Mechanistic Studies of DNA Replication, Lesion Bypass, and Editing

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Austin Tyler Raper, B.S.

The Ohio State University Biochemistry Graduate Program

The Ohio State University

2018

Dissertation Committee

Dr. Zucai Suo, Advisor

Dr. Ross Dalbey

Dr. Michael Poirier

Dr. Richard Swenson

Abstract

DNA acts as a molecular blueprint for life. Adenosine, cytidine, guanosine, and thymidine nucleotides serve as the building blocks of DNA and can be arranged in near-endless combinations. These unique sequences of DNA may encode genes that when expressed produce RNA, proteins, and enzymes responsible for executing diverse tasks necessary for biological existence. Accordingly, careful maintenance of the molecular integrity of DNA is paramount for the growth, development, and functioning of organisms. However, DNA is damaged upon reaction with pervasive chemicals generated by normal cellular metabolism or encountered through the environment. The resulting DNA lesions act as roadblocks to high-fidelity A- and B-family DNA polymerases responsible for replicating DNA in preparation for cell division which may lead to programmed cell death. Additionally, these lesions may fool the polymerase into making errors during DNA replication, leading to genetic mutations and cancer. Fortunately, the cell has evolved DNA damage tolerance as an emergency response to such lesions.

During DNA damage tolerance, a damage-stalled high-fidelity polymerase is substituted for a specialized Y-family polymerase, capable of bypassing the offending DNA lesion, for replication to continue. However, the ability of the specialized polymerase to bypass DNA lesions occurs at the expense of replication fidelity. Hence, tight regulation of polymerase exchange during DNA damage tolerance is imperative to ensure timely bypass of a lesion by the Y-family member, as well as prompt polymerase

replacement by an A- or B-family member to limit DNA replication errors (*i.e.* mutations). Nevertheless, mistakes during DNA damage tolerance that evade DNA repair pathways are intimately connected to mutagenesis and may lead to cancer or numerous other genetic diseases. Until recently, making corrections to erroneous DNA sequences in the cell was prohibitively time-consuming, expensive, and laborious. However, the advent of clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) proteins as a convenient technology for gene editing has opened the door for revolutionary cures to genetic diseases and state-of-the-art research into complex disease states.

Here I have investigated the functions of enzymes and associated cofactors critical for DNA replication, lesion bypass, and editing. Through pre-steady-state gel- and fluorescence-based kinetic techniques, as well as single-molecule fluorescence spectroscopy, I identified and characterized discrete steps of the kinetic mechanisms of these enzymes to better understand how they execute their unique activities, interact with protein partners, and potentially contribute to disease.

Dedication

To Lauren, my captive audience, gentle critic, and fierce advocate.

Acknowledgments

I will begin by thanking my doctoral advisor Dr. Suo for persuading me to join his lab in February of 2014. The first two semesters of graduate school can be quite overwhelming for new students as a result of the heavy course load, laboratory rotations, and looming deadline for selecting an advisor. Dr. Suo expressed intense confidence in my abilities and was gracious in helping me to sort out my research interests and goals during this stressful time. I am thankful for him taking a vested interest in my career. Moreover, I am certain that without his expectations of excellence and constant drive for success, graduate school would have not been as fruitful or rewarding.

I would also like to thank my committee members Drs. Ross Dalbey, Michael Poirier, and Richard Swenson. These accomplished scientists have followed my graduate school progress with enthusiasm and have graciously written strong letters of recommendation on my behalf for many successful attempts at awards or fellowship competitions. During these past few years, they have provided sound research and career advice as well as helped me to complete this dissertation.

I would also like to acknowledge the Ohio State Biochemistry Program and Department of Chemistry and Biochemistry for providing a collaborative atmosphere with excellent facilities and staff for conducting high-level research. Specifically, I thank Drs. Tom Magliery and Jane Jackman for selecting me to represent OSBP in the university-wide Presidential Fellowship competition. Their nomination and advocacy was

crucial to achieving that honor and accompanying funding for my final year of dissertation research. In addition, I thank Dr. Dehua Pei for accepting me into the Chemistry-Biology Interface Training Program which funded my research for two-years.

I am also grateful to the past and current members of the Suo lab who have helped shape me as a scientist and as an individual including Andrew Reed, Anthony Stephenson, Jack Tokarsky, and Walter Zahurancik, as well as Drs. Varun Gadkari, Brian Maxwell, and David Taggart. These gentlemen made graduate school enjoyable and were always quick to offer advice or assistance for any of my endeavors. It was truly an honor to learn, struggle, fail, and triumph with them. I am privileged to call many of them close friends and even brothers. I look forward to growing these relationships for the rest of my life.

I am eternally thankful to my family, including my wife, mother, father, brothers, sister, and in-laws for their constant love and support that has sustained me on this journey. Above all, I would like to thank God for the opportunity of graduate school. Five years ago, I was struggling to find direction after graduating with my bachelor's degree. Fortunately, the Lord has blessed me with an incredible wife that wisely and boldly encouraged me to enroll at The Ohio State University. Together, we embarked on an adventure to Columbus that was heart-wrenchingly difficult, as we left behind many close friends and family, but at the same time permanently rewarding, as we have made new friends and found a church family that feels like home.

Vita

2009–2013……..... B.S. Biochemistry

University of Mount Union, Alliance, Ohio

2013–2018…… Ph.D. Biochemistry

The Ohio State University, Columbus, Ohio

2013–2014…… Ohio State Biochemistry Program Fellow

The Ohio State University, Columbus, Ohio

2014–2015…… Graduate Teaching Associate

Center for Life Sciences Education

The Ohio State University, Columbus, Ohio

2014–2015…… Graduate Teaching Associate,

Department of Chemistry and Biochemistry,

The Ohio State University, Columbus, Ohio

2015–2017…… Chemistry-Biology Interface Program Fellow

The Ohio State University, Columbus, Ohio

2017–2018…… Presidential Research Fellow

The Ohio State University, Columbus, Ohio

Publications
***denotes co-first authorship**

1. **Raper, Austin T.***, Stephenson, A.A., Suo, Z. (2018) Sharpening the Scissors: Mechanistic Details of CRISPR/Cas9 Improve Functional Understanding and Inspire Future Research. *J. Am. Chem. Soc.* In review by invitation.
2. **Raper, Austin T.***, Reed, A.J., Suo, Z. (2018) Kinetic Mechanism of DNA Polymerases: Contributions of Conformational Dynamic and a Third Divalent Metal Ion. *Chem. Rev.* DOI: 10.1021/acs.chemrev.7b00685
3. Stephenson, A.A., **Raper, Austin T.***, and Suo Z. (2018) Bidirectional Degradation of DNA Cleavage Products Catalyzed by CRISPR/Cas9. *J. Am. Chem. Soc.* DOI: 10.1021/jacs.7b13050
4. **Raper, Austin T.***, Stephenson, A.A., and Suo Z. (2018) Functional Insights Revealed by the Kinetic Mechanism of CRISPR/Cas9. *J. Am. Chem. Soc.* DOI: 10.1021/jacs.7b13047
5. Gadkari V.V., Harvey S.R., **Raper, Austin T.***, Chu W, Wang J, Wysocki VH, and Suo Z. (2018) Investigation of Sliding DNA Clamp Dynamics by Single-Molecule Fluorescence, Mass Spectrometry, and Structure-Based Modeling. *Nucleic Acids Research*. DOI: 10.1093/nar/gky125
6. Vyas, R., Reed, A.J., **Raper, Austin T.**, Zahurancik, W.J., Wallenmeyer, P.C., and Suo, Z. (2017) Structural basis for the D-stereoselectivity of human DNA polymerase β. *Nucleic Acids Research.* 45 (10), 6228-6237. DOI: 10.1093/nar/gkx252

7. Reed, A.J., Vyas, R., **Raper, Austin T.**, and Suo, Z. (2017) Structural Insights into the Post-Chemistry Steps of Nucleotide Incorporation Catalyzed by a DNA Polymerase. *J. Am. Chem. Soc.* 139, 465-471. DOI: 10.1021/jacs.6b11258
8. **Raper, Austin T.***, Reed, A.J., Gadkari, V.V., and Suo, Z. (2017) Advances in Structural and Single-Molecule Methods for Investigating DNA Damage Tolerance and Repair. *Chem. Res. Toxicol.* 30, 260-269. DOI: 10.1021/acs.chemrestox.6b00342
9. **Raper, Austin T.**, Suo, Z. (2016) Investigation of Intradomain Motions of a Y-Family DNA Polymerase during Substrate Binding and Catalysis. *Biochemistry.* 55(41):5832-5844. DOI: 10.1021/acs.biochem.6b00878
10. **Raper, Austin T.***, Gadkari V.V., Maxwell B.A., Suo Z. (2016) Single-Molecule Investigation of Response to Oxidative DNA Damage by a Y-Family DNA Polymerase. *Biochemistry.* 55(14):2187-96. DOI: 10.1021/acs.biochem.6b00166

Fields of Study

Major Field: The Ohio State University Biochemistry Graduate Program

Table of Contents

List of Schemes

List of Tables

List of Figures

Chapter 1. Introduction to DNA Replication and DNA Polymerases

Kinetic Mechanism of DNA Polymerases: Contributions of Conformational Dynamics and a Third Divalent Metal Ion

Author Contributions

Austin T. Raper and Andrew J. Reed equally contributed as co-first authors. Dr. Zucai Suo provided some conceptual direction and helped edit the manuscript.

1.1 Abstract

Faithful transmission and maintenance of genetic material is primarily fulfilled by DNA polymerases. During DNA replication, these enzymes catalyze incorporation of deoxynucleotides into a DNA primer strand based on Watson-Crick complementarity to the DNA template strand. Through the years, research on DNA polymerases from every family and reverse transcriptases, has revealed structural and functional similarities, including a conserved domain architecture and purported two-metal-ion mechanism for nucleotidyltransfer. However, it is equally clear that DNA polymerases possess distinct differences that often prescribe a particular cellular role. Indeed, a unified kinetic mechanism to explain all aspects of DNA polymerase catalysis, including DNA binding, nucleotide binding and incorporation, and metal-ion-assisted nucleotidyltransfer (*i.e.* chemistry), has been difficult to define. In particular, the contributions of enzyme conformational dynamics to several mechanistic steps and their implications for replication fidelity are complex. Moreover, recent time-resolved X-ray crystallographic studies of DNA polymerases have uncovered a third divalent metal ion present during DNA synthesis, the function of which is currently unclear and debated within the field. In this review, we survey past and current literature describing the structures and kinetic mechanisms of DNA polymerases from each family to explore every major mechanistic step while emphasizing the impact of enzyme conformational dynamics on DNA synthesis and replication fidelity. This also includes brief insight into the structural and kinetic techniques utilized to study DNA polymerases and RTs. Furthermore, we present the evidences for the two-metal-ion mechanism for DNA polymerase catalysis prior to

interpreting the recent structural findings describing a third divalent metal ion. We conclude by discussing the diversity of DNA polymerase mechanisms and suggest future characterization of the third divalent metal ion to dissect its role in DNA polymerase catalysis.

1.2 Introduction

It is well-known that enzymes evolved for catalysis on nucleic acid substrates often undergo conformational dynamics and engage metal ion cofactors to achieve remarkable catalytic efficiency and reaction specificity.[1-14] In fact, replication of valuable genetic material is entrusted to DNA polymerases, which utilize divalent metal ions to catalyze DNA synthesis. Since their initial discovery in 1950s,[15,16] many DNA polymerases have been identified and phylogenetically classified into distinct A, B, C, D, X, Y, and reverse transcriptase (RT) families based on sequence homology as well as functional and structural analyses.[8,17,18]

As DNA polymerases catalyze the same fundamental reaction (*i.e.* incorporation of deoxyribonucleotide (dNTP) into a nascent DNA primer strand), one could expect these enzymes to share a unified kinetic mechanism describing DNA binding, nucleotide binding, and nucleotide incorporation. However, functional studies have revealed that each polymerase family is often suited to a particular cellular role[19,20] as evident through the utilization of distinct DNA substrates (*i.e.* primer-template DNA, gapped DNA, damage-containing DNA, single-stranded DNA, *etc.*) and wide-ranging nucleotide substrate specificities, which result in varying DNA replication efficiency and fidelity.[8,20-26] Indeed, while some mechanistic steps remain common among DNA polymerases,

researchers have uncovered several events that seem unique to a particular polymerase, or more broadly, a polymerase family. These events are often related to conformational dynamics and may prescribe distinct properties to the polymerase, which dramatically influence DNA and nucleotide binding as well as nucleotide incorporation. In fact, there is substantial debate about the involvement of a particular conformational change in the rate-limiting step of single-nucleotide incorporation and how this step may influence the fidelity of DNA polymerization.[4-6,20,27-29] In this review, we will describe the minimal kinetic mechanism for single-nucleotide incorporation determined by extensive structural and functional studies of DNA polymerases performed by our lab and many others over the last 30 years and we will emphasize the importance of DNA polymerase dynamics to the mechanism of DNA polymerization. Altogether, our comprehensive analysis of DNA polymerase kinetics has led us to the conclusion that a common kinetic mechanism, encompassing all DNA polymerases, likely does not exist and each enzyme should be considered independently.

A two-metal ion mechanism for enzymes that can act on the phosphodiester backbone of DNA or RNA was first postulated by Beese and Steitz in 1991 based on crystal structures of exonucleolytic substrate and product complexes within the active site of the 3′-5′ exonuclease domain of *Escherichia coli* DNA polymerase I (Pol I).[30,31] They postulated that the mechanism of exonucleolytic cleavage would extend to DNA polymerization with each divalent metal ion coordinating essential active site residues and substrate groups as well as providing necessary transition-state stabilization for DNA synthesis thereby reducing the activation energy and facilitating successful

nucleotidyltransfer onto a DNA primer strand. Thus far, the putative roles of the two divalent metal ions during the DNA polymerase-catalyzed reaction have been well-established empirically through biochemical and structural investigations.[1-3,8,24-26,30-37]

Notably, this proposed two-metal ion mechanism (Figure 1.1A) has been heralded as "a mechanism for all polymerases" and draws support from the fact that many crystal structures of DNA or RNA polymerases in complex with nucleic acid and incoming nucleotide (E•DNA•dNTP or E•RNA•rNTP, ternary complex) contain two divalent metal ions in the polymerase active site.[2,26] In a striking example of evolutionary conservation, DNA polymerases from all families have been characterized to follow the same two-metal ion mechanism based on mutational analysis, structural studies, and kinetic investigation.[2,8,25,33,34] However, recent time-resolved (also known as time-lapse, time-dependent, or soak-trigger-freeze) crystallographic studies of the Y- and X-family DNA polymerases, including human DNA polymerases η (hPolη)[38,39], β (hPolβ)[40-45], and μ (hPolμ)[46], have provided substantial evidence to compel an expansion of the two-metal ion mechanism to include a transient (*i.e.* not observed in all time-resolved partial reaction structures), third divalent metal ion, the precise role of which is currently debated (Figures 1.1B and 1.1C).[47,48]

In this review, we aim to briefly summarize the extensive evidence supporting the two-metal ion mechanism for DNA polymerization while highlighting the possibility of a third divalent metal ion and evaluating its involvement in catalysis as well as its biological purpose and significance. This will include a detailed synopsis of the seminal time-resolved X-ray crystallography findings over the last five years that have sparked

renewed interest in the metal ion mechanism including discussion about the evidence, timing, and dynamic nature of the third divalent metal ion. As a result of its transient character, there is some inconsistency with the time at which the third divalent metal ion appears during the reaction with some groups reporting its occupancy during nucleotidyltransfer[38,39,42,45] and others reporting its appearance only in the product complex.[40,41,44,46] Thus, it is unclear if the third divalent metal ion serves a role in transition-state stabilization (Figure 1.1B), product release, catalysis of the reverse reaction (*i.e.* pyrophosphorolysis, Figure 1.1C), or in modulating the chemical equilibrium of nucleotidyltransfer through product-state stabilization. A recent computational analysis of the third divalent metal ion with hPolη[49] supports roles in transition-state stabilization during the forward and reverse reactions. Similarly, our work with hPolβ[42,45] suggests a possible role in transition-state stabilization, while other structural and computational studies completed with hPolβ[40,41,44,50,51] and hPolμ[46] provide evidence for perturbation of the chemical equilibrium by inhibition of pyrophosphorolysis by the third divalent metal ion. Thus, the role of the third divalent metal ion is yet to be fully delineated, and may be unique for each polymerase or polymerase family. We will conclude with a short discussion of the implications that the third divalent metal ion has for the polymerase field including its potential role in the mechanisms of replicative polymerases or as a potential target for antiviral therapies.

1.3 DNA Polymerases and DNA Polymerization

For 5′–3′ nucleic acid synthesis, the terminal 3′-hydroxyl group of a DNA or RNA strand serves as a nucleophile to attack the α-phosphate of a dNTP or ribonucleotide

(rNTP) to form a phosphodiester bond while releasing pyrophosphate (PP_i) as a byproduct (Figure 1.1). In effect, a phosphodiester bond is transferred from the nucleotide to the nascent nucleic acid strand (*i.e.* nucleotidyltransfer). This reaction is catalyzed by enzymes termed DNA/RNA polymerases which bind both DNA/RNA and nucleotide substrates. As DNA and RNA polymerases share certain structural and functional similarities, much of the mechanistic discussion of DNA polymerases focused on in this review may also apply to RNA polymerases. However, for more detailed evaluations of RNA polymerase structure and mechanism, we point the interested readers to several insightful reviews.[52-56]

DNA polymerases take advantage of the specific shape and hydrogen bonding patterns of nucleobase pairs (*i.e.* A:T, G:C) to faithfully recognize and incorporate correct nucleotides during DNA synthesis.[23] In addition to following a conserved two-metal-ion mechanism for nucleotide incorporation (Figure 1.1A),[2,26] DNA polymerases of all families adopt a "right-hand" architecture (with the exception of the X-family members which are left-handed: hPolβ, hPolλ, hPolμ, and terminal deoxynucleotidyl transferase (TdT))[57] consisting of fingers, palm, and thumb subdomains (Figure 1.2).[8,33] Along with these core domains, DNA polymerases may possess auxiliary domains (Figure 1.2) which often help in the execution of a specific biological function. For example, i) members of the A- and B- families often demonstrate high base substitution fidelity during DNA synthesis partially due to their accessory 3′-5′ exonuclease domain, which removes the small number of incorrect nucleotides incorporated during DNA replication; ii) members of the X-family may contain a deoxyribophosphate lyase (dRPase) domain for processing

DNA ends during DNA repair; and iii) members of the Y-family contain a little finger subdomain (also named polymerase associated domain (PAD)) thought to serve a role in damaged DNA binding. These unique accessory subdomains and biochemical characteristics outfit polymerases from a particular family for a specific biological function.[19] Thus, the faithful, efficient, and processive A- and B-family polymerases perform the bulk of leading and lagging DNA strand replication.[58] In contrast, the error-prone and distributive X-family and Y-family DNA polymerases function in DNA repair and DNA damage response, respectively. Thus, the cell has evolved specialized DNA polymerases to perform an array of diverse functions and activities.[19]

To catalyze nucleotidyltransfer, DNA polymerases require divalent metal ion cofactors. The roles of these metal ions in catalysis were elucidated in early structures of the Klenow fragment of *E. coli* DNA polymerase I with single-stranded DNA and dTMP product bound to the 3′-5′ exonuclease domain.[30,31,37] As the enzymatic synthesis and decomposition of nucleic acid molecules are closely related processes, the two-metal-ion mechanism proposed for 3′-5′ exonuclease degradation was extended to DNA polymerization (Figure 1.1A). In the exonuclease active site, one divalent metal ion was coordinated by several carboxylate side chains of surrounding amino acids (Asp355, Glu357, and Asp501), a water molecule, and the 5′-phosphate of the primer terminus. An additional divalent metal ion was coordinated by Asp355, the 5′-phosphate of dTMP, and several water molecules. Through mutation of the coordinating residues to alanine it was determined that these metal ions serve distinct mechanistic roles. Interestingly, it was later discovered that the catalytic subunit of HIV-1 reverse transcriptase[59] (HIV-1 RT)

shares the same "right-hand" domain architecture of Klenow fragment with finger, palm, and thumb domains arranged to form the DNA binding cleft. Moreover, three conserved carboxylate amino acids identified in Klenow fragment were found in HIV-1 RT and their mutation to Ala also abolished catalytic activity.[59] Together, these data strongly supported the two-metal-ion mechanism for phosphoryltransfer reactions, including phosphodiester bond formation and degradation (Figure 1.1A).[60]

This two-metal-ion mechanism for DNA synthesis was later exemplified through the structure of rat DNA polymerase β, an X-family member, bound to primer-template DNA and dideoxy-terminated nucleotide (ddNTP).[61] From this structure, and those of rat DNA polymerase β bound to Mn^{2+} and dATP,[62] a common nucleotidyltransfer reaction mechanism involving two divalent metal ions for all DNA polymerases was postulated (Figure 1.1A).[61] Following polymerase binding at the primer-template junction of a DNA substrate, an incoming nucleotide is bound and positioned in the active site by i) Watson-Crick base-pairing with the templating base; ii) intermolecular contacts between the base, sugar, and phosphates with amino acid residues; and iii) coordination of two divalent metal ions by the three carboxylate residues. One metal ion binds between the primer terminal O3′ atom and the α-phosphate of the incoming dNTP and is often referred to as the A-site (M_A) or catalytic metal ion (Figure 1.1A). The second metal ion is coordinated by the incoming dNTP through the non-bridging oxygen atoms of the α-, β-, and γ-phosphates and is often referred to as the B-site or nucleotide binding metal ion (M_B) as its appearance coincides precisely with the binding of nucleotide (Figure 1.1A). During catalysis, M_A serves as a Lewis acid to lower the pK_a of the primer hydroxyl proton for

abstraction and subsequent in-line nucleophilic attack on the α-phosphate of the dNTP to form a pentacoordinated transition-state with the 3′-oxygen of the primer terminus and four oxygen atoms of the α-phosphate, including one from the PP_i leaving group, occupying each position of the trigonal bipyramid. On the other hand, M_B acts to orient the triphosphate moiety of the bound nucleotide for catalysis and destabilizes the ground state ternary complex of the polymerase to promote catalysis. Furthermore, following nucleophilic attack, M_B stabilizes the pentacoordinated transition-state and neutralizes the developing negative charge on the PP_i leaving group (Figure 1.1A).[2]

Support for the two-metal-ion mechanism of DNA polymerization exists for structurally characterized DNA polymerases from all major families including A,[63-66] B,[67-70] C,[71-74] X,[40,61,62,75-79] and Y[38,80-86] as demonstrated through ternary complex structures of enzyme, DNA, and dNTP with bound divalent metal ions (Figure 1.3). For example, the structures of bacteriophage T7 DNA polymerase[63] as well as *Thermus aquaticus* (*Taq*) DNA polymerase I[64] of the A-family, with DNA, ddNTP, and both divalent metal ions bound, were solved and are consistent with the earlier structural and mechanistic findings with rat DNA polymerase β.[61,62] Crystal structures of the replicative B-family DNA polymerases, including the bacteriophage polymerases T4[67] and RB69[68] also support the two-metal-ion mechanism. Consistently, two metal ions are also found in the active sites of repair and damage bypass DNA polymerases as demonstrated through crystal structures of rat DNA polymerase β,[36,37] discussed above, human DNA polymerase λ (hPolλ),[76] hPolβ,[75,78] and hPolμ[77] of the X-family, as well as hPolη[80] and *Sulfolobus solfataricus* DNA polymerase IV (Dpo4)[87] of the Y-family. Interestingly, RTs

also engage two divalent metal ions for catalysis as demonstrated by the ternary crystal structure of HIV-1 RT.[88] Limited structural evidence from the C- and D-families of DNA polymerases is available as these enzymes are under-represented in the protein data bank. However, a ternary complex crystal structure of a C-family polymerase from *Geobacillus kaustophilus*, as well as a lower resolution (4.6 Å) structure of *E. coli* PolIIIα, suggest a two-metal ion mechanism for nucleotide incorporation.[71,73]

1.4 Kinetic and Structural Mechanism of DNA Polymerases

Throughout the years, mechanistic studies of DNA polymerases from many diverse families, as well as reverse transcriptases, have culminated in a comprehensive kinetic pathway for nucleotide incorporation (Scheme 1.1A).[5,8-12,14,21-23,27,29,31,43,59,61,62,89-112] While particular details of this model may vary between DNA polymerases or systems (*i.e.* kinetically obligated removal or inclusion of elementary steps, see Scheme 1.1), we attest that the polymerase-catalyzed addition of correct nucleotides into a growing DNA primer strand occurs through ten steps (Scheme 1.1A). A DNA polymerase first binds a DNA substrate (Step 1, Scheme 1.1A) containing a primer-template junction to form the binary complex (E•DNA_n*, Scheme 1.1A), Initial DNA binding may place the terminal base pair of the DNA substrate within the polymerase active site (*i.e.* pre-insertion state) thereby occluding dNTP binding. However, DNA translocation (Step 2, Scheme 1.1A) by one nucleotide (E•DNA_n, Scheme 1.1A) to an insertion state correctly positions the templating base and creates the necessary space to bind an incoming dNTP in the subsequent step (Step 3, Scheme 1.1A). Notably, Step 3 includes the association of M_B and possibly M_A. Upon formation of this ground-state or loose ternary complex

(E•DNA_n•dNTP, Scheme 1.1A), many polymerases then undergo a conformational change (Step 4, Scheme 1.1A) of the finger subdomain (or the thumb subdomain for the X-family DNA polymerases) which encloses the newly-formed base pair of the templating nucleotide and the incoming dNTP to form the tight ternary complex (E′•DNA_n•dNTP, Scheme 1.1A). A second conformational change (Step 5, Scheme 1.1A) within the polymerase active site generates the activated ternary complex (E″•DNA_n•dNTP, Scheme 1.1A) wherein reactive groups, including divalent metal ions, catalytic carboxylate residues, 3′-OH of the primer strand, and the α-phosphate of the bound nucleotide, are properly aligned for subsequent nucleotidyltransfer (Step 6, Scheme 1.1A), conventionally referred to as the chemistry step, which extends the primer strand by one nucleotide (E″•DNA_{n+1}•PP_i, Scheme 1.1A). The nucleotide-binding induced conformational changes (Steps 4 and 5, Scheme 1.1A) are reversed in Steps 7 and 8 (Scheme 1.1A) before PP_i is released (Step 9, Scheme 1.1A) from the polymerase active site. Following the reverse conformational changes and PP_i dissociation, the polymerase may translocate by one base pair along the DNA (Step 10_a, Scheme 1.1A) for additional cycles of nucleotide incorporation (*i.e.* processive DNA synthesis) or may dissociate (Step 10_b, Scheme 1.1A) from the DNA substrate (*i.e.* distributive DNA synthesis).

1.4.1 DNA binding and associated polymerase dynamics.

The inclusion and order of the elementary steps in Scheme 1.1A are strongly supported by kinetic, structural, and/or biophysical evidence. Logically, DNA binding (Step 1, Scheme 1.1A) occurs before dNTP binding as the templating information

required for faithful replication is encoded in the DNA. This assertion is supported by inhibitor studies of nucleotide incorporation using PP_i,[24,113] as well as processivity assays, wherein DNA polymerases are observed to incorporate more than one nucleotide per DNA binding event.[90-92,94] In addition, ^{32}P-partioning experiments with *E. coli* Pol I indicated that the reaction followed a specific order in which the polymerase first associated with the DNA then bound dNTP.[114] Lastly, the relative affinity (*i.e.* K_d^{DNA}) of many polymerases for DNA is often in the sub-nanomolar concentration range, while binding affinities for correct or incorrect dNTPs (*i.e.* K_d^{dNTP}) often range from micromolar to millimolar concentrations. Accordingly, DNA polymerases likely spend disproportionately more time bound to DNA than to dNTP, increasing the likelihood of a strict order of substrate binding events. However, a recent structural and biochemical study of the X-family member hPolλ shows a preformed nucleotide binding pocket and reports relatively high affinities for dNTPs with a slight preference for dATP (3.3 μM for dATP and 15-45 μM for the other three dNTPs) in the absence of DNA. This suggests that the hPolλ may in fact bind Mg^{2+}-associated dNTP before DNA,[115] and helps to explain the higher base substitution frequency of hPolλ relative to hPolβ, a close X-family homolog.[109,116,117] While the ability of hPolλ to bind dNTPs prior to DNA is unusual, it has been structurally observed before[62,64,118-120] with the nucleotides often bound in a non-productive conformation. However, dNTP bound crystal structures of truncated hPolλ have shown productive binding of dNTP at the polymerase active site, but the global conformation of the protein has yet to reach the catalytically active state in the absence of DNA.[115] Similarly, dNTP bound crystal and solution NMR structures

(E•dNTP) of African swine fever virus (ASFV) Pol X, an X-family homolog, have revealed dGTP bound in a productive conformation that allows formation of *syn*-dGTP:dG Hoogsteen base pairs upon subsequent DNA binding,[118,121] which is different from the binding of nucleotide in multiple conformations shown for *Thermus thermophilus* Pol X.[122] Importantly, the structural results regarding this unique substrate binding order for ASFV Pol X have been confirmed by steady-state inhibition assays and nucleotide trapping assays.[123] Additionally, modeling based on chemical shift perturbations suggests that nucleotide binding to ASFV Pol X induces a conformational change in the absence of DNA, which further substantiates that dNTP binds first for this viral polymerase.[124] Nevertheless, the binding of nucleotide prior to DNA is likely a rare occurrence and may contribute to the low fidelity of Pol X.[125]

In addition, there may be certain scenarios in which the outcome of nucleotide incorporation is not influenced by the identity of the templating base. For example, a DNA polymerase may prefer to almost exclusively incorporate one particular nucleotide or catalyze template independent nucleotide incorporation such as the Y-family DNA polymerase Rev1, which is known to preferentially incorporate dCTP regardless of the templating base through a "protein template" arginine residue,[83,126-129] or the X-family DNA polymerase TdT, which prefers single-stranded DNA over double-stranded DNA and is incapable of replicating a DNA template,[23] respectively. Moreover, damage to the DNA may result in an unreadable templating base in which case a Y-family DNA polymerase is recruited for non-templated lesion bypass. Accordingly, while certain exceptions to the order of DNA and dNTP substrate binding to a DNA polymerase exist

for specific biological contexts or for specialized DNA polymerases, it is widely accepted that DNA binding typically occurs prior to dNTP binding.

DNA polymerases have been observed to bind to a variety of DNA substrates and the substrate specificity seems to depend on the polymerase family as well as the particular biological function. In general, it is understood that DNA polymerases bind primer-template DNA substrates wherein the 3′-end of the primer strand is recessed relative to the 5′-end of the template strand. However, specialized DNA polymerases such as hPolβ and hPolλ of the X-family prefer to act on gapped DNA substrates containing an upstream primer along with a 5′-phosphorylated or 5′-deoxyribophosphate adducted downstream primer,[79,116,117,130,131] and the Y-family polymerases can tolerate binding to and replicating on damage-containing DNA substrates,[132-139] unlike their replicative polymerase counterparts.[140-143] Several pre-steady-state kinetic assays exist to measure the equilibrium dissociation constant (*i.e.* K_d^{DNA}) for DNA binding by a polymerase including the active site titration. During the active site titration, a fixed amount of a DNA polymerase is titrated with varying amounts of a radiolabeled DNA substrate before being rapidly mixed with correct dNTP to initiate nucleotide incorporation.[144] A burst of product formation is observed at each DNA substrate concentration as dNTP is rapidly bound and incorporated by the pre-formed E•DNA (Scheme 1.1) complex. An important consideration for successful execution of the active site titration is that nucleotide binding and incorporation must be much faster than the binding equilibration of a polymerase and DNA (E + DNA ⇌ E•DNA), otherwise the burst of product formation will not be observed. The concentration of the E•DNA

complex is given by the amplitude of the burst phase which varies as a function of DNA substrate concentration. A quadratic binding equation can then be used to determine both the K_d^{DNA} as well as the active concentration of the polymerase. For example, the DNA binding affinity of the catalytic subunit (*i.e.* p261) of human DNA polymerase ε (hPolε) of the B-family, which is responsible for leading strand DNA replication, was measured to be 79 nM and the enzyme was determined to be only ~16% active.[91] Interestingly, the assay was later repeated with the holoenzyme of hPolε (*i.e.* p261, p59, p17, p12) and the binding affinity increased 2.4-fold to 33 nM, while the enzyme did not gain appreciable activity (18% active).[145] In addition, active site titrations have been useful in determining the DNA binding affinities of *Sulfolobus solfataricus* DNA polymerase B1 (PolB1)[92] and Dpo4[94] as both polymerase demonstrate clear burst phase kinetics. However, when a small or indeterminate burst phase is present, as observed for hPolβ,[110,146] the necessary conditions prescribed for an active site titration are not met (*i.e.* nucleotide binding and/or incorporation is not faster than the binding equilibration of a polymerase and DNA) and therefore a different method must be used to accurately measure DNA binding affinity. An alternative strategy to determine polymerase affinity to DNA involves measuring the microscopic rate constants of k_{off} and k_{on} comprising K_d^{DNA} (*i.e.* $k_{off}/k_{on} = K_d^{DNA}$) through assays designed to monitor the kinetics of polymerase dissociation from or association to a DNA substrate, respectively. Indeed, directly measured values for k_{off} and k_{on} often strongly agree with the measured K_d^{DNA} of a DNA polymerase and provide additional insight into mechanistic steps which may kinetically limit multiple rounds of DNA synthesis.[10,90-92,94,110,145,147,148]

Binding of a DNA substrate by a polymerase is often accompanied with conformational dynamics of the enzyme as well as nucleic acids (Figure 1.2). One striking example of protein dynamics upon DNA binding is demonstrated by *S. solfataricus* Dpo4 of the Y-family (Figure 1.4A). While in the apo state, a crystal structure of Dpo4 reveals that the auxiliary little finger domain interacts with the thumb domain and occupies the DNA binding cleft, thereby occluding the binding of a DNA substrate.[149] Consequently, a major 131° rotation and 1.7 Å translation of the little finger domain is observed in the binary complex crystal structure as the little finger breaks contact with the thumb and establishes new contacts with the finger to vacate the necessary space for DNA to bind (Figure 1.4A).[149] Additionally, high affinity of DNA binding[9,94,134] (K_d^{DNA} = 10 nM) is ensured as both the thumb and finger domains rotate 10° to better contact the DNA in the binary complex structure.[149] This dramatic structural transition was further investigated in fluorescence[149] and stopped-flow Förster resonance energy transfer (FRET) studies monitoring distance changes between[22] and within[10] individual Dpo4 domains. Contrary to the initial hypothesis that Dpo4 must follow an induced fit mechanism for DNA binding, wherein the DNA substrate would induce the conformational change of the little finger domain, it was found that in the absence of DNA, Dpo4 exists in conformational equilibrium between the structurally distinct apo and binary complex configurations (Figure 1.4A) and DNA binding selects for the DNA bound state (*i.e.* shifts conformational equilibrium toward DNA bound state). This was concluded as the rate of conformational transition between apo and binary complex upon DNA binding, monitored through relative distance change between interdomain FRET

probes positioned in the little finger and palm domains, was independent of DNA concentration over two orders of magnitude.[22] This assertion was supported by a nuclear magnetic resonance (NMR) study assigning the backbone chemical shifts of full length Dpo4 in the absence and presence of DNA, which suggested that a minor conformation of apo Dpo4 existed in a conformation consistent with that observed in the Dpo4 binary complex crystal structure.[150] Importantly, it was later shown that the little finger mediates initial DNA binding of Dpo4 through a stopped-flow FRET system reporting on intradomain distance changes of the little finger,[10] as predicted by a previous computational investigation.[151] A similar dramatic structural rearrangement is observed for the related Y-family member human DNA polymerase κ (hPolκ) during DNA binding as its little finger domain moves ~50 Å to intimately contact the DNA major groove and the N-clasp, a unique N-terminal extension, helps encircle the DNA substrate.[152] Conformational dynamics during DNA binding can also be observed for the X-family DNA polymerases hPolβ[61,62] and hPolλ,[76,115] but not hPolμ,[77,153] as the 8 kDa N-terminal dRPase domain of both hPolβ and hPolλ move to engage gapped-DNA substrates (Figure 1.2). Notably, subtle differences in the dRPase domain dynamics between hPolβ and hPolλ may help explain the higher affinity of hPolβ (0.077 – 22 nM)[146,154] for gapped-DNA relative to hPolλ (110 nM),[155] and its role as the primary polymerase for short-patch base excision repair (BER).[21] In contrast to the limited domain motion demonstrated by the X-family polymerases upon DNA binding, the gapped- (or nicked) DNA duplex undergoes a dramatic structural change involving a 90° kink occurring at the 5′-phosphodiester bond of the templating base.[79,153,156] Importantly, this unusual DNA

structure is necessary for thumb domain closure during single-nucleotide gap-filling DNA synthesis by hPolβ and therefore ensures that an important fidelity checkpoint is maintained during nucleotide incorporation. Examination of product complex structures from enzymes involved in BER reveals that the DNA becomes progressively bent as it is sequentially bound and processed by most enzymes of the DNA repair pathway (DNA glycosylase → AP endonuclease → X-family DNA polymerase).[157] Accordingly, DNA repair enzymes may recognize and preferentially bind the bent DNA to facilitate rapid and efficient repair of DNA damage. The higher-fidelity A- and B-family DNA polymerases have also been observed to undergo conformational dynamics upon DNA binding as exemplified by comparison of the apo[158-160] and binary or ternary[64,65,68,161] crystal structures of bacteriophage RB69 DNA polymerase, *Pyrococcus furiosus* DNA polymerase, and *Taq* DNA polymerase I (Figure 1.2). In general, beyond the occasional structuring of disordered regions, DNA binding is typically accompanied by movement of the thumb (or the fingers for the X-family DNA polymerases) domain towards the palm domain in order to wrap around the DNA substrate.

1.4.2 DNA translocation and divalent metal-ion binding to the A- and B-sites.

Following formation of the E•DNA binary complex (Step 1, Scheme 1.1A), nucleotide incorporation into the primer strand of the bound DNA substrate commences upon binding of a dNTP. Importantly, DNA polymerases in the binary complex may exist in non-productive or productive configurations depending on whether the polymerase active site is bound in the pre-insertion or insertion state, respectively. Indeed, a binary complex crystal structure of the Y-family member Dpo4 showed the polymerase in the

pre-insertion state, while a ternary complex structure revealed the polymerase to have translocated by one base pair along the DNA to the insertion state in order to accommodate the incoming correct dNTP.[162] This essential DNA translocation event was later validated and measured to be rapid (>150 s^{-1} at 20 °C) by stopped-flow[11,12] and single-molecule[9,163] FRET studies monitoring distance changes between a DNA substrate and various domains of Dpo4. Interestingly, the single-molecule FRET studies revealed that the polymerase dynamically fluctuates between the pre-insertion and insertion states on the DNA but exclusively populates the insertion state in the presence of correct nucleotide.[9,163] Similar repositioning of the DNA polymerase from the pre-insertion state to the insertion state via DNA translocation is proposed from crystal structures of the Klenow fragment of *Taq* DNA polymerase I[65] and the large fragment of DNA polymerase I from *Bacillus stearothermophilus*.[164] Indeed, as the polymerase transitions between these two states while bound to DNA (Step 2, Scheme 1.1), nucleotide may directly bind to the polymerase-DNA complex at the insertion site or may induce DNA translocation from the pre-insertion site depending on which state is favored at equilibrium.[112]

Once the polymerase has translocated along the DNA to the insertion state, nucleotide binding can commence (Step 3, Scheme 1.1). Binding of dNTP coincides with association of the A- and B-site divalent metal ions. Time-resolved crystallographic studies (Figure 1.5, see Section 1.5.1) of bacteriophage N4 RNA polymerase indicate that M_B binding occurs simultaneously with nucleotide binding and that M_A binding occurs shortly after.[165] This explicit order for divalent metal ion binding to the polymerase (*i.e.*

M_B followed by M_A) is supported by the crystal structure of DNA polymerase ι (hPolι) wherein M_B is clearly associated with the bound dNTP but M_A has yet to bind.[82] However, those authors suggest that M_A may not be necessary for nucleotidyltransfer, as abstraction of the 3′-hydroxyl proton may occur by an active site carboxylate (Glu 127) positioned unusually near to the primer terminus.[82] In addition to binding after the M_B-associated nucleotide, M_A binding is proposed to occur following the conformational change (Step 4, Scheme 1.1A) to the tight ternary complex, as association of M_B-dNTP, not M_A, is sufficient to elicit the conformational change.[112,166-168] While this implies a defined order for the binding of each metal ion during the kinetic mechanism of DNA polymerase-catalyzed nucleotide incorporation (Scheme 1.1A), explicit evidence for the defined sequence of events is lacking and M_A could associate or dissociate at different step(s) of the mechanism.[112] Nevertheless, time-resolved crystallographic studies of hPolβ[40-42,44,45] and hPolμ[46] (see Sections 1.5.4 and 1.5.5, respectively) have demonstrated that following nucleotide incorporation, M_A dissociates prior to M_B indicating that the relative affinity for the divalent metal ion at the A-site is weaker, and further suggesting that M_A likely associates after M_B.

1.4.3 Nucleotide binding, incorporation, and polymerase fidelity.

The apparent affinity of the DNA polymerase binary complex for dNTP (*i.e.* K_d^{dNTP}) and the maximum rate constant of single-nucleotide incorporation (*i.e.* k_{pol}) can be experimentally measured by pre-steady-state kinetic assays.[144] Briefly, DNA polymerase and DNA substrate can be pre-incubated under single-turnover reaction conditions (*i.e.* [E] >> [DNA]) before mixing with various concentrations of correct or

incorrect dNTP. Reactions are quenched at increasing amounts of time and the data are fit to a single-exponential equation ([product] = $A[1 - \exp(- k_{obs}t)]$) to obtain an observed rate constant (k_{obs}) at each concentration of dNTP. The k_{obs} values are then plotted as a function of dNTP concentration and fit to a hyperbolic equation (*i.e.* $k_{obs} = k_{pol}[\text{dNTP}]/(K_d^{dNTP} + [\text{dNTP}])$ to obtain the desired kinetic parameters.[144] Importantly, the measured values of k_{pol} and K_d^{dNTP} for all 16 possible nucleotide incorporations are extremely useful metrics of polymerase efficiency (k_{pol}/K_d^{dntp}), fidelity (calculated as $(k_{pol}/K_d^{dNTP})_{\text{incorrect}} / [((k_{pol}/K_d^{dNTP})_{\text{correct}} + (k_{pol}/K_d^{dNTP})_{\text{incorrect}}]$), and processivity. This latter metric can be calculated as the ratio of k_{pol} to the rate of DNA dissociation (k_{off}, see Section 1.4.1) and describes the average number of bases incorporated by the DNA polymerase during a single DNA binding event, which can be more than 1,500 as observed for highly-processive T7 DNA polymerase (bound to the processivity factor *E. coli* thioredoxin)[90] and human mitochondrial DNA polymerase γ holoenzyme,[169] or less than 20 as observed for the poorly-processive X-family member hPolβ[116] and the Y-family member Dpo4.[94] Importantly, processivity values often increase when the DNA polymerase is associated with accessory subunits or processivity factors (*e.g.* proliferating cell nuclear antigen (PCNA) and β-clamp).[90,108,169-171]

Some of the fastest and most faithful-DNA polymerases exhibit k_{pol} values >200 s^{-1} and bind correct nucleotide with relatively high affinity ($K_d^{dNTP} < 10\ \mu M$), while incorrect nucleotides are bound with ~10- to 100-fold lower affinities and are generally incorporated 100- to 10,000-fold more slowly. Consequently, high fidelity polymerases typically make only one error per ~1 x 10^6 incorporations.[112] For example, the p261

catalytic subunit of hPolε was shown to incorporate correct nucleotides at a rate of 219 – 275 s^{-1} with a high base substitution fidelity of 10^{-4} – 10^{-7} (*i.e.* one error per 10^4 – 10^7 incorporations).[172] More impressively, it was further shown that the 3′-5′ exonuclease activity of hPolε bolstered the overall *in vitro* polymerization fidelity to 10^{-6} – 10^{-11} (*i.e.* one error per 10^6 – 10^{11} incorporations), which unprecedentedly translates to 0.1 – 1 misincorporations per round of human genome replication.[172] This enhancement in overall *in vitro* polymerization fidelity afforded by the exonuclease activity was greater than that observed for the related B-family DNA polymerase PolB1, which demonstrated a two orders of magnitude improvement (*i.e.* 10^{-4} – 10^{-6} to 10^{-6} – 10^{-8}).[173]

For many years, the mechanism by which a DNA polymerase recognizes a mismatch and switches between polymerization and exonuclease modes in order to correct the mismatch remained unclear. It was thought that the exonuclease domain must proofread the nascent DNA for mistakes and, upon identification of a mismatch, must transfer the DNA duplex from the polymerase active site to the exonuclease active site.[95,174,175] This was hypothesized to be a dynamic conformational change between polymerization and editing modes of the polymerase as the distance between the active sites is relatively large (~60 Å *E. coli* Pol IIIα, ~40 Å for *Saccharomyces cerevisiae* Pol ε).[174] Nevertheless, a recent cryo-electron microscopy (cryo-EM) study has revealed the structural basis for mismatch correction by *E. coli* Pol IIIα.[176] Rather than serving an active role as a "proofreader", the exonuclease domain is actually passive, with a terminal mismatch causing the DNA substrate to fray (as supported by NMR analysis of the DNA duplex) resulting in a distorted DNA conformation.[176] Accordingly, the mismatch is essentially

self-correcting as the primer strand from the frayed DNA duplex travels ~55 Å to the exonuclease active site for passive nucleotide excision.[176] This passive mechanism of exonucleolytic cleavage is corroborated by biochemical studies of *S. cerevisiae* Pol ε wherein an extended β-hairpin loop motif, originally thought to serve an active role in mediating a switch between polymerization and editing modes, was shown to have no such effect.[177] Interestingly, relative to matched primer-template termini, the rate of primer extension from a mismatched terminus is slow relative to the rate of exonuclease excision allowing for efficient mismatch removal.[134,172,173,178,179] The structural basis for inefficient polymerization beyond a mismatch stems from a myriad of active site and DNA distortions that misalign reactive groups[180-184] even when the mismatch is several base pairs removed from the primer-template junction (*i.e.* mismatch position *n*-1 to *n*-4).[180] Moreover, binding of a correct nucleotide when the terminal base pair is a mismatch induces distinct structural alterations which ultimately deter nucleotidyltransfer.[184] Together, these structural determinants prevent misincorporations and subsequent extension and push the equilibrium to exonucleolytic removal of the errantly incorporated nucleotide. In contrast to high-fidelity DNA polymerases, moderate-fidelity DNA polymerases[185] such as the X-family members hPolβ and hPolλ lack exonuclease domains (Figure 1.2) and demonstrate poor base substitution fidelity on both non-gapped and gapped DNA substrates ranging from $10^{-2} – 10^{-5}$ (*i.e.* one error per $10^{2} – 10^{5}$ incorporations).[109,117,186,187] Similarly, the error-prone lesion-bypass Y-family DNA polymerases including hPolη[147,188] and Dpo4[93] also lack exonuclease domains

(Figure 1.2) and display comparably poor base substitution fidelities of 10^{-2} – 10^{-4} on undamaged DNA substrates.

1.4.4 Kinetic basis for polymerase fidelity and the rate-limiting step of single-nucleotide incorporation.

The mechanistic basis by which DNA polymerases achieve their remarkable base substitution fidelity has been thoroughly investigated over the years. Based on the seminal findings of Watson and Crick,[189] it was originally thought that DNA polymerases would achieve high base substitution fidelity from the distinct hydrogen bonding patterns between correct versus incorrect base pairs. However, it was quickly discovered that hydrogen bonding alone could not explain the large difference in efficiency between incorporation of correct and incorrect nucleotides.[190] It was later suggested that both the shape of the nascent base pair within the polymerase active site as well as hydrogen bonding contribute to nucleotide specificity.[191-196] Alternatively, it was hypothesized that the difference in free energy between the chemistry of correct versus incorrect nucleotide incorporation alone could explain polymerase fidelity.[6] However, it has been shown that for many polymerases in which the kinetic mechanism has been thoroughly investigated that chemistry is not the rate-limiting step of correct nucleotide incorporation.[5] Accordingly, research now indicates that many factors including but not limited to hydrogen bonding, free-energy differences, base-pair shape complementarity, and polymerase conformational dynamics contribute to high-fidelity DNA synthesis.[5,6,196-198] In fact, even non-catalytic accessory domains[199] as well as the solvent accessibility and water network of a polymerase active site[200] have been implicated or directly shown to

modulate polymerase fidelity. Furthermore, substrate dynamics have also been hypothesized to effect polymerase fidelity. For example, the rare tautomer hypothesis of polymerase fidelity postulates that replication errors occur at low frequencies due to the formation of high energy tautomers of DNA bases which allow incorrect base pairs to form Watson-Crick-like geometries and mislead the polymerase to catalyze a misincorporation. This hypothesis has gained recent support from crystal structures of the *Bacillus stearothermophilus* DNA polymerase I large fragment bound to a dC:dA mismatch[201] and a mutant of hPolλ bound to a dG:dT mismatch,[202] as well as through NMR spectroscopy of DNA duplexes containing site-specific mismatches, which revealed that sequence-dependent tautomerization and ionization of incoming nucleotides within the polymerase active site leads to misincorporations as originally suspected by Watson and Crick.[189,203,204] Moreover, DNA template dynamics associated with incorrect nucleotide incorporation have also been observed.[205] Lastly, DNA polymerases have been demonstrated to monitor base complementarity through sequence independent minor groove interactions.[206-208] Therefore, it is clear that polymerase fidelity is complex and is achieved through a vast array of polymerase and substrate interactions and dynamics.

1.4.4.1 A two-step binding model for DNA polymerase fidelity

Pre-steady-state kinetic studies coupled with pertinent crystal structures of DNA polymerases and RTs provided the first indication that enzyme conformational dynamics were important for the mechanism of DNA polymerization and polymerase fidelity.[4,5,29,61,63,65,68,88,90,209,210] Together, these studies helped define a two-step nucleotide binding mechanism (Scheme 1.1B) involving rapid equilibrium binding of

dNTP (Step 3_a and 3_b, Scheme 1.1B) followed by an open→closed conformational change of the finger (or the thumb for the X-family DNA polymerases) (Step 4_a and 4_b, Scheme 1.1B), supported by comparison of binary and ternary complex structures for many DNA polymerases (Figure 1.2). If rate-limiting, this conformational change would provide the additional selectivity crucial for discriminating against incorrect dNTPs.[27] For many polymerases, kinetic data from experiments studying the incorporation of a S_p-dNTPαS, a nucleotide analog in which the pro-S_p oxygen of the α-phosphate has been substituted with sulfur, indicated that a conformational change, rather than the chemistry of nucleotidyltransfer, was rate-limiting for single-nucleotide incorporation as the sulfur elemental effect (*i.e.* the decrease in the rate of nucleotide incorporation when using α-thio-dNTP versus normal dNTP) was negligible.[5] Briefly, as the A- and B-site metal ions do not interact with the pro-S_p oxygen of a dNTP, the substitution of this atom to sulfur allows kineticists to infer the identity of the rate-limiting step of nucleotidyltransfer from the magnitude of the change in the observed single-nucleotide incorporation rate. Accordingly, significant elemental effects of 4–11-fold (*i.e.* $k_{obs,\ \mathrm{dNTP}}/k_{obs,\ Sp\text{-}dNTP\alpha S}$) were previously considered to indicate that chemistry is rate-limiting for DNA polymerase catalysis, whereas smaller values (*i.e.* < 2-fold) suggest that pre-chemistry conformational changes are rate-limiting.[5,94,103,211] Notably, while a significant sulfur elemental effect was not often observed for correct nucleotide incorporation, experiments performed with incorrect nucleotide frequently revealed that the chemistry step (Step 5_b, Scheme 1.1B) was rate-determining, presumably as a result of misaligned reactive moieties within the polymerase active site.[5] It is important to mention that the sulfur

elemental effect is no longer considered a reliable diagnostic for the rate-limiting step of nucleotide incorporation as intermediate effects (*i.e.* 2–3-fold) are difficult to interpret, while large effects (*i.e.* >10-fold) are thought to arise from disruption of the geometry of the transition-state (*i.e.* steric effects) and therefore no longer exclusively report on the chemistry step.[5] Moreover, the pro-S_p oxygen of the α-phosphate of the incoming dNTP has been hypothesized to not serve a major role in transition-state stabilization and therefore its substitution with sulfur does not adequately probe the chemistry step.[6] More convincing than the sulfur elemental effect, were results obtained through the pulse-chase/pulse-quench experiment, wherein an increase in reaction amplitude of the pulse-chase compared to the pulse-quench is indicative of a rate-limiting pre-chemistry conformational change.[5,99] During the pulse-quench, a pre-incubated solution of polymerase and DNA is mixed with [α-^{32}P]-radiolabeled dNTP for varying amounts of time before quenching. The pulse-chase proceeds similarly, except that before quenching, an excess of cold dNTP is added to the reaction mixture. Accordingly, if a slow-to-form polymerase complex (E′•DNA$_n$•dNTP, Scheme 1.1B) accumulates before the chemistry step (Step 5$_a$ or 5$_b$, Scheme 1.1B) then the chase with excess cold dNTP should cause an increase in reaction amplitude as the reaction is chased forward (*i.e.* an [α-^{32}P]-dNTP-bound polymerase complex can form additional product rather than the [α-^{32}P]-dNTP dissociating out of the active site). Indeed, results of the pulse-chase/pulse-quench experiments for many DNA polymerases identified a rate-limiting pre-chemistry conformational change and supported the two-step binding mechanism for correct nucleotide incorporation.[4,5]

1.4.4.2 The open→closed conformational change is not rate-limiting

While the initial two-step model for nucleotide selection and incorporation was strongly supported by structural and kinetic data, it was later contested as the measured rate of the open→closed conformational transition (Step 4_a, Scheme 1.1B) for polymerases was too rapid to be considered rate-limiting.[7,27,212,213] As this model depended on the assumption that the rate of the open→closed conformational hange (Step 4 , Scheme 1.1A and 1B) must be slow relative to the chemistry step (Step 6, Scheme 1.1A; Step 5, Scheme 1.1B) to afford nucleotide selection specificity, a revised interpretation of the model was necessary.[4] Indeed, studies of T7 DNA polymerase,[27,112] RB69 DNA polymerase,[196,214] and HIV-1 RT[215-217] showed that the rates of the pre-chemistry forward and reverse conformational changes for correct (k_{4a} and k_{-4a}, respectively, Scheme 1.1B) or incorrect nucleotide (k_{4b} and k_{-4b}, respectively, Scheme 1.1B) relative to the rate of chemistry (k_{5a} or k_{5b}, Scheme 1.1B) defined nucleotide specificity. Accordingly, binding of the correct nucleotide rapidly induces a conformational change (*i.e.* large k_{4a}, Scheme 1.1B) to an enzyme complex committed to catalysis (*i.e.* small k_{-4a} relative to k_{5a}, Scheme 1.1B), while binding of the incorrect nucleotide induces a unique conformational change to an enzyme complex ($E^{‡}$•DNA_n•dNTP, Step 3, Scheme 1.1B) which allows rapid release of the incorrect nucleotide (*i.e.* large k_{-4b} relative to k_{5b}, Scheme 1.1B). Altogether, kinetic analysis revealed that the controversy concerning the relative magnitudes of the rates of the pre-chemistry conformational change (k_{4a} and k_{4b}, Scheme 1.1B) versus the chemistry step (k_{5a} and k_{5b}, Scheme 1.1B), and how this could impact nucleotide specificity, was

unfounded if the reverse rate of the pre-chemistry conformational change was slow for correct dNTP (k_{-4a}, Scheme 1.1B), but fast for incorrect dNTP (k_{-4b}, Scheme 1.1B), relative to the chemistry step (k_{5a} and k_{5b}, Scheme 1.1B).[27,112,215,216] This current model suggests that chemistry is fast relative to nucleotide release during correct nucleotide incorporation as the rapid conformational change, promoted by the correct geometry of the base pair within the ground-state ternary complex (E•DNA•dNTP, Scheme 1.1B), ensures the proper alignment of catalytic moieties (E′•DNA_n•dNTP, Scheme 1.1B). On the other hand, chemistry is slow during incorrect nucleotide incorporation as a unique conformational change, prompted by the incorrect geometry of the base pair within the ground-state ternary complex (E•DNA•dNTP, Scheme 1.1B), instigates improper alignment of catalytic groups ($E^{‡}$•DNA_n•dNTP, Scheme 1.1B) and dissociation of nucleotide. In other words, the reversal of the pre-chemistry conformational change is fast relative to chemistry during incorporation of a mismatch, thereby favoring rapid nucleotide dissociation from the polymerase active site prior to nucleotidyltransfer as demonstrated through studies of high-fidelity T7 DNA polymerase[27,112] and RB69 DNA polymerases,[187,203] as well as moderate-fidelity HIV-1 RT.[215-217]

1.4.4.3 Multiple mechanisms of DNA polymerase fidelity

While the latter model (see Section 1.4.4.2)[27,112,196] elegantly explains how high-fidelity DNA polymerases achieve their remarkable substrate specificity for correct nucleotide, it may not extend to the low-fidelity X- and Y-family DNA polymerases. In contrast to the large structural change upon nucleotide binding observed for many A- and B-family DNA polymerases, involving closure of the finger domain,[210] all members of

the Y-family[14,218] and some members of the X-family[57,115,153] do not undergo such a nucleotide-induced conformational change (Figures 1.2 and 1.4B). Despite the structural data, results from the sulfur elemental effect and pulse-chase/pulse-quench experiments for several Y-family DNA polymerases suggest that a pre-chemistry conformational change is rate-limiting for single-nucleotide incorporation.[14,94,103,104] Accordingly, the basis by which error-prone polymerases select for correct nucleotide may involve a distinct mechanism.

Strikingly, stopped-flow fluorescence[219] and FRET[10-12,22] studies of the Y-family polymerase Dpo4 have revealed subtle conformational motions between and within each subdomain (finger, palm, thumb, little finger) during binding and incorporation of a correct nucleotide (Figures 1.4B and 1.4C). While monitoring distance changes between each polymerase domain and the DNA substrate, based on the anti-correlated increases or decreases in the donor and acceptor fluorescent signals, three FRET phases were observed upon mixing the polymerase-DNA binary complex with correct dNTP corresponding to i) rapid DNA translocation by Dpo4 (P_0), ii) synchronized gripping of the DNA substrate by each domain prior to nucleotide incorporation (Figure 1.4B, P_1), and iii) subsequent relaxation of each domain following nucleotidyltransfer (Figure 1.4B, P_2). Interestingly, the slow FRET phase (P_2) vanished during analogous experiments performed using a DNA substrate containing a dideoxy-terminated primer to prevent nucleotide incorporation and therefore must occur following nucleotidyltransfer. The P_1 phase (~15.3 s^{-1}) occurred much faster than the rate-limiting step of single nucleotide incorporation measured by radioactive chemical quench (0.66 s^{-1}). If the rate of

synchronized domain motion (~15.3 s^{-1}) is considered the forward rate for enzyme isomerization (k_{4a}) in Scheme 1.1B, then the rate of chemistry (k_{5a}, Scheme 1.1B) can be calculated as 0.69 s^{-1} from the relationship $k_{5a} \approx k_{4a}k_{pol}/(k_{4a} - k_{pol})$,[27] where k_{pol} is the observed single-turnover rate for correct nucleotide incorporation (0.66 s^{-1}). In contrast to T7 DNA polymerase,[27,112] the forward isomerization rate (k_{4a}, Scheme 1.1B) for Dpo4 is much faster (22-fold) than the surprisingly slow calculated rate of chemistry (k_{5a}, Scheme 1.1B). Consequently, the reverse isomerization rate (k_{-4a}, Scheme 1.1B) must be much slower (0.0017 s^{-1} based on the 410-fold difference between forward and reverse isomerization rates measured for T7 DNA polymerase)[27] than 0.69 s^{-1} in order for Dpo4 to efficiently select the correct nucleotide according to the aforementioned revised model for nucleotide specificity (see Section 1.4.4.2 and Scheme 1.1B). However, given the clear lack of a sulfur elemental effect (1.4) and obvious increase in amplitude (5.5 nM) for the pulse-chase compared to the pulse-quench experiment measured for Dpo4,[94] we hesitate to assign the chemistry step (k_{5a}, Scheme 1.1B) such a slow rate (0.69 s^{-1}) when it appears that some other rate-limiting, pre-chemistry step clearly exists. Indeed, previous kinetic studies of Dpo4 at a range of temperatures (2 – 56 °C) provided four independent lines of kinetic evidence that a pre-chemistry protein conformational change must limit correct nucleotide incorporation.[94,220] Accordingly, nucleotide incorporation for Dpo4 likely proceeds through a mechanism (Scheme 1.1A) requiring two pre-chemistry conformational changes (Steps 4 and 5, Scheme 1.1A). The first conformational change (Step 4, Scheme 1.1A) involves the synchronized domain movements to enhance interaction with the DNA substrate (Figure 1.4B), while the second (Step 5, Scheme

1.1A) is rate-limiting and involves precise alignment of reactive groups achieved through subtle protein motions. Indeed, stopped-flow FRET experiments monitoring distance changes between and within individual domains of Dpo4 during nucleotide binding and incorporation garner support for this model (Figure 1.4C).[10,22] For example, the majority of intradomain FRET pairs demonstrated characteristic P_1 and P_2 phases (*i.e.* anti-correlated phases consistent with interdomain FRET pairs, Figure 1.4B)[11,22] and the P_2 phase was absent during experiments with a dideoxy-terminated primer (Figure 1.4D). However, intradomain FRET pairs positioned within the finger domain showed a unique P_2 phase regardless if nucleotide incorporation was prevented by utility of a dideoxy-terminated primer (Figure 1.4E). Importantly, the unique P_2 phases were in the same direction as P_1 (Figure 1.4E) and the rates of these P_2 phases were on the order of the rate-limiting step of single-nucleotide incorporation measured for Dpo4 by ^{32}P-based assays. We speculate that the observed P_2 phases reflect subtle, collective domain motions necessary to align reactive moieties around the nascent base pair in preparation for rapid nucleotidyltransfer. Consistent with experimental results, these motions should occur whether or not phosphodiester bond formation is prohibited by a terminating primer[10] and may be reflected in the fine adjustments of loops, secondary structural elements, and amino acid side chains near the nucleotide binding pocket as demonstrated through comparison of the binary and ternary crystal structures of Dpo4.[149] Experiments are currently underway to measure the reverse isomerization rate (k_{-4}, Scheme 1.1A) of Dpo4 in order to distinguish between the two competing mechanisms (see Sections 1.4.4.2 and 1.4.4.3) and identify how Dpo4 selects for the correct nucleotide. Future work will also

determine how an incorrect nucleotide may affect the conformational dynamics of Dpo4. Importantly, an additional mechanism for nucleotide specificity by Dpo4 is based on hydrogen-deuterium exchange experiments suggesting that the DNA translocation step may be involved in correct dNTP selection as certain protein motions, speculated to occur during DNA translocation, are only observable in the presence of correct nucleotide.[221] Thus, correct nucleotide binding may stabilize the insertion state relative to incorrect nucleotide by slowing down Dpo4 reverse transition to the pre-insertion state. Similar to the latter model (Scheme 1.1B) for polymerase fidelity (see Section 1.4.4.2),[27,112] this suggests that nucleotide specificity hinges on a reverse step (*i.e.* reverse translocation, k_{-2}) being slow for correct dNTP, but fast for incorrect dNTP, relative to nucleotidyltransfer (Step 5, Scheme 1.1B). This model is supported by single-molecule FRET studies of Dpo4 showing that the correct nucleotide stabilizes the insertion state to a greater extent than incorrect nucleotide.[9,163]

Taken together, it is clear that the mechanisms by which DNA polymerases attain nucleotide specificity are complex and may vary significantly among the polymerase families. As a result, an overarching or unified mechanism to explain these intricate processes for DNA polymerases is likely not possible and we caution that what may appear true for one polymerase may not extend to all. A clear example of polymerase dependent selection mechanisms comes from our recent structural characterization of the inherent *D*-stereoselectivity of several DNA polymerases.[87,222,223] Through structures of the Y-family DNA polymerase Dpo4[87] and the X-family DNA polymerases hPolβ[222] and hPolλ[223] bound to various nucleotide analogs with *L*-stereochemistry we identified several

unique mechanisms by which these polymerases achieve *D*-stereoselectivity. While it was unsurprising that the Y-family polymerase Dpo4 and the X-family polymerases would not have common mechanisms of *D*-stereoselectivity, it was unexpected that hPolβ and hPolλ, which share a high amount of sequence and structural homology, select against *L*-nucleotides in different ways.[87,222,223] Thus, these studies highlight the difficulties in generating a unified mechanism for any aspect of DNA polymerase catalysis and support the necessity to study each polymerase individually.

1.4.5 Post-chemistry steps of nucleotide incorporation.

Many biochemical, biophysical, and structural studies have aimed to deduce the kinetic mechanism and molecular bases for single-nucleotide incorporation and polymerase fidelity through characterization of the steps up to and including nucleotidyltransfer (Steps 1-6, Scheme 1.1A). However, post-chemistry steps involving the reverse (Steps 7 and 8, Scheme 1.1A) of the conformational changes observed during nucleotide binding and incorporation (Steps 4 and 5, Scheme 1.1A), as well as PP_i release (Step 9, Scheme 1.1A) have been seldom examined biochemically and/or structurally. Indeed, isolating post-chemistry events has proven to be difficult leading to the lack of sufficient structural and mechanistic characterization. As a consequence, the order in which PP_i release and the post-chemistry conformational changes occur as well as whether or not the events are cooperative (*i.e.* PP_i release triggers the reverse conformational change or vice versa) is unknown. However, recently the slow incorporation of nucleotide analogs, which closely resemble natural nucleotides but possess *L*-stereochemistry, has been utilized to capture *in crystallo* snapshots of post-

chemistry events by hPolβ.[43] In performing time-resolved X-ray crystallography (see Section 1.5.1) with these analogs, the order of events following the chemistry step were unambiguously defined. Interestingly, hPolβ completed the closed→open conformational changes (Steps 7 and 8, Scheme 1.1A) while the product PP_i remained bound to the polymerase active site. In fact, many of the side chain interactions with the PP_i were maintained despite the domain rearrangement, causing the PP_i to move with the thumb domain away from the incorporated nucleotide during the closed→open conformational transition. Presumably this reopening and movement of PP_i away from the reaction center would facilitate PP_i solvation and dissociation. Surprisingly, the third divalent metal ion previously identified in several time-resolved structural investigations[38-42] had already dissociated following domain reopening thereby directly opposing the hypothesis that the third divalent metal ion plays a role in PP_i dissociation (see Sections 1.2 and 1.5.3). Consistently, recent time-resolved X-ray crystallographic experiments with hPolμ in the presence of Mn^{2+} revealed that the third divalent metal ion dissociates prior to PP_i release and surprisingly showed that the B-site metal ion remains bound following PP_i release, rather than concomitantly dissociating with PP_i as previously purported.[46] Moreover, the post-chemistry structures of hPolβ demonstrate that the next correct nucleotide can bind to the open polymerase conformation to aid PP_i dissociation. This is not unexpected considering that if PP_i were to remain bound at the active site, then the incorporated nucleotide could be removed via pyrophosphorolysis. In this instance, one would expect concerted post-catalytic events including PP_i release, DNA translocation, and dNTP binding (*i.e.* Steps 9, 10_a, and 3, Scheme 1.1A). Altogether, rapid domain opening and the

active displacement of PP_i by the incoming nucleotide ensures forward reaction efficiency during processive DNA synthesis.

In a recent study, PP_i mimetic analogs were used to follow the reverse reaction by time-resolved crystallography (see Section 1.5.1).[224] Consistent with the abovementioned structural findings that PP_i dissociation occurs after opening of the thumb domain,[43] this study demonstrated that during pyrophosphorolysis, PP_i binds to the open form of hPolβ and an open→closed conformational change occurs prior to the reaction.[224] Moreover, structures inform that PP_i fails to support binding of catalytic Mg^{2+} at the A-site and is too far from the reaction center to promote efficient pyrophosphorolysis.[40] Consistently, biochemical and structural analyses with an imidodiphosphate PP_i analog demonstrated that a single atom change (*i.e.* bridging oxygen of PP_i substituted to nitrogen) allows optimal binding of catalytic M_A and positions the analog for efficient catalysis.[224] Interestingly, neither PP_i nor imidodiphosphate was efficient at removing mismatched primer termini, suggesting that pyrophosphorolysis does not act as a fidelity checkpoint during DNA synthesis.[224] Together, these studies have dissected the post-chemistry events of DNA polymerization and have shown that domain reopening occurs prior to PP_i release and the reverse reaction is highly disfavored.

1.5 New Paradigm for DNA Synthesis Catalyzed by DNA Polymerases

1.5.1 Time-resolved X-ray crystallography of DNA polymerase-catalyzed DNA synthesis.

The ability to follow an enzymatic reaction at atomic resolution has been sought after for many years by biochemists and structural biologists.[225] Static crystal structures

of complexes carefully designed to mimic reactant-, intermediate-, and product-states can, at best, only offer hints of the actual reaction mechanism. With the advent of time-resolved X-ray crystallography (Figure 1.5), much of the ambiguity that accompanies the interpretation of static crystal structures is replaced with clear insight into the chemical mechanism of a particular enzyme-catalyzed reaction. Generally, the technique involves preparation and isolation of a crystal containing an enzyme-substrate complex in a pre-catalytic state, followed by reaction initiation by transferring the crystal to a solution containing the reaction activator(s) and cryo-protectant (Figure 1.5A). Next, the reaction is allowed to proceed for a defined time interval before it is freeze-quenched by transferring the crystal to liquid N_2 (Figure 1.5A) for subsequent diffraction experiments. During generation of a structural model, the proportions (*i.e.* occupancies) of two or more states (*i.e.* reactant and product) are modeled and refined to fit the diffraction data (Figure 1.6). It is important to note that time-resolved crystallography is not a single-molecule technique. Rather, the average behavior of numerous molecules within the crystal, in the reactant- or product-states, contributes to the electron density at each time point. This process is repeated for several crystals, each allowed to react for an increasing amount of time. Finally, after solving the structure of the pre-catalytic complex (zero time point), the reaction progress from beginning to the end can be visualized, with each structure of a particular time point serving as a frame in the reaction film (Figure 1.7). For example, recent time-resolved crystallography with DNA polymerases (Figure 1.5) has been successful in following single-nucleotide incorporation[38-46] with an example of the modeling procedure applied during time-resolved crystallography of nucleotidyltransfer

by hPolβ shown in Figure 1.6 and example of the time-resolved snapshots captured by this technique depicted in Figure 1.7. Mechanistic events were able to be temporally resolved *in crystallo*, as reaction rates for single-nucleotide incorporation are observed to be 20- to 100-fold slower for hPolη[38] and hPolβ,[40] compared to rates measured by pre-steady-state kinetic studies of these enzymes in solution.[38,104,226] To obtain these structures, the non-catalytic divalent metal ion Ca^{2+} was exploited to form a stable pre-catalytic complex (Figure 1.5B). Ca^{2+} was then exchanged for the catalytic divalent metal ions Mg^{2+} or Mn^{2+} to start the reaction (Figure 1.5B). Other unique properties of DNA polymerases made implementation of time-resolved X-ray crystallography successful including the relative ease of crystallizing pre-catalytic complexes, the ability to achieve high resolution diffraction data, and the limited impact of conformational heterogeneity or dynamics on crystal integrity during reaction progression.[225] While still a relatively new method (*e.g.* the first reports for DNA polymerases appeared only ~five years ago), researchers have already enjoyed success in utilizing the time-resolved structural technique to uncover new details of structure and function relationships of DNA polymerases. In the coming years, we expect that more details of the DNA polymerase mechanism, extending beyond just the X- and Y-families, will be uncovered as more investigators adapt this powerful methodology.

1.5.2 A third divalent metal ion during nucleotide incorporation.

Despite hundreds of structures of DNA polymerases accumulated through years in the Protein Data Bank, no evidence for a divalent metal ion directly involved in DNA synthesis beyond the well documented M_A and M_B divalent metal ions had been

observed. However, the first time-resolved X-ray crystallographic investigation of hPolη,[38] as well as seven additional studies identify a transient, third divalent metal ion, referred to as the C-site (M_c) or product-associated metal ion, during phosphodiester bond formation (Figure 1.3).[39-46] Thus, traditional pre-catalytic substrate complexes and post-catalytic product complexes of DNA polymerases were not sufficient to structurally capture this apparently dynamic third divalent metal ion. High-resolution diffraction data (~1.5-2.0 Å) permitted for the unambiguous identification of metal ion electron density, coordination geometry, and metal ion-to-ligand coordination distances for M_C (Figure 1.8). Appropriately, M_C was shown to exhibit octahedral coordination geometry, consistent with a bound divalent metal ion, and short metal-to-ligand coordination distances (~2.2 Å), consistent with Mg^{2+} or Mn^{2+}, rather than non-catalytic Ca^{2+} (~2.4 Å). In some cases,[39-42] Mn^{2+} was used to initiate the *in crystallo* reaction, rather than Mg^{2+}, as its stronger signal (*i.e.* Mn^{2+} is more electron rich) allowed for confident assignment of M_C electron density even at 5σ levels (*i.e.* five standard deviations above background) (Figure 1.8D)[40,42] or resulted in anomalous diffraction.[40] M_C was shown to coordinate four water molecules as well as the non-bridging oxygen atom of the α-phosphate and the leaving oxygen atom of the β-phosphate (bridging oxygen between α- and β-phosphates) of the bound nucleotide (Figure 1.1B). In some instances, the number of ligands bound to M_C varied from four to six due to the dynamic nature of coordinating water molecules and presumably the transient nature of M_C. Importantly, these ligands fail to form any protein contacts, but only coordinate to the metal ion, therefore preventing any mutational confirmation of the existence or significance of M_C. Differences in the timing and

occupancy of the third divalent metal ion amongst the time-resolved studies of hPolη,[38,39] hPolβ,[40-42,44,45] and hPolμ[46] are suggestive of its dynamic nature. Furthermore, a third divalent metal ion only appeared with hPolμ when Mn^{2+} was used for metal ion exchange and not with Mg^{2+}.[46] As a consequence of these inconsistencies, the role of the third divalent metal ion in the mechanism of nucleotide incorporation has been highly debated (Figures 1.1B and 1.1C) with hypothesized roles in transition-state stabilization, product release, catalysis of pyrophosphorolysis, or product-state stabilization.[46,47]

1.5.3 Evidence and hypothesized roles for the third divalent metal ion in single-nucleotide incorporation catalyzed by hPol*η*.

In the inceptive time-resolved crystallography study, Nakamura *et al.*[38] followed nucleotide incorporation by hPolη and identified for the first time a third divalent metal ion utilized by a DNA polymerase during catalysis. M_C appeared midway through phosphodiester bond formation (140 s, 60% reactants, 40% products) and remained associated at the active site until the final recorded time point (230 s, 40% reactants, 60% product). Unfortunately, observation of full product and the subsequent release of PP_i was not observed as both the forward DNA synthesis and reverse pyrophosphorolysis reactions became competing at later time points as product occupancy decreased from the penultimate to the final time point. Furthermore, as the competing reactions were occurring simultaneously, it is likely that both activities were aided by M_C through transition-state stabilization and lowering the activation energy barrier for bond formation. Interestingly, appearance of M_C occurred concomitantly with the movement of a positively charged arginine residue (R61), which had flipped away from the α-

phosphate of the bound dNTP, effectively replacing the charge in the active site. This active site configuration led to the hypothesis that the M_C would support chemistry while the subsequent reverse conformational transition of the arginine, or an equivalent positively charged residue in other polymerases, back to its pre-catalytic configuration would act in concert with the third divalent metal ion to actively displace product PP_i. In this way, the side chain is essentially "sweeping out" PP_i from the active site in preparation for DNA translocation and an additional catalytic cycle. Consistently, a molecular dynamics investigation of the pre-catalytic and reaction state side chain conformations of R61 concluded that only the pre-catalytic side chain configuration facilitates nucleotide binding, suggesting that the absence of nucleotide precludes reaction-state configuration. Moreover, following nucleotidyltransfer the side chain must revert back to the pre-catalytic configuration before subsequent rounds of dNTP binding and incorporation.[49] Furthermore, the computational study maintained the notion that M_C, accompanied by the R61 conformational transition, serves as an exit shuttle for PP_i release, and dissociates along with PP_i.[49]

This mechanism of third divalent metal ion-dependent conformational transitions for nucleotide incorporation and product release may be conserved given that a lysine side chain in many polymerases from diverse families,[18,64,68,164,227,228] or an arginine residue in HIV reverse transcriptase,[88] is present at a position similar to R61 of hPolη. Alternatively, this positively charged side chain may be static and therefore occlude the binding of a third divalent metal ion and thus fulfil the roles of transition-state stabilization and/or PP_i release. Hence, it is necessary to investigate these other

polymerase families or RTs through similar methods to determine the precise role of these amino acids and identify whether a third divalent metal ion is utilized. Notably, the X-family repair polymerases, hPolβ, hPolλ, and hPolμ, which typically act on single-nucleotide gapped DNA substrates (see Section 1.4.1), do not possess an analogous positively charged residue to interact with the α-phosphate. This may be a mechanism by which processive DNA synthesis is suppressed to prevent potential misincorporations by these moderate-fidelity enzymes.[79,146-148] While the Y-family polymerases, such as hPolη, are considered low-fidelity enzymes[14,147,229-231] (see Section 1.4.3), the necessity to bypass DNA damage and subsequently extend the DNA primer, which are both difficult tasks for high-fidelity replicative polymerases, may justify the need for the positively charged side chain to aid in processivity, in contrast to the X-family polymerases.

More recently a time-resolved crystallography study of hPolη directly investigated the role of M_C and suggested that it is absolutely essential for catalysis and its binding may kinetically limit the rate of single-nucleotide incorporation.[39] When crystals of the hPolη pre-catalytic ternary complex (E•DNA•dNTP•Ca^{2+}) were soaked with 1 mM Mn^{2+} for varying amounts of time, only the A- and B-site metal ions were occupied by Mn^{2+} and no product formation was detected even after 1,800 s. However, after soaking the crystals in 10 mM Mn^{2+}, the A- and B- sites were readily occupied, while M_C appeared at 30 s coinciding precisely with the appearance of reaction product. The authors were able to determine the relative binding affinities for each metal ion from *in crystallo* metal ion titration experiments, where electron density for each metal ion was examined following crystal soaking at various metal ion concentrations and time points.

It was determined that Mn^{2+} binds to the C-site with an affinity of ~3.2 mM. The apparent weak binding of M_C was further supported by in-solution metal ion titration experiments, wherein the concentration of metal ion necessary to achieve half-maximal reaction rate was determined and yielded a similar affinity (2.7 mM). Importantly, the agreement of the metal ion binding affinities from the two approaches (*i.e.* in-solution and *in crystallo* metal ion titrations) is the first experimental evidence to suggest that M_C is bound by hPolη in solution and is not simply a crystal artifact of the time-resolved crystallography technique.

Examination of the interaction of hPolη with S_p-dATPαS demonstrated that A- and C- site metal ion binding is affected as a direct result of the substitution of the pro-S_p oxygen atom of dATP with a sulfur atom. As the atom at the S_p position is expected to coordinate M_C, the larger atomic radius of sulfur relative to oxygen disrupts the binding and therefore likely explains the observed rate reduction *in crystallo*. In fact, M_C is not observed at all despite product formation (50% product at 600 s), which suggests that M_C may not be absolutely essential for catalysis. However, the authors argue that M_C must be present but is too transient or low occupancy to be observed in the electron density.[39] In addition, mutation of active site residue (R61) of hPolη to alanine, resulted in delayed third divalent metal ion binding and misalignment of the bound dNTP relative to the primer 3′-OH. This is somewhat unexpected considering that R61 in the pre-catalytic ternary crystal structures occupies the space where M_C would bind following catalytic metal ion exchange. Based on this single pre-catalytic rotameric conformation, we expect that the exchange between pre- and post-catalytic side chain configurations of R61 is

slow and may partially limit the rate of M_C binding. Therefore, it is reasonable to hypothesize that mutation of R61 to alanine would facilitate more efficient binding of M_C. However, considering that R61 makes important contacts with both the α- and β-phosphates of the incoming dNTP, and the primer 3′-OH is misaligned relative to wild-type structures, their results may simply reflect the negative impact of the mutation on nucleotide binding as suggested by the significant increase (~2-10-fold) in K_m compared to the wild-type enzyme in steady-state kinetic experiments.[80]

Altogether, the authors use these data to suggest that chemistry (Step 6, Scheme 1.1A) is indeed rate-limiting for hPolη, which contrasts with previous kinetic evidence,[103,104] and that M_C binding may provide the free energy needed to overcome the activation energy barrier for nucleotidyltransfer (*i.e.* transition-state stabilization). However, it is important to consider that hPolη is first crystallized in a ternary complex in the presence of non-catalytic Ca^{2+} and therefore only the bond forming chemistry step is observed. Substrate binding (Step 3, Scheme 1.1A) and any associated conformational rate-limiting steps (Steps 4 and 5, Scheme 1.1A) have presumably already occurred. Thus, the reduction in reaction rate caused by the postponement or disruption to third divalent metal ion binding caused by the S_p-dATPαS or R61A mutation suggest that M_C binding limits the rate of the chemistry step, but likely is not rate-limiting for the entire kinetic pathway for single-nucleotide incorporation (Scheme 1.1A, see Section 1.4.4). Altogether, this work confirmed and measured the binding of M_C to hPolη at reasonable metal ion concentrations (in contrast to work with hPolβ, 200 mM Mg^{2+}/Mn^{2+})[40-42,45] and correlates the M_C binding affinities estimated crystallographically and in-solution. However, this

study falls short of unequivocally limiting the role of M_C to transition-state stabilization, as the temporal resolution afforded by the time-resolved crystallography technique[225] is not sufficient to distinguish the order of M_C binding and nucleotidyltransfer.

1.5.4 Evidence and hypothesized roles for the third divalent metal ion in single-nucleotide incorporation catalyzed by hPol*β*.

Following the seminal study with hPol*η*,[38] time-resolved crystallography was employed to visualize nucleotide incorporation by hPol*β* on a gapped DNA substrate.[40] Interestingly, a third divalent metal ion at the C-site was also observed for hPol*β*[40] during correct nucleotide incorporation in a similar position to the third divalent metal ion in hPol*η* (Figure 1.3).[38] However, in contrast to the results obtained for hPol*η*,[38] M_C only appeared in the product complex structures and coincided with the loss of M_A. This observation suggested that M_C is solely involved in post-chemistry events and perhaps the diverse polymerase families utilize the third divalent metal ion in distinct ways. In fact, the timing of M_C binding suggested a role for M_C in pyrophosphorolysis, wherein it would stabilize the attacking oxygen atom of PP_i following proton abstraction by a water molecule (Figure 1.1C). Consistent with this hypothesis, open nicked DNA binary complex (E•DNA) crystals soaked with Mg^{2+} and PP_i yielded structures of the polymerase in the closed conformation with M_C and PP_i bound to the active site (*i.e.* reactant-state for reverse reaction, pyrophosphorolysis). However, these complexes failed to initiate pyrophosphorolysis, presumably because M_A was a Na^+ ion rather than the catalytic Mg^{2+}. Interestingly, M_C was not observed during incorrect nucleotide incorporation.[40]

The role of M_C in pyrophosphorolysis was further investigated through use of quantum mechanical/molecular mechanical computational methods.[50] It was determined that M_C was beneficial in the initial stages of the chemical reaction (*i.e.* initiating the attack of Pα by Oβ), but became inhibitory as the two reacting atoms (Pα and Oβ) approached a distance of 2.3 Å, likely due to the strict coordination distances and geometry of M_C, thus effectively preventing the transition-state from forming.[50] However, replacement of Mg^{2+} with Na^+ at the C-site resulted in a lowered activation energy barrier, suggesting a mechanism where metal ions may exchange during the reaction pathway to favor reaction completion. In addition, the authors demonstrated that Mg^{2+} binding at the A-site is required for catalysis in either the forward or reverse (*i.e.* nucleotidyltransfer or pyrophosphorolysis) direction and rapid exchange with Na^+ following catalysis effectively pushes the reaction to completion. The inability for Mg^{2+} at the C-site to support the reverse reaction but the requirement of Mg^{2+} at the A-site for forward and reverse catalysis is consistent with the hypothesis that M_C may only be involved in post-chemistry events (*i.e.* PP_i release or conformational changes) for hPolβ and explains the inability of PP_i to support pyrophosphorolysis *in crystallo*.[43,50]

In subsequent time-resolved crystallographic investigations of hPolβ during faithful and unfaithful translesion DNA synthesis across from the major oxidative lesion 8-oxo-7,8-dihydro-2′-deoxyguanine (8-oxoG) (Figure 1.7),[42] M_C was observed in the reaction-state (Figures 1.8A and 1.8B) and post-chemistry (Figure 1.8C and 1.8D) structures with its occupancy similar or equivalent to the product-state occupancy. Notably, as the reaction progressed, M_C moved towards its final position wherein it was

fully coordinated with the reaction products and water molecules during bypass of 8-oxoG with dATP, highlighting the dynamic nature of the third divalent metal ion. This observation is similar to hPolη wherein M_C binding occurs prior to or immediately following reaction initiation,[38,39] and suggests that M_C may diffuse into the hPolβ active site following reaction initiation to associate with reaction intermediates as a means to stabilize the transition-state. This result was further supported in a follow-up time-resolved crystallographic study following hPolβ-catalyzed extension from 8-oxoG containing base pairs which were generated during the bypass[45] (*i.e.* dC:8-oxoG or dA:8-oxoG was the primer-template junction pair). Importantly, during this study,[45] M_C was observed as early as 15% product formation, suggesting an early role in catalysis (*i.e.* transition-state stabilization). Furthermore, the investigation of post-chemistry events for hPolβ-catalyzed nucleotide incorporation, revealed that M_C is not involved in PP_i release (see Section 1.4.5), thus limiting the third divalent metal ion to function in chemical events such as transition-state stabilization or product-state stabilization.

Additional time-resolved studies following the incorporation of 8-oxo-dGTP opposite a template dC or dA[41] or the extension from these incorporation products (*i.e.* 8-oxoG:dC or 8-oxoG:dA were the primer-template junction pair)[44] similarly demonstrated that M_C appears in reaction- and product-state structures. Unexpectedly, the third divalent metal ion appeared in the pre-catalytic structure for incorporation of 8-oxo-dGTP opposite dC presumably as a result of favorable M_C coordination facilitated by the optimal position of the O8 modification of the damaged nucleotide. The inclusion of a third coordinating ligand offered by the O8 atom, in addition to the non-bridging oxygen

atoms on the α- and β-phosphates, likely makes M_C binding more favorable in this structure.[41] These damage-specific interactions coupled with the absence of M_C in all other pre-catalytic structures from time-resolved crystallographic investigations of hPolβ suggest that pre-catalytic M_C binding is likely unique to this damage DNA context and therefore does not represent a common mechanistic feature. Nevertheless, despite the appearance of M_C in reaction intermediate structures, the authors argue that the third divalent metal ion is only involved in post-chemistry events and does not provide transition-state stabilization.[41,44]

To further investigate the role of M_C in the forward nucleotidyltransfer reaction a similar computational investigation as that completed for hPolβ-catalyzed pyrophosphorolysis[50] was performed.[51] As coordination of M_C by the bridging oxygen (O$\alpha\beta$) between the Pα and Pβ of the incoming nucleotide can only occur after the phosphodiester bond (*i.e.* Pα–O$\alpha\beta$) is broken, molecular dynamics was used to determine the position of the Mg^{2+} prior to nucleotidyltransfer. The calculated position of modeled pre-catalytic M_C was similar to that experimentally observed for M_C. However, the modeled pre-catalytic M_C is coordinated by the Pα pro-S_p oxygen of the incoming nucleotide and five water molecules, rather than the experimentally observed coordination by the Pα pro-S_p oxygen of the incorporated nucleotide, O$\alpha\beta$ of PP_i, and four water molecules. Within this system, the simulated activation energy barrier for nucleotidyltransfer was calculated to be 16.6 to 18.1 kcal/mol and, except for a slight repositioning of the product PP_i, very few differences in active site structure were observed relative to the time-resolved studies.[40-42,44,45] Interestingly, a two Mg^{2+} system

in which M_C was omitted and only M_A and M_B were used gave a very similar activation energy barrier of 17.5 to 18.6 kcal/mol, suggesting that M_C does not appreciably aid nucleotidyltransfer. Consistent with these computational predictions, the incorporation of a phosphorothioate nucleotide analog, S_p-dCTPαS, in which the sulfur substitution should ablate M_C binding, was only 3-fold slower than incorporation of dCTP.[146] Moreover, time-resolved crystallography of dCTPαS incorporation did not reveal the presence of a third divalent metal ion following nucleotide incorporation.[51] Taken together, these evidences suggest that the third divalent metal ion does not aid the forward reaction by significantly lowering the activation barrier,[50,51] which is in contrast to hPolη.[39] Alternatively, in a two Mg^{2+} (M_A and M_B) and one Na^+ (M_C) system the activation energy barrier is significantly lowered to 11.6 to 13.2 kcal/mol, suggesting a possible mechanism wherein a Na^+ is initially bound at the C-site to assist nucleotidyltransfer and is subsequently exchanged with Mg^{2+} following incorporation in order to prevent pyrophosphorolysis. This hypothesized metal ion exchange at the C-site is akin to the structurally observed exchange of the A-site Mg^{2+} for Na^+ following nucleotidyltransfer,[40-42,44,45] which also prevents pyrophosphorolysis. These metal ion exchanges within the hPolβ active site may act to favor nucleotidyltransfer while disfavoring pyrophosphorolysis.

While the first time-resolved structural study of hPolβ documenting a third divalent metal ion showed the appearance of M_C only after full product formation,[40] the latter four investigations showed M_C binding coinciding exactly with product formation,[41,42,44,45] which is consistent with the reports of the third divalent metal ion for

hPolη.[38,39] Accordingly, it is unclear for hPolβ if M_C binds prior to and supports nucleotidyltransfer, or binds following nucleotidyltransfer to stabilize the product complex (*i.e.* preventing pyrophosphorolysis).[50] Altogether, conflicting results from the time-resolved[40-42,44,45] and computational studies,[50,51] as well as the aforementioned investigation of hPolβ post-chemistry events[43] (see Section 1.4.5) limits the potential roles of M_C to either transition-state stabilization during nucleotidyltransfer (Figure 1.1B) or suppression of pyrophosphorolysis through stabilization of the product complex. However, similar concentration dependent divalent metal ion soaks to those with hPolη[39] to determine the effect of M_C on enzymatic rate are necessary to experimentally demonstrate whether or not M_C aids nucleotidyltransfer through transition-state stabilization.

1.5.5 Evidence and hypothesized roles for the third divalent metal ion in single-nucleotide incorporation catalyzed by hPolμ.

Similar to studies with hPolβ,[40-45] time-resolved crystallography was used to follow nucleotide incorporation into single-nucleotide gapped DNA by hPolμ,[46] wherein the pre-catalytic ternary complex formed in the presence of non-catalytic Ca^{2+} was soaked with either Mg^{2+} or Mn^{2+} to initiate metal ion exchange and catalysis. In contrast to the time-resolved structural findings with hPolβ[40-42,44,45] and hPolη,[38,39] a third divalent metal ion bound at the C-site could not be observed with Mg^{2+} even after extensive soaking at a high concentration of Mg^{2+} (100 mM) and despite full product formation. As expected, the Mg^{2+} bound at the A-site was eventually replaced by Na^+ at longer time points showing complete product formation. On the other hand, soaking with Mn^{2+}

resulted in appearance of M_C at time points coincident with 40% product formation and beyond, as well as sustained presence of Mn^{2+} at the A- and B-sites at every time point therefore suggesting that Mn^{2+} may be the physiological catalytic divalent metal ion for hPolμ and perhaps other DNA polymerases that exhibit low activity with Mg^{2+}. Notably, the position of M_C and the coordination of M_C by the reaction products, as well as the timing of M_C, are consistent with that observed for hPolβ[40-45] and hPolη.[38,39] Based on the computational work with hPolβ,[50,51] occupancy of the A-site by a divalent metal ion is essential for pyrophosphorolysis whereas the presence of M_C is inhibitory (see Section 1.5.4). As hPolβ and hPolμ share significant structural similarity, the roles of M_A and M_C for the reverse reaction may be conserved between these two polymerases. Accordingly, by analogy to hPolβ (see Section 1.5.4), the A-site metal ion exchange to Na^+ in the presence of Mg^{2+} observed with hPolμ likely precludes pyrophosphorolysis and therefore a third divalent metal ion may not be necessary to suppress this reaction when Mg^{2+} is supplied as the catalytic metal ion. However, the persistence of M_A in the presence of Mn^{2+} necessitates binding of M_C to prevent pyrophosphorolysis.

Single-nucleotide incorporation experiments with S_p-dTTPαS (see Section 1.4.4.1) suggest that M_C could serve a role in nucleotidyltransfer as experiments in the presence of Mg^{2+}, wherein M_C should not be bound, demonstrated a strong elemental effect, whereas this effect was lost in the presence of Mn^{2+} and presumably M_C.[46] However, in their publication, the authors suggest that absence of an elemental effect with Mn^{2+} was not due to binding of M_C as i) C-site divalent metal ion binding occurred following nucleotidyltransfer in the time-dependent structures (*i.e.* occupancy

corresponded exactly with product accumulation), ii) Mn^{2+} is generally considered thio-phobic,[232] and iii) the sulfur substitution of the pro-S_p oxygen would likely disrupt M_C coordination. Thus, similar to their work with hPolβ (see Section 1.5.4),[51] the authors suggest that M_C is not involved in transition-state stabilization during nucleotidyltransfer but rather serves to stabilize the product state to prevent pyrophosphorolysis. Nevertheless, the simultaneous appearance of M_C and reaction products could just as easily imply that the third divalent metal ion is critical for nucleotidyltransfer. Moreover, the expected disruption of M_C binding by the longer P–S bond distance would likely be alleviated by the flexibility of the other coordinating ligands, as four of six are water molecules. Fittingly, it is possible that M_C provides the necessary transition-state stabilization to accelerate the chemistry step in the presence of Mn^{2+} thereby eliminating the elemental effect. In this scenario, the results would be consistent with those of hPolη (see Section 1.5.3).[39]

Similar to our recent investigation of hPolβ-catalyzed post-chemistry events,[43] during time-dependent crystallography of hPolμ, M_C was also observed to dissociate before PP_i.[46] While hPolβ displays a large open→closed conformational change of the thumb domain (Figure 1.2) during nucleotide binding (see Section 1.4.4), such a large change is not observed for hPolμ, which may explain why time-resolved structural capture of the order of M_C and PP_i dissociation was difficult for hPolβ (*i.e.* rapid domain motion of hPolβ results in loss of synchronization of *in crystallo* events and associated electron density, see Section 1.4.5), but readily possible for hPolμ. Altogether, the time-resolved structural study of hPolμ,[46] featuring the third divalent metal ion, parallel those of

hPolβ[40-45] and suggest a conserved role for M_C during X-family polymerase-catalyzed DNA synthesis. However, more work is needed to explicitly delineate the mechanistic function of M_C in transition-state stabilization or preventing pyrophosphorolysis.

1.5.6 Future characterization of the third divalent metal ion.

The role of M_C in single-nucleotide incorporation is not well-defined with compelling evidence to support its involvement in i) stabilizing the transition-state of nucleotidyltransfer, ii) supporting pyrophosphate release, and/or iii) promoting or suppressing pyrophosphorolysis. It is clear from the limited work investigating the third divalent metal ion that its function in DNA polymerase catalysis is complex and may differ between the X- and Y-family polymerases. For example, the complementary time-resolved crystallographic[40-45] and computational studies[50,51] completed for hPolβ (see Section 1.5.4) support the proposed roles of M_C and provide an argument for analogous functions in hPolμ (see Section 1.5.5).[46] However, a potential role in transition-state stabilization for M_C in the X-family DNA polymerases cannot be completely ruled out. Conversely, for hPolη, it has been proposed that M_C is directly involved in transition-state stabilization during nucleotide incorporation.[39,198] Nevertheless, additional work must be completed with other Y-family polymerases to validate the proposed roles of M_C suggested for hPolη (see Section 1.5.3) and to determine if M_C function is conserved for the Y-family polymerases. In addition, computational investigations, such as those performed for hPolβ,[50,51] must be undertaken for hPolμ and hPolη to better substantiate the proposed roles of M_C in these polymerases. Similarly, concentration dependent metal ion soaking as performed for hPolη[39] must be performed for hPolβ and hPolμ to

determine if M_C also acts in transition-state stabilization for these polymerases as argued for hPolη,[39,198] especially considering that high metal ion concentrations of ~200 mM were used for the studies of hPolβ[40-45] and could negatively affect polymerase activity.[93,109,233,234]

Furthermore, the apparent differences between how M_C is utilized between the X- and Y-family polymerases advocates for future research on the A- or B-family replicative polymerases as well as RTs to determine if a third divalent metal ion is used at all, and if so, what apparent role does it serve, and how does this compare to results of hPolη,[38,39] hPolβ,[40-45] and hPolμ.[46] Importantly, if a third divalent metal ion is observed for viral DNA polymerases or RTs and serves a purpose in catalysis (*i.e.* transition-state stabilization as with hPolη), then it may be a potential therapeutic target. For example, the active site of HIV-1 RT is very comparable to that of hPolη and contains an equivalent arginine residue (see Section 1.5.3) that may function similarly with the third divalent metal ion to facilitate nucleotidyltransfer and pyrophosphorolysis (Figure 1.3). As HIV-1 RT is known to remove chain-terminating nucleotide analogs by pyrophosphorolysis,[235] design of antiviral small molecules to specifically block the third divalent metal ion binding may prove to be an effective treatment strategy. Finally, as it stands, the only experimental evidence for the third divalent metal ion comes from X-ray structures capturing *in crystallo* reaction progression. Accordingly, it is possible that these findings may represent an artifact of the structural technique. Therefore, we expect that advanced spectroscopic methods such as electroparamagnetic resonance (EPR) spectroscopy,[236] will be necessary to fully validate and further elucidate the function of

the third divalent metal ion during polymerase catalysis under a more physiological context.

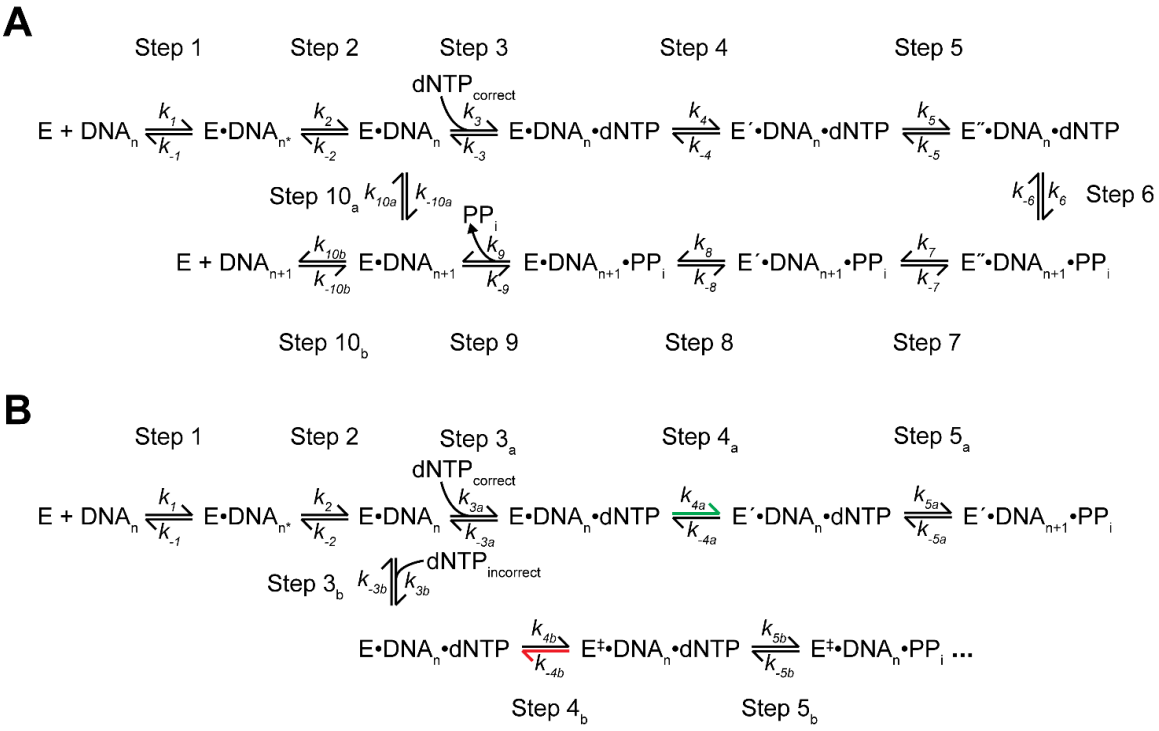
1.6 Concluding Remarks

Despite thousands of published studies investigating the structure and mechanism of DNA polymerases and RTs, it is abundantly clear that there is so much more to learn. Indeed, as the kinetics and conformational dynamics of each step of the DNA synthesis mechanism have been rigorously investigated for many polymerases from all of the diverse families, it is becoming evident that a singular, unified mechanism to describe every unique aspect of polymerase catalysis, including polymerase fidelity, is unrealistic. Thus far, it is evident that conformational dynamics differentially impact various aspects of the catalytic and kinetic mechanism between DNA polymerase and RT families, or even within a family. In the coming years, research to better understand the contributions of polymerase conformational dynamics during DNA binding and translocation, nucleotide binding, selectivity, and incorporation, pyrophosphate binding, and pyrophosphorolysis to the mechanism of DNA polymerization will be paramount.

The skillful application of time-resolved X-ray crystallography to study DNA polymerases has enabled the discovery of a third divalent metal ion during single-nucleotide incorporation. Remarkably, this third divalent metal ion may be important for DNA polymerase and RT catalysis which shifts the long-standing paradigm of two-metal-ion catalysis for DNA polymerization. We are excited at the prospect of identifying the third divalent metal ion in other families of DNA polymerases and RTs to determine if a three-metal-ion mechanism is conserved for DNA synthesis. To date, the third

divalent metal ion has only been captured in two X-family members and one Y-family member, but the significant differences in the proposed function of the third divalent metal ion, already apparent between these two families, implore future research of other polymerase families in hopes of exploiting potential drug targets for developing novel antiviral and antibiotic small molecule therapeutics.

1.7 Schemes



Scheme 1.1: Minimal kinetic mechanisms for nucleotide incorporation.

(A) Kinetic mechanism of nucleotide binding and incorporation with E, E′, and E″ representing different conformations of the DNA polymerase with Step 5 representing an essential, rate-limiting conformational change. (B) Alternative kinetic mechanism wherein incorrect nucleotide is selected against by binding in a unique DNA polymerase conformation designated by $E^{‡}$. Steps 4_a and 5_a occur during correct nucleotide incorporation. Steps 4_b and 5_b occur during incorrect nucleotide incorporation. The green arrow in Step 4_a signifies that the forward rate is highly favored in the presence of correct nucleotide, where E and E′ represent a conformational change upon nucleotide binding. (continued)

Scheme 1.1: continued

In the bottom branch, the red arrow in Step 4_b indicates that the reverse rate is highly favored in the presence of incorrect nucleotide. Following Step 5 the mechanism proceeds as in (A) for both correct and incorrect nucleotides. For (A) and (B) DNA_n^* signifies that the polymerase is bound to the DNA at the pre-insertion site (*i.e.* pre-translocated state).

1.8 Figures

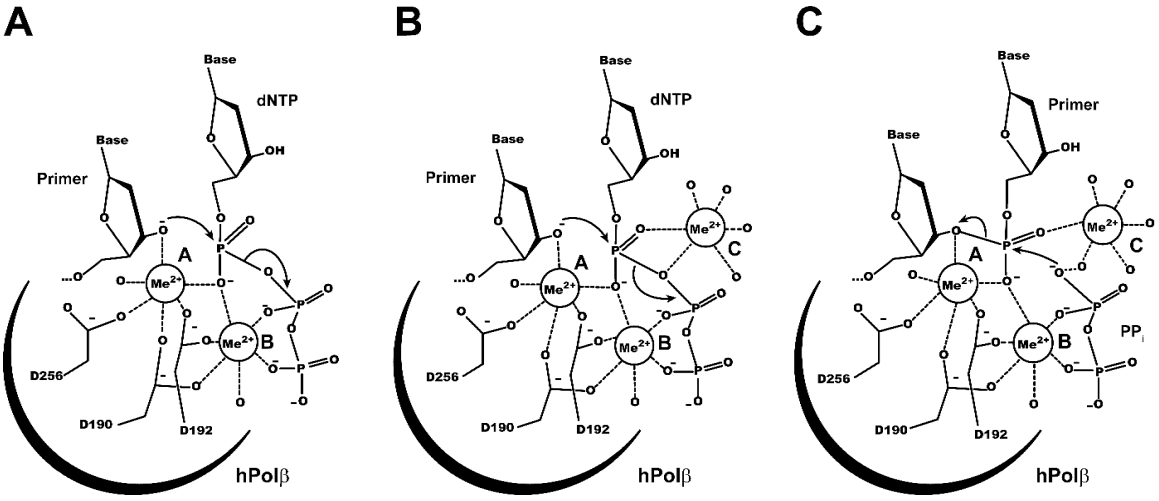


Figure 1.1: Two- vs. three-metal-ion mechanism for DNA polymerase-catalyzed nucleotidyltransfer and third-metal-ion assisted pyrophosphorolysis.

The active site of the well-studied hPolβ was selected to depict the metal ion-based chemical mechanisms. (A) Two-metal-ion mechanism. The 3′-OH of the primer is activated (*i.e.* deprotonated) for an in-line nucleophilic attack on the α-phosphate of the incoming dNTP. The α-phosphate is coordinated by two divalent metal ions (Me^{2+}). The catalytic metal ion at the A-site is also coordinated by the 3′-OH of the primer, active site carboxylate groups (Asp 190, 192, and 256), and a water molecule. The metal ion at the B-site is coordinated by active site carboxylates (Asp 190 and 192), a water molecule, and non-bridging oxygen atoms of the β- and γ-phosphates, to complete the α,β,γ-tridentate coordination of the dNTP. The A-site ion is suggested to activate the primer 3′-

(continued)

Figure 1.1: continued

OH nucleophile and the B-site ion stabilizes the negative charge of the pentacoordinated transition state. (B) Three-metal-ion mechanism. The reaction proceeds as in (A) except that a third divalent metal ion at the C-site appears to perhaps stabilize the transition state, serve as counter-ion to the oxyanion of the PP_i leaving group to aid product release, or participate in the reverse reaction, pyrophosphorolysis. The C-site ion is coordinated by water molecules as well as non-bridging oxygen atom of the α-phosphate and the bridging oxygen between α- and β-phosphates. (C) Third-metal-ion assisted pyrophosphorolysis. The third divalent metal ion may serve a similar role as the A-site metal ion in (A) and (B) to assist in the deprotonation and subsequent stabilization of the O_1 of PP_i. This atom would then attack the nascent phosphodiester bond of the DNA backbone, and the primer 3′-hydroxyl would be protonated to restore the pre-catalytic active site of nucleotide incorporation.

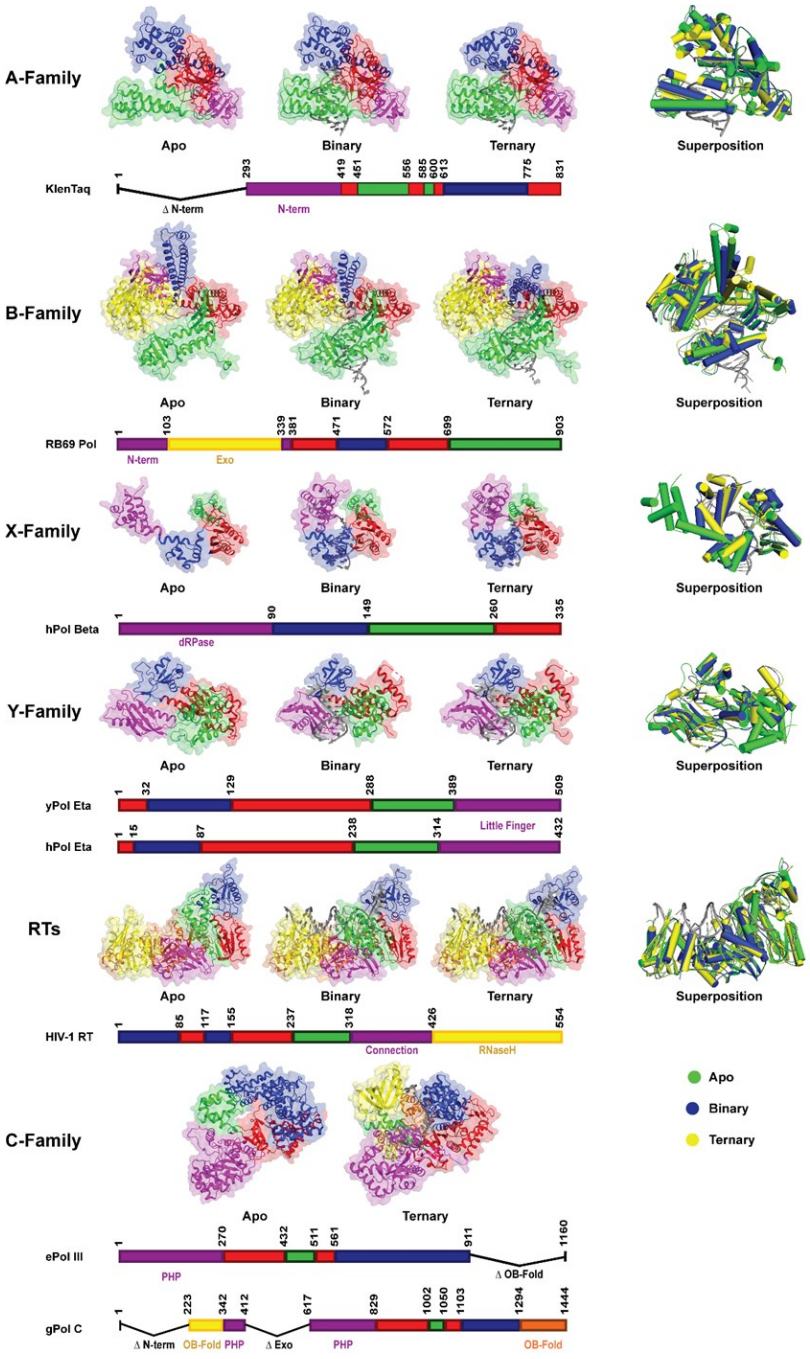


Figure 1.2: Structural comparison of DNA polymerase families.

(continued)

Figure 1.2: continued

binary (DNA bound, E•DNA), and ternary (DNA and nucleotide bound, E•DNA•dNTP) enzyme forms of representative polymerases from each family and superposition of all three forms. The Klenow fragment of Taq DNA polymerase I (KlenTaq) was used for A-family (1KTQ, 4KTQ, and 3KTQ), RB69 DNA polymerase (RB69 Pol) was used for B-family (1IH7, 2P5O, and 3NCI), rat DNA polymerase β (rPol Beta, apo) and human DNA polymerase β (hPol Beta, binary and ternary) were used for X-family (1BPD, 1BPX, and 4KLG), yeast DNA polymerase η (yPol Eta, apo) and human DNA polymerase η (hPol Eta, binary and ternary) were used for Y-family (1JIH, 3TQ1, 4ECX), HIV-1 reverse transcriptase (HIV-1 RT) was used for RTs (1DLO, 3KJV, 3KK2), and E. coli DNA polymerase III (ePol III, apo) and Geobacillus kaustophilus PolC (gPol C, ternary) were used for C-family (4JOM and 3F2D). Each structure is shown as cartoon with transparent surface rendering and individual domains colored. For all structures the thumb, palm, and finger domains are green, red, and blue, respectively. Accessory domains are uniquely colored and named in the associated line diagrams. For the binary and ternary structures, the DNA is shown as gray cartoon. In the ternary structures, the nucleotide is omitted for clarity. The superpositions are shown with cylindrical helices for simplicity of comparison with apo, binary, and ternary structures colored green, blue, and yellow, respectively.

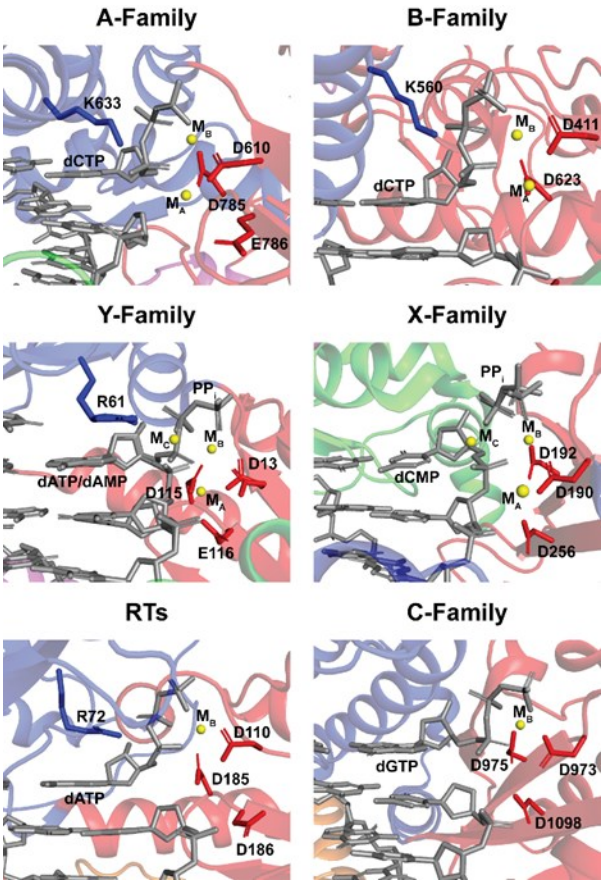


Figure 1.3: Active site comparison of DNA polymerases.

Zoomed views of ternary structures of representative DNA polymerases from the A-family (KlenTaq, 3KTQ), B-family (RB69 DNA polymerase, 3NCI), the X-family (hPolβ, 4KLG), the Y-family (hPolη, 4ECX), the RTs (HIV-1 RT, 3KK2), and the C-family (Geobacillus kaustophilus PolC, 3F2D). The incoming/incorporated nucleotide (dNTP/dNMP+PPi), DNA primer 3′-nucleotide, and active site carboxylates are shown as sticks. Metal ions bound at the active site are shown as yellow spheres. Importantly, in addition to the typical A- and B-site metal ions (MA and MB), the X- and Y-family

(continued)

Figure 1.3: continued

structures have a third divalent metal ion bound (MC). Many polymerases have positively charged residue side chains in the area where a third metal ion may bind and are shown as sticks in blue. Notably, Geobacillus kaustophilus PolC does not have a positively charged residue in this location.

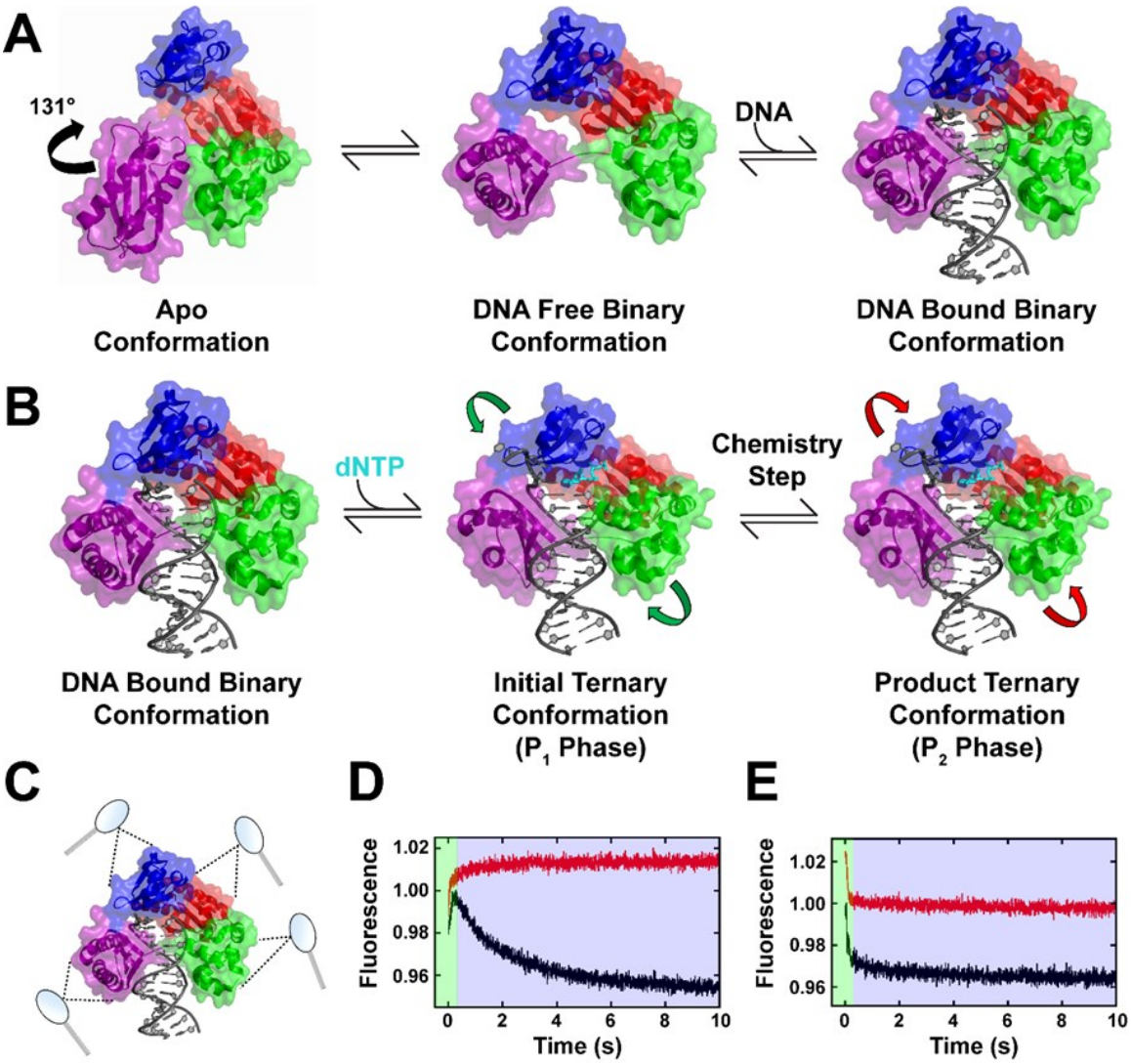


Figure 1.4: Conformational dynamics of Dpo4.

Finger, palm, thumb, and little finger are colored blue, red, green, and purple, respectively. (A) Dynamics of DNA binding. Dpo4 exists in equilibrium between the apo and binary conformations in the absence of DNA. DNA binds to the free binary conformation which may be mediated by the little finger domain.[10,13] (B) Dynamics (continued)

Figure 1.4: continued

during nucleotide binding and incorporation. In contrast to A-, B-, and some X-family DNA polymerases, comparison of binary and ternary crystal structures of Dpo4 demonstrates a lack of significant nucleotide binding associated protein dynamics. However, stopped-flow FRET analyses have uncovered subtle motions for each domain of Dpo4. Green arrows indicate the concerted movement of domains upon nucleotide binding during P_1 phase to grip the DNA substrate. Red arrows depict the relaxation of domains during the P_2 phase (*i.e.* opposite direction of P_1) following nucleotide incorporation.[10-13] (C) Pictorial representation of the intradomain FRET approach to investigate Dpo4 conformational dynamics within each domain (represented by magnifying glasses). Trp residues were site-specifically introduced into each domain to serve as FRET donors, while Cys residues modified with 7-diethylamino-3-(4'-maleimidylphenyl)-4-methylcoumarin) were site-specifically introduced into each domain to serve as FRET acceptors.[10] (D) Stopped-flow trace of little finger intradomain FRET construct (Y274W-K329C^{CPM}). Black trace shows correct nucleotide binding and incorporation on a natural DNA primer and demonstrates characteristic, anti-correlated P_1 (green shaded area) and P_2 (blue shaded area) phases. Red trace shows correct nucleotide binding with a dideoxy-terminated primer.[10] (E) Stopped-flow trace of finger intradomain FRET construct (S22W-K56C^{CPM}) colored as in (D). Note the similar direction of P_1 and P_2 phases regardless of natural or dideoxy-terminated primer.[10] PDBs 2RDI and 2RDJ were used to generate the structural figures in (A), (B), and (C).[149]

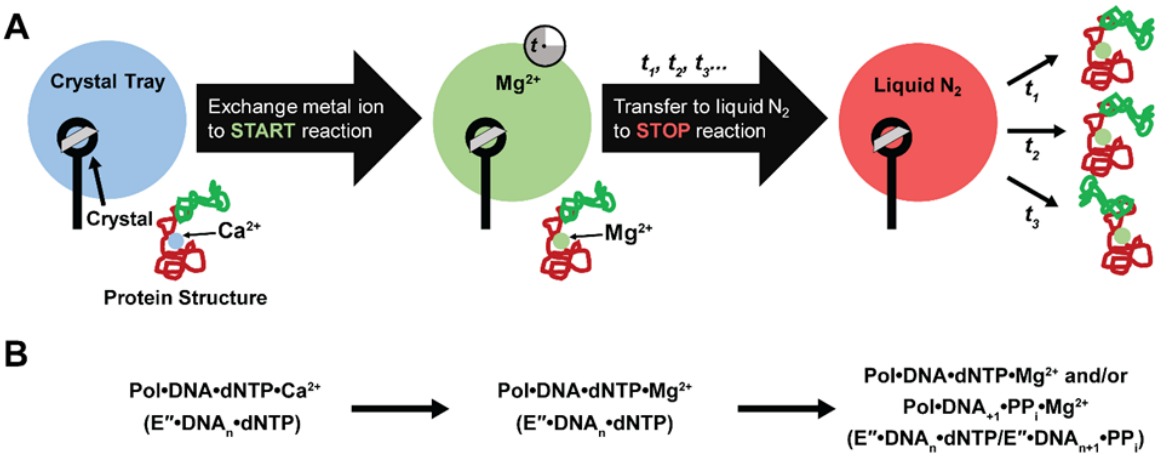


Figure 1.5: Time-resolved crystallography technique.

(A) A crystal in a pre-reactive state is isolated and transferred to a cryo-solution containing the catalytic metal ion. This initiates the reaction *in crystallo* and after varying periods of time, the reaction can be quenched by transferring the crystals to liquid N_2. Diffraction experiments are then performed on the crystals and the diffraction data are used to determine the three dimensional structures. During this process, the electron density of the bond forming and of the bond breaking is modeled as percent occupancy. The F_o-F_c difference map is then used to evaluate how well the model satisfies the experimental electron density.[225] (B) For time-resolved crystallography of a DNA polymerase-catalyzed nucleotidyltransfer reaction, a crystal of the ternary complex formed in the presence of non-catalytic Ca^{2+} is transferred to a cryo-solution containing the catalytic divalent metal ion, Mg^{2+} or Mn^{2+}. The polymerase complexes relevant to panel (A) are depicted and the corresponding enzyme forms relevant to those shown in Scheme 1.1A are shown in parentheses.

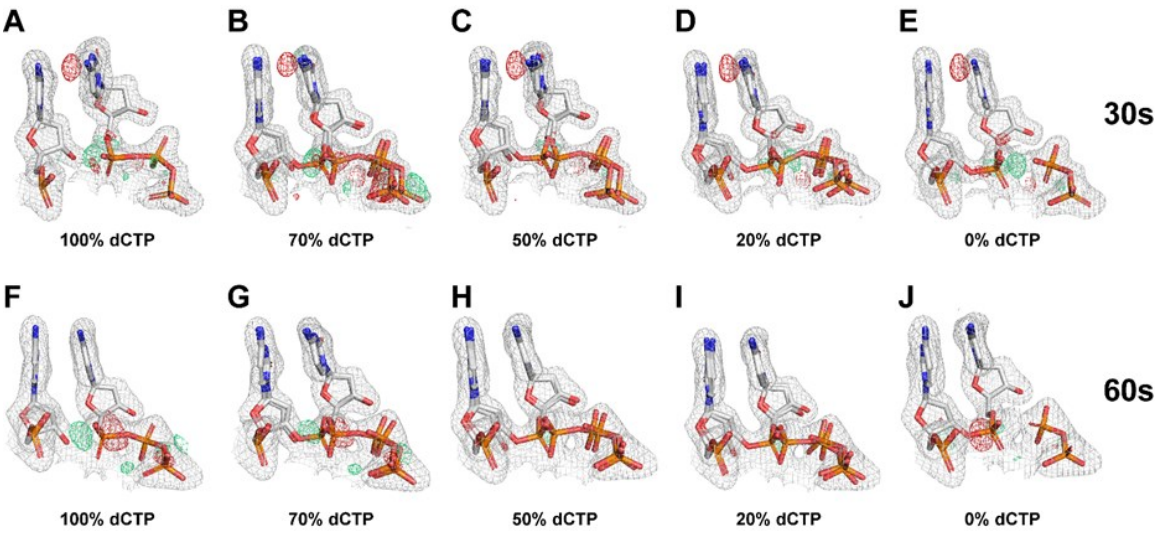


Figure 1.6: Modeling of phosphodiester bond formation.

Modeling of phosphodiester bond formation after 30 or 60 s of Mg^{2+}/Ca^{2+} ion-exchange during time-resolved X-ray crystallography of hPolβ.[42] The $2F_o$-F_c (light blue) maps contoured at 1σ and the F_o-F_c omit maps contoured at either 3σ (green) or -3σ (red) are presented for the primer 3′-terminal nucleotide, incoming dCTP, incorporated dCMP, and PP_i. The modeled occupancy of the reactants is listed below each structure. Strong positive (green) and negative (red) electron density mesh between the primer 3'-OH and the α-phosphate group of dCTP or between the α- and β-phosphate groups of dCTP indicate unsatisfactory modeling, *e.g.* the modeling of the reactants at 100% (A), 70% (B), 20% (D) and 0% (E) occupancies for the 30 s structure and 100% (F), 70% (G), 50% (H), 20% (I), and 0% (J) occupancies for the 60 s structure. In contrast, the absence of any positive or negative electron density with the modeling of the reactants at 50% (C) (continued)

Figure 1.6: continued

and 20% (I) occupancies suggests satisfactory modeling for the 30 and 60 s structures, respectively.[42]

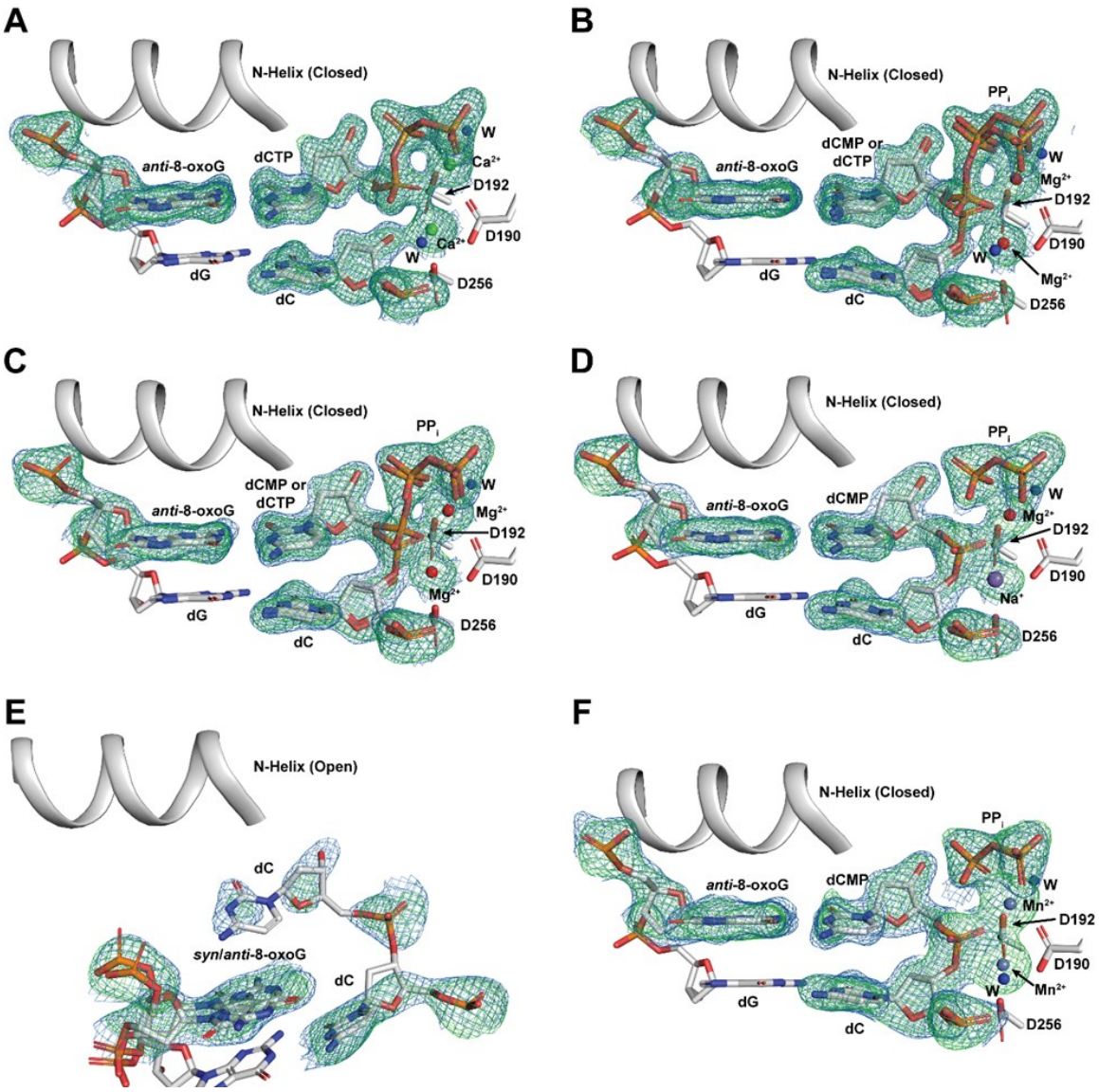


Figure 1.7: *In crystallo* phosphodiester bond formation by hPolβ.

Zoomed active sites show the incorporation of dCTP opposite 8-oxoG. The $2F_o$-F_c (blue mesh) and F_o-F_c (green mesh) maps are shown for the templating 8-oxoG, incoming dCTP, incorporated dCMP, pyrophosphate (PP_i), metal ions at the A- and B-sites, and the

(continued)

Figure 1.7: continued

primer 3′-terminal nucleotide (dC), contoured to 1σ and 3σ levels, respectively. Water molecules are shown as blue spheres. Ca^{2+}, Mg^{2+}, Mn^{2+}, and Na^{+} are shown as green, red, light blue, and purple spheres, respectively. Structures of the hPolβ pre-catalytic ternary complex (A), and structures of the hPolβ reaction intermediate or product ternary complexes following crystal-soaking with either 200 mM Mg^{2+} for 30 s (B), 60 s (C), 80 s (D), and 1 h (E), or 200 mM Mn^{2+} for 35 s (F). In (E), 8-oxoG was modeled in both *anti*- and *syn*-conformations.[42]

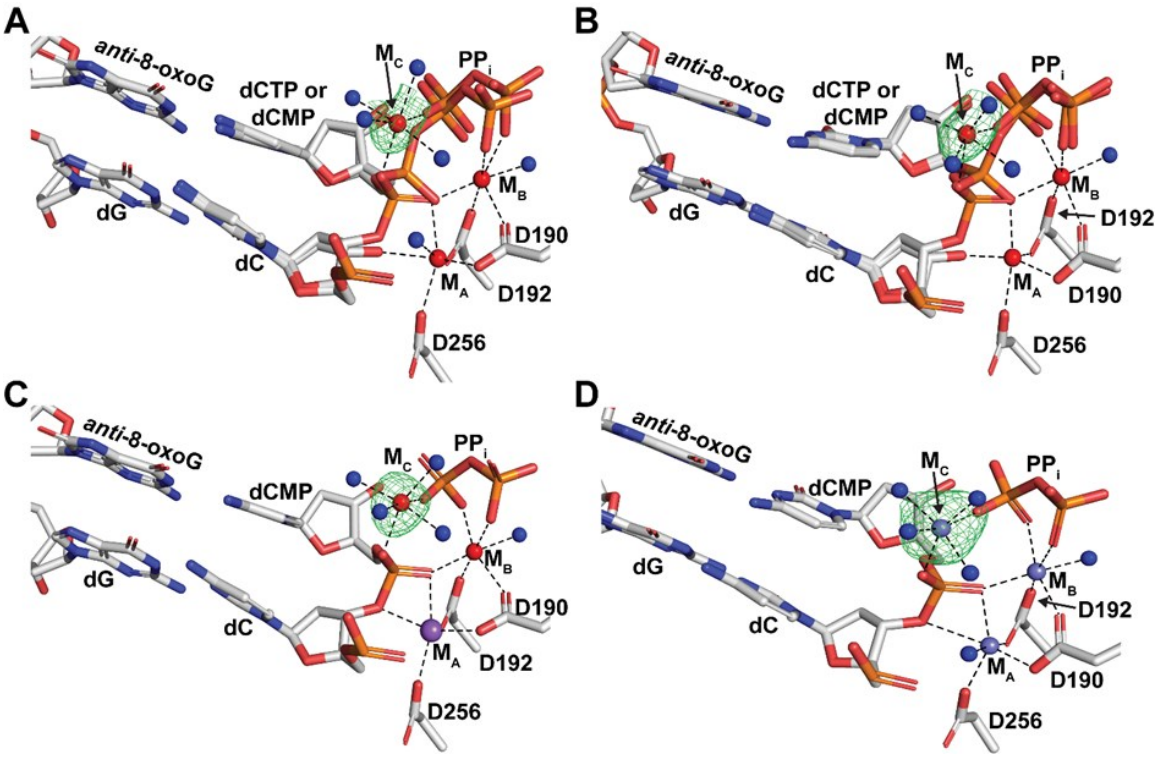


Figure 1.8: Appearance of the third divalent metal ion *in crystallo*

The third divalent metal ion is present in the reaction-state and post-catalytic structures of hPolβ during incorporation of dCTP opposite 8-oxoG. After soaking crystals of the pre-catalytic ternary complex of hPolβ (hPolβ•8-oxoG-DNA•dCTP) with Mg^{2+} for 30, 60, and 80 s, dCTP incorporation had completed by 50% (A), 80% (B), and 100% (C), respectively. (D) Following 35 s soaking with Mn^{2+}, dCTP was 100% incorporated. The F_o-F_c omit maps were contoured at 3σ (A-C) or 5σ (D) to show the electron density (green) of Mg^{2+} or Mn^{2+}, respectively, at the C-site. Coordinating aspartate side chains are shown as stick models, while water molecule ligands (blue), Mg^{2+} (red), Mn^{2+} (light blue), and Na^{+} (purple) are shown as spheres.[42]

Chapter 2. Introduction to DNA Lesion Bypass

Advances in Structural and Single-Molecule Methods for Investigating DNA Lesion Bypass and Repair Polymerases

Reproduced in part with permission from Raper, Austin T., Reed, A.J., Gadkari, V.V., and Suo, Z. (2017) Advances in Structural and Single-Molecule Methods for Investigating DNA Damage Tolerance and Repair. *Chem. Res. Toxicol.* 30, 260-269. DOI: 10.1021/acs.chemrestox.6b00342.

Author Contributions

Austin T. Raper and Andrew J. Reed equally contributed as co-first authors to the writing of the manuscript. Varun V. Gadkari provided initial ideas and helped edit the manuscript. Dr. Zucai Suo provided some conceptual direction and helped edit the manuscript.

2.1 Abstract

Innovative advances in X-ray crystallography and single-molecule biophysics have yielded unprecedented insight into the mechanisms of DNA lesion bypass and damage repair. Time-dependent X-ray crystallography has been successfully applied to view the bypass of 8-oxo-7,8-dihydro-2'-deoxyguanine (8-oxoG), a major oxidative DNA lesion, and the incorporation of the triphosphate form, 8-oxo-dGTP, catalyzed by human DNA polymerase β (hPolβ). Significant findings of these studies are highlighted here and their contributions to the current mechanistic understanding of mutagenic translesion DNA synthesis (TLS) and base excision repair (BER) are discussed. In addition, single-molecule Förster resonance energy transfer (smFRET) techniques have recently been adapted to investigate nucleotide binding and incorporation opposite undamaged dG and 8-oxoG by *Sulfolobus solfataricus* DNA polymerase IV (Dpo4), a model Y-family DNA polymerase. The mechanistic response of Dpo4 to a DNA lesion and the complex smFRET technique are described here. In this perspective, we also describe how time-dependent X-ray crystallography and smFRET can be used to achieve the spatial and temporal resolutions necessary to answer some of the mechanistic questions that remain in the fields of TLS and DNA damage repair.

2.2 Introduction

Over the last few decades, more emphasis has been placed on structural evidence of biochemical events than ever before. While many enzyme mechanisms can be inferred from the results yielded through typical biochemical methods, convincing structural evidence is often necessary to validate a particular model. For example, by solving the

crystal structures of the apo form, the binary complex (E•DNA), the pre-catalytic ternary complex (E•DNA•dNTP), and the product ternary complex (E•DNA•pyrophosphate) of numerous DNA polymerases across all six phylogenetic families (A, B, C, D, X, and Y),[237] which function in DNA replication, DNA damage repair, and translesion DNA synthesis (TLS),[23,61,63,64,67,68,76-78,80-82,84,87,223,238-240] the mechanism of DNA synthesis at the atomic-level was largely determined. Accordingly, more than 1,000 such crystal structures of DNA polymerases provide valuable information on the binding and positioning of substrates (DNA and dNTP) and products (DNA and pyrophosphate), the positioning and coordination of divalent metal ions, the binding conformations of active site residues, and putative protein conformational changes. Although spatially informative, traditional structural techniques are limited to static snapshots of stable states, therefore certain mechanistic details of the transition between reactant and product complexes of polymerase-catalyzed nucleotide incorporation may still be ambiguous or unidentified.

To overcome the disadvantages of conventional X-ray crystallography, time-dependent X-ray crystallography was recently developed and applied to follow polymerase-catalyzed reactions.[38-42,165] This relatively inexpensive and simple to execute technique has allowed several research laboratories to temporally capture the structures of reaction intermediates at atomic level resolutions.[38-42,165] As this elegant technique will likely change the research landscape for structurally characterizing biological processes such as DNA replication, DNA damage repair, and translesion DNA synthesis (TLS), we will highlight several notable time-dependent X-ray crystallographic studies of DNA

polymerases and give our opinions for new research directions in this perspective. Although useful, time-dependent X-ray crystallography is limited to enzyme reactions that can be readily activated *in crystallo*, and may only illuminate specific steps of a multi-step enzyme mechanism. For DNA polymerases, this has been limited to the chemistry step of nucleotide incorporation (Scheme 2.1, Step 6). Therefore, various mechanistic events, such as DNA binding, DNA translocation, nucleotide binding, protein conformational changes (Scheme 2.1, Steps 1, 4, 5, 7, and 8), and product release (Scheme 2.1, Step 9) cannot be observed. Furthermore, while time-dependent crystallography is possible through slowed enzyme catalysis *in crystallo*, any measured rates cannot be reliably compared to solution-based kinetics. Thus, the high-spatial resolution afforded by this technique comes at the cost of temporal resolution.

Consequently, a more versatile approach, capable of accessing faster, more relevant time scales and multiple enzymatic steps, is necessary to fully probe a catalytic mechanism. In recent years, advances in microscope optics, high resolution cameras, and state-of-the-art data acquisition and analysis software have spurned the advent and rapid growth of single-molecule methodologies by which to study biological processes *in vitro*. One particularly informative method, single-molecule Förster resonance energy transfer (smFRET),[241] reports on distance changes between site-specifically placed donor and acceptor fluorophores covalently linked to nucleic acid substrates or enzymes of interest. In this way, the solution behavior of individual molecules can be tracked in real-time to give unparalleled insight into complex enzyme mechanisms as well as protein-protein and protein/DNA interactions. Ensemble steady-state and pre-steady-state kinetic

methodologies,[22] while very informative and valuable, can only be utilized to report on the collective or average behavior of all molecules during a reaction.[241] As a result of this limitation, information on critical minor species or sub-populations of molecules during an enzyme-catalyzed reaction, that may perform distinct functions and/or exist in different conformational states, is lost. Furthermore, by following the reaction trajectories of individual molecules, smFRET has the remarkable ability to show the order of events for a given enzyme-catalyzed reaction without the need for synchronization, which is difficult or impossible for some systems. Thus, smFRET allows for simultaneous spatial and temporal characterization of an enzyme-catalyzed reaction. We have recently applied this technique to study *Sulfolobus solfataricus* DNA polymerase IV (Dpo4), a model Y-family DNA polymerase,[242] as it encountered and bypassed an 8-oxo-7,8-dihydro-2'-deoxyguanine (8-oxoG) lesion, a major oxidative DNA lesion, on a DNA template during DNA synthesis.[9] Structural studies have established that 8-oxoG does not significantly distort the structure of DNA, however its mutagenicity results from its ability to readily form two possible base pairing conformations within the polymerase active site.[243-245] Furthermore, published studies have shown that 8-oxoG forms specific interactions with polymerase active site residues to directly affect lesion bypass activities.[246-251]

Due to the wealth of data that single-molecule techniques can provide, we will give a more thorough description of our recent article in this perspective, as well as comment on future applications of this methodology to the study of DNA replication, TLS, and DNA damage repair. Previously, our laboratory has published many articles on the bypass of a variety of DNA lesions catalyzed by X- and Y-family DNA

polymerases.[11-13,133,134,138,139,226,240,252,253] Since most of these articles have been reviewed recently,[22] here we mainly discuss our work on 8-oxoG bypass catalyzed by Dpo4.[9,134]

2.3 Time-Resolved X-ray Crystallography

DNA polymerases bind DNA and dNTP sequentially, in addition to binding and coordinating divalent metal ions (*e.g.* Mg^{2+} or Mn^{2+}) to three conserved carboxylate residues at the active site in order to catalyze a nucleotidyl transfer reaction. Beginning in 2005, our laboratory attempted to perform time-dependent nucleotide incorporation in the crystals of Dpo4 through soaking its binary complex crystals (E•DNA) with an incoming correct or incorrect dNTP and Mg^{2+} for various times. However, this method resulted in the deterioration of crystal morphology and prevented us from obtaining high-resolution structures of reaction intermediate states. Instead, we collaborated with the laboratory of Paul Carey and utilized low-resolution single crystal Raman crystallography to obtain useful mechanistic information on nucleotide incorporation and protein conformational changes *in crystallo*.[252] In contrast to our misfortune, the laboratory of Katsuhiko Murakami employed similar time-dependent soak-trigger-freeze X-ray crystallography and solved the high-resolution crystal structures of pre-catalytic complex and two reaction intermediate states of bacteriophage N4 RNA polymerase.[165] Their structures reveal that the nucleotide-binding divalent metal ion at the B-site binds first followed by the catalytic divalent metal ion at the A-site. Following A-site divalent metal ion binding a phosphodiester bond was formed between the two bound rNTPs producing a 2-mer RNA transcript at the active site of bacteriophage N4 RNA polymerase, after which the A-site divalent metal ion quickly dissociates.[165] However, they did not observe any

reaction intermediates for bond formation or the third divalent metal ion bound at the C-site (see below).

Concurrently, the laboratory of Wei Yang pioneered a new method for time-dependent X-ray crystallographic studies of polymerases.[38] Rather than the aforementioned soak-trigger-freeze method where nucleotide and catalytic divalent metal ions were soaked simultaneously, their methodology takes advantage of the fact that the binding of Ca^{2+} at a polymerase active site allows for proper binding and positioning of DNA and nucleotide but does not support catalysis.[38] First, a pre-catalytic ternary complex (Scheme 2.1, E′′•DNA•dNTP), that has gone through all elementary steps in a kinetic mechanism (Scheme 2.1, Steps 1-5) prior to phosphodiester bond formation (Step 6)[5,9,11,22] was crystalized in the presence of non-catalytic Ca^{2+}. The Yang laboratory then initiated phosphodiester bond formation *in crystallo* via Ca^{2+} to Mg^{2+} ion exchange. This step was accomplished by removing crystals of the pre-catalytic ternary complex from the crystallization solution and transferring them to a cryo-protectant solution containing an excess of catalytic Mg^{2+} for a variable amount of time (seconds to minutes) before freeze-quenching the reaction in liquid nitrogen (Figure 2.1A). This methodology affords visualization of reaction intermediates in which mechanistic intricacies, such as chemical bond formation and breakage, steric inversion of the SN2 reaction center, changes in side chain rotamer conformations, DNA and dNTP binding conformations, and divalent metal ion coordination can be directly observed.[38] More significantly, the Yang laboratory unprecedentedly observed the appearance of a third Mg^{2+} ion during phosphodiester bond formation (Step 6, Scheme 2.1).[38]

Very recently, the Yang methodology was successfully adapted by Samuel Wilson's group and our laboratory[40-42] to investigate phosphodiester bond formation catalyzed by full-length human DNA polymerase β (hPolβ), an X-family polymerase. Wilson's group employed time-dependent X-ray crystallography to watch how hPolβ incorporated undamaged correct and incorrect dNTPs[40] as well as a damaged nucleotide (8-oxo-dGTP),[41] while we were the first to view both faithful (Figure 2.1B-D) and unfaithful TLS across a DNA lesion (8-oxoG).[42] Notably, the mutagenic bypass of the major oxidative DNA lesion 8-oxoG during base excision repair (BER) *in vivo* could lead to cancer formation.[254-257]

Rotational freedom about the glycosidic bond (*anti/syn*-conformations) and the additional hydrogen bonding capabilities of 8-oxo-dGTP and 8-oxoG on the Hoogsteen face (oxidation of C8 and protonation of N7) result in the dual coding potential of the oxidized nucleotide and the damaged templating base (Figure 2.2). Therefore, *anti*-8-oxo-dGTP and *anti*-8-oxoG can Watson-Crick base pair with cytosine (Figures 2.2A and 2.2C) and *syn*-8-oxo-dGTP and *syn*-8-oxoG can Hoogsteen base pair with adenosine (Figures 2.2B and 2.2D). In the time-dependent studies, both Wilson and our laboratories identified key side chain interactions that support the previously reported dNTP and templating base preference for 8-oxoG bypass and 8-oxo-dGTP incorporation, respectively (Figure 2.2).[226] Notably, R283 of the thumb domain acts as a negative selection factor against 8-oxo-dGTP incorporation opposite dC through electrostatic repulsion between the side chain of R283 and the N2 position of 8-oxo-dGTP (Figure 2.2C).[41] R283 was also shown to stabilize the *syn*-conformation of a templating 8-oxoG

during dATP misincorporation (Figure 2.2B).[42] Interestingly, this structural information is consistent with our previous kinetic characterization of 8-oxoG bypass or 8-oxo-dGTP incorporation by hPolβ, and suggests that R283 modulates the mutagenicity of these processes.[41,42,226]

In addition, in the structures where partial dCTP (Figure 2.1C) or dATP incorporation opposite 8-oxoG was observed the primer 3′-OH moves 0.8-1.3 Å relative to its position in the pre-catalytic ternary structures and initiates the nucleophilic attack on the Pα of the incoming dNTP, resulting in a 1.0-1.1 Å shift of the Pα towards the 3′-OH and the steric inversion of the nonbridging oxygen atoms, as expected for an SN2 reaction.[42] Furthermore, key details for the post-catalytic process were identified including a stacking interaction between the newly incorporated nucleotide and the templating 8-oxoG, suggesting a role in preventing successive nucleotide incorporation and strand displacement on the nicked DNA product.[42]

More interestingly, as with the first time-dependent crystallography study of phosphoryl transfer by hPolη,[38] all three studies on hPolβ revealed three divalent metal ions (Mg^{2+} or Mn^{2+}) at the polymerase active site during phosphodiester bond formation.[40-42] The third divalent metal ion was unambiguously detected when Ca^{2+} was exchanged *in crystallo* with Mn^{2+},[40,42] which yielded anomalous scattering, and suggests that polymerases use a "three-metal-ion mechanism" (Figures 2.1C – 2.1D). These time-dependent X-ray crystallographic studies[38-42,165] clearly established that the binding of the divalent metal ions before phosphodiester bond formation follows the order of the B-, A-, and C-site while their dissociation order after the chemistry step is the A-, C-, and B-site.

These results challenge the long-standing "two-metal-ion mechanism" for the nucleotidyl transfer reaction proposed by Lauren Beese and Tom Steitz nearly two decades ago.[31] The Steitz laboratory solved the first structures of the Klenow fragment of *E. coli* DNA polymerase I and discovered that the exonuclease activity appears to require two Mg^{2+} ions[31] and therefore it was hypothesized that all phosphoryl transfer reactions including polymerase-catalyzed nucleotide incorporation[26,258] likely follow a two-divalent metal-ion mechanism. Their results have since been supported by hundreds of pre-catalytic ternary structures of other DNA polymerases[26,59,61,63,158,207] in which nucleotide and two divalent metal ions are bound at the polymerase active site. In the "two-metal-ion mechanism", the A- and B-site divalent metal ions, are associated with the deprotonation of the primer terminus and the binding of incoming nucleotide, respectively.

However, through the time-dependent X-ray crystallography studies,[38,40-42] the observed third Mg^{2+} ion at the C-site appears in a position bridging the α- and β-phosphates of the incoming nucleotide following reaction initiation (Figure 2.1C) and persisting through completion of nucleotide incorporation (Figure 2.1D), with its role in catalysis still debated and several competing hypotheses proposed. The positioning and timing of the third divalent metal ion have led Wilson's group to initially hypothesize a role in the reverse reaction, pyrophosphorolysis,[40] while others[38,39,41,259] and our laboratory[42] have suggested the third metal ion is involved in transition state stabilization and may aid in pyrophosphate release (Figures 2.1C and 2.1D). In any case, the necessity of a third divalent metal ion for catalysis is hard to discredit with five recent structural studies[38-42] and two computational analyses that detail its presence.[49,50]

While further biochemical or biophysical characterization is necessary to elucidate the role of the third divalent metal ion, it does not coordinate protein residues and therefore it is difficult to definitively assay its role in catalysis through simple mutational studies. However, if the third divalent metal ion serves a catalytic role in the forward reaction, then it should also participate in the reverse reaction. Consistently, the Wilson laboratory was able to crystallize the ground state of the reverse, pyrophosphorolysis reaction with the third divalent metal ion bound, but did not capture any reaction state structures (presumably because the reverse reaction is unfavorable), which lead them to hypothesize its importance in either pyrophosphate release or pyrophosphorolysis.[40] The Yang group was first to attempt to assay the functional role of the third divalent metal ion via measuring metal ion saturation *in crystallo* and then correlating this with the amount of divalent metal ion necessary for nucleotidyl transfer in solution.[39] While this was a reasonable approach, it would be perhaps more convincing to determine its functional role through techniques other than time dependent X-ray crystallography. For example, our laboratory is utilizing electroparamagnetic resonance (EPR) spectroscopy to investigate the functional roles of divalent metal ions on the activity of hPolβ.

Notably, the above mechanistic details as well as others presented in the time-dependent X-ray crystallographic studies[38-42] are unobservable with conventional X-ray crystallography. It will be interesting to see if some of the new mechanistic aspects, such as the "three-metal-ion mechanism", will hold true for other DNA polymerases, especially those in the A, B, C, and D families as well as reverse transcriptases. It will

also be exciting to watch how other types of DNA lesions, *e.g.* bulky single-base lesions, double-base lesions, and tandem lesions, are accommodated and bypassed by the X- and Y-family DNA polymerases. Furthermore, while the mechanistic steps pertaining to the reverse conformational change, pyrophosphate release, and DNA translocation or release (Scheme 2.1, Steps 7-9) have remained elusive to structural characterization, time-dependent X-ray crystallography has the potential to probe these processes to further define the polymerase-catalyzed mechanism of nucleotide incorporation (Scheme 2.1).

2.4 Single-Molecule Förster Resonance Energy Transfer

smFRET can be applied to study distance changes between donor and acceptor fluorescent dyes covalently attached to one or more interacting biomolecules. The donor and acceptor fluorophores must meet several photophysical criteria,[241] and be positioned such that the inter-fluorophore distance is limited to $R = R_0 \pm 0.5R_0$, where R is the inter-fluorophore distance and R_0 is the calculated Förster radius, which is unique to the selected dye pair and depends on several important parameters,[241] not discussed here. When properly designed, a FRET system can be very sensitive to distances changes, yielding low-resolution structural information for distance changes occurring between the donor and acceptor within 10-75 Å.[260]

To apply this technique for the monitoring of single molecules, especially DNA polymerases, one very popular strategy involves the surface immobilization of a DNA substrate that is chemically modified to contain a donor fluorophore and 5′- or 3′-biotin group (Figure 2.3A). Quartz slides and glass coverslips are chemically functionalized and passivated with a lawn of polyethylene glycol molecules (some of which are modified

with biotin bound by NeutrAvidin) as well as sandwiched together and sealed to form a reaction chamber. The modified DNA is then flowed into the chamber and immobilized via the very tight biotin-NeutrAvidin interaction. Prism-based total internal reflection fluorescence (TIRF)[241] is then utilized to greatly reduce fluorescent background in order to achieve signal-to-noise ratios sufficient to image single DNA molecules. Briefly, the donor molecules are specifically excited by a laser of appropriate wavelength that is incident on a Pellin Broca prism at the critical angle necessary to achieve TIRF (Figure 2.3A).[241] Polymerase molecules labeled with an acceptor fluorophore are then introduced into the imaging chamber to bind to the immobilized DNA. Upon polymerase binding to DNA, if the donor and acceptor fluorophores are sufficiently near to one another, energy transfer (i.e. FRET) from the donor to the acceptor can occur (Figure 2.3A). The fluorescence emission of the donor is then spectrally separated from the acceptor by the utility of dichroic mirrors to form distinct donor and acceptor emission channels (Figure 2.3B). Subsequently, both donor and acceptor emissions are incident on an electron multiplying charged coupled device (EMCCD) camera to record the fluorescence intensities over time (Figure 2.3C). The FRET efficiency (E) is then calculated as $E = I_A/(I_A+I_D)$, where I_D and I_A are the fluorescence intensities of the donor and acceptor, respectively. Accordingly, polymerase binding events and subsequent protein conformational changes are reflected as anti-correlated decreases in the donor emission intensity and increases in the acceptor emission intensity. Notably, smFRET has been applied to study various biochemical systems, including DNA replication.[9,111,163,213,261-266]

However, of the many existing DNA polymerase smFRET studies, only a few have explored TLS.[9,265,266]

A recent publication from our laboratory utilized smFRET to study the lesion bypass capabilities of Dpo4.[9] Dpo4, the lone Y-family DNA polymerase in *Sulfolobus solfataricus*, has been extensively studied due to its structural and functional similarities to the human Y-family polymerases.[84,242] We sought to explore the effects of a templating 8-oxoG lesion on the dynamics of Dpo4 during DNA binding as well as nucleotide binding. By examination of single-molecule FRET trajectories of Dpo4 binding to immobilized DNA substrates with a dG base at the templating position, we first observed the dynamic behavior of the polymerase on undamaged DNA. Interestingly, Dpo4 was shown to interconvert between three primary, non-zero FRET efficiency states (Figure 2.4A). In any smFRET analysis, assigning physical significance to observed FRET states is challenging. For our study, we had to carefully analyze all control and experimental data, as well as apply our knowledge of Dpo4 from previous stopped-flow FRET investigations,[11,12,22,267] before developing the proposed model. We assigned the mid-FRET efficiency state (FRET ≈ 0.6) to the polymerase bound to the DNA with the primer/template junction base pair occupying the position of nascent base pair in the active site (pre-insertion state), while in the low-FRET efficiency state (FRET ≈ 0.4), Dpo4 had translocated along the DNA substrate by one base pair to permit the binding of an incoming nucleotide (insertion state). Beyond the location of the polymerase at the primer/template terminus, the low- and mid-FRET states are also associated with distinct conformations of the polymerase as suggested by our dwell time analysis of the low- and

mid-FRET states, wherein the duration of time the polymerase remains at a particular FRET efficiency value before transitioning to another is quantified. We found that the rate of interconversion (2.8-4 s^{-1}) between the low-and mid-FRET states was far too slow to be reporting on the actual DNA translocation event itself, as mistakenly suggested by another single-molecule research group,[163] which has been estimated to occur at rates of >100 s^{-1}.[11,12,268-270] The rate of interconversion actually reports on domain motions, that occur concomitant with translocation along the DNA substrate, as observed through the Cy5-labeled Finger domain in our smFRET study.[9] Accordingly, our FRET analysis was also reporting on the dynamics of the Finger domain during DNA binding as well as DNA translocation, which is consistent with the dynamic behavior of the Finger domain, as well as other domains of Dpo4, previously elucidated by our research group through stopped-flow FRET studies.[11,12,22,151,267] Interestingly, our experiments also detected the existence of a high-FRET state (FRET ≈ 0.8) which we expect to be a unique binding mode of Dpo4 involving the Little Finger domain, although more work is necessary to better characterize this unique finding.

When Dpo4 was bound to DNA containing the 8-oxoG lesion at the templating position (Figure 2.4A), the results were largely unchanged compared to those with the undamaged DNA substrate (Figure 2.4B). However, we did detect subtle differences between the experiments using 8-oxoG-containing DNA compared to those with undamaged DNA including a shift in the number of FRET events at the low-FRET state (50% with the 8-oxoG-containing DNA (Figure 2.4C), versus 35% with the undamaged DNA (Figure 2.4B) as well as increased polymerase binding times and slower

dissociation kinetics from the lesion-containing DNA substrate. Therefore, more polymerase molecules were observed to be in the productive insertion state, ready to accept an incoming nucleotide, as well as spent longer periods of time on the 8-oxoG-containing DNA substrate without dissociating. Accordingly, we suggested that the 8-oxoG lesion modestly stabilized the formation of productive binary complexes. This assertion was further supported by experiments wherein the correct nucleotide, dCTP, was included in the imaging chamber in the presence of non-catalytic Ca^{2+}, to form ternary complexes (E′′•DNA•dCTP, Scheme 2.1) of either Dpo4 with undamaged or 8-oxoG-containing DNA. Interestingly, the presence of the correct nucleotide completely shifted all FRET events to the low-FRET state and significantly enhanced the FRET event durations. In fact, Dpo4 remained bound to DNA 5-10-fold longer in presence of a correct over an incorrect dNTP. Furthermore, we observed that Dpo4, in the presence of dCTP, remained bound to 8-oxoG-containing DNA nearly 2-fold longer than undamaged DNA. In this way, we were able to observe the nucleotide selectivity of a Y-family DNA polymerase on undamaged as well as damage-containing DNA, and preference for the 8-oxoG lesion at the templating position. Notably, Dpo4 has been shown to replicate more efficiently across from 8-oxoG than normal dG,[134] and our data suggest that this occurs as a result of stabilization of the ternary complex of the polymerase. Consistently, these findings are in agreement with previous structural studies that have established that Arg332 in the Dpo4 active site forms specific interactions with 8-oxoG.[162] Our single-molecule investigation of the binding modes and kinetics of Dpo4 is largely consistent with a previous smFRET study of Dpo4 by the Rueda and Romano research laboratories,

wherein Dpo4 binds to undamaged DNA in at least two FRET states, with the low-FRET state stabilized by correct nucleotide binding.[163] Another smFRET study by these groups investigated the effects of the bulky DNA adducts 2-aminofluoene and N-acetyl-2-aminofluorene on Dpo4 binding conformations and activity when the lesions were located at the templating or junction base pair positions.[266] Their results suggest that these bulky lesions significantly distort the structure of the binary complex as compared to results with unmodified DNA or our results with the non-helix distorting lesion, 8-oxoG. Consistent with our results with 8-oxoG, the addition of the correct nucleotide again induced a stable ternary complex (E′′•DNA•dCTP, Scheme 2.1) despite the DNA damage in the template strand. Collectively, their study and ours illuminate key details of the response of Dpo4 to two bulky DNA adducts and a small, major oxidative lesion, respectively, and help to explain their mechanisms of mutagenicity. It is clear from these studies that Dpo4 is better suited to faithfully replicate opposite 8-oxoG, whereas the bulkier amine adducts more significantly affect Dpo4 binary and ternary complexes to contribute to misincorporation events.

For future investigation, it will be interesting to see how the extension step of TLS by Dpo4, which has been shown to be more significantly disrupted by DNA damage,[138,139] is affected by the 8-oxoG lesion at the single-molecule level as well as how the inclusion of bulkier, DNA-helix distorting lesions, such as cyclobutane pyrimidine dimers or nitropolyaromatic hydrocarbon-generated lesions, affect the single-molecule behavior of Dpo4. Furthermore, it remains to be seen if other Y-family polymerases act similar to Dpo4 when they bypass a lesion and then subsequently extend the lesion

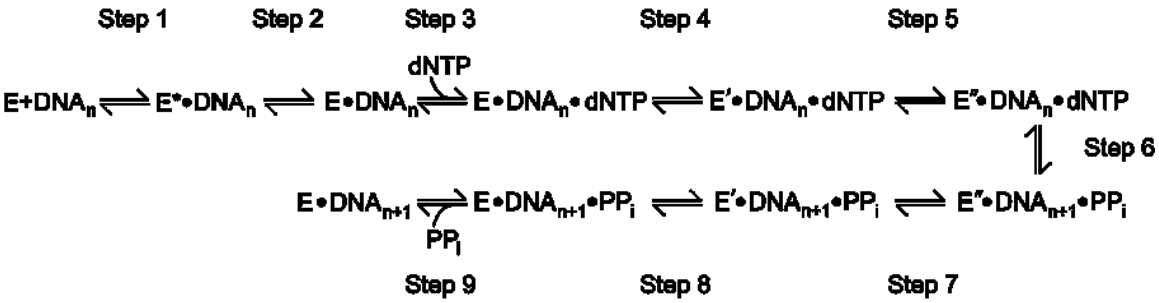
bypass product. Moreover, we are excited at the prospect of utilizing single-molecule spectroscopy to investigate the complexities of polymerase switching during TLS. When a replicative polymerase is stalled by a DNA lesion during DNA replication, it is replaced by a Y-family polymerase, which is expected to bypass the damage, in a process known as polymerase switching.[230,271-273] This event is thought to be mediated by the ring-like clamp protein, proliferating cell nuclear antigen (PCNA) which may simultaneously bind to both the replicative and Y-family DNA polymerases.[274-276] Although polymerase switching is a vital, damage tolerance response for the cell, the details of the process are not well-defined and several competing mechanisms exist to describe its functioning.[230,271-273] Application of smFRET to study polymerase switching will undoubtedly yield key mechanistic insight and clarity into the protein-protein and protein-DNA dynamics at the stalled replication fork.

2.5 Conclusion

Taken together, our detailed analyses of DNA damage tolerance and repair by cutting-edge time-dependent crystallographic and single-molecule methodologies have significantly increased our understanding of these vital life processes. In the coming years, it will be interesting to see how these techniques will continue to inform on the mechanisms of BER enzymes, as well as the putative coordination and interactions between them. Moreover, we are excited about the prospect of further clarifying the underlying mechanisms of mutagenic translesion DNA synthesis by X- and Y-family DNA polymerases by these methods. In all, the technological advances in enzymology described in this review have informed ground-breaking research to yield paradigm-

shifting results, and we expect that future applications will be equally impactful to the field.

2.6 Schemes



Scheme 2.1: Proposed mechanism of nucleotide incorporation catalyzed by a DNA polymerase.

E, E′, and E″ represent three different conformations of the DNA polymerase sampled during correct nucleotide incorporation. E* represents the non-productive, pre-translocated binary complex.

2.7 Figures

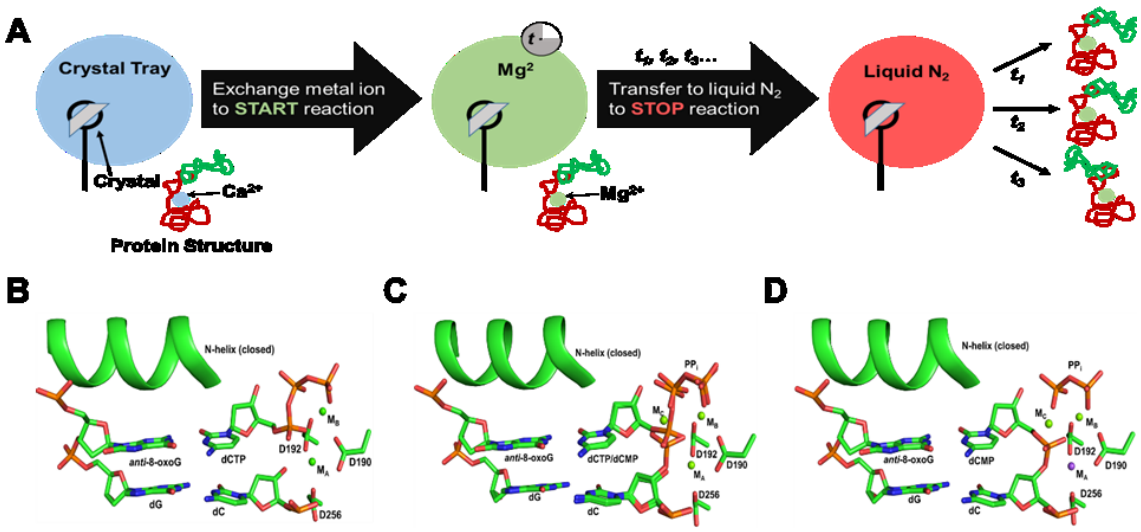


Figure 2.1: Time-dependent X-ray crystallography of DNA lesion bypass.

(A) Cartoon representation of time-dependent X-ray crystallography method. Crystal trays are set up in the necessary conditions to grow protein crystals of pre-catalytic ternary polymerase complexes in the presence of Ca^{2+}. The isolated crystals are transferred to a cryo-solution that contains magnesium (Mg^{2+}) to initiate the reaction *in crystallo* by exchanging the bound Ca^{2+} with Mg^{2+}. The crystals are soaked in the Mg^{2+} cryo-solution for varying times (t_1, t_2, t_3…) and then transferred to liquid N_2 to freeze-quench the reaction before X-ray diffraction experiments. (B-D) Structural progression of single-nucleotide incorporation catalyzed by hPolβ (PDB codes 4RPX, 4RPZ, and 4RQ0).[42] (B) Pre-catalytic ternary complex of hPolβ with correct incoming dCTP paired with templating 8-oxoG with bound A- and B-site divalent metal ions (green spheres, M_A and M_B, respectively). (C) Reaction state structure of hPolβ during phosphoryl transfer

(continued)

Figure 2.1: continued

with the incoming nucleotide modeled partially as dCTP and dCMP to match the increasing electron density (not shown) between the primer 3′-OH and the α-phosphate of dCTP and the decreasing electron density (not shown) between the α- and β-phosphates of dCTP. The third divalent metal ion occupies the C-site (M_C) and is coordinated by the nonbridging oxygen atoms of the α- and β-phosphates and water molecules (not shown). (D) Post-catalytic complex of hPolβ following complete incorporation of dCTP opposite 8-oxoG with M_C and pyrophosphate (PP_i) still bound. Notably, M_A is modeled as Na^+ (purple sphere). (B-D) Each panel shows a cartoon model of the N-Helix of the thumb domain in the closed conformation.

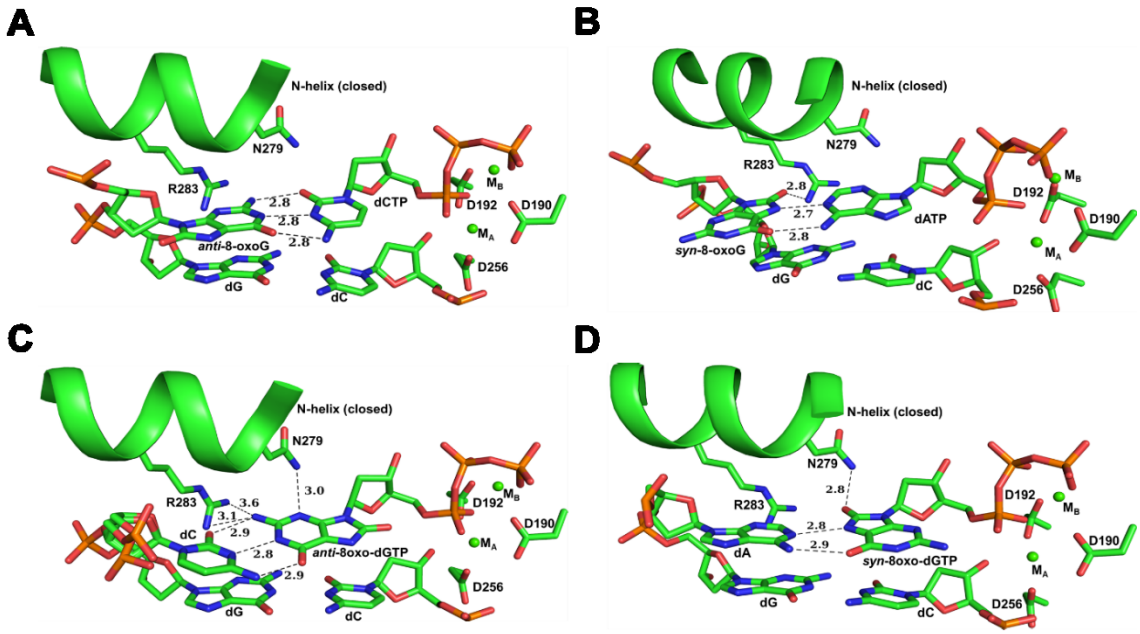


Figure 2.2: Zoomed active site views of hPolβ with templating 8-oxoG or incoming 8-oxo-dGTP.

(A) Templating *anti*-8-oxoG with incoming dCTP (PDB code 4RPX).[42] (B) Templating *syn*-8-oxoG with incoming dATP (PDB code 4RQ3).[42] (C) Incoming *anti*-8-oxo-dGTP with templating dC (PDB code 4UBC).[41] (D) Incoming *syn*-8-oxo-dGTP with templating dA (PDB code 4UAW).[41]

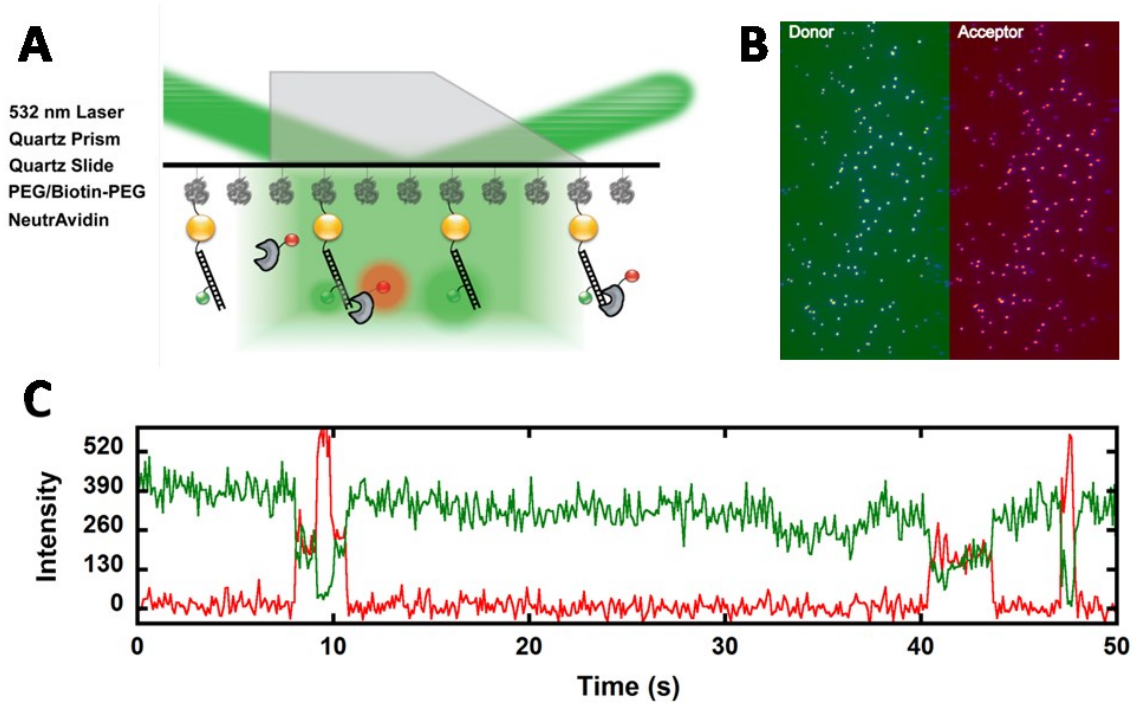


Figure 2.3: Single-molecule FRET.

(A) Modified, donor labeled DNA is immobilized via the biotin-NeutrAvidin interaction. Donor molecules (green spheres) are specifically excited by a laser of appropriate wavelength using total internal reflection fluorescence (TIRF). Acceptor labeled polymerase (red spheres) binding to DNA, results in energy transfer (i.e. FRET). (B) The resulting fluorescence emissions of the donor and acceptor are spectrally separated to form separate donor and acceptor emission channels. (C) Movies of single molecules are recorded and colocalized fluorescence intensities in the donor and acceptor emission channels are extracted by customized software. Anti-correlated changes in the donor and acceptor fluorescence intensities, shown in green and red, respectively, indicate FRET events.

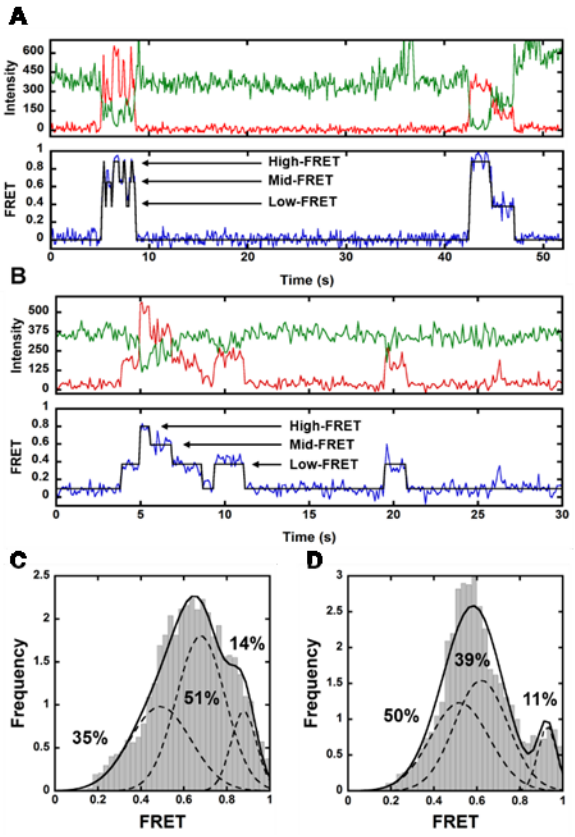


Figure 2.4: Single-molecule analysis of Dpo4 binding to undamaged or 8-oxoG-containing DNA.

(A, B)[9] The top panels show representative single-molecule trajectories of Dpo4 binding to undamaged DNA (A) or 8-oxoG-containing DNA (B) with donor and acceptor fluorescence intensities shown in green and red, respectively. The bottom panels display the corresponding, calculated FRET efficiency in blue and the idealized FRET efficiency from a Hidden Markov analysis in black. Three non-zero FRET states are clearly evident in each trajectory, High-, Mid-, and Low-FRET. FRET efficiency histograms of Dpo4 (continued)

Figure 2.4: continued

binding to undamaged (B) or 8-oxoG-containing DNA (C) with individual Gaussian peak fits for each FRET state shown as overlapping, black dashed lines and the Gaussian sum shown as the solid black line. The respective population percentages for each FRET state are indicated in black.

Chapter 3. Introduction to CRISPR/Cas9 and Gene Editing

Sharpening the Scissors: Mechanistic Details of CRISPR/Cas9 Improve Functional Understanding and Inspire Future Research

Author Contributions

Austin T. Raper and Anthony A. Stephenson wrote the manuscript and generated all figures. Dr. Zucai Suo helped to edit the manuscript.

3.1 Abstract

Interest in CRISPR/Cas9 remains at a high level as new applications of the revolutionary gene-editing tool continue to emerge. While key structural and biochemical findings have illuminated major steps in the enzymatic mechanism of Cas9, several important details remain unidentified or poorly characterized that may contribute to known functional limitations. Here we describe the foundation of research that has led to a fundamental understanding of Cas9 and address mechanistic uncertainties that restrict continued development of this gene-editing platform, including specificity for the protospacer adjacent motif, propensity for off-target binding and cleavage, as well as interactions with cellular components during gene editing. Discussion of these topics and considerations should inspire future research to hone this remarkable technology and advance CRISPR/Cas9 to new heights.

3.2 Introduction

Among recent breakthroughs in scientific research, few have had such an obvious and immediate translation to human health as the discovery of clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) proteins. Shortly after its discovery, this defense system developed by prokaryotes during their endless war against viruses, was repurposed for gene editing in virtually every model organism in each domain of life.[277-281] The CRISPR immune system includes alternating repeat and spacer DNA sequences arranged in a CRISPR array (Figure 3.1A).[281-283] In contrast to the invariant sequence of the repeats, each spacer consists of a unique DNA sequence acquired during previous bacteriophage or plasmid invasion of the host

prokaryote. While there are three main stages to CRISPR-mediated adaptive immunity in prokarya including spacer acquisition, CRISPR RNA (crRNA) biogenesis, and DNA interference, we will mainly focus on the final stage for the popular Class 2, Type II CRISPR/Cas9 system found within *Streptococcus pyogenes*. However, it is important to mention that during the first two stages the Cas9 endonuclease is expressed and the CRISPR array is acquired, transcribed, and processed to generate a dual-RNA guide. The dual-RNA guide is composed of a crRNA that contains a unique spacer sequence as well as the repeat sequence that hybridizes with a transactivating crRNA (tracrRNA) (Figure 3.1B). During the interference stage, Cas9 binds to the dual-RNA guide to form the catalytically-active ribonucleoprotein (RNP). This effector complex must then recognize a three nucleotide protospacer adjacent motif (PAM) on the non-target strand (ntDNA) as well as share sufficient complementarity (~20 base pairs) between the guide region of the crRNA and the target strand (tDNA) of the invading nucleic acid to license double-stranded DNA (dsDNA) cleavage ~3 nucleotides away from the PAM (Figure 3.1B).[281-283] Accordingly, Cas9 can be targeted to cleave virtually any DNA sequence (assuming that the PAM requirement is met) by simply changing the guide region of the crRNA. Moreover, the crRNA and tracrRNA can be fused through a four-nucleotide tetraloop to form the single-guide RNA (Figure 3.1B), thereby simplifying Cas9 to a two-component system for gene editing.[279]

Easily programmable DNA targeting and cleavage is the hallmark of the CRISPR/Cas9 system and sets it apart from previous nucleases engineered for gene editing (*e.g.* zinc-finger nucleases (ZFNs) and transcription activator-like effector

nucleases (TALENs)).[277,284] Whereas previous gene editing platforms were hindered by the complicated engineering of site-specific nucleases, the CRISPR/Cas9 platform overcomes this rate-limiting step. This technological breakthrough in the ease of genetic manipulation has encouraged widespread applications in basic and translational research. CRISPR/Cas9 is perhaps most prominently used to knock-out genes in mammalian cells for the purpose of understanding gene function or manipulating genetic backgrounds for disease models. During gene knock-out experiments, dsDNA breaks generated by Cas9 activate the non-homologous end joining (NHEJ) DNA repair pathway which can cause loss-of-function insertion or deletion mutations (indels) at the repaired cleavage site (*i.e.* frameshift mutation) (Figure 3.2A).[280,284] Alternatively, Cas9-mediated knock-in of a gene can be achieved by providing cells with a DNA fragment (*i.e.* donor DNA) consisting of the desired gene flanked by DNA sequences (*i.e.* homology arms) homologous to the genomic DNA on each side of the Cas9 cut site (Figure 3.2B).[280] This process transpires through a homology directed repair (HDR) pathway and generally occurs with much lower efficiency than gene knock-out.[284,285] As with gene knock-out, the facile ability to perform gene knock-in with Cas9 permits rapid generation of disease models. Moreover, the potential to site-specifically graft functional genes or replace mutant alleles has renewed interest in applications of gene therapy.[277,286,287]

Beyond gene knock-out and knock-in, the site-specific nuclease activity of Cas9 has been harnessed for a variety of remarkable technologies including CRISPR-mediated analog multi-event recording apparatus (CAMERA) for maintaining a biological transcript of environmental conditions,[288] gene-drives for controlling insect

populations,[289] and anti-viral therapy.[290] The utility of Cas9 extends beyond generating targeted DNA strand breaks. In fact, a catalytically inactive (D10A and H840A) mutant of Cas9 (dCas9) can be programmed for specific DNA binding without cleavage.[278,279] Notable utilities of dCas9 involves its fusion to fluorescent molecules for imaging of specific genomic loci in living organisms (Figure 3.2C),[291] fusion to transcriptional activators or repressors for modulating gene expression (Figure 3.2D),[292] and fusion to "DNA base editors" (*e.g.* cytosine or adenine deaminases) to generate single-base changes near a DNA target site (Figure 3.2E).[293-295] Moreover, dCas9 can sterically block transcription initiation and/or elongation to directly facilitate gene knock-down with similar efficiency as RNA interference (Figure 3.2F).[296]

Despite the impressive portfolio of CRISPR/Cas9-mediated applications, several important limitations persist that restrict the potential of this technology. Perhaps the most notorious shortcoming involves the propensity of the Cas9 effector complex to bind and cleave at locations in the genome other than the intended site (*i.e.* off-target sites).[297-299] Off-target cleavage during gene-editing experiments with wild-type Cas9 challenges the safety of CRISPR/Cas9 as an agent for gene therapy. However, several groups have engineered superior variants of Cas9 which appear to display lower off-target cleavage relative to the wild-type enzyme, although more extensive evaluation of these mutants is necessary.[300-303] Notably, while these engineered variants of Cas9 seem to reduce occurrence of off-target cleavage, they are not effective at preventing off-target binding.[300,301,303-306] This is a significant problem as studies report that off-target binding occurs with much higher frequency than off-target cleavage,[307-310] which may confound

experiments utilizing dCas9 for imaging genomic loci (Figure 3.2C), regulating transcription (Figure 3.2D), and precision gene editing with "base editors" (Figure 3.2E). Another shortcoming associated with CRISPR/Cas9 is the restricted target space imposed by the strict PAM requirement. Importantly, homologs and engineered variants of Cas9 with PAM specificities distinct from wild-type *S. pyogenes* Cas9 have been shown to circumvent this problem.[302,311-313] Indeed, one study utilized directed evolution to produce mutants of wild-type Cas9 with broad PAM specificities (*i.e.* recognition of sequences beyond the 5′-NGG-3′ for *S. pyogenes* Cas9), increasing the target space for editing pathogenic single-nucleotide polymorphisms (SNPs) by ~3-fold.[302] However, a sizable amount of target space remains out of reach and the difficulty of gene knock-in experiments are compounded by PAM restrictions within potentially narrow sequence windows for gene knock-in. One more notable obstacle for gene editing with Cas9 is the variability in efficiency of gene knock-in across different cell types and locations in the genome.[283-285,314] It is likely that this efficiency may in large part be derived from the variable activity of competing DNA repair pathways (*i.e.* NHEJ vs. HDR) in cells.[315,316] Gene knock-in efficiency can be significantly enhanced in many cell types by modulating the cellular decision between NHEJ and HDR (*i.e.* knockdown or inhibition of key repair factors).[314] However, it is unknown to what extent Cas9 may influence or disrupt DNA repair processes. Significantly, many solutions to these limitations may come from filling gaps in mechanistic understanding of CRISPR/Cas9 enzymology, which includes the impact of Cas9 on relevant cellular pathways.

3.3 Structure and Mechanism of Cas9-Mediated Interference

Mechanistic work on Cas9 has been greatly facilitated through several key crystal structures (Figure 3.3). Cas9 is a multi-domain protein separated into a recognition lobe (REC lobe), responsible for binding to RNA and DNA, and a nuclease lobe (NUC lobe), responsible for PAM interrogation and cleavage of DNA targets (Figure 3.3A).[282] The REC lobe can be further subdivided into the REC1 and REC2 domains. While REC1 makes extensive contacts with crRNA:DNA heteroduplex and is crucial for DNA cleavage activity, the role of the REC2 domain appears to be more dispensable.[317] Importantly, some variants of Cas9 have markedly smaller REC lobes,[283] which may be more amenable to Cas9-mediated gene therapy using size-restricted viral vector delivery.[282] The NUC lobe can be further subdivided into three primary domains, the C-terminal PAM-interacting domain (CTD), the HNH nuclease domain, and the RuvC nuclease domain.[282] The CTD reads-out the PAM through base-specific interactions of two key arginine residues (R1335 and R1333) with the Hoogsteen faces of the deoxyguanosine nucleotides in the 5′-NGG-3′.[318] Following target recognition and heteroduplex formation, tDNA and ntDNA are cleaved ~3 nucleotides away from the PAM by HNH and RuvC nuclease domains, respectively (Figure 3.1B).[279,319,320] The REC and NUC lobes are connected through a dynamic structural element termed the bridge helix (Figure 3.4A).[282] Remarkably, Cas9 can be split into REC and NUC fragments which exhibit comparable RNA and DNA binding as well as DNA cleavage activity upon reconstitution, underlining the distinct bi-modular functionality of Cas9.[321]

3.3.1 Assembly of the Cas9 effector complex.

Structural and mechanistic work has revealed that Cas9 exists in an auto-inhibited conformation in the apo-state (Figure 3.3A) through occlusion of the nucleic acid binding cleft and HNH active site residues.[322,323] Binding of sgRNA induces a dramatic structural rearrangement of the REC lobe that appears to be mediated by conformational dynamics in an arginine-rich bridge helix (residues 60 – 93) connecting the REC and NUC lobes (Figure 3.4A).[322,323] In fact, mutation of arginine residues within the bridge helix disrupts key ionic interactions with the phosphate backbone of the sgRNA guide region resulting in reduced DNA cleavage ability of Cas9.[317] Moreover, biochemical and biophysical studies suggest that sgRNA truncations lacking the hairpins or, more interestingly, segments of the guide region can inhibit REC lobe conformational dynamics, despite their high affinity binding to Cas9.[323,324] These findings, along with the strong sequence conservation of residues in the bridge helix among Type II Cas9 variants,[317] suggest that the bridge helix is essential for conformationally activating and licensing Cas9 for interference. Interestingly, atomic-force microscopy (AFM) and single-molecule Förster resonance energy transfer (FRET) studies have shown apo-Cas9 to be globally flexible.[325,326] Nevertheless, pre-steady-kinetic and single-molecule FRET work monitoring the dynamic rearrangement of the REC lobe indicate a two-step induced-fit mechanism (Figure 3.4B) for sgRNA binding that is rate-limited by the conformational change (k_2, Scheme 3.1).[320]

3.3.2 DNA interrogation: PAM recognition and on-target DNA binding.

The DNA-target search mechanism following formation of the Cas9•sgRNA effector complex, including recognition of the PAM sequence and rejection of off-target sites, has been well-studied for its relevance in understanding off-target cleavage during gene editing. Single-molecule experiments suggest that Cas9•sgRNA utilizes three-dimensional diffusion (*i.e.* stochastic, dynamic binding and unbinding) to locate target sites, rather than facilitated diffusion (*i.e.* one-dimensional sliding along DNA).[304] However, the target search space is limited by an inherent bias for binding to PAM-containing DNA sequences while also minimizing interactions with sites devoid of the 5′-NGG-3′ motif. As previously mentioned, the PAM is molecularly inspected through base specific interactions with R1335 and R1333 of the CTD (Figure 3.3). Following recognition of the PAM, complementarity between the guide-region of the sgRNA and the DNA target is tested through local, DNA-duplex unwinding beginning with PAM-proximal DNA bases and extending unidirectionally towards the 5′-end of the guide-RNA to generate an R-loop structure.[304,327]

On-target DNA binding has been shown to follow a two-step induced-fit binding mechanism (Figure 3.5A, Scheme 3.1) with kinetics dependent upon divalent cation.[320,328] Interestingly, in the absence of catalytic Mg^{2+}, DNA binding kinetics are ~750-fold slower than in the presence of Mg^{2+}, suggesting a role for divalent cations in DNA unwinding and heteroduplex formation.[320] However, this function of divalent cations is not clearly understood and should be further explored in future work. DNA binding is accompanied by concomitant motion of the HNH domain, which occurs

rapidly following initial DNA association (Figure 3.5A) in the presence of Mg^{2+}.[317,322] Based on comparison of binary (Cas9•sgRNA) and partial ternary (Cas9•sgRNA•DNA with PAM-containing partial duplex DNA) crystal structures, approach of the HNH domain to the scissile phosphate is occluded by the REC2 domain which rotates outward upon Cas9 binding to on-target DNA (Figure 3.3).[318,322]

Notably, several groups have reported a non-productive Cas9 complex (~15% of the population) during DNA binding with evidence suggesting either improperly-formed binary ($Cas9^N$•sgRNA, Scheme 3.1) or ternary ($Cas9^N$•sgRNA•DNA) complexes that must change conformation before DNA binding or cleavage can occur, respectively.[306,320,328] Further work is required to determine the precise identity and physical significance of this non-productive population, but it may originate from failed dynamics of the bridge-helix during sgRNA binding (Figure 3.4A) or a partially formed R-loop during heteroduplex formation.

3.3.3 Fidelity of DNA targeting.

Mismatches between the sgRNA and DNA target can have different effects on DNA interrogation depending on their number and position relative to the PAM. Specifically, PAM-distal mismatches (12 – 20 nucleotides away from the PAM) are permissive, while PAM-proximal mismatches (1 – 8 nucleotides away from the PAM) are prohibitive, to DNA binding.[300,304-306,324,329] Although the equilibrium dissociation constant of fully-matched on-target DNA from the Cas9•sgRNA•DNA ternary complex has been measured to be ~1 – 2 nM, such ultra-stable binding can occur for PAM-containing DNA targets with approximately eight consecutive PAM proximal matches,

regardless of further complementarity at the PAM distal end.[279,304-306,320,324,328] This lack of further contribution to binding beyond the eighth PAM proximal match suggests that the energy barrier required for formation of additional RNA–DNA base pairs falls well below that of DNA rewinding (*i.e.* re-hybridization of tDNA and ntDNA) (Figure 3.5B).[304,306] Importantly, single-molecule studies have shown that requirements for DNA cleavage are stricter than those for tight DNA binding (*i.e.* eight base pairs of PAM-proximal nucleotides are sufficient for tight binding, but ~sixteen matched base pairs (*i.e.* ≤ ~four PAM distal mismatches) are required for robust DNA cleavage) (Figure 3.5B).[300,304-306,324,329] The functional significance of this finding is demonstrated by studies conducted with dCas9 showing a greater abundance of off-target binding sites than those predicted by off-target cleavage assays and represents a major area for improvement.[307-309]

3.3.4 Nuclease domain activation and DNA cleavage.

Following on-target DNA binding and R-loop formation, concerted activation of the HNH and RuvC nuclease domains results in cleavage of the target and non-target DNA strands, respectively (Figure 3.6, Scheme 3.1).[279,300,320,324,329] While the HNH domain strictly cleaves tDNA at a position three nucleotides away from the PAM, RuvC exhibits multiple specificity, cleaving ntDNA at sites 3-4 nucleotides from the PAM with comparable efficiency, resulting in heterogeneous DNA ends following initial cleavage (Figure 3.6C).[319] HNH (His-Asn-His) nucleases are known to cleave nucleic acid substrates through a single-metal-ion mechanism while RuvC nucleases employ a two-metal-ion mechanism, which was confirmed for each Cas9 nuclease domain.[317,328]

Importantly, the cleavage activities of the HNH and RuvC domains of Cas9 occur independently from one another as demonstrated by H840A and D10A nickase mutations which disable target or non-target strand cleavage, respectively, through disruption of metal ion coordination.[279] However, multiple reports have documented conformational communication between HNH and RuvC important for licensing strand scission.[300,320,324,326,329,330] This conformational coupling between the domains appears to be mediated through a dynamic helix located at the junction between HNH and RuvC (residues 909 – 940 of the RuvCIII) (Figure 3.6A). In fact, *in silico* experiments suggest that interactions between this dynamic helix and ntDNA within the RuvC domain drive conformational changes in the HNH domain.[330,331] Pre-steady-state kinetic studies monitoring rapid tDNA and ntDNA cleavage from a pre-formed Cas9•sgRNA•DNA complex indicate that a dynamic equilibrium ($Cas9^{A}$•sgRNA•DNA ⇌ $Cas9^{B}$•sgRNA•DNA, Scheme 3.1) between at least two structurally distinct conformations (Figure 3.6A) of the HNH (HNH^{A}•tDNA ⇌ HNH^{B}•tDNA, Scheme 3.1) and RuvC ($RuvC^{A}$•tDNA ⇌ $RuvC^{B}$•tDNA, Scheme 3.1) domains is established during the initial pre-incubation period in the absence of Mg^{2+}.[320] Based on the forward and reverse rate constants assigned to HNH and RuvC conformational transitions, the unfavorable conformational equilibrium of RuvC (k_{5b}, k_{-5b} Scheme 3.1) is overcome by the favorable equilibrium of the HNH domain (k_{5a}, k_{-5a}, Scheme 3.1) that is stabilized by Mg^{2+} binding (k_6, Scheme 3.1) and leads to product formation (k_{chem}, Scheme 3.1).[320]

3.3.5 Fidelity of DNA cleavage.

Single-molecule FRET studies monitoring motion of the HNH domain demonstrated an intermediate FRET state when Cas9 was bound to an on-target DNA substrate in the absence of Mg^{2+}, in addition to the expected low- (corresponding to HNH in an open conformation far from the cleavage site, Figures 3.3D, 3.3E and 3.5A) and high-FRET (corresponding to HNH docked at the cleavage site, Figures 3.3, 3.6A, and 3.6B) states.[300,329] This intermediate FRET state was suggested to represent a distinct conformation of the HNH domain wherein access to the scissile phosphate is occluded by the REC2 domain and was described as a checkpoint to prevent off-target DNA cleavage as it was predominately observed when DNA targets contained four PAM-distal mismatches even in the presence of catalytic Mg^{2+}.[300,329] However, this intermediate FRET state may be an artifact of insufficient time resolution needed to resolve the potentially rapid transition kinetics of the HNH domain between the open and docked conformations (*i.e.* time-averaged ensemble of FRET states). Conversely, a comparative smFRET analysis of wild-type and engineered high fidelity Cas9 mutants has shown that unwinding and rewinding dynamics of the heteroduplex are perturbed by mismatches and may act as the checkpoint guarding against off-target cleavage (assuming cleavage can only occur after full heteroduplex formation),[306] rather than the purported intermediate conformation of the HNH domain.[300,329] This hypothesis regarding the mechanism of DNA cleavage fidelity further claimed that HNH motion and DNA unwinding do not occur concurrently and was supported by observations[329] that the HNH domain is incapable of achieving its near-cleavage conformation in the absence of Mg^{2+} despite

separate evidence that complete heteroduplex formation can still occur under the same conditions.[306] However, this notion is challenged by an X-ray crystal structure of Cas9•sgRNA•DNA showing full heteroduplex formation with the HNH domain in the near-cleavage conformation even in the absence of Mg^{2+} (Figure 3.6B).[327] Moreover, pre-steady-state FRET studies have demonstrated that Mg^{2+} is not required for HNH to adopt the near-cleavage conformation (Figure 3.6B) but accelerates the kinetics of DNA binding and induces a small (~10 Å) conformational change from the pre-formed near-cleavage state to the reaction state (k_6, Scheme 3.1) .[320] Accordingly, these competing mechanisms of DNA cleavage fidelity, including heteroduplex unwinding/rewinding dynamics and the HNH domain motions, may not be mutually exclusive but part of a two-fold fidelity mechanism. For example, the extent of DNA unwinding (*i.e.* R-loop formation) may correlate to the extent of HNH domain movement towards the scissile phosphate and thus both contribute to the overall cleavage fidelity. Indeed, previous cellular work identified that truncated sgRNAs with guide-regions of ~17 nucleotides in length were capable of reducing mutagenesis at off-target sites following Cas9-mediated gene editing by 5,000-fold without forfeiting on-target editing efficiency.[332] Thus, it may be that the inability to form a fully-extended R-loop when using these truncated sgRNAs inhibits the HNH domain dynamics and, consequently, DNA cleavage. This hypothesis is supported by single-molecule FRET studies showing less excursions of the HNH domain to the high-FRET docked conformation when using truncated sgRNAs only capable of near-complete R-loop formation despite full complementarity with the DNA target.[329] The inhibition of HNH domain dynamics caused by insufficient R-loop formation when

using truncated sgRNAs may exasperate the free-energy penalty caused by mismatches and thus increase overall cleavage fidelity (Figure 3.5B).

3.3.6 Post-cleavage DNA trimming and product release.

Following double-stranded DNA cleavage, Cas9 exhibits a remarkably slow product dissociation rate (k_7, Scheme 3.1) such that on biologically significant time scales Cas9 may be considered a virtual single-turnover enzyme.[319,320] Moreover, several studies indicate that the nuclease domains remain in their activated states following initial DNA cleavage until the HNH domain transitions back to its pre-catalytic conformation upon dissociation of the PAM-distal ntDNA fragment, as supported by the *in silico* results stressing the importance of the ntDNA in conformationally activating the HNH domain.[319,320,329,330] The slow rate of product release, coupled with the persistent activation of the nuclease domains facilitates kinetically-significant, bidirectional (3′→5′ and 5′→3′) post-cleavage trimming (PCT) of the flexible ntDNA products by the RuvC domain (Figure 3.6C).[279,319] Based on the relatively fast kinetics of PCT (~100 – 8000-fold) compared to DNA product dissociation, Cas9 ultimately generates a PAM-proximal DNA fragment with a blunt end and a PAM-distal fragment with a staggered end, 3′-recessed by up to seven nucleotides on the ntDNA strand (Figure 3.6C).[319] This pattern of trimming likely influences the final sequences of indels resulting from repair of DNA breaks generated by Cas9 and may present an important consideration when designing gene editing experiments.

3.4 Future Directions for Cas9 Research

Within a very short amount of time (~six years), an explosion of research has emerged describing the structure and function of Cas9. This body of work has informed the development of engineered variants of Cas9, modified guide-RNAs, and bioinformatics tools for designing, executing, and analyzing CRISPR-based gene editing experiments. Together, these adaptations have greatly improved the safety and efficiency of CRISPR-based technologies. However, mechanistic questions remain in several areas which should be the focus of future research to better understand and improve Cas9 for future clinical applications. As noted above for *S. pyogenes* Cas9, the physical nature of the non-productive complex ($Cas9^{N}$•sgRNA, Scheme 3.1) observed during DNA binding as well as the role of metal ions during DNA binding and nuclease activation are unknown. In addition, we perceive three main topics that should be further explored for *S. pyogenes* Cas9 which are highlighted below. Although beyond the scope of this perspective, there is a large gap between the mechanistic understanding of *S. pyogenes* Cas9 and the relatively understudied homologues belonging to Class 2, Types II, V, and VI (*e.g. Staphylococcus aureus* Cas9, Cas12a, Cas12b, Cas12c, Cas13a, *etc.*) that should also be addressed to potentially expand the CRISPR toolbox.

3.4.1 Increasing the target landscape.

The most critical functional consideration limiting the targetable space for genome engineering with Cas9 is the strict PAM requirement (5′-NGG-3′).[279] On average, a 5′-NGG-3′ sequence occurs once per eight base pairs in the human genome.[333] While this makes many genomic regions accessible to targeting by Cas9 to generate gene

knock-out mutants, the occurrence of 5′-NGG-3′ within A-T rich genomic regions may be much less common making them difficult to target. Furthermore, gene knock-in experiments (Figure 3.2B) wherein the specific location of the DNA break is critical for successful generation of desired knock-in mutants suffers more from restrictions imposed by the 5′-NGG-3′ PAM than gene knock-out experiments. In addition, site-specific base changes using Cas9-mediated base editors (Figure 3.2E) is necessarily restricted to a narrow, optimized sequence window downstream of the PAM nearby the target base pair making many SNPs inaccessible due to the PAM requirement.[293-295,302] As mentioned previously, PAM recognition is achieved through base-specific interactions between the deoxyguanosine nucleotides and two arginine residues within the CTD of Cas9 (Figure 3.1). This atomic-level detail has been exploited to expand the targetable genomic space through rational engineering and structural analysis of Cas9 variants recognizing non-canonical PAMs such as 5′-NAAG-3′, 5'-NGA-3', 5'-NGAG-3', and 5'-NGCG-3'.[311-313] While this approach has generated a toolbox of several Cas9 variants with different PAM specificities, a single Cas9 mutant with broad PAM specificity is highly desirable for multiplexing genome engineering and streamlining gene editing. Remarkably, recent work was successful in using phage-assisted evolution to generate a single variant of Cas9 capable of recognizing a broad spectrum of PAM sequences including 5′-NG-3′, 5′-GAA-3′, and 5′-GAT-3′.[302] Together, these Cas9 variants have increased access to SNPs corrected by C:G→T:A (total of 4,422) conversion from 26% to 77% and those corrected by A:T→G:C conversion (total of 14,969) from 28% to 71%, thereby greatly improving the clinical utility of Cas9 base editors.[302] Despite these improvements, there are still

>5,300 SNPs that are inaccessible necessitating further research in this area including directed evolution, rational engineering, and exploration of Cas enzymes from other species.

3.4.2 Achieving high-fidelity DNA targeting and cleavage.

One limitation of Cas9 that has been a major concern since its initial application in gene editing is off-target site binding and cleavage. As previously mentioned, recent work to rationally engineer Cas9 has resulted in notable reduction in off-target cleavage.[300,301,303] These efforts have mainly focused on reducing non-specific binding energy through disrupting key sequence-independent interactions between Cas9 and the sgRNA:DNA heteroduplex or the orphaned non-target strand such that recognition primarily occurs at the level of complementarity to the RNA guide. Additionally, as the REC2 domain appears to regulate the approach of the HNH domain to the scissile phosphate on tDNA, one unique strategy focused on tuning the conformational cross-talk between these two domains during DNA interrogation to enhance cleavage fidelity by mutating clusters of amino acid residues at the interface of the heteroduplex.[300] In fact, it was shown for this Cas9 variant that, even with fully-matched DNA targets, the activated HNH domain conformation (*i.e.* docked at the tDNA cleavage site) could not be achieved. Thus, it may be that forward transition kinetics of HNH to the activated conformation are slow relative to the reverse kinetics as a result of impaired dynamics of the REC2 mutant, and brief excursions to the activated state facilitate DNA cleavage for on-target DNA but are rare in the presence of mismatches. While all of these Cas9 variants appear to improve DNA cleavage fidelity to varying degrees, off-target DNA

cleavage still occurs and this problem will continue to impede clinical development. Future structural efforts to capture wild-type or engineered Cas9 bound to sgRNA and mismatched DNA may provide clear insight into the mechanism of DNA cleavage fidelity and will be instrumental in developing Cas9 variants exhibiting no off-target effects while retaining high on-target efficiency. However, if the HNH domain is rapidly fluctuating between open and docked conformations in the presence of mismatches, rather than stably positioned in the purported checkpoint conformation, these structures may be difficult to capture.

As previously mentioned, perhaps more troublesome than off-target DNA cleavage by Cas9 is its sequestration at off-target locations (*i.e.* binding but no DNA cleavage).[307-309] While much research has attempted to limit off-target DNA cleavage by Cas9[300,301,303] relatively little has sought to improve its DNA targeting fidelity. Indeed, ultra-stable DNA binding is licensed with as few as eight PAM proximal matches to the guide-RNA while DNA cleavage requires ~17 matches (Figure 3.5B).[304-306] Such binding promiscuity may negatively impact imaging of genomic loci (Figure 3.2C) or transcriptional modulation (Figure 3.2D) with Cas9, and even lower the effective concentration of Cas9•sgRNA during gene-editing experiments leading to undesirable outcomes. It will be exciting to see whether the successful directed evolution strategies recently employed to expand PAM specificity will be exploited to decrease off-target binding and cleavage.[302] Paradoxically, some Cas9 variants intended to alter PAM specificity also exhibited lower off-target effects relative to wild-type Cas9 and the mechanistic basis of this should be delineated.[302] One possibility is that wild-type Cas9

may have evolved to exhibit a specificity sufficient for the small bacterial genome but not suitable to avoid off-target effects in the much larger mammalian genome and is thus poised for greater specificity during directed evolution.[302] Along these lines, bacterial species with larger genomes may possess more specific Cas9 proteins.[334]

3.4.3 Understanding the influence of Cas9 on normal DNA transactions.

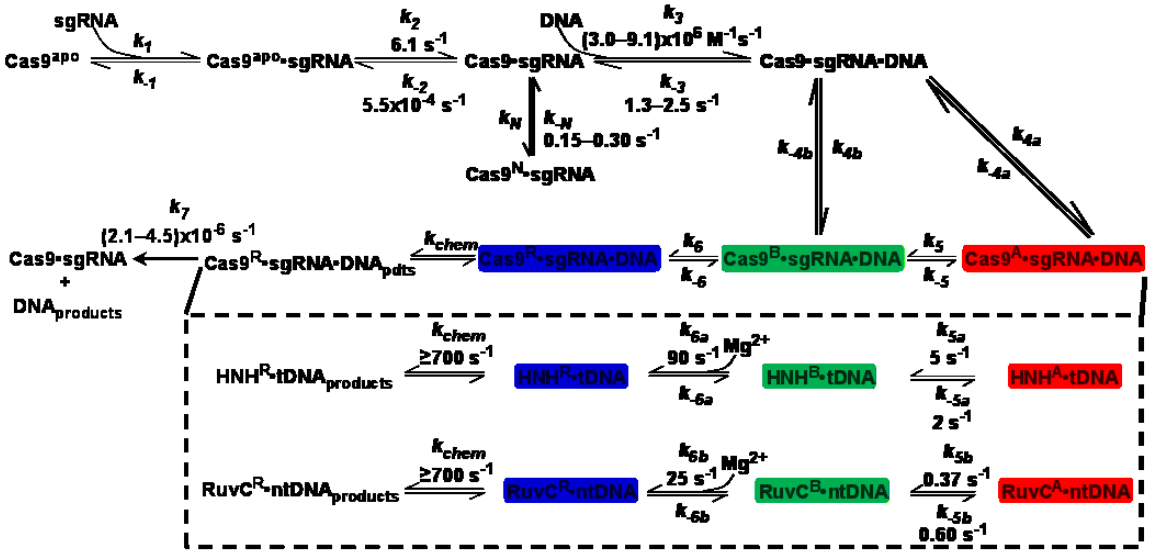
It is very likely that Cas9 interacts with a host of cellular proteins. In fact, ultra-stable binding of Cas9 to DNA ($t_{1/2} \approx 6.5$ hours measured in the absence of Mg^{2+})[320] results in a steric block that prevents binding and/or translocation of proteins along DNA (Figure 3.2F). Even after DNA cleavage has occurred, Cas9 exhibits remarkable long-lived binding (k_7, Scheme 3.1, $t_{1/2} \approx 58$ hours)[320] to the DNA products requiring strong denaturants to disrupt the binding interaction.[304,320] Accordingly, it is expected that Cas9 should block action of DNA repair enzymes on newly-cut DNA and thus delay or inhibit gene editing. However, despite the very slow rate of DNA product release, gene editing can be detected within in as little as 30 minutes of delivery of Cas9 and sgRNA suggesting a cellular mechanism for ejection of Cas9 from DNA cleavage products.[335] It may be that Cas9 is physically ejected from DNA by ATP-driven proteins involved in DNA replication and repair or is targeted for degradation by the proteasome, but such hypotheses remain to be investigated. Moreover, low efficiencies for gene editing and modest toxicity of Cas9•sgRNA observed in certain cell types could result from the steric block imposed by Cas9 on DNA replication or repair and should be explored in future.[334,336] One factor shown to influence efficiency of gene editing is the chromatin state of the DNA at the target site. Indeed, euchromatic DNA targets are more accessible

than highly condensed heterochromatic DNA.[337-339] Importantly, it is unclear how the long-lived DNA binding of Cas9 impacts chromatin remodeling in cells. Through fusion of Cas9 to chromatin remodelers or epigenetic factors it may be possible to increase success of gene editing experiments and allow for site-specific manipulation of the chromatin state.

3.5 Conclusion

Here we have described some of the structural and mechanistic work that has laid the foundation for functional utility of CRISPR/Cas9 in a diverse array of applications. Additionally, we address certain gaps in mechanistic understanding that may prove useful in the development of CRISPR/Cas9 as a safe and effective therapeutic. As the field continues to rapidly evolve, we are excited to see how researchers consider the issues posed here and discover new innovative uses of this revolutionary technology.

3.6 Schemes



Scheme 3.1: Proposed kinetic mechanism of CRISPR/Cas9.

Rate constants were measured using steady-state and pre-steady-state kinetic methods.[320] The dashed box represents conformational heterogeneity in the HNH (HNH^A, HNH^B, HNH^R) and RuvC ($RuvC^A$, $RuvC^B$, $RuvC^R$) domains of Cas9•sgRNA•DNA ($Cas9^A$, $Cas9^B$, $Cas9^R$) ternary complexes. The depiction of Mg^{2+} at certain steps denotes the Mg^{2+}-binding-coupled conformational changes of HNH and RuvC.

3.7 Figures

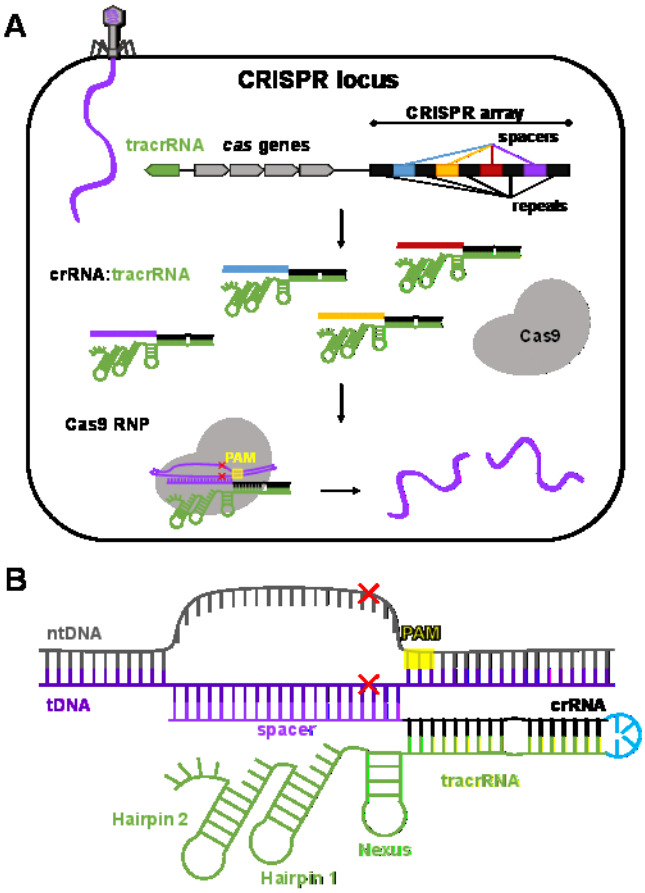


Figure 3.1: Simplified schematic representation of CRISPR-mediated adaptive immunity in prokarya.

(A) Expression of the CRISPR locus occurs to combat viral infection. For Class 2, Type II Cas systems this includes expression of the Cas9 endonuclease and transcription of the CRISPR array, which is composed of repeat sequences (repeats) and unique spacer sequences (spacers) that were acquired during previous viral or plasmid invasion. The transcribed CRISPR array is then processed to generate dual-RNA guides composed of a

(continued)

Figure 3.1: continued

unique crRNA hybridized to a tracrRNA. Cas9 then binds to the dual-RNA guide to generate the ribonucleoprotein (RNP) effector complex. If the invading DNA (double-stranded, purple) contains a protospacer adjacent motif (PAM, yellow) and shares complementarity with the guide region of the crRNA (denoted by blue, gold, red, and purple coloring), the Cas9 RNP will catalyze double-stranded DNA cleavage (denoted by red Xs). (B) Zoomed illustration of the dual-guide RNA base paired with a foreign DNA target. Single-guide RNA is formed by fusing the crRNA and tracrRNA through four nucleotide tetraloop (cyan).

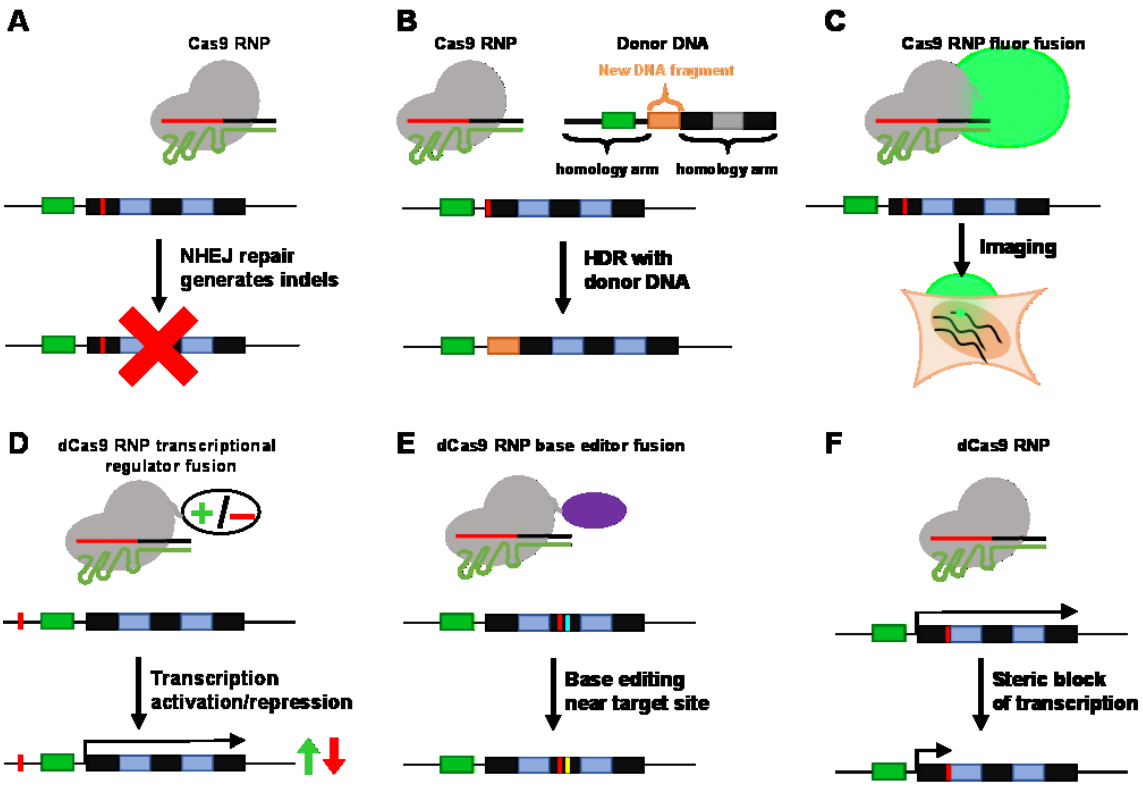


Figure 3.2: Common genetic manipulations in mammalian cells performed with CRISPR/Cas9.

(A) Gene knock-out is achieved by targeting Cas9 to a site (red strip) downstream of the gene promotor (green box) to cleave a crucial exon (black box). Indels could result in deleterious frameshift mutations that eliminate functional gene products (red X). (B) Knock-in of a new DNA fragment through HDR with donor DNA following site-specific cleavage by Cas9. (C) Imaging of genomic loci in cells by targeting Cas9-fluor fusion to a loci interest. (D) Transcriptional control through targeting dCas9, fused to a transcriptional activator (green +) or repressor (red –), to a site (red strip) upstream of a

(continued)

Figure 3.2 continued

promotor region can result in either up-regulation (green arrow) or down-regulation (red arrow) of transcription, respectively. (E) Precision base editing with dCas9 fused to a base editor (purple) to achieve specific mutations (cyan strip changes to yellow strip) near the dCas9 target site (red strip). (F) Transcriptional down-regulation using dCas9 targeted to a gene of interest to block transcriptional activator binding as well as RNA polymerase binding or elongation. Coloring maintained throughout panels A – F. Introns colored blue.

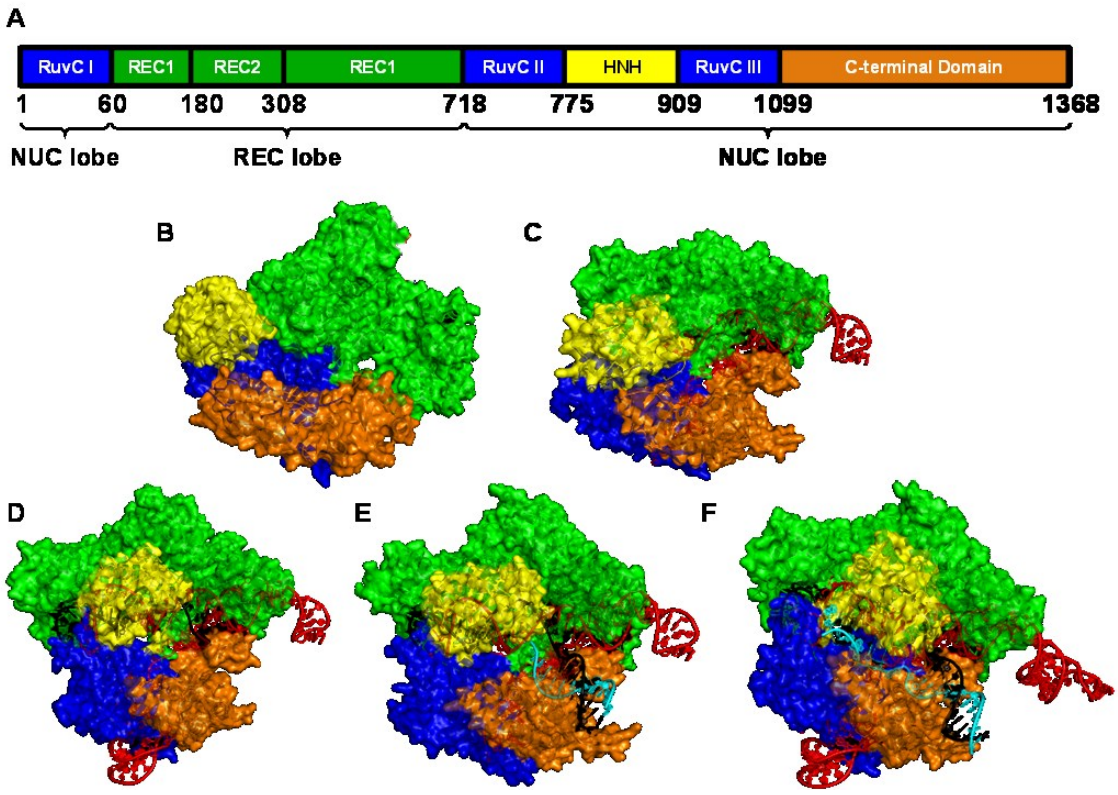


Figure 3.3: Structural snapshots of the CRISPR/Cas9 mechanism.

(A) Simplified diagram of Cas9 domain structure. Crystal structures of (B) apo-Cas9 (PDB: 4CMP), (C) Cas9•sgRNA binary complex (PDB: 4ZT0), (D) Cas9•sgRNA•tDNA partial ternary complex (PDB: 4OO8), (E) Cas9•sgRNA•DNA partial ternary complex with PAM-containing partial duplex DNA (PDB: 4UN3), and (F) catalytically competent Cas9•sgRNA•DNA complete ternary complex (PDB: 5F9R). Cas9 domain coloring as in panel A. sgRNA, tDNA, and ntDNA colored red, black, and cyan, respectively.

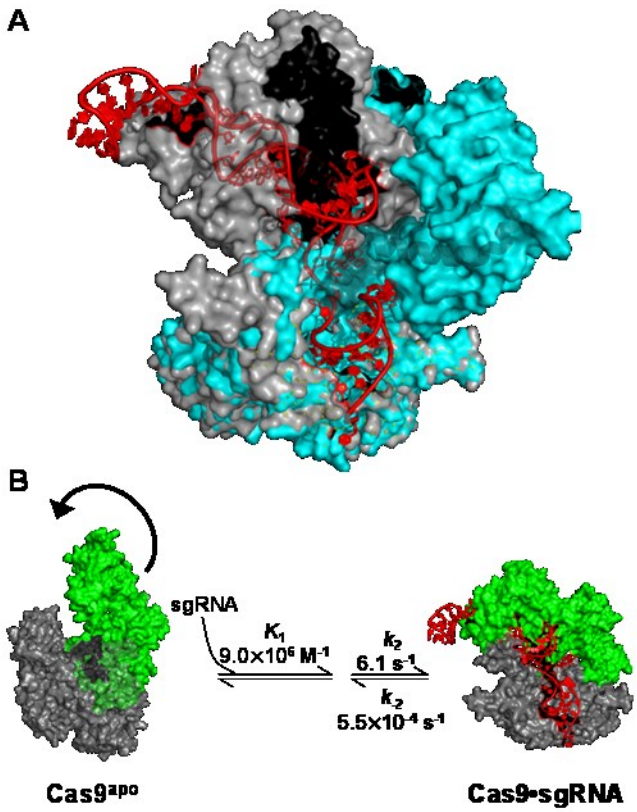


Figure 3.4: Conformational activation of apo-Cas9 upon sgRNA binding.

(A) Superposition of apo-Cas9 (cyan) (4CMP) and sgRNA-bound Cas9 (gray) (4ZT0) aligned via the relatively immobile RuvC domain featuring the dynamic transition of the bridge helix. sgRNA colored red. (B) Two-step induced-fit mechanism for sgRNA binding to Cas9. sgRNA and dynamic REC lobe colored red and green, respectively. Equilibrium (K) and rate constants (k) are derived from Scheme 3.1.[320]

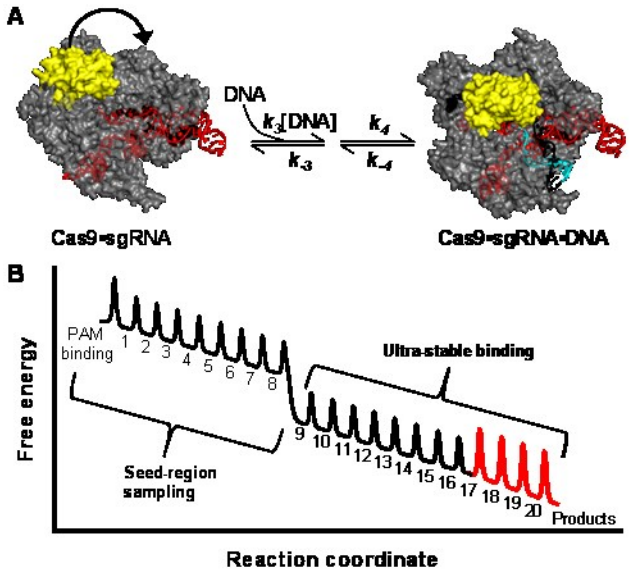


Figure 3.5: DNA interrogation by Cas9•sgRNA.

(A) Two-step induced-fit mechanism for DNA binding. HNH domain and sgRNA colored yellow and red, respectively. Rate constants for DNA binding in the absence and presence of Mg^{2+} were measured using pre-steady-state kinetic methods.[320] (B) Illustrative free energy diagram for DNA interrogation, binding, and cleavage by Cas9•sgRNA. Increasing numbers indicate sequential base pairs formed between sgRNA and tDNA during R-loop formation. Red region represents DNA cleavage-competent state.

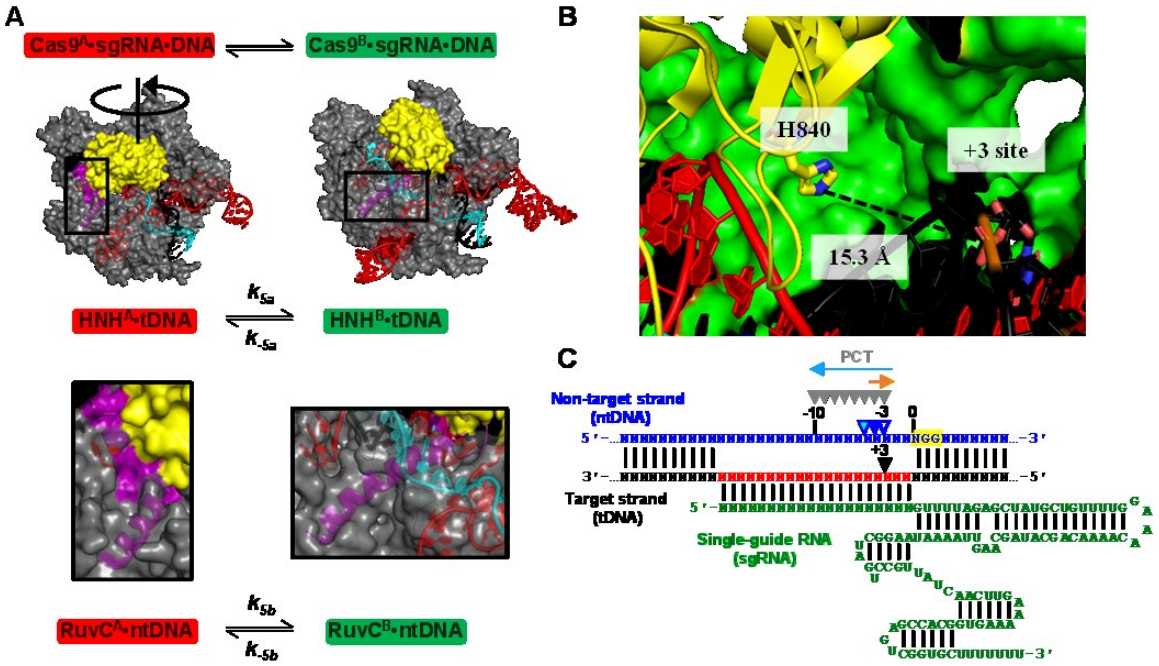


Figure 3.6: Dual nuclease activities of Cas9.

(A) Structural depiction of the transition from early- (Cas9^{A}•sgRNA•DNA, PDB:4UN3) to late- (Cas9^{B}•sgRNA•DNA, PDB:5F9R) stage ternary complexes during Cas9 catalysis. Kinetically favorable conformational changes of the HNH domain (yellow) seem to drive forward the kinetically unfavorable secondary structural changes in the RuvC domain (purple helix, residues 909 – 940) leading to tDNA (black) and ntDNA (cyan) cleavage, respectively.[320] (B) Structural snapshot of the HNH active site in the late-stage ternary complex of Cas9 (Cas9^{B}•sgRNA•DNA, PDB:5F9R) in the absence of Mg^{2+}. The catalytic H840 residue is not quite close enough to the +3 position of the tDNA to coordinate Mg^{2+} for cleavage. (C) Schematic of the sgRNA:DNA heteroduplex with PAM highlighted yellow and 20-base pair complementarity of tDNA and sgRNA colored red. Cleavage

(continued)

Figure 3.6: continued

sites on the tDNA and ntDNA are depicted by black and blue triangles, respectively (filled: preferred cleavage site, open: less favored cleavage site, cyan filled: disfavored site). Directions and extent of RuvC-catalyzed PCT activities are shown by cyan (3′→5′ PCT) and orange (5′→3′ PCT) arrows.

Chapter 4. Single-Molecule Investigation of Response to Oxidative DNA Damage by a Y-Family DNA Polymerase

Reproduced in part with permission from Raper, Austin T., Gadkari V.V., Maxwell B.A., Suo Z. (2016) Single-Molecule Investigation of Response to Oxidative DNA Damage by a Y-Family DNA Polymerase. *Biochemistry.* 55(14):2187-96. DOI: 10.1021/acs.biochem.6b00166.

Author Contributions

Austin T. Raper and Varun V. Gadkari performed all experiments, collected all data, and generated all figures. A.T.R wrote the manuscript with suggestions from V.V.G, Brian A. Maxwell, and Dr. Zucai Suo. B.A.M instructed all authors on the use of the prism-TIR microscope and together with Z.S, provided conceptual direction for the project. All authors contributed to the interpretation of the data.

4.1 Abstract

Y-family DNA polymerases are known to bypass DNA lesions in vitro and in vivo and rescue stalled DNA replication machinery. Dpo4, a well-characterized model Y-family DNA polymerase, is known to catalyze translesion synthesis across a variety of DNA lesions including 8-oxo-7,8-dihydro-2′-deoxyguanine (8-oxo-dG). Our previous X-ray crystallographic, stopped-flow Förster resonance energy transfer (FRET), and computational simulation studies have revealed that Dpo4 samples a variety of global conformations as it recognizes and binds DNA. Here we employed single-molecule FRET (smFRET) techniques to investigate the kinetics and conformational dynamics of Dpo4 when it encountered 8-oxo-dG, a major oxidative lesion with high mutagenic potential. Our smFRET data indicated that Dpo4 bound the DNA substrate in multiple conformations, as suggested by three observed FRET states. An incoming correct or incorrect nucleotide affected the distribution and stability of these states with the correct nucleotide completely shifting the equilibrium towards a catalytically competent complex. Furthermore, the presence of the 8-oxo-dG lesion in the DNA stabilized both the binary and ternary complexes of Dpo4. Thus, our smFRET analysis provided a basis for the enhanced efficiency which Dpo4 is known to exhibit when replicating across from 8-oxo-dG.

4.2 Introduction

One of the most common sources of endogenous DNA damage is aerobic respiration. This essential life process generates oxygen radicals, which are known to cause damage to DNA. For example, 8-oxo-7,8-dihydro-2′-deoxyguanine (8-oxo-dG), a

major oxidative DNA lesion, is formed by the oxidation of the C8 atom of guanine. While structural studies have shown that 8-oxo-dG does not significantly distort the local structure of DNA,[243-245] the lesion remains particularly mutagenic by often being better accommodated in a polymerase active site following rotation about the *N*-glycosidic bond to adopt the unusual *syn*-conformation. While the *anti*-conformation of 8-oxo-dG forms a correct Watson-Crick base pair with correct dCTP, the *syn*-conformation readily forms Hoogsteen interactions with incorrect dATP.[134,246,340] Unique interactions with 8-oxo-dG within the polymerase active site can either promote or limit misincorporation events.[246-251] If left unrepaired, 8-oxo-dG:dA mispairs will result in G→T transversion mutations, which are implicated in cancer induction. [341,342]

DNA lesions can also interfere with faithful DNA replication by stalling high-fidelity replicative DNA polymerases. While 8-oxo-dG does not completely block DNA synthesis, it does cause polymerase pausing in some instances.[185,343-350] It is proposed that this pausing initiates translesion DNA synthesis (TLS) via a polymerase switching mechanism that allows the cell to recruit Y-family DNA polymerases which are known for bypassing lesions *in vitro* and *in vivo*. Notably, the Y-family polymerases have been identified in all three domains of life, *e.g.* four in humans (DNA polymerases η, κ, ι, and Rev1), two in *Escherichia coli* (DNA polymerase IV and V), and one in *Sulfolobus solfataricus* (DNA polymerase IV (Dpo4)).

Pre-steady-state kinetic and crystallographic data indicate that Dpo4 preferentially incorporates dCTP opposite 8-oxo-dG with high efficiency due to stabilization of the *anti*-conformation of the lesion in its active site.[162,250,251,351] Recently, our real-time

stopped-flow FRET studies have revealed how the individual domains of Dpo4 move during substrate binding and catalysis.[11,12,22] Dpo4, akin to all canonical polymerases, contains three core polymerase domains denoted as Finger, Palm, and Thumb. Moreover, characteristic of all Y-family members, Dpo4 also contains an auxiliary domain termed the Little Finger (LF) which is connected to the Thumb domain by a highly flexible 14-amino acid residue peptide linker. The LF domain is known to impart unique DNA binding and lesion bypass capabilities to the Y-family polymerases.[14] While crystallographic investigation suggests that upon nucleotide binding, no large-scale domain movements exist from the binary to ternary structures,[149] our real-time stopped-flow FRET investigations illustrate global conformational changes in Dpo4 during nucleotide binding and incorporation onto undamaged DNA.[11,12,22] These investigations permitted the expansion of the minimal kinetic mechanism for polymerase-catalyzed nucleotide incorporation.

Recently, single-molecule FRET (smFRET) methodologies have been employed to investigate the conformational dynamics of DNA polymerases and have illustrated how domain motions impact substrate recognition and selectivity.[111,163,213,261-264] Specifically, the role of the Finger domain in nucleotide selectivity has been well characterized.[111,261,263] While the Finger domain of Dpo4 has not been exclusively implicated in nucleotide selection, it is part of synchronized, global domain motions which likely play an important role in nucleotide recognition and selectivity. Understanding the mechanism of nucleotide recognition and selectivity is of even greater importance since the Y-family DNA polymerases lack a proof-reading exonuclease

domain. Here, our smFRET study revealed that Dpo4 existed in equilibrium among three FRET states that likely represent distinct structural conformations of Dpo4 and its positions on the undamaged or 8-oxo-dG-containing DNA substrates. The addition of an incoming nucleotide affected the distribution of the FRET states and the DNA binding stability of the polymerase. Our study further established a mechanism by which Dpo4 can faithfully and efficiently bypass a major oxidative lesion.

4.3 Materials and Methods

4.3.1 Preparation of protein and DNA.

We selected a catalytically active mutant of Dpo4 used in previous work with mutations in the Finger domain (C31S, N70C) for the present study.[11,12,22] The engineered Cys mutation allowed for site-specific fluorophore labeling (Figure 4.1B) which was carried out by incubation of Dpo4 with a 15-fold molar excess of Cy5-maleimide (GE Healthcare), overnight at 4˚C in a buffer containing 50 mM Tris (pH 7.2), 150 mM NaCl, 0.5 mM TCEP, and 10% glycerol. Unincorporated free dye was then removed with Micro Bio-Spin columns (Bio-Rad). By measuring the absorbance at 280 nm and 650 nm, the ratio of protein concentration to dye concentration revealed a labelling efficiency of 91% for the reaction.

The 21-mer primer containing a 5′-biotin for surface immobilization as well as a 5-C6-amino-2′-deoxythymidine modification at the 9th base from the 3′ terminus for Cy3-NHS-ester labelling, and control undamaged template were purchased from Integrated DNA Technologies. This labelling was performed according to the manufacturer's protocol (GE Healthcare) (Figure 4.1A). The oligonucleotide containing the 8-oxo-dG

base was purchased from Midland Certified Reagent Company, Incorporated. The primer and template oligonucleotides were annealed to each other by heating to 80˚C for 5 minutes followed by slow cooling to room temperature.

4.3.2 Steady-state fluorescence spectroscopy assays.

Fluorescence spectra were recorded on a Fluoromax-4 (HORIBA Jobin Yvon) at 20˚C. Increasing amounts of Cy5-labeled Dpo4 (0-180 nM) were titrated into a 25 nM solution of the Cy3-labeled DNA substrate (undamaged or damaged) in a buffer containing 25 mM HEPES (pH 7.5), 5 mM $CaCl_2$, 25 mM NaCl and 10 % glycerol. The change in donor quantum yield was plotted against acceptor concentration. The resulting curve was fit to Equation 4.1 to yield the equilibrium dissociation constant for the interaction.

$$\Delta\phi = (\Delta\phi_T/2D_0) \times \left\{\left({K_D}^{DNA} + E_0 + D_0\right) - 0.5\left[\left({K_D}^{DNA} + E_0 + D_0\right)^2 - 4E_0D_0\right]^{1/2}\right\}$$

(Eq. 4.1)

where $\Delta\phi$ is the change in the donor quantum yield, $\Delta\phi_T$ is the maximum change in donor quantum yield, D_0 is the total DNA concentration, E_0 is the total Dpo4 concentration, and ${K_D}^{DNA}$ is the equilibrium dissociation constant.

4.3.3 Single-molecule measurements.

Single-molecule fluorescence experiments were conducted on a custom built prism-type total internal reflection microscope as described previously.[111] Imaging chambers were assembled from quartz slides and coverslips which were cleaned, passivated, biotinylated and coated with Neutravidin (0.2 mg/mL) following a standard, published protocol.[352] The Cy3-labeled DNA was flowed into the imaging chamber at a concentration sufficient to observe single, distinct immobilized molecules (20-40 pM). In

addition to the Dpo4 binding buffer (50 mM HEPES (pH 7.6), 50 mM NaCl, 0.1 mM EDTA, 5 mM $CaCl_2$, and 0.1 mg/mL BSA), experiments were performed in an imaging solution containing an oxygen scavenging system (0.8% (w/v) d-glucose, 1 mg/mL glucose oxidase, and 0.04 mg/mL catalase) and 2 mM Trolox. Ca(II) was used as the divalent metal ion in place of Mg(II) to prevent the incorporation of dNTPs. Ca(II) in the polymerase active site has been shown to closely mimic Mg(II), but prevents or dramatically reduces dNTP incorporation.[353,354] Movies were recorded at 100 ms exposure time resolution, at 20˚C, over several minutes upon introduction of 10-20 nM Cy5-labeled Dpo4 to the imaging chamber. The movies were then processed using IDL (ITT Visual Information Solutions) and analyzed by the use of custom data acquisition and analysis software which included a MATLAB (MathWorks) script necessary for correcting anti-correlated donor and acceptor time trajectories for background (Center for the Physics of Living Cells, University of Illinois at Urbana-Champaign).

4.3.4 Verification of the FRET system.

To confidently correlate changes in FRET with molecular distance changes rather than unrelated photophysical artifacts, we performed several previously described control experiments.[355] We recorded movies of surface immobilized Cy3-DNA following excitation at 532 nm in the presence and absence of WT Dpo4 to show that the donor intensity remains constant until photobleaching between the experiments (Figure 4.2). Further, we recorded acceptor fluorescence upon direct excitation of 10 nM Cy5-Dpo4 with 638 nm laser light. A histogram composed of acceptor intensities from >100 time trajectories revealed one major distribution and, more notably, the absence of higher

intensity distributions eliminating the possibility of multiple Dpo4 molecules binding per DNA substrate (Figure 4.3).

4.3.5 Single-molecule data analysis.

From the donor and acceptor time trajectories, apparent FRET (E_{app}) was

$$\frac{I_A}{I_D+I_A} = E_{app} \qquad \text{(Eq. 4.2)}$$

calculated by Equation 4.2, where I_A is the acceptor intensity and I_D is the donor intensity. Notably, only fluorescence time trajectories with a single photobleaching event were selected for further processing in order to avoid ambiguity in removing background signal. FRET efficiency values from over 200 fluorescence time trajectories with acceptor intensities clearly above background level (as determined by the average background after donor photobleaching) were collected and binned from 0-1 to generate population histograms using a previously described thresholding approach.[111,356] The population distributions were fit to a sum of Gaussian functions using MATLAB, and the percent density of molecules in each respective population was calculated as the total area under each peak. Each histogram was fit with the minimum number of peaks necessary to give a visually acceptable fit. The fit was generally deemed acceptable if the percent error, calculated as the total difference between the experimental bin heights and the computed value of the fit at each bin center, divided by the sum of the bin heights, was less than 10%. However, given the limited signal-to-noise ratio of our single-molecule method and the overlap between FRET distributions, we did not want to bias our analysis to presume a certain number of FRET states. Therefore, to impartially verify our visual assignment of FRET states from the raw trajectory data and the accuracy of the

Gaussian fits, we performed a probability based Hidden Markov modelling (HMM) analysis using HaMMy software.[357] To further avoid biasing the results, the HaMMy software was set to converge on the true number of FRET states in the data from initially guessing the maximum number of FRET states that the program allows (10). We analyzed a subset of single-molecule trajectories (100) for each smFRET experiment by HMM as the HaMMy software was computationally limited in the number of traces that could be processed at one time under our experimental settings (i.e. 10 state model). The HaMMy output files were then additionally processed using accompanying TDP software to generate transition density plots. Kinetic information describing the transitions between states was calculated through the TDP software. Briefly, the transition distributions were fit to Gaussian functions and the calculated peak centers and standard deviations were converted to kinetic rates and rate errors, respectively, by multiplying by the exposure time used during data acquisition (100 ms). HaMMy and TDP were used primarily to complement and verify the results we obtained from the manual analysis of the complete data sets as well as to describe the kinetic rates connecting interconverting FRET states which would be difficult to acquire otherwise.

Dwell time analysis was performed as described previously.[111] Survivor functions were subsequently fit to single (Equation 4.3) or double (Equation 4.4) exponential decay equations.

$$f(t) = A\exp(-kt) \qquad \text{(Eq. 4.3)}$$

$$f(t) = A_1 \exp(-k_1 t) + A_2 \exp(-k_2 t) \qquad \text{(Eq. 4.4)}$$

where *f(t)* is the fraction of molecule remaining bound after time *t*, A_1 and A_2 are the

amplitudes of the phases, and k, k_1, and k_2 are the rate constants of Dpo4 unbinding from the DNA substrate. Fits to the data were evaluated based on the coefficient of determination (R^2) as well as visual inspection. To avoid over interpreting the results, all data were initially assumed to demonstrate single exponential decay behavior. If the R^2 value was less than 0.99 and the fitting curve failed to adequately represent the majority of data, a higher order exponential fit was considered necessary (Eq. 4.4).

4.3.6 Kinetic assays.

Polymerase activity was assessed through a burst kinetic assay at 37˚C in a reaction buffer containing 50 mM HEPES, (pH 7.5 at 37˚C), 50 mM NaCl, 0.1 mg/mL BSA, 0.1 mM EDTA, 5 mM DTT, 10% glycerol, and 5 mM $MgCl_2$. Briefly, 60 nM of 5′-^{32}P-labeled D-1 DNA substrate was pre-incubated with 15 nM of either WT Dpo4 or Cy5-labeled mutant (C31S, N70C) Dpo4 and then rapidly mixed with 100 μM dTTP for increasing amounts of time. All fast reactions were performed using a rapid chemical-quench flow apparatus (Kintek). DNA products were then resolved on a denaturing 17% polyacrylamide gel, scanned using a Typhoon Trio (GE Healthcare), and quantitated with ImageQuant software (Molecular Dynamics). The product formation was then plotted against time (t) and the data were fit to Equation 4.5.

$$[\text{Product}] = A[1-\exp(-k_{burst}t)] + k_{ss}t \qquad \text{(Eq. 4.5)}$$

where A represents the reaction amplitude, k_{burst} the single-turnover nucleotide incorporation rate constant, and k_{ss} the observed steady-state rate constant.

4.4 Results

4.4.1 Design of a FRET system for monitoring Dpo4 interaction with DNA.

To investigate the binding of Dpo4 at the single-molecule level, we adapted a previously designed FRET system to attach a Cy5 fluorophore to the Finger domain of Dpo4 (N70C) and a Cy3 fluorophore to each of the DNA substrates described in Figure 4.1.[11] Notably, Dpo4 is known to bind the blunt-end of DNA with a five-fold lower affinity as compared to the primer/template junction of a double-stranded DNA substrate but does not bind to single-stranded DNA.[358] Therefore, we added a non-complementary frayed end to each of the DNA substrates (Figure 4.1A) used in this article to eliminate off-target binding by Dpo4 which would have complicated data interpretation. Fluorescence spectra recorded upon titration of Cy5-labeled Dpo4 into a solution of Cy3-labeled undamaged DNA or 8-oxo-dG containing DNA (damaged DNA) indicate that the FRET probes are well-positioned to report on the binding of Dpo4 to DNA and that the fluorescent labels and 8-oxo-dG lesion did not affect the DNA binding affinity of Dpo4 (Figure 4.4), as reported previously.[12,134] To determine if the mutations or the label affected the polymerase activity of Dpo4, burst kinetic assays of Cy5-labeled Dpo4 were performed to obtain single-turnover nucleotide incorporation and steady-state rate constants. The rate constants for the labeled enzyme are similar to those of wild-type Dpo4 (Figure 4.5), indicating the mutations and fluorescent label did not alter the polymerase activity of Dpo4.

4.4.2 Investigation of Dpo4 in a binary complex with DNA by smFRET.

To determine if the binding of Dpo4 to single DNA molecules is perturbed by the presence of a non-helix distorting oxidative lesion, 8-oxo-dG, we first conducted smFRET experiments with DNA substrates containing either a dG, or a 8-oxo-dG lesion at the templating position in the DNA (Figure 4.1). Overall, smFRET time trajectories for Dpo4 binding to either DNA substrate were similar, displaying transitions between three, non-zero FRET efficiency values as determined by visual inspection (Figure 4.1D). FRET efficiencies of >200 individual traces from each single-molecule experiment were collected and compiled into population distribution histograms (Figures 4.6A, 4.6B**)**. The Gaussian peak fits of the histograms revealed subpopulations with FRET efficiencies centered at ~0.50 (low-FRET state), ~0.65 (mid-FRET state), and ~0.85 (high-FRET state). Consistently, smFRET investigations of other DNA polymerases, such as *E. coli* DNA polymerase I (Klenow fragment), and *S. solfataricus* Polymerase B1, have also revealed multiple FRET states which suggest various polymerase conformations while bound to DNA.[111,261,355] Interestingly, we observed that the oxidative lesion affected the population distributions represented in the histograms (Figure 4.6B). When bound to the undamaged DNA substrate, most FRET events (51%) were observed to be in the mid-FRET state. However, when an 8-oxo-dG lesion was present, most FRET events (50%) were observed to be in the low-FRET state. Unexpectedly, our data revealed a small, distinct distribution at a FRET efficiency higher than the two previously described.[163] This third state was observed regardless of the DNA substrate used. Given the inherent noise of smFRET data we chose to complement our manual data analysis with a

probability-based Hidden Markov modelling analysis[357] to unbiasedly confirm the number of non-zero FRET efficiency states. The software idealized a subset of FRET trajectories (100) and converged on three, non-zero FRET states as expected from our visual inspection of the single-molecule trajectories (Figure 4.6A, 4.6B). Additional HMM analysis with the program TDP allowed for the construction of two-dimensional transition density plots (TDP) in which initial FRET efficiencies were plotted against final FRET efficiencies for every transition in the data sets (Figure 4.8). Importantly, if the smFRET data contain distinct, reproducible FRET values, then crosspeaks should develop in the TDPs. The corresponding plot in Figure 4.8 clearly shows crosspeaks representing transitions amongst all three states for each of the DNA substrates evaluated. Prominent density in the plots at the high→mid, mid→high, high→low, low→high, mid→low, and low→mid transition areas indicates that Dpo4 was able to freely convert between the three FRET states. Additionally, our TDP analysis gave the shuttling rates for the transitions between the FRET states. Overall, shuttling rate between all states for Dpo4 binding to either DNA substrate were similar (Table 4.1).

To further investigate the response of Dpo4 to undamaged or damaged DNA, a dwell time analysis of the single-molecule binding traces was performed to yield information on the dissociation kinetics of the binary complexes. Dwell time histograms and survivor functions for this analysis are located in Figures 4.9A and 4.10A. The survivor functions were best fit to a double exponential decay equation indicating that binding events for the binary complex (Dpo4•DNA) vary in stability (Table 4.2 and Figure 4.11). Notably, a modest decrease in the fast phase rate constant ($k_{off,1}$) and

increase in the slow phase amplitude (A_2) for Dpo4 dissociation from the damaged DNA substrate relative to the undamaged one suggest a modest increase in the complex stability of Dpo4•DNA caused by the presence of the oxidative lesion.

4.4.3 Effect of dNTPs on Dpo4 binding to undamaged or damaged DNA.

smFRET experiments were performed to probe whether or not the presence of an incoming nucleotide affected the binding of Dpo4 to DNA. Notably, the addition of correct or incorrect dNTP (6 mM) eliminated the high-FRET state that was observed for the binary complex Dpo4•DNA with either the undamaged or damaged DNA substrate (Figures 4.6C, 4.7, 4.12, 4.13). With correct dCTP, smFRET time trajectories clearly displayed a single, sustained FRET state and the population histograms contain one low-FRET distribution (~0.5) (Figures 4.6C and 4.12). Further, HMM and TDP analysis (Figure 4.8) indicate negligible transitions between non-zero FRET states.

Dwell time analysis for Dpo4 binding to either undamaged or damaged DNA in the presence of a correct incoming nucleotide revealed slower (2- to 3-fold) dissociation kinetics compared to the binary complex (Dpo4•DNA) indicating that the correct dNTP stabilizes the ternary complex (Dpo4•DNA•dNTP) (Table 4.2). Dwell time histograms and survivor functions for this analysis are located in Figures 4.9B and 4.10B. Remarkably, the correct nucleotide seemed to have a greater stabilizing effect for Dpo4 binding to the damaged DNA substrate as observed through slower dissociation rates (Table 4.2).

Experiments to observe the effect of an incorrect incoming nucleotide on the conformations and kinetics of Dpo4 binding to undamaged or damaged DNA substrates

were performed to uncover the mechanism of nucleotide selection by Dpo4. smFRET trajectories of Dpo4 binding in the presence of each of the three incorrect nucleotides display short-lived FRET events with two primary FRET efficiencies (Figure 4.13). Population histograms further revealed a bimodal distribution (Figure 4.7). Despite the high concentrations of incorrect nucleotide used in each experiment (6 mM), the bimodal distribution may be a result of an inability fully saturate the enzyme.[93] Dwell time analysis confirmed that Dpo4 binding events to either DNA substrate were transient with fast dissociation kinetics (Table 4.2, Figures 4.9 and 4.10). However, dwell times for Dpo4 binding to the damaged versus to the undamaged DNA substrate in the presence of an incorrect nucleotide were marginally longer lived. While the dwell time histograms for the dissociation of Dpo4 from undamaged or damaged DNA in the presence of incorrect nucleotide fit well to the single exponential decay equation (Equation 4.3), we cannot completely exclude the possibility of longer time scale processes.

4.5 Discussion

Through smFRET investigation, we demonstrated that the binding of Dpo4 to DNA is a complex process. The smFRET trajectories and FRET population histograms for this polymerase binding to an undamaged or a damaged DNA substrate (Figure 4.1A) revealed three FRET states (Figures 4.6A and 3B). Interestingly, our previous stopped-flow FRET studies[11,12,22] have revealed that Dpo4 and DNA forms an initial binary complex (Complex I), which transitions into a ternary complex (Complex II) after an incoming dNTP stimulates the rapid translocation of Dpo4 along the DNA axis by one base pair in order to empty the space within the active site for nucleotide binding

(Scheme 4.1A). Subsequently, synchronized domain motions of Dpo4 help to tighten its "grip" on both the DNA and nucleotide, leading to the formation of Complex III. Prior to phosphodiester bond formation, Complex IV is formed when Dpo4 undergoes the rate-limiting active site rearrangement to reposition key active site residues and properly align all substrates for catalysis (Scheme 4.1A).[93,220] Notably, the binary and ternary crystal structures of Dpo4[149,162] represent Complexes I and IV, respectively, but do not inform on Complexes II and III (Scheme 4.1A). Surprisingly, the primer/template junction base pair in the binary crystal structure occupies the same space as the nascent base pair in the ternary structure,[149] suggesting DNA translocation by one base pair during nucleotide binding. Coupled with the aforementioned stopped-flow FRET[11,12,22] and X-ray crystallographic studies[149,162], we conclude that the mid-FRET state shown in Figures 4.6A and 4.6B likely represents Complex I while the low-FRET state corresponds to the conformation(s) of Dpo4 in Complex III and/or IV wherein DNA has translocated by one base pair and Dpo4 has undergone synchronized domain motions. More importantly, the low-FRET state suggests that Dpo4 might have sampled the conformations in Complexes II-IV throughout the DNA binding process (Dpo4•DNA) in our smFRET experiments, even in the absence of an incoming nucleotide. However, as Complex II is a transient species, existing briefly during the nucleotide binding process, the low-FRET state observed during DNA binding by Dpo4 can only represent Complexes III or IV. Interestingly, the addition of correct nucleotide stabilized the polymerase in the low FRET efficiency state representing Complexes III and IV (Figures 4.6C and 4.12). Notably, Complexes III and IV cannot be distinguished by FRET as they differ by small

active site rearrangements which do not affect the distance between the FRET pair in our system (Figure 4.1).

Consistently, a recently published smFRET study of the binding of Dpo4 to undamaged DNA has revealed two similar FRET states which are correlated to a pre-insertion and an insertion site on the DNA substrate.[163] The pre-insertion site is consistent with our assignment of Complex I. However, those authors have not considered that the low-FRET state represents Complex III and/or IV. In addition, they concluded that the transition rates (1.1-2.7 s^{-1}) between the two FRET states correspond to the rate of DNA translocation.[163] Although our shuttling rates for Dpo4 transitioning between the mid- and low-FRET states with undamaged DNA (2.8-3.3 s^{-1}, Table 4.1) are similar, these rates are too slow to account for the rapid translocation of Dpo4 along DNA. In fact, the translocation was too fast to be accurately measured by our stopped-flow FRET studies at 20 °C, implying a rate of >150 s^{-1}.[11,12] Recent single-molecule investigations[268-270] also suggest that the translocation rate is too fast to be resolved at the frame rates of conventional smFRET cameras (80 ms for the previous study[163] and 100 ms for this study). Accurate rate estimation requires high temporal and spatial resolution such as that obtained in a recent study of phi29 polymerase where single-molecule nanopore technology was used to estimate a DNA translocation rate of >250 s^{-1} at 21 °C.[268-270] Interestingly, the synchronized domain motions of Dpo4 in Step 3 (Scheme 4.1A) occur at rates of ~10 s^{-1} which are reasonably comparable to our shuttling rates (Table 4.1).[11,12,22] Taken together, the shuttling rates obtained here and in the previous study[163] for Dpo4 transitioning between the mid- and low-FRET states are not specifically

reporting on the DNA translocation event, but rather on concomitant domain motions as observed through the Cy5-labeled Finger domain. Thus, our single-molecule investigation yielded results which are consistent with those from ensemble studies[11,12,22] and unveiled a dynamic conformational equilibrium in the domains of Dpo4 that accompanied DNA translocation.

The high-FRET state (0.87) detected in Figure 4.6A was unexpected but confirmed through repeated experiments and unbiased HMM analysis. Analysis of our single molecule data indicates that Dpo4 can bind directly in this high-FRET state as well as access it through either the low-FRET or mid-FRET state (Scheme 4.1B). Notably, the absence of this high-FRET state in the previous smFRET study of Dpo4[163] likely stems from the different fluorophore labelling positions selected for the polymerase and DNA substrates in this work. Although more work is required to characterize this high-FRET state, we can speculate on its identity and significance here. Superposition of the apo and the binary (Dpo4•DNA) crystal structures has revealed large structural changes, especially a dramatic 131° rotation of the LF domain relative to the polymerase core, accompanying DNA binding.[149] The conformational flexibility of the LF domain is facilitated by a highly flexible, 14 amino acid residue linker connecting the LF and Thumb domains of Dpo4. This flexibility is further illustrated in the crystal structure of Dpo4 in complex with a subunit of heterotrimeric proliferating cell nuclear antigen (PCNA) where the LF domain exists in an extended conformation distinct from that in the apo and binary structures.[359] Furthermore, our recent computational analysis of Dpo4 binding to DNA suggests that the LF domain and linker are intimately involved in the

complex, multi-step processes the polymerase utilizes to recognize and bind DNA.[151] Given the inherent plasticity of the LF domain and linker as well as the conformational equilibria between multiple binding states (Schemes 4.1A, 4.2B), we hypothesize that the observed high-FRET state is related to an alternate DNA binding conformation of Dpo4 facilitated by both the dynamic LF domain and the flexible linker (Scheme 4.2A). Interestingly, this high-FRET state is eliminated in the presence of dNTPs, suggesting that this binding mode may be non-productive and the presence of the nucleotide stimulates its conversion into a catalytically productive conformation at the primer/template junction. Alternatively, Dpo4 may sample other non-productive conformations, such as the polymerase binding to DNA in an inverted conformation which would shorten the distance between the Cy5 in the Finger domain and the Cy3 in the DNA substrate (Scheme 4.2B). Consistently, multiple binding orientations for an enzyme bound to a nucleic acid substrate have been previously observed by smFRET.[360,361] Research is currently under way to explicitly characterize this high-FRET state.

By introducing an incoming nucleotide to the binary complex of Dpo4 with undamaged or damaged DNA, we revealed the stringent nucleotide selectivity of Dpo4 through smFRET experiments. Correct dNTP stabilized the binding of Dpo4•DNA as revealed by the complete shift of its FRET distributions to the low-FRET state (Complexes III and IV in Scheme 4.1A) (Figures 4.6C and 4.12) and more than 2-fold slower dissociation kinetics of Dpo4 from undamaged or damaged DNA (Table 4.2). Contrastingly, upon the introduction of an incorrect dNTP, we observed a bimodal

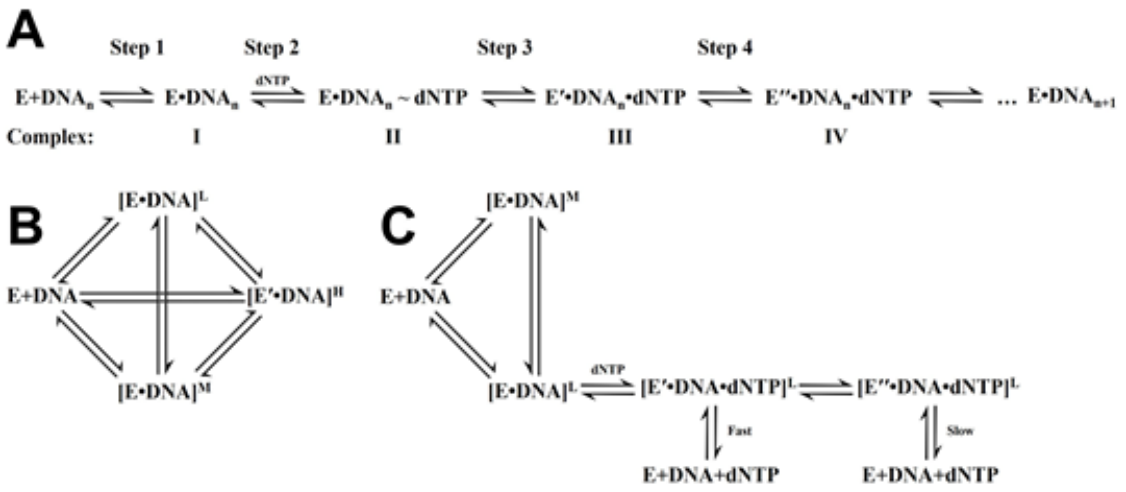
population distribution as a result of incomplete population shifts to the low-FRET state (Figure 4.7). Our dwell time analysis revealed that Dpo4 remained bound to DNA 5- to 10-fold longer in the presence of a correct over an incorrect nucleotide (Table 4.2). Additionally, observed biphasic decay kinetics in the presence of correct dCTP (Figures 4.9B and 4.10B, Table 4.2) suggest the existence of the loose (E′•DNA•dNTP) and tight (E″•DNA•dNTP) ternary complexes, which is consistent with our previously established kinetic mechanism (Scheme 4.1C).[93,94] Interestingly, only fast, single exponential decay kinetics were observed in the presence of an incorrect nucleotide, suggesting the preferred collapse of the loose ternary complex (Figures 4.9 and 4.10, Table 4.2). Thus, our single-molecule analysis indicates that correct nucleotide binding allowed the formation of the catalytically competent, tight complex (E″•DNA•dNTP). Interestingly, two competition experiments were conducted wherein the Dpo4•DNA complex was incubated with equal concentrations of either the three incorrect dNTPs or all four dNTPs. In both experiments, the FRET population distribution was bimodal. However, more FRET events were in the low-FRET state in the presence over absence of correct dCTP (83% versus 58%, Figure 4.14). Taken together, these smFRET experiments demonstrated that Dpo4 was able to differentiate between correct versus incorrect dNTPs during nucleotide binding.

Previously, our ^{32}P-based kinetic assays monitoring the 8-oxo-dG bypass have revealed that Dpo4 exhibits surprisingly high fidelity and efficiency. In fact, the efficiency of correct incorporation opposite 8-oxo-dG is higher than that opposite the undamaged templating dG.[134] Notably, our smFRET time trajectories are similar between

Dpo4 binding to undamaged versus damaged DNA (Dpo4•DNA) but a smaller low-FRET state population was observed with the former (35% versus 50%, Figures 4.6A and 4.6B). More importantly, the dissociation kinetics of ternary complexes with correct dCTP show a larger population of slower dissociation events from the damaged over undamaged DNA substrates (69.8% versus 53.1%, Table 4.2). In the presence of either correct or incorrect nucleotides, ternary complex dissociation is noticeably slower from the damaged than the undamaged DNA substrate (Table 4.2). Consistently, the ternary crystal structure of Dpo4 with DNA containing a templating 8-oxo-dG and an incoming dCTP shows an auxiliary hydrogen bond and an ion-dipole pair between Arg332 and 8-oxo-dG.[162] Overall, our smFRET analysis indicates that Dpo4 preferentially binds DNA containing 8-oxo-dG through additional stabilizing interactions with the lesion. Interestingly, it has been previously hypothesized that Y-family DNA polymerases derive their lesion bypass specificity through unique active site residues which are optimal for their interactions with the templating lesion(s).[228,362] Consistently, our smFRET experiments help to explain the preference of Dpo4 for bypassing 8-oxo-dG which constantly challenges the *Sulfolobus solfataricus* replication machinery *in vivo*.[363]

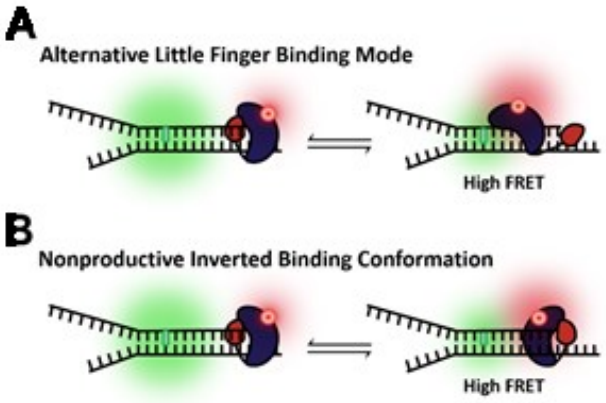
Notably, comparison of FRET efficiency histograms with damaged DNA (Figure 4.7) shows no preference of Dpo4 for any particular incorrect nucleotide, including dATP despite its capability of forming a Hoogsteen base pair with 8-oxo-dG. Consistently, ternary crystal structures show that Arg332 stabilizes the *anti*- over *syn*-conformation of 8-oxo-dG at the active site of Dpo4 through the hydrogen bond and the ion-dipole pair discussed above.[162,250]

4.6 Schemes



Scheme 4.1: Proposed Mechanism of Binary and Ternary Complex Formation.

(A) Previous proposed mechanism of nucleotide incorporation. E, E′, and E″ are different forms of the enzyme. Steps after 4 were not shown for clarity of discussion. (B) Binary complex formation. $[E•DNA]^L$, $[E•DNA]^M$, and $[E•DNA]^H$ refer to low-, mid-, and high-FRET binary complexes, respectively. (C) Ternary complex mechanism. $[E′•DNA•dNTP]^L$ and $[E″•DNA•dNTP]^L$ refer to two different forms of the ternary complex.



Scheme 4.2: Conformational sampling of the high-FRET state.

(A) The LF domain of Dpo4 has been shown to adopt multiple structural conformations. The high-FRET state may be an extended conformation of the LF domain. Proximity between fluorophores increases during transition from a compact binary complex to an extended conformation. The LF binds DNA autonomously allowing the core domains to sample conformational states and DNA surface area as facilitated by the flexible linker region (B) Dpo4 may adopt a nonproductive conformation in which the polymerase is incorrectly oriented on the DNA resulting in high-FRET as the proximity between Cy5 in the Finger domain and Cy3 on the substrate increases.

4.7 Tables

Table 4.1: Shuttling rates from transition density plot

Transition	Undamaged Substrate (s^{-1})	8-oxo-dG modified Substrate (s^{-1})
L→M	2.8 ± 0.7	3.2 ± 0.8
L→H	1.5 ± 0.4	1.0 ± 0.4
M→L	3.3 ± 0.8	4 ± 2
M→H	1.8 ± 0.7	1.8 ± 0.5
H→L	1.8 ± 0.4	5 ± 2
H→M	1.9 ± 0.5	3 ± 1

L = Low-FRET, M = Mid-FRET, H = High-FRET

Table 4.2: Dwell time analysis of Dpo4 binding to DNA

	Average Binding Time (s)	A_1	$k_{off,1}$ (s^{-1})	A_2	$k_{off,2}$ (s^{-1})
Undamaged DNA Substrate					
DNA	1.3	0.51 ± 0.02 (50.5%)	2.5 ± 0.1	0.50 ± 0.02 (49.5%)	0.50 ± 0.01
DNA + dCTP	2.7	0.46 ± 0.02 (46.9%)	6.3 ± 0.5	0.52 ± 0.01 (53.1%)	0.215 ± 0.005
DNA + dATP	0.3	1.01 ± 0.01	5.0 ± 0.1	-	-
DNA + dGTP	0.2	1.02 ± 0.02	6.2 ± 0.2	-	-
DNA + dTTP	0.3	1.02 ± 0.02	5.7 ± 0.2	-	-
Damaged DNA Substrate					
DNA	1.6	0.38 ± 0.04 (38.8%)	1.7 ± 0.01	0.60 ± 0.04 (61.2%)	0.46 ± 0.02
DNA + dCTP	4.2	0.29 ± 0.01 (30.2%)	1.16 ± 0.08	0.67 ± 0.01 (69.8%)	0.185 ± 0.003
DNA + dATP	0.8	0.97± 0.02	1.38 ± 0.04	-	-
DNA + dGTP	0.6	1.04 ± 0.02	2.2 ± 0.07	-	-
DNA + dTTP	0.9	0.92± 0.02	1.13 ± 0.03	-	-

4.8 Figures

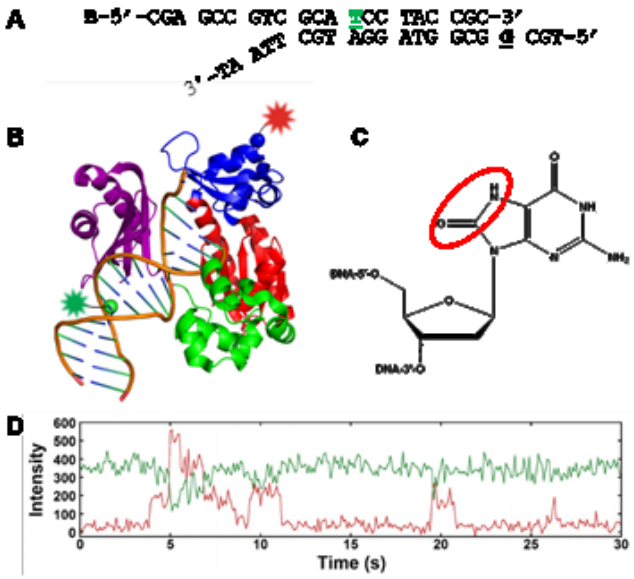


Figure 4.1: Single-molecule FRET analysis of Dpo4 binding to DNA.

(A) DNA substrate used in single-molecule experiments. The primer is biotinylated at the 5′ end (denoted "B") to facilitate surface immobilization. "T" denotes Cy3 attached to 5-C6-amino-2′-deoxythymidine. "G" represents 8-oxo-7,8-dihydro-2′-deoxyguanine (8-oxo-dG) or undamaged dG. (B) The binary complex of Dpo4 and DNA accessed through PDB code 2RDJ. The green star represents a Cy3 donor label on the DNA substrate. The red star signifies the site of Cy5 acceptor labeling in the Finger domain (N70C) of Dpo4. (C) Chemical structure of 8-oxo-dG with chemical modifications at the C8 and N7 positions of guanine circled in red. (D) Representative single-molecule fluorescence trajectory for Dpo4 binding to the DNA substrate containing 8-oxo-dG. Donor and acceptor fluorescence signals are shown in green and red, respectively.

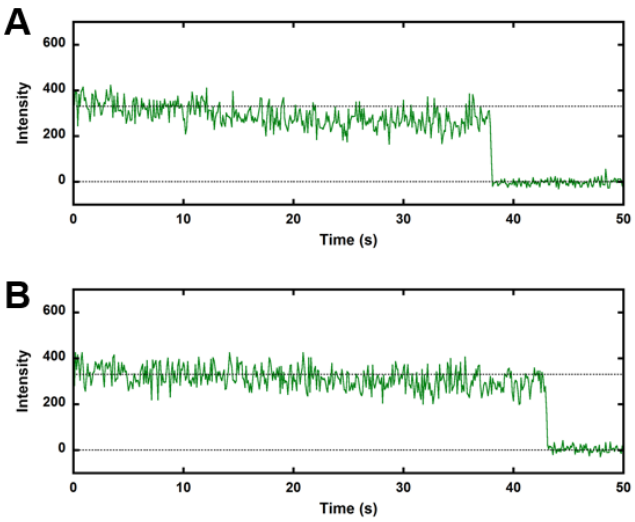


Figure 4.2: Verification of observed FRET.

Potential photophysical phenomenon due to multiple donor molecules immobilized in close proximity or protein induced fluorescence enhancement (PIFE), which may introduce ambiguity in data analysis, must be ruled out to verify the legitimacy of the designed FRET system. PIFE is characterized by an increase in the intensity of a fluorophore upon proximal binding of an unlabeled protein. (A) To verify that our DNA coverage was sufficient to observe single, distinct Cy3-labeled DNA molecules, we directly excited 20 pM of surface immobilized Cy3-labeled undamaged DNA substrate (Figure 4.1) in the absence of any proteins or acceptor fluorophores. Donor fluorescence intensity of single molecules was recorded. Constant fluorescence intensity was observed in the donor channel until one-step photobleaching, whereas none was observed in the acceptor channel (data not shown). Notably, multiple photobleaching events would have suggested too dense of DNA surface coverage. (B) Wild-type Dpo4 was introduced into

(continued)

Figure. 4.1: continued

the chamber containing surface immobilized Cy3-labeled DNA. The donor fluorescence intensity did not change appreciably, and once again no fluorescence intensity above background was observed in the acceptor channel. Introduction of 10 nM Cy5-labeled Dpo4 produced decreases in donor fluorescence intensity with corresponding increases in the acceptor fluorescence intensity as shown in Figure 4.1D**.**

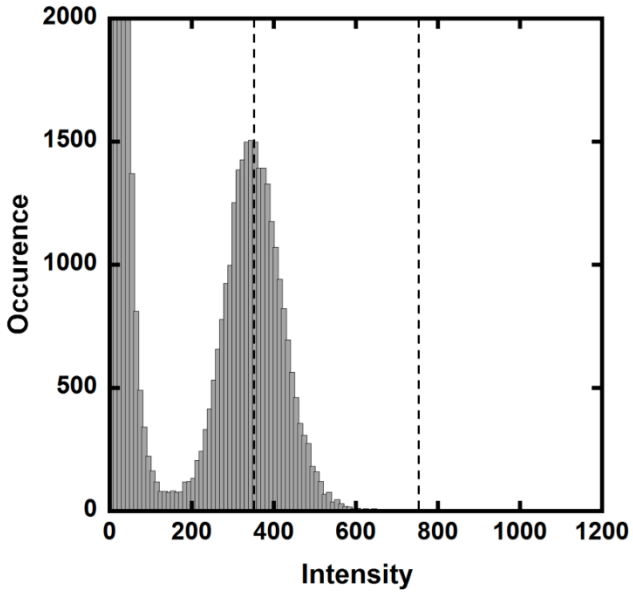


Figure 4.3: Fluorescence assay for multiplicity of binding.

10 nM Cy5-labled Dpo4 was introduced into a chamber containing surface immobilized Cy3-labeled DNA. A 638 nm laser was used to directly excite any Cy5-labeled Dpo4 molecules that were immobilized due to DNA binding. Acceptor channel fluorescence intensity trajectories were recorded, and >100 trajectories were analyzed and used to generate the histogram. A single distribution was observed at an intensity of ~350 (first dashed line) indicating that the Dpo4 molecules bind the DNA substrates in a 1:1 ratio. If multiple Dpo4 molecules were bound to the DNA, a distribution at a higher intensity (~700, second dashed line) due to the additive fluorescence intensities of multiple acceptor fluorophores would have been observed. The large peak at zero intensity is due to background.

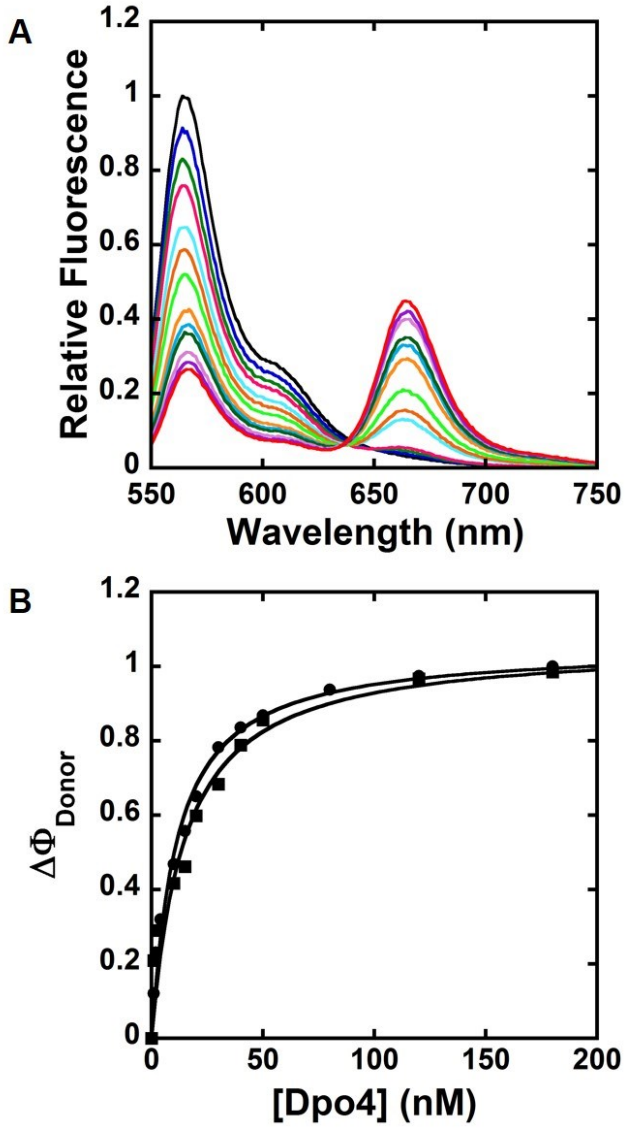


Figure 4.4: Ensemble FRET analysis of Dpo4 binding.

(A) Example fluorescence titration. Cy3-DNA (25 nM) was excited at 532 nM and fluorescence emission spectra were recorded at varying concentrations of Cy5-Dpo4 N70C ranging from 0 nM (*black*) to 180 nM (*pink*). (B) Change in donor quantum yield (continued)

Figure 4.4: continued

($\Delta\Phi_{Donor}$) for the undamaged (●) or 8-oxo-dG modified (■) DNA substrates upon titration with Dpo4. The curves were fit to Eq. 4.1 to yield a K_d of 11 ± 1 nM for Dpo4 binding to undamaged DNA and a K_d of 10 ±1 nM for Dpo4 binding to the 8-oxo-dG modified DNA.

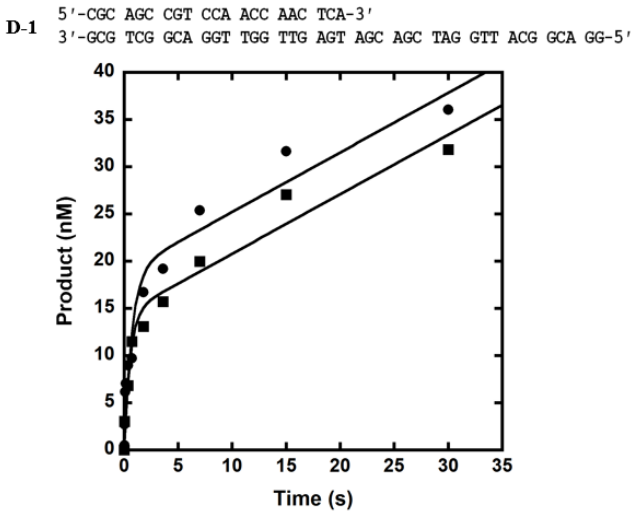


Figure 4.5: Burst Assay to Confirm Dpo4 Activity.

A preincubated solution of either WT (■), or Cy5-labeled mutant (●) Dpo4 (15 nM) and 21/41mer with a 5′-^{32}P-labeled primer (60 nM) was mixed with dTTP (100 μM) in a rapid chemical quench flow apparatus. The reactions were quenched at various times with EDTA to 0.37 M, and products were separated by denaturing PAGE and quantitated using ImageQuant. The data were fit by nonlinear regression to the following burst equation: [product] = A[1-exp(-k_1t) + k_2t] where A is the amplitude of active enzyme, k_1 is the observed burst rate constant, and k_2 is the observed steady-state rate constant. The rate constants for the wild-type Dpo4 were 1.8 ± 0.6 s^{-1} and 0.04 ± 0.01 s^{-1} for the exponential and linear phases, respectively. The rate constants for the Cy5-labeled mutant Dpo4 were 1.4 ± 0.4 s^{-1} and 0.03 ± 0.01 s^{-1} for the exponential and linear phases, respectively. These similar rate constants between wild-type Dpo4 and Cy5-labeled Dpo4 indicate that the fluorescent labeling did not affect the activity of the polymerase.

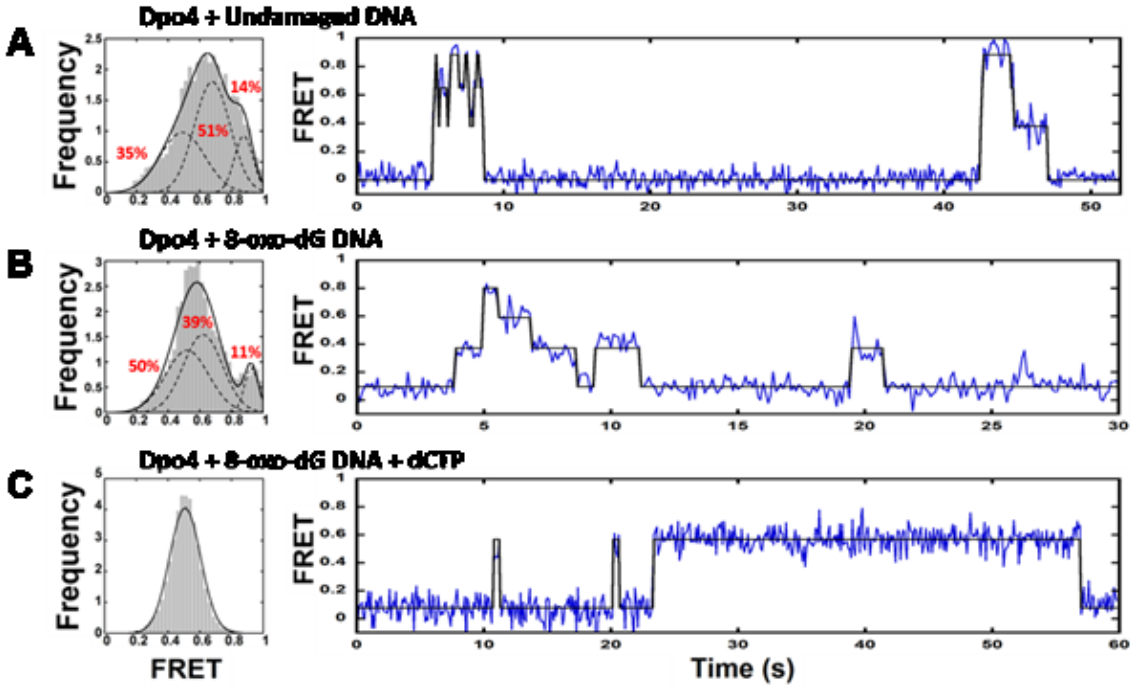


Figure 4.6: FRET from Dpo4 binding to single DNA molecules.

FRET efficiency histograms (left) and representative smFRET trajectories (right) of Dpo4 binding to (A) undamaged DNA (B) 8-oxo-dG modified DNA and (C) 8-oxo-dG modified DNA with 6 mM of correct nucleotide (dCTP). On the left, individual Gaussian peak fits for each state are shown by the dashed black lines with respective population percentage indicated in red. The sum of the individual Gaussians is shown as the solid black line. On the right, representative FRET trajectories are shown in blue and the black lines depict HMM fits to the data.

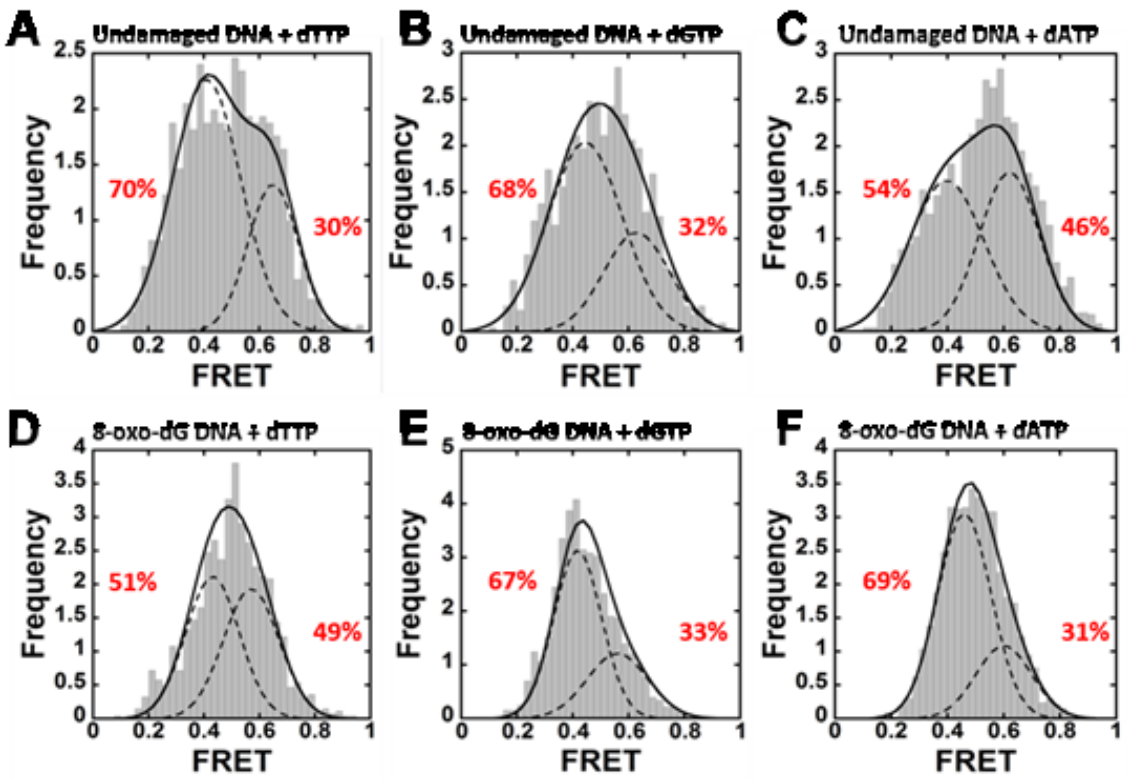


Figure 4.7: FRET histograms of Dpo4 binding DNA with incorrect dNTP.

(A) Population distribution histograms generated from multiple time trajectories of Dpo4 binding to undamaged DNA (top) or 8-oxo-dG modified DNA (bottom) in the presence of 6 mM of either dTTP (A, D), dGTP (B, E), or dATP (C, F). Individual Gaussian peak fits for each state are shown by the dashed black lines with respective population percentage indicated in red. The sum of the individual Gaussians are shown as the solid black line in each.

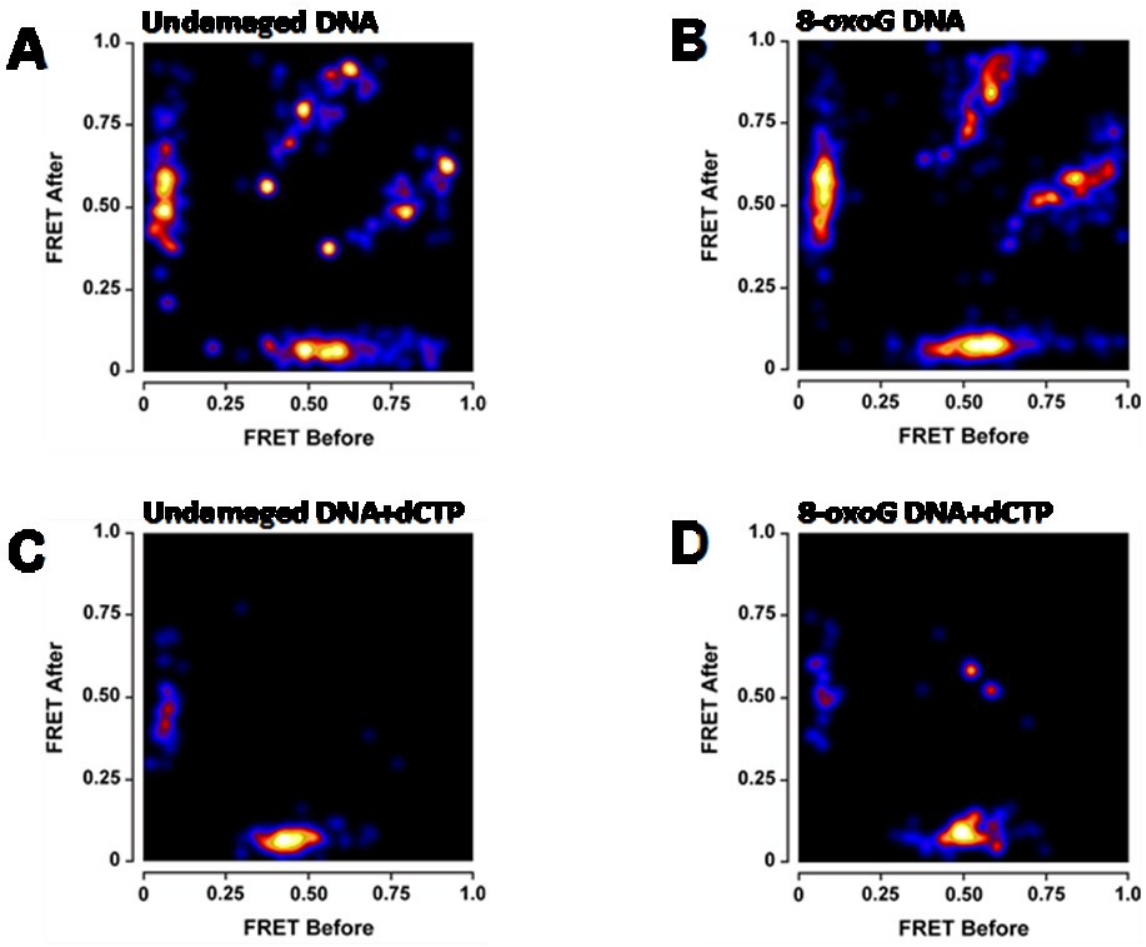


Figure 4.8: Transition density plots.

Transition density plots from experiments monitoring the following complexes: (A) Cy5-Dpo4●Undamaged DNA. (B) Cy5-Dpo4●8-oxoG DNA. (C) Cy5-Dpo4●Undamaged DNA+6 mM dCTP. (D) Cy5-Dpo4●8-oxoG DNA+6 mM dCTP. The transition density plots (TDP) show the transitions between the low, mid, and high FRET states. FRET efficiencies are plotted based on their initial state, and final state. The recurrence of a particular transition by many molecules results in a crosspeak at that position on the plot. (continued)

Figure 4.8: continued

The crosspeaks are represented as colored regions shaded blue to red to yellow to represent an increasing probability of occurrence. Notably, many transitions between low-, mid-, and high-FRET states are observed when Dpo4 binds DNA, however these transitions are nearly eliminated in the presence of the correct nucleotide (dCTP) at saturating concentrations.

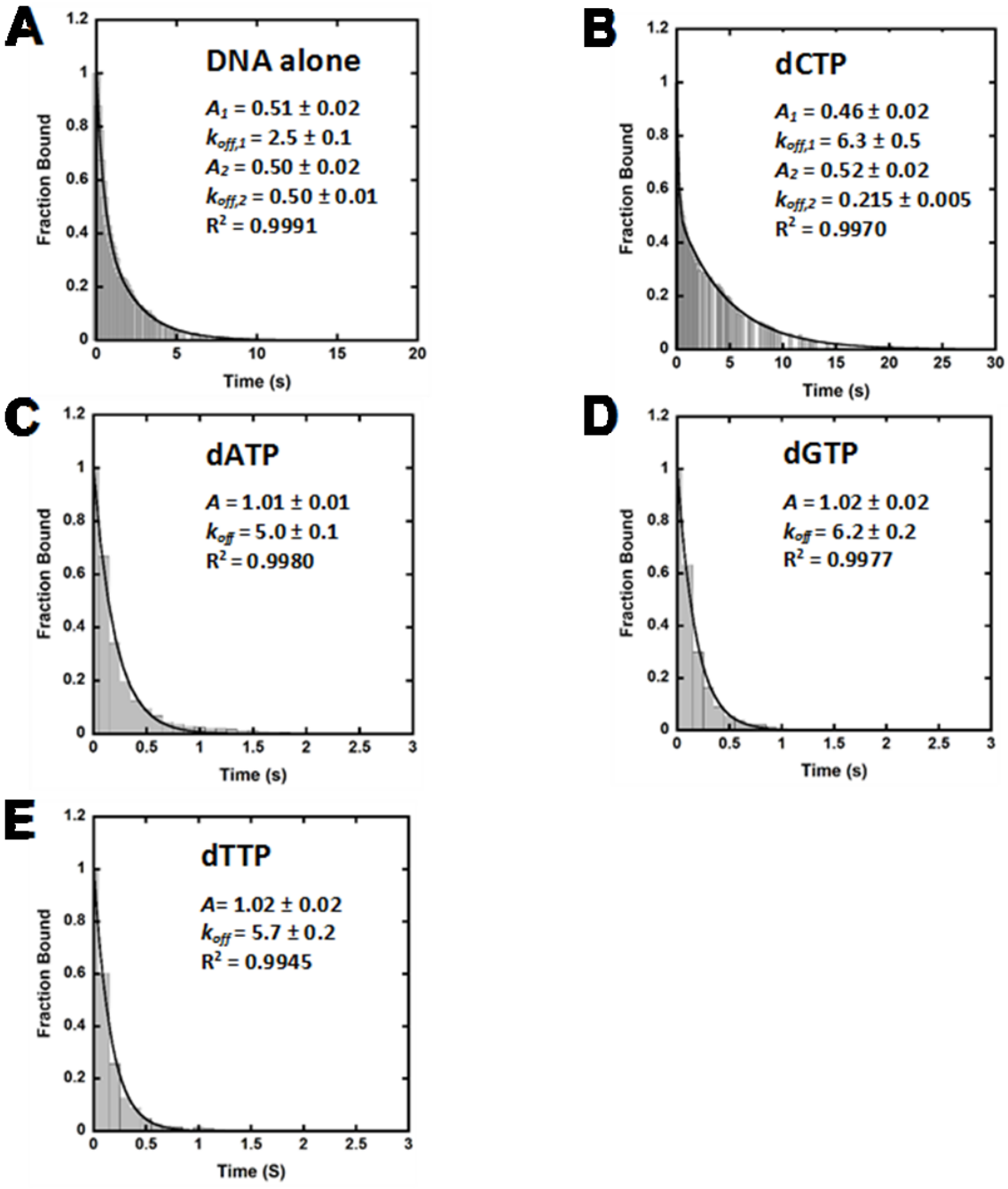


Figure 4.9: Dwell time analysis of Dpo4 binding undamaged DNA.

(continued)

Figure 4.9: continued

Survivor functions for the DNA bound state were computed from histograms of binding times from multiple trajectories with 10 nM Cy5-Dpo4 binding to undamaged DNA with or without nucleotide (Figure 4.1). Double or single exponential fits are depicted as the black solid lines. (A) Binary Complex (B) with 6 mM dCTP (C) with 6 mM dATP (D) with 6 mM dGTP (E) with 6 mM dTTP.

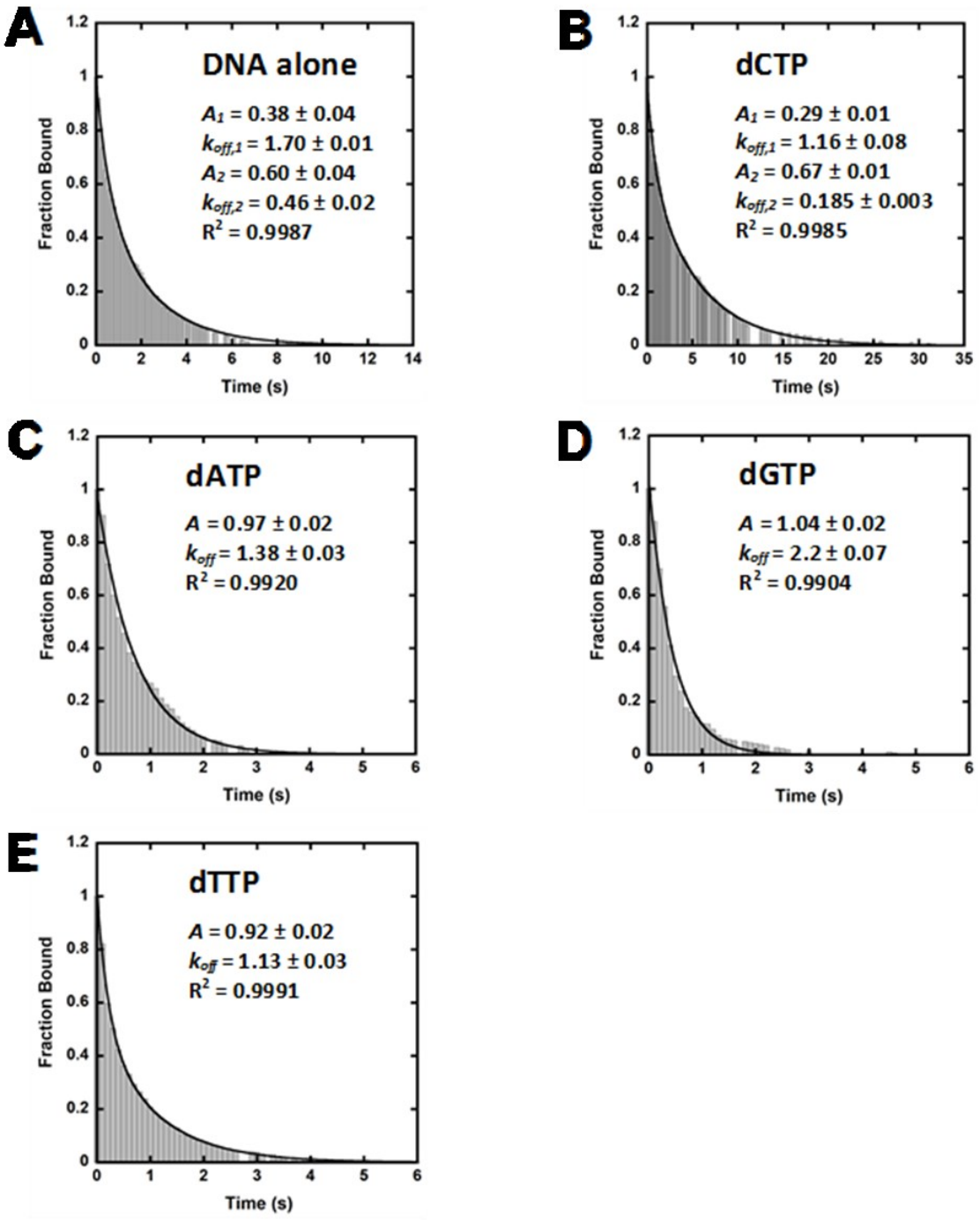


Figure 4.10: Dwell Time Analysis of Dpo4 Binding Damaged DNA.

(continued)

Figure 4.10: continued

Survivor functions for the DNA bound state were computed from histograms of binding times from multiple trajectories with 10 nM Cy5-Dpo4 binding to damaged DNA with or without nucleotide (Figure 4.1).Double or single exponential fits are depicted as the black solid lines. (A) Binary Complex (B) with 6 mM dCTP (C) with 6 mM dATP (D) with 6 mM dGTP (E) with 6 mM dTTP.

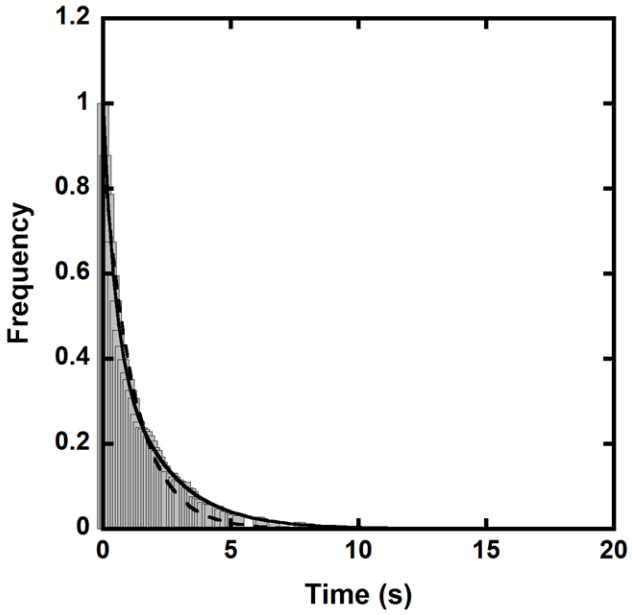


Figure 4.11: Example dwell time survivor function.

Survivor function for the DNA bound state was computed from a histogram of binding times from multiple trajectories with 10 nM Cy5-Dpo4 binding to undamaged DNA. The survivor function was fit to a double exponential decay equation. 50.5% of the population was dissociated in a fast phase with a rate constant of 2.5 ± 0.1 s^{-1}, while 49.5% of the population was dissociated through a slow phase of 0.5 ± 0.01 s^{-1}. Included as a comparsion to the best fit *solid* line, the *dashed* line depicts a single exponential fit to the data.

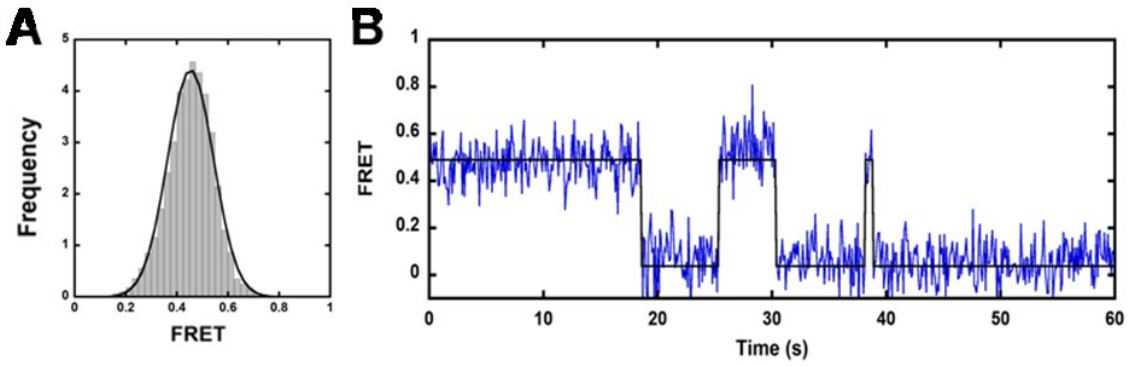


Figure 4.12: FRET from Dpo4 binding to single undamaged DNA molecules in the presence of saturating dCTP.

(A) FRET efficiency histogram with Gaussian peak fit shown as a solid black line for Dpo4 binding to undamaged DNA with saturating concentration of correct nucleotide (dCTP). (B) Representative single-molecule FRET trajectory of Dpo4 binding to undamaged DNA with saturating concentration of correct nucleotide (dCTP) with HMM data fit depicted as solid black line.

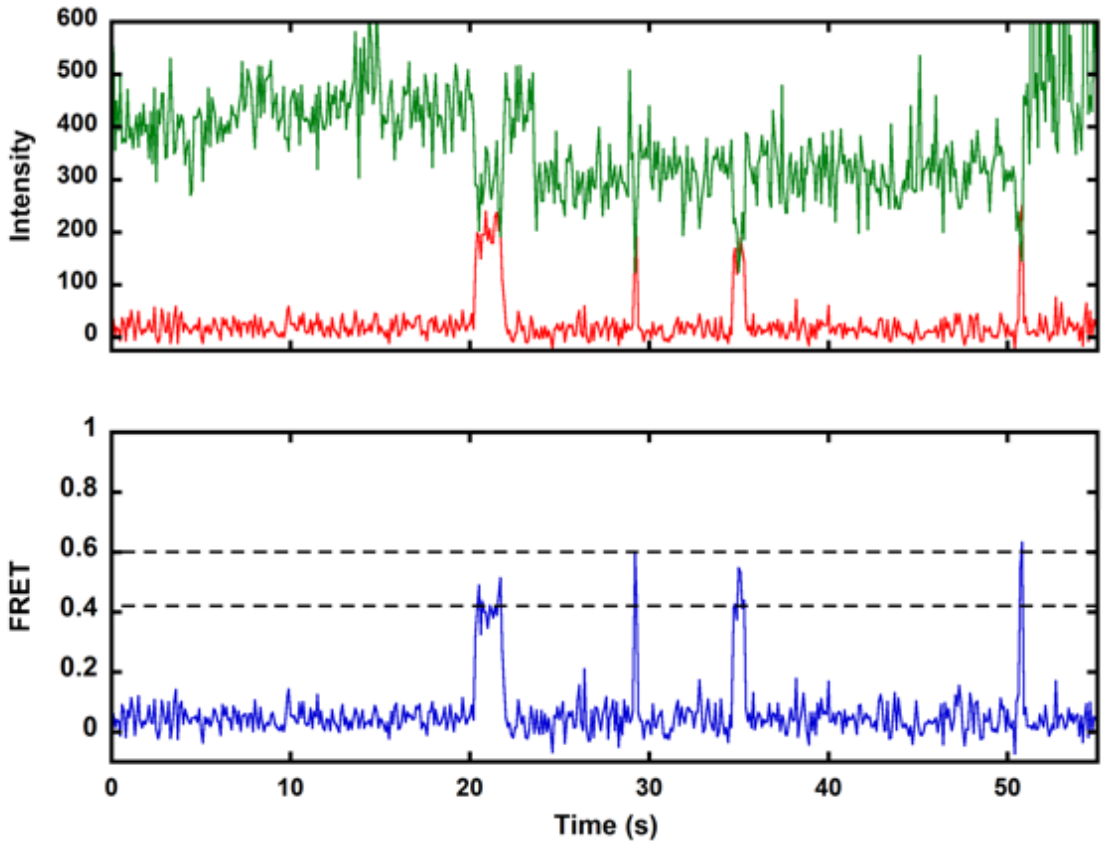


Figure 4.13: Representative FRET trajectory for Dpo4 binding DNA in the presence of saturating incorrect nucleotide.

(A) Representative single-molecule trajectory is shown for a Cy5-labeled Dpo4 molecule bound to Cy3-labeled undamaged DNA, in the presence of 6 mM dTTP. Donor and acceptor intensities are shown in *green* and *red*, respectively. (B) The corresponding FRET signal is shown in *blue*. The dotted black lines represent the two main FRET states observed in this experiment: a low FRET state at ~0.4 and a mid-FRET state at ~0.6. In general the binding events are very transient, resulting in an average binding time in this experiment of 0.3 seconds.

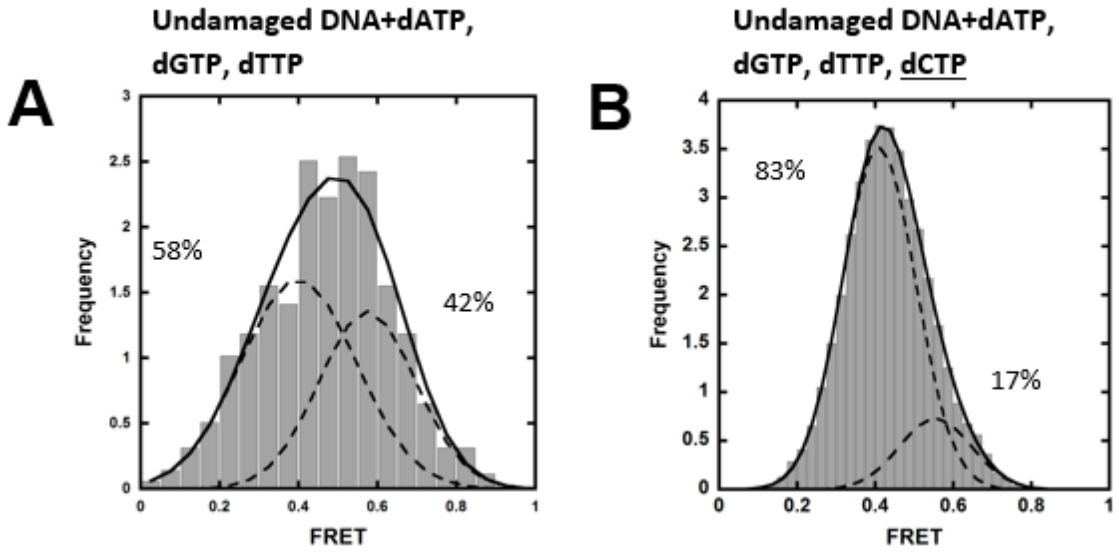


Figure 4.14: Selectivity of Dpo4 for the correct nucleotide.

Single molecule experiments were conducted using the conditions described in the Materials and Methods. Cy5-labeled Dpo4 binding undamaged DNA with either (A) 2 mM dATP, 2 mM dGTP, and 2mM dTTP, or (B) all four nucleotides (1.5 mM each). The total nucleotide concentration in both experiments was 6 mM. (A) In the presence of the three incorrect nucleotides, the population histogram was bimodal as seen previously with the incorrect nucleotide experiments, with 58% of molecules bound in the low FRET state, and 42% bound in the mid FRET state. However, with the addition of all four nucleotides (B), the population of the low FRET state increased 58% to 83%, resembling the trend previously observed with dCTP where almost all molecules are bound in the low FRET state. 17% of molecules still bound in the mid FRET state. This is likely due to the presence of the other three incorrect nucleotides in a 3:1 excess over the correct dCTP.

Chapter 5. Investigation of Intradomain Motions of a Y-Family DNA Polymerase during Substrate Binding and Catalysis

Reproduced in part with permission from Raper, Austin T., Suo, Z. (2016) Investigation of Intradomain Motions of a Y-Family DNA Polymerase during Substrate Binding and Catalysis. *Biochemistry.* 55(41):5832-5844. DOI: 10.1021/acs.biochem.6b00878.

Author Contributions

Austin T. Raper conceived all research, executed all experiments, interpreted all data, generated all figures, and wrote the manuscript. Dr. Zucai Suo helped to edit the manuscript.

5.1 Abstract

DNA polymerases catalyze DNA synthesis through a stepwise kinetic mechanism that begins with binding to DNA, followed by selection, binding, and incorporation of a nucleotide into an elongating primer. It is hypothesized that subtle active site adjustments in a polymerase to align reactive moieties limit the rate of correct nucleotide incorporation. DNA damage can impede this process for many DNA polymerases, causing replication fork stalling, genetic mutations, and potentially cell death. However, specialized Y-family DNA polymerases are structurally evolved to efficiently bypass DNA damage *in vivo*, albeit at the expense of replication fidelity. Dpo4, a model Y-family polymerase from *Sulfolobus solfataricus*, has been well-studied kinetically, structurally, and computationally, yielding mechanistic understanding about how the Y-family DNA polymerases achieve their unique catalytic properties. We previously employed a real-time Förster resonance energy transfer (FRET) technique to characterize the global conformational motions of Dpo4 during DNA binding as well as nucleotide binding and incorporation by monitoring changes in distance between sites on the polymerase and DNA, and even between domains of Dpo4. Here, we extend the utility of our FRET methodology to observe conformational transitions within individual domains of Dpo4 during DNA binding and nucleotide incorporation. The results of this novel, intradomain FRET approach unify findings from many studies to fully clarify the complex DNA binding mechanism of Dpo4. Furthermore, intradomain motions in the Finger domain during nucleotide binding and incorporation, for the first time, report on the rate-limiting step of a single nucleotide addition catalyzed by Dpo4.

5.2 Introduction

Since the initial identification of Y-family DNA polymerases[229], extensive efforts have been made to understand their precise role in DNA replication. In contrast to high-fidelity, processive polymerases from A-, B-, and C-families, members of the Y-family exhibit low-fidelity, distributive DNA synthesis on undamaged DNA[14]. Though these are undesirable traits for a replicative DNA polymerase, they uniquely qualify the Y-family polymerases for the critical cellular responsibilities of immunoglobulin generation, somatic hypermutation, and translesion DNA synthesis (TLS)[23]. Perhaps the most striking role of the Y-family polymerases involves their abilities to bypass DNA damage during TLS, which would otherwise stall DNA synthesis by high-fidelity replicative DNA polymerases and lead to cell death. Much of what is currently known concerning the mechanism of DNA synthesis on undamaged and damaged DNA by the Y-family DNA polymerases comes from structural[84,149,364] and kinetic[14,93,94,134,220] investigations of Dpo4, the model and lone Y-family DNA polymerase from *Sulfolobus solfataricus*.

Akin to other Y-family members, Dpo4 contains a structurally conserved polymerase core consisting of Finger, Palm, and Thumb domains arranged in a characteristic right-hand architecture, as well as a Little Finger (LF) domain that is joined to the polymerase core through a highly flexible peptide linker (Figure 5.1) [84]. Interestingly, there is some evidence to suggest that the LF and linker mediate the association of Dpo4 to DNA[151] and that the LF domain determines the lesion bypass properties of Y-family polymerases[365]. In contrast to many other polymerases, structural comparison of the binary (Dpo4•DNA) and ternary (Dpo4•DNA•dNTP) complexes of the

enzyme reveals that, Dpo4 does not exhibit the open to closed conformational transition of the Finger domain upon nucleotide binding thought to function as a fidelity checkpoint but rather it exhibits a largely preformed active site lacking any major conformational adjustments[149]. Nevertheless, ensemble[11,12,22,219,221] and single-molecule[9,163] techniques have revealed Dpo4 to be dynamic, exhibiting rigid body domain movements as well as more local conformational motions throughout DNA association, nucleotide binding, and nucleotide incorporation steps of catalysis, and these motions were found to be altered by the presence of a lesion in the DNA template[9,12].

Pre-steady-state kinetic characterization of the minimal kinetic mechanism of nucleotide insertion by Dpo4 suggests that a conformational change, likely consisting of subtle active site rearrangements to properly align divalent metal ions and carboxylate moieties as well as reactive groups of the primer and incoming nucleotide, is rate-limiting, rather than phosphodiester bond formation[94]. Unfortunately, there is no explicit evidence for this conformational transition, leading to significant controversy[5]. For A-, B-, C-, and some X-family polymerases, it was thought that the closure of the Finger domain during nucleotide binding may signify the slowest step during single-nucleotide incorporation[29,210], but structural studies indicate that Dpo4 lacks such a change. Our previous, stopped-flow Förster resonance energy transfer (FRET) study[11] sought to capture the rate-limiting conformational change for Dpo4 but instead revealed global dynamic conformational events at several other stages during nucleotide binding and incorporation.

Here we extended our stopped-flow FRET methodology to observe

conformational motions that may exist within individual domains of Dpo4 during DNA and nucleotide binding and catalysis. By generating intradomain FRET constructs consisting of donor and acceptor fluorophores within each domain of Dpo4, we were able to gain significant insight into the DNA binding mechanism of Dpo4, as well as observe dynamic motions within each domain during nucleotide binding and insertion. Slow conformational motions observed for the Finger domain existed independent of phosphodiester bond formation and may be reporting on the elusive rate-limiting step for single-nucleotide incorporation.

5.3 Materials and Methods.

5.3.1 Preparation of protein and DNA.

FRET constructs were generated as previously described[22]. Briefly, donor Trp residues were individually introduced into each domain of Dpo4. For each Trp-containing mutant, a corresponding Cys mutation was generated within the same domain for subsequent labeling with the acceptor fluorophore, 7-(diethylamino)-3-(4′-maleimidylphenyl)-4-methylcoumarin (CPM) (ThermoFisher Scientific). Notably, Dpo4 contains no native Trp residues and the lone native Cys residue was mutated to Ser. All mutations were confirmed by DNA sequencing and are shown in Figure 5.1. Expression, purification, and fluorophore labeling were completed as previously described[22]. Unreacted CPM dye was removed by two consecutive spins through Micro P-6 Bio-Spin columns (Bio-Rad). The protein concentration was determined using the Bradford assay (Bio-Rad) with wild-type (wt) Dpo4 as the standard, and the degree of labeling was assessed by measuring the concentration of CPM by UV/Vis spectrophotometry using an

extinction coefficient of 44800 M^{-1} cm^{-1} at 430 nm. For all CPM-labeled mutants, the ratio of dye concentration to protein concentration revealed a labeling efficiency of ≥90%.

DNA oligonucleotides were purchased from Integrated DNA Technologies (IDT). A 21-mer primer (5′-CGAGCCGTCGCAT̲CCTACCGC-3′) with either a normal, 3′-deoxy-terminated end or a 3′-dideoxy-terminated end was annealed to a 30-mer template (5′-GATGCTGCAGCGG TAGGATGCGACGGCTCG-3′) to generate the DNA^{OH} and DNA^{H} substrates, respectively. All synthetic DNA oligonucleotides were purified by denaturing polyacrylamide gel electrophoresis (PAGE) (17% polyacrylamide, 8 M urea) and the concentrations were determined by UV absorbance at 260 nm. Alexa488-labeled DNA^{OH} was generated as described previously[11,12] through Alexa Fluor 488-NHS ester (Invitrogen) labeling of a 5-C6-amino-2′-deoxythymidine modification (T̲) on the 9th base from the primer 3′ terminus.

5.3.2 Buffers.

Unless otherwise specified, all reactions were performed in buffer R, containing 50 mM HEPES (pH 7.5 at 20˚C), 6 mM $MgCl_2$, 50 mM NaCl, 0.1 mM EDTA, 1 mM DTT, and 10% glycerol.

5.3.3 Steady-state fluorescence spectroscopy assays.

All steady-state fluorescence spectra were recorded on a Fluoromax-4 fluorometer (Jobin Yvon Horiba) at 20˚C. The slit widths for the excitation and emission monochromators were each set to 5 nm. Fluorescence emission spectra of each CPM-labeled Dpo4 mutant (200 nM) in the absence or presence of either DNA (DNA^{OH}, 300

nM) or DNA (300 nM) and correct nucleotide (dTTP, 1 mM) were recorded upon Trp-donor excitation at 290 nm. Upon addition of DNA and dTTP, a five-minute incubation period was observed to allow for reaction equilibration. All reported spectra were corrected for dilution and for the intrinsic fluorescence of buffer components as well as of DNA^{OH} and dTTP, when applicable.

Notably, while the distance between donor and acceptor probes can be calculated from the FRET efficiency, this requires consideration of many complex variables[366,367]. Therefore, to simplify our data analysis, we have calculated $(ratio)_A$ by dividing the acceptor fluorescence due to energy transfer from the excited donor fluorophore by the acceptor fluorescence due to direct excitation. Importantly, changes in $(ratio)_A$ observed upon altering experimental conditions are indicative of conformational change as $(ratio)_A$ is directly proportional to FRET efficiency.

5.3.4 Stopped-flow FRET assays.

Stopped-flow fluorescence spectroscopy experiments were completed at 20˚C using an Applied Photophysics SX20 instrument with a dead time of ~1 ms and monochromator slit widths set to 5 nm. Importantly, a temperature of 20˚C was maintained for all stopped-flow mixing experiments to be consistent with our previous studies and to permit observation of rapid fluorescent phases which may become too fast to be detected at elevated temperatures[11,12,22]. During excitation at 290 nm, reaction components (all reported component concentrations represent final concentrations after mixing) were rapidly mixed within the 20 μL optical cell and subsequent Trp or CPM fluorescence was recorded over time using either a 305 nm or a 420 nm cut-off filter

(Applied Photophysics Ltd.), respectively. The resulting fluorescence traces were fit by nonlinear regression using Pro-Data Viewer software (Applied Photophysics Ltd.) to either single (Eq. 5.1), double (Eq. 5.2), or triple (Eq. 5.3) exponential equations (shown below) depending on the number of observed FRET phases.

$$F(t) = A \cdot \exp(- k_{obs} \cdot t) + \text{constant} \qquad \text{(Eq.5.1)}$$

$$F(t) = A_1 \cdot \exp(- k_1 \cdot t) + A_2 \cdot \exp(- k_2 \cdot t) + \text{constant} \qquad \text{(Eq. 5.2)}$$

$$F(t) = A_1 \cdot \exp(- k_1 \cdot t) + A_2 \cdot \exp(- k_2 \cdot t) + A_3 \cdot \exp(- k_3 \cdot t) + \text{constant} \qquad \text{(Eq. 5.3)}$$

where $F(t)$ signifies the change in fluorescence over time (t); A, A_1, A_2, and A_3 are the phase amplitudes; and k_{obs}, k_1, k_2, and k_3 are the corresponding phase rates. All reported rates and errors represent the average rates and corresponding standard deviations among multiple, replicate stopped-flow mixing experiments.

5.3.5 Stopped-flow anisotropy assays.

Following the manufacturer's protocol, stopped-flow anisotropy measurements were performed using the Applied Photophysics SX20 instrument at 20˚C outfitted with a T-format fluorescence polarization accessory (Applied Photophysics Ltd.) and monochromator slit widths set to 10 nm. For experiments monitoring the binding of wt Dpo4 (500 nM) to Alexa488-labeled DNA^{OH} (500 nM), polarized excitation occurred at 499 nm, and 530 nm cutoff filters were placed in front of each detector. Kinetic parameters from anisotropy measurements were obtained by fitting the data to a single exponential equation (Eq. 5.1 above).

5.3.6 Dpo4 mutant activity assays.

Enzymatic activities of Dpo4 mutants following CPM-labeling were assessed at 20˚C using a burst kinetic assay, as described previously[9,368]. For each CPM-labeled mutant and wt Dpo4, 60 nM of 5′-^{32}P-labeled DNA^{OH} was pre-incubated in buffer R with 10 nM enzyme before rapid mixing with 100 μM dTTP using a rapid chemical-quench flow apparatus (Kintek). Reactions were then quenched at increasing time intervals with the addition of 0.37 M EDTA. Products were then separated by denaturing PAGE (17% polyacrylamide, 8 M urea, 1x TBE running buffer) and quantitated by phosphorimaging (Typhoon Trio, GE Healthcare). Product formation was plotted against time and fit by non-linear regression using KaleidaGraph (Synergy Software) to the burst equation:

$$[P](t) = A \cdot [1 - \exp(- k_{burst} \cdot t) + k_{ss} \cdot t] \quad \text{(Eq. 5.5)}$$

where [P](t) signifies the change in product concentration over time (t), A is the burst phase amplitude, k_{burst} is the single-turnover nucleotide incorporation rate constant, and k_{ss} is the steady-state rate constant.

5.4 Results

5.4.1 Design of an intradomain FRET system.

To detect movements within the individual domains of Dpo4, we modified our previously described FRET system[11,22] wherein a single, site-specifically engineered Trp residue serves as the donor fluorophore, and CPM, conjugated to a single, site-specifically engineered Cys residue, serves as the acceptor fluorophore (Figure 5.1 and Table 5.1). When selecting sites for acceptor labeling, we strove to 1) mutate only surface exposed amino acid residues and 2) optimize the distance to the Trp donor, based on the

Trp-CPM Förster radius (R_0) of 30 Å[369,370]. We sought to keep the distance between donor and acceptor sites within $R_0 \pm 0.5\ R_0$ but had to compromise in several instances due to distance constraints imposed by the individual domains (Table 5.1). In this study, we will refer to unlabeled Dpo4 constructs by their respective Trp and Cys mutations (e.g. Y274W-E291C), and to CPM-labeled Dpo4 constructs in a similar manner but with a superscripted CPM (e.g. Y274W-E291C^{CPM}). ^{32}P-based pre-steady-state kinetic burst assays of wt Dpo4 and each CPM-labeled construct were performed to verify enzymatic activity (Materials and Methods). We found that neither the single-turnover nucleotide incorporation rate constant (k_{burst}, Eq. 5.5) nor the steady-state multiple turnover rate constant (k_{ss}, Eq. 5.5) were appreciably affected by the mutations or labeling when compared to wt Dpo4 (Table 5.2). These results suggest that our labeling procedure did not impact the polymerase activity of Dpo4.

5.4.2 Dpo4 intradomain motion observed by steady-state fluorescence.

The Trp fluorescence of each Dpo4 construct prior to CPM-labeling was investigated by steady-state fluorescence (Figure 5.2). Interestingly, we observed that Trp residues introduced into the Palm (Y118W) and Thumb (Y224W) domains displayed maximum fluorescence emission at 320 nm (Figure 5.2A) while Trp residues introduced into the Finger (S22W) and LF (Y274W) displayed a red-shifted maximum fluorescence emission at 354 nm, suggesting a higher degree of solvent exposure (Figure 5.2B)[149,367,371]. These results qualitatively agree with our expectations based on their locations on Dpo4 in the DNA bound state presented in Figure 5.1, compared to the apo state[149]. The addition of 300 nM of DNA^{OH} (Materials and Methods) resulted in a 16-

35% decrease in fluorescence for all Trp-mutants studied. The subsequent addition of 1 mM dTTP resulted in an additional 34-49% decrease in the fluorescence signal. Unfortunately, given that both DNA and nucleotide absorb at 290 nm, significant excitation light is likely being lost to inner filtering thereby contributing to the decreases in fluorescent signal. However, we would anticipate that if the decreasing signals were due solely to inner filtering then a uniform reduction in signal would be observed for all mutants. Rather, we see signal attenuation to varying degrees among the tested Trp mutants implying protein conformational dynamics upon the addition of DNA and nucleotide.

The emission of Dpo4 mutant Y274W-E291C^{CPM} was recorded upon excitation at 290 nm and revealed significant reduction in Trp emission (Figure 5.2C, compare blue trace with black trace) with a concomitant increase in fluorescence at 472 nm, the emission maximum of CPM (Figure 5.2C, black trace). We calculated $(ratio)_A$ for select Dpo4 mutants (Figure 5.2D), Y274W-E291C^{CPM} (solid bars) and S22W-V62C^{CPM} (hashed bars), from the collected emission spectra of the apo state and following sequential additions of 300 nM DNA^{OH} and 1 mM dTTP. As true FRET efficiency values depend on multiple complex parameters[366,367], we instead used $(ratio)_A$, calculated as the acceptor fluorescence due to energy transfer divided by the acceptor fluorescence due to direct excitation, as a convenient indicator of distance change between the donor and acceptor probes under varying experimental conditions (Materials and Methods). For Y274W-E291C^{CPM} and S22W-V62C^{CPM} the addition of DNA^{OH} resulted in a 32% and a 17% increase in $(ratio)_A$, respectively, indicating that the FRET probes moved closer

together for binary complex formation (Dpo4•DNA^{OH}). Contrastingly, the subsequent addition of 1 mM dTTP resulted in a significant decrease in $(ratio)_A$ for Y274W-E291C^{CPM} (82%) and S22W-V62C^{CPM} (76%). As dTTP strongly absorbs at 290 nm, the dramatic reductions in $(ratio)_A$, as a result of concurrent decreases in Trp and CPM fluorescence intensities, for both Dpo4 mutants likely stems from the inner filtering caused by the high, 1 mM concentration of dTTP needed to saturate nucleotide binding, rather than from nucleotide dependent conformational dynamics. Despite the significant inner filtering, the difference in signal attenuation upon the addition of dTTP may still indicate changes in distance between the donor and acceptor probes as inner filtering should attenuate the total fluorescence to the same degree for each mutant. Steady-state fluorescence spectra for the other Dpo4 mutants were also recorded and produced results (data not shown) consistent with the expected distance changes determined from the crystal structures (Table 5.1). These experiments were repeated with the remaining CPM-labeled Dpo4 mutants listed in Table 5.1 and anti-correlated signal changes were observed for each to clearly report on the efficient energy transfer occurring from the Trp-donor to the CPM-acceptor (data not shown). Taken together, these results suggest that conformational dynamics within individual polymerase domains are taking place during substrate binding and catalysis.

5.4.3 Conformational dynamics of intradomain Little Finger mutants during DNA binding.

To further study the DNA binding process of Dpo4, we monitored fluorescence changes upon rapidly mixing each of the intradomain FRET constructs with DNA^{OH} in a

stopped-flow apparatus at 20˚C. Briefly, 100 nM CPM-labeled Dpo4 mutant was rapidly mixed with 100 nM DNA^{OH} in the stopped-flow apparatus and the fluorescence emission of CPM upon excitation at 290 nm was monitored using a 420 nm cut-off filter. For several of the Dpo4 mutants, including Y118W-K137C^{CPM}, Y224W-K172C^{CPM}, and Y224W-K212C^{CPM}, time-dependent decreases in fluorescence with low signal to noise ratios were observed upon mixing with DNA^{OH} (Figure 5.3). This result suggests that the conformational movements upon DNA binding in the Palm and Thumb domains may not be large enough in magnitude to cause a significant fluorescence signal change. However, experiments repeated with LF mutants Y274W-K329C^{CPM}, Y274W-E291C^{CPM}, and Y274W-R267C^{CPM}, as well as Finger mutants S22W-K56C^{CPM} and S22W-V62C^{CPM}, demonstrated time-dependent increases in CPM fluorescence with high signal to noise ratios upon mixing with DNA^{OH} (Figure 5.4). To gain insight into the conformational changes observed for the LF domain mutants, we monitored the fluorescence response of Y274W-K329C^{CPM} upon rapid mixing with varying concentrations of DNA^{OH} (2-200 nM) (Figure 5.3). To ensure adequate signal change while maintaining pseudo-first order reaction conditions of excess polymerase, 100 nM Y274W-K329C^{CPM} was used for smaller concentrations of DNA^{OH} (2-20 nM) and 1 μM Y274W-K329C^{CPM} was used for larger concentrations of DNA^{OH} (50-200 nM). Interestingly, we found that the observed rates of fluorescence change increased linearly with increasing concentrations of DNA^{OH} (Figure 5.3A). By plotting the observed pseudo-first order rate constants versus their corresponding DNA concentrations, we were able to extract the second-order, bimolecular association rate constant (k_{on}) of 7.8×10^8 M^{-1} s^{-1} which suggests that initial

binary complex formation is nearly diffusion limited (Figure 5.3B)[372]. We performed similar, stopped-flow DNA titration experiments with LF mutant Y274W-E291C^{CPM} and obtained comparable results (Figure 5.6).

To study the reverse conformational change of the LF domain upon dissociation of the binary complex (Dpo4•DNA^{OH}), we performed several trap assays (Figures 5.3C – 5.3E). A pre-incubated solution of Y274W-K329C^{CPM} (100 nM) and DNA^{OH} (100 nM) was rapidly mixed with a 20-fold excess of wt Dpo4 (2 μM), which contains no native Trp residues, and the fluorescence of CPM was monitored upon excitation at 290 nm (Figure 5.3C).The resulting fluorescence decrease was fit to a double exponential equation (Materials and Methods) to yield a fast phase rate of 4.5 ± 0.2 s^{-1} and a slow phase rate of 1.00 ± 0.09 s^{-1}. As an alternative approach to measure the reverse conformational change, the CPM fluorescence upon excitation at 290 nm was monitored for a pre-incubated solution of wt Dpo4 (100 nM) and DNA^{OH} (100 nM) following rapid mixing with a 20-fold excess of Y274W-K329C^{CPM} (2 μM). The resulting fluorescence trace exhibited multi-exponential kinetic behavior, displaying an initial fast phase increase with a rate of 169 ± 20 s^{-1}, followed by a second phase with a rate of 5 ± 1 s^{-1}, and a third, slow phase with a rate of 0.8 ± 0.4 s^{-1}. The two slower phase rates are in agreement with the previous trap assay. While not well resolved, we suspect that the initial fast phase represents the population of DNA^{OH} that remained unbound by wt Dpo4 following pre-incubation and was rapidly bound by the fluorescent Y274W-K329C^{CPM} trap upon mixing. Given that Dpo4 binds to DNA with an equilibrium dissociation constant (K_D) of ~10 nM[9,94,134], we would estimate that ~70% of the DNA^{OH} would be

prebound by wt Dpo4 which would leave ~30 nM of unbound DNA^{OH}. Consistently, the observed fast phase amplitude (42 ± 7%) suggests that ~40 nM DNA^{OH} remained unbound prior to mixing, and the corresponding fast rate (169 ± 20 s^{-1}) was not accurately measured on the time scale of data acquisition, as we would expect for mixing with a high concentration of Y274W-K329C^{CPM} (2 μM) (Figure 5.3B). As the association of Y274W-K329C^{CPM} to DNA^{OH} was limited by the dissociation of wt Dpo4, the remaining slow rates (5 ± 1 s^{-1} and 0.8 ± 0.4 s^{-1}) describe the LF intradomain reverse conformational change upon binary complex dissociation and are consistent with the previous trap experiment (4.5 ± 0.2 s^{-1} and 1.00 ± 0.09 s^{-1}). A trap experiment measuring the Trp emission upon excitation at 290 nm of Y274W-K329C (100 nM) pre-incubated with DNA^{OH} followed by rapid mixing with a 20-fold excess of wt Dpo4 (2 μM) also showed a biphasic increase in acceptor, CPM fluorescence with rates of 9 ± 2 s^{-1} and 0.7 ± 0.3 s^{-1} (Figure 5.3E). This result agrees with the steady-state fluorescence data of Figure 5.2B where the addition of 300 nM DNA^{OH} to apo-Y274W-K329C resulted in a 22% decrease in donor, Trp emission, as well as with the rates for the reverse conformational change upon DNA dissociation collected above (Figures 5.3C and 5.3D). Notably, similar trap experiments were conducted with another LF intradomain mutant, Y274W-E291C^{CPM}, and yielded comparable results (Figures 5.6B and 5.6C).

5.4.4 Conformational dynamics of intradomain Finger mutants during DNA binding.

To gain insight into the conformational changes observed for the Finger domain mutants (Figure 5.4), we monitored the fluorescence response of S22W-V62C^{CPM} upon

rapid mixing with varying concentrations of DNA^{OH} (20-100 nM) (Figure 5.4A). Interestingly, as with LF mutant Y274W-K329C^{CPM} (Figure 5.3), we observed fluorescence changes that were dependent on the concentration of DNA^{OH} as indicated by a comparable linear increase in rates (25 ± 3 s^{-1}, 55 ± 11 s^{-1}, and 90 ± 10 s^{-1}, for 20, 50, and 100 nM, respectively) (Figure 5.4A and inset). However, we also detected an additional, slower fluorescence phase that was independent of DNA^{OH} concentration with an average rate of 1.5 ± 0.3 s^{-1} and amplitude of $26 \pm 2\%$ (Figure 5.4A). This slower phase may represent a conformational change in the Finger domain subsequent to DNA binding, or a population of Dpo4 molecules that must first adjust conformation to allow for DNA binding. To distinguish between these two possibilities, we performed stopped-flow fluorescence anisotropy experiments (Materials and Methods). If a conformational change in the Finger domain were to occur subsequent to DNA^{OH} binding, then no detectable change in fluorescence anisotropy should be observed as this likely small conformational adjustment would not be expected to appreciably perturb the tumbling of the large binary (Dpo4•DNA) complex. However, if instead a population of Dpo4 were to undergo a slow conformational adjustment before binding to DNA, we would observe a time-dependent change in anisotropy with a rate comparable to the slow phase of the previous FRET experiment (Figure 5.4A). Indeed, upon mixing of wt Dpo4 (500 nM) with Alexa488-labeled DNA^{OH} (500 nM) and excitation at 499 nm, we observed a time-dependent change in anisotropy with a rate of 1.2 ± 0.2 s^{-1} (Figure 5.4B), indicating that a population of Dpo4 molecules were slowly associating with the DNA. Notably, at the reagent concentrations (500 nM Dpo4, 500 nM DNA^{OH}) used to achieve sufficient

anisotropy signal, the fast phase association was too rapid to be clearly detected. Control experiments wherein Alexa488-labeled DNA^{OH} was mixed with buffer showed stable anisotropy signal over time (Figure 5.8A), indicating that the anisotropy changes shown in Figure 5.4B resulted from Dpo4 binding to DNA. Furthermore, we attribute the lack of a readily apparent, slow fluorescence phase in the data collected for the LF mutants (Figures 5.4 and 5.5) to poorer signal to noise ratios during data acquisition for those constructs. However, it is clear that at longer time scales, the DNA binding data for the LF constructs significantly deviates from a single exponential fit (Figures 5.4A – 5.4C), thereby further substantiating the slower binding event.

To investigate the reverse conformational change of the Finger domain upon dissociation from DNA, we performed several trap assays as previously completed for LF mutant Y274W-K329C^{CPM}. A pre-incubated solution of S22W-V62C^{CPM} (100 nM) and DNA^{OH} (100 nM) was rapidly mixed with a 20-fold excess of wt Dpo4 (2 μM) and the CPM fluorescence upon excitation at 290 nm was recorded (Fig 4C). Similar to Y274W-K329C^{CPM}, S22W-V62C^{CPM} dissociated with biphasic kinetics. The fast (5.1 ± 0.3 s^{-1}) and slow (0.6 ± 0.1 s^{-1}) phase rates were close to those measured for Y274W-K329C^{CPM} (4.5 ± 0.2 s^{-1} and 1.00 ± 0.09 s^{-1}, respectively). Furthermore, another trap assay wherein a pre-incubated solution of wt Dpo4 (100 nM) and DNA^{OH} (100 nM) was rapidly mixed with S22W-V62C^{CPM} (2 μM) and the CPM fluorescence upon excitation at 290 nm was monitored supported this result (Figure 5.4D). As with the LF mutant, the resulting fluorescence increase was best fit to a triple exponential equation to yield an initial fast phase rate of 94 ± 8 s^{-1} (which is likely a result of S22W-V62C^{CPM} rapidly associating

with unbound DNA^{OH}) and two other phase rates of $5 \pm 1\ s^{-1}$ and $0.5 \pm 0.2\ s^{-1}$, which were comparable to those acquired from Figure 5.4C. Moreover, the presence of a slow phase in the dissociation experiments necessitates a slow association step as observed above (Figure 5.4A), rather than equilibrium binding of Dpo4 (i.e. function of forward and reverse rate constants).Together with the results for the LF mutants, these data further illuminate the complex binding kinetics of Dpo4.

5.4.5 Intradomain conformational dynamics during correct nucleotide binding and incorporation.

To monitor the intradomain motions of Dpo4 during correct nucleotide incorporation, CPM fluorescence was recorded upon excitation at 290 nm following rapid mixing of pre-incubated solutions of CPM-labeled Dpo4 mutants (200 nM) and DNA^{OH} (300 nM) with dTTP (1 mM) (Figure 5.9). Notably, we performed several control experiments to verify that fluorescence changes resulted from distance changes rather than photophysical artifacts such as changes in the environment of the Trp probe or inner filtering (Figure 5.8). Each domain of Dpo4 displayed conformational dynamics upon mixing with nucleotide, demonstrating that in addition to altering its global domain architecture, the polymerase also employs more subtle intradomain motions to achieve catalysis. Within the LF domain, Dpo4 mutants Y274W-$K329C^{CPM}$ (Figure 5.5A) and Y274-$E291C^{CPM}$ (Figure 5.5B) displayed biphasic fluorescence changes during nucleotide insertion consisting of a rapid ($16 \pm 3\ s^{-1}$ and $10 \pm 2\ s^{-1}$, respectively (Table 5.3)) increase phase (P_1) and a slow ($0.46 \pm 0.02\ s^{-1}$ and $0.89 \pm 0.04\ s^{-1}$, respectively (Table 5.3)) decrease phase (P_2). These data suggest that the donor and acceptor probes

rapidly moved near to one another before slowly moving apart. In contrast, LF mutant Y274W-R267C^{CPM} (Figure 5.5C) exhibited a single, rapid decrease phase (P_1) ($6 \pm 1\ s^{-1}$ (Table 5.3)), although with a much smaller amplitude, implying the fast separation of the FRET pair. In the Palm domain, Y118W-K137C^{CPM} (Figure 5.5D) revealed a fast initial decrease phase (P_1) subsequent to a slow increase phase (P_2) upon nucleotide addition. Interestingly, Finger domain mutants S22W-V62C^{CPM} (Figure 5.5E) and S22W-K56C^{CPM} (Figure 5.5F) demonstrated domain motions similar to the Palm domain during nucleotide incorporation, and displayed P_1 and P_2 rates comparable to other intradomain mutants (Table 5.3). However, the motion of P_2, as observed through the Finger domain, was in the same direction as P_1. Thus, while the majority of P_2 phases occur in the opposite direction of their corresponding P_1 phase, indicating a domain relaxation after catalysis, the intradomain motions of the Finger suggest additional gripping motions following P_1. With the Thumb domain mutants Y224W-K172C^{CPM} (Figure 5.5G) and Y224W-K212C^{CPM} (Figure 5.5H) each displayed only a single, fast decrease in fluorescence (P_1). Due to the tight distance constraints imposed by our intradomain analysis, Y274W-R267C^{CPM} and both Thumb domain mutants were at the practical limit of our Trp-CPM FRET system with inter-probe distances near or less than $0.5R_0$ (Table 5.1). Consequently, the signal to noise ratios for data collected with these mutants was less than ideal, and could explain atypical results, including the loss of the P_2 fluorescence phase, observed during nucleotide insertion within their respective domains.

To distinguish between intradomain motions induced by dNTP binding from those associated with the catalysis of nucleotide insertion, we repeated the above

experiments with DNA containing a dideoxy-terminated primer terminus (DNA^{H}, Materials and Methods). Interestingly, when phosphodiester bond formation was prohibited, the P_2 phase vanished (Figure 5.5, red traces) and the rates of P_1 remained unchanged (Table 5.3) upon mixing with dTTP for the majority of Dpo4 mutants. Accordingly, P_1 must represent a pre-catalytic conformational change induced by dNTP binding, while the P_2 motions must be associated with chemistry or post-chemistry events. Consistent with this assignment, the burst phase, single-turnover rates determined in Table 5.2 for the CPM-labeled Dpo4 mutants agree with the observed rates for P_2 at 20˚C (Table 5.3). However, as it has been previously shown[11], the P_2 phase cannot signify the rate-limiting step of nucleotide incorporation, but rather signifies the reverse of the conformational changes that developed during nucleotide binding and insertion, including the rate-limiting fine active site adjustments, as well as the global domain conformational transitions. For several of the Dpo4 mutants, the P_1 phase could not be well-resolved, likely resulting from the poor signal to noise ratios demonstrated by these mutants potentially made worse by reduced nucleotide binding affinities when using the dideoxy-terminated primer[373]. Interestingly, the P_2 phase for both intradomain Finger mutants (S22W-$K56^{CPM}$ and S22W-$V62C^{CPM}$) was still observed (Figures 5.5E and 5.5F) and the phase rates were virtually unaffected (Table 5.3). These results suggest the presence of additional domain motions following the fast conformational change (P_1) despite the dideoxy-terminated primer. Consistently, an interdomain mutant (Y108W-$S307^{CPM}$) from an earlier study[22] demonstrated a similar P_2 phase that was independent of phosphodiester bond formation.

5.5 Discussion

It is well-known that proteins are dynamic macromolecules that experience conformational fluctuations in solution despite static crystal structures that would often suggest otherwise[374-380]. In fact, the conformational dynamics of DNA polymerases have been extensively studied by us[9,11,12,14,22,111] and others[7,381-384], through a variety of techniques and at each major step of the catalytic cycle to help clarify the enzymatic mechanism of DNA synthesis. Our FRET-based stopped-flow[11,12,22] and single-molecule[9] methodologies have allowed for measuring the kinetic rates of the conformational changes between the distinct structural states (*i.e.* apo, binary, ternary) of the model Y-family DNA polymerase, Dpo4[149]. Although it is widely accepted that dNTP binding to the binary complex of DNA polymerases prompts protein conformational changes that substantiate the induced-fit model of nucleotide incorporation and selectivity[4,5,8,27,385], the nature of how a polymerase specifically binds to its DNA substrate is less understood.

5.5.1 Usage of both induced-fit and conformational selection mechanisms by DNA polymerases when binding to DNA.

Given the inherent flexibility of macromolecules like polymerases, two mechanisms[386-388] exist to describe the binding of a DNA polymerase to DNA; 1) an induced-fit mechanism wherein the DNA binds loosely to an inactive form of the polymerase before stimulating conformational changes to an active form and 2) a conformational selection mechanism wherein the enzyme is in an equilibrium between inactive and active forms, the latter of which can bind DNA. Our previous stopped-flow FRET[22] and computational[151] analyses support a complex binding mechanism for Dpo4

binding to DNA that includes aspects of both the induced-fit and conformational selection mechanisms. Such a combination of binding scenarios has been suggested for several other biological interactions as well[387-390]. In general, the mechanism involves multiple, interconverting conformational states of an enzyme each with varying binding competency for the substrate, which may bind to the substrate prior to undergoing conformational adjustments that serve to realize the final enzyme•substrate complex. So while all forms of the enzyme may bind to the substrate, more or fewer conformational adjustments may be necessary to reach the final competent binary complex, depending on the conformational state of the enzyme at the moment of initial substrate encounter. The data collected in this study further support merging of the two binding mechanisms for DNA binding by DNA polymerases.

5.5.2 Expanded DNA binding mechanism of Dpo4.

In our previous study[22], a Trp donor in the Palm and a CPM-acceptor in the LF (Y108W-K329C^{CPM}) provided evidence that the conformational change between the apo and DNA-bound states of Dpo4, may occur even in the absence of DNA. These unique findings prompted a follow-up computational study to examine the binding of Dpo4 to DNA[151]. It was concluded that the DNA binding mechanism is complex, consisting of a Dpo4 conformational equilibrium between three states (Figure 5.6A) with different distributions of populations at each binding stage. While the Dpo4$_A$ and Dpo4$_B$ states refer to those conformations observed in the apo and binary crystal structures, respectively, the Dpo4$_I$ state represents an inevitable intermediate that must form during the transition from Dpo4$_A$ to Dpo4$_B$. Accordingly, the complete binding process consists

of four kinetically connected steps (Steps 1-4 in Figure 5.6B) characterized by their similarity to the final bound state as well as distinct ratios of the Dpo4 conformers. The stopped-flow FRET results presented here provide experimental evidence to validate the computational model and link together previously conflicting reports concerning the DNA binding mechanism of Dpo4. Figure 5.6B depicts the proposed model for Dpo4 binding to DNA. The initial bimolecular encounter of Dpo4 with DNA (Step 1, Figure 5.6B) to form the encounter complex (E_{Enc}•DNA_n) is nearly diffusion limited ($7.8x10^8$ M^{-1} s^{-1}) as estimated in Figures 5.5A and 5.5B. Interestingly, we do not consider it coincidence that this elusive binding step could only be observed through the LF (Figure 5.3) and Finger (Figure 5.4) intradomain FRET systems. It has been previously suggested that the LF and the 14 amino acid residue linker joining it to the polymerase core mediates initial binding to DNA through a "fly-casting" mechanism wherein the positively charged domain and linker facilitate the early capture of the DNA through increased flexibility and electrostatic interactions[9,151,391]. Furthermore, it is known that a disordered loop in the Finger domain, observed in the apo structure of Dpo4, undergoes a transition to ordered upon DNA binding as shown in the Dpo4 binary complex structure[149], and that this transition is tightly coupled to the degree of DNA binding of the LF domain[151]. It is likely that our intradomain FRET system in the Finger domain is sensitive to this transition, resulting in the observed increases in FRET (Figure 4A). Moreover, the lack of appreciable signal changes in the other polymerase domains (Figure 5.3) upon rapid mixing with DNA support this assertion. Remarkably, the aforementioned LF-facilitated DNA binding and the disordered to ordered transition may

be a hallmark of the DNA binding mechanism of the Y-family polymerases.

Following the formation of E_{Enc}•DNA_n (Step 1, Figure 5.6B), our earlier stopped-flow FRET study[22] monitoring the motion between the LF and Palm domains support the inclusion of a succeeding conformational change (Step 2, Figure 5.6B) to form an intermediate complex between Dpo4 and DNA (E_{IS}•DNA_n, Figure 5.6B). During Step 2, the LF domain is transitioning from its position in the apo state ($Dpo4_A$, Figure 5.6A), seen in the apo-Dpo4 crystal structure[149], to an intermediate position ($Dpo4_I$, Figure 5.6A) wherein the complete native contacts of the bound state ($Dpo4_B$, Figure 5.6A) have not yet fully formed. This is then followed by the rate-limiting step (Step 3, Figure 5.6B) of Dpo4 binding to DNA as discovered in this study through the slow phase rate (1.5 ± 0.3 s^{-1}) observed in experiments with the intradomain Finger mutant S22W-V62C^{CPM} that was independent of DNA concentration (see Section 5.4 and Figure 5.4A). We further validated the inclusion of Step 3 (Figure 5.6B) with an experiment to record the time-dependent anisotropy of Alexa488-labeled DNA^{OH} (500 nM) upon rapid mixing with wt Dpo4 (500 nM). Consistently, a slow change in anisotropy was observed (Figure 5.4B) with a rate (1.2 ± 0.2 s^{-1}) that agreed well with the aforementioned slow phase rate from the experiments with the intradomain Finger mutant, indicating the formation of the E•DNA_n* complex (Step 3, Figure 5.6B). During Step 3, the polymerase is making the final transition from the intermediate conformation ($Dpo4_I$, Figure 5.6A) to the bound state conformation ($Dpo4_B$, Figure 5.6 A) observed in the binary crystal structure[149]. The final DNA binding step (Step 4, Figure 5.6B) involves the translocation of the polymerase by one base pair along the DNA (E•DNA_n), to vacate the position of the

primer-template junction base pair at the active site[11,12]. This rapid translocation event (>150 s^{-1}) was demonstrated through the P_0 fluorescence phase of our previous stopped-flow FRET analyses[11,12].

A previous fluorescence study[219] monitoring the emission of a Trp residue (T239W) in the LF of Dpo4 reported a dissociation rate from DNA of ~70 s^{-1}. Although this dissociation rate is much faster than those determined by ^{32}P-based kinetics[94] (0.02 s^{-1}), stopped-flow FRET[22] (3.3 s^{-1}), and single-molecule FRET[9,163] (0.2-2.5 s^{-1}), their particular fluorescence method used may have been reporting on a specific aspect of the DNA binding mechanism. We suggest that the Trp mutant was ideally suited to monitor the rapid dissociation (Step 1, Figure 5.6B) of E_{Enc}•DNA_n (Figure 5.6B) as the Trp probe was likely very sensitive to the changes in polarity that accompany DNA unbinding in the LF domain. Moreover, the stopped-flow FRET study[22] monitoring the conformational motion between the LF and Palm domains was able to detect the reverse conformational change (5 s^{-1}, Step 2, Figure 5.6B) of Dpo4 from the intermediate state ($Dpo4_I$, Figure 5.6A) to the apo state ($Dpo4_A$, Figure 5.6A), and fast fluorescence phases observed in the dissociation experiments (Figures 5.3C, 5.3E and 5.4C) of this study likely report on this event as well (4.5-9 s^{-1}). Interestingly, the fast dissociation rates measured by a single-molecule study[9] of Dpo4 also agree with this assignment (2.5 s^{-1}). The reverse of Step 3 is exemplified through the slow fluorescence phases (0.5-1 s^{-1}) observed in the dissociation experiments of this study (Figures 5.3C – 5.3E and Figures 5.4C, 5.4D) which likely represent the slow, reverse isomerization of Dpo4 to unbind the DNA. The slow rate of DNA dissociation (0.02 s^{-1}) measured through the ^{32}P-based kinetic

experiments[94] must then report on the complete unbinding process from the fully competent bound state (E•DNA_n). Taken together, our novel stopped-flow FRET findings and the results of previous studies[9,11,12,22,94,151,163,219] have uniquely probed different facets of the complex Dpo4-DNA binding process and have collectively permitted the comprehensive elucidation of the complete DNA binding mechanism (Figure 5.6B).

5.5.3 Subtle intradomain motions collectively limit single-nucleotide incorporation.

For the majority of Dpo4 intradomain mutants, two distinct FRET phases could be observed during nucleotide binding and incorporation. Unfortunately, the distance constraints imposed by our intradomain investigation resulted in sub-optimal separations between the FRET donor and acceptor probes for some of the mutants (Table 5.1), which led to poor signal to noise ratios during certain experiments. As such, clear P_1 and P_2 phases could not be observed for all Dpo4 mutants (i.e. Y274W-R267C^{CPM} and Y224W-K172C^{CPM}) (Figure 5.5 and Table 5.3), but this does not necessarily indicate that intradomain motions are not occurring during catalytic events, only that our FRET system is limited in scope.

Our investigation of the conformational motions of individual polymerase domains during nucleotide binding and incorporation is, to the best of our knowledge, the first of its kind. The intradomain FRET changes upon mixing pre-incubated Dpo4 and DNA with dNTP reveal that subtle motions within each polymerase core domain, as well as the LF domain, occur during nucleotide binding and incorporation (Figure 5.5, Table 5.3). Experiments involving the LF intradomain mutants Y274W-K329C^{CPM}, Y274W-

E291C^{CPM}, and Y274W- R267C^{CPM} (Figures 5.5A – 5.5C) expand our earlier discovery[22] of a rotational axis in the LF domain perpendicular to helices M and L. Here, we find that in addition to the putative rotation of the entire LF domain reported previously, α-helix L is also dynamic and likely bends to enhance the polymerase grip on the DNA. The Trp donor, Y274W, is located at the distal end of α-helix L of the LF domain in relative close proximity (10.7 Å, Table 5.1) to the R267C^{CPM} acceptor residue which is positioned in the middle of the helix (Figure 5.1). Based on the observed fluorescence changes, we suspect that during nucleotide insertion, the end of α-helix L containing Y274W bends towards the acceptor residues K329C^{CPM} and E291C^{CPM}, and consequently moves away from R267C^{CPM}, in a manner that would seem to strengthen the polymerase-DNA interactions for subsequent nucleotide incorporation. Alternatively, the loops containing K329C^{CPM} and E291C^{CPM} may swing towards the DNA which would lead to a similar FRET change. Comparison of multiple, ternary complex crystal structures of Dpo4 bound to undamaged DNA[84], or abasic site-containing DNA[364], and dNTP[364] reveal that the mobile LF domain likely enables proper DNA substrate alignment and orientation for catalysis. This observation has allowed for the idea that Dpo4 executes its role as a lesion bypass polymerase through an "induced-grip" approach wherein the LF domain modulates primer-template positioning within the active site to facilitate translesion synthesis[392]. Our intradomain FRET system has thus provided a glimpse of the LF domain dynamics that are applied during nucleotide binding to enhance protein-DNA interactions for efficient catalysis on undamaged DNA. As previously shown for the global conformational motions of Dpo4[12], we expect that the intradomain motions shown

here will be altered by a DNA lesion.

Apart from Y274W-R267C^{CPM} and Y224W-K172C^{CPM}, the remaining intradomain mutants displayed characteristic P_1 and P_2 transitions during correct nucleotide insertion and the P_2 phase was not observed when using a dideoxy-terminated primer for all mutants (Figure 5.9 and Table 5.3), except for those of the Finger domain, suggesting that slow intradomain motions occur following nucleotide incorporation as we rationalized previously[11]. For the Finger intradomain mutants, not only did the P_2 transition occur in the same direction as P_1, it occurred independent of phosphodiester bond formation (Figures 5.5E and 5.5F). The reason is unclear and we speculate that the observed slow intradomain motions in the Finger reflect subtle conformational transitions following the initial binding of the correct nucleotide that function to align active site residues around the replicating base pair[84,364]. As expected, this alignment occurs even when catalysis is prohibited by a chain terminating primer (DNA^H, Figures 5.5E and 5.5F, red trace). Interestingly, the rate of this conformational motion (~0.3 s^{-1}, Table 5.3) is on the order of the rate-limiting step of single-nucleotide incorporation (k_{burst}, Table 5.2) as measured by ^{32}P-based kinetic assays[22,94,220]. Furthermore, it is likely that the dynamic motions observed within each polymerase domain during nucleotide incorporation (P_2) collectively contribute to active site rearrangements (Step 7, Figure 5.6B), which we hypothesize to be the rate-determining step of single-nucleotide incorporation[11]. In fact, subtle adjustments of loops and other secondary structural elements within each domain of Dpo4 between the binary and ternary complexes have been observed structurally[149], as well as distinct changes in rotamer configurations of

amino acid side chains in the Palm (Y10) and Finger (Y48) which reside in the nucleotide-binding pocket of Dpo4.

While the results presented here, and those of our previous stopped-flow FRET studies[11,12,22], have provided significant insight into the solution-state behavior of Dpo4 during substrate binding and catalysis, higher-resolution, solution-state methodologies are necessary to fully evaluate the conformational changes that take place throughout the reaction pathway. Advancements in macromolecular NMR[393] will permit the atomic-level inspection of Dpo4 motions at a multitude of timescales, and the assignments of the backbone nitrogen, carbon, and amide proton resonances of the polymerase core[394] and LF domain[395] will strongly inform these future investigations.

Taken together, the results presented here uncover the complex DNA binding mechanism of a Y-family DNA polymerase which involves aspects of both induced-fit and conformational selection mechanisms. Furthermore, intradomain protein motions were observed throughout nucleotide binding and incorporation, some of which may kinetically limit the rate of correct nucleotide incorporation. Future investigations of the effects of proliferating cell nuclear antigen (PCNA) and DNA damage on the intradomain motions of Dpo4 are of particular intrigue.

5.6 Tables

Table 5.1: Distances between Trp donor residues and Cys-conjugated CPM acceptor fluorophores projected from X-ray crystal structures.

Domain	Mutant	Distance[a] (Å)		
		Apo[b]	Binary[c]	Ternary[d]
Little Finger	Y274W-R267C^{CPM}	10.8	10.7	10.8
	Y274W-E291C^{CPM}	28.3	28.3	28.4
	Y274W-K329C^{CPM}	28.6	28.4	28.4
Thumb	Y224W-K172C^{CPM}	16.3	16.2	16.4
	Y224W-K212C^{CPM}	14.5	14.8	14.8
Palm	Y118W-K137C^{CPM}	26.6	26.1	25.8
Finger	S22W-K56C^{CPM}	22.0	20.9	21.1
	S22W-V62C^{CPM}	24.8	24.5	24.3

[a]Estimated distances are calculated from crystal structures and do not account for the flexible linker utilized to attach the CPM fluorophore to the introduced Cys residues.

[b]Distances calculated between the C_α atom of residues in the apo structure of Dpo4 (PDB entry 2RDI).

[c]Distances calculated between the C_α atom of residues in the Dpo4•DNA binary complex structure (PDB entry 2RDJ)

[d]Distances calculated between the C_α atom of residues in the Dpo4•dideoxy-DNA•dATP ternary complex structure (PDB entry 2AGQ)

Table 5.2: Kinetic parameters of wt Dpo4 and each CPM-labeled Dpo4 mutant measured using the pre-steady-state burst assay at 20˚C.

Enzyme	k_{burst} [a] (s^{-1})	k_{ss} [a] (s^{-1})
wild-type Dpo4	0.5 ± 0.1	0.007 ± 0.001
Y274W-R267C^{CPM}	0.6 ± 0.1	0.004 ± 0.001
Y274W-E291C^{CPM}	0.6 ± 0.1	0.006 ± 0.001
Y274W-K329C^{CPM}	0.4 ± 0.2	0.014 ± 0.003
Y224W-K172C^{CPM}	0.30 ± 0.05	0.012 ± 0.001
Y224W-K212C^{CPM}	0.3 ± 0.2	0.013 ± 0.003
Y118W-K137C^{CPM}	0.5 ± 0.1	0.007 ± 0.002
S22W-K56C^{CPM}	0.3 ± 0.1	0.007 ± 0.008
S22W-V62C^{CPM}	0.3 ± 0.2	0.011 ± 0.002

Burst kinetics assays were performed in which a preincubated solution of a Dpo4 mutant (10 nM) and 5′-^{32}P-labeled DNA^{OH} (60 nM) was rapidly mixed with dTTP (100 μM) and the reaction subsequently quenched at various time points with EDTA (0.37 M). The plot of product concentration vs. time was fit to the burst equation [product] = $A[1-\exp(-k_{burst}t) + k_{ss}t]$ where A is the burst phase amplitude, k_{burst} is the observed burst rate, and k_{ss} is the observed steady-state rate.

Table 5.3: Rates of stopped-flow FRET changes for Dpo4 intradomain mutants.

Domain	Mutant	Phase	Rate (s^{-1})	
			with DNA^{OH}	with DNA^{H}
Little Finger	Y274W-R267C^{CPM}	P_1	6 ± 1	-
		P_2	-	-
	Y274W-E291C^{CPM}	P_1	10 ± 2	12 ± 2
		P_2	0.89 ± 0.04	-
	Y274W-K329C^{CPM}	P_1	16 ± 3	11 ± 2
		P_2	0.46 ± 0.02	-
Thumb	Y224W-K172C^{CPM}	P_1	14 ± 5	-
		P_2	0.17 ± 0.04	-
	Y224W-K212C^{CPM}	P_1	12 ± 1	14 ± 4
		P_2	0.131 ± 0.008	-
Palm	Y118W-K137C^{CPM}	P_1	9 ± 3	5 ± 2
		P_2	0.05 ± 0.02	-
Finger	S22W-K56C^{CPM}	P_1	14 ± 3	18 ± 4
		P_2	0.4 ± 0.2	0.3 ± 0.1
	S22W-V62C^{CPM}	P_1	17 ± 2	15 ± 4
		P_2	0.2 ± 0.1	0.2 ± 0.1

5.7 Figures

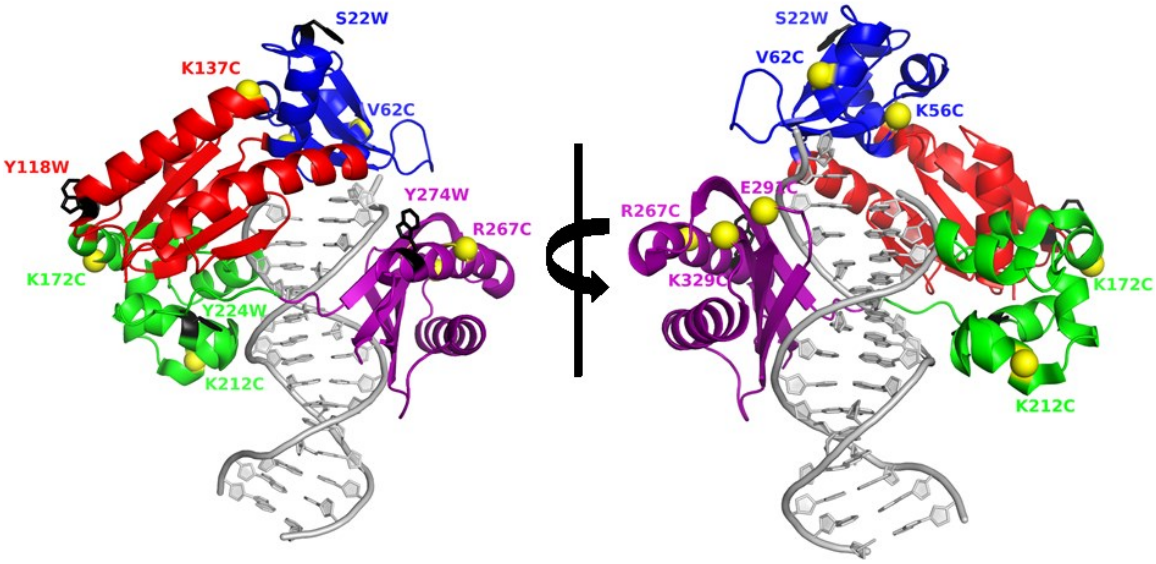


Figure 5.1: Locations of intradomain donor and acceptor FRET pairs mapped onto the X-ray crystal structure of DNA-bound Dpo4.

The Finger, Palm, Thumb, and LF domains are colored blue, red, green, and purple, respectively. Each domain contains one, unique Trp donor residue (black stick models) and one or more sites for CPM acceptor attachment (yellow spheres). The DNA substrate is shown in gray. For clarity, the right structure is rotated 180˚ relative to the left structure.

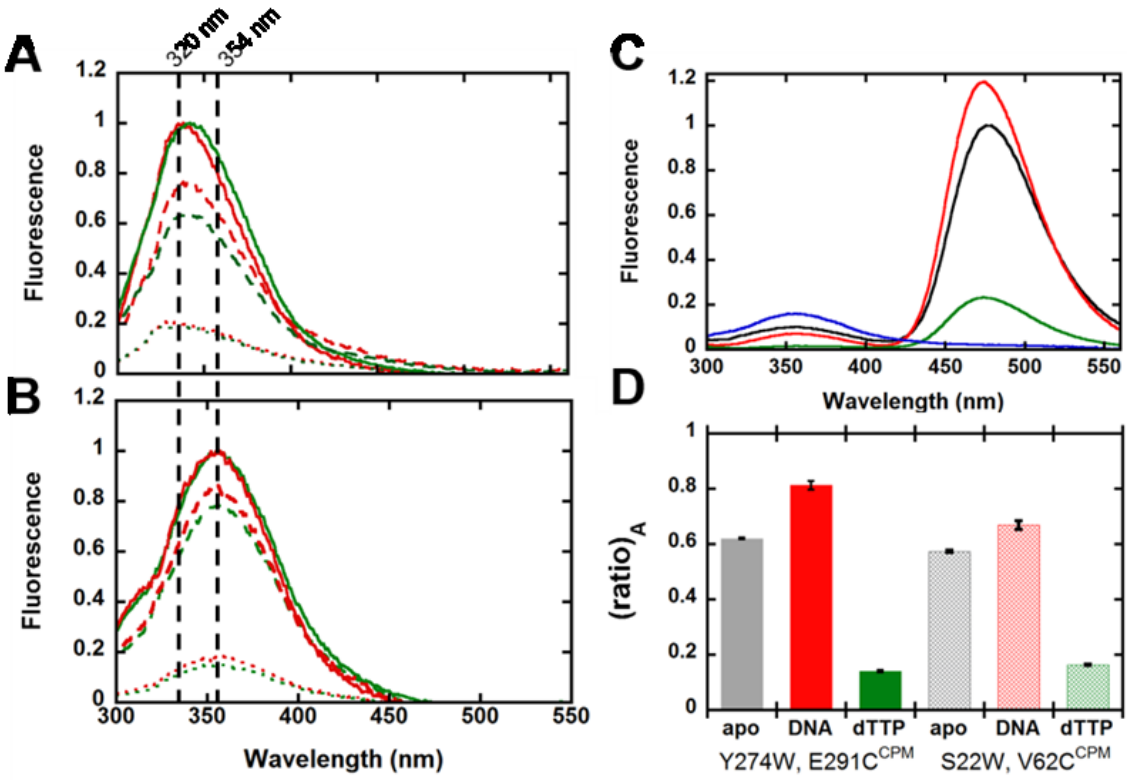


Figure 5.2: Steady-state fluorescence of Trp-containing Dpo4 mutants and select CPM-labeled intradomain FRET constructs.

Emission spectra for Dpo4 mutants (200 nM) were recorded at 20˚C with an excitation wavelength of 290 nm. (A) Emission spectra of mutants Y118W-K137C (green) and Y224W-K172C (red) in the apo state (solid line) and after sequential additions of 300 nM DNA^{OH} (dashed line) and 1 mM dTTP (dotted line). (B) Emission spectra of mutants Y274W-E291C (green) and S22W-V62C (red) in the apo state (solid line) and after sequential additions of 300 nM DNA^{OH} (dashed line) and 1 mM dTTP (dotted line). The black dashed lines at 320 and 354 nm of panels A and B emphasize the differences in the

(continued)

Figure 5.2: continued

emission maximum wavelengths between the various Trp mutants. (C) Emission spectra for CPM-labeled mutant Y274W-E291C^{CPM} in the apo state (black), and after sequential additions of 300 nM DNA^{OH} (red) and 1 mM dTTP (green). The emission spectra of Y274W-E291C before CPM-labeling is shown in blue. (D) The differences in $(ratio)_A$ for select CPM-labeled Dpo4 mutants in the apo state and following sequential additions of 300 nM DNA^{OH} and 1 mM dTTP. Error bars represent the standard deviation calculated from three separate experiments.

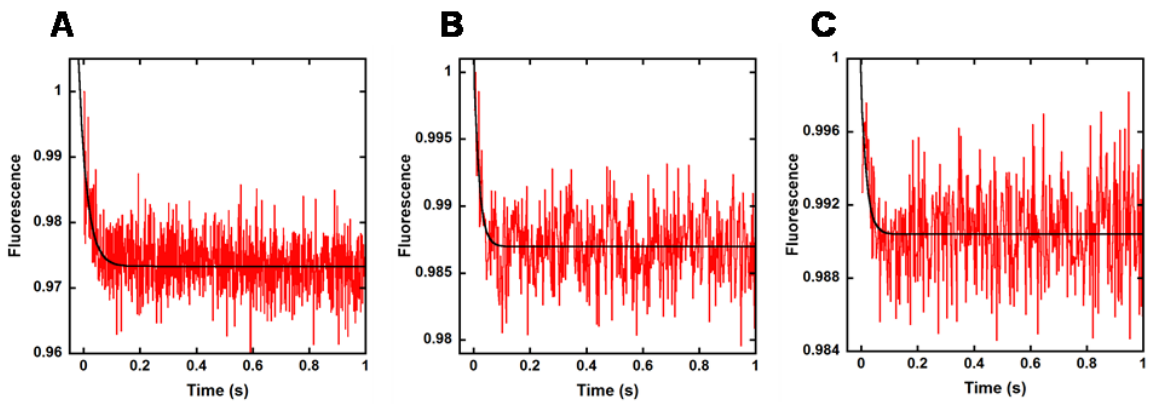


Figure 5.3: Conformational changes upon DNA binding for select intradomain Dpo4 FRET constructs.

100 nM DNA^{OH} was rapidly mixed with 100 nM of (A) Y118W-$K137C^{CPM}$, (B) Y224W-$K172C^{CPM}$, and (C) Y224W-$K212C^{CPM}$ and CPM fluorescence was recorded upon excitation at 290 nm. The black lines are single exponential fits to the data to yield observed association rates of 80 ± 20 s^{-1}, 50 ± 10 s^{-1}, and 30 ± 10 s^{-1}. Errors are standard deviations calculated from ≥ 5 replicate experiments.

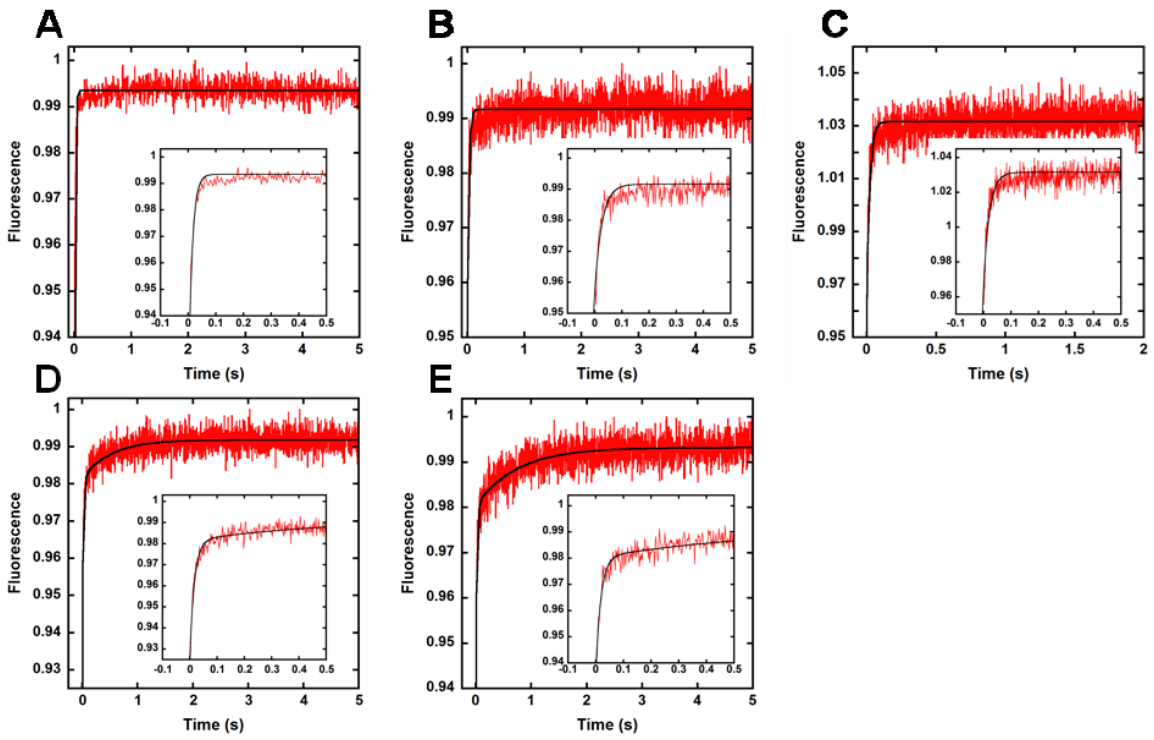


Figure 5.4: Conformational dynamics of select intradomain Dpo4 FRET constructs during DNA binding.

100 nM DNA^{OH} was rapidly mixed with 100 nM of LF mutants (A) Y274W-$R267C^{CPM}$, (B) Y274W-$E291C^{CPM}$, and (C) Y274W-$K329C^{CPM}$ as well as Finger mutants (D) S22W-$K56C^{CPM}$ and (E) S22W-$V62C^{CPM}$ and CPM fluorescence was recorded upon excitation at 290 nm. For traces A-C, the black line depicts single exponential fits to the data yielding observed rates of association of 75 ± 4 s^{-1}, 55 ± 3 s^{-1}, 86 ± 12 s^{-1}, respectively. For traces D and E, the black line depicts double exponential fits to the data yielding the following kinetic parameters: k_1 = 57 ± 8 s^{-1}, A_1 = 0.83 ± 0.01, k_2 = 2.0 ± 0.3 s^{-1}, and A_2 = 0.17 ± 0.01 for trace D; k_1 = 55 ± 11 s^{-1}, A_1 = 0.76 ± 0.01, k_2 = 1.6 ± 0.4 s^{-1}, and A_2 = 0.24 ± 0.01 for trace E. The inset figures feature the fast time points for clarity.

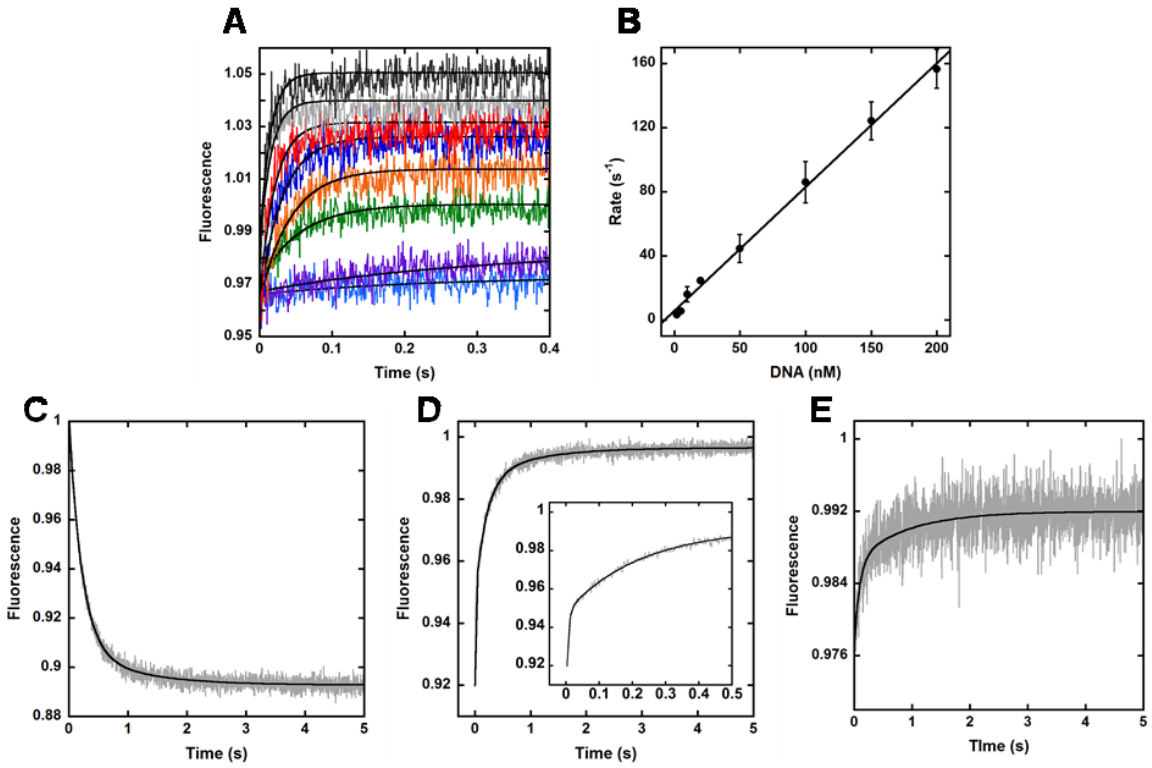


Figure 5.5: Conformational changes in the LF domain upon DNA binding.

(A) Dpo4 LF mutant Y274W-K329C^{CPM} (100 or 1000 nM) was rapidly mixed with varying concentrations of DNA^{OH} (2-200 nM, light blue to black traces, respectively) and the CPM fluorescence was monitored upon excitation at 290 nm. The black lines depict single exponential fits to the data. (B) The observed rate constants (k_{obs}) extracted from (A) were plotted against the corresponding DNA concentrations. Errors bars represent the standard deviations calculated from multiple replicate experiments. The slope of the best fit line to the data signifies the bimolecular association rate constant (k_{on}) for Dpo4 binding to DNA^{OH} (7.8×10^8 M^{-1} s^{-1}). (C) CPM fluorescence of Y274W-K329C^{CPM} (100

(continued)

Figure 5.5: continued

nM) pre-incubated with DNA^{OH} (100 nM) upon excitation at 290 nm following rapid mixing with wt Dpo4 trap (2 μM). (D) CPM fluorescence of Y274W-K329C^{CPM} (2 μM) upon excitation at 290 nm following rapid mixing with a pre-incubated solution of wt Dpo4 (100 nM) and DNA^{OH} (100 nM). The inset features the fast time points to illustrate the multiexponential behavior of the data (E) Trp fluorescence of Y274W-K329C (100 nM) pre-incubated with DNA^{OH} (100 nM) upon excitation at 290 nm following rapid mixing with wt Dpo4 trap (2 μM). The black lines depict double (C and E) or triple (D) exponential fits to the data.

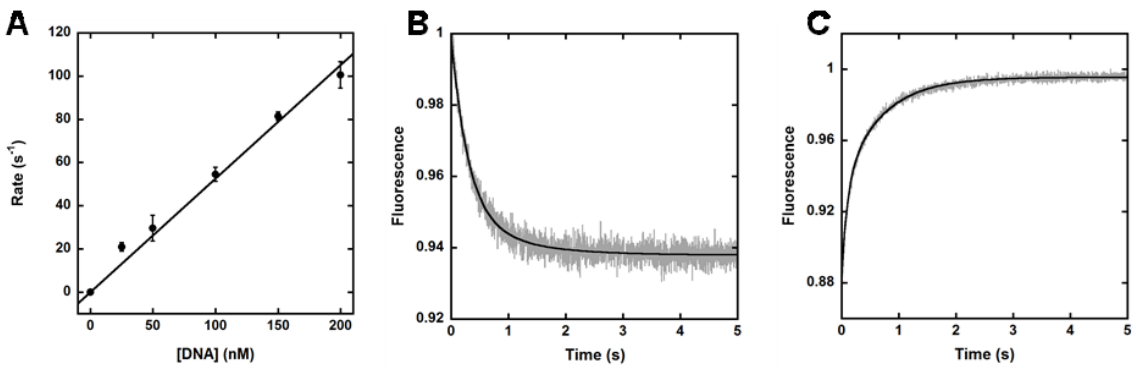


Figure 5.6: Conformational dynamics of Y274W-E291C^{CPM} during DNA association and dissociation.

(A) Observed rates collected for Y274W-E291C^{CPM} binding to varying concentrations of DNA^{OH} (25-200 nM) plotted against DNA concentration. From the slope of the best fit line, the second-order, bimolecular association rate constant (k_{on}) was calculated as 5.3x10^8 M^{-1} s^{-1}. (B) CPM fluorescence of Y274W-E291C^{CPM} (100 nM) pre-incubated with DNA^{OH} (100 nM) upon excitation at 290 nm following rapid mixing with wt Dpo4 trap (2 μM). The data were fit to a double exponential equation (black line) to yield fast and slow phase rates of 3.5 ± 0.2 s^{-1} (A_1 = 0.7 ± 0.1) and 1.3 ± 0.4 s^{-1} (A_2 = 0.3 ± 0.1), respectively. (C) CPM fluorescence of Y274W-E291C^{CPM} (2 μM) upon excitation at 290 nm following rapid mixing with a pre-incubated solution of wt Dpo4 (100 nM) and DNA^{OH} (100 nM). The data were fit to a triple exponential equation (black line) to yield the following kinetic parameters: k_1 = 80 ± 13 s^{-1}, A_1 = 0.31 ± 0.01, k_2 = 5 ± 1 s^{-1}, A_2 = 0.37 ± 0.04, k_3 = 1.2 ± 0.2, and A_3 = 0.32 ± 0.05.

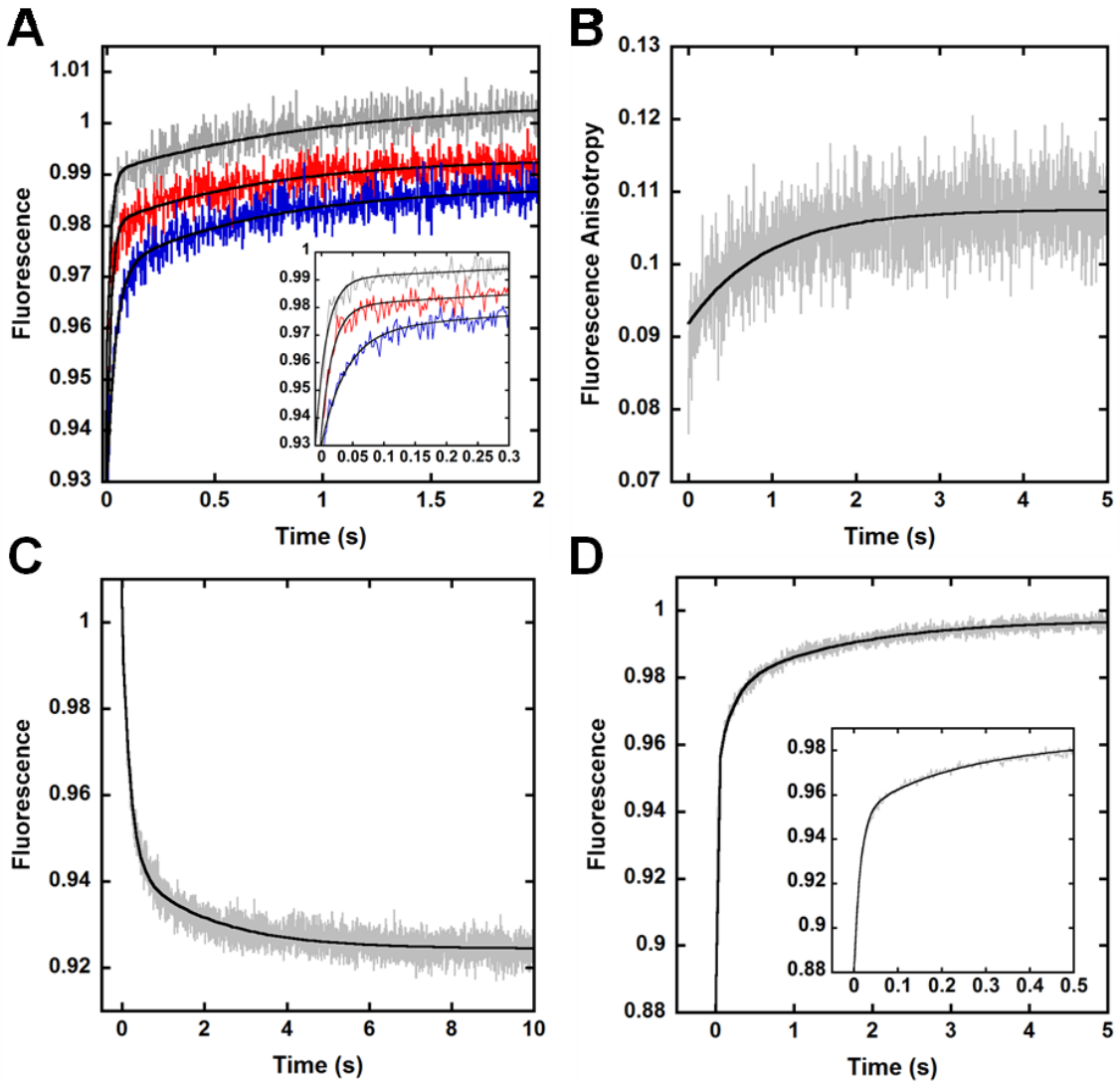


Figure 5.7: Conformational changes in the Finger domain upon DNA binding.

(A) Dpo4 Finger mutant S22W-V62C^{CPM} (100 nM) was rapidly mixed with varying concentrations of DNA^{OH} (20-200 nM, blue to gray traces, respectively) and the CPM fluorescence was monitored upon excitation at 290 nm. The black line depicts a double exponential fit to the data and the inset highlights the concentration dependence of the

(continued)

Figure 5.7: continued

initial fast rate. (B) wt Dpo4 (500 nM) was rapidly mixed with Alexa488-labeled DNA^{OH} (500 nM) and the change in fluorescence anisotropy was measured upon excitation at 499 nm. The black line depicts a single exponential fit to the data. (C) CPM fluorescence of S22W-V62C^{CPM} (100 nM) pre-incubated with DNA^{OH} (100 nM) upon excitation at 290 nm following rapid mixing with a 20-fold excess of wt Dpo4 (2 μM). The black line depicts a double exponential fit to the data. (D) CPM fluorescence of S22W-V62C^{CPM} (2 μM) was monitored upon excitation at 290 nM following rapid mixing with a pre-incubated solution of wt Dpo4 (100 nM) and DNA^{OH} (100 nM). The black line depicts a triple exponential fit to the data and the inset features the fast time points to illustrate the multiexponential behavior of the data.

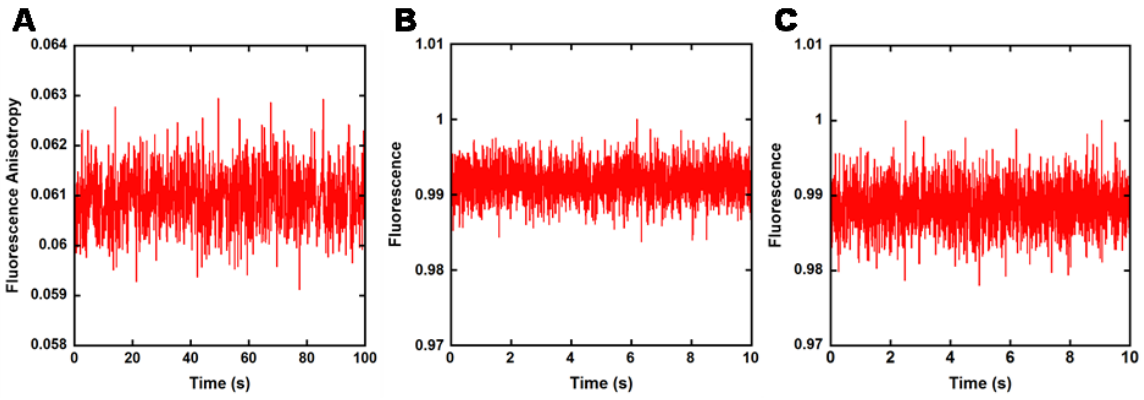


Figure 5.8: Control stopped-flow experiments.

(A) Fluorescence anisotropy recorded upon mixing Alexa488-labeled DNA^{OH} (500 nM) with buffer. Anisotropy did not vary over time indicating that the changes in anisotropy displayed in Figure 5.4B occurred as a result of Dpo4 binding. (B) A pre-incubated solution of Dpo4 intradomain mutant Y274W-K329C (200 nM) and DNA^{OH} (300 nM) was rapidly mixed with dTTP (1 mM) and the Trp fluorescence was monitored upon excitation at 290 nm. No detectable change in fluorescence was observed suggesting that FRET changes observed in Figure 5.5 were due to distance changes between the donor and acceptor probes rather than variations in the local environment of the Trp probe. (C) $Y274W\text{-}K329C^{CPM}$ (200 nM) was rapidly mixed with dTTP (1 mM) and CPM-fluorescence was monitored upon excitation at 290 nm. In the absence of DNA, no detectable change in fluorescence was observed, thereby demonstrating that inner filtering, caused by the large concentration of dTTP, occurred within the dead time of the instrument (~1 ms). Accordingly, the changes in CPM-fluorescence observed in Figure 5.5 serve as proxies for relative distance changes between the donor and acceptor probes.

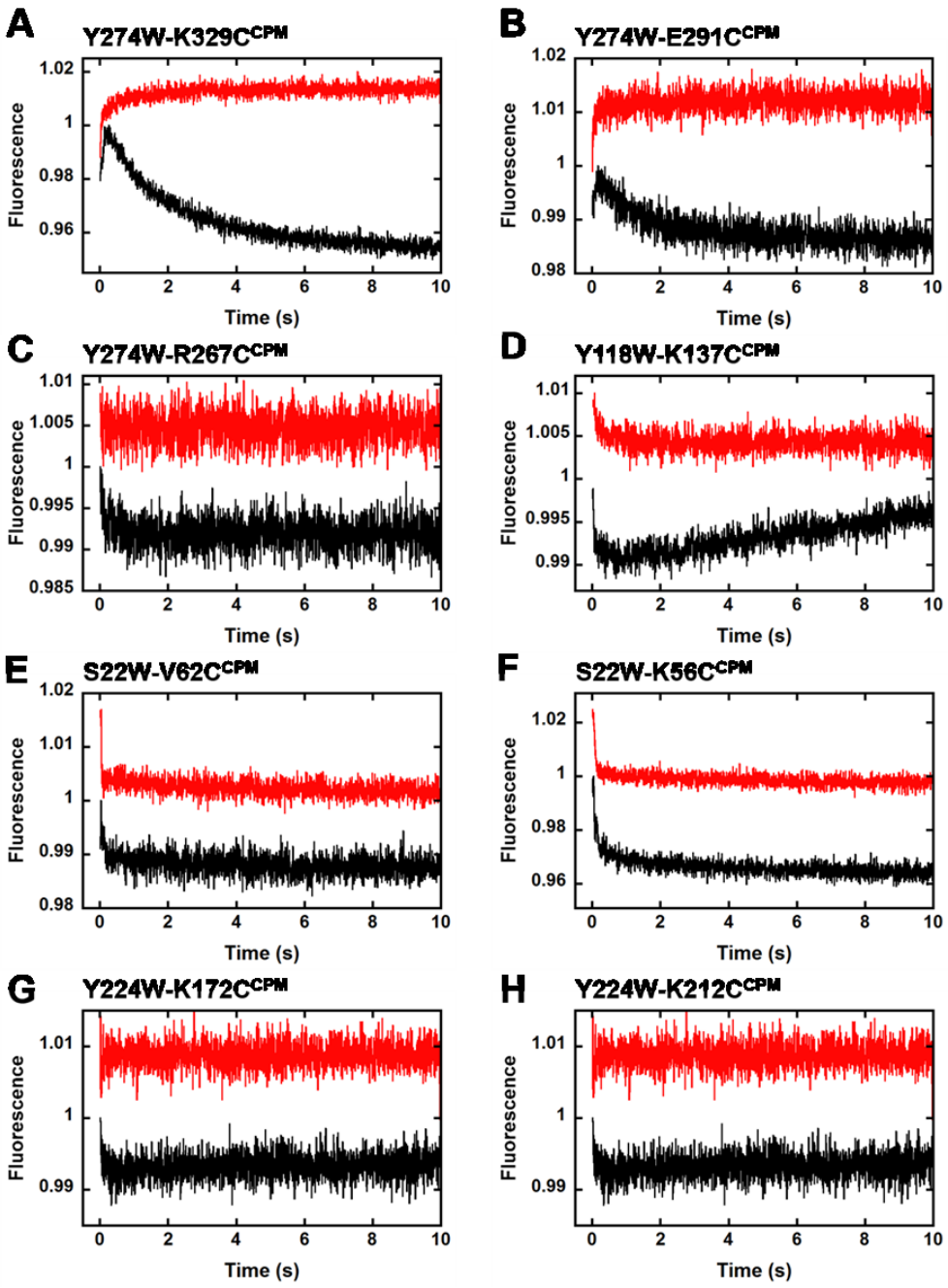


Figure 5.9: Intradomain conformational dynamics during nucleotide binding and incorporation.

(continued)

Figure 5.9: continued

(A-H) Changes in CPM fluorescence upon excitation at 290 nm following rapid mixing of a pre-incubated solution of intradomain Dpo4 mutant (200 nM) and either DNA^{OH} (300 nM, black traces) or DNA^{H} (300 nM, red traces) with dTTP (1 mM) at 20˚C.

Figure 5.10: Complex DNA binding mechanism of Dpo4.

(A) While in the unbound state, Dpo4 is able to interconvert between three conformational states ($Dpo4_A$, $Dpo4_I$, and $Dpo4_B$), each with varying propensity for binding to DNA. Our intradomain analysis has permitted assigning forward and reverse rate constants to these transitions. (B) E_{apo}, E_{Enc}, E_{IS}, E, E′, and E″ represent different conformations of Dpo4 throughout the DNA binding and nucleotide binding and incorporation process. E•DNA* and E•DNA represent non-productive and productive binary complexes, respectively, connected through a DNA translocation event. P_1 and P_2 signify the fluorescence phases observed via our FRET technique.

Chapter 6. Investigation of Sliding DNA Clamp Dynamics by Single-Molecule Fluorescence, Mass Spectrometry, and Structure-Based Modeling

Reproduced in part with permission from Gadkari V.V., Harvey S.R., Raper, Austin T., Chu W, Wang J, Wysocki V.H., and Suo Z. (2018) Investigation of Sliding DNA Clamp Dynamics by Single-Molecule Fluorescence, Mass Spectrometry, and Structure-Based Modeling. *Nucleic Acids Research*. DOI: 10.1093/nar/gky125.

Author Contribution

Varun V. Gadkari, Sophie R. Harvey, and Austin T. Raper contributed equally to this work. V.V.G and A.T.R designed all protein constructs for smFRET. V.V.G generated and purified all protein constructs. V.V.G and A.T.R designed and executed all smFRET experiments. S.R.H designed and performed all mass spectrometry experiments. Wenting Chu performed all *in silico* experiments. A.T.R wrote the manuscript. V.V.G, S.R.H, A.T.R, and W.C all contributed during peer revisions. Drs. Jin Wang, Vicki H. Wysocki, and Zucai Suo helped to edit the manuscript.

6.1 Abstract

Proliferating cell nuclear antigen (PCNA) is a trimeric ring-shaped clamp protein that encircles DNA and interacts with many proteins involved in DNA replication and repair. Despite extensive structural work to characterize the monomeric, dimeric, and trimeric forms of PCNA alone and in complex with interacting proteins, no structure of PCNA in a ring-open conformation has been published. Here, we use a multidisciplinary approach, including single-molecule Förster resonance energy transfer (smFRET), native ion mobility-mass spectrometry (IM-MS), and structure-based computational modeling, to explore the conformational dynamics of a model PCNA from *Sulfolobus solfataricus* (*Sso*), an archaeon. We found that *Sso* PCNA samples ring-open and ring-closed conformations even in the absence of its clamp loader complex, replication factor C, and transition to the ring-open conformation is modulated by the ionic strength of the solution. The IM-MS results corroborate the smFRET findings suggesting that PCNA dynamics are maintained in the gas phase and further establishing IM-MS as a reliable strategy to investigate macromolecular motions. Our molecular dynamic simulations agree with the experimental data and reveal that ring-open PCNA often adopts an out-of-plane left-hand geometry. Collectively, these results implore future studies to define the roles of PCNA dynamics in DNA loading and other PCNA-mediated interactions.

6.2 Introduction

In all three domains of life, DNA replication is an intricate, and stringently coordinated process that ensures faithful and efficient copying of genetic material. Such a complicated process is accomplished by a variety of enzymes including DNA

polymerases [5,25,26,33,228,396-398], glycosylases [399-401], ligases [402,403], and nucleases [404-406], which often interact with an evolutionarily conserved sliding DNA clamp. While encircling duplexed DNA, the sliding DNA clamp functions as a scaffold for DNA replication and repair machinery [407-409]. Moreover, sliding DNA clamps have been identified and studied in all domains of life, *e.g.* the β clamp of *Escherichia coli*, the gene 45 protein (gp45) of T4 bacteriophage, and proliferating cell nuclear antigen (PCNA) in yeast, humans, and the archaeon *Sulfolobus solfataricus* (*Sso*) [410,411].

Previous studies have shown that despite low sequence similarity, all sliding DNA clamps share a similar toroidal structure with a central hole to accommodate a DNA duplex [410]. Despite the highly conserved ring-shaped structure of DNA sliding clamps, the complexity of their subunit composition varies. For instance, the β clamp of *E. coli* is homodimeric, while PCNA clamps in archaeal and eukaryotic organisms are trimeric [410]. Although many PCNA clamps are homotrimeric, *Sso* PCNA is heterotrimeric, consisting of three distinct PCNA monomer subunits: PCNA1, 2, and 3 [411]. As a result, proper subunit arrangement, subunit interfaces [412-414], and exclusive subunit binding partners have been identified for *Sso* PCNA1, PCNA2, and PCNA3 [359,411,414,415]. Previous studies have established that formation of the *Sso* PCNA trimer follows a sequential process, which begins with dimerization of PCNA1 and PCNA2, followed by recruitment of PCNA3 to form the heterotrimer [411-414].

While crystal structures exist for each *Sso* PCNA monomer [413], the PCNA1:PCNA2 dimer [413], and the ring-closed PCNA heterotrimer [412-414], no crystal structure is available for the *Sso* PCNA heterotrimer in a ring-open conformation.

However, biochemical experiments have shown that the *Sso* PCNA heterotrimer opens at the interface between PCNA1 and PCNA3 in order to be loaded onto DNA by a clamp loading accessory protein, replication factor C (RFC) [416]. To date, only the gp45 homotrimer of T4 bacteriophage has been crystallized in a ring-open conformation, but only while in complex with its clamp loader assembly (gp62/gp44 complex) [417]. In this structure, the clamp is open out-of-plane and adopts a right-hand geometry, consistent with the right-handed helix of double-stranded DNA [417]. Paradoxically, fluorescence studies report that the gp45 DNA clamp exists primarily in a ring-open form while in solution [418], yet only the ring-closed form has been crystallized in the absence of the clamp loader [419]. Accordingly, sliding DNA clamps likely undergo conformational dynamics while free in solution, as well as in complex with clamp loaders, DNA, or both.

As sliding DNA clamps serve significant roles in numerous fundamental aspects of DNA metabolism, it is crucial to understand if their conformational dynamics may influence biochemical events. Here, we use a state-of-the-art multidisciplinary approach, including single-molecule Förster resonance energy transfer (smFRET), native ion mobility-mass spectrometry (IM-MS), and structure-based computational modeling, to characterize the dynamics of *Sso* PCNA. Our smFRET results suggest that PCNA exists in dynamic equilibrium between ring-open and ring-closed conformations at the PCNA1:PCNA3 interface, and that ionic strength can modulate the transition kinetics between these states. Moreover, this dynamic open-closed equilibrium of *Sso* PCNA can also be monitored in the gas-phase using IM-MS as demonstrated by two populations of distinct collisional cross sections, which were also sensitive to changes in ionic strength.

Finally, structure-based computational modeling of *Sso* PCNA agreed with the experimentally observed two-state system and its dependence on ionic strength. However, in contrast to the ring-open structure of gp45 in complex with its clamp loader, our computational results suggest an unusual out-of-plane ring-opening of PCNA at the PCNA1:PCNA3 interface to mainly adopt left-hand geometry. This result may have general implications for loading and unloading of PCNA to DNA. Taken together, our results demonstrate the dynamics of an evolutionarily conserved sliding DNA clamp, and implore future efforts to understand how PCNA motions may influence protein-protein and protein-DNA interactions.

6.3 Materials and Methods.

6.3.1 Expression and purifications of proteins.

Sso PCNA plasmids were obtained as generous gifts from the lab of Dr. Stephen D. Bell at Indiana University in Bloomington, IN. Genes coding for *Sso* PCNA1, PCNA2, PCNA3 were expressed by three separate plasmids, while the covalently-linked PCNA heterotrimer (PCNA1-2-3) was expressed as a fusion from a single gene encoding PCNA1, PCNA2, and PCNA3 covalently joined at the PCNA1:PCNA2 and PCNA2:PCNA3 interfaces by amino acid linkers (ASGAGGSEGGGSEGGTSGAT) as described previously [416]. To this construct, site-specific S→C mutations were engineered at residues S64 of PCNA1 and S189 of PCNA3 for fluorophore labeling. Additionally, a gene sequence encoding an AviTag peptide (GLNDIFEAQKIEWHE) to allow site-specific biotinylation by BirA [420] and subsequent surface immobilization for smFRET studies was added to the C-terminus of PCNA1-2-3 before the 6xHis tag.

All expression plasmids were individually transformed into *E. coli* strain Rosetta (DE3) and expressed separately using autoinduction in ZYP-5052 medium [421]. After growth, cells were harvested (4,000 rpm for 20 min), resuspended in Buffer A (50 mM HEPES (pH 8 at 4°C), 100 mM NaCl, and 1 mM DTT), and lysed by French press at 20,000 PSI before clarifying the lysate by ultracentrifugation (40,000 rpm for 40 min). As *Sso* PCNA is thermostable, the clarified lysate was subject to heat shock by incubation at 60°C for 10 min to precipitate contaminating proteins which were subsequently removed by ultracentrifugation (40,000 rpm for 40 min).

Sso PCNA1, PCNA2, and PCNA3 subunits were expressed and purified independently. Each subunit was purified by anion exchange, heparin, and size exclusion chromatography. Briefly, a PCNA subunit was bound to a 5 mL HiTrap Q column (GE Healthcare) and eluted through a linear salt gradient (up to 1 M NaCl). Fractions containing the PCNA monomer were dialyzed against Buffer A, loaded onto a 5 mL HiTrap heparin column (GE Healthcare), and eluted through a linear salt gradient (up to 1 M NaCl). Fractions with negligible nucleic acid contamination (260/280 ratio < 0.6) were dialyzed against Buffer A, concentrated to less than 1 mL, loaded and eluted from a HiPrep 26/60 Sephacryl S-100 column (GE Healthcare). The PCNA heterotrimer was formed by incubating equimolar amounts of purified PCNA1, PCNA2, and PCNA3 at 4°C for one hour in Buffer A, before purification by size-exclusion chromatography, as described above.

Following heat shock, the covalently-linked PCNA1-2-3 protein was bound to Ni^{2+}-NTA resin (Qiagen) and eluted through a linear imidazole gradient (up to 1 M

imidazole) in Buffer A. Following dialysis against Buffer A, the protein was purified by HiTrap heparin and size-exclusion chromatography as described above for the PCNA monomers. The purified PCNA1-2-3 was then biotinylated by following a previously published protocol [420,422] before fluorophore labeling with a 15-fold molar excess of thiol-reactive Cy3- and Cy5-maleimide (Lumiprobe) overnight at 4°C in a buffer containing 50 mM Tris (pH 7.2), 150 mM NaCl, 0.5 mM TCEP, and 10% glycerol. Free dye was removed by size-exclusion chromatography. The extent of fluorophore labeling was verified by measuring absorbance at 280, 552, and 650 nm for PCNA, Cy3, and Cy5, respectively.

6.3.2 Single-molecule measurements.

All single-molecule measurements were conducted on a custom built, prism-type total internal reflection microscope, as described previously [9,111,352]. Imaging chambers were assembled from quartz slides and coverslips that were cleaned, passivated, and biotinylated by following a published protocol [352,423]. Biotinylated, Cy3-Cy5 labeled PCNA1-2-3 was then surface immobilized following the addition of neutravidin (0.2 mg/mL). After rinsing the imaging chamber with T50 buffer (10 mM Tris HCl, pH 8.0, 50 mM NaCl) and addition of imaging buffer (50 mM HEPES [pH 7.6], 0.8% w/v d-Glucose, 2 mM Trolox, 0.1 mg/ml BSA, 1 mg/ml glucose oxidase, 0.04 mg/ml catalase) containing 0, 250, 500, 750, or 1,000 mM NaCl, single-molecule movies were recorded at 25°C using an Andor iXon 897 EM-CCD. Notably, the inclusion of Trolox in the imaging buffer limited acceptor blinking [424]. The movies were recorded over several minutes at a reduced laser power and a slow frame rate (2 frames per second) to limit

observed rapid photobleaching of Cy3 and Cy5 conjugated to PCNA1-2-3 at higher laser powers.

6.3.3 Single-molecule data analysis.

Single-molecule movies were processed using IDL (ITT Visual Information Solutions) and custom MATLAB scripts (Center for the Physics of Living Cells, University of Illinois at Urbana-Champaign) to generate traces of donor and acceptor fluorescence intensities for each detected molecule, as previously described [9,111]. The traces were subsequently processed and background corrected in MATLAB. Traces showing clear anti-correlated donor and acceptor signals, as well as exhibiting single donor photobleaching events, were selected. FRET efficiency values were calculated as apparent FRET (E_{app}) using Equation 6.1, where I_D, and I_A are the donor and acceptor fluorescence intensities, respectively.

$$E_{app} = \frac{I_A}{I_D + I_A} \quad \text{(Eq. 6.1)}$$

For every experiment, FRET efficiency values from >200 FRET trajectories were collected and binned to generate histograms depicting the relative population distributions of FRET events. These data were fit to a sum of Gaussian functions using MATLAB, and percent occupancy of FRET states was calculated as the total area under the individual Gaussian fits. Dwell time analysis was performed as previously described [9,111]. Briefly, the duration of each FRET event for the selected molecules was quantified using a thresholding analysis. The “Low FRET” state was limited by thresholds at FRET efficiencies of 0.2 and 0.65, and the “High FRET” state was limited by thresholds at FRET efficiencies of 0.65 and 1. The resulting dwell times were compiled to generate

survivor functions as previously described [9,111] which were fit to a single exponential decay equation (Equation 6.2), where $f(t)$ is the fraction of molecules in the designated FRET state after time t, A is the amplitude of the function, and k is the decay rate constant associated with the designated FRET state.

$$f(t) = Ae^{-kt} \text{ (Eq. 6.2)}$$

6.3.4 Mass spectrometry.

All mass spectrometry and ion mobility-mass spectrometry studies were performed on an in-house modified Synapt G2 (Waters, Milford, UK) [425]. Typical experimental parameters were as follows; capillary voltage of 1.2-1.4 kV, a sampling cone of 20 V, source temperature of 20°C and trap collision energy of 5 V. Gas flows in the trap, helium cell and IM cell were 4, 120, and 60 mL/min respectively. Experimental collisional cross sections were determined using a calibration procedure, with β-lactoglobulin, avidin, concanavilin A, transthyretin, and serum amyloid P component as standards.[426] In all cases, a trap wave velocity of 250 m/s and height of 3.5 V, ion mobility wave velocity of 300 m/s and height of 17 V, and a transfer wave velocity of 100 m/s and a height of 2 V were used.

6.3.5 Structure-based model.

An initial coarse-grained Cα structure-based model (SBM) was generated from the crystal structure of *Sso* heterotrimeric PCNA (PDB 2HII) (Chains 1, 2, and 3 contain 249, 243, and 243 amino acids, respectively)[414] using the SMOG on-line toolkit, which included one bead on the Cα atom of each residue of the complex [427-429]. The native contact map was built by the Shadow Algorithm [429]. Aiming to sample the ring-open

conformation of heterotrimeric PCNA, the native contacts between chain 1 and chain 3 were not included in this model. The potential energy function consisted of both bonded and nonbonded terms. Additionally, we introduced charge characterization into our SBM to study the electrostatic interactions in the system. As a result, the potential energy form used in this study is given by the following equation:

$$\begin{aligned} V = & \sum_{bonds} \varepsilon_r \left(r - r_0\right)^2 + \sum_{angles} \varepsilon_\theta \left(\theta - \theta_0\right)^2 \\ & + \sum_{dihedrals} K_\phi^{(n)} \left(1 - \cos\left(n \times \left(\phi - \phi_0\right)\right)\right) \\ & + \sum_{contacts} \varepsilon_{ij} \left(5\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - 6\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{10} \right) \\ & + \sum_{non-contacts} \varepsilon_{NC} \left(\frac{\sigma_{NC}}{r_{ij}}\right)^{12} + V_{Debye-H} \end{aligned} \qquad \text{(Eq. 6.3)}$$

In Eq. 6.3, $\varepsilon_r = 100\varepsilon$, $\varepsilon_\theta = 20\varepsilon$, $K_\phi^{(1)} = \varepsilon$ and $K_\phi^{(3)} = 0.5\varepsilon$. The interaction strength of the Lennard-Jones type potential is proportional to the statistical potential reported for the residue types of *i* and *j* by Miyazawa and Jernigan (MJ) [430] to generate the "flavored model" [431]. Therefore, the coefficient of nonbonded contacts, ε_{ij}, is set as follows:

$$\varepsilon_{ij} = \left(\gamma \left(\frac{\varepsilon_{ij}^{MJ}}{\overline{\varepsilon}^{MJ}} - 1 \right) + 1 \right) \qquad \text{(Eq. 6.4)}$$

where ε_{ij}^{MJ} is the original MJ potential, $\overline{\varepsilon}^{MJ}$ is the mean value of the entire set of MJ weights in the complex system, and γ is set to 1.0 corresponding to the "flavored model"

[431].

The electrostatic interaction is calculated by the Debye-Hückel model, which can quantify the strength of charge-charge attractions and repulsions at various salt concentrations:

$$V_{Debye-H\cdots} = \Gamma_{DH} \times K_{coulomb} B(\kappa) \sum_{ij} \frac{q_i q_j \exp(-\kappa r_{ij})}{\varepsilon r_{ij}} \quad \text{(Eq. 6.5)}$$

In Eq. 6.5, $K_{coulomb} = 4\pi\varepsilon_0 = 138.94$ kJ•mol^{-1}•nm•e^{-2} is the electric conversion factor; $B(\kappa)$ is the salt-dependent coefficient; κ^{-1} is the Debye screening length, which is directly influenced by the salt concentration C_{salt} ($\kappa \approx 3.2\sqrt{C_{salt}}$); ε is dielectric constant, which is set to 80 during the simulations; and Γ_{DH} is the energy scaled coefficient, which aims to make the total energy balanceable. In our model, Lys and Arg residues have a positive point charge (+e) while Asp and Glu have a negative point charge (-e). All charges are placed on the Cα atoms. Besides the systems of variable ionic strengths, there is also a system with no electrostatic interactions. Under physiological ionic strengths (C_{salt} ~ 0.15M), κ is 1.24 nm^{-1}. In our simulations, $\Gamma_{DH} = 0.535$ such that V_{DH} for two oppositely charged atoms located at a distance of 0.5 nm matches the native contact energy. For more details regarding the Debye-Hückel model, the authors cite several helpful publications [432-435].

6.3.6 Molecular dynamics simulations.

All molecular dynamic simulations (MD) were performed with Gromacs 4.5.5 [436]. In the SBM, reduced units were used in the potentials. Accordingly, the value of simulation temperature is not the same as the "normal" temperature. To obtain an appropriate simulation temperature, replica-exchange molecular dynamic (REMD) simulations with 24 replicas ranging from about 38 K to 150 K were performed. Each replica was performed for 100 ns. The exchanges were attempted every 5000 steps. As a result, the simulation temperature was set to 85 K. The MD used the Langevin equation with a constant friction coefficient of $\gamma = 1.0$. The cutoff for nonbonded interactions was set to 3.0 nm, and all bonds were constrained using the LINCS algorithm [437]. The MD time step was set to 2.0 fs. For thermodynamical simulations, a 1 μs MD simulation was performed on each salt concentration (including 10, 50, 100, 150, 200, 250, 300, and 500 mM, as well as a system without charge interactions). We define that a native contact is formed if the Cα-Cα distance between any given native atom pair is within 1.2x of its native distance. The native distance is calculated from the initial structure model. Therefore, Q_{inter} is defined as the fraction of the interchain native contact number.

6.3.7 Theoretical collisional cross section determination.

Theoretical CCS were determined for the C alpha models using the projection approximation (PA) and trajectory method (TJM) models in IMPACT [438], for 50,000 frames in each trajectory. Results were filtered and plotted based on the fraction of native contacts between chain 1 and chain 3 ($Q_{inter13}$) and the ion strength, where $Q_{inter13} = 0$ represents the open state, $Q_{inter13} = 1$ represents the closed state.

6.4 Results

6.4.1 Design of covalently-linked Sso PCNA for smFRET.

To understand how sliding DNA clamp conformational dynamics may influence biological processes, we sought to first characterize the dynamics of heterotrimeric PCNA from the model organism *Sulfolobus solfataricus*. This well-studied protein consists of three distinct subunits, PCNA1, PCNA2, and PCNA3, serves as a processivity factor during DNA replication, and is implicated as a key component of polymerase switching during translesion DNA synthesis [230,271,439]. Previous studies have revealed that *Sso* PCNA opens at the PCNA1:PCNA3 interface to be loaded onto DNA by RFC [416]. However, no crystal structure of PCNA in the ring-open conformation is currently available despite structures of the ring-closed conformation alone [412-414] and in complex with protein binding partners [359,414,415]. Accordingly, we wondered if PCNA maintains an equilibrium between ring-open and ring-closed conformations even in the absence of DNA or RFC. To directly probe for this conformational equilibrium, we designed an inter-subunit FRET construct wherein PCNA1 and PCNA3 were modified with site-specific Cys mutations at S64 and S189, respectively, and were simultaneously labeled with Cy3- and Cy5-maleimide (Figures 6.1A and 6.1B). Based on the crystal structure of *Sso* PCNA, the inter-fluorophore distance in the ring-closed conformation is 45 Å which corresponds to a FRET efficiency of ~0.80 (Cy3-Cy5 R_0 = ~54 Å) [440,441]. Consequently, conformational transition of *Sso* PCNA from the ring-closed to the ring-open conformation should cause a decrease in FRET efficiency (Figures 6.1A and 6.1B). Notably, while our labeling strategy can result in a heterogeneous population of labeled

species (*i.e.* Cy3 at both S64 and S189, Cy5 at both S64 and S189, and single-label species), these molecules do not demonstrate anti-correlated donor and acceptor fluorescence signals during smFRET experiments, and thus were either not observed during smFRET imaging or easily identified for exclusion during data processing. Moreover, the two possible dual-labeled PCNA molecules (*i.e.* C64-Cy3/C189-Cy5 PCNA and C64-Cy5/C189-Cy3 PCNA) from Cy3-Cy5 labeling are expected to yield equivalent FRET signals during smFRET experiments as each labeling site was carefully selected (*i.e.* based on the *Sso* PCNA heterotrimer crystal structure, PDB code: 2HII) to avoid secondary structural elements of PCNA while maintaining a high degree of flexibility and solvent accessibility for the conjugated fluorophores. Thus, we assume that the dipole orientation factors (κ^2) [241] for the dual-labeled PCNA molecules with fluorophores positioned at either site should be similar and therefore not affect apparent FRET efficiencies and relative distance changes during our smFRET experiments.

To simplify surface-immobilization for smFRET experiments, PCNA1, PCNA2, and PCNA3 were covalently fused (PCNA1-2-3) through a peptide linker (ASGAGGSEGGGSEGGTSGAT) between PCNA1 and 2 as well as PCNA2 and 3 (Figure 6.1C). This same strategy was previously applied to define the single PCNA subunit interface (PCNA1:PCNA3) for ring-opening, and this linked construct enhanced DNA polymerase processivity in primer extension assays just as wild-type heterotrimeric PCNA [416]. Thus, this covalently-linked PCNA heterotrimer offered the advantage of simplifying the interpretation of smFRET results (*i.e.* by limiting complicating signals from *in situ* PCNA heterotrimer association and dissociation) to exclusively investigate

ring-opening and ring-closing, as well as permitted direct comparison with the previous study [416]. To permit wide-field smFRET imaging, we achieved surface immobilization of *Sso* PCNA by engineering a short AviTag (15 amino acid residues) to the C-terminus of the protein to be biotinylated at a specific Lys residue by BirA [420,422]. The biotinylated PCNA was then bound by NeutrAvidin which was immobilized to the surface of a quartz microscope slide passivated with biotinylated-PEG molecules. Following surface-immobilization, fluorescence intensity from both donor and acceptor channels was simultaneously recorded during Cy3 excitation with a 532 nm laser. Initial experiments yielded single-molecule trajectories with anti-correlated donor and acceptor signals representing two, non-zero FRET efficiency values near 0.80 (high FRET) and 0.55 (low FRET), which we have attributed to the ring-closed (Figure 6.1D) and ring-open (Figure 6.1E) conformations of PCNA, respectively. PCNA was observed to interconvert between the ring-closed and ring-open conformations as demonstrated by FRET fluctuations between the high and low efficiency values (Figure 6.3) in some single-molecule trajectories. The inclusion of Trolox (2 mM) as an anti-blinking agent in the single-molecule imaging buffer, as well as the anti-correlated signal changes and fluctuations to non-zero FRET efficiency values suggests that the observed FRET transitions do not occur as a result of acceptor blinking [424]. However, many single-molecule trajectories demonstrated only high FRET (Figure 6.1D) or low FRET (Figure 6.1E) efficiencies until irreversible photobleaching, suggesting that PCNA ring-opening and ring-closing transition kinetics are slow. Accordingly, reversible fluctuations between FRET efficiency states were only occasionally observed (data not shown) before

photobleaching during the imaging period of our experiments (~120 s). These results verify the utility of our FRET construct and immobilization strategy, and support the hypothesis that *Sso* PCNA is conformationally dynamic even in the absence of DNA or protein partners.

6.4.2 Investigation of Sso PCNA by smFRET.

Previous studies have established that the interfaces between each PCNA monomer are partly stabilized by ionic interactions [412-414]. Moreover, the interacting regions of PCNA1 and PCNA3 at the PCNA1:PCNA3 interface have clear positive and negative surfaces, respectively (Figure 6.2). We thus wondered if PCNA ring-opening and ring-closing detected through our smFRET methodology (Figures 6.1 and 6.3) would be sensitive to the ionic strength of the imaging buffer, which would be observed as changes in the relative distribution of FRET efficiency values and ring-opening/ring-closing transition kinetics upon titration with increasing amounts of NaCl. Accordingly, we performed single-molecule imaging of surface-immobilized, Cy3/Cy5-labeled PCNA1-2-3 in imaging buffer containing different NaCl concentrations (0-1,000 mM). Many single-molecule trajectories (>200) at each NaCl concentration were collected and observed FRET efficiencies were binned to generate FRET population distribution histograms (Figure 6.4). Two FRET populations, corresponding to a high FRET, ring-closed PCNA conformation (E_{app} = 0.7-0.9) and a low FRET, ring-open PCNA conformation (E_{app} = 0.5-0.6) were readily apparent within the FRET distribution histograms (Figures 6.4A – E). The bimodal FRET distributions were fit to a sum of Gaussian functions in MATLAB to quantify the effect of solution ionic strength on the

open-close equilibrium of PCNA. The low- and high-FRET populations (Figures 6.4A – E, dashed red lines) showed a strong dependence on solution ionic strength as low NaCl concentrations favored the ring-closed PCNA conformation (peak center = 0.80) and high NaCl concentrations favored the ring-open PCNA conformation (peak center = 0.55). In fact, by plotting the relative fraction of low-FRET events against NaCl concentration, a clear linear correlation is apparent as increasing ionic strength presumably disfavors the ring-closed PCNA conformation but favors the ring-open PCNA conformation (Figure 6.4F). With no NaCl added to the imaging buffer (Figure 6.4A), a substantial fraction (~25%) of low-FRET events are still observed in the FRET distribution, suggesting that PCNA ring-opening can still spontaneously occur despite weak shielding of electrostatic interactions. Similarly, at high ionic strength (Figure 6.4E, 1,000 mM NaCl), PCNA ring-closing can still spontaneously occur as ~25% of molecules remain in the high-FRET, ring-closed conformation despite significant shielding of electrostatic interactions.

To further investigate the conformational dynamics of *Sso* PCNA, we measured the transition kinetics of PCNA ring-opening and ring-closing through a dwell time analysis of our smFRET data at each NaCl concentration. Briefly, we measured the duration of time each molecule spent at a particular FRET efficiency level before transitioning to a different level or photobleaching. Dwell times of PCNA in the ring-closed (*i.e.* high-FRET) and ring-open (*i.e.* low-FRET) conformations were binned into histograms which were subsequently integrated, normalized, and inverted to generate survivor functions (Figures 6.5A, 6.5B). Single exponential decay fits to the survivor functions of the high-FRET and low-FRET efficiencies yielded rate constants for PCNA

ring-opening (k_{open}) and PCNA ring-closing (k_{close}) at each tested NaCl concentration (Table 6.1). The value for k_{open} was affected by solution ionic strength. Consistent with the decrease in population of the ring-closed PCNA conformer at increasing ionic strength (Figure 6.4), k_{open} became more rapid (0.0258-0.130 s^{-1}) as the NaCl concentration was raised from 0 to 1,000 mM (Figure 6.5C and Table 6.1). In contrast, k_{close} (0.038-0.051 s^{-1}) was not significantly affected by increasing solution ionic strength, suggesting that PCNA ring-closing may occur in two distinct steps, one of which (*i.e.* the protein conformational change associated with ring-closing) is unaffected by solution ionic strength. Consistently, as the solution ionic strength increased, the equilibrium constant (K_{eq}), calculated as k_{open}/k_{close}, also increased thereby reflecting the shift in equilibrium towards the ring-open conformation of PCNA (Table 6.1). In agreement with the lack of reversible fluctuations between FRET efficiency states before photobleaching in smFRET trajectories, the transition rates measured for ring-opening and ring-closing were indeed small (Table 6.1). In fact, the time required for all molecules to complete the reaction ($7*t_{1/2} = 7*(\ln 2/k)$, $t_{1/2}$ = reaction half-life) was 75 s for PCNA opening (*i.e.* calculated based on average k_{open}) and 109 s for PCNA closing (*i.e.* calculated based on average k_{close}), while photobleaching typically occurred within 120 s during smFRET experiments. As a result, the likelihood of observing both slow ring-opening and slow ring-closing within an individual single-molecule trajectory before donor or acceptor photobleaching was minimal.

6.4.3 Ion mobility-mass spectrometry of Sso PCNA.

To further probe the conformational dynamics of *Sso* PCNA and support the conclusions of our smFRET analysis, we studied the structure and conformations of *Sso* PCNA by native mass spectrometry (MS). This unique methodology has been applied to investigate protein-protein interactions in the gas phase as well as to provide low-resolution structural information for proteins and protein complexes, without the need for complicated or potentially disruptive protein labeling [442,443]. Coupling nano-electrospray ionization (nESI), a soft ionization technique that preserves noncovalent interactions for transmission into the mass spectrometer [444-446], with ion mobility-mass spectrometry (IM-MS), we can probe protein shape [447,448] and can even observe differences in molecular conformations [449,450].

To verify that subunit fusion did not alter the conformational dynamics of covalently-linked *Sso* PCNA1-2-3 utilized in the smFRET experiments relative to wild-type *Sso* PCNA heterotrimer, we performed native MS on both constructs. For gas phase analysis, each protein was buffer exchanged into a solution consisting of increasing concentrations of 20% triethylammonium acetate (TEAA), and 80% ammonium acetate (AmAc) (100-500 mM mixed acetate salts) to increase ionic strength. We chose to use a combination of AmAc and TEAA as previous studies have shown that TEAA reduces the charge states of the protein ions and lower charge states are attributed to a more compact and native-like form of the complex [451,452]. While the wild-type PCNA heterotrimer is the predominant species at 100 and 200 mM mixed acetate salts (Figures 6.6A, 6.6B), the mass spectra revealed sample heterogeneity as peaks attributed to monomer and dimer

were apparent. The experimental mass for the wild-type PCNA heterotrimer (Figure 6.6A) was determined to be 82982 ± 41 Da, which is in good agreement with the theoretical mass (82395 Da), when adducting of salt and solvent are considered (common in native MS when attempting to keep structures native). Moreover, masses of monomer and dimer were determined to be 27675 ± 94 Da and 55393 ± 145 Da, respectively, which are consistent with the theoretical masses (PCNA1: 27536 Da, PCNA2: 27436 Da, PCNA3: 27459 Da, PCNA1+2: 54972 Da, PCNA2+3: 54895 Da, PCNA1+3: 54995 Da). The differences in theoretical masses for the monomer and dimer species are relatively small ($\Delta MW_{(|PCNA1\text{-}PCNA2|)}$ = 100 Da; $\Delta MW_{(|PCNA1\text{-}PCNA3|)}$ = 77 Da; $\Delta MW_{(|PCNA2\text{-}PCNA3|)}$ = 23 Da; $\Delta MW_{(|(PCNA1+2)\text{-}(PCNA2+3)|)}$ = 77 Da; $\Delta MW_{(|(PCNA1+2)\text{-}(PCNA1+3)|)}$ = 23 Da; $\Delta MW_{(|(PCNA1+3)\text{-}(PCNA2+3)|)}$ = 100 Da), corresponding to a theoretical $\Delta m/z$ of 3 – 14 for the 7+ charge state of the monomer and a theoretical $\Delta m/z$ of 2 – 9 for the 11+ charge state of the dimer. Accordingly, we cannot confidently assign the identities of the monomer and dimer species present in Figures 6.6A – 6.6C which were acquired using a time-of-flight mass spectrometer and have peak widths (126 and 122, respectively) greater than these theoretical $\Delta m/z$ values. Indeed, given such subtle mass differences, the reported enzyme-catalyzed removal of the N-terminal methionine residue during protein expression in *E. coli*, which occurs with varying efficiency based on protein sequence [453], further complicates assignment of the monomer and dimer species by mass. The higher observed molecular weights in comparison to the theoretical sequences are attributed to the fact that the protein ions will retain salt, solvent, and other adducts during the gentle ionization and desolvation conditions employed here, which were critical for maintaining

the native state of the wild-type PCNA heterotrimer [454]. However, previous biochemical studies of *Sso* PCNA have reported that only PCNA1 and PCNA2 form a dimer and no homodimerzation of any of the PCNA subunits occurs [411]. Hence, the dimer species in our experiments is likely PCNA1+2, while the monomer is PCNA3. The relative intensities of the monomer and dimer species increased at higher ionic strength (Figure 6.6C, 500 mM) suggesting heterotrimer dissociation as stabilizing electrostatic interactions at the subunit interfaces were presumably shielded. As 500 mM mixed acetate salts resulted in significant decomposition of the heterotrimeric PCNA to monomer or dimer constituents, the wild-type *Sso* PCNA was not analyzed at higher ionic strengths.

As depicted for the dominant charge state (14+), two populations of distinct rotationally averaged collisional cross sections (CCS) of 49 and 52 nm^2 were observed suggesting conformational heterogeneity of heterotrimeric PCNA (Figures 6.6D – 6.6F). As the concentration of mixed acetate salts increased from 100 to 500 mM, the more compact population of ions (*i.e.* possessing a smaller CCS of ~49 nm^2) began to diminish while the more extended population of ions (*i.e.* possessing a larger CCS of ~52 nm^2) began to increase (Table 6.2 and Figures 6.6D – 6.6F). These results are consistent with the smFRET data (Figure 6.4) and suggest that ions with smaller CCS represent the ring-closed conformer of PCNA (discussed in more detail in following sections), which is stabilized at low ionic strength, while ions with larger CCS represent the ring-open conformer of PCNA, which is favored at high ionic strength.

We next examined biotinylated *Sso* PCNA1-2-3 by native IM-MS to better

simulate conditions of the smFRET analysis. As expected, the experimental mass of the biotinylated PCNA1-2-3 (Figure 6.7A) is larger (89667 ± 103 Da) than wild-type PCNA heterotrimer (82982 ± 41 Da) which is consistent with the 20 amino acid linkers between PCNA1 and 2 and PCNA2 and 3, as well as the 15 amino acid AviTag conjugated to biotin (total theoretical MW: 88894 Da). The higher measured molecular weight in comparison to the theoretical is likely due to adducts (such as solvent or salt molecules) remaining on the protein during the ionization and desolvation processes. As with the wild-type PCNA heterotrimer, we systematically increased ionic strength by buffer exchanging aliquots of the biotinylated PCNA1-2-3 into increasing amounts of mixed acetate salts (100-1,000 mM) before IM-MS analysis. As expected for a covalent peptide-linked construct, the mass spectra (Figures 6.7 A – 6.7E) obtained at each concentration of mixed acetate salts revealed a homogenous sample of biotinylated PCNA1-2-3 that did not decompose to smaller subunit constituents as ionic strength was increased. The collisional cross section distributions of the two dominant charge states (14+ and 15+) of biotinlylated PCNA1-2-3 (Figure 6.7 F – 6.7O) were similar to the wild-type PCNA heterotrimer and revealed two populations of distinct CCS: a compact species at ~53 nm^2 and an extended species at ~55 nm^2 (Table 6.2). Notably, the CCS values for compact and extended biotinylated PCNA1-2-3 were noticeably larger than for wild-type PCNA heterotrimer (~49 vs. ~53 nm^2 and ~52 vs. ~55 nm^2, respectively) (Table 6.2). We attribute these differences in CCS to the 20 amino acid long linkers between PCNA1 and 2 and PCNA2 and 3, as well as the 15 amino acid long AviTag at the protein C-terminus. As with the wild-type PCNA heterotrimer, increasing ionic strength resulted in reduction

of the intensity of the compact conformational family and concomitant increase in the intensity of the extended conformational family (Figure 6.7), which is consistent with the IM-MS data collected for the wild-type PCNA heterotrimer (Figure 6.6) and indicates that, relative to the wild-type PCNA, neither subunit fusion nor biotinylation of PCNA1-2-3 significantly diminished the ability of the trimer to adopt two distinct conformations. This experimental consistency between the wild-type and covalently-linked constructs verifies the biological relevance of smFRET experiments completed with the covalently-linked PCNA heterotrimer (Figures 6.1 and 6.3). We attribute the compact conformational family (Figures 6.7F – 6.7O), at lower CCS, to PCNA1-2-3 in a ring-closed conformation which transitions to an extended ring-open conformation as ionic strength increases. The collisional cross section distributions between the 15+ and 14+ charge states are marginally different (Figure 6.7). For example, even at lower concentrations of mixed acetate salts (Figure 6.7F, 100 mM), the majority of ions at the 15+ charge state adopt the extended ring-open conformation (~55 nm^2), but the minority ring-closed population continues to diminish as ionic strength increases (Figures 6.7G – 6.7J). This is in contrast to ions of the 14+ charge state which show the majority of ions in the compact ring-closed conformation (52.81 ± 0.15 nm^2) (Figure 6.7K, 100 mM) prior to increasing ionic strength (Figures 6.7L – 6.7O). This difference with charge state is not unexpected as the more compact population would be expected, on average, to accept fewer charges (*i.e.* smaller charge carrying capacity) during the nano-electrospray ionization process than the more extended population (*i.e.* larger charge carrying capacity). Hence, we would expect to see a higher population of the compact

conformational family in the lower charge states (*i.e.* 14+) and a higher population of the extended conformational family in the higher charge states (*i.e.* 15+). Furthermore, while the overall trend observed for the conformational dynamics of biotinylated PCNA1-2-3 at increasing ionic strengths is consistent between the smFRET and IM-MS data, the distributions of PCNA1-2-3 in the ring-open and ring-closed conformations at each salt concentration were slightly different between the two methods. For instance, at concentrations greater than 500 mM NaCl (Figure 6.4) in the smFRET experiments, the majority of PCNA1-2-3 molecules adopt the ring-open conformation. However, for the 15+ and 14+ charge states of the IM-MS experiments (Figure 6.7), concentrations of mixed acetate salts as low as 100 (Figure 6.7F) or 250 mM (Figure 6.7L), respectively, are sufficient to favor the extended ring-open conformation. Accordingly, as choice of monovalent salt (*i.e.* ammonium acetate versus sodium chloride) did not appreciably influence the smFRET results (Figure 6.8), the observed conformational equilibrium of PCNA may be moderately sensitive to IM-MS processes, such as desolvation and ionization, which may positively contribute to PCNA transitions to the extended, ring-open conformation. Nevertheless, these IM-MS results demonstrate a strong correlation between ionic strength and an extended (ring-open) conformation of PCNA1-2-3, and therefore strongly support the in-solution smFRET data of favorable transitions to a low-FRET, ring-open conformation as solution ionic strength increases (Figure 6.4). As a result of the wild-type (Figure 6.6) and covalently-linked (Figure 6.7) PCNA heterotrimer demonstrating similar dynamic properties during IM-MS analysis, the conformational dynamics of PCNA1-2-3 observed by smFRET (Figure 6.4) likely emulate those of the

wild-type PCNA heterotrimer as well.

6.4.4 Molecular dynamics simulation of Sso PCNA.

A powerful tool for understanding the mechanisms of protein folding, binding, as well as conformational changes associated with a biological process [455-459] is generation of a structure-based model (SBM). Accordingly, we used the crystal structure of *Sso* PCNA along with our smFRET and IM-MS experimental findings to develop a SBM of PCNA to simulate ring- opening and ring-closing. A coarse-grained SBM consisting of Cα atoms from the wild-type *Sso* PCNA heterotrimer was generated from PDB code 2HII [414]. The native contacts between PCNA1 and PCNA3 were omitted from the starting model to allow sampling to the ring-open conformer. Using the *Debye-Hückel* model, a series of long-time (1 μs) molecular dynamic simulations were performed at different ionic strengths (10-500 mM), and the ring-open conformation of the wild-type PCNA heterotrimer was observed at higher ionic strengths. Consistent with the collisional cross section distributions of the wild-type PCNA heterotrimer (Figure 6.6), our molecular dynamic simulations suggest that the ring-closed conformation of *Sso* PCNA is more stable at lower salt concentrations based on the root-mean-square deviation (RMSD) (Figure 6.9A) of our SBM with respect to the crystal structure of PCNA in the ring-closed conformation (2HII). Furthermore, the RMSD is shown to increase as the salt concentration is raised from 10 to 500 mM, indicating conformational heterogeneity driven by increased ionic strength (Figure 6.9A). Our simulation tracked the fraction of native contacts at the PCNA1:PCNA3 interface ($Q_{inter13}$) as ionic strength was increased from 10 to 500 mM (Figure 6.9B). With *Sso* PCNA in the fully ring-open conformation,

$Q_{inter13}$ equals 0, while in the fully ring-closed conformation, $Q_{inter13}$ equals 1, indicating that all native contacts are made. Clearly, the fraction of ring-open PCNA conformers is much lower (*i.e.* low probability, low $Q_{inter13}$) at low ionic strength, but increases as ionic strength is raised (Figure 6.9B). Notably, a simulation removing all charge from the system reveals *Sso* PCNA to be in the ring-open conformation as all electrostatic contacts at the PCNA1:PCNA3 interface are eliminated (no charge, Figure 6.9B). By calculating the distance between the center of mass of residues involved in native contacts at the PCNA1:PCNA3 interface at varying ionic strengths, we observed that the distance of the PCNA1:PCNA3 interface lengthens as salt concentration increases (Figure 6.9C) and is greatest when all charges are eliminated (*i.e.* no attractive electrostatic forces) (no charge, Figure 6.9C). Our simulation also reveals that as ionic strength is increased from 10 to 500 mM, the mean number of salt bridges in the SBM decrease as interacting charges become shielded (Figure 6.9D), which provides a basis for the experimentally observed increases in k_{open} (Table 6.1) and the population of ring-open PCNA conformer as salt concentration is increased (Figures 6.4, 6.6, and 6.9). Taken together, our molecular dynamic simulation suggests that the proportion of ring-open PCNA conformer is dependent on ionic strength. Notably, most configurations of *Sso* PCNA from our simulation open out-of-plane with an unexpected left-hand geometry (Figure 6.9E).

6.5 Discussion

Replication and repair of DNA are critical for cell functioning, survival, and proliferation. Accordingly, these complex processes are executed and tightly regulated *in vivo* by a host of enzymes and protein cofactors, including the sliding DNA clamp,

PCNA. This ring-shaped molecule encircles DNA to serve as a processivity factor for replicative DNA polymerases, but has also been implicated as a protein scaffold upon which DNA replication and repair components assemble. As PCNA is often localized at the replication fork, any complications that arise during DNA replication, such as DNA damage or replication fork stalling, can be efficiently resolved by the applicable enzymes recruited to PCNA. Structures of sliding DNA clamps from all three domains of life show a ring-closed conformation, suggesting that a ring-closed→ring-open conformational change is necessary for PCNA loading onto DNA. Indeed, RFC is known to pry open the closed PCNA ring for efficient loading onto DNA at primer/template junctions. Consistently, a structure of gp45 clamp from T4 bacteriophage was solved in a ring-open conformation only when in complex with the gp62/gp44 clamp loader and DNA [416,417,460]. Similarly, a low resolution (12 Å) cryo-electron microscopy structure of archaeal PCNA in complex with RFC and DNA from *Pyrococcus furiosus* also reveals out-of-plane ring-opening [461]. However, biochemical and computational evidence suggest that sliding DNA clamps may sample ring-open and ring-closed conformations [418,462], even in the absence of their respective clamp loaders, despite the aforementioned structural data. In fact, the ring-open conformation of gp45 was shown to dominate in solution by fluorescence spectroscopy [418], while only the ring-closed structure (in the presence of clamp loader) could be solved by X-ray crystallography [419]. For this reason, it was previously hypothesized that crystallization conditions may in fact artificially select for the ring-closed conformation of sliding DNA clamps [462]. Here, using smFRET, IM-MS, and SBM molecular dynamic simulations, we show that *Sso* PCNA exists in equilibrium between

ring-open and ring-closed conformations that are sensitive to ionic strength. To our knowledge, these data are the first to demonstrate passive ring-opening and ring-closing of a sliding DNA clamp in real-time. Importantly, these dynamic conformational transitions may be significant for regulating protein-protein and protein-DNA interactions mediated by PCNA during DNA replication and repair.

The opening of sliding DNA clamps by clamp loader assemblies (*i.e.* gp62/gp44 in T4 bacteriophage, γ-complex in *E. coli*, RFC in archaea and eukaryotes) has been structurally and biochemically evaluated to reveal a largely conserved mechanism [463]. While in complex with ATP, clamp loaders first bind and then open sliding DNA clamps. Next, this complex binds to the primer-template junction of DNA which activates ATP hydrolysis by the clamp loader. ATP hydrolysis results in the closure of the sliding DNA clamp and dissociation of the clamp loader from the complex. While it is suspected that ATP binding to the clamp loader permits the forcible opening of the DNA clamp, an alternative hypothesis suggests that an ATP-stabilized clamp loader conformation preferentially associates with the ring-open conformation of the sliding DNA clamp [417]. For the first time, our smFRET analysis has provided direct evidence for dynamic interconversion of PCNA between ring-open and ring-closed conformations (Figure 6.3) which supports the latter hypothesis for RFC-catalyzed clamp opening (the ATP-stabilized clamp loader associates with the ring-open PCNA). While ensemble FRET has been previously utilized to investigate the conformations of sliding DNA clamps [369,370,418,464,465], these studies could only report on the average behavior of clamp molecules. In contrast, smFRET is not constrained by ensemble avaraging to permit

dynamic visualization of individual molecules. Here, we surface-immobilized PCNA that was simultaenously labeled with Cy3 and Cy5 at positions sensitive to ring-opening and ring-closing at the PCNA1:PCNA3 interface for wide-field smFRET imaging (Figure 6.1). Based on the observed, apparent FRET efficiencies for the high- and low-FRET states (~0.80 and 0.55, respectively) (Figures 6.3 and 6.4), as well as the Förster radius for Cy3-Cy5 (~54 Å) [440,441], the distance change between the ring-closed and ring-open PCNA conformers is roughly 10 Å. This distance is consistent with in-plane ring-opening of ~5-9 Å observed structurally for gp45 and archaeal PCNA in complex with their respective clamp loaders and DNA [417,461]. This supports the notion that single-stranded DNA is first passed through the open ring, followed by PCNA ring-closing and subsequent sliding to the double-stranded primer-template junction, rather than additional ring-opening to permit passage of the wider double-stranded DNA (~20 Å) [417,461]. Notably, while proper data corrections are necessary to measure the absolute inter-fluorophore distance of the low- and high-FRET states, the ~10 Å difference estimated above is a reasonable approximation as both the low-(0.40 – 0.60, center = 0.55) and high-(0.60 – 0.90, center = 0.80) FRET distributions fall squarely in the linear region of the distance-dependency of FRET efficiency for the Cy3-Cy5 pair (Figure 6.10). Our molecular dynamic simulations revealed that most ring-open configurations sampled by PCNA were out-of-plane with left-hand geometry (Figure 6.9E). This result is consistent with previous computational reports [462,466,467] which revealed that PCNA (including yeast and human PCNA [467]) is not biased towards right-handed spiral configurations, which is in contrast to structures of DNA clamps in complex with clamp loaders [417,461,463].

Accordingly, passive loading of PCNA may occur at the single-stranded region of primer-template junctions via left-hand helical geometry. Alternatively, PCNA may be stabilized in the right-hand helical geometry for DNA loading by its binding to RFC.

The novel results of our smFRET investigation were supported and cross-validated by native mass spectrometry of the wild-type *Sso* PCNA (Figure 6.6) and the biotinylated PCNA1-2-3 construct (Figure 6.7). IM-MS revealed that two conformational families existed for each construct with distinct collisional cross sections (Figures 6.6 and 6.7, Table 6.2). These populations are attributed to PCNA heterotrimer in either compact (ring-closed) or extended (ring-open) conformational families and are readily separated by their difference in drift time within the ion mobility cell. In order to further validate these findings we compared the experimental CCS to theoretical CCS determined from the molecular dynamics simulations performed in Figure 6.9. Our MD simulations were performed using coarse grained (C alpha only) structures and to the best of our knowledge, no CCS algorithm exists for CCS determination of coarse grained structures of this type. We therefore could not directly compare the values generated from the theoretical calculations to the experimental values, but could compare the trends observed. Theoretical CCS were determined using both the projection approximation (PA) and the trajectory method (TJM) algorithms in IMPACT and although the values vary significantly for these coarse grained models, the trends remain the same. Figure 6.11 and Table 6.3 summarize the theoretical CCS results for the wild-type PCNA. As expected, the ring-closed form is more compact than the ring-open form with the average percentage difference in CCS between the two forms being 2.1%. This is in line with,

although smaller than, our experimental observations of the wild-type PCNA heterotrimer in which an average 4.7% difference between the two conformational families was observed. It has been previously reported that ring complexes can undergo compaction on transfer into the gas-phase [468], which could in part explain the larger percentage difference between the two conformational families observed experimentally. Interestingly, although the absolute values are significantly different for PA and TJM, the percentage difference calculated between the two conformational families is very similar, 2.1% and 2.2% respectively.

Furthermore, the intensity of each conformational ensemble was shown to be sensitive to the concentration of mixed acetate salts (Figure 6.7), with higher concentrations favoring the extended conformer. The IM-MS data at different mixed acetate salt concentrations agreed with the observed influence of increasing NaCl concentrations on the proportions of FRET events at high and low efficiencies (Figure 6.4) and our molecular dynamic simulations (Figure 6.9), with low ionic strength favoring the closed PCNA conformation and high ionic strength favoring the ring-open PCNA conformation. Moreover, dwell time analysis (Figure 6.5) of the high- and low-FRET events yielded transitions rates for ring-closed→ring-open and ring-open→ring-closed conformational changes, respectively (Table 6.1). As the transition kinetics (Figure 6.5 and Table 6.1) from the ring-closed conformer ($PCNA^{C}$, Scheme 6.1) to the ring-open conformer ($PCNA^{O}$, Scheme 6.1) and occupancy of the low-FRET state (Figure 6.4F) showed an increasing linear dependence on ionic strength, but the reverse transition (*i.e.* $PCNA^{O}$→$PCNA^{C}$) was not sensitive to increasing ionic strength, we

expanded the simple one-step mechanism for PCNA ring-opening/ring-closing (Scheme 6.1A) to include an additional step (Scheme 6.1B). Supported by our molecular dynamic simulations (Figures 6.9B, C, D), as native electrostatic interactions at the PCNA1:PCNA3 interface become sufficiently shielded by electrolyte ions, $PCNA^{C}$ is first converted to $PCNA^{Z}$ which maintains the closed ring conformation but lacks all associated stabilizing contacts. Conversion to $PCNA^{Z}$ (k_1) becomes more rapid as ionic strength increases as illustrated by the slope of 1.0×10^{-4} $mM^{-1}s^{-1}$ for the best fit line to the data in Figure 6.5C which quantitatively describes how the ring-closed→ring-open transition rate scales with increasing NaCl concentrations. Following $PCNA^{Z}$ formation, conformational change to the ring-open state ($PCNA^{O}$, Scheme1B) occurs rapidly (k_2) and is independent of salt concentrations. Thus, k_{open} in Table 6.1 is limited by k_1. During the reverse transition from $PCNA^{O}$ to $PCNA^{C}$, k_{-2} (0.038-0.051 s^{-1}, Table 6.1) describes the rate of the reverse conformational change to the closed ring conformation wherein native electrostatic contacts have not yet re-formed ($PCNA^{Z}$), and is independent of salt concentrations. Moreover, k_{-1} describes the reinstatement of native electrostatic contacts at the PCNA1:PCNA3 interface which presumably does not lead to a measurable FRET change and was not determined in our studies. Accordingly, k_{close} in Table 6.1 equals k_{-2}. Notably, the remaining interfaces of the PCNA heterotrimer (*i.e.* PCNA1:PCNA2 and PCNA2:PCNA3) have been shown to be much more charge-charge compatible, as well as occupy more buried surface area, than the PCNA1:PCNA3 interface, suggesting that disruption through electrostatic shielding of native interactions at these interfaces would occur at higher ionic strengths than for the PCNA1:PCNA3 interface, as supported by our

native mass spectra of the wild-type PCNA heterotrimer (Figure 6.6) which shows dissociation of the trimer at higher ionic strengths [413]. The observed instability of the wild-type PCNA heterotrimer as buffer ionic strength is increased beyond the physiological salt concentrations which are ≤500 mM [469,470] may suggest a mechanism for PCNA subunit dissociation. Indeed, it may be that disassembly of the PCNA heterotrimer favorably occurs from the ring-open state as a result of the loss in electrostatic contacts between PCNA1 and PCNA3, which agrees with the positive correlation between the ring-open state of the PCNA heterotrimer (Figures 6.6 D – 6.6F) and observed subunit dissociation (Figures 6.6 A – 6.6C) at higher buffer ionic strengths. Notably, the low FRET state assigned as the ring-open conformer of covalently fused PCNA during smFRET experiments (Figures 6.1, 6.3, and 6.4) cannot be attributed to the observed subunit dissociation of the PCNA heterotrimer at higher ionic strengths as such a dissociation event would most likely result in FRET efficiencies of less than 0.2 as a result of the 20 amino acid linkers between PCNA1 and PCNA2, as well as PCNA2 and PCNA3. The length of the fully-extended linker is approximately 70 Å (*i.e.* assuming 3.5 Å per amino acid) which would result in a FRET efficiency of ≤ 0.17 (assuming Cy3-Cy5 $R_0 \approx 54$ Å and not including the length of a typical PCNA subunit which would result in an even lower expected FRET value, near zero) upon PCNA subunit dissociation. As a low FRET efficiency threshold of 0.2 was applied during smFRET data analysis (see Section 6.3.3), such near-zero FRET efficiencies, occurring from PCNA subunit dissociation at high ionic strengths, would have been removed from the overall analysis.

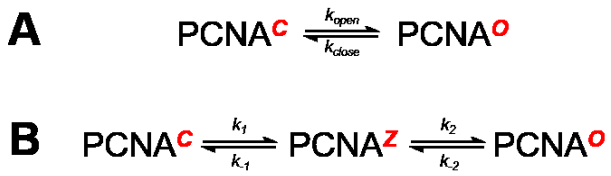
The observed conformational dynamics of PCNA are supported by structural data

collected for *Sso* monomeric, dimeric, and trimeric PCNA [412-414]. Although the angle between each monomer in the crystal structure of the heterotrimer was determined to be 120°, the angle between PCNA1 and PCNA2 in the crystal structure of the heterodimer was determined to be 130°. Accordingly, during formation of the *Sso* heterotrimer (*i.e.* binding of PCNA3 to the PCNA1/PCNA2 heterodimer), a significant spring tension is introduced as PCNA1 and PCNA2 are forced to bend ~10° to accommodate PCNA3. It is thus hypothesized that PCNA ring-opening at the PCNA1:PCNA3 interface occurs to relieve the spring tension induced by PCNA3 [413]. Indeed, a structural model of PCNA3 bound to the PCNA1/PCNA2 heterodimer reveals a ~7 Å gap between PCNA1 and PCNA3 which is consistent with the ~10 Å distance change roughly estimated through our smFRET experiments [413]. As our data reveal that PCNA can open and close even in the absence of RFC, the role of the clamp loader becomes less defined. Consistently, DNA polymerase processivity assays revealed that while *Sso* PCNA was necessary to synthesize full-length DNA product, *Sso* RFC was dispensable, suggesting that *Sso* PCNA is capable of loading to a primed M13 template in the absence of RFC [460]. However, the fastest rate of DNA synthesis and greatest amount of DNA product formation was achieved in the presence of both PCNA and RFC [460]. Accordingly, it seems that RFC may bind to the ring-open conformation of PCNA and guide the sliding DNA clamp to primer-template junctions of DNA substrates. Moreover, *Sso* RFC may be required to accelerate the rate of *Sso* PCNA ring-opening as the passive rate of PCNA ring-opening (k_{open}), even at the highest NaCl concentration, was only 0.13 s^{-1} (Table 6.1) which is considerably slower than the estimated association rate of the binary complex of

RFC/PCNA to DNA (100 s^{-1}) [460]. At a more physiologically relevant salt concentration (250 mM), k_{open} was determined to be even slower (0.0374 s^{-1}, Table 6.1) and the majority of PCNA molecules were observed in the ring-closed conformation (Figure 6.4B), further underlining the need for RFC to accelerate and stabilize the ring-open conformation of *Sso* PCNA for loading onto DNA. As the rate of replication fork progression in a hyperthermophilic organism such as *Sulfolobus solfataricus* is likely very fast, it is crucial that *Sso* PCNA be quickly and efficiently directed to priming sites for Okazaki fragment synthesis during lagging strand replication. Furthermore, MS data demonstrating that the wild-type *Sso* PCNA heterotrimer significantly dissociates into monomer and dimer constituents when the ionic strength is increased to a 500 mM mixed acetate salt threshold (Figure 6.6) underlines the need for RFC to stabilize the PCNA heterotrimer for clamp loading.

Collectively, our interdisciplinary approach has revealed a model PCNA to be a conformationally dynamic protein. The trends in protein conformational dynamics observed in solution during our smFRET experiments are conserved in the gas phase measurements of the solution structures during our IM-MS experiments. Thus, our work further validates native mass spectrometry as an effective tool to study the dynamic motions of proteins. We are eager to apply this methodology to investigate how the observed dynamics exhibited by PCNA may contribute to clamp loading onto DNA as well as key protein-protein and protein-DNA interactions mediated by PCNA during DNA replication, DNA damage response, and DNA repair.

6.6 Schemes



Scheme 6.1: Mechanisms for PCNA opening and closing.

(A) Simple mechanism of PCNA ring-opening and ring-closing. $PCNA^{C}$ denotes the ring-closed conformation of PCNA while $PCNA^{O}$ denotes the ring-open conformation. (B) Expanded mechanism of PCNA ring-opening and ring-closing which shows an intermediate conformation $PCNA^{Z}$, which represents a ring-closed form of PCNA lacking native interactions at the PCNA1:PCNA3 interface.

6.7 Tables

Table 6.1: Kinetic rates of PCNA ring-opening and ring-closing

NaCl (mM)	k_{open} (s^{-1})[a]	k_{close} (s^{-1})[b]	K_{eq} (k_{open}/k_{close})
0	0.0258 ± 0.0004	0.044 ± 0.004	0.59
250	0.0374 ± 0.0007	0.050 ± 0.002	0.75
500	0.0476 ± 0.0008	0.0405 ± 0.0009	1.18
750	0.081 ± 0.002	0.051 ± 0.002	1.59
1,000	0.130 ± 0.002	0.038 ± 0.002	3.42

[a]Dwell time survivor functions to calculate k_{open} (Figure 6.5A) were generated from smFRET trajectories exhibiting ring-closed→ring-open FRET transitions at 0 (n = 39), 250 (n = 50), 500 (n = 84), 750 (n = 80), and 1000 (n = 64) mM NaCl

[b]Dwell time survivor functions to calculate k_{close} (Figure 6.5B) were generated from smFRET trajectories exhibiting ring-open→ring-closed FRET transitions at 0 (n = 20), 250 (n = 23), 500 (n = 77), 750 (n = 68), and 1000 (n = 40) mM NaCl

Table 6.2: Collisional cross sections (CCS) in nm^2 for PCNA

14+ charge state wild-type PCNA heterotrimer					
ionic strength[a,b]	**100 mM**	**200 mM**	**500 mM**		
ring-closed PCNA	49.7 ± 0.31	49.76 ± 0.46	49.67 ± 0.43		
ring-open PCNA	51.86 ± 0.44	52.32 ± 1.12	51.96 ± 0.45		
14+ charge state PCNA1-2-3					
ionic strength[a,b]	**100 mM**	**250 mM**	**500 mM**	**750 mM**	**1000 mM**
ring-closed PCNA	52.81 ± 0.15	53.56 ± 0.91	52.77 ± 0.15	52.70 ± 0.39	52.86 ± 0.45
ring-open PCNA	54.45 ± 0.27	54.59 ± 0.25	54.43 ± 0.09	54.44 ± 0.42	54.50 ± 0.45
15+ charge state PCNA1-2-3					
ionic strength[a,b]	**100 mM**	**250 mM**	**500 mM**	**750 mM**	**1000 mM**
ring-closed PCNA	52.90 ± 0.23	52.81 ± 0.23	53 ± 0.02	53.20 ± 0.38	52.87 ± 0.38
ring-open PCNA	54.53 ± 0.26	54.55 ± 0.22	54.59 ± 0.10	54.54 ± 0.034	54.66 ± 0.33

[a]Ionic strength results from 80% ammonium acetate, 20% triethylammonium acetate mixture
[b]Values reported are the average of three different day repeats and the error represents the standard deviation between the three repeats.

Table 6.3: Average theoretical CCS determined in 100, 200, and 500 mM salt using both the projection approximation (PA) and trajectory method (TJM) algorithms.

	Ionic Strength		
	100 mM	**200 mM**	**500 mM**
PA			
ring-closed PCNA	39.72 ± 0.22 nm^2	39.73 ± 0.22 nm^2	39.80 ± 0.24 nm^2
ring-open PCNA	40.51 ± 0.37 nm^2	40.55 ± 0.40 nm^2	40.63 ± 0.43 nm^2
TJM			
ring-closed PCNA	51.19 ± 0.30 nm^2	51.21 ± 0.31 nm^2	51.26 ± 0.29 nm^2
ring-open PCNA	52.26 ± 0.51 nm^2	52.32 ± 0.55 nm^2	52.43 ± 0.59 nm^2

6.8 Figures

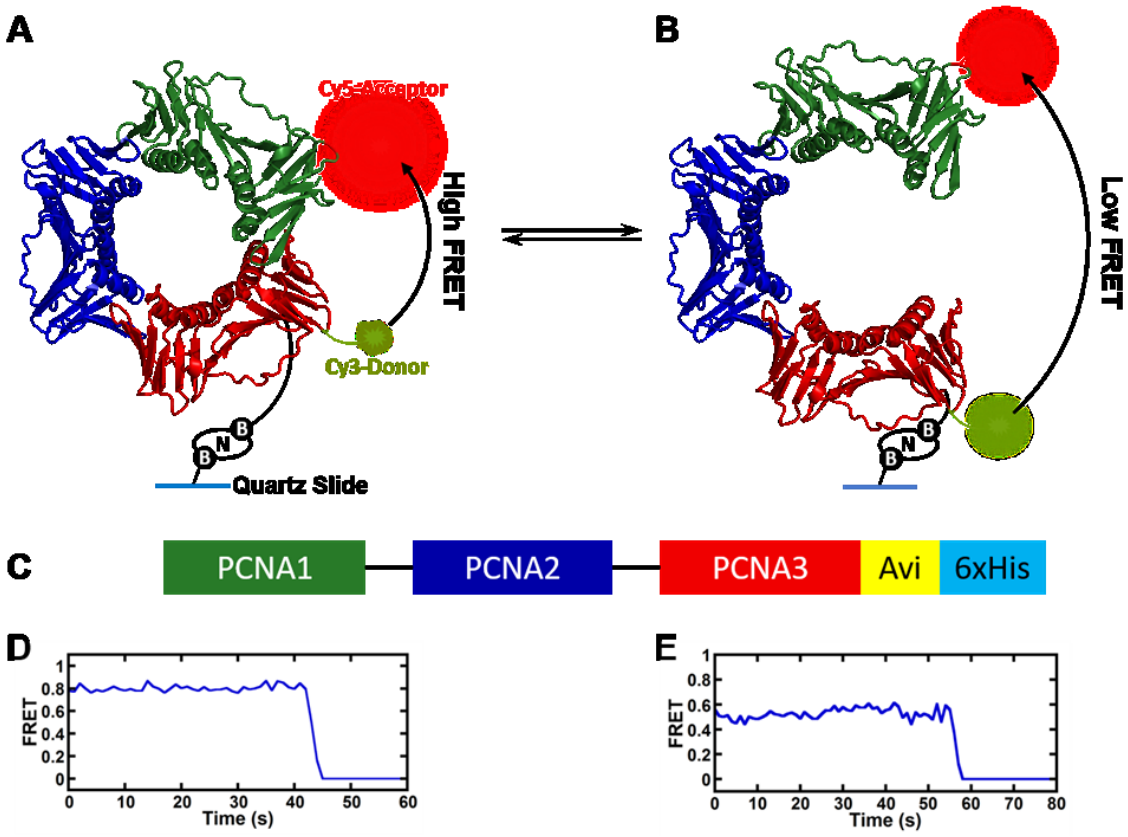


Figure 6.1: Single-molecule FRET system for Sso PCNA conformational dynamics. (A, B) Cy3 donor and Cy5 acceptor fluorescent dyes were introduced at the PCNA1:3 interface to probe conformational changes associated with PCNA ring-opening and ring-closing. (C) PCNA1-2-3 is covalently-linked at the PCNA1:2, and PCNA2:3 interface. (D) Ring-closed PCNA1-2-3 molecules produce high-FRET efficiency signal from 0.60-0.90 while (E) ring-open PCNA1-2-3 molecules produce low-FRET efficiency signal from 0.40-0.60.

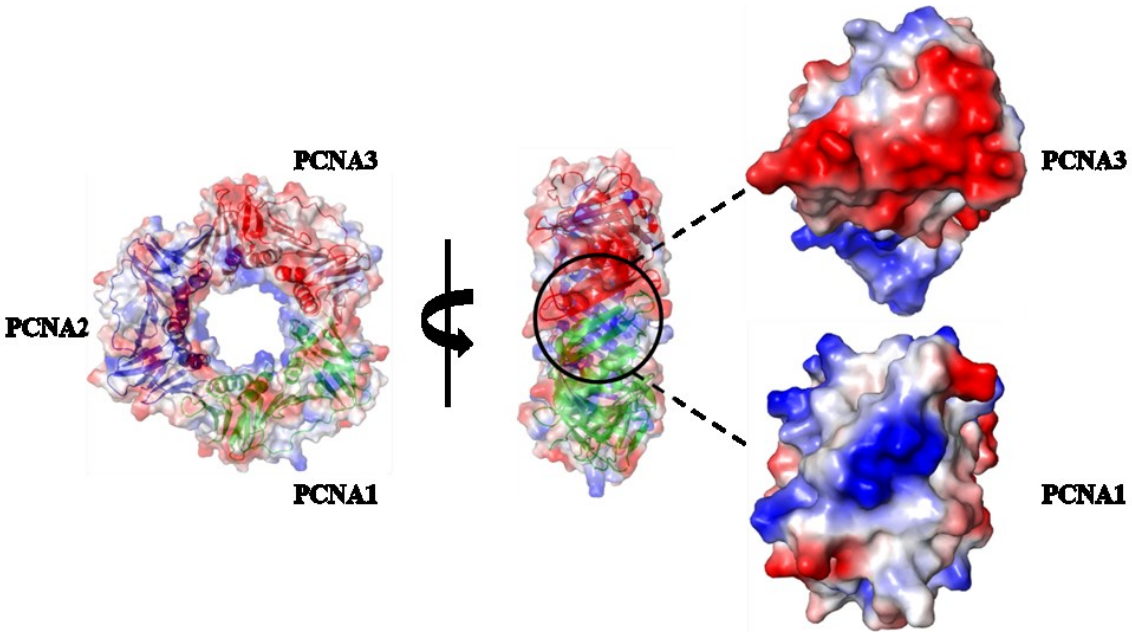


Figure 6.2: Electrostatic potential of the PCNA1:PCNA3 interface.

Structure of heterotrimeric PCNA from *Sulfolobus solfataricus* (PDB code: 2HII) with PCNA1, PCNA2, and PCNA3 colored green, blue, and red, respectively. Vacuum electrostatic protein contact potential for PCNA was generated in PyMOL with darker red colors indicating more negative potential and darker blue colors indicating more positive potential. Individual inspection of PCNA1 and PCNA3 interfaces reveals clear electrostatic compatibility, with PCNA1 possessing an overall positive potential and PCNA3 possessing an overall negative potential.

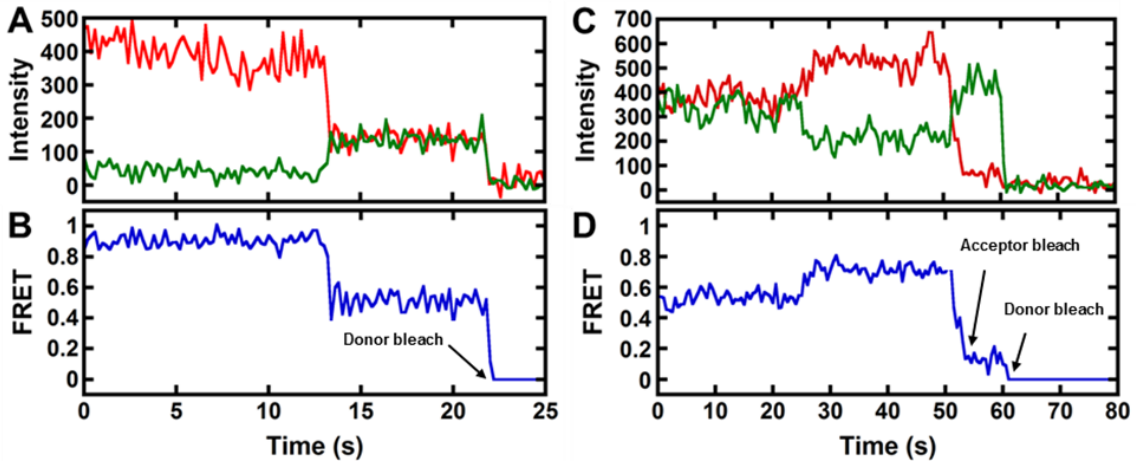


Figure 6.3: PCNA interconverts between its ring-open and ring-closed conformations.

(A) Representative donor (green) and acceptor (red) fluorescence trajectories depicting transition from the PCNA ring-closed (high-FRET) to ring-open (low-FRET) conformation before donor photobleaching. (B) FRET efficiency trajectory (blue) calculated from the donor and acceptor intensities in (A). (C) Representative donor (green) and acceptor (red) fluorescence trajectories depicting transition from the PCNA ring-open (low-FRET) to ring-closed (high-FRET) conformation before acceptor photobleaching. (D) FRET efficiency trajectory (blue) calculated from the donor and acceptor intensities in (C).

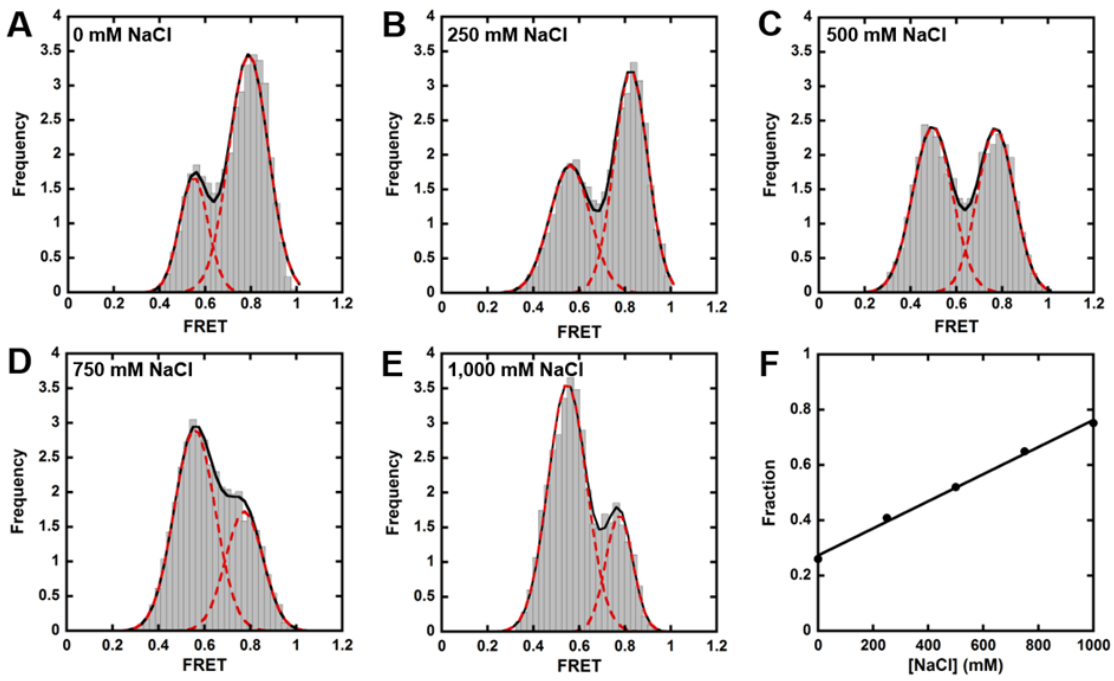


Figure 6.4: FRET distributions change with increasing NaCl concentrations. FRET trajectories for >200 molecules were compiled and binned to generate FRET distribution histograms for each NaCl concentration, (A) 0 mM, (B) 250 mM, (C) 500 mM, (D) 750 mM, and (E) 1,000 mM NaCl. FRET distribution histograms were fit to a sum of Gaussians function (black line) to extract the individual low-FRET (0.40-0.60) and high-FRET (0.60-0.90) populations (red dashed lines). (F) Fraction of PCNA1-2-3 molecules in the ring-open conformation (calculated in A-E) increased linearly as NaCl concentration was increased.

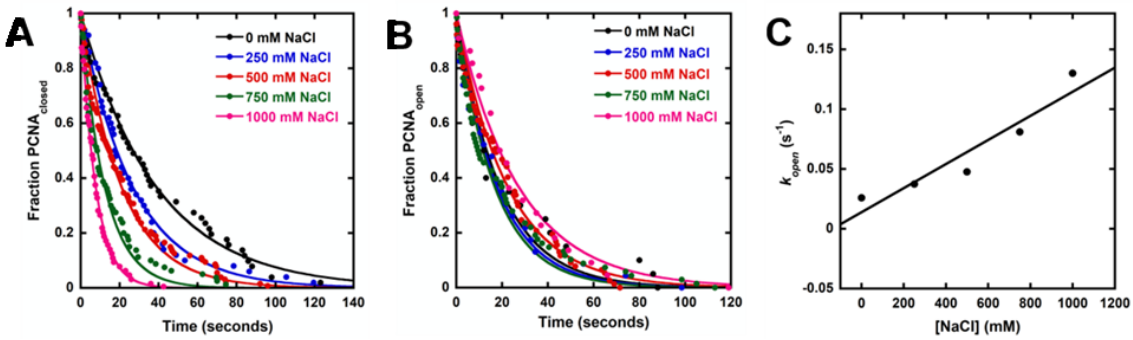


Figure 6.5: Dwell time analysis of PCNA states.

Rates of PCNA (A) ring-opening (k_{open}) and (B) ring-closing (k_{close}) at each NaCl concentration were extracted by fitting the dwell time survivor functions of the high-FRET and low-FRET states, respectively, to a single exponential decay equation. (C) Rates of PCNA ring-opening (k_{open}) were plotted against NaCl concentrations. A linear fit with slope of 1.0 x 10^{-4} $mM^{-1}s^{-1}$ is shown (black line) to describe the dependence of k_{open} values on NaCl concentrations.

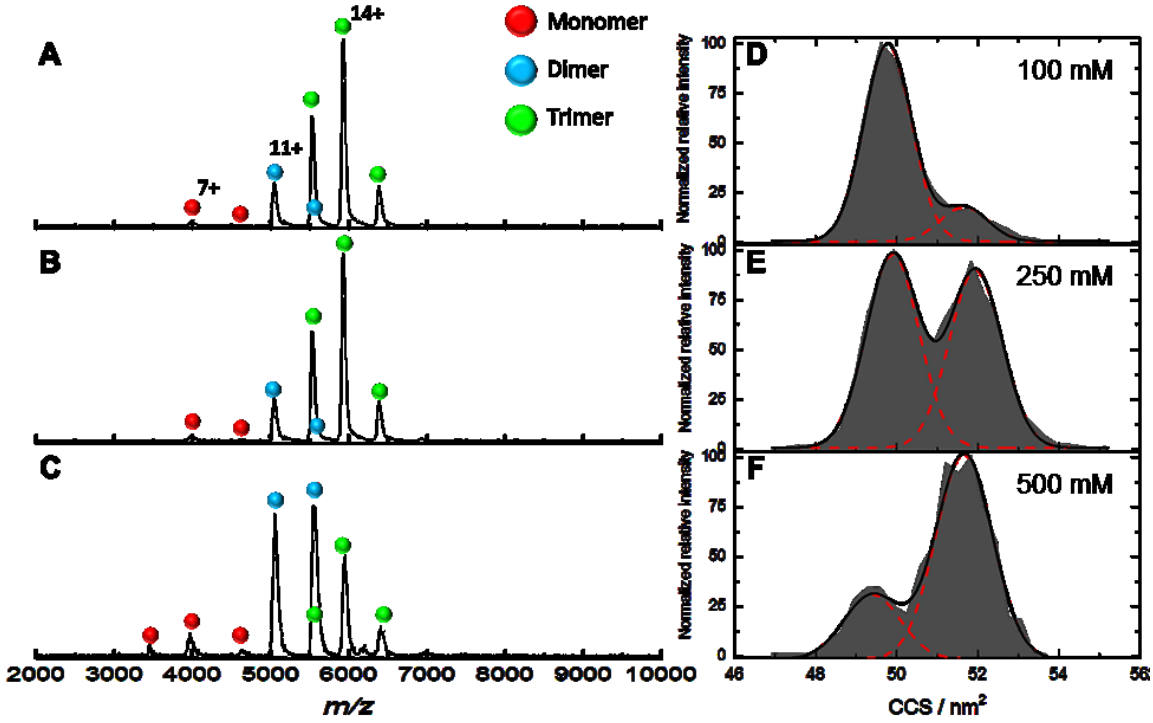


Figure 6.6: Ion mobility-mass spectrometry of wild-type *Sso* PCNA.

Wild-type PCNA trimer was analyzed by IM-MS. Samples were exchanged into buffers containing (A) 100, (B) 200, and (C) 500 mM of 20% triethylammonium acetate (TEAA), and 80% ammonium acetate (AmAc) before introduction to the mass spectrometer by nano-electrospray ionization. Collisional cross section (CCS) distributions for the 14+ charge state of the trimer were obtained from the arrival time distributions using the previously published calibration procedure.[426] In (D) 100, (E) 200, and (F) 500 mM mixed acetate salt, the 14+ trimer showed two distributions. The lower CCS distribution is assigned to a ring-closed PCNA trimer, while the higher CCS

(continued)

Figure 6.6: continued

distribution is expected to be a ring-open form of the PCNA trimer. We observe a transition from the ring-closed to the ring-open PCNA conformation as we increase ionic strength.

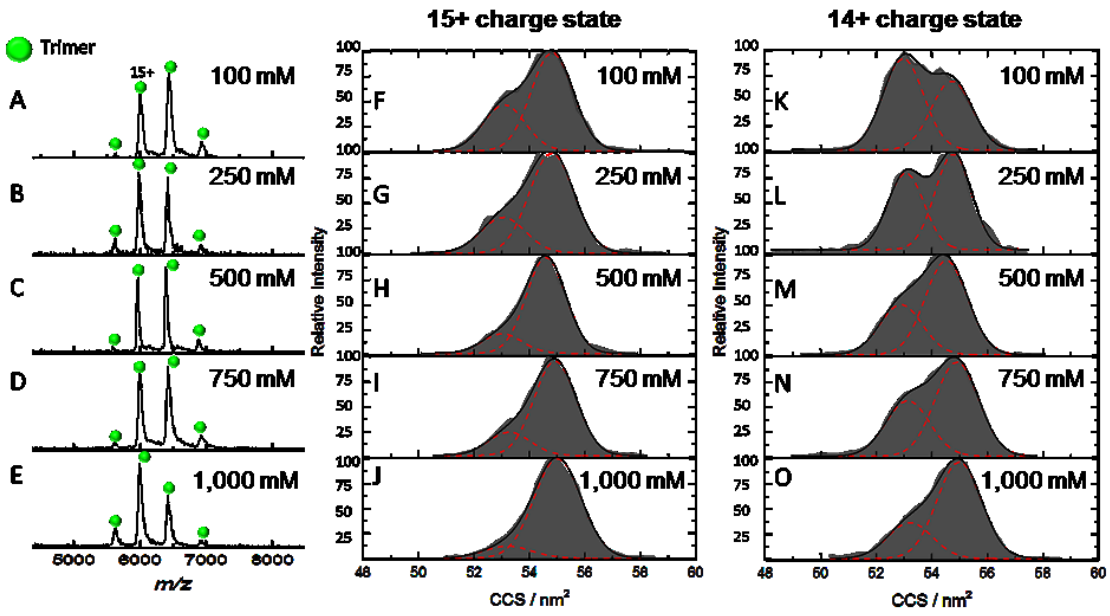


Figure 6.7: Mass spectrometry of covalently-linked *Sso* PCNA.

Biotinylated, covalently-linked PCNA trimer (PCNA1-2-3) was analyzed by mass spectrometry. A mass spectrum is shown for each concentration of 20% triethylammonium acetate (TEAA), and 80% ammonium acetate (AmAc), (A) 100, (B) 250, (C) 500, (D) 750, and (E) 1,000 mM. At all ionic strengths, the dominant species are the 15+ and 14+ charge states of the covalently-linked PCNA trimer. Arrival time distributions were converted to collisional cross section (CCS) distributions, using a previously published calibration procedure [426]. CCS distributions of the 15+ charge state are shown for increasing concentrations of mixed acetate salt, (F) 100, (G) 250, (H) 500, (I) 750, and (J) 1,000 mM. CCS distributions of the 14+ charge state are shown for increasing concentrations of mixed acetate salt, (K) 100, (L) 250, (M) 500, (N) 750, and (O) 1,000 mM.

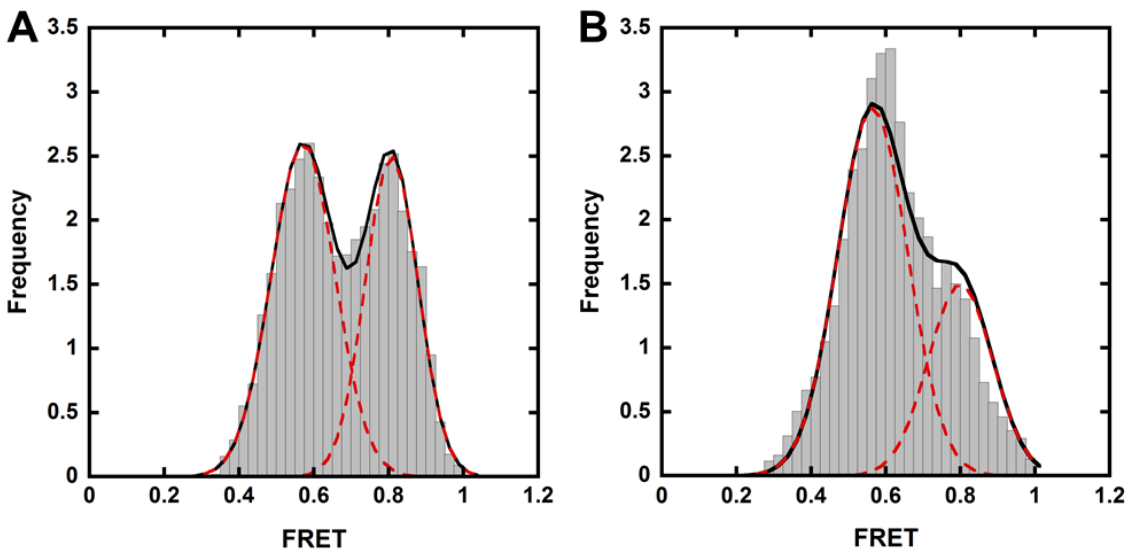


Figure 6.8: Single-molecule FRET experiments using ammonium acetate. smFRET experiments similar to Figure 6.4 were conducted using ammonium acetate, rather than NaCl, to modulate ionic strength. FRET trajectories for >200 molecules were compiled and binned to generate FRET distribution histograms at each ammonium acetate concentration, (A) 500 mM, (B) 1,000 mM. FRET distribution histograms were fit to a sum of Gaussians function (black line) to extract the individual low-FRET (0.40-0.60) and high-FRET (0.60-1.0) populations (red dashed lines).

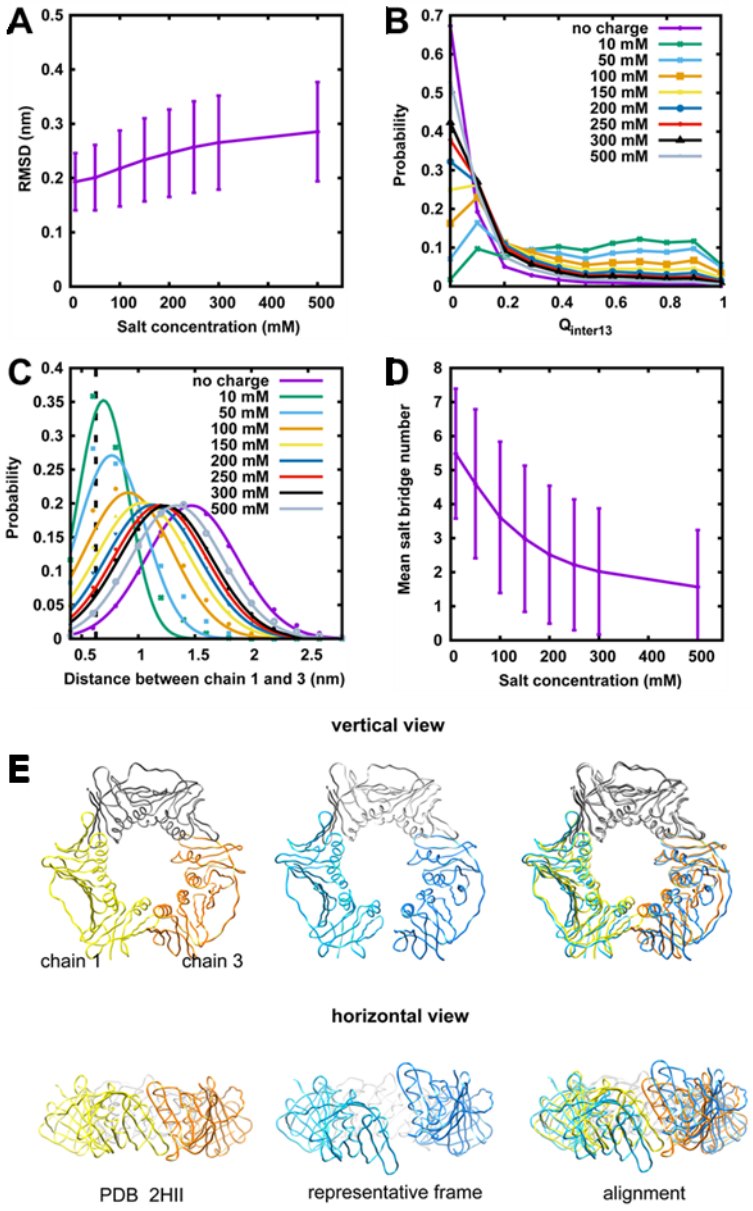


Figure 6.9: Molecular dynamic simulation of the wild-type *Sso* PCNA.

(A) Average RMSD (with standard deviation) of heterotrimeric PCNA at different salt concentrations, with respect to the structure 2HII. (B) Distribution of $Q_{inter13}$ of

(continued)

Figure 6.9: continued

heterotrimeric PCNA at different salt concentrations, as well as heterotrimeric PCNA without charges. (C) Distribution of distance between PCNA1 and PCNA3 at different salt concentrations, as well as without charges. This distance is 0.625 in 2HII (shown with black dashed line). (D) Mean salt bridge number (with standard deviation) between PCNA1 and PCNA3 at different salt concentrations. (E) Comparison between ring-open (middle, blue) and ring-closed (left, yellow) states of heterotrimeric PCNA. According to our simulation results, ring-open PCNA mainly adopts out-of-plane, left-hand spiral geometry.

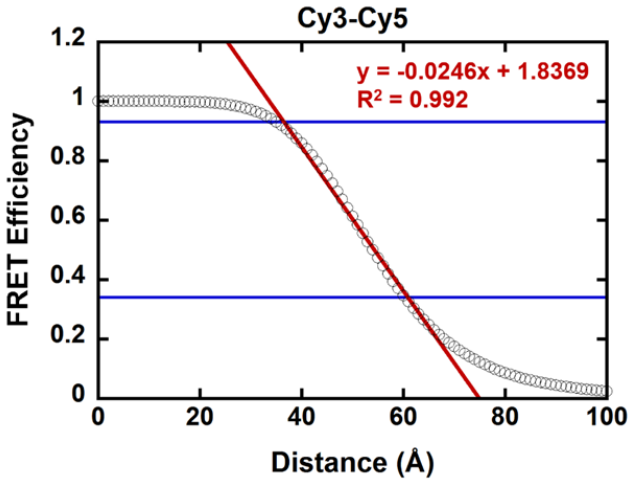


Figure 6.10: Distance-dependency of FRET efficiency for the Cy3-Cy5 FRET pair. Cy3-Cy5 FRET efficiency (black circles) was calculated with respect to distance using the equation $E = 1/((1+(r/R_0)^6)$, where E is FRET efficiency, r is inter-fluorophore distance (Å), and R_0 is the Cy3-Cy5 Förster radius (54 Å). The blue lines set the boundaries for the linear portion of the plot (0.34 – 0.93). The red line depicts the best fit line with a slope of -0.0246 FRET/Å to the linear region of the plot (*i.e.* 0.34 – 0.93). As the apparent FRET efficiencies of the low- (0.40 – 0.60, center = 0.55) and high- (0.60 – 0.90, center = 0.80) FRET states in Figure 6.4 fall squarely within the linear range, it is reasonable to estimate that the observed difference in FRET efficiency between the low- and high-FRET states (0.80 – 0.55 = 0.25) corresponds to a distance of ~10 Å ($FRET = -0.0246\ \frac{FRET}{\text{Å}} * X\text{Å};\ 0.25 = -0.0246\ \frac{FRET}{\text{Å}} * X\text{Å}$).

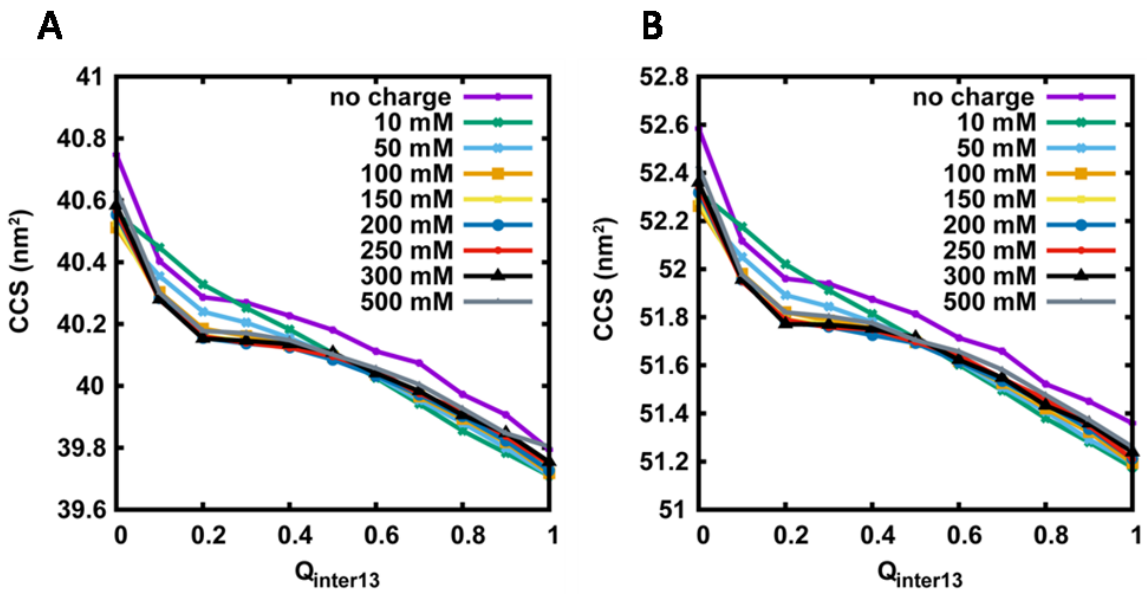


Figure 6.11: Theoretical CCS determined from MD simulations for the wild-type PCNA.

CCS values were determined for 50,000 frames in each trajectory, and filtered according to the fraction of native contacts between chain 1 and chain 3 ($Q_{inter13}$) and the ion strength, where $Q_{inter13} = 0$ represents the open state, $Q_{inter13} = 1$ represents the closed state. A) Determined using the projection approximation algorithm in IMPACT; and B) determined using the trajectory method in IMPACT.

Chapter 7. Functional Insights Revealed by the Kinetic Mechanism of CRISPR/Cas9

Author Contributions

Austin T. Raper and Anthony A. Stephenson equally contributed to this work as co-first authors. A.T.R conceived the research, prepared protein samples, designed all kinetic experiments, conducted kinetic experiments, designed all FRET experiments, performed all FRET experiments, developed the mechanism, generated figures, and wrote the manuscript. A.A.S prepared the RNA/DNA substrates, prepared protein samples, helped to design kinetic assays, conducted kinetic experiments, helped to develop the mechanism, generated figures, and helped to write/edit the manuscript. Dr. Zucai Suo provided insight into kinetic experiments, helped to develop the mechanism, and helped to edit the manuscript.

7.1 Abstract

The discovery of prokaryotic adaptive immunity prompted widespread use of the RNA-guided clustered regularly interspaced short palindromic repeat (CRISPR)-associated (Cas) endonuclease Cas9 for genetic engineering. However, its kinetic mechanism remains undefined and details of DNA cleavage poorly characterized. Here, we establish a kinetic mechanism of *Streptococcus pyogenes* Cas9 from guide-RNA binding through DNA cleavage and product release. Association of DNA to the binary complex of Cas9 and guide-RNA is rate-limiting during the first catalytic turnover, while DNA cleavage from a pre-formed ternary complex of Cas9, guide-RNA, and DNA is rapid. Moreover, an extremely slow release of DNA products essentially restricts Cas9 to be a single-turnover enzyme. By simultaneously measuring the contributions of the HNH and RuvC nuclease activities of Cas9 to DNA cleavage, we also uncovered the kinetic basis by which HNH conformationally regulates the RuvC cleavage activity. Together, our results provide crucial kinetic and functional details regarding Cas9 which will inform gene-editing experiments, guide future research to understand off-target DNA cleavage by Cas9, and aid in the continued development of Cas9 as a biotechnological tool.

7.2 Introduction

The clustered regularly interspaced short palindromic repeat (CRISPR) associated protein (Cas) Cas9 of *Streptococcus pyogenes* (*Sp*) utilizes a dual-RNA guide, or single-guide RNA fusion (sgRNA), to target and cleave DNA[279,280,284,471,472] and has been repurposed for genetic engineering.[203,473] Prerequisites for Cas9-mediated DNA cleavage include homology between the sgRNA and target DNA strand (tDNA), as well as recognition of a protospacer adjacent motif (PAM) on the non-target DNA strand (ntDNA) by Cas9 (Figure 7.1).[279,304,471] Cleavage of tDNA and ntDNA by HNH and RuvC nuclease domains of Cas9, respectively, is thought to occur 3 nucleotides away from the PAM (Figure 7.1).[279,471] Collectively, the current understanding of RNA-guided DNA targeting and cleavage by Cas9 involves 1) sgRNA binding to elicit an active Cas9 conformation,[322,323] 2) PAM-recognition,[304] 3) local DNA duplex unwinding and RNA strand invasion, 4) complete directional unwinding of the DNA to form the RNA-DNA heteroduplex,[327] and 5) coupled conformational changes within Cas9 necessary for subsequent DNA cleavage.[324,471]

Cas9 is a relatively large protein (~158 kDa) with a bilobed architecture consisting of α-helical recognition (REC) and nuclease (NUC) lobes. The NUC lobe contains the HNH and RuvC domains as well as a C-terminal domain (CTD) which interacts with the PAM.[471] Importantly, crystal structures demonstrate that apo-Cas9[323] maintains a conformation that is incompatible with sgRNA binding[322] and must therefore restructure to accommodate sgRNA. However, kinetic details of this structural rearrangement are unclear. Furthermore, close inspection of partial ternary complexes

(*i.e.* Cas9 bound to sgRNA along with single-stranded tDNA or PAM-containing partial duplex DNA)[317,318] or complete ternary complexes (*i.e.* fully-duplexed, on-target DNA, Cas9•sgRNA•DNA)[327] reveal dynamic motions that are suggested to conformationally control DNA cleavage[324] by first activating the HNH domain followed by the RuvC domain. Nevertheless, the mechanistic basis for the order of HNH and RuvC domain activation as well as how conformational motions kinetically regulate DNA strand cleavage remain unknown. Moreover, the individual kinetics of HNH- and RuvC-catalyzed DNA strand cleavage have not been quantitatively measured and may be directly related to Cas9 conformational dynamics.

Detailed kinetic analyses of Cas9-mediated catalytic events are lacking and the few previously reported kinetic parameters[279,304,305,324,471] are either misinterpreted or have less significance outside the context of a complete kinetic mechanism. To fill these gaps, we used radioisotope- and Förster resonance energy transfer (FRET)-based pre-steady-state kinetic techniques as well as single-molecule FRET (smFRET) to characterize each major step of the CRISPR/Cas9 mechanism (Figures 7.2 and 7.3), including catalytic events following initial DNA strand scission.

7.3 Results and Discussion

7.3.1 Kinetics and dynamics of sgRNA binding by Cas9.

During sgRNA (Figure 7.1) binding, Cas9 undergoes major structural rearrangements including dramatic movement of the REC lobe toward the HNH domain (Figure 7.4A).[322-324] To examine the kinetics of this first mechanistic step, we implemented a FRET system ($Cas9^{FRET\text{-}R}$, Figure 7.4A) known to be sensitive to distance

changes associated with Cas9 binding to sgRNA (Figure 7.5A) without affecting catalytic activity (Figure 7.5B).[324] Time-dependent Cy3 and Cy5 fluorescence emissions were recorded following rapid mixing of $Cas9^{FRET\text{-}R}$ (25 nM) with varying concentrations of sgRNA (75 – 1,200 nM) in the presence of $MgCl_2$ (6 mM) while exciting at 532 nm (Figure 7.4B). Strikingly, the resulting time-dependent changes in donor and acceptor signals were monophasic with rates (average: 6.1 s^{-1}) independent of sgRNA concentration (Figure 7.4C), suggesting an induced-fit, two-step binding model in which a rapidly formed initial collision complex ($Cas9^{apo}$•sgRNA, Figures 7.2A and 7.3A) induced a slower conformational change to the binary complex (Cas9•sgRNA, Figures 7.2A and 7.3A). Thus, the observed time-dependent signal changes as in Figure 7.4B reflected the relatively slow ($k_{obs} \approx 6.1\ s^{-1}$) conformational change of $Cas9^{FRET\text{-}R}$ (k_2, Figure 7.2A) following rapid binding (k_1, Figure 7.2A). We next measured the rate of the reverse conformational change (k_{-2}, Figure 7.2A) by monitoring acceptor fluorescence upon mixing the pre-formed binary complex ($Cas9^{FRET\text{-}R}$•sgRNA, 50 nM) with a large excess (2,250 nM) of unlabeled Cas9 (Figure 7.4D). Any sgRNA to dissociate from $Cas9^{FRET\text{-}R}$ following the reverse conformational change was rapidly bound by the large excess of unlabeled Cas9 instead of rebinding to $Cas9^{FRET\text{-}R}$ due to dilution. The resulting increase in acceptor fluorescence as the FRET probes returned to their closer positions in the apo state of $Cas9^{FRET\text{-}R}$ (Figure 7.4A) was monophasic with a rate of $5.5 \times 10^{-4}\ s^{-1}$. Importantly, as the exponential phase was not fully-resolved during the limited time range (0 – 2,000 s) of the stopped-flow mixing experiment, this value reflects an estimate for the rate of the reverse conformational change (k_{-2}, Figure 7.2A). Nevertheless, as the

apparent equilibrium dissociation constant reported for Cas9•sgRNA (K_d^{sgRNA} = 10 pM)[321] is a function[144] of K_1 and K_2 (*i.e.* $K_d^{sgRNA} = (K_1K_2)^{-1} = K_1^{-1}(5.5 \times 10^{-4}\ s^{-1}/6.1\ s^{-1})$) we concluded that the conformational change of Cas9 upon sgRNA binding enhanced the overall binding affinity by at least 10,000-fold (*i.e.* from 100 nM to 10 pM).

To verify the induced fit mechanism for sgRNA binding to Cas9, we performed smFRET experiments with $Cas9^{FRET\text{-}R}$ which was biotinylated via an AviTag[420] for surface-immobilization. Fluorescent labeling and biotinylation of $Cas9^{FRET\text{-}R}$ did not affect its activity relative to wild-type Cas9 (Figure 7.5C). As expected, surface-immobilized apo-$Cas9^{FRET\text{-}R}$ exhibited only high-FRET efficiencies (Figures 7.4E and 7.4F) and did not display transitions to lower-FRET efficiencies during single-molecule imaging suggesting that apo-Cas9 is not conformationally dynamic and sgRNA binding is required to induce the Cas9 conformational change (Figure 7.3A) which led to the observed fluorescence change in Figure 7.4B. Accordingly, Cas9 does not have to adopt the conformation observed in the binary complex crystal structure before sgRNA can associate (*i.e.* sgRNA does not select for a binary complex-like configuration of apo-Cas9). This result is consistent with cryo-electron microscopy (cryo-EM) of apo-Cas9 which demonstrated monodisperse particles with no indication of conformational dynamics at modest resolution (~19 Å) before addition of sgRNA.[323] However, as this FRET system ($Cas9^{FRET\text{-}R}$) specifically reports on conformational motions between the REC lobe and RuvC domain (Figure 7.4A), it does not wholly represent the global conformational dynamics of apo-Cas9.

7.3.2 Complicated DNA binding by Cas9•sgRNA.

We next estimated the apparent equilibrium dissociation constant for Cas9•sgRNA binding to DNA in the absence of $MgCl_2$ (K_d^{DNA} = 2 nM, Figure 7.6A) through an electrophoretic gel-mobility shift assay (EMSA) which was in agreement with a previously reported EMSA measurement.[279] Additionally, as the amplitude of the pre-steady-state burst phase (see Section 7.3.5) reflects the total concentration of ternary complexes at the start of a reaction in a pre-incubated solution of Cas9, sgRNA, and DNA prior to mixing with $MgCl_2$, we could obtain K_d^{DNA} and the active concentration of Cas9•sgRNA by measuring the burst phase amplitude of DNA cleavage product formation at various concentrations of DNA using an active site titration assay (Figure 7.6B). The K_d^{DNA} values measured by the active site titration assay (K_d^{DNA} = 1.82 – 1.85 nM) were consistent with that of the EMSA and Cas9•sgRNA was shown to be >90% active (Figure 7.6B).

To gain insight into the kinetics of DNA binding, we monitored the association kinetics of Cas9•sgRNA to an on-target DNA substrate (Figure 7.1) through a FRET construct ($Cas9^{FRET\text{-}D}$, Figure 7.7A) known to be sensitive to distance changes accompanying DNA binding[324] (Figure 7.5D) without affecting catalytic activity (Figure 7.5E). The pre-formed binary complex ($Cas9^{FRET\text{-}D}$•sgRNA, 75 nM) was rapidly mixed with increasing amounts of the DNA substrate (250 – 1,300 nM) in the presence of $MgCl_2$ (6 mM) and the acceptor fluorescence was recorded during excitation at 532 nm. The resulting time-dependent increases in acceptor fluorescence as the HNH domain moved proximal to the REC lobe confirmed a two-step model for DNA binding and were

biphasic (Figure 7.7B) with the fast exponential phase from each mixing experiment yielding a rate that was linearly dependent on DNA concentration (Figure 7.7C). The slope of the line ($k_{obs} = k_3$[DNA] + k_{-3})[321] yielded the bimolecular association rate constant (k_3, Figure 7.2A) for $Cas9^{FRET\text{-}D}$•sgRNA binding to DNA (9.1 x 10^6 $M^{-1}s^{-1}$) in the presence of $MgCl_2$ to form the ternary complex (Cas9•sgRNA•DNA, Figure 7.2A and 3B), which is similar to a value previously measured by single-molecule spectroscopy (~4 x 10^6 $M^{-1}s^{-1}$),[305] while the intercept of the line gave the rate constant for DNA dissociation (k_{-3}, Figure 7.2A) from the ternary complex (2.5 s^{-1}). In contrast to the conformational change observed for $Cas9^{FRET\text{-}R}$ (Figure 7.4B) which kinetically limited association of sgRNA, the observed linear dependence of the fast phase rate on DNA concentration (Figure 7.7C) suggests that the HNH domain motion related to DNA binding must be faster than the observed rates (k_3[DNA] ≈ k_{obs} ≤ 15 s^{-1}, Figure 7.7C) of DNA association.

The slow exponential phase rates (average = 0.30 s^{-1}, 10 – 15% amplitude, Figure 7.7B) did not display a linear dependence on DNA concentration (Figure 7.7C, bottom panel) which could be potentially attributed to i) a sub-population of the binary complex ($Cas9^{N}$•sgRNA) that had to first change conformation (*i.e.* conversion to Cas9•sgRNA) at the slow rate before binding to DNA (Figure 7.2A), ii) an additional mechanistic step following DNA binding that resulted in a FRET increase at the slow rate k_x (Figure 7.8), iii) a sub-population of ternary complexes ($Cas9^{N}$•sgRNA•DNA) that initially failed to properly associate with the DNA before converting to the productive ternary complex (Cas9•sgRNA•DNA) at the slow rate k_{-N} (Figure 7.9), or, similarly, iv) a sub-population

of ternary complexes that became trapped as a non-productive complex ($Cas9^{N}$•sgRNA•DNA, Figure 7.10) following DNA binding and had to transition back to the productive pathway (*i.e.* to Cas9•sgRNA•DNA) at the slow rate k_{-N}. To narrow down these possibilities, we monitored the association kinetics of unlabeled Cas9•sgRNA to an on-target DNA substrate (Figure 7.1) 5′-radiolabeled on both tDNA and ntDNA. Briefly, Cas9 (50 – 1,000 nM) was pre-incubated with a 3-fold molar excess of sgRNA to form the binary complex (Cas9•sgRNA) before rapid mixing with the DNA substrate (10 nM) in the presence of $MgCl_2$ (6 mM) for various times before quenching with EDTA (0.37 M) (Figure 7.11A). Time-dependent DNA cleavage by HNH and RuvC resulted in a single tDNA product (13-mer) and multiple ntDNA products (27-mer – 20-mer), respectively (Figure 7.11A). Importantly, product formation by RuvC was quantified as the sum of all ntDNA products. The apparent degradation of ntDNA products (Figure 7.11A) likely resulted from the known 3′→5′ exonuclease activity of RuvC which we kinetically and biochemically explored in a recent publication.[319] Similar to the results obtained for $Cas9^{FRET\text{-}D}$ (Figures 7.7B and 7.7C), the association kinetics for HNH (Figure 7.7D) and RuvC (Figure 7.11B) were biphasic with fast phase rates (k_{obs} values in Figures 7.7E and 7.11C, respectively) that were linearly dependent on the concentration of Cas9•sgRNA. This linear dependence of the fast phase rate of DNA product formation versus Cas9•sgRNA concentration suggests that for most molecules (~70%), steps subsequent to DNA binding must be much faster (>>15 s^{-1}) than the values of k_{obs} in the presence of $MgCl_2$ (Figures 7.7E and 7.11C) and thus did not kinetically limit product formation. The slope and intercept of each line yielded the respective bimolecular

association rate constant for Cas9•sgRNA binding to DNA (k_3, Figure 7.2A) and rate constant for DNA dissociation from the ternary complex (k_{-3}, Figure 7.2A) as measured through product formation by HNH (k_3 = 6.7 x 10^6 $M^{-1}s^{-1}$, k_{-3} = 2.5 s^{-1}, Figure 7.7E) or RuvC (k_3 = 3.0 x 10^6 $M^{-1}s^{-1}$, k_{-3} = 1.3 s^{-1}, Figure 7.11C) which agree with the values obtained by stopped-flow fluorescence (k_3 = 9.1 x 10^6 $M^{-1}s^{-1}$, k_{-3} = 2.5 s^{-1}, Figure 7.7C). Based on the true K_d^{DNA} (true $K_d^{DNA} = k_{-3}/k_3$ = 142 – 833 nM) and apparent K_d^{DNA} measured by EMSA and active site titration (~ 1.9 nM, Figures 7.6A and 7.6B) for Cas9•sgRNA and DNA, a second step following initial DNA association must enhance DNA binding by 75 – 450-fold.

The slow phase rates of HNH (0.15 s^{-1}, 28% amplitude, Figure 7.7D) and RuvC (0.17 s^{-1}, 39% amplitude, Figure 7.11B) were independent of Cas9•sgRNA concentration (Figures 7.7E and 7.11C) and of comparable magnitude to the slow phase rates reported for $Cas9^{FRET\text{-}D}$ (0.30 s^{-1}, 15% amplitude, Figure 7.7B) suggesting that both experiments may have been reporting on the same slow mechanistic event. Moreover, the slow phase in the product-dependent assays (Figures 7.7D and 7.11B) clarifies the results of the stopped-flow fluorescence experiments (Figure 7.7B) by excluding the possibility of an additional mechanistic step following DNA binding that limits product formation for all molecules (Figure 7.8). As a result, the mechanistic significance of the slow phase is restricted to either sub-populations (15 – 30%) of binary ($Cas9^N$•sgRNA, Figure 7.2A) or ternary ($Cas9^N$•sgRNA•DNA, Figures 7.9 and 7.10) complexes in a non-productive state that must slowly transition (k_{-N} = 0.15 – 0.30 s^{-1}) to the productive binary (Cas9•sgRNA, Figure 7.2A) or ternary (Cas9•sgRNA•DNA, Figures 7.9 and 7.10) complexes,

respectively, before rapidly cleaving the DNA substrate.

7.3.3 DNA binding kinetics of Cas9•sgRNA are affected by Mg^{2+} concentration.

Similar stopped-flow DNA binding experiments as in Fig 4B were performed in the absence of $MgCl_2$ (Figures 7.11D and 7.11E) and revealed a ~800-fold slower rate of DNA association (1.1 x 10^4 $M^{-1}s^{-1}$) by $Cas9^{FRET\text{-}D}$•sgRNA and ~700-fold slower rate of DNA dissociation (3.8 x 10^{-3} s^{-1}). Accordingly, the divalent metal ion cofactor may accelerate the assembly of the ternary complex. However, the overall stability of the ternary complex was unaffected as the true K_d^{DNA} for DNA binding (true $K_d^{DNA} = k_{-3}/k_3$, Figure 7.2) did not appreciably change (true K_d^{DNA} with $MgCl_2$ = 142 – 833 nM; true K_d^{DNA} without $MgCl_2$ = 345 nM). To further verify that Mg^{2+} affects DNA association by Cas9•sgRNA, we performed experiments similar to Figures 7.7D and 7.11B wherein the 5′-radiolabeled DNA substrate (10 nM) was mixed with Cas9•sgRNA (250 nM) in the presence of varying of amounts of $MgCl_2$ (50 – 5000 μM) (Figure 7.12). Indeed, at the lowest concentration of $MgCl_2$ (50 μM), the fast phase rates of tDNA and ntDNA product formation decreased by at least ~ 50-fold, thereby demonstrating that Mg^{2+} accelerates ternary complex formation by Cas9•sgRNA. Notably, the slow phase rates of product formation (Figures 7.12B and 7.12D, bottom panels) were not sensitive to $MgCl_2$ concentration. Moreover, DNA cleavage experiments (Figures 7.12E and 7.12G) from the pre-formed ternary complex (see Section 7.3.4) conducted with varying amounts of $MgCl_2$ (50 – 5000 μM) yielded DNA cleavage rates that did not depend on $MgCl_2$ concentration (Figures 7.12F and 7.12H). Thus, the decrease in the fast phase k_{obs} values at decreasing concentrations of $MgCl_2$ (Figures 7.12B and 7.12D, top panels) is related to

mechanistic events during DNA association rather than DNA cleavage. Altogether, to account for the distinct changes in DNA binding kinetics caused by $MgCl_2$ concentration, we have included an additional mechanism for ternary complex formation in the absence of $MgCl_2$ (Figure 7.2B).

To monitor DNA dissociation kinetics from the Cas9 ternary complex in the absence of $MgCl_2$, a solution of pre-formed wild-type Cas9•sgRNA•5′-radiolabeled DNA complex (10 nM) was mixed with a large molar excess of a trap consisting of catalytically-inactive dead $Cas9^{279}$ (dCas9, D10A/H840A) pre-bound to sgRNA (dCas9•sgRNA) for varying incubation times (1 – 600 min) before initiating DNA cleavage with $MgCl_2$ (Figure 7.13A). During the variable incubation period in the absence of $MgCl_2$, any 5′-radiolabeled DNA to dissociate from Cas9•sgRNA•DNA was rapidly bound by the large molar excess of dCas9•sgRNA trap and thus protected from wild-type Cas9-catalyzed cleavage, as demonstrated through a control experiment (Figure 7.13B). Importantly, we have demonstrated that Cas9•sgRNA retains nearly full activity after 10 hrs incubation at 37 °C (Figure 7.13C), thus loss in enzyme activity did not contribute to decrease in product formation in these assays. Time-dependent DNA dissociation was found to be biphasic (Figure 7.14A), suggesting that following initial DNA collision to form the encounter complex (Cas9•sgRNA•DNA, Figures 7.2A and 7.2B), the ternary complex may productively exist in at least two conformations ($Cas9^{A}$•sgRNA•DNA and $Cas9^{B}$•sgRNA•DNA, Figures 7.2A and 7.2B). DNA dissociates from the tight complex (k_{-4b}, Figure 7.2B) (76% HNH and 79% RuvC amplitudes, Figure 7.14A) and the less stable complex (k_{-4a}, Figure 7.2B) (24% HNH and

21% RuvC amplitudes, Figure 7.14A) in the absence of $MgCl_2$ with measured rate constants of 3.4 x 10^{-5} and 4.1 x 10^{-3} s^{-1}, respectively, for HNH and 2.5 x 10^{-5} and 2.5 x 10^{-3} s^{-1}, respectively, for RuvC. Importantly, as the rates of the slow phases in both the stopped-flow FRET and product-dependent assays of DNA association did not depend on DNA or Cas9•sgRNA concentration, we can exclude the potential mechanisms involving non-productive ternary complexes (Figures 7.9 and 7.10) as the conversion of Cas9•sgRNA•DNA to $Cas9^N$•sgRNA•DNA, $Cas9^A$•sgRNA•DNA, or $Cas9^B$•sgRNA•DNA should compete with one another resulting in observed DNA concentration- (slow phase k_{obs} in Figure 7.7B = k_{-N} + (k_N[DNA]/(k_{-3}/k_3 + [DNA])) or Cas9•sgRNA concentration- (slow phase k_{obs} in Figures 7.7D and 7.11B = k_{-N} + (k_N[Cas9•sgRNA]/(k_{-3}/k_3 + [Cas9•sgRNA])) dependent rates for the slow phase k_{obs} values. Moreover, as the slow phase of product formation did not change with varying concentration of $MgCl_2$ during DNA association experiments (Figures 7.12B and 7.12D, bottom panels), it must represent a mechanistic step occurring before DNA binding. Thus, the mechanism featuring a sub-population of the binary complex ($Cas9^N$•sgRNA) which must first change conformation (*i.e.* conversion to Cas9•sgRNA) at the slow phase rate before binding to DNA is most appropriate (k_{-N}, Figures 7.2A and 7.2B). Indeed, stopped-flow FRET experiments (Figure 7.11D) of DNA association in the absence of $MgCl_2$ revealed a fluorescence change at a rate (Figure 7.11E, top panel) consistent with the slow phase rates of fluorescence change (Figure 7.7C, bottom panel) and product formation (Figures 7.7E, 7.11C, 7.12B, and 7.12D, bottom panels) measured in the presence of $MgCl_2$.

Importantly, the two ternary complexes detected in Figure 7.14A in the absence of $MgCl_2$ represent different Cas9 conformational states and DNA dissociation from either state is limited by respective slow reverse conformational changes (k_{-4a} and k_{-4b}, Figure 7.2B). Consistently, an analogous stopped-flow assay monitoring the decrease in acceptor fluorescence of $Cas9^{FRET\text{-}D}$•sgRNA•DNA as the HNH domain slowly re-opened to permit DNA release in the absence of $MgCl_2$ revealed a DNA dissociation rate (5.5×10^{-4} s^{-1}) which is similar (within two-fold)[144,474] to an overall DNA dissociation rate ($76\% \times 3.4 \times 10^{-5} + 24\% \times 4.1 \times 10^{-3}$ $s^{-1} = 1.0 \times 10^{-3}$ s^{-1}) from the two ternary complexes ($Cas9^{A}$•sgRNA•DNA and $Cas9^{B}$•sgRNA•DNA, Figure 7.2B) also measured in the absence of $MgCl_2$ within the limited time range of the stopped-flow instrument (Figure 7.14B). This result verifies that DNA substrate release is limited by slow reverse conformational changes of Cas9 (k_{-4a} and k_{-4b}, Figure 7.2B).

Altogether, $Cas9^{A}$•sgRNA•DNA ($K_{eq}^{A} = K_3K_{4a} = (k_3/k_{-3})(k_{4a}/k_{-4a})$) and $Cas9^{B}$•sgRNA•DNA ($K_{eq}^{B} = K_3K_{4b} = (k_3/k_{-3})(k_{4b}/k_{-4b})$) must both contribute to the apparent K_d^{DNA} ($1.9\text{ nM} = K_d^{DNA} \approx (24\%*K_{eq}^{A} + 76\%*K_{eq}^{B})^{-1}$) measured by EMSA (Figure 7.6A) and active site titration (Figure 7.6B) in the absence of $MgCl_2$. Thus, we calculated the rates of the forward conformational changes to $Cas9^{A}$•sgRNA•DNA (k_{4a} = 1.8 – 3.0 s^{-1}, Figure 7.2B) and $Cas9^{B}$•sgRNA•DNA ($k_{4b} = (6.0 – 8.2) \times 10^{-3}$ s^{-1}, Figure 7.2B) in the absence of $MgCl_2$. Based on these calculations, the steps (K_{4a} and K_{4b}) following initial DNA binding increased the overall DNA binding affinity by 182-fold compared to the true K_d^{DNA} (*i.e.* true $K_d^{DNA} = k_{-3}/k_3 = 345$ nM).

7.3.4 Pre-steady-state kinetic studies of Cas9-catalyzed DNA cleavage and HNH domain motion.

To probe the kinetics of DNA cleavage following ternary complex assembly, a pre-incubated solution of Cas9•sgRNA (100 nM) and the DNA substrate (10 nM), 5′-radiolabeled on both tDNA and ntDNA, was rapidly mixed with $MgCl_2$ to initiate DNA cleavage for 0.003 – 60 s before quenching the reactions with EDTA (0.37 M) (Figure 7.15A). Consistent with the existence of two ternary complexes ($Cas9^A$•sgRNA•DNA and $Cas9^B$•sgRNA•DNA, Figures 7.2A and 7.2B) identified from dissociation assays from the binding equilibrium established in the absence of $MgCl_2$ (Figure 7.14), we observed biphasic kinetics for each active site (Figure 7.15B) with HNH cleaving tDNA with an initial fast rate of 90 s^{-1} (67% amplitude) followed by a slow rate of 5 s^{-1} (33% amplitude), and RuvC cleaving ntDNA with fast and slow rates of 25 s^{-1} (38% amplitude) and 0.37 s^{-1} (62% amplitude), respectively. Thus, even the slowest rate of Cas9-catalyzed DNA cleavage from the pre-formed ternary complex measured here (0.37 s^{-1}) occurred 21- to 75- fold more rapidly than previously measured rates for DNA cleavage (0.005 – 0.017 s^{-1}),[279] which must have been primarily reporting on the concentration-dependent kinetics of DNA association to Cas9•sgRNA. Notably, more than one ntDNA product (27-mer, 26-mer, and 25-mer corresponding to cleavage at the -3, -4, and -5, positions, respectively (Figure 7.1)) was observed even at the earliest time point of DNA cleavage accessible by our rapid chemical quench-flow device (0.003 s, Figure 7.15A) suggesting that RuvC may be promiscuous in initial ntDNA cleavage site selection which we have explored further in a recent publication.[319]

The biphasic pattern of DNA cleavage observed in Figure 7.15B may have resulted from conformational changes of the HNH and RuvC domains required for transition from one or more ternary complex configurations (*i.e.* $Cas9^{A}$•sgRNA•DNA, $Cas9^{B}$•sgRNA•DNA, Figure 7.2A) to the reaction state ($Cas9^{R}$•sgRNA•DNA, Figure 7.2A). Indeed, close inspection of the crystal structures of Cas9 bound to sgRNA and tDNA show conformational heterogeneity of the HNH and RuvC domains between two molecules (A and B) within one asymmetric unit (PDB code: 4OO8).[317] Specifically, residues 909-940 of the RuvC domain form two short α-helices in Molecule A while they adopt an extended structure in Molecule B. Moreover, the HNH domain is positioned very far from the scissile bond at the -3 position of tDNA in Molecule A and is therefore likely an early-stage ternary complex that is incapable of tDNA cleavage. In addition to the aforementioned extended α-helix within the RuvC domain observed in Molecule B, the HNH domain was disordered suggesting its dynamic motion towards the tDNA cleavage site *in crystallo*. The conformations of the HNH and RuvC domains described for Molecule A were also observed in a crystal structure of Cas9 (H840A, HNH active site knockout) bound to sgRNA and partial duplex DNA (PDB code: 4UN3) (Figure 7.3C).[318] However, a recent crystal structure of catalytically active Cas9 in a pre-cleavage state bound to sgRNA and full duplex DNA (PDB code: 5F9R)[327], was consistent with Molecule B by demonstrating the extended α-helix within the RuvC domain (Figure 7.3C). As foreshadowed by the domain dynamics inferred from the disordered HNH domain of Molecule B, this structure also revealed a dramatic ~180° rotation and ~20 Å translation of the HNH domain towards the tDNA cleavage site relative to its position in

the partial duplex DNA-bound Cas9 structure (PDB code: 4UN3) (Figure 7.3C) suggesting that this structure represents a late-stage ternary complex poised for DNA cleavage.[13] Together these structures illustrate the dynamics of the HNH and RuvC domains within the ternary complex.

To explore the possibility that observed DNA cleavage kinetics (Figure 7.15B) reflect conformational motions of Cas9, we performed stopped-flow fluorescence experiments with the same FRET construct utilized to examine DNA binding (Figure 7.7A, $Cas9^{FRET\text{-}D}$). This FRET construct should also be sensitive to motion of the HNH domain upon transition from the early-stage to late-stage ternary complex configurations (Figures 7.3C and 7.15D).[324] Upon rapid mixing of $MgCl_2$ with the pre-formed $Cas9^{FRET\text{-}D}$ ternary complex (300 nM), we observed a rapid, biphasic FRET increase (Figure 7.15C) with a fast phase rate of 89 s^{-1} and an amplitude of 77%, as well as a slow phase rate of 4.6 s^{-1} and an amplitude of 23%, as the HNH domain was triggered for catalysis. Remarkably, the kinetics of these conformational transitions (Figure 7.15C) were nearly identical to the product-dependent DNA cleavage kinetics measured for the HNH active site (Figure 7.15B). We thus propose that DNA cleavage kinetics for tDNA and ntDNA are governed by the conformational states of the HNH and RuvC domains, respectively.

Accordingly, the early- and late-stage Cas9 ternary structures correspond to the less stable ($Cas9^{A}$•sgRNA•DNA, Figures 7.2A and 7.3C) and tight ($Cas9^{B}$•sgRNA•DNA, Figures 7.2A and 7.3C) ternary complexes inferred from the dissociation assays pre-equilibrated in the absence of $MgCl_2$, respectively (Figure 7.14). The slow phases of HNH- and RuvC-catalyzed DNA cleavage (Figures 7.15B and 7.15C) must then

correspond to conformational transitions (k_{5a} and k_{5b}, Figures 7.2A and 7.2B) from the early-stage ($Cas9^{A}$•sgRNA•DNA) to the late-stage ($Cas9^{B}$•sgRNA•DNA) ternary complexes (Figure 7.3C). The late-stage ternary structure of Cas9 in the absence of Mg^{2+} shows that the HNH domain is still ~10 Å too far from the scissile bond on tDNA for cleavage to commence (Figure 7.15E).[327] As a result, the catalytic residues of the RuvC domain are also not properly positioned for ntDNA cleavage[327] as supported by conformational coupling between HNH and RuvC reported by a steady-state FRET study.[324] Therefore, we surmise that the fast phases of DNA cleavage correspond to the final ~10 Å translation of the HNH domain from the late-stage ternary complex configuration (Figure 7.15E) toward the scissile bond on tDNA upon Mg^{2+} binding (k_{6a}, Figure 7.2A) which then induces proper alignment of the RuvC domain (k_{6b}, Figure 7.2A) for Mg^{2+} binding and ntDNA cleavage. These steps are highlighted in an expanded mechanism of HNH- and RuvC-catalyzed DNA cleavage beginning from $Cas9^{A}$•sgRNA•DNA shown in Figure 7.2A (dashed box). The two states of the ternary complex ($Cas9^{A}$•sgRNA•DNA and $Cas9^{B}$•sgRNA•DNA, Figures 7.2A and 7.2B) exist in equilibrium between distinct structurally-supported conformations (Figure 7.3C)[318,327] of the HNH and RuvC domains in the absence of Mg^{2+}: an early-stage conformation (HNH^{A}•tDNA and $RuvC^{A}$•ntDNA) which must convert (k_{5a}, k_{5b}, Figures 7.2A and 7.2B) to a late-stage, near-cleavage conformation (HNH^{B}•tDNA and $RuvC^{B}$•ntDNA). From this late-stage conformation, rapid transition (k_{6a} and k_{6b}, Figure 7.2A) to a reaction state (HNH^{R}•tDNA and $RuvC^{R}$•ntDNA) is induced by Mg^{2+} binding, resulting in subsequent tDNA and ntDNA cleavage (HNH^{R}•$tDNA_{Products}$ and $RuvC^{R}$•$ntDNA_{Products}$). We assume

that the binding of Mg^{2+} to HNH and RuvC at the concentration used in our experiments (6 mM) is rapid and the chemistry of DNA strand scission (k_{chem}, Figure 7.2A) from the reaction state ($Cas9^{R}$•sgRNA•DNA, Figure 7.2A) to form DNA products ($Cas9^{R}$•sgRNA•DNA_{Pdts}, Figure 7.2A) is also fast (≥700 s^{-1}) as measured for other endonucleases.[475,476] Thus, it is most likely that k_{6a} and k_{6b} directly report on the Mg^{2+}-binding-coupled conformational changes of HNH and RuvC from the late-stage ternary complex (HNH^{B}•tDNA and $RuvC^{B}$•ntDNA, Figure 7.2A) to their reaction state (HNH^{R}•tDNA and $RuvC^{R}$•ntDNA, Figure 7.2A). This model for DNA cleavage from the pre-formed ternary complex was best exemplified by the stopped-flow fluorescence results of $Cas9^{FRET\text{-}D}$•sgRNA•DNA (Figure 7.15C) which directly reported on HNH domain motion. At the onset of the experiment, the majority of molecules (~80%) rapidly (k_{6a}, Figure 7.2A) closed the distance (~10 Å, Figure 7.15E) to the reaction state ($Cas9^{R}$•sgRNA•DNA) from the late-stage ternary complex configuration ($Cas9^{B}$•sgRNA•DNA) resulting in the rapid FRET increase while the minority population (~20%) had to complete the large conformational change from the early-stage ($Cas9^{A}$•sgRNA•DNA, Figures 7.2A and 7.2B) to the late-stage ($Cas9^{B}$•sgRNA•DNA, Figures 7.2A and 7.2B) ternary complex (Figure 7.15D) as well as the final, small transition to the reaction state resulting in the slower observed rate (k_{5a}, Figure 7.2A).

Based on our model for HNH and RuvC-catalyzed DNA cleavage (Figure 7.2A), we were also able to calculate the rates of the reverse conformational changes of HNH^{B}•tDNA (k_{-5a} = 2 s^{-1}, Figure 7.2A) and $RuvC^{B}$•ntDNA (k_{-5b} = 0.60 s^{-1}, Figure 7.2A) by simple kinetic partitioning using the relative populations of each species established at

equilibrium in the absence of $MgCl_2$ (*i.e.* relative amplitudes of the two observed phases for HNH- (67% for fast phase and 33% for slow phase, $K_{eq}^{HNH} = k_{5a}/k_{-5a}$ = 67%/33%) and RuvC- (38% for fast phase and 62% for slow phase, $K_{eq}^{RuvC} = k_{5b}/k_{-5b}$ = 38%/62%) catalyzed DNA cleavage from the pre-equilibrated ternary complex) and the measured forward rate constants (k_{5a}, k_{5b}, Figure 7.2A) obtained in Figure 7.15B.[144] Remarkably, the forward conformational change of the RuvC domain ($k_{5b,}$ Figure 7.2A) is unfavorable relative to the reverse conformational change (k_{-5b}, Figure 7.2A), indicating that Mg^{2+}-binding-coupled transition of the HNH domain from the late-stage ternary complex (HNH^B•tDNA, Figure 7.2A) to the reaction state (HNH^R•tDNA, Figure 7.2A) is necessary to drive the unfavorable RuvC conformational equilibrium ($RuvC^B$•ntDNA to $RuvC^R$•ntDNA, Figure 7.2A) forward and alluding to a kinetic checkpoint for preventing off-target DNA cleavage by Cas9. For example, as these domain motions are necessary for DNA strand scission, their impediment may offer a unique strategy to conditionally prevent Cas9-catalyzed cleavage of DNA with incomplete complementarity to the sgRNA. Altogether, these kinetic data (Figure 7.2A, dashed box) suggest that the favorable forward conformational equilibria of the HNH domain (Figure 7.3 C) from the non-productive to productive and then to reaction state eventually drive forward the unfavorable conformational equilibrium of the RuvC domain upon binding of Mg^{2+}.

As a consequence of the forward and reverse rates for HNH and RuvC domain motion (Figure 7.2A, dashed box), it is also kinetically possible for $Cas9^B$•sgRNA•DNA to exist with HNH in the late-stage structural conformation (HNH^B•tDNA, Figure 7.2A) before the RuvC domain has achieved the equivalent conformation ($RuvC^B$•ntDNA,

Figure 7.2A). Accordingly, Figure 7.2C summarizes this (Complex II) and three additional configurations of the HNH and RuvC domains which can potentially exist at the onset of the single-turnover kinetic experiments (Figure 7.15). Importantly, one of the four possibilities (Complex III, Figure 7.2C) wherein the RuvC domain has transitioned to the late-stage ternary complex configuration before the HNH domain, is ruled out by both the kinetics in Figure 7.2A and a steady-state FRET study detailing that the HNH domain conformational change is a prerequisite for the RuvC domain motion.[324]

We also point out here the possibility that if Mg^{2+} were to be continuously present for every step of the mechanism (Figure 7.2A), the early-stage ternary complex configuration ($Cas9^{A}$•sgRNA•DNA, Figure 7.2A) may be only transiently populated and the majority of molecules would instead rapidly flux through k_{4b} directly to the late-stage ternary complex configuration ($Cas9^{B}$•sgRNA•DNA, Figure 7.2A). This conclusion is supported by our stopped-flow FRET (Figures 7.7B, 7.7C, 7.11D, and 7.11E) and product-dependent (Figure 7.12) DNA association experiments in the presence of varying amounts of $MgCl_2$ which indicate that DNA association is accelerated by the presence of Mg^{2+}. Importantly, as X-ray crystal structures of Cas9•sgRNA bound to DNA[317,318,327] were all obtained in the absence of $MgCl_2$, the pre-steady-state kinetics measured from the pre-formed ternary complex here (Figures 7.14 and 7.15) reveal the dynamic conformational equilibria that is established in the absence of $MgCl_2$ and correlate the disjointed structural data[317,318,327] with in-solution experiments (Figure 7.3C).

Importantly, as individual knockout of either active site does not preclude cleavage of the opposite strand,[279] HNH and RuvC are not coupled through chemistry but

instead are likely conformationally coupled as supported by our mechanism (Figure 7.2A) and steady-state FRET data.[324] Our results directly show that HNH cleaves faster than RuvC (Figures 7.15A and 7.15B) and therefore expand this hypothesis to suggest that conformational activation of the HNH domain is a prerequisite for cleavage of tDNA and ntDNA, thereby controlling when (*i.e.* timing) DNA cleavage occurs. However, it is the slow rate (0.37 s^{-1}, k_{5b}, Figure 7.2A) for RuvC isomerization which limits the overall rate of double-stranded DNA cleavage from the pre-formed ternary complex.

Notably, we hypothesize that even after initial DNA cleavage, Cas9 remains in the reaction state ($Cas9^{R}$•sgRNA•DNA_{pdts}, Figure 7.2A) as supported by our observed (Figure 7.15A) and the previously reported 3′→5′ exonuclease activity of $RuvC^{279}$ which suggests persistent activated conformations of both HNH (HNH^{R}•tDNA, Figure 7.2A) and RuvC ($RuvC^{R}$•ntDNA, Figure 7.2A). Lack of additional tDNA cleavage by HNH following initial strand scission despite an activated conformation is likely due to base pairing between tDNA products and sgRNA (Figure 7.1) which restricts flexibility of the tDNA products within HNH as compared to the flexible single-stranded ntDNA products within RuvC.

7.3.5 Extremely slow kinetics of multiple-turnovers by Cas9.

To investigate the kinetics of the first and subsequent catalytic turnovers, we performed a burst kinetic assay. Upon rapid mixing of a pre-incubated solution of Cas9•sgRNA (10 nM) and DNA (50 nM), 5′-radiolabeled on both tDNA and ntDNA, with $MgCl_2$ (6 mM), we monitored the early and late time points of DNA cleavage (3 ms – 6 hrs) (Figure 7.16A). Within the first catalytic turnover, we measured biphasic

exponential DNA cleavage kinetics (Figure 7.17A) which were consistent with the single-turnover assay (Figure 7.15B) with respective fast and slow rates of 70 (52% amplitude) and 5 s^{-1} (48% amplitude) for HNH, and 34 (40% amplitude) and 0.17 s^{-1} (60% amplitude) for RuvC. Following the first catalytic turnover, we observed a very slow, linear phase of product formation corresponding to multiple catalytic turnovers which occurred at similar rates of 3.2 x 10^{-6} s^{-1} for HNH and 2.1 x 10^{-6} s^{-1} for RuvC (Figure 7.17A). Consistently, steady-state kinetic assays wherein Cas9•sgRNA (10 nM) was mixed with DNA (50 nM, 5′-radiolabeled on either tDNA (Figure 7.16B) or ntDNA (Figure 7.16C) in the presence of $MgCl_2$ (6 mM) also revealed slow multiple-turnover rates for HNH (4.45 x 10^{-6} s^{-1}) and RuvC (2.1 x 10^{-6} s^{-1}) (Figure 7.17B). Interestingly, these multiple-turnover rate constants measured through the burst and steady-state kinetic assays (k_7, Figure 7.2A) are comparable to the slow rate of DNA substrate dissociation (k_{-4b}, Figure 7.2B) measured in Figure 7.14A and likewise may be limited by slow reverse conformational changes (Figure 7.14B) leading to DNA product release. Indeed, the ~6- to 16-fold slower multiple-turnover rate (k_7 relative to k_{-4b}) for DNA product release could suggest kinetic contribution of multiple slow reverse conformational changes (*i.e.* $Cas9^R \rightarrow Cas9^B \rightarrow Cas9^A \rightarrow Cas9$) from the product state ($Cas9^R$•sgRNA•DNA_{pdts}) which collectively limit DNA product dissociation and enzyme regeneration for subsequent turnovers. Moreover, the DNA product state was likely stabilized by the presence of $MgCl_2$ leading to the slower k_7. Nevertheless, the process of DNA product release (k_7, Figure 7.2A) was the slowest kinetic step during multiple-turnovers (Figure 7.2A), and due to the extremely slow rate ($t_{1/2} = \ln(2)/k_8$ = ~43 – 91 hrs), limits Cas9 as a virtual

single-turnover enzyme during adaptive immunity in bacteria as well as gene-editing in eukaryotes. Following dissociation, the resulting DNA products are no longer suitable substrates for Cas9 and thus re-association (*i.e.* reverse of k_7) is likely not possible. Importantly, later time points from gels of both burst (Figure 7.16A) and steady-state (Figures 7.16B and 7.16C) kinetic assays revealed significant 3′→5′ degradation of ntDNA by RuvC which we quantitatively characterized in a recent publication.[319]

To further test whether Cas9 remains tightly bound to its DNA products, we designed an assay wherein Cas9 and PvuI were used together to cleave a ~5 kb plasmid (derived from pCDNA3.1, see Experimental Section) and generate DNA products of ~1 kb and ~4 kb in size. Briefly, a pre-incubated solution of Cas9•sgRNA (100 nM), PvuI (10 U), and the plasmid (10 nM) containing the same target sequence and PAM as in Figure 7.1 was mixed with $MgCl_2$ (6 mM) for 1 hr at 37 °C before separation by 0.8% native agarose gel electrophoresis. Although samples treated with 1% SDS to denature all protein components revealed the expected ~1 kb and ~4 kb DNA products, samples not treated with the denaturant revealed a single band at ~5 kb (Figure 7.18A). Accordingly, either PvuI or Cas9 remained tightly associated with the DNA product ends following the cleavage reaction and prevented release of the DNA product fragments. Indeed, a similar experiment wherein the PvuI-treated linearized plasmid was gel-purified prior to reaction with Cas9•sgRNA for 5 hrs at 37 °C also revealed the product DNA fragments of expected size (~1 kb and ~4 kb) when treated with 1% SDS, but only a single band in the absence of the denaturant, likely corresponding to Cas9-bound cleaved plasmid (~5 kb) (Figure 7.18B). Thus, Cas9, rather than PvuI, was tightly associated with the product

ends of the DNA plasmid following cleavage even after such a long incubation period (5 hrs) (Figure 7.18B). This result corroborates the findings from the multiple-turnover kinetic assays (Figure 7.17) indicating that DNA product release is the slowest mechanistic step (k_7, Figure 7.2A) and also demonstrates that Cas9 can remain tightly bound to even large DNA products for a substantial amount of time following double-stranded DNA cleavage. Moreover, our findings are consistent with interferometry[477] and single-molecule DNA curtain assays[304] that indicated Cas9 remains tightly associated to its DNA products. In contrast, recent high-speed atomic force microscopy (HS-AFM) of Cas9 demonstrated relatively rapid DNA product release which must be an artifact of the physical disturbances to Cas9 caused by the AFM probe.[325] It is thus clear that during *in vivo* gene-editing, some protein factor(s) must be responsible for displacing Cas9 from the DNA ends generated by DNA cleavage in order for host DNA repair machinery of the non-homologous end joining (NHEJ) or homology-directed repair (HDR) pathways to engage the double-stranded DNA break.

7.4 Experimental Section

7.4.1 Preparation of Sp Cas9 constructs, sgRNA, and DNA substrates.

Sp Cas9 and mutants were purified as previously described.[478] Briefly, wild-type and mutant Cas9 constructs were prepared using pMJ806, a plasmid received as a generous gift from Jennifer Doudna (Addgene plasmid # 39312).[279] D10A/H840A (dCas9) active site mutant was generated by site-directed mutagenesis to abolish Cas9 endonuclease activity.[279]

For fluorescence studies, the two native Cys residues of *Sp* Cas9 (C80 and C574)

were mutated to Ser. Next, site-specific Cys mutations were introduced at positions D435 and E945 to generate a FRET pair capable of monitoring sgRNA binding ($Cas9^{FRET-R}$, Figure 7.4A).[324] Similarly, Cys mutations at positions S867 and S355 were used to generate a FRET pair capable of monitoring DNA binding ($Cas9^{FRET-D}$, Figure 7.7A).[324] Proteins were labeled by incubation with a 15-fold molar excess of Cy3- and Cy5-maleimide (Lumiprobe Corporation) in labeling buffer (20 mM HEPES pH 7.5 at 4 °C, 200 mM KCl, 10% glycerol, and 1 mM TCEP) for 2 hours at 20 °C before transferring to 4ºC overnight. Unreacted dye was removed using size-exclusion chromatography. The degree of protein labeling was assessed by UV/Vis spectroscopy per the manufacturer's protocol with the theoretical molar extinction coefficient for Cas9 of 120,450 M^{-1} cm^{-1} and was found to be >90% Cy3- and >90% Cy5 dual-labeled in all cases. Though this labeling strategy resulted in a heterogeneous population of labeled species (*i.e.* donor and acceptor, donor only, acceptor only, single-label species), our experiments were focused on the directionality and kinetics of conformational motions, rather than specific distance changes. For smFRET experiments, $Cas9^{FRET-R}$ was modified with a 15 amino acid AviTag at the C-terminus. The protein was expressed, purified, and fluorescently labeled, as described above, before site-specific biotinylation with d-biotin and BirA following a published protocol.[420] The extent of protein biotinylation was >95% as determined by a gel-shift assay with streptavidin (data not shown).

sgRNA (Figure 7.1) was prepared by in vitro transcription (IVT) as previously described.[478] Briefly, a 140-mer DNA, based on the S. pyogenes CRISPR and trans-activating RNA sequences fused by a GAAA tetraloop linker, was synthesized by

Integrated DNA Technologies Inc. (IDT), cloned into a plasmid vector, PCR amplified, and used as the template for IVT. The sgRNA was then purified by denaturing PAGE (8 M urea, 10% polyacrylamide, 1x TBE) and precipitated in ethanol for storage at -80 °C. Prior to each experiment, the sgRNA was pelleted, washed, resuspended in RNA annealing buffer (10 mM Tris, 50 mM NaCl, 1 mM EDTA, pH 7.5), and then re-folded by heating to 95 °C for 5 min followed by slow cooling to 20 °C.

DNA substrates (Figure 7.1) were prepared using DNA oligonucleotides purchased from IDT and purified by denaturing PAGE (8 M urea, 10-15% polyacrylamide, 1x TBE). Radiolabeling at the 5′-end of an oligonucleotide was achieved by incubation with [γ-^{32}P]ATP (Perkin-Elmer) and Optikinase (Affymetrix) for 3 hours at 37 °C. Unreacted [γ-^{32}P]ATP was removed using Bio-Spin 6 size-exclusion columns (Bio-Rad). Radiolabeled target and non-target oligonucleotides were annealed by heating to 95 °C for 5 min followed by slow cooling to 20 °C to generate DNA substrates. Varying amounts of radiolabeled target and non-target DNA strands were annealed and analyzed by native-PAGE to determine the ideal ratio for optimal DNA duplex formation. Typical optimized annealing efficiencies were ~90% (data not shown). The primary DNA substrate (40/44-mer) used in this study (Figure 7.1) consisted of a 40-mer target DNA strand, with 20 base pair complementarity to the guide region of the sgRNA, annealed to a 44-mer non-target DNA strand containing the 5′-TGG-3′ PAM sequence.

7.4.2 Pre-steady-state kinetic assays.

All kinetic experiments, unless otherwise specified, were performed at 37 °C in optimized reaction buffer A containing 20 mM HEPES (pH 7.5 at 37 °C), 125 mM KCl,

1 mM EDTA, 1 mM DTT, and 10% glycerol. Fast reactions were performed using a rapid chemical quench-flow apparatus (KinTek). All reported concentrations are final, after mixing. Reactions were terminated using quench solution with a final composition of 0.37 M EDTA, 20% formamide, 0.25% bromophenol blue, and 0.25% xylene cyanol. Notably, 0.37 M EDTA was determined to completely prevent metal ion dependent catalysis (data not shown) as reported for other DNA modifying enzymes.[90,95]

7.4.3 Product and data analysis.

Reaction products were resolved by denaturing PAGE (12% polyacrylamide, 8 M urea, 1x TBE), and quantified using a Typhoon Trio (GE Healthcare) and ImageQuant (Molecular Dynamics). Product formation by HNH and RuvC was simultaneously monitored (5′-radiolabeling of both tDNA and ntDNA) and individually quantified. Unless otherwise specified, RuvC product formation was quantified as the sum of all ntDNA products and all kinetic data were fit by non-linear regression using KaleidaGraph (Synergy Software) to single exponential (Eq. 7.1), double exponential (Eq. 7.2), or the biphasic burst equation (Eq. 7.3).

$$[\text{product}] = A[1 - \exp(-kt)] + \text{Constant} \qquad \text{Eq. 7.1}$$

$$[\text{product}] = A_1[1 - \exp(-k_1t)] + A_2[1 - \exp(-k_2t)] + \text{Constant} \qquad \text{Eq. 7.2}$$

$$[\text{product}] = A_1[1 - \exp(-k_1t)] + A_2[1 - \exp(-k_2t)] + (A_1 + A_2)k_3t \qquad \text{Eq. 7.3}$$

where A, A_1, and A_2 are the exponential phase amplitudes; k, k_1, and k_2 are the observed rate constants; and k_3 is the steady-state rate constant. The constant included in Eq. 7.1 and Eq. 7.2 is a non-zero value for fitting of fluorescence data which has an arbitrary starting voltage (the constant is zero for product-based assays). Eq. 7.1 and Eq. 7.2 are

used when the reaction conditions exclude the possibility of multiple enzyme turnovers (*i.e.* [Enzyme] >> [Substrate]). Eq. 7.3 is used when the reactions conditions permit the possibility of multiple enzyme turnovers (*i.e.* [Enzyme] < [Substrate]). Importantly, experimental conditions for each kinetic assay in this work ensured that first-order reaction kinetics (*i.e.* reaction rate depends on only one reactant concentration) were followed. In this way, mechanistic steps could be probed individually as first-order processes by ensuring that the concentration of one reactant (*i.e.* enzyme or substrate) was in excess over the other reactant.[144] Eq. 7.1 is the integration of the first-order rate law describing the enzyme-catalyzed conversion of substrate to product. Eq. 7.2 was used to describe the conversion of substrate to product through a series of two first-order processes, which is best described using a sum of exponential terms. Eq. 7.3 includes an additional term ($(A_1 + A_2)k_3t$) which is derived from the zero-order rate law governing the kinetics of multiple enzyme turnovers when substrate is in excess over enzyme (k_3 often describes product dissociation from enzyme).[144] Exponential fits to the data were evaluated based on the coefficient of determination (R^2) as well as by visual inspection. Data were initially assumed to display single-exponential kinetics (Eq. 7.). However, if the R^2 was less than 0.98 and the fit failed to adequately describe the majority of data, a higher order exponential function (*i.e.* Eq. 7.2) was used. For more thorough description of kinetic theory, derivation of relevant equations, and experimental techniques to elucidate enzyme kinetic mechanisms, we point the interested readers to several insightful sources[15,31] and previous publications from our laboratory on enzyme kinetic mechanisms.[91,92,94]

7.4.4 Measurement of Cas9•sgRNA affinity to DNA and active enzyme concentration.

To measure the equilibrium dissociation constant of DNA (K_d^{DNA}) from the Cas9 ternary complex (Cas9•sgRNA•DNA), increasing amounts of Cas9 binary complex (Cas9•sgRNA) (0.002 – 100 nM) were added to 5′-radiolabeled DNA (0.1 nM) in the absence of $MgCl_2$ and incubated for 1 hour at 20 °C before ternary complex was resolved from free DNA by gel-shift assay (5% polyacrylamide, 0.5x TBE, 4°C) (Figure 7.6A). The concentration of ternary complex was quantified by phosphorimaging and plotted against the total concentration of Cas9•sgRNA. The data were then fit to a quadratic binding equation, Eq. 7.4,

$$[\text{Complex}] = 0.5(K_d + E_0 + D_0) - 0.5\,[(K_d + E_0 + D_0)^2 - 4E_0D_0]^{1/2} \qquad \text{Eq. 7.4}$$

where E_0 and D_0 are the Cas9•sgRNA and DNA concentrations, respectively, and is valid when [Complex] can be accurately measured and E_0 and D_0 are known.[479]

As the concentration of Cas9•sgRNA•DNA formed during pre-equilibration is given by the amplitude of the first catalytic turnover in the burst kinetic assay (Figure 7.17A), we determined the equilibrium dissociation constant (K_d^{DNA}) of the ternary complex through an active site titration assay wherein a fixed concentration of 2 nM Cas9•sgRNA (pre-formed by incubating Cas9 (2 nM) and sgRNA (6 nM) for 30 min at 20 °C) was titrated with increasing amounts of DNA substrate (0.05-50 nM) radiolabeled at the 5′-ends of tDNA and ntDNA in the absence of $MgCl_2$. The reactions were

incubated for 1 hr at 20 °C before DNA cleavage was initiated by rapid mixing with $MgCl_2$ (6 mM) for 15 s prior to quenching with EDTA (0.37 M). Burst phase product formation was then plotted against corresponding DNA substrate concentrations and the plots were fit to a quadratic equation (Eq. 7.4) to yield K_d^{DNA} and the active enzyme concentration for each active site of Cas9•sgRNA (Figure 7.6B).

7.4.5 Stopped-flow FRET assays.

Stopped-flow fluorescence spectroscopy experiments were performed at 37 °C using an SX20 instrument (Applied Photophysics) with a dead time of ~1 ms. Monochromator slit widths were set to 5 nm for all experiments. Reaction components were rapidly mixed in a 20 μL optical cell during excitation at 532 nm. Emission of Cy3 and Cy5 was separately monitored using Et575/40m and Et685/70m bandpass filters (Chroma), respectively. Stopped-flow kinetic data were fit to single (Eq. 7.1) or double (Eq. 7.2) exponential equations.

7.4.6 Single-molecule FRET assays.

Single-molecule FRET measurements were performed on our custom-built, prism-type TIRF microscopy system, as previously described.[9] Briefly, quartz microscope slides were functionalized and passivated following a published protocol,[480] before imaging chamber assembly and addition of NeutrAvidin (0.2 mg/mL) (ThermoFisher). Apo-Cas9$^{FRET\text{-}R}$ (100 pM) was then surface-immobilized and the imaging chamber was rinsed with T50 buffer (10 mM Tris HCl, pH 8.0, 50 mM NaCl) before addition of imaging buffer (35 mM HEPES, pH 7.5, 75 mM KCl, 2 mM Trolox, 1 mM EDTA, 1 mM DTT, 0.1 mg/mL BSA, 0.8% w/v d-glucose, 1 mg/mL glucose

oxidase (Sigma), 0.004 mg/mL catalase (Calbiochem). Single-molecule movies were then recorded at 25 °C and 2 frames/s using an Andor iXon 897 EMCCD over several minutes. Notably, rapid photobleaching of $Cas9^{FRET\text{-}R}$ limited the temporal resolution of our smFRET experiments, but several short movies recorded at faster frame rate (10 frames/s) did not reveal additional Cas9 dynamics (data not shown). Movies were subsequently processed using IDL (ITT Visual Information Solutions) and custom MATLAB scripts (Center for the Physics of Living Cells, University of Illinois Urbana-Champaign) to map single molecules in the donor and acceptor emission channels.[9] The resulting donor and acceptor fluorescence intensity trajectories were background-corrected in MATLAB and used to calculate apparent FRET efficiency (E_{FRET}) by Eq. 7.5 (Figure 7.4E).

$$E_{FRET} = \frac{I_A}{I_D + I_A} \qquad \text{Eq. 7.5}$$

Molecules demonstrating anti-correlated donor and acceptor signals were collected and the FRET efficiencies were binned to generate FRET population distribution histograms (Figure 7.4F). Molecules demonstrating only donor signal as a result of incomplete fluorophore labeling were removed from analysis using a threshold of $E_{FRET} = 0.2$. Signal following donor or acceptor photobleaching, which resulted in a zero-FRET efficiency state, was removed from each single-molecule trajectory.

7.4.7 Stopped-flow fluorescence measurement of Cas9FRET-R conformational dynamics during sgRNA binding.

Cy3 and Cy5 fluorescence emission were recorded at 37 °C following rapid mixing of $Cas9^{FRET\text{-}R}$ (25 nM) with varying concentrations of sgRNA (75 – 1,200 nM) in

the presence of $MgCl_2$ (6 mM) during excitation at 532 nm (Figure 7.4B).

7.4.8 Stopped-flow fluorescence measurement of reverse conformational changes of Cas9FRET-R•sgRNA during sgRNA dissociation.

$Cas9^{FRET-R}$ (50 nM) was pre-incubated with sgRNA (150 nM) for 30 min at 20 °C to form $Cas9^{FRET-R}$•sgRNA. Cy5 fluorescence emission was recorded at 37 °C following rapid mixing of $Cas9^{FRET-R}$•sgRNA (50 nM) with dCas9 trap (2,250 nM) in the presence of $MgCl_2$ (6 mM) during excitation at 532 nm (Figure 7.4D).

7.4.9 Measurement of DNA association kinetics to Cas9•sgRNA.

We directly measured the rates of DNA cleavage to be fast (Figures 7.15A and 7.15B) by preforming the Cas9•sgRNA•DNA ternary complex before addition of catalytic metal ion (*i.e.* removing kinetics of DNA association from observation). It was thus possible to directly monitor the relatively slower kinetics of DNA association by instead monitoring the rates of DNA cleavage upon mixing various concentrations of Cas9•sgRNA with the DNA substrate in the presence of catalytic metal ion (*i.e.* DNA association is now rate-limiting under these conditions). Accordingly, Cas9 was pre-incubated with a 3-fold molar excess of sgRNA for 30 min at 20 °C to form the Cas9•sgRNA binary complex in the presence of $MgCl_2$ (6 mM). Next, Cas9•sgRNA (50 – 1,000 nM) was rapidly mixed at 37 °C with DNA substrate (10 nM) that was 5′-radiolabeled on both tDNA and ntDNA. Reactions were terminated at various times (0.003 – 60 s) by the addition of EDTA (0.37 M) (example of a gel image in Figure 7.11A). Rates were then extracted from each time course (Figures 7.7D and 7.11B) and plotted against the corresponding concentration of Cas9•sgRNA (Figures 7.7E and

7.11C).

7.4.10 Measurement of Mg^{2+} concentration dependence of DNA association and cleavage kinetics.

The effect of Mg^{2+} concentration on DNA association (Figures 7.12A and 7.12C) and cleavage kinetics (Figures 7.12E and 7.12G) was measured with assays similar to those in Figure 7.7D and Figure 7.15, respectively. For DNA association kinetics, Cas9•sgRNA (250 nM) was mixed with DNA (10 nM) in the presence of various concentrations of $MgCl_2$ (50 – 5000 μM). Reactions were terminated at various times (0.05 – 60 s) by the addition of EDTA (0.37 M). Rates were then extracted from each time course (Figures 7.12A and 7.12C) and plotted against the corresponding concentration of $MgCl_2$ (Figures 7.12B and 7.12D). For DNA cleavage kinetics, pre-formed Cas9•sgRNA•DNA (10 nM) was mixed with various concentrations of $MgCl_2$ (50 – 5000 μM). Reactions were terminated at various times (0.05 – 60 s) by the addition of EDTA (0.37 M). Rates were then extracted from each time course (Figures 7.12E and 7.12G) and plotted against the corresponding concentration of $MgCl_2$ (Figures 7.12F and 7.12H).

7.4.11 Stopped-flow fluorescence measurement of conformational changes upon DNA binding to $Cas9^{FRET-D}$•sgRNA.

$Cas9^{FRET-D}$ (50 nM) was pre-incubated with sgRNA (150 nM) for 30 min at 20 °C to form the $Cas9^{FRET-D}$•sgRNA binary complex in the presence or absence of $MgCl_2$ (6 mM). $Cas9^{FRET-D}$•sgRNA (50 nM) was then mixed with varying concentrations of DNA (250 – 1,300 nM) at 37 °C and Cy5 fluorescence was recorded during excitation at 532

nm (Figure 7.7B).

7.4.12 Measurement of DNA dissociation kinetics from Cas9•sgRNA•DNA.

Wild-type Cas9 (100 nM) and sgRNA (300 nM) were pre-incubated for 30 min at 20 °C to form Cas9•sgRNA. Next, Cas9•sgRNA (100 nM) was pre-incubated with DNA substrate (10 nM) 5′-radiolabeled on both tDNA and ntDNA for 1 hr at 20 °C to form the Cas9•sgRNA•DNA ternary complex (10 nM) in the absence of $MgCl_2$. Cas9•sgRNA•DNA (10 nM) was then mixed with dCas9•sgRNA trap (1,000 nM) (pre-formed by incubation of dCas9 (1000 nM) and sgRNA (3000 nM) for 30 min at 20 °C) at 37 °C for varying amounts of time (1 – 600 min) before addition of $MgCl_2$ (6 mM) for 15 s to initiate DNA cleavage prior to quenching the reaction using EDTA (0.37 M) (Figure 7.14A). During the variable incubation period, any 5′-radiolabeled DNA to dissociate from Cas9•sgRNA•DNA was rapidly bound by dCas9•sgRNA trap and thus protected from wild-type Cas9-catalyzed cleavage (Figure 7.13B).

7.4.13 Stopped-flow fluorescence measurement of conformational changes upon DNA dissociation from Cas9•sgRNA•DNA.

$Cas9^{FRET\text{-}D}$ (25 nM) was pre-incubated with sgRNA (75 nM) for 30 min at 20 °C to form the $Cas9^{FRET\text{-}D}$•sgRNA binary complex. Next, $Cas9^{FRET\text{-}D}$•sgRNA (25 nM) was pre-incubated with DNA substrate (250 nM) for 1 hr at 20 °C to form $Cas9^{FRET\text{-}D}$•sgRNA•DNA. Cy5 fluorescence was recorded upon rapid mixing of $Cas9^{FRET\text{-}D}$•sgRNA•DNA (25 nM) with dCas9•sgRNA trap (2,250 nM) in the absence of $MgCl_2$ during excitation at 532 nm (Figure 7.14B).

7.4.14 Single-turnover kinetic assays of Cas9-catalyzed DNA cleavage.

Cas9 (100 nM) and sgRNA (300 nM) were pre-incubated for 30 min at 20 °C to form Cas9•sgRNA. Next, Cas9•sgRNA (100 nM) was pre-incubated at 20 °C for 1 hr with DNA substrate (10 nM) 5′-radiolabeled on both tDNA and ntDNA to form Cas9•sgRNA•DNA in the absence of $MgCl_2$. Cas9•sgRNA•DNA (10 nM) was then rapidly mixed with $MgCl_2$ (6 mM) at 37 °C to initiate strand scission for varying amounts of time before quenching with EDTA (0.37 M) (Figures 7.15A, 7.15B, 7.5B, 7.5C, and 7.5E).

7.4.15 Stopped-flow fluorescence measurement of $Cas9^{FRET\text{-}D}$ conformational dynamics during DNA cleavage.

$Cas9^{FRET\text{-}D}$ (300 nM) and sgRNA (900 nM) were pre-incubated for 30 min at 20 °C to form $Cas9^{FRET\text{-}D}$•sgRNA. Next, $Cas9^{FRET\text{-}D}$•sgRNA (300 nM) was pre-incubated at 20 °C for 1 hr with DNA substrate (1000 nM) to form $Cas9^{FRET\text{-}D}$•sgRNA•DNA in the absence of $MgCl_2$. Cy5 fluorescence was recorded after mixing $Cas9^{FRET\text{-}D}$•sgRNA•DNA (300 nM) with $MgCl_2$ (6 mM) at 37 °C during excitation at 532 nm (Figure 7.15C).

7.4.16 Measurement of first and subsequent enzymatic turnovers by burst kinetic assay.

Cas9 (10 nM) and sgRNA (30 nM) were pre-incubated for 30 min at 20 °C to form Cas9•sgRNA. Next, Cas9•sgRNA (10 nM) was pre-incubated at 20 °C for 1 hr with an excess of DNA substrate (50 nM) 5′-radiolabeled on both tDNA and ntDNA to form Cas9•sgRNA•DNA in the absence of $MgCl_2$. This solution was then rapidly mixed with $MgCl_2$ (6 mM) at 37 °C to initiate strand scission for varying amounts of time before

quenching with EDTA (0.37 M) (Figure 7.16A). The very slow multiple-turnover rate (k_7, Figure 7.2A) resulted in low product concentration compared to substrate concentration during assays wherein radiolabeled DNA substrate was in excess over Cas9•sgRNA (Figure 7.16) which caused a dynamic range issue for our phosphorimaging detection method making it inaccurate for determining precise total RuvC product concentrations (Figure 7.19). Moreover, in contrast to HNH, RuvC generated more than one product (Figure 7.16) causing the radioactive signal of the products to be spread thin relative to the single substrate band, and thus complicated accurate determination of RuvC product concentrations during multiple-turnover assays (Figure 7.19). However, results from the single-turnover assays (Figures 7.15A, 7.15B, 7.5B, 7.5C, and 7.5E) prescribe that the reaction amplitudes should reach similar values for both HNH and RuvC active sites. Accordingly, to correct for the detection artifact during quantification of RuvC products from multiple-turnover assays, the RuvC burst amplitude was normalized to the HNH burst amplitude (Figure 7.17) as HNH was more accurately quantified in these multiple-turnover assays (*i.e.* a single, distinct product band). Importantly, the extracted rates of RuvC-catalyzed DNA cleavage were not affected by this correction.

7.4.17 Measurement of Cas9-catalyzed steady-state DNA cleavage kinetics.

Cas9 (10 nM) and sgRNA (30 nM) were pre-incubated for 30 min at 20 °C to form Cas9•sgRNA. Next, Cas9•sgRNA (10 nM) was mixed with an excess of DNA substrate (50 nM) 5′-radiolabeled on either tDNA or ntDNA in the presence of $MgCl_2$ (6 mM) for varying amounts of time before quenching with EDTA (0.37 M) (Figures 7.17B,

7.16B, and 7.16C). The RuvC reaction amplitude (Figure 7.17B) was corrected in the same way as for the burst kinetic assay (Figure 7.17A) as mentioned above.

7.4.18 Plasmid DNA cleavage assays.

A plasmid target DNA substrate was generated by sub-cloning the 20-nucleotide DNA target of the sgRNA (Figure 7.1) into pCDNA3.1 (ThermoFisher Scientific) between the NdeI and BamHI restriction sites. For Figure 7.18A, plasmid target DNA (10 nM) was pre-incubated with Cas9•sgRNA (100 nM), PvuI (10 U), or a mixture of Cas9•sgRNA (100 nM) and PvuI (10 U) prior to mixing with $MgCl_2$ (6 mM) for 1 hr at 37 °C. For Figure 7.18B, plasmid target DNA (10 nM) was linearized with PvuI and subsequently gel-purified to remove PvuI prior to pre-incubation with Cas9•sgRNA (100 nM), and mixing with $MgCl_2$ (6 mM) for 5 hr at 37 °C. Reaction products were resolved using non-denaturing 0.8% agarose-TAE (40 mM tris pH 8, 20 mM acetic acid, 1 mM EDTA) gel electrophoresis (Figure 7.18). Where indicated, samples were treated with 1% SDS prior to agarose gel electrophoresis. DNA bands were visualized by ethidium bromide staining (0.5 μg/mL).

7.5 Conclusion

Our establishment of a kinetic mechanism of Cas9 (Figure 7.2) has yielded insight into specific catalytic steps which may prove to be targetable for preventing collateral DNA cleavage at off-target sites during gene-editing. Importantly, using pre-steady-state kinetic assays, we were successful in dissecting the kinetics of transient steps of rapid DNA cleavage following formation of Cas9•sgRNA•DNA (Figure 7.2A, dashed box), which would not be possible by traditional methodologies. The assays presented here will

likely be useful in the development of more efficient and safer Cas9-based gene-editing platforms. For example, we are now evaluating our kinetic mechanism using other DNA targets and guide-RNA sequences as well as investigating the molecular basis and kinetic influence of mismatches between the guide-RNA and tDNA on each mechanistic step to better understand how Cas9 binds and cleaves off-target DNA. We anticipate that DNA targets containing guide-RNA mismatches may affect the kinetics of DNA binding or DNA cleavage (*i.e.* inhibition of individual domain dynamics) depending on the number of mismatches and their proximity to the PAM sequence, and we can directly measure these effects. Moreover, as we found that tDNA and ntDNA cleavage is regulated by the kinetics of HNH and RuvC domain motions, we expect that off-target DNA cleavage can be limited by modulating Cas9 conformational dynamics through protein-engineering efforts. Furthermore, it will be interesting to determine if recently discovered anti-CRISPR proteins[481-483] function by means of inhibition of Cas9 conformational dynamics or by preventing DNA recognition and binding.

7.6 Figures

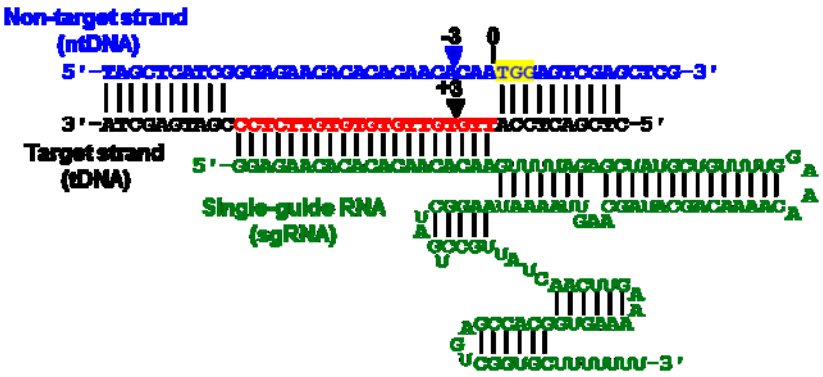


Figure 7.1: Schematic of the sgRNA•DNA heteroduplex.

PAM sequence on ntDNA (44-mer) highlighted in yellow and 20-nucleotide complementarity of tDNA (40-mer) to sgRNA colored red. Expected (canonical) cleavage sites of HNH and RuvC cleavage sites are marked by black and blue triangles, respectively.

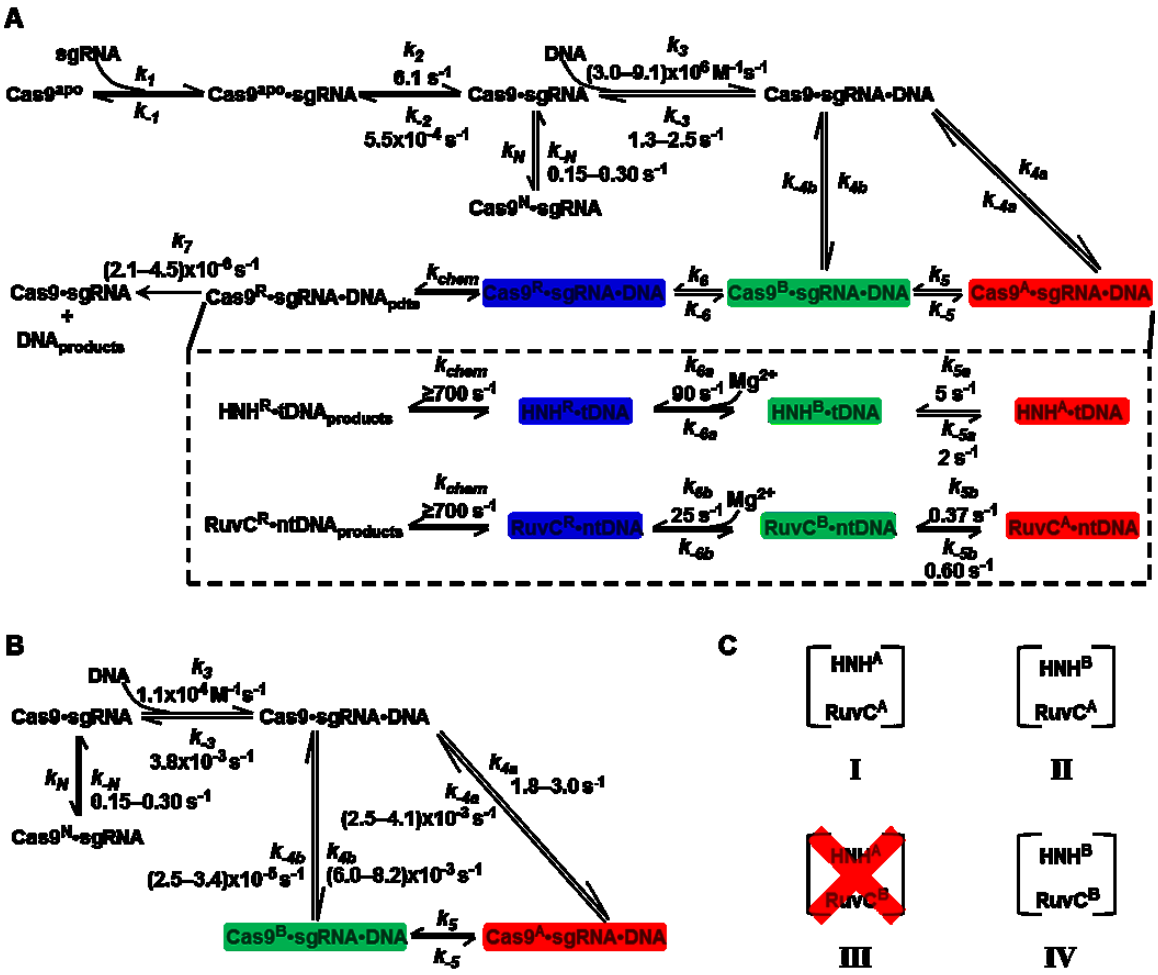


Figure 7.2: Mechanistic basis for Cas9 catalysis.

(A) Proposed kinetic mechanism for Cas9 complex formation and DNA cleavage in the presence of Mg^{2+}. The dashed box depicts kinetic contribution of conformational heterogeneity of the HNH and RuvC domains to DNA cleavage. The depiction of Mg^{2+} at certain steps denotes the Mg^{2+}-binding-coupled conformational changes of HNH and RuvC, rather than second-order metal ion binding. (B) Proposed kinetic mechanism for Cas9•sgRNA binding to DNA in the absence of Mg^{2+}. (C) Possible configurations of

(continued)

Figure 7.2: continued

HNH and RuvC domains of Cas9 ternary complexes pre-equilibrated in the absence of Mg^{2+}

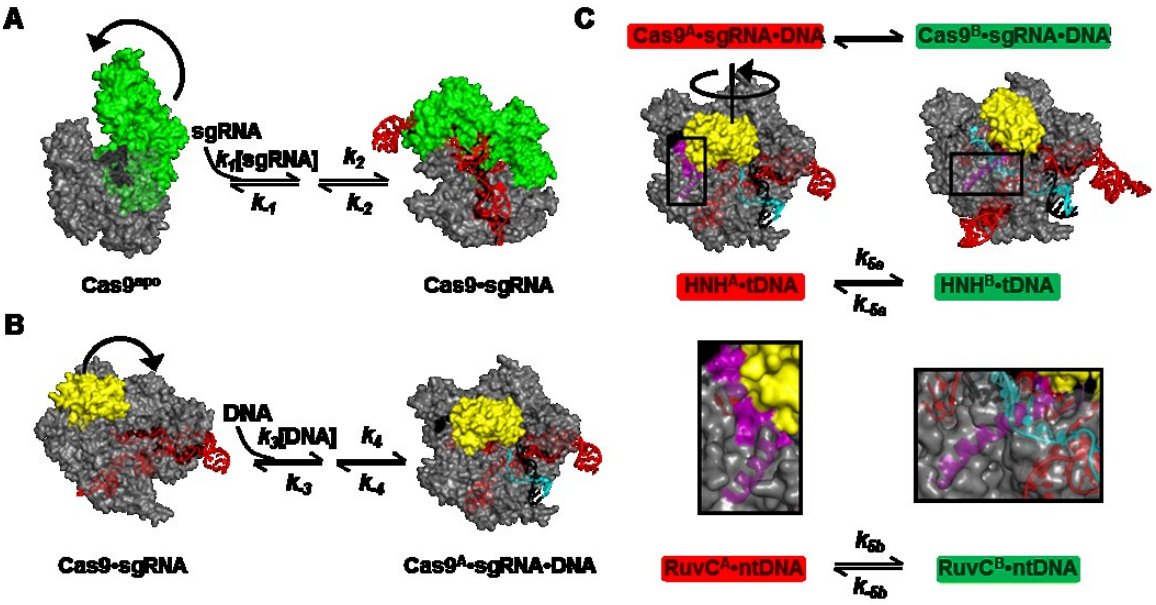


Figure 7.3: Structural bases of observed kinetic events for CRISPR/Cas9.

(A) Structural basis for conformational changes and kinetics observed for Cas9 binding to sgRNA (red). The REC lobe (green) undergoes a slow restructuring upon Cas9 binding to sgRNA (red). (B) Structural basis for conformational changes and kinetics observed for Cas9•sgRNA binding to DNA (tDNA and ntDNA show in black and cyan, respectively). Upon binding of Cas9•sgRNA to DNA, the HNH domain (yellow) swings inward with a fast hinge-motion. (C) Structural basis for early- (Cas9^{A}•sgRNA•DNA) and late-stage (Cas9^{B}•sgRNA•DNA) Cas9 ternary complexes. After the hinge-motion illustrated in Figure 7.3B, the HNH domain (yellow) rotates ~180° and moves ~20 Å to position the HNH active site proximal to the scissile bond on tDNA (black). Additionally, two short helices in the RuvC domain (residues 909-940, purple) are shown to restructure into a single, extended helix. ntDNA is shown in cyan.

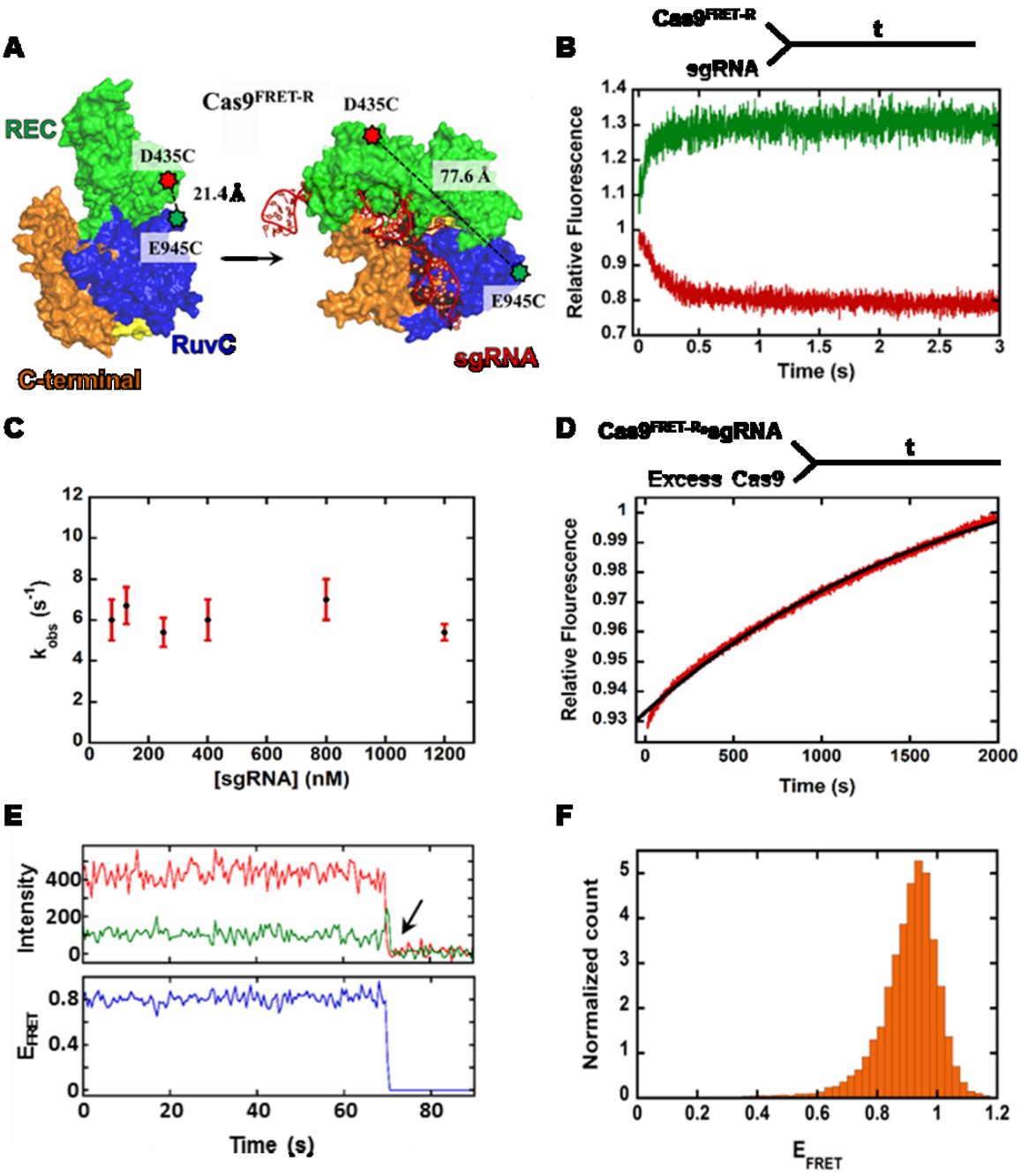


Figure 7.4: Kinetics and dynamics of sgRNA binding to Cas9.

(A) FRET system ($Cas9^{FRET-R}$) to monitor Cas9 conformational change during transition

(continued)

Figure 7.4: continued

from the apo state (PDB code: 4CMP) to the sgRNA-bound state (PDB code: 4ZT0). (B) Representative stopped-flow FRET trace of $Cas9^{FRET\text{-}R}$ (25 nM) conformational change upon sgRNA binding. FRET donor and acceptor signals are shown in green and red, respectively, and were fit to Eq. 7.1. (C) Observed rate constants (k_{obs}) of $Cas9^{FRET\text{-}R}$ (25 nM) conformational change at various concentrations of sgRNA (75 – 1,200 nM). Error bars (red) are standard deviations calculated from five independent experiments. (D) Acceptor fluorescence (red) of $Cas9^{FRET\text{-}R}$•sgRNA (50 nM) upon mixing with a large molar excess of unlabeled Cas9 trap (2,250 nM). The data were fit to Eq. 7.1 (solid black line). (E) Representative single-molecule trajectory of immobilized apo-$Cas9^{FRET\text{-}R}$ depicting high-FRET until donor photobleaching (black arrow). Donor (green) and acceptor (red) fluorescence intensities were used to calculate FRET (blue) using Eq. 7.5. (F) Population distribution histogram featuring FRET efficiencies (E_{FRET}) from multiple (n = 208) single-molecule trajectories of immobilized apo-$Cas9^{FRET\text{-}R}$.

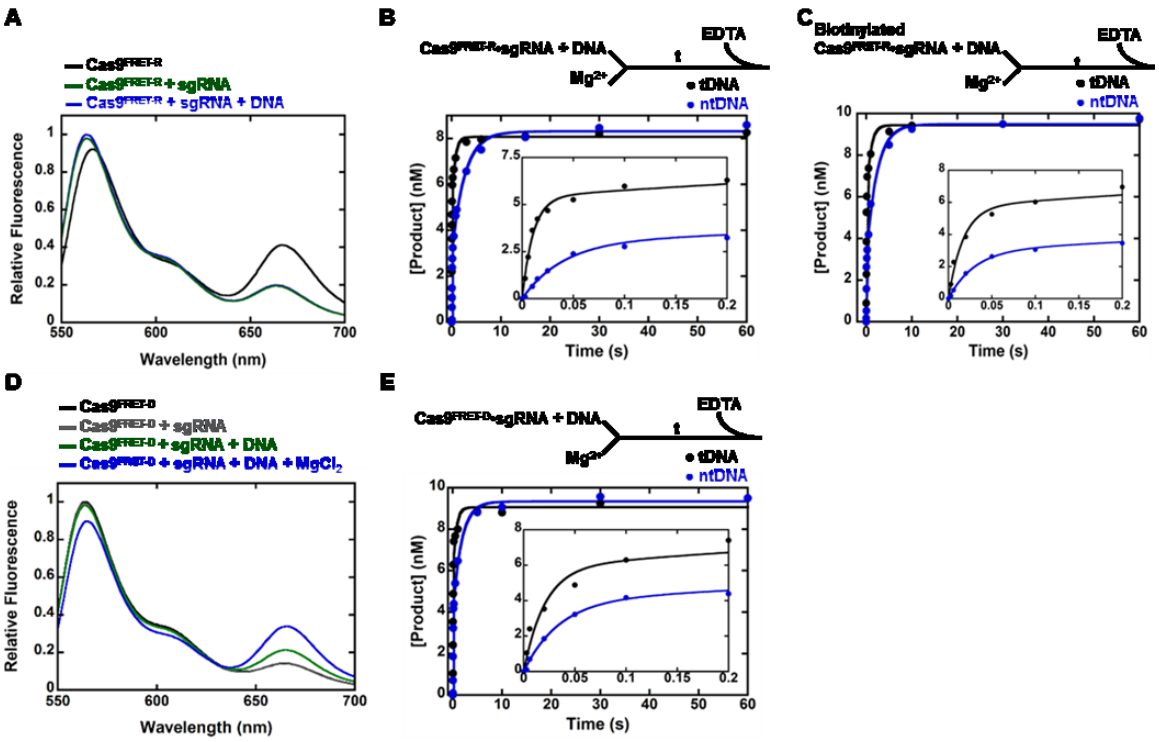


Figure 7.5: Cas9^{FRET-R} and Cas9^{FRET-D} control experiments.

(A) Fluorescence emission spectra of Cas9^{FRET-R} (200 nM) collected before and following sequential addition of sgRNA (300 nM), and DNA (500 nM). (B) Rapid cleavage of DNA (10 nM) by Cas9^{FRET-R}•sgRNA (100 nM). HNH (tDNA, ●) and RuvC (ntDNA, ●) products were fit to Eq. 7.2 (solid lines). The inset features early time points. Fast and slow DNA cleavage kinetics for HNH (k_{fast} = 66 ± 16 s^{-1}, 57% amplitude; k_{slow} = 1.7 ± 0.6 s^{-1}, 43% amplitude) and RuvC (k_{fast} = 29 ± 4 s^{-1}, 43% amplitude; k_{slow} = 0.57 ± 0.08 s^{-1}, 67% amplitude) were similar to those measured for wild-type Cas9•sgRNA (Figure 7.15B). (C) Rapid cleavage of DNA (10 nM) by biotinylated Cas9^{FRET-R}•sgRNA. HNH (tDNA, ●) and RuvC (ntDNA, ●) products were fit to Eq. 7.2 (solid lines). The inset

(continued)

Figure 7.5: continued

features early time points. Fast and slow DNA cleavage kinetics for HNH (k_{fast} = 60 ± 10 s^{-1}, 61% amplitude, k_{slow} = 1.2 ± 0.3, 39% amplitude) and RuvC (k_{fast} = 34 ± 6 s^{-1}, 31% amplitude, k_{slow} = 0.44 ± 0.05 s^{-1}, 69% amplitude) were similar to those measured for wild-type Cas9•sgRNA (Figure 7.15B). (D) Fluorescence emission spectra of $Cas9^{FRET\text{-}D}$ (200 nM) collected before and following sequential addition of sgRNA (300 nM), DNA (500 nM), and $MgCl_2$ (6 mM). (E) Rapid cleavage of DNA (10 nM) by $Cas9^{FRET\text{-}D}$•sgRNA (100 nM). HNH (tDNA, ●) and RuvC (ntDNA, ●) product concentration versus time were fit to Eq. 7.2 (solid lines). The inset features early time points. Fast and slow DNA cleavage kinetics for HNH (k_{fast} = 93 ± 8 s^{-1}, 67% amplitude; k_{slow} = 1.6 ± 0.3 s^{-1}, 33% amplitude) and RuvC (k_{fast} = 28 ± 5 s^{-1}, 36% amplitude; k_{slow} = 0.39 ± 0.05 s^{-1}, 64% amplitude) were similar to those of wild-type Cas9•sgRNA (Figure 7.15B). In Figures 7.5B, 7.5C, and 7.5E product formation by RuvC was taken as the sum of all ntDNA products.

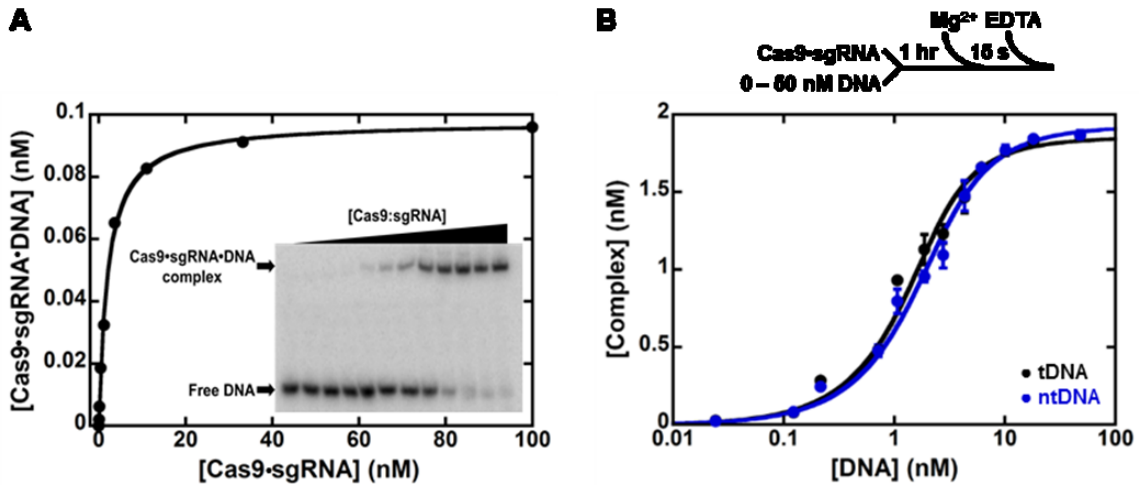


Figure 7.6: Equilibrium DNA binding of Cas9•sgRNA.

(A) Electrophoretic mobility shift assay to measure apparent equilibrium dissociation constant (K_d^{DNA}) of DNA from Cas9•sgRNA•DNA. 5′-radiolabeled DNA (0.1 nM) was incubated for 1 hr with increasing concentrations of Cas9•sgRNA (0.002 – 100 nM) before separation by native PAGE. The data were fit to Eq. 7.4 (solid line) to obtain K_d^{DNA} = 2 nM. Inset: gel image of EMSA showing Cas9•sgRNA binding to 5′-radiolabeled DNA. (B) Active site titration of HNH (tDNA, ●) and RuvC (ntDNA, ●). As the amplitude of the pre-steady-state burst phase (see Section 7.3.5; Figure 7.17A) reflects the concentration of Cas9•sgRNA•DNA at the start of the reaction, we could obtain apparent K_d^{DNA} and the active concentration of Cas9•sgRNA by examining the pre-steady-state kinetics of DNA cleavage at various concentrations of DNA. A pre-incubated solution of Cas9•sgRNA (2 nM) (formed by 30 min incubation of Cas9 with a 3-fold molar excess of sgRNA at 20 °C) and 5′-

(continued)

Figure 7.6: continued

radiolabeled DNA (0.05 – 50 nM) was mixed with $MgCl_2$ (6 mM) for 15 s prior to quenching with EDTA (0.37 M). [Cas9•sgRNA•DNA] was determined from the burst amplitude and plotted against total [DNA]. The data were fit to Eq. 7.4 (solid lines) to obtain K_d^{DNA} and percent activity for HNH (K_d^{DNA} = 1.85 nM, 93% active) and RuvC (K_d^{DNA} = 1.82 nM, 91% active) active sites. Product formation by RuvC was taken as the sum of all ntDNA products. Error bars are standard deviations calculated from three independent experiments.

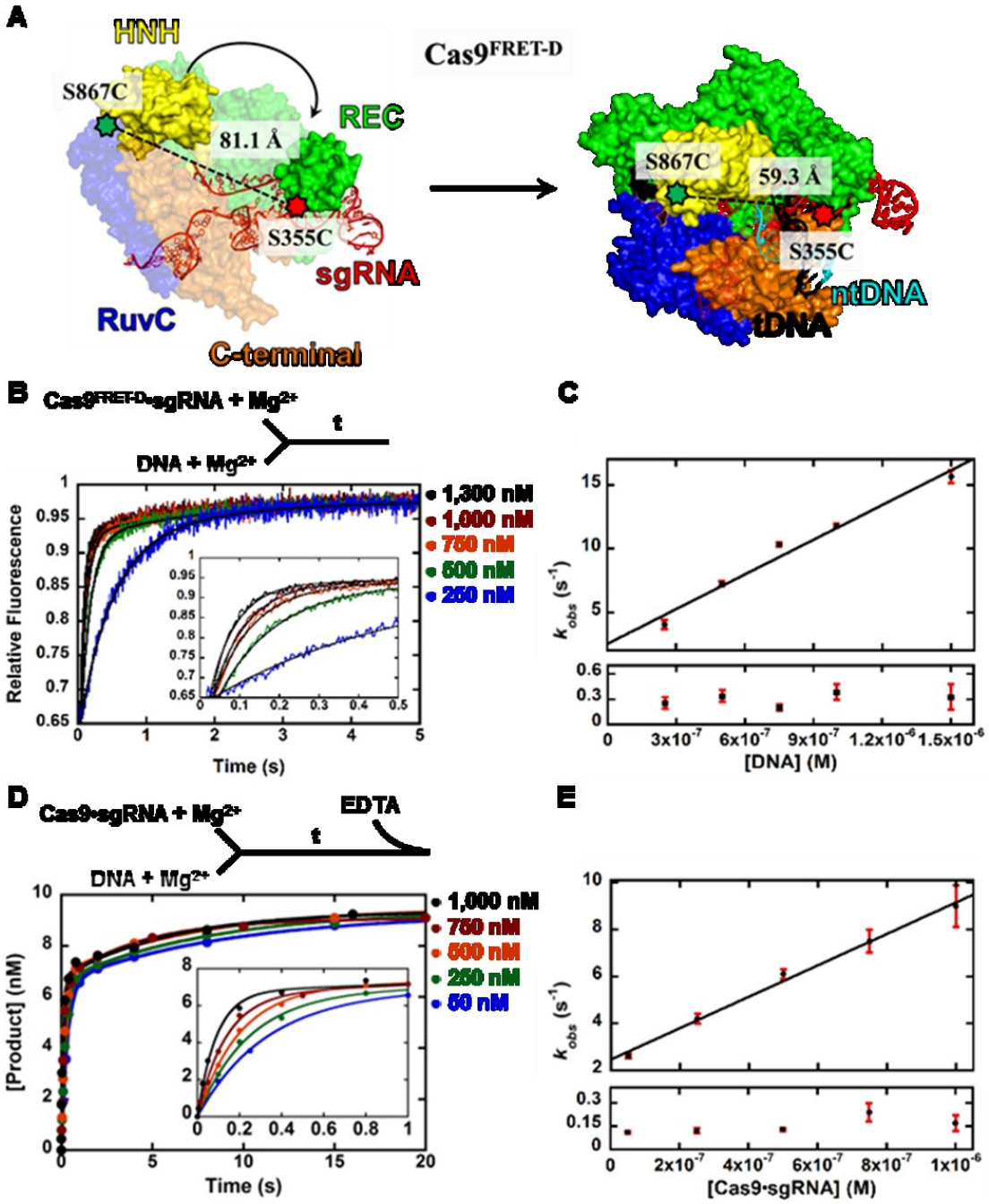


Figure 7.7: DNA association kinetics of Cas9.

(A) FRET system ($Cas9^{FRET-D}$) to monitor HNH domain motion during transition from

(continued)

Figure 7.7: continued

the binary state (PDB code: 4ZT0) to the ternary state (PDB code: 4UN3) of Cas9. DNA binding results in a ~20 Å translation of the HNH domain. (B) Acceptor fluorescence of $Cas9^{FRET\text{-}D}$•sgRNA (75 nM) upon binding to DNA (250 – 1,300 nM) measured by stopped-flow FRET and fit to Eq. 7.2 (solid lines). The inset features the fast exponential phase. (C) k_{obs} values from the fast phase of acceptor fluorescence (top panel) at each DNA concentration measured in Figure 7.7B were fit to a line and the slope yielded the bimolecular association rate constant of Cas9•sgRNA binding to DNA while the y-intercept yielded the DNA dissociation rate constant. k_{obs} values from the slow phase of acceptor fluorescence (bottom panel) at each concentration of DNA were not dependent on DNA concentration. Error bars (red) are standard deviations calculated from five independent experiments. (D) Association of Cas9•sgRNA (50 – 1,000 nM) to DNA (10 nM) measured through time-dependent cleavage of tDNA by HNH. HNH-catalyzed tDNA product formation at each concentration of Cas9•sgRNA was quantified and plotted against time. The data were fit to Eq. 7.2 (solid lines). A representative time course gel image (50 nM Cas9•sgRNA) is provided in Figure 7.11A. (E) k_{obs} values from the fast phase of tDNA cleavage (top panel) at each concentration of Cas9•sgRNA measured in Figure 7.7D were fit to a line and the slope yielded the bimolecular association rate constant of Cas9•sgRNA binding to DNA while the y-intercept yielded the DNA dissociation rate constant. k_{obs} values from the slow phase of product formation (bottom panel) at each concentration of Cas9•sgRNA were not dependent on Cas9•sgRNA concentration. Error bars (red) are fitting errors of Eq. 7.2.

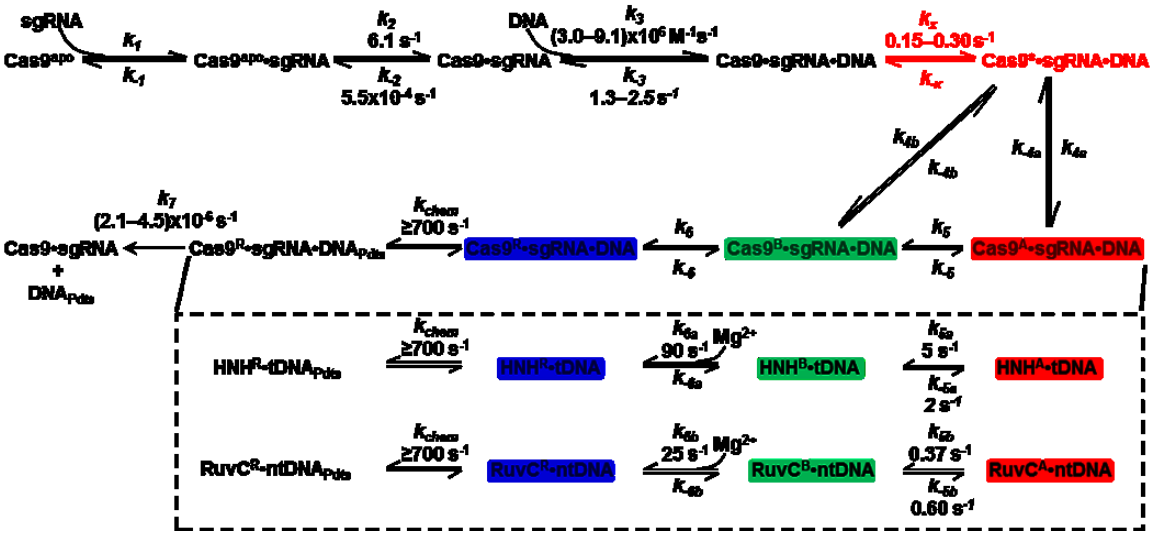


Figure 7.8: Potential competing kinetic mechanism of CRISPR/Cas9 with an addition step following DNA binding.

Steps that deviate from the mechanism in Figure 7.2A are in red. This mechanism was excluded as k_x was not rate-limiting for the entire population of Cas9•sgRNA•DNA during the product-dependent DNA association assay (Figure 7.7D). See discussion in Section 7.3.2.

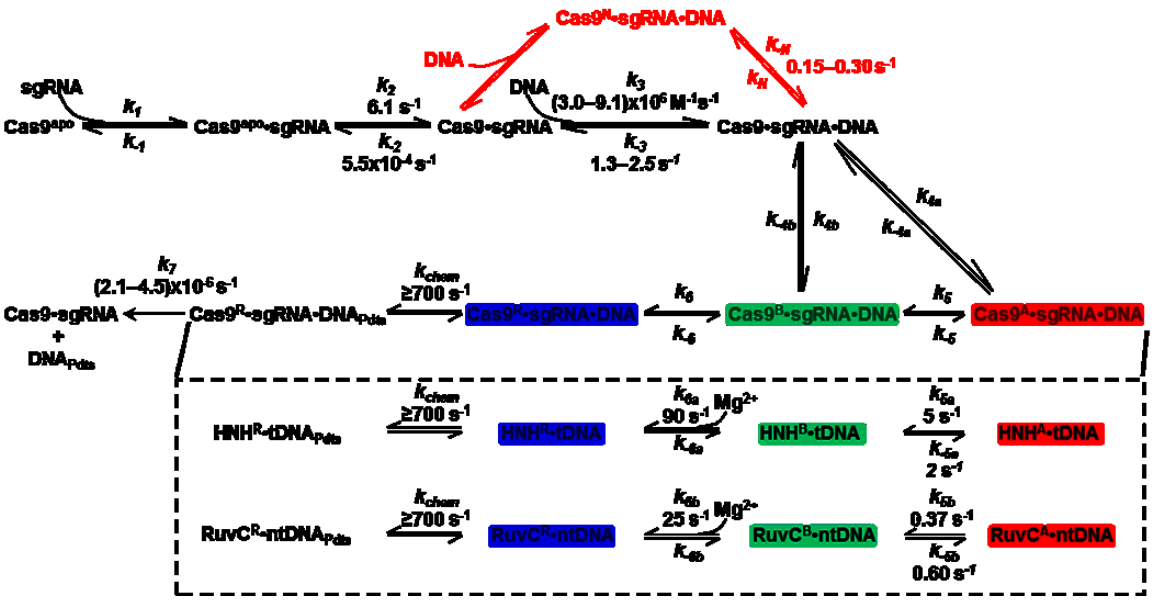


Figure 7.9: Potential competing kinetic mechanism of CRISPR/Cas9 with a non-productive complex (Cas9^{N}•sgRNA•DNA) formed during initial DNA collision. Steps that deviate from the mechanism in Figure 7.2A are in red. This mechanism was excluded as the slow phase k_{obs} values in Figures 7.7C, 7.7E, and 7.11C did not exhibit DNA- or Cas9•sgRNA-concentration dependence and were not affected by $MgCl_2$ concentration (Figures 7.7 and 7.12). See discussion in Section 7.3.3.

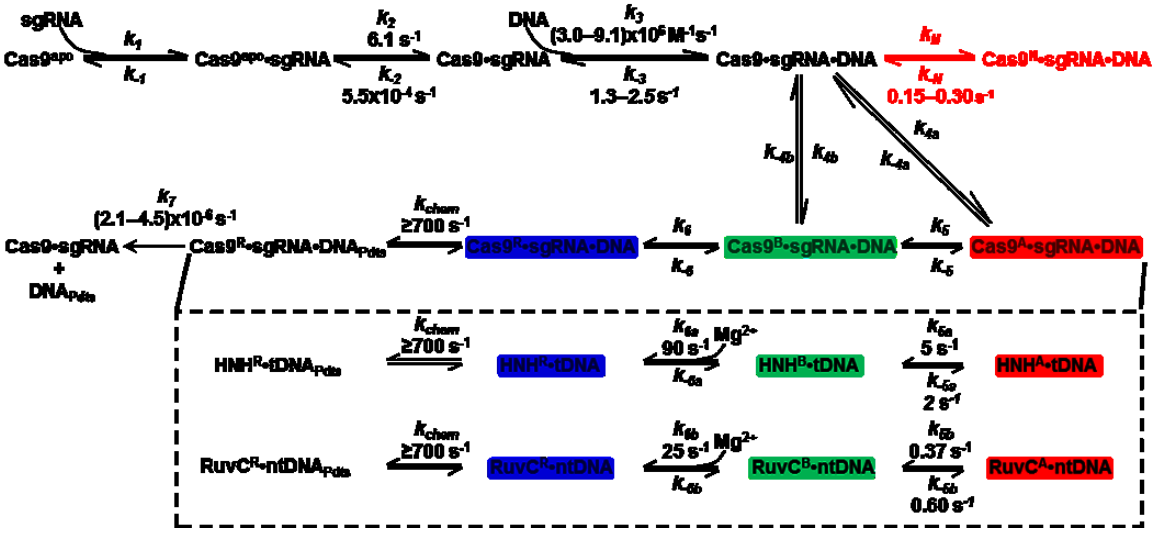


Figure 7.10: Potential competing kinetic mechanism of CRISPR/Cas9 with non-productive ternary complex (Cas9^{N}•sgRNA•DNA) occurring after initial DNA binding.

Steps that deviate from the mechanism in Figure 7.2A are in red. This mechanism was excluded as the slow phase k_{obs} values in Figures 7.7C, 7.7E, and 7.11C did not exhibit DNA- or Cas9•sgRNA-concentration dependence and were not affected by $MgCl_2$ concentration (Figures 7.7 and 7.12). See discussion in Section 7.3.3.

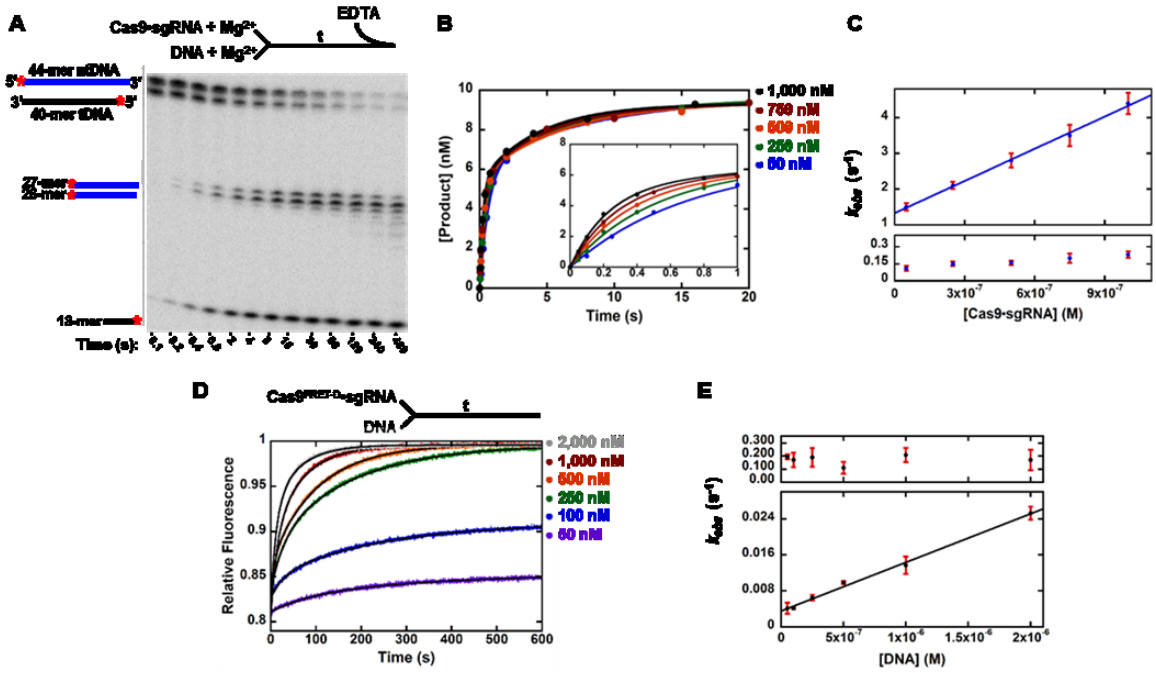


Figure 7.11: Additional DNA association data.

(A) Representative gel image showing cleavage of the DNA substrate (10 nM) upon mixing with a specific concentration of Cas9•sgRNA (50 nM). This experiment was repeated with increasing concentrations of Cas9•sgRNA (50 – 1,000 nM) and adjusted time courses in order to measure the association kinetics of Cas9•sgRNA to DNA. (B) Association of Cas9•sgRNA (50 – 1,000 nM) to DNA (10 nM) measured through RuvC-catalyzed, time-dependent cleavage of ntDNA. RuvC-catalyzed ntDNA product formation at each concentration of Cas9•sgRNA was quantified and plotted against time. The data were fit to Eq. 7.2 (solid lines). Product formation by RuvC was taken as the sum of all ntDNA products. The inset features early time points. (C) k_{obs} of fast phases of ntDNA cleavage (top panel) at each concentration of Cas9•sgRNA measured in Figure

(continued)

Figure 7.11: continued

7.11B were fit to a line and the slope yielded the bimolecular association rate constant of Cas9•sgRNA binding to DNA while the y-intercept yielded the DNA dissociation rate constant. k_{obs} of slow phases of ntDNA cleavage (bottom panel) were independent Cas9•sgRNA concentration. Error bars (red) are fitting errors of Eq. 7.2. (D) $Cas9^{FRET\text{-}D}$•sgRNA (150 nM) was mixed with varying concentrations of DNA (50 – 2,000 nM) in the absence of $MgCl_2$ and the acceptor fluorescence was recorded upon excitation at 532 nm. The data were fit to Eq. 7.2 (solid lines). (E) k_{obs} of slow fluorescence phases (bottom panel) from Figure 7.11D were plotted against the corresponding DNA concentrations and the data were fit to a line to obtain the second-order rate constant of 1.1×10^4 $M^{-1}s^{-1}$ for DNA binding to $Cas9^{FRET\text{-}D}$•sgRNA and the DNA dissociation rate constant of 3.8×10^{-3} s^{-1} in the absence of $MgCl_2$. k_{obs} of fast fluorescence phases (top panel) were independent DNA concentration. Error bars (red) are standard deviations calculated from five independent experiments.

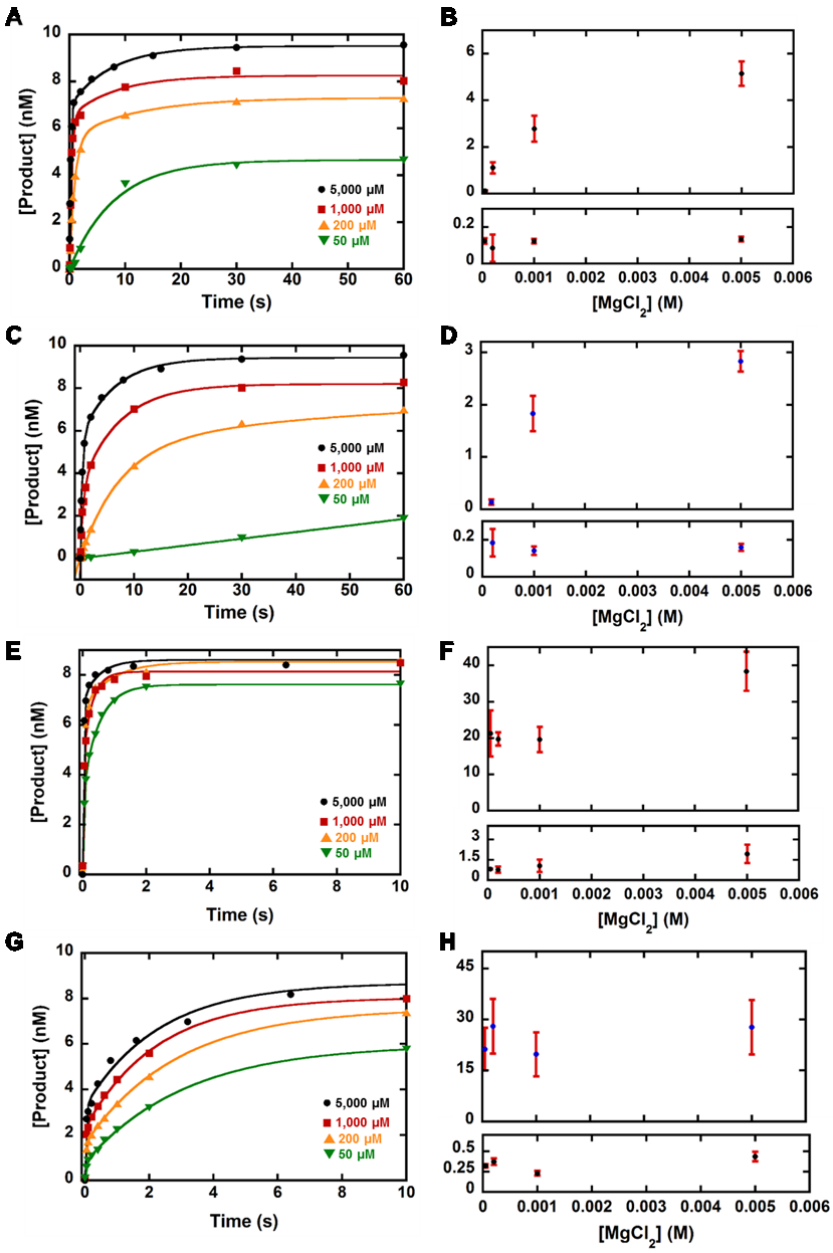


Figure 7.12: Mg^{2+} concentration dependence of DNA association kinetics.

Association of Cas9•sgRNA (250 nM) to DNA (10 nM) measured through time-dependent cleavage of tDNA by HNH (A) or ntDNA by RuvC (C) in the presence of

(continued)

Figure 7.12: continued

various concentrations of $MgCl_2$ (50 – 5000 μM). Reaction mixing scheme as in Figure 7.7D. HNH- or RuvC-catalyzed product formation at each concentration of $MgCl_2$ was quantified and plotted against time. The data were fit to Eq. 7.2 (solid lines). (B, D), k_{obs} values from the fast phases (top panels) and slow phases (bottom panels) of tDNA (B) and ntDNA (D) cleavage at each concentration of $MgCl_2$. While fast phase k_{obs} values showed strong dependence on $MgCl_2$ concentration, k_{obs} values from the slow phase of product formation (B and D, bottom panels) were not dependent on $MgCl_2$ concentration. Error bars (red) are fitting errors of Eq. 7.2. RuvC product formation at 50 μM was too low to be accurately fit to Eq. 7.1 or Eq. 7.2. (E, G) Time-dependent tDNA (E) and ntDNA (G) cleavage by pre-formed Cas9•sgRNA•DNA (10 nM) following mixing with various concentrations of $MgCl_2$ (50 – 5000 μM). Reaction mixing scheme as in Figure 7.15A. HNH- or RuvC-catalyzed product formation at each $MgCl_2$ concentration was quantified and plotted against time. The data were fit to Eq. 7.2 (solid lines). (F, H), k_{obs} values from the fast phases (top panels) and slow phases (bottom panels) of tDNA (F) and ntDNA (H) cleavage at each concentration of $MgCl_2$. k_{obs} values from the fast and slow phases of DNA cleavage from the pre-formed ternary complex showed no dependence on $MgCl_2$ concentration. Error bars (red) are fitting errors of Eq. 7.2.

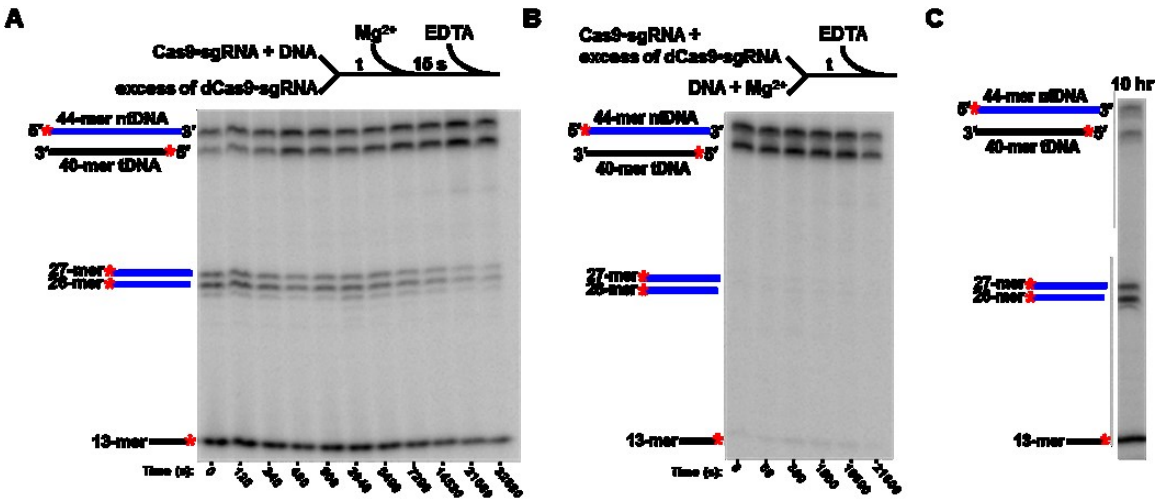


Figure 7.13: Gel image and control experiments for DNA dissociation assays.

(A) DNA dissociation assay wherein Cas9•sgRNA•5′-radiolabeled DNA (10 nM) was mixed with dCas9•sgRNA (1,000 nM) for 1 min to 10 hrs before initiating DNA cleavage for 15 s by the addition of $MgCl_2$ (6 mM). (B) Validation of dCas9•sgRNA as an effective trap. A solution of Cas9•sgRNA (20 nM) and catalytically inactive dCas9•sgRNA (1,000 nM) was mixed with 5′-radiolabeled DNA (10 nM) in the presence of $MgCl_2$ (6 mM) for 1 min to 6 h. The lack of DNA cleavage product accumulation indicates that the large excess of dCas9•sgRNA effectively trapped 5′-radiolabeled DNA, preventing its cleavage by the catalytically active Cas9•sgRNA. (C) A pre-incubated solution of Cas9•sgRNA (100 nM) and DNA (10 nM) was incubated at 37 °C for 10 hrs before initiating DNA cleavage for 15 s by the addition of $MgCl_2$ (6 mM). Robust cleavage of tDNA (87%) and ntDNA (84%) indicates that Cas9 retains nearly full activity after 10 hrs at 37 °C.

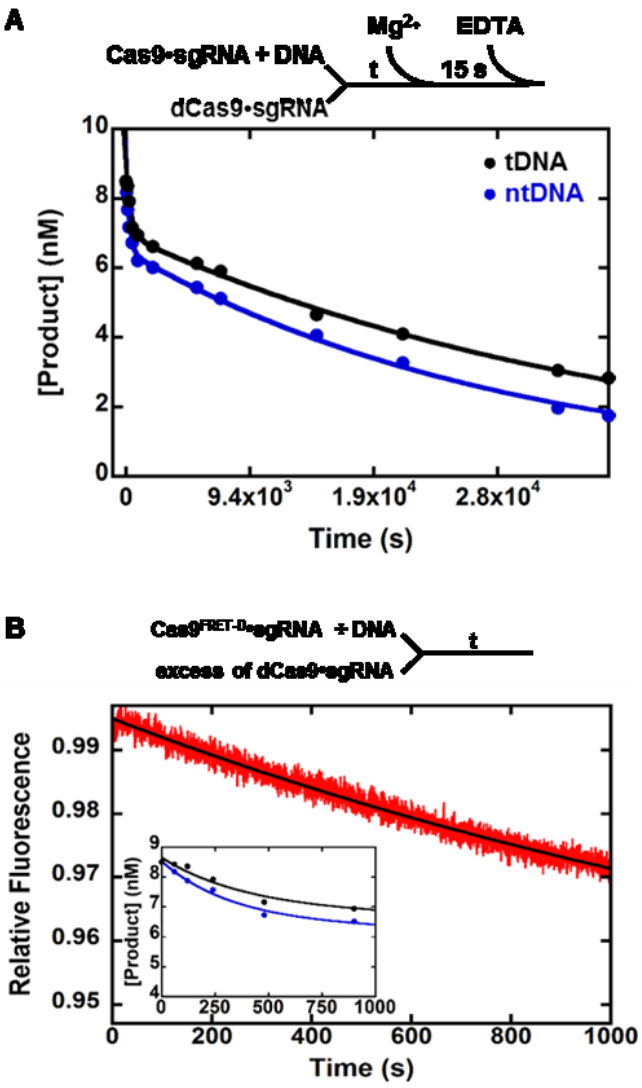


Figure 7.14: DNA dissociation kinetics of Cas9.

(A) DNA dissociation kinetics measured by respective HNH and RuvC cleavage of tDNA (●) and ntDNA (●) remaining in the Cas9•sgRNA•DNA ternary complex (10 nM) following incubation with a large excess of dCas9•sgRNA trap (1,000 nM) for various times (1 – 600 min). Data were fit to Eq. 7.2 (solid lines). (B) Acceptor fluorescence (red) (continued)

Figure 7.14: continued

of $Cas9^{FRET\text{-}D}$•sgRNA•DNA (50 nM) upon mixing with a large molar excess of dCas9•sgRNA (2,250 nM). The data were fit to Eq. 7.1 (black line). Inset features data from Figure 7.14A for comparison of the two plots within the same time range which was limited by the stopped-flow instrument for Figure 7.14B.

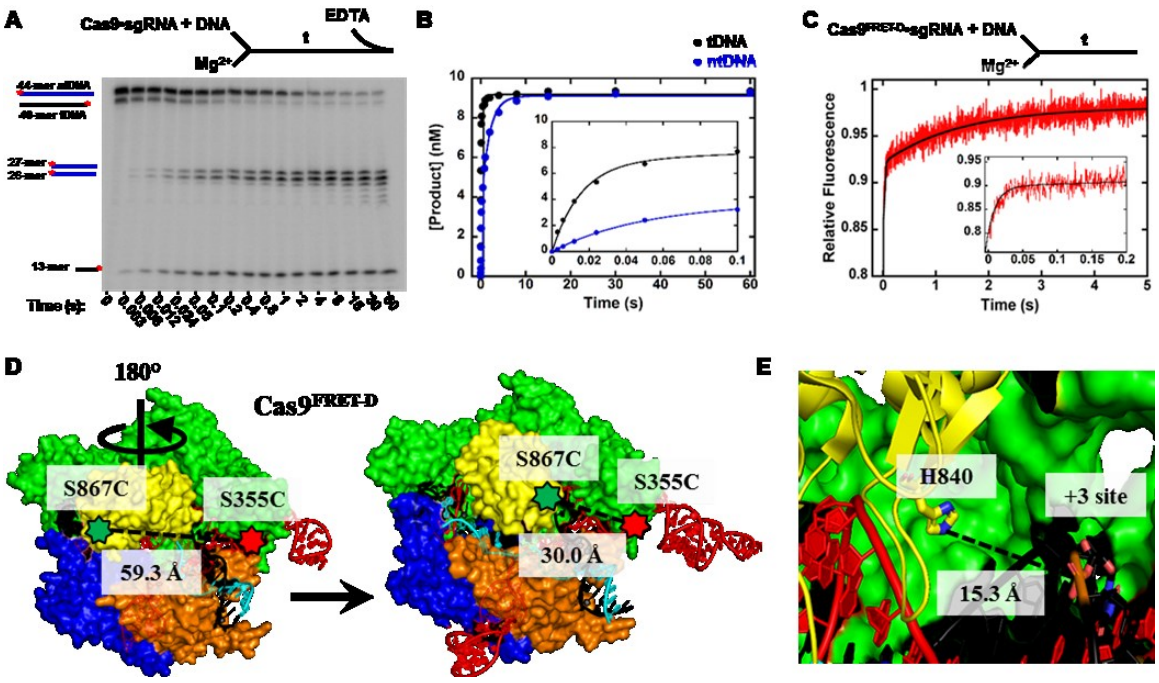


Figure 7.15: Kinetics and conformational dynamics of DNA cleavage by Cas9.

(A) Gel image showing time-dependent (0.003 – 60 s) cleavage of 5′-radiolabeled (*) DNA (10 nM) by Cas9•sgRNA (100 nM) from the pre-formed ternary complex following rapid mixing with $MgCl_2$ (6 mM). (B) Product formation by respective cleavage of tDNA (●) and ntDNA (●) by HNH and RuvC from Figure 7.15A were fit to Eq. 7.2 (solid lines). Product formation by RuvC was taken as the sum of all ntDNA products. Inset features early time points. (C) Acceptor fluorescence of $Cas9^{FRET\text{-}D}$•sgRNA•DNA (300 nM) upon mixing with $MgCl_2$ (6 mM) was fit to Eq. 7.2 (black line). Inset features data re-acquired within a shorter time frame. (D) Following initial DNA binding to form the ternary complex (PDB code: 4UN3), the HNH domain must rotate by 180° (PDB code: 5FR9) in order to bring catalytic residues near the scissile bond on tDNA. This rotation

(continued)

Figure 7.15: continued

brings the FRET probes of $Cas9^{FRET-D}$ nearer by ~30 Å. $Cas9^{FRET-D}$ is colored as in Figure 7.7A. (E) Zoomed view of the HNH active site in the late-stage ternary complex captured in the absence of $MgCl_2$ (PDB code: 5FR9) colored as in Figure 7.7A. The catalytic histidine (H840) is too far from the scissile phosphate at the + 3 position of tDNA to coordinate DNA cleavage.

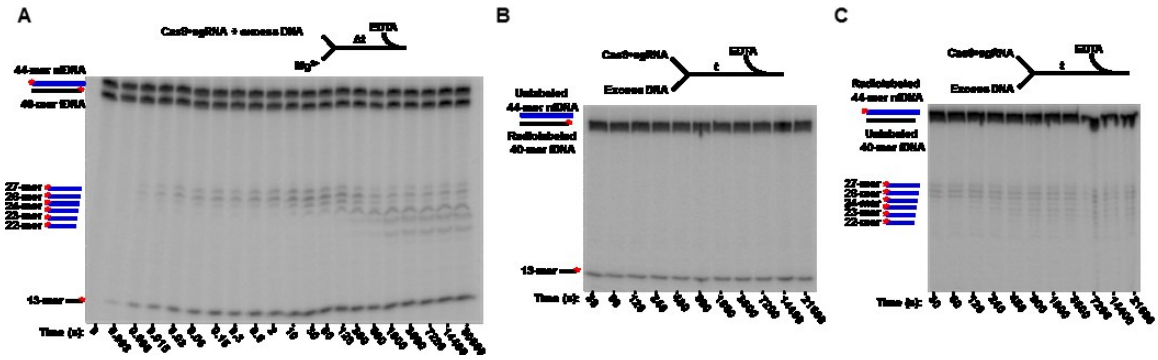


Figure 7.16: Gel images of steady-state and burst kinetic experiments.

(A) Gel image of burst kinetic assay from Figure 7.17A showing time-dependent (0.003 s – 8.5 hrs) cleavage of 5′-radiolabeled (*) DNA (50 nM) by Cas9•sgRNA (10 nM). (B and C) Gel images from the steady-state kinetic assay in Figure 7.17B. The DNA substrate (50 nM, Figure 7.1), which was 5′-radiolabeled on either tDNA (B) to visualize cleavage by HNH, or ntDNA (C) to visualize cleavage by RuvC, was mixed with Cas9•sgRNA (10 nM) in the presence of $MgCl_2$ (6 mM) for 30 s to 6 hrs.

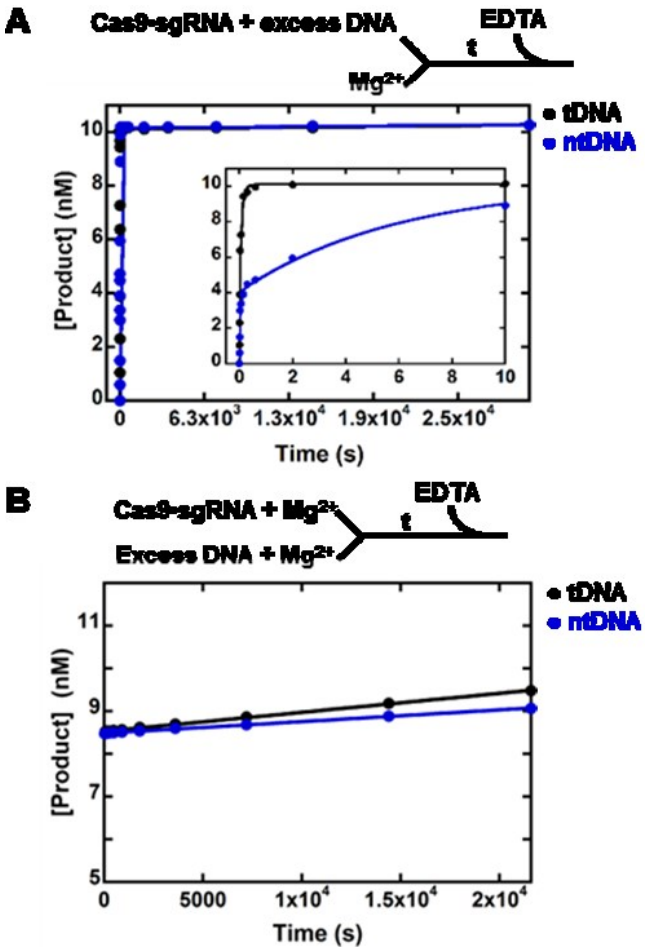


Figure 7.17: Burst and steady-state kinetics of DNA cleavage by Cas9.

(A) Burst assay (3 ms – 8.5 hrs) to monitor first and subsequent catalytic turnovers of Cas9•sgRNA (10 nM) cleaving 5′-radiolabeled DNA (50 nM). Product formation by respective cleavage of tDNA (●) and ntDNA (●) by HNH and RuvC from Figure 7.16A were fit to Eq. 7.3 (solid lines). Product formation by RuvC was taken as the sum of all ntDNA products. Inset features early time points of the first turnover. As expected, fast kinetics of the first turnover were biphasic. (B) Steady-state kinetics measured from

(continued)

Figure 7.17: continued

respective cleavage of tDNA (●) and ntDNA (●) by HNH and RuvC (Figures 7.16B and 7.16C, respectively) after mixing of 5′-radiolabeled DNA (50 nM) with Cas9•sgRNA (10 nM) in the presence of $MgCl_2$ (6 mM). The data were fit to a line.

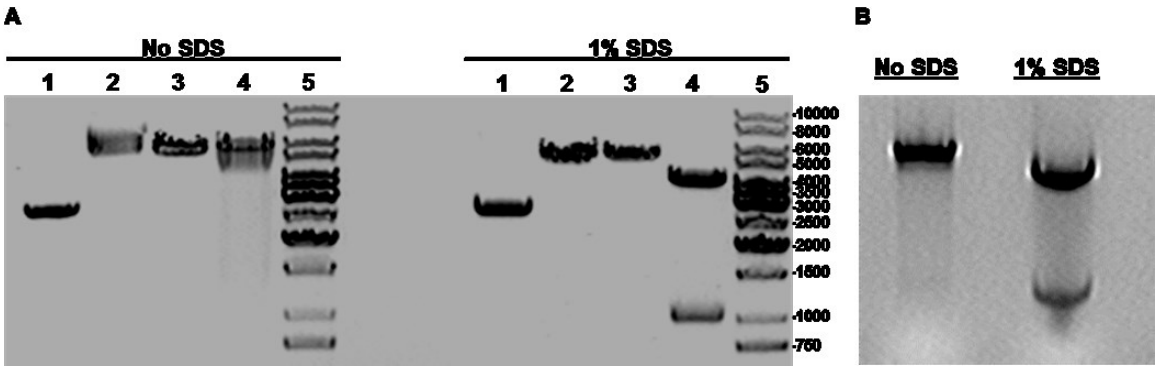


Figure 7.18: Cas9-catalyzed cleavage of plasmid DNA and tight association to plasmid DNA products.

(A) Gel image featuring Cas9- and PvuI-catalyzed cleavage of plasmid DNA. For each reaction, a DNA plasmid (10 nM) containing the same DNA target sequence and PAM as in Figure 7.1 was pre-incubated with either a solution of Cas9•sgRNA (100 nM) (Lane 2), PvuI (10 U) (Lane 3), or Cas9•sgRNA (100 nM) and PvuI (10 U) (Lane 4) prior to mixing with $MgCl_2$ (6 mM) for 1 hr at 37 °C. Reaction products were separated by 0.8% native agarose gel in the absence or presence of 1% SDS and visualized by ethidium bromide staining (0.5 μg/mL). Lane 1 features the plasmid DNA substrate before any treatment. Lane 5 features a 1 kb DNA ladder (GeneRuler, Thermo Fisher Scientific). (B) Gel image featuring Cas9-catalyzed cleavage of pre-linearized plasmid DNA. A plasmid DNA was digested with PvuI for 2 hr at 37 °C prior to agarose gel separation, extraction, and purification. The purified, linearized DNA plasmid (10 nM) was then pre-incubated with a solution of Cas9•sgRNA (100 nM) for 1 hr prior to mixing with $MgCl_2$ (6 mM) for 5 hrs at 37 °C. Reaction products were separated and visualized as in Figure 7.18A.

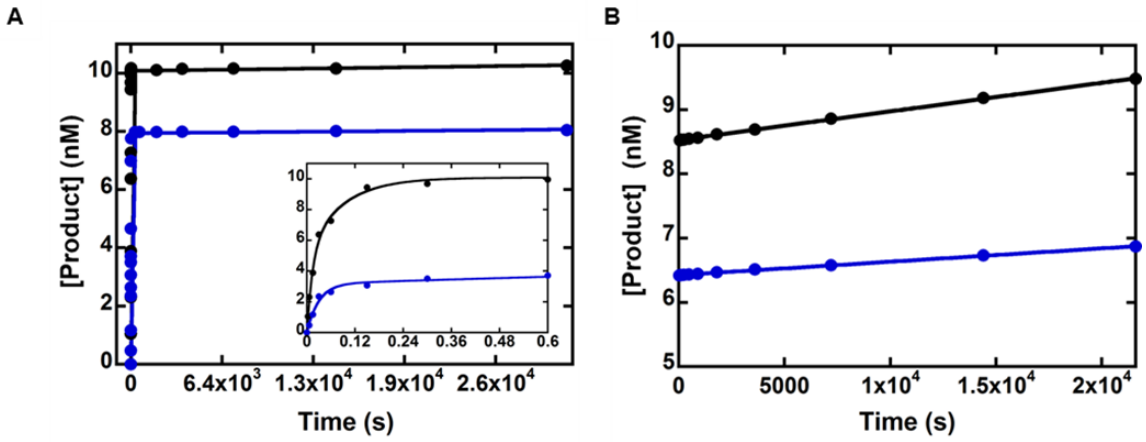


Figure 7.19: Burst and steady-state kinetic assay data before RuvC product normalization.

(A) Initial quantification of the burst kinetic assay in Figure 7.16A. As a result of inherently poor signal-to-noise caused by minimal DNA product signal relative to DNA substrate signal on gels featuring multiple-turnover kinetic assays of Cas9 (Figure 7.16), as well as the dilution of RuvC product radioactive signal across multiple bands due to the known RuvC-catalyzed 3′→5′ degradation of ntDNA product (Figure 7.16), initial quantification of RuvC product concentration yielded artifactually low values relative to HNH product concentration (~2 nM lower). (B) Initial quantification of the steady-state kinetic assay in Figures 7.16B and 7.16C. Similar to Figure 7.19A, inherently poor signal-to-noise and inaccurate quantification of multiple RuvC product bands resulted in seemingly lower RuvC product concentration relative to HNH product concentration (~2 nM lower). For A and B, product formation by RuvC was taken as the sum of all ntDNA products.

Chapter 8. Bidirectional Degradation of DNA Cleavage Products Catalyzed by CRISPR/Cas9

Author Contributions

Austin T. Raper and Anthony A. Stephenson equally contributed to this work as co-first authors. A.T.R conceived the research, prepared the protein, designed experiments, conducted experiments, interpreted data, developed the KinTek model, generated figures, and wrote the manuscript. A.A.S prepared RNA/DNA substrates, designed experiments, conducted experiments, interpreted data, developed the KinTek model, generated figures, and helped to write the manuscript. Dr. Zucai Suo provided insight into kinetic experiments and helped to edit the manuscript.

8.1 Abstract

Since the initial characterization of *Streptococcus pyogenes* CRISPR/Cas9 as a powerful gene-editing tool, it has been widely accepted that Cas9 generates blunt-ended DNA products by concerted cleavage of the target (tDNA) and non-target (ntDNA) strands three nucleotides away from the protospacer adjacent motif (PAM) by HNH and RuvC nuclease active sites, respectively. Following initial DNA cleavage, RuvC catalyzes 3′→5′ degradation of the ntDNA resulting in DNA products of various lengths. Here, we found that Cas9 selects multiple sites for initial ntDNA cleavage and preferentially generates staggered-ended DNA products containing single-nucleotide 5′-overhangs. We also quantitatively evaluated 3′→5′ post-cleavage trimming (PCT) activity of RuvC to find that ntDNA degradation continues up to the -10 position on the PAM distal DNA product and is kinetically significant when compared to extremely slow DNA product release. We also discovered a previously unidentified 5′→3′ PCT activity of RuvC which can shorten the PAM proximal ntDNA product by precisely one nucleotide with a comparable rate as the 3′→5′ PCT activity. Taken together, our results demonstrate that RuvC-catalyzed post-cleavage trimming ultimately generates DNA fragments with heterogeneous ends following initial DNA cleavage including a PAM proximal fragment with a blunt end and a PAM distal fragment with a staggered-end, 3′-recessed on the ntDNA strand. These kinetic and biochemical findings underline the importance of temporal control of Cas9 during gene-editing experiments and help explain the patterns of nucleotide insertions at sites of Cas9-catalyzed gene modification *in vivo*.

8.2 Introduction

The field of genetic engineering was revolutionized by the discovery of clustered regularly-interspaced short palindromic repeats (CRISPRs) and CRISPR-associated (Cas) proteins which can mediate RNA-guided targeting and cleavage of double-stranded DNA.[279,282,284,286,301,472,484-486] The CRISPR/Cas system of *Streptococcus pyogenes* is most widely utilized for gene-editing and consists of the Cas9 endonuclease and a synthetic single-guide RNA (sgRNA).[279,282,284,286,301,472,485] DNA targeting by sgRNA-bound Cas9 is achieved through recognition of a protospacer adjacent motif (PAM) on the non-target DNA strand (ntDNA) followed by heteroduplex formation between the sgRNA and the target DNA strand (tDNA) (Figure 8.1).[318,487,488] Subsequently, consecutive activation of the HNH and RuvC nuclease domains through coupled conformational dynamics leads to rapid cleavage of tDNA and ntDNA at the HNH and RuvC active sites, respectively.[317,324,327] Based on DNA sequencing of Cas9 cleavage products, it was previously concluded that Cas9 generates blunt-ends at a position three base pairs upstream of the PAM.[279] However, molecular dynamics simulations suggest that RuvC may actually generate staggered DNA ends with single-nucleotide 5′-overhangs.[489] While it is widely recognized that, following initial DNA cleavage by Cas9, the ntDNA strand is further processed by the 3′→5′ post-cleavage trimming (PCT) activity of RuvC on the PAM distal DNA fragment, it is unknown if this activity can occur before Cas9 dissociates from its initial DNA cleavage products.[279,304,490,491] Gene-editing with Cas9 is not always successful and can have unintended consequences which represent significant obstacles as this technology moves towards clinical applications (*i.e.* gene therapy). As a

result, comprehensive characterization of Cas9 DNA cleavage and trimming activities is essential to uncovering functional limitations and experimental considerations.

Here we sought to explore the selection of initial cleavage site(s) by RuvC and determine if the PCT activity is kinetically significant within a single Cas9•sgRNA and DNA binding event. We found that RuvC can select up to three positions along the ntDNA strand for initial cleavage but prefers to generate DNA products with single-nucleotide 5′-overhangs (on the tDNA strand), rather than blunt-ends. Surprisingly, we found that, in addition to the known 3′→5′ PCT activity, Cas9 also possesses 5′→3′ PCT activity. We determined the kinetics of RuvC-catalyzed 3′→5′ and 5′→3′ post-cleavage trimming and found them to be significant within a single Cas9•sgRNA and DNA binding event (*i.e.* extensive trimming occurs prior to DNA product release). To our knowledge, this is the first quantitative and mechanistic evaluation of RuvC initial cleavage site selection and post-cleavage trimming that collectively dictate the final types of DNA ends generated by Cas9 during *in vivo* gene-editing before their interaction with protein factors of the non-homologous end joining (NHEJ) or homology-directed repair (HDR) pathways, which mediate gene knock-out and gene knock-in, respectively.[284,314,492]

8.3 Results and Discussion

8.3.1 RuvC selects multiple sites for initial DNA cleavage.

To investigate initial cleavage site selection by RuvC-catalyzed during ntDNA cleavage, we performed single-turnover assays with time points spanning six orders of magnitude (0.003 – 7,200 s) to capture initial DNA cleavage and 3′→5′ PCT activities

(Figures 8.2A and 8.3A). Briefly, Cas9 (100 nM) and sgRNA (300 nM) were pre-incubated for 30 min at 20 °C to form the binary complex (Cas9•sgRNA) before the addition of the DNA substrate (10 nM, Figure 8.1), which was 5′-radiolabeled on ntDNA. Following a 1 hr incubation in the absence of $MgCl_2$ to form the ternary complex (Cas9•sgRNA•DNA), the solution was rapidly mixed with $MgCl_2$ (6 mM) in the presence (Figure 8.2A) or absence (Figure 8.3A) of unlabeled DNA substrate (1,000 nM) to initiate DNA cleavage at 37 °C for various amounts of time before quenching with EDTA (0.37 M). Close inspection of ntDNA cleavage by RuvC (Figures 8.2A and 8.3A) revealed appearance of at least two distinct DNA products (27-mer and 26-mer, Figures 8.2A and 8.3A) even at the earliest time point of acquisition (3 ms). To a lesser extent, a third DNA product (25-mer, Figures 8.2A and 8.3A) was also faintly observed in early time points (<0.024 s, Figures 8.2A and 8.3A). Contrary to the current paradigm that Cas9 first generates a blunt-ended DNA product prior to known 3′→5′ exonuclease degradation of ntDNA by RuvC that results in a pattern of multiple ntDNA products,[279,304] the 27-mer and 26-mer products accumulated with similar kinetics (Figure 8.2B), suggesting that RuvC initially cleaved at the canonical -3 position and an unexpected -4 position on the ntDNA (Figure 8.1). Comparatively, the 25-mer product appeared with slower kinetics (Figures 8.2B and 8.3C) suggesting that initial cleavage at the -5 position (Figure 8.1) was not as favorable. Consistently, the approximate concentration ratio of the initial ntDNA products 27-mer, 26-mer, and 25-mer (1 **:** 1.7 **:** 0.3) was maintained from 0.003 – 60 s and suggested that RuvC-catalyzed ntDNA endonucleolytic cleavage was preferred at the -4 position, rather the -3 and -5 positions

(Figure 8.1). The slower kinetics and unfavorable ratio for initial formation of the 25-mer product likely suggests that the scissile bond of the -5 position is far from the catalytic residues within the RuvC active site. Moreover, the appearance of three endonucleolytic cleavage products suggests that the single-stranded ntDNA is flexible within the active site of RuvC, in contrast to stably bound tDNA, which is base paired to the sgRNA within the HNH active site. As a result, Cas9-catalyzed initial DNA cleavage primarily generated staggered-ended DNA products with single-nucleotide 5′-overhangs, rather than blunt-ends (Figure 8.2C). Importantly, as the 44-mer ntDNA used in all experiments was carefully purified (see Section 8.4.1) and was found to be free of 43-mer or 42-mer DNA impurities by 5′-radiolabeling (Figures 8.2A and 8.3A) and mass spectrometric analyses (Figure 8.4A), it is very unlikely that such strongly accumulated ntDNA cleavage products could have arose from DNA contaminants which are shorter by 1 or 2 nucleotides relative to the 44-mer. The appearance of 26-mer and 25-mer products could arguably have resulted from the previously observed 3′→5′ exonuclease activity of RuvC.[279] To test this assertion, we repeated the single-turnover assay using a 3′-radiolabeled DNA substrate and observed at least two ntDNA products (19-mer and 18-mer, Figure 8.5A), which accumulated with similar kinetics (Figure 8.5B), thereby confirming that RuvC selects multiple sites for initial DNA cleavage as supported by our results with the 5′-radiolabeled DNA substrate (Figures 8.2A and 8.3A) and previously predicted by molecular dynamic simulations.[489] Moreover, only one HNH cleavage product was observed (28-mer, Figure 8.5A). Thus, initial DNA cleavage by Cas9 may result in blunt- or staggered-ended double-strand DNA breaks (Figure 8.5C) as a result of

ntDNA cleavage at one of three positions by RuvC relative to invariant cleavage at the +3 position of tDNA by HNH (Figure 8.1). Our use of pre-steady-state kinetics and a rapid chemical quench-flow apparatus to monitor the earliest time point (3 ms) of rapid Cas9-catalyzed DNA cleavage facilitated discovery of this mechanistic detail (Figures 8.2A, 8.3A, and 8.5A), whereas other studies were limited to comparatively later time ranges (≥ 30 s) wherein 3′→5′ post-cleavage trimming of ntDNA by RuvC would be more extensive.[279,304] Compellingly, Figure 8.5A also revealed that the 3′-radiolabeled 19-mer product was eventually degraded by precisely one nucleotide to become 18-mer, and as we confirmed that our Cas9 protein preparation was free of contaminating nucleases (Figure 8.4B), this result suggested a novel 5′→3′ PCT activity of RuvC.

To further confirm that Cas9 catalyzes 5′→3′ post-cleavage trimming of the ntDNA following initial cleavage, we sequenced the DNA product generated by time-dependent cleavage of a plasmid (Figure 8.6A) containing the same Cas9 target sequence as in Figure 8.1. Briefly, Cas9•sgRNA (300 nM) was pre-incubated with the plasmid DNA substrate (30 nM) for 1 hr to form Cas9•sgRNA•DNA before mixing with $MgCl_2$ (6 mM) to initiate DNA cleavage. Reactions were terminated at various times (15 s – 2 hrs) by vigorous mixing with sodium dodecylsulfate (0.2%). The plasmid products were purified by agarose gel electrophoresis before DNA sequencing analysis (Figure 8.6B). DNA sequencing of the earliest time point (15 s, Figure 8.6B) revealed that initial ntDNA cleavage by RuvC apparently occurred at the -4 position as indicated by the observed penultimate base (*i.e.* 3′ thymidine, indicated by black arrow) of the PAM proximal ntDNA sequencing product. Later time points (5 min – 2 hr, Figure 8.6B) demonstrated

that the relative amplitude of the 3′ thymidine residue decreased over time becoming nearly undetectable after 2 hrs. This indicated that the 5′ adenosine of the PAM proximal ntDNA was slowly removed by the 5′→3′ PCT activity of RuvC in support of the results from the single-turnover assay with 3′-radiolabeled DNA (Figure 8.5A). Notably, DNA sequencing of the PAM distal tDNA (Figure 8.6B) showed the expected cleavage at the -3 position by HNH for all time points.

Taken together, the known 3′→5′ and unexpected 5′→3′ PCT activities of RuvC coupled with insufficient time resolution during past DNA cleavage experiments likely contributed to improper assignment of the preferred ntDNA cleavage site during previously reported DNA sequencing of the PAM proximal and distal products of Cas9-catalyzed DNA cleavage.[279] The experimental consistency between results collected with the relatively small 40/44-mer DNA substrate (Figure 8.5) and the more physiologically-relevant plasmid DNA substrate (Figure 8.6) underlines the validity and likelihood of the newly discovered 5′→3′ PCT activity of RuvC during gene-editing with Cas9. Clearly, the extent of 5′→3′ PCT (one nucleotide, Figs 3A and 4) is less significant relative to 3′→5′ PCT (up to seven nucleotides, Figure 8.2A) which may be partly explained by the lengths of single-stranded PAM proximal versus PAM distal ntDNA. Following initial DNA cleavage, the PAM proximal ntDNA fragment would contain a short single-stranded region (~3 – 5 nucleotides, Figure 8.1) relative to the PAM distal ntDNA fragment (~16 – 18 nucleotides, Figure 8.1). It may be that the shorter single-stranded PAM proximal ntDNA is relatively rigid compared to the longer, likely more flexible, single-stranded PAM distal ntDNA and therefore less susceptible to 5′→3′ PCT by RuvC

(*i.e.* less likely to engage the active site moieties of RuvC for trimming).

8.3.2 The 3′→5′ post-cleavage trimming of ntDNA by Cas9 is significant during a single Cas9•sgRNA and DNA binding event.

For comparison, the single-turnover assays were performed in the presence (Figure 8.2A) or absence (Figure 8.3A) of a large molar excess of unlabeled, double-stranded 40/44-mer trap DNA (1 μM, Figure 8.1). Notably, although single-stranded 44-mer ntDNA binds ~250-fold more weakly (K_d^{DNA} = 490 nM in the absence of $MgCl_2$, Figure 8.7A) to the binary complex Cas9•sgRNA than the 40/44-mer on-target double-stranded DNA substrate ($K_d^{DNA} \approx 2$ nM in the absence of $MgCl_2$, determined in our unpublished manuscript), Cas9•sgRNA is capable of low levels of single-stranded ntDNA cleavage activity (Figure 8.7B). Accordingly, the large molar excess of the trap DNA (40-mer/44-mer) was included to sequester any free Cas9•sgRNA that could potentially degrade short, radiolabeled single-stranded ntDNA products released from the Cas9 product complex during the long incubation period after initial DNA cleavage (Figure 8.7C). Moreover, any dissociated radiolabeled single-stranded ntDNA products were diluted by the trap DNA during the reaction. While processive degradation of the initial ntDNA cleavage products was observed up to the -10 base (20-mer, Figure 8.2A) after 7,200 s in the presence of the trap DNA, minor accumulation of even smaller ntDNA products was observed in the absence of the trap DNA (Figure 8.3A), suggesting that the short single-stranded ntDNA products slowly dissociated from the Cas9 product complex and were subsequently degraded by free Cas9•sgRNA in the absence of the trap DNA. Nevertheless, results in the presence of the trap DNA (Figure 8.2A) demonstrate

that the processive degradation of the ntDNA up to the -10 base is significant within a single Cas9•sgRNA and DNA binding event (*i.e.* ntDNA degradation occurs after initial DNA cleavage but before Cas9•sgRNA can release the initial products of DNA cleavage). Notably, the minor accumulation of small ntDNA products in the absence of trap DNA as observed by 5′-radiolabeling in Figure 8.3A was not detected by 3′-radiolabeling in Figure 8.5A suggesting that the PAM distal ntDNA cleavage product (*i.e.* monitored in Figure 8.3A) may dissociate more readily from the Cas9 product complex than the PAM proximal ntDNA cleavage product (*i.e.* monitored in Figure 8.5A) for subsequent degradation by free Cas9•sgRNA.

8.3.3 Kinetics of RuvC-catalyzed 3′→5′ post-cleavage trimming of ntDNA.

We next quantified the total amount of DNA substrate remaining (*i.e.* 44-mer), as well as the amounts of products formed from the initial ntDNA cleavage and subsequent 3′→5′ post-cleavage trimming (*i.e.* 27-mer – 20-mer) by RuvC for each time point in the absence (Figure 8.3A) or presence (Figure 8.2A) of the trap DNA. As the single-nucleotide excision reactions at multiple positions (*i.e.* 27-mer – 20-mer) appeared to occur with similar rates, simple non-linear regression analysis was not sufficient to directly obtain kinetic parameters for observed 3′→5′ PCT. Thus, we simulated RuvC-catalyzed DNA cleavage with KinTek Explorer global fitting software[493,494] (KinTek Corporation) using the kinetic mechanism and parameters defined in our unpublished manuscript (Figure 8.8). Initial DNA cleavage was modeled as three distinct catalytic events (*i.e.* 44-mer ⇌ 27-mer, 44-mer ⇌ 26-mer, 44-mer ⇌ 25-mer) while subsequent 3′→5′ PCT was modeled as a series of seven single-nucleotide excision events (*i.e.* 27-

mer ⇌ 26-mer ⇌ 25-mer ⇌ 24-mer ⇌ 23-mer ⇌ 22-mer ⇌ 21-mer ⇌ 20-mer). Products shorter than 20-mer in the absence of the DNA trap were not analyzed in Figure 8.3B because they were not likely generated by the 3′→5′ PCT activity as explained below. Visual inspection of the global fit to the data revealed that even at the earliest time points, the kinetic model closely followed the observed depletion and accumulation of substrates and products (Figures 8.3B and 8.9). Consistently, comparison between rates modeled through KinTek (Figure 8.9) and corresponding rates from direct data fitting by non-linear regression analysis (Figure 8.2B) for individual accumulation of 27-mer, 26-mer, and 25-mer products revealed similar values (Table 8.1) and therefore demonstrated the high quality of the fit and the robustness of the simulation. Additionally, initial DNA cleavage kinetics obtained by non-linear regression and global data fitting were consistent regardless of the presence (Figure 8.2B) or absence (Figure 8.3C) of trap DNA (Table 8.1). While the simulated fits of RuvC-catalyzed DNA cleavage in the absence (Figure 8.3B) and presence (Figure 8.9) of the DNA trap were virtually identical at early time points, the late time points of 3′→5′ PCT in the absence of the trap DNA revealed greater accumulation of smaller ntDNA products (*i.e.* 21-mer and 20-mer) than in the presence of the DNA trap, leading to apparently faster ntDNA degradation rates in the absence (0.0015 – 0.033 s^{-1}) than in the presence (0.0005 – 0.017 s^{-1}) of the DNA trap (Table 8.2). These differences are most likely caused by the dissociation of shorter single-stranded ntDNA cleavage products which were further degraded by free Cas9•sgRNA in the absence of the trap DNA. Interestingly, the simulated rates for the sequential 3′→5′ degradation of each ntDNA product (0.0005 – 0.017 s^{-1}, Table 8.2) in

the presence of DNA trap were ~100- to 8,000-fold faster than the rate of DNA product release (k_7 = 2.1 x 10^{-6} – 4.5 x10^{-6} s^{-1}, Figure 8.8) determined in our unpublished manuscript. Since the DNA product release occurred with a $t_{1/2}$ (= ln(2)/rate) of ~43 – 91 hrs, Cas9, within a single enzyme binding event, is projected to catalyze complete processive degradation of the initial ntDNA cleavage products through the -10 position in 5.3 hrs (*i.e.* sum all 7$t_{1/2}$ values; each 7$t_{1/2}$ is considered the time necessary to complete a single excision event). This kinetic conclusion is supported by the results of ntDNA cleavage in the presence of the DNA trap (Figure 8.2A).

Remarkably, Next Generation DNA Sequencing data at sites of Cas9-mediated gene modification[495] have shown that nucleotide insertions during DNA repair are non-random and suggest that the cleavage pattern of RuvC may extend from the -3 to -10 positions of ntDNA (Figure 8.1) which is consistent with our experimental findings (Figure 8.2C). The continuous degradation of ntDNA by RuvC during persistent binding of Cas9 to its DNA products (*i.e.* fast post-cleavage trimming relative to slow DNA product release) is likely the mechanistic basis for these necessary nucleotide insertions at sites of CRISPR/Cas9-mediated gene-editing.

8.3.4 Kinetics of RuvC-catalyzed 5′→3′ post-cleavage trimming of ntDNA.

To determine the kinetics of the novel 5′→3′ PCT activity of Cas9, we quantified the total amount of DNA substrate remaining (*i.e.* 3′-radiolabeled 45-mer), as well as the amounts of products formed from the initial ntDNA cleavage and subsequent 5′→3′ PCT activity (*i.e.* 19-mer – 18-mer) of RuvC for each time point in the single-turnover assay (Figure 8.5A). Similar to Figure 8.9 (*i.e.* 3′→5′ PCT), we simulated the RuvC-catalyzed

initial DNA cleavage and 5′→3′ post-cleavage DNA product trimming (Figure 8.10) using KinTek Explorer global fitting software.[493,494] The initial DNA cleavage was modeled as two distinct catalytic events (*i.e.* 45-mer ⇌ 19-mer, 45-mer ⇌ 18-mer; signal for the 20-mer product was too weak to quantify), while the subsequent 5′→3′ PCT was modeled as a single-nucleotide excision event (*i.e.* 19-mer ⇌ 18-mer). Importantly, the kinetic simulation to the data was considered excellent as visual inspection revealed strong correlation between the modeled and experimental data at both early and late time ranges (Figure 8.10), and the rate constants of initial DNA cleavage derived from the simulation (Figure 8.10) and non-linear regression analysis (Figure 8.5B) were comparable (Table 8.1). As expected, kinetics of initial endonucleolytic cleavage at the -3 and -4 positions obtained by either non-linear regression or global fitting were consistent between the 5′- and 3′-radiolabeling strategies (Figures 8.2B, 8.5B, and Table 8.1). The single-nucleotide excision of the 19-mer product to the 18-mer by 5′→3′ post-cleavage trimming occurred with a rate constant of 0.0026 s^{-1} (Table 8.2) which is comparable (within 6-fold) to those measured for the 3′→5′ PCT activity in the presence of the DNA trap (Table 8.2), suggesting that both PCT activities of RuvC happened within the same time frame. Interestingly, this measured rate for 5′→3′ post-cleavage trimming is in agreement with the timing ($t_{1/2}$ = ~5 min; $7t_{1/2}$ = ~30 min) observed for the disappearance of the penultimate 5′-thymidine during DNA sequencing of a plasmid substrate cleaved by Cas9 for various times in Figure 8.6. In contrast to the degradation of 5′-radiolabeled ntDNA (Figures 8.2A and 8.3A), the 3′-radiolabeled ntDNA was degraded by only a single nucleotide since no other smaller degradation products were observed in Figure

8.5A. Accordingly, the use of a DNA trap was not necessary in the single-turnover assay (Figure 8.5A). Furthermore, the time-dependent accumulation of post-cleavage trimming products from Figures 8.2A and 8.5A suggest that the 3′→5′ or 5′→3′ PCT activities occurred simultaneously.

Contrary to the current paradigm that Cas9 produces blunt-ended DNA products, we have discovered that Cas9 initially generates heterogeneous DNA ends, which are subsequently trimmed in both 3′→5′ and 5′→3′ directions (Figures 8.2C and 8.5C). Furthermore, we have determined the kinetic significance and extent of the RuvC-catalyzed 3′→5′ (Figure 8.2C) and 5′→3′ (Figure 8.5C) post-cleavage trimming within a single Cas9•sgRNA and DNA binding event. Consequently, researchers can now begin to anticipate the sequence context for a given gene modification based on the DNA ends generated by the combined endonuclease and PCT activities of Cas9 (Figures 8.2C and 8.5C) as well as explore the possibility of protein engineering or temporal control of Cas9 for more precise gene-editing. Moreover, as the status of DNA ends following Cas9-catalyzed DNA cleavage may influence the outcome of gene-editing experiments by modulating which repair factors (*i.e.* those of the NHEJ pathway or HDR pathway) are recruited to the double-strand DNA break, the length of time Cas9 remains bound to its DNA product may affect the efficiency of gene insertion as persistent PCT may generate sub-optimal DNA ends for HDR, but the preferred DNA ends for NHEJ. The unique activities of Cas9 characterized here may also be useful in methods of DNA manipulation, including DNA assembly and cloning.

8.4 Experimental Section

8.4.1 Preparation of *Sp* Cas9, sgRNA, and DNA substrates.

Sp Cas9 was purified as previously described.[279,478] *In vitro* transcription (IVT) was used to prepare sgRNA (Figure 8.1) as previously described[478] with a 140-mer DNA template designed from the S. pyogenes CRISPR and trans-activating RNA sequences fused by a GAAA tetraloop linker. Briefly, the sgRNA was synthesized by IVT and subsequently purified by denaturing PAGE (8 M urea, 10% polyacrylamide, 1x TBE) and stored as ethanol precipitates at -80 °C. Prior to each experiment, the sgRNA was pelleted, washed, and then folded in RNA annealing buffer (10 mM Tris, 50 mM NaCl, 1 mM EDTA, pH 7.5) by heating to 95 °C for 5 min and cooling slowly to 20 °C.
DNA oligonucleotide substrates (Figure 8.1) were purchased from Integrated DNA Technologies Inc. (IDT) and subsequently purified by denaturing PAGE (8 M urea, 10-15% polyacrylamide, 1x TBE). Oligonucleotides were 5′-radiolabeled by incubation with [γ-^{32}P]ATP (Perkin-Elmer) and T4 polynucleotide kinase (New England Biolabs) for 3 hrs at 37 °C. Oligonucleotides were 3′-radiolabeled by single-nucleotide extension with [α-^{32}P]-3′-deoxyadenosine triphosphate ([α-^{32}P]-3′-dATP) (Perkin-Elmer) and terminal deoxynucleotidyl transferase for 3 hrs at 37 °C. Bio-Spin 6 size-exclusion columns (Bio-Rad) were used to remove unreacted [γ-^{32}P]ATP or [α-^{32}P]-3′-dATP. Radiolabeled non-target oligonucleotides were annealed to unlabeled target oligonucleotides (10% molar excess) by heating to 95 °C for 5 min followed by slow cooling to 20 °C to generate the 40/44-mer DNA substrate (Figure 8.1) used in Figures 8.2A and 8.3A. To prepare the 3′-radiolabeled DNA substrate (41/45-mer) used in Figure 8.5A, non-radiolabeled 40- and

44-mer were converted into the 3′-radiolabeled 41- and 45-mer, respectively, through TdT-catalyzed addition of [α-^{32}P]-3′-dATP (see above). Varying amounts of 3′-radiolabeled target and non-target DNA strands were annealed and analyzed by native-PAGE to determine the ideal ratio for optimal DNA duplex formation. Typical optimized annealing efficiencies were ~90% (data not shown).

8.4.2 Product and data analysis.

Denaturing PAGE (12% polyacrylamide, 8 M urea, 1x TBE) was used to separate reaction substrates and products. Gels were scanned using a Typhoon RGB (GE Healthcare) and quantified by densitometry in ImageQuant (Molecular Dynamics). RuvC product formation was quantified as the sum of all ntDNA products unless otherwise specified. All kinetic data were fit by either global data fitting (see below) or non-linear regression using KaleidaGraph (Synergy Software). For Figures 8.2B, 8.3C, and 8.5B data were fit to the double exponential equation

$$[\text{product}] = A_1[1 - \exp(-k_1 t)] + A_2[1 - \exp(-k_2 t)] \qquad \text{Eq. 8.1}$$

where A_1 and A_2 are the exponential phase amplitudes, and k_1 and k_2 are the observed rate constants. Equilibrium binding data were fit to Eq. 8.2 (Figure 8.7A),

$$[\text{Complex}] = 0.5(K_d + E_0 + D_0) - 0.5\,[(K_d + E_0 + D_0)^2 - 4E_0D_0]^{1/2} \qquad \text{Eq. 8.2}$$

where E_0 and D_0 are the Cas9•sgRNA and single-stranded ntDNA concentrations, respectively.[479]

8.4.3 Kinetic simulations with KinTek Explorer global fitting software.

Kinetics of time-dependent substrate decay and product formation from Figures 8.2A, 8.2A, and 8.5A were simulated using KinTek Explorer global fitting software. Kinetic models were developed and bounded based upon the kinetic mechanism and measured parameters (*i.e.* rate constants for elementary steps) determined in our unpublished manuscript (Figure 8.8). When simulating kinetics from assays monitoring 5′-radiolabeled DNA substrates and products, initial DNA cleavage was modeled as three distinct catalytic events corresponding to cleavage of the 44-mer ntDNA at the -3, -4, and -5 positions to form 27-mer, 26-mer, and 25-mer ntDNA products, respectively (Figures 8.2A and 8.3A). Subsequent 3′→5′ PCT was modeled as a series of seven single-nucleotide excision events beginning from the 27-mer to ultimately produce the 20-mer (Figures 8.2A and 8.3A). When simulating kinetics from assays monitoring 3′-radiolabeled DNA substrates and products, initial DNA cleavage was modeled as two distinct catalytic events (as a result of weak radioactive signal for cleavage at the -5 position) corresponding to cleavage of the 45-mer ntDNA at the -3 and -4 to generate 18-mer and 19-mer ntDNA products, respectively (Figure 8.5A). Subsequent 5′→3′ PCT was modeled as a single nucleotide excision event beginning from the 19-mer to produce the 18-mer (Figure 8.5A). Kinetic simulations were evaluated based on visual inspection of the goodness of fit, the Chi-squared metric calculated by KinTek Explorer global fitting software, and agreement between the simulated kinetic values for initial DNA cleavage and those directly measured by non-linear regression analysis (Figures 8.2B, 8.3C, 8.5B, and Table 8.1).

8.4.4 Assay for contaminating nuclease activities in Cas9 protein preparation.

To test for contaminating nucleases in our Cas9 protein preparation, apo-Cas9 (100 nM) was mixed with 5′-radiolabeled single- (40-mer ss-tDNA and 44-mer ss-ntDNA) or double-stranded (40/44mer dsDNA) DNA substrates (10 nM) in the presence of $MgCl_2$ (6 mM) for various times (0 – 4 hrs) before quenching with EDTA (0.37 M). Reactions were subsequently resolved by denaturing PAGE (Figure 8.4B).

8.4.5 Measurement of Cas9•sgRNA affinity to single-stranded ntDNA.

The equilibrium dissociation constant for Cas9•sgRNA binding to single-stranded ntDNA was measured by incubating various concentrations of Cas9•sgRNA complex (1 – 1,000 nM) with 5′-radiolabeled ntDNA (0.5 nM) for 1 hr in the absence of $MgCl_2$. Cas9•sgRNA•ntDNA was separated from free ntDNA by native polyacrylamide gel electrophoresis (PAGE) (5% polyacrylamide, 0.5x TBE, 4 °C). The concentration of Cas9•sgRNA•ntDNA was quantified by phosphorimaging and plotted versus the concentration of Cas9•sgRNA (Figure 8.7A). The data were then fit to the quadratic binding equation, Eq. 8.2 (Figure 8.7A).

8.4.6 Single-turnover DNA cleavage assays.

Unless otherwise specified, kinetic experiments were performed at 37 °C in optimized reaction buffer A (20 mM HEPES (pH 7.5 at 37 °C), 125 mM KCl, 1 mM EDTA, 1 mM DTT, and 10% glycerol). A rapid chemical quench-flow apparatus (KinTek) was used to perform fast reactions. All concentrations are reported as final concentrations after mixing. Reactions were terminated using quench solution (0.37 M EDTA, 20% formamide, 0.25% bromophenol blue, and 0.25% xylene cyanol). We

determined 0.37 M EDTA was sufficient to prevent metal ion dependent DNA cleavage (data not shown) as reported for other DNA modifying enzymes.[90,95]

For double-stranded DNA cleavage assays (Figures 8.2A, 8.3A, and 8.5A), Cas9 (100 nM) and sgRNA (300 nM) were pre-incubated for 30 min at 20 °C to form Cas9•sgRNA. Next, Cas9•sgRNA (100 nM) was pre-incubated at 20 °C for 1 hr with DNA substrate (10 nM), 5′-radiolabeled or 3′-radiolabeled on tDNA, ntDNA or both, to form Cas9•sgRNA•DNA in the absence of $MgCl_2$. Cas9•sgRNA•DNA (10 nM) was then rapidly mixed with $MgCl_2$ (6 mM) in the presence (Figure 8.2A) or absence (Figures 8.3A and 8.5A) of unlabeled DNA substrate (1,000 nM) at 37 °C to initiate strand scission for varying amounts of time before quenching with EDTA (0.37 M).

For the single-stranded ntDNA cleavage assay (Figure 8.7B), Cas9 (100 nM) and sgRNA (300 nM) were pre-incubated for 30 min at 20 °C to form Cas9•sgRNA. Next, Cas9•sgRNA (100 nM) was pre-incubated at 20 °C for 1 hr with 5′-radiolabeled single-stranded ntDNA (10 nM), to form Cas9•sgRNA•ntDNA in the absence of $MgCl_2$. Cas9•sgRNA•ntDNA (10 nM) was then rapidly mixed with $MgCl_2$ (6 mM) at 37 °C to initiate strand scission for varying amounts of time (5 – 3,600 s) before quenching with EDTA (0.37 M).

For the ntDNA cleavage assay in Figure 8.7C, Cas9 (100 nM) and sgRNA (300 nM) were pre-incubated for 30 min at 20 °C to form Cas9•sgRNA. Next, Cas9•sgRNA (100 nM) was pre-incubated at 20 °C for 1 hr with 5′-radiolabeled single-stranded ntDNA (10 nM) and a large excess of unlabeled, double-stranded 40/44-mer trap DNA (1,000 nM) in the absence of $MgCl_2$. The solution was then rapidly mixed with $MgCl_2$ (6 mM)

at 37 °C to initiate strand scission for varying amounts of time (0 – 7,200 s) before quenching with EDTA (0.37 M). The absence of ntDNA cleavage indicates that the large excess of trap DNA sequestered Cas9•sgRNA thus protecting radiolabeled ntDNA from cleavage.

8.4.7 Plasmid DNA cleavage assay and DNA sequencing.

Cas9 (300 nM) and sgRNA (900 nM) were pre-incubated for 30 min at 20 °C to form Cas9•sgRNA. Next, Cas9•sgRNA (300 nM) was pre-incubated at 20 °C for 1 hr with a plasmid DNA substrate (30 nM, pTSGAS1, Figure 8.6A), containing the same Cas9 target sequence as in Figure 8.1, to form Cas9•sgRNA•DNA in the absence of $MgCl_2$. Cas9•sgRNA•DNA (30 nM) was then mixed with $MgCl_2$ (6 mM) at 37 °C to initiate strand scission for varying amounts of time. Reactions were terminated by vigorous mixing with sodium dodecylsulfate (0.2%). Plasmid cleavage products were purified by agarose gel electrophoresis (0.8% agarose in TAE) followed by gel extraction (QIAquick Gel Extraction Kit, Qiagen). The extracted plasmid cleavage products were then sent for fluorescence-based Sanger sequencing (Figure 8.6B) by The Ohio State University Comprehensive Cancer Center (OSUCCC) Genomics Shared Resource laboratory (GSR) using custom designed primers (IDT).

8.4.8 MALDI-TOF mass spectrometry.

The 44-mer ntDNA was analyzed by matrix assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry by the Ohio State University Campus Chemical Instrument Center (CCIC) using a 3-hydroxypicolinic acid matrix with 0.1% trifluoroacetic acid (Figure 8.4A). The experiments were performed on a Bruker-

Daltonics UltrafleXtreme MALDI-TOF-TOF mass spectrometer and the data were analyzed using Bruker Compass DataAnalysis 4.2 software.

8.5 Tables

Table 8.1: Comparison of kinetic data fitting methods for determination of RuvC-catalyzed endonucleolytic rate constants at 37 °C

RuvC Endonuclease Activity	ntDNA Position[a]	Rate constant (s^{-1})	
		Non-linear regression fitting[b]	Global data fitting[c]
5′-radiolabeling[d]			
44-mer → 27-mer	-3	18.4	11.1
44-mer → 26-mer	-4	19.4	15.2
44-mer → 25-mer	-5	1.9	3.9
5′-radiolabeling[e]			
44-mer → 27-mer	-3	17.1	9.3
44-mer → 26-mer	-4	15.9	15.3
44-mer → 25-mer	-5	3.3	2
3′-radiolabeling[f]			
45-mer → 19-mer	-4	17.8	18.3
45-mer → 18-mer	-3	23.3	10.1

[a]Relative to the 5′-phosphodiester bond (0 position) of the first nucleotide in the PAM sequence (5′-**T**GG-3′) on ntDNA, refer to Figure 8.1.
[b]Calculated values reflect a weighted sum of the rates for the fast and slow exponential phases of product formation [(k_{fast}*A_{fast}) + (k_{slow}*A_{slow})] that were directly obtained by non-linear regression analysis in Figures 8.2B, 8.3C, and 8.5B.
[c]Values obtained from kinetic simulations in Figures 8.3B, 8.9 and 8.10.
[d]DNA size is based on the 5′-radiolabeling strategy (*i.e.* 40/44-mer DNA substrate), refer to Figure 8.2A. Rates are from the experiment performed with trap DNA.
[e]DNA size is based on the 5′-radiolabeling strategy (*i.e.* 40/44-mer DNA substrate), refer to Figure 8.3A. Rates are from the experiment performed without trap DNA.
[f]DNA size is based on the 3′-radiolabeling strategy (*i.e.* 41/45-mer DNA substrate), refer to Figure 8.5A. Radioactive signal for accumulation of the 20-mer DNA product corresponding to cleavage at the -5 position was too weak to be accurately quantified.

Table 8.2: Rates of RuvC-catalyzed post-cleavage trimming activities at 37 °C simulated by KinTek Explorer global data fitting

3′→5′ PCT[a]	ntDNA Position[b]	Rate constant (s^{-1})[c]	
		With DNA trap	No DNA trap
27-mer → 26-mer	-4	0.011 ± 0.003	0.011 ± 0.002
26-mer → 25-mer	-5	0.017 ± 0.004	0.014 ± 0.001
25-mer → 24-mer	-6	0.0062 ± 0.0007	0.010 ± 0.001
24-mer → 23-mer	-7	0.017 ± 0.004	0.033 ± 0.006
23-mer → 22-mer	-8	0.006 ± 0.004	0.015 ± 0.02
22-mer → 21-mer	-9	0.0005 ± 0.0001	0.006 ± 0.002
21-mer → 20-mer	-10	0.0007 ± 0.0006	0.0015 ± 0.0009
5′→3′ PCT[d]			
19-mer → 18-mer	-3	ND	0.0026 ± 0.0001

[a]DNA size is based on the 5′-radiolabeling strategy (i.e. 40/44-mer DNA substrate), refer to Figure 8.2A
[b]Relative to the 5′-phosphodiester bond (0 position) of the first nucleotide in the PAM sequence (5′-TGG-3′) on ntDNA, refer to Figure 8.1
[c]Substrate decay and product formation from Figures 8.2A, 8.3A, and 8.5A were modeled to obtain reaction rates using KinTek Explorer global fitting software as shown in Figures 8.3B, 8.9, and 8.10. Error values are data fitting errors.
[d]DNA size is based on the 3′-radiolabeling strategy (i.e. 41/45-mer DNA substrate), refer to Figure 8.5

8.6 Figures

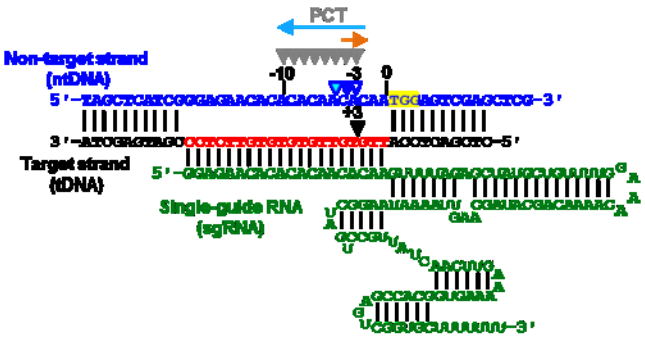


Figure 8.1: Schematic of the sgRNA•DNA heteroduplex showing PCT activities. PAM sequence on ntDNA (44-mer) highlighted in yellow and 20-nucleotide complementarity of tDNA (40-mer) to sgRNA colored red. Experimentally observed cleavage sites are marked by black (HNH) and blue (RuvC) triangles (filled: preferred site, open: canonical site (less favored), cyan filled: disfavored site). Directions and extent of RuvC-catalyzed post-cleavage trimming activities determined here are shown by cyan (3′→5′ PCT) and orange (5′→3′ PCT) arrows.

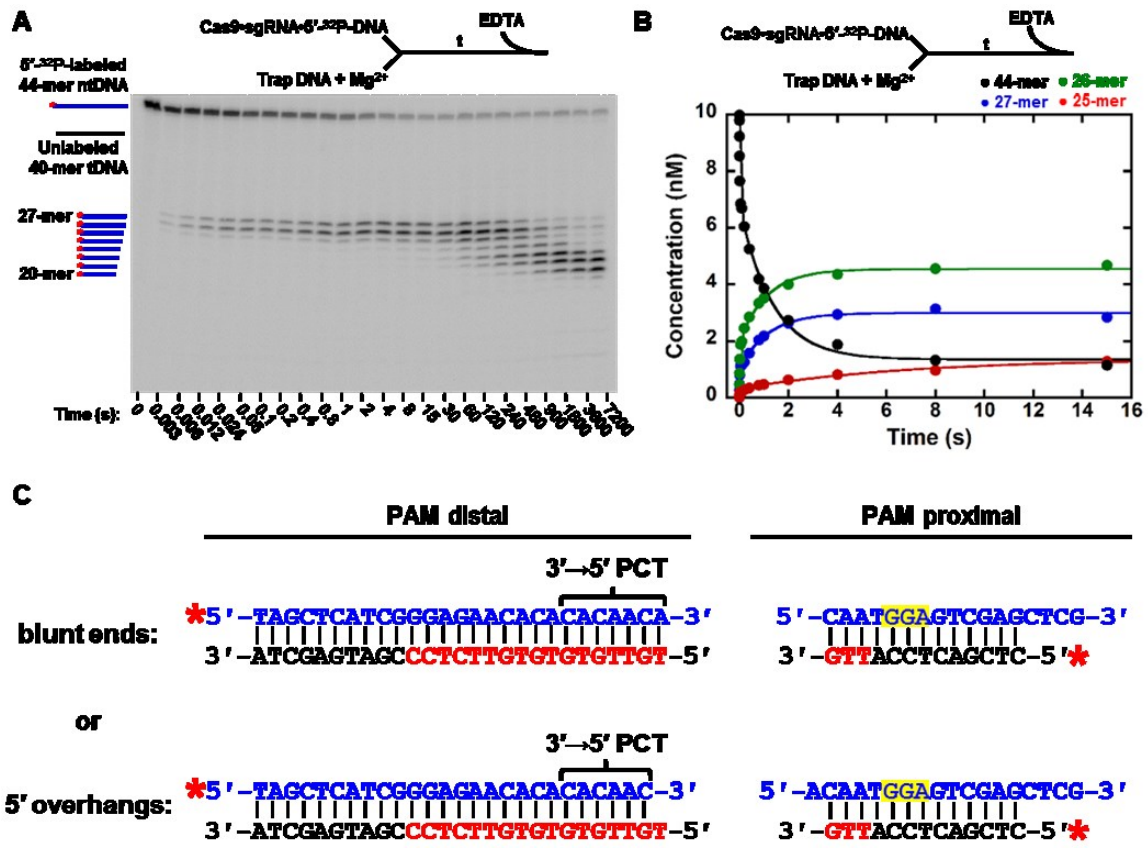


Figure 8.2: Single-turnover DNA cleavage of 5′-radiolabeled DNA by Cas9.

(A) Gel image showing time-dependent (0.003 – 7,200 s) cleavage of DNA (10 nM) by Cas9•sgRNA (100 nM) in the presence of unlabeled trap DNA (1,000 nM). The DNA substrate (Figure 8.1) was 5′-radiolabeled (*) on ntDNA. (B) DNA substrate (44-mer) and products (27-mer – 25-mer) from Figure 8.2A were quantified and fit to Eq. 8.1 using non-linear regression analysis to yield kinetic parameters of the fast and slow phases of DNA cleavage for the 44-mer (k_{fast} = 50 ± 7 s^{-1}, 38% amplitude; k_{slow} = 0.75 ± 0.06 s^{-1}, 62% amplitude), 27-mer (k_{fast} = 54 ± 9 s^{-1}, 33% amplitude; k_{slow} = 0.87 ± 0.09 s^{-1}, 67%

(continued)

Figure 8.2: continued

amplitude), 26-mer (k_{fast} = 44 ± 5 s^{-1}, 43% amplitude; k_{slow} = 0.89 ± 0.09 s^{-1}, 57% amplitude), and 25-mer (k_{fast} = 9 ± 4 s^{-1}, 20% amplitude; k_{slow} = 0.15 ± 0.04 s^{-1}, 80% amplitude). The fit was confined to data from 0 – 16 s to avoid significant contribution from the 3′→5′ PCT activity of RuvC which began at 30 s. (C) 5′-radiolabeled (*) DNA products separated into PAM distal (left) and proximal (right) fragments colored as in Figure 8.1. Initial DNA cleavage is dominant at the +3 position of tDNA by HNH and the -3 and -4 positions of ntDNA by RuvC. Minor cleavage at the -5 position of ntDNA by RuvC is not depicted. Nucleotides removed by 3′→5′ post-cleavage trimming are indicated by brackets.

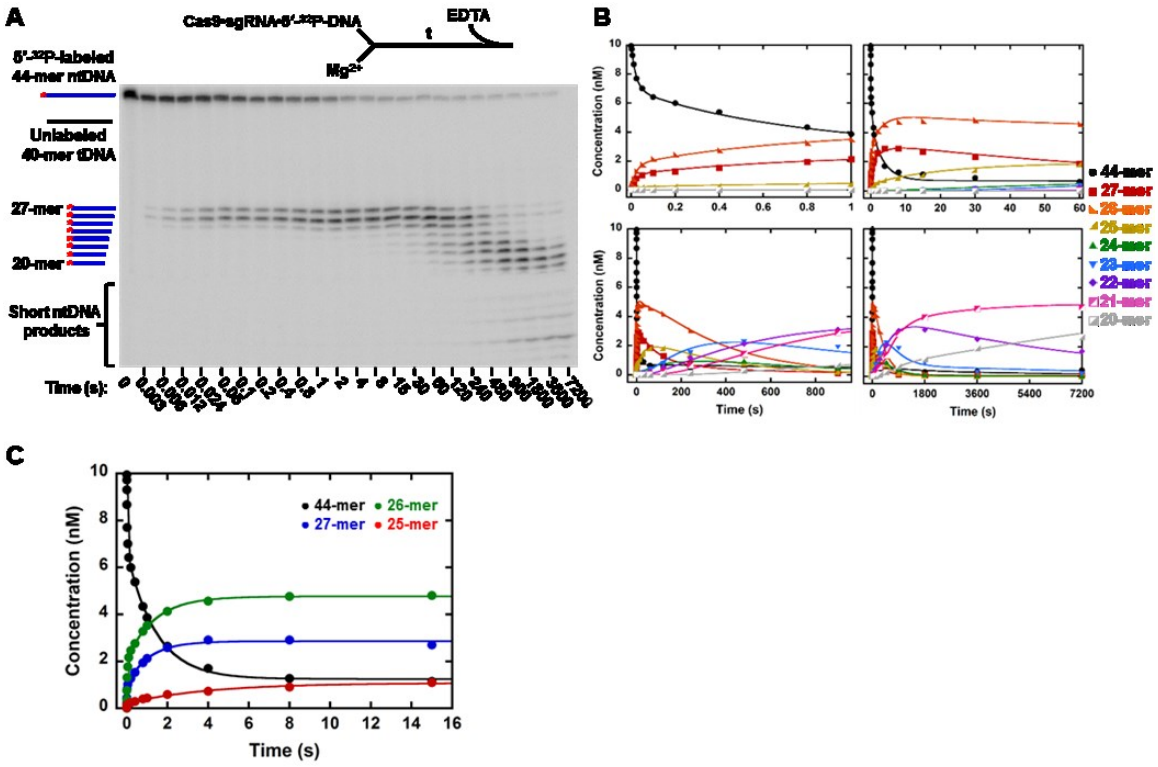


Figure 8.3: Kinetics of RuvC-catalyzed initial DNA cleavage and 3′→5′ post-cleavage trimming.

(A) Gel image showing time-dependent (0.003 – 7,200 s) cleavage of DNA (10 nM) by Cas9•sgRNA (100 nM) in the absence of the trap DNA. The DNA substrate (Figure 8.1) was 5′-radiolabeled (*) on ntDNA. Note the presence of short (<20-mer) ntDNA products. (B) RuvC substrate (44-mer) and products (27-mer – 20-mer) from Figure 8.3A were quantified and cleavage kinetics were simulated using KinTek Explorer global fitting software (solid lines). Separate panels feature different time ranges for clarity. Simulated kinetic parameters for initial DNA cleavage and 3′→5′ post-cleavage trimming (continued)

Figure 8.3: continued

can be found in Table 8.1 and Table 8.2, respectively. (C) DNA substrate (44-mer) and products (27-mer – 25-mer) from Figure 8.3A were quantified and fit to Eq. 8.1 using non-linear regression analysis to yield kinetic parameters of the fast and slow phases of initial DNA cleavage for the 44-mer (k_{fast} = 43 ± 3 s^{-1}, 38% amplitude; k_{slow} = 0.70 ± 0.03 s^{-1}, 62% amplitude), 27-mer (k_{fast} = 50 ± 8 s^{-1}, 33% amplitude; k_{slow} = 0.92 ± 0.09 s^{-1}, 67% amplitude), 26-mer (k_{fast} = 36 ± 2 s^{-1}, 43% amplitude; k_{slow} = 0.74 ± 0.03 s^{-1}, 57% amplitude), and 25-mer (k_{fast} = 15 ± 7 s^{-1}, 21% amplitude; k_{slow} = 0.22 ± 0.04 s^{-1}, 79% amplitude). The fit was confined to data from 0 – 16 s to avoid significant contribution from the 3′→5′ PCT activity of RuvC which began at 30 s.

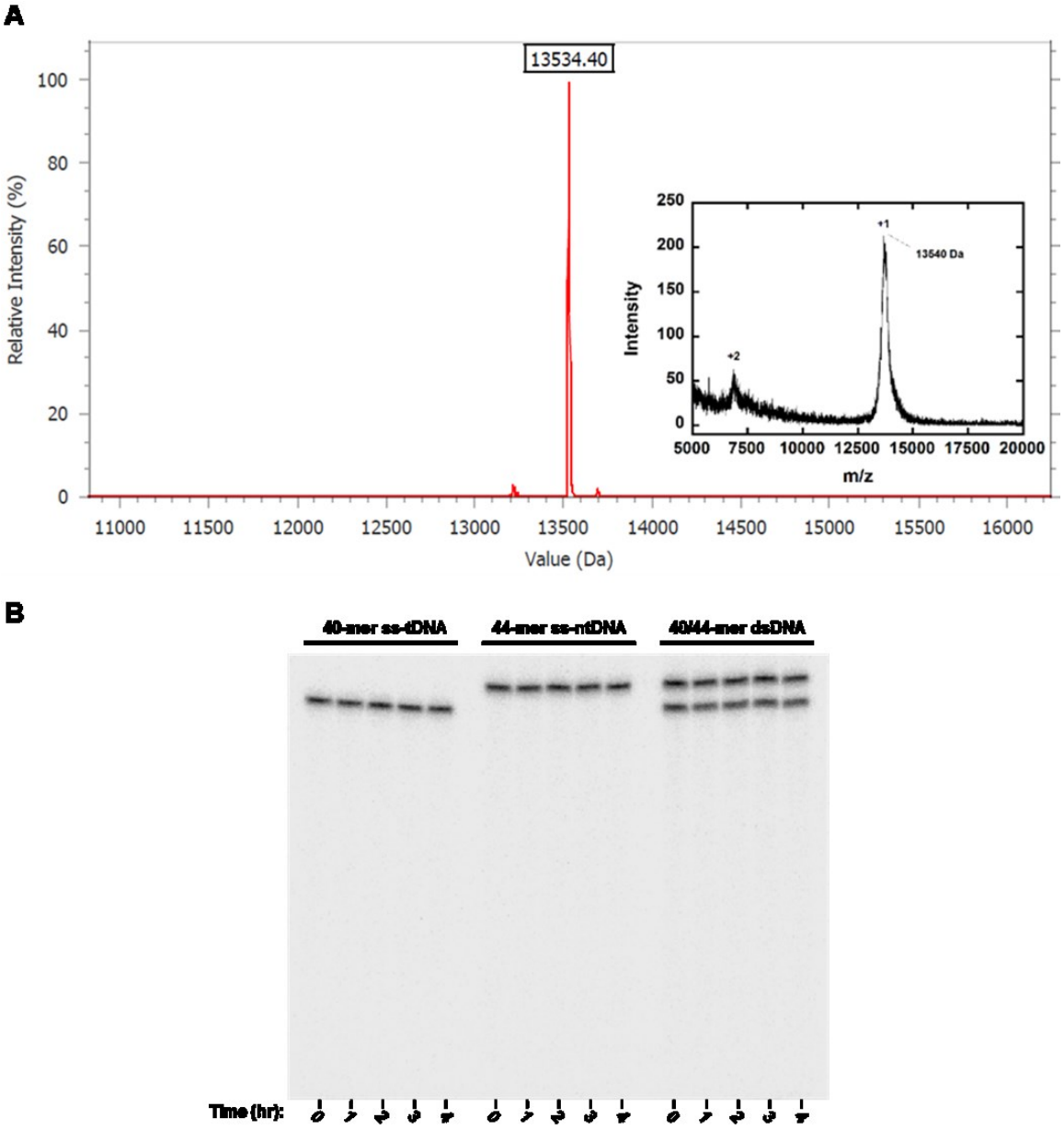


Figure 8.4: Quality controls of DNA and protein preparations.

(A) Mass spectrum of 44-mer ntDNA before purification by denaturing PAGE (see Section 8.4.2) acquired using electrospray ionization mass spectrometry (ESI-

(continued)

Figure 8.4: continued

MS) by Integrated DNA Technologies. Measured mass of 44-mer ntDNA (13534.40 Da) was determined to be in good agreement with the calculated mass (13552.9 Da) and very little 43-mer ntDNA contaminant was detected. Inset shows mass spectrum of the 44-mer ntDNA after purification by denaturing PAGE acquired using matrix assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry and features the absence of lower molecular weight contaminants. (B) Denaturing PAGE analysis of 5′-radiolabeled single- (40-mer ss-tDNA and 44-mer ss-ntDNA) and double-stranded (40/44mer dsDNA) DNA (10 nM) integrities following incubation at 37 °C with apo-Cas9 (100 nM) in the presence of $MgCl_2$ (6 mM) for various times (0 – 4 hrs). No degradation of single- or double-stranded DNA was observed even after 4 hrs suggesting that the Cas9 protein preparation was free of nuclease contamination.

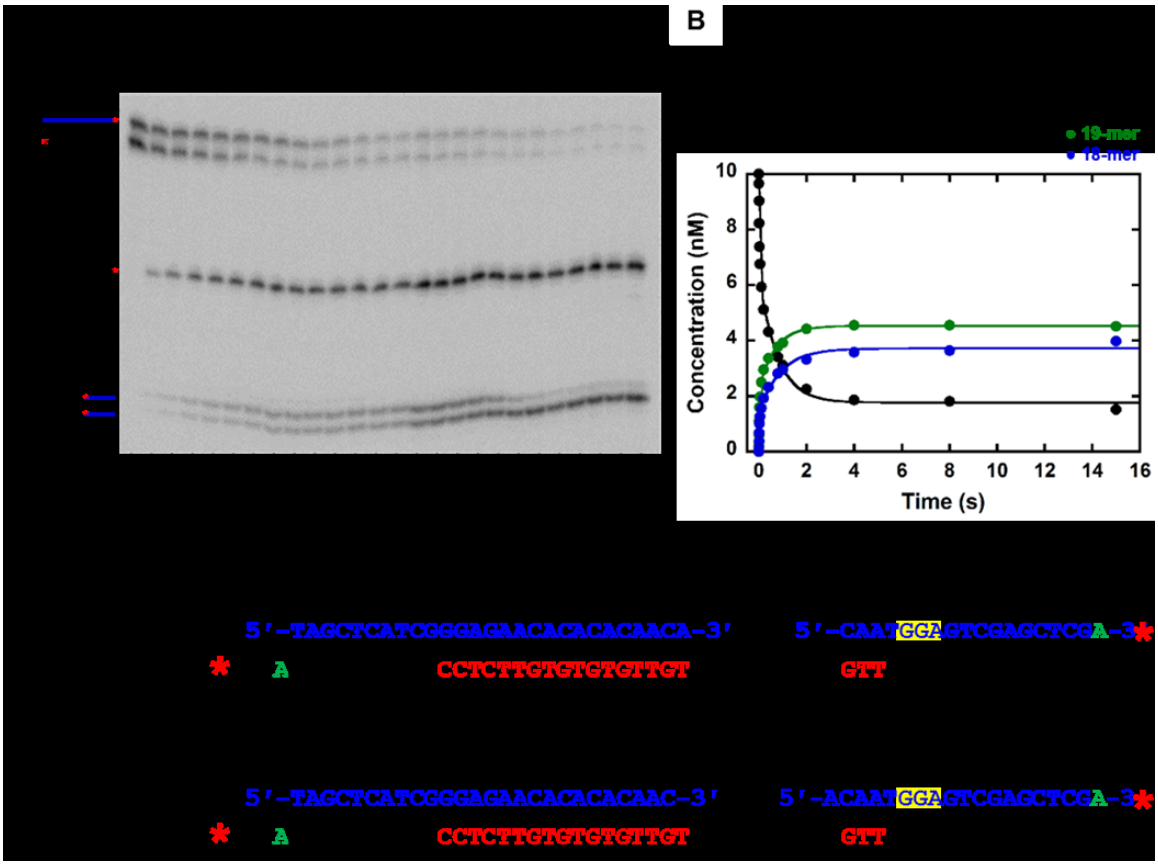


Figure 8.5: Single-turnover DNA cleavage of 3′-radiolabeled DNA by Cas9.

(A) Gel image showing time-dependent (0.003 – 14,400 s) cleavage of 3′-radiolabeled (*) 41/45-mer DNA (10 nM) by Cas9•sgRNA (100 nM). The addition of [α-^{32}P]-3′-dATP to both non-radiolabeled 40- and 44-mer by terminal deoxynucleotidyl transferase (TdT) yielded the 3′-radiolabeled 41- and 45-mer, respectively. Signal for accumulation of a 20-mer DNA product corresponding to cleavage at the -5 position of ntDNA was too weak to accurately quantify. (B) DNA substrate (45-mer) and products (19-mer – 18-mer) from Figure 8.5A were quantified and fit to Eq. 8.1 using non-linear regression analysis to

(continued)

Figure 8.5: continued

yield kinetic parameters of the fast and slow phases of DNA cleavage for the 45-mer (k_{fast} = 45 ± 7 s^{-1}, 44% amplitude; k_{slow} = 1.3 ± 0.1 s^{-1}, 56% amplitude), 19-mer (k_{fast} = 45 ± 9 s^{-1}, 38% amplitude; k_{slow} = 1.1 ± 0.1 s^{-1}, 62% amplitude), and 18-mer (k_{fast} = 46 ± 5 s^{-1}, 49% amplitude; k_{slow} = 1.5 ± 0.2 s^{-1}, 51% amplitude). The fit was confined to data from 0 – 16 s to avoid significant contribution from the 5′→3′ PCT activity of RuvC which began at 30 s. (C) 3′-radiolabeled (*) DNA products separated into PAM distal (left) and proximal (right) fragments colored as in Figure 8.1. Initial DNA cleavage is dominant at the +3 position of tDNA by HNH and the -3 and -4 positions of ntDNA by RuvC. Minor cleavage at the -5 position of ntDNA by RuvC is not depicted. Nucleotide removal by novel 5′→3′ post-cleavage trimming (Figure 8.5A) is indicated by a bracket. The addition of [α-^{32}P]-3′-dATP (green, A) to both non-radiolabeled 40- and 44-mer by TdT yielded the 3′-radiolabeled 41- and 45-mer, respectively (see Section 8.4.1).

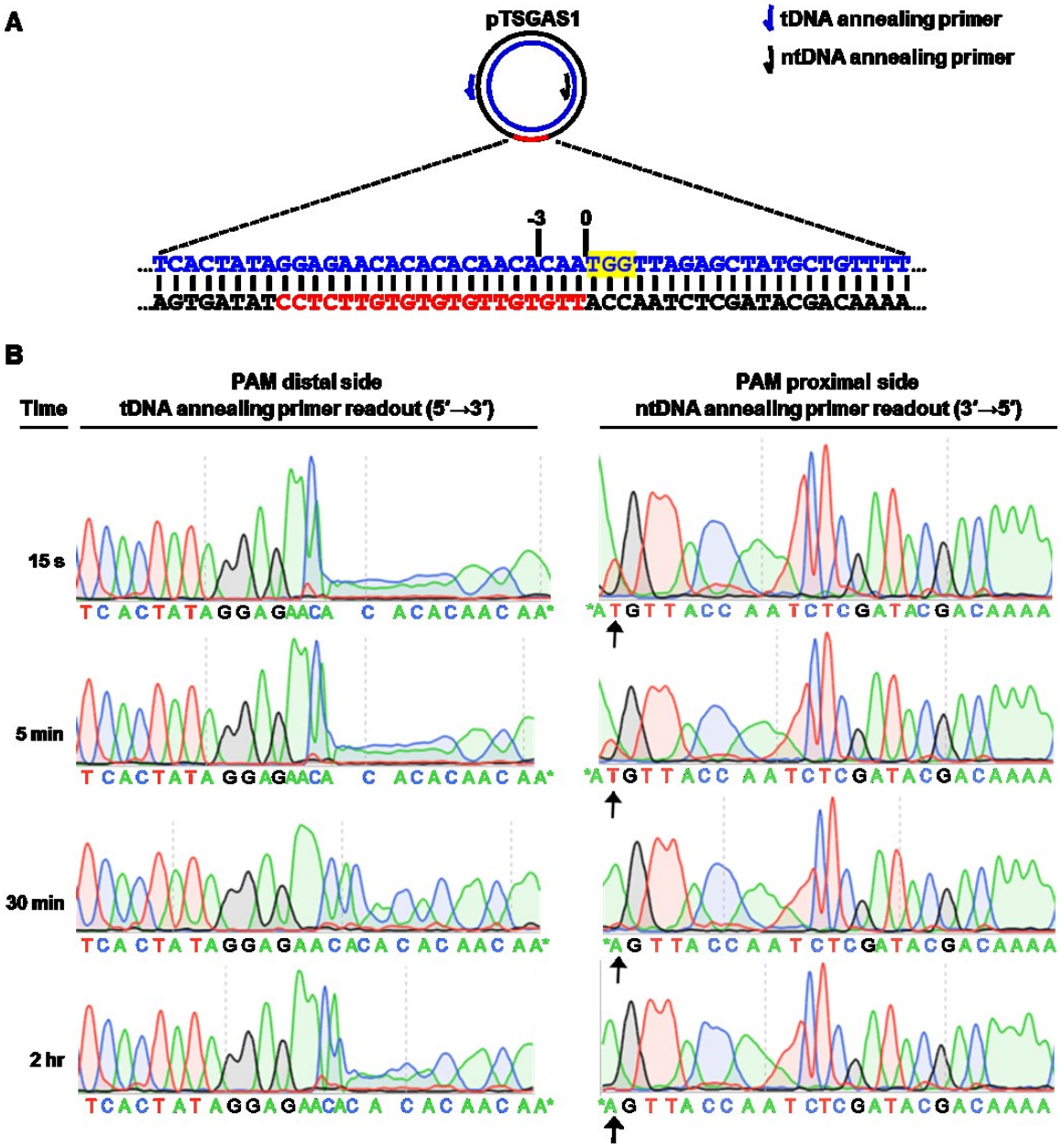


Figure 8.6: Sequencing of plasmid DNA cleaved by Cas9 for various times.

(A) Schematic representation of the plasmid DNA substrate (pTSGAS1) targeted for cleavage by Cas9 (colored as in Figure 8.1). tDNA and ntDNA sequencing primers (blue

(continued)

Figure 8.6: continued

and black arrows, respectively) are shown in their approximate annealing positions on the plasmid. (B) Sanger sequencing chromatograms of plasmid DNA cleavage products generated by Cas9 after various times (15 s – 2 hrs). Termination of primer extension indicates positions of tDNA (left) and ntDNA (right) cleavage. The terminal base A* is a consequence of DNA sequencing using modified *Thermus aquaticus* (*Taq*) DNA polymerase which catalyzes blunt-end addition of adenosine. The ntDNA sequencing readouts (right) are depicted in the 3′→5′ direction for direct comparison to Figure 8.6A. The arrow emphasizes the disappearance of the 3′ thymidine over time and the shift of the terminal base A* by one base at later time points (30 min – 2 hr) as the templating adenosine at the 5′ end of the PAM proximal ntDNA was removed by the 5′→3′ PCT activity of RuvC during the plasmid DNA cleavage assay.

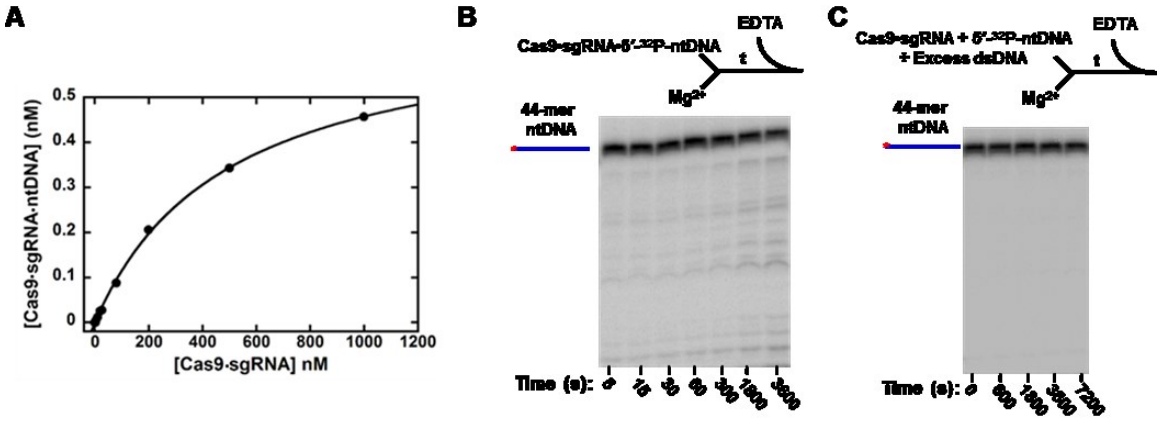


Figure 8.7: Single-stranded ntDNA binding and cleavage by Cas9•sgRNA.

(A) An electrophoretic mobility shift assay was performed to measure the equilibrium dissociation constant (K_d^{DNA}) of single-stranded ntDNA from Cas9•sgRNA•ntDNA. 5′-radiolabeled single-stranded ntDNA (0.5 nM) was incubated for 1 hr with increasing concentrations of Cas9•sgRNA (1 – 1,000 nM) in the absence of $MgCl_2$ before separation by native PAGE. The data were fit to Eq. 8.2 (solid line) to obtain K_d^{DNA} = 490 ± 30 nM. (B) A Gel image showing cleavage of single-stranded ntDNA by Cas9•sgRNA. Briefly, Cas9•sgRNA (100 nM) was pre-incubated with 5′-radiolabeled ntDNA (10 nM) for 1 hr prior to mixing with $MgCl_2$ (6 mM) for various times. Note the non-specific, weak single-stranded DNA cleavage activity of Cas9•sgRNA. (C) Validation of double-stranded on-target DNA as an effective trap for Cas9•sgRNA. Briefly, Cas9•sgRNA (100 nM) was incubated with 5′-radiolabeled ntDNA (10 nM) and a DNA trap (1,000 nM) consisting of unlabeled 40-mer/44-mer DNA (Figure 8.1) for 1 hr prior to mixing with $MgCl_2$ (6 mM) for various times. The lack of ntDNA cleavage products indicates that the

(continued)

Figure 8.7: continued

large excess of the trap DNA effectively sequestered Cas9•sgRNA and prevented cleavage of 5′-radiolabeled ntDNA.

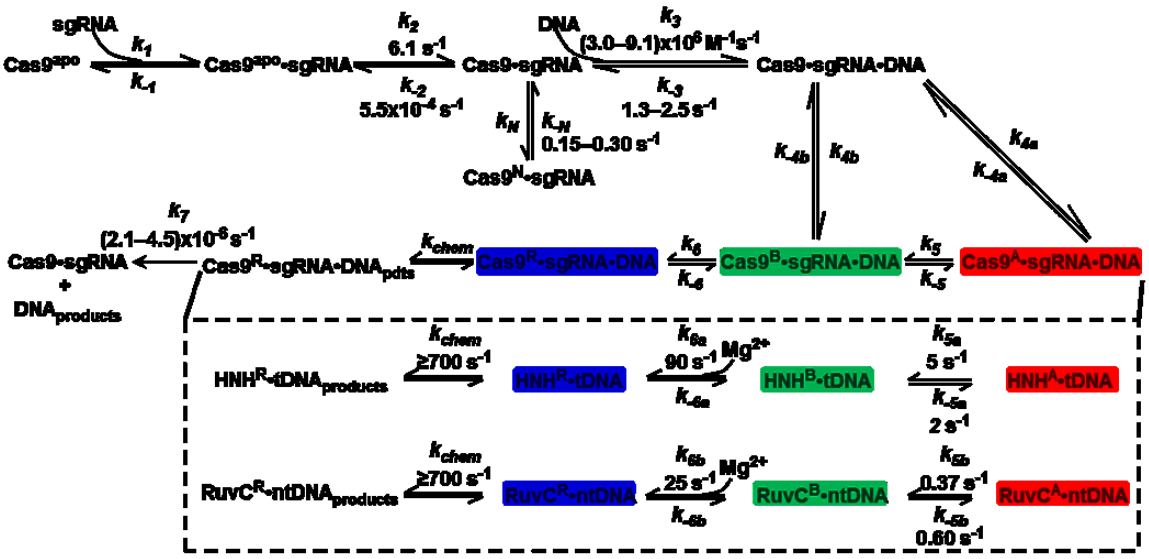


Figure 8.8: Minimal kinetic mechanism of Cas9-catalysis from our preceding publication.

Kinetic parameters for each step of Cas9 catalysis were measured using pre-steady-state kinetic methods as outlined in our preceding publication.[320] The dashed box highlights the kinetic contribution of conformational dynamics of HNH and RuvC domains to DNA cleavage. The depiction of Mg^{2+} at the two steps in the box denotes the Mg^{2+}-binding-coupled conformational changes of HNH and RuvC, rather than second-order Mg^{2+} binding.

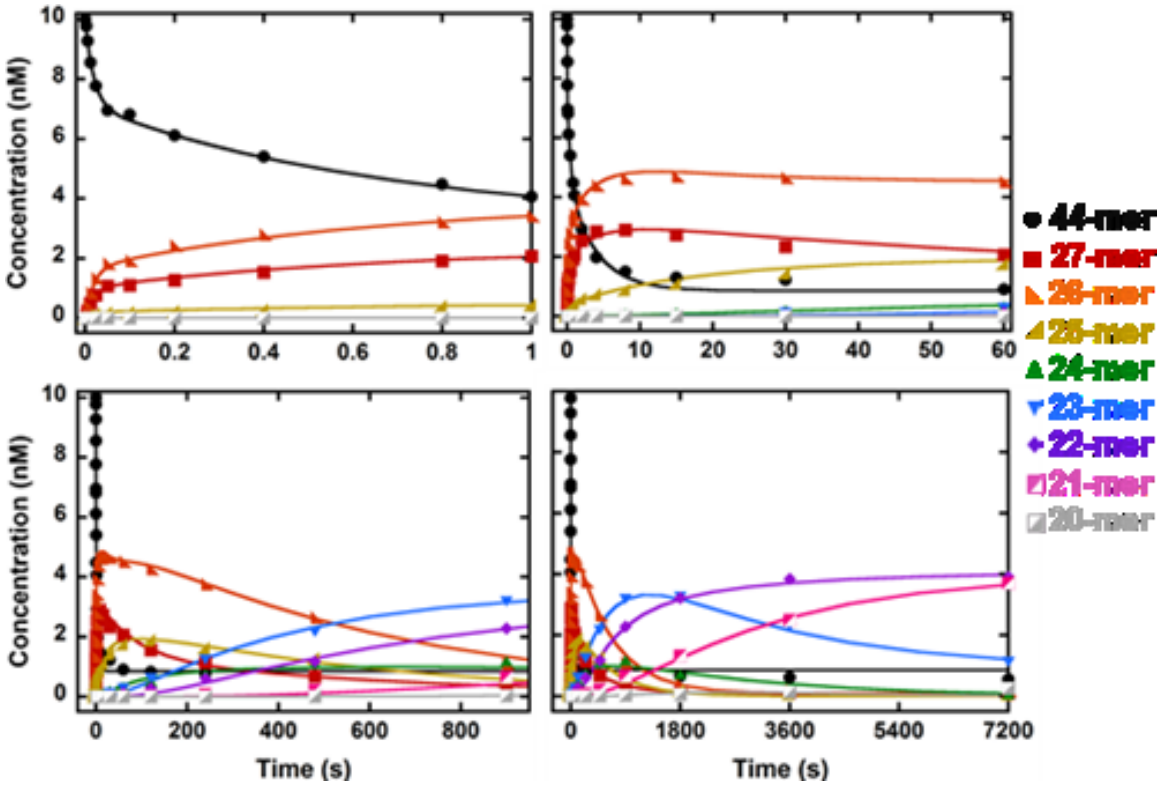


Figure 8.9: Kinetics of initial DNA cleavage and 3′→5′ post-cleavage trimming by RuvC of Cas9 in the presence of the DNA trap.

RuvC substrate (44-mer) and products (27-mer – 20-mer) from Figure 8.2A were quantified and cleavage kinetics were simulated using KinTek Explorer global fitting software (solid lines). Separate panels feature different time ranges for clarity. Simulated kinetic parameters for initial DNA cleavage and 3′→5′ post-cleavage trimming can be found in Table 8.1 and Table 8.2, respectively.

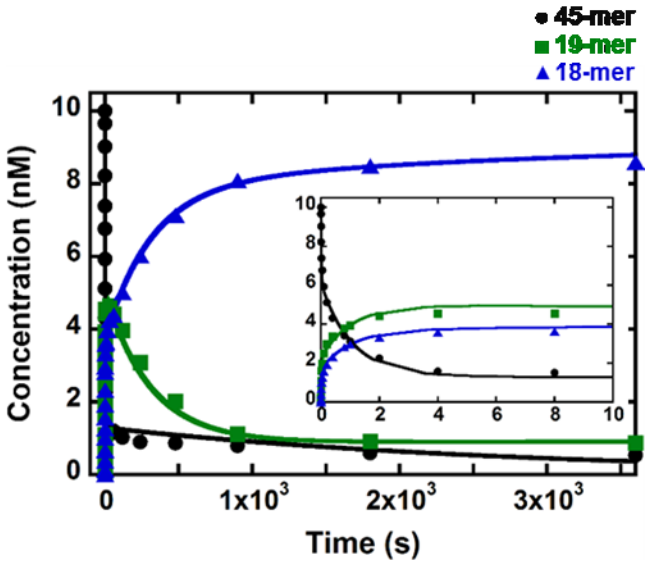


Figure 8.10: Kinetics of initial DNA cleavage and 5′→3′ post-cleavage trimming by RuvC of Cas9.

RuvC substrate (45-mer) and products (18-mer and 19-mer) from Figure 8.5A were quantified and cleavage kinetics were simulated using KinTek Explorer global fitting software (solid lines). Inset features fast time points. Simulated kinetic parameters for initial DNA cleavage and 5′→3′ post-cleavage trimming can be found in Table 8.1 and Table 8.2, respectively.

References

1. Palermo, G. et al. Catalytic metal ions and enzymatic processing of DNA and RNA. *Acc. Chem. Res.* **48**, 220-8 (2015).
2. Yang, W., Lee, J.Y. & Nowotny, M. Making and breaking nucleic acids: two-Mg2+-ion catalysis and substrate specificity. *Mol Cell* **22**, 5-13 (2006).
3. Steitz, T.A. & Steitz, J.A. A general two-metal-ion mechanism for catalytic RNA. *Proc Natl Acad Sci U S A* **90**, 6498-502 (1993).
4. Johnson, K.A. Conformational coupling in DNA polymerase fidelity. *Annu Rev Biochem* **62**, 685-713 (1993).
5. Joyce, C.M. & Benkovic, S.J. DNA polymerase fidelity: kinetics, structure, and checkpoints. *Biochemistry* **43**, 14317-24 (2004).
6. Showalter, A.K. & Tsai, M.D. A reexamination of the nucleotide incorporation fidelity of DNA polymerases. *Biochemistry* **41**, 10571-6 (2002).
7. Rothwell, P.J., Mitaksov, V. & Waksman, G. Motions of the fingers subdomain of klentaq1 are fast and not rate limiting: implications for the molecular basis of fidelity in DNA polymerases. *Mol Cell* **19**, 345-55 (2005).
8. Rothwell, P.J. & Waksman, G. Structure and mechanism of DNA polymerases. *Adv Protein Chem* **71**, 401-40 (2005).
9. Raper, A.T., Gadkari, V.V., Maxwell, B.A. & Suo, Z. Single-Molecule Investigation of Response to Oxidative DNA Damage by a Y-Family DNA Polymerase. *Biochemistry* **55**, 2187-96 (2016).
10. Raper, A.T. & Suo, Z. Investigation of Intradomain Motions of a Y-Family DNA Polymerase during Substrate Binding and Catalysis. *Biochemistry* **55**, 5832-5844 (2016).
11. Xu, C., Maxwell, B.A., Brown, J.A., Zhang, L. & Suo, Z. Global conformational dynamics of a Y-family DNA polymerase during catalysis. *PLoS Biol* **7**, e1000225 (2009).
12. Maxwell, B.A., Xu, C. & Suo, Z. DNA lesion alters global conformational dynamics of Y-family DNA polymerase during catalysis. *J Biol Chem* **287**, 13040-7 (2012).
13. Maxwell, B.A., Xu, C. & Suo, Z. Conformational dynamics of a Y-family DNA polymerase during substrate binding and catalysis as revealed by interdomain Forster resonance energy transfer. *Biochemistry* **53**, 1768-78 (2014).
14. Xu, C., Maxwell, B.A. & Suo, Z. Conformational dynamics of Thermus aquaticus DNA polymerase I during catalysis. *J Mol Biol* **426**, 2901-17 (2014).
15. Bessman, M.J., Kornberg, A., Lehman, I.R. & Simms, E.S. Enzymic synthesis of deoxyribonucleic acid. *Biochim Biophys Acta* **21**, 197-8 (1956).

16. Lehman, I.R., Bessman, M.J., Simms, E.S. & Kornberg, A. Enzymatic synthesis of deoxyribonucleic acid. I. Preparation of substrates and partial purification of an enzyme from Escherichia coli. *J Biol Chem* **233**, 163-70 (1958).
17. Joyce, C.M. & Steitz, T.A. Function and structure relationships in DNA polymerases. *Annu Rev Biochem* **63**, 777-822 (1994).
18. Braithwaite, D.K. & Ito, J. Compilation, alignment, and phylogenetic relationships of DNA polymerases. *Nucleic Acids Res* **21**, 787-802 (1993).
19. Garcia-Diaz, M. & Bebenek, K. Multiple functions of DNA polymerases. *CRC Crit. Rev. Plant Sci.* **26**, 105-122 (2007).
20. Yang, W. & Gao, Y. Translesion and Repair DNA Polymerases: Diverse Structure and Mechanism. *Annu Rev Biochem* (2018).
21. Beard, W.A. & Wilson, S.H. Structure and mechanism of DNA polymerase beta. *Biochemistry* **53**, 2768-80 (2014).
22. Maxwell, B.A. & Suo, Z. Recent insight into the kinetic mechanisms and conformational dynamics of Y-Family DNA polymerases. *Biochemistry* **53**, 2804-14 (2014).
23. Fowler, J.D. & Suo, Z. Biochemical, structural, and physiological characterization of terminal deoxynucleotidyl transferase. *Chem Rev* **106**, 2092-110 (2006).
24. Berdis, A.J. Mechanisms of DNA polymerases. *Chem Rev* **109**, 2862-79 (2009).
25. Steitz, T.A. DNA polymerases: structural diversity and common mechanisms. *J Biol Chem* **274**, 17395-8 (1999).
26. Steitz, T.A. A mechanism for all polymerases. *Nature* **391**, 231-2 (1998).
27. Tsai, Y.C. & Johnson, K.A. A new paradigm for DNA polymerase specificity. *Biochemistry* **45**, 9675-87 (2006).
28. Xiang, Y., Goodman, M.F., Beard, W.A., Wilson, S.H. & Warshel, A. Exploring the role of large conformational changes in the fidelity of DNA polymerase beta. *Proteins* **70**, 231-47 (2008).
29. Wong, I., Patel, S.S. & Johnson, K.A. An induced-fit kinetic mechanism for DNA replication fidelity: direct measurement by single-turnover kinetics. *Biochemistry* **30**, 526-37 (1991).
30. Freemont, P.S., Friedman, J.M., Beese, L.S., Sanderson, M.R. & Steitz, T.A. Cocrystal structure of an editing complex of Klenow fragment with DNA. *Proc Natl Acad Sci U S A* **85**, 8924-8 (1988).
31. Beese, L.S. & Steitz, T.A. Structural basis for the 3'-5' exonuclease activity of Escherichia coli DNA polymerase I: a two metal ion mechanism. *Embo J* **10**, 25-33 (1991).
32. Batra, V.K. et al. Magnesium-induced assembly of a complete DNA polymerase catalytic complex. *Structure* **14**, 757-66 (2006).
33. Brautigam, C.A. & Steitz, T.A. Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes. *Curr Opin Struct Biol* **8**, 54-63 (1998).
34. Nakamura, T.Z., Ye; Yamagata, Yuriko; Hua, Yue-jin; Yang, Wei;. Mechanism of the nucleotidyl-transfer reaction in DNA polymerase revealed by time-resolved crystallography. *Biophysics* **9**, 31-36 (2013).

35. Yang, W. Damage repair DNA polymerases Y. *Curr Opin Struct Biol* **13**, 23-30 (2003).
36. Yang, W. An equivalent metal ion in one- and two-metal-ion catalysis. *Nat Struct Mol Biol* **15**, 1228-31 (2008).
37. Derbyshire, V. et al. Genetic and crystallographic studies of the 3',5'-exonucleolytic site of DNA polymerase I. *Science* **240**, 199-201 (1988).
38. Nakamura, T., Zhao, Y., Yamagata, Y., Hua, Y.J. & Yang, W. Watching DNA polymerase eta make a phosphodiester bond. *Nature* **487**, 196-201 (2012).
39. Gao, Y. & Yang, W. Capture of a third Mg(2)(+) is essential for catalyzing DNA synthesis. *Science* **352**, 1334-7 (2016).
40. Freudenthal, B.D., Beard, W.A., Shock, D.D. & Wilson, S.H. Observing a DNA polymerase choose right from wrong. *Cell* **154**, 157-68 (2013).
41. Freudenthal, B.D. et al. Uncovering the polymerase-induced cytotoxicity of an oxidized nucleotide. *Nature* **517**, 635-9 (2015).
42. Vyas, R., Reed, A.J., Tokarsky, E.J. & Suo, Z. Viewing Human DNA Polymerase beta Faithfully and Unfaithfully Bypass an Oxidative Lesion by Time-Dependent Crystallography. *J Am Chem Soc* **137**, 5225-30 (2015).
43. Reed, A.J., Vyas, R., Raper, A.T. & Suo, Z. Structural Insights into the Post-Chemistry Steps of Nucleotide Incorporation Catalyzed by a DNA Polymerase. *J Am Chem Soc* **139**, 465-471 (2017).
44. Whitaker, A.M., Smith, M.R., Schaich, M.A. & Freudenthal, B.D. Capturing a mammalian DNA polymerase extending from an oxidized nucleotide. *Nucleic Acids Res* **45**, 6934-6944 (2017).
45. Reed, A.J. & Suo, Z. Time-Dependent Extension from an 8-Oxoguanine Lesion by Human DNA Polymerase Beta. *J Am Chem Soc* **139**, 9684-9690 (2017).
46. Jamsen, J.A. et al. Time-lapse crystallography snapshots of a double-strand break repair polymerase in action. *Nat. Commun.* **8**, 253 (2017).
47. Yang, W., Weng, P.J. & Gao, Y. A new paradigm of DNA synthesis: three-metal-ion catalysis. *Cell Biosci* **6**, 51 (2016).
48. Freudenthal, B.D., Beard, W.A. & Wilson, S.H. New structural snapshots provide molecular insights into the mechanism of high fidelity DNA synthesis. *DNA Repair (Amst)* **32**, 3-9 (2015).
49. Genna, V., Gaspari, R., Dal Peraro, M. & De Vivo, M. Cooperative motion of a key positively charged residue and metal ions for DNA replication catalyzed by human DNA Polymerase-eta. *Nucleic Acids Res* **44**, 2827-36 (2016).
50. Perera, L. et al. Requirement for transient metal ions revealed through computational analysis for DNA polymerase going in reverse. *Proc Natl Acad Sci U S A* **112**, E5228-36 (2015).
51. Perera, L., Freudenthal, B.D., Beard, W.A., Pedersen, L.G. & Wilson, S.H. Revealing the role of the product metal in DNA polymerase beta catalysis. *Nucleic Acids Res* **45**, 2736-2745 (2017).
52. Liu, X., Bushnell, D.A. & Kornberg, R.D. RNA polymerase II transcription: structure and mechanism. *Biochim Biophys Acta* **1829**, 2-8 (2013).

53. Svetlov, V. & Nudler, E. Basic mechanism of transcription by RNA polymerase II. *Biochim Biophys Acta* **1829**, 20-8 (2013).
54. Hahn, S. Structure and mechanism of the RNA polymerase II transcription machinery. *Nat Struct Mol Biol* **11**, 394-403 (2004).
55. Kochetkov, S.N., Rusakova, E.E. & Tunitskaya, V.L. Recent studies of T7 RNA polymerase mechanism. *FEBS Lett* **440**, 264-7 (1998).
56. Ebright, R.H. RNA polymerase: structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. *J Mol Biol* **304**, 687-98 (2000).
57. Moon, A.F. et al. The X family portrait: structural insights into biological functions of X family polymerases. *DNA Repair (Amst)* **6**, 1709-25 (2007).
58. Burgers, P.M.J. & Kunkel, T.A. Eukaryotic DNA Replication Fork. *Annu Rev Biochem* **86**, 417-438 (2017).
59. Kohlstaedt, L.A., Wang, J., Friedman, J.M., Rice, P.A. & Steitz, T.A. Crystal structure at 3.5 A resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* **256**, 1783-90 (1992).
60. Steitz, T.A. et al. Two DNA polymerases: HIV reverse transcriptase and the Klenow fragment of Escherichia coli DNA polymerase I. *Cold Spring Harb Symp Quant Biol* **58**, 495-504 (1993).
61. Pelletier, H., Sawaya, M.R., Kumar, A., Wilson, S.H. & Kraut, J. Structures of ternary complexes of rat DNA polymerase beta, a DNA template-primer, and ddCTP. *Science* **264**, 1891-903 (1994).
62. Sawaya, M.R., Pelletier, H., Kumar, A., Wilson, S.H. & Kraut, J. Crystal structure of rat DNA polymerase beta: evidence for a common polymerase mechanism. *Science* **264**, 1930-5 (1994).
63. Doublie, S., Tabor, S., Long, A.M., Richardson, C.C. & Ellenberger, T. Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 A resolution. *Nature* **391**, 251-8 (1998).
64. Li, Y., Kong, Y., Korolev, S. & Waksman, G. Crystal structures of the Klenow fragment of Thermus aquaticus DNA polymerase I complexed with deoxyribonucleoside triphosphates. *Protein Sci* **7**, 1116-23 (1998).
65. Li, Y., Korolev, S. & Waksman, G. 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**, 7514-25 (1998).
66. Korolev, S., Nayal, M., Barnes, W.M., Di Cera, E. & Waksman, G. Crystal structure of the large fragment of Thermus aquaticus DNA polymerase I at 2.5-A resolution: structural basis for thermostability. *Proc Natl Acad Sci U S A* **92**, 9264-8 (1995).
67. Wang, J., Yu, P., Lin, T.C., Konigsberg, W.H. & Steitz, T.A. Crystal structures of an NH2-terminal fragment of T4 DNA polymerase and its complexes with single-stranded DNA and with divalent metal ions. *Biochemistry* **35**, 8110-9 (1996).
68. Franklin, M.C., Wang, J. & Steitz, T.A. Structure of the replicating complex of a pol alpha family DNA polymerase. *Cell* **105**, 657-67 (2001).

69. Hogg, M., Wallace, S.S. & Doublie, S. Crystallographic snapshots of a replicative DNA polymerase encountering an abasic site. *EMBO J* **23**, 1483-93 (2004).
70. Wang, M. et al. Insights into base selectivity from the 1.8 A resolution structure of an RB69 DNA polymerase ternary complex. *Biochemistry* **50**, 581-90 (2011).
71. Evans, R.J. et al. Structure of PolC reveals unique DNA binding and fidelity determinants. *Proc Natl Acad Sci U S A* **105**, 20695-700 (2008).
72. Bailey, S., Wing, R.A. & Steitz, T.A. The structure of T. aquaticus DNA polymerase III is distinct from eukaryotic replicative DNA polymerases. *Cell* **126**, 893-904 (2006).
73. Wing, R.A., Bailey, S. & Steitz, T.A. Insights into the replisome from the structure of a ternary complex of the DNA polymerase III alpha-subunit. *J Mol Biol* **382**, 859-69 (2008).
74. Barros, T. et al. A structural role for the PHP domain in E. coli DNA polymerase III. *BMC Struct Biol* **13**, 8 (2013).
75. Pelletier, H., Sawaya, M.R., Wolfle, W., Wilson, S.H. & Kraut, J. A structural basis for metal ion mutagenicity and nucleotide selectivity in human DNA polymerase beta. *Biochemistry* **35**, 12762-77 (1996).
76. Garcia-Diaz, M., Bebenek, K., Krahn, J.M., Kunkel, T.A. & Pedersen, L.C. A closed conformation for the Pol lambda catalytic cycle. *Nat Struct Mol Biol* **12**, 97-8 (2005).
77. Moon, A.F. et al. Structural insight into the substrate specificity of DNA Polymerase mu. *Nat Struct Mol Biol* **14**, 45-53 (2007).
78. Pelletier, H., Sawaya, M.R., Wolfle, W., Wilson, S.H. & Kraut, J. Crystal structures of human DNA polymerase beta complexed with DNA: implications for catalytic mechanism, processivity, and fidelity. *Biochemistry* **35**, 12742-61 (1996).
79. Sawaya, M.R., Prasad, R., Wilson, S.H., Kraut, J. & Pelletier, H. Crystal structures of human DNA polymerase beta complexed with gapped and nicked DNA: evidence for an induced fit mechanism. *Biochemistry* **36**, 11205-15 (1997).
80. Biertumpfel, C. et al. Structure and mechanism of human DNA polymerase eta. *Nature* **465**, 1044-8 (2010).
81. Irimia, A., Eoff, R.L., Guengerich, F.P. & Egli, M. Structural and functional elucidation of the mechanism promoting error-prone synthesis by human DNA polymerase kappa opposite the 7,8-dihydro-8-oxo-2'-deoxyguanosine adduct. *J Biol Chem* **284**, 22467-80 (2009).
82. Nair, D.T., Johnson, R.E., Prakash, S., Prakash, L. & Aggarwal, A.K. Replication by human DNA polymerase-iota occurs by Hoogsteen base-pairing. *Nature* **430**, 377-80 (2004).
83. Swan, M.K., Johnson, R.E., Prakash, L., Prakash, S. & Aggarwal, A.K. Structure of the human Rev1-DNA-dNTP ternary complex. *J Mol Biol* **390**, 699-709 (2009).
84. Ling, H., Boudsocq, F., Woodgate, R. & Yang, W. Crystal structure of a Y-family DNA polymerase in action: a mechanism for error-prone and lesion-bypass replication. *Cell* **107**, 91-102 (2001).

85. Trincao, J. et al. Structure of the catalytic core of S. cerevisiae DNA polymerase eta: implications for translesion DNA synthesis. *Mol Cell* **8**, 417-26 (2001).
86. Ummat, A. et al. Human DNA polymerase eta is pre-aligned for dNTP binding and catalysis. *J Mol Biol* **415**, 627-34 (2012).
87. Gaur, V. et al. Structural and kinetic insights into binding and incorporation of L-nucleotide analogs by a Y-family DNA polymerase. *Nucleic Acids Res* **42**, 9984-95 (2014).
88. Huang, H., Chopra, R., Verdine, G.L. & Harrison, S.C. Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science* **282**, 1669-75 (1998).
89. Kuchta, R.D., Mizrahi, V., Benkovic, P.A., Johnson, K.A. & Benkovic, S.J. Kinetic mechanism of DNA polymerase I (Klenow). *Biochemistry* **26**, 8410-7 (1987).
90. Patel, S.S., Wong, I. & Johnson, K.A. Pre-steady-state kinetic analysis of processive DNA replication including complete characterization of an exonuclease-deficient mutant. *Biochemistry* **30**, 511-25 (1991).
91. Zahurancik, W.J., Klein, S.J. & Suo, Z. Kinetic mechanism of DNA polymerization catalyzed by human DNA polymerase epsilon. *Biochemistry* **52**, 7041-9 (2013).
92. Brown, J.A. & Suo, Z. Elucidating the kinetic mechanism of DNA polymerization catalyzed by Sulfolobus solfataricus P2 DNA polymerase B1. *Biochemistry* **48**, 7502-11 (2009).
93. Fiala, K.A. & Suo, Z. Pre-steady-state kinetic studies of the fidelity of Sulfolobus solfataricus P2 DNA polymerase IV. *Biochemistry* **43**, 2106-15 (2004).
94. Fiala, K.A. & Suo, Z. Mechanism of DNA polymerization catalyzed by Sulfolobus solfataricus P2 DNA polymerase IV. *Biochemistry* **43**, 2116-25 (2004).
95. Donlin, M.J., Patel, S.S. & Johnson, K.A. Kinetic partitioning between the exonuclease and polymerase sites in DNA error correction. *Biochemistry* **30**, 538-46 (1991).
96. Kati, W.M., Johnson, K.A., Jerva, L.F. & Anderson, K.S. Mechanism and fidelity of HIV reverse transcriptase. *J Biol Chem* **267**, 25988-97 (1992).
97. Beard, W.A. & Wilson, S.H. Structure and mechanism of DNA polymerase Beta. *Chem Rev* **106**, 361-82 (2006).
98. Eger, B.T. & Benkovic, S.J. Minimal kinetic mechanism for misincorporation by DNA polymerase I (Klenow fragment). *Biochemistry* **31**, 9227-36 (1992).
99. Dahlberg, M.E. & Benkovic, S.J. Kinetic mechanism of DNA polymerase I (Klenow fragment): identification of a second conformational change and evaluation of the internal equilibrium constant. *Biochemistry* **30**, 4835-43 (1991).
100. Frey, M.W., Sowers, L.C., Millar, D.P. & Benkovic, S.J. The nucleotide analog 2-aminopurine as a spectroscopic probe of nucleotide incorporation by the Klenow fragment of Escherichia coli polymerase I and bacteriophage T4 DNA polymerase. *Biochemistry* **34**, 9185-92 (1995).

101. Hsieh, J.C., Zinnen, S. & Modrich, P. Kinetic mechanism of the DNA-dependent DNA polymerase activity of human immunodeficiency virus reverse transcriptase. *J Biol Chem* **268**, 24607-13 (1993).
102. Capson, T.L. et al. Kinetic characterization of the polymerase and exonuclease activities of the gene 43 protein of bacteriophage T4. *Biochemistry* **31**, 10984-94 (1992).
103. Washington, M.T., Prakash, L. & Prakash, S. Yeast DNA polymerase eta utilizes an induced-fit mechanism of nucleotide incorporation. *Cell* **107**, 917-27 (2001).
104. Washington, M.T., Johnson, R.E., Prakash, L. & Prakash, S. The mechanism of nucleotide incorporation by human DNA polymerase eta differs from that of the yeast enzyme. *Mol Cell Biol* **23**, 8316-22 (2003).
105. Wohrl, B.M., Krebs, R., Goody, R.S. & Restle, T. Refined model for primer/template binding by HIV-1 reverse transcriptase: pre-steady-state kinetic analyses of primer/template binding and nucleotide incorporation events distinguish between different binding modes depending on the nature of the nucleic acid substrate. *J Mol Biol* **292**, 333-44 (1999).
106. Schermerhorn, K.M. & Gardner, A.F. Pre-steady-state Kinetic Analysis of a Family D DNA Polymerase from Thermococcus sp. 9 degrees N Reveals Mechanisms for Archaeal Genomic Replication and Maintenance. *J Biol Chem* **290**, 21800-10 (2015).
107. Lahiri, I., Mukherjee, P. & Pata, J.D. Kinetic characterization of exonuclease-deficient Staphylococcus aureus PolC, a C-family replicative DNA polymerase. *PLoS One* **8**, e63489 (2013).
108. Einolf, H.J. & Guengerich, F.P. Kinetic analysis of nucleotide incorporation by mammalian DNA polymerase delta. *J Biol Chem* **275**, 16316-22 (2000).
109. Fiala, K.A., Abdel-Gawad, W. & Suo, Z. Pre-Steady-State Kinetic Studies of the Fidelity and Mechanism of Polymerization Catalyzed by Truncated Human DNA Polymerase lambda. *Biochemistry* **43**, 6751-62 (2004).
110. Werneburg, B.G. et al. DNA polymerase beta: pre-steady-state kinetic analysis and roles of arginine-283 in catalysis and fidelity. *Biochemistry* **35**, 7041-50 (1996).
111. Maxwell, B.A. & Suo, Z. Single-molecule investigation of substrate binding kinetics and protein conformational dynamics of a B-family replicative DNA polymerase. *J Biol Chem* **288**, 11590-600 (2013).
112. Johnson, K.A. The kinetic and chemical mechanism of high-fidelity DNA polymerases. *Biochim Biophys Acta* **1804**, 1041-8 (2010).
113. McClure, W.R. & Jovin, T.M. The steady state kinetic parameters and non-processivity of Escherichia coli deoxyribonucleic acid polymerase I. *J Biol Chem* **250**, 4073-80 (1975).
114. Bryant, F.R., Johnson, K.A. & Benkovic, S.J. Elementary steps in the DNA polymerase I reaction pathway. *Biochemistry* **22**, 3537-46 (1983).
115. Liu, M.S. et al. Structural Mechanism for the Fidelity Modulation of DNA Polymerase lambda. *J Am Chem Soc* **138**, 2389-98 (2016).

116. Ahn, J., Kraynov, V.S., Zhong, X., Werneburg, B.G. & Tsai, M.D. DNA polymerase beta: effects of gapped DNA substrates on dNTP specificity, fidelity, processivity and conformational changes. *Biochem J* **331 (Pt 1)**, 79-87 (1998).
117. Brown, J.A., Pack, L.R., Sanman, L.E. & Suo, Z. Efficiency and fidelity of human DNA polymerases lambda and beta during gap-filling DNA synthesis. *DNA Repair (Amst)* **10**, 24-33 (2011).
118. Wu, W.J. et al. How a low-fidelity DNA polymerase chooses non-Watson-Crick from Watson-Crick incorporation. *J Am Chem Soc* **136**, 4927-37 (2014).
119. Ollis, D.L., Brick, P., Hamlin, R., Xuong, N.G. & Steitz, T.A. Structure of large fragment of Escherichia coli DNA polymerase I complexed with dTMP. *Nature* **313**, 762-6 (1985).
120. Beese, L.S., Friedman, J.M. & Steitz, T.A. Crystal structures of the Klenow fragment of DNA polymerase I complexed with deoxynucleoside triphosphate and pyrophosphate. *Biochemistry* **32**, 14095-101 (1993).
121. Chen, Y. et al. Unique 5'-P recognition and basis for dG:dGTP misincorporation of ASFV DNA polymerase X. *PLoS Biol* **15**, e1002599 (2017).
122. Nakane, S., Ishikawa, H., Nakagawa, N., Kuramitsu, S. & Masui, R. The structural basis of the kinetic mechanism of a gap-filling X-family DNA polymerase that binds Mg(2+)-dNTP before binding to DNA. *J Mol Biol* **417**, 179-96 (2012).
123. Kumar, S., Bakhtina, M. & Tsai, M.D. Altered order of substrate binding by DNA polymerase X from African Swine Fever virus. *Biochemistry* **47**, 7875-87 (2008).
124. Maciejewski, M.W. et al. Solution structure of a viral DNA repair polymerase. *Nat Struct Biol* **8**, 936-41 (2001).
125. Showalter, A.K. & Tsai, M.D. A DNA polymerase with specificity for five base pairs. *J Am Chem Soc* **123**, 1776-7 (2001).
126. Brown, J.A., Fowler, J.D. & Suo, Z. Kinetic basis of nucleotide selection employed by a protein template-dependent DNA polymerase. *Biochemistry* **49**, 5504-10 (2010).
127. Nair, D.T., Johnson, R.E., Prakash, L., Prakash, S. & Aggarwal, A.K. Rev1 employs a novel mechanism of DNA synthesis using a protein template. *Science* **309**, 2219-22 (2005).
128. Nelson, J.R., Lawrence, C.W. & Hinkle, D.C. Deoxycytidyl transferase activity of yeast REV1 protein. *Nature* **382**, 729-31 (1996).
129. Howell, C.A., Prakash, S. & Washington, M.T. Pre-steady-state kinetic studies of protein-template-directed nucleotide incorporation by the yeast Rev1 protein. *Biochemistry* **46**, 13451-9 (2007).
130. Chagovetz, A.M., Sweasy, J.B. & Preston, B.D. Increased activity and fidelity of DNA polymerase beta on single-nucleotide gapped DNA. *J Biol Chem* **272**, 27501-4 (1997).
131. Garcia-Diaz, M. et al. DNA polymerase lambda, a novel DNA repair enzyme in human cells. *J Biol Chem* **277**, 13184-91 (2002).
132. Fiala, K.A., Hypes, C.D. & Suo, Z. Mechanism of abasic lesion bypass catalyzed by a Y-family DNA polymerase. *J Biol Chem* **282**, 8188-98 (2007).

133. Brown, J.A., Newmister, S.A., Fiala, K.A. & Suo, Z. Mechanism of double-base lesion bypass catalyzed by a Y-family DNA polymerase. *Nucleic Acids Res* **36**, 3867-78 (2008).
134. Maxwell, B.A. & Suo, Z. Kinetic basis for the differing response to an oxidative lesion by a replicative and a lesion bypass DNA polymerase from Sulfolobus solfataricus. *Biochemistry* **51**, 3485-96 (2012).
135. Kusumoto, R., Masutani, C., Shimmyo, S., Iwai, S. & Hanaoka, F. DNA binding properties of human DNA polymerase eta: implications for fidelity and polymerase switching of translesion synthesis. *Genes Cells* **9**, 1139-50 (2004).
136. Ohashi, E. et al. Error-prone bypass of certain DNA lesions by the human DNA polymerase kappa. *Genes Dev* **14**, 1589-94 (2000).
137. Woodgate, R. A plethora of lesion-replicating DNA polymerases. *Genes Dev* **13**, 2191-5 (1999).
138. Sherrer, S.M. et al. Mechanistic studies of the bypass of a bulky single-base lesion catalyzed by a Y-family DNA polymerase. *J Biol Chem* **284**, 6379-88 (2009).
139. Gadkari, V.V., Tokarsky, E.J., Malik, C.K., Basu, A.K. & Suo, Z. Mechanistic investigation of the bypass of a bulky aromatic DNA adduct catalyzed by a Y-family DNA polymerase. *DNA Repair (Amst)* **21**, 65-77 (2014).
140. Garcia-Gomez, S. et al. PrimPol, an archaic primase/polymerase operating in human cells. *Mol Cell* **52**, 541-53 (2013).
141. Meng, X. et al. DNA damage alters DNA polymerase delta to a form that exhibits increased discrimination against modified template bases and mismatched primers. *Nucleic Acids Res* **37**, 647-57 (2009).
142. Schmitt, M.W., Matsumoto, Y. & Loeb, L.A. High fidelity and lesion bypass capability of human DNA polymerase delta. *Biochimie* **91**, 1163-72 (2009).
143. Higuchi, K. et al. Fate of DNA replication fork encountering a single DNA lesion during oriC plasmid DNA replication in vitro. *Genes Cells* **8**, 437-49 (2003).
144. Johnson, K.A. *In 1 Transient-State Kinetic Analysis of Enzyme Reaction Pathways*, 1-61 (Academic Press, Inc., San Diego, CA, 1992).
145. Zahurancik, W.J., Baranovskiy, A.G., Tahirov, T.H. & Suo, Z. Comparison of the kinetic parameters of the truncated catalytic subunit and holoenzyme of human DNA polymerase varepsilon. *DNA Repair (Amst)* **29**, 16-22 (2015).
146. Vande Berg, B.J., Beard, W.A. & Wilson, S.H. 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**, 3408-16 (2001).
147. Brown, J.A. et al. Pre-Steady-State Kinetic Analysis of Truncated and Full-Length Saccharomyces cerevisiae DNA Polymerase Eta. *J. Nucleic Acids* **2010**(2010).
148. Zhao, L. et al. Elucidation of kinetic mechanisms of human translesion DNA polymerase kappa using tryptophan mutants. *FEBS J* **281**, 4394-410 (2014).
149. Wong, J.H., Fiala, K.A., Suo, Z. & Ling, H. Snapshots of a Y-family DNA polymerase in replication: substrate-induced conformational transitions and implications for fidelity of Dpo4. *J Mol Biol* **379**, 317-30 (2008).

150. Lee, E., Fowler, J.D., Suo, Z. & Wu, Z. Backbone assignment of the binary complex of the full length Sulfolobus solfataricus DNA polymerase IV and DNA. *Biomol. NMR Assign.* **11**, 39-43 (2017).
151. Chu, X. et al. Dynamic conformational change regulates the protein-DNA recognition: an investigation on binding of a Y-family polymerase to its target DNA. *PLoS Comput Biol* **10**, e1003804 (2014).
152. Lone, S. et al. Human DNA polymerase kappa encircles DNA: implications for mismatch extension and lesion bypass. *Mol Cell* **25**, 601-14 (2007).
153. Moon, A.F. et al. Sustained active site rigidity during synthesis by human DNA polymerase mu. *Nat Struct Mol Biol* **21**, 253-60 (2014).
154. Lang, T., Maitra, M., Starcevic, D., Li, S.X. & Sweasy, J.B. A DNA polymerase beta mutant from colon cancer cells induces mutations. *Proc Natl Acad Sci U S A* **101**, 6074-9 (2004).
155. Brown, J.A. et al. Identification of critical residues for the tight binding of both correct and incorrect nucleotides to human DNA polymerase lambda. *J Mol Biol* **403**, 505-15 (2010).
156. Garcia-Diaz, M. et al. A structural solution for the DNA polymerase lambda-dependent repair of DNA gaps with minimal homology. *Mol Cell* **13**, 561-72 (2004).
157. Wilson, S.H. & Kunkel, T.A. Passing the baton in base excision repair. *Nat Struct Biol* **7**, 176-8 (2000).
158. Wang, J. et al. Crystal structure of a pol alpha family replication DNA polymerase from bacteriophage RB69. *Cell* **89**, 1087-99 (1997).
159. Kim, Y. et al. Crystal structure of Thermus aquaticus DNA polymerase. *Nature* **376**, 612-6 (1995).
160. Kim, S.W., Kim, D.U., Kim, J.K., Kang, L.W. & Cho, H.S. Crystal structure of Pfu, the high fidelity DNA polymerase from Pyrococcus furiosus. *Int J Biol Macromol* **42**, 356-61 (2008).
161. Wynne, S.A., Pinheiro, V.B., Holliger, P. & Leslie, A.G. Structures of an apo and a binary complex of an evolved archeal B family DNA polymerase capable of synthesising highly cy-dye labelled DNA. *PLoS One* **8**, e70892 (2013).
162. Rechkoblit, O. et al. Stepwise translocation of Dpo4 polymerase during error-free bypass of an oxoG lesion. *PLoS Biol* **4**, e11 (2006).
163. Brenlla, A., Markiewicz, R.P., Rueda, D. & Romano, L.J. Nucleotide selection by the Y-family DNA polymerase Dpo4 involves template translocation and misalignment. *Nucleic Acids Res* **42**, 2555-63 (2014).
164. Johnson, S.J., Taylor, J.S. & Beese, L.S. 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**, 3895-900 (2003).
165. Basu, R.S. & Murakami, K.S. Watching the bacteriophage N4 RNA polymerase transcription by time-dependent soak-trigger-freeze X-ray crystallography. *J Biol Chem* **288**, 3305-11 (2013).

166. Bakhtina, M., Roettger, M.P., Kumar, S. & Tsai, M.D. A unified kinetic mechanism applicable to multiple DNA polymerases. *Biochemistry* **46**, 5463-72 (2007).
167. Wang, M., Lee, H.R. & Konigsberg, W. Effect of A and B metal ion site occupancy on conformational changes in an RB69 DNA polymerase ternary complex. *Biochemistry* **48**, 2075-86 (2009).
168. Bakhtina, M. et al. Use of viscogens, dNTPalphaS, and rhodium(III) as probes in stopped-flow experiments to obtain new evidence for the mechanism of catalysis by DNA polymerase beta. *Biochemistry* **44**, 5177-87 (2005).
169. Johnson, A.A., Tsai, Y., Graves, S.W. & Johnson, K.A. Human mitochondrial DNA polymerase holoenzyme: reconstitution and characterization. *Biochemistry* **39**, 1702-8 (2000).
170. Graves, S.W., Johnson, A.A. & Johnson, K.A. Expression, purification, and initial kinetic characterization of the large subunit of the human mitochondrial DNA polymerase. *Biochemistry* **37**, 6050-8 (1998).
171. O'Donnell, M.E. & Kornberg, A. Complete replication of templates by Escherichia coli DNA polymerase III holoenzyme. *J Biol Chem* **260**, 12884-9 (1985).
172. Zahurancik, W.J., Klein, S.J. & Suo, Z. Significant contribution of the 3'-->5' exonuclease activity to the high fidelity of nucleotide incorporation catalyzed by human DNA polymerase. *Nucleic Acids Res* **42**, 13853-60 (2014).
173. Zhang, L., Brown, J.A., Newmister, S.A. & Suo, Z. Polymerization fidelity of a replicative DNA polymerase from the hyperthermophilic archaeon Sulfolobus solfataricus P2. *Biochemistry* **48**, 7492-501 (2009).
174. Joyce, C.M. How DNA travels between the separate polymerase and 3'-5'-exonuclease sites of DNA polymerase I (Klenow fragment). *J Biol Chem* **264**, 10858-66 (1989).
175. Reha-Krantz, L.J. DNA polymerase proofreading: Multiple roles maintain genome stability. *Biochim Biophys Acta* **1804**, 1049-63 (2010).
176. Fernandez-Leiro, R. et al. Self-correcting mismatches during high-fidelity DNA replication. *Nat Struct Mol Biol* **24**, 140-143 (2017).
177. Ganai, R.A., Bylund, G.O. & Johansson, E. Switching between polymerase and exonuclease sites in DNA polymerase epsilon. *Nucleic Acids Res* **43**, 932-42 (2015).
178. Johnson, A.A. & Johnson, K.A. Exonuclease proofreading by human mitochondrial DNA polymerase. *J Biol Chem* **276**, 38097-107 (2001).
179. Johnson, A.A. & Johnson, K.A. Fidelity of nucleotide incorporation by human mitochondrial DNA polymerase. *J Biol Chem* **276**, 38090-6 (2001).
180. Johnson, S.J. & Beese, L.S. Structures of mismatch replication errors observed in a DNA polymerase. *Cell* **116**, 803-16 (2004).
181. Xia, S. & Konigsberg, W.H. Mispairs with Watson-Crick base-pair geometry observed in ternary complexes of an RB69 DNA polymerase variant. *Protein Sci* **23**, 508-13 (2014).

182. Trincao, J. et al. Dpo4 is hindered in extending a G.T mismatch by a reverse wobble. *Nat Struct Mol Biol* **11**, 457-62 (2004).
183. Picher, A.J. et al. Promiscuous mismatch extension by human DNA polymerase lambda. *Nucleic Acids Res* **34**, 3259-66 (2006).
184. Batra, V.K., Beard, W.A., Pedersen, L.C. & Wilson, S.H. Structures of DNA Polymerase Mispaired DNA Termini Transitioning to Pre-catalytic Complexes Support an Induced-Fit Fidelity Mechanism. *Structure* **24**, 1863-1875 (2016).
185. McCulloch, S.D. & Kunkel, T.A. The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Res* **18**, 148-61 (2008).
186. Osheroff, W.P., Jung, H.K., Beard, W.A., Wilson, S.H. & Kunkel, T.A. The fidelity of DNA polymerase beta during distributive and processive DNA synthesis. *J Biol Chem* **274**, 3642-50 (1999).
187. Bebenek, K., Garcia-Diaz, M., Blanco, L. & Kunkel, T.A. The frameshift infidelity of human DNA polymerase lambda. Implications for function. *J Biol Chem* **278**, 34685-90 (2003).
188. Johnson, R.E., Washington, M.T., Prakash, S. & Prakash, L. Fidelity of human DNA polymerase eta. *J Biol Chem* **275**, 7447-50 (2000).
189. Watson, J.D. & Crick, F.H. Genetical implications of the structure of deoxyribonucleic acid. *Nature* **171**, 964-7 (1953).
190. Loeb, L.A. & Kunkel, T.A. Fidelity of DNA synthesis. *Annu Rev Biochem* **51**, 429-57 (1982).
191. Kool, E.T. & Sintim, H.O. The difluorotoluene debate--a decade later. *Chem Commun (Camb)* **0**, 3665-75 (2006).
192. Kool, E.T. Hydrogen bonding, base stacking, and steric effects in dna replication. *Annu Rev Biophys Biomol Struct* **30**, 1-22 (2001).
193. Dzantiev, L., Alekseyev, Y.O., Morales, J.C., Kool, E.T. & Romano, L.J. Significance of nucleobase shape complementarity and hydrogen bonding in the formation and stability of the closed polymerase-DNA complex. *Biochemistry* **40**, 3215-21 (2001).
194. Lee, H.R., Helquist, S.A., Kool, E.T. & Johnson, K.A. Importance of hydrogen bonding for efficiency and specificity of the human mitochondrial DNA polymerase. *J Biol Chem* **283**, 14402-10 (2008).
195. Xia, S., Eom, S.H., Konigsberg, W.H. & Wang, J. Structural basis for differential insertion kinetics of dNMPs opposite a difluorotoluene nucleotide residue. *Biochemistry* **51**, 1476-85 (2012).
196. Xia, S. & Konigsberg, W.H. RB69 DNA polymerase structure, kinetics, and fidelity. *Biochemistry* **53**, 2752-67 (2014).
197. Kunkel, T.A. & Bebenek, K. DNA replication fidelity. *Annu Rev Biochem* **69**, 497-529 (2000).
198. Nat. Commun.CRC Critical Reviews in BiochemistryWu, W.-J., Yang, W. & Tsai, M.-D. How DNA polymerases catalyse replication and repair with contrasting fidelity. *Nat. Rev. Chem.* **1**, 0068 (2017).

199. Fiala, K.A., Duym, W.W., Zhang, J. & Suo, Z. Up-regulation of the fidelity of human DNA polymerase lambda by its non-enzymatic proline-rich domain. *J Biol Chem* **281**, 19038-44 (2006).
200. Qin, Y. et al. Direct probing of solvent accessibility and mobility at the binding interface of polymerase (Dpo4)-DNA complex. *J Phys Chem A* **117**, 13926-34 (2013).
201. Wang, W., Hellinga, H.W. & Beese, L.S. Structural evidence for the rare tautomer hypothesis of spontaneous mutagenesis. *Proc Natl Acad Sci U S A* **108**, 17644-8 (2011).
202. Bebenek, K., Pedersen, L.C. & Kunkel, T.A. Replication infidelity via a mismatch with Watson-Crick geometry. *Proc Natl Acad Sci U S A* **108**, 1862-7 (2011).
203. Szymanski, E.S., Kimsey, I.J. & Al-Hashimi, H.M. Direct NMR Evidence that Transient Tautomeric and Anionic States in dG.dT Form Watson-Crick-like Base Pairs. *J Am Chem Soc* **139**, 4326-4329 (2017).
204. Kimsey, I.J. et al. Dynamic basis for dG*dT misincorporation via tautomerization and ionization. *Nature* **554**, 195-201 (2018).
205. Batra, V.K., Beard, W.A., Shock, D.D., Pedersen, L.C. & Wilson, S.H. Structures of DNA polymerase beta with active-site mismatches suggest a transient abasic site intermediate during misincorporation. *Mol Cell* **30**, 315-24 (2008).
206. Minnick, D.T., Astatke, M., Joyce, C.M. & Kunkel, T.A. 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**, 24954-61 (1996).
207. Kiefer, J.R., Mao, C., Braman, J.C. & Beese, L.S. Visualizing DNA replication in a catalytically active Bacillus DNA polymerase crystal. *Nature* **391**, 304-7 (1998).
208. Freudenthal, B.D., Beard, W.A. & Wilson, S.H. DNA polymerase minor groove interactions modulate mutagenic bypass of a templating 8-oxoguanine lesion. *Nucleic Acids Res* **41**, 1848-58 (2013).
209. Spence, R.A., Kati, W.M., Anderson, K.S. & Johnson, K.A. Mechanism of inhibition of HIV-1 reverse transcriptase by nonnucleoside inhibitors. *Science* **267**, 988-93 (1995).
210. Doublie, S., Sawaya, M.R. & Ellenberger, T. An open and closed case for all polymerases. *Structure* **7**, R31-5 (1999).
211. Herschlag, D., Piccirilli, J.A. & Cech, T.R. Ribozyme-catalyzed and nonenzymatic reactions of phosphate diesters: rate effects upon substitution of sulfur for a nonbridging phosphoryl oxygen atom. *Biochemistry* **30**, 4844-54 (1991).
212. Arndt, J.W. et al. Insight into the catalytic mechanism of DNA polymerase beta: structures of intermediate complexes. *Biochemistry* **40**, 5368-75 (2001).
213. Luo, G., Wang, M., Konigsberg, W.H. & Xie, X.S. Single-molecule and ensemble fluorescence assays for a functionally important conformational change in T7 DNA polymerase. *Proc Natl Acad Sci U S A* **104**, 12610-5 (2007).

214. Lee, H.R., Wang, M. & Konigsberg, W. The reopening rate of the fingers domain is a determinant of base selectivity for RB69 DNA polymerase. *Biochemistry* **48**, 2087-98 (2009).
215. Kirmizialtin, S., Nguyen, V., Johnson, K.A. & Elber, R. How conformational dynamics of DNA polymerase select correct substrates: experiments and simulations. *Structure* **20**, 618-27 (2012).
216. Kellinger, M.W. & Johnson, K.A. Nucleotide-dependent conformational change governs specificity and analog discrimination by HIV reverse transcriptase. *Proc Natl Acad Sci U S A* **107**, 7734-9 (2010).
217. Kellinger, M.W. & Johnson, K.A. Role of induced fit in limiting discrimination against AZT by HIV reverse transcriptase. *Biochemistry* **50**, 5008-15 (2011).
218. Yang, W. An overview of Y-Family DNA polymerases and a case study of human DNA polymerase eta. *Biochemistry* **53**, 2793-803 (2014).
219. Beckman, J.W., Wang, Q. & Guengerich, F.P. Kinetic analysis of correct nucleotide insertion by a Y-family DNA polymerase reveals conformational changes both prior to and following phosphodiester bond formation as detected by tryptophan fluorescence. *J Biol Chem* **283**, 36711-23 (2008).
220. Fiala, K.A., Sherrer, S.M., Brown, J.A. & Suo, Z. Mechanistic consequences of temperature on DNA polymerization catalyzed by a Y-family DNA polymerase. *Nucleic Acids Res* **36**, 1990-2001 (2008).
221. Eoff, R.L., Sanchez-Ponce, R. & Guengerich, F.P. Conformational changes during nucleotide selection by Sulfolobus solfataricus DNA polymerase Dpo4. *J Biol Chem* **284**, 21090-9 (2009).
222. Vyas, R. et al. Structural basis for the D-stereoselectivity of human DNA polymerase beta. *Nucleic Acids Res* **45**, 6228-6237 (2017).
223. Vyas, R., Zahurancik, W.J. & Suo, Z. Structural basis for the binding and incorporation of nucleotide analogs with L-stereochemistry by human DNA polymerase lambda. *Proc Natl Acad Sci U S A* **111**, E3033-42 (2014).
224. Shock, D.D., Freudenthal, B.D., Beard, W.A. & Wilson, S.H. Modulating the DNA polymerase beta reaction equilibrium to dissect the reverse reaction. *Nat Chem Biol* **13**, 1074-1080 (2017).
225. Raper, A.T., Reed, A.J., Gadkari, V.V. & Suo, Z. Advances in Structural and Single-Molecule Methods for Investigating DNA Lesion Bypass and Repair Polymerases. *Chem Res Toxicol* **30**, 260-269 (2017).
226. Brown, J.A., Duym, W.W., Fowler, J.D. & Suo, Z. Single-turnover kinetic analysis of the mutagenic potential of 8-oxo-7,8-dihydro-2'-deoxyguanosine during gap-filling synthesis catalyzed by human DNA polymerases lambda and beta. *J Mol Biol* **367**, 1258-69 (2007).
227. Bedford, E., Tabor, S. & Richardson, C.C. The thioredoxin binding domain of bacteriophage T7 DNA polymerase confers processivity on Escherichia coli DNA polymerase I. *Proc Natl Acad Sci U S A* **94**, 479-84 (1997).
228. Beard, W.A. & Wilson, S.H. Structural insights into the origins of DNA polymerase fidelity. *Structure* **11**, 489-96 (2003).
229. Ohmori, H. et al. The Y-family of DNA polymerases. *Mol Cell* **8**, 7-8 (2001).

230. Lehmann, A.R. et al. Translesion synthesis: Y-family polymerases and the polymerase switch. *DNA Repair (Amst)* **6**, 891-9 (2007).
231. Matsuda, T., Bebenek, K., Masutani, C., Hanaoka, F. & Kunkel, T.A. Low fidelity DNA synthesis by human DNA polymerase-eta. *Nature* **404**, 1011-3 (2000).
232. Pecoraro, V.L., Hermes, J.D. & Cleland, W.W. Stability constants of Mg2+ and Cd2+ complexes of adenine nucleotides and thionucleotides and rate constants for formation and dissociation of MgATP and MgADP. *Biochemistry* **23**, 5262-71 (1984).
233. Roettger, M.P., Fiala, K.A., Sompalli, S., Dong, Y. & Suo, Z. Pre-steady-state kinetic studies of the fidelity of human DNA polymerase mu. *Biochemistry* **43**, 13827-38 (2004).
234. Oertell, K. et al. Kinetic selection vs. free energy of DNA base pairing in control of polymerase fidelity. *Proc Natl Acad Sci U S A* **113**, E2277-85 (2016).
235. Hanes, J.W. & Johnson, K.A. A novel mechanism of selectivity against AZT by the human mitochondrial DNA polymerase. *Nucleic Acids Res* **35**, 6973-83 (2007).
236. Nami, F., Gast, P. & Groenen, E.J. Rapid Freeze-Quench EPR Spectroscopy: Improved Collection of Frozen Particles. *Appl Magn Reson* **47**, 643-653 (2016).
237. Hubscher, U., Maga, G. & Spadari, S. Eukaryotic DNA polymerases. *Annu Rev Biochem* **71**, 133-63 (2002).
238. Savino, C. et al. Insights into DNA replication: the crystal structure of DNA polymerase B1 from the archaeon Sulfolobus solfataricus. *Structure* **12**, 2001-8 (2004).
239. Steitz, T.A., Smerdon, S.J., Jager, J. & Joyce, C.M. A unified polymerase mechanism for nonhomologous DNA and RNA polymerases. *Science* **266**, 2022-5 (1994).
240. Vyas, R. et al. Mechanistic Basis for the Bypass of a Bulky DNA Adduct Catalyzed by a Y-Family DNA Polymerase. *J Am Chem Soc* **137**, 12131-42 (2015).
241. Roy, R., Hohng, S. & Ha, T. A practical guide to single-molecule FRET. *Nat Methods* **5**, 507-16 (2008).
242. Boudsocq, F., Iwai, S., Hanaoka, F. & Woodgate, R. Sulfolobus solfataricus P2 DNA polymerase IV (Dpo4): an archaeal DinB-like DNA polymerase with lesion-bypass properties akin to eukaryotic poleta. *Nucleic Acids Res* **29**, 4607-16 (2001).
243. Oda, Y. et al. NMR studies of a DNA containing 8-hydroxydeoxyguanosine. *Nucleic Acids Res* **19**, 1407-12 (1991).
244. McAuley-Hecht, K.E. et al. Crystal structure of a DNA duplex containing 8-hydroxydeoxyguanine-adenine base pairs. *Biochemistry* **33**, 10266-70 (1994).
245. Lipscomb, L.A. et al. X-ray structure of a DNA decamer containing 7,8-dihydro-8-oxoguanine. *Proc Natl Acad Sci U S A* **92**, 719-23 (1995).

246. Beard, W.A., Batra, V.K. & Wilson, S.H. DNA polymerase structure-based insight on the mutagenic properties of 8-oxoguanine. *Mutat Res* **703**, 18-23 (2010).
247. Krahn, J.M., Beard, W.A., Miller, H., Grollman, A.P. & Wilson, S.H. Structure of DNA polymerase beta with the mutagenic DNA lesion 8-oxodeoxyguanine reveals structural insights into its coding potential. *Structure* **11**, 121-7 (2003).
248. Brieba, L.G. et al. Structural basis for the dual coding potential of 8-oxoguanosine by a high-fidelity DNA polymerase. *Embo J* **23**, 3452-61 (2004).
249. Brieba, L.G., Kokoska, R.J., Bebenek, K., Kunkel, T.A. & Ellenberger, T. A lysine residue in the fingers subdomain of T7 DNA polymerase modulates the miscoding potential of 8-oxo-7,8-dihydroguanosine. *Structure* **13**, 1653-9 (2005).
250. Eoff, R.L., Irimia, A., Angel, K.C., Egli, M. & Guengerich, F.P. Hydrogen bonding of 7,8-dihydro-8-oxodeoxyguanosine with a charged residue in the little finger domain determines miscoding events in Sulfolobus solfataricus DNA polymerase Dpo4. *J Biol Chem* **282**, 19831-43 (2007).
251. Rechkoblit, O. et al. Impact of conformational heterogeneity of OxoG lesions and their pairing partners on bypass fidelity by Y family polymerases. *Structure* **17**, 725-36 (2009).
252. Espinoza-Herrera, S.J., Gaur, V., Suo, Z. & Carey, P.R. Following DNA Chain Extension and Protein Conformational Changes in Crystals of a Y-Family DNA Polymerase via Raman Crystallography. *Biochemistry* **52**, 4881−4890 (2013).
253. Tokarsky, E.J. et al. Pre-steady-state kinetic investigation of bypass of a bulky guanine lesion by human Y-family DNA polymerases. *DNA Repair (Amst)* **46**, 20-28 (2016).
254. Kalam, M.A. & Basu, A.K. Mutagenesis of 8-oxoguanine adjacent to an abasic site in simian kidney cells: tandem mutations and enhancement of G-->T transversions. *Chem Res Toxicol* **18**, 1187-92 (2005).
255. Ohno, M. et al. 8-oxoguanine causes spontaneous de novo germline mutations in mice. *Sci Rep* **4**, 4689 (2014).
256. Fraga, C.G., Shigenaga, M.K., Park, J.W., Degan, P. & Ames, B.N. Oxidative damage to DNA during aging: 8-hydroxy-2'-deoxyguanosine in rat organ DNA and urine. *Proc Natl Acad Sci U S A* **87**, 4533-7 (1990).
257. Moriya, M. Single-stranded shuttle phagemid for mutagenesis studies in mammalian cells: 8-oxoguanine in DNA induces targeted G.C-->T.A transversions in simian kidney cells. *Proc Natl Acad Sci U S A* **90**, 1122-6 (1993).
258. Steitz, T.A. DNA- and RNA-dependent DNA polymerases. *Curr Opin Struct Biol* **3**, 31-38 (1993).
259. Freudenthal, B.D., Beard, W.A. & Wilson, S.H. Watching a DNA polymerase in action. *Cell Cycle* **13**, 691-2 (2014).
260. Ha, T. Single-molecule fluorescence resonance energy transfer. *Methods* **25**, 78-86 (2001).
261. Berezhna, S.Y., Gill, J.P., Lamichhane, R. & Millar, D.P. Single-molecule Forster resonance energy transfer reveals an innate fidelity checkpoint in DNA polymerase I. *J Am Chem Soc* **134**, 11261-8 (2012).

262. Christian, T.D., Romano, L.J. & Rueda, D. Single-molecule measurements of synthesis by DNA polymerase with base-pair resolution. *Proc Natl Acad Sci U S A* **106**, 21109-14 (2009).
263. Markiewicz, R.P., Vrtis, K.B., Rueda, D. & Romano, L.J. Single-molecule microscopy reveals new insights into nucleotide selection by DNA polymerase I. *Nucleic Acids Res* **40**, 7975-84 (2012).
264. Santoso, Y. et al. Conformational transitions in DNA polymerase I revealed by single-molecule FRET. *Proc Natl Acad Sci U S A* **107**, 715-20 (2010).
265. Vrtis, K.B., Markiewicz, R.P., Romano, L.J. & Rueda, D. Carcinogenic adducts induce distinct DNA polymerase binding orientations. *Nucleic Acids Res* **41**, 7843-53 (2013).
266. Brenlla, A., Rueda, D. & Romano, L.J. Mechanism of aromatic amine carcinogen bypass by the Y-family polymerase, Dpo4. *Nucleic Acids Res* **43**, 9918-27 (2015).
267. Raper, A.T. & Suo, Z. Investigation of Intradomain Motions of a Y-Family DNA Polymerase during Substrate Binding and Catalysis. *Biochemistry* (2016).
268. Lieberman, K.R., Dahl, J.M., Mai, A.H., Akeson, M. & Wang, H. Dynamics of the translocation step measured in individual DNA polymerase complexes. *J Am Chem Soc* **134**, 18816-23 (2012).
269. Dahl, J.M. et al. Direct observation of translocation in individual DNA polymerase complexes. *J Biol Chem* **287**, 13407-21 (2012).
270. Lieberman, K.R. et al. Kinetic mechanism of translocation and dNTP binding in individual DNA polymerase complexes. *J Am Chem Soc* **135**, 9149-55 (2013).
271. Lovett, S.T. Polymerase switching in DNA replication. *Mol Cell* **27**, 523-6 (2007).
272. Friedberg, E.C., Lehmann, A.R. & Fuchs, R.P. Trading places: how do DNA polymerases switch during translesion DNA synthesis? *Mol Cell* **18**, 499-505 (2005).
273. Jansen, J.G., Fousteri, M.I. & de Wind, N. Send in the clamps: control of DNA translesion synthesis in eukaryotes. *Mol Cell* **28**, 522-9 (2007).
274. Pages, V. & Fuchs, R.P. How DNA lesions are turned into mutations within cells? *Oncogene* **21**, 8957-66 (2002).
275. Becherel, O.J., Fuchs, R.P. & Wagner, J. Pivotal role of the beta-clamp in translesion DNA synthesis and mutagenesis in E. coli cells. *DNA Repair (Amst)* **1**, 703-8 (2002).
276. Indiani, C., McInerney, P., Georgescu, R., Goodman, M.F. & O'Donnell, M. A sliding-clamp toolbelt binds high- and low-fidelity DNA polymerases simultaneously. *Mol Cell* **19**, 805-15 (2005).
277. Carroll, D. Genome Editing: Past, Present, and Future. *Yale J Biol Med* **90**, 653-659 (2017).
278. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A* **109**, E2579-86 (2012).
279. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-21 (2012).

280. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-23 (2013).
281. Klompe, S.E. & Sternberg, S.H. Harnessing "A Billion Years of Experimentation": The Ongoing Exploration and Exploitation of CRISPR–Cas Immune Systems. *The CRISPR Journal* **1**, 141-158 (2018).
282. Jiang, F. & Doudna, J.A. CRISPR-Cas9 Structures and Mechanisms. *Annu Rev Biophys* **46**, 505-529 (2017).
283. Mir, A., Edraki, A., Lee, J. & Sontheimer, E.J. Type II-C CRISPR-Cas9 Biology, Mechanism, and Application. *ACS Chem Biol* **13**, 357-365 (2018).
284. Wang, H., La Russa, M. & Qi, L.S. CRISPR/Cas9 in Genome Editing and Beyond. *Annu Rev Biochem* **85**, 227-64 (2016).
285. Bothmer, A. et al. Characterization of the interplay between DNA repair and CRISPR/Cas9-induced DNA lesions at an endogenous locus. *Nat Commun* **8**, 13905 (2017).
286. Doudna, J.A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096 (2014).
287. Dai, W.J. et al. CRISPR-Cas9 for in vivo Gene Therapy: Promise and Hurdles. *Mol Ther Nucleic Acids* **5**, e349 (2016).
288. Tang, W. & Liu, D.R. Rewritable multi-event analog recording in bacterial and mammalian cells. *Science* **360**(2018).
289. Gantz, V.M. et al. Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito Anopheles stephensi. *Proc Natl Acad Sci U S A* **112**, E6736-43 (2015).
290. Soppe, J.A. & Lebbink, R.J. Antiviral Goes Viral: Harnessing CRISPR/Cas9 to Combat Viruses in Humans. *Trends Microbiol* **25**, 833-850 (2017).
291. Chen, B. & Huang, B. Imaging genomic elements in living cells using CRISPR/Cas9. *Methods Enzymol* **546**, 337-54 (2014).
292. La Russa, M.F. & Qi, L.S. The New State of the Art: Cas9 for Gene Activation and Repression. *Mol Cell Biol* **35**, 3800-9 (2015).
293. Gaudelli, N.M. et al. Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. *Nature* **551**, 464-471 (2017).
294. Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A. & Liu, D.R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420-4 (2016).
295. Nishida, K. et al. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**(2016).
296. Qi, L.S. et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173-83 (2013).
297. Zhang, X.H., Tee, L.Y., Wang, X.G., Huang, Q.S. & Yang, S.H. Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. *Mol Ther Nucleic Acids* **4**, e264 (2015).
298. Fu, Y. et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* **31**, 822-6 (2013).

299. Cho, S.W. et al. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res* **24**, 132-41 (2014).
300. Chen, J.S. et al. Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* **550**, 407-410 (2017).
301. Slaymaker, I.M. et al. Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84-8 (2016).
302. Hu, J.H. et al. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57-63 (2018).
303. Kleinstiver, B.P. et al. High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490-5 (2016).
304. Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C. & Doudna, J.A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62-7 (2014).
305. Singh, D., Sternberg, S.H., Fei, J., Doudna, J.A. & Ha, T. Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9. *Nat Commun* **7**, 12778 (2016).
306. Singh, D. et al. Mechanisms of improved specificity of engineered Cas9s revealed by single-molecule FRET analysis. *Nat Struct Mol Biol* **25**, 347-354 (2018).
307. Wu, X. et al. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat Biotechnol* **32**, 670-6 (2014).
308. Hsu, P.D., Lander, E.S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262-78 (2014).
309. Lin, L. et al. Genome-wide determination of on-target and off-target characteristics for RNA-guided DNA methylation by dCas9 methyltransferases. *Gigascience* **7**, 1-19 (2018).
310. Cui, L. et al. A CRISPRi screen in E. coli reveals sequence-specific toxicity of dCas9. *Nat Commun* **9**, 1912 (2018).
311. Kleinstiver, B.P. et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523**, 481-5 (2015).
312. Anders, C., Bargsten, K. & Jinek, M. Structural Plasticity of PAM Recognition by Engineered Variants of the RNA-Guided Endonuclease Cas9. *Mol Cell* **61**, 895-902 (2016).
313. Hirano, S., Nishimasu, H., Ishitani, R. & Nureki, O. Structural Basis for the Altered PAM Specificities of Engineered CRISPR-Cas9. *Mol Cell* **61**, 886-94 (2016).
314. Chu, V.T. et al. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat Biotechnol* **33**, 543-8 (2015).
315. Brandsma, I. & Gent, D.C. Pathway choice in DNA double strand break repair: observations of a balancing act. *Genome Integr* **3**, 9 (2012).
316. Chapman, J.R., Taylor, M.R. & Boulton, S.J. Playing the end game: DNA double-strand break repair pathway choice. *Mol Cell* **47**, 497-510 (2012).
317. Nishimasu, H. et al. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935-49 (2014).

318. Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513**, 569-73 (2014).
319. Stephenson, A.A., Raper, A.T. & Suo, Z. Bidirectional Degradation of DNA Cleavage Products Catalyzed by CRISPR/Cas9. *Journal of the American Chemical Society* (2018).
320. Raper, A.T., Stephenson, A.A. & Suo, Z. Functional Insights Revealed by the Kinetic Mechanism of CRISPR/Cas9. *J Am Chem Soc* **140**, 2971-2984 (2018).
321. Wright, A.V. et al. Rational design of a split-Cas9 enzyme complex. *Proc Natl Acad Sci U S A* **112**, 2984-9 (2015).
322. Jiang, F., Zhou, K., Ma, L., Gressel, S. & Doudna, J.A. STRUCTURAL BIOLOGY. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science* **348**, 1477-81 (2015).
323. Jinek, M. et al. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343**, 1247997 (2014).
324. Sternberg, S.H., LaFrance, B., Kaplan, M. & Doudna, J.A. Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature* **527**, 110-3 (2015).
325. Shibata, M. et al. Real-space and real-time dynamics of CRISPR-Cas9 visualized by high-speed atomic force microscopy. *Nat Commun* **8**, 1430 (2017).
326. Yang, M. et al. The Conformational Dynamics of Cas9 Governing DNA Cleavage Are Revealed by Single-Molecule FRET. *Cell Rep* **22**, 372-382 (2018).
327. Jiang, F. et al. Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* **351**, 867-71 (2016).
328. Gong, S., Yu, H.H., Johnson, K.A. & Taylor, D.W. DNA Unwinding Is the Primary Determinant of CRISPR-Cas9 Activity. *Cell Rep* **22**, 359-371 (2018).
329. Dagdas, Y.S., Chen, J.S., Sternberg, S.H., Doudna, J.A. & Yildiz, A. A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. *Sci Adv* **3**, eaao0027 (2017).
330. Palermo, G., Miao, Y., Walker, R.C., Jinek, M. & McCammon, J.A. Striking Plasticity of CRISPR-Cas9 and Key Role of Non-target DNA, as Revealed by Molecular Simulations. *ACS Cent Sci* **2**, 756-763 (2016).
331. Palermo, G., Miao, Y., Walker, R.C., Jinek, M. & McCammon, J.A. CRISPR-Cas9 conformational activation as elucidated from enhanced molecular simulations. *Proc Natl Acad Sci U S A* **114**, 7260-7265 (2017).
332. Fu, Y., Sander, J.D., Reyon, D., Cascio, V.M. & Joung, J.K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol* **32**, 279-284 (2014).
333. Karvelis, T. et al. Rapid characterization of CRISPR-Cas9 protospacer adjacent motif sequence elements. *Genome Biol* **16**, 253 (2015).
334. Mali, P., Esvelt, K.M. & Church, G.M. Cas9 as a versatile tool for engineering biology. *Nat Methods* **10**, 957-63 (2013).
335. Rose, J.C. et al. Rapidly inducible Cas9 and DSB-ddPCR to probe editing kinetics. *Nat Methods* **14**, 891-896 (2017).

336. Ousterout, D.G. et al. Multiplex CRISPR/Cas9-based genome editing for correction of dystrophin mutations that cause Duchenne muscular dystrophy. *Nat Commun* **6**, 6244 (2015).
337. Chen, X., Liu, J., Janssen, J.M. & Goncalves, M. The Chromatin Structure Differentially Impacts High-Specificity CRISPR-Cas9 Nuclease Strategies. *Mol Ther Nucleic Acids* **8**, 558-563 (2017).
338. Uusi-Makela, M.I.E. et al. Chromatin accessibility is associated with CRISPR-Cas9 efficiency in the zebrafish (Danio rerio). *PLoS One* **13**, e0196238 (2018).
339. Chen, Y. et al. Using local chromatin structure to improve CRISPR/Cas9 efficiency in zebrafish. *PLoS One* **12**, e0182528 (2017).
340. Marnett, L.J. Oxyradicals and DNA damage. *Carcinogenesis* **21**, 361-70 (2000).
341. Al-Tassan, N. et al. Inherited variants of MYH associated with somatic G:C--->T:A mutations in colorectal tumors. *Nat Genet* **30**, 227-32 (2002).
342. Petitjean, A. et al. Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database. *Hum Mutat* **28**, 622-9 (2007).
343. Furge, L.L. & Guengerich, F.P. Analysis of nucleotide insertion and extension at 8-oxo-7,8-dihydroguanine by replicative T7 polymerase exo- and human immunodeficiency virus-1 reverse transcriptase using steady-state and pre-steady-state kinetics. *Biochemistry* **36**, 6475-87 (1997).
344. Einolf, H.J., Schnetz-Boutaud, N. & Guengerich, F.P. Steady-state and pre-steady-state kinetic analysis of 8-oxo-7,8-dihydroguanosine triphosphate incorporation and extension by replicative and repair DNA polymerases. *Biochemistry* **37**, 13300-12 (1998).
345. Einolf, H.J. & Guengerich, F.P. Fidelity of nucleotide insertion at 8-oxo-7,8-dihydroguanine by mammalian DNA polymerase delta. Steady-state and pre-steady-state kinetic analysis. *J Biol Chem* **276**, 3764-71 (2001).
346. McCulloch, S.D., Kokoska, R.J., Garg, P., Burgers, P.M. & Kunkel, T.A. The efficiency and fidelity of 8-oxo-guanine bypass by DNA polymerases delta and eta. *Nucleic Acids Res* **37**, 2830-40 (2009).
347. Maga, G. et al. 8-oxo-guanine bypass by human DNA polymerases in the presence of auxiliary proteins. *Nature* **447**, 606-8 (2007).
348. Sabouri, N., Viberg, J., Goyal, D.K., Johansson, E. & Chabes, A. Evidence for lesion bypass by yeast replicative DNA polymerases during DNA damage. *Nucleic Acids Res* **36**, 5660-7 (2008).
349. Hogg, M. et al. Kinetics of mismatch formation opposite lesions by the replicative DNA polymerase from bacteriophage RB69. *Biochemistry* **49**, 2317-25 (2010).
350. de Vega, M. & Salas, M. A highly conserved Tyrosine residue of family B DNA polymerases contributes to dictate translesion synthesis past 8-oxo-7,8-dihydro-2'-deoxyguanosine. *Nucleic Acids Res* **35**, 5096-107 (2007).
351. Zang, H. et al. Efficient and high fidelity incorporation of dCTP opposite 7,8-dihydro-8-oxodeoxyguanosine by Sulfolobus solfataricus DNA polymerase Dpo4. *J Biol Chem* **281**, 2358-72 (2006).

352. Selvin, P.R., Ha,T. *Single-molecule Techniques: a Laboratory Manual*, 3-36 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY(2008).
353. Irimia, A. et al. Calcium is a cofactor of polymerization but inhibits pyrophosphorolysis by the Sulfolobus solfataricus DNA polymerase Dpo4. *Biochemistry* **45**, 5949-56 (2006).
354. Zang, H. et al. DNA adduct bypass polymerization by Sulfolobus solfataricus DNA polymerase Dpo4: analysis and crystal structures of multiple base pair substitution and frameshift products with the adduct 1,N2-ethenoguanine. *J Biol Chem* **280**, 29750-64 (2005).
355. Lamichhane, R., Berezhna, S.Y., Gill, J.P., Van der Schans, E. & Millar, D.P. Dynamics of site switching in DNA polymerase. *J Am Chem Soc* **135**, 4735-42 (2013).
356. Blanco, M. & Walter, N.G. Analysis of complex single-molecule FRET time trajectories. *Methods Enzymol* **472**, 153-78 (2010).
357. McKinney, S.A., Joo, C. & Ha, T. Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophys J* **91**, 1941-51 (2006).
358. Fiala, K.A. et al. Mechanism of template-independent nucleotide incorporation catalyzed by a template-dependent DNA polymerase. *J Mol Biol* **365**, 590-602 (2007).
359. Xing, G., Kirouac, K., Shin, Y.J., Bell, S.D. & Ling, H. Structural insight into recruitment of translesion DNA polymerase Dpo4 to sliding clamp PCNA. *Mol Microbiol* **71**, 678-91 (2009).
360. Abbondanzieri, E.A. et al. Dynamic binding orientations direct activity of HIV reverse transcriptase. *Nature* **453**, 184-9 (2008).
361. Liu, S., Abbondanzieri, E.A., Rausch, J.W., Le Grice, S.F. & Zhuang, X. Slide into action: dynamic shuttling of HIV reverse transcriptase on nucleic acid substrates. *Science* **322**, 1092-7 (2008).
362. Zhou, B.L., Pata, J.D. & Steitz, T.A. Crystal structure of a DinB lesion bypass DNA polymerase catalytic fragment reveals a classic polymerase catalytic domain. *Mol Cell* **8**, 427-37 (2001).
363. Bruskov, V.I., Malakhova, L.V., Masalimov, Z.K. & Chernikov, A.V. Heat-induced formation of reactive oxygen species and 8-oxoguanine, a biomarker of damage to DNA. *Nucleic Acids Res* **30**, 1354-63 (2002).
364. Ling, H., Boudsocq, F., Woodgate, R. & Yang, W. Snapshots of replication through an abasic lesion; structural basis for base substitutions and frameshifts. *Mol Cell* **13**, 751-62 (2004).
365. Boudsocq, F. et al. Investigating the role of the little finger domain of Y-family DNA polymerases in low fidelity synthesis and translesion replication. *J Biol Chem* **279**, 32932-40 (2004).
366. Clegg, R.M. Fluorescence resonance energy transfer and nucleic acids. *Methods Enzymol* **211**, 353-88 (1992).
367. Lakowicz, J.R. Principles of Fluorescence Spectroscopy. 954 (2006).

368. Johnson, K.A. 1 Transient-State Kinetic Analysis of Enzyme Reaction Pathways. **20**, 1-61 (1992).
369. Alley, S.C. et al. Sliding clamp of the bacteriophage T4 polymerase has open and closed subunit interfaces in solution. *Biochemistry* **38**, 7696-709 (1999).
370. Trakselis, M.A., Alley, S.C., Abel-Santos, E. & Benkovic, S.J. Creating a dynamic picture of the sliding clamp during T4 DNA polymerase holoenzyme assembly by using fluorescence resonance energy transfer. *Proc Natl Acad Sci U S A* **98**, 8368-75 (2001).
371. Chakraborty, S. et al. Structure and dynamics of the alpha-lactalbumin molten globule: fluorescence studies using proteins containing a single tryptophan residue. *Biochemistry* **40**, 7228-38 (2001).
372. Zhou, G., Wong, M.T. & Zhou, G.Q. Diffusion-controlled reactions of enzymes. An approximate analytic solution of Chou's model. *Biophys Chem* **18**, 125-32 (1983).
373. Astatke, M., Grindley, N.D. & Joyce, C.M. How E. coli DNA polymerase I (Klenow fragment) distinguishes between deoxy- and dideoxynucleotides. *J Mol Biol* **278**, 147-65 (1998).
374. Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450**, 964-72 (2007).
375. Benkovic, S.J. & Hammes-Schiffer, S. A perspective on enzyme catalysis. *Science* **301**, 1196-202 (2003).
376. Koshland, D.E. Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc Natl Acad Sci U S A* **44**, 98-104 (1958).
377. Kumar, S., Ma, B., Tsai, C.J., Sinha, N. & Nussinov, R. Folding and binding cascades: dynamic landscapes and population shifts. *Protein Sci* **9**, 10-9 (2000).
378. Tsai, C.J., Ma, B., Sham, Y.Y., Kumar, S. & Nussinov, R. Structured disorder and conformational selection. *Proteins* **44**, 418-27 (2001).
379. Bosshard, H.R. Molecular recognition by induced fit: how fit is the concept? *News Physiol Sci* **16**, 171-3 (2001).
380. Okazaki, K. & Takada, S. Dynamic energy landscape view of coupled binding and protein conformational change: induced-fit versus population-shift mechanisms. *Proc Natl Acad Sci U S A* **105**, 11182-7 (2008).
381. DeLucia, A.M., Grindley, N.D. & Joyce, C.M. Conformational changes during normal and error-prone incorporation of nucleotides by a Y-family DNA polymerase detected by 2-aminopurine fluorescence. *Biochemistry* **46**, 10790-803 (2007).
382. Zhang, H., Cao, W., Zakharova, E., Konigsberg, W. & De La Cruz, E.M. Fluorescence of 2-aminopurine reveals rapid conformational changes in the RB69 DNA polymerase-primer/template complexes upon binding and incorporation of matched deoxynucleoside triphosphates. *Nucleic Acids Res* **35**, 6052-62 (2007).
383. Purohit, V., Grindley, N.D. & Joyce, C.M. Use of 2-aminopurine fluorescence to examine conformational changes during nucleotide incorporation by DNA polymerase I (Klenow fragment). *Biochemistry* **42**, 10200-11 (2003).

384. Federley, R.G. & Romano, L.J. DNA polymerase: structural homology, conformational dynamics, and the effects of carcinogenic DNA adducts. *J Nucleic Acids* **2010**(2010).
385. Post, C.B. & Ray, W.J., Jr. Reexamination of induced fit as a determinant of substrate specificity in enzymatic reactions. *Biochemistry* **34**, 15881-5 (1995).
386. Zhou, H.X. From induced fit to conformational selection: a continuum of binding mechanism controlled by the timescale of conformational transitions. *Biophys J* **98**, L15-7 (2010).
387. Hammes, G.G., Chang, Y.C. & Oas, T.G. Conformational selection or induced fit: a flux description of reaction mechanism. *Proc Natl Acad Sci U S A* **106**, 13737-41 (2009).
388. Csermely, P., Palotai, R. & Nussinov, R. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends Biochem Sci* **35**, 539-46 (2010).
389. Wlodarski, T. & Zagrovic, B. Conformational selection and induced fit mechanism underlie specificity in noncovalent interactions with ubiquitin. *Proc Natl Acad Sci U S A* **106**, 19346-51 (2009).
390. Grunberg, R., Leckner, J. & Nilges, M. Complementarity of structure ensembles in protein-protein binding. *Structure* **12**, 2125-36 (2004).
391. Shoemaker, B.A., Portman, J.J. & Wolynes, P.G. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc Natl Acad Sci U S A* **97**, 8868-73 (2000).
392. Fleck, O. & Schar, P. Translesion DNA synthesis: little fingers teach tolerance. *Curr Biol* **14**, R389-91 (2004).
393. Kleckner, I.R. & Foster, M.P. An introduction to NMR-based approaches for measuring protein dynamics. *Biochim Biophys Acta* **1814**, 942-68 (2011).
394. Ma, D., Fowler, J.D., Yuan, C. & Suo, Z. Backbone assignment of the catalytic core of a Y-family DNA polymerase. *Biomol NMR Assign* **4**, 207-9 (2010).
395. Ma, D., Fowler, J.D. & Suo, Z. Backbone assignment of the little finger domain of a Y-family DNA polymerase. *Biomol NMR Assign* **5**, 195-8 (2011).
396. Beese, L.S., Derbyshire, V. & Steitz, T.A. Structure of DNA polymerase I Klenow fragment bound to duplex DNA. *Science* **260**, 352-5 (1993).
397. Kunkel, T.A. DNA replication fidelity. *J Biol Chem* **279**, 16895-8 (2004).
398. Garcia-Diaz, M. & Bebenek, K. Multiple functions of DNA polymerases. *CRC Crit Rev Plant Sci* **26**, 105-122 (2007).
399. Dodson, M.L., Michaels, M.L. & Lloyd, R.S. Unified catalytic mechanism for DNA glycosylases. *J Biol Chem* **269**, 32709-12 (1994).
400. Yang, W. Structure and mechanism for DNA lesion recognition. *Cell Res* **18**, 184-97 (2008).
401. Friedman, J.I. & Stivers, J.T. Detection of damaged DNA bases by DNA glycosylase enzymes. *Biochemistry* **49**, 4957-67 (2010).
402. Lehman, I.R. DNA ligase: structure, mechanism, and function. *Science* **186**, 790-7 (1974).

403. Tomkinson, A.E., Vijayakumar, S., Pascal, J.M. & Ellenberger, T. DNA ligases: structure, reaction mechanism, and function. *Chem Rev* **106**, 687-99 (2006).
404. Nishino, T. & Morikawa, K. Structure and function of nucleases in DNA repair: shape, grip and blade of the DNA scissors. *Oncogene* **21**, 9022-32 (2002).
405. Yang, W. Nucleases: diversity of structure, function and mechanism. *Q Rev Biophys* **44**, 1-93 (2011).
406. Finger, L.D. et al. The wonders of flap endonucleases: structure, function, mechanism and regulation. *Subcell Biochem* **62**, 301-26 (2012).
407. Lewis, J.S., Jergic, S. & Dixon, N.E. The E. coli DNA Replication Fork. *Enzymes* **39**, 31-88 (2016).
408. Zhang, D. & O'Donnell, M. The Eukaryotic Replication Machine. *Enzymes* **39**, 191-229 (2016).
409. Noble, E., Spiering, M.M. & Benkovic, S.J. Coordinated DNA Replication by the Bacteriophage T4 Replisome. *Viruses* **7**, 3186-200 (2015).
410. Indiani, C. & O'Donnell, M. The replication clamp-loading machine at work in the three domains of life. *Nat Rev Mol Cell Biol* **7**, 751-61 (2006).
411. Dionne, I., Nookala, R.K., Jackson, S.P., Doherty, A.J. & Bell, S.D. A heterotrimeric PCNA in the hyperthermophilic archaeon Sulfolobus solfataricus. *Mol Cell* **11**, 275-82 (2003).
412. Williams, G.J. et al. Structure of the heterotrimeric PCNA from Sulfolobus solfataricus. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **62**, 944-8 (2006).
413. Hlinkova, V. et al. Structures of monomeric, dimeric and trimeric PCNA: PCNA-ring assembly and opening. *Acta Crystallogr D Biol Crystallogr* **64**, 941-9 (2008).
414. Pascal, J.M. et al. A flexible interface between DNA ligase and PCNA supports conformational switching and efficient ligation of DNA. *Mol Cell* **24**, 279-91 (2006).
415. Dore, A.S. et al. Structure of an archaeal PCNA1-PCNA2-FEN1 complex: elucidating PCNA subunit and client enzyme specificity. *Nucleic Acids Res* **34**, 4515-26 (2006).
416. Dionne, I., Brown, N.J., Woodgate, R. & Bell, S.D. On the mechanism of loading the PCNA sliding clamp by RFC. *Mol Microbiol* **68**, 216-22 (2008).
417. Kelch, B.A., Makino, D.L., O'Donnell, M. & Kuriyan, J. How a DNA polymerase clamp loader opens a sliding clamp. *Science* **334**, 1675-80 (2011).
418. Millar, D., Trakselis, M.A. & Benkovic, S.J. On the solution structure of the T4 sliding clamp (gp45). *Biochemistry* **43**, 12723-7 (2004).
419. Moarefi, I., Jeruzalmi, D., Turner, J., O'Donnell, M. & Kuriyan, J. Crystal structure of the DNA polymerase processivity factor of T4 bacteriophage. *J Mol Biol* **296**, 1215-23 (2000).
420. Fairhead, M. & Howarth, M. Site-specific biotinylation of purified proteins using BirA. *Methods Mol Biol* **1266**, 171-84 (2015).
421. Studier, F.W. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* **41**, 207-34 (2005).

422. O'Callaghan C, A. et al. BirA enzyme: production and application in the study of membrane receptor-ligand interactions by site-specific biotinylation. *Anal Biochem* **266**, 9-15 (1999).
423. Chandradoss, S.D. et al. Surface passivation for single-molecule protein studies. *J Vis Exp* (2014).
424. Rasnik, I., McKinney, S.A. & Ha, T. Nonblinking and long-lasting single-molecule fluorescence imaging. *Nat Methods* **3**, 891-3 (2006).
425. Zhou, M., Huang, C. & Wysocki, V.H. Surface-induced dissociation of ion mobility-separated noncovalent complexes in a quadrupole/time-of-flight mass spectrometer. *Anal Chem* **84**, 6016-23 (2012).
426. Bush, M.F. et al. Collision cross sections of proteins and their complexes: a calibration framework and database for gas-phase structural biology. *Anal Chem* **82**, 9557-65 (2010).
427. Noel, J.K., Whitford, P.C., Sanbonmatsu, K.Y. & Onuchic, J.N. SMOG@ctbp: simplified deployment of structure-based models in GROMACS. *Nucleic Acids Res* **38**, W657-61 (2010).
428. Clementi, C., Nymeyer, H. & Onuchic, J.N. Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *J Mol Biol* **298**, 937-53 (2000).
429. Noel, J.K., Whitford, P.C. & Onuchic, J.N. The shadow map: a general contact definition for capturing the dynamics of biomolecular folding and function. *J Phys Chem B* **116**, 8692-702 (2012).
430. Miyazawa, S. & Jernigan, R.L. Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *J Mol Biol* **256**, 623-44 (1996).
431. Cho, S.S., Levy, Y. & Wolynes, P.G. Quantitative criteria for native energetic heterogeneity influences in the prediction of protein folding kinetics. *Proc Natl Acad Sci U S A* **106**, 434-9 (2009).
432. Azia, A. & Levy, Y. Nonnative electrostatic interactions can modulate protein folding: molecular dynamics with a grain of salt. *J Mol Biol* **393**, 527-42 (2009).
433. Givaty, O. & Levy, Y. Protein sliding along DNA: dynamics and structural characterization. *J Mol Biol* **385**, 1087-97 (2009).
434. Chu, X. et al. Importance of electrostatic interactions in the association of intrinsically disordered histone chaperone Chz1 and histone H2A.Z-H2B. *PLoS Comput Biol* **8**, e1002608 (2012).
435. Wang, Y., Gan, L., Wang, E. & Wang, J. Exploring the Dynamic Functional Landscape of Adenylate Kinase Modulated by Substrates. *J Chem Theory Comput* **9**, 84-95 (2013).
436. Hess, B., Kutzner, C., van der Spoel, D. & Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J Chem Theory Comput* **4**, 435-47 (2008).

437. Hess, B., Bekker, H., Berendsen, H.J.C., Fraaije, J.G.E.M. LINCS: A Linear Constraint Solver for Molecular Simulations. *Journal of Computational Chemistry* **18**, 1463-1472 (1997).
438. Marklund, E.G., Degiacomi, M.T., Robinson, C.V., Baldwin, A.J. & Benesch, J.L. Collision cross sections for structural proteomics. *Structure* **23**, 791-9 (2015).
439. Heltzel, J.M., Maul, R.W., Scouten Ponticelli, S.K. & Sutton, M.D. A model for DNA polymerase switching involving a single cleft and the rim of the sliding clamp. *Proc Natl Acad Sci U S A* **106**, 12664-9 (2009).
440. Lee, S., Lee, J. & Hohng, S. Single-molecule three-color FRET with both negligible spectral overlap and long observation time. *PLoS One* **5**, e12270 (2010).
441. Dietrich, A., Buschmann, V., Muller, C. & Sauer, M. Fluorescence resonance energy transfer (FRET) and competing processes in donor-acceptor substituted DNA strands: a comparative study of ensemble and single-molecule data. *J Biotechnol* **82**, 211-31 (2002).
442. Sharon, M. & Robinson, C.V. The role of mass spectrometry in structure elucidation of dynamic protein complexes. *Annu Rev Biochem* **76**, 167-93 (2007).
443. Marcoux, J. & Robinson, C.V. Twenty years of gas phase structural biology. *Structure* **21**, 1541-50 (2013).
444. Benesch, J.L., Ruotolo, B.T., Simmons, D.A. & Robinson, C.V. Protein complexes in the gas phase: technology for structural genomics and proteomics. *Chem Rev* **107**, 3544-67 (2007).
445. Uetrecht, C. et al. Stability and shape of hepatitis B virus capsids in vacuo. *Angew Chem Int Ed Engl* **47**, 6247-51 (2008).
446. Dyachenko, A., Gruber, R., Shimon, L., Horovitz, A. & Sharon, M. Allosteric mechanisms can be distinguished using structural mass spectrometry. *Proc Natl Acad Sci U S A* **110**, 7235-9 (2013).
447. Uetrecht, C., Rose, R.J., van Duijn, E., Lorenzen, K. & Heck, A.J. Ion mobility mass spectrometry of proteins and protein assemblies. *Chem Soc Rev* **39**, 1633-55 (2010).
448. Konijnenberg, A., Butterer, A. & Sobott, F. Native ion mobility-mass spectrometry and related methods in structural biology. *Biochim Biophys Acta* **1834**, 1239-56 (2013).
449. Jenner, M. et al. Detection of a protein conformational equilibrium by electrospray ionisation-ion mobility-mass spectrometry. *Angew Chem Int Ed Engl* **50**, 8291-4 (2011).
450. Harvey, S.R. et al. Dissecting the dynamic conformations of the metamorphic protein lymphotactin. *J Phys Chem B* **118**, 12348-59 (2014).
451. Hall, Z., Politis, A., Bush, M.F., Smith, L.J. & Robinson, C.V. Charge-state dependent compaction and dissociation of protein complexes: insights from ion mobility and molecular dynamics. *J Am Chem Soc* **134**, 3429-38 (2012).
452. Zhou, M., Dagan, S. & Wysocki, V.H. Impact of charge state on gas-phase behaviors of noncovalent protein complexes in collision induced dissociation and surface induced dissociation. *Analyst* **138**, 1353-62 (2013).

453. Xiao, Q., Zhang, F., Nacev, B.A., Liu, J.O. & Pei, D. Protein N-terminal processing: substrate specificity of Escherichia coli and human methionine aminopeptidases. *Biochemistry* **49**, 5588-99 (2010).
454. Quintyn, R.S., Zhou, M., Yan, J. & Wysocki, V.H. Surface-Induced Dissociation Mass Spectra as a Tool for Distinguishing Different Structural Forms of Gas-Phase Multimeric Protein Complexes. *Anal Chem* **87**, 11879-86 (2015).
455. Ganguly, D. & Chen, J. Topology-based modeling of intrinsically disordered proteins: balancing intrinsic folding and intermolecular interactions. *Proteins* **79**, 1251-66 (2011).
456. Ganguly, D. et al. Electrostatically accelerated coupled binding and folding of intrinsically disordered proteins. *J Mol Biol* **422**, 674-84 (2012).
457. Law, S.M., Gagnon, J.K., Mapp, A.K. & Brooks, C.L., 3rd. Prepaying the entropic cost for allosteric regulation in KIX. *Proc Natl Acad Sci U S A* **111**, 12067-72 (2014).
458. Lu, Q., Lu, H.P. & Wang, J. Exploring the mechanism of flexible biomolecular recognition with single molecule dynamics. *Phys Rev Lett* **98**, 128105 (2007).
459. Turjanski, A.G., Gutkind, J.S., Best, R.B. & Hummer, G. Binding-induced folding of a natively unstructured transcription factor. *PLoS Comput Biol* **4**, e1000060 (2008).
460. Bauer, R.J., Wolff, I.D., Zuo, X., Lin, H.K. & Trakselis, M.A. Assembly and distributive action of an archaeal DNA polymerase holoenzyme. *J Mol Biol* **425**, 4820-36 (2013).
461. Miyata, T. et al. Open clamp structure in the clamp-loading complex visualized by electron microscopic image analysis. *Proc Natl Acad Sci U S A* **102**, 13795-800 (2005).
462. Oakley, A.J. Dynamics of Open DNA Sliding Clamps. *PLoS One* **11**, e0154899 (2016).
463. Kelch, B.A., Makino, D.L., O'Donnell, M. & Kuriyan, J. Clamp loader ATPases and the evolution of DNA replication machinery. *BMC Biol* **10**, 34 (2012).
464. Zhuang, Z., Yoder, B.L., Burgers, P.M. & Benkovic, S.J. The structure of a ring-opened proliferating cell nuclear antigen-replication factor C complex revealed by fluorescence energy transfer. *Proc Natl Acad Sci U S A* **103**, 2546-51 (2006).
465. Kumar, R., Nashine, V.C., Mishra, P.P., Benkovic, S.J. & Lee, T.H. Stepwise loading of yeast clamp revealed by ensemble and single-molecule studies. *Proc Natl Acad Sci U S A* **107**, 19736-41 (2010).
466. Adelman, J.L., Chodera, J.D., Kuo, I.F., Miller, T.F., 3rd & Barsky, D. The mechanical properties of PCNA: implications for the loading and function of a DNA sliding clamp. *Biophys J* **98**, 3062-9 (2010).
467. Kazmirski, S.L., Zhao, Y., Bowman, G.D., O'Donnell, M. & Kuriyan, J. Out-of-plane motions in open sliding clamps: molecular dynamics simulations of eukaryotic and archaeal proliferating cell nuclear antigen. *Proc Natl Acad Sci U S A* **102**, 13801-6 (2005).
468. Ruotolo, B.T. et al. Evidence for macromolecular protein rings in the absence of bulk water. *Science* **310**, 1658-61 (2005).

469. Fagerbakke, K.M., Norland, S. & Heldal, M. The inorganic ion content of native aquatic bacteria. *Can J Microbiol* **45**, 304-11 (1999).
470. Ando, T. & Skolnick, J. Crowding and hydrodynamic interactions likely dominate in vivo macromolecular motion. *Proc Natl Acad Sci U S A* **107**, 18457-62 (2010).
471. Jiang, F. & Doudna, J.A. CRISPR-Cas9 Structures and Mechanisms. *Annu Rev Biophys* (2017).
472. Horvath, P. & Barrangou, R. CRISPR/Cas, the immune system of bacteria and archaea. *Science* **327**, 167-70 (2010).
473. Sternberg, S.H. & Doudna, J.A. Expanding the Biologist's Toolkit with CRISPR-Cas9. *Mol Cell* **58**, 568-74 (2015).
474. Johnson, K.A. A century of enzyme kinetic analysis, 1913 to 2013. *FEBS Lett* **587**, 2753-66 (2013).
475. Maher, R.L. & Bloom, L.B. Pre-steady-state kinetic characterization of the AP endonuclease activity of human AP endonuclease 1. *J Biol Chem* **282**, 30577-85 (2007).
476. Schermerhorn, K.M. & Delaney, S. Transient-state kinetics of apurinic/apyrimidinic (AP) endonuclease 1 acting on an authentic AP site and commonly used substrate analogs: the effect of diverse metal ions and base mismatches. *Biochemistry* **52**, 7669-77 (2013).
477. Richardson, C.D., Ray, G.J., DeWitt, M.A., Curie, G.L. & Corn, J.E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat Biotechnol* **34**, 339-44 (2016).
478. Anders, C. & Jinek, M. In vitro enzymology of Cas9. *Methods Enzymol* **546**, 1-20 (2014).
479. Pollard, T.D. A guide to simple and informative binding assays. *Mol Biol Cell* **21**, 4061-7 (2010).
480. Chandradoss, S.D. et al. Surface passivation for single-molecule protein studies. *J Vis Exp* **1**, e50549 (2014).
481. Rauch, B.J. et al. Inhibition of CRISPR-Cas9 with Bacteriophage Proteins. *Cell* **168**, 150-158 e10 (2017).
482. Bondy-Denomy, J. et al. Multiple mechanisms for CRISPR-Cas inhibition by anti-CRISPR proteins. *Nature* **526**, 136-9 (2015).
483. Pawluk, A. et al. Naturally Occurring Off-Switches for CRISPR-Cas9. *Cell* **167**, 1829-1838 e9 (2016).
484. Marraffini, L.A. & Sontheimer, E.J. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat Rev Genet* **11**, 181-90 (2010).
485. Chen, J.S. & Doudna, J.A. The chemistry of Cas9 and its CRISPR colleagues. *Nature Reviews Chemistry* **1**, 0078 (2017).
486. Chen, W., Zhang, Y., Yeo, W.S., Bae, T. & Ji, Q. Rapid and Efficient Genome Editing in Staphylococcus aureus by Using an Engineered CRISPR/Cas9 System. *J Am Chem Soc* **139**, 3790-3795 (2017).
487. Cencic, R. et al. Protospacer adjacent motif (PAM)-distal sequences engage CRISPR Cas9 DNA target cleavage. *PLoS One* **9**, e109213 (2014).

488. Zhang, Y. et al. Comparison of non-canonical PAMs for CRISPR/Cas9-mediated DNA cleavage in human cells. *Sci Rep* **4**, 5405 (2014).
489. Zuo, Z. & Liu, J. Cas9-catalyzed DNA Cleavage Generates Staggered Ends: Evidence from Molecular Dynamics Simulations. *Sci Rep* **5**, 37584 (2016).
490. Liu, Y. et al. In Vitro CRISPR/Cas9 System for Efficient Targeted DNA Editing. *MBio* **6**, e01714-15 (2015).
491. Jiang, W. et al. Cas9-Assisted Targeting of CHromosome segments CATCH enables one-step targeted cloning of large gene clusters. *Nat Commun* **6**, 8101 (2015).
492. Ma, Y. et al. Increasing the efficiency of CRISPR/Cas9-mediated precise genome editing in rats by inhibiting NHEJ and using Cas9 protein. *RNA Biol* **13**, 605-12 (2016).
493. Johnson, K.A. Fitting enzyme kinetic data with KinTek Global Kinetic Explorer. *Methods Enzymol* **467**, 601-26 (2009).
494. Johnson, K.A., Simpson, Z.B. & Blom, T. Global kinetic explorer: a new computer program for dynamic simulation and fitting of kinetic data. *Anal Biochem* **387**, 20-9 (2009).
495. Shou, J., Li, J. & Wu, Q. Precise And Predictable DNA Fragment Editing Reveals Principles Of Cas9-Mediated Nucleotide Insertion. *bioRxiv* (2017).