1-1-2013

Unveiling Dna Polymerase Synthesis And Proofreading Activities - One Molecule At A Time

Kyle Vrtis
Wayne State University,

UNVEILING DNA POLYMERASE SYNTHESIS AND PROOFREADING ACTIVITIES – ONE MOLECULE AT A TIME

by

KYLE B. VRTIS

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

2013

MAJOR: CHEMISTRY (Biochemistry)

Approved by:

_____

Advisor Date

_____

_____

_____

_____

DEDICATION

This thesis is dedicated to my loving wife, Krish, for her unconditional love and support.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my two brilliant advisors, Dr. Louis Romano and Dr. David Rueda, for all of their guidance and support over the years. My advisors always kept an open door and were open to ideas and discussion. I am incredibly fortunate to have two advisors. Dr. Romano is an expert with biochemical techniques and DNA polymerases. Dr. Rueda is an expert with single molecule techniques and biophysics. Their collective wisdom was a great resource for troubleshooting and for interpreting complex data. Additionally, Dr. Romano and Dr. Rueda have an incredible talent for maintaining funding, which meant I could focus on research without having to worry about how we would pay for chemicals or equipment. I am thankful for my amazing committee, Dr. Christine Chow, Dr. Andrés Cisneros, and Dr. Arun Anantharam, for their constructive criticisms of my research, general advice, and continuous support.

I also need to thank Tom Christian for developing the smFRET approach that was the crux of my research, and for training me on how to carry out various DNA polymerase assays and single molecule experiments. Next, I am thankful to Radek, who conducted many of the experiments in this dissertation with me. Radek made research fun – even when our experiments weren't working. He really does deserve a Ph.D. and I hope he finishes his dissertation soon. Others in the Romano group also need to be acknowledged including: Ashley, Dr. Alfonso, Rick, and Pramodha. Ashley is very knowledgeable in all things DNA polymerases, which meant she was a great resource whenever I needed advice with any type of experiment besides single-molecule

experiments. In my four-plus years at Wayne State, Ashley never turned me down when I asked her for help. Alfonso has worked on the single molecule DNA polymerase project for several years now, and he has been a tremendous asset to our research. He is the go-to person whenever I need advice about single molecule experiments, biophysics, or DNA polymerases. Rick was incredibly talented at fixing instrumentation in the lab, and he was fun to have adventures with. He took me scuba diving and he taught me how to ride a dirt bike. Pramodha is the up and coming scientist of the Romano lab, and I know he will take the single molecule research to the next level. I am also grateful to all of the members of the Rueda lab past and present – Rui, Rajan, Zhoujun, Sharla, May, Marcus, Eric, Bishnu, Gayan, Hansini, Chandani, and Imali – for providing a friendly and productive atmosphere to conduct my experiments. Gayan is a great labmate, who was always willing to lend a hand. I enjoyed my many discussions with Gayan about research because he was very knowledgeable about all things single molecule.

I would like to thank all of my family and friends for their continuous encouragement, love, and support. I owe my deepest gratitude to my parents for always being there for me. They are incredibly dependable and I never doubted they love me. I am also thankful for my siblings – Erin, Cassie, James, and Bridget – I couldn't ask for better people to grow up with. I am thankful for my incredible friend, Noah, for all of the great times we have had over the years.

It has been incredibly difficult being away from my friends and family over the last few years. So difficult, in fact, that I sometimes wondered if the Ph.D. is even worth the time I spent away from home. The answer, of course, is a resounding "yes!" because

it was during my Ph.D. that I met my wife, Krishanthi Karunatilaka. Krish is my role model, best friend, and biggest supporter. She is reliable, trustworthy, funny, intelligent, and caring. I love her, and I am lucky to have her in my life. Finally, I would like to thank Krish's family for all of their support and encouragement over the years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

3-HPA	3-hydroxypicolinic acid
AAF	*N*-acetyl-2-aminofluorene
AF	*N*-2-aminofluorine
BaP	benzo[a]pyrene
ddNTP	2'3'-dideoxyribonucleoside 5'-triphosphate
dG-C8-AAF	*N*-(2'-deoxyguanosin-8-yl)- *N*-acetyl-2-aminofluorene
dG-C8-AF	*N*-(2'-deoxyguanosin-8-yl)-2-aminofluorine
DNA	deoxyribonucleic acid
DNA Pol I	*E. coli* DNA polymerase I
dNTP	2'-deoxyribonucleoside 5'-triphosphate
Dpo4	*Sulfolobus solfataricus* DNA polymerase IV
E. coli	*Escherichia coli*
ESI	electrospray ionization
FRET	Förster resonance energy transfer
HMM	hidden Markov model
HPLC	high performance liquid chromatography
k_{off}	dissociation rate
k_{on}	association rate
K_D	dissociation constant

KF	Klenow fragment (truncated *E. coli* DNA polymerase I)
MALDI	matrix assisted laser desorption ionization
NTP	either a ribo- or 2'- deoxy-nucleoside 5'-triphosphate
PAGE	polyacrylamide gel electrophoresis
PIFE	protein-induced fluorescence enhancement
RNA	ribonucleic acid
rNTP	ribonucleoside 5'-triphosphate
SDS	sodium dodecyl sulphate
smFRET	single molecule Förster resonance energy transfer
smPIFE	single molecule protein-induced fluorescence enhancement
Tris-HCl	tris-hydroxymethyl aminomethane hydrochloride

Chapter 1: Introduction

1.1. DNA polymerases

1.1.1. The genetic code and DNA polymerases

We recently celebrated the 6oth anniversary of the publication of one the most revolutionary scientific discoveries of the 20th century. On April 25, 1953 James D. Watson and Francis H. C. Crick published the correctly solved structural model of the carrier of the genetic code, DNA(1). They described the structure with remarkable detail: two antiparallel strands with a phosphate backbone on the outside and bases facing inward, 3.4 Å distance between base pairs, 10 base pairs per helical turn, and that adenine pairs with thymine and guanine pairs with cytosine. The model derived from Watson and Crick was simple and elegant, especially compared to other, three-helix models at the time(1-3). The complementarity of the two strands meant that if the sequence of one strand was known, the sequence of the other strand could easily be elucidated. Those accurate findings alone were Nobel Prize worthy, and the prophetic statements that followed were equally impressive.

They noted, “It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.” Later that year, Watson and Crick published another paper in which they hypothesized the mechanism by which DNA could be replicated. In short, one strand of the DNA could be used as a template to guide the synthesis of a new, complementary strand(4).

Three years after the structure of DNA was resolved, Author Kornberg published an article describing an enzyme that could incorporate deoxyribonucleotide triphosphates (dNTP) to a growing strand of DNA while using the original strand as a template(5). These enzymes, collectively known as DNA polymerases, are responsible for the faithful replication of the genetic code in viruses and all domains of life including archaea, bacteria, and eukaryotes.

1.1.2. The model polymerase, DNA polymerase I – Klenow fragment

From the date of its discovery in 1956 by Arthur Kornberg, *E. coli* DNA Polymerase I (Pol I) became one of the most intensely studied enzymes(5). *E. coli* Pol I has three separate enzymatic activities: 5'-3' polymerase, 5'-3' exonuclease, and 3'-5' exonuclease. The Klenow fragment (KF) of Pol I lacks the 5'-3' exonuclease domain, yet retains the full 5'-3' polymerase and 3'-5' exonuclease activities(6,7). The polymerase active site (pol site) is structurally highly conserved and resembles the shape of a right hand with distinct fingers, palm and thumb subdomains (**Figure 1**) (8). The static thumb and palm domains are responsible for stabilizing the interaction with the DNA duplex and hosting the catalytic residues necessary for phosphodiester bond formation, respectively(9,10). The fingers domain is dynamic and plays a critical role in binding of the incoming dNTP and subsequent alignment of the active site for the chemistry step(11,12). In crystal structures from close homologues of KF, this alignment requires the rotation of the fingers domain 40-46° toward the DNA to form the closed ternary

complex (**Figure 1a**), in which the dNTP is enclosed within the active site and appropriately positioned for phosphodiester bond formation(13-15).

1.1.3. Nucleotide selection from the pol site

A DNA polymerase must select the complementary dNTP to the templating base from a pool of NTPs each round of incorporation. In *E. coli*, this pool of nucleotides contains approximately 35-fold excess of incorrect NTPs (dNTPs and rNTPs) over the correct dNTP(16). *In vitro* the presence of mismatched dNTPs destabilizes the KF-DNA binary complex(12); however, why this would be advantageous for the polymerase *in vivo* remains unclear (see Chapter 3 for more information). Nonetheless, exonuclease-deficient KF can replicate with remarkable fidelity, only averaging an error once every 10^5 incorporations(17).

The fingers domain of unliganded KF freely rotates between the open and closed conformations. When KF is complexed with DNA, the open conformation is stabilized (**Figure 1a**)(18,19). It is from this conformation that the polymerase first checks the identity of an incoming nucleotide for complementarity to the templating base. If the incoming nucleotide is not complementary to the templating base, the fingers adopt an ajar conformation (**Figure 3c**)(18,19), which precludes the necessary alignment of the incoming NTP with the 3'-OH of the primer strand for the chemistry step. The ajar state is a recent development and it has been detected in a crystal structure of Bst pol(20) as well as in single molecule studies in KF(18,19) and Klentaq(21). If complementarity is met, the polymerase will further rotate the fingers domain to the closed conformation

(**Figure 1a**)(18,19). In the closed conformation, the α-phosphate of the incoming dNTP will be properly aligned with the 3'-OH of the primer terminus for phosphodiester bond formation(22). As the ajar state is most heavily populated with incorrect nucleotides, it likely a fundamental fidelity checking step in the nucleotide selection pathway(18,19).

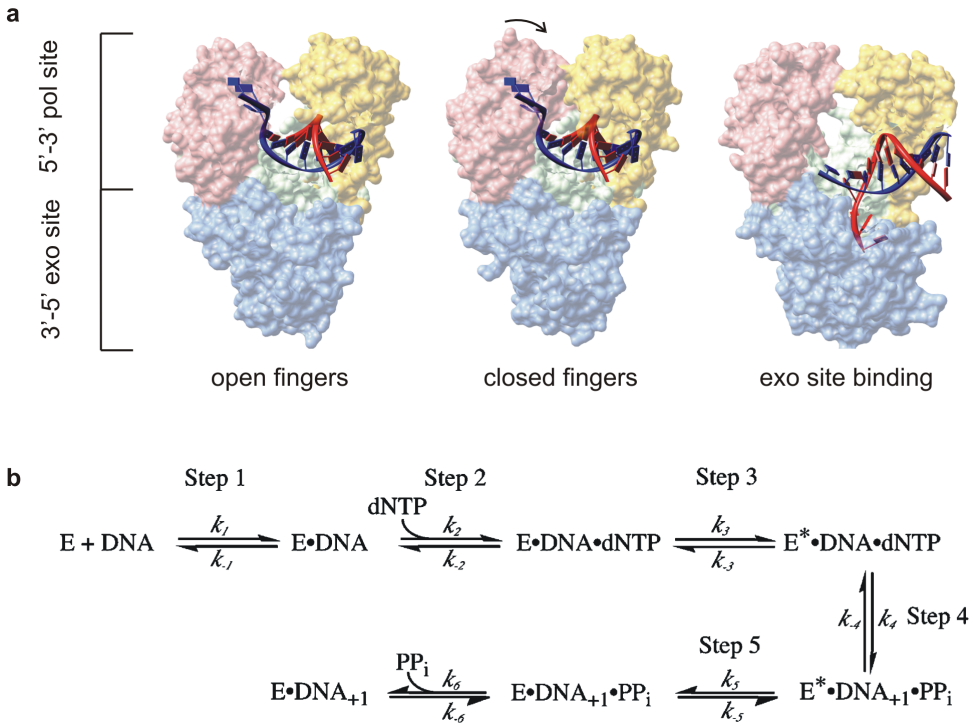

Figure 1. General reaction pathway for dNTP incorporation includes fingers closing motion. (**a**) DNA polymerase structures with DNA bound at either the pol site (open fingers conformation (PDB ID 1L3S) or closed fingers conformation (PDB ID 2HVI)) or exo site (PDB ID 1KLN). Template directed DNA synthesis occurs from the pol site, which includes the fingers (light red), thumb (yellow), and palm (light green) domains. Following a rare misincorporation, the primer strand (red) can be transferred to the exo site (light blue) to cleave off the incorrect nucleotide. (**b**) Minimal reaction pathway for nucleotide incorporation. Step 1 represents polymerase binding to the DNA, which is the only step not repeated during processive replication. During step 2, a dNTP is preselected from pool of NTPs. If the dNTP is correct, it triggers a conformational change that enables catalysis to take place (step 3). Upon phosphoryl transfer reaction completion (step 4), the complex reverses the previous conformational rearrangement (step 5) and the pyrophosphate is released. Adapted from (105).

1.1.4. Proofreading from the 3'-5' exo site

On the rare occasion in which KF misincorporates a mismatch, the terminal base pairs melt, KF transfers the primer strand to the 3'→5' exonuclease site (exo site) ~35 Å away from the polymerization site (**Figure 1a, exo site binding**), and KF catalyzes the excision of the incorrect 3' terminal nucleotide(23). This intrinsic 3'-5' exonuclease activity enhances the fidelity of KF ~7-fold(17). KF binds the DNA exclusively at its exonuclease site when two or more consecutive mismatches occur at the primer-template terminus(24). KF exonuclease activity can be abolished, while preserving the integrity of the site for binding DNA, by a single point mutation of a catalytic aspartate residue (D424A)(25).

1.2. Chemical carcinogen-induced mutagenesis

1.2.1. Carcinogens and cancer

Cells have an intrinsic system to control cell growth. This elaborate system involves the activation and suppression of biomolecules that signal the cell to proliferate, remain quiescent, or die. Cancer propagates from an accumulation of errors in genes that control cell growth, resulting in unregulated cell growth and division(26).

Interestingly, only 10 % of cancers are caused by inherited genetic defects, the remaining cases are caused by lifestyle and environmental factors (e.g. chemical carcinogens)(27). Many chemical carcinogens are metabolized within cells to form adducts on DNA bases(28). If left unrepaired, these adducts can have dire consequences during DNA replication.(28)

The degree of inhibition of a DNA polymerase by a carcinogenic adduct on the DNA can vary depending on the sequence context, type of DNA polymerase, type of lesion, and whether the polymerase bypasses the lesion from a running start (polymerase initially binds before the adduct position) or a standing start (polymerase initially binds at the adduct position)(29-32). However, most high fidelity DNA polymerases cannot bypass adducted DNA bases efficiently(28,33-35). Instead, lower fidelity DNA polymerases, such as Y-family DNA polymerases, carry out the trans-lesion synthesis so that replication can proceed(33-36). The frequency of mutagenesis near these sites can be quite high.

1.2.2. AAF and AF

N-acetyl-2-aminofluorene (AAF) is a synthetic compound that was originally developed as a pesticide(37). Fortunately, scientists discovered that AAF is carcinogenic before the compound could go to market(37). It has since been found that AAF is carcinogenic in many animals including: mice, rats, fish, dogs, cats, frogs, and bats(37,38). The types of cancers found in these animals included: kidney, liver, ear duct, testes, skin, breast, and intestinal cancer(37,38).

We now know that AAF is a potent, arylamine mutagen that, following metabolic activation *in vivo*, forms two different adducts at the C8 position of guanine bases: 2-aminofluorene (AF-dG) and *N*-acetyl-2-aminofluorene (AAF-dG)(39-41). Interestingly, while AF-dG and AAF-dG have similar chemical structures, they have drastically different effects on duplex DNA conformation and DNA synthesis.

AF-dG predominantly exists in the *anti* conformation in duplex DNA. In this conformation, the AF adduct is displaced into the major groove while dG remains base paired(42,43). On the other hand, AAF-dG, tends to be in the *syn* conformation due to a steric clash that occurs between AAF's acetyl group and the ribose sugar when dG-AAF is in the *anti* conformation(44). In the *syn* conformation, the AAF adduct is inserted into the duplex DNA while the guanine base sits in the major groove(45).

1.3. Single molecule microscopy investigation of DNA polymerases

Portions of this section were written with Alfonso Brenlla as part of a review paper.

1.3.1. Single molecule strategies to probe polymerase dynamics

Researchers have probed nearly every element of DNA replication with a wide array of single molecule tools. These techniques can reveal key transient intermediate states in reaction pathways and heterogeneity within biomolecule populations that would otherwise remain hidden in ensemble-averaged experiments (46-48). A partial list of single molecule approaches used to probe DNA polymerases is included in Table 1 along with example designs in **Figure 2**. Dozens of variations of these experiments exist, thus we will only highlight a few frequently used approaches in this review. Applications of these techniques will be discussed in later sections. The majority of single molecule experiments used to study DNA polymerases fall into two categories: single molecule fluorescence spectroscopy and single molecule force spectroscopy.

Precise spatial information can be obtained using single molecule fluorescence microscopy by attaching fluorescent reporter dyes to molecules of interest (48,49). In fact, by strategically conjugating donor and acceptor fluorophores on a DNA polymerase

or DNA, binding dynamics as well as conformational changes within the polymerase can be probed by single molecule Forester resonance energy transfer (smFRET) (18,19,21,50-54). Additionally, binary and ternary complex formation can be resolved without labeling the polymerase by protein-induced fluorescence enhancement (PIFE) (52,53,55-57). Advances in fluorescence microscopy equipment and analysis now enable researchers to track individual molecules in a living cell in three dimensions with a spatial accuracy well below the diffraction limit of light (58-60). Live cell imaging has allowed researchers to determine the stoichiometry of polymerases at the replisome under physiological conditions (61,62). Moreover, the assembly, collaboration, and functions of replisome components have been interrogated via rolling circle and flow stretching assays (63-65).

Single molecule force experiments manipulate biomolecules with external forces such as laminar flow, laser traps, or magnetic traps (66). Laminar flow is applied to stretch DNA for visualization of replisome-mediated DNA synthesis (65). Optical and magnetic tweezers force experiments tease out mechanical forces exerted as molecules undergo motion (67). In contrast to fluorescence experiments, biomolecules are not directly visualized; although, the biomolecule position may be implied from the force trajectories. These experiments have identified several hidden, elementary steps in polymerization and proofreading.

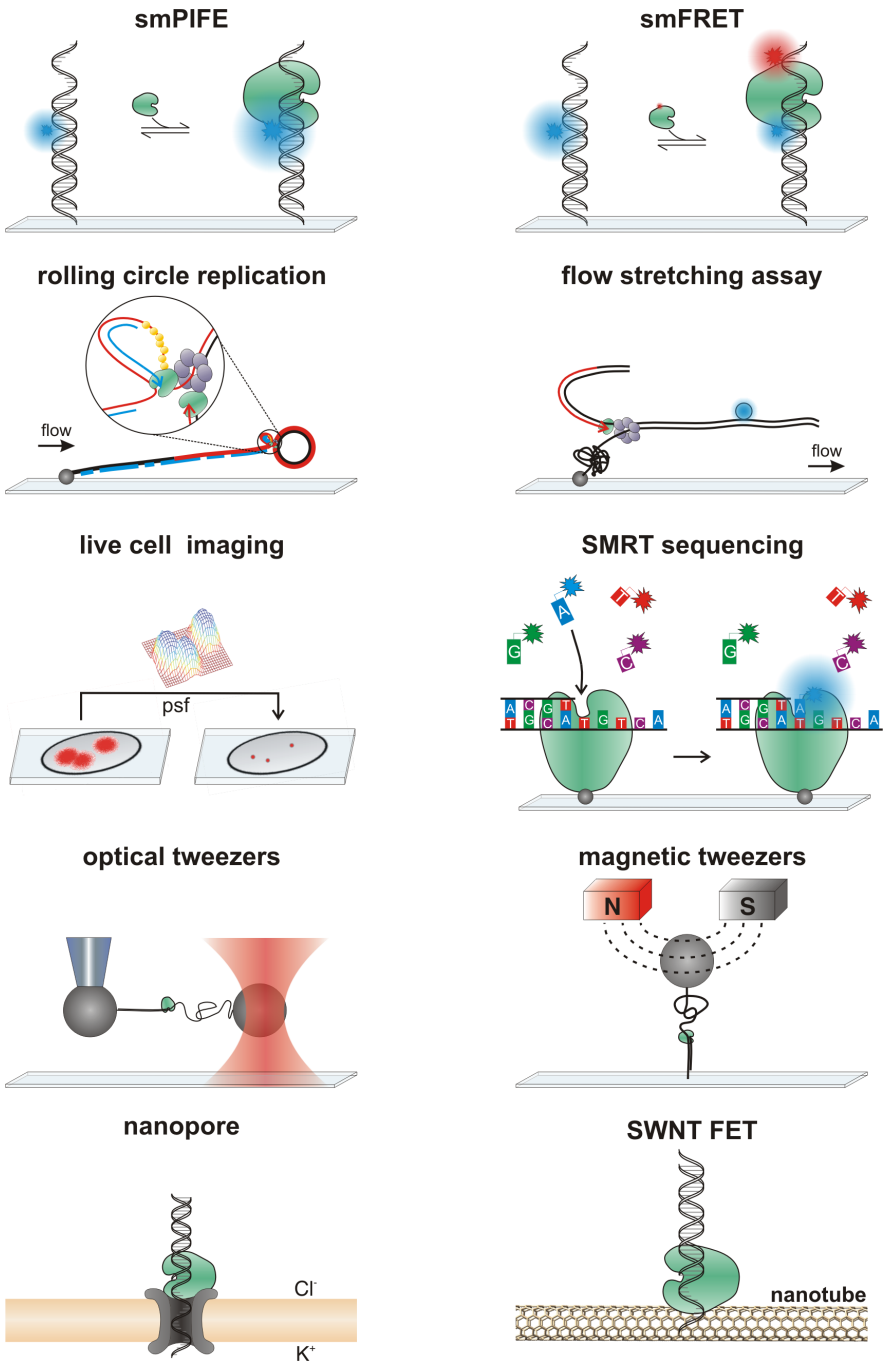

Figure 2. Single molecule strategies used to study DNA polymerases. See Table 1 for a detailed explanation of each approach.

Table 1. Single molecule approaches to study DNA polymerases

Technique	Description
smPIFE	smPIFE occurs when a DNA polymerase binds in close proximity to a fluorophore. The protein changes the local environment about the fluorophore in a favorable manner such that distinct spikes in fluorescence intensity are observed.
smFRET	smFRET reports the efficiency of energy transfer between an excited donor fluorophore and a nearby acceptor fluorophore. Energy transfer efficiency is strongly dependent on the orientation and distance between the donor and acceptor fluorophores. Consequently, smFRET can be used to measure short distance changes in the 3-8 nm range.
Rolling circle replication	Rolling circle replication uses closed, circular ssDNA as a template to make a long, continuous strand of newly synthesized DNA (leading strand). The leading strand is then used as a template for lagging strand DNA synthesis. By tethering one end of the DNA to a slide and flowing reaction buffer across the sample, the newly synthesized dsDNA can be detected with intercalating dyes as the replisome moves farther from the point of origin.
Flow stretching	Flow stretching assays use a hydrodynamic drag force to stretch the dsDNA while the ssDNA remains a random coil. As the replisome progresses the amount of tethered ssDNA increases and “pulls” the dsDNA closer to the tether. Replisome activity is determined from dsDNA displacement over time, which is visualized with a quantum dot or bead attached to the dsDNA.
Live cell imaging	Live cell imaging directly measures the real-time localization of fluorescently labeled biomolecules in the cell. Fluorescence intensities and centroids obtained from point spread functions (psf) of localized fluorescent spots reveal the stoichiometry and positions of the biomolecules at the replication fork, respectively.
SMRT sequencing	SMRT sequencing detects the fluorescence from distinguishable fluorescently-tagged dNTPs as the nucleotides are incorporated. DNA polymerases are surface immobilized in separate zero-mode wave guides to allow simultaneous reading of thousands of sequencing reactions in real-time.
Optical tweezers	Optical tweezers use a tightly focused laser beam to apply force to a dielectric bead attached to DNA tethered to a second bead that is held in place by surface immobilization, a pipette, or a second optical trap. The distance between bases in ssDNA and dsDNA differ at a given tension; consequently, the conversion between the two forms will change the total end-to-end distance of the DNA. Therefore, if the tension and the end-to-end length of the DNA are known, the fractions of ssDNA and dsDNA can be calculated. Polymerization and proofreading are detected as changes in the fractions of ssDNA and dsDNA over time.
Magnetic tweezers	Magnetic tweezers use magnets to generate a homogenous magnetic field to apply a lateral force to magnetic bead-conjugated DNA tethered a slide. Moving the magnets closer to the magnetic bead can modulate the force applied to the magnetic bead. Polymerization and proofreading are detected as changes in the position (normal to the slide surface) of the magnetic bead over time.
Nanopores	Nanopore experiments use patch clamp amplifier devices to monitor the ionic current passing through a nanopore over time. The electric field pulls the negatively charged DNA through the nanopore. DNA association with the nanopore decreases current passing through the nanopore. DNA polymerase-bound DNA decreases the current through the nanopore to a lesser extent than DNA only. Therefore the amplitude of the current directly reports the presence or absence of a DNA polymerase on the DNA.
SWNT FET	By conjugating a DNA polymerase to a single-walled carbon nanotube (SWNT) field effect transistor (FET) device, the fingers closing conformational change can be measured with submillisecond resolution. Polymerase motions transduce detectable electrical signals along the nanocircuit.

DNA polymerase activity has also been measured at the single molecule level by measuring the change in current generated during nanopore and single-walled carbon nanotube (SWNT) field effect transistor (FET) experiments with submillisecond resolution (22,68). Nanopore experiments measure the distinct decreases in current passing through the nanopore when DNA only or binary complexes bind to the nanopore (68-70). SWNT FET experiments monitor the change in current along the SWNT as DNA polymerases undergo conformational changes during polymerization (22).

1.3.2. Ajar fingers state selects for complementary NTPs

DNA polymerases replicate DNA with a fidelity rate far surpassing that expected based on the thermodynamics of complementary base pairing alone (71). This feat is particularly impressive considering the physiological pool of incorrect rNTPs and dNTPs far exceeds the correct dNTP (16). It is now clear that for most high fidelity DNA polymerases screening against incorrect nucleotides involves a series of elementary steps and a large conformational change in the fingers domain from an open to a closed fingers state (72). Protease digests of *E. coli* DNA polymerase I Klenow fragment (KF) ternary complexes with incorrect and correct dNTPs demonstrated that a conformational change consistent with fingers closure readily occurs with correct dNTPs, but is impeded with incorrect dNTPs (73). These results suggest fingers closure is used as a step in the nucleotide selection mechanism to select against incorrect nucleotides. X-ray crystallography studies have provided snapshots of polymerase structures that are believed to correspond to the open and closed fingers states. For bacteriophage T7 and

Klentaq DNA polymerases, the fingers closing resulted in 41° and 46° inward rotations of the O-helix of the fingers, respectively (13,14). Fingers closing properly aligns the α-phosphate of the incoming dNTP with the 3'-OH of the primer terminus for phosphodiester bond formation (13,74). Ensemble experiments with Pol β, Klentaq, and Klenow concluded the rate of fingers closing was too fast to be rate limiting (12,75,76). Furthermore, selection for complementarity between the incoming NTP and the templating base was conducted from the open fingers conformation, while selection for the correct deoxyribose sugar occurred as the fingers closed (12).

Formation of a closed ternary complex increases the stability of the polymerase on the DNA presumably due to an increase in the number of contacts among the polymerase side chains and the DNA (14,74,77). The increased stability of the ternary complex has been detected in numerous ensemble studies (77) as well as smPIFE (52), smFRET (53,78), and nanopore (68,70) experiments. Conversely, noncomplementary dNTPs and rNTPs destabilize polymerase binding to the DNA (52,70). A complementary rNTP slightly stabilizes polymerase binding to the DNA, which suggests that the polymerase is able to go farther along the reaction pathway with a complementary base than a noncomplementary one and the polymerase selects first form complementarity and second for the correct sugar (52,70). To our knowledge, there is no crystal structure of an incorrect nucleotide ternary complex. This may be because the complex is not stable enough to capture by crystallization. Furthermore, kinetic analysis suggests that mismatch discrimination doesn't occur from a single state, but from a mixture of states

(72). Single molecule spectroscopy is ideal for revealing the heterogeneous populations present during nucleotide selection.

Single molecule PIFE was the first single molecule approach used to directly monitor fingers opening and closing in real-time. Luo, *et al.* found the intensity of the Cy3 fluorophore linked to the 5' end of a template strand three bases upstream from the primer-template terminus increased whenever KF, T4, or T7 DNA polymerases bound to the DNA (55). Additionally, with T7 DNA polymerase, formation of a closed ternary complex with an incoming dNTP further increased the Cy3 intensity. The fingers closure rate was faster in the presence of correct dNTPs than incorrect dNTPs. This supports a mechanism by which screening against noncomplementary bases occurs prior to fingers closing, thus precluding formation of a closed ternary complex with incorrect dNTPs. While smPIFE is capable of monitoring fingers dynamics, it lacks the ability to precisely measure distances.

By using smFRET, Santoso *et al.* also characterized the fingers closing dynamics of KF, a close homologue of T7 polymerase (51). smFRET has been described as a spectroscopic ruler for its ability to measure small distance changes in the 20-80 Å range (79). The authors measured the fingers movements in real-time using smFRET by conjugating the two fluorophores of a FRET pair to the mobile region of the fingers domain and to a static position on the polymerase, respectively (**Figure 3a**). These smFRET experiments were designed to have the distance between the FRET pair fluorophores longer in the open fingers state than in the closed fingers state. FRET efficiency increases as the distance between the two fluorophores decreases; therefore,

the closed fingers state has higher FRET than the open fingers state. This approach revealed that in the absence of DNA or nucleotides, KF readily oscillated its fingers between the open and closed conformations in the millisecond timescale. Upon KF binding to the DNA, the open fingers state predominated (~84%) over the closed state (~16%) (**Figure 3b**). Stabilization of the open fingers state situates the polymerase in a conformation that is readily accessible to incoming nucleotides. Formation of the correct ternary complex shifted the equilibrium to the closed fingers complex (95%) (**Figure 3b**).

Interestingly, the low FRET state shifted to slightly higher values with the formation of any ternary structure (correct or incorrect) due to the fingers pausing at a unique conformation between the open and closed states (19). This partially closed structure, termed ajar, was also identified by a similar smFRET strategy in the Klenow fragment of *Thermus aquaticus* DNA polymerase I (Klentaq) (21). This conformation is believed to be the same structure crystallized by Wu and Beese with BF (the large fragment from *Bacillus stearothermophilus* DNA polymerase I), in which the fingers domain was captured in an ajar state (**Figure 3c**) (20). Existing evidence points towards this ajar state as a structural checkpoint from which the polymerase probes incoming nucleotides for complementarity to the templating base prior to fingers closure. In support of this idea, two separate smFRET studies with KF and one with Klentaq found that the ajar fingers conformation was preferentially populated (~69%) when an incorrect pol-DNA-dNTP structure was formed (**Figure 3b**) (18,19,21).

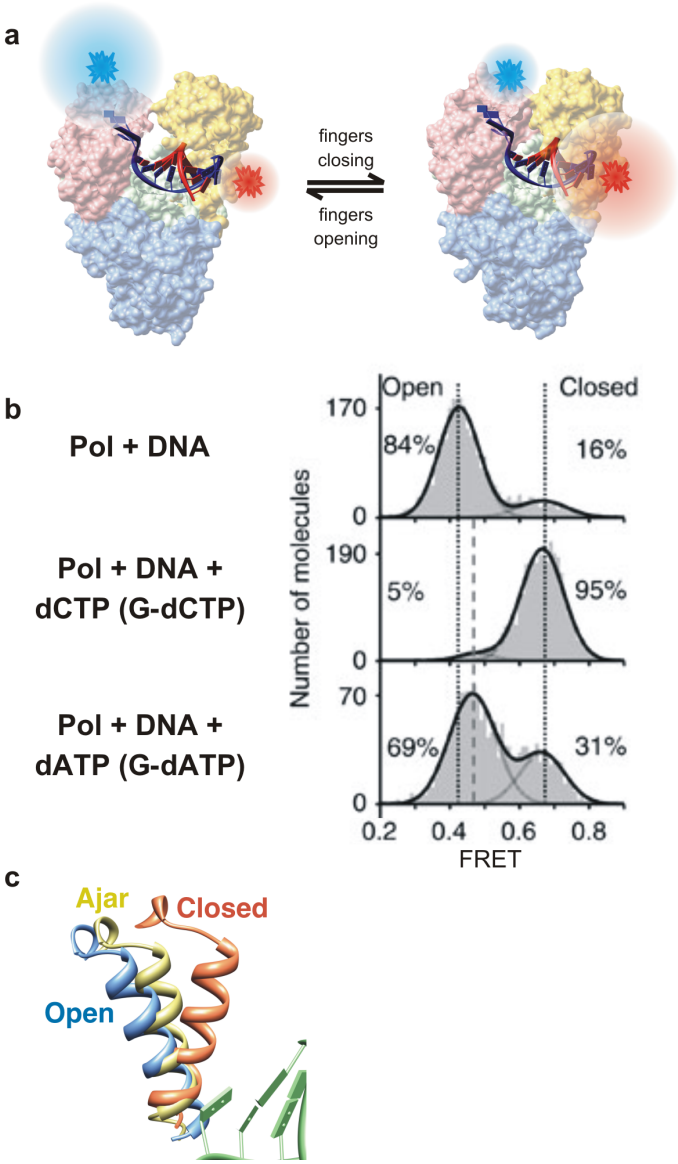

Figure 3. Ajar fingers state detected at the single molecule level. (**a**) Schematic of the single molecule design used to monitor fingers movement by FRET. As the fingers domain closes, the distance between the donor (blue) and acceptor (red) fluorophores decreases, which results in a higher efficiency of resonance energy transfer (FRET). (**b**) FRET histograms for binary, correct ternary (G-dCTP), and incorrect ternary (G-dATP) complexes. The two dotted lines mark the center of the FRET distributions for open and closed fingers states, as indicated. The dashed line marks the center of the FRET distribution for the ajar fingers state. Adapted from ref. (39). (**c**) Structural comparison of the proposed open (blue, PDB ID 1L3S), ajar (yellow, PDB ID 3HPO), and closed (orange, PDB ID 2HVI) fingers conformations.

Berezhna *et al*. observed KF fingers motion by attaching an acceptor fluorophore to a mobile portion of the fingers domain and a donor fluorophore to a static position on the DNA (18). This design allowed the authors to not only detect the fingers open, ajar, and closed states, but also binding at the exo site. Intriguingly, these studies revealed an unexpected behavior of the polymerase in which it bound to the exo site more frequently in the presence of incorrect nucleotides (13% exo site for binary complex, 25% for incorrect ternary complex). The authors suggest that the increase in exo site binding may be attributed to the exo site serving a secondary function in nucleotide selection prior to incorporation. Alternatively, the increase in exo site binding relative to pol site binding may also be attributed to destabilization of pol site binding by the incorrect dNTP.

A similar smFRET design was used with a B family, replicative DNA polymerase from *Solfolobus solfataricus*, DNA polymerase B1 (PolB1) (53). This design also allowed the real-time detection of fingers movement and exo site binding via the changes in FRET; however, the ajar state was not detected with PolB1. This could be because the FRET peaks corresponding to the open and closed states had significant overlap, thus making it difficult to resolve an intermediate FRET state, or it could be that PolB1 uses an alternative mechanism for nucleotide selection. Notably, PolB1 has a unique extended fingers structure, which may be involved in discrimination against incorrect nucleotides.

Y-family polymerases present unique structural and catalytic features among all polymerases families, suggesting that the dNTP incorporation pathway may be different with respect to other replicative high-fidelity polymerases. Recent smFRET experiments by Brenlla, *et al.* with the acceptor fluorophore on an immobile portion of the low-

fidelity, Y-family repair DNA polymerase, *Solfolobus solfataricus* Dpo4, revealed the polymerase has an interesting approach to nucleotide incorporation (**Figure 4a**) (Alfonso's data, submitted). In the absence of dNTPs, the Dpo4 binary complex rapidly shuttled between preinsertion and insertion conformations (**Figure 4b**, top). This is in contrast to smFRET studies with the A family DNA polymerase, KF, in which the KF-DNA binary complex remained stationary. Interestingly, the correct Dpo4 ternary complex is stabilized in the insertion conformation (**Figure 4b**). This suggests a mechanism whereby the correct dNTP facilitates translocation along the DNA from the preinsertion conformation to the insertion conformation, an intriguing feature that has not been detected with any other DNA polymerase at the single molecule level.

The differences observed at the single molecule level among different DNA polymerase families suggest while the fingers closing motion may be a universal mechanism for nucleotide incorporation, the role of this motion may not be highly conserved. These changes in the mechanisms may have emerged out of necessity to perform particular roles in the cells. For example, Dpo4 is a lesion bypass polymerase with a wide active site while T7 is a replicative DNA polymerase with a tight active site.

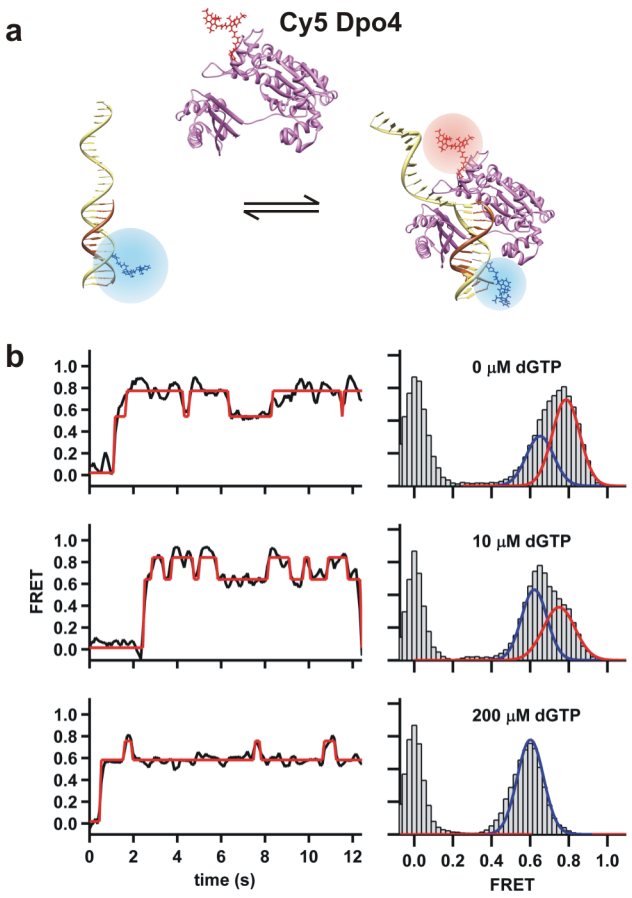

Figure 4. Correct dNTP stabilizes the insertion complex for Dpo4. (**a**) Schematic of single molecule FRET design used to observe Dpo4 dynamics on the DNA. Upon polymerase binding to the DNA, energy is transferred from the donor fluorophore (blue) on the DNA to the acceptor fluorophore (red) on the DNA polymerase. (**b, left**) Since the efficiency of FRET is distance and orientation dependent, DNA polymerase dynamics on the DNA can be monitored in real time from fluctuations in FRET trajectories. (**b**) Dpo4 shuttles between a pre-insertion site (FRET ~0.60) and an insertion site (FRET ~0.80). The insertion site is stabilized by formation of the complementary ternary complex (dGTP is the correct nucleotide).

1.3.3. Fidelity-checking following incorporation

In addition to the nucleotide selection step that precedes the chemistry step, DNA polymerases have also developed checkpoints following phosphodiester bond formation to assess the nascent base pair for complementarity (80). This fidelity-checking step can

be vaguely described kinetically: the polymerization rate is faster than proofreading when the polymerase is extending complementary DNA, and polymerization is slower than proofreading when the terminal base pair is a mismatch (77,81). However, the structural and mechanistic basis for fidelity-checking has remained unclear from ensemble studies.

In order to capture the moments immediately following an incorporation in real-time at the single molecule level, Christian, *et al*. designed an smFRET approach to observed individual incorporations (50). By attaching a donor fluorophore on the DNA and an acceptor fluorophore on an immobile portion of KF (**Figure 5a**), the FRET was determined for incrementally increasing primer-template lengths from a 15mer/28mer primer-template to an 18mer/28mer primer-template (**Figure 5b**). The differences in FRET between each of the primer-templates were large enough so that the polymerase position on the DNA could be tracked with base-pair resolution from a FRET trajectory.

Incorporation of a single nucleotide to the 15mer/28mer primer-template resulted in a change in FRET from ~0.68 (15mer FRET) to ~0.62 (16mer FRET) as the polymerase translocated from the 15mer to 16mer position (**Figure 5c**, top). A second incorporation to the 15mer/28mer primer-template resulted in another drop in FRET to ~0.47 (17mer FRET) (**Figure 5c**, bottom). Intriguingly, immediately following each incorporation was a dip in the FRET to a value that was consistent with the polymerase binding to the n+2 position, before the FRET returned to the n+1 site (**Figure 5b**). The unexpected decrease in FRET beyond the next primer-template position was attributed to the polymerase translocating approximately one base pair farther along the template than the next preinsertion site to place the nascent base pair into a fidelity-checking site. Accordingly,

if complementarity is met, the polymerase will return to the next pre-insertion site; however, if a mismatch is present the polymerase will transfer the primer from the fidelity-checking site to the exonuclease site for proofreading. Recent computational work with KF and *Bacillus stearothermophilus* DNA polymerases identified several residues that could be involved in this putative fidelity-checking activity (82). Future mutational studies in combination with smFRET should shed light on the fidelity-checking mechanism.

Optical and magnetic tweezers experiments have found that T7 DNA polymerase and KF organize multiple template bases in the pol site during each incorporation cycle (83,84). This organization has been associated with the movement of the polymerases during the rate-limiting step of the incorporation cycle. An interesting possibility is that this motion is related to the fidelity-checking step observed with Klenow by smFRET. If so, these developments immediately support a mechanism by which the polymerase performs fidelity-checking immediately following each incorporation.

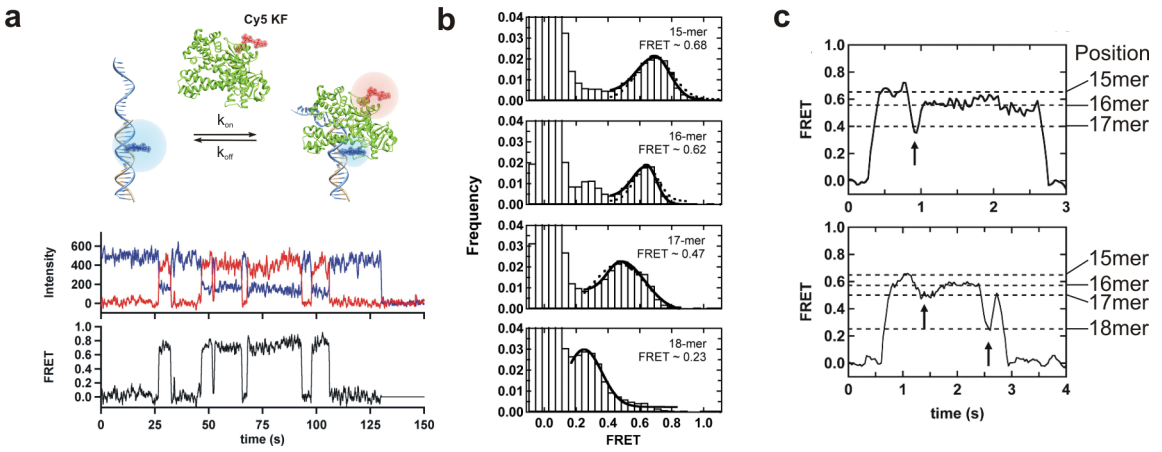

Figure 5. Fidelity-checking step identified in the incorporation cycle. (**a, top**) Schematic smFRET assay used to monitor polymerase dynamics on the DNA with single-base pair resolution. The design is nearly identical to the set up used in Figure 4a, except KF is used instead of Dpo4. (**a, bottom**) Representative intensity and FRET traces for polymerase binding to the DNA. The top trajectories show the donor (blue) and acceptor (red) fluorescence intensities over time. KF binding to the DNA resulted in an anticorrelated drop in donor intensity along with a concurrent increase in acceptor intensity. Donor photobleaching occurred at ~130 s. The bottom trajectory is the calculated FRET from the donor and acceptor intensities. (**b**) FRET distributions for KF binding to incrementally increasing primer-template lengths (from 15-mer to 18-mer). KF binds most stably at the primer-template terminus and the donor position on the template was fixed. Therefore, by incrementally increasing the length of the primer, acceptor-labeled KF bound farther from the donor on the DNA and the FRET decreased. (**c**) Two example FRET traces from KF incorporating dNTPs to the 15-mer primer template. Incorporation events were clearly distinguished as drops in the FRET over time. Dashed lines mark different primer-template lengths. Interestingly, following each incorporation, the FRET dropped (arrows) to a value that was consistent with the transient translocation past the next insertion site – possibly due to fidelity-checking. Adapted from reference (35).

1.3.4. Pol to exo site switching through an intermediate state

DNA polymerases with an intrinsic exonuclease activity catalyze 5'-3' polymerization and 3'-5' proofreading from spatially separated active sites (23,71,81). The physical separation of the two sites allows polymerases to catalyze DNA synthesis

without substantial competition from proofreading, which would slow down DNA synthesis. However, the separation necessitates a transfer mechanism for shuttling the DNA between the two sites when proofreading is necessary.

A recent smFRET study by Lamichhane *et al*. attached a donor fluorophore to the DNA and an acceptor fluorophore to the thumb domain of KF to monitor the polymerase dynamics on the DNA in real-time (85). The authors were able to directly differentiate between pol and exo site binding from the distinct FRET values for each site. KF directly associated and dissociated from DNA with a single mismatch at the primer-template terminus from both sites. Furthermore, the polymerase transferred the DNA between the pol and exo sites without dissociation. These results suggested there are two mechanisms by which the polymerase can bind to the DNA at the exo site: (1) intermolecular transfer in which the polymerase dissociates and another polymerase binds to the DNA from its exo site, or (2) intramolecular transfer between the pol and exo sites without dissociation from the DNA. Interestingly, dNTPs in solution increased the rate at which the primer was transferred to the exo site, which implies a role for dNTPs in proofreading.

Vrtis, *et al*. were also able to easily differentiate between KF pol and exo site binding using smPIFE and smFRET (54). KF bound to a single mismatch primer-template from an intermediate state that was distinct from normal pol and exo site binding. Notably, this intermediate state was also detected with primer-templates with either *N*-acetyl-2-aminofluorene (AAF) or 2-aminofluorene (AF) carcinogenic adducts at the duplex DNA terminus. The fact that the intermediate state was detected with three different aberrant DNA termini suggests the intermediate state likely plays a role in proofreading. To

further support this hypothesis, direct transitions between the intermediate and exo sites were observed without polymerase dissociation from the AAF-adducted DNA. Moreover, simulations revealed that the intermediate state was an obligatory intermediate in the mechanism by which KF bound to AAF-modified DNA at the exo site.

Optical tweezers experiments provided additional validation for an intermediate state between pol and exo site binding (84,86). The measured polymerization rates were comparable to ensemble values when the tension applied to the DNA was low (<10 pN) (86). DNA synthesis velocity decreased as the tension increased until ~37 pN, at which point the polymerase stalled on the DNA. High tension (>46 pN) applied along the length of the template strand by an optical trap promoted a force-induced exo activity (FIEA) with T7 and ϕ29 DNA polymerases (84,86). FIEA could be reversed by decreasing the tension below ~37 pN. Intriguingly, the net velocity of the polymerase was ~ 0 nt/s when the tension applied to the template was between ~37 to ~46 pN (86). This stalling behavior was attributed to an off-pathway obligatory intermediate state between pol and exo site binding (86).

The intermediate state was also detected during strand-displacement DNA synthesis using a magnetic trap approach (87). A DNA hairpin template was tethered between a glass surface and a magnetic bead captured by magnetic tweezers (**Figure 6a**). DNA synthesis through the duplex DNA hairpin required the polymerase to unwind dsDNA to expose the ssDNA template in order to extend the primer, a process that is difficult for most replicative polymerases. Polymerization and exonuclease activities were tracked from individual DNA polymerases from monitoring the change in the DNA extension

over time (**Figure 6b**). As mechanical force was applied to each end of the hairpin (to assist melting of the hairpin base pairs), polymerization was stimulated. At low force, T7 and T4 holoenzymes were incapable of extending the primer strand through the hairpin. In fact, the regression pressure from the DNA fork upstream of the holoenzymes shifted the equilibrium from polymerization to proofreading, and the holoenzymes processively degraded the primer. Prominent pauses were detected during both elongation and degradation.

Thorough analysis of the pause durations as a function of force led to a new kinetic scheme for switching between polymerization and degradation during strand displacement DNA synthesis (Figure 6c). Importantly, this model includes an inactive obligatory intermediate state between pol and exo site binding, which is consistent with the smFRET (54) and optical tweezers (84,87) studies with KF, T7, and ϕ29 DNA polymerases, respectively. The reverse pathway from the exo to the pol site could pass through the intermediate state, or the primer could directly transfer from the exo to the pol site (87).

The ability to manipulate the degree of regression pressure ahead of the holoenzymes using magnetic tweezers was used to cause a DNA polymerase to switch between polymerization and degradation. This principle forms the basis to quickly achieve multiple reads of the same DNA sequence during single molecule DNA sequencing (technique termed cyclic polymerase assay, CPA) (87). *In vitro* magnetic tweezers assays have shown that the regression pressure from the upstream DNA fork is released by a helicase unwinding the DNA ahead of the polymerase (88). Therefore, the processive

exonuclease activity induced by upstream fork regression pressure may act to restore symmetry following disassembly of a stalled replisome. These results emphasize the importance of studying DNA polymerase function in the context of an assembled replisome.

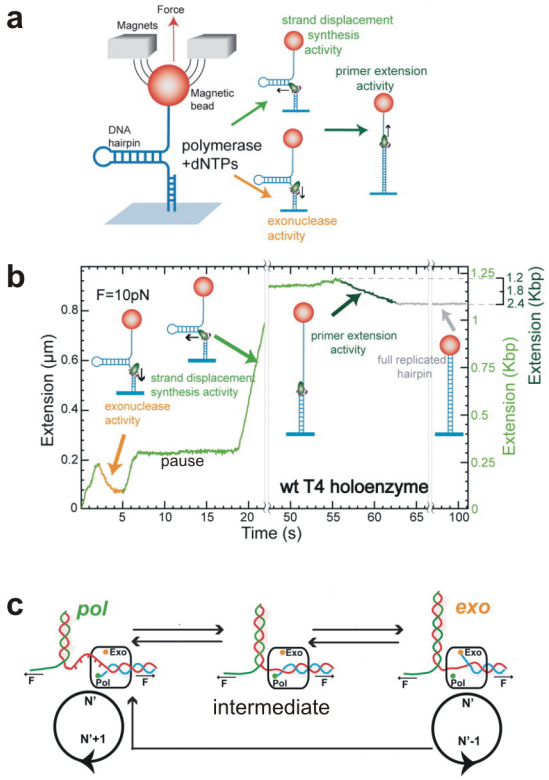

Figure 6. Magnetic tweezers detection of an intermediate state between pol and exo site binding. (**a**) Schematic representation the magnetic tweezers set-up used to observed strand displacement DNA synthesis. DNA was tethered between a slide and a magnetic bead. Magnets trapped the bead and applied force to stretch the DNA. Once the polymerase synthesized DNA up to the loop of the DNA, the assay switched from measuring strand displacement synthesis to primer extension synthesis. (**b**) The change in the extension of the DNA over time reports the polymerase activity on the DNA. Polymerization (light green) and exonuclease (yellow) activities in the hairpin were tracked by increases or decreases in the DNA extension over time, respectively. At this tension, primer extension activity (dark green) resulted in a decrease of the DNA extension over time. Distinct pause events were detected during polymerization and proofreading. (**c**) Proposed kinetic scheme for the primer transfer pathway in strand displacement. Shuttling from the pol to exo site passed through an obligatory intermediate, while the exo to pol site transfer can sometimes bypass the intermediate state. The dissociation pathway through the intermediate state and the inactive polymerization intermediate were removed from this scheme for simplicity. Adapted from reference (76).

1.3.5. Exploring replisome dynamics

Replicative DNA polymerases perform most of their activity *in vivo* as part of a multiprotein complex called the replisome. This molecular machinery carries out replication of a whole genome in minutes with a high level of accuracy. The 5'→3' directionality of enzymatic DNA synthesis implies that one of the nascent DNA strands is elongated continuously in the same direction as the growing replication fork (leading strand), while the other strand is extended in the opposite direction (lagging strand) in short segments known as Okazaki fragments (OFs) (89). As all DNA polymerases require a primer to catalyze the incorporation of dNTPs, OF maturation begins with a primase reading specific sequences in the template and synthesizing short, complementary RNA oligos. Next, a lagging strand DNA polymerase extends the primed oligos until the beginning of the previous OF is reached. This cycle is repeated every 1-2 kb for prokaryotic organisms (90) or every 100-200 bases for eukaryotes (91). Remarkably, leading and lagging strand synthesis takes place in a coordinated fashion, ensuring that both parental DNA templates are duplicated at the same time. Despite the fact OFs were discovered in the 1960s, it was not until recently that development of new single molecule techniques allowed the elucidation of leading and lagging strand coordination mechanism and characterization of the replisome dynamics in general.

T7 and *E. coli* replisomes are the most thoroughly studied to date at the single molecule level. The bacteriophage T7 replisome has been widely used as a model system due to its simplicity; in fact, it can be reconstituted *in vitro* with only four proteins (92). Although significantly more complex, the *E. coli* replisome can also be reconstituted in

vitro from 13 purified proteins (93). In 2007, Kim *et al*. (65) developed a single molecule flow stretching DNA assay that allowed multiplexed monitoring of DNA synthesis in real time. A DNA fork is surface-immobilized on a microscope slide and a constant laminar flow exerts a small force on the DNA large enough to keep dsDNA stretched while low enough to allow ssDNA coiling (**Figure 7b**). Replisome components were introduced in solution and assembled onto the fork. DNA synthesis changed the ratio between ssDNA and dsDNA, thus altering the overall length of the construct. DNA replication progress is tracked from the position of a bead or a quantum dot attached to the DNA. As the relation between the number of base pairs and end-to-end length of DNA was known, rates and processivity values were calculated from individual traces. The rolling circle replication assay is an alternative approach that also enables observation of DNA synthesis rates and processivity in real time at the single molecule level (64). In this experimental setup, a linear DNA oligo with a free 3'-OH group is surface-immobilized on a microscope slide. A complementary circular template DNA strand is annealed to the immobilized oligo. As DNA polymerases and accessory proteins are introduced in solution, DNA replication starts. The polymerase bound to the DNA junction on the circular template corresponds to the leading strand polymerase. As the leading strand is synthesized, it is unwound from the circular template, leaving behind a tail of ssDNA (**Figure 2**). If the enzymes involved in lagging strand synthesis are present, the displaced ssDNA can be primed and replicated in discrete OFs. Immobilized DNA is stretched by a constant laminar flow and visualized using fluorescent dsDNA intercalating dyes such as Sytox Orange or YO-PRO1.

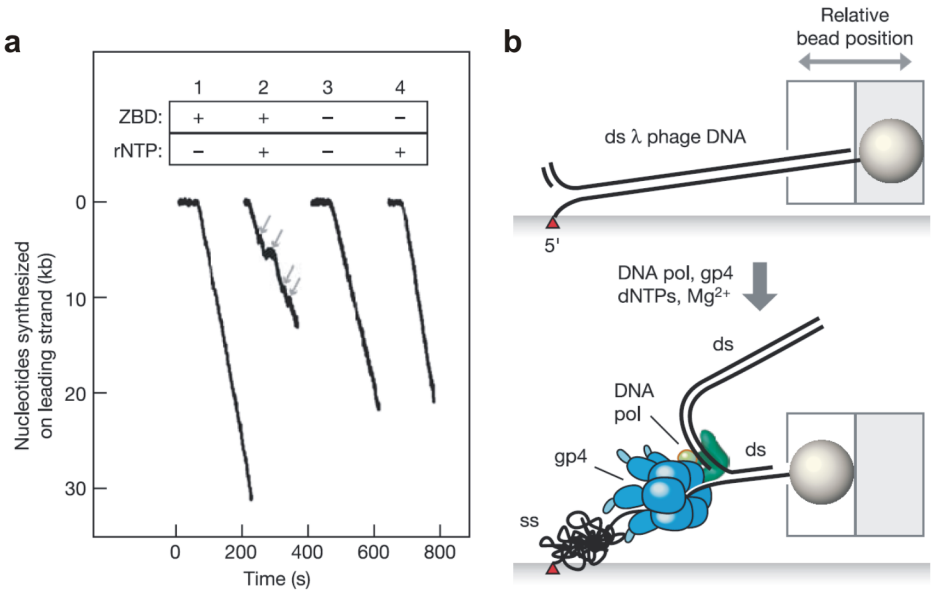

Figure 7. Leading strand synthesis of the T7 replisome measured with flow stretching assay. (**a**) Representative traces showing DNA replication progress in real time. The presence of ribonucleotides (rNTP) and gp4 zinc-binding domain (ZBD) are known to be required for primase activity, which results in pauses (shown as arrows in the second trace). (**b**) Schematic representation of the single-molecule setup. The DNA is stretched by a laminar flow. Upon introduction of replisome components and dNTPs, leading strand synthesis begins and the overall length of the construct is decreased. Figure adapted from (83).

In a pioneering work by Lee *et al*., the authors used a single molecule flow-stretching assay to measure DNA replication rates in real time (94). Leading strand synthesis for the T7 replisome halted while RNA primers were synthesized in the lagging strand. Replication stalling only took place when both primase and rNTPs (required for primer synthesis) were present (**Figure 7a**). These results suggested that primase activity was required to induce a conformational change that stalled Pol_{lead}. Pausing events preceded lagging strand synthesis and loop formation, thus supporting a model whereby primase activity induces Pol_{lead} stalling. Using a similar setup, Hamdan *et al*. characterized the

dynamics governing lagging strand synthesis (95). They monitored loop length and lag times between loop synthesis events at different rNTP concentrations. A thorough statistical analysis of individual loop lengths and lag times suggested that primer synthesis and OF completion both trigger loop release. Combined with the work by Lee *et al*., these two papers provided some initial understanding on the dynamics governing leading and lagging strand synthesis coordination.

Loparo *et al*. studied the T7 replisome machinery to investigate the polymerase exchange mechanism (96). The authors designed an elegant flow stretching assay with fluorescently labeled polymerases and a quantum dot attached to the DNA. The simultaneous tracking of the quantum dot position and the quantification of the fluorescence intensity at the replication fork allowed DNA elongation rates and T7 DNA polymerase stoichiometry at the replisome to be simultaneously determined in real time. Up to three polymerase units bound to an active replisome were observed. However, the presence of three polymerases at the replisome does not establish that polymerase switching occurs during DNA replication. In order to characterize Pol_{lead} exchange kinetics, the authors used a mutant with slower polymerization rates. Therefore, a change in DNA synthesis rate was observed when mutant and wt polymerases exchanged. In the final proposed model, once a polymerase binds the replisome, it can dissociate or exchange with the active Pol_{lead} at roughly the same rates (~ 0.02 s^{-1}).

The *E. coli* replisome is significantly more complex than the T7 replisome. However, it more closely resembles other prokaryotic or eukaryotic replisomes, thus making it an attractive model system to characterize. Tanner *et al*. measured DNA replication rates in

the absence (ssDNA as substrate) and presence (dsDNA as substrate) of DnaB helicase, finding that DnaB increased processivity by a factor of ~7.5 (97). This result pointed towards a stabilizing effect when the number of DNA-protein contacts was increased. On the other hand, DnaG primase binding decreased processivity by a factor of ~3 independently of DnaG activity. This destabilization could have been due to the loss of contacts between the helicase and the DNA-polymerase complex as it is known that primase interacts with the helicase during primer synthesis (98). Challenging bulk studies, the authors found that DnaG binding to the replisome was cooperative. A similar study by Yao *et al*. found that lagging strand synthesis decreases fork progression rate (factor ~0.8) but increases processivity (factor ~1.6) (99). The authors also measured an increase in processivity upon increasing the polymerase concentration in solution. Using a rolling circle replication assay, Tanner *et al*. showed that β-clamps from a finished OF can be reused for the next one (100). This example constitutes further evidence of the flexibility of the replisome: if essential components cannot be captured from solution, the replisome may reutilize the prebound ones in order to continue DNA replication.

The recent development of live-cell single molecule techniques opened the possibility of characterizing the replisome in physiological conditions. Fluorescent-fusion proteins consist of a functional protein fused with a fluorescent one. By using slimfield microscopy, fused protein positions can be tracked in living cells with millisecond time resolution. Reyes-Lamothe *et al*. tracked and quantified the number of polymerases in the replisome of *E. coli* cells (61). They found that three polymerases could be associated with the replisome, instead of the historically accepted two. This finding helps to explain

how replisomes maintain high processivities while exchanging polymerases in a dynamic fashion: active polymerase exchange may only take place when a replacement polymerase is already present at the replisome. In follow-up work by Lia *et al*. using a very similar setup, the authors monitored OF synthesis dynamics by using fluorescent chimeric ssDNA binding proteins (62). They found that Pol_{lag} is likely renewed for each OF *in vivo*. In case the binding of a new Pol_{lag} from the solution is delayed, the third Pol III in the replisome is readily available and DNA replication stalling is prevented. Recent work by Georgescu *et al*. provided strong supporting evidence that *E. coli* tripolymerase replisomes are more efficient than the ones with only two polymerases (101). An *in vitro* rolling circle replication assay with DiPol and TriPol replisomes showed that TriPol replisome processivity is higher by a factor of ~2 even in the presence of excess Pol III in solution. In addition, DiPol replisomes leave more ssDNA gaps in the lagging strand. These features provide an explanation for the unexpected TriPol replisome observed *in vivo*.

The spatial organization of sister replisomes emerging from an origin of replication is subject to debate as to whether or not the sister replisomes remain physically coupled during replication (102-106). Yardimci, *et al.* applied a flow stretching assay to directly visualize replication products and proteins via fluorescence microscopy in a vertebrate model system using *Xenopus leavis* egg extracts (103,107). If the sister replisomes remain coupled during DNA replication, the dsDNA will be "pulled" toward the origin of replication during DNA synthesis (103). By tethering both ends of the dsDNA to the surface, "pulling" should be inhibited due to an accumulation of tension along the

surface-tethered DNA. Interestingly, their data showed sister replisomes efficiently synthesized DNA in opposing directions without physical coupling (103). This result was also reproduced with large T antigen, a DNA helicases from simian virus 40 (104). Unexpectedly, this experimental design also showed that large T antigen could bypass a protein barrier covalently attached to the DNA (104).

1.3.6. Non-canonical DNA may alter the dNTP incorporation pathway

DNA damage usually blocks or slows down primer extension by DNA polymerases (33). Regarding undamaged DNA, specific hot spot sequences increase the probability of misincorporations and therefore mutagenesis (108). The general mechanism for DNA elongation described in the introduction may present small variations depending on the DNA substrate context. Naturally modified or damaged bases, mismatches, sequence context or the presence of secondary structures may alter the polymerase binding mode and the dNTP incorporation pathway. Single molecule techniques enable us to monitor polymerase binding modes and reaction kinetics in real time, thus allowing the observation of alternative reaction pathways and intermediates otherwise hidden in bulk measurements. This section will focus on different reaction/binding modes observed as a function of the DNA substrate.

Single molecule real time sequencing (SMRT) is a third generation DNA sequencing technique that monitors dNTP incorporation events in real time at the single molecule level (109). Highly processive ϕ29 DNA polymerases are surface-immobilized with DNA and phosphate-labeled dNTPs in solution (**Figure 8a**). The conjugated dyes are different for each one of the four dNTPs. During a DNA elongation cycle, a cognate dNTP binds

to the DNA/Pol complex and the concomitant fluorescent light (characteristic for each dNTP) can be detected above the background noise as a pulse. Upon dNTP incorporation, the cleaved phosphate with the fluorescent dye diffuses away leaving an unmodified DNA substrate behind ready for a new elongation cycle. Simultaneously reading thousands of molecules and consensus-alignment of the output data allow sequencing whole genomes within a few hours. Additionally, kinetic data for dNTP incorporations over a wide range of DNA sequences are obtained. The pulse width (PW, **Figure 8b**) reflects the time between dNTP binding to the complex and cleaved pyrophosphate leaving. For unmodified DNA, the pulse width does not vary significantly with dNTP identity, showing an average value of ~100 ms. This result suggests that the incorporation pathway is very similar for all four dNTPs. The interpulse duration (IPD, **Figure 8b**) corresponds to DNA translocation plus dNTP binding times. IPDs show an increase at DNA regions with secondary structures such as hairpins. As PWs are mostly determined by a chemical process, it is not surprising that PWs remain almost independent of DNA secondary structures.

SMRT sequencing technology allows detection of several DNA modifications (110-112). Modified bases cause changes in elongation kinetics, altering PWs and IPDs. Different modifications affect polymerases in different ways, allowing discrimination between them. Similarly to DNA secondary structure, modified bases increased IPDs more than PWs. In general, bigger DNA structure distortions induced longer IPDs. Variations in IPDs were usually observed from positions -3 to +7 with respect to the modified base. Those positions mostly overlap with the polymerase footprint, suggesting

that the distortion in the DNA structure caused by the modified base hinders DNA translocation after each elongation cycle. Methylated bases such as N6-methyladenine, 5-methylcytosine and 5-hydroxymethylcytosine induce an ~5-fold increase in IPDs or a small decrease in IPDs (110,112). Damaged bases can dramatically distort DNA structure and not surprisingly, it is observed they can lead to bigger changes in IPDs. Oxidized bases such as 8oxoG and 8oxoA increase IPDs up to ~10-fold, whereas some alkylated bases show ~40-fold increase in IPDs. The most dramatic effect was observed with thymine dimers, with ~80-fold increase in IPDs (111).

Studies with unmodified DNA also showed pausing at specific sequences or regions with secondary structure. Schwartz *et al*. carried out single molecule fluorescence experiments using KF with a labeled DNA substrate (113). The authors monitored the elongation rates at a DNA hairpin and found that elongation speeds were significantly faster (one order of magnitude) than reported by bulk. Faster rates for single molecule experiments when compared to bulk were also observed in similar systems (114). KF pauses at Pyr-C-G motifs were observed and attributed to template rearrangement during dNTP incorporation cycle. Kim *et al*. also quantified elongation rates for HIV-1 RT using a DNA flow stretching assay (115). HIV-RT synthesizes DNA at two significantly different rates. The slow rate correlated well with DNA hairpin regions, suggesting that dsDNA unwinding becomes the rate limiting step for HIV-1 RT in the presence of secondary structures.

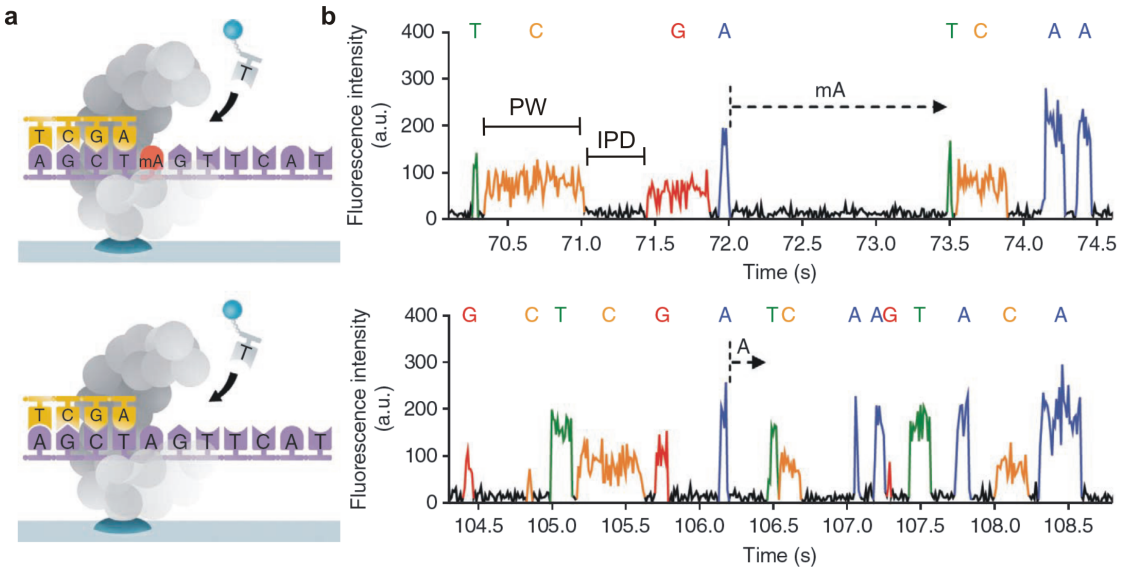

Figure 8. Modified bases induce a change in dNTP incorporation kinetics that can be monitored by SMRT. (**a**) Scheme representing polymerase activity with a DNA substrate containing either a methylated adenosine (top) or an unmodified adenosine (bottom) as the templating base. (**b**) Representative SMRT fluorescence traces showing multiple dNTP incorporations. Methylated adenine (top trace) induces a large change in inter-pulse duration (IPD) when compared to unmodified adenine (bottom trace). The variation in the pulse width (PW) is more modest. Figure adapted from (99).

Chapter 2: Experimental Procedures

Portions of the text in this chapter were reprinted or adapted with permission from:

Vrtis, K.B.*, Markiewicz, R.P.*, Rueda, D., and Romano, L.J. (2012). Single-molecule microscopy reveals new insights into nucleotide selection by DNA polymerase I. Nucleic acids research *40*, 7975-7984. *co-first authors

Vrtis, K.B., Markiewicz, R.P., Romano, L.J., Rueda, D. (2013) Carcinogenic DNA adducts induce distinct DNA polymerase binding orientations. Nucleic acids research 41, 7843-7853.

2.1. Klenow fragment purification and labeling

Plasmid pXS106, carrying the KF_{exo-} (D424A) gene, and *E. coli* strain CJ376 were generous gifts from Dr. C. Joyce (Yale). Enzyme was purified as described(116). Purity exceeded 98% as assessed by SDS-PAGE and ESI-MS. KF containing a single native cysteine (C907) was incubated in 50mM Tris-HCl buffer, pH 7.0, 120 μM tris(2-carboxyethyl)phosphine (TCEP) with 5- to 10-fold molar excess of Cy5 maleimide (GE healthcare) for 1 hr at room temperature. Reaction was stopped with 10 mM dithiothreitol

(DTT). Cy5 labeled enzyme was separated from the free dye on polyacrylamide Bio-Gel P6 spin columns. These stringent reaction conditions yield ~70% KF conjugation to Cy5, and minimize doubly labeled species. Enzyme activity was measured by primer extension on 20% polyacrylamide gel. Cy5 labeling did not affect enzyme activity(50).

2.2. Dpo4 purification and labeling

Plasmid p1914, carrying the *Sulfolobus solfataricus* DNA polymerase IV (Dpo4) gene, and *E. coli* strain RW382 were generous gifts from Dr. Roger Woodgate (NICHD). Dpo4 was purified with only slight modifications to the method previously described(117) and stored at -20 ºC in Dpo4 storage buffer (40 mM Tris HCl pH 7.5, 100 mM NaCl, 0.2 mM EDTA, 1 mM DTT, 50 % glycerol). Dpo4 concentration was determined by the absorbance at 280 nm (22,000 $M^{-1}cm^{-1}$ extinction coefficient). Dpo4 labeling was performed as described for KF.

2.3. Checking polymerase labeling efficiency by UV-Vis and ESI-MS

We checked the polymerase labeling efficiency by UV-Vis and ESI-MS. UV-Vis measurements were collected with a NanoDrop® ND-1000 spectrophotometer and the protein concentrations were determined in accordance with manufacture's instructions (GE Healthcare). Briefly, to obtain the 280 nm absorbance used to calculate the protein concentration, 5 % of the absorbance at 650 nm was subtracted from the absorbance at 280 nm, which corrected for the Cy5 absorbance at 280 nm. The concentrations of dye

and KF were calculated using the extinction coefficients (ε_{280} = 58,800 $M^{-1}cm^{-1}$ (KF), ε_{650} = 250,000 $M^{-1}cm^{-1}$ (Cy5)). The labeling ratio is [dye]/[protein].

The samples were prepared for ESI-MS in two steps. First, Bio-Spin 6 Tris Columns (Bio-Rad, cat. # 732-6227) were used to buffer exchange the samples into 10 mM ammonium acetate according to manufacturer's instructions. Second, Amicon ultra-0.5 centrifugal filter units with ultracel-3 membrane (Millipore) were used to buffer exchange the samples into 1 mM ammonium acetate according to manufacturer's instructions. ESI-MS measurements were acquired with a Waters LCT Premier™ time-of-flight spectrometer (Waters, Milford, MA). Approximately 20 µM protein in 1 mM ammonium acetate were diluted 1:2 in a solution of methanol, 0.2 % acetic acid. Samples (5-10 µL) were injected into the ESI-MS with a Waters Acquity AutoSampler. Capillary voltage was set at 3.4 kV, and desolvation temperature was set to 320 ºC. Mass spectra were obtained by scanning an *m/z* range between 350 to 2000 at 1 s per scan. Mass spectra were deconvoluted using MassLynx™ V4.1 software (Waters). All measured masses were within 0.01 % of the theoretical values.

2.4. Factor Xa protease digestion

KF and Factor Xa (Factor Xa Activated, from bovine plasma, Sigma, cat. # F9305 50UG) stocks were diluted in KF storage buffer to 6.3 µM and 0.36 µg/µL, respectively. The protease reactions were carried out by mixing 8 µL 6.3 µM KF, 2 µL 0.36 µg/µL Factor Xa, and 10 µL 2x reaction buffer (100 mM Tris, pH 8.0, 200 mM NaCl, and 2 mM $CaCl_2$) for 4 h at 37 ºC. The reactions were stopped with 20 µL 2x SDS loading buffer

(100 mM Tris-HCl, pH 7.5, 2 mM DTT, 4 % (w/v) sodium dodecyl sulfate, 0.04 % (w/v) bromophenyl blue, 20 % (v/v) glycerol), heated at 65 ºC for 3 min, and ran (15 µL/lane) on 10-20 % Bio-rad gradient gels. Gels were scanned for Cy5 on a Typhoon 9210 Variable Mode Imager (GE Healthcare).

2.5. DNA oligonucleotide purification and labeling

Primers with 5'-Cy3 or 5'-biotin, and amino-modified C6-dT templates were custom synthesized by Eurofins MWG Operon. DNA oligonucleotides were purified by HPLC chromatography on analytical C18 column. Purified oligonucleotides were desalted and their purity was assessed by MALDI-TOF MS. Purified DNA oligonucleotides were homogenous. All experiments were performed with the long template unless otherwise stated.

Amino-modified C6-dT 28-mer and 33-mer templates (Figure 2a) were labeled with Cy3 NHS ester (GE healthcare) using N,N-diisopropyl-N-ethylamine (DIPEA) as the basic agent. Briefly, a reaction containing 10 µl dioxane, 10 µl dimethyl formamide (DMF), 3 µl DIPEA, and 1 nmol of desalted, amine-modified DNA oligonucleotide was incubated with 10 nmol of the Cy3 NHS ester for two hours at room temperature. Upon 10-fold dilution with buffer containing 0.1 M triethylammonium acetate pH 7.5, 5 % acetonitrile, the reaction mix was directly injected onto a C18 analytical column and the Cy3-modified oligonucleotides were collected (yields ≥90%). Purified oligonucleotides were desalted, characterized by MALDI-TOF MS and stored at -20 ºC in 10 mM Tris, pH 7.5, 1 mM EDTA buffer.

Dideoxy-terminated primers were synthesized enzymatically by single nucleotide extension reaction catalyzed by deoxynucleotydyl transfarase (TDT). DNA oligonucleotide (1.2 nmol), TDT (45 units) and 100-fold excess of appropriate dideoxy-nucleotide-5'-triphosphate were incubated in the manufacturer's buffer (USB Affymetrix, Inc.) for 6 hours at 37 ºC.

2.6. Modifying template DNA with AAF and AF

AAF and AF-modified templates were prepared as previously described(118). Briefly, 20 nmol an oligonucleotide containing a single guanine was incubated with 500 nmol of 2-(*N*-acetoxy-*N*-acetyl)aminofluorene (AAAF) for 1 h at 37 °C in a degassed solution containing 20% ethanol and 2 mM sodium citrate, pH 6.8. The reaction was stopped by removal of the excess AAAF with water-saturated ether. The AAF-modified template was purified by HPLC using a C18 column. To convert the AAF-modified template to an AF-modified template, the AAF-DNA was incubated in 1 M NaOH, 0.25 M β-mercaptoethanol for 1 h at 37°C. The reaction was stopped by neutralization of the solution with HCl. The AF-modified DNA was purified by HPLC using a C18 column.

All templates were labeled at the amino-modified C6-dT with Cy3 NHS ester (GE Healthcare) as previously described (52). Dideoxy-terminated primers were enzymatically synthesized by terminal deoxynucleotidyl transferase (TdT) (USB Affymetrix, Inc.). DNA oligonucleotide (1 nmol), 45 units TdT, and 100 nmol of the appropriate dideoxy-nucleotide-5'-triphosphate were incubated in manufacturer's

reaction buffer (USB Affymetrix, Inc.) for 6 hours at 37 ºC. Dideoxy products were HPLC purified by reverse phase chromatography on a C18 column.

The purity and structure of all DNA oligonucleotides were confirmed by MALDI-TOF MS.

2.7. Polymerase activity assays

KF (100 pM) was incubated with 15 nM primer-template (16mer:33mer, 1:3 ratio) in reaction buffer (50 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 1 mM DTT and 100 μM dNTPs) for the indicated times at 37 °C. The primers were labeled with Cy3 at the 5' end for detection. Reactions were stopped by mixing reaction aliquots with an equal volume of loading buffer (10 mM EDTA, 1 mg/mL bromophenol blue, in 10 mL of formamide) and heating the samples at 80 °C for 75 s. Samples were run on 20% denaturing polyacrylamide gels for ~16 hours at 800 V. Gels were scanned on a Typhoon 9210 Variable Mode Imager (GE Healthcare). Slight deviations from this method may have been used for some experiments; see text for details.

2.8. Single molecule measurements

Quartz slides and cover slips were prepared as described(48,50,119). DNA was surface-immobilized by washing the slide with streptavidin (0.2 mg/ml) followed by incubation with the biotinylated primer-template duplex (20 pM) for 8 min. Data was acquired on a home-built, prism-based total internal reflection fluorescence microscope, as previously described(50). Measurements were performed in 50mM Tris-HCl, pH 7.5,

10 mM Mg, 1 mM DTT, 50 μg/ml BSA, and an oxygen scavenging system (4 % wt/vol glucose, 0.04 mg/ml glucose oxidase, 0.008 mg/ml catalase). Enzyme concentration and additives are as indicated. The apparent FRET efficiencies were calculated by dividing acceptor intensity (I_A) by the sum of donor and acceptor intensities ($I_D + I_A$). smPIFE and smFRET measurements were performed at ~50 ms time resolution. Experiments in the presence of the next correct dNTP were carried out at ~100 ms time resolution and decreased laser power due to the long binding events. Single-molecule time trajectories were analyzed with either five or seven point moving average. smPIFE and smFRET histograms were created from over 100 single-molecule time trajectories, except for the NaCl titration, which includes at least 50 trajectories per concentration point. Up to 5% of the molecules for any given experiment were excluded from the analysis due to aberrant photophysical or binding behavior. Slightly larger numbers of molecules were excluded for experiments containing the next correct nucleotide.

2.9. Applying a hidden Markov model to traces

Hidden Markov models (HMMs) probabilistically determine the FRET state at a given time from defined model parameters (transition probability matrix and emission probability functions), prior state information, and the current observation (current FRET value)(120). The most probabilistic hidden path is determined using the Viterbi algorithm and the given model parameters(120).

We empirically determined the defined model parameters used as initial guesses for the HMM. As initial guesses, we used transition probabilities of 0.02 for transitions to either of the higher FRET states from 0 FRET, 0.05 for transitions to 0 FRET from either

of the higher FRET states, and 0.10 for transitions between 0.5 and 0.63 FRET. The emission probabilities were calculated from normalized probability distributions centered at the three FRET states (0, 0.5, and 0.63) with 0.1 FRET standard deviations. The FRET states for each event in the TDP were determined from the mean FRET between transitions.

2.10. Simulating transitions between states

Simulations were used to determine if intermediate state binding was an obligatory step for polymerase association to the exo site or dissociation from the exo site for an AAF-modified primer-template (see chapter 5). All simulations were carried out with 10,000 one-binding event traces for transitions between the 0 and 0.63 FRET states. The traces consisted of the following: 15 time steps at either 0 or 0.63 FRET, followed by intermediate time steps of varying length (n), and finishing with $105 - n$ time steps at the post-transition FRET state (0 or 0.63) for a total trace length of 120 time steps. The FRET at the given time step was randomly chosen from a normalized probability distribution centered at the given FRET state with a FRET deviation of 0.1. Each trace was averaged with a five-point moving average as was used to analyze the experimental data. The traces were analyzed by the HMM described above and the length of the intermediate FRET state (0.50) pauses determined by the HMM were enumerated.

Simulations of direct transition between the two FRET states ($n = 0$) usually resulted in either 2 or 3 time step intermediate pauses due to averaging and exposure time integration. Therefore, the majority of the experimental intermediate pause length events

4 time steps or longer are the result of real intermediate state binding between the 0 and 0.63 FRET states. The rates for the intermediate state pauses were determined by fitting single exponential curves to the experimental data four time steps and longer. These rates were then used to simulate obligatory intermediate state pauses between the 0 to 0.63 FRET states.

Chapter 3: controlling the promiscuity of maleimide dyes for protein labeling

3.1. Introduction

Conjugation of reporter molecules to proteins is commonly used to probe the answers to various biological questions(121). Covalent attachment of dyes to proteins is typically accomplished through either lysine amino groups or cysteine sulfhydryl groups (122-124). Dyes functionalized with amino-reactive groups, such as *N*-hydroxysuccinimides (NHS-esters), tend to be promiscuous due to the abundance of reactive amines on a protein(124). While site-specific labeling is not requisite for some fluorescence applications (provided the original activity of the protein is retained), site specificity is an important factor when seeking precise spatial information(125). An array of methods have been developed to site-specifically label proteins including: unnatural amino acid mutagenesis(126), enzymatic labeling of specific peptide sequences(127), N- and C-terminal affinity tags(128), self-splicing inteins(129), and fluorescent protein fusions(125,130). These approaches can be technically difficult, impractical, and laborious, which is why the most common approach to site-specifically label proteins *in vitro* is by using a thiol-reactive dye, which will readily couple with surface accessible cysteine residues(125).

Cysteines tend to be less prevalent in proteins than other amino-group residues(131), and the process of removing and adding cysteines via site-directed mutagenesis has become a straightforward and routine technique(132). While labeling

cysteines via maleimide chemistry is a fairly rudimentary process, if care is not taken to control the reaction conditions nonspecific coupling to unintended amino acids can occur(124). This, in turn, can lead to a drastic misinterpretation of the fluorescence results.

Several papers have been published describing methods to conjugate thiol-reactive fluorophores to cysteines(50,52,122,124,131,133-136). The recommended protocols vary with regards to incubation time (10 min to overnight), temperature (4 ºC to room temperature), pH (6.8 – 7.7), and excess dye concentration (1- to 100-fold). Additionally, some procedures used precipitated proteins(122,131), affinity columns(124,133,135), phenyl arsenic oxide(135), or group 12 metal ions (Cd^{2+} and Zn^{2+})(134) to achieve site-specific and efficient labeling. A possible reason for these variations is the fact that proteins are a diverse group of biomolecules; hence some proteins may require very unique labeling conditions. Nonetheless we have established general reaction conditions to site-specifically introduce a single label that should be applicable to most single cysteine proteins. We also demonstrate how individual reaction parameters affect the reaction yield and how these parameters can be adjusted, as necessary.

While ultraviolet-visible spectrometry (UV-Vis) is currently the most common approach to check labeling efficiency, it lacks essential information about the subpopulations present in the solution. To overcome this issue, we determined the extent of labeling with electrospray ionization mass spectrometry (ESI-MS). Furthermore, we demonstrate with a well-characterized DNA polymerase single molecule

system(50,52,54) how slightly altering the labeling conditions can lead to dramatic misinterpretation of fluorescence results.

3.2. Results

3.2.1. *ESI-MS reveals promiscuity of maleimide dye conjugation to single cysteine proteins*

The most common approach to estimate protein-dye conjugation labeling efficiency is by measuring UV-Vis absorption spectra. Unlabeled protein will give a single, absorbance peak at 280 nm (**Figure 9a**). Cy5 dye-conjugated protein has an additional absorbance peak at 650 nm (**Figure 9b**).

Klenow fragment of *E. coli* DNA polymerase I has only one native cysteine residue (C907), which makes it an excellent candidate for labeling via a thiol-reactive probe such as Cy5 maleimide. Manufacture's protocol (GE Healthcare) for labeling 1 mg of IgG antibody with Cy5 maleimide involves incubating the antibody with ~50-fold molar excess of dye overnight, buffer pH between 7.0-7.5, at 4ºC. These conditions yield 0.5-3.5 dye molecules per IgG. When we incubate KF under more stringent conditions (~10-fold molar excess of dye, pH 7.5, 6 h, and 4ºC), which we will hitherto refer to as "heavy" labeling conditions, we get ~1.4 dye molecules per protein by UV-Vis (**Figure 9b**).

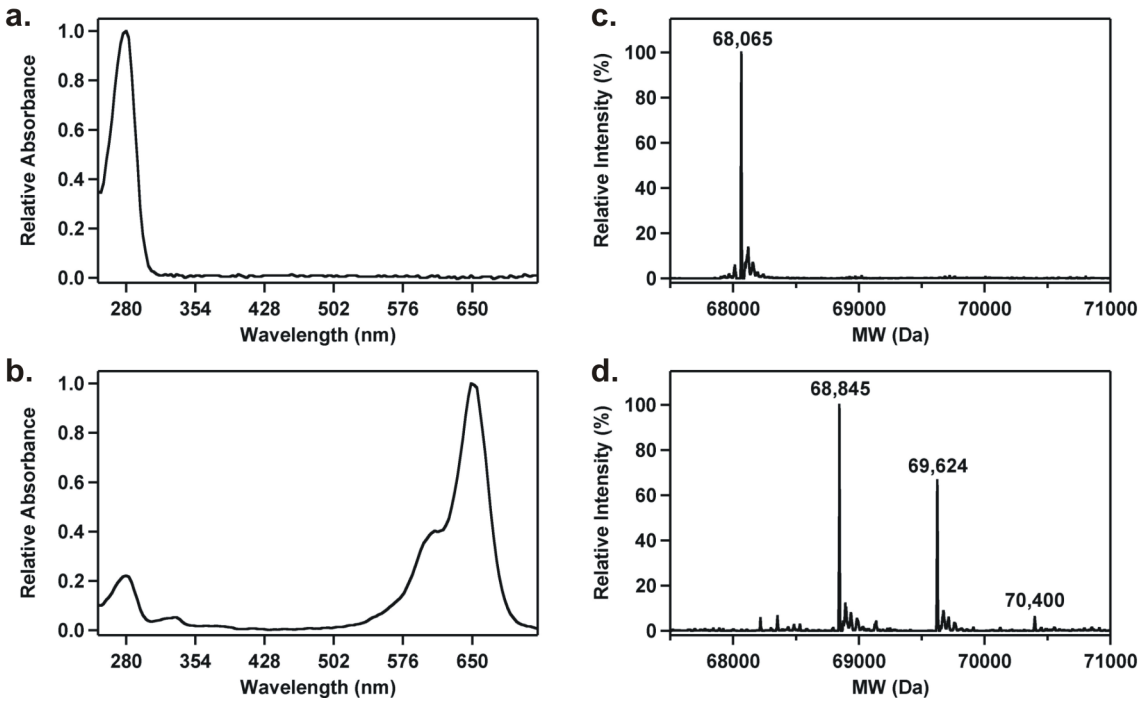

Figure 9. ESI-MS reveals promiscuity of maleimide dye conjugation to single cysteine proteins. UV-VIS and ESI mass spectra of unlabeled KF (**a** and **c**) and multiply labeled KF (**b** and **d**). The multiply labeled KF was prepared by incubating unlabeled, single cysteine KF with 10-fold molar excess dye, Tris-HCl, pH 7.5, at 4 ºC for 6 h (heavy labeling conditions). The theoretical masses of unlabeled, singly labeled, doubly labeled, and triply labeled KF are 68,062 Da, 68,841 Da 69,620 Da, and 70,399 Da, respectively. The deconvoluted ESI-MS spectra were obtained from the m/z spectra in Figure 2.

There are two intrinsic problems with determining labeling efficiency by UV-Vis. First, you need to be able to accurately determine the extinction coefficients for the protein and the dye. The change in the local environment about the dye may change the spectral properties of the dye, as was shown to occur with Cy3b(131). Therefore, the extinction coefficient used for determining the free dye concentration may not be appropriate to determine the dye concentration when it is conjugated to a protein. Second, UV-Vis spectra lack information about subpopulations present in the solution. For

example, when the dye:protein ratio is 1.4, the additional 40 % of dye can come from free, unreacted dye or it can be the result of labeling the protein with multiple dyes.

ESI-MS is a more accurate way to measure labeling efficiency because it does not depend on the spectral properties of the reagents, and it provides information about the subpopulations of unlabeled and labeled proteins in the solution. Unlabeled KF is 68,065 Da (theoretical value 68,062 Da) (**Figure 9c** and **Figure 10a**). The heavy labeled KF, as determined by UV-Vis, has three predominant subpopulations corresponding to singly labeled KF (68,845 Da, theoretical 68,841 Da), doubly labeled KF (69,624 Da, theoretical 69,620 Da), and triply labeled KF (70,400 Da, theoretical 70,399 Da) (**Figure 9d** and **Figure 10b**). The proportions of singly, doubly, and triply labeled KF were 52 %, 43 %, and 4 %, respectively. The presence of multiply labeled species of KF demonstrates that nonspecific conjugation of dye to a single cysteine protein occurs under conditions less stringent than manufacturer's instructions for labeling proteins with maleimide dyes. Furthermore, the higher concentration of dye than protein measured by the UV-Vis is predominantly the consequence of multiply labeled proteins, not excess free dye in solution.

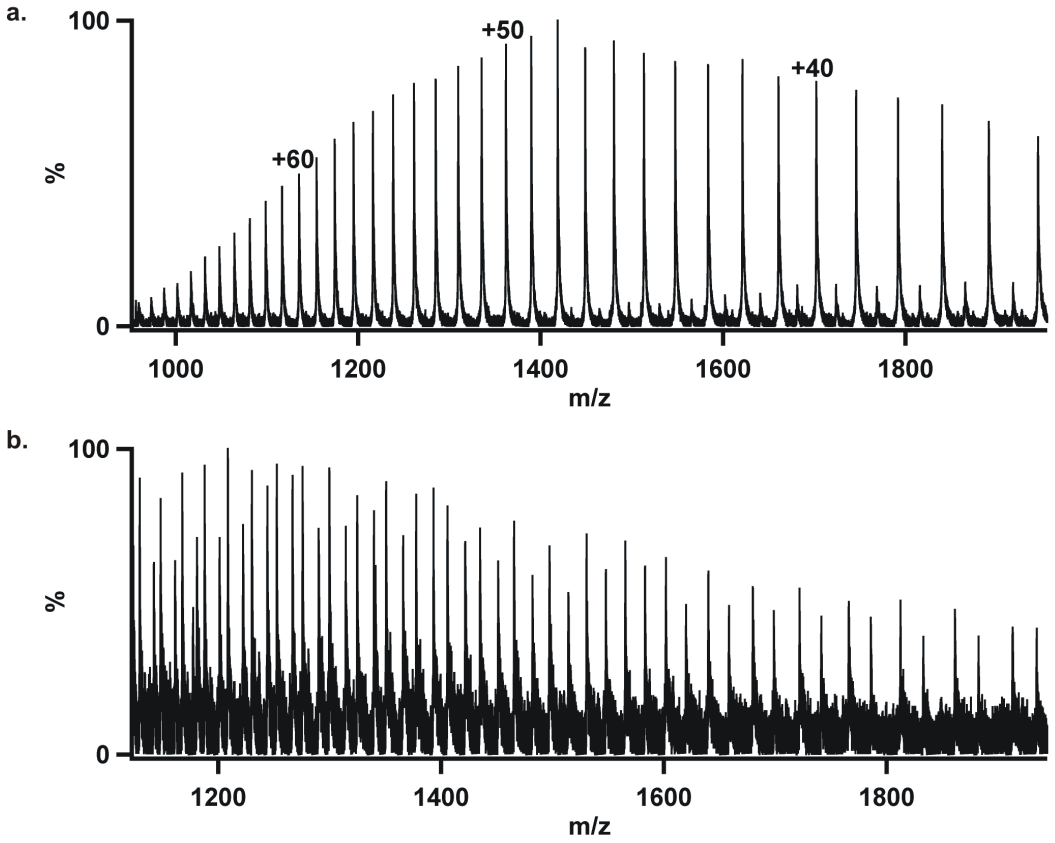

Figure 10. ESI-MS m/z spectra of (**a**) unlabeled labeled and (**b**) multiply labeled (heavy labeling conditions) KF. **Figure 9b** and **Figure 9d** are the deconvoluted spectra for (**a**) and (**b**), respectively.

3.2.2. Controlling the labeling reaction

It is often preferred, if not necessary, to limit the conjugation of dyes to proteins to one dye per protein for fluorescence experiments. Incubation time, dye concentration, and pH are three reaction variables that are easily manipulated to limit the extent of the reaction. By decreasing the incubation time to 30 min, the Cy5 concentration to 4-fold molar excess over KF concentration, or the pH to 7.0 we were able to limit the production of doubly-labeled species to less than 15 % while completely preventing the formation of

species with more than two conjugates (**Figure 11a-c**). Simultaneously decreasing the incubation time to 30 min and the pH to 7.0 ("light" labeling conditions), gave 22 % unlabeled, 74 % singly labeled, and, most importantly, less than 4 % doubly-labeled species. Under identical labeling conditions, *Sulfolobus solfataricus* DNA polymerase IV (Dpo4), which has only one cysteine, was nearly 100 % singly-labeled (**Figure 12**). ESI-MS has clearly confirmed that decreasing incubation time, dye concentration, or pH can limit the extent of the reaction to less than one dye per protein molecule, which is ideal for fluorescence applications that require accurate distance and orientation information.

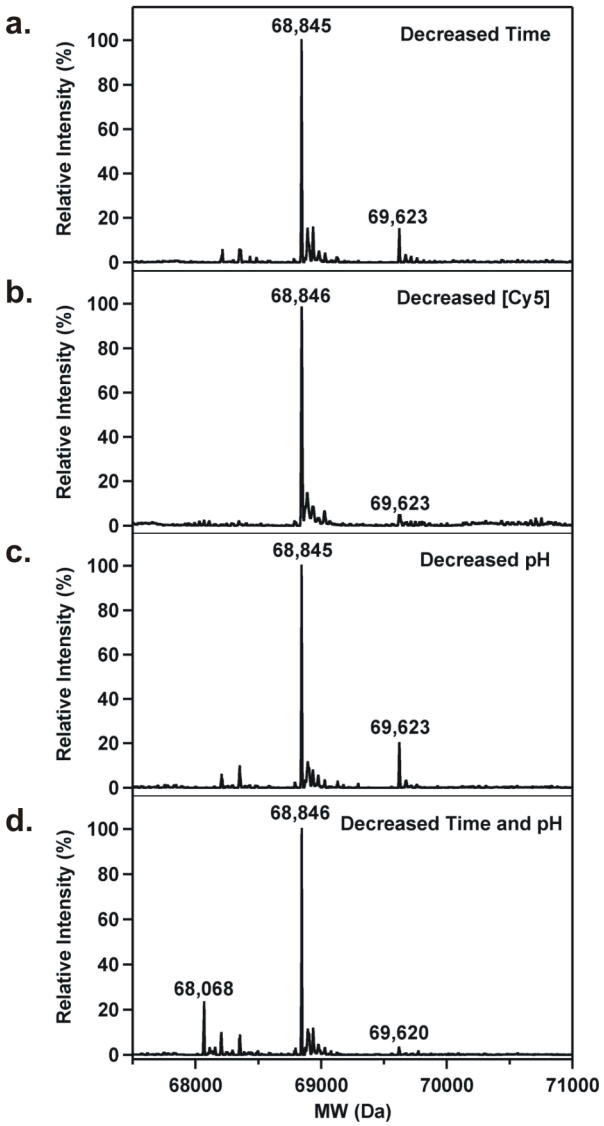

Figure 11. Labeling reaction variables adjusted to decrease the extent of conjugation. Degree of labeling limited by decreasing the (**a**) incubation time to 30 min, (**b**) Cy5 maleimide dye concentration to 4-fold excess, (**c**) or pH to 7.0. (**d**) Multiply labeled species were further limited by simultaneously decreasing the time to 30 min and the pH to 7.0 (light labeling conditions). Unless specifically noted, the other reaction variables were identical to the heavy labeling conditions used to obtain multiply labeled KF (from **Figure 9d**). The theoretical masses of unlabeled, singly labeled, doubly labeled, and triply labeled KF are 68,062 Da, 68,841 Da 69,620 Da, and 70,399 Da, respectively.

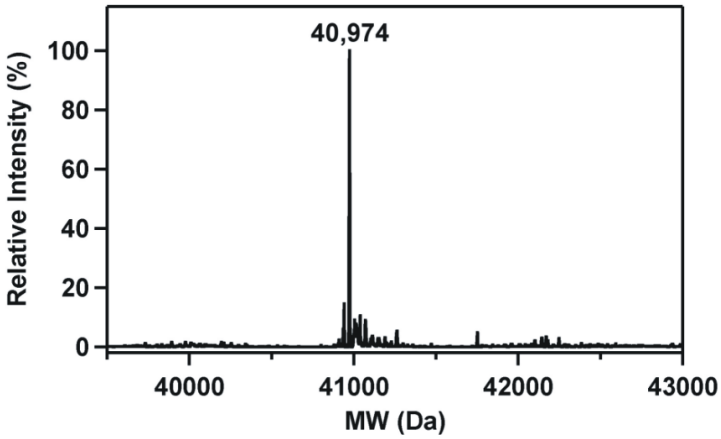

Figure 12. Deconvoluted ESI-MS spectra of singly labeled Dpo4 (theoretical mass 40,973 Da). Dpo4 was nearly 100 % singly labeled by the light labeling conditions used to label KF in Figure 2d (10-fold molar excess dye, Tris-HCl, pH 7.0, at 4 ºC for 30 min).

3.2.3. Additional label(s) added to N-terminal fragment

Maleimides readily react with the thiol groups on cysteine side chains; however, at pH higher than 8, other basic amino acids such as lysine are susceptible to maleimide conjugation. Additionally, the local environment around a basic amino acid may increase the side chain reactivity at neutral pH. We digested labeled proteins with Factor Xa protease to determine the regions where the labels were added to KF. Factor Xa cleaves KF into two fragments: 1) a 57 kDa N-terminal fragment, and 2) a 12 kDa C-terminal fragment, which includes the single cysteine amino acid (**Figure 13a**). The 12 kDa C-terminal fragment was the only fragment labeled under the light labeling conditions (**Figures 14d** and **13b, Cys**).

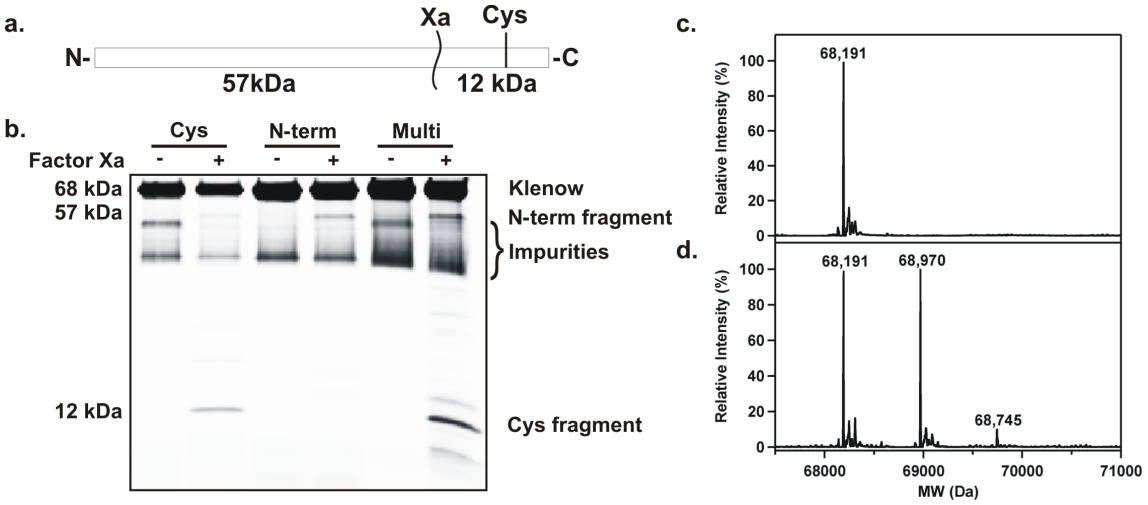

Figure 13. Nonspecifically conjugated dyes added to the N-terminal fragment. (**a**) Cleavage of KF with Factor Xa should produce two fragments: 1) a 57 kDa N-terminal fragment, and 2) a 12 kDa C-terminal fragment. (**b**) Factor Xa digest of unlabeled, singly labeled at Cys (Cys), N-terminal fragment labeled (N-term), and multiply-labeled (Multi) KF. Singly labeled KF (Cys) was prepared with the light labeling conditions (Figure 2d). The EM blocked, N-terminally labeled KF did not have a 12 kDa band, which confirms the native cysteine residue was blocked by the EM and it suggests the nonspecifically conjugated dyes were not added to the C-terminal fragment. (**c**) N-terminally labeled KF (theoretical mass 68,187 Da) was prepared by first blocking the single cysteine with EM using the light labeling conditions used to obtain singly labeled KF (10-fold molar excess dye, Tris-HCl, pH 7.0, at 4 ºC for 30 min). (**d**) Next, the EM, blocked KF was incubated with Cy5 maleimide dye under the heavy label conditions used to obtain multiply labeled protein (10-fold molar excess dye, Tris-HCl, pH 7.5, at 4 ºC for 6 h). The theoretical masses of singly and doubly labeled EM-blocked KF are 68,966 Da and 69,745 Da, respectively.

The single KF cysteine was blocked with *N*-ethylmaleimide (EM) (**Figure 13c**) prior to labeling with Cy5 to determine where the additional, nonspecific labels were added to KF. No Cy5 labels were added to the C-terminal fragment when the single cysteine was first blocked by EM (**Figure 13b, N-term** and **13d**). Instead, the Cy5 labels were only present on the N-terminal fragment. This suggests the nonspecifically labeled

amino acids were predominantly on the N-terminal fragment. As expected, Factor Xa digestion (**Figure 13b, Multi**) revealed that both the N- and C-terminal fragments were labeled for the heavy labeled KF (**Figure 9b,d**).

3.2.4. Improper labeling produces misleading smFRET results

Single molecule fluorescence has reached new levels of acceptance in recent years because it can reveal heterogeneous subpopulations, transient intermediate states, and complex kinetic information that would otherwise be obscured by ensemble measurements(49,54,137). smFRET uses the photophysical behavior of a FRET pair to measure changes in orientation and distance on the 25-100 Å molecular scale(49). This technique has been used to monitor KF dynamics on DNA with single nucleotide resolution(50,52,54).

KF catalyzes 5'→3' template directed DNA synthesis from its pol site, and it excises misincorporated nucleotides from its 3'→5' exonuclease site located ~35 Å from the pol site (**Figure 14a**)(23). It remains highly debated as to whether or not KF stably binds to its exonuclease site with complementary DNA(24,52,138,139). By placing a Cy3, donor, fluorophore on the DNA and a Cy5, acceptor, fluorophore(s) on the polymerase, we were able to observe polymerase binding to the DNA in real-time (**Figure 14a**). Upon KF binding to the DNA, energy is transferred between the Cy3 on the DNA and the Cy5 on the polymerase in an anticorrelated matter, such that Cy3 intensity decreases along with a simultaneous increase in Cy5 intensity (**Figure 14b, top**). The calculated FRET shows discreet spikes when the polymerase binds to the DNA

(**Figure 14b, bottom**). We determined the FRET states that were prevalent during the binding assay by plotting FRET histograms from over 100 traces.

When the heavy labeled KF batch (from **Figure 9d**) bound to the DNA, multiple FRET states were observed (**Figure 14b,c**). The zero FRET state corresponds to unbound DNA. Considering KF has two active sites, the higher FRET states (0.32 and 0.76) would likely be misinterpreted as pol site binding and exo site binding, respectively. However, when the singly labeled polymerase batch bound to the DNA (from **Figure 11d**) only one FRET state, 0.32, was observed. This suggests that only pol site binding occurs when the polymerase binds to complementary DNA. Moreover, the second FRET state (0.76) detected with the heavy labeled KF was due to the presence of multiple acceptor fluorophores on individual polymerases. When multiple acceptor fluorophores are localized around one donor fluorophore, the apparent FRET will be higher(140,141). This is because multiple acceptors in close proximity to a single donor increase the probability of successful resonance energy transfer(140). Consequently, mislabeling the protein with multiple acceptor fluorophores completely changes the interpretation of the smFRET results.

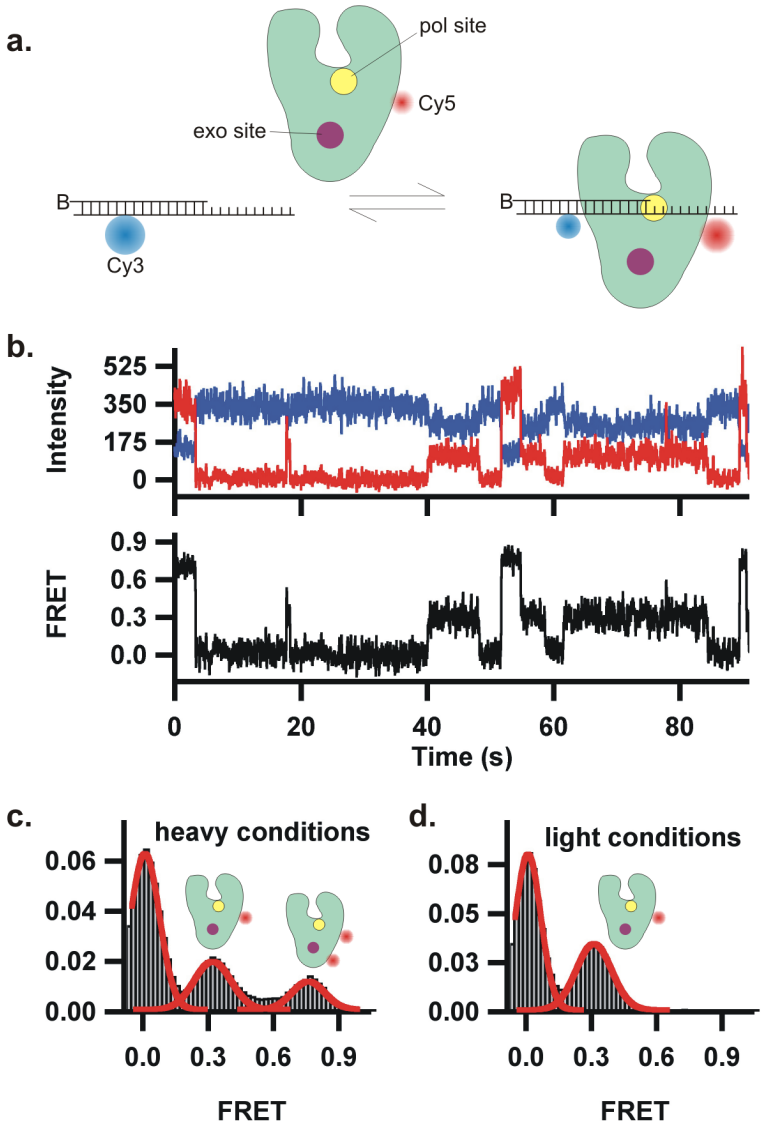

Figure 14. smFRET reveals that multiply labeled KF gives aberrant, high FRET state. (**a**) Schematic of smFRET design. Cy3 labeled DNA is surface immobilized to a quartz slide via a biotin (B)-streptavidin bridge. As Cy5-labeled KF binds to the Cy3-labeled primer-template, energy is transferred from the Cy3 (blue) to the Cy5 (red). (**b, top**) Representative Cy3 (blue) and Cy5 (red) intensity trajectories for heavy labeled KF binding to DNA. (**b, bottom**) FRET trajectory for the molecule shown in **b, top**. FRET histograms from (**c**) multiply labeled (heavy conditions) and (**d**) singly labeled (light conditions) KF batches. Singly labeled KF gives a FRET distribution at ~0.32, while the multiply labeled KF gives an additional FRET distribution at ~0.76.

3.3. Discussion

Oftentimes site-directed mutagenesis is used to ensure there is only a single cysteine present on a protein. In principle, the cysteine side chain can then be targeted to label the protein with one, thiol-reactive fluorescent probe. While we only tested this approach with single cysteine proteins, other groups have shown that it is possible to site-specifically label multi-cysteine proteins if the other cysteines are not solvent exposed(124) or if the solvent exposed cysteines have substantially different reactivities(136). Alternatively, one can first reversibly protect the cysteine of interest, block other reactive cysteines with a nonfluorescent compound (such as EM), deprotect the cysteine of interest, and then label it(134,135). Our labeling approach or a slight alteration thereof, should be appropriate for the final conjugation of the dye for these multi-cysteine approaches as well.

If care is not taken to control the reaction conditions, other amino acids may be labeled in addition to the cysteine. As we have shown, slightly altering the labeling conditions can lead to improper labeling of the protein and to false evaluation of fluorescence data. Consequently, we strongly recommend checking the degree of labeling by mass spectrometry to detect populations with multiple labels. If excessive labeling occurs, decrease the pH, incubation time, and/or dye concentration to decrease the labeling yield.

Chapter 4: New insights into nucleotide selection by dna polymerase i

Portions of the text in this chapter were reprinted or adapted with permission from:

Vrtis, K.B.*, Markiewicz, R.P.*, Rueda, D., and Romano, L.J. (2012). Single-molecule microscopy reveals new insights into nucleotide selection by DNA polymerase I. Nucleic acids research *40*, 7975-7984. *co-first authors

4.1. Introduction

The incorporation of a nucleotide by a DNA polymerase is an extraordinarily accurate process that is repeatedly required for the faithful duplication of the genome prior to cell division. Although this single catalytic step has been thoroughly studied for over five decades, it is still not completely understood and continues to be the subject of a great deal of interest. *E. coli* DNA polymerase I, first discovered in 1956(5), has served as a model for the elucidation of this catalytic mechanism, and it remains one of the most intensively studied polymerases. Proteolysis of DNA polymerase I produces the Klenow fragment (KF) which contains the polymerase and the 3'-5' exonuclease (proofreading) domains(6,7). Numerous DNA polymerase crystal structures have shown that many of

the structural features first observed for KF are present in high fidelity polymerases isolated from a variety of sources(10). It is therefore not surprising that these polymerases also share many mechanistic features.

The first step in the catalytic cycle is the association of the polymerase with the primer-template complex (**Figure 15a**). This is the only step not repeated during processive DNA synthesis. Once bound, the polymerase must select the dNTP complementary to the templating base from a pool of four dNTPs and four rNTPs and this step provides most of the accuracy observed in this synthetic process. For high-fidelity polymerases, it is well established that, once the correct dNTP base pairs with the templating base in the polymerase active site, catalysis involves a nucleotide-induced conformational change of the so called fingers region of the polymerase(142). Recent single-molecule studies have shown that the unliganded polymerase molecule undergoes rapid conformational dynamics between the open and closed form, while for the ternary KF-DNA-dNTP complex the closed conformation dominates (51). This conformational change results in the formation of a tight binding pocket around the nascent base pair that aligns the 3'-OH of the primer with the a-phosphate of the dNTP allowing phosphodiester bond formation to proceed(11-15,73). On the rare occasion in which KF incorporates a mismatched nucleotide, it has been proposed that the terminal base pairs melt, transferring the primer strand to the exonuclease site located ~35 Å away from the polymerization site, and allowing the proofreading activity to excise the incorrect nucleotide (23). Several studies using a correctly paired primer template have observed

significant levels of binding of the primer to the exonuclease site (24,138,143,144), while a recent report suggests that this occurs less than 3% of the time (139).

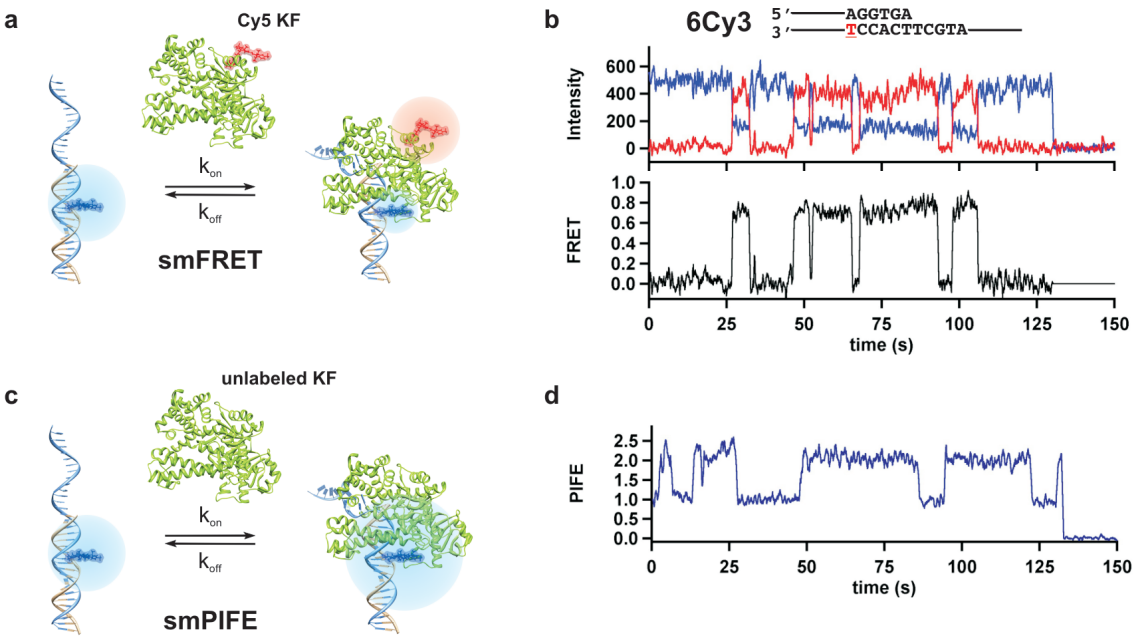

Figure 15. Real time single-molecule measurements of KF binding to DNA. (**a**) Schematic of single-molecule FRET design. Upon KF binding to the primer-template, energy is transferred from the Cy3 (blue) on the DNA template to the Cy5 (red) conjugated to KF (**b, top**). The structure of the primer-template (the 6Cy3 label refers to the fact that there are six nucleotides between the Cy3 and the primer terminus) is shown above. Representative donor (blue) and acceptor (red) time trajectories for KF binding to 6Cy3 (**b, bottom**) FRET trace for the trajectory shown in **b, top**. Photobleaching of Cy3 occurred at ~130 s. (**c**) Schematic of single-molecule PIFE design. Upon unlabeled KF binding to the primer template shown in **b, top**, the fluorescence intensity of the Cy3 is enhanced about two fold. (**d**) Representative PIFE trace for KF binding to 6Cy3. Photobleaching of Cy3 occurred at ~130 s.

In this study, we have monitored the formation of the KF-DNA binary complex in real time and measured the response to internal and external factors using single-molecule Förster resonance energy transfer (smFRET) and single-molecule protein-induced fluorescence enhancement (smPIFE)(56). By measuring the dynamics for polymerase-DNA binding under a variety of solution conditions, we quantify the stabilizing effect of the next correct nucleotide, as well as, the destabilizing effect of incorrect dNTPs and rNTPs. We find that the polymerase can reject incorrect nucleotides in the presence of the correct dNTP without significantly destabilizing the complex. These results provide further evidence for a mechanism in which nucleotide selection occurs predominantly in the open complex prior to fingers closing and the formation of the tight binding pocket. Finally, we were also able to distinguish between binding of the primer terminus to the polymerase or exonuclease domain and, challenging most ensemble-averaged studies, found that a complementary primer remains bound exclusively to the template at the polymerase active site.

4.2. Results

4.2.1. Tracking KF on DNA template with base pair resolution

Similar to our prior studies(50), we have used smFRET to observe the dynamics of a Cy5-labeled DNA polymerase (Cy5-KF) on a Cy3-labeled DNA template (**Figure 15a**). smFRET enables monitoring transient events from individual molecules that are otherwise hidden in ensemble-averaged experiments(50,137,145), therefore revealing KF-DNA binding and dissociation in real time. The nomenclature “nCy3” indicates the

number of base pairs (n) between the Cy3 position on DNA and the primer-template junction (**Figure 16a**). Control experiments confirm stoichiometric labeling and proper function of the labeled samples.

In the absence of KF, Cy3 emission appears constant, and the resulting FRET ratio is zero (**Figure 17a**). In the presence of Cy5-KF, the donor intensity exhibits random decreases accompanied by anti-correlated increases in the acceptor intensity, indicating Cy5-KF binding to 6Cy3 DNA (**Figure 15b**). The apparent FRET ratio for the binary complex is ~0.77 (**Figure 15b**). During the experimental time window of ~150 s, several binding events were observed and, within experimental noise, each binding event reached the same FRET ratio.

To track the position of Cy5-KF on the DNA, we used multiple nCy3 constructs with n ranging from 5-11 (**Figure 16a**). For each nCy3 binary complex, the resulting histograms reveal the corresponding FRET ratios ranging from 0.81 for 5Cy3 to 0.33 for 11Cy3 (**Figure 16b**). No transitions were observed between these different states, indicating that KF binds the primer-template junction in a well-defined orientation. The different FRET ratios obtained for binding to 7Cy3 to 10Cy3 were large enough to allow us to measure binding position with single nucleotide resolution (**Figure 16d**).

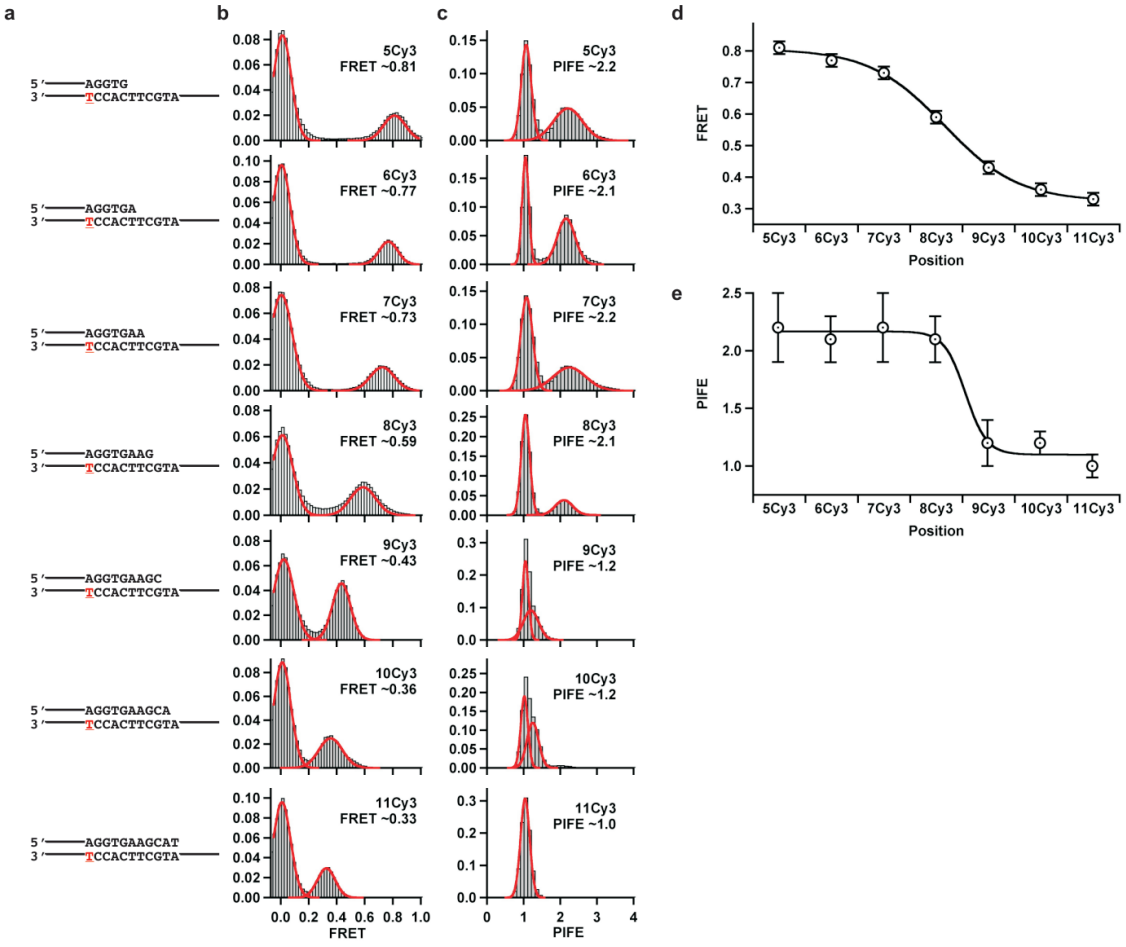

Figure 16. Tracking the position and footprint of KF on the primer-template with smFRET and smPIFE. (**a**) The structure of DNA duplex is shown for each primer-template. The Cy3 is conjugated to the thymine shown in red by an amine linker. (**b**) smFRET efficiency histograms for KF binding to each duplex. (**c**) smPIFE histograms for KF binding to each duplex. (**d**) smFRET efficiencies from (**b**) were plotted as a function of the distance between the Cy3 and the primer-template terminus. (**e**) PIFE values from (**c**) were plotted as a function of the distance between the Cy3 and the primer-template terminus. The errors for (**d**) are estimated to be ± 0.02, and the errors for (**e**) are one-half the width at half maximum amplitude of the peaks. The errors are attributed to the noise of the experiments as opposed to subtle movement of the polymerase on the DNA.

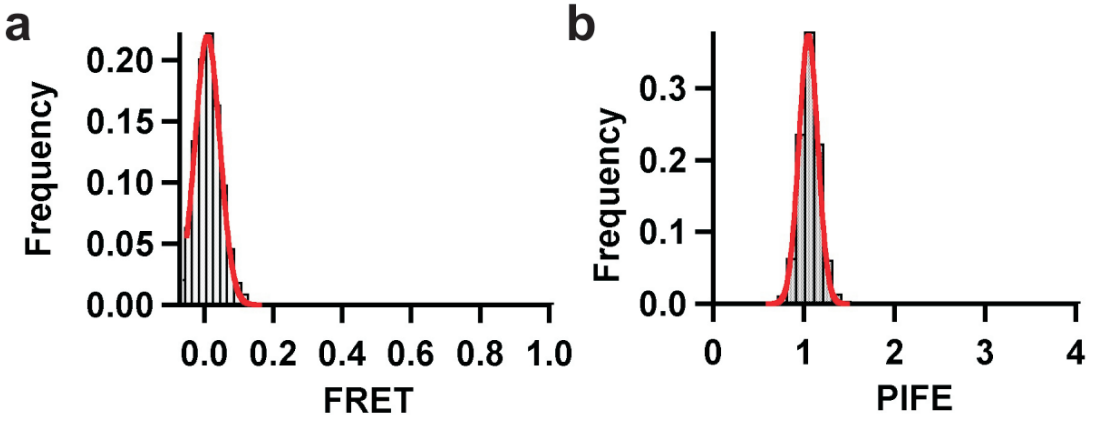

Figure 17. FRET and PIFE distributions in the absence of KF. The traces remained static at 0 and 1 in the absence of KF for FRET (**a**) and PIFE (**b**), respectively. Data were fit to Gaussian distributions.

Next, we tested if the observed FRET efficiencies for different binary complexes are independent of the DNA sequence and the length of the DNA primer-template duplex. In these experiments, we used an AT rich template with a shorter duplex region (**Figure 18a**, short template, and **b**) and found that the FRET ratios for 7-11Cy3 DNAs were essentially identical to those observed for the long template (**Figure 18c,d**). The implications of these results are that the polymerase binds to a primer-template junction with an orientation and position that is independent of the sequence context or primer-template length.

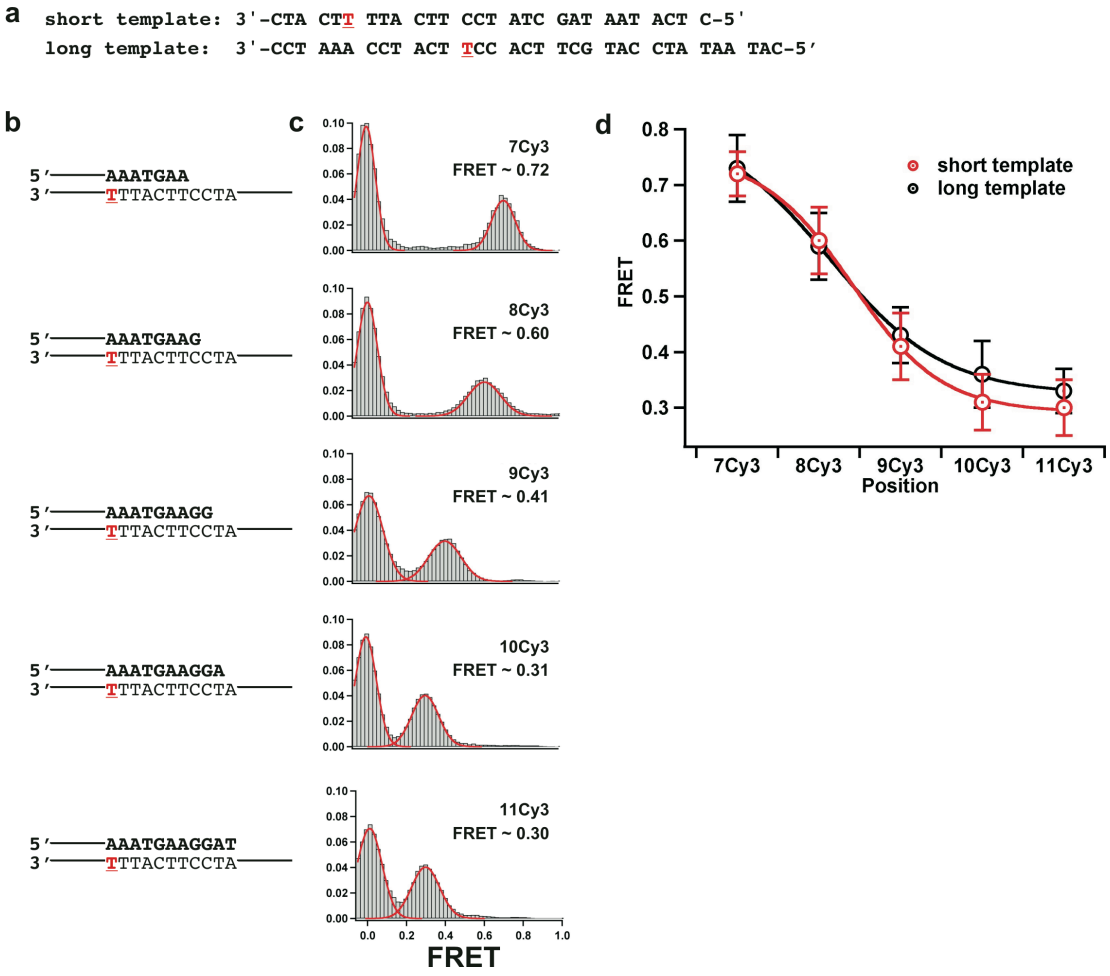

Figure 18. KF binding position and orientation is primer-template length and sequence independent. (**a**) DNA sequences for short and long primer-templates. Cy3 is conjugated to red, underlined thymine (**b**) Short template DNA duplex is shown for each primer-template. These duplexes have a different sequence and 7 fewer base pairs in the double stranded region of the DNA and 5 fewer bases at the 5' end of the template compared with the long template sequence used for the data shown in **Fig. 16**. (**c**) FRET efficiency histograms for KF binding to each primer-template. (**d**) FRET efficiencies from (**c**) (short template, red) and **Fig. 15b** (long template, black) were plotted as a function of number of nucleotides between the Cy3 and the primer-template terminus. The errors for (**d**) are one-half the width at half maximum amplitude of the peaks.

4.2.2. KF-DNA binary complex dynamics by smPIFE

In the presence of unlabeled KF, binding and dissociation can also be directly observed by smPIFE of the Cy3 label on the DNA primer template(55,56). Binding of KF to 6Cy3 DNA results in sudden fluorescence intensity increases (almost double) caused by the change in environment around the fluorophore when KF binds the DNA (**Figure 15c,d**). These fluctuations are never observed in the absence of KF (**Figure 17b**). Therefore, smPIFE can be used to visualize KF binding and dissociation by normalizing the Cy3 fluorescence intensity to 1.0 in the protein-free state (**Figure 15d**). It is noteworthy that in the smFRET experiments, we can readily distinguish between FRET and PIFE (**Figure 19**), and therefore, PIFE does not affect our ability to use smFRET to study KF binding.

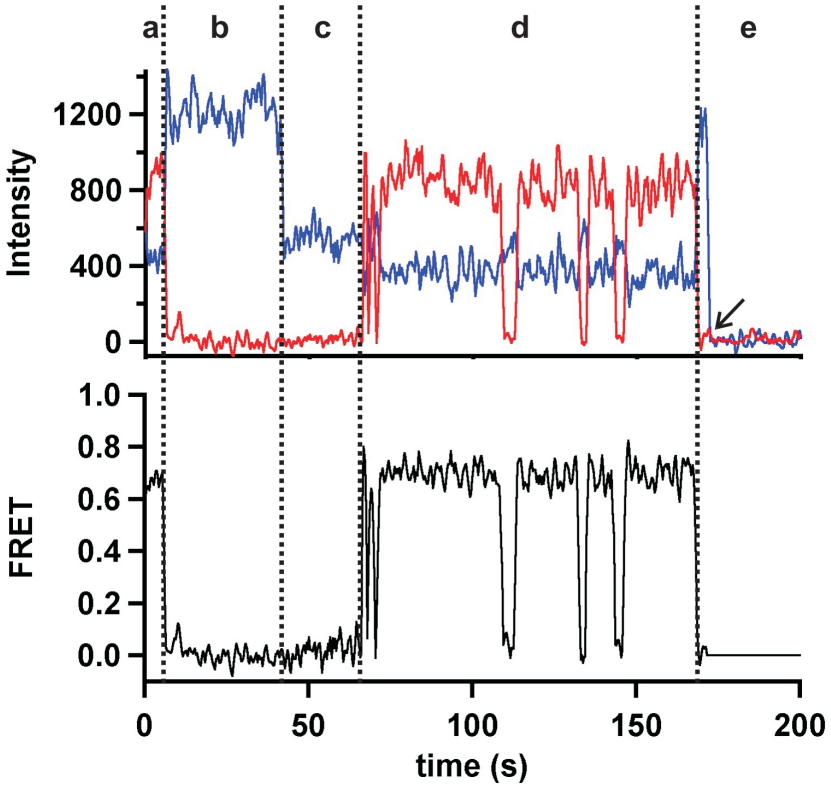

Figure 19. PIFE observed in FRET traces. Two donor intensities at ~550 and ~1,200 were observed in the absence of acceptor fluorescence. The lower donor intensity (550) is unbound DNA; the higher value (1,200) is attributed to Cy3 fluorescence enhancement due to PIFE when the polymerase is bound to the DNA in close proximity to the Cy3. We attribute these PIFE events to either the binding of an unlabeled polymerase or by photobleaching or blinking of a DNA bound Cy5-labeled polymerase. (**a**) At 0 s the polymerase is bound to the DNA producing a donor intensity of 450 and an acceptor intensity of 950. FRET occurs until ~5 s, at which time the Cy5 either blinks or photobleaches leading to a donor intensity of ~1,200, which is caused by KF binding-induced PIFE. (**b**) The polymerase remains bound to the DNA until ~44 s at which time the donor intensity drops to 550, the level observed for unbound DNA, because KF has dissociated from the DNA. (**c**) The DNA remains unbound until ~65 s. (**d**) From ~65-170 s multiple association-dissociation events occur with a stable acceptor and no visible donor intensity attributable to PIFE. (**e**) At ~170 s the Cy5 either blinks or photobleaches, but the polymerase remains bound to the DNA leading to a PIFE-induced donor intensity of ~1,200. About two seconds thereafter the Cy3 photobleaches (arrow) and no further donor or acceptor fluorescence intensity is observed.

We determined the fluorescence enhancement for binary complex formation with primer-template junctions from 5-11Cy3 (**Figure 16c,e**). The observed PIFE remains constant at ~2.1 between 5-8Cy3, and drops abruptly to ~1.2 for 9 and 10Cy3. Beyond 10Cy3 no fluorescence enhancement was observed. These results indicate that the KF footprint on the DNA covers ~8 base pairs from the primer-template junction, in agreement with previous results(146,147).

Although smPIFE trajectories contain no distance information, they present an advantage over smFRET because Cy5 blinking and photobleaching do not interfere with the PIFE data analysis. Therefore, we used smPIFE to characterize the association and dissociation kinetics of the binary complex. **Figure 20a** shows representative smPIFE time trajectories with corresponding distributions at [KF] from 0.1 to 5 nM. Dwell time analysis of >100 trajectories at each concentration produces the corresponding pseudo-first order binding (k'_{on}) and dissociation (k_{off}) rate constants (**Figure 20b**). As expected for a binary reaction the off rate (k_{off} = 0.37 s^{-1}) is concentration independent, while k'_{on} increases linearly with [KF] (**Figure 20c**). A linear fit to the latter yields the diffusion limited second order dissociation rate constant (k_{on} = 10^8 $M^{-1}s^{-1}$) and the calculated dissociation constant (K_D = 3.7 nM) (**Figure 20c**), comparable to prior results(50,148-150).

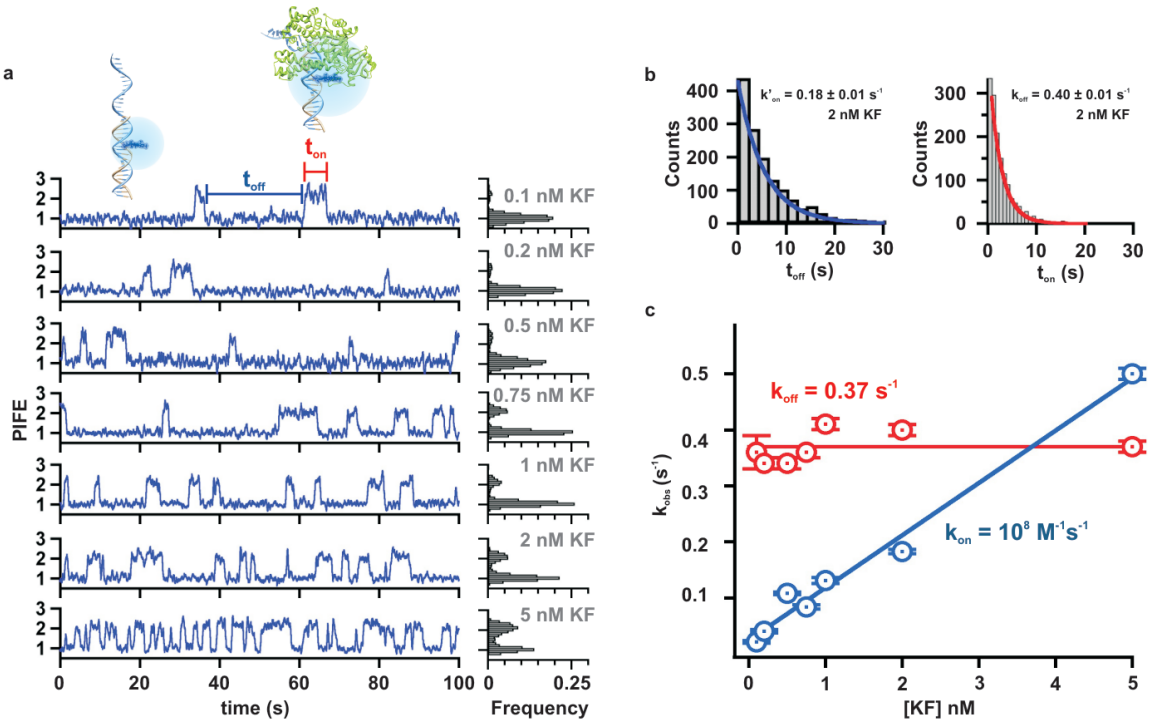

Figure 20. Concentration dependence of KF binding to 8Cy3 observed by smPIFE. (**a**) Representative smPIFE traces and histograms of KF binding to 8Cy3 at increasing concentrations of KF. Unbound DNA has a PIFE value of approximately 1 until KF binds to the DNA, at which time the PIFE increases to ~2 as shown by the two cartoons. Dwell time analysis was performed to determine the rate of association (k'_{on}) and rate of dissociation (k_{off}) at each KF concentration. (**b**) Representative on and off dwell time distributions and single exponential fit for 2 nM KF binding to 8Cy3. The on dwell times were determined from the times between binding events (t_{on}), and the off dwell times were measured as the time when KF was bound to the DNA (t_{off}). (**c**) The off (red) and pseudo-first order on (blue) rates as a function of KF concentration. The off rates were independent of KF concentration, while the pseudo-first order on rates had a linear dependence with KF concentration.

4.2.3. Electrostatic competition between KF and NaCl for the DNA

The KF-DNA binary complex is stabilized by electrostatic interactions between the negatively charged phosphate backbone of the DNA and the positively charged basic residues in the polymerase binding cleft(148,151). It is therefore not surprising that ionic strength influences the binary complex stability(152). We used smPIFE to quantify the effect of monovalent ions on KF binding to 8Cy3 DNA in low (0 mM NaCl) and high (100 mM NaCl) salt concetrations. Representative trajectories (**Figure 21a**) clearly show fewer and shorter-lived binding events in high salt. Dwell time analysis in 0.01-150 mM NaCl (**Figure 21b**), reveals that salt concentrations above 10 mM accelerate KF dissociation (k_{off}) and slow down association (k'_{on}), indicating that salt electrostatically shields the DNA and KF from interacting with each other. One possible interpretation is that high [Na^+] result in both fewer reactive collisions and a less stable binary complex, and thus, slower binding and faster dissociation. We observe a linear relationship between the natural log of the inverse K_D and the natural log of the NaCl concentration (**Figure 22**). The slope of this plot suggests that, under our conditions, one Na^+ ion is released upon binding of the protein to the DNA (153). A previous study suggested that 2.8 K^+ ions may be released upon binding (154). The difference between two results may be due to the lower overall ionic strength used in the prior study. Finally, at salt concentrations higher than 100 mM, some of the binding events likely become too short to be detected with our time resolution (~50 ms). Nonetheless, our results are consistent with a model where electrostatic interactions between the DNA and KF play an important role in stabilizing the binary complex(148).

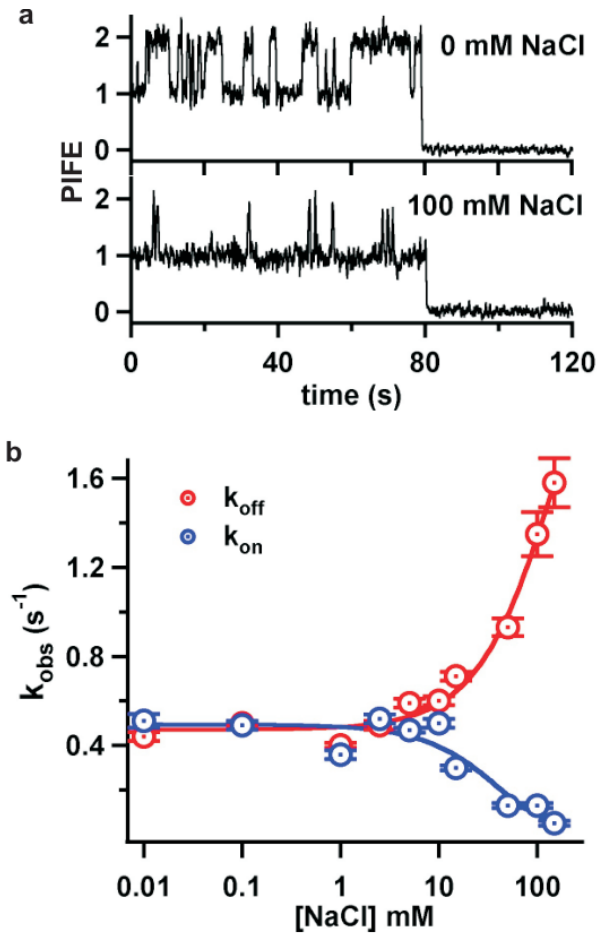

Figure 21. NaCl electrostatically competes with KF for binding to 8Cy3 DNA. (**a**) Comparison of smPIFE trajectories of KF (5 nM) binding to DNA at 0 mM NaCl (top) and 100 mM NaCl (bottom). In both cases photobleaching of Cy3 occurred at ~80s. (**b**) Association (blue) and dissociation (red) rates at increasing concentrations of NaCl. The fits are single exponential and the error reported is the error of the single exponential fits of the dwell times at each NaCl concentration.

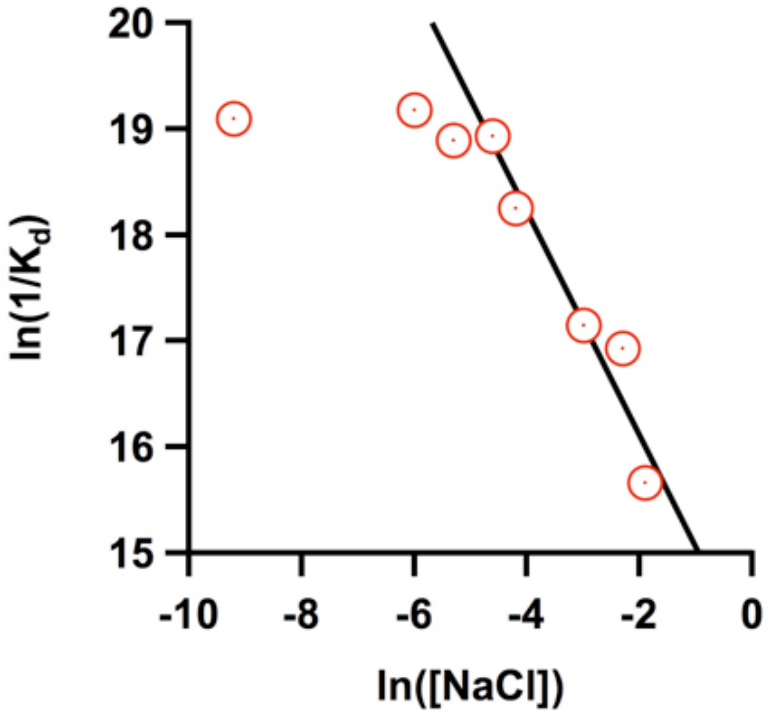

Figure 22. KF releases one monovalent ion upon binding DNA. The influence of monovalent ions on the stability of the KF-DNA binary complex was determined by plotting the natural log of the inverse K_Ds as a function of the natural log of the NaCl concentration. The slope of the linear region of the linkage plot (-1.0 ± 0.2) gives the number of ions released upon KF binding to the DNA. This binding was not significantly influenced by NaCl at concentrations lower than that shown in the linear region of the plot.

4.2.4. NTPs modulate the KF-DNA complex stability

We next used smPIFE to study the effect of correct and incorrect NTPs on the KF binding dynamics. For these experiments, we used a non-extendable 2',3'-dideoxy terminated primer (8Cy3dd) to prevent nucleotide incorporation. The observed FRET ratio is identical to 8Cy3, indicating that KF binds the dideoxy primer in the same position and orientation as the 3'-OH terminated primer. Dwell time analysis also shows that KF associates with the dideoxy primer with approximately the same rate constant

(k'_{on}), although the lack of the 3'-OH did result in a two-fold decrease in the dissociation rate (k_{off} = 0.18 s^{-1}) (cf. **Figure 20** and **Figure 23**). A two-fold decrease in k_{off} corresponds to ~0.4 kcal/mol stabilization of the binary complex.

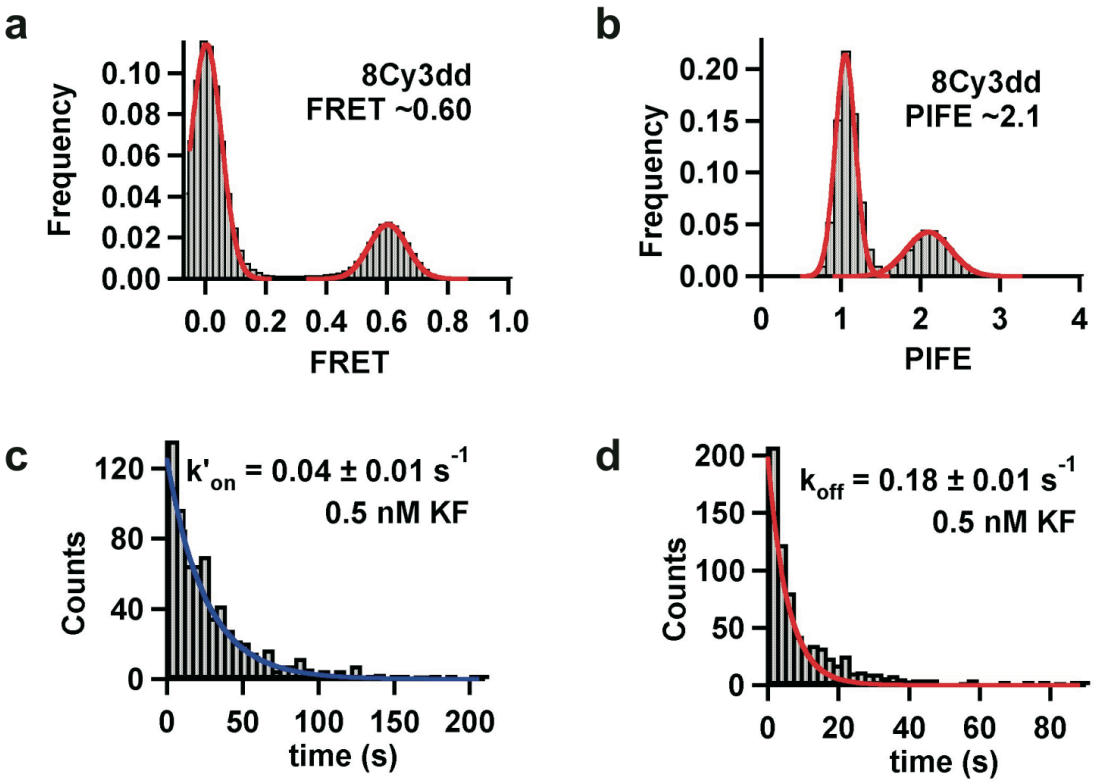

Figure 23. Lack of a 3' OH does not change the binding position or pseudo-first order k'_{on}. (**a** and **b**) FRET and PIFE values are essentially identical to the primer-template containing a 3' OH terminus (*cf.* **Fig. 9**). Each FRET and PIFE histograms was built from over 100 molecules and the data fit to Gaussian distributions. **(c)** Dwell time distribution for the pseudo first order on rate (0.04 s^{-1}) determined using PIFE was essentially identical to the value determined (0.05 s^{-1}) from the best fit line shown in **Fig. 20** for binding to the primer-template containing a 3' OH terminus. **(d)** Off rate for KF binding (0.18 s^{-1}) determined using PIFE was about one half the value determined (0.37 s^{-1}) for binding to the primer-template containing a 3' OH terminus.

In the presence of the next correct nucleotide, in this case dCTP, (**Figure 24**) numerous binding events longer than 10 min were observed, indicating a more stable ternary complex. The resulting dissociation rate constant decreases ten-fold (k_{off} = 0.016 s^{-1}), while the association rate remains constant (**Figure 25**). These results show that the correct nucleotide stabilizes the ternary complex by ~1.4 kcal/mol. In the presence of dCTP, the calculated dissociation constant K_D = 0.16 nM is in good agreement with prior results(155-157). While this prolonged binding would not be expected to occur during DNA synthesis when the primer contains a 3'-OH, these long binding events suggest the correct nucleotide induces the formation of the stable closed ternary complex that is necessary for phosphodiester bond formation to occur(142).

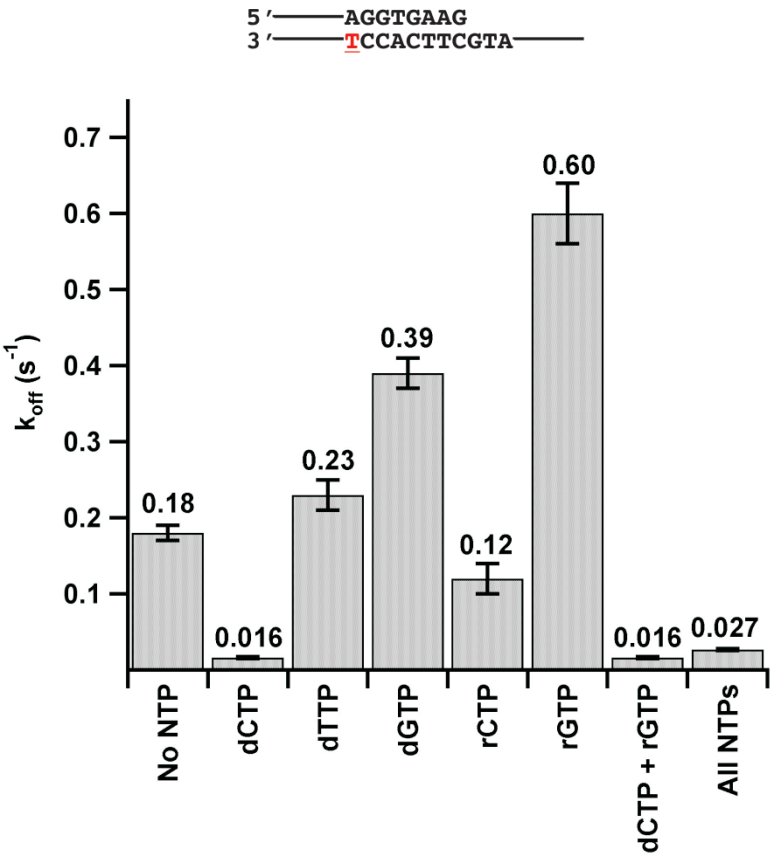

Figure 24. KF rejects incorrect NTPs in favor of the correct dNTP. k_{off} was measured as shown in Figure 3c in the absence or presence of the indicated dNTP or rNTP. Except when all eight NTPs were present, the concentration of nucleotide was 200 μM. When all NTPs were present, the concentration of the dNTPs was 200 μM and the concentration of the rNTPs was 2 mM. All off rates are the averages from titrations of at least five different concentrations of KF (see **Figure 25**), with the exception of the "dCTP + rGTP" and "All NTPs" samples in which the off rates were only measured at 0.5 nM KF. The errors are the standard deviations of the averages, with the exception of the "dCTP + rGTP" and "All NTPs" samples in which the errors reported are the errors of the single-exponential fits to the dwell time distributions.

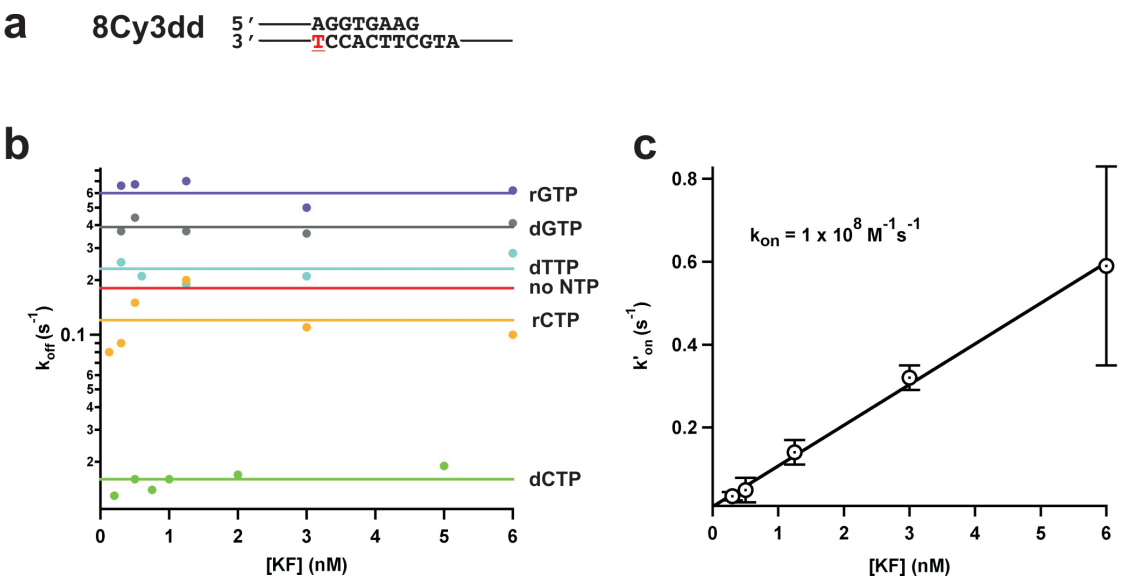

Figure 25. Dissociation and association rates in the presence or absence of NTPs. (**a**) Dissociation rates were determined at increasing concentrations of KF in the presence of 200 μM rGTP, dGTP, dTTP, rCTP, dCTP, or no NTP as indicated in the right hand margin. Each data point is calculated from a dwell time distribution created from greater than 100 molecules. The lines are the average of the distribution of points for each condition. All errors from the fits of the single-exponentials of the dwell times are less than 10% of the value of the rate (errors not shown for simplicity). (**b**) The association rates were determined from the averages of each of the association rates measured for each of the experimental conditions in (**a**) for each KF concentration shown. The error reported is the standard deviation of the averages. The true second order k_{on} determined from the slope is 10^8 $M^{-1}s^{-1}$, which is diffusion limited and identical to the value determined for binding to 8Cy3 in the absence of nucleotides (**Figure 20c**).

Similarly, we tested the effect of an incorrect pyrimidine (dTTP) and purine (dGTP). In both cases, the association rate remains constant, and only changes in the dissociation constant were observed (**Figure 24** and **Figure 25**). Our data reveal that a purine-purine mismatch induces greater destabilization (~0.5 kcal/mol) than a pyrimidine-purine mismatch (~0.1 kcal/mol). Purine-purine mismatches are known to differ significantly from the standard Watson-Crick base pair shape, and therefore, may cause a greater steric

clash in the active site during nucleotide selection than pyrimidine-purine mismatches(156,158,159).

In the cell, the concentration of rNTPs is much higher than that of dNTPs(16) and therefore DNA polymerases must also be capable of discriminating between ribo- and deoxynucleotides. In the presence of rGTP, we observe faster dissociation rate constants but similar binding rate constants compared to all other NTPs tested (**Figure 23**). We find that KF dissociates about 1.5-fold faster in the presence of rGTP compared with dGTP, indicating that the 2'-OH of the incoming nucleotide further destabilizes the ternary complex by ~0.3 kcal/mol. rCTP caused a slightly slower dissociation rate constant compared with no nucleotide, but dissociated about 8-fold faster than in the presence of dCTP (**Figure 24** and **Figure 25**). This suggests that rCTP can partially stabilize the ternary complex by ~0.2 kcal/mol. Interestingly, the observed partial stabilization in the presence of rCTP is not a simple compensation between the proper base pairing stabilization (~1.4 kcal/mol) and the 2'-OH destabilization (~0.3 kcal/mol), indicating that additional interactions are involved in the initial discrimination step.

4.2.5. KF readily rejects incorrect NTPs in favor of the correct dNTP

The observed stabilization by the correct dNTP and destabilization by incorrect NTPs, raises the question of how the polymerase selects the correct dNTP in presence of a pool of more highly concentrated incorrect NTPs without the incorrect NTP causing dissociation. To address this question, we measured the association and dissociation rate constants of KF in presence of an equimolar mixture of dCTP and rGTP, the correct

nucleotide and the most destabilizing nucleotide, respectively. Under these conditions we observed an identical dissociation rate constant compared with what was observed for dCTP alone (k_{off} = 0.016 s^{-1}) (**Figure 24**). When a near-physiological mixture of dNTPs (200 μM) and rNTPs (2 mM) was present, the dissociation rate increased to 0.027 s^{-1}. The small change in k_{off} as compared with the presence of only the correct nucleotide is likely due to the increased ionic strength from the counter ions associated with the 8.8 mM nucleotide concentration, which we estimate leads to a ~2-fold rise in k_{off} (**Figure 21)**. Taken together, these results suggest that when a mixture of nucleotides is present the incorrect nucleotide is rejected at a step preceding the proposed steric clash that leads to a faster dissociation rate constant in presence of a single mismatched nucleotide.

4.2.6. Exonuclease-site binding only observed with mismatched termini

Structural analysis of KF and its close homologues has shown that the presence of a mismatched primer terminus leads to the movement of the 3'-nucleotide from the polymerase to the exonuclease domain(8,23). Following excision of the mismatched terminal nucleotide, the primer strand presumably reanneals to the template and DNA synthesis resumes. It is interesting, and somewhat surprising, that a crystal structure of KF bound to fully complementary primer-templates shows binding to the exonuclease site(144) and that several ensemble measurements have shown that a fully paired primer binds to the exonuclease site 20 to 40% of the time(24,138,143). To test if different smFRET or smPIFE values could be observed for exonuclease site primer binding, we prepared a 9Cy3 primer-template containing two mismatched nucleotides at the primer

terminus (**Figure 26a**). Prior ensemble averaged studies have shown that such a primer binds exclusively to the exonuclease site(24,139). Our smPIFE data show PIFE values increased from 1.2 for the fully paired primer-template (9Cy3, **Figure 18**) to 1.8 for the double mismatch (9Cy3mm, **Figure 26**). Similarly, the smFRET efficiency increased from 0.43 to 0.59. This increase in FRET is consistent with the Cy5 on the polymerase moving closer to the Cy3 on the template and is identical to the value obtained for a fully paired 8Cy3 primer-template (**Figures 18** and **26**).

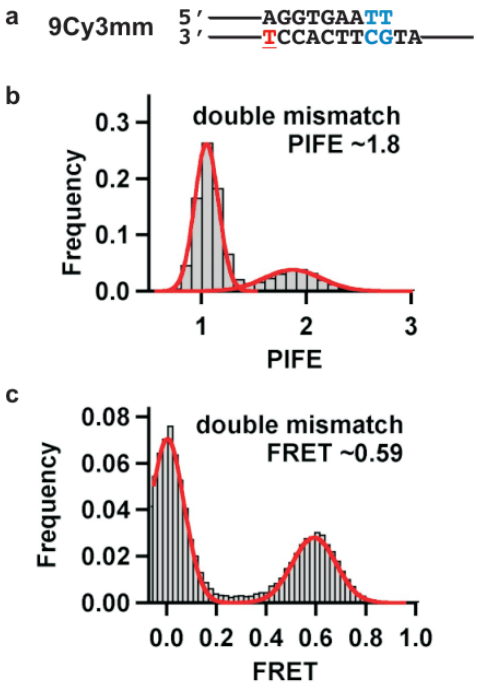

Figure 26. Exonuclease site binding is induced by a double mismatch. (**a**) The double mismatched primer-template structure. (**b**) The PIFE histogram for KF binding to the double mismatched primer-template shown in (**a**). (**c**) The FRET histogram for KF binding to the double mismatched primer-template shown in (**a**).

Overall, our results indicate that, in the presence of a fully paired primer, the DNA binds exclusively in the polymerase site. In the presence of a mismatched template, the DNA binds in a different orientation consistent with binding in the exonuclease site.

4.3. Discussion

We have used smFRET and smPIFE to measure and characterize the binding of KF to DNA primer-templates in real time. KF binding to templates with increasingly longer primers shows that the observed FRET ratios decrease with each addition to the length of the primer, as previously shown(50). Similar experiments measuring smPIFE showed that the PIFE remains unchanged with increasing primer length until the primer-template junction was nine base pairs downstream from the Cy3, at which point the PIFE abruptly decreased. As the donor fluorescence enhancement for PIFE requires the interaction between the Cy3 and the DNA polymerase, the sudden decrease in PIFE suggests that the Cy3 is no longer within the footprint region of the polymerase. Our experimentally determined KF footprint of eight base pairs is in agreement with previously established ensemble values(146,147).

DNA polymerases must efficiently select against the incorporation of incorrect nucleotides during replication to prevent unacceptably high levels of mutagenesis. For KF, it has been suggested that selection against mispaired rNTPs and dNTPs occurs while the fingers domain is in the open conformation(12,160). Selection against the complementary rNTP is thought to occur as the fingers domain closes(12), at which point the 2'-OH of the ribose sugar has been reported to be sterically blocked by Glu-710 in the

polymerase active site(161). In agreement with other ensemble studies(68,73,156), we also find the polymerase-DNA complex was destabilized in the presence of an incorrect dNTP or rNTP, with the notable exception of the rNTP complementary to the templating base, for which we observed stabilized polymerase binding. Another single-molecule study suggested the fingers domain of the polymerase was in a "partially closed" state in the presence of the complementary rNTP, which also could lead to increased stability of the polymerase-DNA-complementary rNTP complex(51).

The destabilizing effect of mismatched nucleotides is difficult to square with the requirement for a DNA polymerase to remain bound to the DNA template during processive DNA synthesis. While it is easy to understand how the next correct dNTP causes an enhanced stability of the polymerase-DNA complex by triggering a conformational change from the open binary to the more stable closed ternary complex, the destabilizing effect of non-complementary dNTPs and rNTPs, which are present in large excess over the properly paired dNTP(16), is counterintuitive, especially considering the evidence that most mispaired nucleotides seem to be rejected prior to the closing of the fingers(12). However, when we measured the dissociation rates for KF in the presence of mixtures of correct and incorrect nucleotides, even where the incorrect nucleotides are present in large excess, we found that the k_{off} remains essentially unchanged compared with that observed when only the correct nucleotide is present.

We conclude from these results that there appears to be a different mechanism in place when only an incorrect nucleotide is present compared with a mixture of correct and incorrect. Although we cannot be sure of the mechanism causing these effects

without further structural characterization of the polymerase complex in the presence of an incorrect nucleotide, one possible explanation for this observation that is consistent with the prior studies(12,160) is that the presence of only the incorrect nucleotide allows the movement of the templating base from the pre-insertion to insertion site, possibly at a much slower rate compared with a correct nucleotide. In this scenario, when a mixture of nucleotides is present, incorrect nucleotides are rejected at the pre-insertion site until a correct nucleotide binds. However, in the absence of the correct nucleotide, eventually a mispaired nucleotide results in the movement of the nascent base pair to the insertion site, causing the initiation of fingers closure and a steric clash in the active site induced by the mispair and resulting in an enhanced dissociation rate of the polymerase. In addition, the dissociation rate constants and K_Ds in the presence of different mispairing dNTPs(73) show that the larger the size of the mispair the less stable the binding of KF to the primer-template. Thus, it appears that the larger the size of the mispair, the greater the steric clash when the fingers closure initiates.

Under the rare circumstances that misincorporation of a non-complementary nucleotide occurs, there is substantial evidence that several terminal base pairs melt and the single-stranded primer DNA is transferred to the exonuclease site(23,144). This placement allows for the excision of the incorrect terminal 3' nucleotide followed by the return of the primer strand to the polymerase site for continued synthesis. KF binding at the exonuclease site is thought to cause the movement of the polymerase upstream along the duplex DNA by about 2 or 3 nucleotides(162).

Crystallographic and several ensemble biochemical studies have suggested that even when the primer template is properly paired, the primer strand can show significant levels of partitioning to the exonuclease site(24,138,144). Here, we have shown that we can use both smFRET and smPIFE to detect exonuclease site binding using a double mismatched primer-template, which has been shown to bind exclusively to the exonuclease site(24,139). Using these techniques, we have attempted to detect either exonuclease site binding for a paired primer-template or dynamic exchange between exonuclease and polymerase site binding. In our hands, we see no evidence that a properly paired template is positioned outside of the polymerase active site. This raises the question of what factors might lead to the differences observed for the ensemble studies and these single molecule results.

When the crystal structure of KF was first published, it was surprising that it showed the primer terminus bound exclusively to the exonuclease site (144). However, the co-crystals of the DNA and KF were formed at high ionic strength and it is possible these conditions reduced the frequency of polymerase site binding. Fluorescence depolarization studies were later used to measure the partition coefficient for movement between the polymerase site and exonuclease site and found that up to 14% of the primer termini were in the exonuclease site, depending on the sequence at the 3-terminus of the primer(24). Fluorescence and circular dichroism measurements of primers containing 2-aminopurines as the two 3'-terminal base pairs showed about a 43% occupancy in the exonuclease site(138). However, these studies were performed in the presence of Ca^{2+} rather than Mg^{2+} and it is unclear if the 2-aminopurine contributed to higher levels of exonuclease

binding. Finally, surface plasmon resonance has been used to measure the dissociation rates from DNA bound to the polymerase site and exonuclease site(139). Similar to our work, these studies were carried out with DNA containing no modifications near the primer-template junction and under standard polymerase buffer conditions and showed that about 97% of the polymerase molecules had the primer bound in the polymerase site.

The work herein provides a comprehensive assessment of DNA polymerase binding dynamics with DNA at the single-molecule level. We have utilized two distinct approaches, smFRET and smPIFE, to observe the interactions between the polymerase and the DNA in real time. We found these two approaches naturally complement each other by individually providing unique information about the system. smFRET enables us to track the position of the polymerase on the DNA with single-base pair resolution and can be used to distinguish between polymerase and exonuclease site binding. smPIFE, which is not affected by acceptor bleaching or blinking, was used to determine the binding kinetics in absence and presence of dNTPs and rNTPs, and was able to precisely measure the binding footprint of KF on DNA. In the future, this method can easily be used to study other DNA binding proteins, to characterize the potentially mutagenic interactions of DNA polymerases with mispaired primer termini or carcinogenic adducts linked to the DNA, and to investigate further the mechanism by which DNA polymerases maintain their remarkable fidelity.

Chapter 5: carcinogenic adducts induce distinct polymerase binding states

Portions of the text in this chapter were reprinted or adapted with permission from:

Vrtis, K.B., Markiewicz, R.P., Romano, L.J., Rueda, D. (2013) Carcinogenic DNA adducts induce distinct DNA polymerase binding orientations. Nucleic acids research 41, 7843-7853.

5.1. Introduction

Most cancers propagate from an accumulation of mutations in genes that control cell growth(26,163). It is estimated that only 5-10% of cancer cases are attributed to inherited genetic defects, with the remaining cases resulting from lifestyle and environmental factors, such as chemical carcinogens(27). Arylamines are a well-studied class of carcinogen found in numerous occupational settings, tobacco smoke and chemical dyes(164-166). *N*-acetyl-2-aminofluorene is a potent, arylamine mutagen that, following metabolic activation *in vivo*, forms two different adducts at the C8 position of guanine

bases (**Figure 27**): 2-aminofluorene (AF-dG) and *N*-acetyl-2-aminofluorene (AAF-dG)(39-41).

While AF-dG and AAF-dG adducts differ by only an acetyl group, the two adducts have drastically different effects on the structure of duplex DNA(42-44,118,167). AF-dG adducts in DNA are most stable in the *anti* conformation (**Figure 27b**), which positions the AF moiety into the major groove allowing the guanine to participate in canonical base pairing with cytosine(42,43). The acetyl group from the AAF moiety destabilizes the *anti* conformation of the AAF-dG adduct apparently due to a steric clash between the AAF actetyl group and the deoxyribose(44), which leads to a distorted DNA structure having the AAF-dG in a *syn* conformation (**Figure 27b**). This structure places the bulky AAF moiety in the duplex DNA region exposing the guanine base to solvent in the major groove(45). It is thought that these structural differences lead to different effects on DNA synthesis and are the cause of the distinct mutation profiles that have been observed(74,168): AF-dG adducts cause high-fidelity polymerases to pause before bypass occurs(169,170) and primarily induce substitution mutations(171), while AAF-dG adducts are strong blocks to DNA synthesis(170) and predominantly induce frameshift mutations(168,171,172). While AF and AAF are amongst the most thoroughly studied carcinogenic DNA adducts(173), the molecular mechanism by which either adduct inhibits DNA synthesis and promotes mutagenesis remains unclear.

DNA polymerases are structurally analogous to a right hand, complete with fingers, thumb, and palm domains (pol site) (**Figure 27a**)(8). During 5'-3' template-directed polymerization, the formation of a phosphodiester bond between the incoming

deoxynucleoside 5'-triphosphate (dNTP) and the 3'-OH of the growing primer strand is catalyzed in the pol site. Additionally, most replicative polymerases are associated with a 3'-5 proofreading exonuclease activity(77,174,175). This activity can be part of the polymerase itself (exo site) or be located on a separate subunit, but, most significantly, in each case it is positioned a significant distance away from the pol site (**Figure 27a**). As the name suggests, the role of the proofreading activity is to increase the accuracy of the polymerase by excising misincorporated nucleotides(10,17,77). DNA transfer between the pol and exo sites must be carefully regulated because unnecessary nucleotide excision from the growing DNA strand would needlessly slow DNA synthesis and consume essential chemical energy in the cell. The mechanism by which DNA polymerases transfer the DNA from the pol to the exo site remains elusive.

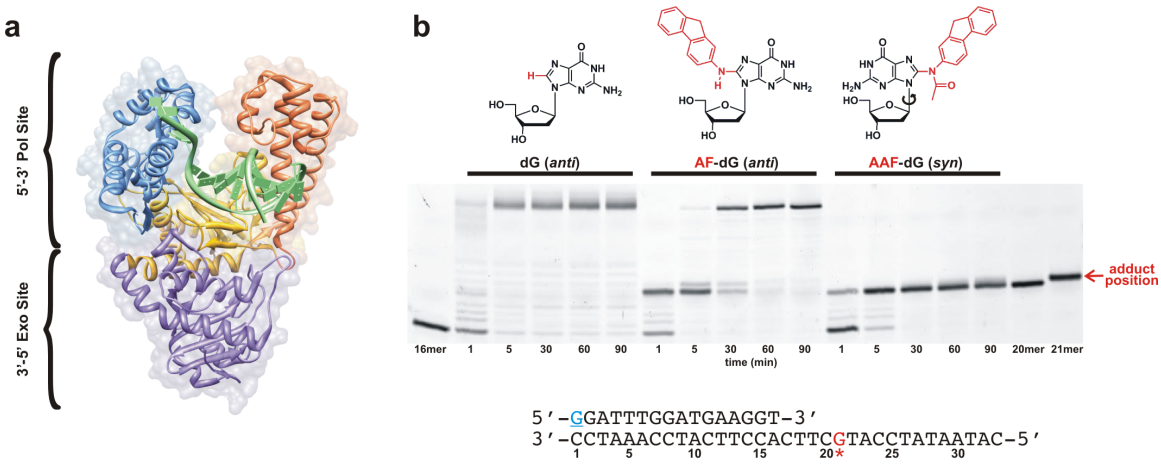

Figure 27. Carcinogenic adducts induce polymerase stalling on the DNA. (**a**) DNA polymerase structure. The fingers (blue), thumb (orange), and palm (yellow) domains encompass the pol site, where 5'-3' template directed DNA synthesis occurs. Misincorporated nucleotides can be excised at the exo site (purple), which increases the overall fidelity of the polymerase. Crystal structure from *Bacillus stearothermophilus* DNA polymerase I (PDB ID 1l3s), a close structural homolog of KF. (**b**) The chemical structures of deoxyguanosine (dG), *N*-(deoxyguanosin-8-yl)-2-aminofluorene (AF-dG), and *N*-(deoxyguanosin-8-yl)-*N*-acetyl-2-aminofluorene (AAF-dG) are displayed in their most stable conformations in duplex DNA. Below the structures is a running start DNA polymerase extension assay for the extension of the primer-template shown below the gel. The primer has a 5'-Cy3 attached to the underlined, blue G. The template either has an unmodified dG, a AF-dG, or a AAF-dG at position 21 (red G in template). Three different reactions were carried out with the indicated primer-templates, with aliquots being removed from the reaction mixture and stopped by addition of an equal volume of loading buffer (10 mM EDTA, 1 mg/mL bromophenol blue, in 10 mL of formamide) at the indicated time points. The samples were run on a 20 % denaturing polyacrylamide gel and scanned for Cy3 on a Typhoon 9210 Variable Mode Imager (GE Healthcare).

In the present study we have used two powerful real-time single-molecule approaches to monitor the interactions between *E. coli* DNA polymerase I (Klenow fragment, KF) and DNA primer-templates containing either an AF-dG or AAF-dG adduct. To help elucidate the mechanism by which these adducts affect DNA synthesis, we have determined the effects of these adducts on DNA polymerase binding to a primer-template

in which the primer terminates either before or across from the adduct position. Our data show that either adduct linked to the template in the single-stranded DNA (ssDNA) region does not disrupt the binding orientation of the DNA polymerase. However, positioning either adduct on the templating guanine of the terminal base pair causes the polymerase to bind in two distinct orientations: one that is consistent with the primer positioned in the exo site and a second, previously unreported, intermediate orientation that is distinct from either pol or exo site binding. The presence of the next correct dNTP rescues pol site binding in the case of the AF-dG adduct, while this change is not observed for the AAF-dG adduct. This distinction possibly provides further evidence for why AAF-dG adducts cannot be bypassed by high-fidelity polymerases while AF-dG adducts are bypassed after a brief stall at the adduct position. The intermediate orientation was also observed in the presence of a single mismatched primer, raising the intriguing possibility that this orientation represents a key intermediate in the polymerase proofreading mechanism.

5.2. Results

5.2.1. Carcinogenic adducts induce polymerase stalling on the DNA

The effect of an adduct on DNA synthesis depends on the sequence context within which the adduct is located, the structure and orientation of the adduct in the DNA template, and the properties of the DNA polymerase. To measure the degree of inhibition by either an AF-dG or AAF-dG adduct positioned in our particular primer-template system (**Table 2**), we measured running start DNA synthesis on the same AF or AAF-

modified DNA template sequences used in the single-molecule studies. KF fully extends the 16mer/33mer unmodified primer-template within 5 min (**Figure 27b**). Using a template containing an AF-dG adduct (**Table 2**), we found that KF transiently stalls one base before and across the adduct (**Figure 27b**) and then fully extends the primer within 60 min. However, with the analogous AAF-dG-modified primer-template, the polymerase is completely blocked one base before the adduct and no evidence for extension is observed after extended incubations (**Figure 27b**). Similar results were obtained using a template in which the AF or AAF-dG adduct was initially located at the templating position or across from the primer terminus (**Figures 28** and **29**).

Table 2**.** DNA sequences for ensemble and single-molecule experiments

Extension gel assay oligonucleotides[1]	
16mer-Cy3 primer	5’-[Cy3]-GGA TTT GGA TGA AGG T-3’
20mer-Cy3 primer	5’-[Cy3]-GGA TTT GGA TGA AGG TGA AG-3’
21mer-Cy3 primer	5’-[Cy3]-GGA TTT GGA TGA AGG TGA AGC-3’
33mer template	3’-CCT AAA CCT ACT TCC ACT TCG TAC CTA TAA TAC-5’
Single molecule oligonucleotides[2,3]	
20mer-biotin primer	5’-[Biotin]-GGA TTT GGA TGA AGG TGA AG-3’
21mer-biotin primer	5’-[Biotin]-GGA TTT GGA TGA AGG TGA AGC-3’
21mer single T mm primer	5’-[Biotin]-GGA TTT GGA TGA AGG TGA AGT-3’
21mer double TT mm primer	5’-[Biotin]-GGA TTT GGA TGA AGG TGA ATT-3’
33mer-Cy3 template	3’-CCT AAA CCT ACT TCC ACT TCG TAC CTA TAA TAC-5’

[1] Underlined red G was modified with either AF or AAF (see methods).

[2] Cy3 is linked to the underlined blue T (see methods).

[3] Green T’s highlight mismatched (mm) nucleotides on the primer.

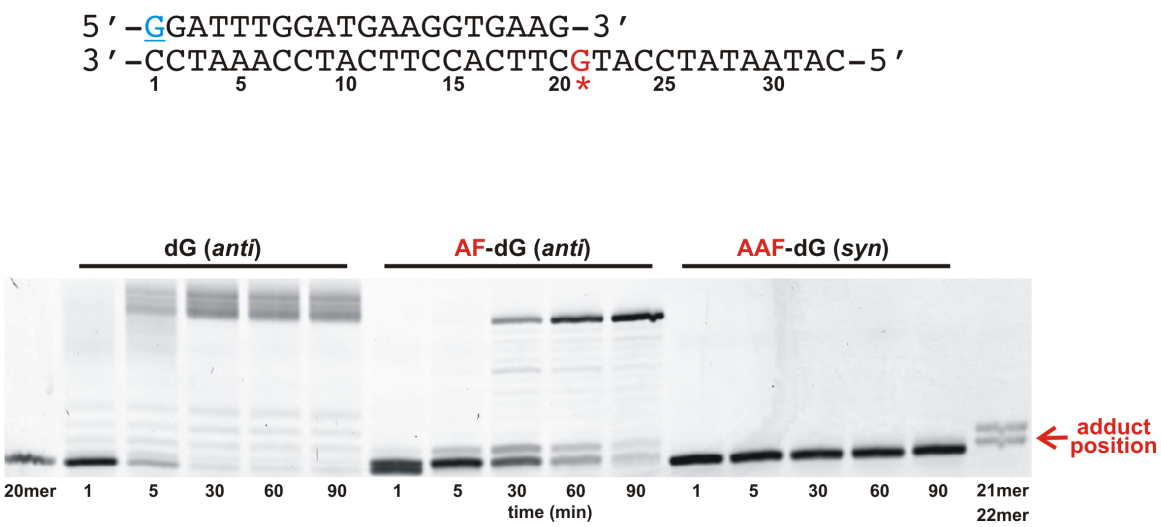

Figure 28. DNA extension reactions on modified DNA using a standing start. The reactions were carried out identically to extension reactions shown in Figure 1 except that the primers used in these reactions terminated one base before the adduct position (20mer primer).

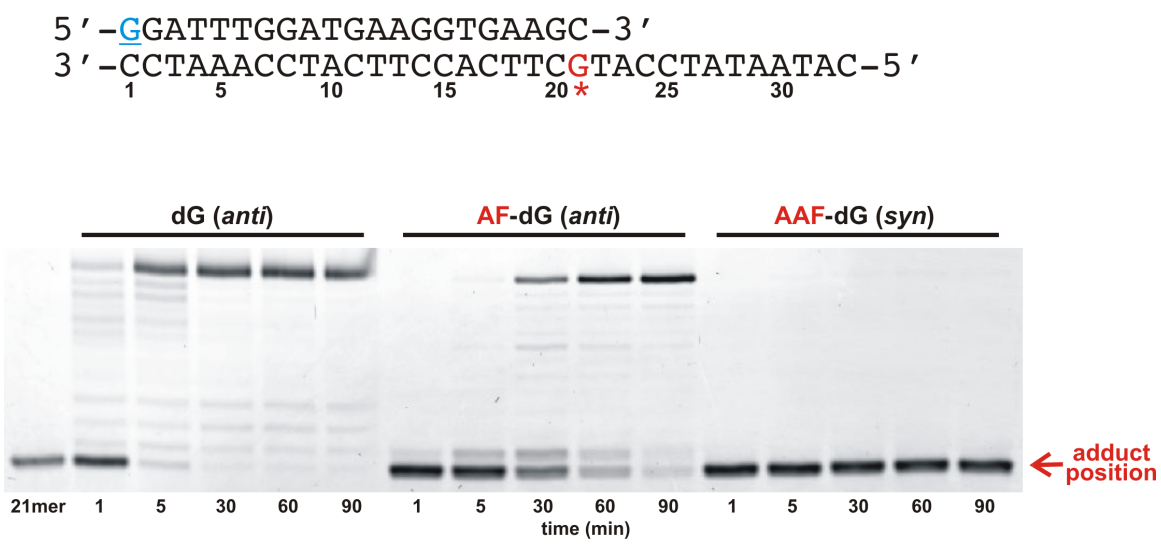

Figure 29. Polymerase can not extend AAF primer when the adduct is at duplex DNA terminus. The standing start reactions were carried out identically to the 16mer extension shown in Figure 1; however, the initial primers used in these reactions terminated across the adduct (21mer primer). The unmodified dG primer-template and the AF-modified primer-template were nearly fully extended by 5 min and 90 min, respectively. No extension occurred on the AAF modified primer-template.

The mechanisms that lead to these polymerization results can be explained by several scenarios, some of which might be generalized to other bulky adduct structures. First, it is possible that the bulky adducts induce the polymerase to dissociate from the DNA prior to dNTP incorporation; second, the adduct might prevent the proper alignment of the templating base in the pol site; or third, the bulky adduct could prevent the formation of a closed ternary complex by interfering with fingers closing or dNTP binding. Biochemical(176) and crystallographic(118) evidence has suggested that an AAF-dG adduct in the templating position inhibits incorporation using this latter scenario by preventing the movement of the fingers to form a closed ternary complex. However, the mechanism by which an AF-dG adduct causes a polymerase to stall near the adduct site or how an AAF-dG adduct inhibits extension remains unknown. Because of the limitations of ensemble-averaged experiments to measure dynamic processes that might distinguish these possible mechanisms, we have turned to single-molecule fluorescence resonance energy transfer (smFRET) and protein induced fluorescence enhancement (smPIFE) methods to monitor the binding position and dynamics of individual polymerase on AF and AAF-modified DNA templates in real time(50,52).

5.2.2. Polymerase-DNA binding dynamics monitored by smFRET and smPIFE

Single-molecule approaches reveal transient events from individual molecular interactions in real time and unveil heterogeneity within subpopulations of molecules that would otherwise remain hidden in ensemble-averaged experiments(49,137,177). By strategically labeling the DNA and polymerase, we can monitor polymerase binding

dynamics with single-base pair resolution(50,52). In our assay, free Cy3-labeled DNA primer-template exhibits only Cy3 fluorescence (I_D) (**Figure 30a**). Upon Cy5-labeled polymerase binding, I_D decreases concurrently with an anticorrelated increase in Cy5 fluorescence intensity (I_A) (**Figure 30a**). The apparent FRET efficiency is calculated as FRET = $I_A/(I_A+I_D)$ (**Figure 30a**). Because FRET efficiency depends on relative fluorophore distance and orientation, it reports on polymerase position and orientation on the DNA.

We have also shown that binding dynamics can be monitored with unlabeled polymerase by smPIFE(52). Polymerase binding near Cy3 changes the local fluorophore environment (viscosity and polarity), thus increasing its fluorescence quantum yield, resulting in subsequent fluorescence increases(52,55,56,178). We normalize PIFE to 1.0 for unbound DNA, such that relative PIFE increases report on polymerase binding events (**Figure 30b**). Although smPIFE does not provide precise distance information, it offers the advantage that it is not susceptible to acceptor blinking or photobleaching, thereby providing an accurate method to measure binding kinetics(52). Using smFRET and smPIFE in tandem provides a clearer picture of polymerase dynamics on the DNA than can be obtained using either method alone.

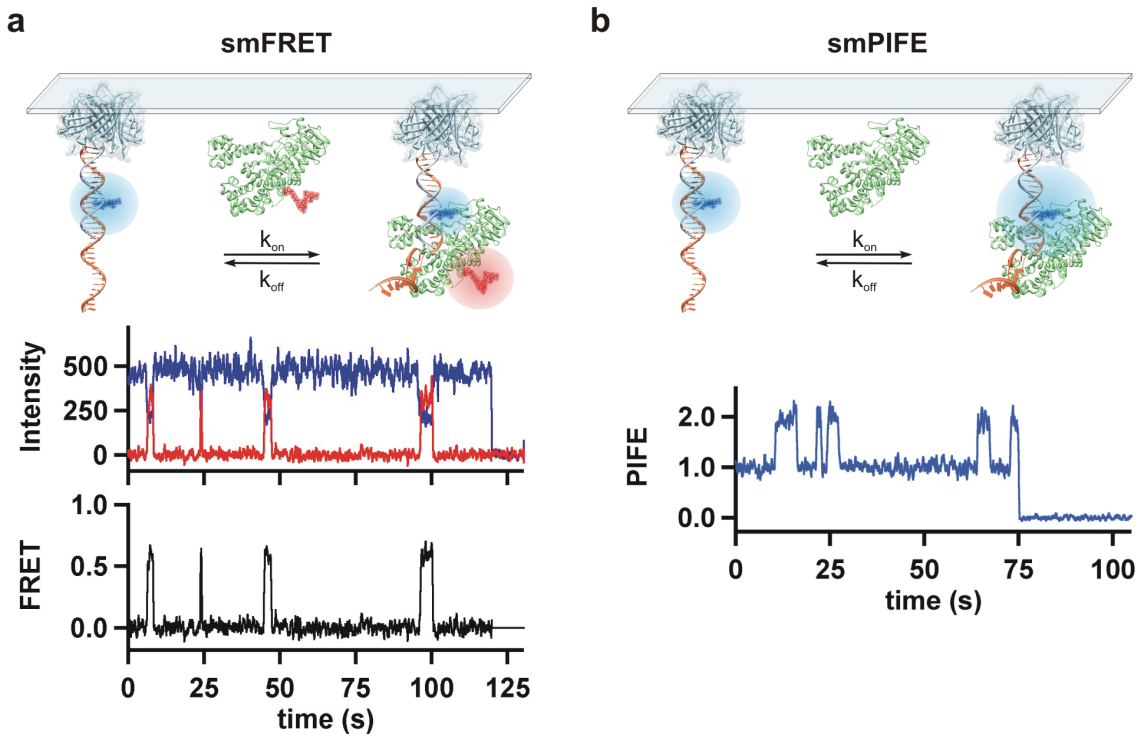

Figure 30. Two single-molecule approaches to monitor polymerase interactions with the DNA in real time. (a, top) Schematic of single-molecule FRET design. Upon polymerase binding to the DNA, energy is transferred from the donor Cy3 (blue) on the DNA template to the acceptor Cy5 (red) conjugated to KF. (**a**, **bottom**) Representative donor (blue) and acceptor (red) time trajectories for KF binding to the unmodified primer-template shown in Figure 3., and the FRET trajectory (black) calculated from the donor and acceptor intensities (FRET = $I_A/(I_A+I_D)$). (**b**, **top**) Schematic of single-molecule PIFE design. When the polymerase binds to the DNA in close proximity to the Cy3, the fluorescence intensity of the Cy3 is enhanced due to a change in the local environment about the fluorophore. (**b**, **bottom**) Representative PIFE trace (blue) for KF binding to the same primer-template used for (**a**). The Cy3 photobleached at ~75 s.

5.2.3. Adducts bound to the templating base in ssDNA stabilize the binary complex

To test whether bulky adducts impede polymerization by destabilizing the polymerase-DNA binary complex or by inducing polymerase misalignment, we determined binding rate constants and conformations using smFRET and smPIFE in the absence and presence of AF- and AAF-dG adducts on the templating base (**Figure 31a**). In the absence of an adduct, the observed FRET and PIFE values were 0.59 and 2.0, respectively (**Figure 31b**), in good agreement with prior results(52). The high PIFE value indicates that Cy3, located 8 nucleotides from the primer-template junction (**Figure 31a**), is within the molecular footprint of the polymerase. Interestingly, in the presence of either an AF or AAF-dG adduct at the templating base, the PIFE and FRET distributions remain unchanged within experimental error (**Figure 31c,d**). These results indicate that neither of these bulky adducts on the templating base affect the global polymerase position or orientation on the DNA primer-template (**Figure 31e**).

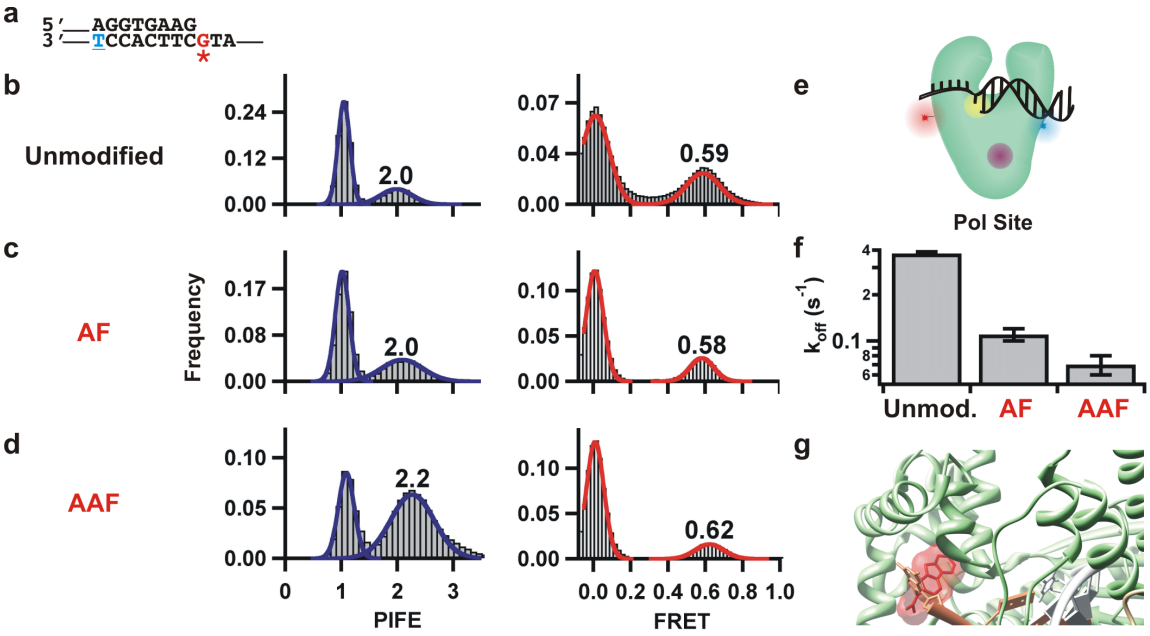

Figure 31. Carcinogenic adducts in ssDNA stabilize pol site binding. (**a**) Primer-template design for investigating the ensemble stalling trend observed (Figure 27b and Figure 28) one position before the adduct. The Cy3 is conjugated to the underlined, blue thymine in the template by an amine linker. For AF and AAF lesion templates, the adduct is attached to the red guanine. (**b-d**) The PIFE and FRET efficiency histograms for KF binding to (**b**) unmodified, (**c**) AF-modified, or (**d**) AAF-modified primer-templates. (**e**) Schematic of the primer-template bound at the pol site (yellow circle) of the DNA polymerase. During FRET, energy is transferred from the Cy3 (blue) to the Cy5 (red) dye; PIFE experiments are identical except they lack the Cy5 dye. When the polymerase binds the DNA at the pol site, the templating base and the 3' primer terminus are within the pol site. (**f**) Comparison of polymerase dissociation rates for the unmodified, AF, and AAF primer-templates. (**g**) Crystal structure with the AAF moiety intercalated into the fingers domain (PDB ID 1X9M).

We next used the distribution of dwell times in the high PIFE state to determine the dissociation rate constants (k_{off}, **Figure 31f**). In the absence of adducts, the polymerase dissociates with a rate constant $k_{off} = 0.40 \pm 0.01\ s^{-1}$, in good agreement with prior results(52). In the presence of an AF-dG adduct, k_{off} decreases 4-fold ($0.10 \pm 0.01\ s^{-1}$), indicating that the bulky adduct stabilizes the binary complex by 0.8 ± 0.1 kcal mol^{-1}. In the presence of an AAF-dG adduct, the k_{off} decreases even further ($0.07 \pm 0.01\ s^{-1}$), indicating that AAF stabilizes the binary complex by 1.0 ± 0.1 kcal mol^{-1}. In agreement with our prior results(155) these data show that the bulky adducts do not induce polymerase dissociation before nucleotide incorporation, but rather stabilize the binary complex. This additional pol site stabilization and the inability to incorporate nucleotides across from the AAF adduct may be due to the intercalation of the bulky fluorene ring into the polymerase's fingers domain, as was observed by biochemical studies(176) and in a T7 DNA polymerase co-crystal structure with a primer-template having an AAF-dG adduct located at the same position (**Figure 31g**)(118).

5.2.4. Adducts at duplex DNA terminus induce multiple binding states

We then determined the effect of positioning these adducts across from the primer terminus on polymerase binding orientation and kinetics. On an unmodified primer-template with a primer one nucleotide longer, KF binding to the pol site yields FRET and PIFE values of 0.4 and 1.2, respectively, (**Figure 32a**). As we have previously shown(52), this FRET value is consistent with the polymerase being one base pair further from the Cy3 donor and the PIFE value indicates that the KF footprint no longer interacts

with the Cy3. The presence of an AF-dG adduct across from the primer terminus caused the FRET and PIFE values to increase to 0.51 and 1.9, respectively (**Figure 32b**). These changes in FRET and PIFE suggest that this adduct induces the polymerase to bind in an orientation different from that which occurs with an unmodified template, an orientation that causes the Cy3 to be within the polymerase footprint. In the presence of a similarly positioned AAF-dG adduct (**Figure 32c**), the observed PIFE value is also increases to 1.9, but the FRET distribution reveals two populations centered at 0.50 and 0.63, indicating that the polymerase binds in two different orientations, one of which (0.50) is similar to that observed with the AF-dG adduct and another novel one (0.63).

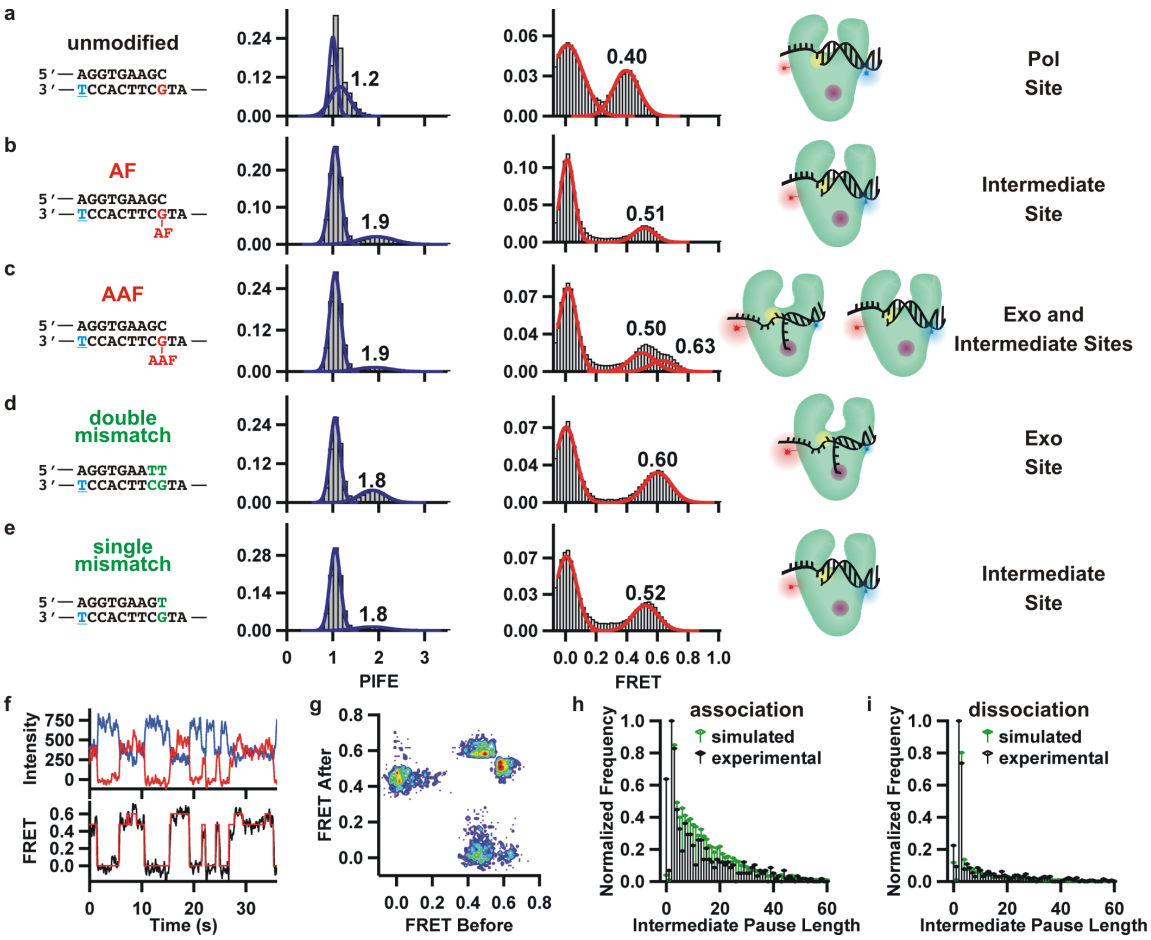

Figure 32. Carcinogenic adducts at duplex terminus induce distinct binding states. (**a-e**) Primer-template sequence, PIFE histogram, FRET histogram, and DNA polymerase binding states for (**a**) unmodified, (**b**) AF-modified, (**c**) AAF-modified, (**d**) double mismatch, and (**e**) single mismatch DNA. Cy3 is conjugated to blue thymine. Mismatches are shown in green. For the polymerase structures, Cy3 is shown in blue, Cy5 in red, the pol site yellow, and the exo site purple. (**f**) Representative donor (blue) and acceptor (red) time trajectories for KF binding to the AAF-modified primer-template shown in (**c**), and the calculated FRET trajectory (black) from the donor and acceptor intensities. The red line in the FRET trace is HMM analysis (see Methods for details). (**g**) Transition density plot for the AAF-modified primer-template shown in (**c**). (**h**) Initial exo site association to an AAF-modified primer-template occurs through an obligatory intermediate state. Pause lengths measured from 10,000 simulated traces in which the intermediate site (~0.50 FRET) was an obligatory binding step between the 0 and 0.63 FRET states (green bars) closely resemble the experimental exo site association pause lengths (black bars). (**i**) KF exo site dissociation from an AAF-modified primer-template terminus can occur directly from the exo site (60%) or through the intermediate site (40%). Simulated pause lengths for 40% dissociation from the exo site through the intermediate site and 60% direct

dissociation from exo site closely fit the experimental dissociation pause lengths. See **Figure 35** and Chapter 2 Methods for HMM and simulation details.

Taken together, these data imply that either an AF or AAF-dG adduct at the primer-template terminus base pair prevents normal pol site binding and causes the polymerase to be positioned closer to the Cy3. These results challenge an existing crystal structure of an analogous DNA polymerase, *Bacillus stearothermophilus* DNA polymerase I fragment (BF), bound to AF-modified DNA at this position, which reveals no change in the polymerase binding location relative to the duplex DNA terminus(167). A possible explanation for this discrepancy is the fact that BF lacks the 3'-5' exo site that is present in KF(179). These alternative binding orientations may reflect the primer binding to the polymerase at a secondary position, such as at the exo site, or at a site that participates in proofreading, possibly explaining why a these adducts either block or hinder DNA synthesis (**Figures 28** and **30**)(180).

5.2.5. Polymerase binds to adducts and single mismatch with an intermediate conformation

Prior structural and biochemical studies have identified two active sites in KF, the pol site located in the palm domain and the exo site located ~35 Å from the pol site(23). It is possible that the altered FRET and PIFE values observed when the adduct is present in the terminal base pair could be accounted for by the transfer of the primer strand to the exo site. To test for this possibility we compared these FRET and PIFE values with those obtained when the primer terminus contained mismatched nucleotides (**Figure 32d,e**).

Prior studies have shown that substrates containing two terminal mismatched base pairs bind almost exclusively to the exo site of KF(24,52,139). The observed PIFE (1.8) and FRET (0.60) values with the double mismatch substrate (**Fig 32d**) are both significantly higher than the unmodified complementary DNA (**Figure 32a**), in agreement with our prior results(52). These values resemble the second population (0.63 FRET) observed with the complementary AAF-adducted primer-template (**Figure 32c**), suggesting that this population corresponds to exo site binding. An AAF-adducted double mismatched primer-terminus DNA also yields very similar FRET (0.63) and PIFE (1.9) values (**Figure 33**), in support of this assignment. Furthermore, an AF adduct combined with the double mismatch yields 1.7 PIFE and 0.59 FRET, in very good agreement with unmodified, exo site binding (**Figure 33b**), yet distinct from the complementary AF-adducted primer-template.

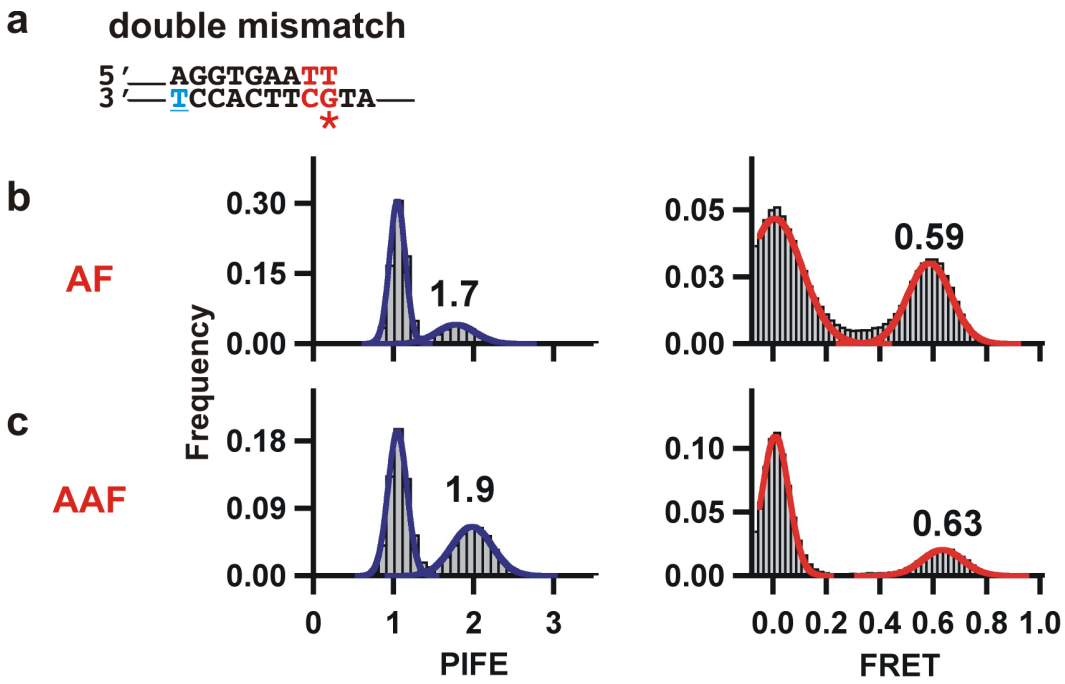

Figure 33. A double mismatch induces exo site binding for both an AAF and AF-modified primer-template (**a**) Double mismatch primer-template sequence with the adducts at the duplex terminus. The double mismatch at the terminus was used to induce exo site binding. The Cy3 is conjugated to the underlined, blue thymine in the template by an amine linker. The adducts are attached to the red, asterisk guanine. (**b** and **c**) PIFE (left) and FRET (right) histograms for polymerase binding to (**b**) AF- or (**c**) AAF-adducted primer-templates. The AAF PIFE and FRET values match the high FRET state from Figure 32c.

Using the same length primer-template, we also observed a FRET value of ~0.50 for both the AF-dG (**Figure 32b**) and AAF-dG adduct (**Figure 32c**), a value that does not correspond to either the pol or exo KF binding orientation. This intermediate site binding could be reproduced using a primer-template that contained a single mismatch (**Figure 32e**). It is possible that this intermediate state results from the rapid transfer of the primer terminus between the pol and exo sites at a rate that is faster than our time resolution (~50 ms). This would lead to a FRET value that is the average of pol and exo orientations. We

rule out this possibility because it would also yield an average PIFE value of ~1.5, which we do not observe (**Figure 32e**).

One possible explanation for this intermediate site binding is that it corresponds to a distinct binding orientation that is an intermediate in the proofreading process. This assignment is supported by the single-molecule FRET trajectories for the AAF-modified primer-template, which reveal direct transitions between the intermediate site and the exo site without polymerase dissociation (**Figure 34** and **Figure 32f**). Analysis of the trajectories using a hidden Markov Model (HMM) yields transition rate constants from the intermediate site to the exo site, $k_{exo} = 2.1 \pm 0.1\ s^{-1}$, and back to the intermediate site, $k_{int} = 4.3 \pm 0.4\ s^{-1}$. Each transition between the unbound DNA (0 FRET), intermediate site (~0.50 FRET) and exo site (~0.63 FRET) was counted and binned into a transition density plot (TDP) (**Figure 32g**)(120), which reveals four major transition peaks. Two of them correspond to association (**Figure 32g,** 0 FRET before, 0.5 FRET after) and dissociation (0.5 FRET before, 0 FRET after) transitions to and from the intermediate site, whereas the other two (0.63 before, 0.5 after and 0.5 before, 0.63 after) correspond to shuttling between the exo and intermediate sites. The lack of prominent peaks representing direct exo site binding (0 before, 0.63 after or 0.63 before, 0 after) raises the intriguing possibility that binding of the primer terminus to the intermediate site is an obligatory step prior to initial binding at the exo site or dissociation from the exo site.

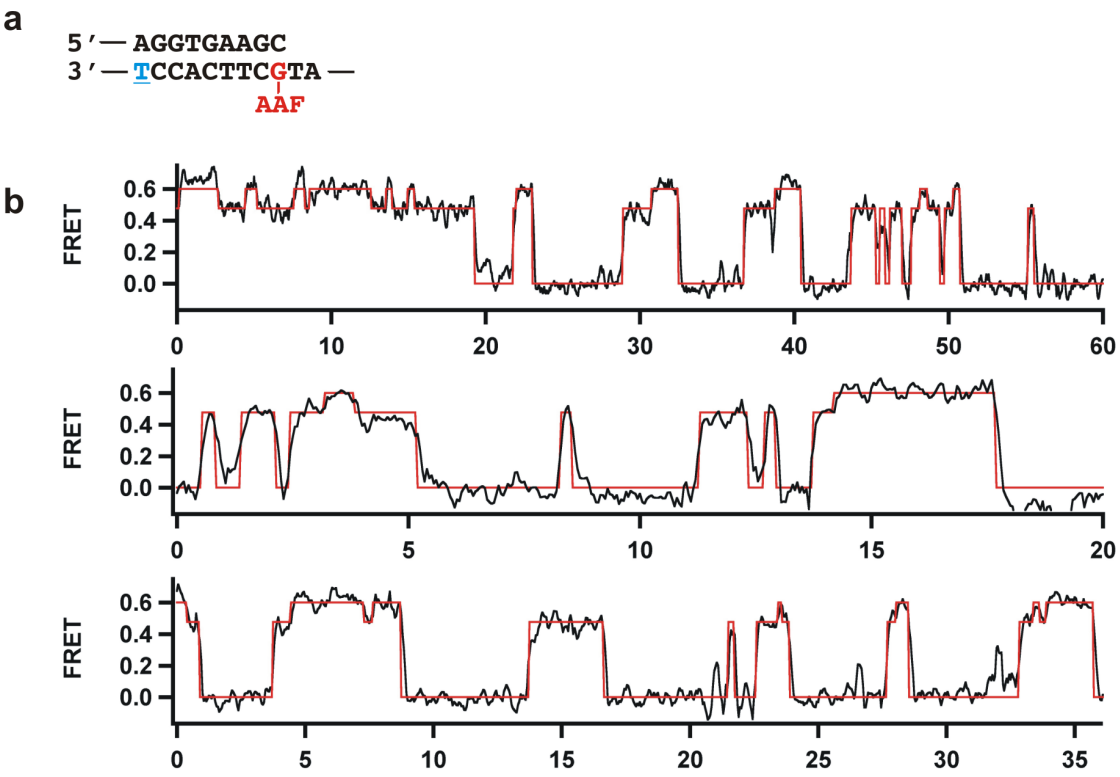

Figure 34. KF transitions between intermediate and exo states. (**a**) AAF-modified primer-template sequence. The Cy3 is conjugated to the underlined, blue thymine in the template by an amine linker. The AAF adduct is attached to the red guanine. (**b**) Three representative FRET trajectories (black) calculated from the donor and acceptor intensities (FRET = $I_A/(I_A+I_D)$). Red lines in the traces are the FRET trajectories calculated from HMM (see Methods for details).

To distinguish between real intermediate state pausing and artefacts in the HMM analysis caused by signal averaging and exposure time integration, we carried out a series of simulations (**Figure 35**). Shown in **Figure 32h** is a comparison between the experimental results and a simulation for the AAF-modified primer-template in which binding to the exo site occurs through an obligatory intermediate. The fact that these simulations so closely model the experimental data supports that binding of the primer terminus to the intermediate site is an obligatory step that precedes binding to the exo

site. However, for polymerase dissociation, the model best fits the experimental data when 60% of dissociations occur from the exo site and 40% from the intermediate site (**Figure 32i**). Sixty percent direct dissociation is likely an upper limit due to acceptor blinking and photobleaching.

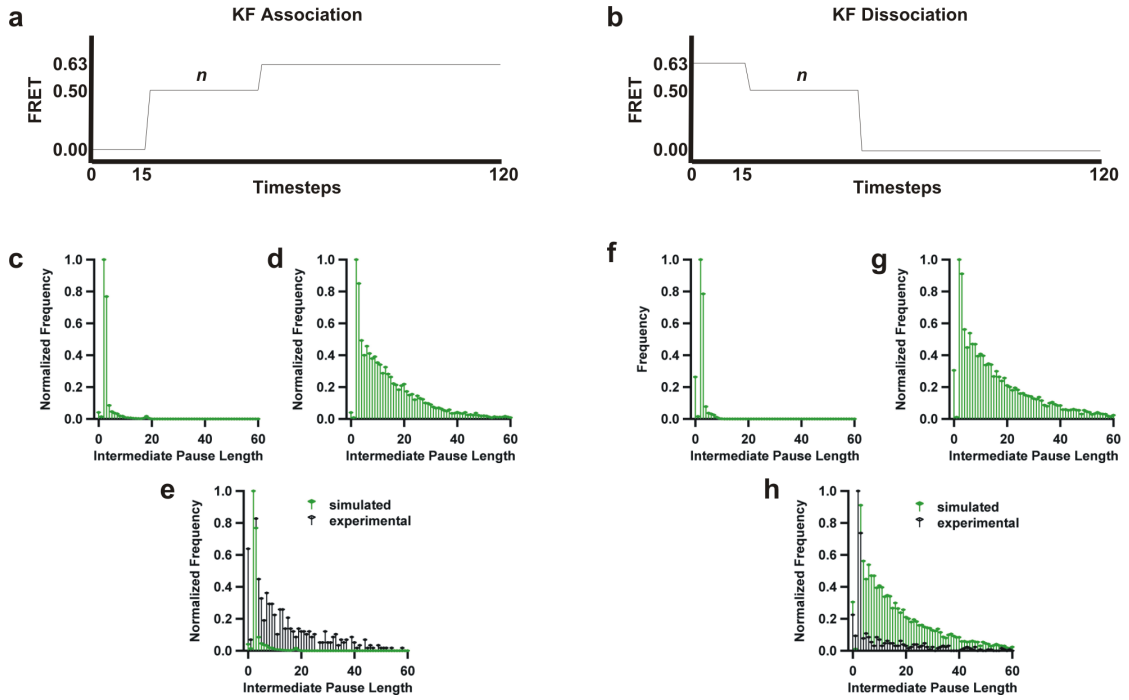

Figure 35. Simulating intermediate site pause length for KF association to the exo site and dissociation from the exo site. (**a** and **b**) Steps used to simulate KF association to exo site and dissociation from exo site. *n* is the number of intermediate state timesteps simulated. (**c** and **f**) Simulated results from 10,000 traces for direct association to exo site or dissociation from exo site ($n = 0$). (**d** and **g**) Simulated results from 10,000 traces for association to the exo site through an obligatory intermediate or dissociation from the exo site with an obligatory intermediate. (**e**) Simulation of KF direct dissociation results from (**c**) plotted with the experimental association results. (**f**) Simulation of KF dissociation through an obligatory intermediate from (**g**) plotted with the experimental association results. See Chapter 2 Methods for details.

These data suggest a model for the stalling of KF when either an AF or AAF-dG is positioned across from the primer terminus. In both cases the adduct induces the polymerase to bind away from the pol site, either into the intermediate or exo site, leading to a cessation of DNA synthesis. At that point the models diverge for the AAF and AF adduct. The AAF-dG adduct structure apparently does not allow synthesis to continue, while for AF synthesis eventually continues after a delay that leads to a pause in the extension reaction (**Figure 27b**). The question then becomes whether extension for the AF-modified template is allowed to continue from the intermediate state or if the presence of a dNTP causes a transition of the primer terminus to the pol site. These two scenarios can be distinguished by carrying out single-molecule experiments in the presence of the next correct dNTP.

5.2.6. Correct dNTP rescues pol site binding for AF, but not AAF

To carry out experiments in the presence of dNTPs, dideoxy-terminated primers must be used to prevent extension. Dideoxy-termination of the primers did not significantly change the observed PIFE or FRET values for the binary complex for either the unmodified or adduct-containing templates (compare **Figure 32a,b, Figure 36b,c,f,g** and **Figure 37b**), indicating that the polymerase binds in the same pol site orientation in the absence of a 3'-OH on the primer terminus.

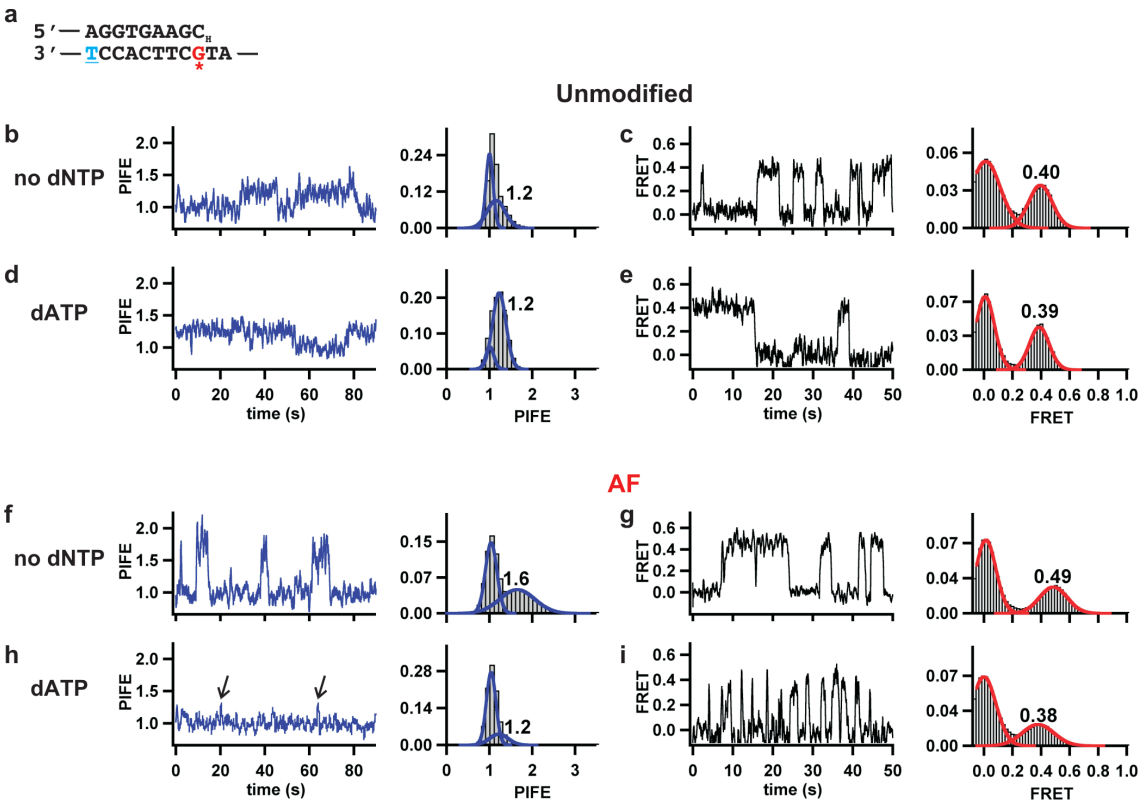

Figure 36. Correct dNTP rescues pol site binding for AF. (**a**) Dideoxy-terminated primer-template sequence used to observe polymerase binding to the DNA with the nucleotides present in solution. The lack of an 3'-OH in the primer prevents nucleotide incorporation. The Cy3 is conjugated to the underlined, blue thymine in the template by an amine linker. For the AF experiments, the adduct is attached to the red, asterisks guanine. (**b-i**) An example PIFE or FRET trace and the PIFE or FRET histogram for DNA polymerase binding to (**b-e**) unmodified or (**f-i**) AF-modified DNA in the absence or presence of dATP, as indicated. The PIFE and FRET states do not change upon addition of the correct dNTP (dATP) for unmodified DNA; however, for AF-modified DNA the PIFE and the FRET decrease upon addition of the correct dNTP.

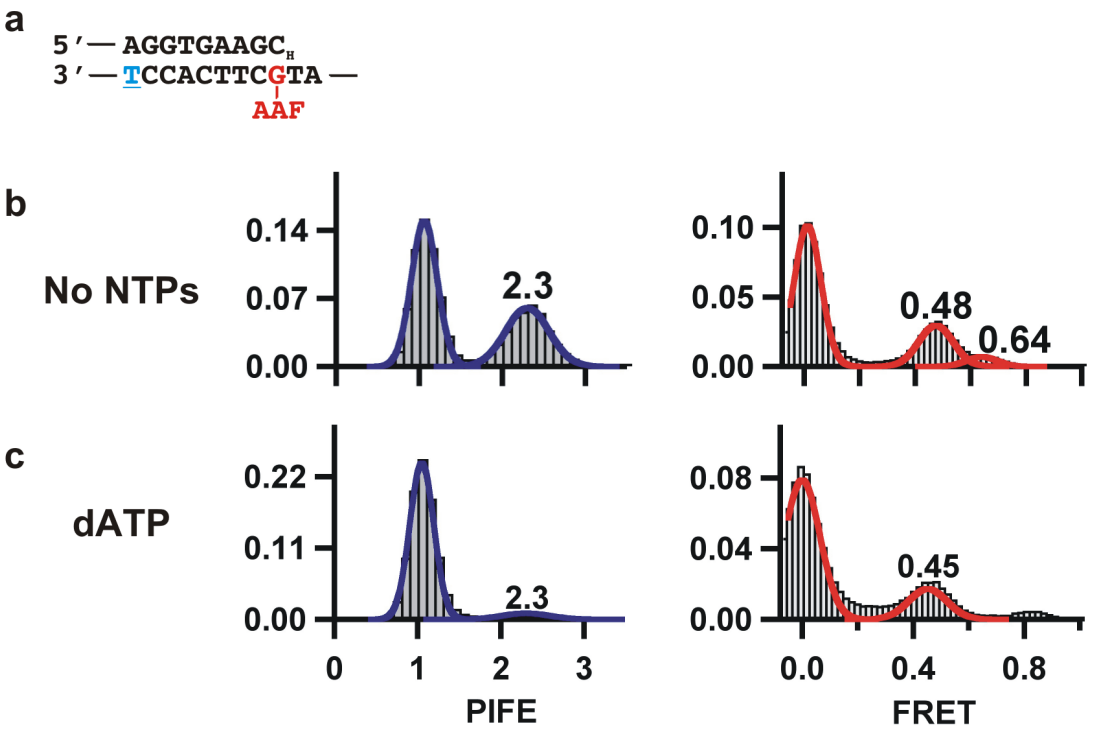

Figure 37. AAF adduct prevents proper alignment. (**a**) Dideoxy-terminated primer-template sequence used to observe polymerase binding to the DNA with the nucleotides present in solution. The lack of the 3' OH at the primer terminus prevents nucleotide incorporation. The Cy3 is conjugated to the underlined, blue thymine in the template by an amine linker. The AAF adduct is attached to the red, guanine. (**b** and **c**) PIFE and FRET histograms for polymerase binding to the AAF-modified primer-template in the presence or absence of the correct dATP, as indicated. The 0.64 FRET state observed without NTPs is not present with dATP in solution. Additionally, a small ~0.8 FRET state is observed in the presence of dATP.

As we previously reported(52), addition of the next correct nucleotide (dATP in this case) does not change the PIFE or FRET values for an unmodified primer-template, indicating that the nucleotide does not change the position or orientation of the primer-template in the polymerase active site (**Figure 36b**). However, with the template containing an AF-dG adduct across from the primer terminus, both PIFE and FRET values obtained were essentially identical to those obtained using the unmodified

template (**Figure 36**). This change was not observed with the analogous AAF-dG-modified template (**Figure 37c**). These results suggest that for the template containing an AF-dG adduct, the presence of the next correct nucleotide in the active site is capable of rescuing pol-site binding by displacing the primer from the intermediate site.

The addition of an incorrectly paired nucleotide did not cause the same change for the AF-modified template as was observed when the next correct nucleotide was added (**Figure 38b**). This suggests that it is the correct base pairing that triggers the conversion of the intermediate site binding to the pol site binding. Further support for this is provided by the observation that the presence of the next correct ribonucleotide (rATP) resulted in the same change as was observed for dATP (cf. **Figure 36h** with **Figure 38c**). It is interesting that KF cannot incorporate rATP in this template (**Figure 39**) suggesting that the polymerase is discriminating against the ribose sugar at a subsequent step. Prior studies have shown that when an rNTP is in the active site, the closing of the fingers domain is sterically blocked by the interaction of the 2'-OH with the glutamate at position 710(161).

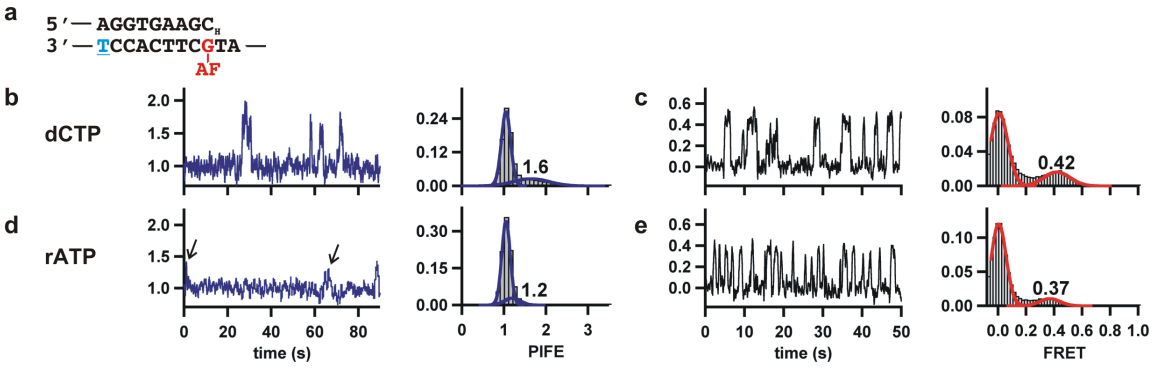

Figure 38. Complementary base rescues pol site binding for AF. (**a**) Dideoxy-terminated primer-template sequence used to observe polymerase binding to the DNA with the nucleotides present in solution. The lack of the 3' OH at the primer-terminus prevents nucleotide incorporation. The Cy3 is conjugated to the underlined, blue thymine in the template by an amine linker. The AF adduct is attached to the red, guanine. (**b-e**) Examples of PIFE or FRET traces and the PIFE or FRET histograms for DNA polymerase binding to AF-modified DNA in the presence of dCTP or rATP, as indicated. Interestingly, the PIFE and FRET for polymerase binding to the AF-modified primer-template in the presence of rATP decrease to values that resemble the polymerase binding to the unmodified primer-template in the presence of the correct dATP.

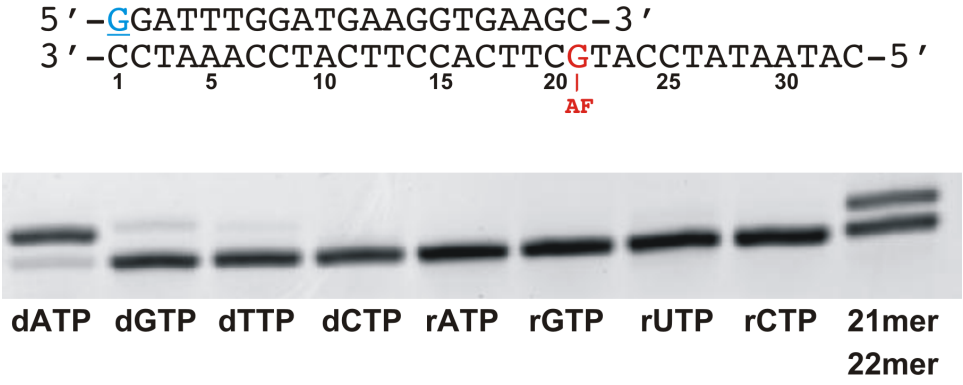

Figure 39. KF only incorporates the correct nucleotide on an AF-modified primer. Single nucleotide incorporation reactions were carried out under identical conditions to the extension reactions (methods) except that only one nucleotide was added to the reaction, as indicated. Reactions were stopped after 60 min and run on the gel.

Finally, we have previously shown that the correct nucleotide decelerates ternary complex dissociation by ~10-fold compared to the binary complex alone(52,155,176). In the presence of the AF-adduct, the observed dissociation constant is $k_{off} = 2.5 \pm 0.1\ s^{-1}$, ~150-fold faster than the unmodified ternary complex(52). Thus, the AF adduct destabilizes the ternary complex by ~3.0 ± 0.3 kcal mol^{-1}, which provides an interesting explanation for why the polymerase requires over 30 minutes to complete elongation across AF (**Figure 27b** and **Figures 28** and **29**).

5.3. Discussion

Gaining an understanding of how a DNA polymerase interacts with DNA adducts formed by chemical carcinogens is an important goal because these interactions are the basis for the mutagenic effects of these DNA lesions. We have used two complementary single-molecule methods, smFRET and smPIFE, to characterize these interactions for two related arylamine DNA adducts that we have positioned in a primer-template construct at either the templating position or across from the primer terminus. The two carcinogens employed here represent two interesting case scenarios; one stalls polymerization completely (AAF), while the other slows but does not block synthesis (AF).

Prior studies have shown that the DNA-distorting AAF-dG adduct inhibits DNA synthesis when positioned at the templating position by interfering with the closing of the fingers domain of either KF or T7 DNA polymerase(118,176), two related high-fidelity polymerases. AF-dG does not prevent the formation of the closed complex for KF(176)

and presumably this represents one reason why this adduct can be bypassed by most polymerases. Much less is known about the structures that form during bypass of an AF-dG adduct, structures that presumably lead to mutations. Also unknown is the structural basis for stalling before and past the adduct location, nor how these effects compare with the AAF-dG adduct, which cannot be bypassed by KF. Even though the AAF moiety has been shown to intercalate within the fingers domain of T7 DNA polymerase when this adduct is located at the templating position(118), our data show that KF binds in the functional pol site orientation when either an AAF or AF-dG adduct is located at this position.

However, when the primer is extended one position so that either adduct is across from the primer terminus, structures are formed that are different from that observed with unmodified DNA. For the AAF-dG case, two binding orientations are observed, one of which we identify as exo site binding and a second that has a FRET value in between that observed for pol and exo site binding. For the AF-dG adduct, only this latter intermediate binding site orientation is observed. Our data is consistent with a model in which these adducts cause the polymerase to follow one of two distinct pathways depending on the adduct structure. In the presence of a non-distorting adduct that stalls the polymerase but eventually can be bypassed, such as AF-dG, the polymerase binds the DNA at this newly-identified, non-catalytic intermediate site. Binding of a nucleotide that can correctly base pair with the templating base eventually rescues the polymerase from this state, but polymerization occurs slowly because the adducted ternary complex is destabilized ~3.0 kcal mol^{-1} compared to unmodified DNA, and therefore multiple binding and dissociation

cycles likely occur prior to phosphodiester bond formation.

In the presence of an adduct that completely stalls DNA synthesis, such as AAF-dG, our single-molecule trajectories show that following incorporation across from the adduct by a lesion bypass polymerase, the polymerase binds this structure first in the intermediate site orientation and then the DNA is shuttled to the exo site without KF dissociation. In these studies, we used an exonuclease inactive mutant, but in a wild type protein, it is likely that the polymerase would then cleave bases from the primer 3'-end. Addition of the correct nucleotide did not rescue pol site binding, consistent with the fact that KF and other high-fidelity polymerases cannot extend primer-templates where the primer terminates across from the AAF-dG adduct. Presumably, it is the structural differences between the AAF and AF adducts and the ability of the AF-dG adduct to form a dG:dC base pair that allows the polymerase to follow the pathway that leads to extension of the primer and adduct bypass(42-44).

Future studies will test whether this mechanism applies generally to other chemical modifications and to other DNA polymerases, including those specialized in trans-lesion synthesis. We also plan to characterize further the intermediate binding state, which may be an intermediate in the process by which an incorrectly paired nucleotide is removed by the proofreading exonuclease activity.

Chapter 6: optimization of single molecule dna synthesis assay

6.1. Introduction

DNA polymerases replicate DNA with remarkable fidelity in order to preserve the integrity of the DNA from one generation to the next. Accordingly, some DNA polymerases, including the Klenow fragment (KF) of *E. coli* DNA polymerase I, are able to catalyze template-directed DNA synthesis with error frequencies below 1 every 10^5 bases(17). Polymerases ensure proper nucleotide selection by a series of checkpoints prior to and following nucleotide incorporation(149,181,182).

Prior to incorporation, the polymerase checks incoming nucleotides for complementarity to the templating base from the pol site(71,183). For nearly all DNA polymerases, the pol site consists of three structural domains that are shaped like a right hand: fingers, thumb, and palm domains(8,10). If complementarity is met, the polymerase will then close its fingers domain to properly align the incoming nucleotide with the 3'-OH terminus of the primer stand(12). During this step, the DNA polymerase selects for the correct sugar. A "steric gate" that sterically prevents 2'-OH sugars from aligning properly with the primer terminus screens against ribose sugars(161). If the incoming nucleotide is complementary to the templating base and it contains a deoxyribose sugar, then phosphodiester bond formation can occur(12).

Many high fidelity DNA polymerases also have 3'-5' exonuclease activity in addition to 5'-3' polymerization activity. Following a misincorporation, the DNA polymerase can transfer the primer strand to the exonuclease site (exo site) located ~35 Å from the pol site (for KF)(23), cleave off the incorrect terminal nucleotide, and reanneal the primer with the template to continue DNA synthesis without the mismatch in the DNA. This proofreading activity increases the overall fidelity of the polymerase by ~7-fold(17).

Unnecessary excision of correct nucleotides would slow down DNA synthesis and waste energy(71). Therefore, it stands to reason that a DNA polymerase should have a post-incorporation checkpoint that checks the complementarity of the DNA. This could be kinetic checkpoint, where the rate of proofreading is faster than the rate of incorporation of a nucleotide to a mismatched terminus, but substantially slower than the rate of incorporation to a complementary primer-template terminus(71,81). Additionally, there could be a physical fidelity-checking site on the DNA polymerase that checks for complementarity following each incorporation(50).

A recent single molecule FRET study found direct evidence for a fidelity-checking site(50). In this work, single incorporation events were tracked with base pair resolution. Approximately 66% of the incorporations were immediately followed by a sharp dip in the FRET trajectory, possibly due to the primer-template terminus being shuttled to the fidelity-checking site.

Several concerns remain with regard to the fidelity-checking site observed in the single molecule study. First, why were transitions to the site slower than the *in vitro* rate of incorporation? KF can incorporate 7-50 nucleotides per second *in vitro*(71,113), yet

the polymerase dwelled in the fidelity-checking site for ~ 200 ms. This may be due to the polymerase being less active on the surface immobilized DNA used for single molecule acquisition or to slower incorporation of the initial nucleotides. Second, could the dip in the FRET be from fast polymerase dissociation and reassociation? Ensemble studies have shown KF's processivity is ~250 nucleotides(71). However, the polymerase concentration used for the single molecule experiments was high enough so that a second polymerase could rebind within the acquisition time of the instrument. Perhaps, under single molecule conditions, the polymerase is nonprocessive. After averaging, this would look like a dip in FRET that could be misinterpreted as shuttling the DNA to another site on the polymerase. Finally, evidence shown for the incorporations were selected traces from the experiments. While the traces supported the fidelity-checking hypothesis, synchronization of many incorporation events to a single graph would more convincingly display the fidelity-checking events.

To thoroughly address these issues, we need to optimized the conditions used to observe single incorporations and improved the analysis method. We detected single incorporation events by FRET and confirmed the incorporations by PIFE. We decreased the dNTP concentration used for the experiments from 100 μM by at least 10,000-fold to slow down the incorporation rate and simplify the identification of single incorporations. We improved the analysis of the incorporations by writing a Matlab code to apply a hidden Markov model (HMM) to the traces and to synchronize the moment of incorporation for each of the molecules. In agreement with the previous single molecule study, our synchronized plots also revealed a clear dip in FRET following an

incorporation. We attribute this dip to fidelity checking. Our improved assay and analysis will be instrumental for future characterization of the fidelity-checking activity for KF, as well as other DNA polymerases.

6.2. Results

6.2.1. Real-time, single molecule tracking of DNA synthesis

Direct excitation of unbound, surface immobilized Cy3-labeled DNA with a 532 nm laser gives a high Cy3 fluorescence signal. Upon Cy5-labeled KF binding to the primer-template, energy is transferred from the Cy3 to the Cy5 with an anticorrelated drop in Cy3 fluorescence intensity along with a concurrent increase in Cy5 intensity. The corresponding FRET, calculated as the ratio of the acceptor intensity to the sum of the donor and acceptor intensities is zero for unbound DNA and greater than zero when KF binds to the DNA.

FRET is a distance dependent phenomenon and DNA polymerases bind stably at the primer-template junction (3'-OH end of the primer). Consequently by keeping the position of the Cy3 on the template constant while incrementally increasing the primer length, we should see distinct drops in the FRET for each increase in primer length due to the polymerase binding farther from the Cy3. In fact, this is exactly the trend observed between the 19mer and 22mer positions. KF binding to the shortest primer-template, 19mer (**Figure 40b**), in the absence of dNTPs gave a FRET distribution at 0.66 that did not substantially change over time (**Figure 40c**). FRET for the longer templates also remained constant over time at 0.58, 0.38 and 0.34 for the 20mer, 21mer and 22mer

primer-templates, respectively (**Figure 40a**). In good agreement with our previously published results, these FRET values differ enough for us to visualize the polymerase position on the DNA with single nucleotide resolution.

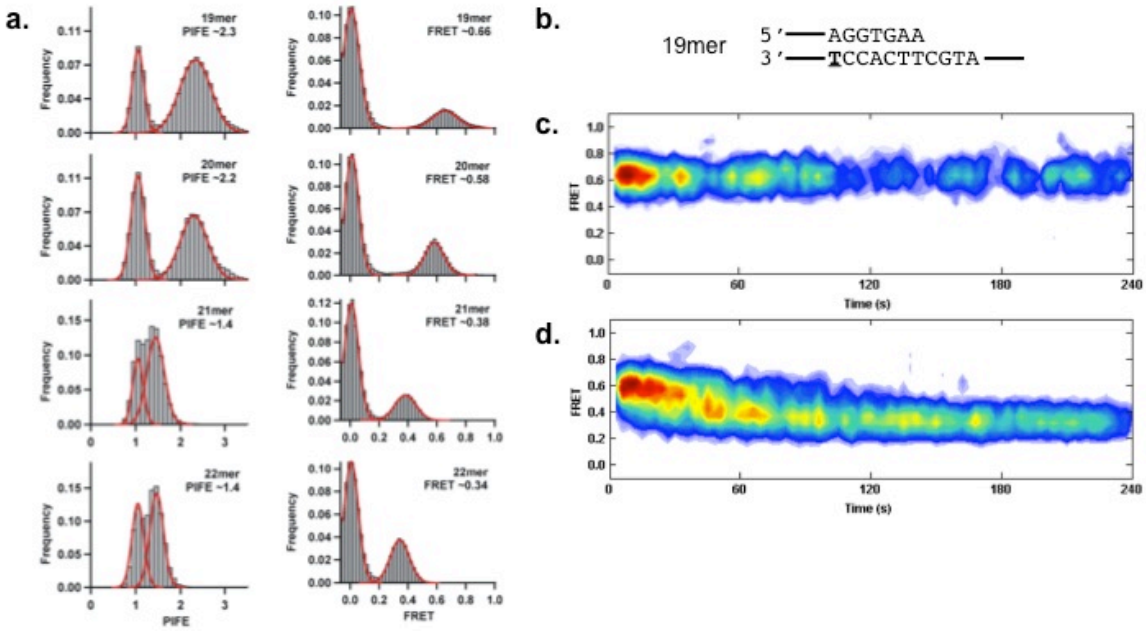

Figure 40. Monitoring DNA synthesis at the single molecule level. (**a**) Single molecule PIFE (left) and FRET (right) histograms for KF binding to 19, 20, 21, and 22mer primer-templates. (**b**) 19mer primer template used for binding and incorporation experiments shown in **c** and **d**. The bold, underlined T is the Cy3-labeled, amino-modified thymine. (**c**) FRET from all KF binding events to the 19mer primer-template shown over time with no dNTPs in the solution. (**d**) FRET from all KF binding events to the 19mer primer-template shown over time with 5 nM dGTP, dCTP, and dATP in the solution. FRET decreases over time as the polymerase synthesizes new DNA and moves farther away from the Cy3. The number of binding events observed over time decreases due to Cy3 photobleaching, which prevents further observation of binding events.

Therefore, we predicted that by adding the next three correct nucleotides (dCTP, dGTP, and dATP) to the binding buffer with the 19mer primer-template (**Figure 40b**), we should see evidence for the polymerase synthesizing DNA over time by observing a decrease in FRET. We observed a clear drop in FRET over time, by synchronizing over one hundred traces (**Figure 40d**). This demonstrates KF's polymerization activity was not hindered by the surface immobilization of the DNA on the slide.

6.2.2. 20mer to 21mer incorporation clearly distinguishable by smPIFE and smFRET

Our ability to track the polymerase position with single nucleotide resolution enabled us to also observe individual incorporations of nucleotides to the primer-template. For instance, a single incorporation to a 20mer primer-template should result in a discreet change in FRET from ~0.58 (20mer) to ~0.38 (21mer). The 20mer to 21mer incorporation occurred in the trajectory shown in **Figure 41** at ~96 s.

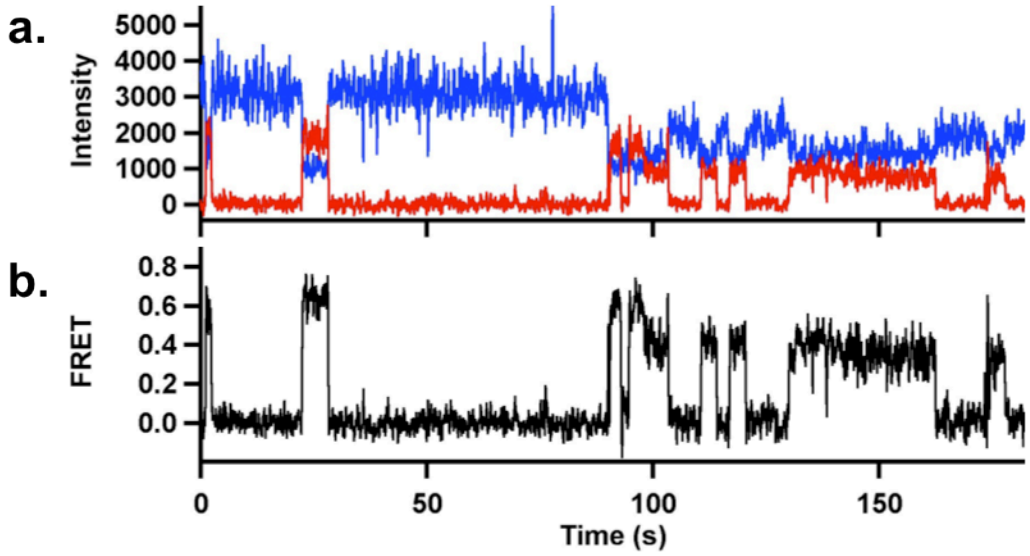

Figure 41. Moment of incorporation from 20mer to 21mer clearly apparent in individual single molecule trajectories. (**a**) Representative Cy3 donor (blue) and Cy5 acceptor (red) time trajectories for KF synthesizing new DNA on the 19mer primer-template shown in **Figure 33b**. (**b**) Corresponding FRET trajectory for the intensity trace shown in **a**. FRET was calculated as the ratio of the acceptor intensity to the sum of the donor and acceptor intensities. Incorporation to a 20mer primer-template and translocation to the 21mer position is detected by the change in Cy3 intensity (blue, ~3,300 before and ~1,800 after) and a change in FRET (~0.58 before and ~0.38 after).

In addition to FRET, we also use smPIFE to monitor the polymerase position on the DNA. PIFE requires a close interaction between KF and Cy3; therefore, when the Cy3 is beyond the reach of the polymerase footprint, PIFE abruptly drops. Accordingly, the 19mer and 20mer primer-templates had high fluorescence enhancement (~2.2, **Figure 40a**), while the 21mer and 22mer primer-templates, which are beyond the reach of KF's footprint had low PIFE (~1.4, **Figure 40a**). These PIFE values are in good agreement with our previously published results(52).

While PIFE does not provide the same spatial resolution as FRET, the abrupt drop that occurs following the translocation from 20mer (PIFE ~2.2, **Figure 40a**) to 21mer

(PIFE ~1.4) should be apparent in the single molecule trace. The fluorescence enhancement occurs whenever unlabeled KF or KF with and inactive Cy5 binds to the DNA. Because we use a high concentration of polymerase, we predominantly see polymerase-bound DNA, thus we either have PIFE or FRET occurring throughout the duration of most of the trajectories. Following the 20mer to 21mer incorporation, the donor intensity is decreased when FRET is not visible due to the decrease in PIFE (**Figure 41a**, high donor intensity before 96 s, low donor intensity after 96 s). The addition of PIFE to our single molecule incorporation assay strengthens our interpretation of the polymerase position on the DNA. While we also observed many incorporations to the 19mer and 21mer primer-templates by FRET, these incorporations are not accompanied by a change in PIFE (**Figure 40a**, 19mer and 20mer have same PIFE, and 21mer and 22mer have same PIFE). We will focus on the 20mer to 21mer incorporation for further analysis because we can easily detect this transition both fluorescence phenomena: PIFE and FRET.

6.2.3. KF conducts fidelity-checking following incorporation of a dNTP

We applied a hidden Markov model to the traces with visible changes in PIFE and FRET for the incorporation from 20mer to 21mer. As mentioned above, the 20mer to 21mer incorporation should result in a change in FRET from ~0.58 to ~0.38. The incorporation moments were determined with a HMM and synchronized to time zero. Hence, the time before the incorporation was negative and the time after the incorporation was positive.

The synchronized trajectory for the incorporation revealed a clear change in the FRET from ~0.58 to ~0.38, as expected (**Figure 42**). Intriguingly, immediately following the incorporation, the FRET dropped to ~0.33 before returning to ~0.38 (**Figure 42**). Considering the FRET for the 22mer position is ~0.32, the drop in FRET to ~0.33 suggests KF transiently shuttled the primer-template terminus to a position that is approximately one base pair distance (~3.4 Å) away from the core of the pol site. The duration of the displacement of the primer-template away from the core of the pol site was ~200 ms. This shuttling behavior is in good agreement with the previously observed dips in FRET detected by Christian, *et al*(50). Considering the transient dip occurs between the polymerase translocation from 20mer to 21mer, we also attribute the dip in FRET to fidelity checking.

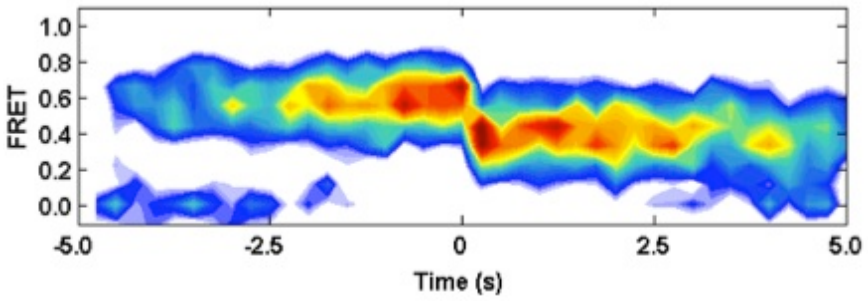

Figure 42. Dip in FRET corresponding to fidelity-checking is apparent following 20mer incorporation. Traces (n =21) were synchronized so that the time before the incorporation is negative and the time after the incorporation is positive. See text for details.

6.3. Discussion

DNA polymerases must replicate DNA with high fidelity in order to maintain genome integrity. The additional proofreading activity many polymerases possess increases the overall fidelity of these polymerases by excising misincorporated bases from the exo site. We have developed an assay that can track the polymerase position on the DNA with single base pair resolution and can distinguish between pol and exo site binding in real-time(50,52). With this single molecule fluorescence approach we monitored individual polymerases as they synthesized new DNA.

Our fidelity-checking results are in good agreement with the work by Christian, *et al*(50). We improved the previous assay by decreasing the dNTP concentration and by synchronizing incorporation events from many molecules together to characterize the distinctive FRET dips observed following incorporations.

Our work suggests the moderately high fidelity polymerase, KF, assesses the complementarity of the newly formed base pair by transferring the primer-template terminus to a putative fidelity-checking site. Presumably, following a misincorporation, the polymerase would then transfer the primer strand from the fidelity-checking site to the 3'→5' exonuclease site. We found that when the polymerase incorporates the correct dNTP, it will return the primer template terminus to the pol site after checking complementarity at the fidelity-checking site.

Graham, *et al.* identified many potential amino acid residues that may be involved in fidelity-checking with advanced computational analysis(82). Several conserved residues in our putative fidelity-checking region displayed altered interactions with three mispairs

compared to complementary DNA. A thorough site-directed mutagenesis study with these amino acids, in combination with a fidelity assay and our single molecule assay, should be able to test whether or not these residues are involved in fidelity-checking.

The dwell time of the primer-template terminus at the putative fidelity-checking site was ~200 ms. If the checking occurs after each incorporation, then the rate of checking is too slow to account for KF's *in vitro* synthesis rate of ~7-50 nucleotides per second(71,113). One possibility is that an incoming dNTP helps shuttle the DNA from the fidelity-checking site to the pol site. In our experiments we used very low concentrations of dNTPs to slow down the polymerization rate, but future experiments could increase the dNTP concentration to see if the dwell time at the fidelity-checking site decreases as a result of dNTP-enhanced shuttling of the DNA back to the pol site. A second possibility is that the incorporation of the first few nucleotides is slower than the incorporation of later nucleotides. Our assay could also test whether or not this is the case by starting with shorter primers and high dNTP concentration, and measuring whether or not the fidelity-checking dwell time changes after several incorporations.

The single molecule assay and analysis optimization presented here immediately advances our ability to characterize the newly identified fidelity-checking site. Further experiments with other DNA polymerases will determine whether or not the fidelity-checking mechanism is unique to KF, or if it is a general DNA polymerase mechanism. Bacillus stearothermophilus fragment (BF), another A family polymerase and a close homolog to KF, shares many residues with KF at the fidelity-checking site(82); however, BF lacks a 3'→5' exonuclease site(179). If the fidelity-checking activity is related to

proofreading, it stands to reason that BF would not need this active site. Perhaps the site is left over from an evolutionary common ancestor, or perhaps the fidelity checking works by another mechanism for BF. For example, binding a mismatched pair at the fidelity-checking site may destabilize BF association with the DNA. BF dissociation followed by association of another 3'→5' exonuclease-containing DNA polymerase to excise the misincorporated nucleotide would be a novel approach (to our knowledge) for mismatch repair.

Conclussions and future directions

We optimized and applied the single molecule fluorescence assay originally developed by Christian, *et al*.(50) to provide new insights into DNA polymerase synthesis and proofreading. We have optimized nearly every aspect of the assay. First, we determined the optimal conditions for labeling single cysteine proteins. We learned through these experiments that slightly altering the labeling conditions could lead to drastic misinterpretation of fluorescence data. Second, we added single molecule PIFE to our arsenal of single molecule techniques. smPIFE was a tremendous asset to our research. With smPIFE we were able to reliably measure binding kinetics, strengthen our interpretation of the polymerase position on the DNA, and convincingly identify a novel intermediate state for KF binding. Finally we upgraded our analysis approaches to include HMM analysis for determining the most probabilistic path in noisy single molecule traces, simulations for confirming models for polymerase kinetics, and synchronization analysis for revealing the fidelity-checking state following each incorporation.

Application of the optimized single molecule fluorescence approaches permitted the evaluation of many steps involved in DNA polymerase-catalyzed DNA synthesis and proofreading. Through simple binding assessment we found that monovalent ions, incorrect nucleotides, mismatched primer-template termini, and carcinogenic DNA adducts at duplex DNA termini destabilize the KF-DNA binary complex. On the other hand, the correct dNTP and aminofluorene adducts attached to the templating base

stabilized the KF-DNA complex. Our assay distinguished between pol and exo site binding, which allowed us to determine how mismatches and carcinogenic adducts induce proofreading. Moreover, by using PIFE and FRET in tandem, we were able to identify a new intermediate state in the proofreading mechanism. Finally, the capability to track the DNA polymerase position on the DNA with single-nucleotide resolution allowed us to monitor multiple incorporation cycles from individual DNA polymerases in real time. These incorporation experiments support the hypothesis that a fidelity-checking step occurs following each phosphodiester bond formation.

As single molecule experiments provide a unique and fresh perspective to nearly any biological system, the number potential avenues to explore in the future seems limitless. The future directions I present here are ones that I believe are relatively easy to conduct (now that the single molecule assays are developed) and are aligned with the laboratory's current funding sources.

Ultimately, our lab works toward understanding the fundamental mechanism(s) of mutagenesis. We have already characterized KF interactions with AF/AAF on a random DNA sequence. Now, there are three primary avenues to take the research while using the same single molecule techniques: 1) change the polymerase, 2) change the adduct, and 3) change the DNA sequence.

E. coli DNA polymerase I is a model DNA polymerase; however, it is not known for conducting trans-lesion DNA synthesis. I believe studying eukaryotic DNA polymerases, especially human DNA polymerases, could be of broad interest to the chemical carcinogenesis and DNA polymerase fields. Alfonso's current research with Dpo4, albeit

Dpo4 is not a human DNA polymerase, is also a good route because Dpo4 is a hot topic in the DNA polymerase field and it is known for carrying out trans-lesion DNA synthesis.

AF and AAF are model carcinogens, but they are not naturally occurring. Natural progression of this research is to change the carcinogenic adduct to one that has been shown to cause a human cancer. Pramodha is already synthesizing an excellent example of such a carcinogen modification, Beno-α-pyrene (BaP), in the lab. BaP, a common, naturally occurring carcinogen, has been directly linked to causing mutations in the p53 gene in lung cancer tumors(184,185). p53 is a tumor suppressor protein and it is of considerable interest in cancer, molecular biology, physiology, and clinical biology fields. Additionally, our assay could also be easily adapted to research the interactions between DNA polymerases and other naturally occurring modified bases such as 8-oxo-guanine and 5-methylcytosine.

DNA mutations, even those induced by carcinogenic adducts, are known to occur in "mutational hotspots." That is, certain sequences of DNA are more prone to mutations than others. BaP, for example, is known to induce mutations in 3 different mutational hotspots in the p53 gene(184). What is unique about these sequences? Comparison of polymerase binding to a random DNA sequence, to a mutational hotspot sequence without a carcinogenic adduct, and a mutational hotspot sequence with a carcinogenic adduct may reveal the mechanism by which these hotspots induce mutagenesis.

Misincorporation of an incorrect base by KF is a rare occurrence (about once every 10,000 bases). Therefore it is currently extremely difficult for us to detect a misincorporation in real time. However, by using a carcinogenic adduct in a mutational

hotspot the odds of a mutation occurring would increase dramatically. This may be the best approach to solve an incredibly important and persistent enigma: How does mutagenesis occur? Here is the ultimate, high impact experiment: monitoring, in real-time, trans-lesion DNA synthesis past BaP by a human DNA polymerase in a p53 mutational hotspot. While the technology currently exists to conduct this experiment, there are many technical hurdles to overcome before the experiment would be possible. Good luck!

REFERENCES

1. Watson, J.D. and Crick, F.H. (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, **171**, 737-738.
2. Pauling, L. and Corey, R.B. (1953) Structure of the nucleic acids. *Nature*, **171**, 346.
3. Pauling, L. and Corey, R.B. (1953) A Proposed Structure For The Nucleic Acids. *Proc Natl Acad Sci U S A*, **39**, 84-97.
4. Watson, J.D. and Crick, F.H. (1953) Genetical implications of the structure of deoxyribonucleic acid. *Nature*, **171**, 964-967.
5. Kornberg, A., Lehman, I.R. and Simms, E.S. (1956) Polydesoxyribonucleotide synthesis by enzymes from Escherichia coli. *Federation Proceedings*, **15**, 291-292.
6. Brutlag, D., Atkinson, M.R., Setlow, P. and Kornberg, A. (1969) An active fragment of DNA polymerase produced by proteolytic cleavage. *Biochemical and biophysical research communications*, **37**, 982-989.
7. Klenow, H. and Henningsen, I. (1970) Selective elimination of the exonuclease activity of the deoxyribonucleic acid polymerase from Escherichia coli B by limited proteolysis. *Proc Natl Acad Sci U S A*, **65**, 168-175.
8. Ollis, D.L., Brick, P., Hamlin, R., Xuong, N.G. and Steitz, T.A. (1985) Structure of large fragment of Escherichia coli DNA polymerase I complexed with dTMP. *Nature*, **313**, 762-766.

9. Minnick, D.T., Astatke, M., Joyce, C.M. and Kunkel, T.A. (1996) A thumb subdomain mutant of the large fragment of Escherichia coli DNA polymerase I with reduced DNA binding affinity, processivity, and frameshift fidelity. *J Biol Chem*, **271**, 24954-24961.

10. Steitz, T.A. (1999) DNA polymerases: structural diversity and common mechanisms. *J Biol Chem*, **274**, 17395-17398.

11. Astatke, M., Grindley, N.D. and Joyce, C.M. (1995) Deoxynucleoside triphosphate and pyrophosphate binding sites in the catalytically competent ternary complex for the polymerase reaction catalyzed by DNA polymerase I (Klenow fragment). *J Biol Chem*, **270**, 1945-1954.

12. Joyce, C.M., Potapova, O., Delucia, A.M., Huang, X., Basu, V.P. and Grindley, N.D. (2008) Fingers-closing and other rapid conformational changes in DNA polymerase I (Klenow fragment) and their role in nucleotide selectivity. *Biochemistry*, **47**, 6103-6116.

13. Doublie, S., Tabor, S., Long, A.M., Richardson, C.C. and Ellenberger, T. (1998) Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 A resolution. *Nature*, **391**, 251-258.

14. Li, Y., Korolev, S. and Waksman, G. (1998) Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of Thermus aquaticus DNA polymerase I: structural basis for nucleotide incorporation. *EMBO J*, **17**, 7514-7525.

15. Johnson, S.J., Taylor, J.S. and Beese, L.S. (2003) Processive DNA synthesis observed in a polymerase crystal suggests a mechanism for the prevention of frameshift mutations. *Proc Natl Acad Sci U S A*, **100**, 3895-3900.

16. Mathews, C.K. (1972) Biochemistry of deoxyribonucleic acid-defective amber mutants of bacteriophage T4. 3. Nucleotide pools. *J Biol Chem*, **247**, 7430-7438.

17. Bebenek, K., Joyce, C.M., Fitzgerald, M.P. and Kunkel, T.A. (1990) The fidelity of DNA synthesis catalyzed by derivatives of Escherichia coli DNA polymerase I. *J Biol Chem*, **265**, 13878-13887.

18. Berezhna, S.Y., Gill, J.P., Lamichhane, R. and Millar, D.P. (2012) Single-molecule Forster resonance energy transfer reveals an innate fidelity checkpoint in DNA polymerase I. *Journal of the American Chemical Society*, **134**, 11261-11268.

19. Hohlbein, J., Aigrain, L., Craggs, T.D., Bermek, O., Potapova, O., Shoolizadeh, P., Grindley, N.D., Joyce, C.M. and Kapanidis, A.N. (2013) Conformational landscapes of DNA polymerase I and mutator derivatives establish fidelity checkpoints for nucleotide insertion. *Nature communications*, **4**, 2131.

20. Wu, E.Y. and Beese, L.S. (2011) 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*, **286**, 19758-19767.

21. Rothwell, P.J., Allen, W.J., Sisamakis, E., Kalinin, S., Felekyan, S., Widengren, J., Waksman, G. and Seidel, C.A. (2013) dNTP-dependent conformational

transitions in the fingers subdomain of Klentaq1 DNA polymerase: insights into the role of the "nucleotide-binding" state. *J Biol Chem*, **288**, 13575-13591.

22. Olsen, T.J., Choi, Y., Sims, P.C., Gul, O.T., Corso, B.L., Dong, C., Brown, W.A., Collins, P.G. and Weiss, G.A. (2013) Electronic measurements of single-molecule processing by DNA polymerase I (Klenow fragment). *Journal of the American Chemical Society*, **135**, 7855-7860.
23. Beese, L.S., Derbyshire, V. and Steitz, T.A. (1993) Structure of DNA polymerase I Klenow fragment bound to duplex DNA. *Science*, **260**, 352-355.
24. Carver, T.E., Jr., Hochstrasser, R.A. and Millar, D.P. (1994) Proofreading DNA: recognition of aberrant DNA termini by the Klenow fragment of DNA polymerase I. *Proc Natl Acad Sci U S A*, **91**, 10670-10674.
25. Derbyshire, V., Freemont, P.S., Sanderson, M.R., Beese, L., Friedman, J.M., Joyce, C.M. and Steitz, T.A. (1988) Genetic and crystallographic studies of the 3',5'-exonucleolytic site of DNA polymerase I. *Science*, **240**, 199-201.
26. Hanahan, D. and Weinberg, R.A. (2011) Hallmarks of cancer: the next generation. *Cell*, **144**, 646-674.
27. Anand, P., Kunnumakkara, A.B., Sundaram, C., Harikumar, K.B., Tharakan, S.T., Lai, O.S., Sung, B. and Aggarwal, B.B. (2008) Cancer is a preventable disease that requires major lifestyle changes. *Pharmaceutical research*, **25**, 2097-2116.
28. Guengerich, F.P. (2006) Interactions of carcinogen-bound DNA with individual DNA polymerases. *Chemical reviews*, **106**, 420-452.

29. Eckert, K.A. and Opresko, P.L. (1999) DNA polymerase mutagenic bypass and proofreading of endogenous DNA lesions. *Mutation research*, **424**, 221-236.

30. Singh, J., Su, L. and Snow, E.T. (1996) Replication across O6-methylguanine by human DNA polymerase beta in vitro. Insights into the futile cytotoxic repair and mutagenesis of O6-methylguanine. *J Biol Chem*, **271**, 28391-28398.

31. Efrati, E., Tocco, G., Eritja, R., Wilson, S.H. and Goodman, M.F. (1997) Abasic translesion synthesis by DNA polymerase beta violates the "A-rule". Novel types of nucleotide incorporation by human DNA polymerase beta at an abasic lesion in different sequence contexts. *J Biol Chem*, **272**, 2559-2569.

32. Michaels, M.L., Reid, T.M., King, C.M. and Romano, L.J. (1991) Accurate in vitro translesion synthesis by Escherichia coli DNA polymerase I (large fragment) on a site-specific, aminofluorene-modified oligonucleotide. *Carcinogenesis*, **12**, 1641-1646.

33. Lange, S.S., Takata, K. and Wood, R.D. (2011) DNA polymerases and cancer. *Nature reviews. Cancer*, **11**, 96-110.

34. Friedberg, E.C., Lehmann, A.R. and Fuchs, R.P. (2005) Trading places: how do DNA polymerases switch during translesion DNA synthesis? *Molecular cell*, **18**, 499-505.

35. Nohmi, T. (2006) Environmental stress and lesion-bypass DNA polymerases. *Annual review of microbiology*, **60**, 231-253.

36. Rechkoblit, O., Kolbanovskiy, A., Malinina, L., Geacintov, N.E., Broyde, S. and Patel, D.J. (2010) Mechanism of error-free and semitargeted mutagenic bypass of

an aromatic amine lesion by Y-family polymerase Dpo4. *Nat Struct Mol Biol*, **17**, 379-388.

37. (2011) 2-Acetylaminofluorene. *Report on Carcinogens, Twelfth Edition*.

38. (1998) Evidence on the carcinogenicity of 2-aminofluorene. *Reproductive and Cancer Hazard Assessment Section*.

39. Kriek, E., Miller, J.A., Juhl, U. and Miller, E.C. (1967) 8-(N-2-fluorenylacetamido)guanosine, an arylamidation reaction product of guanosine and the carcinogen N-acetoxy-N-2-fluorenylacetamide in neutral solution. *Biochemistry*, **6**, 177-182.

40. Irving, C.C. and Veazey, R.A. (1969) Persistent binding of 2-acetylaminofluorene to rat liver DNA in vivo and consideration of the mechanism of binding of N-hydroxy-2-acetylaminofluorene to rat liver nucleic acids. *Cancer research*, **29**, 1799-1804.

41. Miller, J.A. (1970) Carcinogenesis by chemicals: an overview--G. H. A. Clowes memorial lecture. *Cancer research*, **30**, 559-576.

42. Broyde, S. and Hingerty, B. (1983) Conformation of 2-aminofluorene-modified DNA. *Biopolymers*, **22**, 2423-2441.

43. Eckel, L.M. and Krugh, T.R. (1994) 2-Aminofluorene modified DNA duplex exists in two interchangeable conformations. *Nature structural biology*, **1**, 89-94.

44. O'Handley, S.F., Sanford, D.G., Xu, R., Lester, C.C., Hingerty, B.E., Broyde, S. and Krugh, T.R. (1993) Structural characterization of an N-acetyl-2-

aminofluorene (AAF) modified DNA oligomer by NMR, energy minimization, and molecular dynamics. *Biochemistry*, **32**, 2481-2497.

45. Grunberger, D., Nelson, J.H., Cantor, C.R. and Weinstein, I.B. (1970) Coding and conformational properties of oligonucleotides modified with the carcinogen N-2-acetylaminofluorene. *Proc Natl Acad Sci U S A*, **66**, 488-494.

46. Shi, J., Dertouzos, J., Gafni, A. and Steel, D. (2008) Application of single-molecule spectroscopy in studying enzyme kinetics and mechanism. *Methods Enzymol*, **450**, 129-157.

47. Zhuang, X., Kim, H., Pereira, M.J., Babcock, H.P., Walter, N.G. and Chu, S. (2002) Correlating structural dynamics and function in single ribozyme molecules. *Science*, **296**, 1473-1476.

48. Zhao, R. and Rueda, D. (2009) RNA folding dynamics by single-molecule fluorescence resonance energy transfer. *Methods*, **49**, 112-117.

49. Ha, T. (2001) Single-molecule fluorescence resonance energy transfer. *Methods*, **25**, 78-86.

50. Christian, T.D., Romano, L.J. and Rueda, D. (2009) Single-molecule measurements of synthesis by DNA polymerase with base-pair resolution. *Proc Natl Acad Sci U S A*, **106**, 21109-21114.

51. Santoso, Y., Joyce, C.M., Potapova, O., Le Reste, L., Hohlbein, J., Torella, J.P., Grindley, N.D. and Kapanidis, A.N. (2010) Conformational transitions in DNA polymerase I revealed by single-molecule FRET. *Proc Natl Acad Sci U S A*, **107**, 715-720.

52. Markiewicz, R.P., Vrtis, K.B., Rueda, D. and Romano, L.J. (2012) Single-molecule microscopy reveals new insights into nucleotide selection by DNA polymerase I. *Nucleic acids research*, **40**, 7975-7984.

53. Maxwell, B.A. and Suo, Z. (2013) Single-molecule investigation of substrate binding kinetics and protein conformational dynamics of a B-family replicative DNA polymerase. *J Biol Chem*, **288**, 11590-11600.

54. Vrtis, K.B., Markiewicz, R.P., Romano, L.J. and Rueda, D. (2013) Carcinogenic adducts induce distinct DNA polymerase binding orientations. *Nucleic acids research*.

55. Luo, G., Wang, M., Konigsberg, W.H. and Xie, X.S. (2007) Single-molecule and ensemble fluorescence assays for a functionally important conformational change in T7 DNA polymerase. *Proc Natl Acad Sci U S A*, **104**, 12610-12615.

56. Hwang, H., Kim, H. and Myong, S. (2011) Protein induced fluorescence enhancement as a single molecule assay with short distance sensitivity. *Proc Natl Acad Sci U S A*, **108**, 7414-7418.

57. Hwang, H. and Myong, S. (2013) Protein induced fluorescence enhancement (PIFE) for probing protein-nucleic acid interactions. *Chemical Society reviews*.

58. Pavani, S.R., Thompson, M.A., Biteen, J.S., Lord, S.J., Liu, N., Twieg, R.J., Piestun, R. and Moerner, W.E. (2009) Three-dimensional, single-molecule fluorescence imaging beyond the diffraction limit by using a double-helix point spread function. *Proc Natl Acad Sci U S A*, **106**, 2995-2999.

59. Backlund, M.P., Lew, M.D., Backer, A.S., Sahl, S.J., Grover, G., Agrawal, A., Piestun, R. and Moerner, W.E. (2012) Simultaneous, accurate measurement of the 3D position and orientation of single molecules. *Proc Natl Acad Sci U S A*, **109**, 19087-19092.

60. Xu, K., Zhong, G. and Zhuang, X. (2013) Actin, spectrin, and associated proteins form a periodic cytoskeletal structure in axons. *Science*, **339**, 452-456.

61. Reyes-Lamothe, R., Sherratt, D.J. and Leake, M.C. (2010) Stoichiometry and architecture of active DNA replication machinery in Escherichia coli. *Science*, **328**, 498-501.

62. Lia, G., Michel, B. and Allemand, J.F. (2012) Polymerase exchange during Okazaki fragment synthesis observed in living cells. *Science*, **335**, 328-331.

63. van Oijen, A.M. and Loparo, J.J. (2010) Single-molecule studies of the replisome. *Annual review of biophysics*, **39**, 429-448.

64. Tanner, N.A., Loparo, J.J., Hamdan, S.M., Jergic, S., Dixon, N.E. and van Oijen, A.M. (2009) Real-time single-molecule observation of rolling-circle DNA replication. *Nucleic acids research*, **37**, e27.

65. Kim, S., Blainey, P.C., Schroeder, C.M. and Xie, X.S. (2007) Multiplexed single-molecule assay for enzymatic activity on flow-stretched DNA. *Nature methods*, **4**, 397-399.

66. van Oijen, A.M. (2007) Single-molecule studies of complex systems: the replisome. *Molecular bioSystems*, **3**, 117-125.

67. Greenleaf, W.J., Woodside, M.T. and Block, S.M. (2007) High-resolution, single-molecule measurements of biomolecular motion. *Annual review of biophysics and biomolecular structure*, **36**, 171-190.

68. Garalde, D.R., Simon, C.A., Dahl, J.M., Wang, H., Akeson, M. and Lieberman, K.R. (2011) Distinct complexes of DNA polymerase I (Klenow fragment) for base and sugar discrimination during nucleotide substrate selection. *J Biol Chem*, **286**, 14480-14492.

69. Benner, S., Chen, R.J., Wilson, N.A., Abu-Shumays, R., Hurt, N., Lieberman, K.R., Deamer, D.W., Dunbar, W.B. and Akeson, M. (2007) Sequence-specific detection of individual DNA polymerase complexes in real time using a nanopore. *Nature nanotechnology*, **2**, 718-724.

70. Hurt, N., Wang, H., Akeson, M. and Lieberman, K.R. (2009) Specific nucleotide binding and rebinding to individual DNA polymerase complexes captured on a nanopore. *Journal of the American Chemical Society*, **131**, 3772-3778.

71. Echols, H. and Goodman, M.F. (1991) Fidelity mechanisms in DNA replication. *Annu Rev Biochem*, **60**, 477-511.

72. Johnson, K.A. (2010) The kinetic and chemical mechanism of high-fidelity DNA polymerases. *Biochim Biophys Acta*, **1804**, 1041-1048.

73. Dzantiev, L. and Romano, L.J. (2000) A conformational change in E. coli DNA polymerase I (Klenow fragment) is induced in the presence of a dNTP complementary to the template base in the active site. *Biochemistry*, **39**, 356-361.

74. Federley, R.G. and Romano, L.J. (2010) DNA polymerase: structural homology, conformational dynamics, and the effects of carcinogenic DNA adducts. *J Nucleic Acids*, **2010**.

75. Rothwell, P.J., Mitaksov, V. and Waksman, G. (2005) Motions of the fingers subdomain of klentaq1 are fast and not rate limiting: implications for the molecular basis of fidelity in DNA polymerases. *Molecular cell*, **19**, 345-355.

76. Vande Berg, B.J., Beard, W.A. and Wilson, S.H. (2001) DNA structure and aspartate 276 influence nucleotide binding to human DNA polymerase beta. Implication for the identity of the rate-limiting conformational change. *J Biol Chem*, **276**, 3408-3416.

77. Johnson, K.A. (1993) Conformational coupling in DNA polymerase fidelity. *Annu Rev Biochem*, **62**, 685-713.

78. Gill, J.P., Wang, J. and Millar, D.P. (2011) DNA polymerase activity at the single-molecule level. *Biochemical Society transactions*, **39**, 595-599.

79. Stryer, L. and Haugland, R.P. (1967) Energy transfer: a spectroscopic ruler. *Proc Natl Acad Sci U S A*, **58**, 719-726.

80. Joyce, C.M. and Benkovic, S.J. (2004) DNA polymerase fidelity: kinetics, structure, and checkpoints. *Biochemistry*, **43**, 14317-14324.

81. Reha-Krantz, L.J. (2010) DNA polymerase proofreading: Multiple roles maintain genome stability. *Biochim Biophys Acta*, **1804**, 1049-1063.

82. Graham, S.E., Syeda, F. and Cisneros, G.A. (2012) Computational prediction of residues involved in fidelity checking for DNA synthesis in DNA polymerase I. *Biochemistry*, **51**, 2569-2578.

83. Maier, B., Bensimon, D. and Croquette, V. (2000) Replication by a single DNA polymerase of a stretched single-stranded DNA. *Proc Natl Acad Sci U S A*, **97**, 12002-12007.

84. Wuite, G.J., Smith, S.B., Young, M., Keller, D. and Bustamante, C. (2000) Single-molecule studies of the effect of template tension on T7 DNA polymerase activity. *Nature*, **404**, 103-106.

85. Lamichhane, R., Berezhna, S.Y., Gill, J.P., Van der Schans, E. and Millar, D.P. (2013) Dynamics of site switching in DNA polymerase. *Journal of the American Chemical Society*, **135**, 4735-4742.

86. Ibarra, B., Chemla, Y.R., Plyasunov, S., Smith, S.B., Lazaro, J.M., Salas, M. and Bustamante, C. (2009) Proofreading dynamics of a processive DNA polymerase. *EMBO J*, **28**, 2794-2802.

87. Manosas, M., Spiering, M.M., Ding, F., Bensimon, D., Allemand, J.F., Benkovic, S.J. and Croquette, V. (2012) Mechanism of strand displacement synthesis by DNA replicative polymerases. *Nucleic acids research*, **40**, 6174-6186.

88. Manosas, M., Spiering, M.M., Ding, F., Croquette, V. and Benkovic, S.J. (2012) Collaborative coupling between polymerase and helicase for leading-strand synthesis. *Nucleic acids research*, **40**, 6187-6198.

89. Ogawa, T. and Okazaki, T. (1980) Discontinuous DNA replication. *Annual review of biochemistry*, **49**, 421-457.

90. Marians, K.J. (1992) Prokaryotic DNA replication. *Annual review of biochemistry*, **61**, 673-719.

91. Masai, H., Matsumoto, S., You, Z., Yoshizawa-Sugata, N. and Oda, M. (2010) Eukaryotic chromosome DNA replication: where, when, and how? *Annual review of biochemistry*, **79**, 89-130.

92. Hamdan, S.M. and Richardson, C.C. (2009) Motors, switches, and contacts in the replisome. *Annual review of biochemistry*, **78**, 205-243.

93. O'Donnell, M. (2006) Replisome architecture and dynamics in Escherichia coli. *The Journal of biological chemistry*, **281**, 10653-10656.

94. Lee, J.B., Hite, R.K., Hamdan, S.M., Xie, X.S., Richardson, C.C. and van Oijen, A.M. (2006) DNA primase acts as a molecular brake in DNA replication. *Nature*, **439**, 621-624.

95. Hamdan, S.M., Loparo, J.J., Takahashi, M., Richardson, C.C. and van Oijen, A.M. (2009) Dynamics of DNA replication loops reveal temporal control of lagging-strand synthesis. *Nature*, **457**, 336-339.

96. Loparo, J.J., Kulczyk, A.W., Richardson, C.C. and van Oijen, A.M. (2011) Simultaneous single-molecule measurements of phage T7 replisome composition and function reveal the mechanism of polymerase exchange. *Proc Natl Acad Sci U S A*, **108**, 3584-3589.

97. Tanner, N.A., Hamdan, S.M., Jergic, S., Loscha, K.V., Schaeffer, P.M., Dixon, N.E. and van Oijen, A.M. (2008) Single-molecule studies of fork dynamics in Escherichia coli DNA replication. *Nature structural & molecular biology*, **15**, 998.

98. Corn, J.E. and Berger, J.M. (2006) Regulation of bacterial priming and daughter strand synthesis through helicase-primase interactions. *Nucleic acids research*, **34**, 4082-4088.

99. Yao, N.Y., Georgescu, R.E., Finkelstein, J. and O'Donnell, M.E. (2009) Single-molecule analysis reveals that the lagging strand increases replisome processivity but slows replication fork progression. *Proc Natl Acad Sci U S A*, **106**, 13236-13241.

100. Tanner, N.A., Tolun, G., Loparo, J.J., Jergic, S., Griffith, J.D., Dixon, N.E. and van Oijen, A.M. (2011) E. coli DNA replication in the absence of free beta clamps. *EMBO J*, **30**, 1830-1840.

101. Georgescu, R.E., Kurth, I. and O'Donnell, M.E. (2012) Single-molecule studies reveal the function of a third polymerase in the replisome. *Nature structural & molecular biology*, **19**, 113-116.

102. Wessel, R., Schweizer, J. and Stahl, H. (1992) Simian virus 40 T-antigen DNA helicase is a hexamer which forms a binary complex during bidirectional unwinding from the viral origin of DNA replication. *Journal of virology*, **66**, 804-815.

103. Yardimci, H., Loveland, A.B., Habuchi, S., van Oijen, A.M. and Walter, J.C. (2010) Uncoupling of sister replisomes during eukaryotic DNA replication. *Molecular cell*, **40**, 834-840.

104. Yardimci, H., Wang, X., Loveland, A.B., Tappin, I., Rudner, D.Z., Hurwitz, J., van Oijen, A.M. and Walter, J.C. (2012) Bypass of a protein barrier by a replicative DNA helicase. *Nature*, **492**, 205-209.

105. Weisshart, K., Taneja, P., Jenne, A., Herbig, U., Simmons, D.T. and Fanning, E. (1999) Two regions of simian virus 40 T antigen determine cooperativity of double-hexamer assembly on the viral origin of DNA replication and promote hexamer interactions during bidirectional origin DNA unwinding. *Journal of virology*, **73**, 2201-2211.

106. Duderstadt, K.E., Reyes-Lamothe, R., van Oijen, A.M. and Sherratt, D.J. (2013) Replication-Fork Dynamics. *Cold Spring Harbor perspectives in biology*.

107. Yardimci, H., Loveland, A.B., van Oijen, A.M. and Walter, J.C. (2012) Single-molecule analysis of DNA replication in Xenopus egg extracts. *Methods*, **57**, 179-186.

108. Rogozin, I.B. and Pavlov, Y.I. (2003) Theoretical analysis of mutation hotspots and their DNA sequence context specificity. *Mutation research*, **544**, 65-85.

109. Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B. *et al.* (2009) Real-time DNA sequencing from single polymerase molecules. *Science*, **323**, 133-138.

110. Flusberg, B.A., Webster, D.R., Lee, J.H., Travers, K.J., Olivares, E.C., Clark, T.A., Korlach, J. and Turner, S.W. (2010) Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nature methods*, **7**, 461-465.

111. Clark, T.A., Spittle, K.E., Turner, S.W. and Korlach, J. (2011) Direct detection and sequencing of damaged DNA bases. *Genome integrity*, **2**, 10.

112. Song, C.X., Clark, T.A., Lu, X.Y., Kislyuk, A., Dai, Q., Turner, S.W., He, C. and Korlach, J. (2012) Sensitive and specific single-molecule sequencing of 5-hydroxymethylcytosine. *Nature methods*, **9**, 75-77.

113. Schwartz, J.J. and Quake, S.R. (2009) Single molecule measurement of the "speed limit" of DNA polymerase. *Proc Natl Acad Sci U S A*, **106**, 20294-20299.

114. Ortiz, T.P., Marshall, J.A., Meyer, L.A., Davis, R.W., Macosko, J.C., Hatch, J., Keller, D.J. and Brozik, J.A. (2005) Stepping statistics of single HIV-1 reverse transcriptase molecules during DNA polymerization. *The journal of physical chemistry. B*, **109**, 16127-16131.

115. Kim, S., Schroeder, C.M. and Xie, X.S. (2010) Single-molecule study of DNA polymerization activity of HIV-1 reverse transcriptase on DNA templates. *J Mol Biol*, **395**, 995-1006.

116. Joyce, C.M. and Derbyshire, V. (1995) Purification of Escherichia coli DNA polymerase I and Klenow fragment. *Methods Enzymol*, **262**, 3-13.

117. Boudsocq, F., Iwai, S., Hanaoka, F. and Woodgate, R. (2001) Sulfolobus solfataricus P2 DNA polymerase IV (Dpo4): an archaeal DinB-like DNA

polymerase with lesion-bypass properties akin to eukaryotic poleta. *Nucleic acids research*, **29**, 4607-4616.

118. Dutta, S., Li, Y., Johnson, D., Dzantiev, L., Richardson, C.C., Romano, L.J. and Ellenberger, T. (2004) Crystal structures of 2-acetylaminofluorene and 2-aminofluorene in complex with T7 DNA polymerase reveal mechanisms of mutagenesis. *Proc Natl Acad Sci U S A*, **101**, 16186-16191.

119. Lamichhane, R., Solem, A., Black, W. and Rueda, D. (2010) Single-molecule FRET of protein-nucleic acid and protein-protein complexes: surface passivation and immobilization. *Methods*, **52**, 192-200.

120. McKinney, S.A., Joo, C. and Ha, T. (2006) Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophys J*, **91**, 1941-1951.

121. Lakowicz, J.R. (2006) *Principles of fluorescence spectroscopy*. 3rd ed. Springer, New York.

122. Vives, E. and Lebleu, B. (2003) One-pot labeling and purification of peptides and proteins with fluorescein maleimide. *Tetrahedron Lett*, **44**, 5389-5391.

123. Kalab, P. and Soderholm, J. (2010) The design of Forster (fluorescence) resonance energy transfer (FRET)-based molecular sensors for Ran GTPase. *Methods*, **51**, 220-232.

124. Modesti, M. (2011) Fluorescent labeling of proteins. *Methods Mol Biol*, **783**, 101-120.

125. Kapanidis, A.N. and Weiss, S. (2002) Fluorescent probes and bioconjugation chemistries for single-molecule fluorescence analysis of biomolecules. *Journal of Chemical Physics*, **117**, 10953-10964.

126. Seo, M.H., Lee, T.S., Kim, E., Cho, Y.L., Park, H.S., Yoon, T.Y. and Kim, H.S. (2011) Efficient single-molecule fluorescence resonance energy transfer analysis by site-specific dual-labeling of protein using an unnatural amino acid. *Analytical chemistry*, **83**, 8849-8854.

127. Rashidian, M., Dozier, J.K. and Distefano, M.D. (2013) Enzymatic Labeling of Proteins: Techniques and Approaches. *Bioconjug Chem*.

128. Sandtner, W., Bezanilla, F. and Correa, A.M. (2007) In vivo measurement of intramolecular distances using genetically encoded reporters. *Biophysical Journal*, **93**, L45-47.

129. Yang, J.Y. and Yang, W.Y. (2009) Site-Specific Two-Color Protein Labeling for FRET Studies Using Split Inteins. *Journal of the American Chemical Society*, **131**, 11644-+.

130. Tsien, R.Y. (1998) The green fluorescent protein. *Annu Rev Biochem*, **67**, 509-544.

131. Kim, Y., Ho, S.O., Gassman, N.R., Korlann, Y., Landorf, E.V., Collart, F.R. and Weiss, S. (2008) Efficient site-specific labeling of proteins via cysteines. *Bioconjug Chem*, **19**, 786-791.

132. Carrigan, P.E., Ballar, P. and Tuzmen, S. (2011) Site-directed mutagenesis. *Methods Mol Biol*, **700**, 107-124.

133. Joo, C. and Ha, T. (2012) Labeling proteins for single-molecule FRET. *Cold Spring Harbor protocols*, **2012**, 1009-1012.

134. Puljung, M.C. and Zagotta, W.N. (2011) Labeling of specific cysteines in proteins using reversible metal protection. *Biophysical Journal*, **100**, 2513-2521.

135. Kuiper, J.M., Pluta, R., Huibers, W.H., Fusetti, F., Geertsma, E.R. and Poolman, B. (2009) A method for site-specific labeling of multiple protein thiols. *Protein science : a publication of the Protein Society*, **18**, 1033-1041.

136. Ratner, V., Kahana, E., Eichler, M. and Haas, E. (2002) A general strategy for site-specific double labeling of globular proteins for kinetic FRET studies. *Bioconjug Chem*, **13**, 1163-1170.

137. Karunatilaka, K.S., Solem, A., Pyle, A.M. and Rueda, D. (2010) Single-molecule analysis of Mss116-mediated group II intron folding. *Nature*, **467**, 935-939.

138. Datta, K., Johnson, N.P., LiCata, V.J. and von Hippel, P.H. (2009) Local conformations and competitive binding affinities of single- and double-stranded primer-template DNA at the polymerization and editing active sites of DNA polymerases. *J Biol Chem*, **284**, 17180-17193.

139. Tsoi, P.Y., Zhang, X., Sui, S.F. and Yang, M. (2003) Effects of DNA mismatches on binding affinity and kinetics of polymerase-DNA complexes as revealed by surface plasmon resonance biosensor. *The Analyst*, **128**, 1169-1174.

140. Fabian, A.I., Rente, T., Szollosi, J., Matyus, L. and Jenei, A. (2010) Strength in numbers: effects of acceptor abundance on FRET efficiency. *Chemphyschem*, **11**, 3713-3721.

141. Koushik, S.V., Blank, P.S. and Vogel, S.S. (2009) Anomalous surplus energy transfer observed with multiple FRET acceptors. *PLoS One*, **4**, e8031.

142. Doublie, S., Sawaya, M.R. and Ellenberger, T. (1999) An open and closed case for all polymerases. *Structure*, **7**, R31-35.

143. Datta, K., Johnson, N.P. and von Hippel, P.H. (2010) DNA conformational changes at the primer-template junction regulate the fidelity of replication by DNA polymerase. *Proc Natl Acad Sci U S A*, **107**, 17980-17985.

144. Freemont, P.S., Friedman, J.M., Beese, L.S., Sanderson, M.R. and Steitz, T.A. (1988) Cocrystal structure of an editing complex of Klenow fragment with DNA. *Proc Natl Acad Sci U S A*, **85**, 8924-8928.

145. Guo, Z., Karunatilaka, K.S. and Rueda, D. (2009) Single-molecule analysis of protein-free U2-U6 snRNAs. *Nat Struct Mol Biol*, **16**, 1154-1159.

146. Joyce, C.M. and Steitz, T.A. (1987) DNA-Polymerase-I - from Crystal-Structure to Function Via Genetics. *Trends Biochem Sci*, **12**, 288-292.

147. Allen, D.J., Darke, P.L. and Benkovic, S.J. (1989) Fluorescent oligonucleotides and deoxynucleotide triphosphates: preparation and their interaction with the large (Klenow) fragment of Escherichia coli DNA polymerase I. *Biochemistry*, **28**, 4601-4607.

148. Halford, S.E. (2009) An end to 40 years of mistakes in DNA-protein association kinetics? *Biochemical Society transactions*, **37**, 343-348.

149. Kuchta, R.D., Mizrahi, V., Benkovic, P.A., Johnson, K.A. and Benkovic, S.J. (1987) Kinetic mechanism of DNA polymerase I (Klenow). *Biochemistry*, **26**, 8410-8417.

150. Polesky, A.H., Steitz, T.A., Grindley, N.D. and Joyce, C.M. (1990) Identification of residues critical for the polymerase activity of the Klenow fragment of DNA polymerase I from Escherichia coli. *J Biol Chem*, **265**, 14579-14591.

151. Warwicker, J., Ollis, D., Richards, F.M. and Steitz, T.A. (1985) Electrostatic field of the large fragment of Escherichia coli DNA polymerase I. *J Mol Biol*, **186**, 645-649.

152. Datta, K. and LiCata, V.J. (2003) Salt dependence of DNA binding by Thermus aquaticus and Escherichia coli DNA polymerases. *J Biol Chem*, **278**, 5694-5701.

153. Lohman, T.M. and Mascotti, D.P. (1992) Thermodynamics of ligand-nucleic acid interactions. *Methods in enzymology*, **212**, 400-424.

154. Datta, K. and LiCata, V.J. (2003) Salt dependence of DNA binding by Thermus aquaticus and Escherichia coli DNA polymerases. *The Journal of biological chemistry*, **278**, 5694-5701.

155. Dzantiev, L. and Romano, L.J. (1999) Interaction of Escherichia coli DNA polymerase I (Klenow fragment) with primer-templates containing N-acetyl-2-aminofluorene or N-2-aminofluorene adducts in the active site. *J Biol Chem*, **274**, 3279-3284.

156. Dzantiev, L., Alekseyev, Y.O., Morales, J.C., Kool, E.T. and Romano, L.J. (2001) Significance of nucleobase shape complementarity and hydrogen bonding in the

formation and stability of the closed polymerase-DNA complex. *Biochemistry*, **40**, 3215-3221.

157. Lone, S. and Romano, L.J. (2003) Mechanistic insights into replication across from bulky DNA adducts: a mutant polymerase I allows an N-acetyl-2-aminofluorene adduct to be accommodated during DNA synthesis. *Biochemistry*, **42**, 3826-3834.

158. Kool, E.T., Morales, J.C. and Guckian, K.M. (2000) Mimicking the Structure and Function of DNA: Insights into DNA Stability and Replication. *Angew Chem Int Ed Engl*, **39**, 990-1009.

159. Trostler, M., Delier, A., Beckman, J., Urban, M., Patro, J.N., Spratt, T.E., Beese, L.S. and Kuchta, R.D. (2009) Discrimination between right and wrong purine dNTPs by DNA polymerase I from Bacillus stearothermophilus. *Biochemistry*, **48**, 4633-4641.

160. Rothwell, P.J. and Waksman, G. (2007) A pre-equilibrium before nucleotide binding limits fingers subdomain closure by Klentaq1. *J Biol Chem*, **282**, 28884-28892.

161. Astatke, M., Ng, K., Grindley, N.D. and Joyce, C.M. (1998) A single side chain prevents Escherichia coli DNA polymerase I (Klenow fragment) from incorporating ribonucleotides. *Proc Natl Acad Sci U S A*, **95**, 3402-3407.

162. Joyce, C.M. and Steitz, T.A. (1994) Function and structure relationships in DNA polymerases. *Annu Rev Biochem*, **63**, 777-822.

163. Hanahan, D. and Weinberg, R.A. (2000) The hallmarks of cancer. *Cell*, **100**, 57-70.

164. Vineis, P. (1994) Epidemiology of cancer from exposure to arylamines. *Environmental health perspectives*, **102 Suppl 6**, 7-10.

165. Yu, M.C., Skipper, P.L., Tannenbaum, S.R., Chan, K.K. and Ross, R.K. (2002) Arylamine exposures and bladder cancer risk. *Mutation research*, **506-507**, 21-28.

166. Skipper, P.L., Tannenbaum, S.R., Ross, R.K. and Yu, M.C. (2003) Nonsmoking-related arylamine exposure and bladder cancer risk. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, **12**, 503-507.

167. Hsu, G.W., Kiefer, J.R., Burnouf, D., Becherel, O.J., Fuchs, R.P. and Beese, L.S. (2004) Observing translesion synthesis of an aromatic amine DNA adduct by a high-fidelity DNA polymerase. *J Biol Chem*, **279**, 50280-50285.

168. Shibutani, S. and Grollman, A.P. (1993) On the mechanism of frameshift (deletion) mutagenesis in vitro. *J Biol Chem*, **268**, 11703-11710.

169. Michaels, M.L., Johnson, D.L., Reid, T.M., King, C.M. and Romano, L.J. (1987) Evidence for in vitro translesion DNA synthesis past a site-specific aminofluorene adduct. *J Biol Chem*, **262**, 14648-14654.

170. Vooradi, V. and Romano, L.J. (2009) Effect of N-2-acetylaminofluorene and 2-aminofluorene adducts on DNA binding and synthesis by yeast DNA polymerase eta. *Biochemistry*, **48**, 4209-4216.

171. Tebbs, R.S. and Romano, L.J. (1994) Mutagenesis at a site-specifically modified NarI sequence by acetylated and deacetylated aminofluorene adducts. *Biochemistry*, **33**, 8998-9006.

172. Fuchs, R.P. (1984) DNA binding spectrum of the carcinogen N-acetoxy-N-2-acetylaminofluorene significantly differs from the mutation spectrum. *J Mol Biol*, **177**, 173-180.

173. Hoffmann, G.R. and Fuchs, R.P. (1997) Mechanisms of frameshift mutations: insight from aromatic amines. *Chemical research in toxicology*, **10**, 347-359.

174. Bebenek, K. and Kunkel, T.A. (2004) Functions of DNA polymerases. *Adv Protein Chem*, **69**, 137-165.

175. Cann, I.K. and Ishino, Y. (1999) Archaeal DNA replication: identifying the pieces to solve a puzzle. *Genetics*, **152**, 1249-1267.

176. Dzantiev, L. and Romano, L.J. (2000) 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*, **39**, 5139-5145.

177. Li, G.W. and Xie, X.S. (2011) Central dogma at the single-molecule level in living cells. *Nature*, **475**, 308-315.

178. Mujumdar, R.B., Ernst, L.A., Mujumdar, S.R., Lewis, C.J. and Waggoner, A.S. (1993) Cyanine dye labeling reagents: sulfoindocyanine succinimidyl esters. *Bioconjug Chem*, **4**, 105-111.

179. Aliotta, J.M., Pelletier, J.J., Ware, J.L., Moran, L.S., Benner, J.S. and Kong, H. (1996) Thermostable Bst DNA polymerase I lacks a 3'-->5' proofreading exonuclease activity. *Genetic analysis : biomolecular engineering*, **12**, 185-195.

180. Miller, H. and Grollman, A.P. (1997) Kinetics of DNA polymerase I (Klenow fragment exo-) activity on damaged DNA templates: effect of proximal and distal template damage on DNA synthesis. *Biochemistry*, **36**, 15336-15342.

181. Bryant, F.R., Johnson, K.A. and Benkovic, S.J. (1983) Elementary steps in the DNA polymerase I reaction pathway. *Biochemistry*, **22**, 3537-3546.

182. Bermek, O., Grindley, N.D. and Joyce, C.M. (2013) Prechemistry nucleotide selection checkpoints in the reaction pathway of DNA polymerase I and the roles of Glu710 and Tyr766. *Biochemistry*.

183. Kuchta, R.D., Benkovic, P. and Benkovic, S.J. (1988) Kinetic mechanism whereby DNA polymerase I (Klenow) replicates DNA with high fidelity. *Biochemistry*, **27**, 6716-6725.

184. Denissenko, M.F., Pao, A., Tang, M. and Pfeifer, G.P. (1996) Preferential formation of benzo[a]pyrene adducts at lung cancer mutational hotspots in P53. *Science*, **274**, 430-432.

185. Alexandrov, K., Rojas, M. and Satarug, S. (2010) The critical DNA damage by benzo(a)pyrene in lung tissues of smokers and approaches to preventing its formation. *Toxicology letters*, **198**, 63-68.

ABSTRACT

UNVEILING DNA POLYMERASE SYNTHESIS AND PROOFREADING ACTIVITIES – ONE MOLECULE AT A TIME

by

KYLE B. VRTIS

December 2013

Advisor: Dr. Louis J. Romano

Major: Chemistry (Biochemistry)

Degree: Doctor of Philosophy

DNA polymerases maintain the genome integrity from one generation to the next by faithfully synthesizing new DNA and by participating in DNA repair processes. The Klenow fragment of *E. coli* DNA polymerase I has served as a model polymerase for decades because it is straightforward to purify, well-characterized kinetically, and it is able to carry out DNA synthesis and proofreading. In 2009, Christian, et al. developed a single molecule Förster resonance energy transfer (smFRET) approach to monitor the polymerase position on the DNA with single base pair resolution.

We have worked to optimize nearly every aspect of that approach and to apply the technique to characterize DNA polymerase dynamics on the DNA with unprecedented detail. First, we improved the cysteine-labeling protocol to ensure the labeling was site-specific and efficient. Second, we applied a new single molecule fluorescence technique, single molecule protein induced fluorescence enhancement (smPIFE), to our system to accurately measure binding dynamics and to strengthen our interpretation of the

polymerase position on the DNA. Third, determined how carcinogenic DNA adducts disrupt polymerase binding on the DNA by using smFRET and smPIFE. Accordingly, we designed the experiment to distinguish between pol and exo site binding in real-time, and we identified a novel intermediate state in the proofreading mechanism. Finally, we tracked individual polymerases as they incorporated bases on the DNA. These experiments support the existence of a fidelity-checking step following each incorporation. The assays and applications described herein lay the groundwork for future DNA polymerase mechanistic studies.

AUTOBIOGRAPHICAL STATEMENT

Kyle Vrtis

EDUCATION

Ph.D.	Chemistry – Major in Biochemistry, 3.94 GPA	2009-Present
	Wayne State University, Detroit, Michigan	
	Advisors – Dr. Louis Romano and Dr. David Rueda	
B.S.	Chemistry – Major in Biochemistry, Magna Cum Laude	2004-2008
	Northern Michigan University, Marquette, Michigan	
	Advisor – Dr. Lesley Putman	

PRESENTATIONS

Oral

- 2012 Graduate Research Symposium, Wayne State University, Detroit, Michigan — October, 2012
- 2012 Midwest Single Molecule Workshop, University of Michigan, Ann Arbor, Michigan — July, 2012

Poster

- 2013 Graduate Exhibition, Wayne State University, Detroit, Michigan — March, 2013
- 2013 Annual Biophysical Society Meeting, Pennsylvania Convention Center, Philadelphia, Pennsylvania — February, 2013
- Gordon Research Conference: Single Molecule Approaches to Biology, Mount Snow Resort, West Dover, Vermont — July, 2012

PUBLICATIONS

1. **Vrtis, K.B.***, Markiewicz, R.P.*, Rueda, D., and Romano, L.J. (2012). Single-molecule microscopy reveals new insights into nucleotide selection by DNA polymerase I. Nucleic acids research *40*, 7975-7984. *co-first authors, (*Feature Article – top 5% of papers; NAR cover image*).
2. **Vrtis, K.B.**, Markiewicz, R.P., Romano, L.J., Rueda, D. (2013) Carcinogenic DNA adducts induce distinct DNA polymerase binding orientations. Nucleic acids research 41, 7843-7853. (*NAR cover image*).
3. **Vrtis, K.B.***, Markiewicz, R.P.*, Romano, L.J., Rueda, D. General methodology to ensure site-specific and efficient dye conjugation to proteins. (Manuscript in preparation). *co-first authors.
4. **Vrtis, K.B.**, Rueda, D., Romano, L.J., Unveiling DNA polymerase synthesis and proofreading activities – one molecule at a time. (Manuscript in preparation).