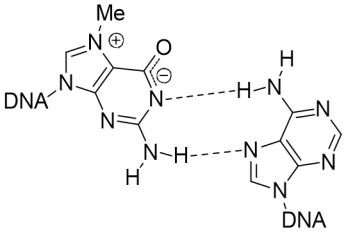
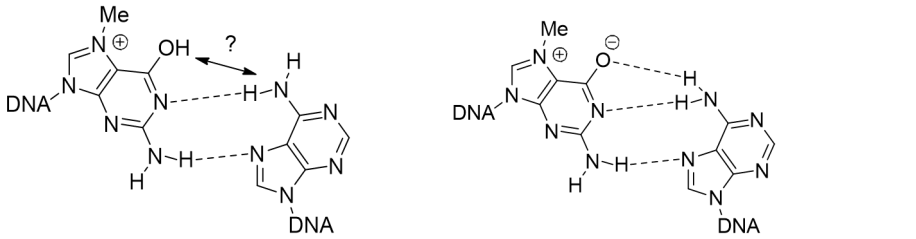


Figure 2.8 N7meG(*anti*):A(*syn*) (zwitterionic form)

The deprotonation of N1 atom from guanine base can also be resulted from the tautomerization (enolization), as shown below (Figure 2.8). In the tautomer form of guanine, the O6 carbonyl oxygen would now be converted into hydroxyl oxygen. The H bond distance between the O6 of N7meG and the N6 amine of adenine is 3.5Å, which is roughly within the average H bond distance range. The H bond existing between the -OH and -NH_2 is also unlikely because the hydrogen shielding effect over the hydroxyl oxygen. It is possible, however, when the O6 is taking the ionized form. But in this way, the N6 amine from pairing adenine would have to rearrange both of its N-H bonds to a disfavored position towards N7meG, as shown below on the right. Plus the shortened H bond distance between the N6 of adenine and N1 of N7meG, it is likely that the ionized zwitterionic form with negative charge distributed near N1 atom of N7meG is the major

case. Although it cannot be completely ruled out of a possible form N7meG base pair with adenine in this new shifted pattern, the possibility, at least judging from the structure we have, is small.



N7meG(*anti*):A(*syn*) (tautomer form) N7meG(*anti*):A(*syn*) (from N1H deprotonation)

Figure 2.9 Other possibilities of N7meG:dA base pair

As can be seen from the Figure 2.5E, compared to the normal G:A pair with ideal coplanarity, the N7meG:A base pair does not exhibit such perfectness. Compare base pair parameters with the previous N7meG:C and normal G:C base pair, it is obvious that the N7meG:A base pair great distorts the coplanarity. The huge opening distortion is created by its *syn* conformation of adenine. But even the buckle and propeller distortion is larger than the N7meG:C base pair.

Table 2.8 Local base pair parameters of N7meG:A

Local base pair parameters	Shear	Stretch	Stagger	Buckle	Propeller	Opening
N7meG:A	-1.99	-3.13	-0.70	11.25	19.77	86.11
G:A	0.31	-4.77	-0.020	11.44	10.53	85.78

The surrounding base pairs of N7meG:A base pair shows the similar deformed coplanarity (Figure 2.5G), with prominent increase in the buckle and opening distortion even compared to the N7meG:C, as shown below:

Table 2.9 Neighbor base pair parameters

Local base pair parameters	Shear	Stretch	Stagger	Buckle	Propeller	Opening
A5:T12 (N7meG:A)	0.05	-0.14	-0.06	9.96	-12.47	-0.01
T3:A14 (N7meG:A)	0.25	-0.20	0.39	7.45	-5.32	12.30
A5:T12 (N7meG:C)	0.21	-0.16	0.28	2.46	-17.39	-1.01
T3:A14 (N7meG:C)	-0.17	-0.12	0.23	7.49	-16.72	1.78
A5:T12 (G)	0.16	0.03	0.35	2.90	-16.46	7.06
T3:A14 (G)	0.75	-0.18	-0.23	-3.19	-10.66	5.24

The prominent local and surrounding distortion induced by N7meG:A base pair indicates that the in-between conformations of these bases on the strand might be distorted as well. The local base pair step parameters show that there are significant tilt and roll distortions around the N7meG:A base pair and on both strands. The tilt angles are especially huge, suggesting the N7meG:A would severely distort the surrounding parallel base pair conformations (Figure 3F).

Table 2.10 Local base pair step parameters of N7meG:A

Local base pair step parameters	Shift	Slide	Rise	Tilt	Roll	Twist
T12A13/N7meG4A5	0.60	-3.67	2.37	-173.25	12.23	0.08
A13A14/T3N7meG4	-3.01	-1.28	-0.77	133.39	-114.73	23.07
T12C13/N7meG4A5	1.61	0.67	2.91	10.05	0.08	39.60
C13A14/T3N7meG4	-1.63	0.36	3.64	-8.23	4.03	32.99
T12A13/G4A5	0.46	-3.42	2.13	-174.50	1.48	31.97
A13A14/T3G4	-0.92	3.26	-0.61	-134.43	110.45	154.01

Our thermodynamic data shows that Tm of the duplex containing N7meG:A base pair lowers about 5 degree when compared to that of the normal G:A base pair. The

change of ΔH and ΔS are in agreement with the structural data that N7meG:A base pair is neither stable nor it can bring the order to the duplex when compared to the N7meG:C base pair. Although this novel shifted Hoogsteen base pair brings in more overlapping area around the N7meG:A base pair (as shown in the chart below), the aforementioned distortions of base pair and between base pair are so large that they probably override this stacking effect. This huge distortion may be the reason that the presence of N7meG:A destabilizes the DNA duplex from every thermodynamic parameters.

Table 2.11 Neighbor bases overlapping area of N7meG:A

Neighbor bases overlapping	i1-i2	i1-j2	j1-i2	j1-j2	sum
T12A13/N7meG4A5	5.04 (1.22)	0.00 (0.00)	0.00 (0.00)	1.54 (0.00)	6.58 (1.22)
A13A14/T3N7meG4	3.69 (2.04)	0.00 (0.00)	0.00 (0.00)	6.92 (1.62)	10.60 (3.67)
T12C13/N7meG4A5	5.48 (1.17)	0.00 (0.00)	0.00 (0.00)	1.65 (0.00)	7.13 (1.17)
C13A14/T3N7meG4	0.00 (0.00)	0.00 (0.00)	0.12 (0.00)	6.96 (2.04)	7.08 (2.04)
T12A13/G4A5	5.90 (1.31)	0.00 (0.00)	0.00 (0.00)	4.68 (1.76)	10.59 (3.06)
A13A14/T3G4	1.97 (1.25)	0.00 (0.00)	0.00 (0.00)	0.70 (0.00)	2.67 (1.25)

As can be seen from the Figure 3G, compared to normal G:A base pair in the polβ binary structure, while adenine is almost centered at the same position as cytosine, the N7meG moved at least 1.7Å towards the major groove side to be able to make the shifted Hoogsteen base pair with the adenine. This also widens the major groove by ~1Å (P-P distance around 18.6Å surrounding N7meG). So primarily, it is possible that the presence of N7meG:A base pair with the novel shifted Hoogsteen base pair pattern serves as the initial driving force that causes all these structural and responding thermodynamic changes.

Structure of the N7meG base pair with the thymine

Of the eight possible mismatches, the G-T mismatch is well known as the most common one. It can be easily created by the cytosine methylation, and spontaneous deamination, especially in the CpG contexts. The two processes are continuously connected to each other[55]. This mismatch can also be formed through erroneous incorporation of dNTPs by DNA polymerases during replication or through hetero-duplex formation after homologous recombination[78]. It can be efficiently repaired from the DNA polymerase with proofreading functions, post-replicative mismatch repair pathway (MMR) or glycosylases that excise the thymine base specifically[79]. The efficient repair of all these pathways does not guarantee its instability. In fact, as aforementioned, G:T mismatch can be quite stable. Thus its detection and potential local distortion still remains as the hot spot for mismatch structure studies in the DNA duplex[80].

The highest presence probability, the importance in mismatch repairing recognition and the aforementioned fact that inefficient correct insertion across N7meG in the polymerase active site, leads to a natural question of what would be the situation of the N7meG:T mismatch in the DNA duplex. As can be seen from the occurrence of the G:T mismatch. The C in many N7meG:C base pair could undergo methylation and deamination to be converted into N7meG:T mismatch. However, beyond these many sources, the influence of N7meG:T mismatch in the duplex remains still unknown.

Several studies have been done to explore the potential influence of the normal unmethylated G:T mismatch (post-extension in DNA duplex). Among them, the majority is done through the NMR solution study[81][82]. Only two studies are with X-ray crystallography[83][84]. All of which show the wobble base pair pattern of the G:T mismatch in the DNA duplex, as shown in the figure (Figure 2.9)[80] below:

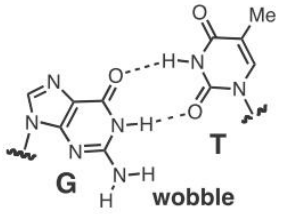


Figure 2.10 G:T wobble base pair

For the G:T mismatch formation in the active site of the polymerase, there have been several studies. Study of *Bacillus stearothermophilus* DNA polymerase I fragment (BF) shows wobble G:dTTP in the presence of the active-site Mg^{2+}[47]. The B-family DNA polymerase RB69 structural study also shows wobble G:dTTP base pairs in the active site[48]. In the active site of Y-family DNA polymerases human polη, T:dGTP also forms a wobble base pair[49]. In the study of the X-family an error-prone DNA polymerase polλ, T:dGTP forms a Watson–Crick-like base pair in the presence of the active-site Mg^{2+} [53]. But as aforementioned, it still forms wobble base pair in the DNA duplex after insertion. Similar case has been observed in the polβ active site[80], with ionized form of Watson-Crick pattern of G:dTTP mismatch.

From all these, we can see that the G:T mismatch exclusively follows the wobble base pair in the DNA duplex after its formation in the active site. And even in the active site of the polymerase, which technically cannot be termed as G:T mismatch (since the nucleotidyl transfer reaction is not finished), the G:T base pair majorly takes the wobble pattern. For the N7meG case, however, this wobble pattern would remain questionable, since as aforementioned, there would be a decrease of p*K*a of the N1 atom on guanine because of the N7 methylation, as seen previously in the N7meG:A mismatch. If this is true, then the wobble base pair pattern of G:T mismatch would lose one H bond, endangering its stability.

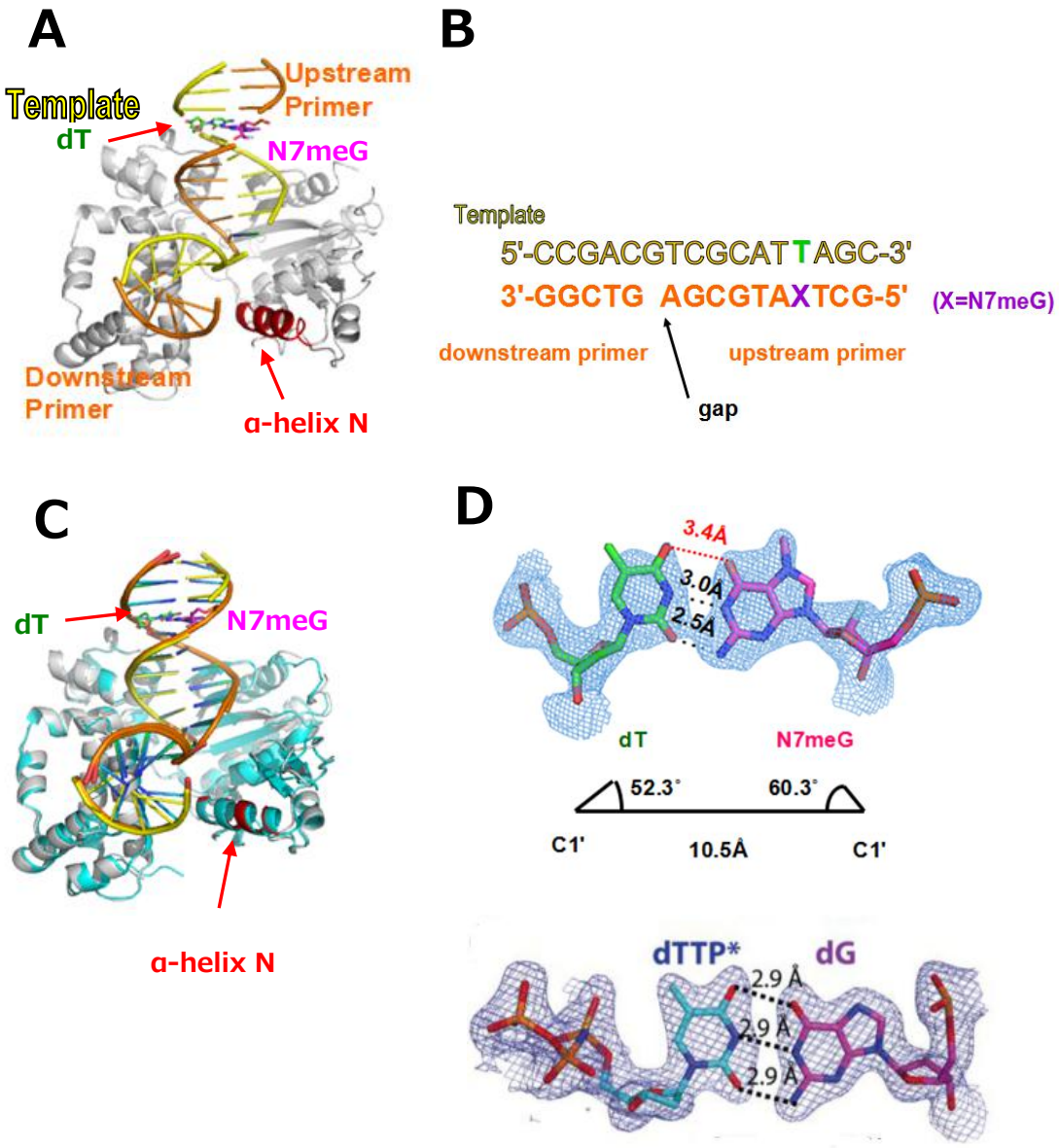


Figure 2.11. N7meG:T in HGC system. (A) Overall structure of polβ-HGC-N7meG:T. N7meG is shown in purple. (B) Paradigm of 'guest' DNA duplex sequence. (C) polβ-HGC-N7meG:T complex structure alignment with known polβ gapped binary structure (PDB ID = 1BPX, in cyan). (D) Electron density contoured at 1 σ level with base pair parameters shown. Note H bonds distance difference compared to the dTTP*:dG mismatch in the active site of polβ below[80].

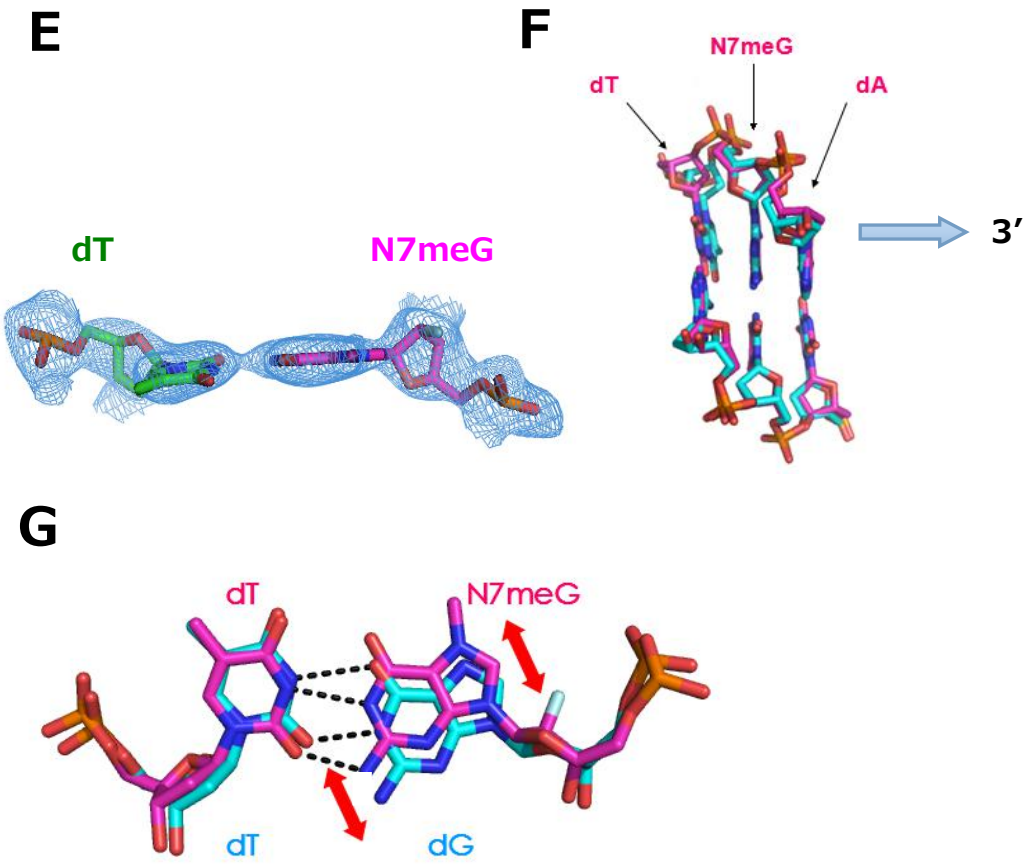


Figure 2.11. N7meG:T in HGC system. (E) View from the minor groove side, showing propeller distortion of the base pair. N7meG is shown in purple. (F) Distortions brought by N7meG:T in the surrounding duplex, as shown in purple, normal G:C in surrounding duplex is shown in cyan (PDB ID = 1BPX). (G) Shift of N7meG (in purple, N7meG:T) compared to the normal G:T in HGC.

To answer all these above mentioned problems, the HGC structure of polβ with N7meG:T mismatch has been solved, and refined to 2.22Å, as shown in the Figure 2.10A. Since there is no incoming dNTP, the polβ is taking its open conformation as indicated by the α-helix N as usual. The HGC-polβ-N7meG:T is quite similar to that of the binary gapped normal G:C structure mentioned above (Figure 2.10C, PDB ID=1BPX,

r.m.s.d.=0.709Å), indicating that the N7meG:T mismatch does not severely distort the whole structure. As can be seen from the Figure 2.10D, the N7meG:T mismatch can be seen clearly with electron density at 1 σ . However, the base pairing pattern of this N7meG:T in the DNA duplex is different from all the G:T mismatch in the DNA duplex. Instead of taking the wobble base pair pattern, which is the H bonds formed between N3 and O2 of thymine to the O6 and N1 of guanine, the N7meG:T is taking the Watson-Crick base pair pattern: H bonds formed between the N3 and O2 of thymine to the N1 and N2 exocyclic amine of N7meG. The similar pattern are only observed in the active site of polβ, when dT is instead an incoming dTTP* analog, at the presence of Mn^{2+}[80] as previously mentioned; and in the active site of pol λ , where G is replaced by the incoming dGTP at the Mg^{2+} presence as aforementioned[53]. And in the pol λ structure, after the active site polymerization, the G:T mismatch in the DNA duplex still remains as the wobble structure. So it is justifiably to say that this a first N7meG:T mismatch structure that takes the Watson-Crick like base pairing pattern in DNA duplex. Additionally, even compared to the active site formation structures, the N7meG:T mismatch is unique in that there is no H bond formation between the O6 of N7meG and O4 of T, whereas in both the aforementioned active site structures, there is about 2.9Å distance between the two atoms.

This lack of H bond between the O6 of N7meG and the O4 of T rules out the possibility that the N7meG:T base pair is a Watson-Crick one with N7meG undergoes the tautormerization, as shown below (Figure 2.11 left). The only pattern left is the ionized one of N7meG. Could it be the tautomer of ionized form of N7meG with O6 possessing the charge while N1 still remains in deprotonated form (below Figure 2.11 right)? If in this case, there would likely exist a H bond between O6 of N7meG and N3 of pairing T. However, this H bond is not observed in our structure.

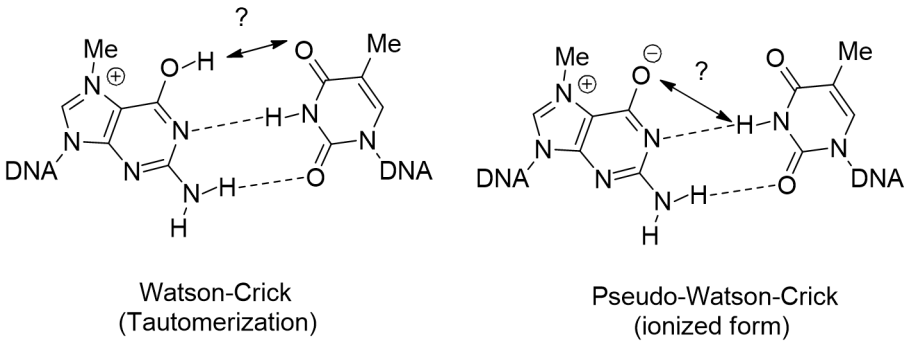


Figure 2.12 N7meG:T base pair possibilities

The most likely H bonding pattern is shown below. This agrees with the previously mentioned p*K*a decrease of the N1H on the N7meG when N7 is methylated. As said, the zwitterionic form of N7meG is totally possible when the N1 p*K*a is lower than the crystallization pH, the N1H simply deprotonates to become a good H bond acceptor, as shown below (Figure 2.12):

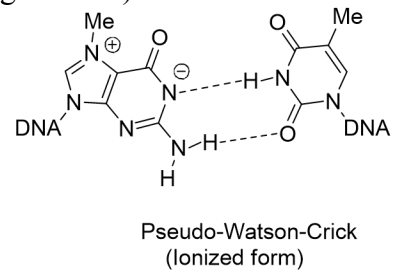


Figure 2.13 N7meG:T pseudo-Watson-Crick base pair

We can see that the N7meG:T base pair is with shorter H bond distance towards the minor groove side (Figure 2.10D[80]). This is not seen in either active site formation structures mentioned above, suggesting a special influence on the DNA duplex structure. Thus the N7meG:T base pair is the more opened pseudo-Watson-Crick pattern, which is observed for the first time for G:T mismatch in the DNA duplex. With this new pattern, the λ angles and C1'-C1' distances now changes a little bit of this base pair to λ1: 60.3˚, λ2: 52.3˚, and the C1'-C1' distance is now 10.5Å.

As can be seen from the Figure 4E, there is a little propeller distortion between the N7meG and T. When compared the base pair parameters to the N7meG:C or even the G:C, we can see that the N7meG:T is actually distorted relatively in a smaller way and only in this kind of distortion:

Table 2.12 Local base pair parameters of N7meG:T

Local base pair parameters	Shear	Stretch	Stagger	Buckle	Propeller	Opening
N7meG:T	0.75	-0.26	0.46	1.29	-13.72	3.12
N7meG:C	-0.14	-0.02	-0.33	10.68	-15.59	14.08
G:T	2.32	-0.64	-0.19	3.86	-9.03	-0.49

Decreased propeller, buckle and opening distortion also decreases the possible deformation of the neighboring residues (Figure 2.10F), as shown below:

Table 2.13 Neighbor base pair parameters of N7meG:T

Local base pair parameters	Shear	Stretch	Stagger	Buckle	Propeller	Opening
A5:T12 (N7meG:T)	0.23	0.040	0.14	3.04	-18.05	-1.62
T3:A14 (N7meG:T)	0.07	-0.49	0.10	-2.28	-11.52	-2.02
A5:T12 (N7meG:C)	0.21	-0.16	0.28	2.46	-17.39	-1.01
T3:A14 (N7meG:C)	-0.17	-0.12	0.23	7.49	-16.72	1.78
A5:T12 (G:T)	-0.12	-0.03	-0.11	3.40	-12.24	3.37
T3:A14 (G:T)	-0.11	-0.20	0.01	-1.54	-6.12	-0.65

As we can see, base pair distortion along the Y axis (the strand I) are reduced, with just one neighboring base pair T3:A14 showing increased opening angle, which does not affect so much of the in-between base pair structure:

Table 2.14 Local base pair step parameters of N7meG:T

Local base pair step parameters	Shift	Slide	Rise	Tilt	Roll	Twist
T12T13/N7meG4A5	0.75	0.17	3.41	-0.73	-2.31	40.08
T13A14/T3N7meG4	-1.28	-0.58	3.40	-3.19	2.17	31.17
T12C13/N7meG4A5	1.61	0.67	2.91	10.05	0.08	39.6
C13A14/T3N7meG4	-1.63	0.36	3.64	-8.23	4.03	32.99
T12T13/G4A5	0.49	0.10	3.22	3.12	1.46	47.12
T13A14/T3G4	-0.51	-0.16	3.29	-0.14	7.26	28.24

As can be seen from this table below, any distortion exists in-between the base pairs has been reduced to almost perfect conditions, which is also true for the base pair overlapping area that estimates the base stacking surrounding N7meG:T base pair:

Table 2.15 Neighbor bases overlapping area of N7meG: T

Neighbor bases overlapping	i1-i2	i1-j2	j1-i2	j1-j2	sum
T12T13/N7meG4A5	7.19(1.30)	0.00(0.00)	0.00(0.00)	3.77(0.78)	10.96(2.09)
T13A14/T3N7meG4	0.04(0.00)	0.00(0.00)	1.56(0.18)	4.31(1.06)	5.91(1.24)
T12C13/N7meG4A5	5.48(1.17)	0.00(0.00)	0.00(0.00)	1.65(0.00)	7.13(1.17)
C13A14/T3N7meG4	0.00(0.00)	0.00(0.00)	0.12(0.00)	6.96(2.04)	7.08(2.04)
T12T13/G4A5	7.90(2.26)	0.00(0.00)	0.00(0.00)	5.29(2.58)	13.19(4.84)
T13A14/T3G4	0.00(0.00)	0.00(0.00)	0.11(0.00)	0.83(0.00)	0.94(0.00)

As can be seen from table above, the overlapping area around the N7meG:T is larger even than that of the normal G:T and N7meG:C base pairs.

All these parameters shown above indicate an astonishingly stabilizing effect of N7meG:T in the duplex DNA, which is even better a stable factor than normal G:C base pair. These data obtained from the structure goes well with the thermodynamic parameters shown above of stabilizing changes in ΔH, ΔS, ΔG and even the Tm. These changes can only be derived from the N7 methylation on guanine. As can be seen from the Figure 4G, upon N7 methylation, the N7meG:T both shift towards the major groove side. Although the major groove width has been widened by the presence of N7meG:T (~1Å), the overall effect of N7meG:T with its more opened pseudo-Watson-Crick base pair pattern improves the stability of the DNA duplex, even compared to the normal G:C base pair.

Structure of the N7meG base pair with the guanine

Among all the 8 mismatches, purine-purine mismatches can exist in a various configurations depending on environmental conditions[85][86]. One famous example of the G:G mismatch is the G-quadruplex (G-quartet), which is the nucleic acid sequences that are rich in guanine and are capable of forming a four-strand structure, among which the Gs are connected to each other through Hoogsteen base pairs to form a square and planar structure which is called guanine tetrad. These planes can stack on one another, with the center cation further stabilizing the whole quadruplex structure[87]. There are various forms of such G quartet, such as the ones formed intra-molecularly or multi-molecularly and they widely exist in DNA, RNA with parallel or anti-parallel directions[88]. The G quartets have been confirmed of its structural obstruction function in transcriptional regulation and the most well known immunoglobulin heavy chain switching[89].

Among the guanine involved mismatches, the G:A mismatch has its 'fame' for high escaping rate from the repair detection and thus the low repairing efficiency. Usually the homo purine-purine mismatch would be among those of the high efficient repairing rate, since the bulkier base pair would create severe distortion or bulge in the DNA duplex[90]. However, it has been confirmed by several studies that the G:G mismatch itself is among the stablest known mismatch pairs[91]. In comparing the stability by the base pair life span, G:G mismatch actually far exceeds others, as aforementioned, and even is comparable to A:T, the canonical and correct Watson-Crick base pair[91]. This suggests that the G:G mismatch induced distortion in the DNA duplex might be quite local, which is in contrast to its highly efficient repair rate. The key to the seemingly conflicting phenomena probably lies in the place where the delicacy of the repair recognition machinery also resides.

As can be seen from all these mentioned, N7meG:G mismatch can occur naturally around, not to mention the correct insertion towards N7meG is quite inefficient and that many G:G mismatch occurs readily from the erroneous replication single DNA strand. Herein, we present the structural and thermodynamic study of the first N7meG:G mismatch.

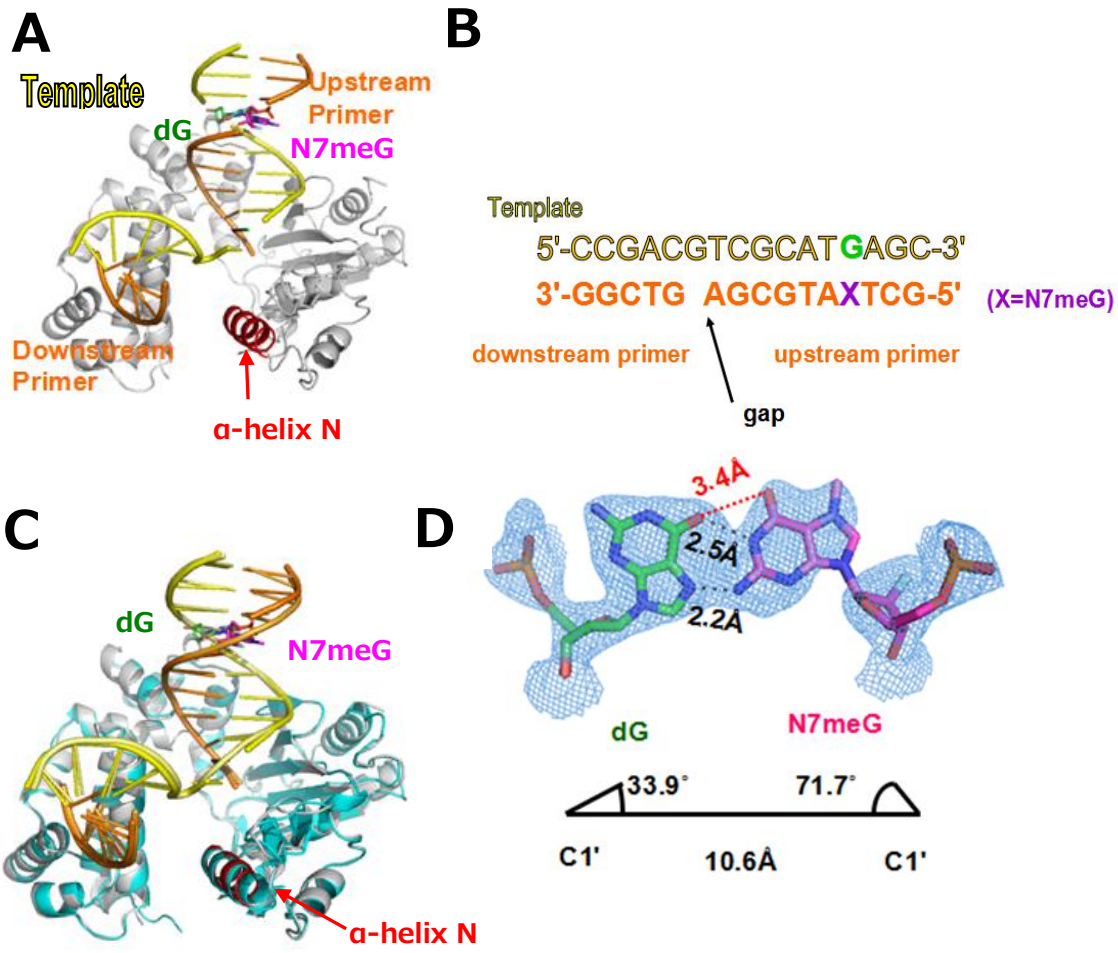


Figure 2.14. N7meG:G in HGC system. (A) Overall structure of polβ-HGC-N7meG:G. N7meG is shown in purple with pairing dG shown in green. (B) Paradigm of 'guest' DNA duplex sequence. (C) polβ-HGC-N7meG:G complex structure alignment with known polβ gapped binary structure (PDB ID = 1BPX, in cyan). (D) Electron density contoured at 1 σ level with base pair parameters shown.

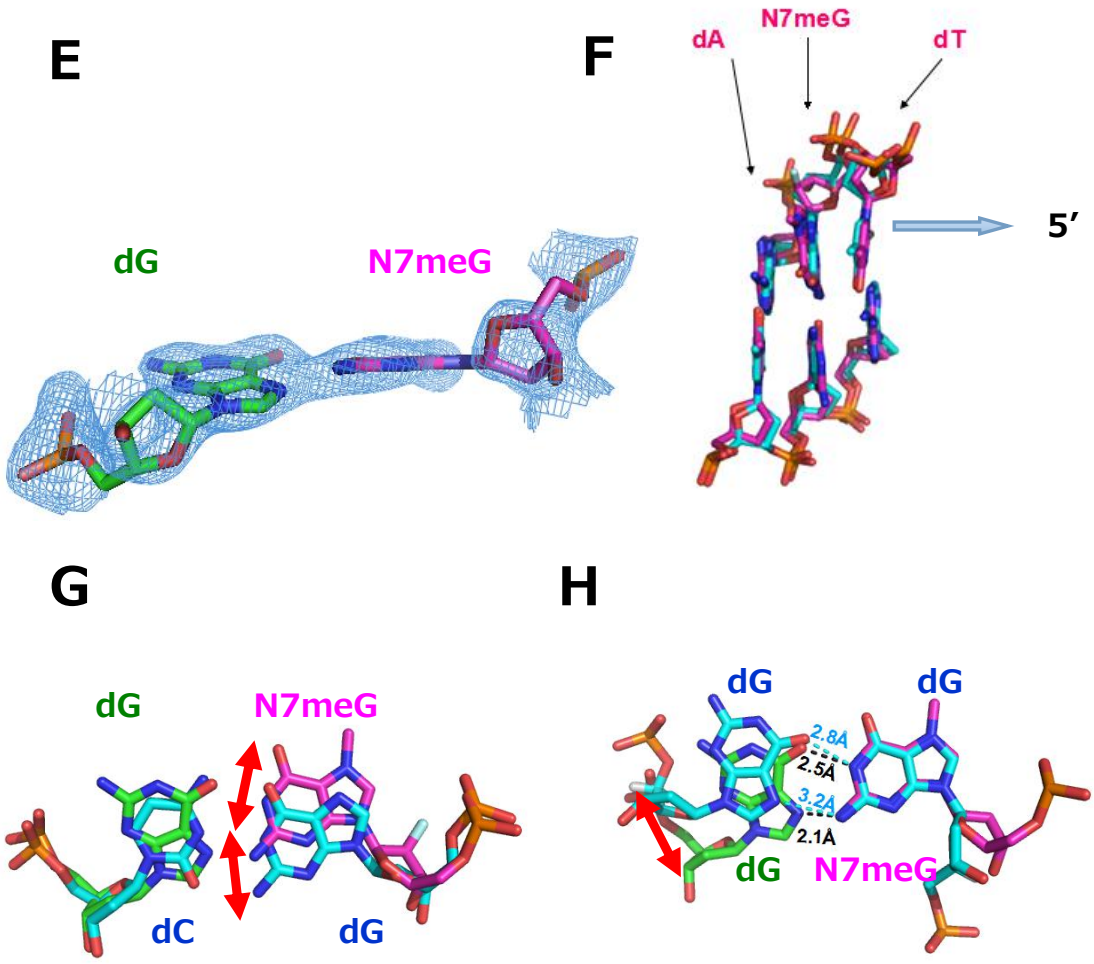


Figure 2.14. N7meG:G in HGC system. (E) View from the minor groove side, showing propeller distortion of the base pair. N7meG is shown in purple. (F) Distortions brought by N7meG:G in the surrounding duplex, as shown in purple, normal G:C in surrounding duplex is shown in cyan (PDB ID = 1BPX). (G) Shift of N7meG (in purple, N7meG:G) compared to the normal G (PDB ID = 1BPX, in cyan, G:C). (H) N7meG (purple):G (green) base pair compared to G:G (cyan) base pair observed in G quartet (PDB ID = 2HY9[93]), note the H bond distance shortened in N7meG:G base pair.

As can be seen from Figure 2.12A, the HGC polβ-N7meG:G structure has been solved and refined to 2.45Å. The same as aforementioned, the polβ is taking the open

conformation as indicated by the α-helix N, since there is no incoming dNTP. The electron density of N7meG:G mismatch containing DNA duplex can be clearly seen indicating the relative stability of N7meG:G mismatch in the duplex structure. The overall HGC complex is very similar to that of the previous mentioned gapped binary structure containing the G:C pair (PDB ID=1BPX, r.m.s.d.=0.674Å).

Scrutinizing the N7meG:G base pair, as shown in Figure 2.12D, their electron density at 1 σ is clear enough to assign the molecular models. It is obvious that the N7meG is taking the *anti* conformation while the pairing G is taking the *syn* conformation. This *anti*:*syn* conformation has been observed before, either from NMR solution structures[85][90] or X-ray crystallographic structures[92]. Although *anti*:*anti* has been observed and suggested in an equilibrium with *anti*:*syn* base pair pattern, the majority of observed is *anti*:*syn*, probably because of its less bulging effect in the duplex than that of *anti*:*anti*. In all the observed G(*anti*):G(*syn*) forms, including the G quartet[93], the H bonds are based paired in the similar way of N1 to O6 and N2 to N7. As shown by the Figure 2.12D above, it is very clear that the N7meG:G mismatch observed in the polβ HGC structure is taking the canonical G(*anti*):G(*syn*) base pair patterns, with N1 to O6 and N2 to N7. However, there is obvious difference.

The G(*anti*):G(*syn*) base pair observed in the previous studies all have the H bond distances larger than the H bonds distances in the N7meG:G mismatch structure. In an early study[92], the distance between the N1 to the O6 (*anti* to *syn*, the same as followings) is around 2.8Å, and that N2 to N7 is around an average of 2.7Å. This is the similar case for the H bonds between the G(*anti*):G(*syn*) mismatch in the G quartet[93], with N1 to O6 distance now 2.8Å, and that the N2 to N7 distance is 3.2Å. However, in the similar Hoogsteen base pair of N7meG(*anti*):G(*syn*), the N1 to O6 distance is shortened to 2.5Å,

and N2 to N7 is 2.2Å. The comparison of H bond distances between the G(*anti*):G(*syn*) in G quartet (PDB ID: 2HY9) and N7meG(*anti*):G(*syn*) is shown in Figure 5H.

The shortened H bond between the N1 of N7meG and the O6 of the pairing guanine suggests that although the Hoogsteen base pair pattern is the same, the N1 position of the N7meG is likely to take the deprotonated form, while the O6 of the pairing guanine is probably taking a tautomer form of hydroxyl group instead of the carbonyl group (as shown below, Figure 2.13). So in this way O−H---:N is stronger than N−H---:O[94][95] with shorter H bond distance.

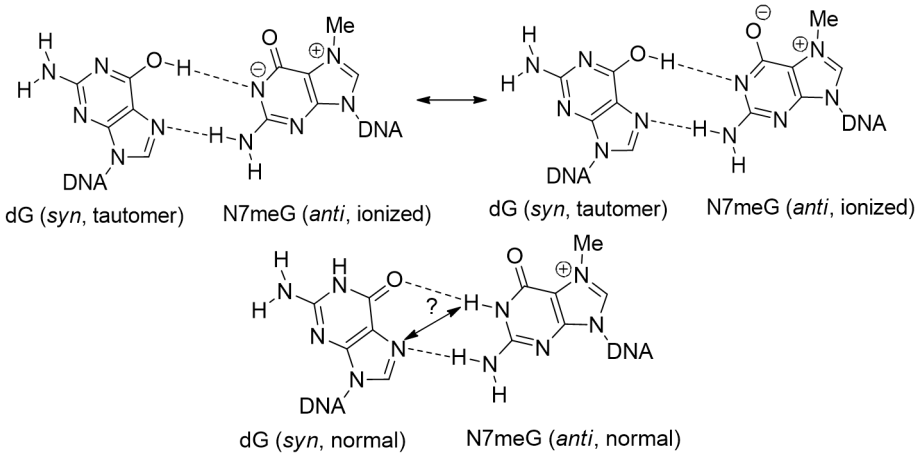


Figure 2.15 Possibilities of N7meG:G base pair pattern

This probability can also easily find other proofs, since we have seen that N7 methylation can and actually made the deprotonation of N1-H on the guanine, as from previous mentioned mismatches and studies. The N1 (G in *anti*)to N7 (G in *syn*) distance has been proved to be around proper H bond distance or ever shorter, in the previous study of G(*anti*):G(*syn*) structure in the B DNA[92], because of the possible tilted N1H to N7 H bond existence. However, this is not the case in the N7meG:G mismatch structure, as shown above, the N1 (N7meG) to N7 (dG) distance is out of the upper allowed limit of

the H bond distance, further suggesting that the possibility of N1 deprotonation upon N7 methylation.

There is also one small possibility that N7meG is taking the tautomer form with O6 now -O6-H, as shown below (Figure 2.14). However, by the structure we have, there is no H bond formed between the O6 atoms of N7meG:G base pair, not to mention the shielding effect of hydrogen atom from the -OH group.

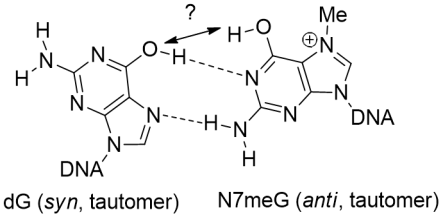


Figure 2.16 N7meG:G base pair tautomer pattern

Summarily, it is likely that N7meG is taking the ionized form with N1-H deprotonated to form the H bond to the tautomer pairing dG O6H. However, we can never completely rule out the possibility that the N1 still is bonded to carbonyl O6 through its N1-H hydrogen. But this possibility is small based on the structure presented here.

As can be seen from the Figure 5D, because of the distorted Hoogsteen base pair, the λ angles of this base pair are now λ1: 33.9˚, λ2:71.7˚, and the C1'-C1' distance is 10.6Å, which does not change much because bulge is shrunken by the *anti-syn* base pair pattern. However, the N7meG:G local base parameters still changes significantly, as can be seen from the following table:

Table 2.16 Local base pair parameters of N7meG:G

Local base pair parameters	Shear	Stretch	Stagger	Buckle	Propeller	Opening
N7meG:G	-1.94	-3.14	-0.58	14.48	15.17	77.37
N7meG:C	-0.14	-0.020	-0.33	10.68	-15.59	14.08
G:G	-1.70	-3.31	-0.76	13.87	6.75	80.49

The *syn* conformation of pairing G explains the huge increase of the open distortion observed. Additionally, we can see that the N7meG:G mismatch deforms the ideal coplanarity greatly by buckle and propeller like distortions (Figure 2.12E).

Although the neighboring base pairs does not seem to have that much distortions as compared to N7meG:C, they still exhibit obvious distortion concerning the coplanarity (Figure 2.12F):

Table 2.17 Neighbor base pair parameters of N7meG:G

Local base pair parameters	Shear	Stretch	Stagger	Buckle	Propeller	Opening
A5:T12 (N7meG:G)	0.17	-0.04	-0.03	6.14	-16.57	-2.27
T3:A14 (N7meG:G)	0.03	-0.16	0.67	5.10	-9.39	1.13
A5:T12 (N7meG:C)	0.21	-0.16	0.28	2.46	-17.39	-1.01
T3:A14 (N7meG:C)	-0.17	-0.12	0.23	7.49	-16.72	1.78
A5:T12 (G:G)	0.16	-0.07	0.18	4.58	-15.15	1.38
T3:A14 (G:G)	0.04	-0.11	0.17	-1.02	-12.33	0.60

The in-between base pair parameters are also deviated from normal ones or even from the N7meG:C, suggesting the neighboring environment is also in great distortion (Figure 2.12F):

Table 2.18 Local base pair step parameters of N7meG:G

Local base pair step parameters	Shift	Slide	Rise	Tilt	Roll	Twist
T12G13/N7meG4A5	1.68	-3.62	1.88	-174.57	14.85	52.75
G13A14/T3N7meG4	-3.47	-0.86	-1.40	129.08	-115.21	-7.21
T12C13/N7meG4A5	1.61	0.67	2.91	10.05	0.080	39.60
C13A14/T3N7meG4	-1.63	0.36	3.64	-8.23	4.03	32.99
T12G13/G4A5	1.16	-3.51	1.82	-175.10	19.19	35.52
G13A14/T3G4	-2.30	-1.99	-2.65	130.34	-117.82	76.82

Our thermodynamic data suggests that N7meG:G mismatch destabilizes the duplex from every aspect: decreased ΔH, ΔS and ΔG, and additional Tm lowering, as compared to the normal G:G mismatch. The above data can fairly explain that the N7meG:G is not only destabilizing itself compared to normal G:G pair, but also induces severe distortions of neighboring environment (major groove width also increases by 1Å compared to the normal G:C containing duplex). Although the stacking seems improved from the base overlapping are shown below, the improvement is limited compared to the severe distortion listed above, so it is justifiable to observe a Tm decrease of N7meG:G compared to that of the normal G:G in the DNA duplex.

Table 2.19 Neighbor bases overlapping area of N7meG:G

Neighbor bases overlapping	i1-i2	i1-j2	j1-i2	j1-j2	sum
T12G13/N7meG4A5	6.10(1.29)	0.00(0.00)	0.00(0.00)	1.51(0.00)	7.61(1.29)
G13A14/T3N7meG4	1.07(0.78)	0.00(0.00)	0.00(0.00)	6.58(1.26)	7.65(2.04)
T12C13/N7meG4A5	5.48(1.17)	0.00(0.00)	0.00(0.00)	1.65(0.00)	7.13(1.17)
C13A14/T3N7meG4	0.00(0.00)	0.00(0.00)	0.12(0.00)	6.96(2.04)	7.08(2.04)
T12G13/G4A5	6.39(1.73)	0.00(0.00)	0.00(0.00)	1.93(1.14)	8.31(1.73)
G13A14/T3G4	0.42(0.35)	0.00(0.00)	0.47(0.00)	4.63(1.09)	5.13(1.44)

It is obvious that the N7 methylation is the driving force for all these destabilizations. A rough comparison to the normal G:C base pair, we can see that the methylation on N7 shifts the guanine almost 2Å towards the major groove side, while the pairing G is staying at the similar position. This minor shift might be the initial cause. Further and more accurate conclusion needs the N7meG:G structure to be compared to the G:G structure with the same sequence.

2.4 SUMMARY OF CHAPTER II

In summary, N7meG can base pair with all four bases no matter the base pair pattern is Watson-Crick or not. Compared to the normal G base pairs and mismatches, because of the N7 methylation, the N7meG is in an N1 deprotonated condition. This changes almost all the known base pair patterns of G:X (X=A, T, G) in the DNA duplex. While the most lesions exert the destabilization effect on the duplex, the N7meG:T,

taking a novel and more opened pseudo-Watson-Crick base pair pattern, is surprisingly stabilizing the duplex from both structural and thermodynamic data. Undoubtedly, all these new knowledge would be very useful for unveiling the mismatch repair mystery on the N7-alkyl G lesions in the genomic DNA. And additionally, these data have proven the N1 deprotonation possibility. Associating with the fact that dTTP insertion to the N7meG in the template strand of polβ model is quite minimal, the data in this chapter suggests that dTTP insertion across N7meG might be a high energy and locally rate limiting step in maintaining the genomic integrity, as we can see from the mismatch data, once inserted, N7meG:T mismatch remains quite stable in the DNA duplex structure.

Since the G:T mismatch can also be converted from the G:C base pair by a naturally and spontaneously process. The stable and the pseudo-Watson Crick pattern of N7meG:T in DNA might imply a new gene regulation pathway by methylation: in which the sequence can be changed by mismatch. Undoubtedly, this requires to explore the percentage of N7meG:T in the whole genome and whether the N7 methylation affects the C to T conversion.

Chapter III: Aromatic and flexible N7 benzyl substituent of guanine displays different and metal dependent translesion synthesis pattern of human DNA polβ

3.1 BACKGROUND INTRODUCTION

Published evidence points to alkylator-DNA adduct formation as a key event in the carcinogenesis. Alkylators can come in various aromatic bulkiness, flexibility, and with additional functional groups causing significant damages via direct adduction on DNA. As introduced in previous chapters, N7 of guanine base is the major alkylation site *in vivo* due to its prominent nucleophilicity on the base ring and relatively unprotected status by hydrogen bonds. Alkylation damages are repaired by alkyladenine DNA glycosylase (AAG) in humans and AlkA in *E. coli*. Many DNA-alkylator adducts have half-lives of several hours to several days in DNA duplex, so, if not removed by alkylation-specific DNA glycosylases, these lesions can be a significant block to various DNA polymerases[14]. In addition, the presence of aflatoxin-N7G adduct and acridine half-mustard-N7dG adduct have been shown to induce G to T and G to A mutations, respectively[50][51]. The presence of N7-methyl group, the smallest alkyl group, on guanine has been shown to slow replication by DNA polymerase ~300-fold[14]. Therefore, compared to smaller and fixed methyl guanine, we hypothesize that a bulkier and flexible substituent on N7 of guanine may have more severe effects, such as delaying or even stopping DNA replication.

To test our hypothesis, we choose to study the effect of N7-benzyl guanine on DNA replication. We use N7-benzyl guanine for the following reasons. First, the benzyl group is bulkier than methyl group and contains an aromatic ring moiety, which sets as a good model for bulky carcinogens such as benzopyrene and anticancer drugs such as pluramycin antibiotics, and many flexible N7 alkylators such as styrene and butadiene

oxides. Second, benzyl-donating compounds widely exist around us, both industrially and environmentally. Benzyl halides are widely used as precursors as plasticizer, flavorants, and perfumes. Benzyl alcohol is widely used as solvent and well known in cosmetics, preservatives and flavoring industry. Furthermore, there are numerous benzyl derivatives used as food additives and contained in the industrial waste. Third, numerous studies have shown severe carcinogenic potentials of benzylating compounds both *in vivo* and *in vitro*. Benzyl halides have been demonstrated causing base-pair mutations in *Salmonella typhimurium*, inducing sister chromatid exchange in Chinese hamster cells [96], affecting the replication and repair mechanisms in fungi [97], inducing skin cancers [98][99], stomach carcinomas and bronchiolar adenoma[100] in mice, thyroidal carcinomas, hemangiosarcoma, hepatic and forestomach carcinoma[100] in rats. Benzyl alcohol could induce malignant lymphomas in mice, squamous cell papilloma in rats [101], and is highly teratogenic for chick embryo [102]. N-Nitrosobenzylmethylamine (NBzMA), a well-known esophageal carcinogen in laboratory animals [103][104][105], is believed to cause direct *in vivo* benzylation on DNA study [106]. Lastly, despite these many studies indicating the mutagenecity of benzyl group, molecular-level mechanism study for this mutagenic lesion has not been reported,.

Although the half life of N7 benzylation on guanine is relatively stable which can be compared to N7 methylation[107][108], N7 adduction phosphoramidites has long been known unstable for site-specific DNA oligo synthesis. To overcome this difficulty, we applied the same 2'-fluorine-mediated transition-state destabilization strategy[14]. So that for the first time, N7bnG is synthesized and inserted site specifically into the DNA template strand.

To evaluate the effect of N7-benzylguanine on DNA replication, we use human DNA polymerase β (polβ) as a model enzyme. As aforementioned, the X-family DNA polymerase polβ is an error-prone DNA polymerase lacking an intrinsic proofreading 3' to 5' exonuclease activity. Polβ participates in base-excision DNA repair by filling a short nucleotide gap generated by the action of DNA glycosylases and endonuclease. Furthermore, as aforementioned, polβ has been implicated in translesion synthesis of various types of DNA lesions such as 8-oxoguanine, cisplatin-GG intra-strand cross-link adducts, and UV-induced cyclobutane pyrimidine dimers .

Herein, we report the first synthesis of N7-benzylguaine-containing DNA, kinetic experiments for polβ catalysis across N7-benzylguaine, and three ternary structures of polβ incorporating dCTP or dTTP opposite templating N7-benzylguanine. Our studies provide insights into the effect of flexible and aromatic N7-alkyl-dG lesion on DNA replication.

3.2 EXPERIMENTAL PROCEDURE

To evaluate the effect of N7bnG on the polymerase activity of polβ, we first synthesized 2'-F N7bnG analogue (as the scheme shown below). We then determined kinetic parameters for the incorporation of dCTP and dTTP opposite the chemically stable N7bnG by polβ and solved three ternary structures of polβ in complex with the templating N7bnG paired with incoming nonhydrolyzable 2'-deoxycytidine-5'-triphosphate (dCTP*) or 2'-deoxythymidine-5'-triphosphate (dTTP*) analogues. Different from N7meG translesion synthesis, we report herein the first kinetic study which shows that dCTP insertion across N7bnG is severely slowed down upon using Mg^{2+}, but remains normal upon Mn^{2+}; and that there is some insertion of dTTP upon Mn^{2+}

presence. To explain this on a molecular level, we solved crystal structures of polβ inserting dCTP* and dTTP* opposite the templating N7bnG, which reveals for the first time the influence of Bn group hydrophobicity, Mn^{2+} ion dependence and new minor groove interactions in N7bnG base pairing.

Synthesis of 2'-F N7bnG (N7bnG) analogue

The N7bnG phosphoramidite was prepared starting from a 2'-fluoro ribose derivative (Scheme shown below). After phenoxyacetylation, the derivative 2 is submitted to benzylation (b). N7 position of guanine can be easily benzylated in short time at room temperature by benzyl halide, which also gives good yield. Then the 2'-fluoro N7 benzylated guanosine undergoes dimethoxytritylation (c) and phosphitylation (d) to give the final 2'-fluoro N7 benzylated guanosine phosphoramidite, which is ready for the site-specific DNA oligo synthesis. (The specific steps are shown below)

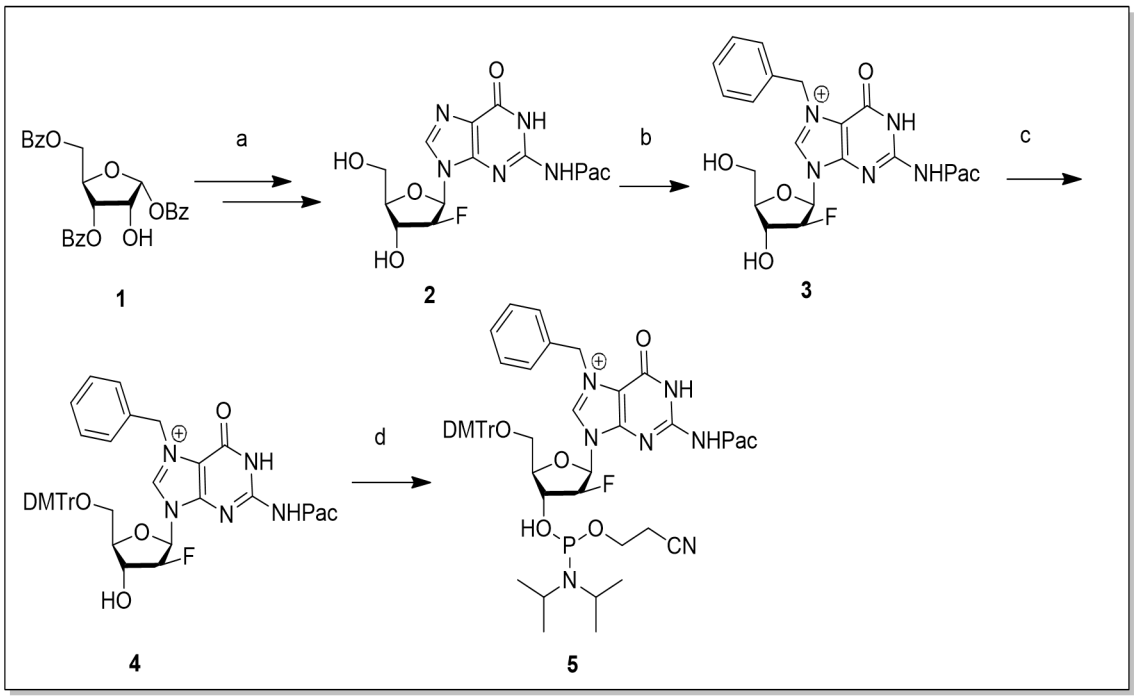


Figure 3.1 Synthesis of N7bnG amidite

7-benzyl-9-(3-fluoro-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-6-oxo-2-(2-phenoxyacetamido)-6,9-dihydro-1H-purin-7-ium (3)

To a suspension of 2′-fluoro deoxyguanosine derivative 2 (205mg, 0.76 mmol) in DMF (5 mL) was added benzyl bromide (1.8 mL, 15.2 mmol). After stirring at room temperature overnight, the resulting solution was poured into Et_2O and the resulting precipitate was collected by filtration and concentrated to afford 246 mg (90%) of benzylguanosine 3. 1H NMR (400 MHz, DMSO-d6) δ 10.15 (1H, br, s), 9.70 (1H, s), 9.53 (1H, s), 7.52-7.22 (7H, m), 6.94-6.82 (3H, m), 6.40 (1H, dd, J=3.6, 12.0 Hz), 6.13 (1H, d, J=3.8 Hz), 5.73 (2H, d, J=16 Hz), 5.34 (1H, dt, J= 3.6, 52.1 Hz), 5.19 (1H, t, J= 6.0 Hz), 4.93 (1H, d, J=9.6 Hz), 4.46 (1H, dd, J= 4.0, 12.8 Hz), 4.01 (1H, m), 3.68 (2H, m); HRMS calculated for C25H25FN5O6 (M+) 510.1784, found 510.1783.

7-benzyl-9-(5-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-3-fluoro-4-hydroxytetrahydrofuran-2-yl)-6-oxo-2-(2-phenoxyacetamido)-6,9-dihydro-1H-purin-7-ium (4)

To a pyridine solution of N7-benzyl deoxyguanosine 3 (125mg, 0.25 mmol) was added 4,4'-dimethoxytrityl chloride (125 mg, 0.37 mmol). And the mixture was stirred at the room temperature for 2 hours, which was then diluted with DCM and washed with saturated $NaHCO_3$ and brine. After drying over anhydrous Na_2SO_4, the organic layer was filtered, concentrated and purified by silica gel chromatography to give 161mg (80%) of tritylated compound 4. 1H NMR (400 MHz, DMSO-d6) δ 11.95 (1H, br, s), 9.90 (1H, s), 9.34 (1H, s), 7.50-7.14 (14H, m), 6.98-6.76 (7H, m), 6.43 (1H, dd, J=4.0, 12.4 Hz), 6.06 (1H, d, J=4.2 Hz), 5.69 (2H, d, J=9.2 Hz), 5.33 (1H, dt, J= 3.2, 52.4 Hz), 4.97 (1H, d, J=8.8 Hz), 4.59 (1H, dd, J=4.6, 16.8 Hz), 4.02 (1H, m), 3.72 (2H, m); HRMS calculated for C46H43FN5O6 (M+) 812.3109, found 812.3090.

7-benzyl-9-(5-((bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-4-(((2-cyanoethoxy)(diisopropylamino)phosphino)oxy)-3-fluorotetrahydrofuran-2-yl)-6-oxo-2-(2-phenoxyacetamido)-6,9-dihydro-1H-purin-7-ium

To a DCM solution of tritylated compound 4 (125mg, 0.15 mmol) was added 4, 5-dicyanoimidazole (12.7 mg, 0.11 mmole) and (2-cyanoethyl n n n' n'-tetraisopropylphosphorodiamidite) (85.2 μL, 0.3 mmol). After stirring at the room temperature for 1h, the mixture was concentrated and submitted to silica gel chromatography (+0.5% triethylamine) to afford 120.4 mg (78%) of N7-benzylguanine phosphoramidite 5. 31P NMR (400 MHz, CD3CN) δ 150.67, 150.37; HRMS calculated for C55H60FN7O9P (M+) 1012.4169, found 1012.4165.

DNA sequences used for kinetic and X-ray crystallographic studies

Oligonucleotides were synthesized by Midland Certified Reagent Co. (Midland, TX, USA) and Integrated DNA Technologies (Coralville, IA, USA). They were purified by the manufacturers and confirmed by MALDI-TOF mass spectrometry. The sequences used for kinetic assays were: upstream primer: 5'-FAM/GGGGGCTCGTAAGGATTC-3', downstream primer, 5'-phosphate/AGTCGG-3' and templat: 5'-CCGACT(X)GAATCCTTACGAGCCCCC-3', X = dG, 2FdG or N7bnG). The DNA sequences used for co-crystallization were 5'-CCGAC(N7bnG)TCGCATCAGC-3' for template, 5'-GCTGATGCGA-3' for the upstream primer and 5'-phosphate/GTCGG-3' for the downstream primer. The template DNA was prepared via solid-phase DNA synthesis using ultramild deprotection conditions (50 mM K_2CO_3 in MeOH, 25˚C, 24 h). The upstream, downstream primers and template were annealed to give a single-nucleotide gapped DNA, which was used for kinetics an the crystallization of both binary and ternary polβ complex structures[14].

Expression and purification of human DNA polymerase β

Polβ used for kinetic and crystallographic studies was expressed and purified from *Escherichia coli* as described previously[14].

Steady-state kinetics of nucleotide incorporation opposite N7bnG by polβ

Steady-state kinetic parameters for dCTP insertion opposite templating dG, 2'-fluoro-2'-deoxyguanosine (FdG) and N7meG by polβ were determined previously[14]. DNA substrates containing a single-nucleotide gap opposite the templating N7bnG were prepared by annealing the template oligonucleotide (5'-CCGACT(X)GAATCCTTACGAGCCCCC-3'; X = N7bnG) with the upstream (5'-FAM/GGGGGCTCGTAAGGATTC-3') and the downstream primers (phosphate/AGTCGG-3') at 95° C for 3 min followed by slow cooling to 22° C. Polβ activities were determined using the reaction mixture containing 50 mM Tris-HCl pH 7.5, 5 mM MgCl2, 100 mM KCl, 80 nM substrate DNA and varying concentration of the incoming nucleotide. The nucleotidyl transfer reactions were initiated by adding polβ. After incubated at 37˚C for 2 min, it was quenched by adding a stop solution containing 95% formamide, 45 mM Tris-borate, 20 mM ethylenediaminetetraacetic acid , 0.1% bromophenol blue and 0.1% xylene cyanol. The nucleotidyl transfer reaction products were resolved on 20% denaturing urea polyacrylamide gels, and the product formation was analyzed using a PhosphorImager (GE). The efficiency of nucleotide insertion was calculated as k_{cat}/K_m. The relative efficiency of dNTPs incorporation opposite N7bnG was determined as $f = (k_{cat}/K_m)_{[dNTP:N7bnG]}/(k_{cat}/K_m)_{[dNTP:dG]}$.

Co-crystallization and structure determination of polβ-DNA binary and ternary complexes

The binary and ternary polβ complexes with non-hydrolyzable dCTP or dTTP (Jena Bioscience) in the presence of 4mM $MgCl_2$ or $MnCl_2$ with templating N7bnG were crystallized using the similar conditions described previously[14]. Diffraction data were collected at 100 K at the beamline 5.0.3 at the Advanced Light Source, Lawrence Berkeley National Laboratory. All diffraction data were processed using HKL2000. Structures were solved by molecular replacement using a gapped binary complex structure with an open conformation (PDB ID: 3IBS) and a ternary complex structure with a closed conformation (PDB ID: 2FMS) as the search models. The model was built using COOT[156] and refined using Phenix[157].

3.3 EXPERIMENTAL RESULTS AND DISCUSSIONS

Table 3.1 Data collection and refinement statistics

PDB CODE	N7BG-gap-Mg^{2+} binary (4YMM)	N7BG C-Mg^{2+} ternary (4YMN)	N7BG C-Mn^{2+} ternary (4YMO)	N7BG T-Mn^{2+}ternary (4YN4)
Data Collection				
space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Cell Constants				
a (Å)	54.426	54.907	50.728	54.774
b	78.966	79.020	79.170	79.179
c	54.838	55.019	55.634	55.031
α (°)	90.00	90.00	90.00	90.00
β	106.34	107.92	107.24	106.53
γ	90.00	90.00	90.00	90.00
resolution (Å)[a]	20-2.20 (2.24-2.20)	20-2.58 (2.62-2.58)	20-2.15 (2.19-2.15)	20-2.25 (2.29-2.25)
R_{merge} [b] (%)	0.074 (0.439)	0.110 (0.545)	0.108 (0.425)	0.081 (0.411)
$\langle I/\sigma \rangle$	25.2 (3.02)	15.7 (2.48)	14.2 (2.12)	21.6 (2.68)
completeness (%)	100.0 (99.5)	100.0 (99.7)	100.0 (99.8)	99.9 (98.8)
redundancy	5.5 (5.2)	5.6 (5.2)	4.7 (4.1)	5.5 (5.2)
Refinement				
R_{work} [c]/R_{free} [d] (%)	19.9/25.7	20.3/26.9	19.6/22.6	20.5/25.5
unique reflections	22625	13902	22917	21437
Mean B Factor ($Å^2$)				
protein	38.69	41.40	24.69	37.36
ligand	38.81	34.62	35.36	36.77
solvent	36.08	34.93	27.95	33.48
Ramachandran Plot				
most favored (%)	96.6	95.0	97.8	94.4
add. allowed (%)	3.4	5.0	2.2	5.6
RMSD				
bond lengths (Å)	0.009	0.009	0.006	0.007
bond angles (degree)	1.445	1.618	1.471	1.487

[a] Values in parentheses are for the highest resolution shell.

[b] $R_{merge} = \Sigma|I-\langle I \rangle|/ \Sigma I$ where I is the integrated intensity of a given reflection.

[c] $R_{work} = \Sigma|F(obs)-F(calc)|/\Sigma F(obs)$.

[d] $R_{free} = \Sigma|F(obs)-F(calc)|/\Sigma F(obs)$, calculated using 5% of the data.

Insertion kinetics results

To assess whether N7bnG at the templating position affects the polymerase activity of polβ, we determined the kinetic parameters for polβ-catalyzed insertion of dCTP and dTTP opposite N7bnG (Table 3.1 below). In the absence of the 5'-phosphorylated downstream primer (5'-phosphate/AGTCGG-3'), the catalytic efficiency (k_{cat}/K_m) for dCTP insertion opposite the templating dG is very low. The use of the downstream primer dramatically increases the insertion efficiency (~1100-fold), which is consistent with previous kinetic results. We previously reported that 2'-fluorination has a little effect on polβ catalysis and that upon N7 methylation the insertion of incoming dCTP efficiency decreased by ~300-fold compared to that of temlating G[14]. In the presence of the active-site Mg^{2+}, N7-benzylguanine adduct reduces the catalytic efficiency of dCTP incorporation ~6000-fold as compared to templating dG, which is 20-fold less than that of templating N7meG. This suggests that the bulky N7-benzyl group deters the insertion of dCTP more strongly than N7-methyl group does. In the presence of the active site Mn^{2+}, insertion efficiency of dCTP* opposite templating N7-benzylG decreased by only 2 fold as compared to that opposite templating N7-methylG . This suggests that there might be metal- or benzyl group-dependent conformational difference during nucleotide incorporation. In addition to the N7BndG:dCTP kinetics, we wanted to examine the effect of the N7-benzylation on the formation of G:T mismatches, which comprise about 60% of polβ-induced spontaneous base substitutions[52]. Unlike N7meG case, where there is no dTTP insertion in the presence of the active site Mg^{2+} and Mn^{2+}, for the N7bnG at templating position, dTTP insertion is observed in the presence of Mn^{2+}, although it is about 10-fold less efficient than dCTP* insertion. This kinetic difference suggests that structurally flexible and aromatic benzene group, unlike methyl

group in N7meG, may induce changes in the active site conformation in a way to facilitate the incorporation of incoming dTTP opposite the lesion.

Table 3.2 Kinetics of dNTP insertions across templating N7bnG by polβ

Template:dNTP	pH	K_m (μM)	k_{cat} ($10^{-3}s^{-1}$)	k_{cat}/K_m ($10^{-3}s^{-1}\mu M^{-1}$)	f[a]
dG:dCTP[b]	7.5	269.55 ± 18.87	7.56 ± 0.38	0.03	-
dG:dCTP[c]	7.5	0.59 ± 0.03	20.38 ± 0.50	34.54	1
FdG:dCTP[c]	7.5	0.3 ± 0.07	12.52 ± 0.53	36.82	1.1
N7bn:dCTP[b,d]	7.5	-	-	-	-
N7bnG:dCTP (Mg^{2+})[c]	7.5	20.24 ± 0.06	0.11 ± 0.02	0.0054	1.57×10^{-4}
N7bnG:dCTP (Mn^{2+})[c]	7.5	6.15 ± 0.09	0.32 ± 0.03	0.052	1.51×10^{-3}
N7bnG:dTTP (Mg^{2+})[c,d]	7.5	-	-	-	-
N7bnG:dTTP (Mn^{2+})[c]	7.5	23.20 + 0.05	0.10 ± 0.02	0.0043	1.25×10^{-4}

[a]Relative efficiency: $(k_{cat}/K_m)_{[dNTP:N7bnG]}/(k_{cat}/K_m)_{[dCTP:dG]}$.

[b]A recessed DNA (i.e. DNA without the downstream primer) was used as a substrate.

[c]Single-nucleotide gapped DNA was used as a substrate.

[d]Nucleotide incorporation was not observed or minimal.

Binary structure of polβ in complex with DNA bearing a single-nucleotide gap opposite N7bnG

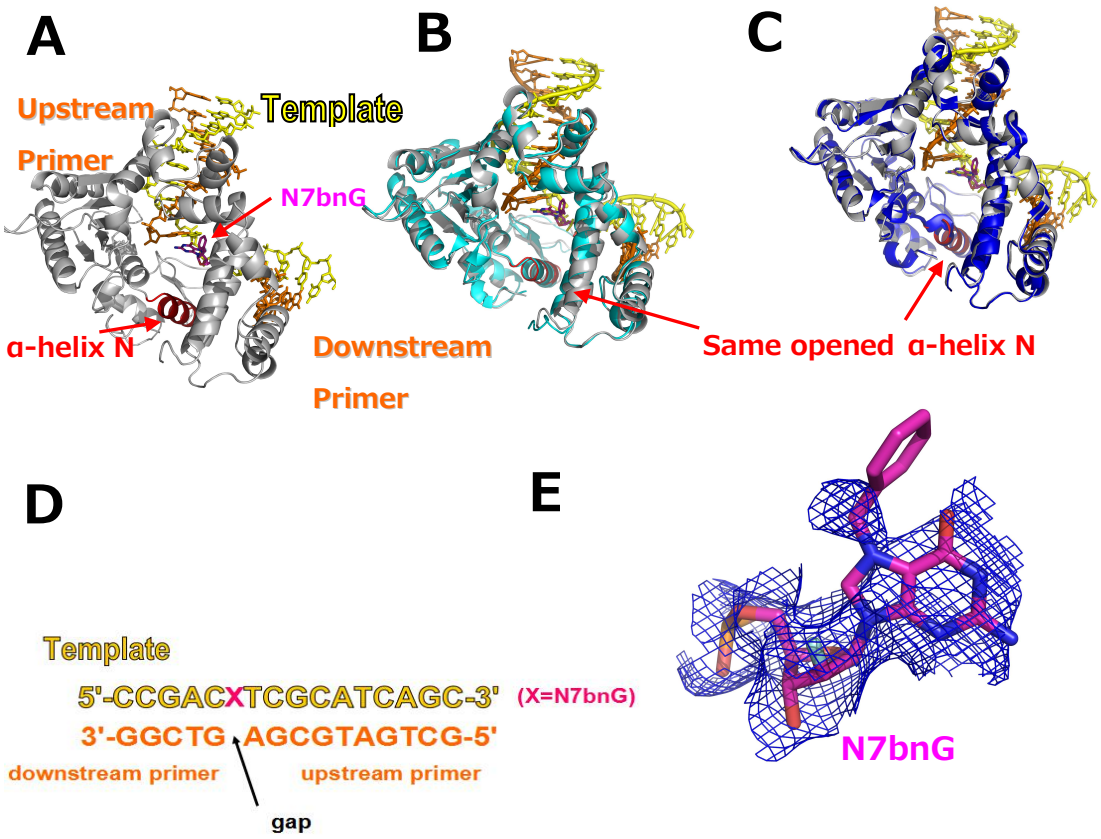


Figure 3.2 N7bnG binary structure. (A) Overall view of polβ-N7bnG gapped binary structure. Template strand shown in yellow, primer strands in brown, and protein in gray. N7bnG is shown in purple. (B) Alignment of the known gapped polβ binary structure(PDB ID: 1BPX, in cyan) with polβ-N7bnG. (C) Alignment of the known gapped polβ-N7meG binary structure(PDB ID: 4O5C, in blue) with polβ-N7bnG. (D) Sequence of N7bnG template with upstream and downstream primers. (E) Electron density of N7bnG contoured at 1 σ .

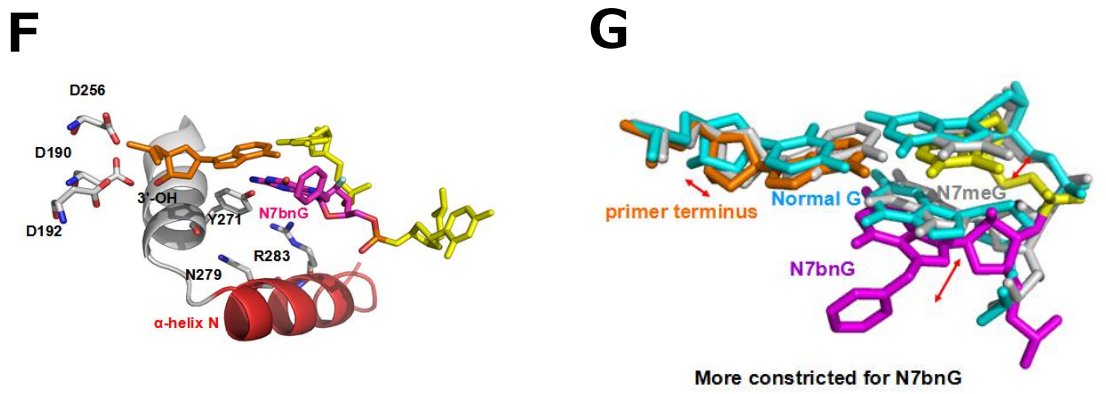


Figure 3.2 N7bnG binary structure. (F) Active site view of polβ-N7bnG. (G) Shifts of N7bnG with surrounding nucleobases (purple, brown, yellow) compared to N7meG (gray, PDB ID: 4O5C), normalG (cyan, PDB ID: 1BPX).

To evaluate whether N7bnG at the templating position perturbs the conformation of DNA and protein, we solved a binary structure of polβ bound to DNA containing a single-nucleotide gap opposite the templating N7bnG (PDB ID: 4YMM. Figure 3.1A). The gapped binary structure, refined to 2.2 Å resolution, shows that N7bnG does not induce any substantial distortion of the polβ overall conformation and is similar to that of the published gapped structure of normal templating dG (Figure 3.1B, PDB ID: 1BPX, RMSD = 0.775 Å) in which, polβ assumes an open conformation of α-helix N. The crucial minor-groove interacting amino acids Asn279 and Arg283 do not engage in hydrogen bonding with DNA (Figure 3.1F). The templating N7bnG adopts an anti-base conformation and C2'-endo sugar pucker, as does normal dG in the template. Although the 2'-F is evident on the electron density map, the N7-Bn moiety is not so clear (Figure 3.1E), indicating its flexibility endowed by the sp^3 hybridized methylene carbon. However, compared to normal gapped dG structure and N7meG (PDB ID: 4O5C, RMSD = 0.327 Å), there are some interesting structural changes with N7bnG binary structure

(Figure 3.1G): i) Compared to normal dG, N7meG base shifts towards minor groove, whereas N7bnG base moves far away from the minor groove, causing a distance of 2.7 Å between the N atom of exocyclic amines. Thus a previously observed hydrogen bond between Tyr 271 side chain to exocyclic amine of N7meG does not exist (Figure 3.1F). ii) Along with this movement, phosphate backbone and the 5'and 3' sides of N7bnG shifts far away from the minor groove of normal dG, where as N7meG shifts slightly toward the minor groove. iii) Interestingly, although N7bnG moves in opposite direction compared to N7meG on the templating strand, on the upstream primer strand, the 3'-OH at the terminus moves similarly as that of N7meG towards the templating strand, with 1.3 Å distance from that of the normal dG structure. Overall, the N7bnG gapped binary complex structure indicates that templating N7bnG induces a slight conformational change of DNA around the lesion. Compared to methylation, in the binary structure without incoming nucleotide, the bulkier benzylation on N7 position seems to create a more perturbed active site.

Ternary structure of polβ incorporating dCTP analog opposite templating N7bnG in the presence of the active site Mg^{2+}

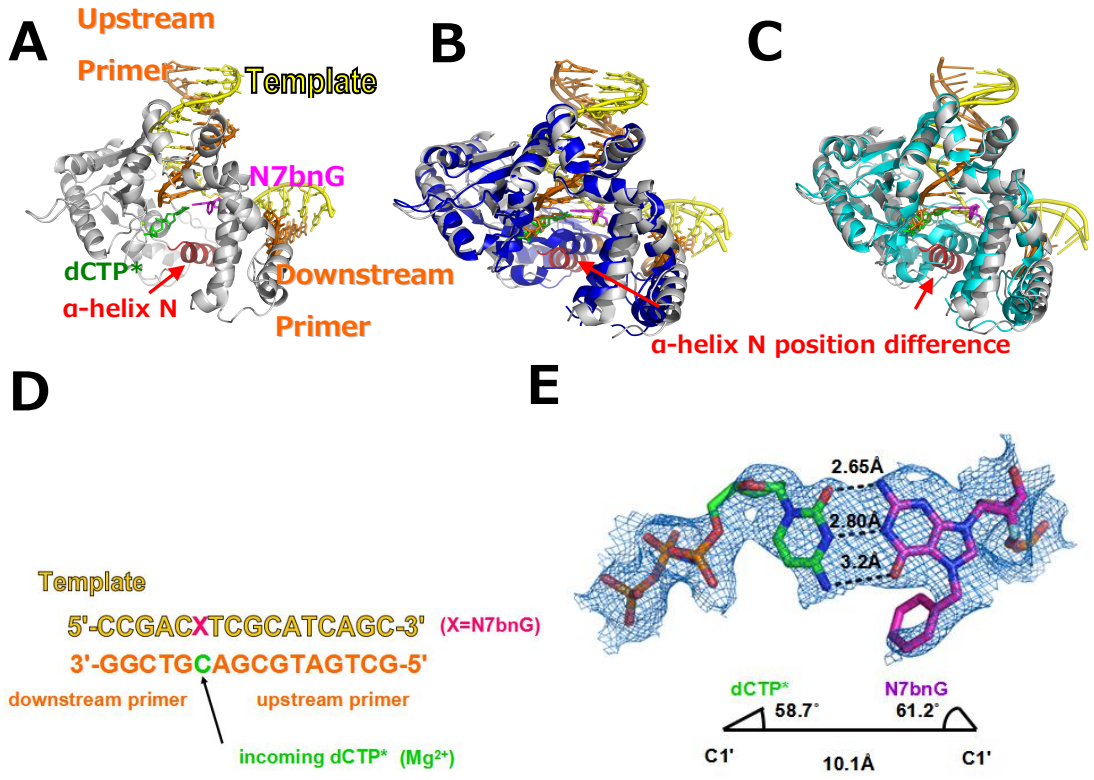


Figure 3.3 N7bnG:C structure. (A) Overall view of polβ-N7bnG:dCTP*-Mg^{2+} structure. Template strand shown in yellow, primer strands in brown, and protein in gray. N7bnG is shown in purple. (B) Alignment of the polβ-N7meG:dCTP* ternary structure(PDB ID: 4O5K, in blue) with polβ-N7bnG:dCTP*-Mg^{2+}, α-helix N is in red in N7bnG structure. Note its position difference to N7meG structure (in blue). (C) Alignment of the known gapped polβ-dG:dCTP-Mg^{2+} binary structure(PDB ID: 2FMS, in cyan) with that of N7bnG structure(α-helix N is in red). (D) Sequence of N7bnG template with upstream and downstream primers, and with incoming dCTP*. (E) Electron density of N7bnG:dCTP* contoured at 1 σ .

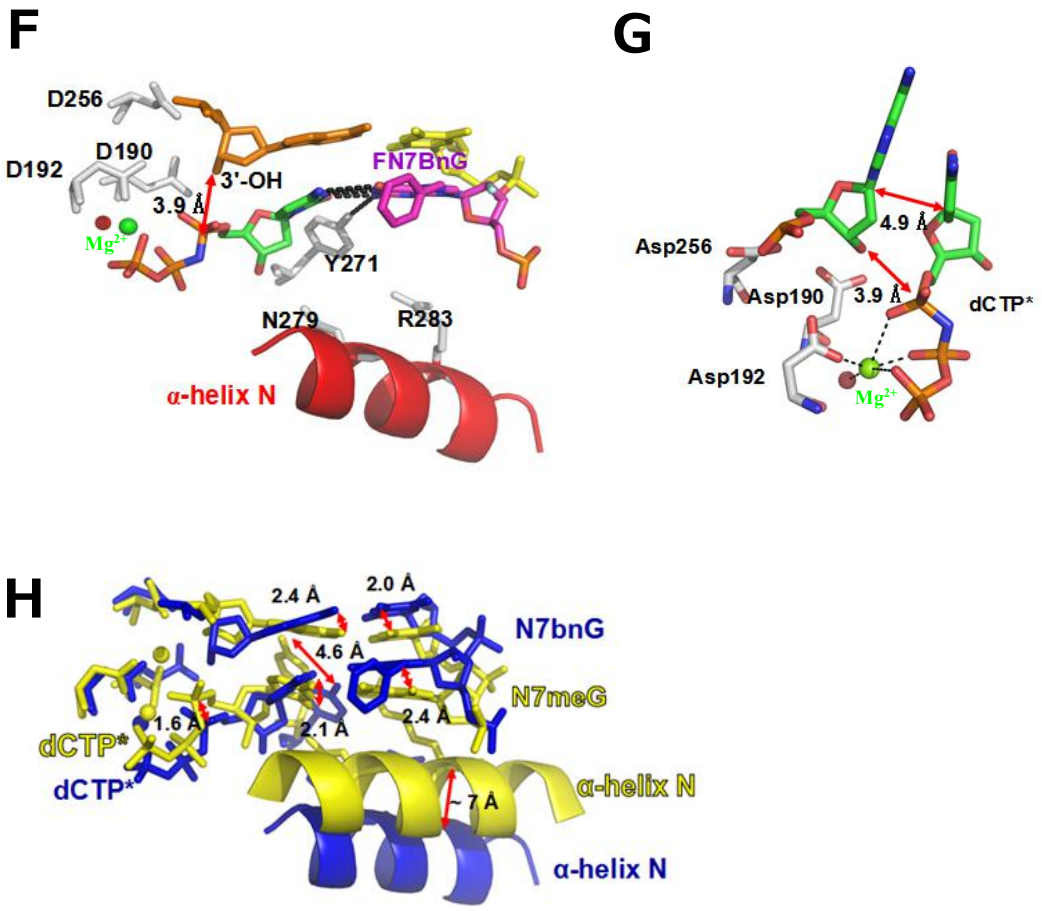


Figure 3.3 N7bnG:C structure. (F) Active site view of polβ-N7bnG:dCTP*-Mg^{2+}. (G) Coordination of nucleotidyl transfer reaction area of polβ-N7bnG:dCTP*-Mg^{2+}, binding Mg^{2+} in green sphere, water in red sphere, note there is no catalytic metal ion. (H) Active site shifts of polβ-N7bnG:dCTP*-Mg^{2+} (blue) compared to polβ-N7meG:dCTP*-Mg^{2+} (PDB ID: 4O5K, yellow).

To gain insight into the observed slow insertion of dCTP opposite N7bnG by polβ, we solved ternary structures (PDB ID: 4YMN, 4YMO) of polβ with templating N7bnG base-paired with the incoming nonhydrolyzable dCTP analog dCMPNPP (dCTP*

hereafter) in the presence of the active-site Mg^{2+} or Mn^{2+}. The polβ-N7bnG:dCTP*-Mg^{2+} ternary complex structure was refined to 2.6 Å resolution.

Interestingly, the overall structure of the N7bnG:dCTP*-Mg^{2+} ternary complex (Figure 3.2A) is very different from dG:dCTP*-Mg^{2+} ternary structure (PDB ID: 2FMS, RMSD = 2.078 Å, Figure 3.2C) and N7meG:dCTP*-Mg^{2+} ternary structure (PDB ID: 4O5K RMSD = 2.115 Å Figure 3.2B). Whereas N7meG:dCTP*-Mg^{2+} ternary structure is essentially identical to the normal dG:dCTP*-Mg^{2+} structure, N7bnG:dCTP*-Mg^{2+} structure is largely different from N7meG:dCTP*-Mg^{2+} structure. First, α-helix N is ~7 Å away relative to the closed conformation of N7meG or normal dG, taking an open protein conformation(Figure 3.2B). Second, despite roughly good nucleotidyl transfer reaction distance between the 3'-OH of primer terminus and α phosphor atom of incoming dCTP*, the critical catalytic Mg^{2+} is not present in the active site (Figure 3.2G). Therefore, catalytically critical coordination of the catalytic metal ion is absent and the Asp 256 is taking the catalytically incompetent conformation (Figure 3.2F). This causes the incoming dCTP* shifting away from the 3' primer terminus by additional 1.6 Å, thus rotates the cytosine ring 1.6 Å away from the ideal coplanar position, tilted to the small extent of the coplanarity of the base pairing (Figure 3.2F). However, since the incoming dCTP* is still making the traditional Watson-Crick base pair with N7bnG, so the tilted dCTP* base ring shifts the all the active site nucleobases compared to the N7meG:dCTP*-Mg^{2+} structure (Figure 3.2H). Lastly, there are no minor groove interactions of DNA with Asn279 or Arg283 on the α-helix N. And the Tyr271 is not making the hydrogen bond to the primer terminus to stabilize it as in the closed conformation structure. Instead, it is hydrogen bonded to the N7bnG (Figure 3.2F). The open conformation of polβ-N7bnG:dCTP*-Mg^{2+} ternary complex structure lacks all these crucial minor groove interactions that are observed in the closed conformation.

In summary, unlike N7meG, N7bnG:dCTP*-Mg^{2+} ternary complex takes the open conformation due to lack of the crucial catalytic ion and loss of minor groove interactions. Although these lead to some minor conformation change of the base pair. However, the N7bnG: dCTP* base pair is still taking the near ideal Watson-Crick base pair, which explains its kinetics of 10 fold decrease of insertion efficiency compared to N7meG. Compared to gapped binary structure, the benzyl group density in this ternary structure can be observed, it is bending towards the O6 atom of the base ring.

Ternary structure of polβ incorporating dCTP analog opposite templating N7bnG in the presence of the active site Mn^{2+}

As aforementioned, when Mn^{2+} is present, dCTP insertion to N7bnG increases about 5 fold compared to Mg^{2+} presence. To gain insights into the insertion efficiency difference between N7bnG:dCTP-Mg^{2+} and N7bnG:dCTP-Mn^{2+} complexes, we determined ternary structure of polβ-N7bnG:dCTP* in the presence of Mn^{2+}. The structure was refined to 2.15 Å resolution.

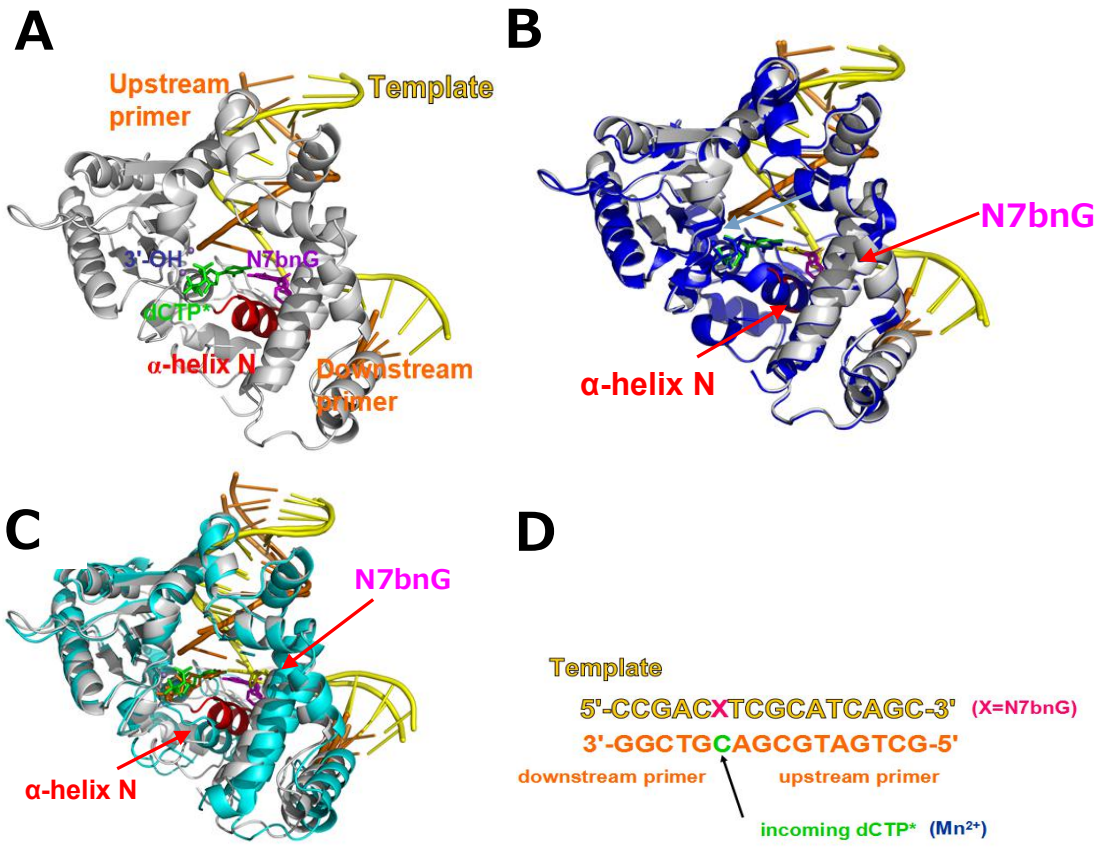


Figure 3.4 N7bnG:dCTP*-Mn^{2+} structure. (A) Overall view of polβ-N7bnG:dCTP*-Mn^{2+} structure. Template strand shown in yellow, primer strands in brown, and protein in gray. N7bnG is shown in purple. (B) Alignment of the polβ-N7meG:dCTP* ternary structure(PDB ID: 4O5K, in blue) with polβ-N7bnG:dCTP*-Mn^{2+}, α-helix N is in red in N7bnG structure. Note its position is almost the same to N7meG structure (in blue). (C) Alignment of the known gapped polβ-N7bnG:dCTP*-Mg^{2+} (in cyan) with that of Mn^{2+} structure(α-helix N is in red, note its position difference). (D) Sequence of N7bnG template with upstream and downstream primers, and with incoming dCTP*.

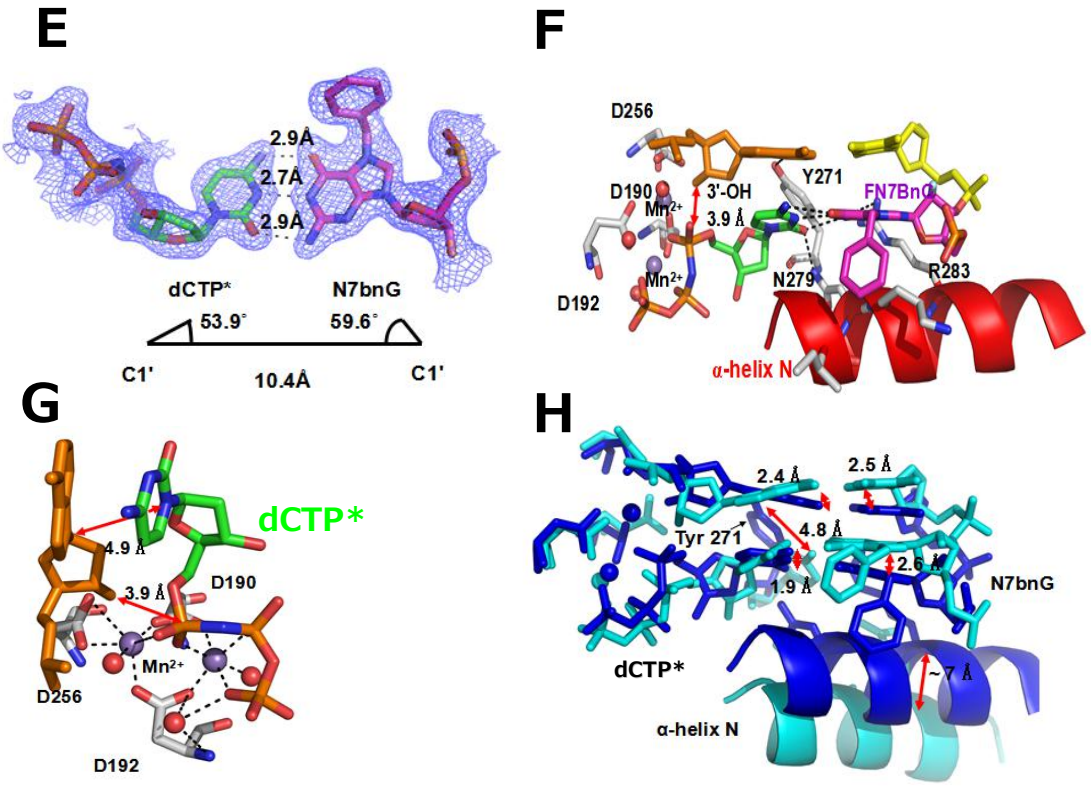


Figure 3.4 N7bnG:dCTP*-Mn^{2+} structure. (E) Electron density of N7bnG:dCTP* at presence of Mn^{2+} contoured at 1 σ . (F) Active site view of polβ-N7bnG:dCTP*-Mn^{2+}. Waters shown in red spheres. (G) Coordination of nucleotidyl transfer reaction area of polβ-N7bnG:dCTP*-Mn^{2+}, waters in red sphere. (H) Active site comparison of polβ-N7bnG:dCTP*-Mg^{2+} (cyan) with polβ-N7meG:dCTP*-Mn^{2+} (blue).

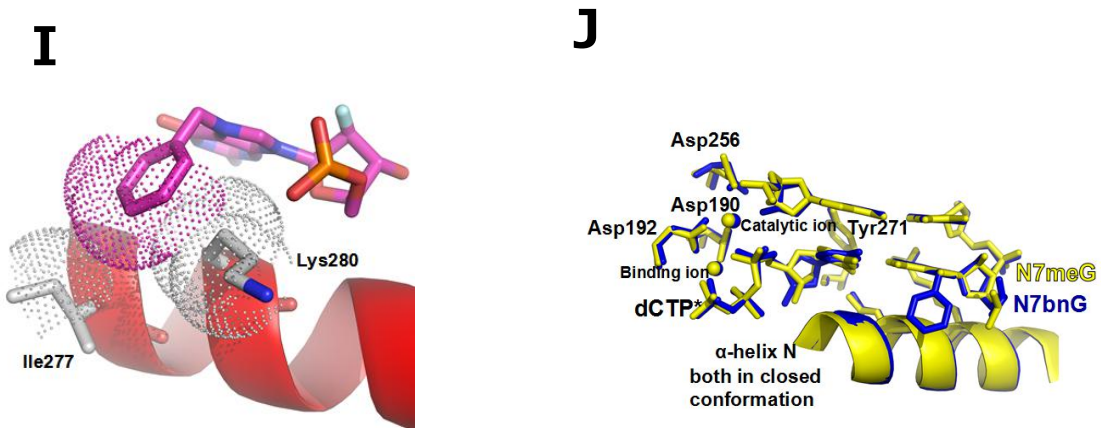


Figure 3.4 N7bnG:dCTP*-Mn^{2+} structure. (I) Bn group of N7bnG is making the hydrophobic interactions with Ile277 and carbon atoms on Lys280 side chain, as the vdw radius indicated by dots. (J) Comparison of N7bnG:dCTP*-Mn^{2+} active site (blue) with that of the N7meG:dCTP*-Mg^{2+} (yellow).

Substitution of the active-site Mn^{2+} for Mg^{2+} triggers an open to closed conformational transition of the polβ-N7bnG:dCTP* complex (Figure 3.3A, 3.3B). The N7bnG:dCTP*-Mn^{2+} structure is now essentially identical to the N7meG:dCTP*-Mg^{2+} structure (Figure 3.3B and 3.3J, PDB ID: 4O5K, RMSD = 0.333Å). Unlike the N7bnG:dCTP*-Mg^{2+} structure (Figure 3.2C), the N7bnG:dCTP*-Mn^{2+} structure reveals the presence of the catalytic Mn^{2+} and coordination of the catalytic metal ion with the three catalytic Asp residues: Asp 190, 192 and 256, the primer terminus 3'-OH, the phosphate group of incoming dCTP* and a water molecule (Figure 3.3F). The nucleotidyl transfer reaction distance between Pα of the incoming nucleotide and the 3'-OH of the primer terminus is 3.9 Å and the base ring of dCTP* is positioned to form an ideal coplanar Watson-Crick base pair to N7bnG (Figure 3.3F). Compared to the N7bnG:dCTP*-Mg^{2+} structure, Tyr271 now rotates ~ 4.6 Å and is hydrogen-bonded to minor-groove edge of primer terminus to stabilize it (Figure 3.3H). The electron density

of Bn group of N7bnG is very clear in this Mn^{2+} ternary structure (Figure 3.3E), indicating Bn group fixation by surrounding interactions. Compared to Bn group in Mg^{2+} structure, the Bn group on the N7 position now rotates towards α-helix N (Figure 3.3H). And rather than bending towards O6 atom as in the gapped binary structure, it is now positioned downward in a small hydrophobic pocket created by Ile277 and the alkyl β, γ, δ carbons of Lys280 from the α-helix N (Figure 3.3I). The Bn group and the alkyl side chains of Ile277 and Lys280 are within the hydrophobic interaction range, indicating hydrophobic interactions among them. For minor groove recognition, Asn279 and Arg283 engages respectively in hydrogen bonding with minor groove edge of incoming dCTP* and templating N7bnG (Figure 3.3F).

As can be seen from the comparison (Figure 3.3H), The replacement of the active site Mg^{2+} with Mn^{2+} induces drastic conformational change in the polβ-N7bnG:dCTP* complex structure by promoting completion of the coordination of the catalytic metal ion. This essential catalytic metal ion draws near all the correct coordinations, promotes the minor groove recognition of Tyr271 and corrects all the insertion imperfects from the incoming dCTP* side (Figure 3.3F, 3.3G). On the other hand, the Bn group rotation from the base ring plane towards α-helix N, brings it to a closed conformation such that crucial minor groove recognition of Asn 279 and Arg 283 with incoming dCTP* and N7bnG are now established (Figure 3.3F). This corrects all the insertion imperfects from the N7bnG side: the position of N7bnG in the Mn^{2+} structure now shifts about 2.5 Å towards the minor groove and positions exactly the same as N7meG in dCTP*-Mg^{2+} structure (Figure 3.3J). So by the help of Mn^{2+} catalytic ion and Bn group rotation, the ideal coplanarity and Watson-Crick base pair between dCTP* and N7bnG are now established. This catalytically favorable conformation seen in the polβ-N7bnG:dCTP-Mn^{2+} structure well

explains the 10 fold increased insertion efficiency of dCTP* when the active site Mg^{2+} is replaced by Mn^{2+} (Table 3.0).

Ternary structure of polβ incorporating dTTP analog opposite N7bnG in the presence of Mn^{2+}

As described in kinetic studies (Table 3.0), we observed, albeit inefficient, insertion of dTTP* opposite templating N7bnG in the presence of Mn^{2+}. The insertion efficiency is comparable to that of the dCTP* insertion opposite N7bnG in the presence of Mg^{2+}. To gain insight into this discovery, the N7bnG:dTTP*-Mn^{2+} ternary structure is solved to 2.2 Å resolution.

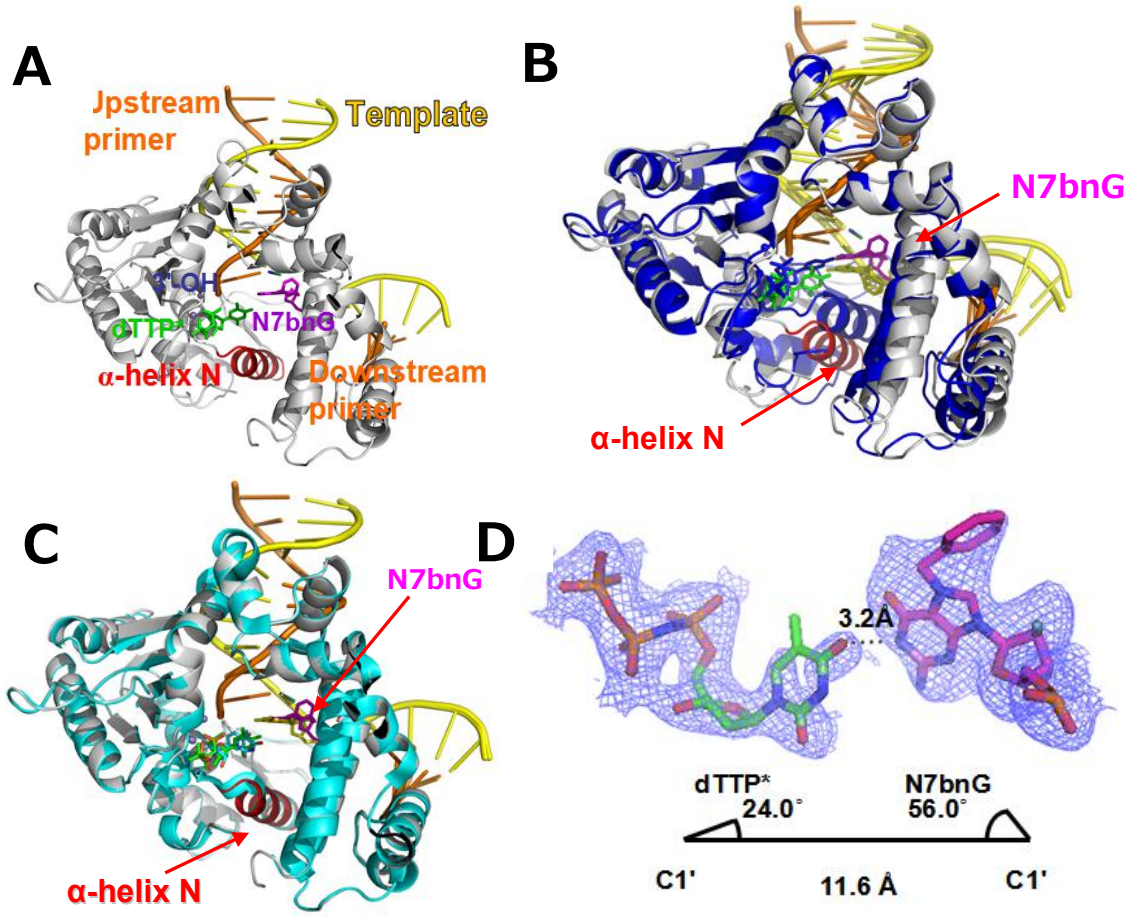


Figure 3.5 N7bnG:dTTP*-Mn^{2+} structure. (A) Overall view of polβ-N7bnG:dTTP*-Mn^{2+} structure. Template strand shown in yellow, primer strands in brown, and protein in gray. N7bnG is shown in purple. (B) Alignment of the polβ-N7bnG:dCTP*-Mn^{2+}ternary structure (in blue) with polβ-N7bnG:dTTP*-Mn^{2+}, α-helix N is in red in dTTP* structure. Note its position is different to that in dCTP*-Mn^{2+} structure (in blue). (C) Alignment of the known gapped polβ-N7meG:dTTP*-Mn^{2+} (PDB ID: 4P2H, in cyan) with polβ-N7bnG:dTTP*-Mn^{2+} structure. (α-helix N is in red, note their positions are essentially the same open conformation). (D)Electron density of N7bnG:dTTP* at presence of Mn^{2+} contoured at 1 σ .

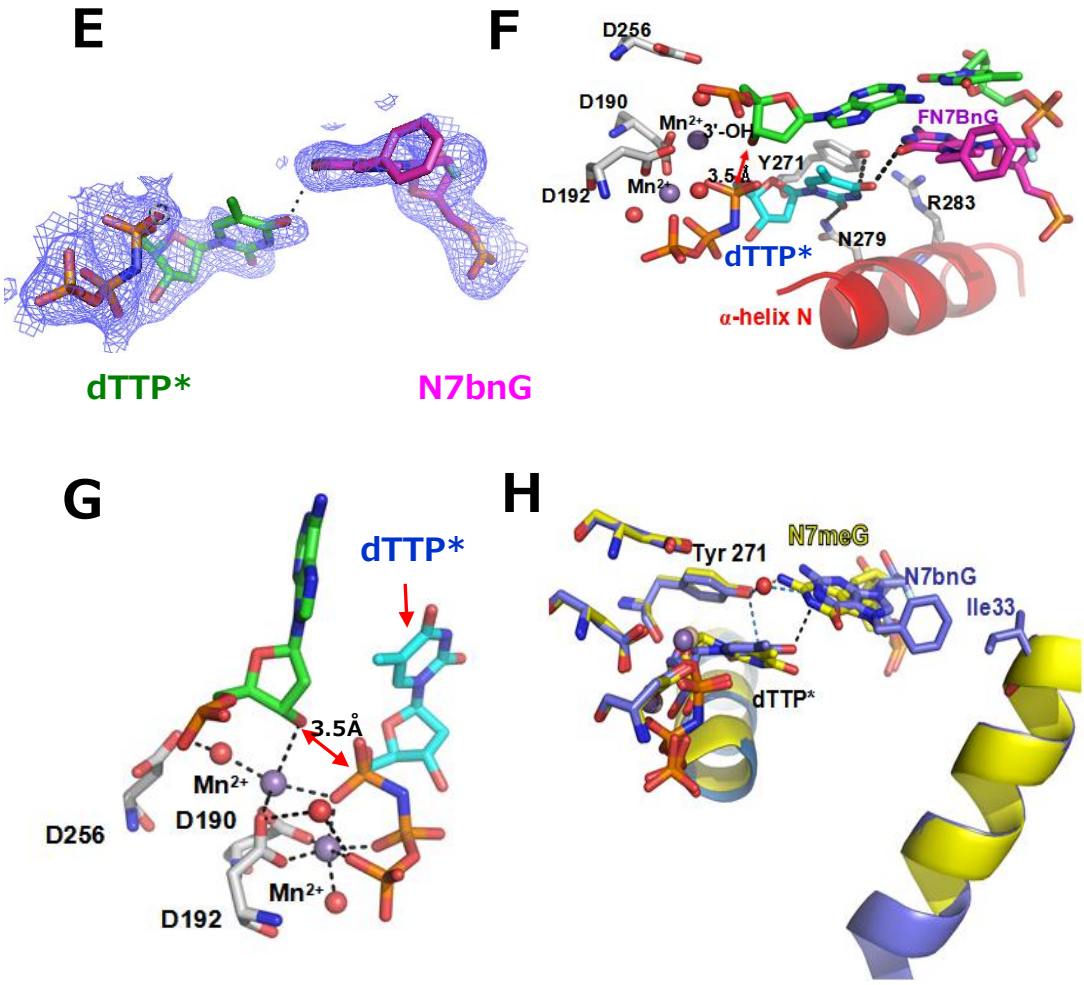


Figure 3.5 N7bnG:dTTP*-Mn^{2+} structure. (E) N7bnG:dTTP* base pair viewed with electron density at 1 σ from the major groove side. (F) Active site view of polβ-N7bnG:dTTP*-Mn^{2+} structure. (G) Coordination of nucleotidyl transfer reaction area of polβ-N7bnG:dTTP*-Mn^{2+}, waters in red sphere. (H) Active site comparison of polβ-N7meG:dTTP*-Mn^{2+} (PDB ID: 4P2H, in yellow) to polβ-N7bnG:dTTP*-Mn^{2+} (blue). Note that more staggered base pair conformation of N7meG:dTTP* compared to more relaxed N7bnG:dTTP* conformation (1 H bond in between).

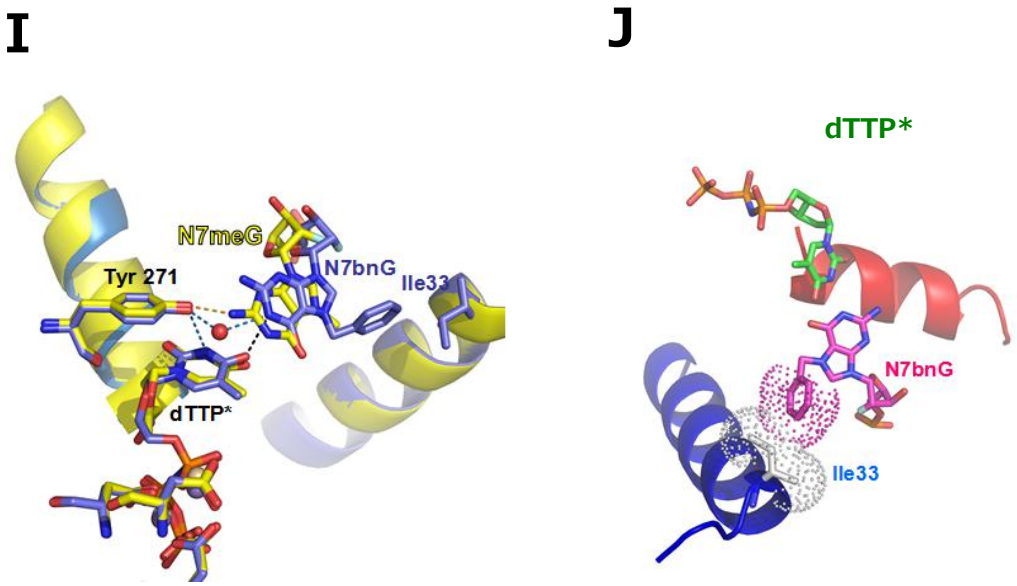


Figure 3.5 N7bnG:dTTP*-Mn^{2+} structure. (I) top view of (H). (J) Hydrophobic interaction of N7bnG Bn group with the side chain of Ile33, with vdw radius shown in dots.

The overall of this ternary structure is very similar to that of the N7MeG: dTTP*-Mn^{2+} (Figure 3.4C, PDB ID: 4P2H, RMSD: 0.280Å), with α-helix N in the near open conformation. However, we cannot observe any dTTP* insertion across the N7MeG. On the opposite, we have observed the insertion of dTTP* towards N7BnG, which is even comparable to the dCTP* insertion when Mg^{2+} is present. Considering that Bn group has the methylene carbon which is in the same position of N7 methyl group, this insertion difference is confusing.

However, after taking a scrutiny at the active site, we find that there exists the hydrogen bond (3.2 Å) formed by O4 atom from the incoming thymine base ring to the N1 atom of the N7bnG guanine base, as can be seen from the Figure 3.4D. In the N7meG: dTTP*-Mn^{2+} structure, this hydrogen bond does not exist, the distance between O4 and N1 is beyond the upper limit. Furthermore, N7meG:dTTP* base pair is stacking

with each other thereby creating a staggered overlapped conformation. In contrast, N7bnG: dTTP* base pair shows a conformation with less overlapping and more spreading towards a coplanar like conformation between the two (Figure 3.4H). As can be seen from the comparison (Figure 3.4H and 3.4I), less stacking is partially created by the rotation of the N7bnG guanine base towards the minor groove as compared to the N7MeG guanine base. When aligned together (Figure 3.4H and 3.4I), these two base rings are on the same plane, but the base ring of N7bnG rotates around overlapped N1 about 46.9˚ (C8-N1-C8') towards the minor groove as compared to N7meG base ring. The only difference between the two are the N7 adduction groups. In the N7bnG: dTTP*-Mn^{2+} structure, Bn group swings towards the other side, dragging the base ring to minor groove. As can be seen in the structure (Figure 3.4J), Bn group is having the hydrophobic interactions with the side chains of Ile33. PISA interface calculation confirmed a buried surface area of ~7.5 $Å^2$ between the two. This hydrophobic interaction from Ile 33 is located well on the surface of its α-helix by a position that is equal and far away from the neighboring hydrophilic side chains of His34, Tyr36 and Gln31, creating a hydrophobic channel surrounded by the hydrophilic environment of helix side chains and solvent (Figure 3.4J).

The hydrophobic interaction of the Bn group with Ile 33 not only relaxes the overlapped staggering base pair towards a more coplanar like conformation, but it also helps to stabilize and draws near the incoming dTTP*. As can be seen from the Figure 4H and 4I, because of the rotation dragged by the Bn group, the N1 atom of the N7bnG base is making the hydrogen bond to a nearby water molecule. This water molecule is then hydrogen bonded to OH group of the Tyr271, enable which to make the hydrogen bond to N3 of the incoming dTTP* and stabilize it. This hydrogen bonding stabilization is not observed in N7meG: dTTP*(Mn^{2+}) structure. Additionally, compared to N7meG:

dTTP* (Mn^{2+}) structure, this hydrogen bonding network draws the thymine ring of dTTP*~1 Å more towards the base pair plane, making O4 atom of dTTP* in the optimal position to form the hydrogen bond with the N1 atom from the N7bnG guanine base (Figure 3.4H and 3.4I). In the N7meG: dTTP* (Mn^{2+}) structure, since the N7meG base ring is not rotated towards the minor groove, Tyr271 directly makes the H bond with N2 of the N7meG guanine ring with no water mediating in between. This causes a small shift of Tyr271 and incoming dTTP*. Thus there is no dTTP* stabilization by Tyr271 and no hydrogen bonding between dTTP* and N7meG.

Polβ' s active site does not promote the formation of zwitterionic N7-alkylguanine

It has been suggested that N7 guanine alkylation may increase the acidity of N1 of guanine. Thus the decreased p*K*a of N1 atom would enable it to be changed to a hydrogen bond acceptor that weakens the base pair with dCTP*. In this case, the zwitterionic form of N7meG would prefer wobbled base pair with incoming dCTP* through its Watson-Crick face. However, we did not observe this for N7meG in the previous study; nor did we observe such wobble base pairing in N7bnG case. In the N7 methyl adduction case, methyl group is relatively small and thus with limited ability to pull away and stabilize the lone pair electron from N7 atom. Compared to N7 methylation, benzylation introduces a benzene ring with a methylene carbon, enhances the ability to delocalize and pull away the lone pair electrons from N7 atom. However, even at the Mg^{2+} presence, where there lacks the important catalytic ion, incoming dCTP* is still forming the classic Watson-Crick base pair with N7bnG. The ionization effect of N1 deprotonation that is expected to be brought by N7 benzylation is not shown. And in the binary structure of N7bnG, where there is no incoming dCTP*, no water molecule or metal ions could be

observed. This suggests at least in the active site of polβ, N7 adduction probably would not introduce an ionizing effect so strong that would deprotonate N1 atom to make the N7 alkyl guanine base to be in a zwitterionic form. The positive charge brought by N7 adduction probably is consumed or delocalized largely around the 5 member ring of the guanine base, with little effect extending to the other part.

We have previously studied other guanine lesions as the templating base for dCTP* insertion. When there are lesions occurring on the Watson-Crick face, esp.: O6MeG[109], incoming dCTP* cannot form Watson-Crick base pair with the guanine. However, when the lesion takes place on the imidazole ring of guanine (e.g.: 8-Br-dG, or 8-oxo-dG), incoming dCTP* still forms the Watson-Crick base pair with guanine. These, combined with N7 alkylation studies suggest that templating G:C base pairing will generally take the traditional Watson-Crick pattern unless Watson-Crick face of templating guanine is affected.

Implication into the aromatic and flexible effects of guanine N7-substituents

Our recent study supports N7 methylguanine as a benign DNA lesion. This is partially due to its small size, as the similar case of N7 ethyl adduction, which also shows no cytotoxic or mutagenic effects as aforementioned. Unlike N7meG, N7-alkylguanine produced by bulky alkylating agents such as aflatoxins, pluramycins, altromycins, azinomycins, *S*-(2-haloethyl)glutathione, and leinamycin have been shown to be genotoxic[50]. These group are all aromatic bulky adductions and some even with flexibility. However due to site specific insertion difficulty, the mutagenic mechanisms of these bulky adducts are largely unknown. N7 Bn adduct provides an excellent study model for these cases in that it is both aromatically bulky and flexible.

As can be seen from the figure below, unlike methyl group, the Bn group on N7 guanine changes its position to have different hydrophobic interactions when facing different insertion conditions. And even when the same dCTP* is inserted, Bn group changes the position in the presence of different metal ions. This is not observed either in N7meG. In N7meG, at Mg^{2+} presence, polβ takes on an ideal closed conformation already. There is no interaction between the methyl group to any other residues in polβ. Even for the N7meG:dTTP* mismatches using both Mg^{2+} and Mn^{2+}, there is essentially no structural difference of N7meG or the active site.

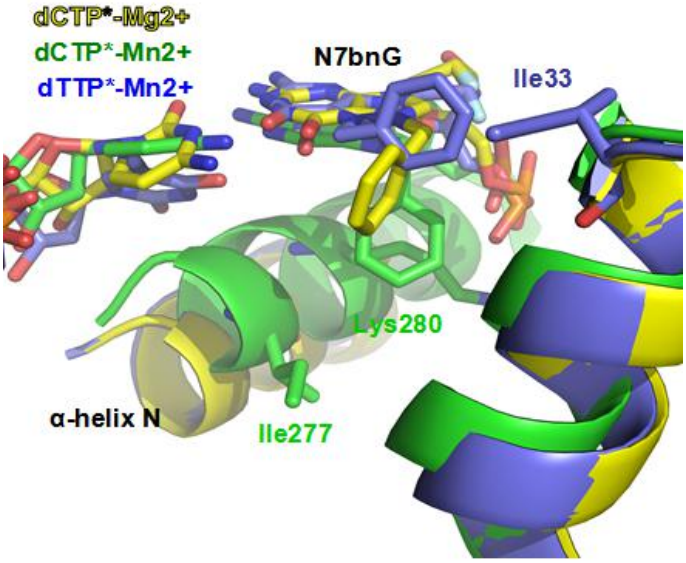


Figure 3.6

N7bnG alignments in yellow: -dCTP*-Mg^{2+}, green:-dCTP*-Mn^{2+}, blue: -dTTP*-Mn^{2+}

The flexibility of Bn group provides hydrophobic interaction with different residues of polβ. Thus, compared to N7meG, N7 benzyl adduct shows a totally different insertion pattern both structurally and kinetically. Aromaticity influence of N7 adduct has been shown by N7-aflatoxin B1-dG adduct in Y-family DNA polymerase Dpo4[50]. However, aflatoxin is less flexible than Bn and stacks only between the bases without any contacts to the polymerase. Herein, this N7bnG in polβ is the first example of flexibility

and aromaticity of N7 adduct mediating different active site conformation changes along with correspondent kinetics difference. It provides us with the new recognition that N7 bulky adducts may involve interactions with specific polymerase residues that may generate different results which needs a case by case analysis.

Metal dependent conformational activation of DNA polymerase β

As can be seen from the above, substituting Mn^{2+} for Mg^{2+} in the N7bnG:dCTP* complex increases the dCTP insertion efficiency, and triggers an open to a closed conformational reorganization. By contrast, N7meG:dCTP*-Mg^{2+} complex adopts the closed conformation. Metal dependent conformational activation of DNA polymerase has been rarely reported. Our previous studies of polβ in complex with mismatched O6MeG:dTTP[109] and dG:dTTP[80] complexes also showed metal-dependent conformational activation of the enzyme, but such activation has not been observed with correct G:T base pair.

This Bn group mediated open to closed conformation transition is not seen either in the N7meG:dCTP*-Mg^{2+} structure. The reason might be that the methyl group is smaller in size so its presence does not interrupt minor groove recognition or much restricted closed conformation of the active site. Plus it does not possess any similar aromaticity or flexibility as the Bn group to interact with surrounding polymerase environment. So in the N7meG: dCTP*-Mg^{2+} structure, α-helix N of polβ takes closed conformation. However, in the Mg^{2+} case of Bn adduct when dCTP* is the incoming nucleotide, the bulky presence of Bn group distorts the DNA in the active site and make it shift about 2 Å on average compared to the N7meG: dCTP*-Mg^{2+} structure. This entirely abrogates the previous closed conformation observed in N7meG. Although dCTP* still can base pair with N7bnG (Mg^{2+}) since Watson-Crick face is not interfered,

the insertion efficiency drops 10 fold compared to N7meG. N7bnG salvages this at the presence of Mn^{2+}, which, unlike the Mg^{2+} case where there is no catalytic ion, catalytic Mn^{2+} ion transforms the Asp256 from catalytically incompetent form into an catalytically competent form, as shown above. This corrects all the coordinations of the nucleotidyl transfer reactions site which shifts the upstream primer terminus and incoming dCTP* into the right positions with Tyr271 now stabilizing the primer terminus. Together with the Bn group rotation that forms the hydrophobic interaction with α-helix N. The Mg^{2+} to Mn^{2+} replacement activates the active site of polβ holding the dCTP:N7bnG base pair from open to closed conformation, which also brings a 10-fold increase in the insertion efficiency.

Effects of bulky N7-alkyl-dG adduct on DNA replication

N7bnG study represents the first human DNA polymerase TLS study of N7 alkyl adduction that is relatively bulky and flexible. Our published N7meG study indicated that small N7meG is not mutagenic. However, N7bnG study indicates that some bulky N7 adducts significantly delay the correct insertion of dCTP under Mg^{2+} condition. There are some recent studies of TLS using bulky N7 adduct. However, they are not even similar to this study of N7bnG. In the study of Dpo4 with N7-aflatoxin B1-dG adduct[50], a Y family polymerase, bulky aflatoxin adduct is not flexible but rather fixed and stacked between the bases. Similar principle may be the case for the bulky acridine N7 dG adduct induced G to A transition. Bn group is quite flexible in different insertion conditions and is not stacking between the DNA bases. This provides a new type of influence on DNA replication brought by N7-adduction. So, it can be justifiably expected that with numerous other N7 aryl or arylalkyl adducts which possess a variety of aromatic bulkiness, flexibility and maybe other functions, minor groove recognition residues could

develop new binding patterns that does not necessarily discriminate the mismatch. This study also suggests that, with more mismatch accommodating polymerases such as the A-family polymerase *Bacillus stearothermophilus* DNA polymerase I large fragment[47], the B-family RB69[48] and the Y-family DNA polymerase pol η [49], there could be a variety of protein-DNA interactions which features the flexibility, aromaticity, and bulkiness of the various N7 substituent groups.

3.4 SUMMARY OF CHAPTER III

From the structures and kinetic results shown above concerning the N7bnG in the polβ, it can be seen that the insertion across the DNA lesions, especially the N7 site derivatives, needs case by case analysis. From comparison to the N7meG, the bulkiness of the N7 substitution would likely to decrease the insertion efficiency. This probably relates to the fact that polβ active site has more spacial and recognition restraints. So that the bulky adduct, even at N7 position, where it cannot interfere much of the Watson-Crick base pair, still can affect the surrounding minor recognition residues from the α-helix N. It is possible that in a more open and less restricted active site, such as that of pol η , the influence brought by the N7 adduct bulkiness would be less.

The rotation of the benzyl group makes it possible for hydrophobic contacts with different surrounding areas from the polymerase. However, it is hard to conclude that whether these are the causes or the results of the increased stability. A possible way to further explore is to mutate the contacting residues from the polymerase. From a general view, this kind of hydrophobic interaction can also be eliminated if another shape of aromaticity instead of benzyl group is adducted on the N7 position. All these supports the idea that the analysis for N7 adducts polymerization might need case by case analysis.

Chapter IV: Structural and kinetic study of mutagenicity concerning 8-chloroguanine lesion in human DNA polymerase β model

4.1 BACKGROUND INTRODUCTION

One of the important types of DNA damage besides alkylation is halogenation. The aromatic base ring of DNA made of conjugated nitrogen and carbon is a good substrate for this reaction. Hot spots include carbon-8 position of purines and carbon-5 position of pyrimidines.

Halogen and their related compounds widely exist in the daily life. Metal halides, hydrogen halides, interhalogen, organic halogen and poly halogen compounds find intensive applications in almost every aspect of modern society: Chlorine and bromine are used as disinfectants for drinking water, swimming pool water, dirty wounds, dirty dishes, and other surfaces that require thorough cleaning. Sodium hypochlorite, the product from chlorine, is the well known active ingredient of most bleaches. Ethylene halides has been used as gasoline additives when the anti-lead measure scavenge the lead by forming the volatile lead halide[110]. Another well known application of halogen related industry is plastic manufacturing. Vinyl halides can be used in the production of polyethylene, polyvinylchloride or polypropylene. And specific highly halogenated molecules can also be added to participate in the polymerization process[111]. In drug discovery, the incorporation of halogen atoms into a lead candidate results in analogs that are usually more lipophilic and less water-soluble. Therefore, halogen atoms are used to improve penetration through lipid membranes and tissues. So it is suggested that there is a tendency for some halogenated drugs to accumulate in human body [112].

Chlorine is one important element of halogen. The simple substance form is not only widely used as the active ingredient for water and swimming pool disinfection, but is also used in million tons annually in bleaching industry, paper industry (Besides

bleaching, chlorine is used to break down the lignan that holds the wood fibers together). All of these creates amazingly significant amount of potential chlorine pollution [113].

The simple substance form of chlorine can react with H_2O to generate hypochlorous acid:

$$Cl_2 + H_2O \rightleftharpoons HCl + HClO \quad (1)$$

Hypochlorous acid is mainly used as the effective ingredient of disinfectant. For example: *Escherichia coli* exposed to this acid will lose viability in miliseconds due to inactivation of many vital systems [114]. LD50 of hypochlorous acid is 0.0104–0.156 ppm and 2.6 ppm can cause 100% growth inhibition in several minutes [115]. The mode of such disinfection has been suggested of several possibilities: i) It has been proposed that HOCl blocks uptake of nutrients by inactivating transport proteins. Venkobachar et al.[116] found that succinic dehydrogenase was inhibited *in vitro* by HOCl. Albrich et al.[117]'s study found that HOCl would destroy cytochromes and also iron-sulfur clusters. ii) Barrette et al.[118] found that HOCl exposed cells have lost the ability to regulate their adenylate level. Later study showed that ATP synthetase is involved in loss of ability to regenerate ATP[119]. iii) HOCl has been suggested to make some post-translational modifications of protein. Not only the oxidation of cysteine and methionine is observed, but also through the damage of the Hsp33, a chaperone known to be activated by oxidative heat stress, HOCl could also easily induce protein aggregation[120].

One important aspect of HOCl disinfection mechanism is its inhibition effect on DNA replication. It has been shown that DNA synthesis could be precipitously decreased or even prominently inhibited[121]. And the reason has been suggested to be associated with the damage of crucial membrane proteins involved in DNA replication machinery [122].

As can be seen from the aforementioned, HOCl is frequently generated in our daily life by wide production and use of chlorine plus industrial or household application of its disinfection effect. All these serves as the exogenous source of HOCl.

Besides the disinfection effect, recent studies have shown that exogenous HOCl can also lyse the cells and damage the DNA inside: Michael JD. et al.[123] has found that by forming cell derived radicals, HOCl would lyse the monocytes and macrophages instantly. Various research has shown that HOCl generated by the chlorination in the swimming pool or drinking water could cause both genotoxic and carcinogenic effects[124][125][126]. Additionally, hypochlorous acids and ionized salt form have been confirmed of damaging DNA nucleobases directly *in vitro*[127][128]. Summarily, all these studies indicate that great deal of exogenous HOCl pollution can possibly cause some carcinogenic effect on our bodies.

Besides the exogenous source of reactive halogen species HOCl, there is also an endogenous source of HOCl generation. In the electron transport chain, electrons are passed through a series of proteins via oxidation-reduction reactions, with each acceptor along the chain possessing a greater reduction potential than the previous. The final destination for this electron trip on this chain is an oxygen molecule. Under normal conditions, the oxygen is reduced to produce water; however, there is still about 0.1–2% of electrons leaked through the chain, oxygen is then incompletely reduced to give the superoxide radical ($\cdot O_2^-$). These incomplete reductions occur mostly in Complex I and Complex III[129]. One important enzyme is activated which is called SOD (SuperOxide Dismutase). This enzyme can catalyze the dismutation of superoxide, as shown below (M represents metal element such as Cu, Mn, Fe and Ni):

$$M^{(n+1)+} - SOD + O_2^- \rightarrow M^{n+} - SOD + O_2 \quad (2)$$

$$M^{n+} - SOD + O_2^- + 2H^+ \rightarrow M^{(n+1)+} - SOD + H_2O_2 \quad (3)$$

This dismutation of superoxide is considered as the detoxification process, which is one of the defense mechanism the cell uses to counter the imbalance of the 'oxidative stress' and ameliorate the harmful effects of these ROS (Reactive Oxygen Species). In the peroxisomes of eukaryotic cells, the enzyme catalase converts H_2O_2 to water and oxygen, and thus completes the detoxification initiated by SOD.

Activation of various leukocytes *in vivo* also results in the generation of H_2O_2. In tissues with inflammation or infection, various cells in the host immune system are activated, including nitric-oxide synthase and NADPH (Nicotinamide Adenine Dinucleotide Phosphate) oxidase. NADPH oxidase is a multicomponent membrane bound complex, and is needed for the bactericidal action of phagocytes. This enzyme is the most efficient enzyme to produce significant amount of H_2O_2. When at rest, heterodimer of two membrane embedded polypeptides (p22-phox and gp91-phox), which also contains two heme groups, transfer electrons from cytosolic NADPH across the membrane to one molecular of oxygen. Upon phagocyte activation, polypeptides (p47-phox, p67-phox and p40phox) translocate to the inner side of the plasma membrane to form a fully oxidase active enzyme complex. Then the charge compensation begins when gp91-phox polypeptide serves as an H^+ channel[130].

Upon further activation, when defense activity is severely needed, which is often the case when there is large or chronic inflammation situations, the respiratory burst further enhances the toxicity of hydrogen peroxide, converting it into hypochlorous acid. This is made possible by the enzyme called myeloperoxidase (MPO).

$$H_2O_2 + Cl^- \xrightarrow{MPO} HOCl + OH^- \qquad (4)$$

MPO protein is a 150 kDa dimer peroxidase consisting of two light chains and two variable-weight glycosylated heavy chains bound to a prosthetic heme group. It can

form the pentagonal pyramid conformation and has a calcium binding site which is indispensable for maintaining the active site structure.

The peroxidation of halide ions is a general two-step reaction starts with 2 electron oxidation of enzyme by the starting ROS H_2O_2 to form Compound I. This Compound I contains a porphyrin radical with an oxygen bound ferryl iron. It then undergoes the two-electron oxidation of substrate halides ions to form the related hypo-form products[140].

In addition to its main function of producing the hypohalogen acids or salts, MPO can also use its Compound I from the first step generated by H_2O_2 to oxidize a number of other substrates by a single-electron oxidation mechanism. This includes a variety of aromatic alcohols and amines, in which, tyrosine oxidized into tyrosyl radical is an example[141].

The peroxidation of chloride ions carried out by MPO has been testified as a pH tolerating reaction, though drastic change of H^+ would affect its kinetics. In fact, the concentration of starting materials H_2O_2 and most importantly Cl^- would all interfere with the rate of the peroxidation.

Other halogen ions such as Br^- is also the substrate of MPO for making HOBr acid or OBr^-. However, physiological concentration of bromide ion is more than 1000 times lower than that of the chloride ion[142].

Another enzyme capable of peroxidation of halide ions is called eosinophil peroxidase (EPO). It uses this as the similar defense mechanism also well known in its antimicrobial activity. However, many studies have shown that EPO is less active than MPO with chloride as the substrate. It prefers the bromide ion over the chloride ion, particularly under some circumstances when chloride ion concentration is low (as in some secreted fluids where chloride ion is trapped and retained)[143].

As aforementioned, bromide ion concentration is at least 1000 times lower than the chloride ion. The physiological concentrations of other substrates for MPO or EPO are also considerably low: compared to 0.1 M chloride, there is only 0.02-0.1 mM bromide, 0.1-0.6 M iodide, and 0.02-0.12 mM thiocyanate[145]. So it is relatively reasonable to state that HOCl is the major product from MPO or EPO *in vivo*.

One important discovery made for HOCl or hypo-acid form of other halide ion is that they can easily perform transhalogenation reactions using nucleobases at the physiological level[145]. This suggests the bactericidal effect of the MPO or EPO system as in the host defense system: reactive halogen species mutate DNA bases to kill the pathogens[146].

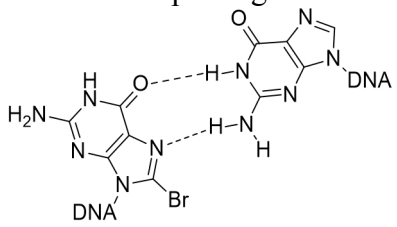
However, halogenation has been known with non-selectivity against DNA bases whether they are coming from the host or not. So this halogenation raises the question as whether HOCl can produce such DNA lesions inside our bodies and if so, whether such lesions are mutagenic or not.

It has long been known that inflammation and cancer has the relationship more than just association, especially the chronic and severe types of inflammation. An epidemiological study has given an astonishing percentage: at least 20% of all cancers are caused by chronic inflammation[147]. Additionally, many of the organ specific carcinogenesis have been shown to have a cause-and effect relationship attributed to local inflammation[148][149]. Some of such famous infection-cancer examples include viral hepatitis and liver cancer, *Helicobactor pylori* and gastric cancer and *Schisotosoma haematobium* and bladder cancer.

The underlying mechanism has been proposed that tissue cells would respond to the infection by releasing cytokines or other small molecules to activate the immune cells (mostly leukocytes). These defensive cells, such as macrophages, neutrophils and

lymphocytes are then recruited and infiltrated to the site of infection. If the infection persists to the state of chronic inflammation, the tissue cells would undergo cell cycle arrest and apoptosis. During this period, great amount of reactive oxygen, halogen and nitrogen species are piled up on scene in order to eliminate the infection[150]. And from the above mentioned, it is easy to see that HOCl is the major reactive halogen species in this chronic inflammation site. As aforementioned, HOCl possesses the ability to damage the DNA, forming 5-halo-dC and 8-halo-dG and -dA lesions[151]. Plus the fact that the similar reactive halogen species, HOBr, has been testified to be mutagenic in damaging DNA by forming 8BrG: G base pair [152]. So it is interesting to explore the structural and mutagenic potentials of DNA nucleobase damage caused by HOCl.

Guanine has been regarded as the most easily oxidized and damaged base in DNA (E^o 1.29 V vs. NHE) [153] and is thus the major target for HOCl either coming from outside or at the local chronic inflammation site. As said, our lab has previously showed the structural basis for 8BrG induced G to C transversion mutation: In the BrG: dGTP ternary structure, BrG adopts *syn* conformation and forms Hoogsteen base pairing with the incoming dGTP analog, as shown in the paradigm [152] below (Figure S1.1):



syn-8BrG:*anti*-dG

Figure 4.1

Chlorine has one less layer of electrons than bromine, however, atomic radius difference is not as big as that between chlorine and fluorine. Compared to 114 pm of bromine, chlorine has 99 pm, with less than 15% of difference. Because its bulkiness,

8BrG has the steric clash with the oxygen in its *anti* form. Via the rotation of the glycosidic linkage between the C1' and N9, 8BrG can acquire the *syn* conformation as shown above. Since the chlorine radius is smaller than bromine but to a small extent. So it is interesting to see whether 8ClG will have the similar case or not.

To explore the potential mutagenicity of 8ClG, human polymerase β is employed as the study model. As introduced previously, polβ is a DNA repair enzyme that fills short nucleotide gaps in DNA produced during base-excision DNA repair pathway. As aforementioned, it belongs to X family and lacks the intrinsic 3' to 5' exonuclease activity. The protein itself contains the N-terminal lyase domain (8 kDa) and C-terminal polymerase domain (31 kDa) which can be structurally divided into thumb, fingers and palm domains.

To evaluate the effect of 8ClG on the polymerase activity of polβ, kinetic parameters for the incorporation of dCTP, dGTP, dATP, dTTP opposite the chemically stable 8ClG by polβ are determined. Besides the binary 8ClG structure, four ternary structures are solved in which templating 8ClG exists with each of the four incoming nonhydrolyzable deoxy triphosphate nucleotide analogs. Herein, we report the first kinetic study that shows the effect of incoming dNTPs across the templating 8ClG in polymerase β model. We also present the first crystal structural study in polβ inserting incoming dNTPs opposite the templating 8ClG, which reveals for the first time the base-pairing nature of 8ClG facing different dNTPs in the confines of a DNA polymerase.

4.2 EXPERIMENTAL PROCEDURES

Synthesis of 8ClG phosphoramidite

The synthesis of 8ClG phosphoramidite follows the published scheme[154]. Briefly, as shown below, after the dehydration of deoxyguanosine monohydrate by pyridine (Py.) coevaporation, it is protected by isobutyryl group. The protected product is then chlorinated at the C8 position by *N*-Chlorosuccinimide (NCS) in tetrahydrofuran (THF). After deprotection by sodium methoxide (NaOMe) in methanol (MeOH), the compound is dimethoxytritylated and phosphitylated as all other amidites to generate 8ClG phosphoramidite.

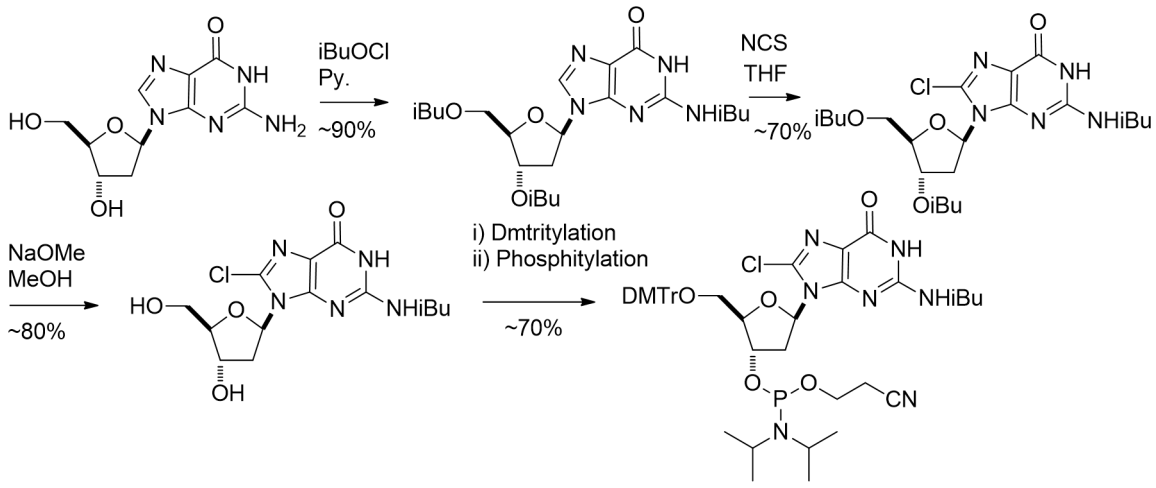


Figure 4.2 Synthesis of 8ClG amidite

DNA Sequences

Oligonucleotides were purchased from Integrated DNA Technologies (IDT) or synthesized from the above amidites by Midland Certified Reagent Co. (Midland, TX). They were purified by the manufacturer, and their sequences were confirmed by MALDI-TOF mass spectrometry. DNA substrates used for crystallographic studies consisted of a 16-mer template, a complementary 10-mer primer, and 5-mer downstream

oligonucleotides. The template DNA sequence used for crystallization was 5'-CCGAC[8ClG]TCGCATCAGC-3'. The upstream primer sequence was 5'-GCTGATGCGA-3'. The downstream oligonucleotide sequence was 5'-pGTCGG-3', and the 5' terminus was phosphorylated. The DNA sequence almost identical to a published ternary complex structure (PDB ID 1BPY) was used to minimize sequence-dependent structural differences. The oligonucleotides were mixed and annealed to give a 1 mM mixture of gapped DNA as decribed [14].

Steady State Kinetics

Steady-state kinetic parameters for dNTP (dCTP, dGTP, dATP, dTTP) insertion opposite templating 8ClG by polβ were determined using the same procedure as described previously[14].

Protein Expression and Purification

Polβ was expressed and purified from *Escherichia coli* with minor modifications of the method described previously[14].

Protein-DNA Co-crystallization

The binary polβ complex containing templating 8ClG in a single-nucleotide gapped DNA was prepared under conditions similar to those described previously[152]. Polβ was complexed with a single-nucleotide gapped DNA containing a 16-mer template (5'-CCGAC[8ClG]GCGCATCAGC-3'), a complementary 10-mer primer (5'-GCTGATGCGC-3'), and a 5-mer downstream oligonucleotide (5'-pGTCGG-3'). The resulting polβ-DNA complex was used to obtain binary and ternary complex crystals in the absence or presence of an incoming nucleotide, respectively. The ternary polβ-DNA

complex co-crystals with nonhydrolyzable dNTP analogs paired with templating 8ClG in a single-nucleotide gap at the active site were grown in a solution containing 50 mM imidazole, pH 7.5, 14 – 23% PEG3400, and 350 mM sodium acetate as described previously[152]. Crystals were cryo-protected in mother liquor supplemented with 12% ethylene glycol and were flash-frozen in liquid nitrogen.

Data Collection and Refinement

Diffraction data were collected at 100 K using either a Rigaku MicroMax-007 HF microfocus x-ray generator with R-Axis IV++ imaging plate area detector or the beamline 5.0.3 Advanced Light Source at Berkeley Center for Structural Biology. All diffraction data were processed using HKL 2000[155]. The structures of the binary polβ complex with templating 8ClG in a single-nucleotide gapped DNA and the ternary complex of polβ with templating 8ClG paired with dNTP analog were solved by molecular replacement with polβ with a single-nucleotide gapped DNA (PDB code 1BPX) as the search model. The model was built using COOT and refined using PHENIX software [156][157]. MolProbity was used to make Ramachandran plots[158].

Table 4.1 Data collection and refinement statistics.

PDB code	**8ClG - gapped binary**	**8ClG• C- Mg^{2+} ternary**	**8ClG• G- Mg^{2+} ternary**	**8ClG •C- Mn^{2+} ternary**	**8ClG• G- Mn^{2+} ternary**	**8ClG• A- Mn^{2+} ternary**	**8ClG• T- Mg^{2+} ternary**
Space Group	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Cell constants							
a (Å)	54.772	50.829	54.985	50.77	54.713	54.786	54.73
b	79.668	80.243	79.622	80.09	79.620	78.993	78.93
c	54.967	55.836	55.179	55.62	55.262	54.771	54.62
α (°)	90.00	90.00	90.00	90.00	90.00	90.00	90.00
β	105.60	107.80	107.88	107.0	108.78	105.91	105.9
γ	90.00	90.00	90.00	90.00	90.00	90.00	90.00
Resolution (Å)[a]	20-2.18 (2.22-2.18)	20-1.93 (1.96-1.93)	20-2.33 (2.37-2.33)	20-2.04 (2.08-2.04)	20-2.17 (2.21-2.17)	20-2.05 (2.09-2.05)	20-2.09 (2.13-2.09)
R_{merge} [b] (%)	0.090 (0.548)	0.051 (0.247)	0.096 (0.488)	0.076 (0.43)	0.071 (0.316)	0.070 (0.252)	0.089 (0.399)
<I/σ>	20.8 (1.86)	31.2 (4.31)	20.1 (2.36)	21.5 (2.50)	21.4 (3.29)	25.5 (3.82)	21.5 (2.86)
Completeness (%)	100 (99.9)	100 (100)	100 (100)	99.8 (98.0)	98.8 (93.5)	99.9 (97.9)	99.2 (97.5)
Redundancy	5.4 (5.1)	4.7 (4.2)	5.6 (5.4)	5.4 (4.4)	5.6 (5.5)	4.6 (3.8)	4.9 (4.7)
Refinement							
R_{work} [c]/R_{free} [d] (%)	21.8/26.8	19.1/23.9	19.7/24.3	19.4/24.1	20.9/25.9	19.7/25.4	21.2/25.8
Unique reflections	23552	32119	19481	27033	23751	28154	26349
Mean B factor ($Å^2$)							
Protein	36.3	23.3	41.2	28.0	30.7	28.6	32.2
Ligand	34.5	27.7	37.6	31.1	34.0	33.8	30.9
Solvent	33.3	32.6	38.0	30.8	30.6	27.4	31.2
Ramachandran plot							
Most favored (%)	96.3	98.5	97.2	97.5	97.5	96.0	95.0
Additional allowed (%)	3.4	1.5	2.8	2.5	2.5	4.0	4.7
RMSD							
Bond lengths (Å)	0.004	0.005	0.006	0.013	0.005	0.005	0.004
Bond angles (°)	0.886	1.183	1.291	1.682	1.216	1.444	1.288

Table 4.0 Data collection and refinement statistics.

[a] Values in parentheses are for the highest resolution shell.

[b] R_{merge} = Σ|I-<I>|/ ΣI where I is the integrated intensity of a given reflection.

[c] R_{work} = Σ|F(obs)-F(calc)|/ΣF(obs).

[d] R_{free} = Σ|F(obs)-F(calc)|/ΣF(obs), calculated using 5% of the data.

4.3 EXPERIMENTAL RESULTS AND DISCUSSIONS

Kinetics results

Table 4.2. Kinetic parameters of dNTPs insertion opposite the 8ClG lesion

Template:dNTP	pH	K_m (μM)	k_{cat} ($10^{-3}s^{-1}$)	k_{cat}/K_m ($10^{-3}s^{-1}\mu M^{-1}$)	f [a]
dG:dCTP[b]	7.5	269.55 ± 18.87	7.56 ± 0.38	0.03	
dG:dCTP[c]	7.5	0.59 ± 0.03	20.38 ± 0.50	34.54	1
8ClG:dCTP(Mg^{2+})[c]	7.5	6.52 ± 0.81	3.59 ± 0.23	0.55	1.6×10^{-2}
8ClG:dCTP(Mn^{2+})[c]	7.5	1.31 ± 0.22	16.52 ± 1.01	12.61	0.37
8ClG:dGTP(Mg^{2+})[c]	7.5	22.96 ± 0.23	0.87 ± 0.05	0.038	1.1×10^{-3}
8ClG:dGTP(Mn^{2+})[c]	7.5	3.92 ± 0.09	4.52 ± 0.34	1.15	3.3×10^{-2}
8ClG:dATP(Mn^{2+})[c]	7.5	60.70 ± 3.8	0.96 ± 0.32	0.016	4.5×10^{-4}
8ClG:dTTP(Mg^{2+})[c]	7.5	90.23 ± 7.2	0.29 ± 0.07	0.0032	9.3×10^{-5}

[a] Relative efficiency: $(k_{cat}/K_m)_{[dNTP:8ClG]}/(k_{cat}/K_m)_{[dCTP:dG]}$.

[b] A recessed DNA (i.e. DNA without the downstream primer) was used as a substrate.

[c] Single-nucleotide gapped DNA was used as a substrate.

Structure of gapped binary complex of polβ with templating 8ClG

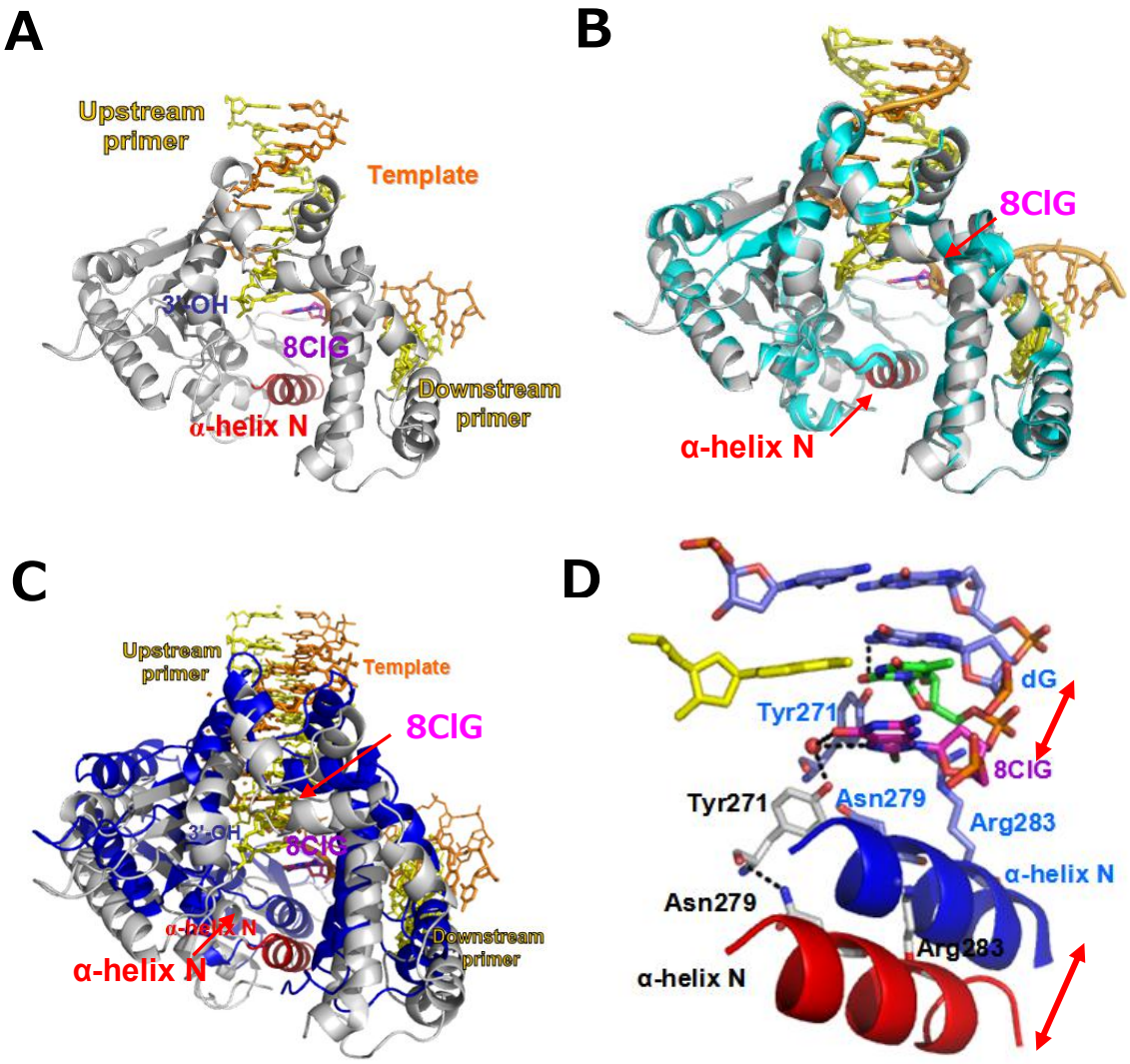


Figure 4.3 Polβ-8ClG gapped binary structure. (A) Overall view of polβ-8ClG gapped binary structure. (B) Alignment of polβ-8ClG binary structure with published polβ-8BrG binary structure (PDB ID: 4M2Y, r.m.s.d. = 0.741 Å). Note that the α-helix N positions are essentially the same. (C) Alignment of polβ-8ClG binary structure with published polβ normal dG binary structure (PDB ID: 1BPX, r.m.s.d. = 0.794 Å). Note the α-helix N position difference. (D) Active site comparison of those from (C).

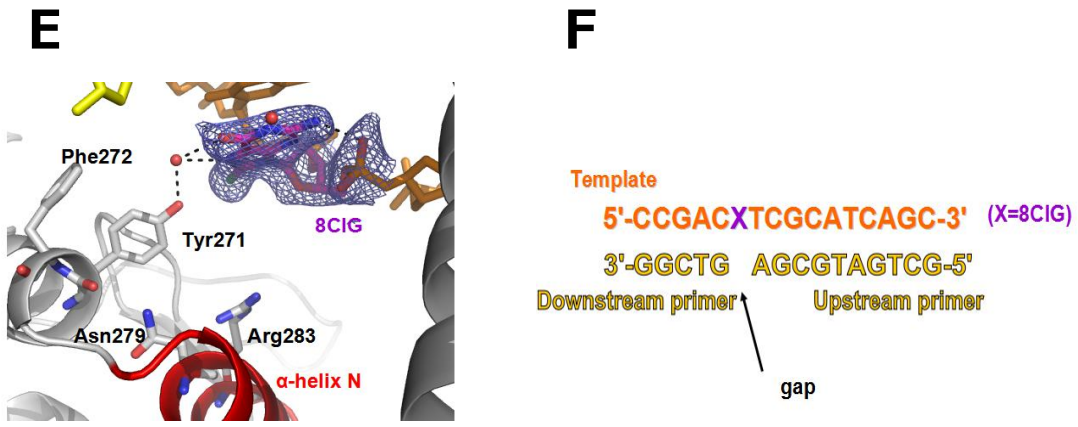


Figure 4.3 Polβ-8ClG gapped binary structure. (E) Electron density of 8ClG at 1 σ contour. Note the water mediated H bond network and itself stabilization from the backbone. 8ClG is in syn conformation. (F) DNA sequence used for crystallization of the gapped polβ complex.

The X-ray structure of a single-nucleotide gapped binary complex of polβ bound to 8ClG-containing DNA. The structure of the gapped binary complex, refined to 2.2 Å resolution, is very similar to that of a published gapped binary structure of 8BrG (PDB ID: 4M2Y, r.m.s.d. = 0.741 Å) (Fig. 4.2A and 4.2B) [152]. Polymerase is in an open conformation since the minor groove recognition residues Asn279 and Arg283 on the α-helix N, are being located ~10 Å from templating 8ClG.

However, when compared to the published open conformation of binary structure of normal dG (PDB ID: 1BPX), although the overall has not been changed much (r.m.s.d. = 0.794 Å), the active site overall has moved together about 7Å. And the minor groove recognition residues Asn279 and Arg283 are in different conformation from α-helix N (Figure 4.1C and 4.1D). This difference is triggered by the 8ClG in the templating strand. Unlike G or 8oxoG, unpaired 8ClG at templating position preferentially adopts *syn* conformation with clear electron density (Figure 4.1F). This completely changes the

conformation of the templating strand. By alignment, it moves away from the dG position as in binary structure (PDB ID: 1BPX) about 6 Å. Similar to 8BrG binary structure, 8ClG is also stabilized in this binary structure by the water mediated H bonds with O6 and N7 from its Hoogsteen edge and the -OH from the side chain of Tyr271 (Figure 4.1E). 8ClG fixes itself also by the intra-molecular H bond between N2 and the 5'-phosphate oxygen (Figure 4.1E).

8oxoG binary structure is known to have the mixture conformations of both *syn* and *anti*[159]. When compared together with this 8ClG and previous dG, 8BrG binary structures, it becomes clear that the size of 8 position adducts affects the conformation of the base. Oxygen atom has the radius of 73 pm which is less than chlorine (99pm) and bromine (114pm). Thus it is possible that the smaller size of oxygen atom on the guanine base does not clash that much with oxygen atom of the sugar. And conformations can exist in both forms. While bigger substitution of chlorine or bromine, this clash is much more prominent as to an extent that *syn* conformer is preferred. Additionally, because of the *syn* conformation, the exocyclic N2 amine is positioned near the phosphate backbone (Figure 4.1E). This contributes to a H bond made between the 5'-phosphate oxygen and the donor amine, as can be seen in both 8ClG and 8BrG structures. While taking the *anti* conformation, as can be seen from the normal dG structure (Figure 1D, PDB ID: 1BPX), the N2 amine is far away from phosphate backbone but points toward the opposite direction, makes H bond instead to Tyr271 (Figure 4.1D). Needless to say, this intra-molecular H bond further stabilizes the *syn* conformation of the 8ClG(and 8BrG) as well as the distorted templating strand.

Ternary structure of polβ with templating 8ClG paired with an incoming dCTP Analog

As shown in the kinetics results, among all other dNTPs, for the incoming dCTP analog, Polβ with templating 8ClG shows efficient catalysis. To further elucidate this, pre-catalytic ternary structures of polβ in complex with an incoming nonhydrolyzable dCMPNPP (hereafter dCTP*) paired with templating 8ClG are obtained upon different metal ions.

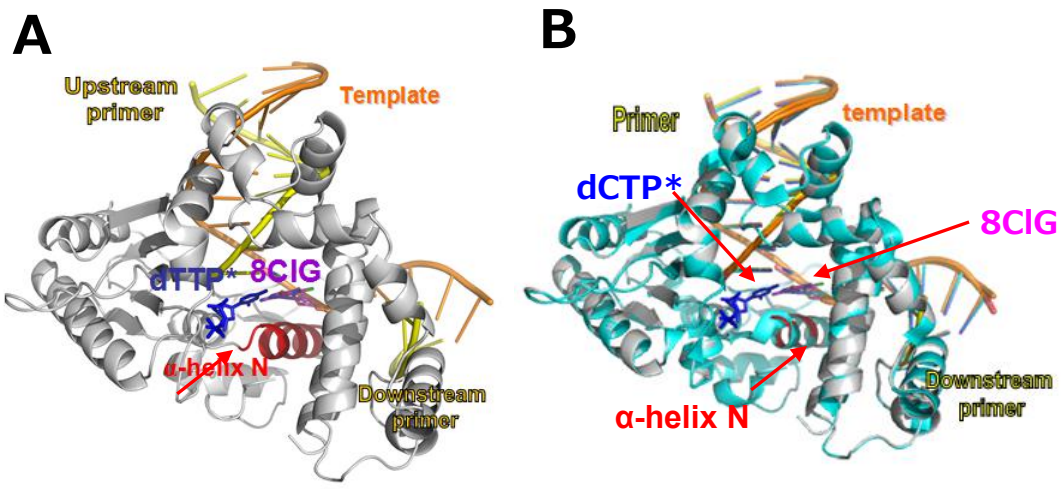


Figure 4.4 Polβ-8ClG:dCTP*-Mg^{2+} structure. (A) Overall view of polβ-8ClG:dCTP*-Mg^{2+} structure. 8ClG in purple. Incoming dCTP* in blue. (B) Alignment of polβ-8ClG:dCTP*-Mg^{2+} structure with published polβ-8BrG:dCTP*-Mg^{2+} structure (PDB ID: 4NLK, r.m.s.d.=0.215Å). Note that the α-helix N positions are essentially the same.

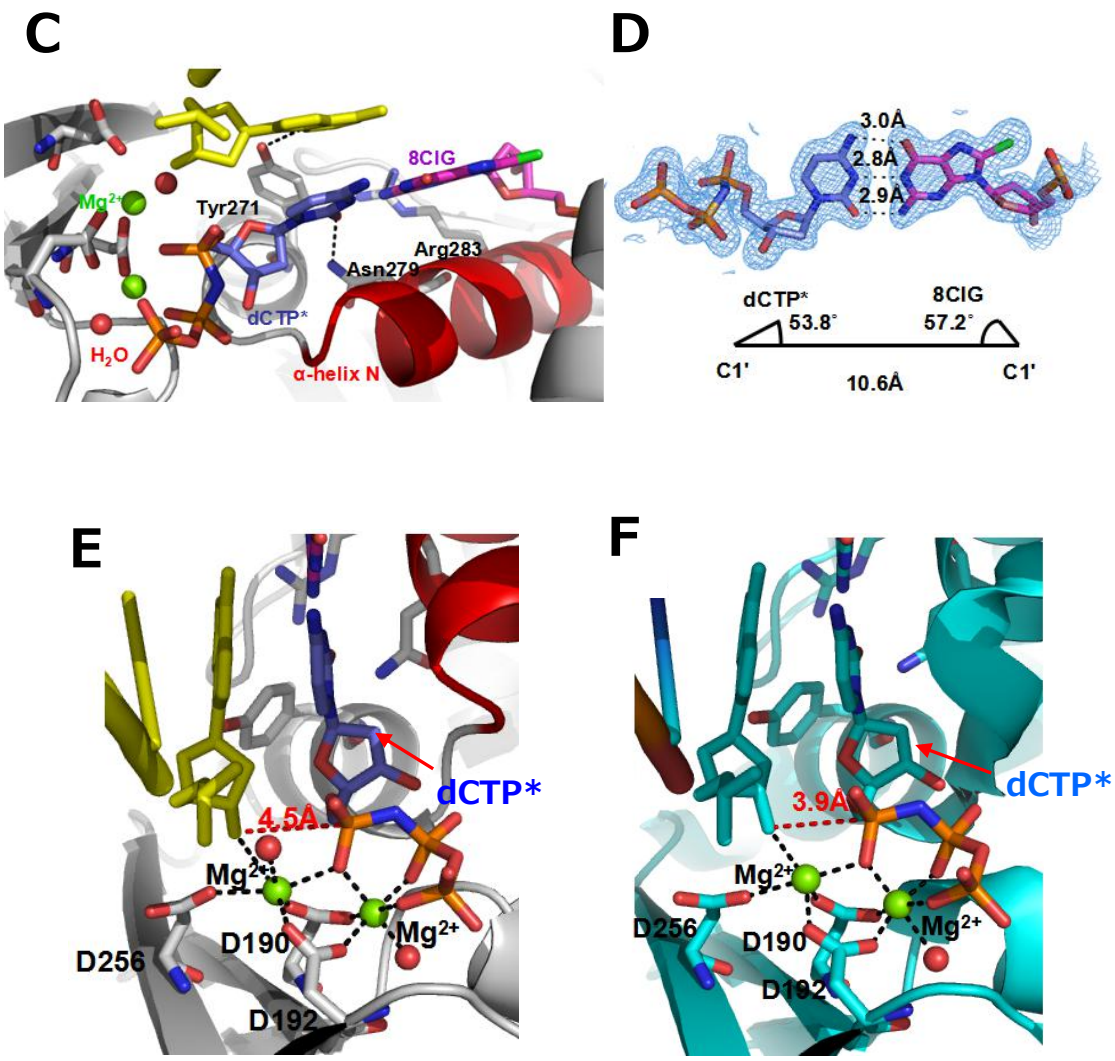


Figure 4.4 Polβ-8ClG:dCTP*-Mg^{2+} structure. (E) Coordination view of polβ-8ClG:dCTP*-Mg^{2+} structure. Nucleotidyl transfer reaction distance is shown in red. Water in red spheres. (F) Coordination view of polβ-8BrG:dCTP*-Mg^{2+} structure. Nucleotidyl transfer reaction distance is shown in red. Water in red spheres. Note the 3'-OH position and water coordination difference from (E).

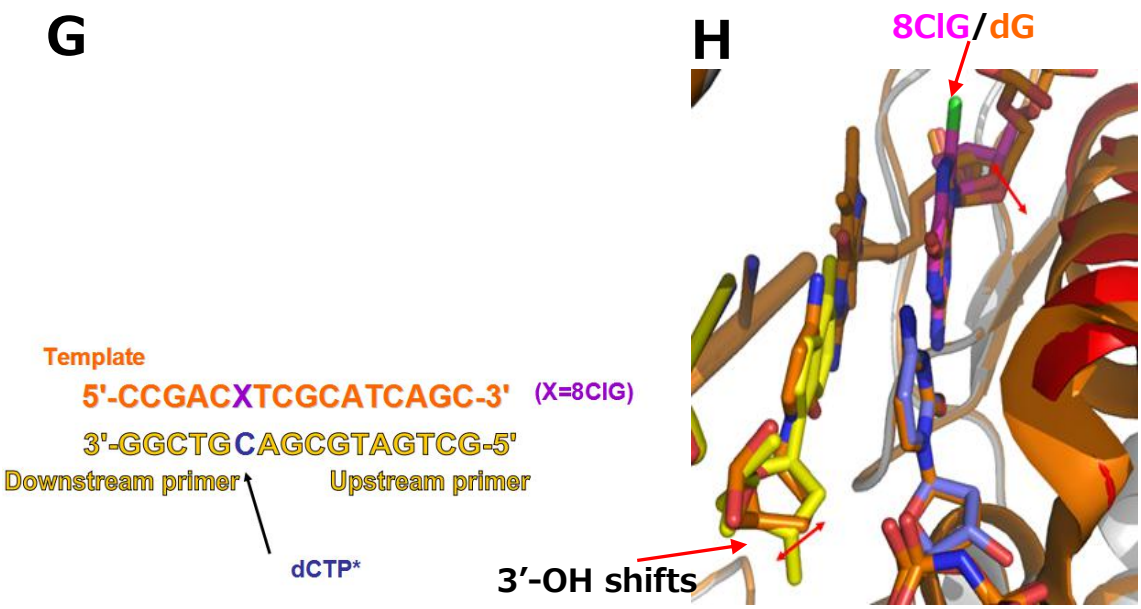


Figure 4.4 Polβ-8ClG:dCTP*-Mg^{2+} structure. (G) DNA sequence used for crystallization of the polβ-8ClG:dCTP*-Mg^{2+} structure. (H) 3'-OH and backbone shift shown by the red arrow heads of polβ-8ClG:dCTP*-Mg^{2+} structure (colors other than brown) from normal dG:dCTP structure (in brown, PDB ID: 2FMP).

The precatalytic 8ClG:dCTP* (Mg^{2+}) ternary structure is refined to 1.96Å. This ternary structure is essentially identical to the 8BrG:dCTP*(Mg^{2+}) structure (PDB ID: 4NLK, r.m.s.d.=0.215Å): i) α-helix N shifted from the open conformation of the gapped binary structure towards the active site sandwiching the incoming dCTP* to form base pair with templating 8ClG, taking a closed conformation (Figure 4.2C). ii) Because of the α-helix N movement, the minor groove recognition residues Asn279, Arg283, and Tyr271 successfully engage in H-bonding interactions with the incoming nucleotide, templating base, and primer terminus respectively (Figure 4.2C). iii) The incoming dCTP* is forming the classic Watson-Crick base pair with the 8ClG, while the 8ClG is taking the *anti* conformation. The λ angles for dCTP* and 8ClG are 53.8° are 57.2°,

respectively. And the distance between the two C1' atoms is 10.6Å (Figure 4.2D). iv) The two magnesium metals can be seen coordinated with proper distance to the surrounding phosphate groups from incoming dCTP*, 3 coordination residues: Asp 190, 192 and 256, two water molecules and also the 3'-OH of the primer terminus (in yellow, Figure 4.2E).

8ClG:dCTP*-Mg^{2+} ternary structure is very similar to that of the 8BrG ternary structure: both templating G takes the *anti* conformation instead of *syn* as in both binary structures. This suggests that the incoming dCTP*, by forming the correct Watson-Crick base pair with 8ClG, tends to choose the *anti* form of 8ClG, which is also seen in the 8oxoG cases, as mentioned above. The change of *syn* to *anti* conformation involves the break and formation of the H bonds of 8ClG. By comparing the binary gapped and the ternary dCTP* precatalytic structures, during the conversion from *syn* to *anti* conformation, 4 H bonds are broken with more than 6 H bonds formed. This does not include the coordination contacts made from the primer side by 3 Asp residues and metal ions. So base pairing with incoming dCTP* of *anti* 8ClG is thermodynamically more stable than its binary *syn* conformation, at least from the local perspective.

The catalytic magnesium ion in 8BrG structure does not acquire the perfect octahedral coordination geometry, however in the 8ClG:dCTP*-Mg^{2+} ternary structure the missing water molecule is clearly coordinating the catalytic metal ion. There is also the difference of nucleotidyl transfer reaction distance: in the 8BrG:dCTP* ternary structure (PDB ID: 4NLK), the 3′-OH of the primer–terminal nucleotide is in position for in line attack on the α-phosphate (Pα) of the incoming nucleotide. The angle of C3'-3'OH-Pα is 91.5°. And the distance between the 3′-OH and the α-phosphate is 3.94Å. While in the 8ClG:dCTP* ternary structure, the primer terminus 3′-OH is not optimally positioned for in-line nucleophilic attack on the Pα of the incoming nucleotide: the

distance between the 3'-OH and Pα is elongated to 4.54Å and the angle of C3'-3'OH-Pα decreases to 73.1°. Compared the two structures, 8ClG:dCTP* probably represents the earlier stage of nucleophilic attack, in which the incoming dCTP* has just been positioned rightly to be in the range by catalytic metal with water there to coordinate; whereas the 8BrG:dCTP* structure probably represents the stage after this positioning: 3'OH maybe already deprotonated to 3'O which is now positioned in line to attack the Pα. The water molecule missed here maybe the one coordinating the catalytic metal ion in the 8ClG:dCTP* structure, which may represent the base performing the deprotonation of 3'-OH.

Compared to the normal dG, insertion efficiency for the 8ClG is relatively low for incoming dCTP, as shown in the kinetic results. It is, however, comparable to the previously reported efficiency of templating 8BrG. This can be reasonably explained by the structures: Both 8ClG and 8BrG with dCTP* structures showed near ideal Watson-Crick base pairing, with minor groove DNA-protein interactions and coordinated catalytic and binding metal ions. However, the nucleotidyl transfer reaction distance of both exceeds the ideal distance of 3.4Å. For 8ClG:dCTP*, the 3'OH is not in line at all for nucleophilic attacking. For 8BrG:dCTP*, the 3'OH, is partially taking the in line attacking position. All these difference originates from the fact that 8 position of guanine base is chlorinated or brominated. As can be seen from the alignment of 8ClG:dCTP* with that of the normal dG structure[41] (Figure 4.2H), we can see that the introduction of 8-chlorination results in 2Å movement of the surrounding phosphate backbone. And the 3' primer terminus sugar has thus rotated about 37.1° (C2'(Cl structure)-C3'(Br structure)-C2'(Br structure)). So it is probably that this backbone movement induces difference of the whole DNA duplex during the dCTP* insertion, and thus lowers the catalytic efficiency than the templating 8BrG case.

Ternary structure of polβ with templating 8ClG paired with an incoming dGTP analog

Polβ-8ClG:dGTP*-Mg^{2+} ternary structure has been solved to the resolution of 2.37Å to elucidate the higher insertion efficiency of incoming dGTP towards templating 8ClG with human DNA polymerase β.

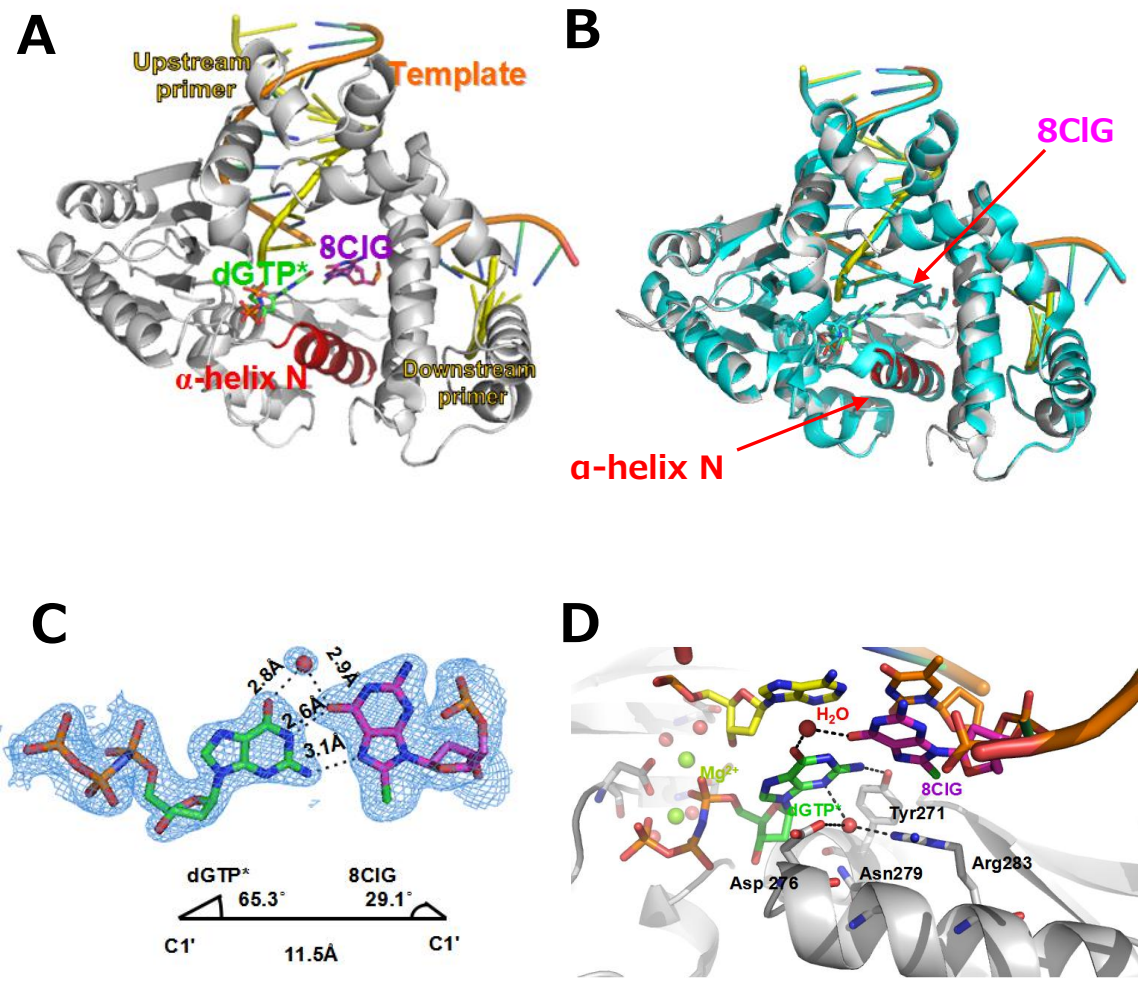


Figure 4.5 Polβ-8ClG:dGTP*-Mg^{2+} structure. (A) Overall view of polβ-8ClG:dGTP*-Mg^{2+} structure. 8ClG in purple. Incoming dGTP* in green. (B) Alignment of polβ-8ClG:dGTP*-Mg^{2+} structure with published polβ-8BrG:dGTP*-Mg^{2+} structure (in cyan, PDB ID: 4M47, r.m.s.d.= 0.223Å). Note that the α-helix N positions are essentially the same. (C) Electron density of 8ClG:dGTP* at 1 σ contour. (D) Active site view of polβ-8ClG:dGTP*-Mg^{2+} structure showing the minor groove recognition. Note that water is not coordinated to Asn279 but to Asp 276.

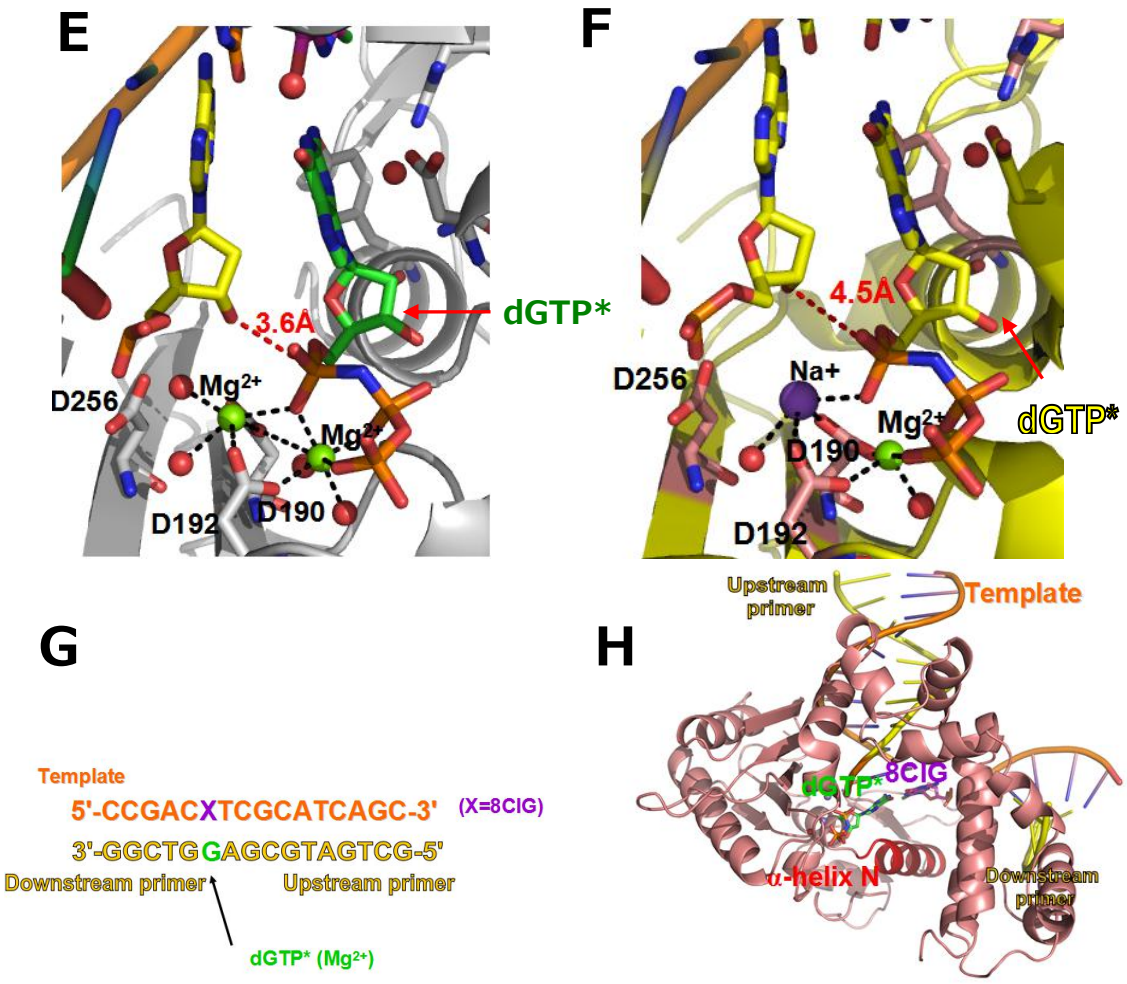


Figure 4.5 Polβ-8ClG:dGTP*-Mg^{2+} structure. (E) Coordination view of polβ-8ClG:dGTP*-Mg^{2+} structure. Nucleotidyl transfer reaction distance is shown in red. Water in red spheres. (F) Coordination view of polβ-8BrG:dGTP*-Mg^{2+} structure. Nucleotidyl transfer reaction distance is shown in red. Water in red spheres. Note the 3'-OH position difference and the replacement of catalytic Mg^{2+} with Na^{+}. (G) DNA sequence used for crystallization of the polβ-8ClG:dGTP*-Mg^{2+}/Mn^{2+} structure. (H) Overall view of polβ-8ClG:dGTP*-Mn^{2+} structure, 8ClG in purple, incoming dGTP* in green.

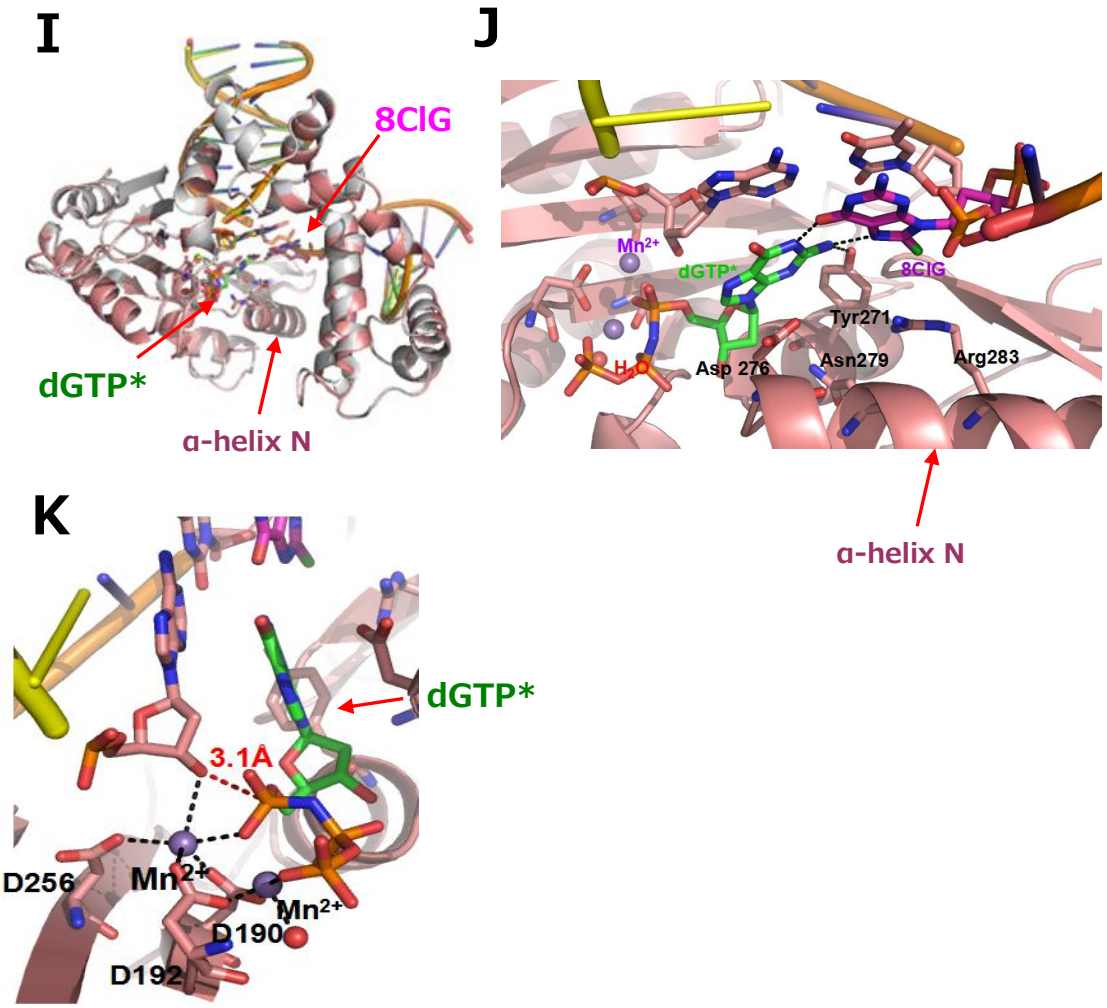


Figure 4.5 Polβ-8ClG:dGTP*-Mg^{2+} structure. (I) Alignment of polβ-8ClG:dGTP*-Mn^{2+} (all in brownish pink) with polβ-8ClG:dGTP*-Mg^{2+} (gray, with multicolors showing the active site). (J) Active site view of polβ-8ClG:dGTP*-Mn^{2+}, note there lack both the water coordinating the base pair and the minor groove recognition, as compared to (D). (K) Coordination of polβ-8ClG:dGTP*-Mn^{2+} structure active site. Note the catalytic Mn^{2+} is making coordination to 3'-OH and Asp256 side chain. Thus the nucleotidyl transfer reaction distance is shortened, as shown in red.

The overall structure of 8ClG:dGTP*-Mg^{2+} is very similar to that of the 8BrG:dGTP* structure (Figure 3B, PDB ID: 4M47, r.m.s.d.= 0.223Å). The overall polβ takes the near close conformation, with α-helix N taking the intermediate conformation. The incoming dGTP* is base pair with the templing 8ClG. Similar to 8BrG: dGTP* base pair, the incoming dGTP* is taking the *anti* conformation to have the two H bonds formed from the Hoogsteen face to the *syn* conformer of the templating 8ClG, which is also stabilized similarly by a water molecule bridging the two via its H bonding towards the O6 carbonyl oxygens of both guanine bases (Figure 4.3C). Obviously the geometry of the *anti*-dGTP*: *syn*-8ClG is considerably different from that of a canonical C:G base pair, however it is similar to the previously 8BrG case: C1' to C1' distance is elongated to 11.5Å. And λ angles deviates to 65.3° and 29.1°.

As aforementioned, minor groove interactions from the α-helix N is crucial to stabilizing the primer terminus, the incoming dNTP and the templating base and relates tightly to the insertion efficiency. Here in the 8ClG:dGTP*-Mg^{2+} ternary structure, the Tyr271 is not stabilizing the primer terminus. Instead it makes contact by H bond to stabilize the incoming dGTP* previously seen in the 8BrG case. This incoming dGTP* is further stabilized by a water molecule which is also supported by H bond from Arg 283 of the α-helix N. However, different from the 8BrG structure this water molecule is not further supported by Asn 279 side chain, but from the Asp 276 side chain from the same α-helix N. By alignment, this minor difference is caused by the minor movement of α-helix N, which is shifted more towards the closed conformation in 8ClG than in the 8BrG case (Figure 4.3D). Also, the H bond seen in the 8BrG structure bridging between the 3'OH of the incoming dGTP* and the carbonyl oxygen of the Phe 272 backbone is not observed in the 8ClG structure. This is also caused by the minor movement of α-helix N and slight shift of the incoming dGTP* in the 8ClG structure.

The major difference between the 8ClG:dGTP*-Mg^{2+} ternary structure to that of the 8BrG lies in the metal ion coordinations. As can be seen from the comparison (Figure 4.3E and F), the binding magnesium ions are almost in the identical position both structures and are both well coordinated octahedrally with the surrounding Asp 190, Asp192, phosphate group of the incoming dGTP* and a water molecule. However, in the 8BrG structure, there is no magnesium ion at the supposed catalytic position. It is replaced by a sodium ion which has a rather bad coordination because of its natural valence. In the 8ClG structure, a clear magnesium metal ion is seen at the position. This magnesium catalytic ion is coordinated by the Asp 190, Asp 192, Pα of the incoming dGTP* and two water molecules. As shown by the graphs (Figure 4.3E), this magnesium ion is still not well coordinated, since it still lacks the H bonds to the 3'OH of the primer terminus and the Asp 256 side chain. However, because of its valence, it shifts the Asp 192 about 2Å to a better position, making the catalytic metal ion a better coordinated situation than that of the 8BrG.

An additional difference between the 8ClG:dGTP* (Mg^{2+}) structure and that of the 8BrG is the nucleotidyl transfer reaction distance. The distance between the 3'-OH and the Pα oxygen is shortened by almost 1Å in the 8ClG than that of the 8BrG (Figure 4.3E and F). Furthermore, the 3'-OH of the primer terminus in 8ClG is taking the in line nucleophilic attacking position while the 3'-OH of the 8BrG is taking a relatively detached position.

The structural difference between the 8ClG:dGTP* (Mg^{2+}) and that of the 8BrG indicates that dGTP insertion is more efficient in the 8ClG case: closer α-helix N, presence of catalytic metal ion with right valence, shortened nucleotidyl transfer reaction distance with in line position of 3'-OH of primer terminus. All these originates from the difference of the 8 position adduction radius. The change of 8Br to 8Cl causes the

surround phosphate backbone shifting about 0.5Å, which then generates a serial of conformation change involving the upstream primer terminus and the surrounding polβ residues.

In order to further elucidate the enhanced promutagenicity of 8ClG over 8BrG. The 8ClG:dGTP* structure is solved with Mn^{2+} metal ion with 2.21Å resolution (Figure 4.3H). As can be seen from the graph, the overall polβ-8ClG:dGTP*-Mn^{2+}structure is very similar to that of the Mg^{2+} bound one (Figure 4.3I, r.m.s.d.=0.256Å). The base pair pattern is still the *anti*-dGTP* with *syn*-8ClG via its Hoogsteen face of two H bonds. However, there is no observed water molecule bridging the carbonyl oxygens of the two. The α-helix N is taking the same intermediate position towards sandwiching the nascent base pair. Tyr271 is still making the H bond to stabilized the incoming dGTP*. But the there is no observed water molecule that could mediate the minor groove recognition from the α-helix N to further stabilized the incoming dGTP* (Figure 4.3J).

The major difference between the Mn^{2+} bound and the Mg^{2+} bound structures is observed of the metal ions coordination, especially the catalytic metal ion. As shown in the Figure 3K, in polβ-8ClG:dGTP*-Mn^{2+} structure, both the catalytic and binding metal ions have been successfully replace by Mn^{2+}. Where as the binding metal ion remains the same position and make coordination around, the catalytic Mn^{2+} ion is now drawing the coordination of Asp256 and the 3'-OH of the primer terminus, both of which are now changed into the active and rightful coordination position. Although this catalytic Mn^{2+} lacks the water molecules to form a perfect octahedral coordination, we can see that the 3'OH is taking the near perfect in line attacking position towards the Pα of the incoming dGTP*. And subsequently, the nucleotidyl transfer reaction distance is further shortened to 3.13Å, indicating an improvement in processing nucleophilic attack. All these are consistent with the kinetics that the insertion efficiency of dGTP can be further enhanced

by the presence of Mn^{2+} ion, which are also well confirmed in other papers of polβ mediated catalysis as well.

Summarily, it seems that the C8 halogenation takes the Hoogsteen H bonds to base pair with the incoming dGTPs. The insertion efficiency increases from $8Br^-$ to $8Cl^-$, since the bulkiness of C8 halogenation affects the duplex conformation. This small shift in the structure can cause a serial of changes which ultimately changes the catalytic metal ion coordinations and its mediation of nucleotidyl transfer. Changing Mg^{2+} to Mn^{2+} demonstrates that the Hoogsteen base pair between the 8ClG:G base pair is still stable, and does not necessarily depend on the surrounding residues' supports. Mn^{2+} can enhance the insertion efficiency significantly because of its coordination ability to position the Asp 256 and 3'-OH into the catalytically active states.

Ternary structure of polβ with templating 8ClG paired with an incoming dATP analog

The kinetics results show that for the templaitng 8ClG, although it prefers the binding of dGTP and dCTP, there is still some insertion of dATP, especially under the Mn^{2+} presence. To elucidate the mechanism, the polβ-8ClG:dATP*-Mn^{2+} ternary structure is solved to 2.09Å resolution.

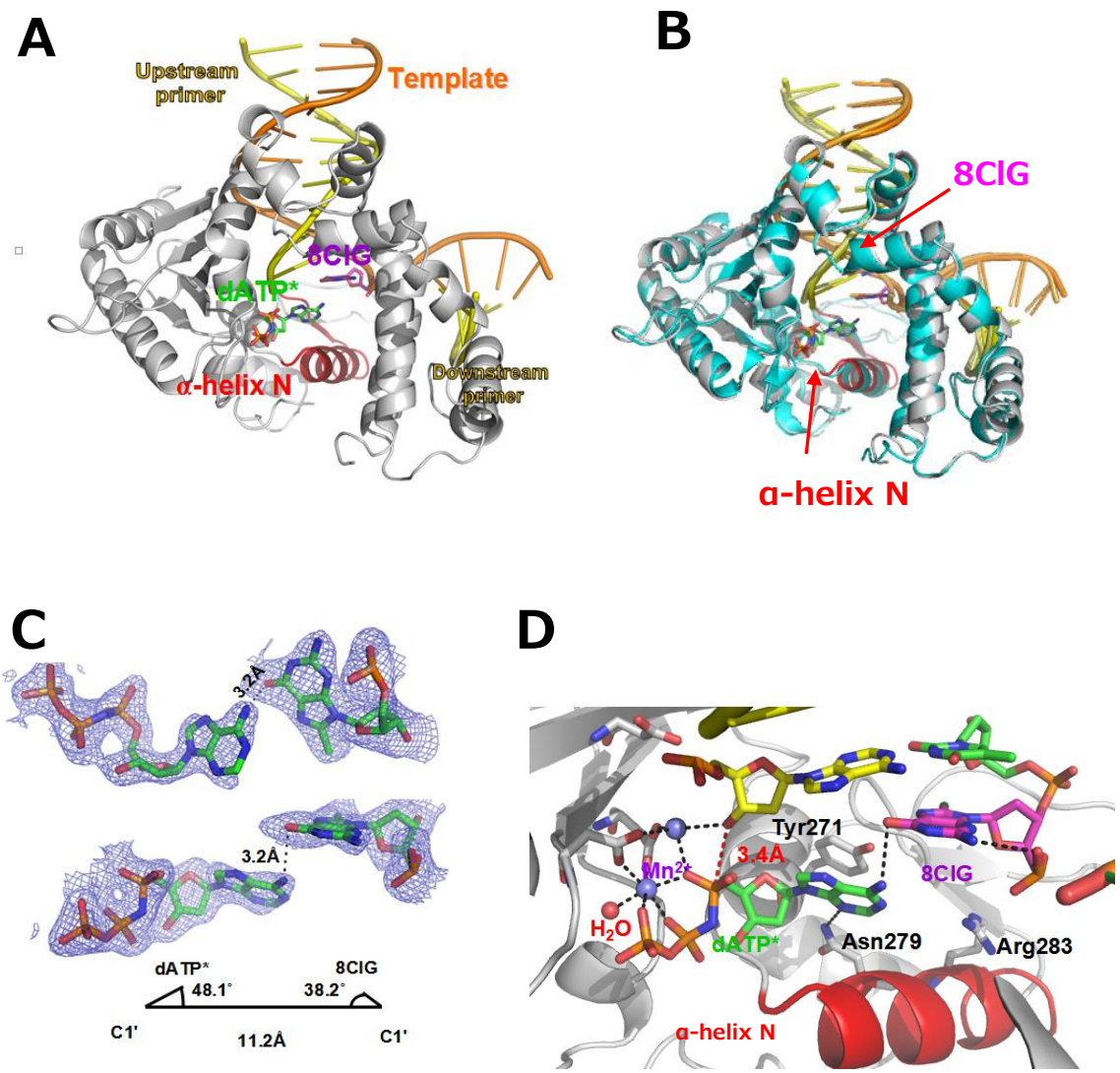


Figure 4.6 Polβ-8ClG:dATP*-Mn^{2+} structure. (A) Overall view of polβ-8ClG:dATP*-Mn^{2+} structure. 8ClG in purple. Incoming dATP* in green. (B) Alignment of polβ-8ClG:dATP*-Mn^{2+} structure with published polβ normal dG gapped binary structure (in cyan, PDB ID: 1BPX, r.m.s.d.=0.871Å). Note that the α-helix N position from 8ClG is a little nearer towards closed conformation. (C) Electron density of 8ClG:dATP* at 1 σ contour. Note the H bond in between and the stagger conformation. (D) Active site view of polβ-8ClG:dATP*-Mn^{2+} structure showing the minor groove recognition from Asn279.

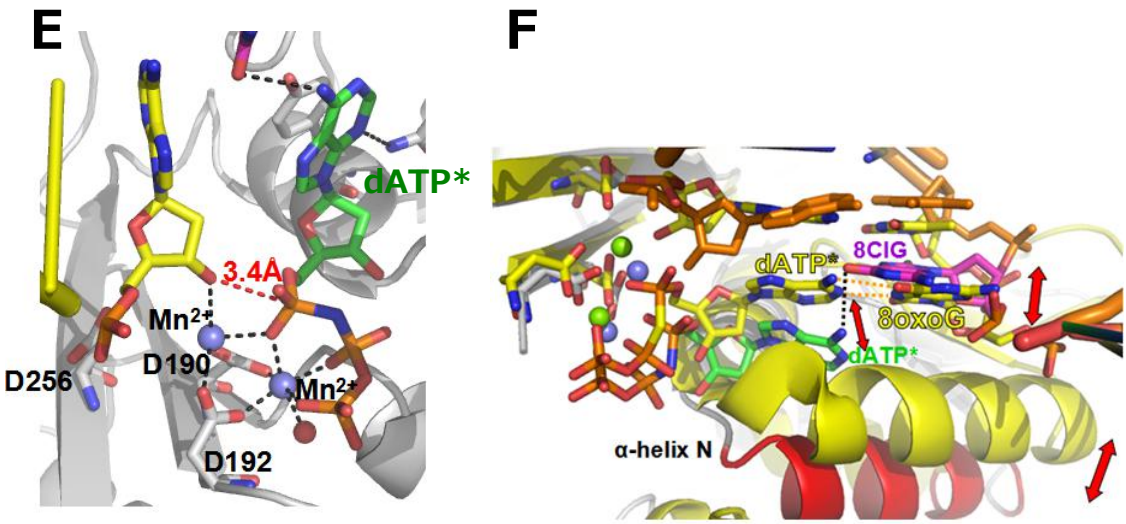


Figure 4.6 Polβ-8ClG:dATP*-Mn^{2+} structure. (E) Coordination view of polβ-8ClG:dATP*-Mn^{2+} structure. Nucleotidyl transfer reaction distance is shown in red. Water in red spheres. (F) Comparison of active site of polβ-8ClG:dATP*-Mn^{2+} structure (colors other than yellow) with polβ-8oxoG:dATP*-Mg^{2+} structure (yellow). Mn^{2+} in dark gray while Mg^{2+} in green spheres.

The overall polβ 8ClG:dATP*(Mn^{2+}) ternary structure is very similar to the published binary structure of gapped normal templating dG (Figure 4.4A and 4.4B, PDB ID: 1BPX, r.m.s.d.=0.871Å), suggesting the open conformation which agrees with the kinetic result of a low efficiency of dATP insertion.

By comparison (Figure 4.4B), the α-helix N of the 8ClG:dATP*-Mn^{2+} is taking the conformation between the intermediate and the totally open states. This correlates well with the base pairing condition of incoming dATP* with the templating 8ClG: they are of the staggering conformation with far less coplanarity than the dGTP*:8ClG base pair. Only one H bond is observed between dATP*:8ClG, which is bridging the exocyclic amine group of the dATP* and the exocyclic carbonyl oxygen of the 8ClG; while the *syn* conformation of 8ClG is stabilized by the H bond between its exocyclic amine group and

the phosphate oxygen of its backbone (Figure 4.4C). The staggering conformation of the dATP*:8ClG severely distorts the geometry of an ideal base pairing, with λ1=48.1°, λ2=38.2°, and the C1'-C1' distance 11.2Å.

Since the α-helix N of polβ is not totally in the open conformation, there is one minor groove recognition interaction: the Asn279 side chain makes the H bond to the N3 of the incoming dATP* to stabilize it. Neither Tyr271 stabilizes the primer terminus nor the Arg283 stabilizes the templating 8ClG (Figure 4.4D).

Two Mn^{2+} metal ions can be seen in the active site. Although the binding ion is ideally octahedrally coordinated, the catalytic metal ion misses the coordination from the Asp256 side chain. The 3'-OH is taking the in line nucleophilic attacking position and the distance between the 3'-OH to the Pα oxygen is about the proper range of 3.42Å (Figure 4.4E).

The kinetics result indicates that the incoming dATP, even in the presence of Mn^{2+}, is not so efficient, especially the binding affinity roughly presented by K_m. From the structure, the explanation is clear: there cannot be the Hoogsteen H bonds between the dATP and 8ClG, so the 8ClG:dATP* remains in a rather detached staggering conformation. This can be seen clearly by the alignment of 8ClG with 8oxoG towards the incoming dATP insertion (Figure 4.4F, PDB ID: 3RJF, r.m.s.d.=2.10Å)[159]. 8oxoG is making the perfect coplanarity with incoming dATP via Hoogsteen H bonds. And its α-helix N, minor groove interactions, catalytic and binding metal ions' coordinations are all consistent with ideal base pairing state. 8ClG is different from 8oxoG in that chlorination on the C8 atom would make the N7 still retains its saturation, so there lacks the needed H bond donor if Hoogsteen base pair were to form, since the N2 atom on the adenine is also saturated. On the contrary, for 8oxoG, because of the oxygen presence on the C8, N7 now becomes the NH, which is able to perform the H bond donor role in the Hoogsteen

base pair. So for the 8ClG:dATP base pair, even though the use of the Mn^{2+} has created the most possibilities for catalytic coordination and nucleotidyl transfer reaction, the staggering like conformation of the incoming dATP* and 8ClG severely decreases the insertion efficiency.

Ternary structure of polβ with templating 8ClG paired with an incoming dTTP analog

Our kinetics data shows that there is some insertion of dTTP to the templating 8ClG, although the efficiency is the lowest among all 4 dNTPs. To gain more insight into this never studied issue, the polβ-8ClG:dTTP*-Mg^{2+} ternary structure is solved and refined to 2.13Å.

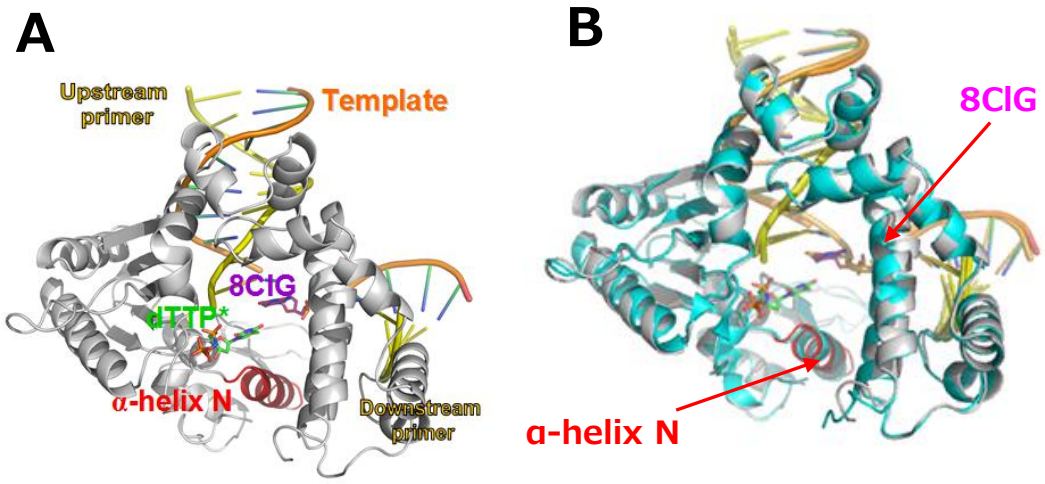


Figure 4.7 Polβ-8ClG:dTTP*-Mg^{2+} structure. (A) Overall view of polβ-8ClG:dTTP*-Mg^{2+} structure. 8ClG in purple. Incoming dTTP* in green. (B) Alignment of polβ-8ClG:dTTP*-Mg^{2+} structure with published polβ normal dG gapped binary structure (in cyan, PDB ID: 1BPX, r.m.s.d.=0.29Å).

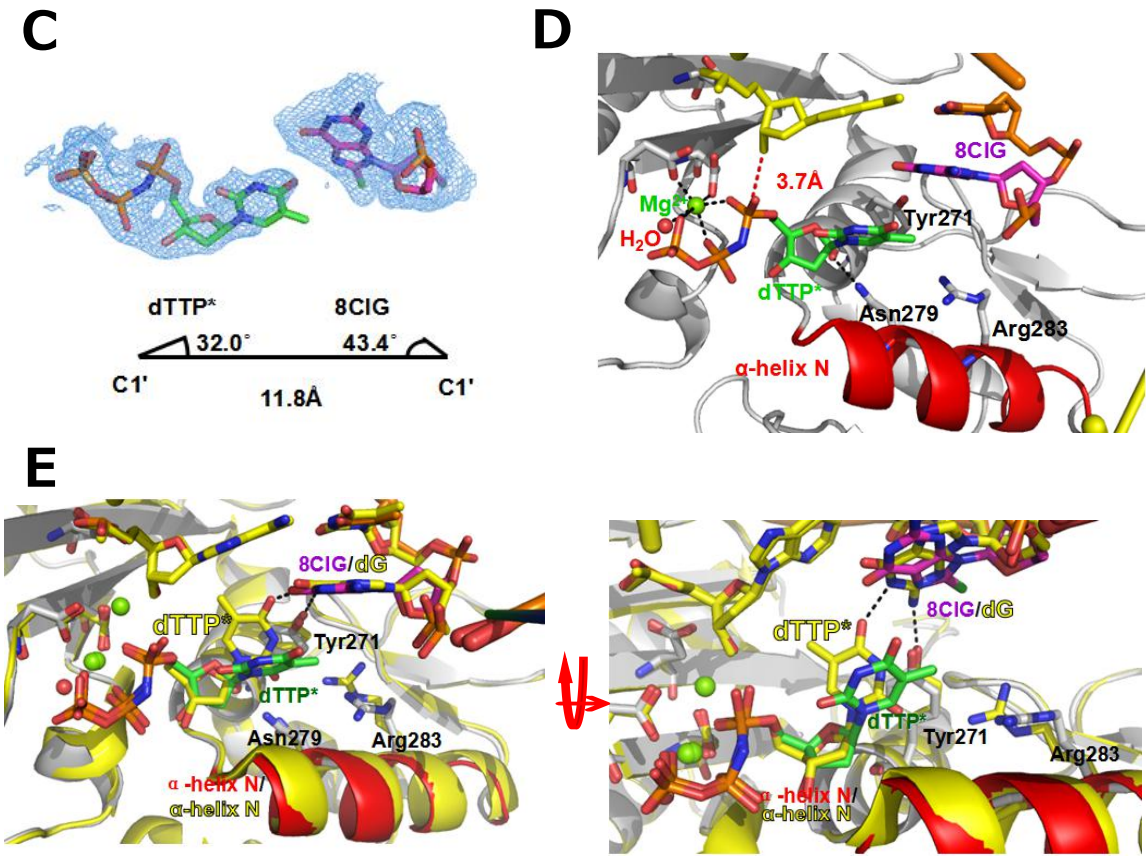


Figure 4.7 Polβ-8ClG:dTTP*-Mg^{2+} structure. (C) Electron density of 8ClG:dTTP* at 1 σ contour. Note there is no H bond. (D) Active site view of polβ-8ClG:dTTP*-Mg^{2+} structure. Note there is no minor groove interaction but H bond between Tyr271 to Asn279. (E) Active site comparison of polβ-8ClG:dTTP*-Mg^{2+} structure (color other than yellow) with polβ-G:dTTP*-Mg^{2+} (yellow, PDB ID: 4PGQ, r.m.s.d.=0.277Å) .

The overall structure is very similar to the binary gapped structure of 8ClG (PDB ID: 1BPX, r.m.s.d.=0.29Å). The α-helix N of the 8ClG:dTTP*-Mg^{2+} is taking the total open conformation (Figure 4.5A and 4.5B). The templating 8ClG base is in *syn* conformation. And the 8ClG and dTTP* are making no actual base pair or H bond contacts, and remain as a detached and staggered geometry. This is reflected by the geometric parameters, the λ 1=31.98°, λ 2=43.38°, and C1' to C1' distance is 10.79Å

(Figure 4.5C). And there is no contacts between the DNA and the minor groove recognition residues. In the active site, only one binding metal ion is observed, and coordinated from the surrounding side chains of Asp190, Asp192, phosphate groups of incoming dTTP* and a water molecule. Although the 3'-OH of the primer terminus is taking the in line nucleophilic attacking position towards Pα, The nucleotidyl transfer reaction distance is 3.73Å, exceeding the proper range of 3.4Å (Figure 4.5D).

By comparison of the active site structures, there is a clear explanation for the less insertion efficiency observed of dTTP to dATP for templating 8ClG. 8ClG:dATP* enjoys one H bond contact; closer position of α-helix N (~0.7Å); minor groove interaction of Asn279 with the incoming dATP*; presence of the catalytic metal ion with coordination; better in-line attacking position of 3'-OH of primer terminus and shorter nucleotidyl transfer reaction distance. In the dTTP* case, neither of these exists.

We have previously solved the G(templating):T mismatch complex structure of polβ in the presence of Mg^{2+} (PDB ID: 4pGQ)[80]. By alignment, the 8ClG:dTTP*-Mg^{2+} ternary structure is similar to this G:T-Mg^{2+} mismatch structure complex (r.m.s.d.=0.277Å). However, the insertion efficiency of dTTP to the templating 8ClG is lower than the normal dG. This can be explained by the structural difference from the two active sites: i) In G:T-Mg^{2+} mismatch structure, one H bond is observed between the incoming dTTP* and dG, while none observed in 8ClG case. ii) In G:T-Mg^{2+} mismatch structure, Tyr271 stabilizes the dG by H bond, while no minor groove interactions observed for 8ClG. iii) In G:T-Mg^{2+} mismatch structure, although the side chain of Asp 256 is not in the catalytic active position, there exists the catalytic metal ion coordinating the 3'-OH of the primer terminus and the Pα. In contrast, 8ClG:dTTP*-Mg^{2+} has none of such. All these difference stems from the C8 chlorination of the guanine, which facilitates its staying in the *syn* conformation. As can be seen from the structural comparison

(Figure 4.5E), the *syn* conformer of 8ClG rotates the exocyclic amine to the far away side abolishing the possibility of its potential H bond to the carbonyl oxygen of the incoming thymine, where as the amine of the *anti* conformer dG is within the range of making the H bond to the carbonyl oxygen of the incoming dTTP*. This is also reflected by the base pair geometry. While the 8ClG:dTTP* is basically a detached and staggered base pair, the G:dTTP* is, as reported, the pseudo-propeller twist conformation, in which the H bond associates the two bases.

In summary, the 8ClG made by the chlorination on deoxyguanosine preferably takes the *syn* conformation because of the clash of chlorine atom to the oxygen on the sugar ring. Because of the relatively smaller radius of chlorine to bromine, the templating 8ClG can induce the small shifts of the DNA duplex, which would bring a series of active site changes that improve the insertion efficiency of incoming dGTP* than the templating 8BrG. For the non-mutagenic dCTP* insertion, the 8ClG base takes the *anti* conformation, and thus classic Watson-Crick base pair is formed. However, the insertion efficiency is lower than that of the 8BrG. Still the small changes brought by the radius difference may account for this.

Templating 8ClG shows low insertion efficiency for both incoming dATP and dTTP, both kinetically and structurally. Structurally, the C8 chlorination is not able to change the N7 to NH, thus Hoogsteen base pair is hard to be maintained simply by one H bond. The observed H bond is bridging the amine from dATP* to the O6 of the 8ClG. Additionally to 8ClG:dTTP* ternary structure, the incoming dATP* is also stabilized by the minor groove interaction with well coordinated catalytic metal ion and short nucleotidyl transfer reaction distance. So it is reasonable that insertion of dATP* is more efficient than the insertion of dTTP*, where there is neither base pair H bonds, nor the other catalytic improving geometries.

4.4 SUMMARY OF CHAPTER IV

As aforementioned, 8ClG is the major product during the chronic inflammation, and can be produced easily from daily and industrial contacts. This study has demonstrated, in the polβ model, the 8ClG can be more mutagenic than the 8BrG. Since the polβ is mainly involved in BER and remains as a high fidelity DNA polymerase in the human body, the strong mutagencity of 8ClG, especially considering its easy production, probably requires more attention for the future cancer prevention.

More specifically, from the 3'-OH coordination by the catalytic metal ion in the study of this chapter, it seems that the 3'-OH position, although under coordination, can take two conformations, one is in a better in-line attacking position than the other one. Though the exact reason for this is unknown. It is likely that the small shift induced by the base pairing may be the reason. As can be seen from the above structures, the bigger difference between the base structure, the bigger difference of the positions of 3'-OH.

It can also be drawn from the structures of 8ClG with incoming dATP* and dTTP* that the coplanarity between the base pairs are crucial to the insertion efficiency. And obviously, the stable and acceptable coplanarity is formed by two H-bonds, as in the case of 8oxoG and dATP*. And two H-bonds may be enough already, as comparing the 8ClG:dGTP* under Mn^{2+} and Mg^{2+}, in which the former one does not possess a third water mediated H-bond as that of the latter one. However, the Mn^{2+} presence still improves the insertion efficiency of the incoming dGTP*.

Bibliography

[1]Esteller M. Epigenetic gene silencing in cancer: the DNA hypermethylome. Hum Mol Genet 2007; 16: R50-R59.

[2]Sedivy JM., et al. Aging by epigenetics-a consequence of chromatin damage? Exp Cell Res 2008; 314:1909-1917.

[3]Hill-Perkins M., et al. Site-specific mutagenesis in vivo by single methylated or deaminated purine bases. Mutat. Res 1986; 162, 153–163.

[4]Lindahl,T. Instability and decay of the primary structure of DNA. Nature 1993; 362, 709–715.

[5]Jaruga,P., et al. Repair of products of oxidative DNA base damage in human cells. Nucleic Acids Res 1996; 24, 1389–1394.

[6]Marnett,L.J. Oxy radicals, lipid peroxidation and DNA damage. Toxicology 2002; 181/182, 219–222.

[7]Liehr,J.G. Is estradiol a genotoxic mutagenic carcinogen? Endocr. Rev. 2000; 21,40–54.

[8]Thierry D., et al. Bipyrimidine Photoproducts Rather than Oxidative Lesions Are the Main Type of DNA Damage Involved in the Genotoxic Effect of Solar UVA Radiation. Biochemistry 2003; 42 (30), 9221–9226.

[9]Ringdahl E., et al. Treatment of knee osteoarthritis. Am Fam Physician. 2011; 83(11):1287-92.

[10]Nahas R., et al. Complementary and alternative medicine for the treatment of major depressive disorder. Can Fam Physician. 2011; Jun;57(6):659-63.

[11]Bryan, G.T. et al. Metabolism of 4(5)-(3,3-dimethyl-1-triazeno)-imidazole-5(4)-carboxamide to 4(5)-aminoimidazole-5(4)-carboxamide in man. Biochem.Pharmacol. 1970; 19, 2043.

[12]Jiaping X. et al. Mechanisms of Cancer Induction by Tobacco-Specific NNK and NNN. Cancers 2014; 6(2), 1138-1156.

[13]Lawley, P.D. et al. DNA adducts from chemotherapeutic agents. Mutation Res. 1996; 355, 13

[14]Koag MC, Kou Y, Ouzon H, Lee S. Transition-state destabilization reveals how human DNA polymerase β proceeds across the chemically unstable lesion N7-methylguanine. Nucleic Acids Res. 2014 Jul; 42(13):8755-66.

[15] Marnett, L. J., et al. Endogenous DNA adducts: potential and paradox. Chem. Res. Toxicol.1993; 6, 771 - 785.

[16]Tomasz, M., Olson, J., and Mercado, C. M. Mechanism of the isotopic exchange of the C-8 hydrogen of purines in nucleosides and deoxyribonucleic acid. Biochemistry 1972; 11, 1235 - 1241.

[17]Lawley P.D., et al. Acidic dissociation of 7:9-dialkylguanines and its possible relation to mutagenic properties of alkylating agents. Nature 1961; 192:1081-1082.

[18]Lawley P.D., et al. Further studies on the alkylation of nucleic acids and their constituent nucleotides. Biochem. J. 1963; 89:127–138.

[19]Wang, M., et al. A cyclic N7, C-8 guanine adduct of N-nitrosopyrrolidine (NPYR): formation in nucleic acids and excretion in the urine of NPYR-treated rats. Chem. Res. Toxicol, 1997; 10, 772 - 778.

[20]Mao, H., et al. An intercalated and thermally stable FAPY adduct of aflatoxin B 1 in a DNA duplex: Structural refinement from 1 H NMR. Biochemistry 1998; 37, 4374 - 4387.

[21]Wiederholt, CJ., et al. Fapy-dG instructs Klenow Exo- to misincorporate deoxyadenosine. J. Am. Chem. Soc. 2002; 124, 7278 - 7279.

[22]Bergdorf,L.T., et al. Synthesis, stability, and conformation of the formamidopyrimidine G DNA lesion. Chem. Eur. J. 2002; 8, 293 - 301.

[23]Humphreys, W. G., et al. Structure of formamidopyrimidine adducts as determined by NMR using specifically 15 N labeled guanosine. Chem. Res. Toxicol. 1991; 4, 632 -636.

[24]Crine, P., et al. A study of DNA spontaneous degradation. Biochim. Biophys. Acta 1976; 442, 50 - 57.

[25]Zoltewicz, J. A., et al. Kinetics and mechanism of acid-catalyzed hydrolysis of some purine nucleosides. J. Am. Chem. Soc. 1970; 92, 1741 - 1750.

[26]McHugh, P. J., et al. Novel reagents for chemical cleavage at abasic sites and UV photoproducts in DNA. Nucleic Acids Res. 1995; 23, 1664 - 1670.

[27]Osborne, M. R., et al. Preparation of a methylated DNA standard, and its stability on storage. Chem. Res. Toxicol. 2000; 13, 257 - 261.

[28]Margison, G. P., et al. Methylated purines in the deoxyribonucleic acid of various Syrian-golden-hamster tissues after administration of a hepa-tocarcinogenic dose of dimethylnitrosamine. Biochem. J. 1976; 157,627 - 634.

[29]Johnson, S. P., et al. In vitro repair of phosphoramide mustard induced DNA interstrand cross-links. Proc. Am. Assoc. Cancer Res. 1990; 41, 98.

[30]Seongmin L., et al. Synthesis and Structure of Duplex DNA Containing the Genotoxic Nucleobase Lesion N7-Methylguanine. J. Am. Chem. Soc., 2008; 130 (35), pp 11570–11571.

[31]Szyfter K., et al. Tobacco smoke-associated N7-alkylguanine in DNA of larynx tissue and leucocytes. Carcinogenesis 1996;17:501-506.

[32]Lawley P.D., Brookes P. Further studies on the alkylation of nucleic acids and their constituent nucleotides. Biochem. J. 1963; 89:127–138.

[33]Sowers L.C., et al. DNA base modification: ionized base pairs and mutagenesis. Mutat. Res. 1987;177:201-218.

[34]Watanabe S., et al. Methylated DNA-binding domain 1 and methylpurine – DNA glycosylase link transcriptional repression and DNA repair in chromatin. Proc. Natl. Acad. Sci. U.S.A. 2003;100:12859-12864.

[35]Hayashibara K.C., Verdine G.L. Template-directed interference footprinting of cytosine contacts in a protein-DNA complex: potent interference by 5-aza-2 ' -deoxycytidine. Biochemistry 1992;31:11265-11273.

[36]Barnes DE, Lindahl T. Repair and genetic consequences of endogenous DNA base damage in mammalian cells. Annu Rev Genet. 2004; 38:445-76.

[37]Makridakis N.M., Reichardt J.K.V. Translesion DNA polymerases and cancer. Front. Genet. 2012;3:1-8.

[38]Sawaya M.R., et al. Crystal structures of human DNA polymerase beta complexed with gapped and nicked DNA: evidence for an induced fit mechanism. Biochemistry 1997; 36:11205-11215.

[39]Freudenthal B.D., et al. Observing a DNA polymerase choose right from wrong. Cell 2013;154:157-168.

[40] Osheroff W.P., et al. The fidelity of DNA polymerase β during distributive and processive DNA synthesis. J. Biol. Chem. 1999;274:3642-3650.

[41]Batra V., et al. Magnesium-induced assembly of a complete DNA polymerase catalytic complex. Structure 2006;14:757-766.

[42]Lin P., et al. Incorrect nucleotide insertion at the active site of a G:A mismatch catalyzed by DNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 2008;105:5670-5674.

[43]Abashkin Y.G., et al. Quantum chemical investigation of enzymatic activity in DNA polymerase β. A mechanistic study. J. Phys. Chem. B 2001;105:287-292.

[44]Batra V.K., et al. Amino acid substitution in the active site of DNA polymerase β explains the energy barrier of the nucleotidyl transfer reaction. J. Am. Chem. Soc. 2013;135:8078-8088.

[45]Batra V.K., et al. Structures of DNA polymerase beta with active-site mismatches suggest a transient abasic site intermediate during misincorporation. Mol. Cell 2008; 30:315-324.

[46]Koag M., Lee S. Metal-dependent conformational activation explains highly promutagenic replication across O6-methylguanine by human DNA polymerase β. J. Am. Chem. Soc. 2014;136:5709-5721.

[47]Wang W., et al. Structural evidence for the rare tautomer hypothesis of spontaneous mutagenesis. Proc. Natl. Acad. Sci. U.S.A. 2011;108:17644-17648.

[48]Xia S., Wang J., Konigsberg W.H. DNA mismatch synthesis complexes provide insights into base selectivity of a B family DNA polymerase. J. Am. Chem. Soc. 2013;135:193-202.

[49]Zhao Y., et al. Mehchanism of somatic hypermutation at the WA motif by human DNA polymerase η. Proc. Natl. Acad. Sci. U.S.A. 2013;110:8146-8151.

[50]Banerjee S., et al. Bypass of aflatoxin B1 adducts by the Sulfolobus solfataricus DNA polymerase IV. J. Am. Chem. Soc. 2011;133:12556-12568.

[51]Sahasrabudhe S.R., et al. Induction of GŸC to AŸT transitions by the acridine half-mustard ICR-191 supports a mispairing mechanism for mutagenesis by some bulky mutagens. Biochemistry 1990; 29:10899-10905.

[52]Beard W.A., et al. Influence of DNA structure on DNA polymerase β active site function. J. Biol. Chem. 2004; 279:31921-31929.

[53]Bebenek K., et al. Replication infidelity via a mismatch with Watson-Crick geometry. Proc. Natl. Acad. Sci. U.S.A. 2011; 108:1862-1867.

[54]Alvarez-Salgado F., et al.NMR assessment of the global shape of a non-labelled DNA dodecamer containing a tandem of G-T mismatches. Magn Reson Chem. 2006 Dec;44(12):1081-9.

[55]Riggs AD, Jones PA. 5-methylcytosine, gene regulation, and cancer.Adv Cancer Res. 1983; 40:1-30.

[56]Fersht AR., et al. Kinetic basis of spontaneous mutation. Misinsertion frequencies, proofreading specificities and cost of proofreading by DNA polymerases of Escherichia coli. J Mol Biol. 1982 Mar 25;156(1):37-51.

[57]Kettani A. Interlocked mismatch-aligned arrowhead DNA motifs. Structure. 1999 Jul 15;7(7):803-15.

[58]Sundquist WI, Klug A. Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops.Nature. 1989 Dec 14;342(6251):825-9.

[59]Beal PA, Dervan PB. Second structural motif for recognition of DNA by oligonucleotide-directed triple-helix formation. Science. 1991 Mar 15;251(4999):1360-3.

[60]Esposito D., et al. H-NS oligomerization domain structure reveals the mechanism for high order self-association of the intact protein. J Mol Biol. 2002 Dec 6;324(4):841-50.

[61]Cognet JA., et al. Solution conformation of an oligonucleotide containing a G.G mismatch determined by nuclear magnetic resonance and molecular mechanics. Nucleic Acids Res. 1991 Dec 25;19(24):6771-9.

[62]Li Y., et al. NMR and molecular modeling evidence for a G.A mismatch base pair in a purine-rich DNA duplex. Proc Natl Acad Sci U S A. 1991 Jan 1;88(1):26-30.

[63]Wing, R., et al. Crystal structure analysis of a complete turn of B-DNA. Nature 1980; 287, 755 - 758.

[64]Horace R. D., et al.Structure of a B-DNA dodecamer: Conformation and dynamics. Proc. Nati. Acad. Sci. April 1981; Vol. 78, No. 4, 2179-2183.

[65]Drew H., et al. High-salt d(CpGpCpG), a left-handed Z' DNA double helix. Nature. 1980 Aug 7;286(5773):567-73.

[66]Ferre-D’ A., et al. Crystallization and structure determination of a hepatitis delta virus ribozyme: use of the RNA-binding protein U1A as a crystallization module. J. Mol. Biol. 2000; 295, 541–556.

[67]Bowman BR., et al. Structure of the E. coli DNA glycosylase AlkA bound to the ends of duplex DNA: a system for the structure determination of lesion-containing DNA. Structure. 2008; Aug 6;16(8):1166-74.

[68]Lu L., et al. Structure determination of DNA methylation lesions N1-meA and N3-meC in duplex DNA using a cross-linked protein-DNA system. Nucleic Acids Res. 2010 Jul; 38(13):4415-25.

[69]Zheng G., et al. 3DNALandscapes: A Database for Exploring the Conformational Features of DNA, Nucleic Acids Res 2010; 38(Database issue), D267-D274.

[70]Xiang-Jun L., Wilma K. O. 3DNA: a software package for the analysis, rebuilding and visualization of three‐dimensional nucleic acid structures. Nucl. Acids Res. 2003; 31 (17): 5108-5121.

[71]John SL., and Donald H. The Thermodynamics of DNA Structural Motifs. Annual Review of Biophysics and Biomolecular Structure 2004; Vol. 33: 415-440.

[72]Frank S., and Simone B. Mismatch formation in solution and on DNA microarrays: how modified nucleosides can overcome shortcomings of imperfect hybridization caused by oligonucleotide composition and base pairing. Mol. BioSyst., 2008, 4, 232-245.

[73]H. R., et al. Single-stranded DNA: replacement of canonical by base-modified nucleosides in the minihairpin 5'-d(GCGAAGC)-3' and constructs with the aptamer 5'-d(GGTTGGTGTGGTTGG)-3'. Helv. Chim. Acta, 2004; 87, 536.

[74]I. Hirao, et al. Extraordinarily stable mini-hairpins: electrophoretical and thermal properties of the various sequence variants of d(GCGAAAGC) and their effect on DNA sequencing. Nucleic Acids Res., 1992; 20, 3891.

[75]D.J. Patel., et al, Dynamics of DNA duplexes containing internal G.T, G.A, A.C, and T.C pairs: hydrogen exchange at and adjacent to mismatch sites, Fed. Proc. 43 1984; 2663 – 2670.

[76]Sanchez AM., et al. Initiation of repair of A/G mismatches is modulated by sequence context. DNA Repair (Amst). 2003; Aug 12;2(8):863-78.

[77]Brown T., et al. Molecular structure of the G.A base pair in DNA and its implications for the mechanism of transversion mutations. Proc Natl Acad Sci U S A. 1986; Apr;83(8):2402-6.

[78]Modrich, P. Mechanisms and biological effects of mismatch repair. Annu. Rev. Genet. 1991; 25, 229 – 253.

[79]Petronzelli, F., et al. Biphasic kinetics of the human DNA repair protein MED1 (MBD4), a mismatch-specific DNA N-glycosylase. J. Biol. Chem. 2000; 275, 32422 – 32429.

[80]Koag MC., et al. The spontaneous replication error and the mismatch discrimination mechanisms of human DNA polymerase β. Nucleic Acids Res. 2014; 42(17):11233-45.

[81]Isaacs RJ., et al. Structural differences in the NOE-derived structure of G-T mismatched DNA relative to normal DNA are correlated with differences in (13)C relaxation-based internal dynamics. J Mol Biol. 2002; May 24;319(1):191-207.

[82]Allawi HT., SantaLucia J. NMR solution structure of a DNA dodecamer containing single G.T mismatches. Nucleic Acids Res. 1998; Nov 1;26(21):4925-34.

[83] Hunter, W. N., et al. The structure of guanosine – thymidine mismatches in B-DNA at 2.5-A resolution. J. Biol. Chem. 1987; 262, 9962 – 9970.

[84]Katarzyna B., et al. Replication infidelity via a mismatch with Watson – Crick geometry. Proc Natl Acad Sci U S A. 2011; Feb 1;108(5):1862-7.

[85]Cognet, J. A. H., et al. Solution conformation of an oligonucleotide containing a G: G mismatch determined by nuclear magnetic resonance and molecular mechanics, Nucleic Acids Res. 1991; 19,6771 -6779.

[86]Borden, K. L. B., et al. Conformational properties of the G : G mismatch in d(CGCGAATTGGCG), determined by NMR, Biochemistry 1992; 31, 541 1-5422.

[87]Campbell NH., Neidle S. G-quadruplexes and metal ions. Met Ions Life Sci. 2012; 10:119-34.

[88]Bochman ML., et al. DNA secondary structures: stability and function of G-quadruplex structures. Nat Rev Genet. 2012; Nov;13(11):770-80.

[89]Maizels N., Gray LT. The G4 genome.PLoS Genet. 2013; Apr;9(4):e1003468.

[90]Lane AN., Peck B. et al. Conformational flexibility in DNA duplexes containing single G.G mismatches. Eur J Biochem. 1995; Jun 15;230(3):1073-87.

[91]Bhattacharya PK., et al. 1H NMR determination of base-pair lifetimes in oligonucleotides containing single base mismatches.Nucleic Acids Res. 2002; Nov 1;30(21):4740-50.

[92]Skelly JV., et al. Crystal structure of an oligonucleotide duplex containing G.G base pairs: influence of mispairing on DNA backbone conformation. Proc Natl Acad Sci U S A. 1993; Feb 1;90(3):804-8.

[93]Dai J., et al. Structure of the intramolecular human telomeric G-quadruplex in potassium solution: a novel adenine triple formation. Nucleic Acids Res. 2007; 35(7):2440-50.

[94]Larson, J. W.; McMahon, T. B. Gas-phase bihalide and pseudobihalide ions. An ion cyclotron resonance determination of hydrogen bond energies in XHY- species (X, Y = F, Cl, Br, CN). Inorganic Chemistry 1984; 23 (14): 2029 – 2033.

[95]Emsley, J. Very Strong Hydrogen Bonds". Chemical Society Reviews 1980; 9 (1): 91 – 124.

[96]Hemminki K., et al. Reactivity, SCE induction and mutagenicity of benzyl chloride derivatives.J Appl Toxicol. 1983 Aug;3(4):203-7.

[97]Parry JM, Wilcox P. The genetic toxicology in fungi of 4-chloromethylbiphenyl (4CMB), 4-hydroxymethylbiphenyl (4HMB) and benzyl chloride (BC). Survey of the results of the U.K.E.M.S. collaborative genotoxicity trial 1981. Mutat Res. 1982 Jan-Feb;100(1-4):185-200.

[98]Coombs MM. Attempts to initiate skin tumours in mice in the 2-stage system using 4-chloromethylbiphenyl (4CMB), 4-hydroxymethylbiphenyl (4HMB), and benzyl chloride (BC). Report of the experiment at 10 months. Mutat Res. 1982 Jan-Feb;100(1-4):403-5.

[99]Fukuda K., et al. Carcinogenicity of benzyl chloride, benzal chloride, benzotrichloride and benzoyl chloride in mice by skin application. Gan. 1981 Oct;72(5):655-64.

[100]Lijinsky W. Chronic bioassay of benzyl chloride in F344 rats and (C57BL/6J x BALB/c)F1 mice. J Natl Cancer Inst. 1986 Jun;76(6):1231-6.

[101]Nair B. Final report on the safety assessment of Benzyl Alcohol, Benzoic Acid, and Sodium Benzoate. Int J Toxicol. 2001;20 Suppl 3:23-50.

[102]Thomas H. S., Ronald J. L. Catalog of Teratogenic Agents. JHU Press, 2004.

[103]Druckrey H., et al. Organotropic carcinogenic effects of 65 various N-nitroso-compounds on BD rats. Z Krebsforsch. 1967;69(2):103-201.

[104]Stinson SF., et al. Pathology of esophageal neoplasms and associated proliferative lesions induced in rats by N-methyl-N-benzylnitrosamine. J Natl Cancer Inst. 1978 Dec;61(6):1471-5.

[105]Sander J, Schweinsberg F. Tumorinduction in mice with methylbenzylnitrosamine in low doses. Z Krebsforsch Klin Onkol Cancer Res Clin Oncol. 1973 May 7;79(3):157-61.

[106]Peterson LA. N-Nitrosobenzylmethylamine is activated to a DNA benzylating agent in rats. Chem Res Toxicol. 1997 Jan;10(1):19-26.

[107] Moschel, R. C., et al. Substituent-induced effects on the stability of benzylated guanosines: model systems for the factors influencing the stability of carcinogen-modified nucleic acids. J. Org. Chem. 1984; 49, 363 - 372.

[108] Muller, N., and Eisenbrand, G. The influence of N7-substitutents on the stability of N7-alkylated guanosines. Chem.-Biol. Interact. 1985; 53, 173 - 181.

[109]Koag MC., Lee S. Metal-dependent conformational activation explains highly promutagenic replication across O6-methylguanine by human DNA polymerase β. J Am Chem Soc. 2014 Apr 16;136(15):5709-21.

[110]Alaee M., et al. An overview of commercially used brominated flame retardants, their applications, their use patterns in different countries/regions and possible modes of release. Environ Int. 2003 Sep;29(6):683-9.

[111]Weil, Edward D.; Levchik, Sergei. A Review of Current Flame Retardant Systems for Epoxy Resins. Journal of Fire Sciences2004; 22: 25.

[112]Thomas, G. Medicinal Chemistry an Introduction. John Wiley & Sons, West Sussex, UK. 2000; ISBN 978-0-470-02597-0.

[113]Delzell, E., et al. Interpretive Review of the Potential Adverse Effects of Chlorinated Organic Chemicals on Human Health and the Environment. Regulatory Toxicology and Pharmacology 20 (1, Part 2 of parts) 1994; S1-S1056.

[114]Fair, G. M., et al. The behavior of chlorine as a water disinfectant". J. Am. Water Works Assoc. 1948; 40: 1051 – 1061.

[115]Chesney JA., et al. Bacterial glutathione: a sacrificial defense against chlorine compounds. J Bacteriol. 1996 Apr;178(7):2131-5.

[116]Venkobachar C., et al. Mechanism of disinfection. Water Research. Vol 9, Issue 1, January 1975, 119 – 124.

[117]Albrich JM., et al. Biological reactivity of hypochlorous acid: implications for microbicidal mechanisms of leukocyte myeloperoxidase. Proc Natl Acad Sci U S A. 1981 Jan;78(1):210-4.

[118] Barrette Jr, WC., et al. Hypochlorous acid-promoted loss of metabolic energy in Escherichia coli. Infection and immunity 1987; 55 (10): 2518 – 25.

[119]Barrette Jr, WC; et al. General mechanism for the bacterial toxicity of hypochlorous acid: abolition of ATP production. Biochemistry 1989; 28 (23): 9172 – 8.

[120]Winter, J., et al. Bleach Activates a Redox-Regulated Chaperone by Oxidative Protein Unfolding. Cell 2008; 135 (4): 691 – 701.

[121]McKenna, SM; Davies, KJ. The inhibition of bacterial growth by hypochlorous acid. Possible role in the bactericidal activity of phagocytes. The Biochemical journal 1988; 254 (3): 685 – 92.

[122]Rosen, H et al. Differential effects of myeloperoxidase-derived oxidants on Escherichia coli DNA replication. Infection and immunity 1998; 66 (6): 2655 – 9.

[123]Hawkins CL., et al. Hypochlorite- and hypobromite-mediated radical formation and its role in cell lysis. Arch Biochem Biophys. 2001 Nov 15;395(2):137-45.

[124]Liviac D., et al. Genotoxicity of Water Concentrates from Recreational Pools after Various Disinfection Methods Environ. Sci. Technol. 2010; 44 (9), pp 3527 – 3532.

[125]McDonald TA and Komulainen H. Carcinogenicity of the chlorination disinfection by-product MX. J Environ Sci Health C Environ Carcinog Ecotoxicol Rev. 2005; 23(2):163-214.

[126]Richardson SD., et al. Occurrence, genotoxicity, and carcinogenicity of regulated and emerging disinfection by-products in drinking water: a review and roadmap for research. Mutat Res. 2007; 636(1-3):178-242.

[127]Masuda M., et al. Chlorination of guanosine and other nucleosides by hypochlorous acid and myeloperoxidase of activated human neutrophils. Catalysis by nicotine and trimethylamine.J Biol Chem. 2001 Nov 2;276(44):40486-96.

[128]Whiteman M., et al. Hypochlorous acid-induced base modifications in isolated calf thymus DNA. Chem Res Toxicol. 1997 Nov;10(11):1240-6.

[129]Li X., et al. Targeting mitochondrial reactive oxygen species as novel therapy for inflammatory diseases and cancers.J Hematol Oncol. 2013 Feb 25;6:19.

[130]Hancock, J.T., et al. Role of Reactive Oxygen Species in Cell Signaling Pathways. Biochemical and Biomedical Aspects of Oxidative Modification 2001; 29(2):345-350.

[140]Dolphin, D., et al. Compounds I of Catalase and Horse Radish Peroxidase: π - Cation Radicals. Proc. Natl. Acad. Sci. U.S.A 1971; 68, 614−618.

[141]Heinecke JW., et al. Tyrosyl radical generated by myeloperoxidase catalyzes the oxidative cross-linking of proteins. J Clin Invest. 1993 Jun;91(6):2866-72.

[142]Thomas EL., et al. Oxidation of bromide by the human leukocyte enzymes myeloperoxidase and eosinophil peroxidase. Formation of bromamines. J Biol Chem. 1995 Feb 17;270(7):2906-13.

[143]Mayeno AN., et al. Eosinophils preferentially use bromide to generate halogenating agents. J Biol Chem. 1989 Apr 5;264(10):5660-8.

[144]Ramsey PG., et al. Arming of mononuclear phagocytes by eosinophil peroxidase bound to Staphylococcus aureus. J Immunol. 1982 Jan;128(1):415-20.

[145]Henderson JP., et al. Production of brominating intermediates by myeloperoxidase. A transhalogenation pathway for generating mutagenic nucleobases during inflammation. J Biol Chem. 2001 Mar 16;276(11):7867-75.

[146]Sassa A., Mutational specificities of brominated DNA adducts catalyzed by human DNA polymerases. J Mol Biol. 2011 Mar 11;406(5):679-86.

[147]Bouvard V., et al. A review of human carcinogens--Part B: biological agents. Lancet Oncol. 2009 Apr;10(4):321-2.

[148]Whitcomb DC. Inflammation and cancer.V. Chronic pancreatitis and pancreatic cancer. Am J Physiol Gastrointest Liver Physiol 2004; 287:G315 – G319.

[149]Olliver JR., et al. Risk factors, DNA damage, and disease progression in Barrett's esophagus. Cancer Epidemiol Biomarkers Prev. 2005 Mar;14(3):620-5.

[150]Coussens LM, Werb Z. Inflammation and cancer. Nature 2002;420:860 – 7.

[151]Ohshima H, et al .Chemical basis of inflammation-induced carcinogenesis. Arch Biochem Biophys 2003;417:3 – 11.

[152]Koag MC., et al. Structural basis for promutagenicity of 8-halogenated guanine. J Biol Chem. 2014 Feb 28;289(9):6289-98.

[153]Steenken S, Jovanovic SV. How easily oxidizable is DNA? One-electron reduction potentials of adenosine and guanosine radicals in aqueous solution. J Am Chem Soc 1997; 119:617 – 18.

[154]Hamm ML., et al. Base pair stability of 8-chloro- and 8-iodo-2'-deoxyguanosine opposite 2'-deoxycytidine: implications regarding the bioactivity of 8-oxo-2'-deoxyguanosine. J Am Chem Soc. 2005 Sep 7; 127(35):12220-1.

[155]Otwinowski, Z., and Minor, W. Processing of x-ray diffraction data. Methods Enzymol. 1997; 276, 307 – 326.

[156]Emsley P., Cowtan K. COOT: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. 2004; 60, 2126 – 2132.

[157]Adams P. D., et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. 2010; 66, 213 – 221.

[158]Davis I. W., et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res. 2007; 35, W375 – W383.

[159]Batra, V. K., et al. Binary complex crystal structure of DNA polymerase beta reveals multiple conformations of the templating 8-oxoguanine lesion. Proc. Natl. Acad. Sci. U.S.A. 2012; 109, 113 – 118.