1-1-2017

Real-Time Investigation Of Bulky Lesion Bypass By Y-Family Dna Polymerase, Dpo4, Using Single Molecule Fret

Pramodha Liyanage
Wayne State University,

REAL-TIME INVESTIGATION OF BULKY LESION BYPASS BY Y-FAMILY DNA POLYMERASE, DPO4, USING SINGLE MOLECULE FRET

by

PRAMODHA S. LIYANAGE

DISSERTATION

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

2017

MAJOR: CHEMISTRY (Biochemistry)

Approved By:

_____

Advisor Date

_____

_____

_____

_____

DEDICATION

This thesis is dedicated to my Family.

ACKNOWLEDGEMENTS

It is with greatest respect and admiration I express my deepest sense of gratitude to my advisor, Prof. Louis Romano for his optimistic support and confidence extended towards me during my work for this dissertation. It would also like to express my sincere gratitude and appreciation to my former adviser Prof. David Rueda for his continues motivation and directions. I am extremely fortunate to be skilled under Dr. Romano and Dr. Rueda who are experts with biochemical and single molecule techniques. Their collective insight greatly inspired me to proceed with the work to a successful completion of my PhD. I gratefully appreciate my graduate committee Prof. Bernhard Schlegel , Prof. Andrew Feig and Prof. Athar Ansari for continues support and valuable advice.

I would take this opportunity to express my sincere thanks to the Romano's lab single molecule group members Tom, Radek, Kyle, and Alfonso for their assistance and guidance. It was a great resource having them around when I need an advice on troubleshooting. Even after leaving the lab they never turned me down when I asked for help.

I would like to thank all the academic and non-academic staff, Department of Chemistry Wayne State University for providing me the knowledge and facilities to carry out my studies.

I am ever thankful to my wife for bearing with me throughout these difficult years balancing the family and supporting me towards the successful completion of my project. It was not easy for both of us to fallow our doctoral studies with two kids.

Finally, with heartfelt gratitude and love I remember my beloved parents for all the inspirations, guidance and sacrifices they made eagerly for my success.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

TEEA	triethyl ammonium acetate
HPLC	High pressure liquid chromatography
MALDI TOF	Matric assisted laser desorption ionization time of flight
ACN	Acetonitrile
3-HPA	3-Hydroxypicolinic acid
DMF	Dimethylformamide
DIPEA	N,N-Diisopropylethylamine
Dpo4	DNA polymerase IV
B[*a*]P	Benzo[*a*]pyrene
2-AAF	2-acetylaminofluorene
TIRF	Total internal reflation fluorescence microscope
smFRET	single molecule Förster resonance energy transfer
smPIFE	single molecule protein induced energy transfer
Tris-HCL	Tris-hydroxymethyl aminomethane hydrochloride
IBC	Insertion binary complex
PIBC	Preinsertion binary complex
KF	Klenow fragment
BF	Bacillus fragment
K_{off}	Dissociation rate constant
dNTP	2’-deoxyribonucleoside 5’-triphosphate
E. coli	Escherichia coli
Cy3	Cyanine 3

Cy5	Cyanine 5
BPDE	Benzo[*a*]pyrene diol epoxide

CHAPTER 1: INTRODUCTION

1 DNA damage and cancer

Transformation of genetic information from parents to their offspring is the fundamental requirement for the existence of all living beings on earth. Therefore, the genetic information transfer process must be done with greater accuracy and efficiency than that of all other biological processes. As the holder of genetic information, DNA must be protected from threats that cause DNA damage. To accomplish this goal, cells have developed various defensive, repair and surveillance mechanisms in the cell [1-3]. However, all these mechanisms have limitations that cause some of the DNA damage and create health problems such as cancer, one of the main health issues that humans face today as it is second only to human death by cardiovascular diseases [4]. Conversion of a normal healthy cell into a tumorigenic cell is a complex and catastrophic process. Often mutations cause such cellular transformations that by resulting the activation of oncogenes or inactivation of tumor suppressor genes.

Somatic mutations are generated by various exogenous and endogenous factors, low fidelity in DNA replication and imperfect DNA repair machinery [5-8]. Tobacco smoking, exposure to UV light, certain food (grilled, fried and red meet), obesity, infections, stress and lack of physical activity are some of the lifestyle factors that may have a link with cancer initiation [9]. Among all the cancer types that have been diagnosed in humans, more than 25% of cancers are directly related to smoking [10] and It has been determined that benzo[*a*]pyrene (B[*a*]P) is one of the key DNA damaging chemicals in cigarette smoke. B[*a*]P has an ability to chemically modify the DNA, and it has been shown to disrupt DNA replication. One mechanism that is used to

overcome these types of damages is employed by Y-family polymerases. This research is focused on studying Dpo4's conformational changes and related mutagenic properties during processing B[*a*]P-modified DNA in real time using smFRET.

1.1 Chemical carcinogenesis

There has been a long history of research on chemical carcinogens. In 1761John Hill made the First recorded observation of a chemical carcinogenesis when he linked increased levels of nasal cancer with the use of snuff. However, the first dated study of chemical carcinogenesis was done by Sir Percival Pott in 1875. Pott observed a common occurrence of scrotum cancer among chimney sweepers in England and inferred that scrotum cancer was caused by excessive exposure to soot. His observation is also considered the first report of occupational-related cancer. Pott's observation laid the foundation for the discovery of many environment carcinogens that are now known to cause cancer in humans today [11]. There are massive levels of carcinogens in our environment and it is unavoidable that some of these chemicals will enter the human body by air, food or water.

Most known chemical carcinogens can be generally classified into two groups based on their functions [12]: ether genotoxic or epigenetic carcinogens. Genotoxic carcinogens directly react with DNA to generate mutations. Epigenetic carcinogens do not react with DNA directly, but they interfere with the DNA expression and cell proliferation [13]. One of the major obstacles to identifying carcinogenic chemicals and their relationship to cancer is the long latency period between chemical exposure and cancer development [11]. Some of the known carcinogenic compounds, such as 2-napthylamine, benzidine and vinyl chloride, have a 20-30-year latency period.

In 1970, Millers found that many unreactive carcinogenic chemicals follow a common mechanism in carcinogenesis that requires cellular metabolism [11, 14]. They showed that 2-AAF must be metabolically activated to reactive species before DNA adducts can form. Later, many metabolic pathways and various enzymes were identified as being involved in carcinogenesis.

Some of the enzymes involve in these pathways introduce very reactive organic functional groups to the carcinogenic compounds that are known as activators. In contrast some other enzymes convert these functional groups into less reactive, water soluble compounds, which are known as inactivators. Therefore, carcinogenic chemical exposure leads to initiate two competitive metabolic reactions in the cells. The activated compounds have an infinite number of possibilities to react with other cellular molecules before inactivating. If these activated molecules react with DNA recruiting barriers for DNA replication, the cells get one step close for cancer initiation. DNA repair pathways play a key role in eliminating these carcinogenic adduct modified DNA to prevent them to be processed by polymerases to introduce mutations. The carcinogenic adducts escaped form DNA repair may induce mutations in the cells which may cause for cellular death or ultimately cancer. The processes involved in chemical carcinogenesis are summarized in the Figure 1 [11].

.

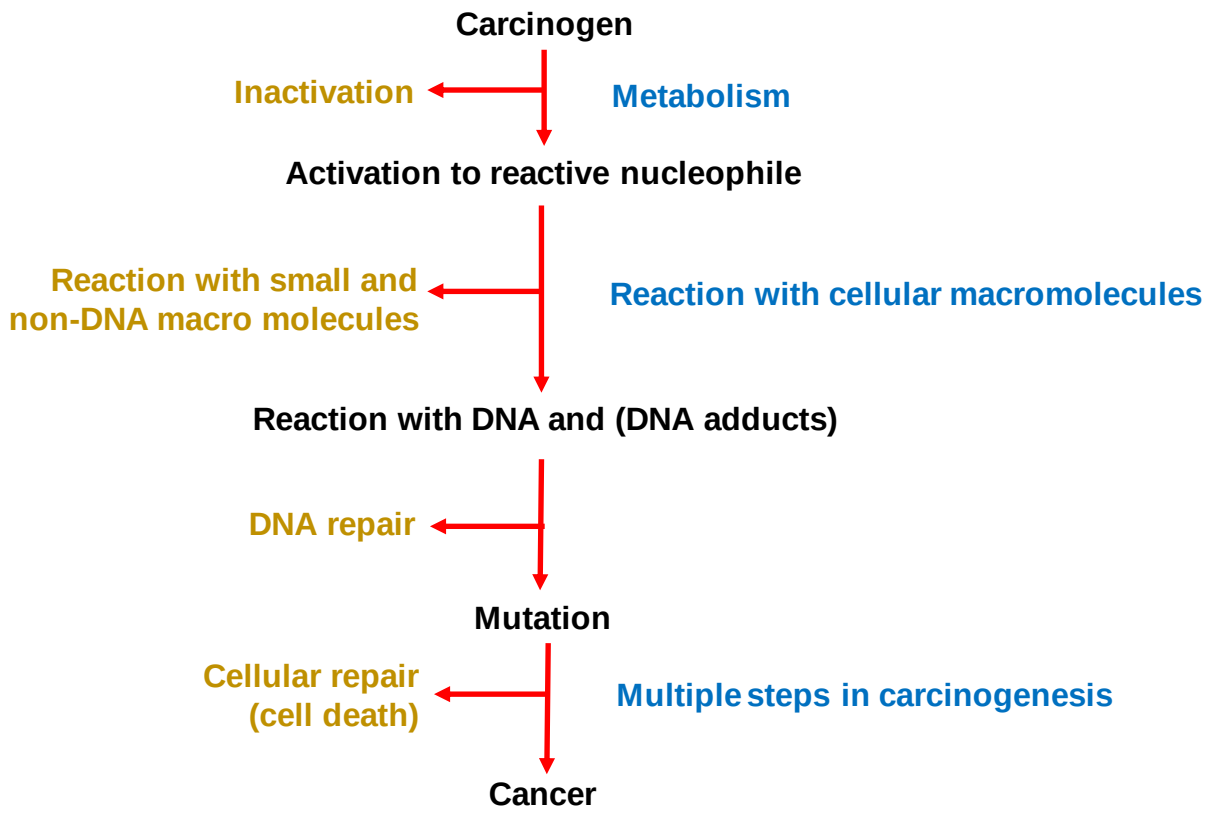


Figure 1. The process of chemical carcinogens leading to cancer. Chemical exposure causes competition between metabolic activation and inactivation. The activated chemical carcinogens may escape from the inactivation to react with biological macro molecules. These adducted molecules may slip through the reaper processes to cause mutations. Some of the mutations disrupt the cellular processes leading to cell death, and others tend to abolish the cell's regulatory process to convert it into a cancer. The figure was adapted from [11].

1.2 Benzo[*a*]pyrene

As mention above, many skin cancer cases were reported among workers in chimney cleaning, later correlation course found in other industries such as paraffin refining, shale oil and coal tar industries [15, 16]. Although scientists had tried to reproduce skin cancers by applying those industrial raw materials using animal models to understand the carcinogenicity, those experiments were not successful. After many years of effort, J.W. Cook, C.L. Hewett and I. Hieger in Kennaway's lab were finally able

to identify the benzo[*a*]pyrene (B[*a*]P) as a cause of the coal tar associated cancer (Figure 2) [15]. The B[*a*]P, which belongs to a group of potentially toxic compounds known as polycyclic aromatic hydrocarbons (PAHs) [16], is the first discovered and mostly studied exogenous cancer-causing agent [17]. Today, major sources of PAHs, B[*a*]P in particular, are byproducts of incomplete combustion of fossil fuels, including emissions from combustion engines, coke ovens, volcanic eruptions, forest fires and soot, smoke from coal or tar furnaces [18]. Among all the B[*a*]P sources, cigarette smoke is particularly important because over a million people die each year from lung cancers, which have been identified to show a direct association with cigarette smoking [19, 20].

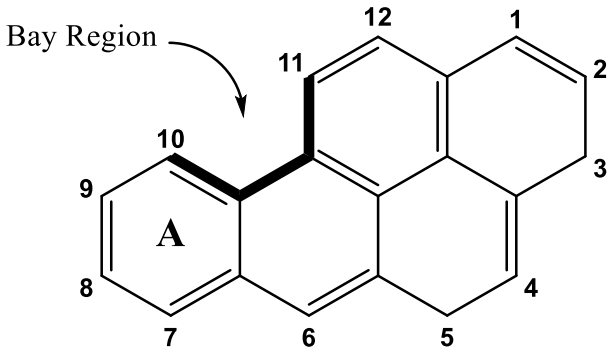


Figure 2. Structure of Benzo[*a*]pyrene (B[*a*]P). The highlighted bonds represent the boundary of the Bay-region.

Once burned, a cigarette contains about 20-40 ng of B[*a*]P, which has been proposed as the key carcinogenic compound responsible for lung cancer found in smokers [20, 21] An occurrence of a higher percentage (~60%) of mutations, such as G to T transversion, has been identified in human lung tumors. These transversion mutations are the hallmark of mutations caused by PAHs, particularly B[*a*]P [20].

1.2.1 Benzo[*a*]pyrene metabolism

B[*a*]P is a hydrophobic, planer, aromatic hydrocarbon that is chemically inert and insoluble in an aqueous environment (Figure 2). Therefore, once B[*a*]P is ingested and absorbed into the body, it has to be converted to a water-soluble compound in order to facilitate excretion. This process is called the metabolic activation and as a result of extensive research, three potential metabolic activation pathways have been identified. They are the bay-region dihydrodiol epoxide pathway, the one-electron oxidation pathway and the dihydrodiol dehydrogenase pathway [22, 23].

The B[*a*]P metabolic activation by the bay-region dihydrodiol epoxide pathway is thought to be the major activation route using metabolomics, DNA binding and carcinogenesis experiments. It comprises three enzymes-mediated biochemical reactions. The first reaction is the oxidation of C^7-C^8 double bond in B[*a*]P ring “A” (Figure 2). Epoxidation of these prochiral aromatic carbon atoms produce two optically active isomers. They can be identified as 7R,8S-epoxide and (Figure 2, compound I) 7S,8R-epoxide (not shown). Both isomers initiate a series of biochemical reactions. However, only 7R,8S-epoxide produces biologically important major products. Therefore, 7S,8R-epoxide and its subsequent reactions are not shown in Figure 3. This oxidation reaction is initiated by cytochrome P450 enzymes (CYPs) particularly monooxygenase enzymes [22-24].

The second reaction is the hydrolysis of B[*a*]P-7,8-epoxide to form trans B[*a*]P-7,8-diol which is conducted by microsomal epoxide hydrolase (EH). Hydroxyl group addition takes place via an inversion of configuration converting 7R,8S-epoxide into 7R,8R-diol (Figure 3, compound II) [25].

The third reaction is the oxidation of C^9-C^{10} double bond in B[*a*]P ring “A” to produce vicinal diol-epoxide, B[*a*]P-7,8-diol-9-10-epoxide (BPDE) which is driven by cytochrome p450 enzymes (Figure 3). This reaction further adds stereoisomeric complexity to B[*a*]P metabolic activation by making another two pairs of enantiomers which are named as (+)-*anti*- BPDE and (-)-*syn*- BPDE isomers (Figure 3, compound III and IV).

The product of the third reaction can be further hydrolyzed to make B[*a*]P-7,8,9,10-tetrol which is a highly hydrophilic compound that can be excreted from the body [22, 26]. Based on quantitative metabolism studies, it has been identified that (±)-*anti*- BPDE (compound II) is the major and most biologically active BPDE isomer, possibly due to its longer half-life [15].

1.2.2 Structural characteristics of B[*a*]P isomers

An epoxide is a highly reactive organic group because it can efficiently undergo nucleophilic addition reactions. Therefore, BPDE formed during the B[*a*]P metabolic activation process can react with various cellular macromolecules such as DNA, RNA and protein. It was first thought that instead of reacting with DNA, the planer BPDE would intercalate between two base pairs because it could be stabilized by hydrophobic interactions with adjacent bases in the core of the DNA helix [27]. However, it was confirmed that B[*a*]P can form a covalent bond with DNA, mostly reacting with purine residues [28].

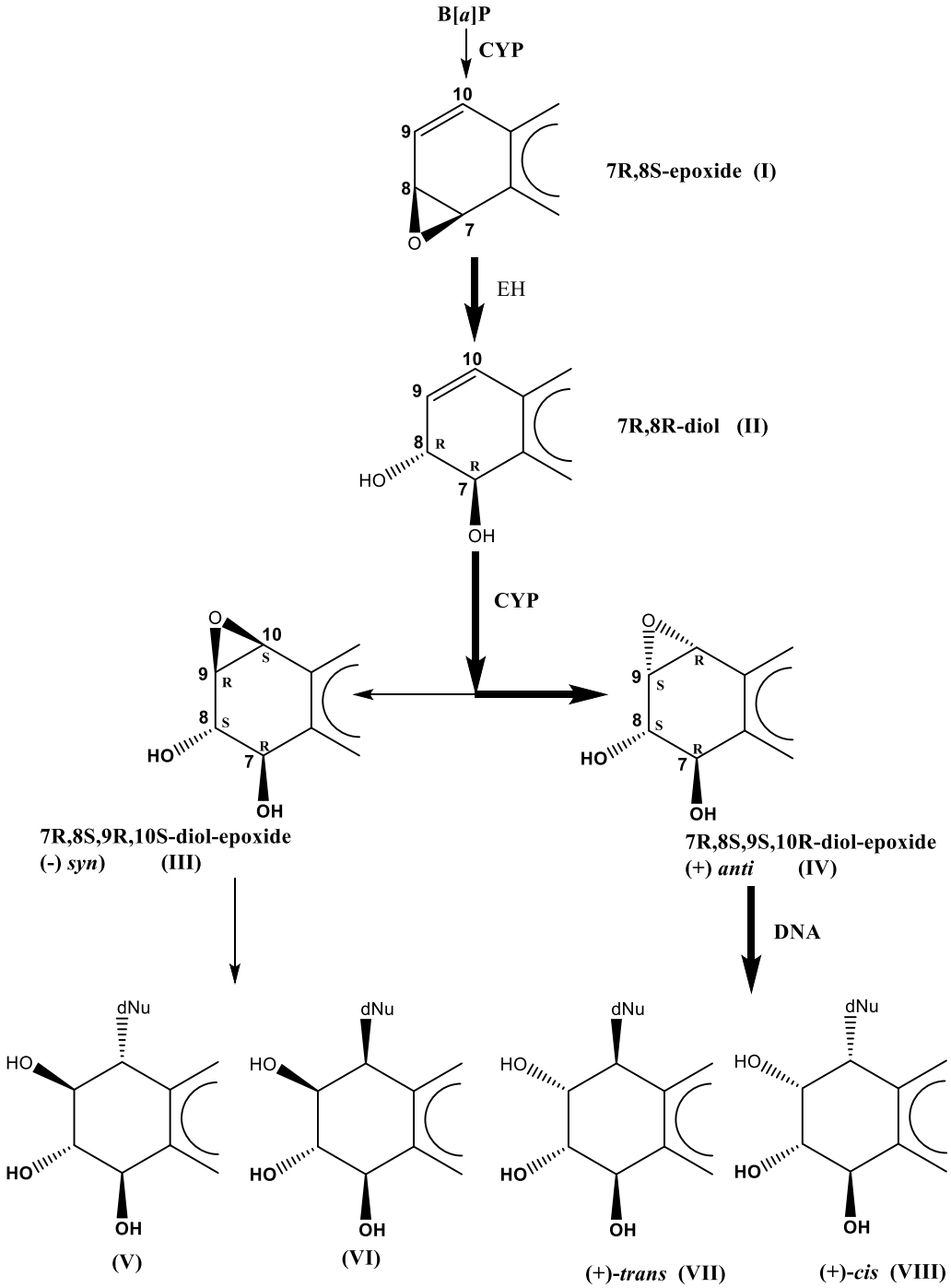


Figure 3. The reaction scheme shows the formation of major DNA adducts by the B[*a*]P metabolic activation pathway (highlighted arrows).

The nucleophilic addition reaction with BPDE and dG leads to a racemic mixture of enantiomers distinguished as *trans* and *cis* dG adducts [29]. It has been identified that (+)-*trans* (compound VII) and (+)-*cis* (compound VIII) DNA adducts are the most abundant isomers (Figure 3). It has also been shown that BPDE reacts with the exocyclic amine group (N^2) at C^6 of the deoxy adenine (dA) and at the and N^7 position of guanine to yield minor products. It is thought that these adducts have an insignificant impact on biology [28].

B[*a*]P adducts and their structural properties on DNA seem to play a critical role in numerous cellular events including DNA replication, lesion recognition and DNA repair [28]. The magnitude of the DNA damage and related biological consequences can be attributed to the conformational diversity of B[*a*]P adducts on DNA. Each isomer forms different conformations on DNA that leads to generate totally different mutagenic properties.

1.2.3 B[*a*]P isomers conformations in the DNA

Many solution NMR studies were done in the 1990's to understand the structural properties of the B[*a*]P adduct in DNA. Among other researchers, N. E. Geacintov, Suse Broyde and D. J. Patel were the pioneers in the field of B[*a*]P-DNA structure determination by solution NMR. They characterized many structures by placing B[*a*]P isomers in various places in the DNA, such as duplex DNA, primer-template junction and in front of an abasic site. Dramatically different B[*a*]P adduct conformation were found in different DNA construct for various B[*a*]P isomers. For instance, the (+)-trans-dG adduct in the duplex DNA has shown that the pyrenyl ring system in the minor groove of the DNA over the sugar group of the G18. The B[*a*]P ring system is pointing

towards the 5’ end of the DNA template. The Watson-Crick base pairing between the modified dG* and dC is not disrupted (Figure 4A) [30]. In the same duplex DNA, the (-)-trans-dG adduct orients 3’ end of the template while adduct resting on the minor groove over the sugar group of C17 (Figure 4B).

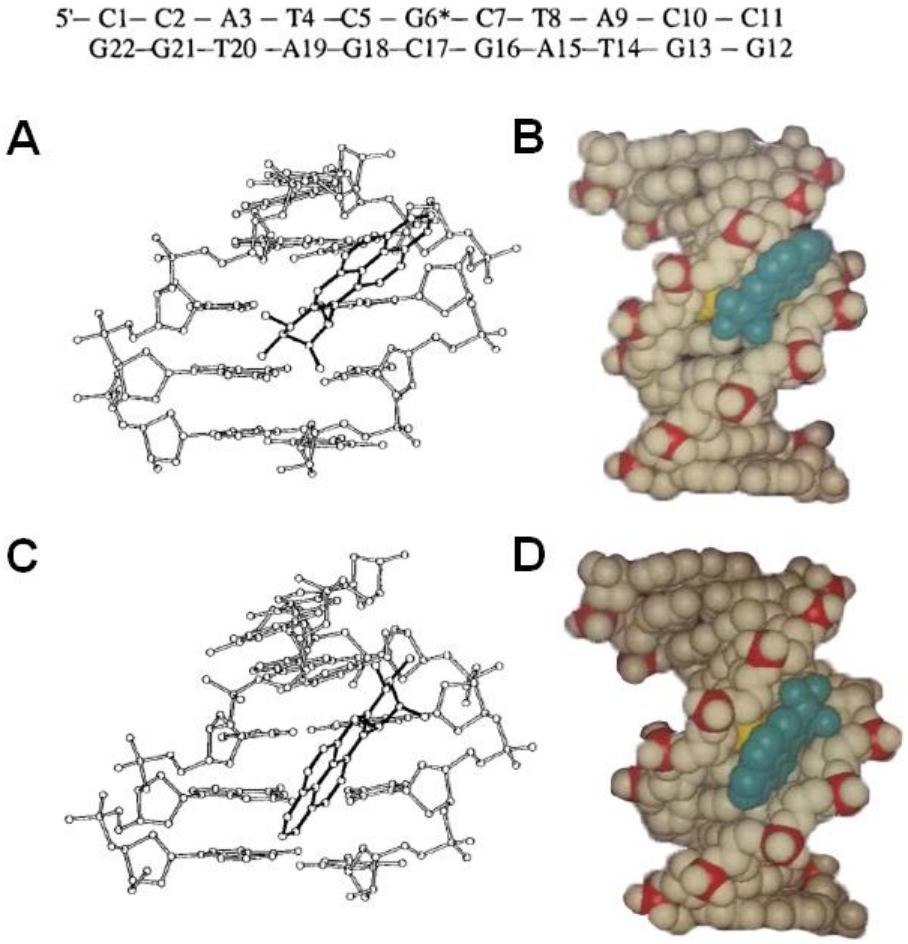


Figure 4. Solution NMR structures of the B[*a*]P adducts. (A) (+)-*trans*-B[*a*]P-N^2-dG adduct (B) The space filling model of the (+)-*trans*-B[*a*]P-N^2-dG adduct shows minor groove B[*a*]P in blue color. (C) (-)-*trans*-B[*a*]P-N^2-dG adduct (D) The space filling model of the (-)-*trans*-B[*a*]P-N^2-dG adduct. B[*a*]P adduct is shown in blue. And the modified G is in yellow. The figure is adapted with permission from [28], Appendix A1.

However, the corresponding cis adducts represent remarkably different conformations from those of the trans adduct. Their solution NMR structure show that the B[*a*]P ring system is intercalated inside the duplex DNA in both cis isomers [31]. The modified dG* is displaced within the minor groove. Furthermore, the normal dC:dG* base pairing is totally disrupted, and the dC is displaced to from the helix (Figure 5).

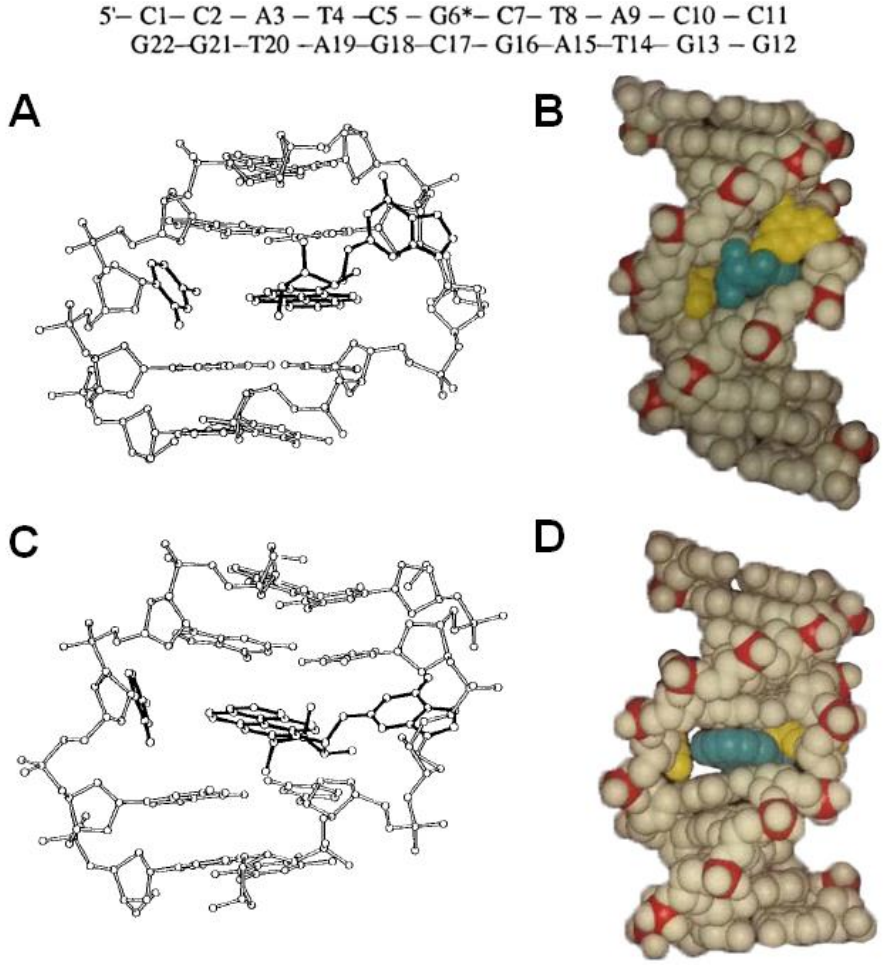


Figure 5. Solution NMR structures of the B[a]P adducts. (A) (+)-*cis*-B[*a*]P-N^2-dG adduct (B) The space filling model of the (+)-*cis*-B[*a*]P-N^2-dG adduct shows minor groove B[*a*]P in blue color. (C) (-)-*cis*-B[*a*]P-N^2-dG adduct (D) The space filling model of the (-)-*cis*-B[*a*]P-N^2-dG adduct. B[*a*]P adduct is shown in blue. And the modified G is in yellow. The figure is adapted with permission from [28] Appendix A1.

The bulky DNA adducts are known to deform the DNA structure or around the damage site. These distorted sites on the DNA tend to form misaligned and

frameshifted intermediates in the DNA replication. which causes base deletions to generate (deletion duplex, Figure 6 sequence) [28]. The B[*a*]P adduct conformations in the deletion duplex are significantly different from the full-duplex. The solution NMR structures of both sterioisomeric *trans* adducts have shown the same modified dG* displaced intercalated conformation (Figure 6).

In contrast, the (+)-*cis*-B[*a*]P-N^2-dG adduct has a nearly identical conformation in both the duplex and deletion duplex as shown in Figure 6. Interestingly, both the (+)-*trans*-B[*a*]P-N^2-dG and (+)-*cis*-B[*a*]P-N^2-dG isomers in the deletion duplex share similar structural features except for the position of the benzylic ring [32].

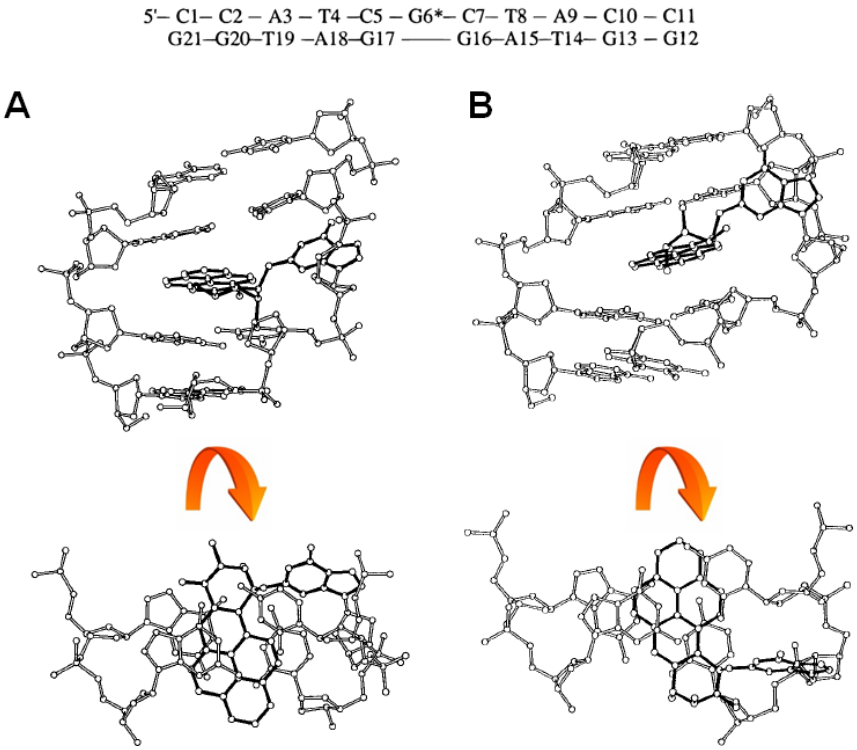


Figure 6. Solution NMR structures of the B[*a*]P adducts. (A) (+)-*trans*-B[*a*]P-N^2-dG adduct. The top-down view is shown below (B) (+)-*cis*-B[*a*]P-N^2-dG adduct. The top-down view is shown below. The figure is adapted with permission from [28], Appendix A1.

DNA primer-template junctions are a very important system in DNA replication. The polymerases add dNTPs to the 3' end of the primer, complementary to the

templating base. The smooth replication machinery is blocked or disrupted if the templating base is modified by a DNA adduct. Therefore, the B[*a*]P adduct conformations at the primer-template junction are important to understand its effects on the primer extension. The solution NMR structure of the (+)-*trans*-dG at the primer-template junction shows a stacking conformation with the terminal base pair in the duplex region. This stacking conformation does not disturb the base pairing in the duplex region (Figure 7) [28]. However, they were not be able to fully resolve the structure for the (-)-*trans*-dG at the primer-template junction due to the unstructured nature at the damage site. In contrast to the (+)-*trans*-dG adduct, the (-)-*trans*-dG adduct may disrupt the base pairing in the duplex region around the primer-template junction [28].

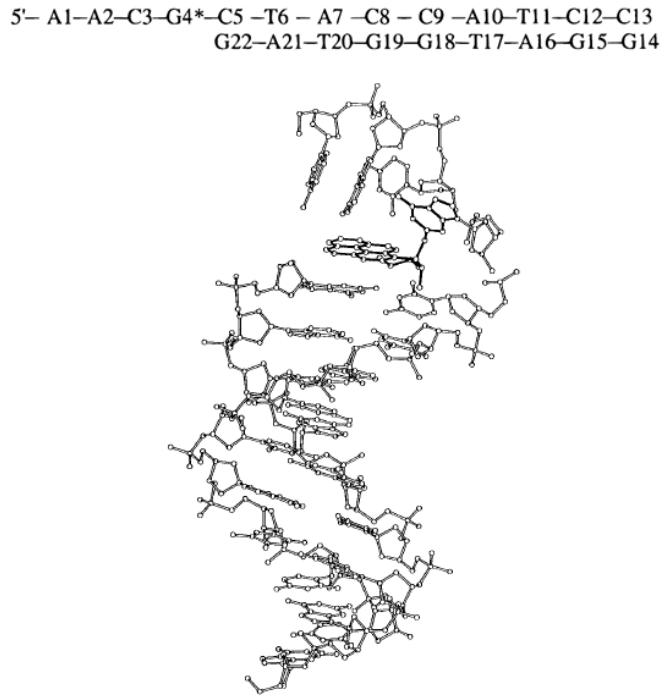


Figure 7. Solution NMR structures of the (+)-trans-B[a]P-N^2-dG adduct at the primer template junction. The figure is adapted with permission from [28], Appendix A1.

1.3 Y-family DNA polymerase

Replication of whole genomic DNA is an essential process that must be completed before cell division. To achieve this goal, replication of billions of nucleotides regardless of the DNA sequence is governed by high fidelity polymerases. This is a fast, highly accurate process only specialized high-fidelity polymerases can accomplish with an extremely low error rate. However, these polymerases are vulnerable to the presence of damage in the DNA which can cause them to stall at the damage site [33], which may lead to collapse of the replication fork. The genomic DNA is often exposed to various exogenous and endogenous DNA damaging agents that are frequently “road blocks” for replicative polymerases [17, 34]. This damage includes DNA strand breaks, base hydrolysis, base modification, and the deamination and alkylation of bases. The majority of these DNA lesions are repaired by several DNA-repair pathways. However, some lesions cannot be repaired or escape from the repair machinery [35]. Two alternative routes have been discovered to overcome these replication-barriers and to maintain genomic integrity in every organism [33, 36, 37]. One is homologous recombination (HR), which includes a network of interrelated pathways to repair double stand DNA (dsDNA) breaks and inter strand crosslinks. Furthermore, HR provides greater support in recovering a stalled or broken replication fork [38]. The second pathway is to use specialized DNA polymerases to replicate across from and past a DNA adduct [33]. The DNA damage site may have structural and chemical properties that are bulky, unrepairable and unfavorable for proper Watson−Crick (WC) base pairing and base stacking. In that case, these specialized polymerases are able to synthesize past the adduct because their active site is able to tolerate the DNA damage

[33]. This damage tolerance DNA replication is called translesion synthesis (TLS), and most specialized polymerases engaged in TLS are part of the Y-family of polymerases [33-37].

All the Y-family polymerases that have been discovered so far are classified into six major groups. These group are represented by *E. coli* pol IV, V human pol η, ι, κ and Rev1. All the Y-family polymerases contain two major parts: a catalytic region and a regulatory region [33]. The catalytic region consists of four subdomains, which include the finger, palm, thumb and little finger (LF). Both replicative and TLS polymerases contain finger, palm and thumb subdomains, but relatively smaller thumb and finger subdomains are found in Y-family polymerases, which allows the polymerase to accommodate a bulky DNA adduct. This also causes the DNA substrate to be more solvent-exposed at the active site of the polymerase. The finger subdomain of replicative polymerases undergoes a large open to closed conformational change when a nucleotide binds to the DNA-polymerase complex. Y-family polymerases are thought to not undergo these large structural changes but may undergo smaller localized rearrangements [39].

The LF domain is found only in Y-family polymerases, and it has the most diverse DNA sequence among the Y-family members [33]. The thumb and LF subdomains hold the DNA substrate. While palm subdomain performs the catalysis, this catalytic region contains the important three carboxylates, Asp 7, Asp 105, Glu 106. These three amino acids are highly conserved in all the Y-family members, and they coordinate with two Mg^{2+} ions during catalysis. Also, it has been discovered that some Y-family polymerases display a pre-formed active site, where the fingers are closed in

the absence of either DNA or dNTP [40, 41]. The most important feature of Y-family polymerases is the large solvent-accessible active site which can accommodate bulky DNA adducts. This spacious active site allows dNTPs and Mg^{2+} ions to move in and pyrophosphate to come out freely without larger conformational changes of the subdomains of the polymerase [33].

All high-fidelity polymerases are adapted to replicate normal DNA with uniform WC base pairs. Therefore, cells do not need many different types of replicative polymerases. However, there are many different types of DNA lesions that can form upon exposure to a damaging agent, and these require many specialized polymerases to accomplish TLS. *E. coli* contains five DNA polymerases, but only one of them (Pol III) is a replicative polymerase while three are Y-family polymerases. For comparison, five of 17 DNA polymerases are replicative polymerases in humans. Although Y-family polymerases are essential for TLS, they have a high error rate when coping with DNA presumably because of the looser active site and missing proofreading activity that is present in replicative polymerase complexes [33].

1.3.1 Y-Family polymerases in action- TLS

Y-family polymerases have an ability to tolerate a variety of structurally different and potentially mutagenic DNA lesions during DNA synthesis. Despite their overall fidelity, carcinogenic adduct bypass ability, and structural properties, Y-family polymerase have a higher degree of specificity for particular type/types of lesions. Here, I summarize the potential diversity and carcinogenic specificity of different Y-family polymerases in TLS. One of the remarkable discoveries about Y-family polymerases is the linkage of the pol η with the xeroderma pigmentosum (XP) [42, 43]. XP is a rare,

genetic disorder, which makes the Individuals extremely sensitive to ultra violet (UV) light. The UV light exposure can cause dermatological cancer by creating cyclobutene pyrimidine dimers (CPDs) in DNA. The pol η can replicate CPDs containing DNA with a higher accuracy and processivity [44] yet, patients with XP were identified to carry non-functional pol η. When the mutations in the gene of pol η (*POLH)* in humans produce the inactive polymerase, other Y-family polymerases might be recruited to replicate CPD DNA. However, they perform error prone replication of CPDs by making mutations, which eventually leads to the development of skin cancer. In addition to CPDs, pol η can bypass cisplatin adducts which are widely used as chemotherapeutic drugs [45].

Apurinic and Apyrimidinic sites are other common DNA lesion types which are generated by lost DNA bases due to various reasons. It is estimated that more than 10^3 AP sites are created in a single mammalian cell per day [46, 47]. The DNA replication by high fidelity polymerases is impeded by AP sites and many studies show that the replicative polymerases cannot efficiently bypass AP sites. Therefore, Y-family polymerases usually process AP sites in the cells, and tend to add dATP across the AP site which is often referred to as the “A”-rule [48]. Studies show that both Y-family pol η and pol κ can successfully bypass AP sites in an error-prone manner [47, 49].

The pol κ deleted mouse embryonic cells have shown a higher sensitivity for B[*a*]P adducts but are not sensitive to UV light, indicating pol κ is not involved in bypassing the CPDs [50, 51]. It has been shown that the human pol κ is the only Y-family polymerase which has the ability to bypass the B[a]P adduct error free [52]. The pol κ gene expression is controlled by the aryl-hydro carbon receptor and therefore the

cells exposed to the PAHs, such as B[a]P, activate the pol κ expression, which prevents PAH-induced mutagenesis [50]. Various DNA adducts can upregulate pol κ expression in the P53 dependent pathway. If P53 is non-functional, there are alternative pathways to activate pol κ in response to DNA damaging agents. Therefore, pol κ is very important in TLS in mammals because it can be activated by both P53 dependent and independent pathways [53, 54].

Pol ι is a unique TLS polymerase, and its other homologs have been found only in higher eukaryotes, such as drosophila, mice and human [35, 55]. Pol ι shows tissue specific expression in mammals, and is highly expressed in human and mice testes, probably indicating a role in spermatogenesis [56]. Pol ι is highly error prone in replicating DNA, and it prefers to incorporate purines rather than pyrimidines in both modified and unmodified DNA [57]. Pol ι has the ability to process non-structured and distorted DNA lesions by incorporating nucleotides using Hoogsteen base pairing [35, 58]. Though it can incorporate nucleotides across different types of DNA lesions such as CPDs, 8-oxo-dG and PAHs modified DNA, the extension after the lesion is strongly inhibited [35, 59, 60]. Therefore, translesion synthesis of pol ι must be accompanied by another Y-family polymerase. Interactions of the Pol ι were observed with the REV1 polymerase in the process of co-localization with the DNA replication machinery in UV-induced cells [61]. Another interesting finding about pol ι is that both the pol $\eta^{-/-}$, pol $\iota^{-/-}$ cells have shown extreme UV sensitivity compared to only pol $\eta^{-/-}$ cells. However, significantly higher-level of mutations that can cause XP appeared in pol $\eta^{-/-}$ cells, suggesting that pol ι replicates CPDs in the absence of pol η, but generating many replication errors [35, 62, 63]. There is also strong evidence for the linkage between pol

ι and carcinogenesis since the over-expression of pol ι has been observed in several breast cancer cell lines [64].

1.3.2 DNA polymerase IV (Dpo4)

Dpo4 is isolated from the thermophilic archaeon bacterium, *Sulfolobus solfataricus*, which shows in hot springs at 80^0C. Although most organisms have several Y-family polymerases, *Sulfolobus solfataricus* has only one called DNA polymerase IV (Dpo4). Therefore, it should have the ability to process a broad range of DNA lesions and because of this Dpo4 is one of the most studied Y-family polymerases. In addition, its extraordinary thermal stability allows for easy purification and handling.

More than 100 crystal structures have been solved for apo-Dpo4 and other Dpo4-DNA structures such as binary complexes and ternary complexes, demonstrating significantly different domain arrangements between the apo-Dpo4 structure and the Dpo4-DNA structures [65]. In the binary and ternary complexes, the little finger domain wraps around the DNA, causing the little finger and finger domain sit closer to each other. However, in the apo-Dpo4 structure the little finger domain is stacked with the thumb domain (Figure 8A) indicating the it has the flexibility to move and rotate independently. This mobility of this domain is supported by the linker, which tethers the LF to the core domain. Dpo4 is in the identical conformation in both binary and ternary crystal structures (Figure 8D) [65] and it is believed that the Dpo4 global conformation remains constant in both binary and ternary structures, but the detailed analysis of the amino acid residues shows some local conformational changes. For instance, Arg51, which stays out of the active site in the apo-Dpo4 moves into the active site in the binary complex, providing electrostatic stabilization to the bound DNA. The Dpo4 nucleotide

binding site is occupied by Tyr10 in the apo and binary structures but It moves away to provide a space for the incoming dNTP in the ternary complex. [65].

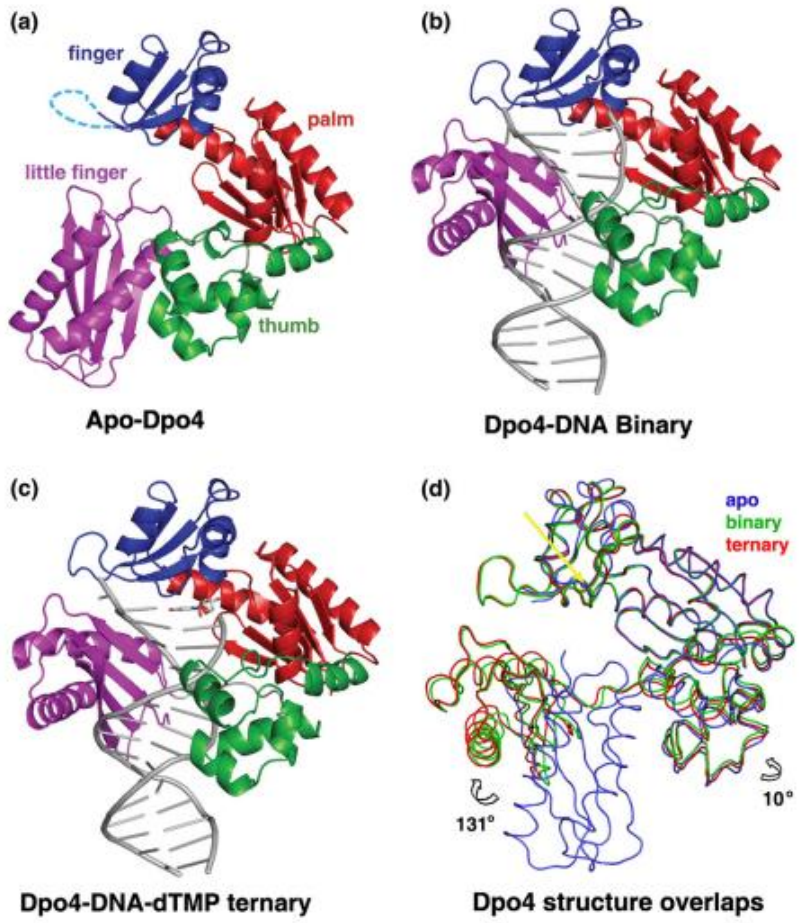


Figure 8. The Dpo4 crystal structures. **(A)** the color coded Dpo4 domains are shown in the apo-Dpo4. **(B)** The Dpo4-DNA binary complex structure **(C)** The crystal structure of the Dpo4-DNA-dNTP ternary complex. **(D)** The overlay of the apo, binary and ternary Dpo4 conformations. The figure is adapted with permission from [65], Appendix A2.

Dpo4 active site draws major attention compared to all the other interesting properties. Consistent with the other polymerases, the Dpo4 active site also contains three highly conserved carboxylates surrounded by the palm and finger domains (Figure 9) [66]. In 2001 the Wei Yang group produced two types of Dpo4 crystal structures. In the first crystal structure, the templating base is the dT that is stabilized by the V32, A42 and G58 whereas the ddATP is the correct incoming nucleotide (Figure 9A) which is associated with A44, A57 and T45, R51 by van der Waals interactions and H-bonding respectively [66]. The sugar ring and phosphate groups are stabilized by the Y12, D105, K159, D7 and the metal ion while tA4 in the template stays outside the helix. Even though the templating base and the incoming nucleotide are properly aligned for the base pairing for nucleotidal reaction to occur, the space between the LF and fingers domain permits the active site exposed to the solvent, creating it less protected. In the second crystal form, the templating base and the incoming dNTP are dG. Surprisingly, Dpo4 has translocated to the tC3 bypassing the tG4. The ddGTP, pT12 and tT2 have shown significant directional changes to support this skipped-template structure [66].

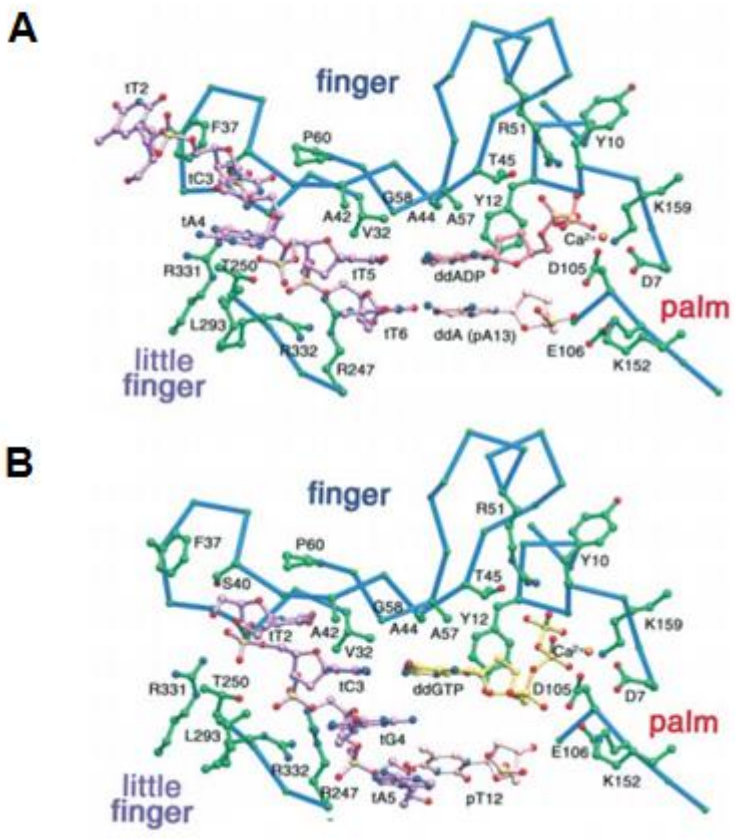


Figure 9. Dpo4 active site structure in the presence of cognate and non-cognate nucleotides. **(A)** Dpo4 active site conformation in the presence of the next correct nucleotide (dTTP). **(B)** The Dpo4 active site interactions with a wrong nucleotide (dGTP). The figure is adapted with permission from [66]. Appendix A3.

1.4 Single-molecule FRET

Conventional biochemical experiments provide ensemble average results because all molecules are considered to be a single homogenous population. However, this is not the case in reality. Complex biological reactions cannot be simply averaged or synchronized because often they are multi-step and several subpopulations can be present at a given time. The best way to cope with this problem is to monitor the dynamics of a large population of individual molecules at specific time points over the course of a reaction [67]. Use of optical methods to detect and study biological

molecules was first developed by the Hirschfeld lab in the 1970s [68]. There have been numerous advancements in the technology of single-molecule detection, and this method now has the tools to study the structural, functional and dynamic features of complicated biological reactions at the single-molecule level, which provides reaction details that cannot be seen in ensemble average experiments [67-70].

Early single-molecule experiments used a single dye (fluorophore), which allowed the localization and quantification of the labeled macromolecules. These experiments were taken to the next level by using two or three fluorophores, which gives much more power to study macromolecular interactions, translocation and movements. The fluorescence intensity fluctuation, polarization and fluorescence (Förster) resonance energy transfer (FRET) are some of the parameters measured in single molecule spectroscopy [71, 72]. In physical theory, ground state (S_0) electrons absorb photons to different vibrational exited state sates (S_1) (Figure 10A) [73]. Some of the electrons lose their energy and fall to the lowest vibrational energy level of the exited state due to vibrational relaxation. The energetically unstable electrons in the exited state relax back to the ground state by emitting excess energy as photons. The excitation and relaxation may not occur between the same energy levels. Usually less energy is emitted during relaxation with a longer wavelength. This process is called fluorescence (Figure 10A). The emission energy from the relaxing molecule (donor) can be transferred to another chromophore (acceptor) if it is in close proximity, causing non-radiatively by coupling of dipole-dipoles (Figure 10B). Therefore, FRET is a non-radiative electromagnetic phenomenon where the quantum energy transfer between two fluorophores is measured [73]. The fundamental requirement for FRET is the overlap of

the emission spectrum of the donor and the excitation spectrum of the acceptor. FRET efficiency depends on the distance between the donor and acceptor fluorophores (Figure 10C). The FRET efficiency can be calculated by the Förster equation, $E = (1 + (R / R_0)^6)^{-1}$ where R_0 is the radius when E=0.5 and R is the inter fluorophore distance. The distance dependency of FRET is very useful as a molecular ruler to estimate the intermolecular distances within the range of 20-70 Å (Figure 10C).

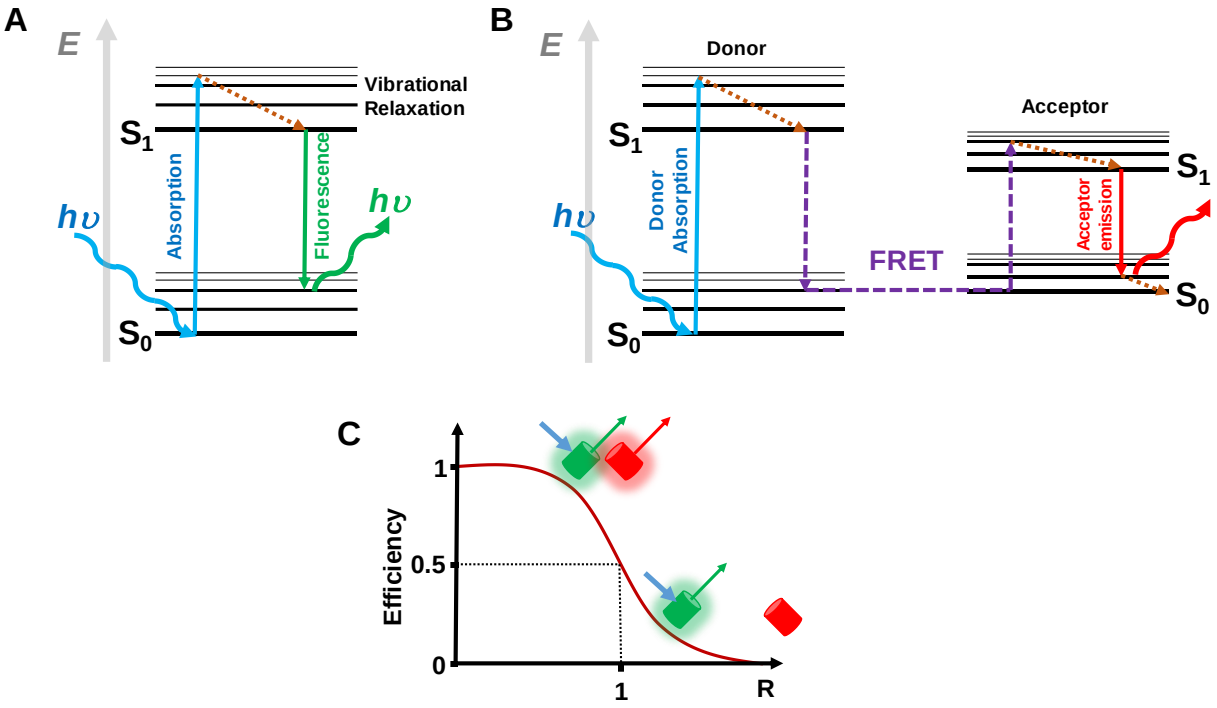


Figure 10. The fundamentals of the Florence and FRET. The fundamentals of the Florence and FRET. (A) the theory behind the fluorescence process. (B) The nonradiative fluorescence energy transfer between the donor and acceptor. (C) The distance dependency of the FRET phenomena. The figure is adapted from [73].

The apparent FRET efficiency (E_{FRET}) can be estimated by measuring the ratio of the emission intensities of the donor and acceptor fluorophores [$E_{FRET} = (I_A/I_A + I_D)$]. The change in FRET can be directly related to the real time conformational dynamics of individual macromolecules [74]. Therefore, the use of the smFRET technique to answer fundamental biological questions has exploded since its discovery. Currently it has been

used in various systems to study macromolecular interactions and dynamics, such as protein-protein, protein-DNA, protein-RNA, RNA folding, protein folding, mortar proteins, and ion-channels, and its uses continue to expand in many applications [74, 75].

1.4.1 Use of smFRET in studying DNA polymerase mechanism

DNA polymerase mechanism involves several steps with different reactions rates. Over the past few decades the polymerase mechanism has been extensively elaborated with conventional biochemical techniques and structural studies. The advancement of crystallographic techniques has generated many thousands of structures of various polymerases allowing us to visualize the different stages of the polymerase mechanism [76]. However, these structures generate a snapshot of a multistep, dynamic reaction. Over the last decade, various single-molecule techniques have been used to probe the DNA replication process. These techniques have a huge advantage over the conventional methods because the transition intermediates, population heterogeneity and their binding and dissociation kinetics can be easily measured using single-molecule methods. Furthermore, these techniques have the freedom to change the positions of the fluorophores on the polymerase and also to study the dynamic changes in real time.

Several FRET-based single-molecule approaches have illuminated knowledge about polymerase conformational changes in the DNA replication mechanism. One of the early studies from our lab with *E. coli* DNA polymerase I (Klenow fragment, KF) was able to study the DNA replication mechanism in a single-base pair resolution [77]. They used a FRET system with Cy3 label on DNA and Cy5 labeled on KF. The observed that FRET states for the KF-DNA interactions clearly represented the stages of the known

polymerase mechanism. One of the observed results is that KF binding to the mismatch DNA causes the polymerase to change the conformation to the exonuclease site. They observed a major conformational change, which was not known before, after the nucleotide incorporation; presumably it was assigned as a proofreading step [77]. The observed KF binding conformations in that study are summarized in the Figure 11.

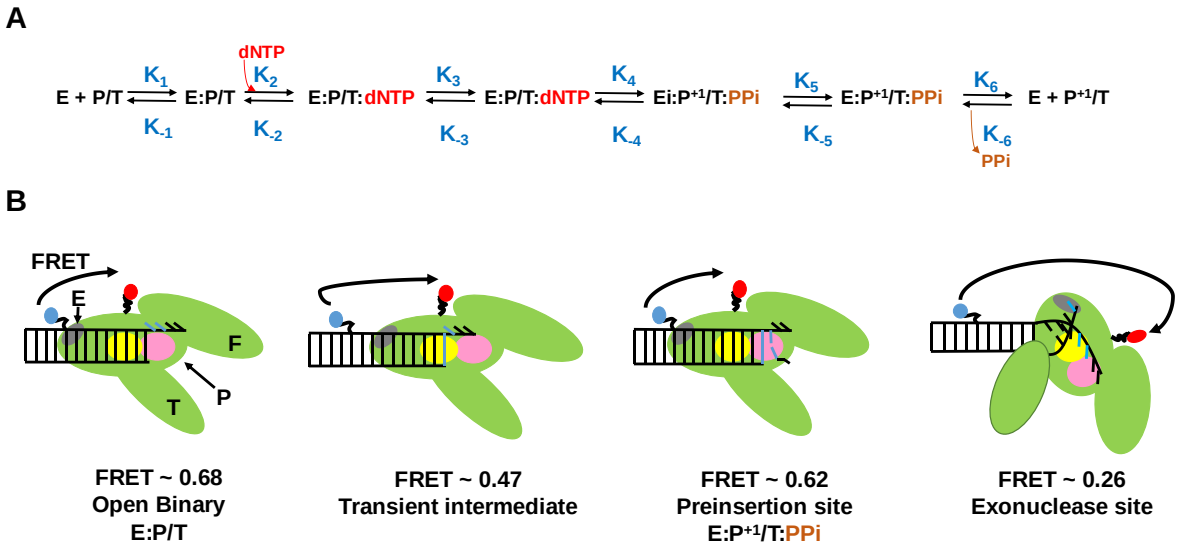


Figure 11. KF binding conformations to the DNA. **(A)** KF's kinetic mechanism during DNA synthesis. **(B)** The KF (green) binding conformations to the DNA. The domains of the KF are labeled as the corresponding letters (Thumb-T, fingers-F, Palm-P). The exonuclease domain is marled in orange. The active site (purple) and preinsertion site (yellow) are located on the palm domain. First, KF binding to the DNA forms an open binary complex (FRET ~0.68). KF translocases to the transient intermediate site after incorporating nucleotides (FRET~ 0.47). Then polymerase moves backward and positions at the preinsertion site (FRET~0.62). The correct dNTP selection occurs after the preinsertion site. If there is a mismatch at the primer template junction, the KF, translocates to the exonuclease site (FRET 0.26). The figure was adapted from [77].

The KF nucleotide selectivity was thoroughly studied using smFRET and smPIFE in our lab [78]. The KF binding to DNA is stabilized in the presence of the correct nucleotide and dNTP selection appeared to occur before forming the closed ternary complex. Because more noncognate nucleotides are present than correct nucleotides

for any incorporation step. Therefore, incorrect nucleotide rejection might occur in a step that prevents KF dissociation [78].

The KF finger closing mechanism was studied by *Santoso et al.* in 2010 [79] in a study that placed both fluorophores on the KF, with one on the finger domain and one on the thumb domain. The change in FRET was directly proportional to the finger closing and opening moments. The high FRET and low FRET efficiencies were identified as finger closing and opening, respectively. Their results indicate that the finger closing might not be represented by a successful dNTP incorporation; instead, other mechanisms such as fidelity checking can occur before the dNTP incorporation. An incomplete finger closing was observed in the presence of a mismatch in the primer template junction which supports their initial hypothesis [79].

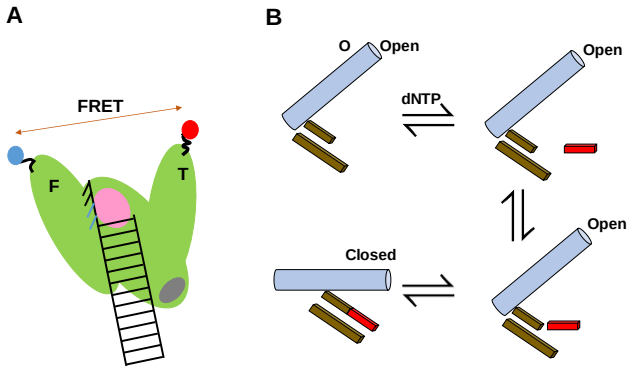


Figure 12. The conformational changes of the finger domain during DNA binding and nucleotide selection. **(A)**The positions of the fluorescent labels on finger (F) and thumb (T) domains. **(B)** The proposed mechanism for the KF pre-chemistry steps. The KF-DNA binary complex, the O-helix (O) in the open conformation. KF forms an initial DNA-KF-dNTP complex in the presence of nucleotide which followed by a fidelity checking step. If the correct dNTP is selected, the finger domain closes to form the stable ternary complex. The figure was adapted from [79].

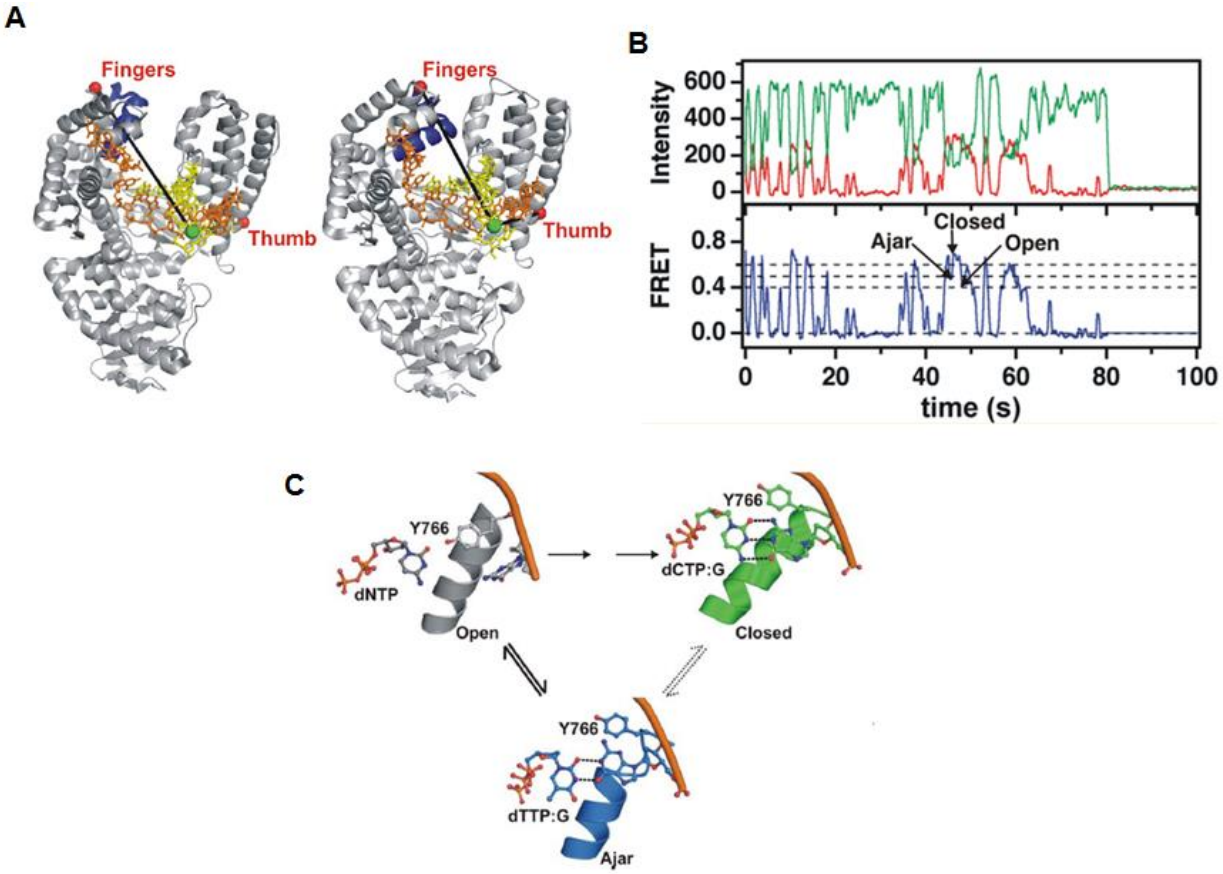


Figure 13. Dpo4 conformational changes during DNA binding and nucleotide incorporation. **(A)** The positions of the fluorophores on Dpo4. The two structures show the change of positions of them during finger movements. **(B)** The characteristic intensity and FRET traces with captured Dpo4 conformational changes. **(C)** The proposed reaction pathway of dNTP incorporation at the KF active site. KF is in the open conformation of the DNA-KF binary complex. Upon KF binding to the correct nucleotide, the open conformation goes to the closed conformation quickly. The open conformation is in equilibrium with ajar which check the nucleotide binding to the active site. The incorrect nucleotide causes for rapid conformational change form ajar to open. The figure is adapted with permission [80], Appendix A4.

This intermediate, fidelity checking conformation observed in the crystal structure in 2011 [81] and, David Miller's group apparently observed this also fidelity checking step using smFRET [80].

1.4.2 Adaptation of smFRET technique to study Dpo4

Since its discovery hundreds of biochemical and structural studies have been carried out on Dpo4 to understand its properties, function and role in TLS, but no studies have determined the Dpo4 binding conformations and related binding kinetics in real-time using single molecule techniques. In 2013, our lab adapted the smFRET technique to characterize the real-time Dpo4 binding conformations and related binding kinetics [82].

We used the Cy3/Cy5 fluorophore system to label the DNA and the Dpo4 respectively as was used with KF. However, in contrast to the previous single molecule experiments with KF, the DNA-Dpo4 binary complex was formed to sample two FRET states that these were not static conformations as appear to correspond to two Dpo4 binding conformations. Interestingly, transitions between these two FRET states were observed. The addition of dNTPs resulted in only a single FRET distribution (low FRET state, Figure 14). The corresponding Dpo4 binding kinetics also reveal the formation of a relatively stable ternary complex in the presence of the next correct nucleotide. The observed FRET states were in good agreement with previously published crystal structure information of the Dpo4 [82]. The new findings of the Dpo4 binding conformation in the binary complex and ternary complex are summarized in the Figure 14.

In 2015, the single molecule method developed to study conformational dynamics of Dpo4 binding to normal DNA was taken into the next level to study carcinogenic adduct modified DNA. When we studied the conformational changes that occur during the bypass of AF and AAF adducts in DNA [83]. In contrast to the unmodified DNA, the smFRET results revealed that the binding conformation of the binary complexes is different in AF/AAF modified DNA indicating an altered structure at the adduct site. However, in the presence of dNTPs, Dpo4 forms a ternary complex conformation similar to the unmodified DNA. Furthermore, that study reveals different dNTP disincorporation pathways for AF and AAF modified DNA [83].

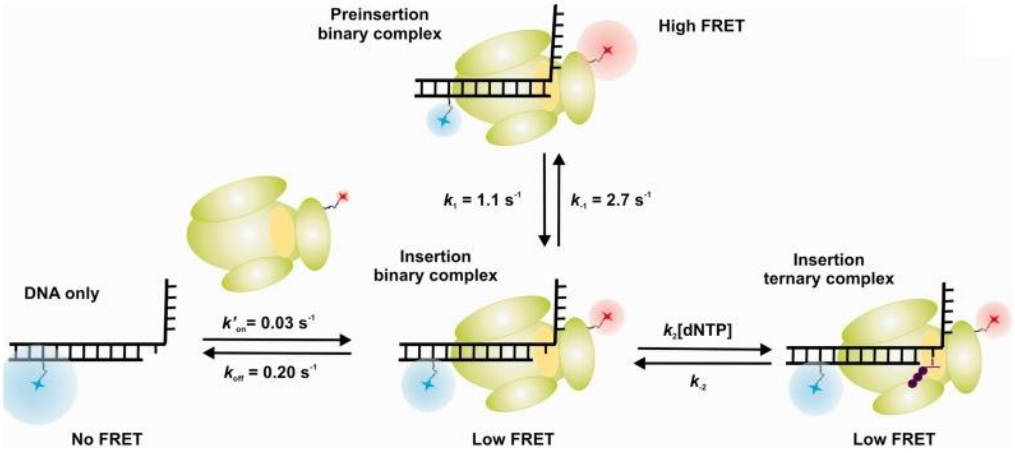


Figure 14. The proposed mechanistic pathway of Dpo4 binding to unmodified DNA. The figure was adapted with permission form [82], appendix A5.

CHAPTER 2: EXPERIMENTAL PROCEDURES

Some sections of this chapter were taken from the manuscript, *Liyanage et al.*

2 Materials and methods

2.1 Materials

2.1.1 Benzo[a]pyrene

A racemic mixture of activated B[*a*]P ((±)-*anti*-B[*a*]PDE) was perched from National Cancer Institute, Chemical Reference Standard Repository, Kansas City, MO.

2.1.2 DNA oligonucleotides

Custom synthesized DNA oligonucleotides including Cy5-DNA, Biotin-DNA and amino modified C6-dT were purchased from either Eurofins genomics (Louisville, KY 40299) or Integrated DNA Technologies (Coralville, Iowa). All the DND sequences are shown in Table 1.

2.1.3 MALDI-TOF MS

ZipTip pipette tips were purchased form Merk Millipore Ltd. Tullagreen. The 3HPA used for matrix preparation was bought form Sigma Aldrich.

2.1.4 DNA polymerase IV

Archaeon bacterium *Sulfolobus solfotericus* DNA polymerase IV gene carrying plasmid (p1914) transformed *E. coli* strain (RW 382) was generously gifted by Dr. Roger Woodgate (NICHD).

Table 1. The DNA sequences used for all the primer extension assays and single *molecule experiments*

Experiment	Name	Sequence
To make B[*a*]P template	11 mer	**5' CCATAGATACC 3'**
	15 mer	**5' CpATGTCGTTTTGGTG 3 '**
	26 mer scaffold	**5' CACCAAAACGACATGGGTATCTATGG 3'**
Single molecule FRET	20 mer primer	**5' [Biotin]-CACCAAAACGACATGGGTAT 3'**
	21 mer primer	**5' [Biotin]-CACCAAAACGACATGGGTATC 3'**
	21 mer primer mismatch	**5' [Biotin]-CACCAAAACGACATGGGTATT 3'**
	26 mer modified template	**5' CCATAGATACCCATGTCGTTTTGGTG 3'**
	26 mer unmodified template	**5' CCATAGATACCCATGTCGTTTTGGTG 3'**
Primer extension assays	16 mer primer	**5' [Cy5]-CACCAAAACGACATGG 3'**
	20 mer primer	**5' [Cy5]-CACCAAAACGACATGGGTAT 3'**
	21 mer primer	**5' [Cy5]-CACCAAAACGACATGGGTATC 3'**
	21 mer primer mismatch	**5' [Cy5]-CACCAAAACGACATGGGTATT 3'**

Cy3 is attached to blue underline T

Red G is the B[*a*]P modified base

2.1.5 Single molecule FRET

The quartz slides were purchased from TED PELLA, Inc. The amino propyl silane was purchased from UTC Inc. Bristo, PA for the quartz slide silanization reaction. The m-Peg and Biotin-Peg used for the pegilation reaction during single molecule slide preparation were purchased from Laysan Bio Inc. Arab, AL 35016. The NeutrAvidin used for immobilizing DNA was purchased from Life Technologies, Eugene, OR.

2.2 Methods

2.2.1 DNA purification

All the DNA oligonucleotides were HPLC purified (Virion pro Star) using a C18 analytical column. The HPLC buffer systems and gradients are listed in the Table 1. All the buffers were degassed before running the HPLC. The HPLC column was washed (100% Buffer B) and equilibrated (100% Buffer A) at least for 15 min prior to each injection. The purified DNA quantified and stored in -20^{o}C freezer.

Table 2. The HPLC buffers used for DNA purification.

Experiment	Buffer A	Buffer B	Gradient (Buffer B)
11 mer	0.1 M TEEA, 5% ACN	95% ACN	5-21% over 30 min
B[*a*]P –11 mer	0.1 M TEEA	65% CAN 35% ,0.1 M TEEA	0-65 over 80 min
Cy3-15mer	0.1 M TEEA, 5% ACN	95% ACN	5-30% over 40 min
Cy3-26mer	0.1 M TEEA, 5% ACN	95% ACN	5-30% over 40 min 30-100% over 50min
DHPLC	0.1 M TEEA, 5% ACN	65% ACN	10-28% over 60 min

2.2.2 DNA purity checking

The purified and B[a]P modified DNA quality was checked by MALDI-TOF MS. A 100 pmol of the DNA was mixed with 15 μL of 0.1 M TEEA to make the DNA solution. In the desalting process, first ZipTip was washed three times with 50% ACN. Then the tip was equilibrated with 0.1 M TEEA. The DNA sample was aspirated more than 20 times into the equilibrated ZipTip for maximum DNA binding to the column. Then the tip was washed three times with 0.1 M TEEA. Finally, the tip was washed three times by purified water. Meantime, 2 μL of 3HPA solution (about 50 mg of 3HPA was dissolved in 20 μL of 25 mM ammonium acetate) spotted on to the MALDI plate. When the 3HPA spot is completely dried, the DNA was eluted (elution solution 50% ACN) directly onto the spot. When the spot is completely dried, the plate was taken to the instrument

(Bruker) for analysis. The MALDI spectra for the (+Cis) and (+trans) modified 11 mer DNA are shown in Figure 15.

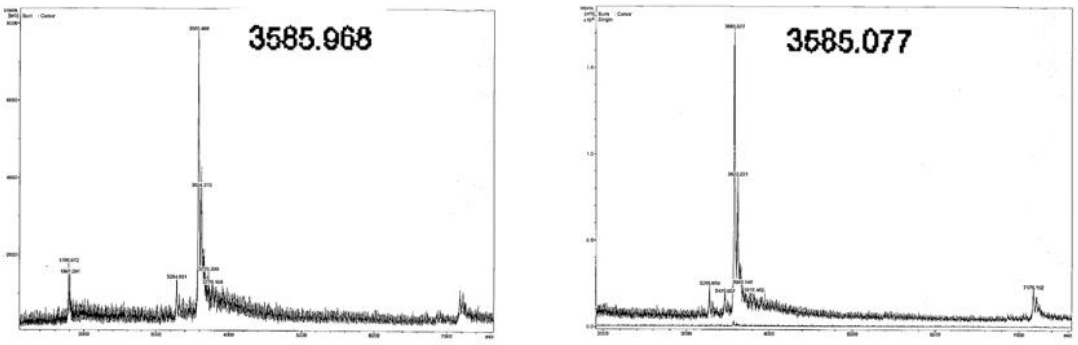


Figure 15. The MALDI spectra for the (+)-cis-B[a]P-11mer (left) and (+)-trans-B[a]P-11mer (right).

2.2.3 DNA Cy3 labeling for smFRET experiments

To make the Cy3 labelled oligonucleotides the amino modified C6-dT containing DNA was reacted with Cy3 mono reactive dye (GE Healthcare Bio-Sciences). The reaction scheme is shown in Figure 16A. The Cy3 dye was dissolved in 40 μL DMF and stored under argon gas in -20^{o}C. Usually 5-10 nM DNA was reacted with 15 μL of Cy3 in a solution contain Dioxane (10 μL), DMF (10 μL) DIPEA (3 μL) For ~1 hr at room temperature. The reaction was quenched by adding 10 μL of DDT). The Cy3 labeled DNA was purified from the unlabeled DNA using HPLC (the buffer are shown in Table 1, the HPLC chromatogram is in Figure 16B). The 26 mer oligonucleotide was Cy3 labeled to make the unmodified template for smFRET experiments (Amino modified dT is at 14th poison on the template). However, The Cy3 labeled B[*a*]P modified 26 mer template was made differently.

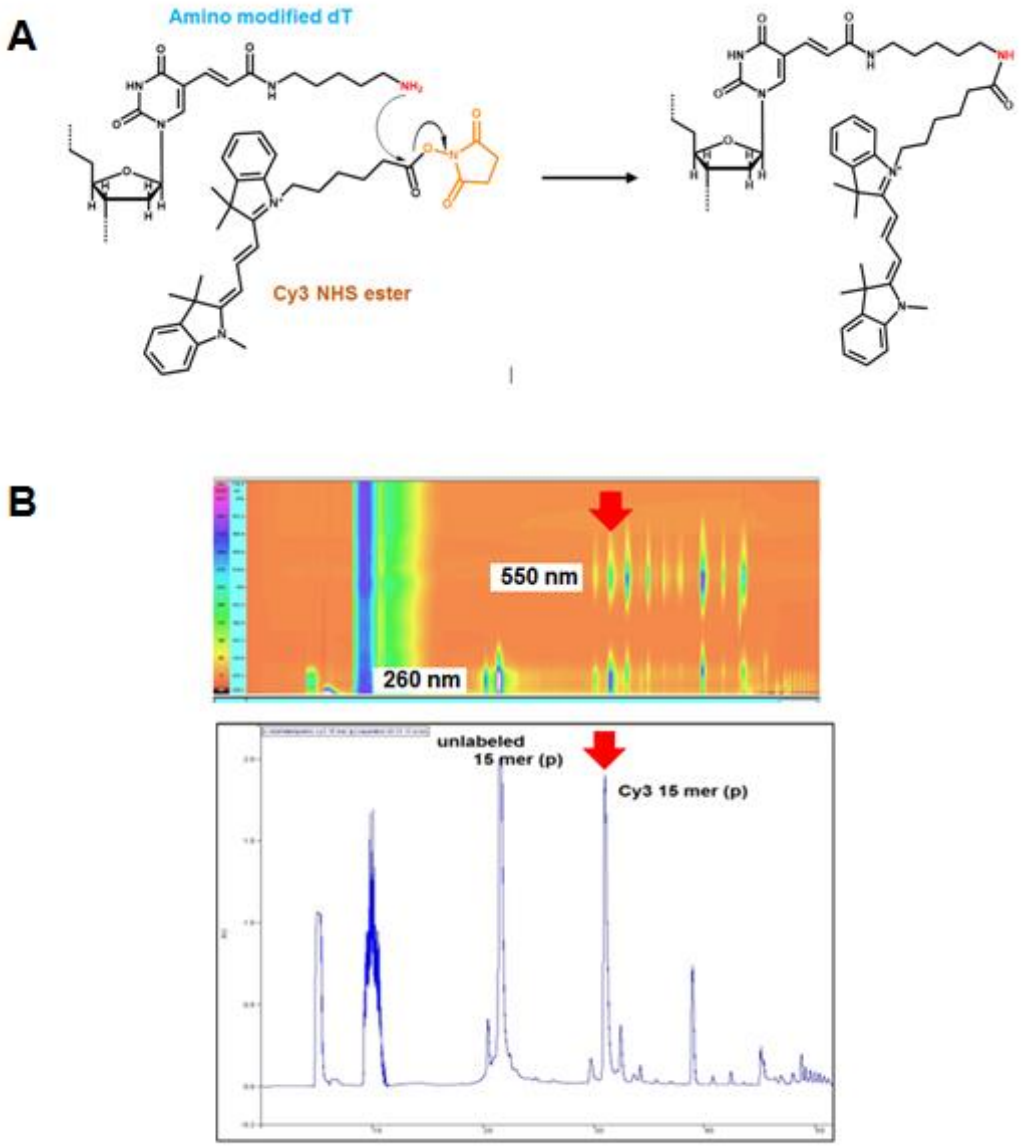


Figure 16. Cy3 labelled DNA HPLC purification. **(A)** The Cy3 labelling reaction. **(B)** The HPLC chromatogram for Cy3 labeled DNA purification. The Aurora spectrum on the top shows the 260 nm and 550 nm absorption. The red line indicates the Cy3 labelled DNA peak.

2.2.4 Preparation of B[*a*]P modified DNA template.

The BPDE was dissolved in a solution containing 19:1 THF:TEA extremely carefully inside the hood. A continuous stream of argon gas was flowed over the tube to minimize the contact with oxygen. B[*a*]P-modified template was prepared as described [84] (Figure 17). Briefly, a racemic mixture of (±)-*anti*-B[*a*]PDE was incubated with an 11-mer containing a single dG. The reaction yields four isomeric products that were separated by reverse phase HPLC (Figure 18). The selected B[*a*]P-modified 11-mer was ligated to a Cy3-labeled 15-mer using T4 DNA ligase and the final product (26-mer) was purified by reverse phase HPLC with a heated column (DHPLC). The 26-mer unmodified template and B[*a*]P modified template were heat annealed to the appropriate primer depending on the experiment.

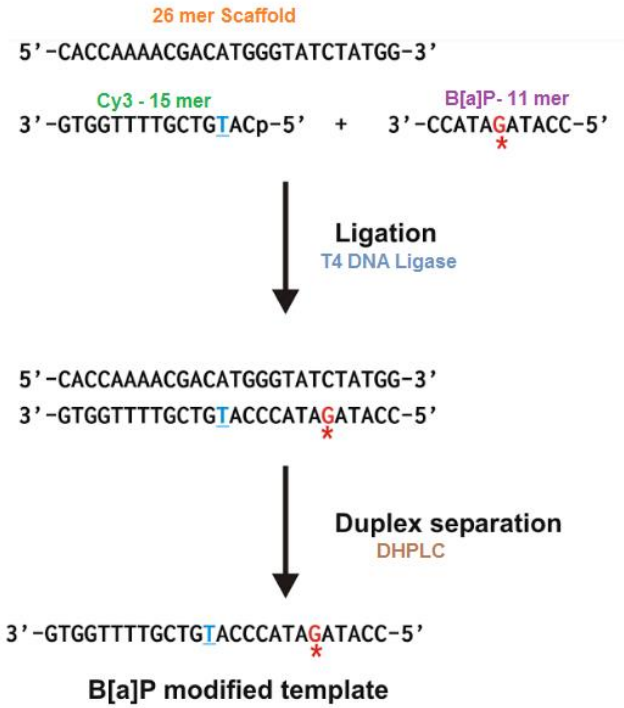


Figure 17. Preparation of B[a]P modified DNA.

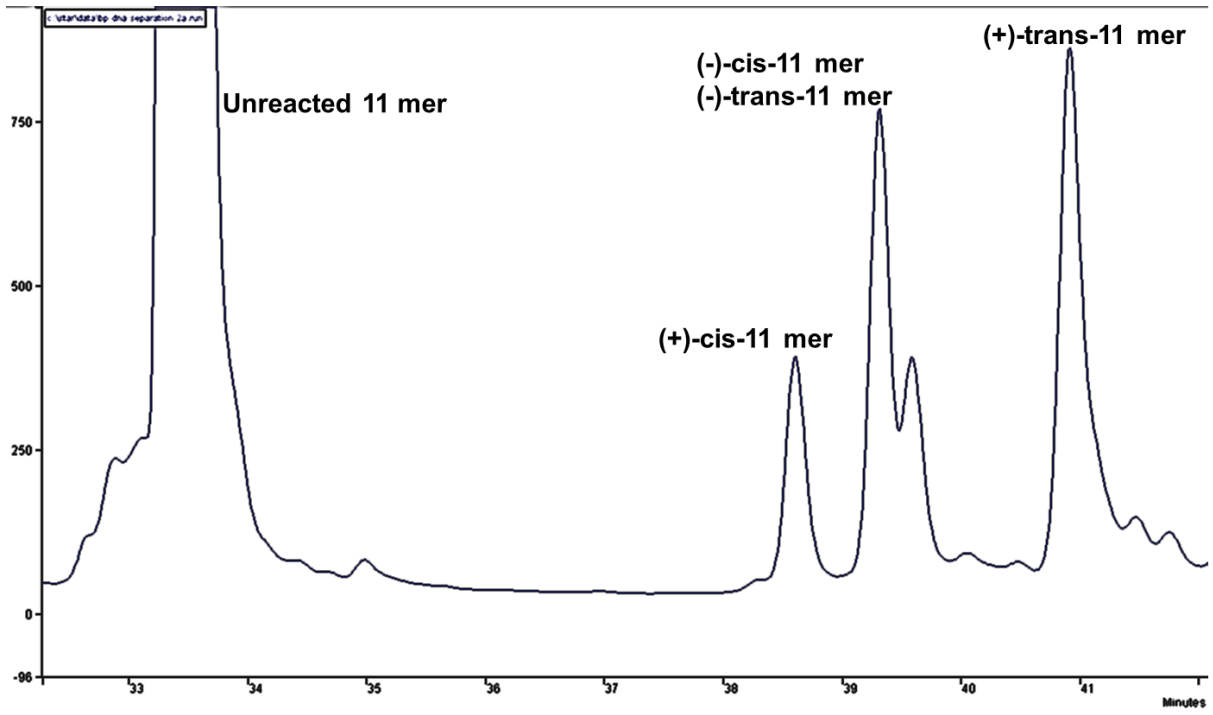


Figure 18. B[*a*]P isomers separation by HPLC.

2.2.5 Primer extension and single nucleotide incorporation assays.

Cy5-labeled primers and B[*a*]P modified or unmodified templates (15 nM) were incubated with Dpo4 (2 nM for unmodified template and 50 nM for B[*a*]P modified template) in reaction buffer (50 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 1 mM DTT, 25 □g/ml bovine serum albumin (BSA), 100 μM dNTPs, (±)10% DMSO) at 37°C. 10 μL aliquots of the reaction mixture were taken for each time point and mixed with 2x DNA loading buffer (95% formamide, 1 mg/ml bromophenol blue, 10 mM EDTA) to quench the extension reaction. The samples were loaded to a 20% polyacrylamide denaturing gel and run for ~16 hrs at 800V. Gels were scanned for Cy5 fluorescence in a Typhoon 9210 Variable Mode Imager (GE Healthcare).

Single nucleotide incorporation assays were carried out at 37^0C in a buffer containing 50 mM Tris-HCl, pH 7.5, 50 mM NaCl, 5 mM Mg, 0.025 mg/ml BSA, 10 mM DTT, 10 nM primer/template. Dpo4 concentration in the both unmodified and B[*a*]P

modified template containing reactions was 10 nM. However, incubation time for unmodified template containing reaction was 1 min whereas modified DNA reaction was 20 min. At the end of the incubation time, the reaction was quenched with 2X DNA loading dye. The gel electrophoresis is done similar the primer extension assay.

2.2.6 Dpo4 purification and Cy5 labeling

Dpo4 purification was carried as described [82, 85]. The native single cysteine in Dpo4 was labeled with a Cy5 maleimide ester (GE healthcare) as described [82].

2.2.7 Single molecule experiments

DNA was immobilized on a PEGylated quartz slide surface *via* biotin-streptravidin linkage. The PEGylated slide was prepared as previously described [69, 82, 86]. Single molecule experiments were done as follows: PEGylated slides were incubated with a streptavidin solution (0.02 mg/ml) for 7 min. After washing, 10-15 pM biotinylated primer-template was introduced and incubated for another 7 min. After a final wash, Cy5-labeled Dpo4 (10-15 nM) was introduced in binding buffer containing 50 mM Tris-HCl, pH 7.5, 35 mM $CaCl_2$, 25 μg/ml BSA, oxygen scavenging system (dihydroxybenzoic acid and photocatechuate dioxygenase from *Pseudomonas sp.*) and 1 mM Trolox. Calcium ions were used for all smFRET experiments to prevent DNA synthesis in the presence of dNTPs. Single molecule experiments were conducted at 21ºC in a home-built prism-based total internal reflection microscope. All the smFRET experiments were carried out at 80 ms time resolution. The traces were analyzed with 5 point moving average. The apparent FRET efficiencies were calculated as $FRET = I_A/(I_D + I_A)$, where I_D and I_A are the of donor and acceptor fluorescence intensities,

respectively. All smFRET histograms were built using >100 trajectories. The calculated FRET values are associated with an error of ± 0.02.

2.2.8 Single molecule microscopy

A home-built total internal reflection microscopy (TIRF) was used to measure the FRET intensity. The detailed experimental set up is explained in [86]. Briefly, a 532 nm green laser light is directed onto the single molecule experimental slide through a quartz, Pellin-Broca prism. The optics were arranged to get the incident beam angle greater than the critical angle, there by an evanescence wave is crated on the slide. The evanescence wave excites the fluorescence probes which are immobilized to the quartz slide. The donor (Cy3) and acceptor (Cy5) emissions were collected by a water immersion objective lens which is fixed to an inverted microscope. The Cy3 and Cy5 emissions were separately recorded by a CCD camera (Ixon+, Andor).

MD simulations

The crystal structure of the binary complex of DpO4 (pdbid 2IA6[18]) with the BPA adduct was used as the starting template for all simulations. MolProbity [87] was employed to check and protonate the structure. The DNA substrate sequence was modified to be similar to the primer used experimentally (5'-CATGGGTATC, 3'-GTACCCATA(B[*a*]P-*cis*-G)ATA), and the adduct in the crystal structure was modified from the trans conformation to the cis conformation, with the backbone of the DNA helix remaining in the same location as the crystal structure. The final C base in the 5' DNA strand was also modified to investigate several different systems and structural aspects. A total of nine systems with different nucleotide substitutions, adduct orientations and solvents were created to investigate the impact of DMSO on the system—six were in

pure water, three were in 90% water and 10% DMSO. All molecular dynamics simulations and structural analysis was done with the AMBER 14 simulation package,[88-90] and PCA and NMA were done with in the ProDy module in VMD [91, 92].

The simulations in pure water comprise two binary insertion complexes and one binary pre-insertion complex. All of the insertion binary complexes in water have the 3'-B[a]P adduct stacked within the DNA helix and the modified G nucleobase stacked within the minor groove [93]. The two sub-systems for the insertion binary complex had modified bases at the ending 5' position opposite the adduct, which was changed from C to G in the second system. The corresponding nucleobase was deleted completely from the third to simulate the pre-insertion complex. These systems were compared in order to investigate the effect of different nucleobases on the stability of the complex. Two additional simulations were run in 10% DMSO with the adduct flipped out into solvent; one with the original experimental primer given above and one with the final nucleobase at the end of the 5' strand (across from the adducted base location) deleted. Further changes were not performed for the DMSO since the flipped-out conformation for the B[a]P-cis-G adducted base does not have substantial interaction with the corresponding nucleobase. Additionally, the TYR-274 amino acid in the protein chain was changed to the primary Dunbrack rotamer (53.1377%) in order to eliminate steric clashes with the flipped out DNA base in the 10% DMSO systems [94].

Partial charges for the B[a]P-cis-G adduct were calculated using RESP fitting on the RED development servers [95, 96] and the standard procedure for creating modified nucleotides for proper linkages between nucleobases. An additional angle parameter

was taken from Mocquet 2007, and the partial charges generated in this work compared favorably with the partial charges from those results [97]. The guanine geometry parameters were taken from the OL15 DNA force field [98], and the remaining geometry parameters for the adduct were generated from the GAFF force field [99]. Parameters for the Ca^{2+} ions present in the crystal structure [100] and parameters for the DMSO solvent box[101] (where appropriate) were taken from the literature.

Molecular dynamics simulations on all systems were performed with the pmemd.cuda program from AMBER16, with the ff14SB force field for all of the protein parameters,[88-90] the OL15 force field for the DNA parameters[98] and a 1 fs timestep. Long-range electrostatics were treated with sPME with an 8Å cutoff for all nonbonded interactions.[102] All simulations were performed in the NVT ensemble with the Berendsen thermostat and barostat[103], and SHAKE was applied to all bonds involving hydrogen atoms.

Prior to solvation, all systems were neutralized to a net charge of 0 with Na+ ions in AMBER16's tleap program. The water systems were solvated in a rectangular box of TIP3P water using a 12Å pad between the surface of the protein and the edge of the box [104]. For the systems with 10% DMSO, in order to ensure an even distribution of DMSO molecules within the box, the AMBER16 program AddToBox was used to generate the 10% DMSO/90% TIP3P water solvent boxes, with the parameters of the box given to meet or exceed 12 Å from the surface of the protein and the edge of the box. RMSD, RMSF, correlation analysis, and solvation shell tracking were all calculated with AMBER16's cpptraj program. PCA and NMA were done with in the ProDy module

in VMD [91, 92]. 33,000 snapshots from each trajectory were analyzed with ProDy's PCA algorithm, with 100 PC modes generated.

CHAPTER 3: BULKY LESION, (+)-*cis*-B[*a*]P-N^2-dG BYPASS REQUIRES DPO4 BINDING IN DISTINCT CONFORMATIONS

Some sections of this chapter were taken from the manuscript by Pramodha Liyanage *et al.* “Bulky Lesion Bypass Requires Dpo4 Binding in Distinct Conformations” under review, JACS.

3 Background and significance

Extensive crystallographic studies have revealed that Dpo4 contains a spacious active site able to accommodate two templating bases simultaneously. Also, because it has unusually small finger and thumb domains, it has a very open active site with few DNA contacts leading to DNA slippage during synthesis and resulting in high levels of mismatch and frame-shift mutations [18, 105]. In high fidelity DNA polymerases, correct dNTP incorporation is accomplished by a large conformational change in the thumb and fingers domain, while crystal structures show that most Y-family polymerases, such as Dpo4, do not exhibit large conformational rearrangements in the ternary complex, which is also thought to affect the accuracy of nucleotide incorporation [106].

Replication of B[*a*]P-modified DNA is highly error prone, and the error signature depends on the organism studied, the specific DNA sequence around the adduct position and the stereochemistry of the adduct [107]. In some cases, more than one Y-family polymerase is required for successful primer extension across and after the adduct position [16, 60]. Once translesion synthesis has occurred, the bulky adduct can result in a distortion in the DNA that can activate the nucleotide excision repair (NER) pathway that will maintain the genomic integrity but fix a mutation if inaccurate DNA synthesis had occurred [108].

In the present study, smFRET and single nucleotide incorporation assays were used to characterize Dpo4 bypass of a DNA template containing (+)-*cis*-B[*a*]P-N^2-dG adduct. Our data show that Dpo4 binds to the B[*a*]P-modified DNA in two distinct conformations when the modified base is at the templating position. These conformations were identified as pre-insertion binary complex I and insertion binary complex I, each of which has the B[*a*]P adduct intercalated into the DNA helix, blocking nucleotide incorporation from occurring.

A recent study about the cytosolic dynamics which was carried out using confocal microcopy estimated the cytosolic dielectric constant as ~21.9, which is lower than that of the dielectric constant of water, ~78 [109]. Therefore, presumably solvation of B[*a*]P adducts is supported by the cytosolic micro environment stabilizing the hydrophobic pyrene ring system. Furthermore, they suggested a solvent system, 18% DMSO and tracetin, to be used in vitro to imitate the cytosolic polarity [109]. We used DMSO to mimic the cytosolic microenvironment to study the Dpo4 binding conformations and bypass mechanism *in vitro*. However, we could use only up to 15% DMSO to maintain the proper refractive index in the smFRET experiments.

Dpo4 can slowly extend the primer across from and after the adduct in the presence of DMSO by making a productive binary complex in which the (+)-*cis*-B[*a*]P-N^2-dG adduct is likely flipped out and exposed to the solvent. The smFRET experiments in the presence of DMSO reveal that Dpo4 also binds to form two complexes (insertion binary complex II and preinsertion binary complex II) and gel analysis shows that DMSO enhances the rate of incorporation across from the adduct. In the presence of DMSO, Dpo4 predominantly incorporates dG across from the adduct, producing a G:G

mismatch. Finally, when the DNA primer terminates across the adduct, DMSO has a more prominent effect on the Dpo4 binding conformation because in its presence we observe a single FRET value. Accurate incorporation occurs at the position following the adduct for both the modified and unmodified templates.

3.1 Results

3.1.1 Stalled Dpo4 primer extension is rescued by DMSO

Successful DNA adduct bypass depends on various factors, such as the nature of the adduct, its orientation within the DNA helix, the flanking DNA sequence and the properties of the DNA polymerase. To investigate the blocking potential of the B[*a*]P adduct, we first performed running start primer extension assay by Dpo4 using either an unmodified or (+)-*cis*-B[*a*]P-N^2-dG-modified template (Figure 20). The unmodified primer-template DNA construct results in efficient DNA synthesis, and the starting product is fully extended within 1 min (Figure 19). Lowering the Dpo4 concentration reveals replication intermediates at early time points and full extension after 10 minutes (Figure 1B, left panel). By contrast, the B[*a*]P-modified template results in Dpo4 stalling one base before the adduct (Figure 20C). After 30 minutes of incubation, some incorporation across the adduct takes place, but fully extended product is not observed even after two hours (Figure 20C).

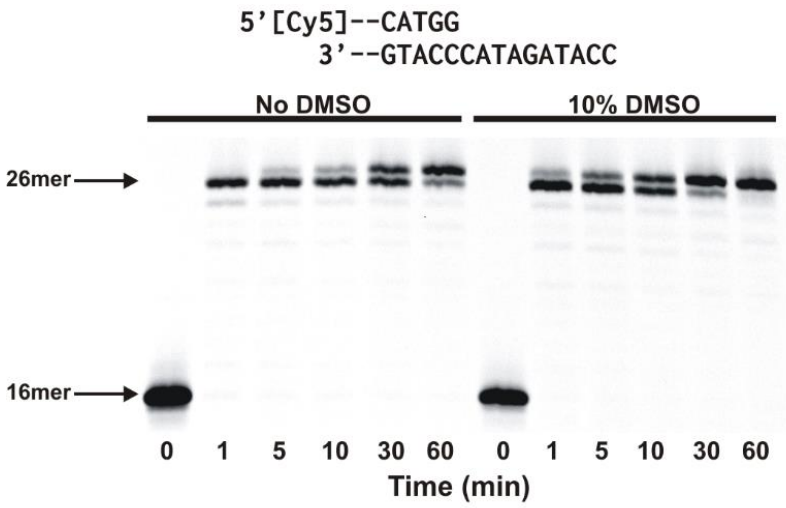


Figure 19. Dpo4 running start primer extension assay with unmodified template. The 16mer primer is fully extended to 26mer product in one minute. The sequence used for the extension assay is shown on the top of the gel.

Previous work using a (+)-*cis*-B[*a*]P-N^2-dA adduct has shown that the B[*a*]P aromatic moiety can adopt a stacked conformation within the DNA helix, similar to what is observed for the (+)-*cis*-B[*a*]P -N^2-dG adduct, effectively blocking replication by Dpo4 [16]. In that study, it was shown that adduct stacking could be reduced by the addition of an organic solvent, such as DMSO, that reduced the dielectric constant of the media presumably causing the adduct to adopt an alternate, more solvent exposed conformation and allowing increased adduct bypass [16]. To test whether a reduction in adduct stacking can also enhance bypass synthesis with the (+)-*cis*-B[*a*]P -N^2-dG adduct, we added 10% DMSO to the reaction buffer and measured DNA extension past the adduct.

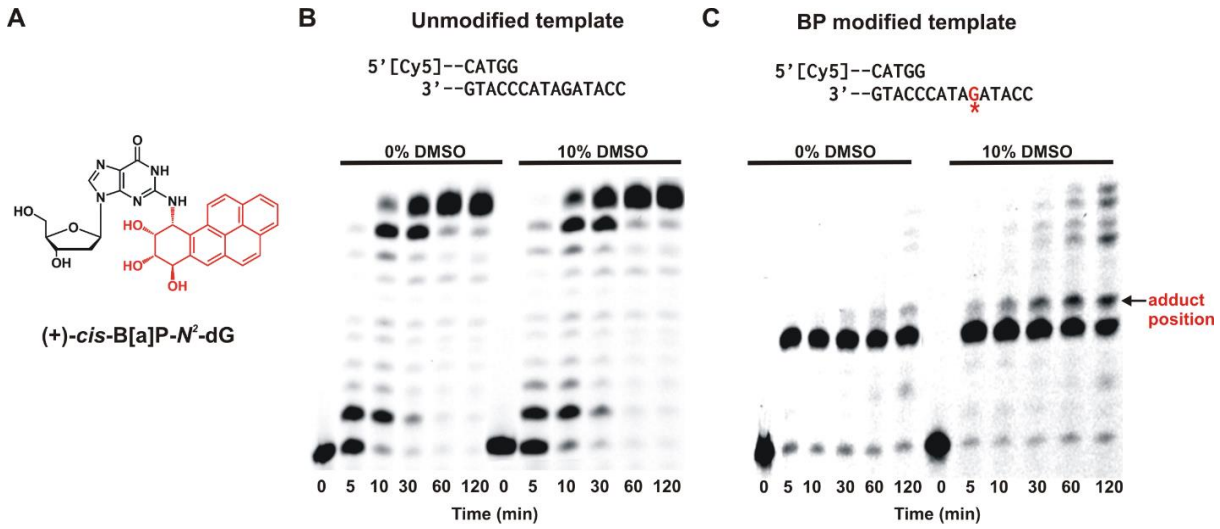


Figure 20. DMSO rescues stalled Dpo4 primer extension. **(A)** Chemical structure of the carcinogenic adduct (+)-cis-B[*a*]P-N^2-dG isomer. **(B)** Primer extension assay with unmodified template. DNA construct used for the assay is shown on top with 5'-Cy5 labeled primer). Primer extension reaction was initiated with 2 nM Dpo4 and aliquots were quenched at the indicated time points before loading onto a denaturing acrylamide gel. The enzyme completes the extension in 30 min both in the absence and presence of DMSO. **(C)** Primer extension assay with B[*a*]P modified template. Position of the adduct is indicated by red G with an asterisk. Reaction was initiated with 50 nM Dpo4. The adduct stalls Dpo4 one nucleotide before the adduct position (arrow). 10% DMSO rescues stalled Dpo4 after 60 min.

In agreement with the prior study, the presence of 10% DMSO allows Dpo4 to extend across and past the B[*a*]P adduct faster than in the absence of the organic solvent (Figure 20C). Addition of 10% DMSO did not significantly alter DNA extension kinetics for the unmodified template (Figure 20B). These data suggest that DMSO induces a conformational rearrangement of the B[*a*]P adduct that results in the formation of a more catalytically active complex.

3.1.2 B[*a*]P adduct induces different Dpo4 binding conformations compared to unmodified DNA

It has been shown that different Y-family polymerases display different efficiencies and fidelities for the same B[*a*]P adduct during TLS [52]. However, most of these ensemble-averaged experiments do not provide information about the

polymerase binding dynamics in the population or sub-populations. Single-molecule techniques such as smFRET enable the monitoring of transient events and individual molecular populations [110]. To characterize the conformation and dynamics of Dpo4 on the adducted DNA using smFRET, we labeled the DNA with a FRET donor (Cy3) and Dpo4 with a FRET acceptor (Cy5, Figure 21A). This labeling strategy has previously enabled us to characterize Dpo4 binding kinetics while monitoring the global conformation of the binary complex [82, 83]. The resulting single-molecule trajectories show free DNA as Cy3-only fluorescence, and Dpo4 binding as anti-correlated increases in Cy5 fluorescence and decreases in Cy3 (Figure 21B). In agreement with our previous work, [82] the DNA-Dpo4 binary complex samples two FRET distributions for the unmodified DNA construct (Figure 21B, right). The main population is centered at ~0.8 FRET, while the minor population is centered ~0.6 FRET. Based on our previous work,[82] we assign the high FRET peak (0.8) to a pre-insertion binary complex, and the low FRET peak (0.6) to the insertion binary complex. Single-molecule trajectories with the (+)-*cis*-B[*a*]P -N^2-dG-modified template also exhibit a similar behavior with two possible binding configurations (Figure 21C). Analysis of ~100 single-molecule trajectories shows a bimodal FRET distribution, albeit with overall lower FRET values (peaks centered at ~0.5 and ~0.75, Figure 21C, right), indicating that the (+)-*cis*-B[*a*]P-N^2-dG adduct slightly distorts the polymerase binding orientation on the DNA template.

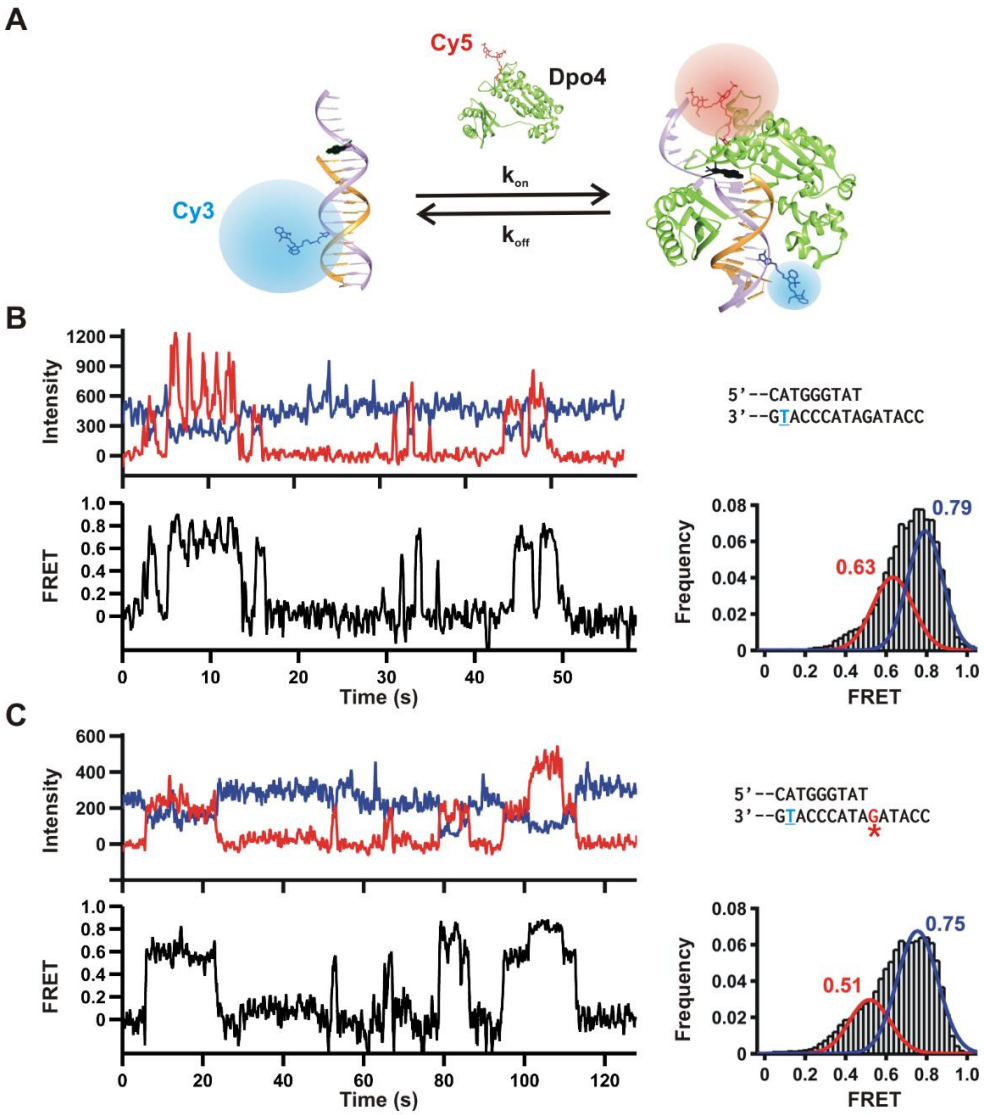


Figure 21. Single-molecule Dpo4 binding dynamics on unmodified and B[a]P-modified templates. **(A)** Schematic diagram of the smFRET setup to monitor the Dpo4 binding dynamics to DNA. Cy3-labeled primer-template is immobilized onto a quartz slide trough biotin-streptravidin and with Cy5-labeled Dpo4 in solution. Upon Dpo4 binding, FRET takes place between Cy3 and Cy5. **(B)** Characteristic smFRET Dpo4 binding trajectory to the unmodified DNA. DNA sequence is on the right (see materials and methods). Underlined T indicates Cy3 position. Dpo4 binding decreases Cy3 fluorescence (I_D, blue) while Cy5-signal (I_A, red) increases simultaneously. Apparent FRET efficiencies (bottom) are as calculated as $I_A/ (I_A + I_D)$. Time-binned smFRET histogram from >100 trajectories; bimodal distribution reveals two binding conformations. **(C)** smFRET Dpo4 binding trajectories to the B[a]P modified DNA. Red G indicates adduct location. Time-binned smFRET histogram from >100 trajectories yields a bimodal distribution revealing two binding conformations.

3.1.3 Dpo4 binds to B[*a*]P-modified DNA in a catalytically active conformation in the presence of DMSO

There is no structural information for the (+)-*cis*-B[*a*]P-N^2-dG adduct in complex with Dpo4. Early solution NMR studies of the (+)-*cis*-B[*a*]P-N^2-dG adduct in duplex DNA have shown that the B[*a*]P hydrocarbon moiety intercalates into the helix [28]. Crystal structures of the (+)-*cis*-B[*a*]P-N^2-dA DNA adduct in complex with Dpo4 show two possible conformations where B[*a*]P moiety is either stacked within the DNA helix or flipped out to the major groove [16]. The latter orientation is energetically unfavorable because the hydrophobic moiety is exposed to solvent, but presumably it can be stabilized in the presence of organic solvents, which has been also shown to improve Dpo4 lesion bypass synthesis [16]. As shown above, the presence of DMSO also enhances Dpo4 bypass of the (+)-*cis*-B[*a*]P-N^2-dG adduct. Therefore, we hypothesize that the (+)-*cis*-B[*a*]P-N^2-dG adduct behaves akin to the (+)-*cis*-B[*a*]P-N^2-dA adduct in the binary complex. To test this, we measured the binary complex orientation using smFRET in the presence of DMSO (Figure 23A).

Addition of DMSO does not significantly affect the FRET distribution for the unmodified template, and the relative area under each peak remains essentially unchanged throughout the DMSO titration (Figure 23B), suggesting that the equilibrium between the insertion and pre-insertion binary complexes is not altered by DMSO. In the presence of the B[*a*]P-modified template, increasing the DMSO concentration favors the low FRET conformation (designated insertion binary complex II) and shifts the FRET distribution towards lower values (Figure 23B). We also measured the polymerase complex dissociation kinetics by analyzing the dwell-time distributions (Figure 23). For

the unmodified template, Dpo4 dissociates with a rate constant k_{off} = 0.22 s^{-1}, which agrees well with our previous data using a different DNA construct (0.20 s^{-1}) [82]. This value remains unchanged in the presence of up to 15% DMSO (Figure 22C).

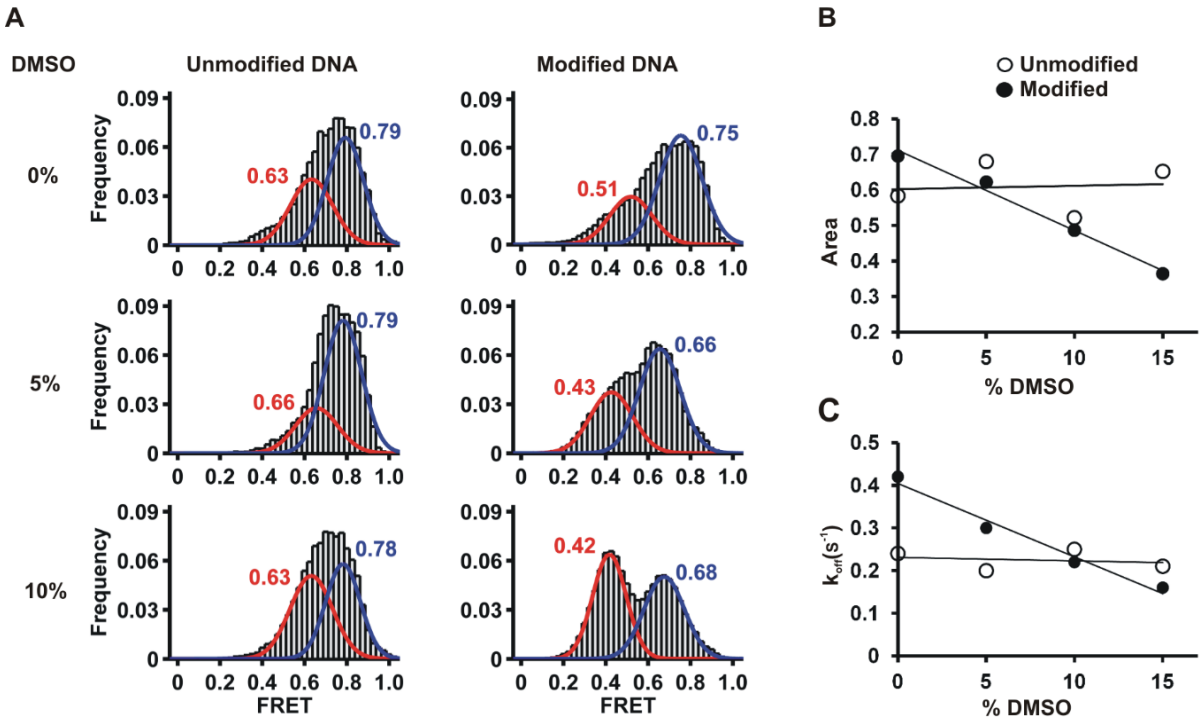


Figure 22. Dpo4 binding conformation of the (+)-*cis*-B[*a*]P-N^2-dG modified template. **(A)** smFRET histograms with increasing DMSO to monitor its effect on Dpo4 binding for the unmodified and modified DNAs. All histograms are fit to binomial distributions. **(B)** Relative fraction area of high FRET state as a function of DMSO percentage in the buffer. Solid lines represent linear fits to the data. DMSO stabilizes the low FRET conformation for the adducted DNA but the unmodified DNA. **(C)** Dpo4 dissociation rate constant (k_{off}) as function of DMSO concentration. DMSO stabilized the modified binary complex but not with the unmodified DNA.

However, in the presence of the B[*a*]P-modified template Dpo4 dissociates approximately two-fold faster (k_{off} = 0.42 s^{-1}), indicating that the adduct destabilizes the binary complex by about 0.32 kcal/mol. Addition of DMSO decreases k_{off} linearly (Figure 21C) to 0.21, in agreement with the idea that the organic solvent favors the B[*a*]P solvent-exposed conformation, thus re-stabilizing the binary complex.

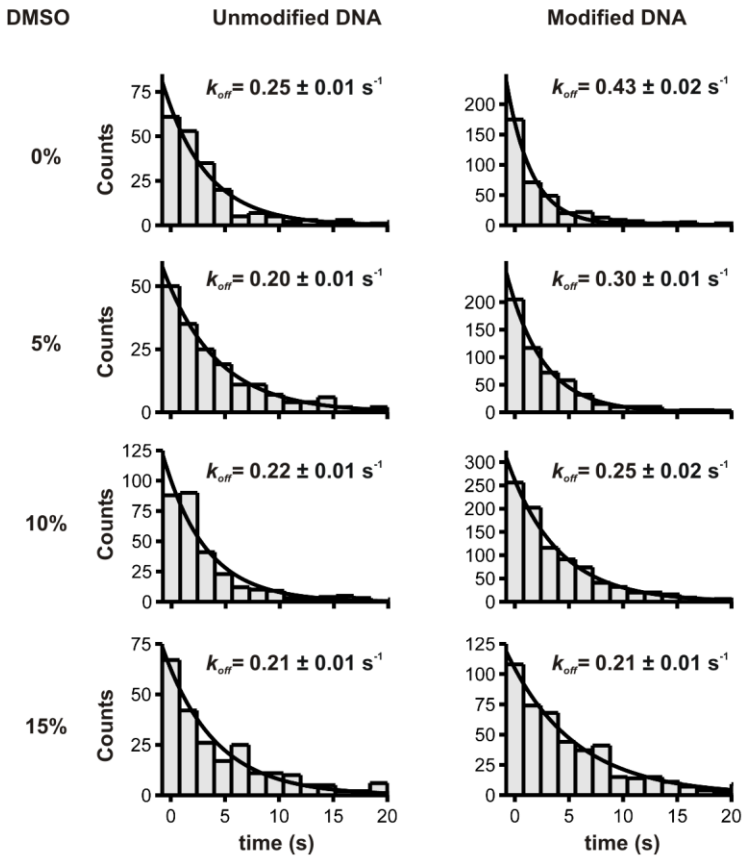


Figure 23. Dwell time distributions for Dpo4-DNA binary complex. The dissociation constants (k_{off}) were calculated by fitting data to single exponential decays. The corresponding smFRET experiments were carried out as a function of DMSO concentration shown in Figure 7C.

3.1.4 Dpo4 favors dG misincorporation across (+)-*cis*-B[*a*]P-N^2-dG adduct

Y-family polymerases do not have proofreading activity, which is an inherent property of high fidelity polymerases. Therefore, Y-family polymerases have higher error rates (in the 10^{-2} - 10^{-3} range) when copying normal DNA [111]. To study the preference of Dpo4 nucleotide incorporation in the ternary complex, we first used a single-nucleotide incorporation assay with unmodified DNA (Figure 24). Dpo4 mainly incorporates the next correct nucleotide, dC, across from the dG in the unmodified template in the presence or absence of DMSO. However, other non-complimentary bases can also be incorporated depending on the Dpo4 concentration and reaction time (data not shown).

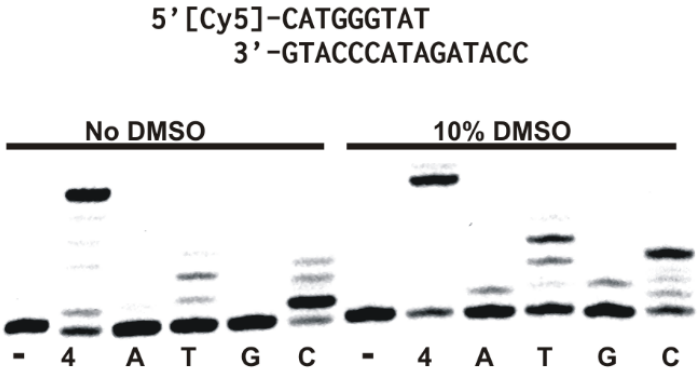


Figure 24. Dpo4 incorporates the next correct nucleotide in the unmodified DNA construct. Single nucleotide incorporation assay shows the tendency of dNTP incorporation across the dG, templating base in the 20mer/26mer primer-template. Unmodified DNA sequence used for this assay is shown on the top. Dpo4 mainly incorporates dC in both in absence and presence of DMSO. Lanes 1 and 7 represent control experiments in absence of dNTPs in the experiments (labeled as -). Lanes 2 and 8 represent experiments with all 4 dNTPs (labeled as 4). Other lanes contain only the designated dNTP.

We next carried out single-nucleotide incorporation assays to characterize the Dpo4 error signature across (+)-*cis*-B[*a*]P-*N*^{2}-dG. In the absence of DMSO, DNA extension proceeds inefficiently with minor amounts of dG and dA mostly being incorporated (Figure 25A left). The same reaction in 10% DMSO proceeds to a much larger extent, with dG being preferentially incorporated across from the adduct (Figure 25A right), thus resulting in the formation of a G:G mismatch. These results suggest that the DMSO-induced solvent exposed conformation of (+)-*cis*-B[*a*]P-*N*^{2}-dG adduct is more catalytically active but does not adopt a conformation permitting correct dNTP incorporation.

To investigate the possible ternary complex conformations leading to this error signature, we carried out smFRET experiments in the presence of each nucleotide. To prevent dNTP incorporation, we used Ca^{2+} as divalent metal ion instead of $Mg^{2+,}$ which has been used in the past for inhibiting nucleotide incorporation by Dpo4 [82, 112]. In agreement with our previous results, binding to an unmodified template in the presence of each of the four dNTPs produces a single FRET distribution (Figure 25B, left panels, red). Addition of 10% DMSO (Figure 25B, blue) did not significantly alter peak positions with the exception of dGTP and dTTP, both of which change slightly. The lack of significant changes in the FRET distributions suggests that the overall conformation of the polymerase complex remains largely unchanged in the presence of DMSO.

In the presence of the (+)-*cis*-B[*a*]P-*N*^{2}-dG adduct and each dNTP (Figure 25B, right panels), the ternary complexes sample mostly a single conformation, with FRET values consistently larger than the unmodified template, suggesting that the ternary complex structure is somewhat distorted by the presence of the adduct. When 10%

DMSO is present (blue histograms), FRET distributions for dATP and dCTP remain unchanged (centered at ~0.62), whereas the FRET distributions for dGTP and dTTP, shift towards lower values (~0.53). This value is similar to that of the unmodified template with the correct nucleotide (dCTP) indicating that, in the presence of DMSO, the B[*a*]P-modified ternary complex adopts a similar conformation to the cognate dG base in presence of either dGTP or dTTP.

3.1.5 B[*a*]P adduct at duplex DNA terminus is bypassed in an error-free manner

We next determined how a (+)-*cis*-B[*a*]P-N^2-dG adduct bound to the terminal base pair of the primer-template affects the binary complex conformation. As expected for the unmodified primer-template in the absence of DMSO, Dpo4 samples the two conformations that we have previously assigned to the pre-insertion (FRET ~0.8) and insertion binary complexes (FRET ~0.6, Figure 26A). In the presence of 10% DMSO, the FRET distribution does not significantly change and still exhibits two main populations, showing again that the organic solvent does not alter the equilibrium between the insertion and pre-insertion binary complexes (Figure 26B). Symmetric transition density plots at both 0% and 10% DMSO show reversible shuttling between the insertion and pre-insertion conformations (Figure 26A and B).

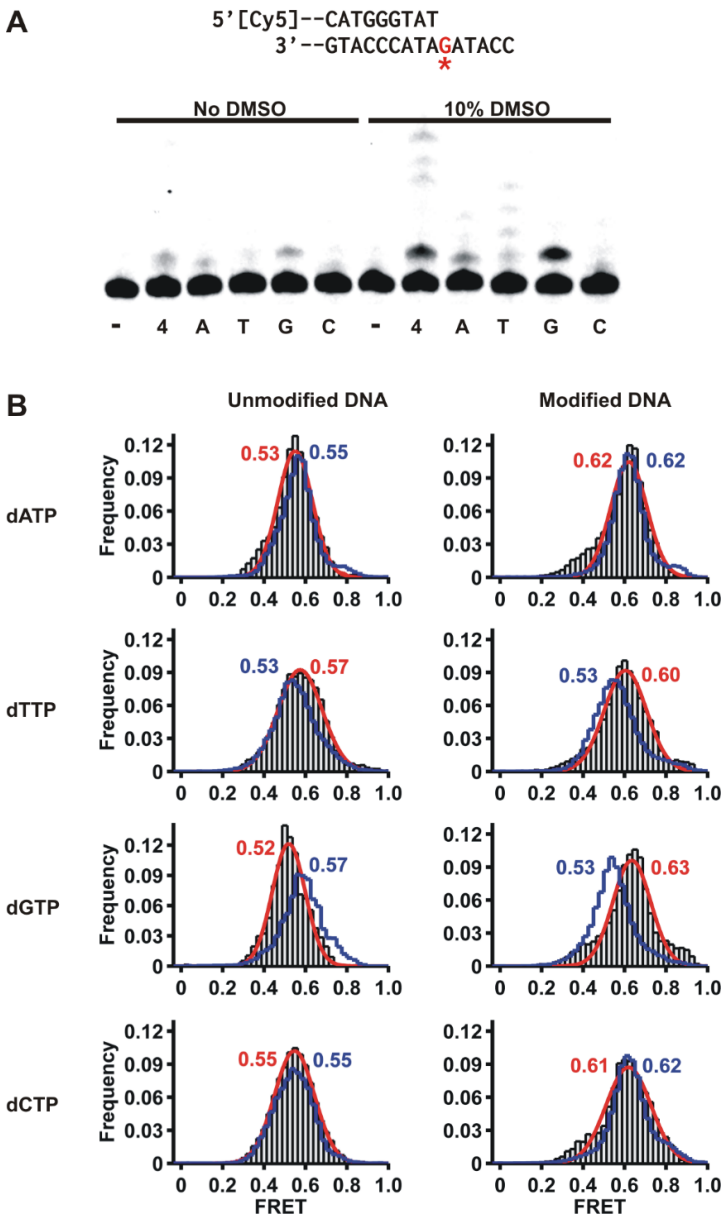


Figure 25. Dpo4 binding dynamics to the DNA with adducted dG as templating base in the presence of each nucleotide. **(A)** Modified DNA sequence used for this single nucleotide incorporation assay (top). Lanes 1 and 7 are control experiment in the absence of dNTPs. Lanes 2 and 8 are control experiments in the presence of all dNTPs. The other lanes contain only the specified nucleotide. In the absence of DMSO no incorporation takes place. In the presence of DMSO dG is incorporated across the adduct **(B)** smFRET distributions for Dpo4 binding to unmodified (left) and B[a]P modified (right) DNA in the presence of 1 mM dNTP as indicated, in the absence (red) and presence (blue) of 10% DMSO. In the presence of nucleotides all histograms are fit to single Gaussian distributions.

When the B[*a*]P adduct is present at the DNA duplex terminus and in the absence of DMSO, the FRET distribution solution shows two peaks centered near ~0.7 and ~0.5, presumably corresponding to slightly perturbed conformations of the pre-insertion complex I and insertion binary complex I as observed in Figure 21 (Figure 26C). Addition of 10% DMSO results in a dramatic change of the FRET distribution that only exhibits a single peak centered ~0.55. This result shows that in 10% DMSO Dpo4 samples only the B[*a*]P solvent-exposed conformation, which results in a global conformation akin to the insertion binary complex II (Figure 26D).

Extension of DNA with (+)-*cis*-B[*a*]P-N^2-dG adduct at the duplex terminus by Dpo4 takes place mostly error-free in this sequence context (Figure 28A), as dT is the only nucleotide showing significant incorporation across from dA. The degree of dT incorporation is very slightly enhanced by the presence of 10% DMSO, suggesting that the solvent exposed conformation for (+)-*cis*-B[*a*]P-N^2-dG adduct may be favored even in the absence of organic solvent. The same DNA primer with unmodified template shows similar dT incorporation across the dA (Figure 27). In agreement with these results, the FRET distribution in the presence of the next correct nucleotide dT is virtually identical in 0% and 10% DMSO (Figure 27). Moreover, this FRET value (0.55) matches the one observed for the unmodified DNA construct in the presence of dT, suggesting that the conformation of the cognate ternary complexes is very similar for unmodified and B[*a*]P modified DNA constructs. Addition of incorrect nucleotides dA, dC and dG for the B[*a*]P-modified DNA results in higher FRET values (~0.6) when compared to the cognate dT (Figure 28B). The tertiary complex binding conformation only resembles the ones observed with the unmodified DNA in the presence of the

correct dTTP. As we have previously shown, for unmodified DNA all four dNTPs result in similar FRET distributions (~0.54), suggesting that cognate and non-cognate ternary complexes present an overall similar structural arrangement.

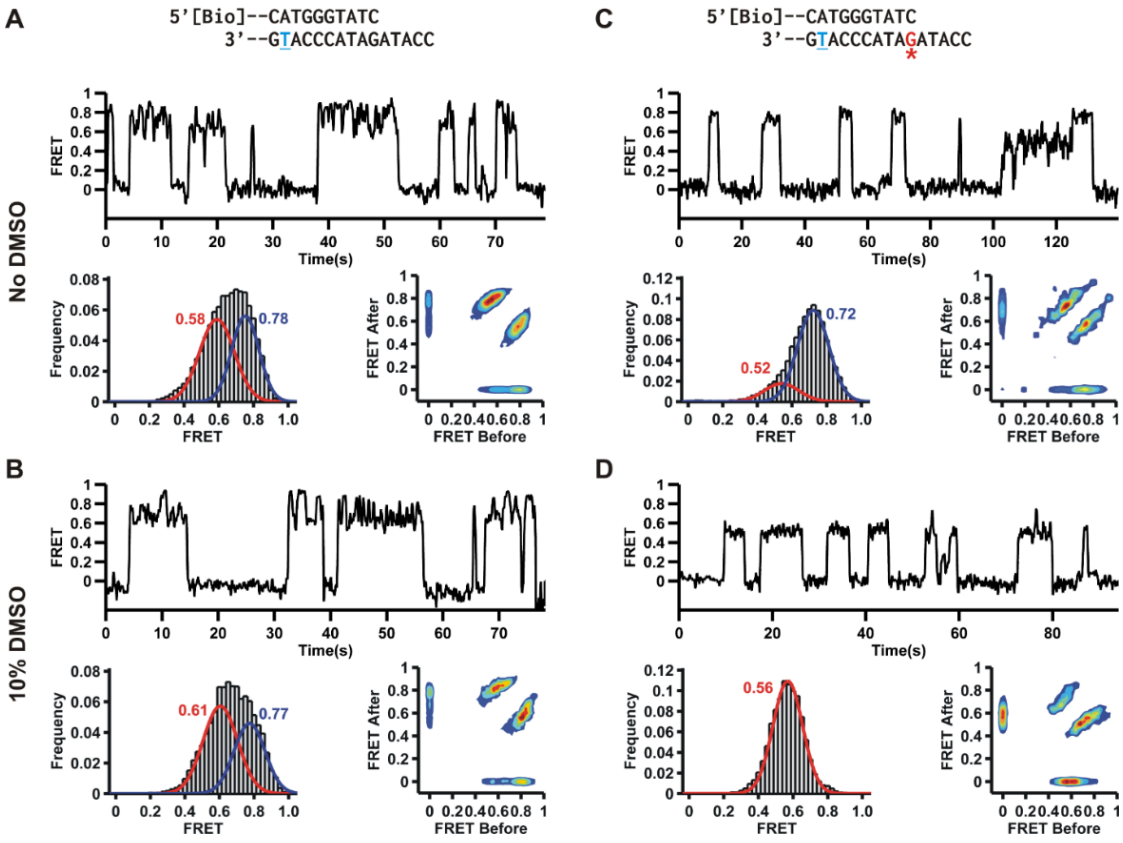


Figure 26. Dpo4 binding dynamics to the DNA where the primer terminates across the adduct. **(A)** Dpo4 binding to the unmodified DNA in the absence of DMSO. Characteristic FRET trajectory (top), histogram (bottom left) with bimodal fit and transition density plot (TDP, bottom right). **(B)** Dpo4 binding to the unmodified DNA in the presence of 10% DMSO. Characteristic FRET trajectory, and resulting histogram and transition density plot (TDP). **(C)** Dpo4 binging to the adducted DNA in the absence of DMSO. Characteristic FRET trajectory, and resulting FRET histogram and TDP. **(D)** Dpo4 binging to the adducted DNA in the presence of 10% DMSO. Characteristic FRET trace, FRET histogram and TDP.

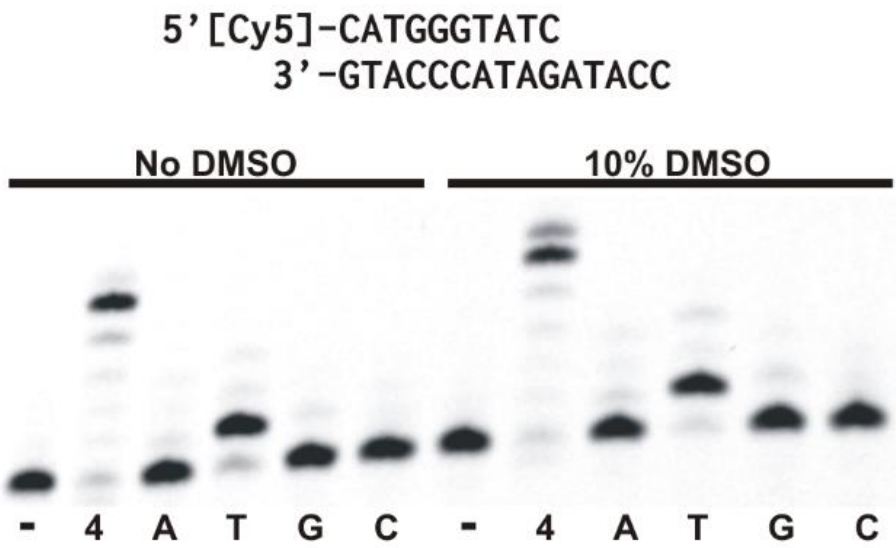


Figure 27. Single nucleotide incorporation assays on 21mer/26mer unmodified DNA primer-template. Cy5-labeled DNA primer shows the extended products on the gel. Dpo4 incorporates the next correct dNTP, dT in the DNA construct shown on top of the gel. In this single nucleotide incorporation assay, lanes 1and 7 (dash line) corresponding to control experiments without dNTPs. Lanes 2 and 8 contain all four dNTPs (labeled as 4). Other lanes contain the designated dNTP below the lane. The primer extension reaction was quenched after 1 min.

3.1.6 Misincorporated dG forms a stable conformation across the (+)-*cis*-B[*a*]P-N^2-dG adduct.

Incorporation of dG across the (+)-*cis*-B[*a*]P-N^2-dG adduct by Dpo4 is strongly supported by biochemical and single molecule data (Figure 25). To further investigate the experimental results, we performed several molecular dynamics simulations of various systems. All simulations have RMSDs below 2.5 Å on average, indicating that the protein structures were fairly stable in the simulations. The details of the RMSD results the is shown in the Figure 29.

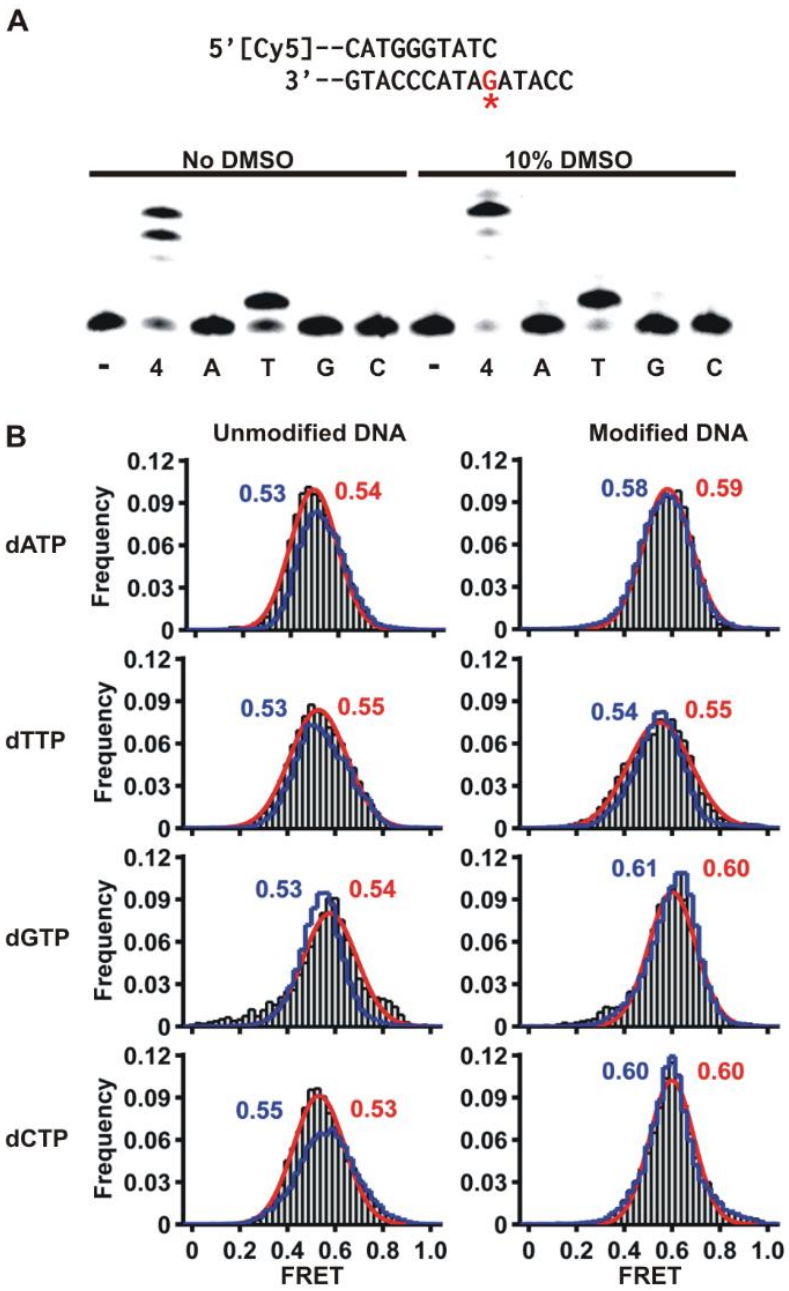


Figure 28. Dpo4 incorporates the correct dNTP after the adduct position. **(A)** Single nucleotide primer extension of the adducted DNA in the absence and presence of DMSO. After the adduct position there is no observable difference in Dpo4 synthesis with DMSO. **(B)** FRET efficiency histograms for Dpo4 biding to unmodified (left) and B[a]P modified DNA (right) in the presence of indicated nucleotide (1 mM). The FRET distributions in the presence of 10% DMSO are shown in blue stairs and in grey bars (with Gaussian fit in red) in the absence of DMSO.

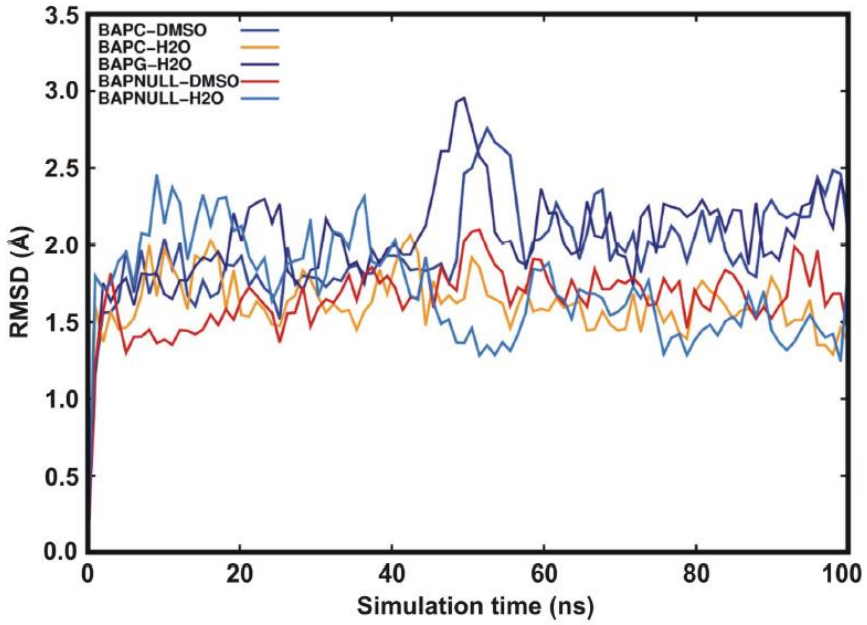


Figure 29. The RMSD of the protein backbone over time for each trajectory.

To confirm the stability of the (+)-*cis*-B[*a*]P-N^2-dG:dG (B[*a*]P-dG:dG) mismatch, and to show that the force field parameters for the computational simulations were consistent with those in the literature, we investigated the relative structure of the DNA in three similar systems with different nucleotides across from the adduct (Figure 30). Figure 29A denotes a structure where the (+)-*cis*-B[*a*]P-N^2-dG has no corresponding nucleobase, i.e. the binary preinsertion complex. In this case, the adduct over the course of the simulation becomes stacked within the minor groove and distorts the attached nucleobase and backbone structure substantially. Structures in Figure 29B and Figure 29C are consistent with previous work done by Broyde *et al*.; although the previous work was focused on systems where the adduct was mid-DNA chain rather than at the primer-template junction [108]. The (+)-*cis*-B[*a*]P-N^2-dG:dC (B[*a*]P-dG:dC) complex shows that the adduct remains stacked within the major groove and the

adducted G base remains stacked within the minor groove for the duration of the simulation (Figure 30B).

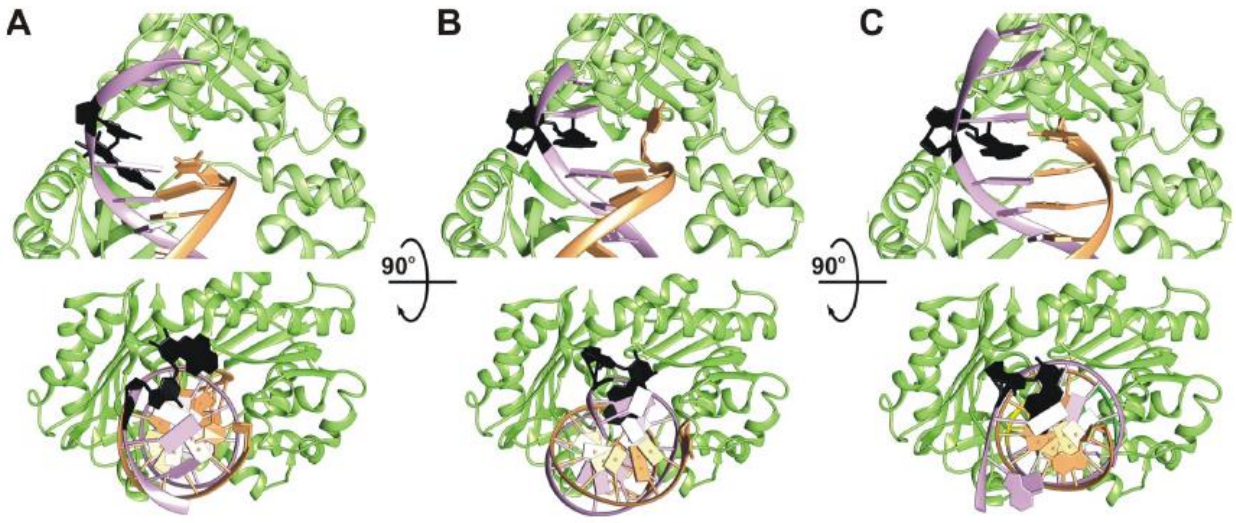


Figure 30. Representative structures of the DNA helix structures from the end of three MD simulations. **(A)** The structure of B[a]PG without a corresponding nucleobase at the end of 100 ns of MD in a pure water system. **(B)** The structure of B[*a*]PG-dC at the end of 100 ns of MD in a pure water system. **(C)** The structure of B[*a*]PG-dG at the end of 100 ns of MD in a pure water system.

Broyde's results show the dC base to be in a similar, flipped-out orientation, but with a hydrogen bond not seen here—most likely the fact that the dC base is at the end of the DNA primer makes it even more mobile and less likely to form a hydrogen bond with the backbone. The (B[*a*]P-dG:dG) system (Figure 30C) also is consistent with Broyde's results, where the adduct is stacked within the major groove, the adducted dG base is stacked within the minor groove, and the corresponding dG base orientation is also stacked within the major groove and on top of the adduct, which is again consistent with the literature. This is also in good agreement with the experimental and previous computational results that the dG incorporation across from the adduct is the most stable and preferred structure [108].

In addition, to further investigate the movement and structure of the protein, several sets of dynamic structural analyses were performed. Cross-correlation analysis was done on the B[*a*]P-dG:dC and B[*a*]P-dG:dG systems in water and 10% DMSO to help isolate differences in the relative movements of the residues in each trajectory. Additionally, two different sets of residue-wise correlation analysis subtractions (difference correlation) were performed—first, between two systems in water with differing nucleotides across from the adducted base (dC and dG) and second between one system in DMSO with a solvent-exposed conformation and one system in water, with the same nucleotide across from the adducted base (dC). Finally, the difference correlation associated with the adducted base was extracted and mapped onto the protein and DNA (Figure 31).

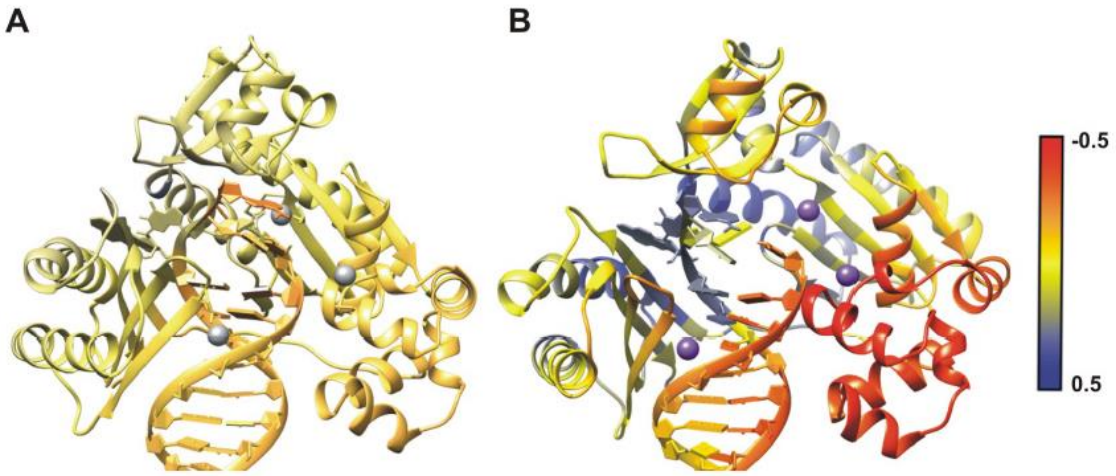


Figure 31. Difference correlation analysis relative to the adducted B[a]P-dG residue. Difference correlation analysis relative to the adducted B[*a*]P-dG residue. Red denotes an increase in anticorrelation, yellow denotes no difference in correlation, and blue denotes an increase in correlation. The color scale has been rescaled to a range of -0.5 to 0.5 for clarity. **(A)** The difference in correlation between B[*a*]P-dG:dC in pure water and B[*a*]P-dG:dG in pure water relative to the adducted residue with the calcium ions colored in gray. **(B)** The difference in correlation between B[*a*]P-dG:dC in 10% DMSO/90% water with a solvent-exposed adduct conformation and B[*a*]P-dG:dC in pure water with the calcium ions colored in purple.

The difference in correlation between the two systems in pure water is fairly negligible, indicating that the movement of the protein and DNA residues relative to the adduct is essentially the same regardless of the corresponding nucleotide, despite the differences in DNA conformation (Figure 31). In DMSO, however, there are very large differences between the correlation relative to the adducted base, especially for the section of protein around the adducted base and near the active site. The correlated movement of the palm domain of Dpo4 relative to the adducted base in its flipped out orientation is reasonable considering the change in orientation. Also interesting is the large increase in anti-correlated movement of the thumb region of Dpo4, since that particular region is associated with the translocation of the DNA. To further investigate the large-scale motions of the protein we also performed principal coordinate analysis (PCA), which shows substantial differences in the largest movements of the protein between the (B[*a*]P-dG:dC) system in water versus in 10% DMSO. PCA is a useful tool to isolate and show large protein movements. PCA was performed on 33000 snapshots for each trajectory. 100 total modes were generated, and the cross-correlation plots for those modes were subtracted in the same way described in above for the distance correlation plots (Figure 32). These plots show some interesting features and strong differences; notably, there are again fewer differences between B[*a*]P-dG:dC in water and B[*a*]P-dG:dG in water versus B[*a*]P-dG:dC in water and B[*a*]P-dG:dC in DMSO. In light of the experimental results that show a return to function for Dpo4 with the adduct in a solvent-exposed conformation, the substantial differences in the large motions of the protein support the concept that not only is the structure of the protein different, but that the overall movement of the protein has drastically changed. A representative

depiction of the first (and largest) principal coordinate (PC) for B[*a*]P-dG:dC in water and B[*a*]P-dG:dG in DMSO can be seen in Figure 31A and Figure 31B respectively. The two modes contain not only very different movements, but also the B[*a*]P-dG:dC in water's first PC has substantially less movement overall. This is also consistent with the experimental result that when the adduct is stacked within the major groove of the DNA helix, the protein is blocked from functioning properly. That said, it is important to recognize that the first PC only represents one large-scale motion of the protein—while more difficult to visually interpret, the overall correlation differences are more telling, since they show substantial differences in many of the PCs rather than just one.

To investigate the effect of solvation in 10% DMSO on the adduct, a radial distribution function (RDF) for the C3B atom in the adduct (near the middle of the B[*a*]P adduct, with the same atom naming scheme as in PDB 2IA6) and the sulfur atoms in the DMSO solvent was calculated (Figure 33B). The DMSO RDF adopts two distinct conformations, with either the sulfur or the methyl groups facing the B[*a*]P rings, which in this case can be seen most clearly above the adduct (Figure 33A).

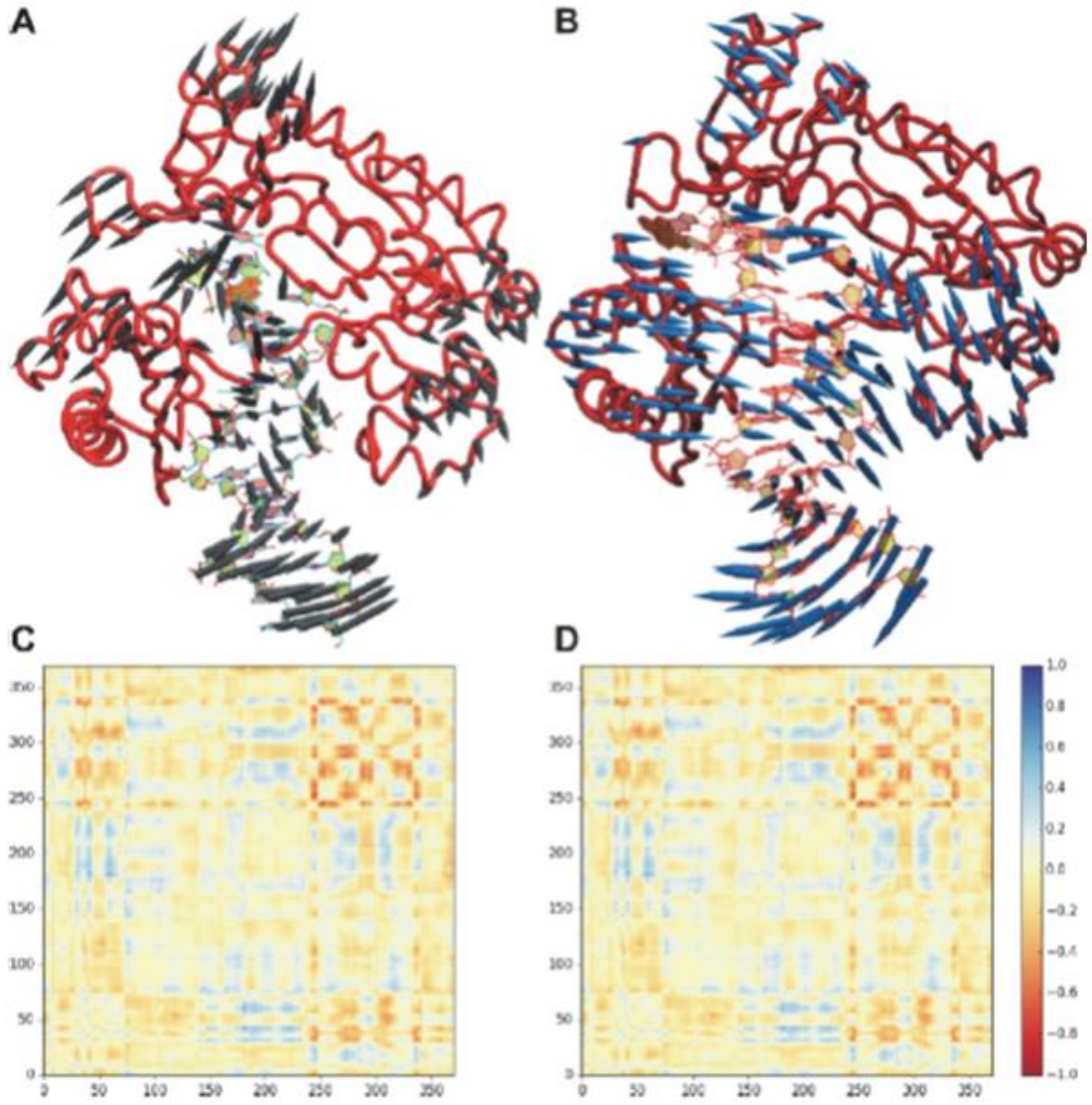


Figure 32. Structural representations from principal coordinate analysis Structural representations from principal coordinate analysis and difference correlation matrices generated from the first 100 PC modes and normalized to +/- 1. **(A) T**he first principal coordinate (PC) mode of B[a]P-dG:dC in water**. (B)** The first PC mode of B[a]P-dG:dG in DMSO in the solvent-exposed orientation. **(C)** The subtraction of the normal mode correlation matrices of B[a]P-dG:dC in water from B[a]P-dG:dG in water **(D)** The subtraction of the normal mode correlation matrices of B[a]P-dG:dC in DMSO subtracted from B[a]P-dG:dC in water.

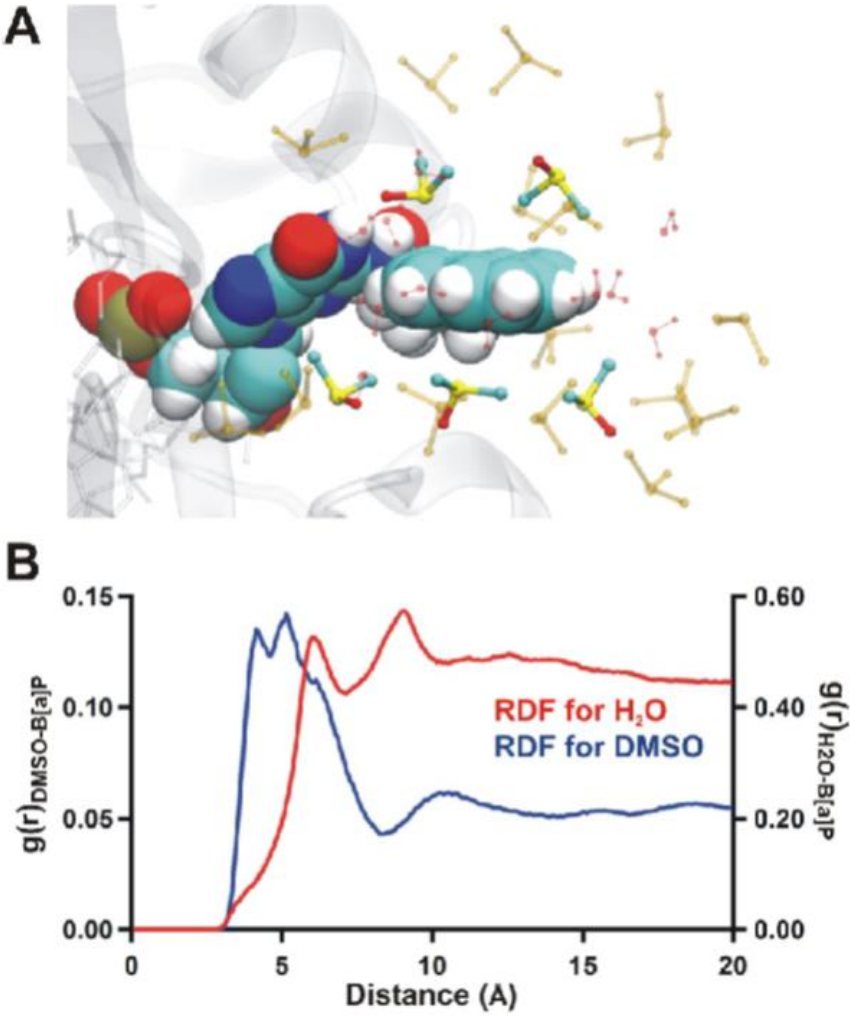


Figure 33. (A) Representative structure of the solvent-exposed adduct in 10% DMSO/90% water after 100 ns of MD simulation. The adduct is represented in VDW, the first solvation shell of DMSO (within 5.15 A) is represented in ball and, the second solvation shell of DMSO (within 10.15 A) is represented by ball and stick and colored transparent yellow, and the first solvation shell of water (within 5.15 A) is represented by ball and stick and colored transparent red. **(B)** Calculated average radial distribution function over the entire 100 ns MD trajectory between the C3B atom in the B[a]P adduct and the S atom in the DMSO solvent molecules, and between the C3B atom in the B[a]P adduct and the O atom in water.

On average, it is observed that when the sulfur from DMSO is pointing towards the B[*a*]P the distance between this atom and C3B of B[*a*]P is around 4.5-5 Å. Conversely, when the methyl groups of DMSO are pointing towards the B[*a*]P the distance is increased to 5-5.5 Å. This helps explain the splitting observed in the first peak, since the different conformations result in slightly different distances between the two molecules.

There are two distinct peaks at 5.15 and 10.15 Å, indicating two solvation shells of DMSO around the adducted base. Figure 33A shows a representative structure with for the solvation around B[*a*]P. The colored and solid DMSO molecules are within 5.15 Å of the adduct, whereas the transparent yellow DMSO molecules are within 10.15 Å of the adduct. Additionally, all water molecules within 5.15 Å from the adduct are displayed in transparent red. Our results indicate that there are no water molecules interacting with the top or bottom of the adduct; only a few along the equatorial hydrogens on the side of the ring structure, which is represented in the RDF of water as the first maxima. Taken together, these results suggest that the adduct is microsolvated by DMSO. Our simulation results are consistent with the smFRET results presented above, which indicate that the solubility of the B[*a*]P is increased in 10% DMSO and therefore that a solvent-exposed conformation would be much more favorable.

2.3 Discussion

Exposure to exogenous mutagens is one of the main causes of somatic mutations, some of which eventually lead to cancer. One means for understanding the mutagenic outcome of DNA damage is to visualize how DNA polymerases interact with lesions formed by various carcinogenic DNA adducts. B[*a*]P shows complicated

carcinogenic properties because it forms different isomeric DNA adducts that each shows distinct mutational patterns in DNA. Most research about B[*a*]P carcinogenesis has focused on the more abundant (+)-*trans*-B[*a*]P but we believe that a complete understanding of the mutagenic properties of B[*a*]P requires an in-depth characterization of all isomers resulted from BPDE reaction with DNA bases. Therefore, in this work we have used smFRET, primer extension assays and molecular dynamic simulations to study the mechanistic implications of bypass of a bulky lesion, (+)-*cis*-B[*a*]P-N^2-dG by the model Y-family polymerase Dpo4.

Even though Y-family polymerases are essential for TLS of bulky B[*a*]P-DNA adducts, it is not known if a single Y-family polymerase can bypass all the B[*a*]P stereoisomers [52]. Y-family polymerases have been shown to incorporate both correct and incorrect dNTPs across the B[*a*]P adduct [113]. The ability to successfully synthesize past the adduct also depends on the nature of the lesion [52]. Many studies have shown that primer extension after the adduct is severely inhibited or slowly rescued and may require the action of a second polymerase [52, 114, 115]. For example, Woodgate *et al.* have found evidence suggesting two Y-family polymerases are required for successful B[*a*]P-dA bypass: Pol ι incorporates mostly dT across a B[*a*]P-dA adduct and Pol κ extends past the terminal B[*a*]P-dA:dT base pair [60].

Although the precise mechanism by which bulky adducts inhibit high-fidelity DNA polymerases is not known, there is evidence that these structures may prevent nucleotide incorporation by precluding the large finger-closing movement required for dNTP incorporation [110, 116]. Y-family polymerases do not generally undergo such large rearrangements during their catalytic cycle and also present wider active sites that

can presumably accommodate a bulky adduct structure. Therefore, it is believed that inhibition of Y-family polymerases activity by adducted DNA bases occurs through major alteration of the cognate DNA/Pol complex structure [18]. The most important and non-mutually exclusive distortion mechanisms include stacking of an aromatic moiety within the DNA helix, flipping the adducted base to a *syn* conformation or misalignment/rotation of the primer-template DNA strands. Even though Dpo4 is a versatile polymerase able to bypass a wide variety of DNA damage, the strong roadblock posed by (+)-*cis*-B[*a*]P-N^2-dG suggests that the DNA/Pol structure is severely distorted by the adduct. The structures generated by MD simulations have revealed the stacked and distorted adduct conformations at the primer template junction (Figure 30), as discussed in the results section. Similar to other B[*a*]P adducts; the longer pause occurs during dNTP incorporation opposite the damaged base. No crystal structures exist with a (+)-*cis*-B[*a*]P-N^2-dG adduct as the templating base, but our data suggest that the aromatic moiety of B[*a*]P is present in one of two conformations: either stacked within the DNA helix or in a solvent-exposed, catalytically active conformation (Figure 34). Interestingly, even though different B[*a*]P isomers show distinct error signatures, the DNA synthesis-blocking mechanism seems to share similar features among all B[*a*]P adducts and it implies some degree of B[*a*]P stacking in the DNA helix (Figure 34).

One remarkable feature of Dpo4 synthesis is that it incorporates mostly dG across from a (+)-*cis*-B[*a*]P-N^2-dG adduct, a quite uncommon base insertion, and does not show any tendency to incorporate the correct nucleotide dC. When the dC base is placed opposite B[*a*]P adduct, it flips out from the helix showing an unfavorable

interaction with the adducted base (Figure 30B). This indicates that, although the (+)-*cis*-B[*a*]P-N^2-dG adduct is a minor reaction product, it has the potential to be a strong mutagen with an unusual mutational signature. Moreover, the fact that this mutation was not observed in *E. coli* or COS7 cells [117] suggests that pol κ (the eukaryotic Dpo4 homologue) may not be responsible for synthesis across from the (+)-*cis*-B[*a*]P-N^2-dG adduct. Also, the fact that synthesis past the adduct is accurate may imply that this polymerase takes on that synthetic role. Unlike what is observed using the (+)-*trans*-B[*a*]P-N^2-dG adduct, Dpo4 does not use the downstream templating base for damage bypass with the *cis* adduct. This suggests that, for the (+)-*cis*-B[*a*]P-N^2-dG adduct, a misaligned structure with the adducted base looped-out (and not base-paired) is catalytically inefficient or not favored. Therefore, in this case we conclude that nucleotide incorporation takes place with the adducted dG sitting in the active site, and adopting alternate rotameric and/or tautomeric conformations to enable dG:dG pairing. In agreement with this idea, recent molecular dynamics simulations have shown that dG is the most stable nucleotide that can pair with a (+)-*cis*-B[*a*]P-N^2-dG adduct.[108] It appears that the dG:dG base pair is stabilized by some of the major groove contacts of the purine ring system. Also, the least nucleotide excision repair efficiency is observed when dG is the partner of the (+)-*cis*-B[*a*]P-N^2-dG adduct suggesting that this mispaired structure is the least distorting.

When the primer is extended one nucleotide and the (+)-*cis*-B[*a*]P-N^2-dG adduct sits at the primer-template junction, our data is also consistent with a (+)-*cis*-B[*a*]P-N^2-dG stacked solvent–exposed conformation, which is presumably more catalytically active. In contrast with our prior studies with the aromatic amine-dG adducts, Dpo4

extends past the (+)-*cis*-B[*a*]P-N^2-dG adduct in an error-free manner, suggesting that the solvent-exposed conformation does not dramatically distort DNA at the Dpo4 active site. Shown in Figure 34 is a model that incorporates the various binding conformations we observe into a possible mechanistic pathway that allows Dpo4 to bypass the (+)-*cis*-B[*a*]P-N^2-dG adduct.

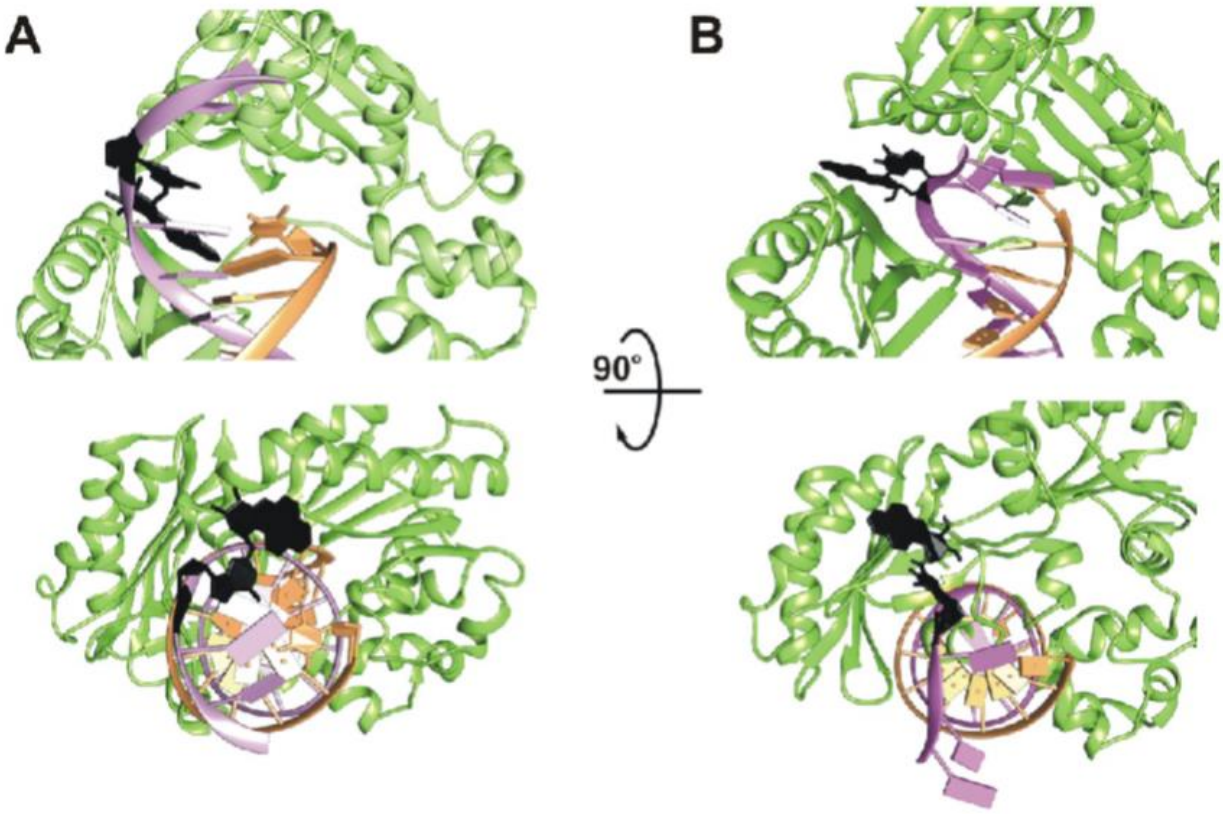


Figure 34. Comparison of the adduct conformation in the presence and absence of DMSO. **(A)** A representative structure of Dpo4-DNA binary complex. The adduct is stacked in the minor groove. **(B)** A representative structure of Dpo4-DNA binary complex in the presence of DMSO. The adduct is flipped outside of the helix and exposed to the solvent.

When Dpo4 binds to the (+)-*cis*-B[*a*]P-N^2-dG-modified DNA, we observe two distinct Dpo4 binding conformations that are in dynamic equilibrium (insertion binary complex I

and preinsertion binary complex I, Figure 35). Dpo4 prefers to form the preinsertion complex I, where the terminal base pair occupies the dNTP binding site,[82] but we also observe a substantial fraction of the complex in the insertion I conformation. This latter structure places the (+)-*cis*-B[*a*]P-N^2-dG in the templating position so that the B[*a*]P is in the active site blocking incorporation (Figure 35), as evidenced by our primer extension results (Figure 20).

Previously, it has been observed that the use of organic solvents could enhance the bypassing of the B[*a*]P adduct by Dpo4 [16]. The proposed reasoning for this observation was that hydrophobic properties of the organic solvent could stabilize the solvent exposed pyrene ring system. Since Dpo4 active site is open to the solvent, it is possible these hydrophobic molecules directly interact with the adduct within the active site. The cytosolic hydrophobicity is much lower than that of the aqueous buffers used for the *in vitro* experiments [109]. Adding DMSO can increase the hydrophobicity of the solution mimicking the micro environmental polarity of the cytosol and the local environment in a replication complex.

In the presence of DMSO, smFRET experiments provide evidence for the existence of two new binding conformations that we label preinsertion binary complex II and insertion binary complex II, both of which appear to have the (+)-*cis*-B[*a*]P-N^2-dG adduct flipped out in a solvent exposed orientation (Figure 35). These two conformations are also in equilibrium, with insertion binary complex II being most stable. Because this structure has neither the terminal nucleotide nor the B[*a*]P adduct in the active site, this complex is able to incorporate a nucleotide across from the modified base, but because the adduct alters the templating base orientation, incorporation is

highly inaccurate. MD simulations for the binary complex in the presence of 10% DMSO also shows the formation of the solvent exposed conformation of the (+)-*cis*-B[*a*]P-N^2-dG adduct in the primer-template junction (Figure 35). Furthermore, the solvent exposed B[*a*]P adduct seems to be sandwiched between the LF and finger domain of the Dpo4 (Figure 35).

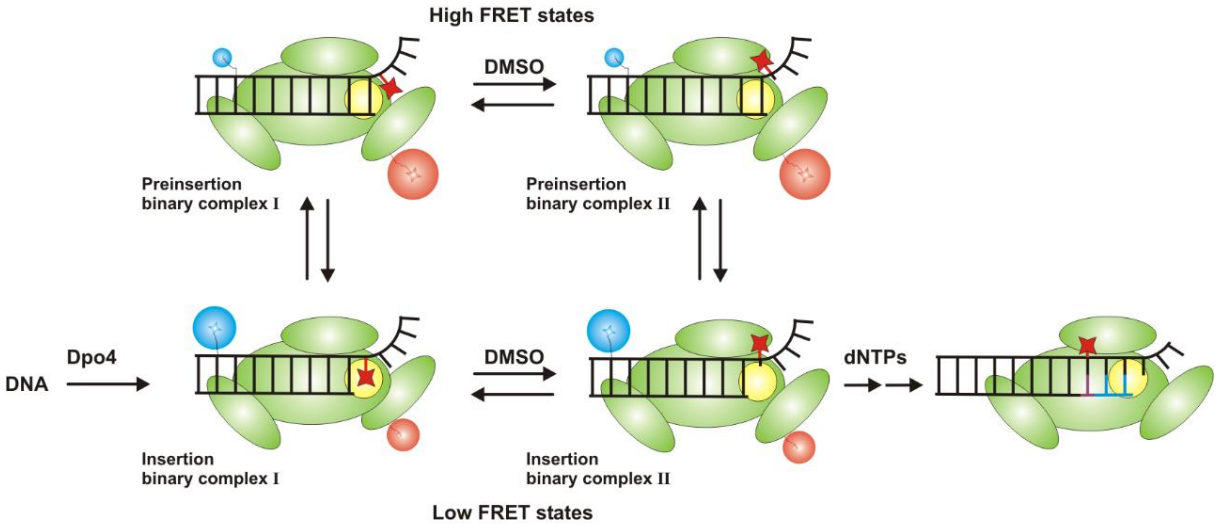


Figure 35. Proposed model shows different Dpo4 binding conformations on (+)-*cis*-B[*a*]P-N^2-dG (red star) modified DNA and possible mechanistic implications that permit Dpo4 for bypassing the adduct. When Cy3 (blue) labelled DNA binds to Cy5 (red) labelled Dpo4 (green), it forms the insertion binary complex I where the adduct locates Dpo4 active site (yellow) in the intercalated orientation. The adduct bypassing is prohibited in this conformation. At the same time, Dpo4 samples high FRET state, preinsertion binary complex I where the terminal base of the primer is located in the Dpo4 active site which is again unfavorable conformation for primer extension. We found that in the presence of DMSO, Dpo4 forms a binding conformation which permits Dpo4 to bypass the adduct forming insertion binary complex II. The adduct is in the exposed conformation and Dpo4 active site is opened for dNTP incorporation. Even with DMSO, Dpo4 is equilibrated between two FRET states forming another high FRET conformation, preinsertion complex II. In the absence of DMSO, preinsertion binary complex I generates a higher population whereas insertion binary complex II forms a higher population with DMSO. Formation of insertion binary complex II can follow either pathway. Upon addition of dNTPs, the primer extension can occur from insertion binary complex II. The G:G mismatch bond is shown in red line and newly formed complementary bases are shown in blue lines.

Using the same model shown in Figure 35 to interpret the analogous Dpo4 binding conformations for the DNA construct where the (+)-*cis*-B[*a*]P-N^2-dG adduct is located at the primer terminus with a dC partner base, we observe both insertion and preinsertion complex I, with the preinsertion complex predominating. The addition of DMSO leads to the predominant formation of insertion binary complex II, presumably because the B[*a*]P adduct is preventing translocation to the preinsertion complex.

CHAPTER 4: THE DPO4 BINDING CONFORMATIONS DURING BYPASSING THE (+)-trans-B[*a*]P-*N*^{2}-dG ADDUCT.

Some sections of this chapter were taken from the manuscript by Pramodha Liyanage *et al.* “Bypass of bulky lesions requires Dpo4 to bind in a productive conformation”

4 Background and significance

The (+)-*trans*-B[*a*]P-*N*^{2}-dG adduct is known to reside in the minor groove of the B-DNA whereas (+)-*cis*-B[*a*]P-*N*^{2}-dG adduct is intercalated in the DNA helix. These conformational differences within the DNA explain their different carcinogenic properties. Previous studies have shown that Dpo4 can bypass the (+)-*trans*-B[*a*]P-*N*^{2}-dG adduct less efficiently and, it is highly mutagenic causing all three base substitution mutations [118]. The crystal structure has been resolved for the Dpo4 and (+)-*trans*-B[*a*]P-*N*^{2}-dG modified DNA binary complex where the DNA primer terminates across the adduct (Figure 36) [18]. Two types of crystal structures were obtained with different adduct orientations. One of them has a pyrene ring system intercalated into the DNA double helix which displaces the modified dG into the minor groove (Figure 36A). This adduct conformation has disrupted the base pairing at the primer-template junction. In contrast to stacked B[*a*]P orientation, the second crystal form shows that the modified dG is flipped out and the B[*a*]P adduct is buried in the gap between the LF and core domains producing a more open active site for nucleotide incorporation (Figure 36B). The adduct flipped-out orientation may reduce the structural distortion at the primer-template junction.

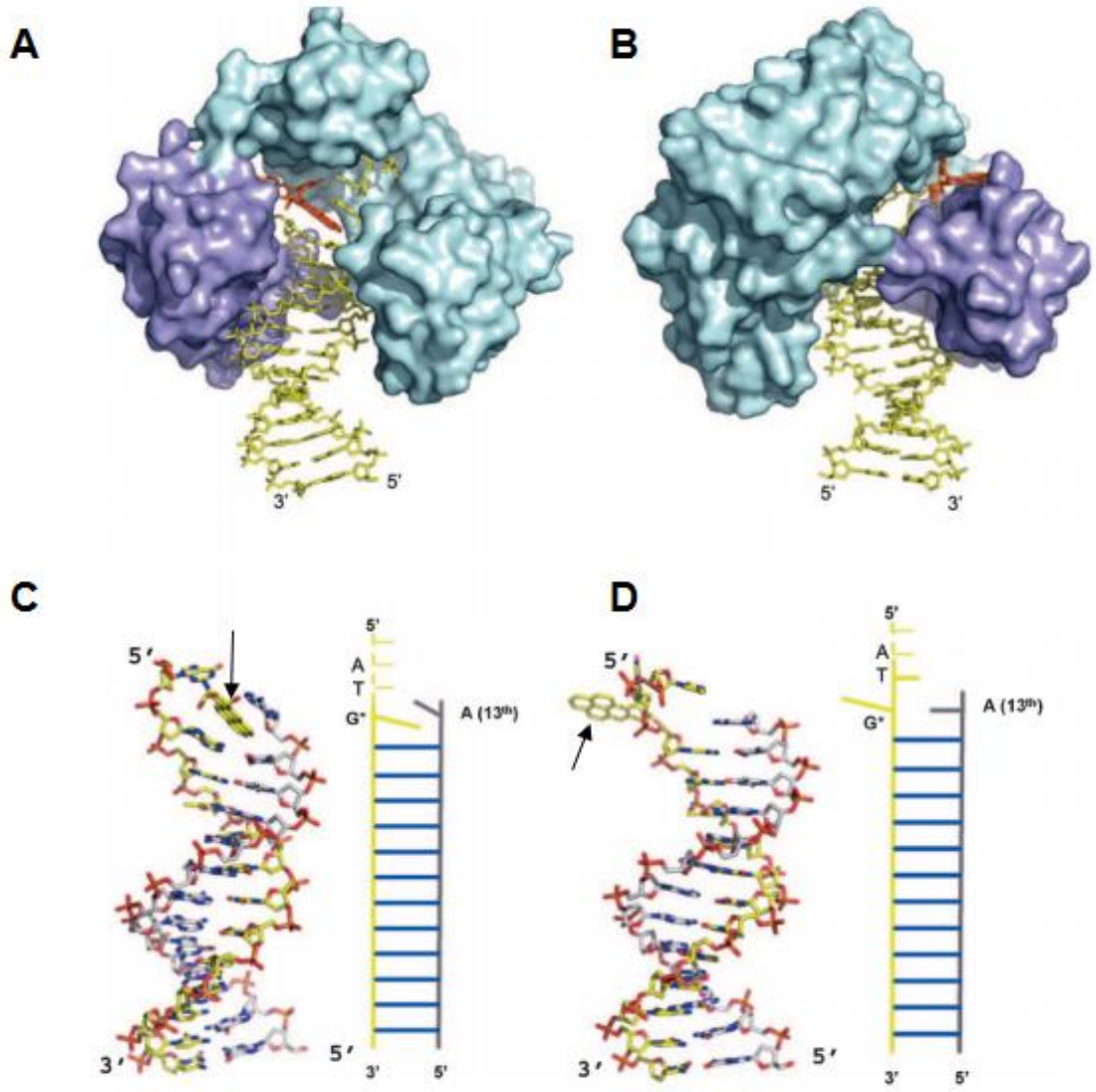


Figure 36. Dpo4 crystal structure of the (+)-*trans*-B[*a*]P-N^2-dG modified DNA bound Dpo4. **(A)** The adduct (Red) is in the intercalated conformation within the Dpo4 active site. **(B)** The adduct's solvent exposed conformation in the Dpo4 where it is sandwiched between the little finger and core domains. **(C)** A schematic representation of the intercalated conformation of the adduct without polymerase. The black arrow indicates the position of the (+)-*trans*-B[*a*]P-N^2-dG adduct. **(D)** A schematic representation of the exposed conformation of the adduct without polymerase. The black arrow indicates the position of the (+)-*trans*-B[*a*]P-N^2-dG adduct. This figure was adapted form [18].

When the second crystal was co-crystallized with the dTTP in the ternary complex, they observed that a looped out modified dG and misaligned the primer-template, leading to successful DNA synthesis (Figure 37) [18]. This study provided the basic structural information for a -1-frameshift mutation and template misalignment.

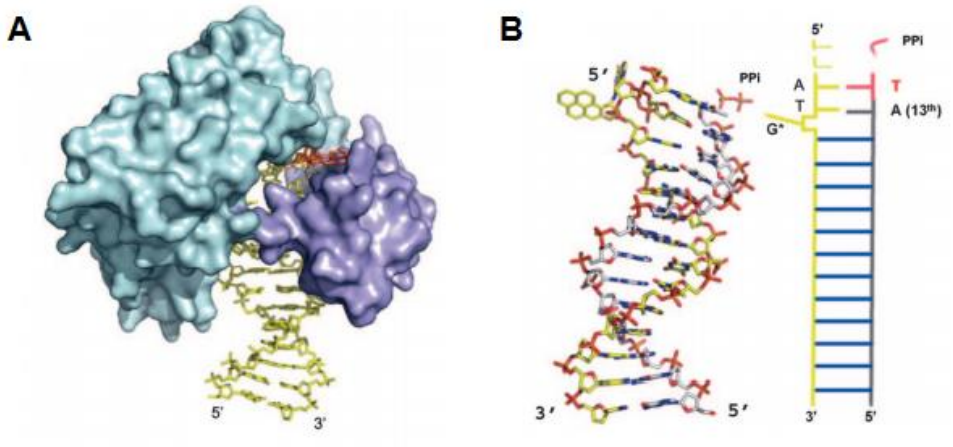


Figure 37. Ternary complex crystal structure with Dpo4 and (+)-*trans*-B[*a*]P-N^2-dG modified DNA. **(A)** The adduct is placed in the structural gap of the Dpo4 little finger and core domain. **(B)** A schematic representation of the exposed conformation of the adduct in the ternary complex (The structure of the Dpo4 is not shown). The modified dG is looped out and the terminal base of the primer (dA) base pair with dT next to the modified dG. The DNA sequence supports for this base pairing. The incoming dT binds to the next templating base, dA. This figure was adapted form [18].

However, the Dpo4 global conformation is nearly identical in both crystal types observed in Figure 36 A and B. Therefore, we cannot get any information about Dpo4 binding conformations and relative movements out of these static crystal structure. Here we use smFRET, to obtain the Dpo4 binding dynamics to the (+)-*trans*-B[*a*]P-N^2-dG modified DNA in real time. We monitor the Dpo4 binding conformations before and after the B[*a*]P adduct using different DNA primers.

Our data show that Dpo4 binds to the B[*a*]P modified DNA in two different conformations in the binary complex. The predominant conformation is in equilibrium with the nonproductive, less stable conformation. In the Dpo4 predominant

conformation, the B[*a*]P adduct may exist in two different positions in the DNA. The B[*a*]P intercalated conformation blocks the dNTP incorporation while he solvent exposed adduct conformation facilitates error prone DNA synthesis past the B[*a*]P adduct. Furthermore, our data supports the previously discovered Dpo4 5' replication slippage and the-1 frameshift mutation mechanism.

4.1 Results

4.1.1 B[*a*]P adduct induced Y-family polymerase stalling slowly rescued by Dpo4

The previous studies show that the (+)-*trans*-B[*a*]P-N^2-dG adduct is a strong block for high fidelity DNA polymerases [119, 120]. Therefore, Y-family polymerases must be involved in processing this adduct *in vivo*. First, we did the running start primer extension assay to study dNTP incorporation across and past the adduct by Dpo4 in our sequence. Both unmodified and B[*a*]P modified 26mer DNA template was used with a 16mer primer. Dpo4 could extend the primer very efficiently with unmodified template, and 100% extended product appeared within one minute (Figure 38B). The (+)-*trans*-B[*a*]P-*N^2*-dG adduct is positioned in the 21^{st} position on the template. Dpo4 could extend the B[*a*]P modified template-primer up to the 20^{th} base efficiently. But Dpo4 showed a strong pause one base before the adduct position. The blocked replisome slowly recovered over time, and the fully extended primer started to appear after 30 min (Figure 38C). However, only 22% bypass product was generated at 120 mins. Incorporation of dNTP across the adduct seem to be more difficult for Dpo4 than replication after adduct.

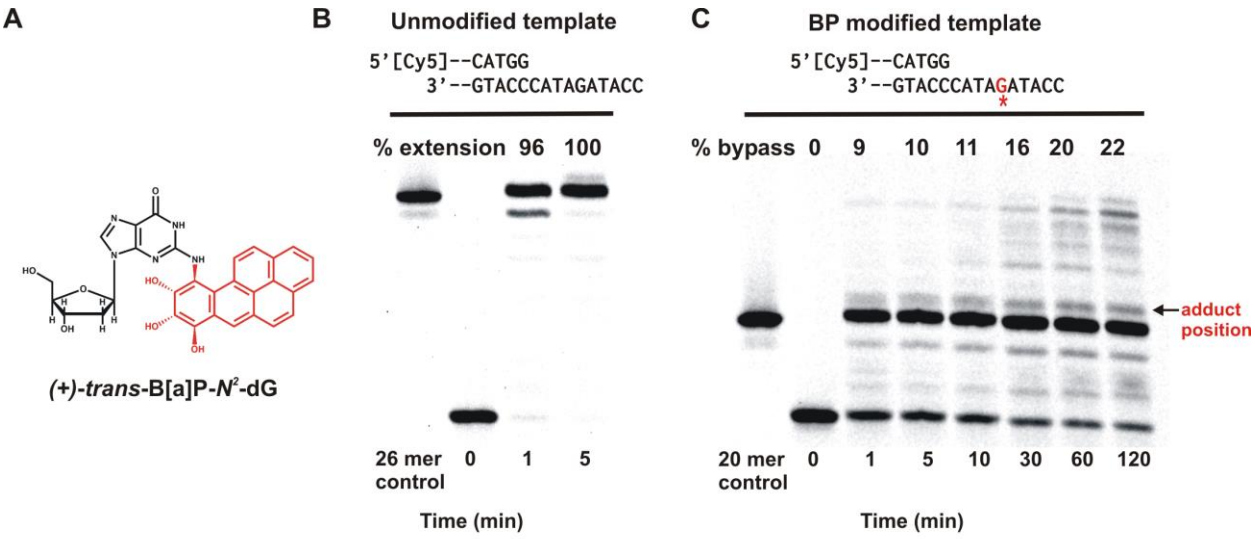


Figure 38. Dpo4 slowly extend the primer after carcinogenic adduct. Dpo4 slowly extend the primer after carcinogenic adduct. (**A**) Chemical structure of the (+)-trans-B[a]P-N^2-dG carcinogenic adduct. **(B)** Running start primer extension assay for unmodified DNA construct. The 16 mer primer with 5'-Cy5 labelled is hybridized to unmodified DNA (26 mer). A portion of the DNA sequence is shown on the top. The primer extension assays were initiated by adding Dpo4 (20 nM) to the reaction and at 37°C. The 0^{th} time refers to the reaction right before adding Dpo4. The percent extension is reported above each band with respect to the intensity of the 0^{th} time band. **(C)** The primer extension assay for the (+)-trans-B[a]P-N^2-dG modified template. This assay was carried out under the same conditions as unmodified template. The modified DNA primer-template is shown on the top of the gel. The red "G" indicates the position of the adduct on the template which is located at 5 nucleotides away from the primer terminus. A strong pause appeared at the 20^{th} position during DNA synthesis.

4.1.2 B[*a*]P adduct effects on Dpo4 binding conformation

To study the Dpo4-DNA binding conformations and related biophysical phenomena, we designed the smFRET experiment similar to (+)-*cis*-B[*a*]P-N^2-dG adduct that we discussed in the chapter 3. According to the experimental set up, the relative positions of the fluorophores will be changed depending on the binding, translocation and corresponding conformational changes of the Dpo4 (Figure 39A). The recoded fluorescence intensity fluctuation, and the corresponding FRET changes are shown in Figure 39B, left.

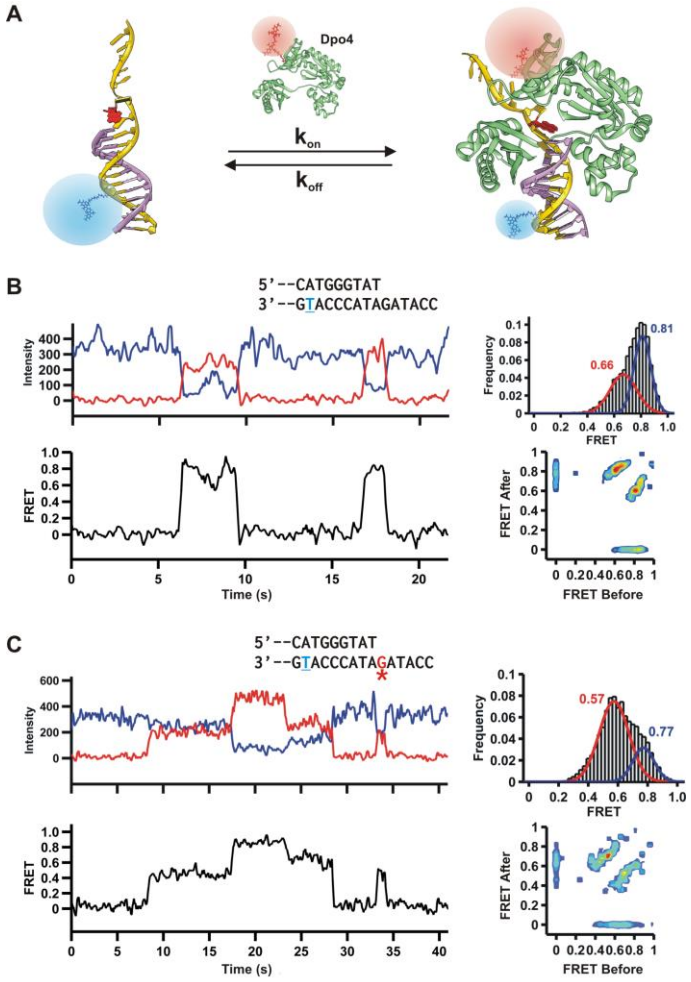


Figure 39. Experimental set up to study DNA-Dpo4 interactions. **(A)** B[a]P modified, Cy3 labeled DNA template is hybridized to a 3' biotinylated primer. The DNA construct is immobilized to a quartz slide by a biotin-streptavidin linkage. The Cy5 labelled Dpo4 is free floating in the buffer. When Dpo4 binds to the DNA, FRET occurs between fluorophores. **(B)** Dpo4 binding conformation on the unmodified template. A part of the unmodified DNA construct is shown on the top of the figure. The representative fluorescence donor (blue) and accepter (red) intensity trace (top) and the corresponding FRET trace is shown (below). The smFRET histogram shows the Dpo4 binding conformations to unmodified template. The data is fitted to two Gaussian functions. Below the histogram, the transition density plot (TDP) is given for Dpo4 binding to the unmodified template. **(C)** A part of the B[a]P modified DNA construct is shown on the top. The representative intensity trace and the FRET trace are for the modified template (left). Sm FRET histogram and TDP for modified template is shown on the right.

First, we used the 8Cy3 (the Cy3 label is 8 nucleotides away for the primer template junction) unmodified DNA construct to study the effect of the Dpo4 binding conformation. The binding of Dpo4 to the unmodified template generated two distinct peaks similar to a previously published study [82]. The histogram in Figure 38B shows the two FRET peaks which are centered at FRET 0.81 and FRET 0.66, presumably two different Dpo4 binding orientations in DNA. The histogram was best fitted for two Gaussian distributions that show two asymmetric sub-populations. The prominent high FRET peak revels that the Dpo4 binds to DNA in a conformation which has a greater preference over the other low FRET state conformation. The corresponding transition density plots further confirm the occurrence of the prominent high FRET conformation and their transition between the two FRET stets (Figure 38B, bottom right). The appearance of two FRET states with the interaction of Dpo4 and DNA and the analogues Dpo4 binding orientations were thoroughly explained in *Brenlla et al.* [82, 83]. The high FRET and low FRET Dpo4 binding conformations were assigned as the preinsertion binary complex (PBC) and incretion binary complex (IBC), respectively [82]. In the PBC, the terminal base pair occupies at the Dpo4 active site that blocks the dNTP incorporation. Therefore, the PBC seems to be a nonproductive binary complex whereas the IBC is considered as a productive binary complex where the templating base occupies at the Dpo4 active site. The next nucleotide can be incorporated across the templating base in the IBC.

The smFRET experiments with the (+)-*trans*-B[*a*]P-N^2-dG modified template also show two distinct FRET states indicating two possible Dpo4 binding conformations (Figure 39C). However, they are broader and slightly lower compared to the FRET

values observed for the unmodified template, indicating that Dpo4 binding is altered by the modified DNA at the adduct site. The most interesting feature is their population distribution. In the modified template, the lower FRET peak is significantly higher and broader than that of the high FRET peak. This suggests that Dpo4 prefers to bind to the low FRET conformation.

4.1.3 Dpo4 ternary complex binding conformations with dC and dT are similar in both unmodified and B[*a*]P modified templates, respectively.

It is commonly observed among classical DNA polymerases that they stall just before the B[*a*]P adduct. Some polymerases are able to incorporate dNTPs across the adduct in a mostly error prone or rarely error free manner, but further extension is strongly inhibited [52]. It has been seen previously that the nucleotide selectivity of Dpo4 depends on the DNA sequence while bypassing the B[*a*]P adduct [16, 18]. Therefore, we carried out a single nucleotide incorporation assay to visualize the dNTP selectivity in our sequence. First, we used a 20mer/26mer unmodified DNA primer-template to study the dNTP selectivity. Dpo4 synthesizes the unmodified DNA in the presence of all four nucleotides very efficiently (Figure 40A, left). The fully extended 26mer product appeared in the gel within a minute. In the presence of only a single nucleotide, Dpo4 incorporated dC, the next correct base across the unmodified dG. Although a small amount of dT incorporation can be seen in the gel, dA and dG were not incorporated at all under the reaction conditions.

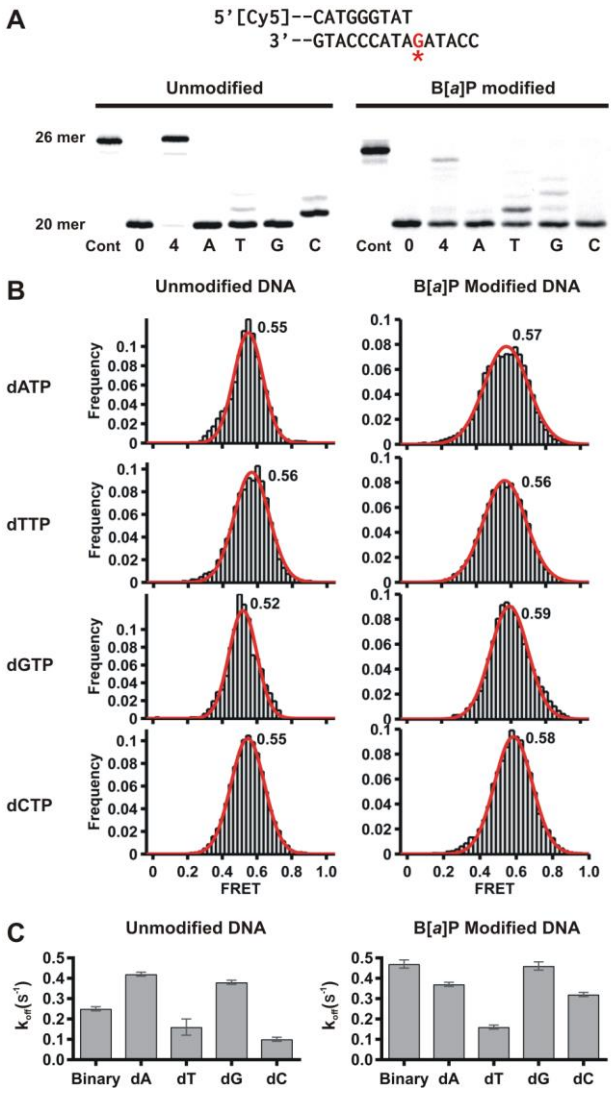


Figure 40. Ternary complex Dpo4 binding conformations with the modified dG as the templating base. **(A)** The B[*a*]P modified DNA sequence used for the single nucleotide incorporation assay is shown on the top. The lane 1 and 7 depicts the 26 mer control DNA band. The lane 2 and 8 are control experiments with all four dNTPs. The other lanes corresponding to the single nucleotides as indicated. The Dpo4 concentration is 10 nM for the both modified and unmodified experiments. However, the reaction time for the unmodified template is 1 min, and for the modified template is 20 min. the next correct nucleotide, dC is incorporated in the unmodified template. Dpo4 predominantly incorporated dT, which is the complementary base next to the modified dG. **(B)** Single molecule FRET histograms for the ternary complex Dpo4 binding conformations. The histograms in the left and right are for the unmodified and modified DNA respectively. **(C)** The Dpo4 dissociations rate constant for the binary complex and ternary complex conformations are plotted for the unmodified template (left) and modified template (right).

However, the single nucleotide incorporation assay with a longer reaction time and a high concentration of Dpo4 shows the incorporation of all non-cognate dNTPs in multiple times [111, 121]. Nucleotide incorporation by Dpo4 in the (+)-*trans*-B[*a*]P-N^2-dG modified template strongly depends on the DNA sequence [18]. Previous studies have revealed that Dpo4 tends to incorporate the dNTP across the adduct not complimentary to the modified dG but to the base 5' to the B[*a*]P modified dG [18].In agreement with this, Figure 40A (right) shows Dpo4 incorporated dT complimentary to the dA which is at 5'to the modified dG. The reaction with all four dNTPs (Figure 40A-right) showed the fully extended product, but the intensity and the migration of the band are comparatively lower than that of the unmodified DNA. The observation may indicate the inefficient activity of Dpo4 and its single nucleotide shorter product in the presence of a carcinogenic adduct, revealing the compatibility of this out come with the previously published data [18].

Then we conducted the smFRET experiments with individual nucleotides to learn the significance of Dpo4 binding conformations. These experiments were carried out in the presence of Ca^{2+} to prevent the dNTP incorporation. Unlike the DNA-Dpo4 binary complex, we observed a single FRET state for the Dpo4-DNA-dNTP ternary complex similar to a previously published study (Figure 40B) [82]. The single FRET state indicates that Dpo4 forms a unique conformation in the ternary complex. The smFRET histogram with unmodified template in the presence of dCTP shows a peak centered at FRET 0.55. Interestingly, Dpo4 showed a similar FRET value for the B[*a*]P modified template in the presence of dTTP, which is the preferred base across the adduct (Figure 40B). The stability of the Dpo4-DNA interaction in both binary and ternary complexes

can be measured by the polymerase dissociation rate constant (k_{off}). We performed the dwell time analysis to calculate k_{off} values of the ternary complex for all the smFRET experiments, and the calculated k_{off} values for both unmodified and modified templates are plotted in Figure 39C. The most stable conformation was represented by the lowest dissociation rate constant. The lowest k_{off} value, 0.10 ± 0.01 was obtained for the ternary complex with dC (Figure 40C, right). Interestingly, that is the correct nucleotide incorporated by Dpo4 in the unmodified template (Figure 40A, right). Dpo4 forms a relatively stable binding conformation in the ternary complex with dC compared to the binary complex (k_{off} = 0.25 ± 0.01). All the non-cognate dNTPs received higher k_{off} values compared to the dC indicating a relative binding instability in the ternary complex (Figure 40C, right). The dwell time histograms for the modified DNA show the lowest k_{off} value for the dT, which is the preferred nucleotide across the adduct (Figure 40C, right). The other nucleotides show relatively higher k_{off} values indicating the formation of less stable ternary complexes. In addition, an smFRET experiment with B[*a*]P modified template shows k_{off} 0.47± 0.02 in the binary complex whereas unmodified template gives k_{off} 0.25 ± 0.01. The structural constrain imposed by the bulky adduct at the modified DNA causes relative instability in the binary complex, giving a higher k_{off} value.

4.1.4 Dpo4 binds to the B[*a*]P modified dG:dT mis-paired DNA similar to the unmodified DNA

Our next aim was to the study the Dpo4 binding conformations when the primer terminates across the adduct in real time. To do that, we used a 21mer DNA primer with a terminal dC, which is the complimentary base for the dG in the template. In this DNA construct, since the Cy3 label is located 9 nucleotides away from the primer terminus, it

is called the 9Cy3. Dpo4 showed two distinct FRET states (FRET = 0.55, 0.74) for the unmodified, 9Cy3, binary complex which are lower than that of the unmodified, 8Cy3 binary complex (Figure 41). Since Dpo4 binds to a primer that is one base longer in the 9Cy3 DNA construct, the relative positions of the fluorophores would be longer, giving smaller FRET values. Surprisingly, the B[*a*]P modified-9Cy3 binary complex generated a FRET distribution similar to the B[*a*]P modified-8Cy3 binary complex (Figure 41).

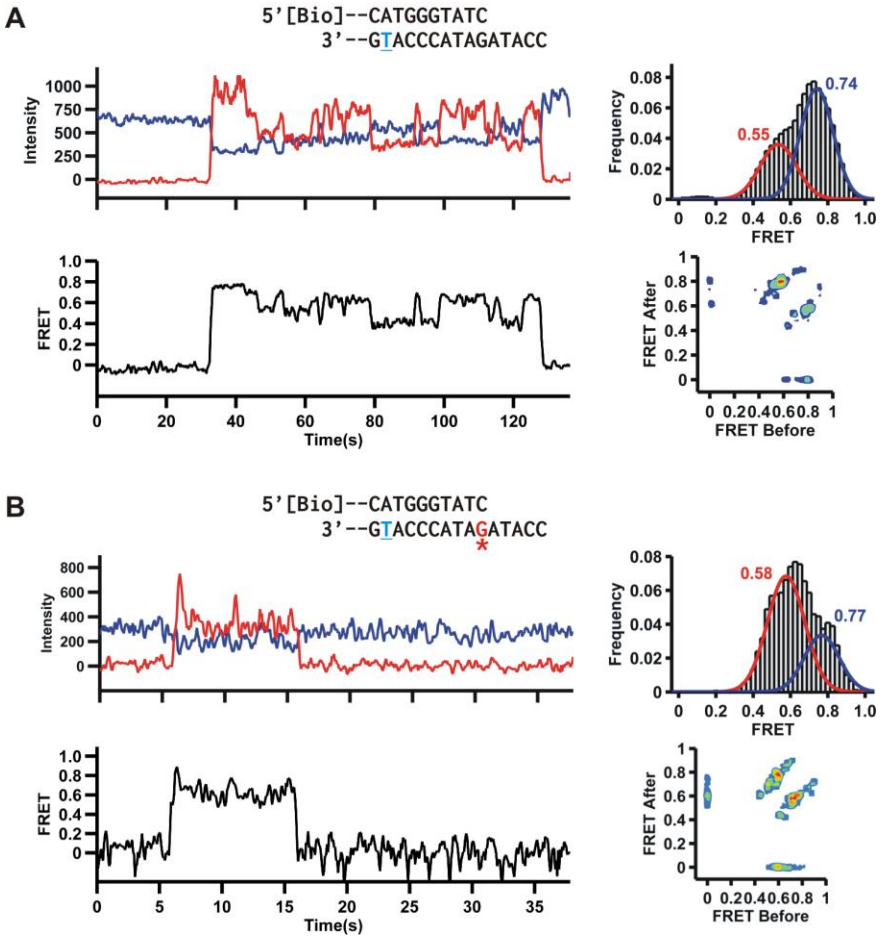


Figure 41. Binary complex Dpo4 binding conformation of the 9Cy3 DNA construct. (A) The characteristic traces, histogram and TDP for unmodified template. (B) The characteristic traces, histogram and TDP for B[*a*]P modified template.

One possible explanation for this observation is that the modified dG*:dC base pair might create a massive destabilization at the DNA primer template junction which would prevent the translocation of the Dpo4 on the DNA. On the other hand, Dpo4 naturally incorporates dT instead of dC across the B[*a*]P adduct (Figure 40A). Further, the primer extension by Dpo4 was strongly inhibited for the 9Cy3 primer with terminal dC (Figure 42).

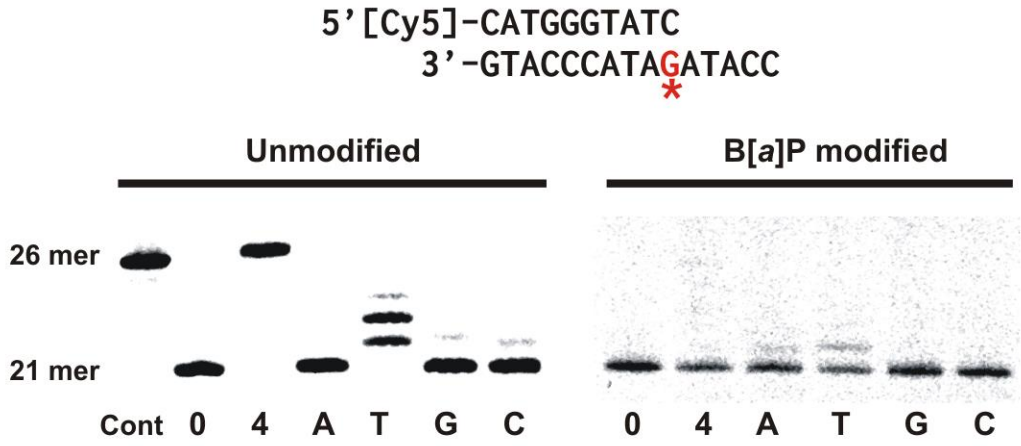


Figure 42. Single nucleotide incorporation in the 9Cy3 DNA construct. The DNA primer terminates with the correct dNTP, dC (Sequence). Dpo4 incorporates the correct dT across the unmodified dG (lift). The nucleotide incorporation is strongly inhibited in the midwifed template.

Therefore, to study Dpo4's nucleotide incorporation ability after the adduct position, we used a 21mer premier with a terminal dT (Figure 43). However, there is a G:T mismatch at the primer-template junction in this DNA construct (named as 9Cy3mm). The unmodified 9Cy3mm binary complex shows a typical two FRET efficiency distribution (FRET = 0.52 and 0.71, Figure 41A). The decreased FRET values are consistent with the movement of Dpo4 from 8Cy3 to 9Cy3 DNA. The observed FRET efficiency values for the B[a]P modified 9Cy3mm binary complex is very similar to

the unmodified DNA (FRET=0.55 and 0.71, Figure 43B). However, the FRET distribution is rather different when comparing the unmodified and modified binary complexes. Dpo4 prefers to sample the high FRET (0.71) preinsertion conformation in the unmodified template possibly due to the fact that this DNA construct contains a nucleotide mismatch in the primer- template junction. However, Dpo4 binds to both FRET states almost equally in the B[*a*]P modified 9Cy3mm binary complex. We used Hidden Markov Model (HMM) to check the transition between these two states. The corresponding transition density plot illustrates the Dpo4 shuttle between the two states (Figure 43).

The dwell time analysis of the 9Cy3mm binary complexes generated the dissociation rate constant for both templates (Figure 44). We obtained a relatively lower k_{off} value for the B[*a*]P modified template (k_{off} = 0.52±0.03) compared to the unmodified template (k_{off} =0.64±0.04). The nucleotide mismatch at the primer-template junction may destabilize the Dpo4-DNA binary complex conformation in the unmodified template.

4.1.5 The next correct nucleotide incorporation after the adduct gives the greatest stability in the ternary complex.

Next, we wanted to study Dpo4's dNTP incorporation ability and corresponding binding conformations in the ternary complex for the 9Cy3mm DNA construct. The fully extended primer appeared in less than 5 minutes for the 9Cy3mm unmodified template (Figure 45A). The single nucleotide incorporation gel picture reveals the selectivity of the nucleotides during the DNA synthesis.

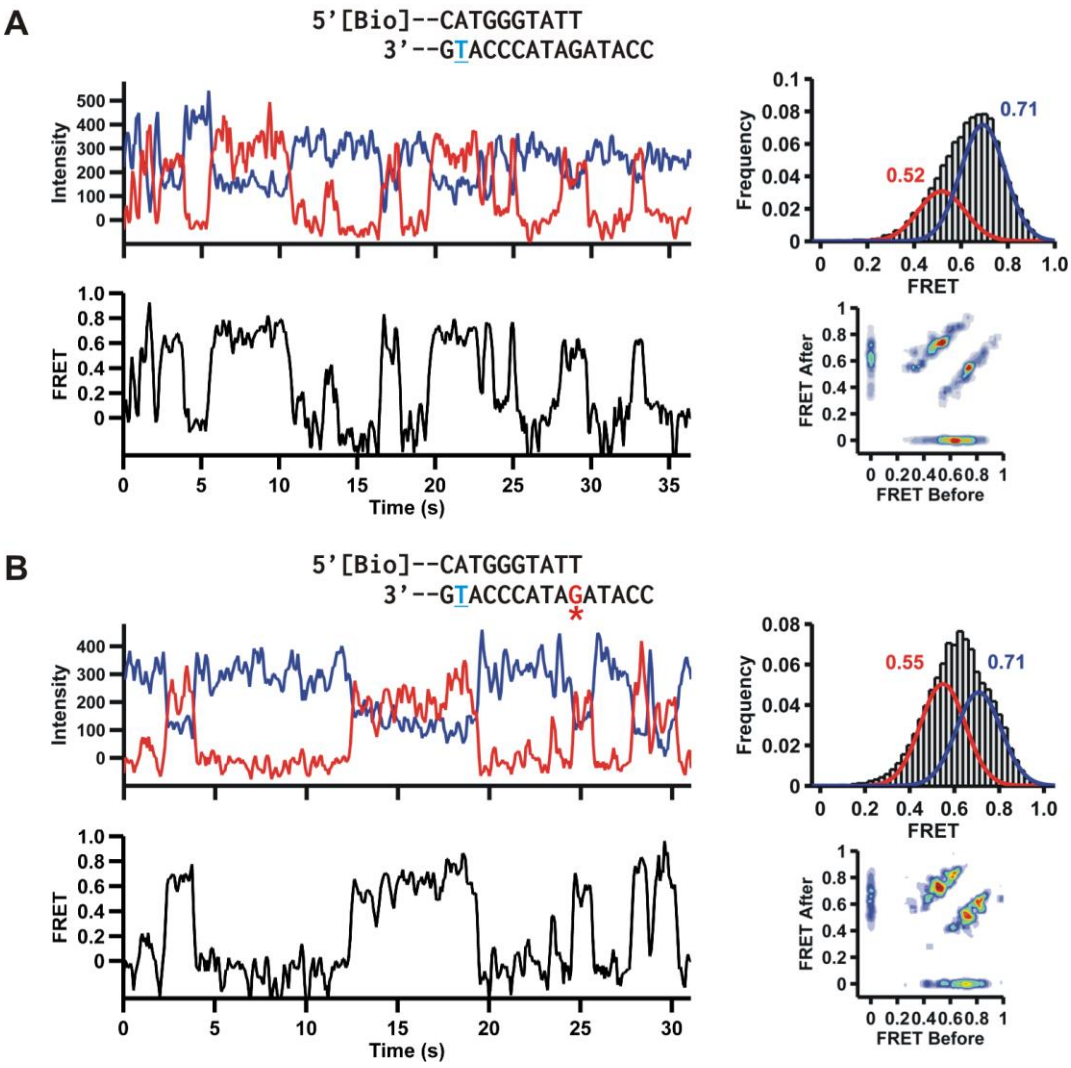


Figure 43. The smFRET result for the Dpo4 binding dynamics when the primer terminates across the adduct. The smFRET result for the Dpo4 binding dynamics when the primer terminates across the adduct. **(A)** The DNA sequence (top) is for the 9 Cy3 unmodified DNA construct. The characteristic fluorescence intensity trace and the corresponding FRET trace are shown on the left. The resulting FRET histogram is best fit for double Gaussians (top right). The TDP (bottom right) shows the dynamic Dpo4 movements between the two FRET states shown in the histogram. **(B)** The 9Cy3 B[a]P modified DNA sequence (top) used for the Dpo4-binary complex smFRET experiments. The intensity trace, FRET trace, histogram and the TDP are shown.

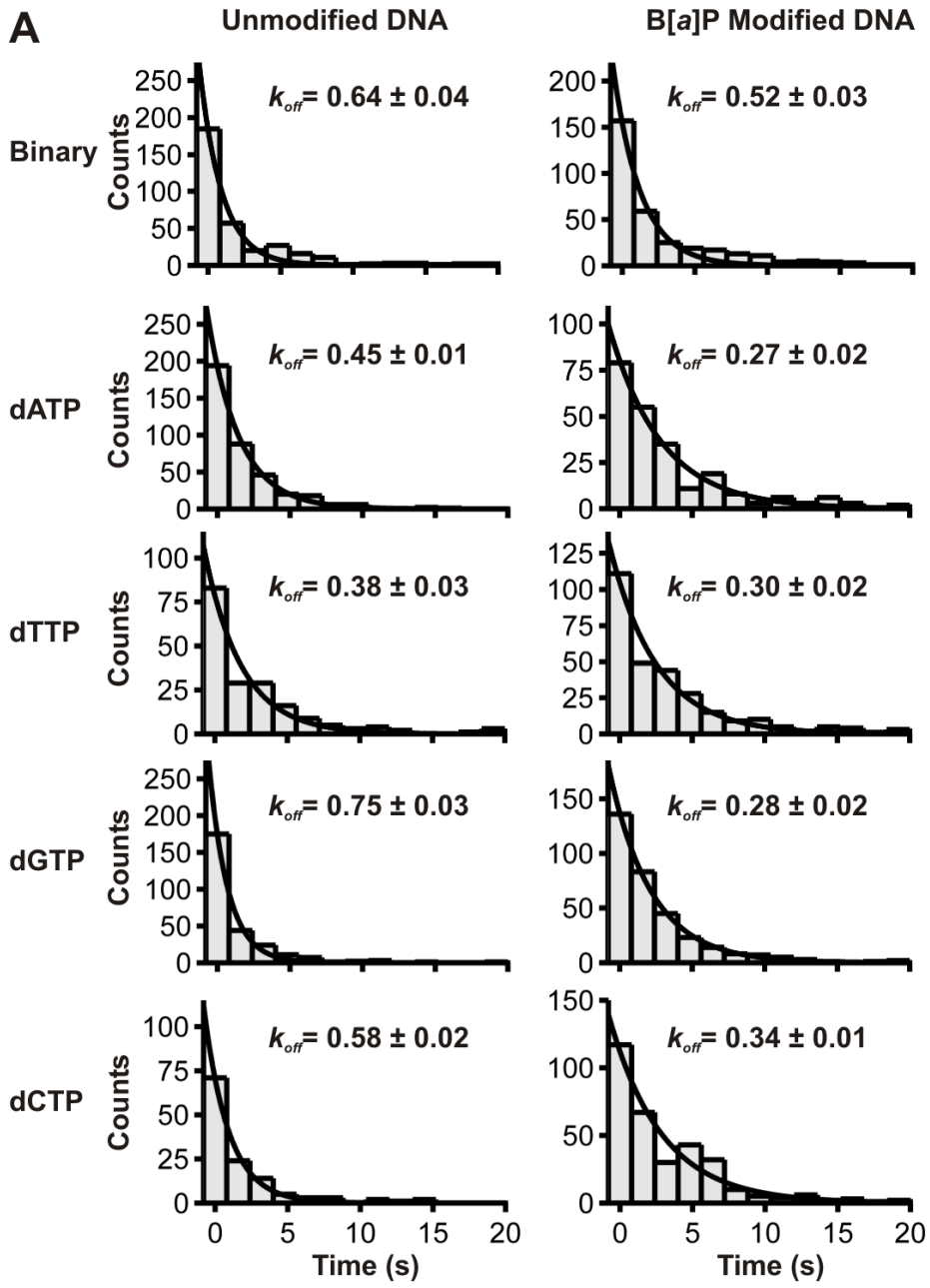


Figure 44. The dwell time analysis of the 9Cy3mm DNA construct.

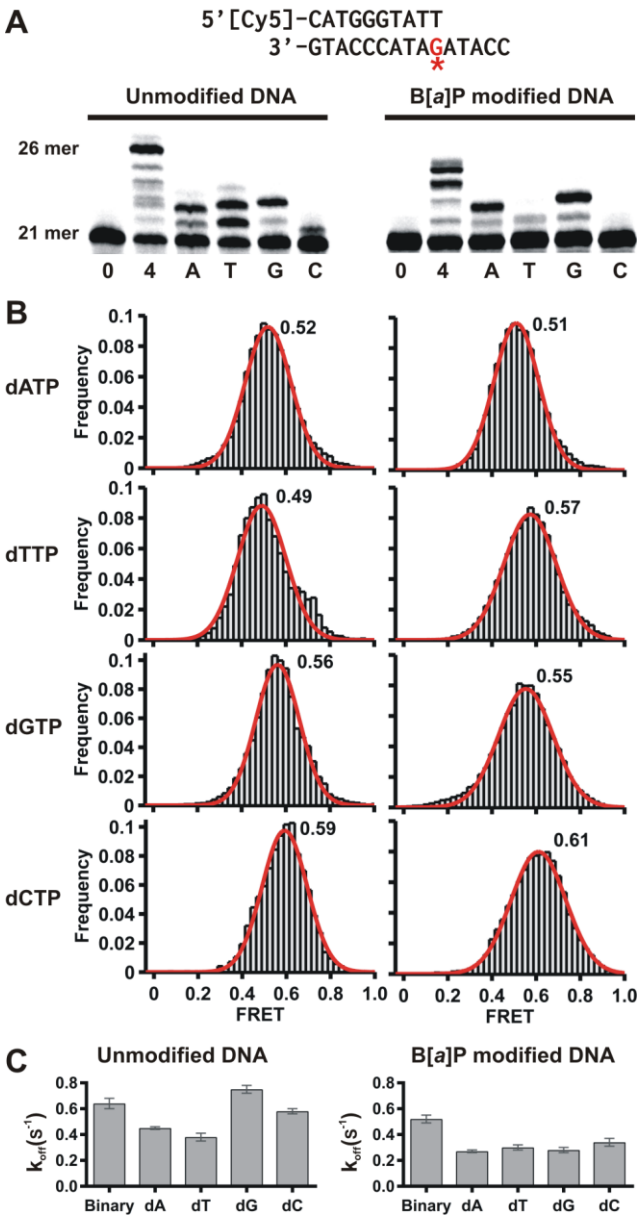


Figure 45. The Dpo4 binding conformations for the G:T mismatch DNA construct (9Cy3mm). **(A)** The DNA sequence shown on the top used for the single nucleotide incorporation assay. The lanes 1and 7 are for the control experiment without dNTPs. The control experiment with all four dNTPs shows the lane 2 and 8. The other lanes for the experiments with a single nucleotide as indicated. **(B)** the FRET histograms for Dpo4 binding conformations in the 9Cy3mm DNA construct. The histograms were best fit for a single Gaussian disturbing. The histograms in the left column are for the unmodified template whereas the histograms in the right column for the B[a]P modified DNA. **(C)** The Dpo4 dissociation rates constant for the binary complex and ternary complex are shown in the graphs.

It shows that Dpo4 incorporates all the non-cognates in the unmodified DNA, which is consistent with its low processivity and fidelity (Figure 45A). The B[*a*]P modified 9Cy3mm DNA has shown the extension of the primer to be one nucleotide shorter than the fully extended product. This result indicates that the terminal dT in the primer may base pair with the dA 5'on the template next to the lesion while looping out the modified dG with B[*a*]P adduct.

We performed the smFRET experiments to evaluate the effect of single nucleotides in Dpo4 binding to the 9Cy3mm ternary complex (Figure 45B). Consistent with previous results, Dpo4 binds to both DNA constructs with a single conformation generating one FRET distribution. Though the gel picture shows the incorporation of non-cognate nucleotides in the unmodified template, the smFRET histograms show a clear discrimination of Dpo4 binding conformation with the different nucleotides. The lowest FRET distribution (FRET = 0.49) is given for the Dpo4 ternary complex with dT which is the next correct dNTP. The highest FRET value (FRET = 0.59) was obtained for the ternary complex with dC, which is not incorporated according to the gel picture. The other non-cognate nucleotides, dA and dG, were given intermediated FRET values, (FRET = 0.52, 0.56 respectively) and they have been incorporated in the unmodified DNA. The low FRET insertion binary complex is believed to be the productive conformation that leads for successful DNA synthesis. We observe throughout our smFRET experiments that the lowest FRET state is given for the next correct dNTP in the ternary complex. Furthermore, the dwell time analysis of the ternary complex unmodified DNA gives the lowest k_{off} value (0.38±0.03) for the dT, indicating the

formation of the most stable ternary conformation compared to the other dNTPs (Figure 43C, left).

The smFRET experiments were carried out under the same conditions as the unmodified DNA to determine the Dpo4 binding conformations in the 9Cy3mm, B[*a*]P modified ternary complex (Figure 45B, left). The lowest FRET value was obtained for the Dpo4-DNA-dA ternary complex. Probably the base, dA, is the one being incorporated in the misaligned, modified DNA construct. Moreover, the lowest k_{off} value also obtained for the dA containing ternary complex (Figure 45C, right). The overall result confirms that the Dpo4 ternary complex is favorable for the dA incorporation.

4.1.6 Discussion

The (+)-*trans*-B[*a*]P-N^2-dG adduct tends to disrupt the essential minor groove polymerase contacts introducing a distortion in the damage site which cannot be tolerated by high fidelity polymerases. However, Y-family polymerases, including Dpo4, can bypass this adduct less efficiently [122, 123]. In contrast to the (+)-*trans*-B[*a*]P-N^2-dG adduct, the (+)-*cis*-B[*a*]P-N^2-dG adduct conformation in the DNA, completely blocks the DNA synthesis.

Using well established 5’ slippage mechanism, Dpo4 slowly bypasses the (+)-*trans*-B[*a*]P-N^2-dG adduct in our DNA sequence, which has AG*A at the primer-template junction (Figure 35) [18, 118]. The dT incorporation is significantly favored across the adduct in the 8Cy3 ternary complex with a B[*a*]P modified template indicating that Dpo4 incorporates incoming dTTP with dA, which is the second template base 5’ of the modified dG (Figure 38). This mechanism is supported by the Dpo4’s large active site which can accommodate two bases simultaneously [105, 124].

The crystal structure of the Dpo4 binary complex reveals that the last base pair of the primer-template junction is located at the nucleotide binding site [125]. In this conformation, the Dpo4 active site is blocked, and dNTP incorporation is prevented. Therefore, during a successful nucleotide incorporation, Dpo4 has to move one base forward to align with the proper templating base. In the absence of nucleotides, Dpo4 might move back and forth to find the proper alignment of the templating base before catalysis. These dynamic movements which were not seen in the crystal structure were characterized in this study in real time using smFRET.

Consistence with the previously published study,[82] our data shows that Dpo4 shuttles between two FRET states in the 8Cy3 unmodified binary complex (Figure 36B). The predominant high FRET state (FRET = 0.81) is considered as PIBC where Dpo4's active site is blocked by the terminal base pair of the DNA. The PIBC is in equilibrium with the low FRET IBC where the active site is open for nucleotide incorporation. The TDP also gives direct evidence about the transition of Dpo4 between these two FRET states.

The 8Cy3, (+)-*trans*-B[*a*]P-N^2-dG modified binary complex shows a totally inverted FRET distribution (Figure 39C). The Dpo4 prefers to bind to the low FRET (FRET = 0.57) state, and we propose two possible Dpo4 binding conformations, IBC1 and IBC2 for the low FRET state (Figure 46). In IBC1 the adduct is intercalated to the DNA, and it blocks the Dpo4 active site. In IBC2, presumably Dpo4 is located in the same position on the DNA, but the adduct might change its intercalated/stacked conformation to the exposed/outside conformation. The exposed adduct might release the distortion of the primer template junction. Most importantly, the Dpo4 active site is

vacant in the IBC2. As we propose that the position of the Dpo4 is the same in both IBC1 and IBC2, the crystal structure of the (+)-*trans*-B[*a*]P-N^2-dG adduct with Dpo4 also shows the same Dpo4 conformation in both crystals [18].

The IBC1 is in equilibrium with PIBC1 where Dpo4 translocates backward on the DNA, which gives a high FRET state (FRET = 0.77). The Dpo4 active site is occupied by the terminal base pair in the PIBC1. The same way, IBC2 is in equilibrium with PIBC2 blocking the Dpo4 active as shown in Figure 46.

We propose that the low FRET, IBC1 and IBC2 are relatively stable compared to their PIB complexes, which may be short lived and/or unable to get in the crystals. The small population in the high FRET state in the Figure 37C histogram (FRET = 0.77) supports the idea of the instability of the PIB complexes. However, the resolution limits of the FRET experiments do not capture the interconversion dynamics between IBC1 and IBC2. As we see a very broad transition in the TDP (Figure 39C) compared to the unmodified DNA, these transitions may be buried in there.

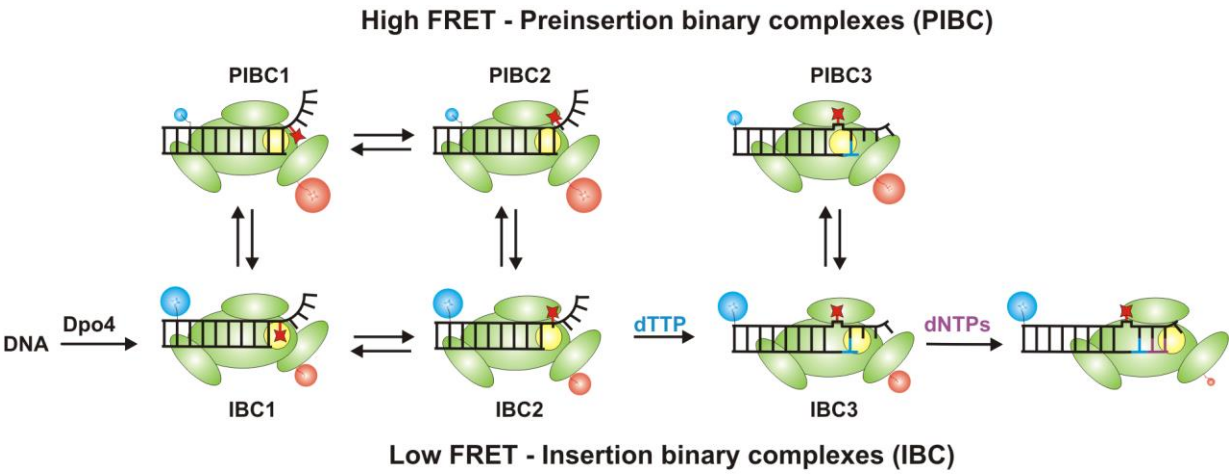


Figure 46. The proposed model shows the possible Dpo4 binding conformations during bypassing the (+)-trans-B[a]P-N2-dG adduct. The proposed model shows the possible Dpo4 binding conformations during bypassing the (+)-trans-B[a]P-N^2-dG adduct. When Dpo4 binds to the B[a]P modified DNA, it forms two possible conformations in the binary complex, the insertion binary complex (IBC) and preinsertion binary complex (PIBC). The B[a]P adduct can passion in two different ways in the IBC. Therefore, IBC can have two sub conformations, IBC1 and IBC2. In the IBC1 the adduct is intercalated into the DNA. The adduct places exposed or outside the helix in the IBC2. These two conformations are in equilibrium with the corresponding high FRET preinsertion conformations, PIBC1 and PIBC2. Out of all four possible confrontations, only IBC2 has an open active site for nucleotide incorporation. Therefore, we propose that only IBC2 will lead for a successful dNTP incorporation. As Dpo4 uses a template slippage mechanism for dNTP incorporation, in our template, dTTP is incorporated complementary to the next base 5' to the modified dG. The B[a]P modified dG is skipped and looped out causing to make -1 frame shift mutation. After dT incorporation, the next binary complex also can be in two possible conformation such as IBC3 and PIBC3. Again, IBC3 has the open active sine which allows for continuing DNA synthesis.

The Dpo4 binding conformation in the ternary complex depicts a significant change compared to the binary complex. A single FRET state is formed in the ternary complex, indicating a single Dpo4 binding conformation regardless of the nucleotide. This new FRET state in the ternary complex is closer to the low FRET state in the IBC, which indicates that the translocated Dpo4 active site is ready for the dNTP incorporation as it is blocked in the PIBC. The 8Cy3 ternary complex gives nearly identical FRET values for all the dNTPs (Figure 40B). The potential reason for this

observation would be due to the fact that the Dpo4 incorporates all the noncognate dNTPs in the unmodified DNA. However, the dissociation rate constants for the unmodified template clearly shows the greatest ternary complex stabilization for the incorporation of dCTP across the unmodified dG (k_{off} = 0.10±0.02, Figure 40C left). The same phenomenon was observed for the 8Cy3 B[a]P modified ternary complex during the incorporation of dT across the adduct having the lowest dissociation rate constant (k_{off} = 0.16±0.02, Figure 40C, right).

The 21 mer primer used in the 9Cy3mm DNA construct supports the 5'-DNA slippage mechanism, which is thoroughly explained in *Xu et al* [105]. Hence, it is possible that modified dG is looped out by allowing terminal dT to base pair with the dA which is the 5' to the modified dG. The Dpo4 binding conformations in the B[*a*]P modified 9Cy3 binary complex show FRET states similar to the unmodified DNA (Figure 41). The normal, unstrained nature of the DNA after the adduct would be the reason for the similar Dpo4 conformation after the adduct. Furthermore, Dpo4 tends to be highly error prone again in the modified DNA, similar to the unmodified DNA (Figure 43A). However, the Dpo4 shows the greatest stabilization again for the dT and dA incorporation having the lowest dissociation rate constants in the unembodied and modified DNA, respectively (Figure 43C).

According to the proposed mechanistic pathway (Figure 44), the primer extension is inhibited by intercalated (+)-*trans*-B[*a*]P-N^2-dG in the IBC1, so only IBC2 can proceed to the next step for dNTP incorporation. Dpo4 shows a major pause before the adduct during replication (Figure 36) possibly due to the interconversion of all these conformations until it forms IBC2. Our data shows a strong support for the adduct-

lopped-out, Dpo4's 5' slippage mechanism. The Dpo4 successfully incorporates dTTP bypassing the (+)-*trans*-B[*a*]P-N^2-dG adduct to from the IBC3 using the 5' slippage mechanism. Though IBC3 is in equilibrium with PIBC3, only the IBC3 can continue with DNA synthesis.

CHAPTER 5: CONCLUSIONS

Understanding the relationship between the exogenous chemicals and cancer is an important step in cancer prevention [126]. B[*a*]P is a well-known environmental carcinogen which is known to produce DNA adducts that blocks the DNA synthesis by high fidelity, replicative polymerases. Hence, low fidelity Y-family polymerases are recruited to the damage site to bypass bulky, B[*a*]P adducts, and to restore DNA synthesis. In this study, we use *Sulfolobus solfataricus* DNA polymerase IV (Dpo4) as a model Y-family polymerase which is used extensively to understand the B[*a*]P bypass mechanism. We carried out primer extension assays and the smFRET experiments to study Dpo4 conformational dynamics during bypassing two isomers of B[*a*]P DNA adducts, (+)-*cis*-B[*a*]P-N^2-dG and (+)-*trans*-B[*a*]P-N^2-dG.

The use of smFRET technique in studying Dpo4-DNA interactions was first developed by Alfonso Brenlla in our lab. This method was successfully applied to study our system, Dpo4-B[*a*]P modified DNA, to provide mechanistic insights into the Dpo4's trans lesion mechanism. First we made the B[*a*]P modified DNA template using the protocol which was used by Tom Christian in 2009. Though we were able to successfully separate all four B[*a*]P isomeric product by HPLC, we decided to continue experiments with (+)-*cis*-B[*a*]P-N^2-dG and (+)-*trans*-B[*a*]P-N^2-dG adducts because they are the most abundant isomeric products in nature.

It has been seen previously that the B[*a*]P adducts can change their conformations between the solvent exposed/outside and the intercalated/inside. The exposed conformation can be stabilized in two ways. First one is the stabilization of the B[*a*]P hydrophobic ring system between the structural gap, little finger and core domain,

in the Dpo4 [18]. If the B[*a*]P adduct stereochemistry does not support for it to locate in the structural gap, the hydrophobic adduct is forced to stay in the intercalated/inside conformation unless its solvent exposed conformation is stabilized by the microenvironment around the adduct.

In our study, the (+)-*cis*-B[*a*]P-N^2-dG adduct completely blocks the DNA synthesis by Dpo4. Presumably the structural properties of the adduct does not allow it to be positioned and stabilized in the structural gap of the Dpo4 (there is no crystal structure for (+)-*cis*-B[*a*]P-N^2-dG and Dpo4). Therefore, in the second way, the adduct exposed conformation must be stabilized by other factors, which are not known specifically, presumably water soluble macro molecules, metabolites, molecular crowding around the adduct and other proteins assist in TLS for bypassing DNA adduct *in vivo*. Therefore, we used DMSO to mimic the cytosolic hydrophobic properties in our experiments.

Most importantly, Dpo4 binding conformation was changed as we decreased the solvent polarity by adding DMSO to the binding buffer, and Dpo4 was able to synthesize DNA past the adduct in the presence of DMSO. These observations confirm the change of adduct conformation during successful DNA synthesis, and our results elucidate the possible mechanistic pathway how Dpo4 bypass the (+)-*cis*-B[*a*]P-N^2-dG adduct *in vivo*. We were able to reproduce the single molecule results of the Dpo4-DNA interactions which was published in *Brenlla et al.* for the unmodified DNA template. Dpo4 showed rather different binding conformation in the modified template due to the distortion imposed by the adduct. The nucleotide incorporation assay with single nucleotide showed that Dpo4 predominantly inserts dG across the (+)-*cis*-B[*a*]P-N^2-dG adduct

forming an unusual G:G mismatch. We fully characterized all these Dpo4 binding conformations of the (+)-*cis*-B[*a*]P-N^2-dG adduct with the Dpo4 bypass mechanism proposing a possible model as shown in Figure 35.

We were fortunate to start a collaboration with Prof. Andrés Cisneros to study our system computationally. Their lab did some computational calculations and MD simulations on the Dpo4-DNA system in the presence and absence of DMSO. Our biochemical and single molecule results were strongly supported by their computational results. The stabilization of the solvent exposed conformation of the adduct in the presence of DMSO was one of the interesting computational results that we obtained comparable to our experimental data. The correlation analysis with and without DMSO also generated a very promising data that are nicely align with experimental data.

In contrast, the (+)-*trans*-B[*a*]P-N^2-dG adduct conformation on the DNA does not completely blocks the DNA synthesis by Dpo4, but it prefers to incorporate nucleotides across the adduct not complimentary to the modified base, but to the base 5' to the modified dG, which is dA in our template. The Dpo4 bypass mechanism follows the -1 frameshift mutation while pushing out the (+)-*trans*-B[*a*]P-N^2-dG adduct to the structural gap between the little finger and thumb domains as seen in the crystal structure. The smFRET experimental results show four possible Dpo4 binding conformations (IBC1, IBC2, PIBC1, PIBC2) for the binary complex (Figure 45). However, we were able to capture only two of them due to the resolution limit of the smFRET experiments, and/or fast transition between these complexes (IBC1 and IBC2). As we seen in the smFRET experiments, Dpo4 predominantly binds to a low FRET conformation where (+)-*trans*-B[*a*]P-N^2-dG adduct is in the exposed conformation stabilizing between the structural

gap. These conformational dynamics reveal that modified DNA bound Dpo4 is restricted to conformation limiting its freedom to move around the primer-template junction. Dpo4 shows a greater stabilization in the ternary complex binding conformation for the nucleotides that are being incorporated during DNA synthesis. Our results provide a mechanistic basis for the stepwise process by which a Y-family polymerase extends across and past a bulky adduct in DNA.

APPENDIX

A1. Copyright permission for figure **4-7.**

Home | Create Account | Help

ACS Publications
Most Trusted. Most Cited. Most Read.

Title:	NMR Solution Structures of Stereoisomeric Covalent Polycyclic Aromatic Carcinogen–DNA Adducts: Principles, Patterns, and Diversity
Author:	Nicholas E. Geacintov, Monique Cosman, Brian E. Hingerty, et al
Publication:	Chemical Research in Toxicology
Publisher:	American Chemical Society
Date:	Feb 1, 1997

Copyright © 1997, American Chemical Society

LOGIN

If you're a copyright.com user, you can login to RightsLink using your copyright.com credentials. Already **a RightsLink user** or want to learn more?

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

A2. Copyright permission for figure **8.**

Title: Snapshots of a Y-Family DNA Polymerase in Replication: Substrate-induced Conformational Transitions and Implications for Fidelity of Dpo4

Author: Jimson H. Wong,Kevin A. Fiala,Zucai Suo,Hong Ling

Publication: Journal of Molecular Biology

Publisher: Elsevier

Date: 29 May 2008

Logged in as:
Pramodha Liyanage
LOGOUT

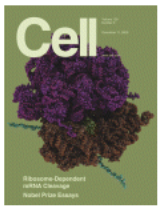
Order Completed

Thank you for your order.

This Agreement between Pramodha S Liyanage ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

A3. Copyright permission for figure **9.**



Title: Crystal Structure of a Y-Family DNA Polymerase in Action A Mechanism for Error-Prone and Lesion-Bypass Replication

Author: Hong Ling,François Boudsocq,Roger Woodgate,Wei Yang

Publication: Cell

Publisher: Elsevier

Date: 5 October 2001

Logged in as:
Pramodha Liyanage
Account #:
3001145779
LOGOUT

Order Completed

Thank you for your order.

This Agreement between Pramodha S Liyanage ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

Printable details.

License Number	4098910527588
License date	Apr 30, 2017
Licensed Content Publisher	Elsevier
Licensed Content Publication	Cell
Licensed Content Title	Crystal Structure of a Y-Family DNA Polymerase in Action A Mechanism for Error-Prone and Lesion-Bypass Replication
Licensed Content Author	Hong Ling,François Boudsocq,Roger Woodgate,Wei Yang
Licensed Content Date	5 October 2001
Licensed Content Volume	107
Licensed Content Issue	1
Licensed Content Pages	12
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Order reference number	
Original figure numbers	Figure 6
Title of your thesis/dissertation	REAL-TIME INVESTIGATION OF BULKY LESION BYPASS BY Y-FAMILY DNA POLYMERASE, DPO4, USING SINGLE MOLECULE FRET
Expected completion date	May 2017
Estimated size (number of pages)	140
Elsevier VAT number	GB 494 6272 12
Requestor Location	Pramodha S Liyanage 4647 Chrysler Dr Apt 202 DETROIT, MI 48201 United States Attn: Pramodha S Liyanage
Publisher Tax ID	98-0397604
Total	0.00 USD

A4. Copyright permission for figure **13.**

Title: Single-Molecule Förster Resonance Energy Transfer Reveals an Innate Fidelity Checkpoint in DNA Polymerase I

Author: Svitlana Y. Berezhna, Joshua P. Gill, Rajan Lamichhane, et al

Publication: Journal of the American Chemical Society

Publisher: American Chemical Society

Date: Jul 1, 2012

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

If credit is given to another source for the material you requested, permission must be obtained from that source.

A5. Copyright permission for figure **14.**

Title: Nucleotide selection by the Y-family DNA polymerase Dpo4 involves template translocation and misalignment
Author: Brenlla, Alfonso; Markiewicz, Radoslaw P.
Publication: Nucleic Acids Research
Publisher: Oxford University Press
Date: 2013-11-21

Logged in as:
Pramodha Liyanage
Account #:
3001145779
LOGOUT

Review Order

Please review the order details and the associated terms and conditions.

No royalties will be charged for this reuse request although you are required to obtain a license and comply with the license terms and conditions. To obtain the license, click the Accept button below.

Licensed Content Publisher	Oxford University Press
Licensed Content Publication	Nucleic Acids Research
Licensed Content Title	Nucleotide selection by the Y-family DNA polymerase Dpo4 involves template translocation and misalignment
Licensed Content Author	Brenlla, Alfonso; Markiewicz, Radoslaw P.
Licensed Content Date	2013-11-21
Licensed Content Volume	42
Licensed Content Issue	4
Type of Use	Thesis/Dissertation
Requestor type	Educational Institution/Non-commercial/ Not for-profit
Format	Print and electronic
Portion	Figure/table
Number of figures/tables	1
Will you be translating?	No
Order reference number	
Title of your thesis / dissertation	REAL-TIME INVESTIGATION OF BULKY LESION BYPASS BY Y-FAMILY DNA POLYMERASE, DPO4, USING SINGLE MOLECULE FRET
Expected completion date	May 2017
Estimated size(pages)	140
Requestor Location	Pramodha S Liyanage 4647 Chrysler Dr Apt 202 DETROIT, MI 48201 United States Attn: Pramodha S Liyanage
Publisher Tax ID	GB125506730
Total	0.00 USD

REFERENCES

1. Zhou, B.B. and S.J. Elledge, *The DNA damage response: putting checkpoints in perspective.* Nature, 2000. **408**(6811): p. 433-9.

2. Horvath, P. and R. Barrangou, *CRISPR/Cas, the immune system of bacteria and archaea.* Science, 2010. **327**(5962): p. 167-70.

3. Elledge, S.J., *Cell cycle checkpoints: preventing an identity crisis.* Science, 1996. **274**(5293): p. 1664-72.

4. Benigni, R. and C. Bossa, *Mechanisms of chemical carcinogenicity and mutagenicity: a review with implications for predictive toxicology.* Chem Rev, 2011. **111**(4): p. 2507-36.

5. Alexandrov, L.B., et al., *Signatures of mutational processes in human cancer.* Nature, 2013. **500**(7463): p. 415-21.

6. Vogelstein, B. and K.W. Kinzler, *Carcinogens leave fingerprints.* Nature, 1992. **355**(6357): p. 209-10.

7. Vayssier-Taussat, M., et al., *Effects of tobacco smoke and benzo[a]pyrene on human endothelial cell and monocyte stress responses.* Am J Physiol Heart Circ Physiol, 2001. **280**(3): p. H1293-300.

8. Hollstein, M., et al., *p53 mutations in human cancers.* Science, 1991. **253**(5015): p. 49-53.

9. Anand, P., et al., *Cancer is a preventable disease that requires major lifestyle changes.* Pharm Res, 2008. **25**(9): p. 2097-116.

10. Jemal, A., et al., *Cancer statistics, 2007.* CA Cancer J Clin, 2007. **57**(1): p. 43-66.

11. Cohen, S.M. and L.L. Arnold, *Chemical carcinogenesis.* Toxicol Sci, 2011. **120 Suppl 1**: p. S76-92.

12. Weisburger, J.H. and G.M. Williams, *The distinction between genotoxic and epigenetic carcinogens and implication for cancer risk.* Toxicol Sci, 2000. **57**(1): p. 4-5.

13. Weisburger, J.H. and G.M. Williams, *Carcinogen testing: current problems and new approaches.* Science, 1981. **214**(4519): p. 401-7.

14. DeBaun, J.R., E.C. Miller, and J.A. Miller, *N-hydroxy-2-acetylaminofluorene sulfotransferase: its probable role in carcinogenesis and in protein-(methion-S-yl) binding in rat liver.* Cancer Res, 1970. **30**(3): p. 577-95.

15. Phillips, D.H., *Fifty years of benzo(a)pyrene.* Nature, 1983. **303**(5917): p. 468-72.

16. Ling, H., et al., *Crystal structure of a benzo[a]pyrene diol epoxide adduct in a ternary complex with a DNA polymerase.* Proc Natl Acad Sci U S A, 2004. **101**(8): p. 2265-9.

17. Chandani, S. and E.L. Loechler, *Molecular modeling benzo[a]pyrene N2-dG adducts in the two overlapping active sites of the Y-family DNA polymerase Dpo4.* J Mol Graph Model, 2007. **25**(5): p. 658-70.

18. Bauer, J., et al., *A structural gap in Dpo4 supports mutagenic bypass of a major benzo[a]pyrene dG adduct in DNA through template misalignment.* Proc Natl Acad Sci U S A, 2007. **104**(38): p. 14905-10.

19. Pfeifer, G.P., et al., *Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers.* Oncogene, 2002. **21**(48): p. 7435-51.

20. Denissenko, M.F., et al., *Preferential formation of benzo[a]pyrene adducts at lung cancer mutational hotspots in P53.* Science, 1996. **274**(5286): p. 430-2.

21. Kozack, R., et al., *Toward an understanding of the role of DNA adduct conformation in defining mutagenic mechanism based on studies of the major adduct (formed at N(2)-dG) of the potent environmental carcinogen, benzo[a]pyrene.* Mutat Res, 2000. **450**(1-2): p. 41-59.

22. Xue, W. and D. Warshawsky, *Metabolic activation of polycyclic and heterocyclic aromatic hydrocarbons and DNA damage: a review.* Toxicol Appl Pharmacol, 2005. **206**(1): p. 73-93.

23. Harvey, R.G., *Polycyclic aromatic hydrocarbons : chemistry and carcinogenicity*. Cambridge monographs on cancer research. 1991, Cambridge ; New York: Cambridge University Press. xv, 396 p.

24. Gelboin, H.V., *Benzo[alpha]pyrene metabolism, activation and carcinogenesis: role and regulation of mixed-function oxidases and related enzymes.* Physiol Rev, 1980. **60**(4): p. 1107-66.

25. Levin, W., et al., *An enantiomeric interaction in the metabolism and tumorigenicity of (+)- and (-)-benzo[a]pyrene 7,8-oxide.* J Biol Chem, 1980. **255**(19): p. 9067-74.

26. Dipple, A., et al., *7,12-dimethylbenz[a] anthracene--DNA binding in mouse skin: response of different mouse strains and effects of various modifiers of carcinogenesis.* Carcinogenesis, 1984. **5**(8): p. 1087-90.

27. Geacintov, N.E., et al., *Non-covalent intercalative binding of 7,8-dihydroxy-9,10-epoxybenzo(a)pyrene to DNA.* Biochem Biophys Res Commun, 1981. **100**(4): p. 1569-77.

28. Geacintov, N.E., et al., *NMR solution structures of stereoisometric covalent polycyclic aromatic carcinogen-DNA adduct: principles, patterns, and diversity.* Chem Res Toxicol, 1997. **10**(2): p. 111-46.

29. Hruszkewycz, A.M., et al., *DNA polymerase action on benzo[a]pyrene-DNA adducts.* Carcinogenesis, 1992. **13**(12): p. 2347-52.

30. Cosman, M., et al., *Solution conformation of the major adduct between the carcinogen (+)-anti-benzo[a]pyrene diol epoxide and DNA.* Proc Natl Acad Sci U S A, 1992. **89**(5): p. 1914-8.

31. Cosman, M., et al., *Solution conformation of the (+)-cis-anti-[BP]dG adduct in a DNA duplex: intercalation of the covalently attached benzo[a]pyrenyl ring into the helix and displacement of the modified deoxyguanosine.* Biochemistry, 1993. **32**(16): p. 4145-55.

32. Cosman, M., et al., *Solution conformation of the (+)-cis-anti-[BP]dG adduct opposite a deletion site in a DNA duplex: intercalation of the covalently attached benzo[a]pyrene into the helix with base displacement of the modified deoxyguanosine into the minor groove.* Biochemistry, 1994. **33**(38): p. 11518-27.

33. Yang, W., *An Overview of Y-Family DNA Polymerases and a Case Study of Human DNA Polymerase η.* Biochemistry, 2014. **53**(17): p. 2793-2803.

34. Sholder, G., A. Creech, and E.L. Loechler, *How Y-Family DNA polymerase IV is more accurate than Dpo4 at dCTP insertion opposite an N2-dG adduct of benzo[a]pyrene.* DNA Repair, 2015. **35**: p. 144-153.

35. Guo, C., et al., *Y-family DNA polymerases in mammalian cells.* Cell Mol Life Sci, 2009. **66**(14): p. 2363-81.

36. Boiteux, S. and S. Jinks-Robertson, *DNA repair mechanisms and the bypass of DNA damage in Saccharomyces cerevisiae.* Genetics, 2013. **193**(4): p. 1025-64.

37. Budzowska, M. and R. Kanaar, *Mechanisms of dealing with DNA damage-induced replication problems.* Cell Biochem Biophys, 2009. **53**(1): p. 17-31.

38. Li, X. and W.D. Heyer, *Homologous recombination in DNA repair and DNA damage tolerance.* Cell Res, 2008. **18**(1): p. 99-113.

39. Sale, J.E., A.R. Lehmann, and R. Woodgate, *Y-family DNA polymerases and their role in tolerance of cellular DNA damage.* Nat Rev Mol Cell Biol, 2012. **13**(3): p. 141-152.

40. Yang, W. and R. Woodgate, *What a difference a decade makes: Insights into translesion DNA synthesis.* Proceedings of the National Academy of Sciences, 2007. **104**(40): p. 15591-15598.

41. Yang, W., *Damage repair DNA polymerases Y.* Curr Opin Struct Biol, 2003. **13**(1): p. 23-30.

42. Masutani, C., et al., *The XPV (xeroderma pigmentosum variant) gene encodes human DNA polymerase eta.* Nature, 1999. **399**(6737): p. 700-4.

43. Johnson, R.E., et al., *hRAD30 mutations in the variant form of xeroderma pigmentosum.* Science, 1999. **285**(5425): p. 263-5.

44. Johnson, R.E., S. Prakash, and L. Prakash, *Efficient bypass of a thymine-thymine dimer by yeast DNA polymerase, Poleta.* Science, 1999. **283**(5404): p. 1001-4.

45. Chaney, S.G., et al., *Recognition and processing of cisplatin- and oxaliplatin-DNA adducts.* Crit Rev Oncol Hematol, 2005. **53**(1): p. 3-11.

46. Lindahl, T., *Instability and decay of the primary structure of DNA.* Nature, 1993. **362**(6422): p. 709-15.

47. Kokoska, R.J., S.D. McCulloch, and T.A. Kunkel, *The efficiency and specificity of apurinic/apyrimidinic site bypass by human DNA polymerase eta and Sulfolobus solfataricus Dpo4.* J Biol Chem, 2003. **278**(50): p. 50537-45.

48. Strauss, B.S., *The 'A rule' of mutagen specificity: a consequence of DNA polymerase bypass of non-instructional lesions?* Bioessays, 1991. **13**(2): p. 79-84.

49. Ohashi, E., et al., *Error-prone bypass of certain DNA lesions by the human DNA polymerase kappa.* Genes Dev, 2000. **14**(13): p. 1589-94.

50. Ogi, T., et al., *Polkappa protects mammalian cells against the lethal and mutagenic effects of benzo[a]pyrene.* Proc Natl Acad Sci U S A, 2002. **99**(24): p. 15548-53.

51. Makridakis, N.M. and J.K. Reichardt, *Translesion DNA polymerases and cancer.* Front Genet, 2012. **3**: p. 174.

52. Rechkoblit, O., et al., *trans-Lesion synthesis past bulky benzo[a]pyrene diol epoxide N2-dG and N6-dA lesions catalyzed by DNA bypass polymerases.* J Biol Chem, 2002. **277**(34): p. 30488-94.

53. Velasco-Miguel, S., et al., *Constitutive and regulated expression of the mouse Dinb (Polkappa) gene encoding DNA polymerase kappa.* DNA Repair (Amst), 2003. **2**(1): p. 91-106.

54. Wang, Y., et al., *Elevated expression of DNA polymerase kappa in human lung cancer is associated with p53 inactivation: Negative regulation of POLK promoter activity by p53.* Int J Oncol, 2004. **25**(1): p. 161-5.

55. Vaisman, A., et al., *poliota-dependent lesion bypass in vitro.* Mutat Res, 2002. **510**(1-2): p. 9-22.

56. McDonald, J.P., et al., *Novel human and mouse homologs of Saccharomyces cerevisiae DNA polymerase eta.* Genomics, 1999. **60**(1): p. 20-30.

57. Tissier, A., et al., *poliota, a remarkably error-prone human DNA polymerase.* Genes Dev, 2000. **14**(13): p. 1642-50.

58. Nair, D.T., et al., *Replication by human DNA polymerase-iota occurs by Hoogsteen base-pairing.* Nature, 2004. **430**(6997): p. 377-80.

59. Vaisman, A. and R. Woodgate, *Unique misinsertion specificity of poliota may decrease the mutagenic potential of deaminated cytosines.* EMBO J, 2001. **20**(22): p. 6520-9.

60. Frank, E.G., et al., *Translesion replication of benzo[a]pyrene and benzo[c]phenanthrene diol epoxide adducts of deoxyadenosine and deoxyguanosine by human DNA polymerase iota.* Nucleic Acids Res, 2002. **30**(23): p. 5284-92.

61. Tissier, A., et al., *Co-localization in replication foci and interaction of human Y-family members, DNA polymerase pol eta and REVl protein.* DNA Repair (Amst), 2004. **3**(11): p. 1503-14.

62. Wang, Y., et al., *Evidence that in xeroderma pigmentosum variant cells, which lack DNA polymerase eta, DNA polymerase iota causes the very high frequency and unique spectrum of UV-induced mutations.* Cancer Res, 2007. **67**(7): p. 3018-26.

63. Tissier, A., et al., *Misinsertion and bypass of thymine-thymine dimers by human DNA polymerase iota.* EMBO J, 2000. **19**(19): p. 5259-66.

64. Yang, J., et al., *Altered DNA polymerase iota expression in breast cancer cells leads to a reduction in DNA replication fidelity and a higher rate of mutagenesis.* Cancer Res, 2004. **64**(16): p. 5597-607.

65. Wong, J.H., et al., *Snapshots of a Y-family DNA polymerase in replication: substrate-induced conformational transitions and implications for fidelity of Dpo4.* J Mol Biol, 2008. **379**(2): p. 317-30.

66. Ling, H., et al., *Crystal structure of a Y-family DNA polymerase in action: a mechanism for error-prone and lesion-bypass replication.* Cell, 2001. **107**(1): p. 91-102.

67. Ha, T., *Single-Molecule FRET.* Single Molecules, 2001. **2**(4): p. 283-284.

68. Weiss, S., *Fluorescence Spectroscopy of Single Biomolecules.* Science, 1999. **283**(5408): p. 1676-1683.

69. Lamichhane, R., et al., *Single-molecule FRET of protein–nucleic acid and protein–protein complexes: Surface passivation and immobilization.* Methods, 2010. **52**(2): p. 192-200.

70. Ha, T., *Single-Molecule Fluorescence Resonance Energy Transfer.* Methods, 2001. **25**(1): p. 78-86.

71. Lu, H.P., L. Xun, and X.S. Xie, *Single-Molecule Enzymatic Dynamics.* Science, 1998. **282**(5395): p. 1877-1882.

72. Ha, T., et al., *Single-molecule fluorescence spectroscopy of enzyme conformational dynamics and cleavage mechanism.* Proceedings of the National Academy of Sciences, 1999. **96**(3): p. 893-898.

73. Hochreiter, B., A.P. Garcia, and J.A. Schmid, *Fluorescent proteins as genetically encoded FRET biosensors in life sciences.* Sensors (Basel), 2015. **15**(10): p. 26281-314.

74. Roy, R., S. Hohng, and T. Ha, *A practical guide to single-molecule FRET.* Nat Meth, 2008. **5**(6): p. 507-516.

75. Hirata, E. and E. Kiyokawa, *Future Perspective of Single-Molecule FRET Biosensors and Intravital FRET Microscopy.* Biophysical Journal.

76. Raper, A.T., et al., *Advances in Structural and Single-Molecule Methods for Investigating DNA Lesion Bypass and Repair Polymerases.* Chem Res Toxicol, 2017. **30**(1): p. 260-269.

77. Christian, T.D., L.J. Romano, and D. Rueda, *Single-molecule measurements of synthesis by DNA polymerase with base-pair resolution.* Proc Natl Acad Sci U S A, 2009. **106**(50): p. 21109-14.

78. Markiewicz, R.P., et al., *Single-molecule microscopy reveals new insights into nucleotide selection by DNA polymerase I.* Nucleic Acids Res, 2012. **40**(16): p. 7975-84.

79. Santoso, Y., et al., *Conformational transitions in DNA polymerase I revealed by single-molecule FRET.* Proc Natl Acad Sci U S A, 2010. **107**(2): p. 715-20.

80. Berezhna, S.Y., et al., *Single-molecule Forster resonance energy transfer reveals an innate fidelity checkpoint in DNA polymerase I.* J Am Chem Soc, 2012. **134**(27): p. 11261-8.

81. Wu, E.Y. and L.S. Beese, *The structure of a high fidelity DNA polymerase bound to a mismatched nucleotide reveals an "ajar" intermediate conformation in the nucleotide selection mechanism.* J Biol Chem, 2011. **286**(22): p. 19758-67.

82. Brenlla, A., et al., *Nucleotide selection by the Y-family DNA polymerase Dpo4 involves template translocation and misalignment.* Nucleic Acids Res, 2014. **42**(4): p. 2555-63.

83. Brenlla, A., D. Rueda, and L.J. Romano, *Mechanism of aromatic amine carcinogen bypass by the Y-family polymerase, Dpo4.* Nucleic Acids Res, 2015. **43**(20): p. 9918-27.

84. Arghavani, M.B., J. SantaLucia, Jr., and L.J. Romano, *Effect of mismatched complementary strands and 5'-change in sequence context on the thermodynamics and structure of benzo[a]pyrene-modified oligonucleotides.* Biochemistry, 1998. **37**(23): p. 8575-83.

85. Boudsocq, F., et al., *Sulfolobus solfataricus P2 DNA polymerase IV (Dpo4): an archaeal DinB-like DNA polymerase with lesion-bypass properties akin to eukaryotic poleta.* Nucleic Acids Res, 2001. **29**(22): p. 4607-16.

86. Zhao, R. and D. Rueda, *RNA folding dynamics by single-molecule fluorescence resonance energy transfer.* Methods, 2009. **49**(2): p. 112-7.

87. Vincent B. Chen, W.B.A.I., Jeffrey J. Headd, Daniel A. Keedy, Robert M. Immormino, Gary J. Kapral, Laura W. Murray, Jane S. Richardson and David C. Richardson, *MolProbity: all-atom structure validation for macromolecular crystallography.* Acta Crystallographica, 2010. **D**(66): p. 12-21.

88. Case, D.A., et al., *AMBER 2016*. 2016, University of California: San Francisco.

89. Salomon-Ferrer, R., et al., *Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald.* Journal of Chemical Theory and Computation, 2013. **9**(9): p. 3878-3888.

90. Maier, J.A., et al., *ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB.* Journal of Chemical Theory and Computation, 2015. **11**(8): p. 3696-3713.

91. Bakan, A., L.M. Meireles, and I. Bahar, *ProDy: Protein Dynamics Inferred from Theory and Experiments.* Bioinformatics, 2011. **27**(11): p. 1575-1577.

92. Humphrey, W., A. Dalke, and K. Schulten, *VMD: Visual molecular dynamics.* Journal of Molecular Graphics, 1996. **14**(1): p. 33-38.

93. Mu, H., et al., *The role of structural and energetic factors in regulating repair of a bulky DNA lesion with different opposite partner bases().* Biochemistry, 2013. **52**(33): p. 10.1021/bi4009177.

94. Shapovalov, M.V. and R.L. Dunbrack, Jr., *A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions.* Structure, 2011. **19**(6): p. 844-58.

95. Dupradeau, F.Y., et al., *The R.E.D. tools: advances in RESP and ESP charge derivation and force field library building.* Phys Chem Chem Phys, 2010. **12**(28): p. 7821-39.

96. Vanquelef, E., et al., *R.E.D. Server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments.* Nucleic Acids Res, 2011. **39**(Web Server issue): p. W511-7.

97. Mocquet, V., et al., *The human DNA repair factor XPC-HR23B distinguishes stereoisomeric benzo[a]pyrenyl-DNA lesions.* Embo j, 2007. **26**(12): p. 2923-32.

98. Cheatham, T.E., 3rd and D.A. Case, *Twenty-five years of nucleic acid simulations.* Biopolymers, 2013. **99**(12): p. 969-77.

99. Wang, J., et al., *Development and testing of a general amber force field.* Journal of Computational Chemistry, 2004. **25**(9): p. 1157-1174.

100. M. Bradbrook, G., et al., *X-Ray and molecular dynamics studies of concanavalin-A glucoside and mannoside complexes Relating structure to thermodynamics of binding.* Journal of the Chemical Society, Faraday Transactions, 1998. **94**(11): p. 1603-1611.

101. Fox, T. and P.A. Kollman, *Application of the RESP Methodology in the Parametrization of Organic Solvents.* The Journal of Physical Chemistry B, 1998. **102**(41): p. 8070-8079.

102. Essmann, U., et al., *A smooth particle mesh Ewald method.* The Journal of Chemical Physics, 1995. **103**(19): p. 8577-8593.

103. Berendsen, H.J.C., et al., *Molecular dynamics with coupling to an external bath.* The Journal of Chemical Physics, 1984. **81**(8): p. 3684.

104. Jorgensen, W.L., et al., *Comparison of simple potential functions for simulating liquid water.* The Journal of Chemical Physics, 1983. **79**(2): p. 926-935.

105. Xu, P., et al., *Nucleotide selectivity opposite a benzo[a]pyrene-derived N2-dG adduct in a Y-family DNA polymerase: a 5'-slippage mechanism.* Biochemistry, 2008. **47**(9): p. 2701-9.

106. Maxwell, B.A., C. Xu, and Z. Suo, *Conformational dynamics of a Y-family DNA polymerase during substrate binding and catalysis as revealed by interdomain Forster resonance energy transfer.* Biochemistry, 2014. **53**(11): p. 1768-78.

107. Hanrahan, C.J., et al., *Sequence specific mutagenesis of the major (+)-anti-benzo[a]pyrene diol epoxide-DNA adduct at a mutational hot spot in vitro and in Escherichia coli cells.* Chem Res Toxicol, 1997. **10**(4): p. 369-77.

108. Mu, H., et al., *Role of structural and energetic factors in regulating repair of a bulky DNA lesion with different opposite partner bases.* Biochemistry, 2013. **52**(33): p. 5517-21.

109. Ghosh, S., et al., *Dynamics in cytoplasm, nucleus, and lipid droplet of a live CHO cell: time-resolved confocal microscopy.* Langmuir, 2013. **29**(25): p. 7975-82.

110. Vrtis, K.B., et al., *Carcinogenic adducts induce distinct DNA polymerase binding orientations.* Nucleic Acids Res, 2013. **41**(16): p. 7843-53.

111. Vaisman, A., et al., *Fidelity of Dpo4: effect of metal ions, nucleotide selection and pyrophosphorolysis.* EMBO J, 2005. **24**(17): p. 2957-67.

112. Irimia, A., et al., *Calcium is a cofactor of polymerization but inhibits pyrophosphorolysis by the Sulfolobus solfataricus DNA polymerase Dpo4.* Biochemistry, 2006. **45**(19): p. 5949-56.

113. Starostenko, L.V., et al., *Human DNA polymerases catalyze lesion bypass across benzo[a]pyrene-derived DNA adduct clustered with an abasic site.* DNA Repair (Amst), 2014. **24**: p. 1-9.

114. Ohashi, E., et al., *Interaction of hREV1 with three human Y-family DNA polymerases.* Genes Cells, 2004. **9**(6): p. 523-31.

115. Lehmann, A.R., et al., *Translesion synthesis: Y-family polymerases and the polymerase switch.* DNA Repair (Amst), 2007. **6**(7): p. 891-9.

116. Dzantiev, L. and L.J. Romano, *Differential effects of N-acetyl-2-aminofluorene and N-2-aminofluorene adducts on the conformational change in the structure of DNA polymerase I (Klenow fragment).* Biochemistry, 2000. **39**(17): p. 5139-45.

117. Fernandes, A., et al., *Mutagenic potential of stereoisomeric bay region (+)- and (-)-cis-anti-benzo[a]pyrene diol epoxide-N2-2'-deoxyguanosine adducts in Escherichia coli and simian kidney cells.* Biochemistry, 1998. **37**(28): p. 10164-72.

118. Perlow-Poehnelt, R.A., et al., *The spacious active site of a Y-family DNA polymerase facilitates promiscuous nucleotide incorporation opposite a bulky carcinogen-DNA adduct: elucidating the structure-function relationship through experimental and computational approaches.* J Biol Chem, 2004. **279**(35): p. 36951-61.

119. Shibutani, S., et al., *Translesional synthesis on a DNA template containing a single stereoisomer of dG-(+)- or dG-(-)-anti-BPDE (7, 8-dihydroxy-anti-9, 10-epoxy-7,8,9,10-tetrahydrobenzo[a]pyrene.* Biochemistry, 1993. **32**(29): p. 7531-7541.

120. Hsu, G.W., et al., *Structure of a high fidelity DNA polymerase bound to a benzo[a]pyrene adduct that blocks replication.* J Biol Chem, 2005. **280**(5): p. 3764-70.

121. Shcherbakova, P.V. and I.J. Fijalkowska, *Translesion synthesis DNA polymerases and control of genome stability.* Front Biosci, 2006. **11**: p. 2496-517.

122. Wood, R.D., et al., *DNA damage recognition and nucleotide excision repair in mammalian cells.* Cold Spring Harb Symp Quant Biol, 2000. **65**: p. 173-82.

123. Yang, W., *An overview of Y-Family DNA polymerases and a case study of human DNA polymerase eta.* Biochemistry, 2014. **53**(17): p. 2793-803.

124. Eoff, R.L., J.Y. Choi, and F.P. Guengerich, *Mechanistic Studies with DNA Polymerases Reveal Complex Outcomes following Bypass of DNA Damage.* J Nucleic Acids, 2010. **2010**.

125. Rechkoblit, O., et al., *Stepwise translocation of Dpo4 polymerase during error-free bypass of an oxoG lesion.* PLoS Biol, 2006. **4**(1): p. e11.

126. Kwack, S.J. and B.M. Lee, *Correlation between DNA or protein adducts and benzo[a]pyrene diol epoxide I-triglyceride adduct detected in vitro and in vivo.* Carcinogenesis, 2000. **21**(4): p. 629-32.

ABSTRACT

REAL-TIME INVESTIGATION OF BULKY LESION BYPASS BY Y-FAMILY DNA POLYMERASE, DPO4, USING SINGLE MOLECULE FRET

by

PRAMODHA LYANAGE

August 2017

Advisor: Dr. Louis J. Romano

Major: Chemistry (Biochemistry)

Degree: Doctor of Philosophy

DNA is constantly exposed to various DNA damaging agents that are generated by various internal and external sources. Some of this damage may not be able to be repaired by cellular machineries causing DNA replication to be blocked. Once the replication fork is blocked by a DNA adduct, damage tolerance DNA polymerases, mainly Y-family, are able to restore the DNA replication by synthesizing past the DNA adduct. Benzo[*a*]pyrene (B[*a*]P) is one of the most studied environmental carcinogens. It is known to make covalent DNA adducts after metabolic activation and the bulkiness of the B[*a*]P adducts impose a strong barrier to high-fidelity polymerases. However, many Y-family polymerases are able to bypass these adducts. In this study, we used a model Y-family polymerase isolated from *Sulfolobus solfataricus* called DNA polymerase IV (Dpo4), to study the B[*a*]P modified DNA bypass mechanism using single molecule FRET (smFRET). The B[*a*]P reaction with DNA produces four different isomeric products and we selected the most abundant of them, (+)-*cis*-B[*a*]P-N^2-dG and (+)-*trans*-B[*a*]P-N^2-dG, to study.

Our data show that (+)-*cis*-B[*a*]P-N^2-dG adduct blocks the DNA replication when it is in the intercalated conformation in the DNA. Use of DMSO in the reaction buffer supports for the stabilization of the adduct solvent exposed conformation which removes the obstacle in the Dpo4 active site for DNA synthesis. In contrast, (+)-*trans*-B[*a*]P-N^2-dG adduct does not completely block the DNA replication, but instead causes frame-shift mutation leading to a -1 deletions in the DNA product. Consistence with the previously published studies, Dpo4 moves between two conformations in the binary complex, but single conformation is formed in the ternary complex. However, Dpo4 binding conformations in the normal DNA and modified DNA are clearly different indicating its competence in translesion synthesis.

AUTOBIOGRAPHICAL STATEMENT

Pramodha Liyanage

Education

PhD	Chemistry -major Biochemistry Wayne State University, Detroit, MI 48201. Advisor – Prof. Louis J. Romano	2010-present
BSc	Chemistry The Open University of Sri Lanka.	2005
BSc	Chemistry College of Chemical Sciences, Institute of Chemistry Ceylon	2005

Presentations

Posters - 3rd Midwest single molecule workshop (2014).
University of Illinois at Urbana-Champaign, IL 61801.
- 16th Annual Chemistry graduate research symposium (2014).
Wayne State University, Detroit, MI 48201

Oral -18th Annual Chemistry graduate research symposium (2016).
Wayne State University, Detroit, MI 48201.

Publications

Review Articles

- König SL, **Pramodha Liyanage**, Sigel RK, Rueda David., Helicase-mediated changes in RNA structure at the single-molecule level. *RNA Biol., 2013 Jan;10(1):133-48*.

Research Articles

- **Pramodha Liyanage**, Alice R. Walker, Alfonso Brenlla, G. Andrés Cisneros, Louis J Romano* and David Rueda*
"Bulky Lesion Bypass Requires Dpo4 Binding in Distinct conformations". Under review in JACS (Journal of American Chemical Society)

- **Pramodha Liyanage**, David Rueda and Louis J Romano., "Investigation of Benzo[*a*]Pyrene lesion bypass mechanism by Dpo4, using single molecule FRET", (Manuscript in preparation).