UNIVERSITY OF CALIFORNIA,
IRVINE

Detection of DNA Polymerase I (Klenow Fragment) Activity Using Single-Molecule Electronic Nanocircuits

and

Toward the *In Vitro* Evolution of High-Affinity Binding Partners for Membrane Proteins From Phage-Displayed Affinity Reagent Libraries

DISSERTATION

submitted in partial satisfaction of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in Chemistry

by

Kaitlin Michelle-Peck Pugliese

Dissertation Committee:
Professor Gregory A. Weiss, Chair
Professor Jennifer A. Prescher
Professor David L. Van Vranken

2016

DEDICATION

To

My mentors,

For everything they have taught me

My friends and family,

For accepting me for who I am and supporting my independence

My husband,

For giving me the gift of unconditional love and support every single day

"In the middle of difficulty lies opportunity"

– Albert Einstein

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I want to thank the American Chemical Society for permission to include Chapter 2 of my dissertation, which was originally published in the Journal of the American Chemical Society. Financial support was provided by the University of California, Irvine, NSF Grants DMR-1104629 and ECCS-1231910, NIH NCI Grant R01 CA133592-01, and NIH NIGMS Grant 1R01GM106957-01.

Thank you to the University of California, Irvine for accepting me into their Chemistry department. The opportunity that UCI provided to me allowed me to develop confidence in myself as a scientist.

I am forever grateful to my committee chair, Professor Gregory A. Weiss, for accepting me into his laboratory. I knew the day that I met him that he was the boss for me, and it was more than just the fact that we are both redheads. Greg's enthusiasm is incredibly contagious. He has made me a better writer, communicator, marketer, interviewee, interviewer, researcher, teacher, and so much more. I thank him for giving me to opportunity to substitute lecture in his classes of 250+ students. The best compliment that I could receive when substituting for his lectures was that I came anywhere close to resembling him as a teacher. Greg has been incredibly supportive of my goals.

Thank you to my committee members, Professor Jennifer Prescher and Professor David Van Vranken. Jenn's implementation of the "Life After the Ph.D." seminar series has opened my eyes to many career opportunities and is one of the reasons that I am so proud to have attended UCI for graduate school. Having the opportunity to learn chemical biology from Jenn was a true pleasure and I cannot wait to tell people about it when she becomes even more famous. Dave's undivided attention when discussing my research progress at our annual

meetings meant the world to me. Regardless of how busy his schedule was, he always sat with me for at least an hour with true interest in what I was working on and I appreciated that tremendously.

I want to thank Professor Phil Collins for the quick acceptance of me into our collaboration and allowing me to work in his laboratory. He has challenged me in many ways and has been an incredible source to learn from in the field of physics. He always treated me with respect, and this encouraged me to have conviction in my ideas.

I want to thank Dr. O. Tolga Gul and Dr. Yongki Choi for being great colleagues during the many hours of measurements during which we sat together staring at a screen. I want to thank Dr. Max Akhterov for also being a great colleague and developing the high-bandwidth system for measuring protein dynamics. I want to thank Dr. Patrick Sims for the last minute experiments that he completed for our paper when he really didn't have to. I want to thank Dr. Denys Marushchak for his meticulousness and collaboration, and Mackenzie Walker-Turvey for teaching me the same things over and over again. I want to thank the other members of the Collins lab for always being willing to answer questions, especially when we had to learn how to perform AFM.

I want to thank all the past and current members of the Weiss laboratory. Dr. Lucie Lee had the reputation of training graduate students incredibly well and I am so grateful to have been one of those students. Overlapping with Dr. Glenn Eldridge, Dr. Agi Hajduczki, Dr. Rosemarie Vithayathil, Dr. Cathie Overstreet, Dr. Issa Moody, Dr. Tom Yuan, Andrew Gansmiller, Johnny Rodriguez, and Rick Hooy was a great opportunity and allowed me to learn so much. Thank you to Dr. Jessica Arter Williams for being a great friend and for your continued support. I want to thank Dr. Tivoli Olsen for leaving her baby in my hands. I appreciate her confidence in me to

take the project forward and I am very grateful to have had such a great foundation from which to build this thesis. It was such a great pleasure to have undergraduates Chinelo Okany and Paul Hetman work with me for some time, and I hope I was able to teach them something useful. I want to thank Callum Ormonde and Mark Balk – their time in the lab was short but was plenty enjoyable.

I want to thank Luz Meneghini for her thoughtfulness – she always thinks of me when she notices an opportunity that I would be interested in and it means so much to me. She has also been a great friend and someone I could commiserate with about membrane proteins. I have never met anyone quite like Dr. Amanda Gilliam, and her particular sense of humor is something that I do not forget. I want to thank Johnathon Truong for being such a great sport this last year. Johnathon has made a significant contribution to Chapter 4 of my dissertation and I was impressed with his ability to remain calm when those >20 cell pellets were left at room temperature overnight. I hope I have had a positive effect on him, because the time spent training and working with him has been a pleasure. Thank you to Mariam Iftikhar for being my bay-mate, and, in effect, my complaining partner. I think this is an essential colleague to have in graduate school, and even better when you just have to turn your chair. Thank you to Josh Smith for your self-deprecating humor and other bits of comic relief – he is never really in a bad mood and I appreciate that about a person. Thank you to (Dr.!!) Kritika Mohan for your support as we both went on this journey together – through advancement and now through finally completing our Ph.Ds. It cannot be understated the value of having someone going through the exact stages that you are and I truly appreciated her answering all of my annoying questions while we both worked on our theses. I want to thank Joshua Edgar, Shae Schlegel, Jeffrey Briggs, Kevin Johnston, Marcus Woodworth, and Andrew Nguyen for not calling me the lab nag – at least not

to my face. I want to thank Rebekah Dyer for joining the lab and being a great source of positivity – I think she will be incredibly successful in the lab and I wish her the best of luck. Thank you to Dr. Sudipta Majumdar for being the lab Jedi – something I think every lab needs. His reputation preceded him and he did not disappoint. Although I just missed him in the lab when I started graduate school, I am incredibly grateful that I could overlap with him toward the end of my Ph.D. He has given me great advice and guidance. I want to thank Josh Britton for being my spin class and In and Out partner and sharing my guilty pleasure of Bravo reality television. I want to thank Dr. Mark Richardson for the many lunches in which we have talked at length about a range of topics. Discussions with Mark have given me a great deal of confidence in myself and I will be forever grateful for that. I want to also thank him for not really minding if many of our conversations consisted of me talking about 95% of the time about my woes – he has become an incredible friend.

I want to thank Emma Flores for everything she has done for the GPS-BIOMED program. This program has exposed me to so many professionals and careers, and has given me an immense amount of knowledge. I want to also thank Bri McWhorter, who gave me great advice about my elevator pitch that I have carried with me into interviews and other professional settings.

I want to thank Dr. Robert Lin, Mona Wood, and Abraham Chiu for being an amazing group of people with which I could share my career interests. We have spent hours doing case interviews and discussing our goals, and I only wish that I found them sooner. I can't wait to follow their successes.

I want to thank the Prescher, Martin, Heyduk, Chamberlin, Nowick, and Edinger labs for being great neighbors, letting us use their equipment, and being filled with helpful colleagues.

I want to thank Kevin Stockdale and Heather McCoy at KUCI 88.9 FM for giving me the opportunity to do a radio show and to be involved with an interesting new group of people.

I want to thank everyone in Iota Sigma Pi for all the work they have done to help advance women in science. Professor Liz Jarvo was incredibly helpful in obtaining funding when we put together a conflict resolution workshop, and made the rewarding experience possible.

I want to thank everyone in the Department of Chemistry at UCI – the faculty, the administrative staff, and facilities managers. I never left the Department office without a smile on my face. I want to thank Professor A. Richard Chamberlin and Professor Andrej Luptak for their participation in reading my 2nd year report, and Professor James Nowick, Professor Scott Rychnovsky, and Professor Michelle Khine for being part of my advancement committee. I also want to thank James for his involvement in the Chemical Biology summer program, which I was able to experience in Summer 2010 for the beginning of my Ph.D.

I want to thank Professor John Welch for originally giving me the opportunity to do research. I knocked on his door one fateful day when I realized I should be doing research if I wanted any hope of getting into graduate school. He answered the door, invited me in, talked with me for over an hour, and convinced me to go to graduate school. His belief in me has been instrumental in my development, including my acceptance into graduate school. I am forever grateful to the opportunities he has provided for me. He has taught me to speak well and always encouraged me to "center my chi" when I started to get crazy. I want to thank Dr. Paul Savoie, Dr. Halimah Sayahi, and Dr. Seiichiro Higashiya for being my original graduate student/post-doc mentors. I remember learning more my first week in a research lab than I had learned during all my instructional labs, and these three were responsible for that. I want to thank everyone else in

the Department of Chemistry at the University at Albany for giving me an excellent foundation in chemistry.

I want to thank all of the friends and colleagues that I have met in graduate school – Olivia Cromwell, Lindsay Cameron, Joanna Laird, Rachel Steinhardt, Domarin Khago, Greg Williams, Janice Wong, Mary Beth Daub, and many others that have been such a joy to learn from and interact with. I look forward to all of their successes.

I want to thank Brittany Bogardus, Kara Risley, Sarah Makarchuk, Katie Mutz, and Ashley Smith for being my best friends for over 15 years. These are the people that really know me and with whom I can be my true self.

I want to thank the people that have become my family and immediately treated me as such – Cathy Pugliese, Carl Pugliese, Peter Pugliese, Michael Ciarlo, and Tim Bishop. I am incredibly lucky to have people that have immediately accepted into their lives and have done everything to help me.

I want to thank my family for loving and supporting me even though I never call them. My sisters Keegan and Kara and my brother Kyle have taught me so much about being a big and a little sister and they completely accept me for who I am. They have supported my independence and I hope I have made them proud. My parents are literally the best parents in the world. They both work incredibly hard and have given me all of the opportunities that they possibly could. I am grateful to them for putting me in sports and dance class from 5 years old. These activities have taught be how to be coached, how to perform, how to be confident, and how to work on a team. These skills have been tremendously useful in my life. No matter how busy they were, I don't remember them missing a game. Their incredible support and trust of

me, and their acceptance of my independence have allowed me to flourish into the person I am today.

Finally, I want to thank my husband, Steve. Graduate school has been hard and my failures have often made me unpleasant. Steve has been there through it all and understood all of the late nights or times when I said I would be done in an hour, which ended up being three hours. Going home to Steve is my greatest joy in life. I am literally obsessed with him. I am so lucky to have a partner that accepts my ambition and wants to support me in every way possible. He gives me something to look forward to every single day and makes me feel like the smartest person alive.

CURRICULUM VITAE

KAITLIN M. PUGLIESE

kaitlin251@gmail.com
315.420.3222
6307 Palo Verde Road
Irvine, CA 92617

EDUCATION

Ph.D., Chemistry 02/2016
University of California, Irvine (Irvine, CA)

M.S. coursework, Chemistry 08/2009 – 05/2010
University at Albany, SUNY (Albany, NY)

B.S., Chemistry (Forensic Chemistry emphasis), Cum Laude 08/2005 – 05/2009
University at Albany, SUNY (Albany, NY)

RESEARCH EXPERIENCE

Graduate Student Researcher 07/2010 – present
University of California, Irvine - Department of Chemistry (Irvine, CA)

Thesis Advisor: Professor Gregory A. Weiss

- Engineered enzyme variants and studied catalysis of single enzymes using electronic nano-circuits
- Progressed development of a selection strategy for membrane proteins using phage-display
- Prepared insoluble proteins for refolding by a vortex fluid device
- Screened small molecule library and identified potential inhibitors of the HIV-1 Vif-Gag interaction

Graduate Student Researcher 09/2009 – 05/2010
University at Albany, SUNY - Department of Chemistry (Albany, NY)
Advisor: Professor John T. Welch

- Screened possible inhibitors of anti-leishmanial targets SIRT1 and SIRT2 in preliminary assay
- Expressed/purified *Mtb* FAS I and analyzed inhibition by pyrazinamide analogs using NMR spectroscopy

Research Scientist 05/2009 – 09/2009
Central New York Research Corporation (Syracuse, NY)
Advisors: Professor John T. Welch and Dr. Michael Cynamon

- Synthesized and purified 5-chloropyrazinamide for clinical animal trials against *Mtb* FAS I

Undergraduate Researcher 01/2009 – 05/2009
University at Albany, SUNY - Department of Chemistry (Albany, NY)
Advisor: Professor John T. Welch

- Synthesized and analyzed pentafluorosulfanyl-functionalized β-lactam derivatives

ADDITIONAL LEADERSHIP, COMMUNICATION, AND WORK EXPERIENCE

Laboratory Management, *Weiss Laboratory* 06/2011 – present

- Manage laboratory group job organization, meeting schedules, and petty cash accounting
- Developed new process that reduced disposable item consumption by 50% and increased throughput 400%
- Negotiated service contracts and created a sustainable system for maintenance of essential instrumentation
- Previously acted as safety representative, managing standard operating procedures and ensuring safety

Vice President, *Iota Sigma Pi – Calcium Chapter* 06/2011 – 06/2015

- Volunteered at various activities to engage young people in chemistry, including National Chemistry Week at the Santa Ana Zoo, Girls Inc., Ask-a-Scientist Night, and the UC Irvine student-run Wayzgoose festival
- Organized successful conflict resolution workshop reaching several departments across campus
- Diversified fundraising and increased efficiency of surveys, elections, and ordering merchandise
- Previously acted as Treasurer (2011-2012) and Secretary (2012-2013)

Founding Member and President, *Advanced Degree Consulting Club – UCI* 07/2015 – present

- Organize weekly meetings with post-doctoral associates and graduate students to foster consulting skills
- Established connections with former and current consultants to enhance student development

Project Consultant, *Oxbridge Biotech Roundtable* 06/2014 – 10/2014

- Interviewed several biotech entrepreneurs and rewrote a report about SBIR grants for a project aimed to improve government funding of biotech startups

Co-Coordinator, *Graduate Student and Post-Doctoral Colloquium (UCI)* 06/2012 – 06/2014

- Organized monthly symposia featuring UCI Chemistry student and post-doctoral research

Teaching Assistant, *Department of Chemistry Teaching Program (UCI)* 09/2010 – 06/2014

- Led general and organic chemistry laboratory sections and ensured student safety
- Substitute-lectured for Organic Chemistry and Intro to Chemical Biology courses
- Recognized for outstanding contributions to the teaching program (May, 2012)
- Received consistently strong evaluations from students

Public Affairs Radio Show Host, *"Red Does Research" on KUCI 88.9FM* 01/2013 – 12/2013

- Produced weekly radio show, communicating science to the public through relatable discussions

Graduate Student Mentor, *UCI Competitive Edge Summer Research Program* Summer, 2012

- Advised incoming graduate students in an effort to promote diversity retention in STEM

UNDERGRADUATE MENTORING

Johnathon Truong, *Undergraduate Researcher*, UCI 04/2015 – present

Paul Hetman, *Undergraduate Researcher,* UCI 02/2013 – 03/2014

Chinelo Okany, *NSF-REU Researcher,* UCI Summer, 2012

Sadia Azar, *Undergraduate Researcher,* SUNY Albany 12/2009 – 05/2010

Tom Fredriksen, *Undergraduate Researcher,* SUNY Albany 08/2009 – 11/2009

PUBLICATIONS

O. Tolga Gul*, **K. M. Pugliese***, Y. Choi, P. Collins, and G. A. Weiss. "Base counting with single nucleotide resolution using DNA Polymerase I (Klenow Fragment) electronic nanocircuits." *Manuscript in preparation.*

K. M. Pugliese*, O. Tolga Gul*, T. J. Olsen, Y. Choi, P. Collins, and G. A. Weiss. "Processive Incorporation of Deoxynucleoside Triphosphate Analogs by Single-Molecule DNA Polymerase I (Klenow Fragment) Nanocircuits". *J. Am. Chem. Soc.,* **2015**, *137*, 9587-9594.

T. Z. Yuan, C. F. G. Ormonde, S. T. Kudlacek, S. Kunche, J. N. Smith, W. A. Brown, **K. M. Pugliese**, T. J. Olsen, M. Iftikhar, C. L. Raston, and G. A. Weiss. "Shear-Stress-Mediated Refolding of Proteins from Aggregates and Inclusion Bodies." *ChemBioChem,* **2015**, *16*, 393-396.

H. Sayahi, **K. M. Pugliese**, O. Zimhony, W. R. Jacobs, Jr., A. Shekhtman, and J. T. Welch. "Analogs of the Antituberculous Agent Pyrazinamide Are Competitive Inhibitors of NADPH Binding to *M. tuberculosis* Fatty Acid Synthase I." *Chem. Biodivers.,* **2012**, *9*, 2582-2596.

S. C. Ngo, J. Lin, P. R. Savoie, E. M. Hines, **K. M. Pugliese** and J. T. Welch. "Preparation and Reactions of Aliphatic 2-Pentafluorosulfanyl Aldehydes." *E. J. Org. Chem*., **2012**, 4902-4905.

Z. Ahmad, S. Tyagi, A. Minkowski, D. Almeida, E. L. Nuermberger, **K. M. Peck**, J. T. Welch, A. S. Baughn, W. R. Jacobs, Jr., and J. H. Grosset. "Activity of 5-chloro-pyrazinamide in mice infected with *Mycobacterium tuberculosis* or *Mycobacterium bovis*." *Indian J. Med. Res.*, **2012**, *136*, 124-130.

*Authors contributed equally

POSTERS and PRESENTATIONS

K. M. Pugliese, O. T. Gul, M. Akhterov, P. G. Collins, G. A. Weiss. "Observing Processive Incorporation of Deoxynucleoside Triphosphates with DNA Polymerase I (Klenow Fragment) Electronic Nanocircuits." Inaugural Gordon Research Seminar and Gordon Research Conference on Enzymes, Coenzymes, & Metabolic Pathways, Waterville Valley, NH, July 11-17, 2015.

K. M. Pugliese, O. T. Gul, T. J. Olsen, Y. Choi, P. G. Collins, and G. A. Weiss. "Single-Molecule Electronic Measurements of DNA Polymerase I Klenow Fragment." 248th ACS National Meeting & Exposition, San Francisco, CA, August 10-14, 2014.

M. Iftikhar, **K. M. Pugliese**, T. J. Olsen, Y. Choi, P. C. Sims, I. S. Moody, O. T. Gul, B. Corso, G. A. Weiss, and P. G. Collins. "Listening to Single Proteins via Electronic Circuits." 248th ACS National Meeting & Exposition, San Francisco, CA, August 10-14, 2014.

K. M. Pugliese. "Engineering of Vif for Targeted Degradation and In Vitro Evolution of Binding Partners for ShuA." UC Irvine Chemistry Department Post-Doc and Graduate Student Colloquium, Irvine, CA, April 27, 2012.

J. T. Welch, H. Sayahi, A. Shekhtman and **K. Peck**, "Saturation Transfer Difference NMR Studies of the Inhibition of *M. tuberculosis* FAS I by Pyrazinamide and Pyrazinamide Analogs". Symposium on Encouraging Development of Therapeutics for Neglected Diseases, Philadelphia, PA, June 16-17, 2010.

J. T. Welch, H. Sayahi, S. Puttamadappa, **K. M. Peck** and A. Shekhtman. "5-Fluoropyrazinoic Acid Derivatives: Inhibition of *Mycobacterium tuberculosis*; Fatty Acid Synthase I". 19th International Symposium on Fluorine Chemistry, Jackson Hole, WY, August 23-38, 2009.

J. T. Welch, S. C. Ngo, P. R. Savoie, **K. M. Peck** and L. Zhong "Preparation and Reactions of Aliphatic Pentafluorosulfanyl Aldehydes and Ketones". 19th International Symposium on Fluorine Chemistry, Jackson Hole, WY, August 23-38, 2009.

AWARDS AND HONORS

3rd Place – Inaugural Elevator Pitch Competition, UC Irvine	2015
Recognition for Contributions to the Chemistry Department Teaching Program, UCI	2012
Chemistry Departmental Honors, University at Albany	2009
Presidential Scholar, University at Albany	2005-2009
Dean's List, University at Albany	2005-2009

AFFILIATIONS

Inaugural Trainee, Graduate Professional Success in the Biomedical Sciences	2015 – present
Iota Sigma Pi, National Honor Society for Women in Chemistry – Ca Chapter	2011 – present
American Chemical Society (ACS)	2009 – present

PROFESSIONAL REFERENCES

Gregory A. Weiss, Ph.D.
University of California, Irvine
Department of Chemistry
Irvine, CA 92697
(949) 824-5566
gweiss@uci.edu

John T. Welch, Ph.D.
University at Albany
Department of Chemistry
Albany, NY 12222
(518) 442-4455
jwelch@uamail.albany.edu

ABSTRACT OF THE DISSERTATION

Detection of DNA Polymerase I (Klenow Fragment) Activity Using

Single-Molecule Electronic Nanocircuits

and

Toward to *In vitro* Evolution of High-Affinity Binding Partners to Membrane Proteins

From Phage-Displayed Affinity Reagent Libraries

By

Kaitlin Michelle-Peck Pugliese

Doctor of Philosophy in Chemistry

University of California, Irvine, 2016

Professor Gregory A. Weiss, Chair

DNA polymerases are central biocatalysts in the life cycle of all organisms. Structural transitions of this enzyme during the incorporation of deoxynucleoside triphosphate (dNTP) substrates govern the accurate replication of DNA. Biophysical tools have emerged as useful techniques for understanding these structural transitions, and are continually being improved. Electronic-based techniques, in particular, provide many advantages over other biophysical techniques, including long duration monitoring and enhanced time resolution. For example, recently developed single-walled carbon nanotube field-effect transistors that measure DNA polymerase I Klenow Fragment (KF) activity serve as a valuable method for label-free detection

of biocatalysis. In my Ph.D. work, I have sought to develop this technique further to better understand enzyme activity and take advantage of our observations for biotechnological applications.

Incorporation of dNTP analogs with KF-functionalized nanocircuits challenged the molecular recognition of these substrates by KF. For some dNTP analogs, an alternative motion of KF during dNTP incorporation was observed and may be valuable in the discrimination of DNA bases in a template. The nanocircuits also accurately identified DNA template lengths when such templates were composed of highly repetitive sequences. Accumulation of several data sets further enabled precise measurement of DNA template lengths. Initial experiments of dNTP incorporation with a KF nanocircuit measured at high bandwidths demonstrated that the accurate measurement of DNA template lengths required the lower time resolution. The higher bandwidth experiments revealed previously unobserved KF motions that lasted for short durations, making this technique unsuitable for counting DNA bases. However, improvement to KF nanocircuits measured at high bandwidth can provide a powerful tool for understanding transient KF dynamics.

To better understand and improve KF nanocircuits, a variety of KF variants were designed and generated. Many of these variants exhibited polymerase activity at the single-molecule level, and are in various stages of progression for their measurement with polymerase-functionalized nanocircuits. The experiments necessary to determine their viability in improving and understanding the electronic signals generated by KF were outlined for future Weiss lab members.

Membrane proteins are a class of biomolecules that also benefit from the improved understanding of their molecular mechanisms. *In vitro* evolution of high affinity binding

partners with stable, rigid phage-displayed protein libraries were attempted for the *Helicobacter pylori* inner membrane protein Ure-I, but were unsuccessful due to Ure-I's rapid degradation. An alternative method for *in vitro* evolution for high affinity binding partners to the *Shigella dysenteriae* outer membrane protein ShuA was optimized to maintain ShuA in a native membrane environment.

Chapter 1: Conformational Transitions of DNA Polymerases During Nucleotide Incorporation and Opportunities for Applications in Biotechnology

1.1 Abstract

DNA polymerases achieve high fidelity through an intricate mechanism of correct nucleotide selection. Kinetic, mutational, and structural analysis initially identified the conformational transitions involved in this stage of the polymerase catalytic cycle. Real-time biophysical techniques developed over the last 15 years have elucidated the more dynamic structural changes that correspond to molecular recognition. Most recently, electronic-based methods for the sensitive detection of enzyme motion have observed the incorporation of nucleotides in real-time at ultrafast time scales. Some of these techniques have been adapted for use in DNA sequencing and have the potential to serve both fundamental and applied needs. Many opportunities exist to improve electronic biosensors in such a way that the distinct polymerase motions that correspond to base discrimination can be hijacked for DNA sequencing and identification of epigenetic modifications.

1.2 Introduction

DNA polymerases accurately copy genetic information through a set of carefully choreographed, reiterative steps. In the presence of a primed DNA template, DNA polymerases use Mg^{2+} ions to aid phosphoryl transfer of an incoming deoxynucleoside triphosphate (dNTP) onto the 3' hydroxyl of the primer (Figure 1-1).[1] Enzyme selection of the correct dNTP substrate from an assortment of closely related molecules is a principal component of this mechanism.[2] As a result, DNA polymerases have long served as general models for molecular recognition.[3] A complete picture of this part of the DNA polymerase catalytic cycle can help to explain the aberrant enzyme activity that leads to mutation and to guide drug design to counteract resultant diseases.

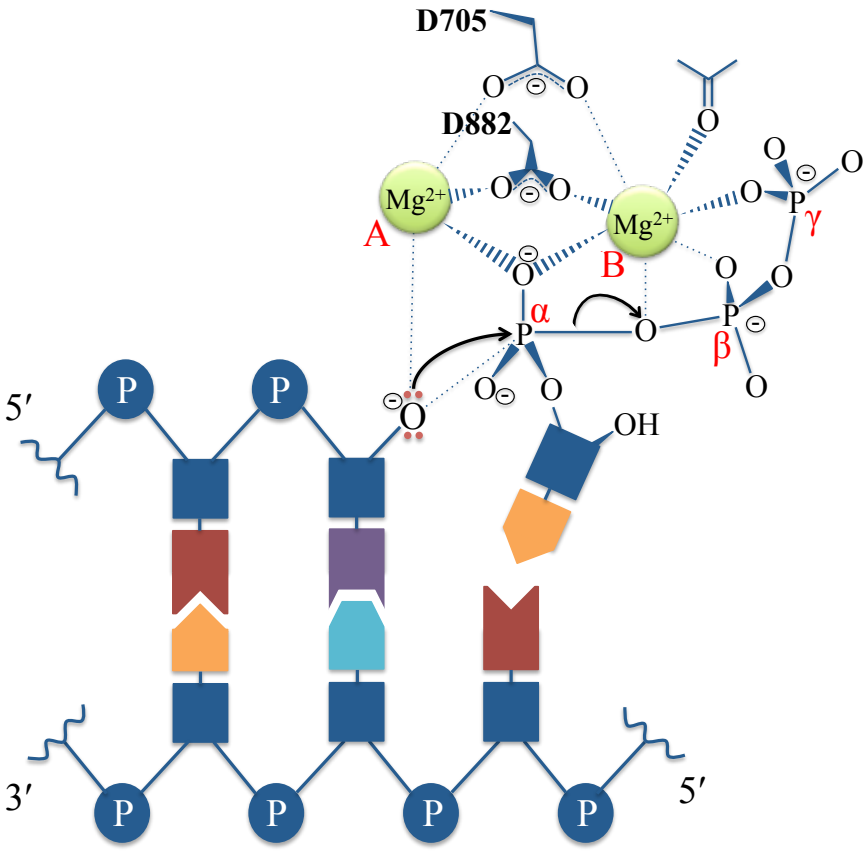


Figure 1-1. Nucleotidyl transfer by DNA polymerase I Klenow Fragment (KF) (description on following page).

Figure 1-1. Nucleotidyl transfer by DNA polymerase I Klenow Fragment (KF). DNA bases are shown schematically in orange, dark red, purple, and light blue. The phosphate/deoxyribose backbone is shown schematically in dark blue. Two metal ions in the active site stabilize the transition state. The 3'-OH of the template base attacks the α-phosphate of the incoming dNTP upon activation by metal ion A. Metal ion B coordinates to the β- and γ-phosphate, and stabilizes the bridging oxygen between the α- and β-phosphate. KF residues D705 and D882 also make contacts with the metal ions. Adapted from Brautigam and Steitz.[169]

DNA polymerases and other enzymes exist as a broad distribution of conformational states along a dynamic energy landscape.[4] A-family polymerases, including *Escherichia coli* DNA Polymerase I, use structural flexibility to achieve high fidelity (low error frequency).[5,6] This class of enzymes contains a large proteolytic fragment that retains polymerase activity and resembles a right hand with "fingers," "palm," and "thumb" domains (Figure 1-2).[7] These domains work in concert to successfully attain fidelity through specific conformational motions along the reaction pathway.[8–10] The conformational fluctuations by DNA polymerase that are associated with correct and incorrect dNTP recognition, in particular, comprise a fundamental aspect of this enzyme's catalytic cycle.[11–14] Therefore, it is essential that these structural changes and the nature of polymerase motion be fully understood. Notably, such knowledge can be exploited in biotechnological applications that require nucleobase discrimination, such as DNA sequencing.

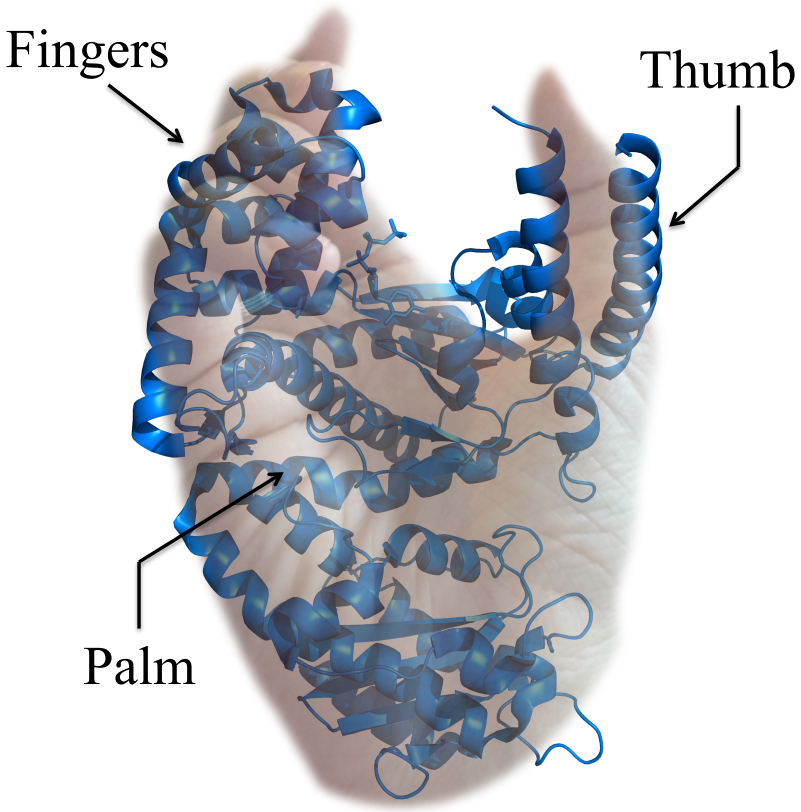


Figure 1-2. Structure of DNA polymerase overlaid with a right hand. DNA polymerase I Klenow Fragment (KF) and other A-family polymerases resemble a right hand with "fingers," "palm," and "thumb" domains (PDB ID: 1KFD).

Intense research during the last 30 years has uncovered key aspects of the mechanistic, kinetic, and structural details of the DNA polymerase reaction.[15] Determination of A-family polymerase structures established the formation of multiple distinct complexes of the enzyme and its substrates during nucleotide incorporation.[3] This structural information led to hypotheses linking the different complexes to discrete steps determined by kinetic analysis.[10] The structures inspired and guided the design of biophysical techniques for experimental analysis of the polymerase reaction.[15] Most recently, results from single-molecule experiments have been especially informative about the transient motions involved in nucleotide incorporation.[16] A comprehensive picture of the catalytic cycle, however, has been limited by the inability to experimentally determine ultrafast motions likely associated with nucleotide selection. Computer simulations have attempted to theorize these rapid motions using crystal structures, but more advanced experimental techniques are required to support such data.

Single-molecule methodologies that use DNA polymerases to read a DNA sequence are highly sensitive and require limited sample preparation.[17] Like techniques to study polymerase motion, sequencing technologies also require extraordinary time resolutions to read single bases within a DNA strand. Although sequencing DNA with single DNA polymerase molecules allows rapid detection and long reads by taking advantage of the enzyme's activity, such methods achieve poor accuracy.[18] More advanced technologies are emerging to improve this limitation, some of which could link nucleobase discrimination directly to enzyme motion during incorporation. By taking advantage of the enzyme's biophysical characteristics during the substrate recognition stage of polymerase catalysis, sequencing approaches could serve both basic research and

applied methodology. Distinct polymerase conformations that arise during correct versus incorrect substrate recognition could not only be better understood, but could also provide highly accurate information about a DNA sequence. Furthermore, alternative changes in polymerase structure that correspond to the recognition of modified bases could identify epigenetic modifications.

In this chapter, I will describe discoveries of the important conformational transitions involved in DNA polymerization with a focus on the steps immediately preceding and following nucleotide incorporation. With an established basis of the noncovalent changes worth further discussion in regard to nucleotide selectivity, I will introduce the development of experimental and theoretical techniques that serve to understand these changes. This technical description will highlight the advantages and limitations of each technique in addition to their relevance in nucleobase discrimination. Finally, I will summarize the current and emerging polymerase-driven DNA sequencing technologies that may also be used to fundamentally understand molecular recognition by DNA polymerases. This survey will also assess improvements that could be made to emerging sequencing technologies to enable fundamental study of DNA polymerase conformational dynamics.

1.3 Indirect evidence of conformational transitions through kinetic evaluation

Initial insights into important conformational transitions along the DNA polymerase I reaction pathway were determined by kinetic assays. In the majority of these kinetic studies, the large proteolytic fragment, or Klenow fragment (KF), of *E. coli* DNA polymerase I was used to model the class of enzymes responsible for catalyzing DNA polymerization. KF retains both 5'-3' polymerase and 3'-5' exonuclease activities,

but loses 5'-3' exonuclease activity.[19] In these kinetic assays, chemical-quench methods with labeled substrates monitored polymerase reactions with pre-steady-state or steady-state kinetics. These chemical-quench experiments demonstrated that single or processive nucleotide turnover, respectively, include non-covalent, rate-determining steps that later became the subject of substantial interest.

A kinetic mechanism (Figure 1-3) first postulated the presence of two partially rate-limiting steps, one before and one after chemical bond formation, that are of approximately equal magnitude.[20] The rapid incorporation of a single dNTP followed by slower subsequent incorporation in a processive cycle suggested this scheme. Further examination of assay results with single nucleotide turnover assigned the first partially rate-limiting step to a conformational change following nucleotide binding.[21–23] The second partially rate-limiting step was initially assigned to a conformational change prior to pyrophosphate (PP_i) release that only occurred after incorrect incorporation.[24] However, using single-turnover PP_i-exchange experiments, this particular conformational change was found to occur following correct incorporation as well.[23] The physical nature of the second conformational change was hypothesized to be either intramolecular DNA transfer to the exonuclease domain or translocation along the DNA template. The latter hypothesis was based on the observed competitive inhibition of PP_i and the incoming dNTP.

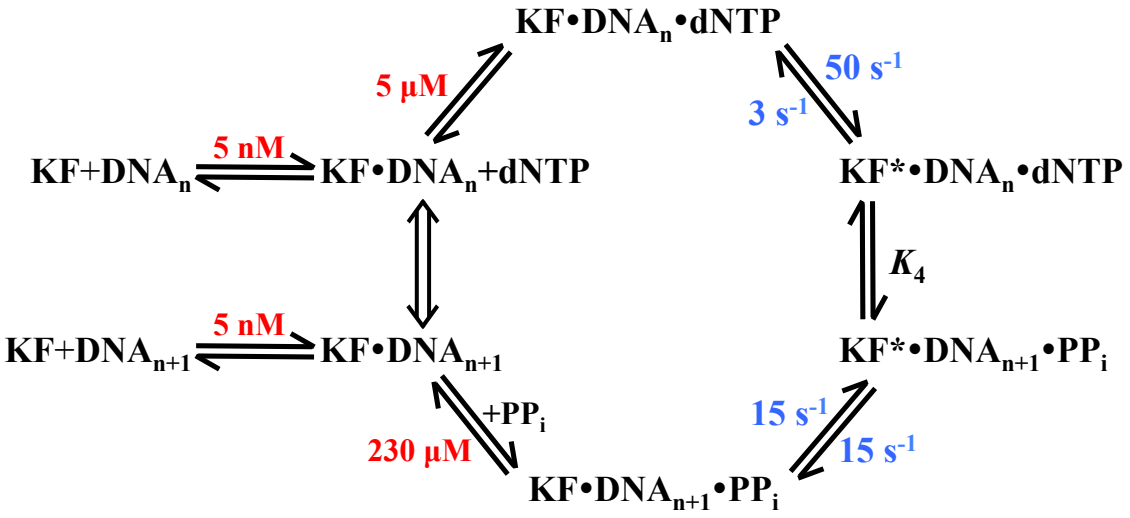


Figure 1-3. Kinetic mechanism for DNA polymerase I Klenow Fragment. The kinetic pathway for correct nucleotide incorporation includes conformational changes before and after nucleotide incorporation noted with an asterisk. Dissociation constants for DNA, dNTP, and PP_i are shown in red. Kinetic rates are shown in blue and represent dATP incorporation. Adapted from Kuchta *et al.*[22] and Dahlberg *et al.*[23]

The second rate-limiting step differs importantly for correct versus incorrect nucleotide incorporation, whereas the first rate-determining conformational change is identical in kinetic magnitude.[25] Differences between correct and incorrect incorporation in the second rate-determining conformational change play a pivotal role in the tremendous fidelity of DNA polymerases. Specifically, this second conformational change is equally rate-limiting compared to the first conformational change in correct incorporation, as mentioned previously.[23] Alternatively, the second conformational change becomes the dominant rate-limiting step for misincorporation, allowing time for internal pyrophosphorolysis or exonucleolytic cleavage.[25,26] Compared to correct

incorporation, an unfavorable energy barrier for conversion of the immediate incorrect product ternary complex with the enzyme and primer-template (P/T) to the species that releases PP_i causes the slower rate.[25] These incorporation results reinforced earlier work in which the differentiation by the exonuclease domain between correct or incorrectly paired nucleotides was not responsible for base discrimination.[24] Conclusions from experiments comparing correct versus incorrect incorporation support a potential use for non-hydrolyzable analogs; such analogs would extend the lifetime of the product ternary complex to discriminate template bases by single turnover rates.

Conformational changes before and after chemical bond formation have also been identified as critical to the catalytic cycle of T7 DNA polymerase, an A-family homolog of KF. In studies with this high-fidelity polymerase, specific mutations at the N-terminus of the protein abolished exonuclease activity and allowed stable misincorporations for examination of substrate selectivity.[27,28] Results demonstrated that T7 polymerase, like KF, retains a mismatched DNA ternary complex for sufficient time to excise the error. Additionally, T7 polymerase selects against incorrect substrates by an induced-fit mechanism prior to the first conformational change that precedes chemical bond formation.[28] The induced-fit model was further supported by tryptic digestion analyses of both T7 and KF that compared the stability of the binary complex with the P/T to ternary polymerase-P/T-dNTP complexes with correct or incorrect active site geometry.[29–31] An amino acid substitution in the exonuclease-deficient KF active site points to the polymerase domain as more critical than the exonuclease domain for accurate discrimination.[32] As such, this residue (KF I709) may indirectly introduce conformational changes in the enzyme.[33] Subsequent studies focused on identifying the

physical nature of the conformational change preceding the phosphoryl transfer step through the use of an altered substrate or enzyme.

One strategy to better understand the conformational transition that precedes chemical bond formation is through the use of either dNTP substrate analogs or altered P/T DNA. The large amount of polymerase/dNTP binding energy required to fit unnatural base pairs to exact Watson-Crick geometry indicate that such perfect geometry is not necessary for incorporation.[34] As a result, the polymerase's flexible active site likely exhibits a conformational range that, in one example, includes desolvation of the incoming dNTP.[35] This observation may explain why the first rate-determining conformational change for an incorrect dNTP that bypasses induced-fit discrimination is kinetically indistinct[25] from the conformational change for a correct dNTP. Highlights of the literature describing the importance of base pair geometry in experiments with dNTP analogs is further discussed in the introduction of Chapter 2.[36] For altered DNA templates, carcinogenic DNA adducts blocked DNA synthesis to different extents based on positioning in the KF active site and subsequent inhibition and/or significant destabilization of the conformational change.[37–39] Single-turnover experiments of T4 polymerase catalyzing translesion DNA synthesis provided further evidence that this conformational change preceding chemistry was rate-limiting.[40] Finally, a combination of a lesion-containing DNA template and dNTP analogs demonstrated that dNTP hydrophobicity and aromaticity influence the conformational change step.[41]

Another strategy to examine the first conformational change employs mutational analyses of the active site, experiments of which are further discussed in Chapter 4 and rely on the crystal structures described in the next section. Polymerase variants that

probe changes in the active site have supported earlier kinetic studies and can generate useful information about the mechanisms of nucleotide discrimination. For example, mutation of several KF active site residues led to the deduction of two kinetically distinct steps prior to phosphoryl transfer that includes the chemical step[42] and supported earlier work.[23] The binding sites for both the dNTP substrate and PP_i product were identified with site-directed mutagenesis.[43] In another study, an active-site lysine residue (KF K758) was suggested to play a role in the conformational changes just before and just after the chemical step.[44] Results from E480 and Y530 mutations in the T7 polymerase active site, which are highly conserved with respect to KF and other A-family polymerases, indicate involvement of these residues in ground state nucleotide binding.[45]

Experiments with dideoxynucleoside triphosphates (ddNTPs) and ribonucleoside triphosphates (rNTPs) revealed the amino acid residues that are involved in sugar discrimination. Conservative and non-conservative mutations of KF E710, F762, and/or Y766 led to the conclusion that the conformational change transition state prior to chemical bond formation is the site of discrimination against ddNTPs.[46] These residues are also important in discrimination against rNTPs, during which F762 primarily constrains the nucleotide while E710 acts as a "steric gate" that blocks the 2'-OH.[47] These studies, together with the conformations that can be visualized through crystal structures, can lead to conclusions about molecular dynamics and guide strategies for base discrimination.

1.4 The use of crystallography to observe static conformational transitions

Various structures of A-family DNA polymerases have been determined by X-ray crystallography. These structures, at high resolution, provide static pictures of

polymerases at different stages of incorporation, and can include details about the interplay between multiple functional domains. For example, X-ray crystal structures showed that, in order to be proofread and possibly excised, newly synthesized DNA must be shuttled from the polymerase domain to the exonuclease domain. Analysis of static polymerase complexes have elucidated important aspects of catalysis, and led to a better understanding of the biophysical mechanisms of polymerization. Furthermore, high-resolution crystal structures have been instrumental in guiding the development of more recent techniques.

The secondary structure of *E. coli* DNA polymerase I was initially predicted in 1982[48] and supported hypotheses for the presence of three functional domains.[49–51] Initial crystal structures confirmed that residues 324-907, the 68 kDa KF, contained only two of these domains.[52] In this structure, deoxythymidine monophosphate (dTMP), consistent with the product of 3'-5' exonucleolysis, was bound at the N-terminus.[53] Point mutants of KF that abolish exonuclease activity and stably bind the DNA substrates required for exonucleolysis were also crystallized.[54–56] In a separate structure, the C-terminus was bound to duplex DNA and appeared to contain a binding site for dNTP, both polymerase substrates.[57] The separation of 30-35 Å between these catalytic domains, as shown in crystal structures, suggested enzyme motions necessary to achieve the required fidelity.

Mechanistic details gleaned from structures of KF can provide insight into the KF domains' separate activities, and lead to insights about DNA partitioning from the polymerase domain to the exonuclease domain. The relationship of known polymerase structural elements to their function has been comprehensively reviewed.[12,58] For example, structures of KF complexed with either dNTP or PP_i identified positively

charged amino acid residues likely involved in binding the phosphate moieties during nucleotide incorporation.[59] However, since this structure was only a binary complex without the DNA_n substrate or DNA_{n+1} product, the position of the dNMP portion during catalysis could not be concluded.[59] Analysis of the exonuclease domain structure, on the other hand, clarified how DNA is denatured and stabilized for exonucleolysis.[55,56] Specifically, regions of disorder in the absence of DNA move in concert with two helices toward the exonuclease domain and become more structured to form the editing complex.[57] Mutational analysis, inspired by known crystal structures, determined that the J-helix changed in structure to control intramolecular transfer.[60] The motions required for shuttling from the polymerase domain to the exonuclease domain can be only suggested with available crystal structures. However, lack of an available crystal structure for the KF-P/T-dNTP ternary complex left many unanswered questions about the motions required for nucleotide incorporation.

Two A-family DNA polymerases homologous to *E. coli* DNA polymerase I were also extensively characterized through crystallographic studies: the large, or Klenow-like, fragments of *Thermus aquaticus* DNA polymerase I (KlenTaq) and of *Bacillus stearothermophilus* DNA polymerase I (BF). T7 DNA polymerase, which is also a subject of discussion throughout this chapter, has also been crystallized at high resolution.[61] Neither KlenTaq nor BF contain 3'-5' exonuclease activity and, therefore, only exhibit 5'-3' polymerase activity. Crystal structures of KlenTaq and BF provided the pictures necessary to better understand nucleotide selection by DNA polymerases.

High-resolution crystals of KlenTaq-P/T or BF-P/T binary complexes showed the same amino acid residues in homologous positions of the active site compared to KF.[62–67]

Ternary structures of the enzymes complexed with both P/T and a ddNTP[66–69] provided additional details about nucleotide incorporation. These complexes "trapped" the enzymes in a new conformation that revealed a large reorientation of the fingers subdomain necessary for catalysis and was referred to as the "closed" conformation (Figure 1-4). Additionally, "closed" complexes of BF bound to DNA containing a dideoxy-terminated primer[67] were visualized as successive replication events and demonstrated the path taken during translocation of DNA along the palm cleft. Comparison of the "open" and "closed" crystal structures revealed important intra-protein and protein-P/T interactions that later served as a guide for fluorescent and other experimental techniques to further study KF.

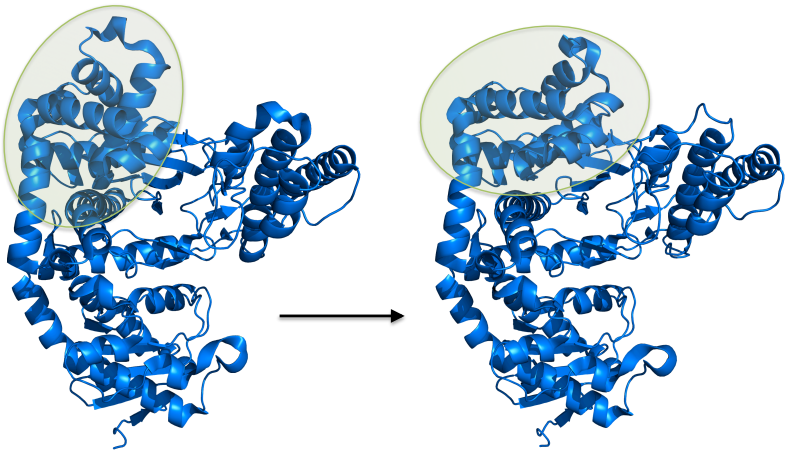


Figure 1-4. Comparison of "open" and "closed" crystal structures of DNA polymerase. The *Bacillus stearothermophilus* DNA polymerase large fragment was one of the first A-family polymerases to be crystallized in its "closed" structure. Upon dNTP binding, the fingers subdomain (circled in green) undergoes a large rearrangement prior to phosphoryl transfer. Once crystal structures of these distinct complexes were determined, the specific helices involved in this rearrangement could be rationally targeted in fluorescence-based assays.

The most recently solved DNA polymerase crystal structures present a more thorough picture of nucleotide selection through the use of non-cognate substrates. For example, rNTPs, which are present in the cell at high concentrations alongside dNTPs, result in non-productive conformations of the O-helix in the fingers subdomain.[69] Alternatively, possibly one of the most important discoveries was that of an "ajar" BF conformation in the presence of a mismatched dNTP, intermediate between the "open" and "closed" conformations of the O-helix (Figure 1-5).[70] Complexes with rare tautomers of incoming dNTPs,[71] bulky dNTP analogs,[72] and unnatural base pairs[73,74] also resulted in the "ajar" conformation along with other varied perturbations of the O-helix. Taken together, these results indicated that the various structures of polymerases required to accommodate such substrates are involved in the pathway toward nucleotide discrimination. Static crystal structures, however, do not provide adequate information about real-time protein motions. Instead, these homologous protein structures have guided experimental design, such as in fluorescent-based techniques, for elucidating the dynamic motions of specific incorporation. Distinct motions are likely required for specific dNTP incorporation and, if determined, can be applied to base discrimination.

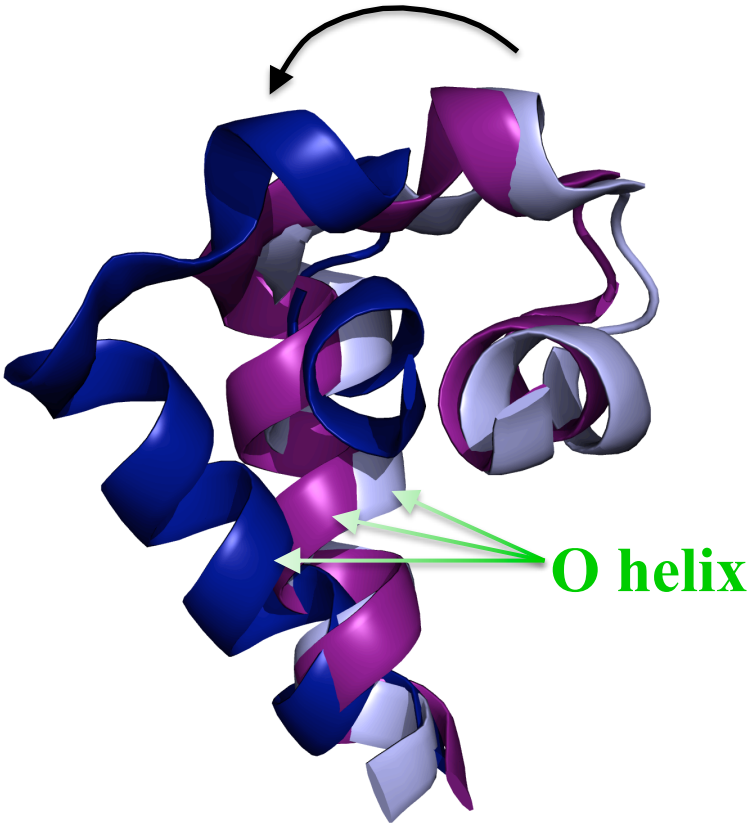


Figure 1-5. Structures of the O helix during nucleotide incorporation for the *Bacillus stearothermophilus* DNA polymerase large fragment. The discovery of an "ajar" conformation (magenta, PDB ID: 3HT3) between the "open" (light blue, PDB ID: 1L3T) and "closed" (dark blue, PDB ID: 1LV5) states provided new insights into the conformational pathway that precedes chemical bond formation.

1.5 From static to dynamic: Molecular dynamics simulations and isotopic labeling

Experimental limitations inherent to study of the rapid, non-rate-limiting steps of the DNA polymerase reaction prevent complete structural characterization of the intermediates. Even the millisecond timescale of state-of-the-art fluorescent experimental techniques can be insufficient for observing some subdomain motions and restricts necessary access to a multidimensional enzymatic process. Techniques with enhanced time resolutions, elongated time scales, and minimal non-native modifications are constantly being developed to further elucidate the dynamic processes of enzyme catalysis.[36,75]

Molecular dynamics (MD) can model protein motions using computer simulation starting with DNA polymerase crystal structures. The approach can identify the driving forces behind conformational transitions associated with catalysis. MD simulations have explored crucial aspects of DNA polymerization, including nucleotide selection,[76] fingers-opening,[77] fingers-closing,[78] and translocation.[79] Simulations of HIV reverse transcriptase, for example, linked fidelity with preferential correct dNTP recognition rather than incorrect dNTP rejection,[76] supporting single-molecule Förster Resonance Energy Transfer (smFRET) data discussed below.[80] *In silico* fluctuations of BF in the absence of substrate quantify the "open" to "ajar" transition as requiring less than 20 ns. This BF simulation suggested that the "ajar" intermediate acts as an energetic barrier to opening, and promotes correct dNTP incorporation.[77] A recent crystal structure of an "open" ternary BF complex allowed simulation of the events in active site assembly, which appears to include desolvation of the dNTP[78] and supports kinetic data referenced previously.[35] Following incorporation, the polymerase translocates along DNA through a

sequence of conformational changes and requires gating by the O1 helix. Pyrophosphate release then facilitates the fingers-opening transition.[79] Overall, MD simulations can uncover significant aspects of enzymatic catalysis not observable by conventional experimental techniques.

Different types of isotopic labeling of enzymes can assess structural changes within the enzyme, and can also identify amino acid residues involved in catalysis. Isotopic labeling is a valuable technique that results in minimal perturbation of enzyme activity and can be used in a range of applications. Evaluation of an isotopically labeled polymerase with nuclear magnetic resonance (NMR) spectroscopy is advantageous compared to crystallography in that it can be performed under near physiological conditions. In an NMR study of KlenTaq, matched and mismatched enzyme-P/T-ddNTP ternary complexes were compared, and revealed distinct conformations. Binary enzyme-P/T "open" complexes were close in resemblance to both mismatch ternary complexes and ternary complexes formed during nucleotide recognition opposite abasic DNA sites.[81] A completely different technique that uses isotopic labeling, kinetic isotope effect experiments, identified that certain active site residues participate in a general acid catalysis mechanism of nucleotide transfer with active site metal ions.[82] Single-molecule kinetic isotope experiments can identify key functional groups necessary to catalysis, and could be extended to DNA polymerases.[83] Conclusions drawn from various isotopic labeling experiments can guide alternative techniques, which more closely examine base discrimination.

1.6 Observing molecular dynamics with fluorescence-based experiments

Fluorescent techniques emerged as powerful tools to understand the biophysical characteristics of enzyme dynamics around the year 2000. Specifically, previously inaccessible transient intermediates preceding phosphoryl transfer in the polymerase reaction pathway could be observed in real-time. Results for fluorescent assays provide a clearer picture of DNA polymerase conformational motion with respect to itself or to its substrate. This information is obtained through results with a single, conformationally sensitive fluorophore, or a pair of donor/acceptor fluorophores that monitor changes in distance through Förster Resonance Energy Transfer (FRET). Once the interaction to be explored has been defined, crystallographic information can suggest the placement or choice of fluorophores.

One type of assay measures changes in fluorescence for inherently fluorescent or fluorescent dye-labeled incoming nucleotides. Fluorescent cytosine analogs, for example, were used to demonstrate active site flexibility through measurement of natural and unnatural dNTP dissociation constants. "Closed" state formation caused an observed increase in fluorescence anisotropy.[84] However, environmental sensitivity of these analogs restricts their usefulness in valuable single-molecule studies, which will be described in the next section. A recently developed method used *B. stearothermophilus* polymerase and coumarin-labeled nucleotides to rapidly measure incorporation kinetics in an automated system.[85] Although this system has the potential to improve current sequencing technologies, which depend on fluorescently-labeled nucleotides, conformational transitions cannot be directly observed.

A second type of single fluorophore assay monitors conformational transitions of polymerases using DNA substrates with the fluorescent base analog 2-aminopurine (2-AP) in the template.[86] Stacking interactions with neighboring bases quench 2-AP fluorescence.[87] Therefore, an increase in basal fluorescence likely corresponds to a "closed" complex, determined by X-ray crystallography, in which the 2-AP is flipped out from the P/T helix (Figure 1-6).[15,67] A subsequent decrease in fluorescence then corresponds to translocation following nucleotide incorporation, which again alters the 2-AP environment.

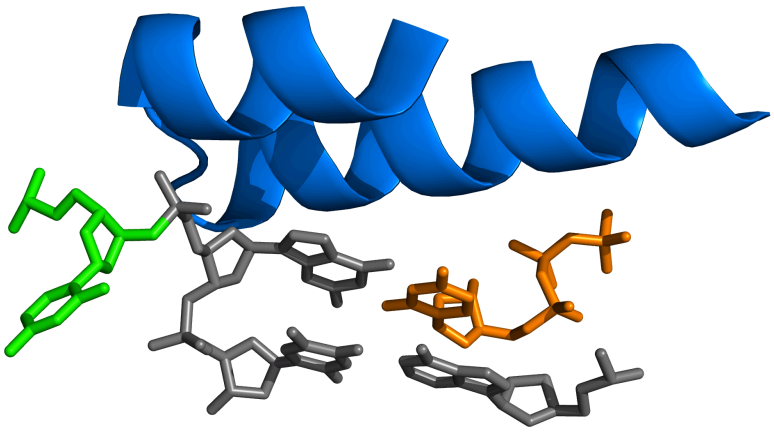


Figure 1-6. Source of changes in 2-aminopurine (2-AP) fluorescence. Fluorescence of 2-AP is quenched by stacking interactions with neighboring bases. A ternary complex of *Bacillus stearothermophilus* DNA polymerase large fragment (blue, PDB ID: 1LV5) with an incoming dNTP (orange) and primer-template DNA (gray) show that the next template base (green) is flipped out from the primer-template. When 2-AP is placed at this template position in single-probe fluorescence experiments, a large increase in fluorescence is observed upon dNTP binding. The flipping out of 2-AP away from any neighboring bases is the likely source of this increase in fluorescence and provides a measure of the fingers-closing transition.

The 2-AP probe is useful in that it does not disrupt DNA structure and is very sensitive to its surrounding environment. Experiments with T4 polymerase and P/T DNA containing 2-AP demonstrated high selectivity in the nucleotide-binding step as the result of a pocket induced by P/T binding; the subsequent fingers-closing transition provided little discrimination, supporting kinetic studies described earlier.[25,88] Different observed changes in fluorescence during misincorporation by KF as compared to correct incorporation indicated that the conformations for each respective case are distinct from one another.[89] A later study supported this conclusion, and completely different behavior was observed for mispaired dNTPs versus complementary dNTPs.[90] This result might represent the BF intermediate structures previously observed with mismatches.[70] The variant E710A, however, recognized properly paired bases as mismatches and exhibited similar fluorescent behavior.[90] While powerful, experiments with 2-AP do not always clarify the structural changes that cause the changes in fluorescence.

Another example of a single-probe fluorescence approach is a protein-fluorophore conjugate that, like 2-AP, responds to changes in its immediate surrounding environment. This type of assay has been used mainly in studies of the nucleotide-induced fingers-closing transition of T7 polymerase. Single cysteine variants are required for site-specific incorporation of the fluorophore onto the protein. In both cases presented here, the fluorescent label 7-diethylamino-3-((((2-maleimidyl)-ethyl)amino)carbonyl) coumarin (MDCC) was conjugated to an engineered cysteine at residue 514 on the outside surface of the mobile fingers subdomain (Figure 1-7). One study determined that mutation of a glycine-alanine-glycine (GAG) hinge flanking the O1 helix limited the flexibility of the enzyme required to recognize and incorporate correct nucleotides.[91]

Another study using a single-fluorophore probe suggested a "new paradigm" in polymerase specificity that may extend to other enzymes in which the incorrect substrate causes a misalignment of catalytic residues. Specifically, the authors determined that a mismatch induces a "fully open" state due to misaligned residues and does not lie between the "open" and "closed" states.[92] The results contrast with known BF structural data that places a mismatched dNTP-enzyme-P/T ternary complex within the pathway from "open" to "closed,"[70] and, therefore, these states require structural determination. Due to the sensitivity of the MDCC fluorophore, however, the different states observed may reflect subtle changes in positioning of residue 514.

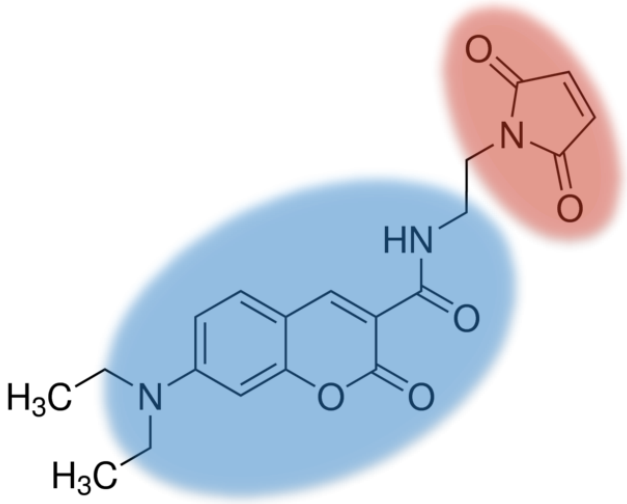


Figure 1-7. Chemical structure of the fluorescent probe 7-diethylamino-3-((((2-maleimidyl)-ethyl)amino)carbonyl) coumarin (MDCC). MDCC is a commonly used probe that is conjugated to DNA polymerases for use in fluorescence-based assays. Single-cysteine variants of DNA polymerases allow site-specific labeling through covalent bond formation with the MDCC maleimide functional group (red). MDCC is fluorescent due to its highly conjugated π-system (blue). Conformational transitions of polymerases cause changes in the local environment and fluorescence of MDCC.

FRET-based fluorescence assays permit distance measurements by specific placement of two fluorophores that undergo energy transfer only when in close proximity to one another. After crystallographic discovery of the necessary fingers-closing transition, several studies investigated whether this conformational change was the rate-limiting step preceding phosphoryl transfer. Placement of one fluorophore on the mobile fingers-subdomain and the other on the P/T DNA enabled these investigations; fluorescent base analogs also demonstrated potential as acceptor fluorophores on the P/T.[93] One system that monitored the fingers-closing transition in KlenTaq revealed that this structural change was faster than the rate-limiting step.[94] This observation was supported by experiments with KF, which also detected a possible base discrimination step prior to fingers-closing.[95] It was suggested that the rate-limiting step is instead a re-organization of the active site following fingers-closing, and is involved in metal binding.[95,96] Follow-up experiments with the KlenTaq system found that the step before dNTP binding differed for all four nucleotides,[97] suggesting an applied method of base discrimination that depends on the templating base. Experiments with fluorophores placed on various domains of full-length Taq polymerase uncovered additional conformational changes, which extend beyond the fingers domain.[98]

Attachment of two fluorophores to two different sites within the same protein can monitor structural changes in fingers-opening, the enzyme-dNTP binary complex, or the apo enzyme.[99] However, these transient states are dynamic. Ensemble measurements are only useful for the fingers-closing transition, after which there is a conformational convergence of individual molecules. Therefore, the intramolecular FRET system is advantageous at the single-molecule level when compared to intermolecular FRET.

1.7 Clearer pictures of conformational dynamics: Single-molecule techniques

As illustrated thus far, enzyme dynamics play an important role in substrate selection and catalysis by enzymes. Conformational fluctuations occur over a broad range of time scales in equilibrium and cannot be detected through ensemble-averaged experiments.[100,101] Methods to measure enzyme activity at the single-molecule level provide previously unresolved details about the dynamics that contribute to enzymatic reactivity.[102–106] Single-molecule biophysical methods to measure DNA polymerase reactions are especially powerful for answering unresolved questions about nucleotide selectivity and DNA replication fidelity.[16,107]

Single-molecule FRET (smFRET) provides measurements of conformational motions along reaction pathways at millisecond timescales.[108–110] When applied to KF, the apo enzyme or KF-dNTP complex fluctuated rapidly on this timescale, and indicated sampling of the "open" and "closed" states as a manner of intrinsic conformational flexibility.[111–113] In the case of the dynamic enzyme-P/T binary complex, the rate of fingers-closing was slower than the rate of fingers-opening. However, upon incubation with increasing complementary dNTP concentrations, the rate of fingers-closing became faster than fingers-opening and supported an induced fit model of recognition.[113] Unique intermediate FRET signatures, suggested to be fidelity checkpoints, occurred between the "open" and "closed" states in the presence of mismatched dNTPs.[80,111,113,114] Importantly, these intermediate FRET signatures could not be observed in ensemble measurements.[90] The results, also observed in KlenTaq experiments,[115] suggested that such signatures were representative of the "ajar" crystal structures previously observed for BF.[70] Complementary nucleotides also produced an intermediate FRET signature that

is short-lived and barely detectable, indicating the transient presence of this state along the catalytically relevant reaction pathway.[114] Furthermore, observation of this intermediate step, previously unobserved for correct nucleotides, underscores the importance of using this sensitive single-molecule technique.

Lifetime-resolved smFRET measurements can indicate local changes in the enzyme environment of intermediate or final reaction products using a double-substituted fluorogenic substrate.[116,117] While this technique would certainly be appreciated in the investigation of phosphoryl transfer and pyrophosphate release by DNA polymerase, careful design of substrates will be required for successful catalysis. SmFRET systems, which measure dynamics in nucleotide-induced polymerase domain distance fluctuations, are therefore most relevant to applications for base discrimination.

Experiments using single-molecule fluorescence to observe polymerase motions also expand beyond the normal nucleotide incorporation step and generate different insights into how the enzyme achieves fidelity. Fluorescence measurements with an incorrectly paired nucleotide resulted in intramolecular partitioning to the exonuclease domain,[114,118] and appeared faster than intermolecular transfer.[118] A separate step caused a drop in the FRET ratio after nucleotide incorporation and indicated yet another previously unobserved motion that may be important in proofreading.[119] This separate step may contribute to the slower rate of fingers-opening after correct nucleotide incorporation as compared to the dNTP-induced fingers-closing rate.[113] Although data from smFRET about steps after correctly matched nucleotide incorporation is extremely valuable, such data is not discriminatory for the incoming base. Additionally, limited

time resolution with smFRET can conceal rapid subdomain dynamics that are important to fully understand the mechanism of nucleotide incorporation.

A more sensitive single-molecule technique known as optical tweezers uses a highly focused laser beam to manipulate and study biomolecules attached to beads.[120] By measuring mechanical force exertion during the rate-limiting step for T7 polymerase, this technique enabled determination of two template bases inside the polymerase's active site during each round of catalysis.[121] Optical tweezers were also used to study translocation in Phi (ϕ) 29 DNA polymerase, a B-family polymerase that is structurally and functionally similar to KF.[122] Experiments with ϕ29 polymerase were designed to determine how mechanical motion is coupled to catalysis,[123] and could provide insight into A-family polymerase mechanics. By combining optical tweezers with glass nanocapillaries, a P/T-bound RNA polymerase molecule was detected and demonstrated the potential of this new tool to examine DNA polymerases under physiological conditions. Compared to nanopores, nanocapillaries are cheaper and more easily produced but much less well-characterized with respect to polymerase-P/T complexes.

Nanopores, unlike optical tweezers, do not require surface immobilization of substrate molecules; instead, these devices are comprised of a thin membrane containing a biological or solid-state nanopore through which captured ssDNA can affect changes in ionic current.[124,125] Recent nanopore systems were developed to measure motion of motor proteins with sub-angstrom resolution. This technique is approximately 7-8 times more sensitive in spatial resolution than FRET or optical tweezers.[126] A-family polymerases, including KF, T7, and Taq, have been captured atop a nanopore in an applied electric field to allow detection of polymerase-dependent DNA extension (Figure

1-8). The rate of polymerization by the DNA polymerase controls ssDNA translocation through the pore.[127–131] These nanopore-based experiments have recognized enzyme-P/T binary complex and enzyme-P/T-dNTP ternary complex discrimination,[127] incorrect dNTP binding,[128] and pre-equilibrium kinetics.[132] Perhaps one of the most impressive studies using nanopores measured distinct rNTP- versus dNTP-bound ternary complexes, which demonstrated distinct intermediate complexes with increasing stability. This study established the first step of complementary base recognition, followed by a deoxyribose recognition state, then fingers-closing, and finally an active site rearrangement.[133] Much focus, however, has been toward developing this technique for DNA sequencing following the achievement of single-base resolution.[134]

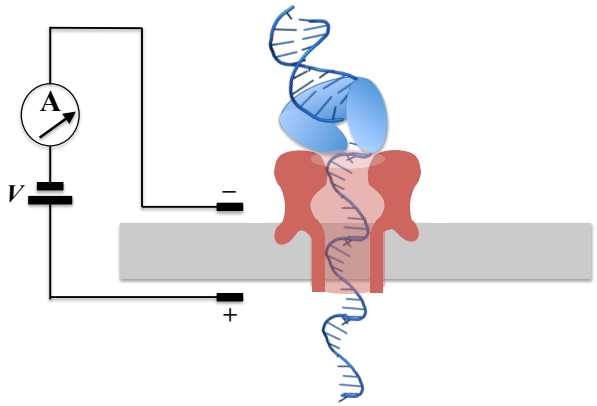


Figure 1-8. Schematic diagram of a nanopore device with a captured DNA polymerase. Protein nanopores (red) can be inserted into an insulating lipid-bilayer or solid-state membrane that separates two compartments containing buffer solutions. An amplifier applies an electric field and measures ionic current across the membrane. A DNA polymerase (blue) captured atop the nanopore controls DNA translocation across the membrane as the enzyme polymerizes the DNA. Various complexes of the DNA polymerase and its substrates can be monitored via ionic changes in current. Adapted from Lieberman *et al*.[151]

A final example of a real-time, single-molecule method developed to understand polymerase conformational motion is an electronic biosensor. Biosensors, which use immobilized enzymes, may be better suited to resolve steps on the polymerase reaction pathway than methods studying enzymes in solution; biosensors do not require labels and are sensitive enough to measure various interactions of the protein with different concentrations of substrates. Recently, the dNTP-induced fingers-closing transition was observed with a commercial biosensor that also revealed previously unidentified tight binding states for Taq and KF.[135]

Another biosensing device was developed in our lab and has enabled detection of conformational transitions during catalysis by several enzymes,[136,137] including KF (Figure 1-9).[75] This technique relies on voltage-gating of single-walled carbon nanotube field-effect transistors (SWCNT-FETs) by charged amino acids on the enzyme's surface that move within 1 nm of the nanotube during catalysis.[138] Although the exact nature of the observed conformation for KF is unclear, distinct states have been observed for dNTP analogs and will be further discussed in Chapter 2.[36] Despite a recent report that touted the short probe-able distance from a sensor as a limitation,[139] this distance-limited sensitivity can in fact be an advantage to reduce noise from extraneous protein motion. The location of attachment is significant for the thermal stability of immobilized enzymes[140] and rapid, stable immobilization in a reproducible, specific orientation is required for consistent, sensitive detection.[141] Biosensors are a promising technique for the dual capability of conformational analysis and base discrimination, but require additional development to be commercially viable.

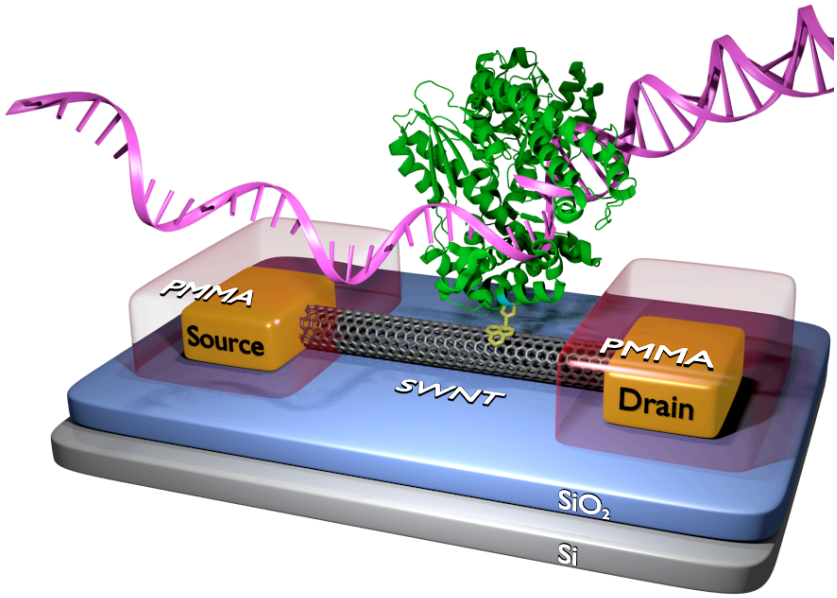


Figure 1-9. Schematic diagram of a single-walled carbon nanotube field effect transistor (SWCNT-FET) non-covalently bioconjugated to a single molecule of DNA polymerase I (KF). KF is attached to a SWCNT-FET through a single cysteine introduced in the "fingers" subdomain. A pyrene-maleimide linker (yellow) adheres to the SWCNT-FET through π-π stacking and covalently attaches to the single cysteine to immobilize the KF. The SWCNT-FET is grown on SiO_2, connects to source and drain metal electrodes, and is passivated with a polymer (PMMA, red).

1.8 Single-molecule sequencing, additional applications, and necessary improvements

High-performing DNA sequencing methods require the rapid discrimination of individual bases along a translocating strand of DNA. In addition to speed and cost-effectiveness, successful sequencing techniques are dependent upon sample preparation, time resolution, accuracy, run length, throughput, and data analysis. Single-molecule DNA sequencing technologies have emerged as formidable tools for sensitive and distinct detection of specific DNA bases, saving time and resources. Furthermore, direct observation of polymerase motion for DNA sequencing takes advantage of the enzyme's inherent processivity and fidelity. By enabling single-molecule methodology that is driven by a DNA polymerase, fundamental enzyme characterization and impactful biotechnological applications can be simultaneously realized by a single technique.

Single-molecule real-time sequencing (SMRT), developed by Pacific Biosciences,[142] enables exceptional processivity, speed, and accuracy with the use of ϕ29 polymerase.[143] In order to drive real-time sequencing with a DNA polymerase, SMRT requires reliable immobilization and successful incorporation of labeled substrates while preserving polymerase activity. Additionally, sophisticated instrumentation permits single-molecule detection within a nanophotonic structure. SMRT has also been useful beyond traditional sequencing in the identification of epigenetic methylation patterns.[144,145] However, one study that compared SMRT to an earlier generation sequencing platform found higher error rates with SMRT[18] and signified a need for improved accuracy. Additionally, SMRT detects distinct polymerization states and

produces kinetic information about each individual base incorporation, but does not directly observe polymerase motion.

Another example of a single-molecule DNA sequencing technique uses atomic force microscopy (AFM) to measure polymerase mechanics. In this technique, a DNA polymerase is conjugated to an AFM tip, captures a primer-template DNA, and is then dipped into a pool of one of the four dNTPs. Correct dNTP binding induces a distinct enzyme-P/T-dNTP ternary complex via its fit into the polymerase active site and causes an observable rupture of the AFM tip.[146] The incorrect dNTP does not produce any signal. Advantages of this system include the retained activity of the DNA polymerase when immobilized to the AFM tip and the fast rate required for rapid detection. However, rupture events for each base are not differentiated and do not provide any useful information about how the enzyme specifically recognizes one base over the other.

Nanopore sequencing is receiving much attention as a valuable tool for DNA sequencing, especially after DNA polymerases were employed to control DNA translocation speed through the pore. Phi 29 polymerase has been particularly useful in overcoming the challenges of using nanopores for sequencing through its lower turnover rate and enhanced polymerase complex stability.[147] Further improvements to this sequencing methodology include the use of an alternative biological nanopore to enable nucleotide-specific current resolution.[148,149] Nanopore devices based on ϕ 29 have also facilitated understanding of deoxyribose discrimination,[150] processive synthesis,[151] dynamics of translocation,[152,153] and details about partitioning to the exonuclease domain.[154,155] Additionally, like SMRT, nanopores have recently enabled detection of epigenetic methylation patterns.[156] In order for nanopores to be a feasible option for

DNA sequencing, algorithms have been designed to translate electronic data into readable sequences.[157,158]

Several recent developments hold great promise for the future improvement of nanopores and other electronic techniques toward a commercially viable DNA sequencing technique. To be successful, these techniques require high-fidelity proteins, high throughput, significant signal to noise ratios, reliable base discrimination, and ease of fabrication. In regard to qualifying A-family polymerases for use in nanopores, unnatural amino acid replacement of 32 proline residues within KlenTaq generates a highly active protein.[159] To overcome the challenge of inaccurate replication of repetitive sequences, a recently engineered T4 polymerase with enhanced and accurate processivity could be used.[160] To parallelize nanopore sequencing and increase throughput, arrays of nanopores can be generated.[161] To overcome low signal to noise ratios, a new biological nanopore has been implemented in protein-free devices and may extend to devices with enzyme-P/T complexes.[162] Another strategy improved base discrimination with truncated pores for exosequencing and has the potential to be applied to polymerase-based nanopore devices.[163] Solid-state nanopores based on silicon are easy to fabricate and may also extend to polymerase-complexed devices.[164]

Two strategies could be adapted to observe polymerases within a nanopore and couple sequencing capability with a tool that allows further understanding of enzyme catalysis. In one strategy, nanopores with internally bound proteins[165] could evaluate a single enzyme over long time scales by measuring small changes in the protein's conformation. In a second strategy, translocation of a protein-P/T complex through a nanopore could sequentially evaluate various single-molecules and permit determination

of an activity distribution from single-molecule populations.[166,167] The challenges that may arise from protein detection within nanopores could be alleviated through the use of nanochannels. Specifically, conical nanochannels, which can be reproducibly fabricated with a sensitive tip region that responds to the analytes, provide a superior signal to noise ratio and time resolution for transport of proteins.[168]

1.9 Conclusions

Many years of research have been devoted to the complete characterization of the DNA polymerase catalytic cycle to further understand enzyme specificity. Important conformational transitions contribute to the fidelity of these enzymes, and tools to study such structural changes have been subject to extensive technical development. Single-molecule biophysical techniques in particular provide a clearer picture of the precise nature of nucleotide selection. The development of more sophisticated experimental techniques have led to the use of DNA polymerases to drive polymerization for sequencing applications. Improvements to these techniques are required to further enhance the understanding of fundamental catalysis while concurrently making these tools suitable for inexpensive, commercially practical DNA sequencing. The sensitivity of electronic devices, such as nanopores and biosensors, illuminates a promising strategy to satisfy this dual capability.

1.10 References

1. Steitz, T. A. DNA- and RNA-dependent DNA polymerases. *Curr. Opin. Struct. Biol.* **3,** 31–38 (1993).

2. Echols, H. & Goodman, M. F. Fidelity mechanisms in DNA replication. *Annu. Rev. Biochem.* **60,** 477–511 (1991).

3. Arnold, E., Ding, J., Hughes, S. H. & Hostomsky, Z. Structures of DNA and RNA polymerases and their interactions with nucleic acid substrates. *Curr. Opin. Struct. Biol.* **5,** 27–38 (1995).

4. Frauenfelder, H., Sligar, S. & Wolynes, P. The energy landscapes and motions of proteins. *Science* **254,** 1598–1603 (1991).

5. Steitz, T. A. DNA polymerases: Structural diversity and common mechanisms. *J. Biol. Chem.* **274,** 17395–17398 (1999).

6. Kool, E. T. Active site tightness and substrate fit in DNA replication. *Annu. Rev. Biochem.* **71,** 191–219 (2002).

7. Kohlstaedt, L., Wang, J., Friedman, J., Rice, P. & Steitz, T. Crystal structure at 3.5 A resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* **256,** 1783–1790 (1992).

8. Beard, W. A. & Wilson, S. H. Structural insights into the origins of DNA polymerase fidelity. *Structure* **11,** 489–496 (2003).

9. Kunkel, T. A. DNA replication fidelity. *J. Biol. Chem.* **279,** 16895–8 (2004).

10. Joyce, C. M. & Benkovic, S. J. DNA polymerase fidelity: kinetics, structure, and checkpoints. *Biochemistry* **43,** 14317–24 (2004).

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

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

13. Johnson, K. A. The kinetic and chemical mechanism of high-fidelity DNA polymerases. *Biochim. Biophys. Acta* **1804,** 1041–8 (2010).

14. Wu, S., Beard, W. A., Pedersen, L. G. & Wilson, S. H. Structural comparison of DNA polymerase architecture suggests a nucleotide gateway to the polymerase active site. *Chem. Rev.* **114,** 2759–74 (2014).

15. Joyce, C. M. Techniques used to study the DNA polymerase reaction pathway. *Biochim. Biophys. Acta* **1804,** 1032–40 (2010).

16. Gill, J. P., Wang, J. & Millar, D. P. DNA polymerase activity at the single-molecule level. *Biochem. Soc. Trans.* **39,** 595–9 (2011).

17. Morey, M. *et al.* A glimpse into past, present, and future DNA sequencing. *Mol. Genet. Metab.* **110,** 3–24 (2013).

18. Mosher, J. J., Bernberg, E. L., Shevchenko, O., Kan, J. & Kaplan, L. A. Efficacy of a 3rd generation high-throughput sequencing platform for analyses of 16S rRNA genes from environmental samples. *J. Microbiol. Methods* **95,** 175–81 (2013).

19. Klenow, H. & Henningsen, I. Selective elimination of the exonuclease activity of the deoxyribonucleic acid polymerase from *Escherichia coli* B by limited proteolysis. *Proc. Natl. Acad. Sci.* **65,** 168–175 (1970).

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

21. Mizrahi, V., Henrie, R. N., Marlier, J. F., Johnson, K. A. & Benkovic, S. J. Rate-limiting steps in the DNA polymerase I reaction pathway. *Biochemistry* **24,** 4010–4018 (1985).

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

23. 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,** 4835–4843 (1991).

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

25. Eger, B. T. & Benkovic, S. J. The minimal kinetic mechanism for misincorporation by DNA polymerase I (Klenow fragment). *Biochemistry* **31,** 9227–9236 (1992).

26. Mizrahi, V., Benkovic, P. A. & Benkovic, S. J. Mechanism of the idling-turnover reaction of the large (Klenow) fragment of *Escherichia coli* DNA polymerase I. *Proc. Natl. Acad. Sci. U. S. A.* **83,** 231–5 (1986).

27. 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,** 511–525 (1991).

28. 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,** 526–537 (1991).

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

30. 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,**

3215–3221 (2001).

31. Ramanathan, S., Chary, K. V & Rao, B. J. Incoming nucleotide binds to Klenow ternary complex leading to stable physical sequestration of preceding dNTP on DNA. *Nucleic Acids Res.* **29,** 2097–105 (2001).

32. Shinkai, A. & Loeb, L. A. In vivo mutagenesis by *Escherichia coli* DNA polymerase I. Ile(709) in motif A functions in base selection. *J. Biol. Chem.* **276,** 46759–64 (2001).

33. Shinkai, A., Patel, P. H. & Loeb, L. A. The conserved active site motif A of *Escherichia coli* DNA polymerase I is highly mutable. *J. Biol. Chem.* **276,** 18836–42 (2001).

34. Barsky, D., Kool, E. T. & Colvin, M. E. Interaction and solvation energies of nonpolar DNA base analogues and their role in polymerase insertion fidelity. *J. Biomol. Struct. Dyn.* **16,** 1119–34 (1999).

35. Zhang, X., Motea, E., Lee, I. & Berdis, A. J. Replication of a universal nucleobase provides unique insight into the role of entropy during DNA polymerization and pyrophosphorolysis. *Biochemistry* **49,** 3009–23 (2010).

36. Pugliese, K. M. & Gul, O. T. *et al.* Processive incorporation of deoxynucleoside triphosphate analogs by single-molecule DNA polymerase I (Klenow fragment) nanocircuits. *J. Am. Chem. Soc.* **137,** 9587–94 (2015).

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

38. Dzantiev, L. & Romano, L. J. Differential effects of *N*-acetyl-2-aminofluorene and *N*-2-aminofluorene adducts on the conformational change in the structure of DNA polymerase I (Klenow fragment). *Biochemistry* **39,** 5139–5145 (2000).

39. Alekseyev, Y. O., Dzantiev, L. & Romano, L. J. Effects of benzo[a]pyrene DNA adducts on *Escherichia coli* DNA polymerase I (Klenow fragment) primer−template interactions: Evidence for inhibition of the catalytically active ternary complex formation. *Biochemistry* **40,** 2282–2290 (2001).

40. Berdis, A. J. Dynamics of translesion DNA synthesis catalyzed by the bacteriophage T4 exonuclease-deficient DNA polymerase. *Biochemistry* **40,** 7180–7191 (2001).

41. Devadoss, B., Lee, I. & Berdis, A. J. Enhancing the 'A-rule' of translesion DNA synthesis: promutagenic DNA synthesis using modified nucleoside triphosphates. *Biochemistry* **46,** 13752–61 (2007).

42. Polesky, A. H., Dahlberg, M. E., Benkovic, S. J., Grindley, N. D. & Joyce, C. M. Side chains involved in catalysis of the polymerase reaction of DNA polymerase I from *Escherichia coli*. *J. Biol. Chem.* **267,** 8417–8428 (1992).

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

44. Kaushik, N., Pandey, V. N. & Modak, M. J. Significance of the O-helix residues of *Escherichia coli* DNA polymerase I in DNA synthesis: dynamics of the dNTP binding pocket. *Biochemistry* **35,** 7256–66 (1996).

45. Donlin, M. J. & Johnson, K. A. Mutants affecting nucleotide recognition by T7 DNA polymerase. *Biochemistry* **33,** 14908–14917 (1994).

46. 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).

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

48. Brown, W., Stump, K. & Kelley, W. *Escherichia coli* DNA polymerase I. Sequence characterization and secondary structure prediction. *J. Biol. Chem.* **257,** 1965–1972 (1982).

49. Klenow, H., Overgaard-Hansen, K. & Patkar, S. A. Proteolytic cleavage of native DNA polymerase into two different catalytic fragments. Influence of assay conditions on the change of exonuclease activity and polymerase activity accompanying cleavage. *Eur. J. Biochem.* **22,** 371–381 (1971).

50. Jacobsen, H., Klenow, H. & Overoaard-Hansen, K. The N-terminal amino-acid sequences of DNA polymerase I from *Escherichia coli* and of the large and the small fragments obtained by a limited proteolysis. *Eur. J. Biochem.* **45,** 623–627 (1974).

51. Que, B. G., Downey, K. M. & So, A. G. Mechanism of selective inhibition of 3' to 5' exonuclease activity of *Escherichia coli* DNA polymerase I by nucleoside 5'-monophosphates. *Biochemistry* **17,** 1603–1606 (1978).

52. Setlow, P. & Kornberg, A. Deoxyribonucleic acid polymerase: Two distinct enzymes in one polypeptide. II. A proteolytic fragment containing the 5'-3' exonuclease function. Restoration of intact enzyme fuctions from the two proteolytic fragments. *J. Biol. Chem.* **247,** 232–240 (1972).

53. 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,** 762–766 (1985).

54. Derbyshire, V. *et al.* Genetic and crystallographic studies of the 3',5'-exonucleolytic site of DNA polymerase I. *Science* **240,** 199–201 (1988).

55. 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.* **85,** 8924–8928 (1988).

56. 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).

57. Beese, L., Derbyshire, V. & Steitz, T. Structure of DNA polymerase I Klenow fragment bound to duplex DNA. *Science* **260,** 352–355 (1993).

58. Joyce, C. M. & Steitz, T. A. DNA polymerase I: from crystal structure to function via genetics. *Trends Biochem. Sci.* **12,** 288–292 (1987).

59. 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,** 14095–14101 (1993).

60. Tuske, S., Singh, K., Kaushik, N. & Modak, M. J. The J-helix of *Escherichia coli* DNA polymerase I (Klenow fragment) regulates polymerase and 3'- 5'-exonuclease functions. *J. Biol. Chem.* **275,** 23759–68 (2000).

61. Doublié, 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).

62. 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).

63. Eom, S. H., Wang, J. & Steitz, T. A. Structure of *Taq* polymerase with DNA at the polymerase active site. *Nature* **382,** 278–81 (1996).

64. Kiefer, J. R. *et al.* Crystal structure of a thermostable *Bacillus* DNA polymerase l large fragment at 2.1 Å resolution. *Structure* **5,** 95–108 (1997).

65. 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).

66. 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,** 7514–7525 (1998).

67. 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).

68. Li, Y. & Waksman, G. Crystal structures of a ddATP-, ddTTP-, ddCTP, and ddGTP- trapped ternary complex of Klentaq1: insights into nucleotide incorporation and selectivity. *Protein Sci.* **10,** 1225–33 (2001).

69. Wang, W., Wu, E. Y., Hellinga, H. W. & Beese, L. S. Structural factors that determine selectivity of a high fidelity DNA polymerase for deoxy-, dideoxy-, and ribonucleotides. *J. Biol. Chem.* **287,** 28215–26 (2012).

70. Wu, E. Y. & Beese, L. S. The structure of a high fidelity DNA polymerase bound to a mismatched nucleotide reveals an 'ajar' intermediate conformation in the nucleotide selection mechanism. *J. Biol. Chem.* **286,** 19758–67 (2011).

71. 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).

72. Obeid, S., Baccaro, A., Welte, W., Diederichs, K. & Marx, A. Structural basis for the synthesis of nucleobase modified DNA by *Thermus aquaticus* DNA polymerase. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 21327–31 (2010).

73. Betz, K. & Malyshev, D. *et al.* KlenTaq polymerase replicates unnatural base pairs by inducing a Watson-Crick geometry. *Nat. Chem. Biol.* **8,** 612–4 (2012).

74. Betz, K. & Malyshev, D. *et al.* Structural insights into DNA replication without hydrogen bonds. *J. Am. Chem. Soc.* **135,** 18637–43 (2013).

75. Olsen, T. J. & Choi, Y. *et al.* Electronic measurements of single-molecule processing by DNA polymerase I (Klenow fragment). *J. Am. Chem. Soc.* **135,** 7855–60 (2013).

76. 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).

77. Miller, B. R., Parish, C. A. & Wu, E. Y. Molecular dynamics study of the opening mechanism for DNA polymerase I. *PLoS Comput. Biol.* **10,** e1003961 (2014).

78. Miller, B. R., Beese, L. S., Parish, C. A. & Wu, E. Y. The closing mechanism of DNA polymerase I at atomic resolution. *Structure* **23,** 1609–20 (2015).

79. Golosov, A. A., Warren, J. J., Beese, L. S. & Karplus, M. The mechanism of the translocation step in DNA replication by DNA polymerase I: a computer simulation analysis. *Structure* **18,** 83–93 (2010).

80. Hohlbein, J. *et al.* Conformational landscapes of DNA polymerase I and mutator derivatives establish fidelity checkpoints for nucleotide insertion. *Nat. Commun.* **4,** 2131 (2013).

81. Holzberger, B., Pszolla, M. G., Marx, A. & Möller, H. M. KlenTaq DNA polymerase adopts unique recognition states when encountering matched, mismatched, and abasic template sites: an NMR study. *Chembiochem* **13,** 635–9 (2012).

82. Castro, C. & Smidansky, E. D. *et al.* Nucleic acid polymerases use a general acid for nucleotidyl transfer. *Nat. Struct. Mol. Biol.* **16,** 212–8 (2009).

83. Pudney, C. R. *et al.* Enzymatic single-molecule kinetic isotope effects. *J. Am. Chem. Soc.* **135,** 3855–64 (2013).

84. Sandin, P. *et al.* Highly efficient incorporation of the fluorescent nucleotide analogs tC and tCO by Klenow fragment. *Nucleic Acids Res.* **37,** 3924–33 (2009).

85. Walsh, M. T., Roller, E. E., Ko, K.-S. & Huang, X. Measurement of DNA polymerase incorporation kinetics of dye-labeled nucleotides using total internal reflection fluorescence microscopy. *Biochemistry* **54,** 4019–21 (2015).

86. 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,** 9185–9192 (1995).

87. Rachofsky, E. L., Osman, R. & Ross, J. B. A. Probing structure and dynamics of DNA with 2-aminopurine: Effects of local environment on fluorescence. *Biochemistry* **40,** 946–956 (2001).

88. Fidalgo da Silva, E., Mandal, S. S. & Reha-Krantz, L. J. Using 2-aminopurine fluorescence to measure incorporation of incorrect nucleotides by wild type and mutant bacteriophage T4 DNA polymerases. *J. Biol. Chem.* **277,** 40640–9 (2002).

89. Purohit, V., Grindley, N. D. F. & 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).

90. Bermek, O., Grindley, N. D. F. & Joyce, C. M. Prechemistry nucleotide selection checkpoints in the reaction pathway of DNA polymerase I and roles of glu710 and tyr766. *Biochemistry* **52,** 6258–74 (2013).

91. Jin, Z. & Johnson, K. A. Role of a GAG hinge in the nucleotide-induced

conformational change governing nucleotide specificity by T7 DNA polymerase. *J. Biol. Chem.* **286,** 1312–22 (2011).

92. Tsai, Y.-C. & Johnson, K. A. A new paradigm for DNA polymerase specificity. *Biochemistry* **45,** 9675–87 (2006).

93. Stengel, G. *et al.* Conformational dynamics of DNA polymerase probed with a novel fluorescent DNA base analogue. *Biochemistry* **46,** 12289–97 (2007).

94. 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).

95. Joyce, C. M. *et al.* Fingers-closing and other rapid conformational changes in DNA polymerase I (Klenow fragment) and their role in nucleotide selectivity. *Biochemistry* **47,** 6103–16 (2008).

96. Bermek, O., Grindley, N. D. F. & Joyce, C. M. Distinct roles of the active-site Mg^{2+} ligands, Asp882 and Asp705, of DNA polymerase I (Klenow fragment) during the prechemistry conformational transitions. *J. Biol. Chem.* **286,** 3755–66 (2011).

97. Rothwell, P. J. & Waksman, G. A pre-equilibrium before nucleotide binding limits fingers subdomain closure by Klentaq1. *J. Biol. Chem.* **282,** 28884–92 (2007).

98. Xu, C., Maxwell, B. A. & Suo, Z. Conformational dynamics of *Thermus aquaticus* DNA polymerase I during catalysis. *J. Mol. Biol.* **426,** 2901–17 (2014).

99. Allen, W. J., Rothwell, P. J. & Waksman, G. An intramolecular FRET system monitors fingers subdomain opening in Klentaq1. *Protein Sci.* **17,** 401–408 (2008).

100. Yang, H. *et al.* Protein conformational dynamics probed by single-molecule electron transfer. *Science* **302,** 262–6 (2003).

101. Min, W. *et al.* Fluctuating enzymes: lessons from single-molecule studies. *Acc. Chem. Res.* **38,** 923–31 (2005).

102. Lu, H. P., Xun, L. & Xie, X. S. Single-molecule enzymatic dynamics. *Science* **282,** 1877–1882 (1998).

103. Xie, X. S. & Lu, H. P. Single-molecule enzymology. *J. Biol. Chem.* **274,** 15967–15970 (1999).

104. Xie, S. Single-molecule approach to enzymology. *Single Mol.* **2,** 229–236 (2001).

105. Smiley, R. D. & Hammes, G. G. Single molecule studies of enzyme mechanisms. *Chem. Rev.* **106,** 3080–94 (2006).

106. Lu, H. P. Sizing up single-molecule enzymatic conformational dynamics. *Chem. Soc. Rev.* **43,** 1118–43 (2014).

107. Deniz, A. A., Mukhopadhyay, S. & Lemke, E. A. Single-molecule biophysics: at the interface of biology, physics and chemistry. *J. R. Soc. Interface* **5,** 15–45 (2008).

108. Ha, T. & Ting, A. Y. *et al.* Single-molecule fluorescence spectroscopy of enzyme conformational dynamics and cleavage mechanism. *Proc. Natl. Acad. Sci.* **96,** 893–898 (1999).

109. Roy, R., Hohng, S. & Ha, T. A practical guide to single-molecule FRET. *Nat. Methods* **5,** 507–16 (2008).

110. Blank, K., De Cremer, G. & Hofkens, J. Fluorescence-based analysis of enzymes at the single-molecule level. *Biotechnol. J.* **4,** 465–79 (2009).

111. Santoso, Y. & Joyce, C. M. *et al.* Conformational transitions in DNA polymerase I revealed by single-molecule FRET. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 715–20 (2010).

112. Torella, J. P., Holden, S. J., Santoso, Y., Hohlbein, J. & Kapanidis, A. N. Identifying molecular dynamics in single-molecule FRET experiments with burst variance analysis. *Biophys. J.* **100,** 1568–77 (2011).

113. Evans, G. W., Hohlbein, J., Craggs, T., Aigrain, L. & Kapanidis, A. N. Real-time single-molecule studies of the motions of DNA polymerase fingers illuminate DNA synthesis mechanisms. *Nucleic Acids Res.* **43,** 5998–6008 (2015).

114. Berezhna, S. Y., Gill, J. P., Lamichhane, R. & Millar, D. P. Single-molecule Förster resonance energy transfer reveals an innate fidelity checkpoint in DNA polymerase I. *J. Am. Chem. Soc.* **134,** 11261–8 (2012).

115. Rothwell, P. J. & Allen, W. J. *et al.* dNTP-dependent conformational transitions in the fingers subdomain of Klentaq1 DNA polymerase: insights into the role of the 'nucleotide-binding' state. *J. Biol. Chem.* **288,** 13575–91 (2013).

116. Terentyeva, T. G., Hofkens, J., Komatsuzaki, T., Blank, K. & Li, C.-B. Time-resolved single molecule fluorescence spectroscopy of an α-chymotrypsin catalyzed reaction. *J. Phys. Chem. B* **117,** 1252–60 (2013).

117. Zheng, D. & Lu, H. P. Single-molecule enzymatic conformational dynamics: spilling out the product molecules. *J. Phys. Chem. B* **118,** 9128–40 (2014).

118. 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).

119. 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).

120. Jagannathan, B. & Marqusee, S. Protein folding and unfolding under force. *Biopolymers* **99,** 860–9 (2013).

121. Wuite, G. J., Smith, S. B., Young, M., Keller, D. & Bustamante, C. Single-molecule studies of the effect of template tension on T7 DNA polymerase activity. *Nature* **404,** 103–6 (2000).

122. Bernad, A., Blanco, L. & Salas, M. Site-directed mutagenesis of the YCDTDS amino acid motif of the Φ29 DNA polymerase. *Gene* **94,** 45–51 (1990).

123. Morin, J. A. *et al.* Mechano-chemical kinetics of DNA replication: Identification of the translocation step of a replicative DNA polymerase. *Nucleic Acids Res.* **43,** 3643–52 (2015).

124. Hornblower, B. *et al.* Single-molecule analysis of DNA-protein complexes using nanopores. *Nat. Methods* **4,** 315–317 (2007).

125. Deamer, D. Nanopore analysis of nucleic acids bound to exonucleases and polymerases. *Annu. Rev. Biophys.* **39,** 79–90 (2010).

126. Derrington, I. M. *et al.* Subangstrom single-molecule measurements of motor proteins using a nanopore. *Nat. Biotechnol.* **33,** 1073–5 (2015).

127. Benner, S. *et al.* Sequence-specific detection of individual DNA polymerase complexes in real time using a nanopore. *Nat. Nanotechnol.* **2,** 718–24 (2007).

128. Hurt, N., Wang, H., Akeson, M. & Lieberman, K. R. Specific nucleotide binding and rebinding to individual DNA polymerase complexes captured on a nanopore. *J. Am. Chem. Soc.* **131,** 3772–8 (2009).

129. Wilson, N. A. *et al.* Electronic control of DNA polymerase binding and unbinding to single DNA molecules. *ACS Nano* **3,** 995–1003 (2009).

130. Chu, J., González-López, M., Cockroft, S. L., Amorin, M. & Ghadiri, M. R. Real-time monitoring of DNA polymerase function and stepwise single-nucleotide DNA strand translocation through a protein nanopore. *Angew. Chemie* **122,** 10304–10307 (2010).

131. Olasagasti, F. *et al.* Replication of individual DNA molecules under electronic control using a protein nanopore. *Nat. Nanotechnol.* **5,** 798–806 (2010).

132. Wang, H., Hurt, N. & Dunbar, W. B. Measuring and modeling the kinetics of individual DNA-DNA polymerase complexes on a nanopore. *ACS Nano* **7,** 3876–86 (2013).

133. Garalde, D. R. *et al.* Distinct complexes of DNA polymerase I (Klenow fragment) for base and sugar discrimination during nucleotide substrate selection. *J. Biol. Chem.* **286,** 14480–92 (2011).

134. Cockroft, S. L., Chu, J., Amorin, M. & Ghadiri, M. R. A single-molecule nanopore device detects DNA polymerase activity with single-nucleotide resolution. *J. Am. Chem. Soc.* **130,** 818–20 (2008).

135. Langer, A. *et al.* Polymerase/DNA interactions and enzymatic activity: multi-parameter analysis with electro-switchable biosurfaces. *Sci. Rep.* **5,** 12066 (2015).

136. Choi, Y. & Moody, I. S. *et al.* Single-molecule lysozyme dynamics monitored by an electronic circuit. *Science* **335,** 319–24 (2012).

137. Sims, P. C. *et al.* Electronic measurements of single-molecule catalysis by cAMP-dependent protein kinase A. (2013).

138. Choi, Y. & Olsen, T. J. *et al.* Dissecting single-molecule signal transduction in carbon nanotube circuits with protein engineering. *Nano Lett.* **13,** 625–31 (2013).

139. Laborde, C. & Pittino, F. *et al.* Real-time imaging of microparticles and living cells with CMOS nanocapacitor arrays. *Nat. Nanotechnol.* **10,** 791–5 (2015).

140. Ogorzalek, T. L. *et al.* Molecular-level insights into orientation-dependent changes in the thermal stability of enzymes covalently immobilized on surfaces. *Langmuir* **31,** 6145–53 (2015).

141. Nishida, H. *et al.* Self-oriented immobilization of DNA polymerase tagged by titanium-binding peptide motif. *Langmuir* **31,** 732–40 (2015).

142. Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., & Rank, D. *et al.* Real-time DNA sequencing from single polymerase molecules. *Science* **323,** 133–8 (2009).

143. Blanco, L. *et al.* Highly efficient DNA synthesis by the phage phi 29 DNA polymerase. Symmetrical mode of DNA replication. *J. Biol. Chem.* **264,** 8935–8940 (1989).

144. Flusberg, B. A. *et al.* Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods* **7,** 461–5 (2010).

145. Murray, I. A. *et al.* The methylomes of six bacteria. *Nucleic Acids Res.* **40,** 11450–62 (2012).

146. Kim, Y. *et al.* Reading single DNA with DNA polymerase followed by atomic force microscopy. *J. Am. Chem. Soc.* **136,** 13754–60 (2014).

147. Lieberman, K. R. *et al.* Processive replication of single DNA molecules in a

nanopore catalyzed by phi29 DNA polymerase. *J. Am. Chem. Soc.* **132,** 17961–72 (2010).

148. Manrao, E. A. *et al.* Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nat. Biotechnol.* **30,** 349–53 (2012).

149. Wang, Y., Yang, Q. & Wang, Z. The evolution of nanopore sequencing. *Front. Genet.* **5,** 449 (2014).

150. Dahl, J. M., Wang, H., Lázaro, J. M., Salas, M. & Lieberman, K. R. Kinetic mechanisms governing stable ribonucleotide incorporation in individual DNA polymerase complexes. *Biochemistry* **53,** 8061–76 (2014).

151. 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).

152. 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).

153. Dahl, J. M., Wang, H., Lázaro, J. M., Salas, M. & Lieberman, K. R. Dynamics of translocation and substrate binding in individual complexes formed with active site mutants of phi 29 DNA polymerase. *J. Biol. Chem.* **289,** 6350–61 (2014).

154. Dahl, J. M. *et al.* Direct observation of translocation in individual DNA polymerase complexes. *J. Biol. Chem.* **287,** 13407–21 (2012).

155. Lieberman, K. R., Dahl, J. M. & Wang, H. Kinetic mechanism at the branchpoint between the DNA synthesis and editing pathways in individual DNA polymerase complexes. *J. Am. Chem. Soc.* **136,** 7117–31 (2014).

156. Zeng, T. & Liu, L. *et al.* Detection of 5-methylcytosine and 5-hydroxymethylcytosine in DNA via host–guest interactions inside α-hemolysin nanopores. *Chem. Sci.* **6,** 5628–5634 (2015).

157. Szalay, T. & Golovchenko, J. A. De novo sequencing and variant calling with nanopores using PoreSeq. *Nat. Biotechnol.* **33,** 1087-1091 (2015).

158. Berlin, K. & Koren, S. *et al.* Assembling large genomes with single-molecule sequencing and locality-sensitive hashing. *Nat. Biotechnol.* **33,** 623–630 (2015).

159. Holzberger, B. & Marx, A. Replacing 32 proline residues by a noncanonical amino acid results in a highly active DNA polymerase. *J. Am. Chem. Soc.* **132,** 15708–13 (2010).

160. Reha-Krantz, L. J., Woodgate, S. & Goodman, M. F. Engineering processive DNA polymerases with maximum benefit at minimum cost. *Front. Microbiol.* **5,** 380 (2014).

161. Huang, S., Romero-Ruiz, M., Castell, O. K., Bayley, H. & Wallace, M. I. High-throughput optical sensing of nucleic acids in a nanopore array. *Nat. Nanotechnol.* **10,** 986-991 (2015).

162. Farimani, A. B., Heiranian, M. & Aluru, N. R. Electromechanical signatures for DNA sequencing through a mechanosensitive nanopore. *J. Phys. Chem. Lett.* **6,** 650–7 (2015).

163. Ayub, M., Stoddart, D. & Bayley, H. Nucleobase recognition by truncated α-hemolysin pores. *ACS Nano* **9,** 7895–903 (2015).

164. Rodríguez-Manzo, J. A., Puster, M., Nicolaï, A., Meunier, V. & Drndić, M. DNA translocation in nanometer thick silicon nanopores. *ACS Nano* **9,** 6555–6564 (2015).

165. Tagliazucchi, M. & Szleifer, I. How does confinement change ligand-receptor binding equilibrium? Protein binding in nanopores and nanochannels. *J. Am. Chem. Soc.* **137**, 12539-12551 (2015).

166. Plesa, C., Ruitenberg, J. W., Witteveen, M. J. & Dekker, C. Detection of individual proteins bound along DNA using solid-state nanopores. *Nano Lett.* **15,** 3153–8 (2015).

167. Yu, J.-S. & Lim, M.-C. *et al.* Identifying the location of a single protein along the DNA strand using solid-state nanopores. *ACS Nano* **9,** 5289–98 (2015).

168. Wang, C. & Fu, Q. *et al.* Atomic layer deposition modified track-etched conical nanochannels for protein sensing. *Anal. Chem.* **87,** 8227–33 (2015).

169. 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).

Chapter 2: Processive Incorporation of Deoxynucleoside Triphosphate Analogs by Single-Molecule DNA Polymerase I (Klenow Fragment) Nanocircuits

2.1 Abstract

DNA polymerases exhibit a surprising tolerance for analogs of deoxyribonucleoside triphosphates (dNTPs), despite the enzymes' highly evolved mechanisms for the specific recognition and discrimination of native dNTPs. Here, individual DNA Polymerase I Klenow Fragment (KF) molecules were tethered to a single-walled carbon nanotube field-effect transistor (SWCNT-FET) to investigate accommodation of dNTP analogs with single-molecule resolution. Each base incorporation produced a change in current with its duration defined by τ_{closed}. Under V_{max} conditions, the average time of τ_{closed} was similar for all analog and native dNTPs (0.2 to 0.4 ms), indicating no kinetic impact on this step due to analog structure. Accordingly, the average rates of dNTP analog incorporation were largely determined by durations with no change in current defined by τ_{open}, which includes molecular recognition of the incoming dNTP. All α-thio-dNTPs were incorporated more slowly, at 40 to 65% of the rate for the corresponding native dNTPs. During polymerization with 6-Cl-2APTP, 2-thio-dTTP, or 2-thio-dCTP, the nanocircuit uncovered an alternative conformation represented by positive current excursions that do not occur with native dNTPs. The results suggest the enzyme applies a dynamic stability-checking mechanism for each nascent base pair, and exemplifies the potential of SWCNT-FETs for both fundamental and biotechnological applications.

2.2 Introduction

To ensure their survival, all organisms rely on DNA polymerases to correctly recognize deoxynucleoside triphosphates (dNTPs) and successfully catalyze their incorporation into new strands of DNA. The fidelity of DNA polymerases relies predominantly on the correct geometry of the nascent base pair.[1–3] By forming base pairs with a shape and size similar to the canonical A•T and G•C base pairs, isosteric and hydrophobic dNTP analogs incapable of hydrogen bonding with native, complementary bases have been successfully incorporated by DNA polymerases.[4–6] Studies with other dNTP analogs substantiate a requirement for shape complementarity[7–9] and demonstrate that additional factors, including stereochemistry, sterics, electronic effects, base stacking, and hydrophobic interactions contribute to the fidelity of DNA polymerases.[8,10–15]

Experiments with dNTP analogs can also illustrate the prerequisite for snug fit in the DNA polymerase active site.[8,12,14] For example, recent crystal structures of the A-family DNA polymerase fragment KlenTaq demonstrate a mutually induced fit to a Watson-Crick geometry by both the nascent base pair and the polymerase during accommodation of unnatural or non-native pairing,[16] illustrating plasticity in the enzyme's molecular recognition. Specifically, the O helix of A-family DNA polymerases adopts multiple distinct conformations in the active site to discriminate against imperfect substrates.[17] For example, O helix residues F762 and Y766 in the DNA Polymerase I Klenow Fragment (KF) active site discriminate amongst dNTPs through base specificity to maintain fidelity,[18–20] and are involved in the active site tightness around the nascent base pair.[6,21]

Detailed evaluation of unnatural dNTP polymerization beyond a single base pair could provide accurate kinetic information about this non-native polymerase activity, and uncover general insights into enzyme molecular recognition and specificity. As reported here, such information can be elucidated by single-molecule techniques; observation of intermediate steps and transition states are otherwise lost through averaging in ensemble populations.[22,23] DNA polymerization experiments employing single-molecule Förster resonance energy transfer (smFRET) have revealed conformational flexibility and insights into the fidelity mechanism of KF.[24–27] Additionally, the observed kinetic and thermodynamic differences between correct and mismatched dNTP-DNA-polymerase complexes quantify the driving forces for correct nucleotide incorporation into DNA templates.[26–28]

The experiments reported here differ from previous fluorescence-based studies of DNA polymerases. Despite its power to capture new information about enzyme dynamics, FRET requires a fluorescently labeled protein and/or substrate. Photobleaching and the flux of photons between fluorophores limit both the duration and time resolution, respectively, of FRET and smFRET experiments. Furthermore, smFRET experiments typically examine single dNTP or dideoxy-NTP (ddNTP) incorporation events. Such experiments offer the important advantage of avoiding the potentially distorted homopolymeric dsDNA products. In the experiments reported here, however, homopolymeric templates must be used to determine the kinetics of individual dNTPs and their analogs in multiple turnover events. The approach enables these first single-molecule studies to explore unnatural dNTP analog incorporation beyond the base recognition step, expanding this evaluation to processive incorporation events.

Recently, we described a new approach to single-molecule enzymology and applied it to three enzymes. In this technique, an individual protein is bioconjugated to a single-walled carbon nanotube field effect transistor (SWCNT-FET; Figure 1-9). The approach uncovered new insights into the number of steps, kinetic parameters, and processivity of the well-studied T4 lysozyme.[29,30] Engineered lysozyme variants revealed that significant differences in electronic signal resulted from the electrostatic charges of side chains close to the attachment site.[31] Tremendously dynamic rates of cAMP-dependent protein kinase A (PKA) demonstrated the role of protein kinases as a highly variable molecular switch.[32] KF conjugated to SWCNT-FETs uncovered significant kinetic and conformational differences between the enzyme-catalyzed formation of A•T/T•A and G•C/C•G sets of base pairs.[33]

The sensitivity of the KF nanocircuit to small differences in base pair recognition is surprising since the native Watson-Crick base pairs have similar sizes. By coupling to key active site residues, the location of attachment at position 790 in the "fingers" subdomain can allow transfer of active site dynamics to charged surface residues near the SWCNT; though the approach reported here requires attachment to the SWCNT, conjugation at this position causes only minimal perturbation of KF's activity compared to solution-phase DNA polymerization assays.[33] Combining the SWCNT-FET sensitivity with KF's malleable active site recognition, we hypothesized that the KF nanocircuit technique might also be responsive to the unnatural structures of dNTP analogs. Translation of the associated small conformational changes during base recognition and subsequent incorporation into a reliable measurement could reveal new aspects of the roles for base pair structure, both steric and electronic, during DNA polymerization.

Here, the SWCNT-FET technique was used to examine KF kinetic and conformational dynamics during native and analog dNTP incorporation. Using dNTP analogs with phosphodiester or nucleobase modifications, DNA polymerization was monitored with single base pair resolution and then statistically analyzed to reveal differences in incorporation kinetics and conformations. Analogs significantly altered the time required for nucleotide recognition, but not the kinetics of KF's closed conformation. However, signals revealed a highly plastic aspect of the closure as the enzyme accommodated various analog base pairs with different structures, hydrogen bonding patterns, and electron densities.

2.3 Experimental section

SWCNT-FETs were fabricated[29] and functionalized with a single-cysteine variant of exonuclease-deficient KF (D355A/E357A/L790C/C907S).[33,34] Purification of KF to >95% ensured its homogeneity (Figure 2-1A). A fluorescence-based assay confirmed activity of the bulk enzyme prior to attachment (Figure 2-1B).[33,35] Attachment of KF to SWCNT-FETs was accomplished by soaking the devices in a solution of *N*-(1-pyrenyl)maleimide (1 mM in ethanol, 30 min), followed by incubation with KF (300 nM KF in a standard KF activity buffer of 20 mM Tris, 50 mM NaCl, 10 mM $MgCl_2$, 100 μM TCEP, pH 8.0). Atomic force microscopy after data collection confirmed attachment of a single KF molecule to each device (Figure 2-2). Such devices are referred to simply as KF nanocircuits.

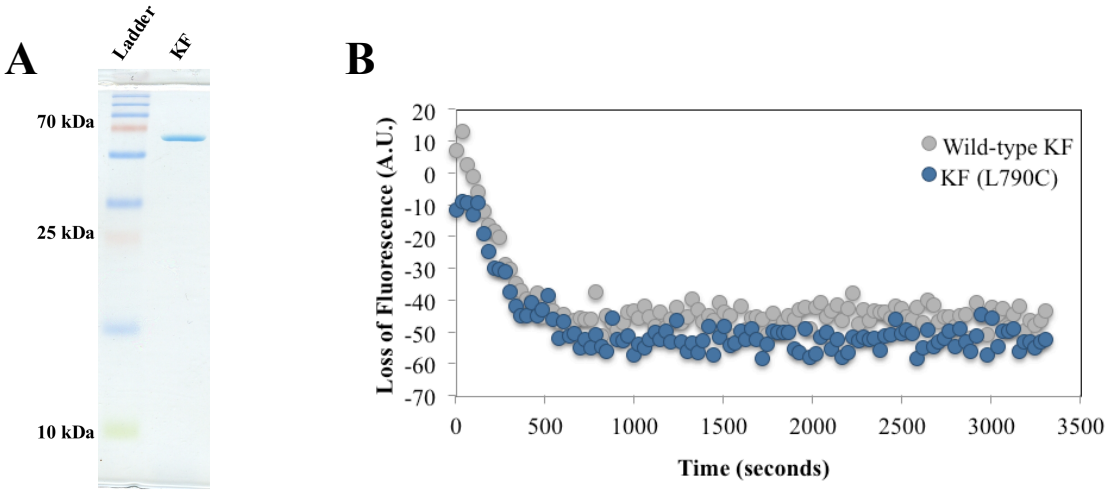


Figure 2-1. Representative 15% SDS-PAGE gel and fluorescence-based activity assay after over-expression and purification of KF (L790C). **(A)** KF was purified to >95% homogeneity and migrated at its expected mass of ≈68 kDa. **(B)** Activity of KF (L790C) was compared to activity of wild-type KF under steady-state conditions. The primer extension reaction occurs in the presence of dATP, dTTP, dCTP, and dGTP. The raw data was subtracted from background, which measured activity in the absence of dNTPs.

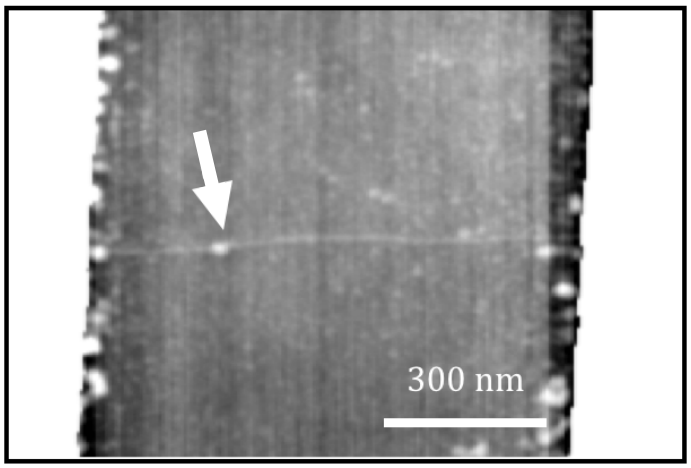
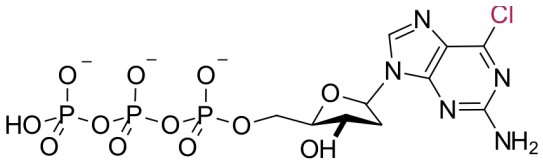
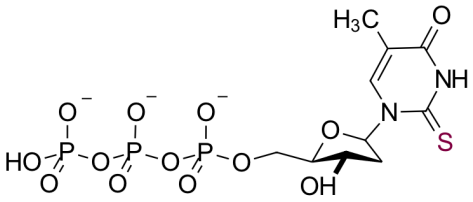


Figure 2-2. Single nanocircuit with KF attachment. Atomic force microscopy shows the 1-2 nm diameter of the SWCNT-FET with a single KF attachment (7 nm, arrow).

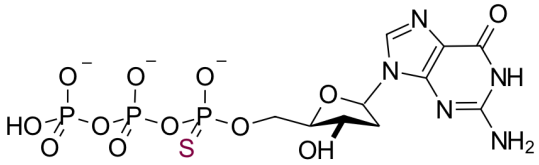
Experiments used the homopolymeric DNA templates $(A)_{42}$, $(T)_{42}$, $(G)_{42}$, or $(C)_{42}$ mixed with complementary dNTP analogs. Each template was fused to an M13 priming site and mixed with an M13 forward primer in a 1:1 stoichiometric ratio; for hybridization, the mixture was heated in a thermal cycler to 95 °C for 5 to 10 min followed by cooling to 65 °C then further cooling with a gradient of 5 °C every five min until reaching room temperature. KF nanocircuits were immersed in activity buffer with the annealed template-primer at 100 nM concentrations. Native or analog dNTPs were added to the buffer in excess, ensuring V_{max} conditions for KF catalysis. To compensate for possibly reduced affinity of dNTP analogs, the experiments applied higher concentrations of analogs (Figure 2-3, 100 μM, Trilink Biotechnologies) than the native dNTPs (10 μM, Fisher). The supplier provided structural characterizations of the dNTP analogs.



6-Cl-2APTP



2-thio-dTTP



S_p α–thio-dGTP

Figure 2-3. Chemical structures of representative dNTP analogs. Chemical modifications from the native dNTPs are highlighted in red.

Measurements consisted of monitoring the source-drain current, $I(t)$, through the SWCNT-FET while the attached KF molecule interacted with its surrounding environment. The drain electrode was biased at 100 mV, and the electrolyte, which served as a gate electrode, was held at or near 0 V. Incubation of the device with any template-primer and its complementary dNTPs transduced fluctuations, $\Delta I(t)$, whereas these fluctuations were absent with non-complementary dNTPs or in control measurements missing the template-primer or KF attachment.[33] $I(t)$ fluctuations were amplified (Keithley 428), digitized at 100 kHz, and stored as uninterrupted, 600 s data sets for later analysis. Between measurements, the KF nanocircuits were rinsed twice with activity buffer, incubated in buffer for 5 min, then rinsed twice with buffer before introducing another nucleotide and template-primer. Each KF molecule was monitored with multiple analogs, their corresponding native dNTPs, and nucleotide-free buffer in order to collect directly comparable data sets, confirm typical KF activities, and reproduce the types of $\Delta I(t)$ excursions reported previously.[33]

2.4 Signals generated by dNTP analogs upon incubation with KF nanocircuits

Figures 2-4A and 2-4B show representative $\Delta I(t)$ signals produced by a KF nanocircuit processing a $(C)_{42}$ DNA template in the presence of dGTP. The device produced uninterrupted sequences of negative $\Delta I(t)$ excursions, shown at three different magnifications. Each $\Delta I(t)$ excursion indicated the formation of one base pair, and the kinetic parameters derived from $\Delta I(t)$ data sets were consistent with previous single-molecule analysis of KF motions[33] and ensemble KF incorporation rates.[36,37] As observed previously, results with G•C or C•G base pair formation were essentially identical to one another; A•T/T•A base pair formation also provided very similar

polymerization kinetics, dynamics and $\Delta I(t)$ values compared to each other. Measurements with the native dNTPs provided baseline values for comparison with dNTP analogs.

Commercially available dNTP analogs were incorporated into DNA through KF polymerization in both ensemble (Figure 2-5) and single-molecule assays. Measurements with α-thio-dNTP, or dNTPαS, analogs produced $\Delta I(t)$ data sets that appeared quite similar to the native dNTPs, but with different incorporation rates (Figure 2-4C). When measured with KF nanocircuits, incorporation of 6-chloro-2-aminopurine-drTP, or 6-Cl-2-APTP, opposite both $(C)_{42}$ and $(T)_{42}$ DNA templates caused $\Delta I(t)$ signals with inverted amplitude reflecting a different KF conformation (Figure 2-4D). This analog incorporated more slowly; for example, opposite $(C)_{42}$, 6-Cl-2APTP produced $\Delta I(t)$ excursions at 80% of the rate of dGTP. $\Delta I(t)$ records with the 2-thio-dNTP analogs produced mixed behaviors in which KF activity produced negative $\Delta I(t)$ excursions during one minute, positive $\Delta I(t)$ excursions during another minute, and, more rarely, mixtures of both behaviors along a single template strand (Figures 2-4E,F).

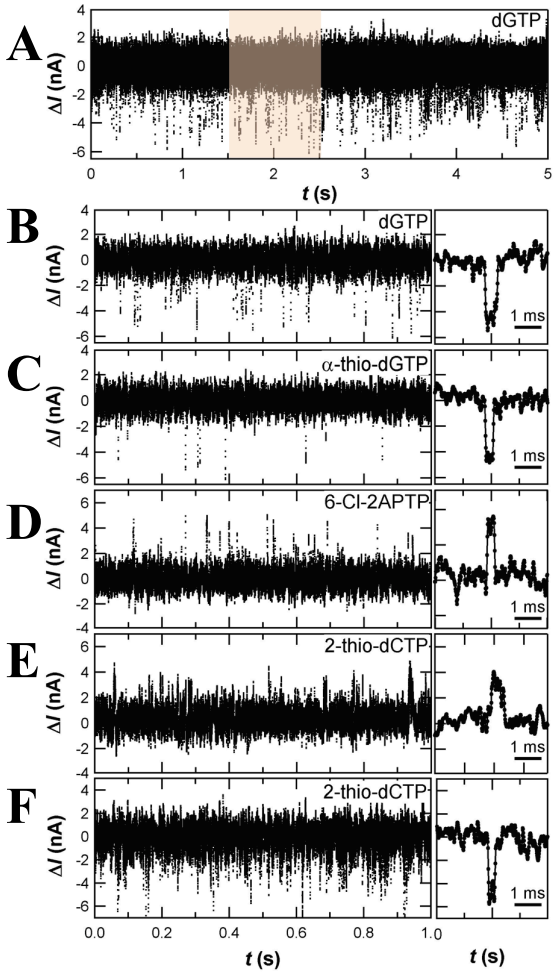


Figure 2-4. Δ*I*(t) excursions during native and analog dNTP incorporation. High and low current states correspond to conformational dynamics of the enzyme during processive nucleotide incorporation. (**A**) In the presence of a $(C)_{42}$ DNA template and its complementary native dGTP, Δ*I*(t) excursions occur during each base incorporation. (**B**) Magnification of the highlighted region in (**A**) illustrates the Δ*I*(t) events corresponding to single base incorporations. For comparison, representative 1 s data sets are shown for the same KF nanocircuit incorporating the dNTP analogs (**C**) α-thio-dGTP, (**D**) 6-Cl-2APTP, and (**E,F**) 2-thio-dCTP. To the right of each data set, the magnified view depicts one typical Δ*I*(t) excursion.

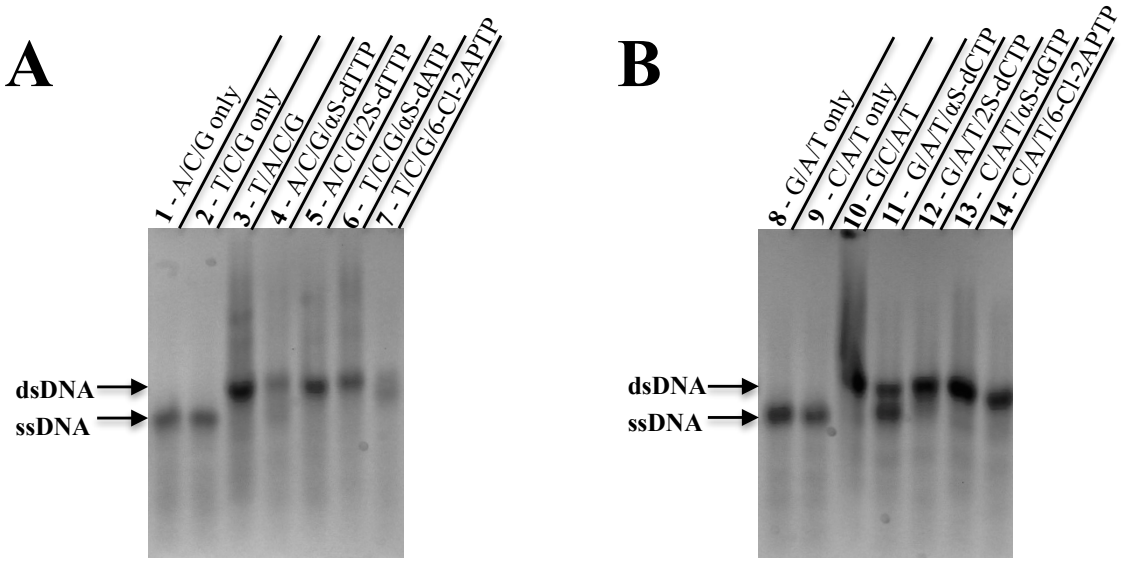


Figure 2-5. Ensemble assay showing incorporation of dNTP analogs with the templates described in Table 1. Polymerization products with dNTP analogs and the A/T incorporation template (**a**) or the G/C incorporation template (**b**) were electrophoresed on a 5% high-resolution agarose gel. Negative control reactions with only 3 dNTPs, omitting dTTP (**1**), dATP (**2**), dCTP (**8**), and dGTP (**9**), contained no dsDNA. Positive control reactions with all four dNTPs showed conversion to dsDNA with both the A/T incorporation template (**3**) and the G/C incorporation template (**10**). Reactions with dNTP analogs (**4**-**7** and **11**-**14**) omitted their native dNTP counterpart and contained the remaining 3 native dNTPs. Opposite the A/T incorporation template, α-thio-dTTP (**4**) and 2-thio-dTTP (**5**) incorporated opposite the template base A, and α-thio-dATP (**6**) and 6-Cl-2APTP (**7**) incorporated opposite the template base T. Opposite the G/C incorporation template, α-thio-dCTP (**11**) and 2-thio-dCTP (**12**) incorporated opposite the template base G, and α-thio-dGTP (**13**) and 6-Cl-2APTP (**14**) incorporated opposite the template base C. After visualization, the image colors were inverted, then changed to black and white.

In our previous report with native dNTPs, the time constants for the experimental baseline current, τ_{open} here, were referred to as τ_{hi}. Time constants representing a native dNTP incorporation event all occurred with lower current, and were referred to as τ_{lo}.[33] Since positive, negative, or mixtures of both positive and negative $\Delta I(t)$ excursions are reported here, time constants for either direction of excursions are termed τ_{closed}. Distributions of τ_{open} and τ_{closed} were derived from each record of polymerization data.

2.5 Statistical analysis of dNTP analog incorporation compared to native dNTPs

Figure 2-6 shows example distributions for incorporation of dGTP substrates into $(C)_{42}$ DNA templates. The distributions from native and analog dGTP τ_{closed} events were nearly indistinguishable except for rare events in the tails, for which we have the poorest statistics (Figure 2-6A). To draw comparisons between native and analog dNTPs, we focused on the mean time constant $<\tau>$ of the primary, Poissonian component of these distributions. All of the mean values for $<\tau_{closed}>$ were in close agreement around 0.3±0.1 ms. By comparison, the distributions and mean values of $<\tau_{open}>$ were clearly different. For example, KF spent 63.6±2.8 ms in its open conformation when processing α-thio-dGTP, which is 56% longer than the 40.8±0.6 ms observed for native dGTP (Figure 2-6B).

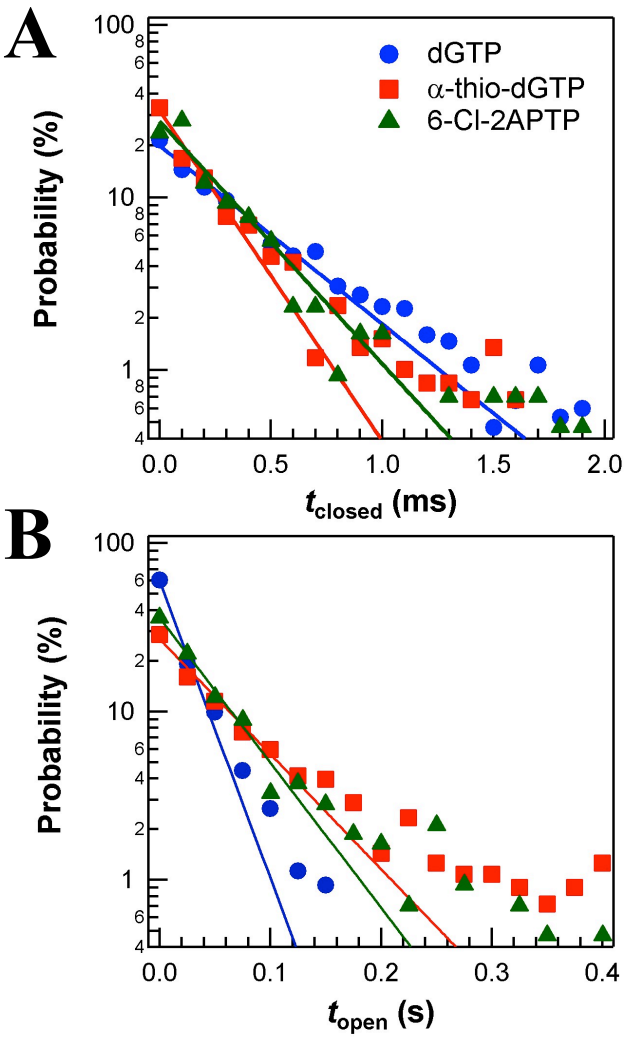


Figure 2-6. Direct comparison of the probability distributions of $<\tau_{open}>$ and $<\tau_{closed}>$ durations during incorporation of the indicated dNTPs from >50 s data sets. For both (a) τ_{closed} and (b) τ_{open}, the homopolymeric $(C)_{42}$ was the DNA template used. Single-exponential fits for each nucleotide are shown as solid lines.

The kinetic parameters $<\tau_{closed}>$, $<\tau_{open}>$, and the average rate of incorporation k were analyzed for the four homopolymeric templates with native and analog dNTPs (Table 2-1). As with the case described above, every combination produced identical τ_{closed} distributions with $<\tau_{closed}>$ values in the range of 0.3±0.1 ms. While a similar effect was previously observed for the four native dNTPs,[33] the extension of this result to dNTP analogs having different nucleobase sizes, electronic properties, hydrogen bonding, or substitution at the α-phosphodiester was unexpected.

Table 2-1. Kinetics of native and analog dNTP incorporation by KF.[a]

Template	Nucleotide	$<t_{open}$ (ms)>	$<t_{closed}$ (ms)>	k (1/s)
$(T)_{42}$	dATP	58.9 ± 1.2	0.34 ± 0.18	16.8 ± 0.4
	α-thio-dATP	145.9 ± 8.4	0.38 ± 0.21	6.8 ± 0.4
$(A)_{42}$	dTTP	69.6 ± 2.3	0.33 ± 0.12	14.3 ± 0.5
	α-thio-dTTP	152.1 ± 6.6	0.29 ± 0.13	6.6 ± 0.3
	2-thio-dTTP	61.1 ± 3.2[b]	0.23 ± 0.14[b]	16.3 ± 0.9[b]
$(G)_{42}$	dCTP	42.8 ± 5.0	0.35 ± 0.20	23.2 ± 3.2
	α-thio-dCTP	68.8 ± 4.6	0.33 ± 0.19	14.5 ± 1.0
	2-thio-dCTP	32.3 ± 1.1[b]	0.41 ± 0.15[b]	30.6 ± 1.2[b]
$(C)_{42}$	dGTP	40.8 ± 6.0	0.40 ± 0.20	24.3 ± 4.3
	α-thio-dGTP	63.6 ± 2.8	0.21 ± 0.15	15.7 ± 0.7
	6-Cl-2APTP	50.5 ± 1.4	0.20 ± 0.12	19.7 ± 0.6

[a]Average values ± standard deviation

[b]Similar values were observed for both up- or down-switching events.

On the other hand, τ_{open} was far more sensitive to dNTP identity. The mean duration of $<\tau_{open}>$ ranged from 23 ms with native dCTP to 145 ms with α-thio-dATP. Among the four native dNTPs, $<\tau_{open}>$ was longer for dTTP or dATP incorporation than for dGTP or dCTP incorporation. This hierarchy was preserved within longer $<\tau_{open}>$ durations measured for all four α-thio-dNTPs. The α-thio substitution increased $<\tau_{open}>$ by 50% in the case of dGTP and dCTP, whereas the increase was more than 100% for dTTP and dATP.

The average KF processing rate for dNTP incorporation was calculated as $k = (<\tau_{open}> + <\tau_{closed}>)^{-1}$. τ_{open} largely determines k, because it is at least 60 times longer than τ_{closed}. At its fastest, KF incorporated 2-thio-dCTP at more than 30 s^{-1}. The increase in τ_{open} described above for α-thio-dNTPs reduced k to 15 s^{-1} for α-thio-dCTP and α-thio-dGTP and 7 s^{-1} for α-thio-dATP and α-thio-dTTP. Rates for 6-Cl-2APTP incorporation compared most favorably to the slowest rates observed for native dGTP incorporation. Conversely, 2-thio-dTTP and 2-thio-dCTP incorporation appeared slightly faster than incorporation of their native counterparts.

Similar results were reproduced using a dozen different KF molecules. Each KF was attached to a different SWCNT-FET and measured independently. For comparison, a non-homopolymeric template measured with dNTP analogs resulted in similar kinetics (data not shown). As mentioned previously, our experiments applied 100 μM of dNTP analogs to ensure steady state conditions; for comparision, 10 μM α-thio-dATP with the $(T)_{42}$ template did not affect DNA polymerization. Due to static disorder, some KF molecules processed faster or slower than the ensemble average, but without any significant change to the relative comparison of analog to native dNTPs.

2.6 Significance of dNTP analog nucleobase structure on incorporation kinetics

The dNTP analogs were chosen for their ability to be incorporated into DNA templates by DNA polymerases and variations in sizes, structures, and reactivity. We examined either substitution at the α-phosphate or nucleobase. The first type of analog, α-thio-dNTP, substituted a non-bridging, α-phosphoryl oxygen atom with sulfur to introduce a new stereocenter and alter the reactivity at this crucial site. The second category of dNTP analogs, halogen or sulfur substitution on the nucleobase, changes the size and electronic structure of the base pair; some analogs also alter the hydrogen bonding available for base pairing. For example, 6-Cl-2-APTP (Figure 2-3), has two hydrogen bonding profiles, allowing its incorporation opposite both T and C bases.[9,15,38] Compared to dATP, 6-Cl-2-APTP replaces the 6-amino group with chlorine, but introduces a 2-amino functionality; this configuration ultimately provides the same number of Watson-Crick hydrogen bonds complementary to T as dATP. When used as a dGTP analog, 6-Cl-2-APTP has different tautomerization, which changes the N-1 from a hydrogen bond donor to an acceptor. In this case, replacement of oxygen with chlorine dramatically decreases the strength of the hydrogen bonding.[39] Like 6-Cl-2APTP, sulfur-substituted analogs 2-thio-dTTP and 2-thio-dCTP also form larger base pairs due to the increased bond length of the thiocarbonyl.[40]

The single-molecule experiments carried out in this study illustrate and shed new light on the well-appreciated plasticity of DNA polymerases like KF. This class of enzymes can accommodate even dramatically modified incoming dNTPs. However, we directly observe conformational motions required by the enzyme to maintain fidelity when faced with certain altered dNTPs. Reflecting the limits for such accommodations,

DNA polymerases are known to exhibit strong sensitivity to minor changes in dNTP size and shape.[8,12] Our analysis benefits from comparing single molecule data with native and analog dNTPs during numerous processive incorporation events. This analysis begins with the kinetics of the two observed enzyme conformations during catalysis, which were captured by τ_{open} and τ_{closed}.

Events taking place during τ_{open} include the rate-limiting step of dNTP recognition, which is sensitive to both nucleobase and backbone modifications. Successful recognition and binding of the appropriate nucleotide triggers KF's activation and closure.[41] Previous FRET-based experiments with the related T7 DNA polymerase have identified a "fully open" conformational state resulting from mismatch recognition.[42] However, using the L790C attachment site, the SWCNT-FET records no KF motions and no signals in the presence of mismatched dNTPs. The absence of intermediate states or mismatch-associated motions suggests that our attachment site is insensitive to this initial fidelity checkpoint. Thus, $\Delta I(t)$ excursions result from a catalytically committed conformation, and are not restricted to simply the global motion of the enzyme opening and closing.

We propose that the differences observed in τ_{open} largely reflect the mechanisms for recognizing and binding unnatural dNTPs. Long tails in the distributions for α-thio-dGTP and 6-Cl-2APTP compared to native dGTP may have been responsible for the $<\tau_{open}>$ increase (Figure 2-6B). In fact, these tails can be fit to second exponentials with time constants of 200 ms, about five times longer than $<\tau_{open}>$ for native dGTP. Similar long tails were observed with all tested dNTP analogs, illustrating the challenges faced by the enzyme when incorporating non-natural substrates. Steps other than recognition

potentially take place during the τ_{open} reported here; covalent bond formation is one possible example that would occur too quickly, even with the slowed reactivity of α-thio-dNTPs, to be detectable as rate-limiting.[43] Faster rates of incorporation observed with the 2-thio analogs can result from more stable base pair formation, effectively shortening $<\tau_{open}>$ values. The larger size of the 2-thio-dCTP sulfur atom at the hydrogen bonding interface with the template G base does not appear to affect the ability of 2-thio-dCTP to base pair efficiently. These results agree with the previously observed increase in polymerization efficiency with 2-thio-dTTP and 4-thio-dTTP compared to dTTP incorporation.[14,44]

The 6-Cl-2APTP analog, with much weaker hydrogen bonding and consequent imperfect base pairing compared to dGTP, exemplifies the challenges of base pairing recognition during KF-catalyzed DNA polymerization. Longer $<\tau_{open}>$ values for 6-Cl-2APTP versus dGTP incorporation opposite a $(C)_{42}$ DNA template illustrate the willingness of DNA polymerases to accept unnatural dNTPs in part by lengthening the time allotted for recognition. The $<\tau_{open}>$ value, and thus the rate of incorporation, observed during 6-Cl-2APTP polymerization opposite $(C)_{42}$ fell between the values measured for native dGTP and dATP incorporation opposite complementary, homopolymeric templates. Thus, despite its altered tautomerization and consequent loss of at least one base pairing hydrogen bond when compared to dGTP, 6-Cl-2APTP can still be incorporated more quickly opposite $(C)_{42}$ than native dATP opposite $(T)_{42}$. Notably, the base pairing hydrogen bond in the minor groove remains unchanged when 6-Cl-2APTP is considered a dGTP analog, and could govern the relatively faster rates observed for dGTP, dCTP, and 6-Cl-2APTP opposite a $(C)_{42}$ DNA template.

In addition to recognition and binding, prolonged $<\tau_{open}>$ values for α-thio-dNTP incorporation could result from the reduced stability of the newly synthesized DNA. KF-catalyzed processing of homopolymeric templates can result in distorted dsDNA.[45] Furthermore, α-thio-dNTPs are particularly prone to form less stable binary complexes with unfavorable DNA backbone interactions, which progressively slows the catalytic rate of KF.[46,47] More pronounced effects on this step, compared to experiments with the respective native dNTPs, were observed during α-thio-dATP/α-thio-dTTP versus α-thio-dGTP/α-thio-dCTP incorporation. Such results may indicate sequence-dependent DNA instability, which underscores an important caveat to these studies with homopolymeric templates. Alternatively, this difference could suggest that the α-thio substitution further interferes with the mechanism that causes $<\tau_{open}>$ to be longer for native A•T/T•A base pairs. Some of the variation in τ_{open} associated with α-thio-dNTP incorporation could result from the weakly inhibitory R_p stereoisomer ($K_i \approx 30$ μM), present at an approximately 1:1 ratio with the S_p stereoisomer in the commercial synthesis of this analog.[48] This inhibition is about an order of magnitude weaker than the K_m for the native dNTP,[10] and thus can be expected to affect $<\tau_{open}>$ values only modestly.

2.7 Potential source of alternative KF motions during dNTP analog incorporation

During τ_{closed}, KF undergoes a distinct conformational change corresponding to formation of one phosphodiester bond between the incoming nucleotide and the nascent dsDNA. In substrate-limited experiments, the number of $\Delta I(t)$ excursions matched the number of overhanging template bases;[33] thus, the conformational change during τ_{closed} must occur for each successful, processive nucleotide incorporation. Earlier, the short and equal duration of $<\tau_{closed}>$ for native dNTPs supported a model in which τ_{closed} results

from the covalent bond-forming step itself.[33] Here, we reevaluate this assignment due to three observations with dNTP analogs. First, the direction of $\Delta I(t)$ excursions was reversed for some dNTP analogs. Second, incorporation of 2-thio-dNTP analogs produced mixtures of both positive and negative $\Delta I(t)$ excursions. Third, as shown in Table 2-1, the invariance in $<\tau_{closed}>$ extended to all analogs tested despite substitutions at the electrophilic α-phosphate or the likely alternative conformations needed to accommodate substitutions on the nucleobase.

In this electronic technique, the underlying SWCNT-FET is extremely sensitive to electrostatic gating by the protein's charged surface residues within 1 nm of the attachment site.[31] Previous work proved that different variants of the same enzyme can exhibit either positive or negative $\Delta I(t)$ excursions depending on the charge of the SWCNT-adjacent residues and the directions of their motion.[31] KF and its charged residues electrostatically gating the SWCNT-FET remain invariant in this study. Therefore, variable $\Delta I(t)$ excursions indicate that the residues adjacent to the KF attachment site are adopting different motions in response to certain dNTP analogs during a catalytically competent cycle. Such motions are likely transmitted from the KF active site through allostery, but they are not necessarily the motions of covalent bond formation. In fact, we conclude that the covalent step could not proceed by the same mechanism and with the same $<\tau_{closed}>$ duration but with two opposing motions. Instead, the relevant residue motions responsible for τ_{closed} are likely independent of both initial molecular recognition and the chemical step of KF catalysis.

KF is attached to the SWCNT-FET through the protein's L790C sidechain in the "fingers" subdomain, linking the electrostatic gating motions of relevant charged residues

to catalytically committed motions during τ_{closed}. We propose that each τ_{closed} event results from the active site O helix itself or a particular O helix residue twisting in two possible directions during the observed stage of successful nucleotide incorporation. This proposed twisting is inferred by considering active site residue motions during known stages of nucleotide incorporation and their effect on the theoretical proximity of charged residues to the SWCNT-FET. For example, smFRET experiments with KF reveal an intermediate conformation of the active site O helix between the open and closed states;[27] a potentially analogous "ajar" conformation was observed in the crystal structures of the KF homolog *Bst* Pol I. The C-terminus of the *Bst* Pol I O helix kinks on the pathway to closure such that a large shift of the KF Y766 equivalent is accompanied by a subtle rotation of the KF F762 equivalent. The rotation of the KF F762 equivalent continues until enzyme closure.[49–51]

By comparing crystal structures of KF and *Bst* Pol I, we identified charged residues adjacent to the SWCNT-FET that could move in response to rotations by Y766 and F762 in the KF active site (Figure 2-7). Thus, we hypothesize that the source of $\Delta I(t)$ excursions is likely additional motions of Y766 and/or F762 after enzyme closure and base incorporation that continue to propagate to charged residues near the SWCNT-FET. Indeed, an additional KF conformational change when the nascent base pair moves to the KF post-insertion site has been observed by smFRET following successful nucleotide incorporation,[24] and is possibly the motion measured by τ_{closed}. Significant interactions imparted by aromatic active site residues could include π-π stacking with the newly formed base pair. Such a motion would assess the electronic configuration of the base

pair and interrogate the fidelity of the bond formation step without requiring hydrophilic interactions, which are altered by the dNTP analog's substitutions.

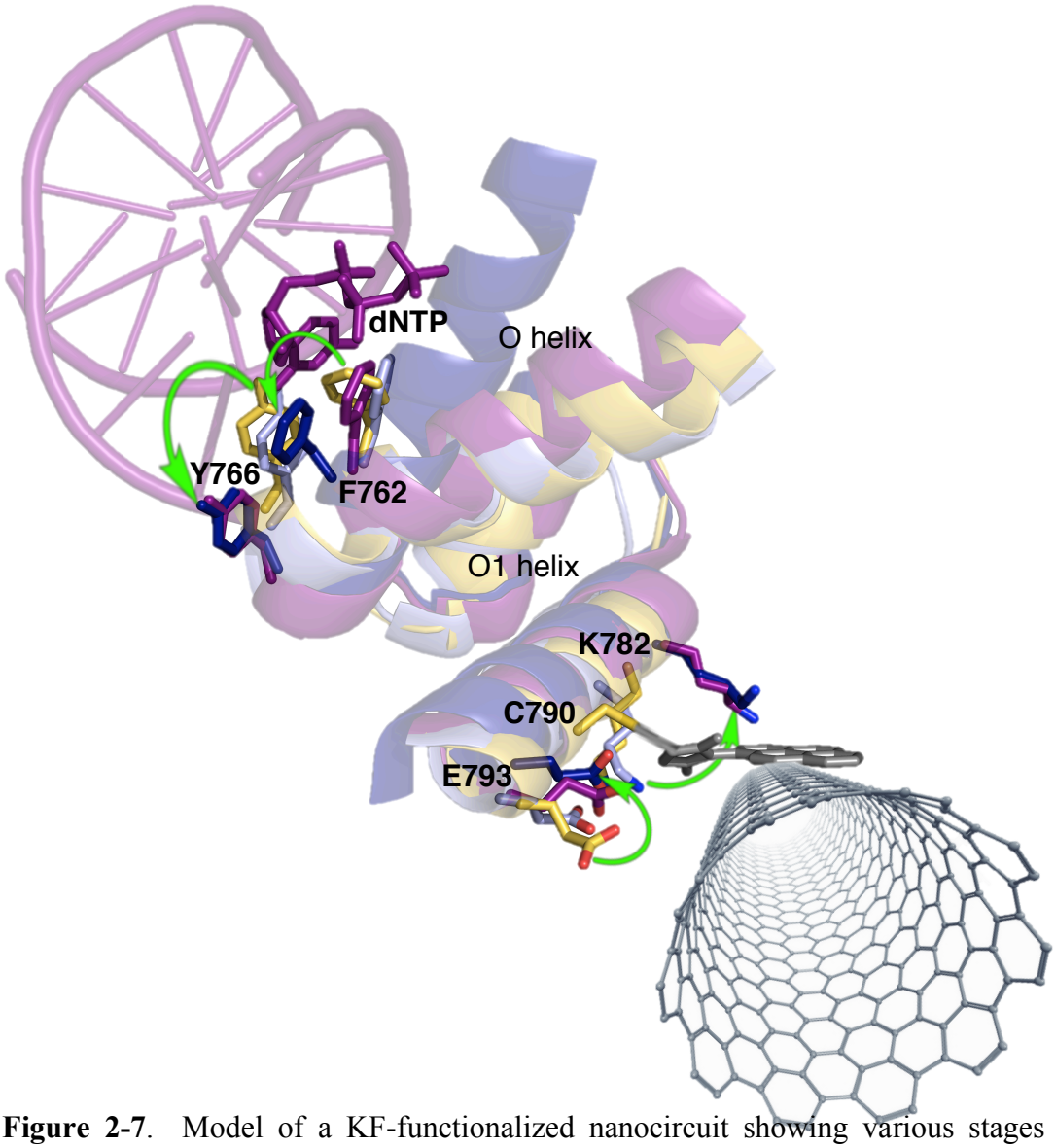


Figure 2-7. Model of a KF-functionalized nanocircuit showing various stages of dNTP incorporation. The description is continued on the following page.

Figure 2-7. The open conformation of KF (yellow, PDB ID: 1KFD) can be structurally aligned with the following structures: the homologous *Bst* Pol I fragment in its open conformation without dNTP bound (light blue, PDB ID: 1L3T), its closed conformation (dark blue, PDB ID: 1LV5), and its ajar conformation with a matched base pair (magenta, PDB ID: 3HT3). The equivalent of KF F762 makes small rotations during each stage of nucleotide incorporation shown, while the equivalent of KF Y766 moves dramatically only from the open to "ajar" state. Representative charged residues near the attachment site show variable movement that correspond to the active site residues, and likely continue to move in concert with F762 and Y766 as the enzyme closes and checks the nascent base pair's electronic structure. Green curved arrows show the direction of movement for each labeled residue from the open to closed states. Negative and positive $\Delta I(t)$ excursions likely result from these rotations during incorporation of native and non-native dNTPs, respectively.

The proposed O helix twisting mechanism explains key observations of DNA polymerases, including results from our previous single-molecule work. The KF nanocircuit reveals larger $\Delta I(t)$ excursions for the A•T/T•A set than the G•C/C•G set of base pairs. Structural results have suggested that A and T template bases are most deeply buried in the DNA polymerase active site,[21] and, therefore, the swiveling of the O helix could be maximized. KF E710 and Y766 or homologous, active site glutamate and tyrosine residues have been implicated in a mechanism for stabilization of A•T/T•A base pairs over G•C/C•G base pairs.[52] Thus, the hydrogen bonding interaction between KF E710 and KF Y766 prior to nucleotide incorporation could influence the size and shape of the active site and may play an important role in the τ_{closed} step of dNTP analog recognition. Further insights will require structural analysis, mutagenesis, and modeling.

Similar results during incorporation of 2-thio-dTTP and 2-thio-dCTP illustrate KF's preferential recognition of the base pair's electronic structure. Although the sulfur substitution only affects a Watson-Crick hydrogen bond acceptor in the 2-thio-dCTP analog, both 2-thio-dNTP analogs result in mixtures of positive and negative $\Delta I(t)$ excursions and thus both cause similar KF motions during incorporation. The sulfur substitution for the 2-thio-dNTP analogs is minor compared to the more dramatic electronic variations introduced into 6-Cl-2APTP, but the enzyme responds in a similar, although non-exclusive, manner. The observed mixtures of both negative and positive $\Delta I(t)$ excursions suggest that KF accesses both native and alternative motions, respectively, during incorporation of the 2-thio-substituted dNTPs. An apparent memory effect locks the enzyme into one motion or the other for tens of seconds, implicating an

additional conformational change that is energetically bistable in the special case of 2-thio-dNTPs.

Finally, we have also considered shuttling of the nascent DNA to the inactive exonuclease (*exo*) domain as a possible source of positive $\Delta I(t)$ excursions. Upon melting of an unstable primer terminus due to imperfect base pairing, DNA shuttles to and from an inactive *exo* domain, and KF undergoes distinct conformational changes.[24,53–55] However, such transitions occur distant from the attachment site and positive $\Delta I(t)$ excursions observed here do not change durations of $<\tau_{closed}>$. Accordingly, shuttling to the *exo* domain seems inconsistent with the observation of positive $\Delta I(t)$ excursions. Similar to the conformational steps known to occur during mismatched dNTP recognition, shuttling to the *exo* domain must take place during τ_{open}. In summary, the $\Delta I(t)$ excursions reported here only occur during a committed catalytic cycle, and likely represent an adaptable KF motion consistent with a swiveling O helix testing the electronic integrity of the newly formed DNA base pair.

2.8 Conclusion

The experiments reported here with dNTP analogs challenge the limits of nucleotide incorporation by DNA polymerases, including the stereochemistry at the electrophilic phosphate, the hydrogen bonding capability of the incoming base, and the mechanisms of fidelity checking. Since most dNTP analogs increase average $<\tau_{open}>$ and the broadness of its kinetic distributions, the rate-determining dNTP recognition step appears highly sensitive to even minor variation in substrate structure. However, dramatic substitutions at the reactive site of bond formation fail to impact the durations of $<\tau_{closed}>$. The direction of the $\Delta I(t)$ excursions, on the other hand, switches to positive or

a mixture of both negative and positive signals with base-modified dNTP analogs. Since these dNTP analogs have functionalities at the bond formation center identical to native substrates, we interpret the dramatic changes in $\Delta I(t)$ direction to result from fidelity checking by KF before opening to process the next substrate. Such events can be readily distinguished from native dNTP incorporation events and provide direct observation of the enzyme accommodating unnatural dNTPs by reversing the direction of its dynamic error checking. Such insights could lay the groundwork for a range of applications from DNA sequencing to drug design.

2.9 Materials and methods

Materials

Reagents purchased commercially include antibiotics (Fisher Scientific), Ni-IMAC resin (Bio-Rad Laboratories), cell lines (Stratagene), deoxynucleoside triphosphates (Fisher Scientific), deoxynucleoside triphosphate analogs (Trilink Biotechnologies), enzymes (New England Biolabs or Fermentas), oligonucleotides (Fisher), high-resolution agarose (The Nest Group) and 96-well fluorescence plates (Nunc). All other chemicals were purchased commercially from Acros Organics, EMD, Fisher Scientific, or Sigma Aldrich. All reagents were used as received.

Expression and purification of KF

The following expression and purification protocol applied buffers and plasmids from a previously described procedure.[33] A pET28c plasmid containing a gene encoding KF(D355A/E357A/C907S/L790C),[33,34] referred to hereafter as KF, was used to transform $CaCl_2$-competent BL21(DE3) *E. coli* cells by heat shock. Following overnight growth on solid media, a single colony was used to inoculate 25 mL LB media supplemented with

40 μg/mL kanamycin for growth in liquid media overnight at 37 °C with shaking. LB (1 L) supplemented with 40 μg/mL kanamycin was inoculated with 10 mL of the overnight culture and incubated with shaking at 37 °C for several hours. Once the cells reached late log phage (OD_{600} = 0.9), KF expression was induced by the addition of 1 mM IPTG. After 3-4 h of protein expression at 37 °C with shaking, cells were harvested by centrifugation (6000 rpm, 20 min, 4 °C) and resuspended in lysis buffer (20 mM Tris, 50 mM NaCl, 10 mM BME, pH 8.0).

Cells were lysed by sonication and the cell debris was collected by centrifugation (15,000 rpm, 45 min, 4 °C). Following filtration through a 0.45 μm pore filter, the lysate supernatant was allowed to bind to Ni-IMAC resin overnight at 4 °C. KF was eluted in the lysis buffer with 250 mM imidazole, concentrated, and then treated with TEV protease for two days at 4 °C. The mixture was centrifuged then filtered through a 0.45 μm filter prior to size exclusion chromatography in TBS (20 mM Tris, 50 mM NaCl, 100 μM TCEP, pH 7.9) on a Bio-Rad Biologic DuoFlow FPLC. KF purity was assessed by SDS-PAGE (Figure 2-2A).

Ensemble activity of KF and dNTP analog incorporation

Oligonucleotides used to test activity

Table 2-2 lists the oligonucleotides used to test KF activity, dNTP analog incorporation, and for measurements with the nanocircuit. Upon receipt, HPLC-purified oligonucleotides were solubilized in water to 100 μM. Bold regions indicate the M13 priming site. Italicized regions indicate restriction sites. [2AP] indicates 2-aminopurine.

Table 2-2. Oligonucleotides used for activity and electronic measurements

Oligonucleotide	**Sequence (5′-3′)**	**Use**
M13F	TGTAAAACGACGGCCAGT	M13 forward primer
ActAssay Template	TCGAGCTATCTCTAAAGC[2AP] GCTAACTATCGAGCTATCGCGA A**ACTGGCCGTCGTTTTACA**	Standard activity assay template containing 2-aminopurine
A/T Incorporation Assay Template	CCTAACGCAGATAGACGTTGTT TAGAGCCCGGGTCGGCCAT**ACT GGCCGTCGTTTTACA**	Test incorporation of dATP or dTTP analogs in Figure 2-5A
G/C Incorporation Assay Template	CCTAACGCAGATAGACGTTGTT TAGAGATTTAAATTCGGCC**ACT GGCCGTCGTTTTACA**	Test incorporation of dCTP or dGTP analogs in Figure 2-5B
$(A)_{42}$	$(A)_{42}$**ACTGGCCGTCGTTTTACA**	Test native and analog dTTP incorporation on nanocircuit
$(T)_{42}$	$(T)_{42}$**ACTGGCCGTCGTTTTACA**	Test native and analog dATP incorporation on nanocircuit
$(G)_{42}$	$(G)_{42}$**ACTGGCCGTCGTTTTACA**	Test native and analog dCTP incorporation on nanocircuit
$(C)_{42}$	$(C)_{42}$**ACTGGCCGTCGTTTTACA**	Test native and analog dGTP incorporation on nanocircuit

Ensemble assay for KF Activity

To confirm activity of KF(L790C) versus wild-type KF, a previously described assay was adapted as follows.[33,35] A randomized DNA template containing both 2-aminopurine (ActAssay template in Table 2-2) and an M13 priming site (bold) was annealed to the M13F primer by heating the mixture to 65 °C and slow-cooling to room temperature for 1 h. A comparable decrease in fluorescence was observed for KF(L790C) and wild-type KF (both 1 μM) upon incubation with the primer-template mixture (25 μM) and dNTPs (250 μM). The raw fluorescence data was corrected by subtraction of background, which was measured in the absence of dNTPs. The excitation and emission wavelengths employed in this experiment were 305 and 365 nm, respectively.

Ensemble assay for dNTP analog incorporation

To confirm incorporation of dNTP analogs, randomized DNA templates (Table 2-2) were polymerized by KF after hybridization to an M13F primer. Positive control reactions contained KF (1 μM), dNTPs or dNTP analogs (100 μM), and A/T or G/C incorporation template-primer (5 μM) in 10 mM Tris, 50 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT, pH 7.9. Reactions to test dNTP analog incorporation contained 100 μM analog in place of its native dNTP, and negative control reactions omitted either the analog or its native dNTP. Reactions were kept at 25 °C for 2 h in a thermal cycler before electrophoresis on a 5% high resolution agarose gel.

2.10 References

1. Echols, H. & Goodman, M. F. Fidelity mechanisms in DNA replication. *Annu. Rev. Biochem.* **60,** 477–511 (1991).

2. Kool, E. T. Active site tightness and substrate fit in DNA replication. *Annu. Rev. Biochem.* **71,** 191–219 (2002).

3. Kunkel, T. A. DNA replication fidelity. *J. Biol. Chem.* **279,** 16895–8 (2004).

4. Schweitzer, B. A. & Kool, E. T. Hydrophobic, non-hydrogen-bonding bases and base pairs in DNA. *J. Am. Chem. Soc.* **117,** 1863–1872 (1995).

5. Morales, J. C. & Kool, E. T. Efficient replication between non-hydrogen-bonded nucleoside shape analogs. *Nat. Struct. Biol.* **5,** 950–4 (1998).

6. Potapova, O. *et al.* DNA polymerase catalysis in the absence of Watson-Crick hydrogen bonds: analysis by single-turnover kinetics. *Biochemistry* **45,** 890–8 (2006).

7. Chiaramonte, M., Moore, C. L., Kincaid, K. & Kuchta, R. D. Facile polymerization of dNTPs bearing unnatural base analogues by DNA polymerase alpha and Klenow fragment (DNA polymerase I). *Biochemistry* **42,** 10472–81 (2003).

8. Sintim, H. O. & Kool, E. T. Remarkable sensitivity to DNA base shape in the

DNA polymerase active site. *Angew. Chem. Int. Ed. Engl.* **45,** 1974–9 (2006).

9. Devadoss, B., Lee, I. & Berdis, A. J. Enhancing the 'A-rule' of translesion DNA synthesis: promutagenic DNA synthesis using modified nucleoside triphosphates. *Biochemistry* **46,** 13752–61 (2007).

10. Burgers, P. M. & Eckstein, F. A study of the mechanism of DNA polymerase I from *Escherichia coli* with diastereomeric phosphorothioate analogs of deoxyadenosine triphosphate. *J. Biol. Chem.* **254,** 6889–6893 (1979).

11. Gao, J., Liu, H. & Kool, E. T. Expanded-size bases in naturally sized DNA: evaluation of steric effects in Watson-Crick pairing. *J. Am. Chem. Soc.* **126,** 11826–31 (2004).

12. Kim, T. W., Delaney, J. C., Essigmann, J. M. & Kool, E. T. Probing the active site tightness of DNA polymerase in subangstrom increments. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 15803–8 (2005).

13. Kincaid, K. *et al.* Exploration of factors driving incorporation of unnatural dNTPs into DNA by Klenow fragment (DNA polymerase I) and DNA polymerase alpha. *Nucleic Acids Res.* **33,** 2620–8 (2005).

14. Sintim, H. O. & Kool, E. T. Enhanced base pairing and replication efficiency of thiothymidines, expanded-size variants of thymidine. *J. Am. Chem. Soc.* **128,** 396–7 (2006).

15. Zhang, X., Motea, E., Lee, I. & Berdis, A. J. Replication of a universal nucleobase provides unique insight into the role of entropy during DNA polymerization and pyrophosphorolysis. *Biochemistry* **49,** 3009–23 (2010).

16. Betz, K. & Malyshev, D. *et al.* KlenTaq polymerase replicates unnatural base pairs by inducing a Watson-Crick geometry. *Nat. Chem. Biol.* **8,** 612–4 (2012).

17. Wang, W., Wu, E. Y., Hellinga, H. W. & Beese, L. S. Structural factors that determine selectivity of a high fidelity DNA polymerase for deoxy-, dideoxy-, and ribonucleotides. *J. Biol. Chem.* **287,** 28215–26 (2012).

18. Carroll, S. S., Cowart, M. & Benkovic, S. J. A mutant of DNA polymerase I (Klenow fragment) with reduced fidelity. *Biochemistry* **30,** 804–813 (1991).

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

20. Hohlbein, J. *et al.* Conformational landscapes of DNA polymerase I and mutator derivatives establish fidelity checkpoints for nucleotide insertion. *Nat. Commun.* **4,**

2131 (2013).

21. Bermek, O., Grindley, N. D. F. & Joyce, C. M. Prechemistry nucleotide selection checkpoints in the reaction pathway of DNA polymerase I and roles of glu710 and tyr766. *Biochemistry* **52,** 6258–74 (2013).

22. Min, W. *et al.* Fluctuating enzymes: lessons from single-molecule studies. *Acc. Chem. Res.* **38,** 923–31 (2005).

23. Lu, H. P. Sizing up single-molecule enzymatic conformational dynamics. *Chem. Soc. Rev.* **43,** 1118–43 (2014).

24. 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).

25. Santoso, Y. & Joyce, C. M. *et al.* Conformational transitions in DNA polymerase I revealed by single-molecule FRET. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 715–20 (2010).

26. 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).

27. Berezhna, S. Y., Gill, J. P., Lamichhane, R. & Millar, D. P. Single-molecule Förster resonance energy transfer reveals an innate fidelity checkpoint in DNA polymerase I. *J. Am. Chem. Soc.* **134,** 11261–8 (2012).

28. Garalde, D. R. *et al.* Distinct complexes of DNA polymerase I (Klenow fragment) for base and sugar discrimination during nucleotide substrate selection. *J. Biol. Chem.* **286,** 14480–92 (2011).

29. Choi, Y. & Moody, I. S. *et al.* Single-molecule lysozyme dynamics monitored by an electronic circuit. *Science* **335,** 319–24 (2012).

30. Choi, Y. *et al.* Single-molecule dynamics of lysozyme processing distinguishes linear and cross-linked peptidoglycan substrates. *J. Am. Chem. Soc.* **134,** 2032–5 (2012).

31. Choi, Y. & Olsen, T. J. *et al.* Dissecting single-molecule signal transduction in carbon nanotube circuits with protein engineering. *Nano Lett.* **13,** 625–31 (2013).

32. Sims, P. C. *et al.* Electronic measurements of single-molecule catalysis by cAMP-dependent protein kinase A. **135**, 7861-8 (2013).

33. Olsen, T. J. & Choi, Y. *et al.* Electronic measurements of single-molecule

processing by DNA polymerase I (Klenow fragment). *J. Am. Chem. Soc.* **135,** 7855–60 (2013).

34. Derbyshire, V. *et al.* Genetic and crystallographic studies of the 3',5'-exonucleolytic site of DNA polymerase I. *Science* **240,** 199–201 (1988).

35. 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,** 9185–9192 (1995).

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

37. 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,** 4835–4843 (1991).

38. Patro, J. N., Urban, M. & Kuchta, R. D. Role of the 2-amino group of purines during dNTP polymerization by human DNA polymerase alpha. *Biochemistry* **48,** 180–9 (2009).

39. Politzer, P., Lane, P., Concha, M. C., Ma, Y. & Murray, J. S. An overview of halogen bonding. *J. Mol. Model.* **13,** 305–11 (2007).

40. Biedermann, M., Hartung, H., Dölling, W. & Verjus, P. (R)- and (S)-2-(4-bromophenyl)-2-oxoethyl 2-methoxymethylpyrrolidine-1-dithiocarboxylate. *Acta Crystallogr. Sect. C Cryst. Struct. Commun.* **54,** 507–509 (1998).

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

42. Tsai, Y.-C. & Johnson, K. A. A new paradigm for DNA polymerase specificity. *Biochemistry* **45,** 9675–87 (2006).

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

44. Sismour, A. M. & Benner, S. A. The use of thymidine analogs to improve the replication of an extra DNA base pair: a synthetic biological system. *Nucleic Acids Res.* **33,** 5640–6 (2005).

45. Leslie, A. G. W., Arnott, S., Chandrasekaran, R. & Ratliff, R. L. Polymorphism of DNA double helices. *J. Mol. Biol.* **143,** 49–72 (1980).

46. Eckstein, F. & Jovin, T. M. Assignment of resonances in the phosphorus-31 nuclear magnetic resonance spectrum of poly[d(A-T)] from phosphorothioate substitution. *Biochemistry* **22,** 4546–4550 (1983).

47. Mizrahi, V., Henrie, R. N., Marlier, J. F., Johnson, K. A. & Benkovic, S. J. Rate-limiting steps in the DNA polymerase I reaction pathway. *Biochemistry* **24,** 4010–4018 (1985).

48. Ludwig, J. & Eckstein, F. Rapid and efficient synthesis of nucleoside 5'-*O*-(1-thiotriphosphates), 5'-triphosphates and 2',3'-cyclophosphorothioates using 2-chloro-4*H*-1,3,2-benzodioxaphosphorin-4-one. *J. Org. Chem.* **54,** 631–635 (1989).

49. 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).

50. Trostler, M. *et al.* Discrimination between right and wrong purine dNTPs by DNA polymerase I from *Bacillus stearothermophilus*. *Biochemistry* **48,** 4633–41 (2009).

51. Wu, E. Y. & Beese, L. S. The structure of a high fidelity DNA polymerase bound to a mismatched nucleotide reveals an 'ajar' intermediate conformation in the nucleotide selection mechanism. *J. Biol. Chem.* **286,** 19758–67 (2011).

52. Donlin, M. J. & Johnson, K. A. Mutants affecting nucleotide recognition by T7 DNA polymerase. *Biochemistry* **33,** 14908–14917 (1994).

53. Joyce, C. How DNA travels between the separate polymerase and 3'-5'-exonuclease sites of DNA polymerase I (Klenow fragment). *J. Biol. Chem.* **264,** 10858–10866 (1989).

54. Tuske, S., Singh, K., Kaushik, N. & Modak, M. J. The J-helix of *Escherichia coli* DNA polymerase I (Klenow fragment) regulates polymerase and 3'-5'-exonuclease functions. *J. Biol. Chem.* **275,** 23759–68 (2000).

55. 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).

Chapter 3: Direct Measurement of DNA Lengths Without Amplification Using Single-Molecule DNA Polymerase I (Klenow Fragment) Nanocircuits

3.1 Abstract

In order for a polymerase-based technique to be viable for biotechnological applications, DNA template lengths must be accurately counted. Single-molecule electronic biosensors functionalized with DNA polymerases are particularly advantageous for such applications, as these devices directly monitor polymerization with repetitive DNA templates. Here, single-walled carbon nanotube field-effect transistors (SWCNT-FETs) were functionalized with single-molecules of DNA polymerase I Klenow Fragment (KF), and incubated with DNA templates of varied lengths and repeat motifs. KF nanocircuits accurately counted the number of bases in various repetitive and randomized DNA templates. In mixtures of templates with lengths that differed by a single repeat motif, KF nanocircuits identified the presence of the two different templates. Higher bandwidth experiments measuring DNA polymerization of homopolymeric templates revealed the presence of additional enzyme dynamics that could complicate analysis. Thus, a limited time resolution is required to identify single base incorporations for the accurate measurement of DNA template lengths. Further measurements of DNA polymerization at higher time resolutions could be more directly applicable to providing insight into transient KF motions.

3.2 Introduction

Observation of enzyme catalysis at the single-molecule level can resolve important features of individual turnover events otherwise hidden in a heterogeneous population of enzymes.[1–5] Specifically, the variable dynamics of enzymatic efficiency,[6,7] processivity,[8–10] and kinetics[7,11–13] can be examined through observation of individual molecules; furthermore, details about intramolecular distances and forces,[14–16] molecular orientation,[17,18] substrate binding and release mechanisms,[19–21] and substrate preferences[9,22,23] can be revealed. To achieve this, fluorescent and electronic single-molecule techniques often observe the dwell times of various enzyme states during catalysis.[13,24–26] For example, a recent electronic technique functionalizes single-walled carbon nanotube field-effect transistors (SWCNT-FETs) with individual enzyme molecules.[6,8,9,27–29] Using this technique, enzyme catalysis can be observed in the presence of the required substrates and cofactors by tracking the enzyme's conformational states.

Experiments with enzyme-functionalized SWCNT-FETs, or electronic nanocircuits, highlight the technique's versatile ability to gain new information about both distributive and processive enzyme catalysis. Nanocircuits functionalized with cAMP-dependent protein kinase A (PKA), for example, measured isolated phosphoryl transfer events and exposed PKA as an imperfect enzyme. Specifically, PKA efficiently and directly phosphorylated its substrate without initially transitioning between its apo and intermediate states 77% of the time.[6] Analogous experiments with T4 lysozyme-functionalized nanocircuits measured processive events of this enzyme and demonstrated its discrimination between two closely related substrates. Lysozyme hydrolyzed linear

peptidoglycan almost twice as fast as the wild-type cross-linked peptidoglycan due to less non-productive motion with the linear substrate.[9] Electronic monitoring of single enzymes with SWCNT-FETs also offers fast time resolution and long duration measurements. Such properties can shed new light onto the conformational dynamics of processive enzyme catalysis and the number of successive bonds processed.

Recently, DNA Polymerase I Klenow Fragment (KF), a moderately processive and well-characterized polymerase from *E. coli*, was engineered for SWCNT-FET attachment.[28,29] In KF nanocircuits, DNA polymerization in the presence of limited template concentrations was diffusion-dependent and paused to allow binding of the next template. The pauses represented the idle enzyme waiting for the next round of processive catalysis and clearly separated clusters of polymerase activity. In the presence of deoxynucleoside triphosphates (dNTPs) and other necessary cofactors, 42 catalytic events were distinctly resolved for a homopolymeric DNA template containing 42 unpaired nucleotides.

DNA templates with repetitive sequences challenge accurate replication by polymerases and can lead to problems in the analysis of their amplified fragments.[30–32] Specifically, the human genome includes stretches of repeat motifs,[33–36] which are susceptible to inaccurate replication. Errors during their synthesis can lead to disease and inaccurate analysis of these target regions of DNA.[37–40] Thus, techniques to examine these significant areas of the genome must eliminate amplification. The single-molecule base pair resolution achieved with this technique can allow individual template identification by direct observation of template polymerization with unamplified templates. By combining observation of processive catalysis with separation of

distributive events, mixtures of templates with various numbers of repetitive motifs can be distinguished.

Here, in experiments performed with Dr. O. Tolga Gul and Dr. Yongki Choi, we have observed the polymerization of DNA templates containing designed or biologically significant repeat motifs with KF-functionalized nanocircuits. Thorough analysis was applied to the electronic processing of a lone template to obtain a narrow Gaussian distribution centered on the number of possible catalytic events, or unpaired nucleotides. Two templates differing by a single tetranucleotide repeat motif were distinguished in a mixture using this technique. Cumulative long duration measurements from multiple mixtures equaling 1 h of data provided more apparent depictions of template separation and demonstrated that bin size is a key determinant of resolution for this analysis. Interestingly, experiments with a high bandwidth microfluidic flow cell designed by Dr. Max Akhterov[41,42] observed shorter duration events dispersed between base incorporations. The observation of additional events in higher-resolution experiments demonstrates the requirement for limited resolution of DNA polymerization to accurately count the number of base incorporations. This higher-resolution technique also showed that the transition of KF from "open" to "closed" and "closed" to "open" occurred rapidly and could not be fully resolved.

3.3 Experimental section

A fluorescence-based ensemble assay confirmed activity of a purified single-cysteine variant of exonuclease-deficient KF (Figure 2-1). Prior to KF functionalization, SWCNT-FETs were fabricated and immersed in a solution of *N*-(1-pyrenyl)maleimide (1 mM in ethanol, 30 min). For KF attachment (Figure 3-1A), SWCNT-FETs were then

incubated with 300 nM KF in 20 mM Tris, 50 mM NaCl, 10 mM $MgCl_2$, 100 μM TCEP, pH 8.0, a standard DNA Polymerase I activity buffer. Atomic force microscopy confirmed single KF molecule attachment to each device (Figure 3-1B).

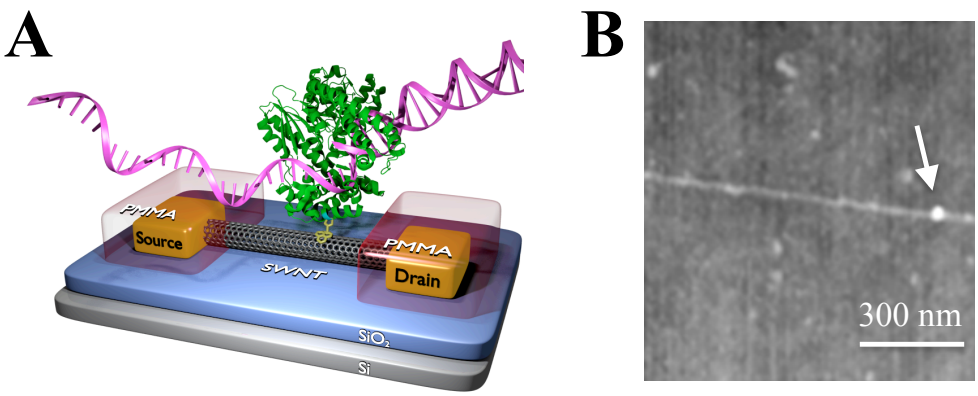


Figure 1. Single-molecule attachment of DNA Polymerase I Klenow Fragment (KF) to SWCNT-FETs. a) Schematic representation of a single KF molecule attached to a carbon nanotube. b) Single-molecule attachment was confirmed using atomic force microscopy.

Counting experiments used heteropolymeric templates containing 2 components fused together: an M13 priming site and 36 to 44 overhanging bases with different repeat sequences. The M13 priming site of each template was hybridized to an M13 forward primer in a 1:1 stoichiometric ratio by heating to 95 °C for 5 to 10 minutes followed by slow-cooling to room temperature. Following hybridization, overhanging bases for each DNA template were composed of the following repeats: $(A_3C_6)_4$, $(C_3G_6)_4$, $(B_{10}A)_4$, $(CTTT)_9$, $(CTTT)_{10}$, or $(CTTT)_{11}$, in which B is the standard IUB code used to represent either G, T, or C. To measure catalytic activity, KF nanocircuits were soaked in activity buffer with 0.5 nM primer-template and 10 μM native complementary dNTPs (Fisher).

For measurements, the source-drain current $I(t)$ through the SWCNT-FET was monitored during KF catalysis. The electrolyte, or gate source, was held at or near 0 V while the source-drain electrode was biased at 100 mV. Current fluctuations $\Delta I(t)$ were absent in control experiments when the device was missing either the KF attachment, the template-primer, or complementary dNTPs. A Keithley 428 potentiostat amplified currents of 5 to 50 nA with a bandwidth of 40 kHz, and $\Delta I(t)$ fluctuations were digitized at 100 kHz. Experiments with increased time resolution achieved bandwidth of 1 MHz. Each experiment's measurements were stored as 600 s data sets for future analysis. Before a new measurement, the KF nanocircuits were extensively rinsed with activity buffer, incubated in buffer for 10 minutes, and then rinsed again with buffer. Each KF molecule was monitored with multiple templates and the corresponding natural dNTPs.

3.4 Templates and template concentrations used in the experiments

The hybridized primer-templates used in these experiments were designed to assess different challenging aspects of KF-catalyzed DNA polymerization. For example,

two example repeat motifs used here alternated between two nucleotides; the $(C_3G_6)_4$ template contains a G/C rich stretch of DNA and the $(A_3C_6)_4$ template introduces the slower incorporation of dTTP compared to dGTP. Another example motif tested here offers a random sequence of $(B_{10}A)_4$, in which the B_{10} represents a random sequence of G, T, and C. Finally, three additional templates were designed with 9, 10, or 11 repeats of the sequence CTTT, a biologically relevant repeat sequence located on chromosome 4 of the human alpha fibrinogen gene. Such an assortment of repeat motifs can possibly highlight any discrepancies arising through polymerization of highly repetitive templates. DNA primer-template concentrations were optimized to ensure clear separation of the individual template reads when applied to KF-functionalized nanocircuits. Polymerization of all templates measured here applied template-primer concentrations 1 order of magnitude (0.1 nM) below the DNA dissociation constant for KF (1-5 nM).[43] This lower substrate concentration produced a typical duration of idle enzyme from 10 to 100 s between clusters of $\Delta I(t)$ excursions during KF-catalyzed polymerization. Average rates of polymerization, *k,* observed for repetitive templates tested here (i.e., 14.2 s^{-1}) agreed favorably with our previous analysis. Since a 10 s data set typically contained $\Delta I(t)$ excursions for only one template molecule, multiple 600 s measurements were acquired and accumulated to achieve significant statistics.

3.5 Discrimination guidelines for base incorporation

Distinct criteria were used to identify the $\Delta I(t)$ excursions belonging to a single template read. Each template molecule was typically read within a 10 s window and the characteristic cluster of $\Delta I(t)$ excursions lasted 2 to 6 s for the of 36 to 44 overhanging bases reported here. For example, a DNA template containing 44 bases with 11 repeats

of the CTTT motif required a roughly 3 s duration. For our analysis, we first segmented template reads into single second durations containing various numbers of individual $\Delta I(t)$ excursions (Figure 3-2A). Durations of baseline current, τ_{zero}, between $\Delta I(t)$ excursions belonging to the same template molecule determine the average rate of catalysis and last an average of ~70 ms for the $(CTTT)_{11}$ template. Similarly, every template examined has an average τ_{zero} value of < 0.5 s. Consequently, any $\Delta I(t)$ excursions separated from a cluster by a baseline current of 0.5 s or longer were not considered to belong to the template (orange, Figure 3-2A).

Establishment of a true $\Delta I(t)$ excursion must meet criteria that include noise discrimination and a minimum number of data points, determined by our 100 kHz data acquisition. The single second duration in Figure 3-2A containing 14 current excursions is magnified in Figure 3-2B to demonstrate the first criterion. The corresponding binary output is identified as a horizontal green line with the level finding algorithm and overlaid on the observed signals. Each transistor applied in our experiments is assigned a current threshold dependent upon noise level, which must be surpassed by the $\Delta I(t)$ excursion in question (i.e., the cyan horizontal dashed line at -3 nA in Figure 3-2B). For the second criterion of a required data point minimum, a digital discriminator excluded high frequency noise and $\Delta I(t)$ excursions above 10 kHz. True $\Delta I(t)$ excursions contained at least 10 data points that fell below the current threshold indicated, lasting over a 100 μs duration. Figure 3-2C demonstrates an example $\Delta I(t)$ excursion that did not meet the 100 μs duration criterion, and is excluded despite being present among 6 real events in Figure 3-2B. Transient signals that satisfied the criterion, exceeding both the set noise threshold and required number of data points, were representative of a single base incorporation

event (Figure 3-2D). Flexible or intermediate KF motions likely induce small-amplitude and short-duration events. Further investigation of such signals will require higher bandwidths and lower background noise to determine if and how they correlate to KF conformational dynamics.

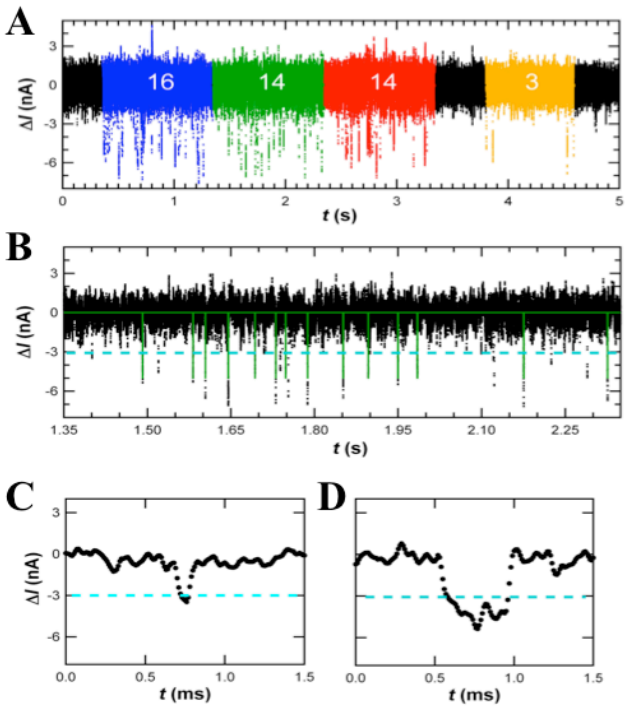


Figure 3-2. Criteria for identification of individual $(CTTT)_{11}$ template molecules. (**A**) Clusters of $\Delta I(t)$ excursions that represent polymerization of a single template molecule can be analyzed by segmenting into single second durations (blue, green, and red). Current excursions taking place >0.5 s after a previous excursion are not considered to be part of the same template (orange). (**B**) The green segment in (**A**) containing 14 excursions is magnified and a threshold current is set for each individual device (-0.3 nA here, dashed line in cyan). (**C**) Current excursions are not considered a true base incorporation event if they do not fall below the set threshold or if they contain <10 data points below the threshold. (**D**) An excursion meeting both criteria are considered a true base incorporation event.

3.6 Determination of template lengths from single oligonucleotides and mixtures

Single templates or stoichiometric ratios of two differently sized templates with variable repeating units were next examined using KF nanocircuits (Figure 3-3). Figure 3-3A shows data collected over a longer time period than Figure 3-2A with the same $(CTTT)_{11}$ template. In 3-3A, two 6 s clusters of $\Delta I(t)$ excursions are separated by 2 s of baseline current. Histograms of single templates were fitted to a single Gaussian to determine average template lengths with a full-width half maximum (FWHM) (Figure 3-3B,C). In some cases, multiple clusters of polymerase activity can be observed within a 10 s data set. For example, within 8 s, processing of both $(CTTT)_9$, a 36-mer template, and $(CTTT)_{10}$, a 40-mer template, were observed upon incubation of KF nanocircuits with a 1:1 molar ratio of the two templates (Figure 3-3D). Histograms of this and other mixtures demonstrated distinction by the device of two templates differing by 4 to 8 bp (Figure 3-3E,F). Accurate counting of DNA templates required a histogram bin size of 2 bp to allow separation when analyzing a small number of total counts. In mixtures containing equal concentrations of $(CTTT)_{10}$ with $(CTTT)_{11}$, the shorter template appeared to be counted more frequently. The experimental bias is likely indicative of an upper limit of KF processivity that is being approached with templates having > 40 overhanging bases.

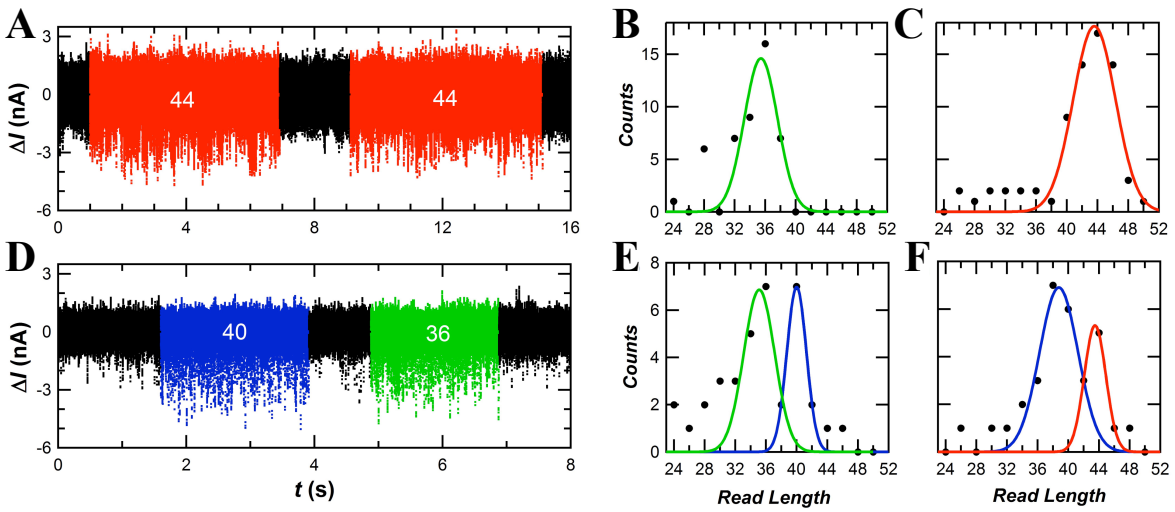


Figure 3-3. Determination of DNA alone or in a mixture. (**A**) A representative data set of a single template, $(CTTT)_{11}$, is shown with regions of template processing highlighted in red. Distributions of read lengths and their respective counts are fit to a Gaussian curve to determine the peak position, or DNA size. For example, (**B**) a fit to the distributions of read lengths for the template $(A_3C_6)_4$ resulted in a peak position of 35.5 ± 0.7 and (**C**) a similar analysis for $(CTTT)_{11}$ resulted in a peak position at 43.6 ± 0.2. (**D**) Two templates containing different numbers of the same repeat, $(CTTT)_9$ + $(CTTT)_{10}$, can both be observed in data sets when mixed in a 1:1 ratio and applied to KF nanocircuits. Distributions for each templates were individually fitted to a Gaussian to determine the DNA sizes in solution, as shown for the mixtures of (**E**) $(CTTT)_9$ + $(CTTT)_{10}$ or (**F**) $(CTTT)_{10}$ + $(CTTT)_{11}$.

3.7 Reduction of peak position uncertainty and systematic undercounting

Peak positions and full-width half-maximums (FWHMs) for single templates and mixtures measured here are summarized in Table 3-1. Limited event occurrences and the 2 bp histogram bin size for analysis causes histogram peak positions to systematically undercount the template length by an average of -0.6 bp. However, even these relatively short data collection times (~10 min) resulted in minor peak position uncertainties, typically 0.2 to 0.3 bp, and FWHM error values < 1 bp. Collecting data for longer periods provides more template-associated clusters of $\Delta I(t)$ excursions, and we hypothesized that redundant reads can improve resolution by reducing uncertainty.

Length (base)	Template	Peak Position (bp)	FWHM (bp)	Bin Size
36	$(A_3C_6)_4$	35.5 ± 0.7	6.2 ± 1.0	2
	$(C_3G_6)_4$	35.3 ± 0.2	5.5 ± 0.8	2
40	$(CTTT)_{10}$	40.0 ± 0.2	3.6 ± 0.4	2
	$(CTTT)_{10}$	38.8 ± 0.3	7.1 ± 1.0	2
44	$(CTTT)_{11}$	43.6 ± 0.2	7.8 ± 0.7	2
	$(CTTT)_{11}$	43.3 ± 0.3	6.4 ± 0.9	2
	$(B_{10}A)_4$	43.5 ± 0.3	4.0 ± 1.0	2
36 & 40	$(CTTT)_9$ & $(CTTT)_{10}$	34.5 ± 1.1 & 40.0 ± 0.2	10.1 ± 3.6 & 3.6 ± 0.4	2
40 & 44	$(CTTT)_{10}$ & $(CTTT)_{11}$	38.8 ± 0.3 & 43.5 ± 0.3	7.1 ± 1.0 & 4.0 ± 1.0	2
36 & 44	Varying sequences	35.33 ± 0.25 & 43.26 ± 0.26	5.5 ± 0.7 & 6.4 ± 0.8	2
		35.84 ± 0.16 & 43.83 ± 0.06	4.8 ± 0.4 & 5.5 ± 0.1	1

Table 3-1. Fitting parameters to peak positions and FWHMs for single templates and mixtures.

Accumulation of several data sets, which compile a large number of total counts, allows for improved analysis of template mixtures (Figure 3-4). Data sets of template mixtures compiled for Figure 3-4 contain varying sequences of 36 and 44 bases. The histograms represent an extended data set containing nearly 1000 template reads over 200 minutes of single-molecule signals. Applying the same analysis used for previous mixtures (bin size = 2 bp), systematic undercounting of template lengths was not improved by simply increasing the number of counts (Figure 3-4A). However, the large number of counts allows for greater precision of template lengths in the mixture by permitting a smaller bin size. When the bin size = 1 bp, Gaussian peaks are fully separated and the systematic undercounting is reduced from -0.6 bp to -0.2 bp (Figure 3-4B). This analysis also results in less peak position uncertainty and smaller FWHM values and errors. Peak positions for the mixtures are centered around 35.84 ± 0.16 and 43.83 ± 0.06 bp. The FWHM values of the peaks are reduced to 4.8 ± 0.4 and 5.5 ± 0.1 bp, respectively. Consequently, we are confident that bin size is the foremost contribution to the errors stated in Table 3-1.

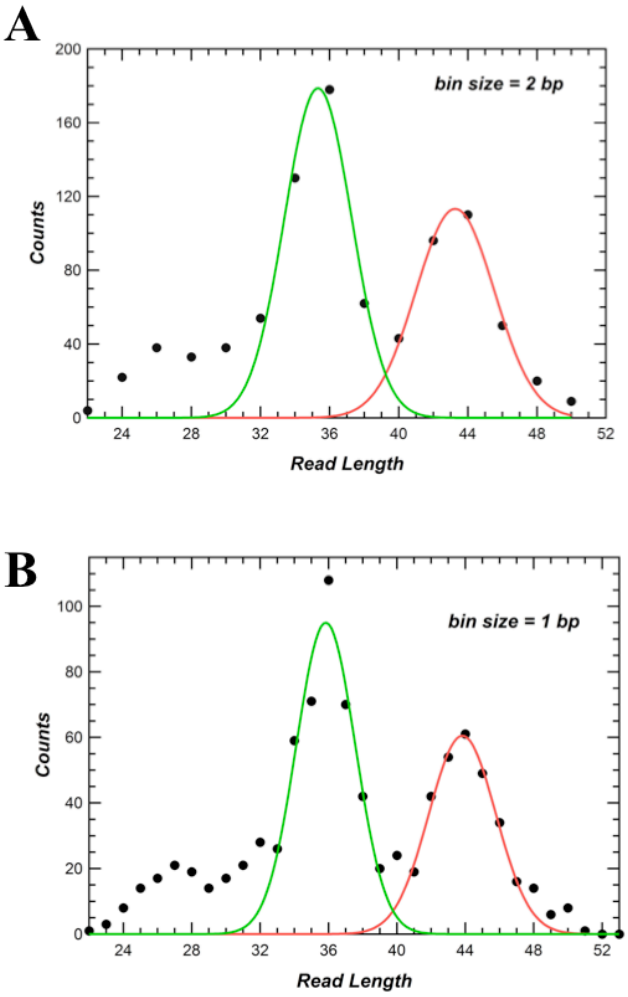


Figure 3-4. Accumulation of data sets from multiple mixtures containing the same combination of DNA lengths. Data was compiled from measurements with $(A_3C_6)_4$ + $(CTTT)_{11}$, $(C_3G_6)_4$ + $(CTTT)_{11}$, $(A_3C_6)_4$ + $(B_{10}A)_4$, and $(C_3G_6)_4$ + $(B_{10}A)_4$. Next, (**A**) the peaks were individually fitted to a single Gaussian and plotted with a bin size of 2 bp to determine peak position. (**B**) Peak positions can be further resolved by plotting with a bin size of 1 bp.

3.8 Subdomain motions observed at higher time resolutions

In a set of experiments using a microfluidic flow cell with high bandwidth, several short-duration (<100 μs) $\Delta I(t)$ excursions were observed (Figure 3-5). These short duration events occurred, on average, in a 5:1 ratio with the regular duration (0.1-1 ms) $\Delta I(t)$ excursions representing base incorporation (Figure 3-5A). This observation indicates that transient KF motion, possibly related to the dynamics of the binary complex, can obscure the identification of base incorporation events in more resolved experiments. Additionally, previously unobserved dynamics during the longer-duration lower current states (>1 ms) were resolved in these experiments (Figure 3-5B). The motions observed during the longer duration events could be further evidence of induced fit by the enzyme of the nascent base pair into the correct geometry before translocation, as hypothesized previously.[29] KF transitions between "open" and "closed" current states were nearly identical to this updated technique's time resolution (~2 μs) and could not be fully resolved (Figure 3-6).

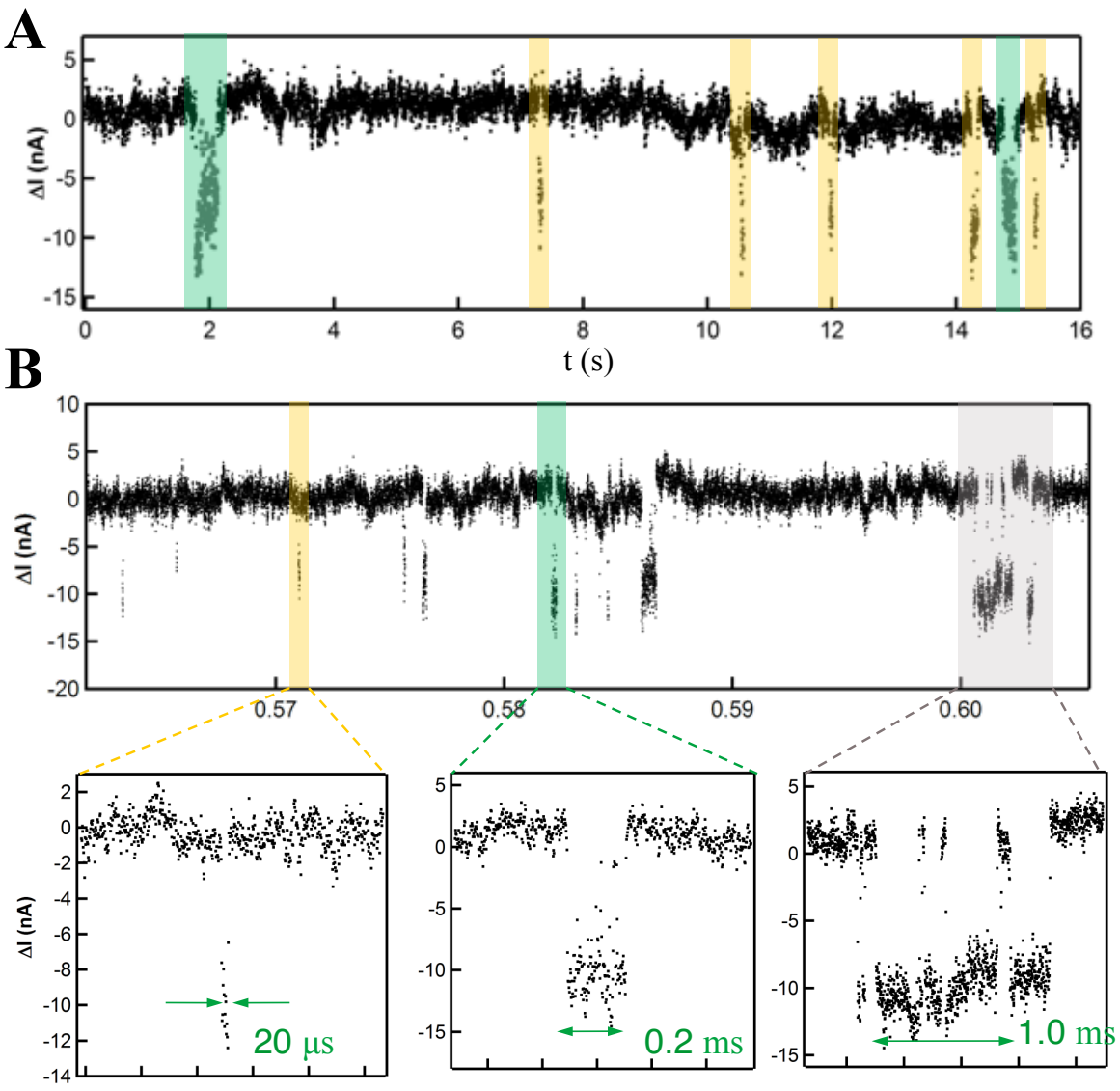


Figure 3-5. High-bandwidth measurements of KF. (**A**) Short duration events lasting <100 µs (yellow) were observed at an approximate 5:1 ratio with the normal duration events (0.1-1ms, green) in 1 MHz bandwidth measurements. (**B**) Events could be categorized as short (yellow), regular (green), and long (gray). In longer duration events, additional enzyme dynamics were observed that likely correspond to the enzyme fitting the nascent base pair into the correct geometry.

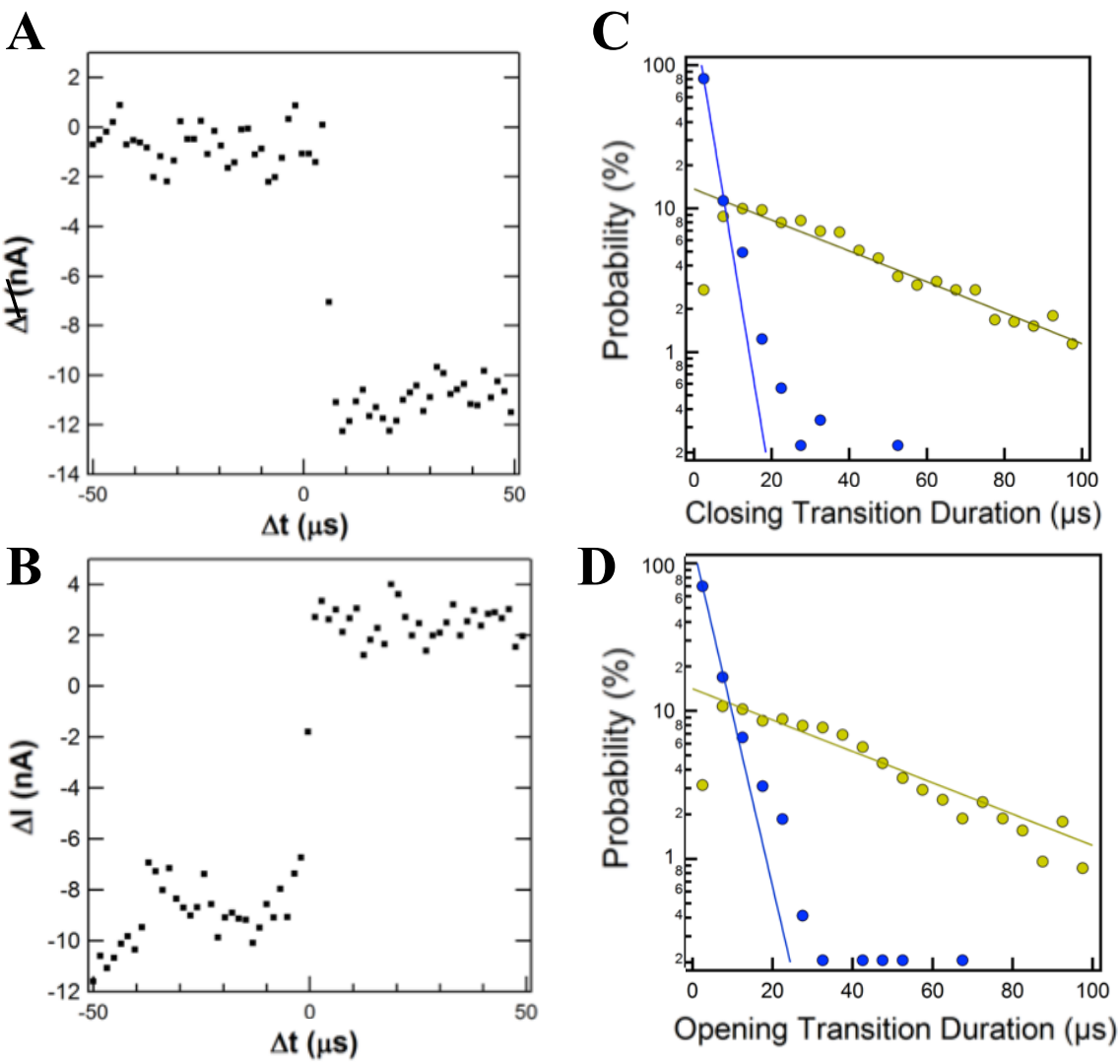


Figure 3-6. Closing and opening transitions of KF. High-bandwidth measurements were not able to fully resolve the (**A**) closing or (**B**) opening transitions of the enzyme. (**C**) The majority of events had transition durations from "open" to "closed" of 3 ± 2 μs compared to lysozyme's transition duration of 41 ± 5 μs. (**D**) "Closed" to "open" transitions of KF were similar to the reverse transition at 4 ± 2 μs compared to lysozyme's opening transition of 38 ± 2 μs.[41]

3.9 Conclusion

In summary, we have demonstrated an important application of SWCNT-FETs bearing single KF molecules, which could readily be extended to other processive enzymes. In electronic measurements with KF nanocircuits, strict guidelines set forth here regard permissible $\Delta I(t)$ excursions as associated with single base pair incorporation events. Such criteria are shown to accurately identify the number of overhanging bases in a DNA template when alone in solution or mixed with a template differing by 4 or more bases. Furthermore, templates can be accurately distinguished by length regardless of sequence composition. Longer duration measurements and accumulation of multiple data sets are required to produce a more precise examination of the template lengths. The results and analysis described here represent a new method for identification of substrates in a mixture through observation of enzyme catalysis.

The time resolution of measurements is essential for accurate base counting. In a technique that applied a higher bandwidth, short duration events that likely corresponded to transient enzyme motion occurred approximately five times as often as true base incorporations. Transient dynamics were also observed within the longest duration events. Analysis of KF transitions could not fully resolve the duration of these motions. Nevertheless, additional experiments using this updated, high-bandwidth technique can be extremely useful to better understand KF motion throughout polymerization, but require enhanced signal to noise ratios.

3.10 Materials and methods

Materials

Reagents purchased commercially include antibiotics (Fisher Scientific), Ni-IMAC resin (Bio-Rad Laboratories), cell lines (Stratagene), deoxynucleoside triphosphates (Fisher Scientific), oligonucleotides (Fisher), and 96-well fluorescence plates (Nunc). All other chemicals were purchased commercially from Acros Organics, EMD, Fisher Scientific, or Sigma Aldrich. All reagents were used as received.

Expression and purification of KF

The following expression and purification protocol applied buffers and plasmids from a previously described procedure.[28,29] A pET22b plasmid containing a gene encoding KF(D355A/E357A/C907S/L790C),[28,29,44] referred to hereafter as KF, was used to transform $CaCl_2$-competent BL21(DE3) *E. coli* cells by heat shock. Following overnight growth on solid media, a single colony was used to inoculate 25 mL LB media supplemented with 50 μg/mL carbenicillin for growth in liquid media overnight at 37 °C with shaking. LB (1 L) supplemented with 50 μg/mL carbenicillin was inoculated with 10 mL of the overnight culture and incubated with shaking at 37 °C for several hours. Once the cells reached late log phage (OD_{600} = 0.9), KF expression was induced by the addition of 1 mM IPTG. After 3-4 h of protein expression at 37 °C with shaking, cells were harvested by centrifugation (6000 rpm, 20 min, 4 °C) and resuspended in lysis buffer (20 mM Tris, 50 mM NaCl, 10 mM BME, pH 8.0).

Cells were lysed by sonication and the cell debris was collected by centrifugation (15,000 rpm, 45 min, 4 °C). Following filtration through a 0.45 μm pore filter, the lysate supernatant was allowed to bind to Ni-IMAC resin overnight at 4 °C. KF was eluted in

the lysis buffer with 250 mM imidazole, concentrated, and then treated with TEV protease for two days at 4 °C. The mixture was centrifuged then filtered through a 0.45 μm filter prior to size exclusion chromatography in TBS (20 mM Tris, 50 mM NaCl, 100 μM TCEP, pH 7.9) on a Bio-Rad Biologic DuoFlow FPLC. KF purity was assessed by SDS-PAGE (Figure 2-1A).

Ensemble activity of KF and dNTP analog incorporation

Oligonucleotides used to test activity

Table 3-2 lists the oligonucleotides used to test KF activity and for measurements with the nanocircuit. Upon receipt, HPLC-purified oligonucleotides were solubilized in water to 100 μM. Bold regions indicate the M13 priming site. [2AP] indicates 2-aminopurine.

Table 3-2. Oligonucleotides for ensemble and single-molecule measurements of KF.

Oligonucleotide	**Sequence (5′-3′)**	**Use**
M13F	TGTAAAACGACGGCCAGT	M13 forward primer
ActAssay Template	TCGAGCTATCTCTAAAGC[2AP] GCTAACTATCGAGCTATCGCGA A**ACTGGCCGTCGTTTTACA**	Standard activity assay template containing 2-aminopurine
$(A)_{42}$	$(A)_{42}$**ACTGGCCGTCGTTTTACA**	Test native dTTP incorporation on nanocircuit
$(A_3C_6)_4$	CCCCCCAAACCCCCCAAACCCC CCAAACCCCCCAAA**ACTGGCC GTCGTTTTACA**	Test ability of device to count repetitive 36mer template
$(C_3G_6)_4$	GGGGGGCCCGGGGGGCCCGGG GGGCCCGGGGGGCCC**ACTGGC CGTCGTTTTACA**	Test ability of device to count repetitive 36mer template
$(B_{10}A)_4$	ATTCGCGTTTCATCGGCTCCCG ATTTGCGCCGGAGTTCTCGCGC **ACTGGCCGTCGTTTTACA**	Test ability of device to count randomized 44mer template
$(CTTT)_9$	CTTTCTTTCTTTCTTTCTTTCTTT CTTTCTTTCTTT**ACTGGCCGTC GTTTTACA**	Test ability of device to count repetitive 36mer template

Oligonucleotide	Sequence (5′-3′)	Use
$(CTTT)_{10}$	CTTTCTTTCTTTCTTTCTTTCTTT CTTTCTTTCTTTCTTT**ACTGGCC GTCGTTTTACA**	Test ability of device to count repetitive 40mer template
$(CTTT)_{11}$	CTTTCTTTCTTTCTTTCTTTCTTT CTTTCTTTCTTTCTTTCTTT**ACT GGCCGTCGTTTTACA**	Test ability of device to count repetitive 44mer template

Ensemble assay for KF Activity

To confirm activity of KF(L790C) versus wild-type KF, a previously described assay was adapted as follows.[28,45] A randomized DNA template containing both 2-aminopurine (ActAssay template in Table 3-2) and an M13 priming site (bold) was annealed to the M13F primer by heating the mixture to 65 °C and slow-cooling to room temperature for 1 h. A comparable decrease in fluorescence was observed for KF(L790C) and wild-type KF (both 1 μM) upon incubation with the primer-template mixture (25 μM) and dNTPs (250 μM). The raw fluorescence data was corrected by subtraction of background, which was measured in the absence of dNTPs. The excitation and emission wavelengths employed in this experiment were 305 and 365 nm, respectively.

3.11 References:

1. Xie, S. Single-molecule approach to enzymology. *Single Mol.* **2,** 229–236 (2001).

2. Smiley, R. D. & Hammes, G. G. Single molecule studies of enzyme mechanisms. *Chem. Rev.* **106,** 3080–94 (2006).

3. Deniz, A. A., Mukhopadhyay, S. & Lemke, E. A. Single-molecule biophysics: at the interface of biology, physics and chemistry. *J. R. Soc. Interface* **5,** 15–45 (2008).

4. Blank, K., De Cremer, G. & Hofkens, J. Fluorescence-based analysis of enzymes at the single-molecule level. *Biotechnol. J.* **4,** 465–79 (2009).

5. Lu, H. P. Sizing up single-molecule enzymatic conformational dynamics. *Chem. Soc. Rev.* **43,** 1118–43 (2014).

6. Sims, P. C. *et al.* Electronic measurements of single-molecule catalysis by cAMP-dependent protein kinase A. (2013).

7. Amrute-Nayak, M. & Lambeck, K.-A. *et al.* ATP turnover by individual myosin molecules hints at two conformers of the myosin active site. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 2536–41 (2014).

8. Choi, Y. & Moody, I. S. *et al.* Single-molecule lysozyme dynamics monitored by an electronic circuit. *Science* **335,** 319–24 (2012).

9. Choi, Y. *et al.* Single-molecule dynamics of lysozyme processing distinguishes linear and cross-linked peptidoglycan substrates. *J. Am. Chem. Soc.* **134,** 2032–5 (2012).

10. Szafran, M. J., Strick, T., Strzałka, A., Zakrzewska-Czerwińska, J. & Jakimowicz, D. A highly processive topoisomerase I: studies at the single-molecule level. *Nucleic Acids Res.* **42,** 7935–46 (2014).

11. Juul, S. *et al.* Detection of single enzymatic events in rare or single cells using microfluidics. *ACS Nano* **5,** 8305–10 (2011).

12. Terentyeva, T. G., Hofkens, J., Komatsuzaki, T., Blank, K. & Li, C.-B. Time-resolved single molecule fluorescence spectroscopy of an α-chymotrypsin catalyzed reaction. *J. Phys. Chem. B* **117,** 1252–60 (2013).

13. Langer, A. *et al.* Polymerase/DNA interactions and enzymatic activity: multi-parameter analysis with electro-switchable biosurfaces. *Sci. Rep.* **5,** 12066 (2015).

14. Yang, H. *et al.* Protein conformational dynamics probed by single-molecule electron transfer. *Science* **302,** 262–6 (2003).

15. Pudney, C. R. *et al.* Enzymatic single-molecule kinetic isotope effects. *J. Am. Chem. Soc.* **135,** 3855–64 (2013).

16. Jagannathan, B. & Marqusee, S. Protein folding and unfolding under force. *Biopolymers* **99,** 860–9 (2013).

17. Rosenberg, S. A., Quinlan, M. E., Forkey, J. N. & Goldman, Y. E. Rotational motions of macro-molecules by single-molecule fluorescence microscopy. *Acc. Chem. Res.* **38,** 583–93 (2005).

18. 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).

19. 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).

20. Zheng, D. & Lu, H. P. Single-molecule enzymatic conformational dynamics: spilling out the product molecules. *J. Phys. Chem. B* **118,** 9128–40 (2014).

21. Rivas-Pardo, J. A., Alegre-Cebollada, J., Ramírez-Sarmiento, C. A., Fernandez, J. M. & Guixé, V. Identifying sequential substrate binding at the single-molecule level by enzyme mechanical stabilization. *ACS Nano* **9,** 3996–4005 (2015).

22. Craig, D. B., Eggertson, M. J., Chikamatsu, M. & Horwood, C. A. Single molecule assay of *Escherichia coli* β-galactosidase using two competing substrates simultaneously, DDAO-β-D-galactoside and resorufin-β-D-galactoside. *Anal. Lett.* **44,** 1835–1841 (2011).

23. Bétous, R. *et al.* Substrate-selective repair and restart of replication forks by DNA translocases. *Cell Rep.* **3,** 1958–69 (2013).

24. Roy, R., Hohng, S. & Ha, T. A practical guide to single-molecule FRET. *Nat. Methods* **5,** 507–16 (2008).

25. Deamer, D. Nanopore analysis of nucleic acids bound to exonucleases and polymerases. *Annu. Rev. Biophys.* **39,** 79–90 (2010).

26. Ochoa, M. A., Zhou, X., Chen, P. & Loring, R. F. Interpreting single turnover catalysis measurements with constrained mean dwell times. *J. Chem. Phys.* **135,** 174509 (2011).

27. Choi, Y. & Olsen, T. J. *et al.* Dissecting single-molecule signal transduction in carbon nanotube circuits with protein engineering. *Nano Lett.* **13,** 625–31 (2013).

28. Olsen, T. J. & Choi, Y. *et al.* Electronic measurements of single-molecule processing by DNA polymerase I (Klenow fragment). *J. Am. Chem. Soc.* **135,** 7855–60 (2013).

29. Pugliese, K. M. & Gul, O. T. *et al.* Processive incorporation of deoxynucleoside triphosphate analogs by single-molecule DNA polymerase I (Klenow fragment) nanocircuits. *J. Am. Chem. Soc.* **137,** 9587–94 (2015).

30. Kunkel, T. A. Misalignment-mediated DNA synthesis errors. *Biochemistry* **29,** 8003–8011 (1990).

31. Bebenek, K., Abbotts, J., Wilson, S. & Kunkel, T. Error-prone polymerization by

HIV-1 reverse transcriptase. Contribution of template-primer misalignment, miscoding, and termination probability to mutational hot spots. *J. Biol. Chem.* **268,** 10324–10334 (1993).

32. Kunkel, T. A., Patel, S. S. & Johnson, K. A. Error-prone replication of repeated DNA sequences by T7 DNA polymerase in the absence of its processivity subunit. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 6830–4 (1994).

33. Schmid, C. W. & Deininger, P. L. Sequence organization of the human genome. *Cell* **6,** 345–358 (1975).

34. Batzer, M. A. & Deininger, P. L. Alu repeats and human genomic diversity. *Nat. Rev. Genet.* **3,** 370–9 (2002).

35. Alkan, C., Coe, B. P. & Eichler, E. E. Genome structural variation discovery and genotyping. *Nat. Rev. Genet.* **12,** 363–76 (2011).

36. de Koning, A. P. J., Gu, W., Castoe, T. A., Batzer, M. A. & Pollock, D. D. Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet.* **7,** e1002384 (2011).

37. Epplen, C., Santos, E. J., Mäueler, W., van Helden, P. & Epplen, J. T. On simple repetitive DNA sequences and complex diseases. *Electrophoresis* **18,** 1577–85 (1997).

38. Petruska, J., Hartenstine, M. J., & Goodman, M. F. Analysis of strand slippage in DNA polymerase expansions of CAG/CTG triplet repeats associated with neurodegenerative disease. *J. Biol. Chem.* **273,** 5204–5210 (1998).

39. Mirkin, S. M. Expandable DNA repeats and human disease. *Nature* **447,** 932–40 (2007).

40. Budworth, H. & McMurray, C. T. A brief history of triplet repeat diseases. *Methods Mol. Biol.* **1010,** 3–17 (2013).

41. Akhterov, M. V. *et al.* Observing lysozyme's closing and opening motions by high-resolution single-molecule enzymology. *ACS Chem. Biol.* **10,** 1495–501 (2015).

42. Akhterov, M. V. *et al.* Microsecond-resolution recording of T4 lysozyme observes a brownian ratchet. *Biophys. J.* **108,** 31a (2015).

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

44. Derbyshire, V. *et al.* Genetic and crystallographic studies of the 3',5'-exonucleolytic site of DNA polymerase I. *Science* **240,** 199–201 (1988).

45. 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,** 9185–9192 (1995).

Chapter 4: Toward the Improvement of Single-Molecule DNA Polymerase I Klenow Fragment Nanocircuits for Fundamental and Biotechnological Applications

4.1 Abstract

Single-walled carbon nanotube field-effect transistors (SWCNT-FETs) can be functionalized with single DNA polymerase I Klenow Fragment (KF) molecules to generate electronic signals that correspond to base incorporations. These devices can accurately measure DNA template lengths and generate signals with deoxynucleoside triphosphate (dNTP) analogs that are unique when compared to native dNTP incorporation. Here, experiments with untested substrates and newly generated KF variants aim to further enhance the potential of KF-functionalized nanocircuits for biotechnological applications. Incubation of KF nanocircuits with incorrect substrates or inhibitors attempted to measure the effect of these molecules on electronic signals, but require further experimentation for reproducibility. Several KF variants with different properties, ranging from altered activity and surface charge to increased processivity, were generated to elucidate the source of currently observed KF signals and/or improve these signals. A comprehensive outline for future experiments can guide further development of KF nanocircuits for fundamental and applied methodologies.

4.2 Introduction

Single-walled carbon nanotube field-effect transistors (SWCNT-FETs) can be functionalized with enzymes to create powerful tools that reveal important aspects of biocatalysis.[1–6] For devices functionalized with DNA polymerase I Klenow Fragment (KF), however, the precise nature of these signals is unclear.[3,6] As a result, such signals are interpreted with speculation, and require further examination to elucidate the conformational changes being observed. Furthermore, a full command of the basis for KF-generated electronic signals can aid in the development of the device for biotechnological applications. Thus, for both signal interpretation and biotechnological applications, the enzyme can be engineered to improve the signal to noise ratio and substrate processivity. Experiments testing different substrates with the current KF variant (L790C) or engineered KF variants are necessary to achieve these goals.

Several strategies to understand the observed signals of enzyme activity with respect to SWCNT-FETs have been planned. First, untested incorrect substrates or catalytic inhibitors will perturb the observed signal for structure-activity relationship experiments. Second, KF variants that are known to alter enzyme activity will be examined. Third, point-charge mutations of amino acid residues residues proximal to the nanocircuit will determine the residues responsible for the observed signals or introduce new signals.

Effects from incorrect substrates or inhibitors of KF on the electronic signals generated with SWCNT-FETs could elucidate the origin of these signals. Such reagents may generate unique signals like those previously seen with dNTP analogs[6] or change the kinetics of the currently observed signals. In earlier experiments with KF-functionalized

SWCNT-FETs, signals were not observed when nanocircuits were incubated with homo-oligomeric DNA and non-complementary dNTPs.[3] This result demonstrates that the observed conformational change only occurs upon formation of complementary Watson-Crick base pairs between the DNA template and incoming dNTP. However, the dependence of this change on correct deoxyribose recognition, which is known to occur sequentially after the base recognition step,[7] is unknown. The use of complementary ribonucleoside triphosphates (rNTPs) could further clarify where the observed conformational change lies along the KF reaction pathway. At high concentrations, pyrophosphate (PP_i), the byproduct of correct incorporation, is a KF inhibitor and drives the reverse pyrophosphorolysis reaction to re-generate dNTP.[8] Incubation of KF-functionalized SWCNT-FETs with excess PP_i could prevent the enzyme from re-opening, and reveal more details about the observed signals.

Comparison of dNTP incorporation with α-thio-dNTP incorporation by KF variants could determine which current state includes chemical bond formation. Mutational analysis of polymerases previously determined the amino acid residues involved in chemical catalysis. For example, mutation of KF E883 or Q849 to alanine caused the chemical step of nucleotide incorporation to become rate-limiting. This new rate-limiting step resulted in an elemental effect when dNTP incorporation was compared to incorporation of α-thio-dNTP; the replacement of oxygen with sulfur results in a less electrophilic reaction center on the phosphorus.[9] Such an effect should be distinct from the different reaction rates of dNTPs and corresponding α-thio-dNTPs previously observed with KF nanocircuits.[6] Thus, incorporation of dNTPs versus α-thio-dNTPs by a

KF E883A or Q849A variant could generate an elemental effect, and place phosphodiester bond formation in a particular current state.

Exonucleolytic cleavage of a newly incorporated dNTPs by KF could add new dynamics to the currently observed signals. The KF variant previously used in electronic measurements lacks 3′-5′ exonuclease activity due to the mutations D355A and E357A;[10] this variant allows the accurate counting of DNA template bases (see Chapter 3). Distinct conformational changes occur to move primer-template DNA to the exonuclease domain,[11–14] and these changes are potential sources of electronic noise. Newly incorporated dNTP analogs, in particular, possibly cause the DNA to undergo shuttling to the inactive exonuclease domain.[6] Subsequent exonucleolysis of dNTP analogs could generate more distinctive electronic signals with a KF variant that has restored 3′-5′ exonuclease activity.

Active site mutations of KF could probe the nature of the observed conformational states through changes in the enzyme's molecular recognition. KF residues E710, F762, and Y766 contribute to a snug fit around the nascent base pair in the active site through their involvement in base and ribose discrimination (Figure 4-1).[15–24] Mutation of any of these three residues, therefore, adversely affects the fidelity of KF.[15–21,23,24] Y766F KF increases occupancy of the closed conformation for the apo-enzyme,[23] and these dynamics in the absence of substrate could be observed by SWCNT-FETs functionalized with this variant. The KF residue K758 is thought to be involved in the conformational change of the ternary complex, which precedes phosphoryl transfer.[17] Mutation of this residue could impact the observed signals by altering the direction, speed, or magnitude. Specific results with low-fidelity KF mutants incorporating either

incorrect dNTPs or incorrect riboses provide further inspiration for electronic-based measurements.

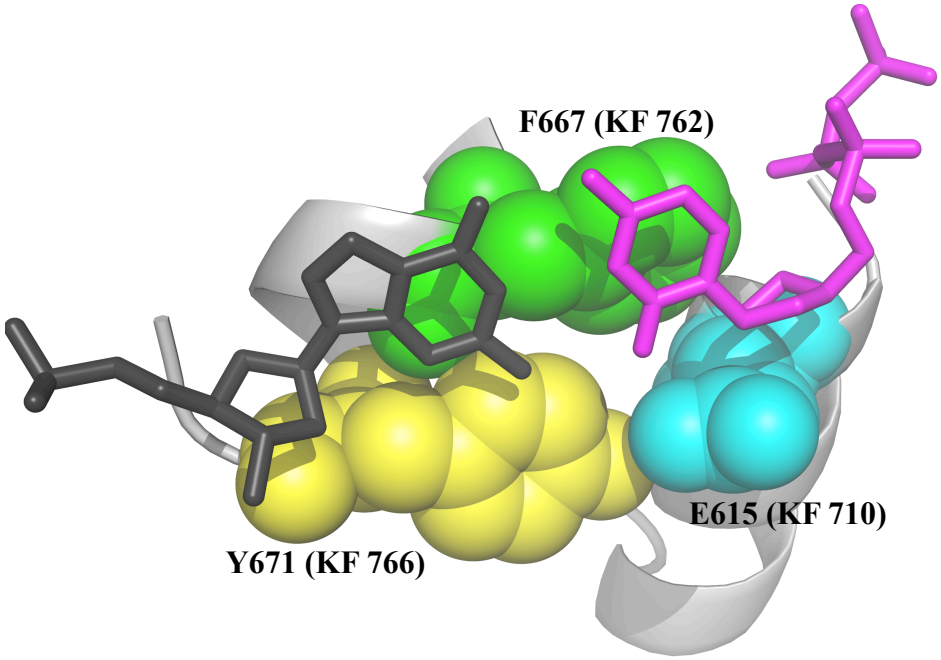


Figure 4-1. Snug fit of A-family polymerase active site in ternary complex. A crystal structure of KlenTaq (PDB ID: 3KTQ) demonstrates the contributions of KF E710 (cyan), F762 (green), and Y766 (yellow) to a snug fit of the nascent base pair. The template base is shown in dark gray and the incoming base is shown in magenta. Adapted from Potapova *et al*.[22]

Incorporation of specific non-complementary nucleotides by SWCNT-FETs functionalized with low-fidelity KF variants could alter molecular recognition or kinetic rates. For example, E710A KF has reduced fidelity for incorporation of incoming pyrimidine mismatches,[21] and Y766S KF is especially prone to misincorporation of dTTP opposite template base G.[15] Functionalization of SWCNT-FETs with either variant and subsequent incubation with these mismatched substrates could affect the direction of the observed signal, similar to the conformational change observed during dNTP analog incorporation.[6] Also, Y766S KF reduces the rate of extension past the dTTP-G mismatch[15] and Y766F KF reduces correct incorporation 5-fold.[23] Therefore, these variants could measurably and predictable impact the kinetics observed with SWCNT-FETs upon incubation with matched or mismatched substrates.

Recognition of incorrect ribose sugars by SWCNT-FETs functionalized with KF variants could alter a step of molecular recognition distinct from nucleotide discrimination. E710A KF and F762A KF are each permissive for rNTPs and use all four rNTPs to a similar degree.[19] KF E710D also permits rNTP incorporation, but, similar to wild-type KF, incorporates rCTP and rGTP more readily than rATP and rUTP. The double mutant E710A/F762Y and the point mutants E710A or F762A enable incorporation of dNTPs containing a 3′-OH instead of the correct 2′-OH.[19] KF F762 mutations allow di-deoxynucleoside triphosphate (ddNTP) incorporation[16] in two distinct ways; F762Y makes ddNTPs better substrates for incorporation and F762A makes dNTPs worse substrates.[20] Motivated by these results, SWCNT-FETs functionalized with KF variants could be incubated with specific substrates containing incorrect ribose

sugars, and attribute a particular stage of the observed electronic signals to sugar recognition.

Determination of the charges responsible for electronic signals with KF-functionalized SWCNT-FETs could lead to a better understanding of the conformational transition that is being observed. Charged amino acid residues that are on the surface of enzymes, within 1 nm of the carbon nanocircuit, and mobile during catalysis are responsible for the signals generated with KF-functionalized SWCNT-FETs.[4] Some of the residues that are potentially responsible for these signals were previously hypothesized.[6] Thorough examination of KF crystal structures determined that charged residues in close proximity to the carbon nanotube include E740, E752, K782, E783, K786, D789, and E793 (Figure 4-2A). Each residue could be mutated to the opposite charge and sequentially measured with native, complementary substrates after functionalization on the SWCNT-FET. A polarity inversion observed with a particular KF point-charge variant could implicate that amino acid residue in the generation of electronic signals.

Introduction of an unnatural charge onto a domain that undergoes a large conformational motion with respect to the SWCNT-FET could generate a novel electronic signal. An overlay of an A-family polymerase *Bacillus stearothermophilus* (BF) large fragment open crystal structure with a closed BF crystal structure depicts a large domain motion near the nanotube (Figure 4-2B). The loop of this domain has closest proximity to the nanotube and contains mostly neutral residues. L744 of this loop has historically been used in Förster Resonance Energy Transfer (FRET) assays[23,25–28] and, therefore, is a promising candidate for installation of a new charge. If the residue is

close enough to the nanotube to affect its conductivity, any new signals generated could be directly attributed to a specific motion. Furthermore, multiple current stages could reveal a clearer picture of the KF conformational landscape during dNTP incorporation.

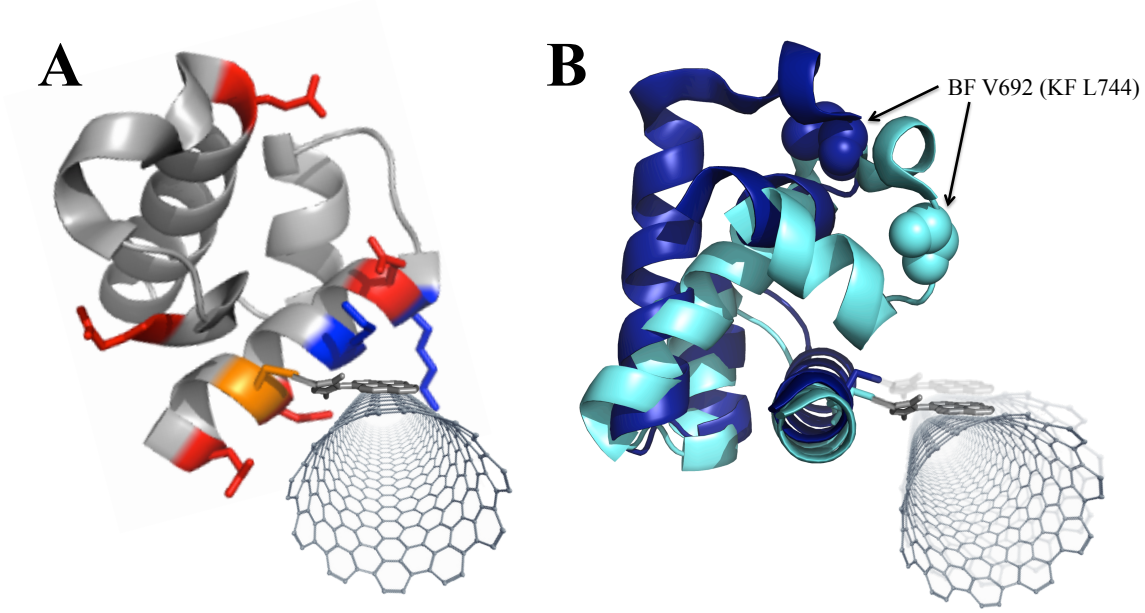


Figure 4-2. Schematic figures to depict point charge mutants of polymerases for nanotube measurements. (**A**) Positively charged residues of KF (gray, PDB ID: 1KFD) are highlighted in red and negatively charged residues are highlighted in blue. The attachment site is highlighted in orange. The charged residues appear close in proximity to the expected position of the carbon nanotube and may be responsible for the observed signals. (**B**) The open (cyan, PDB ID: 1L3T) and closed (dark blue, PDB ID: 1LV5) crystal structures of BF represent the open to closed transition of KF with respect to the carbon nanotube. Arrows show that the equivalent position of KF L744 moves further from the nanotube upon the fingers-closing transition. Installation of a positive or negative charge could introduce a new signal and provide additional information about enzyme motion with respect to our measurements.

Several additional strategies have been planned to improve the detection of KF activity with SWCNT-FETs. First, functionalization of the SWCNT-FET at different sites on the enzyme could explore different stages of dNTP incorporation. Second, introduction of a new domain could improve the overall activity of the functionalized enzyme for potential biotechnological applications. Third, installation of a large number of identical charges proximal to the carbon nanotube could increase signal significantly over background noise.

Alternative attachment sites to conjugate KF to the nanocircuit could explore different motions of the enzyme. Two particularly interesting residues to explore in this regard are KF L744 and K550. As mentioned previously, the KF residue L744 was used in several FRET-based studies due to the residue's location on the highly mobile fingers domain (Figure 4-2B). The KF residue K550 has also been used in many FRET-based studies[14,23,25,27,28] as an anchor point due to its minimal movement during the "open" to "closed" transition (Figure 4-3).[29,30] Also, due to its location at the base of the thumb domain, nanocircuits attached to KF at K550 could be more susceptible to motions that correspond to DNA shuttling to the exonuclease domain or to exonuclease activity. The use of the KF residues L744 and K550 in previous FRET studies implies minimal perturbation of enzyme activity when functionalized at these sites with a large fluorophore. Accordingly, these residues serve as excellent candidates for attachment to nanocircuits.

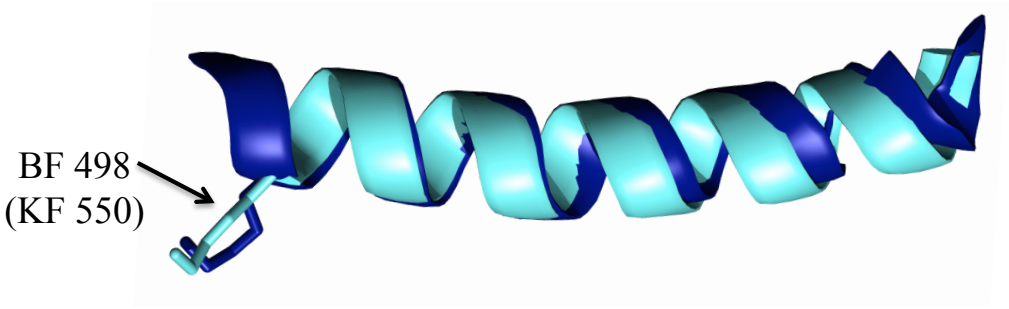


Figure 4-3. Polymerase crystal structures depicting the position of a possible new attachment site. BF open (cyan, PDB ID: 1L3T) and closed (dark blue, PDB ID: 1LV5) crystal structures represent the minimal motion of KF residue 550 during this transition.

In order to develop KF nanocircuits for biotechnological applications, the enzyme must be optimized to overcome significant limitations. One such limitation is the short read length, or processivity, of KF. Specifically, KF incorporates only 40-50 dNTPs before dissociation from the DNA primer-template.[31] Longer read lengths facilitate the determination of full DNA sequences by providing more unique sequence information per DNA fragment. Fusion to KF of the double-stranded DNA binding protein Sso7d from *Sulfolobus solfataricus* could enhance the enzyme's processivity to enable longer read lengths.[32]

A second limitation in biosensor-based measurements is the presence of background noise that can make signals difficult to interpret. As described previously, signals generated from enzyme-functionalized nanocircuits are susceptible to moving surface charges proximal to the nanotube.[4] Consequently, the presence of multiple oppositely charged amino acids may cancel out their relative gating effects and prevent a high signal to noise ratio. "Supercharging" of enzyme surfaces successfully enhances the solubility of an enzyme without affecting its activity.[33] This strategy can also be applied to the improvement of signal to noise ratios with KF-functionalized nanocircuit by the concentration of either positive or negative charge proximal to the SWCNT-FET.

All of the aforementioned strategies have been explored to different extents to understand and improve signals generated by KF nanocircuits. KF L790C was used in preliminary experiments performed with Dr. Denys Marushchak to study the effect of rNTPs or PP_i on electronic nanocircuits. Table 4-1 shows all of the KF variants that were generated and their purposes. For these variants, DNA plasmids were first generated to contain the desired mutations. Next, the variant enzymes were expressed, purified, and

characterized through fluorescence-based activity assays; Johnathon Truong completed much of this work under my direction. Sso7d-KF nanocircuit measurements were performed with Dr. O. Tolga Gul. Experimental hypotheses with electronic nanocircuits for each of the remaining variants were outlined for future work (Table 4-2).

Variants	**Purpose**
Sso7d-KF	To introduce double-stranded DNA binding domain to increase enzyme processivity
R781E/K782E/R794E/R804E E783R/E793R/E800R/E803R	To supercharge the protein surface proximal to carbon nanotube in an effort to enhance signal to noise ratio
Q849A E883A	To determine which current level includes chemical bond formation
L744C K550C	To introduce a new SWCNT attachment site
L744E L744R	To introduce charge onto a domain that shows a large motion during the open to closed transition
F762A Y766F Y766S Y766A E710A	To decrease enzyme fidelity and observe motion associated with incorrect incorporation
E740R E752R K782E E783R K786E D789R E793R	To change a charged residue proximal to the SWCNT to an opposite charge and determine whether the residue contributes to the observed signal
A355D/A357E	To re-introduce exonuclease activity

Table 4-1. Engineered KF variants and their respective purposes.

Variant	Hypothesis
R781E/K782E/ R794E/R804E E783R/E793R/ E800R/E803R	• Control measurements with poly$(dN)_{42}$ + dNTP could demonstrate increased signal to noise ratio
Q849A E883A	• The current level within which the chemical bond formation takes place could be determined by the observation of an elemental effect when α-thio-dNTP incorporation is compared to the corresponding dNTP incorporation
L744C K550C	• Control measurements with poly$(dN)_{42}$ + dNTP could transduce $I(t)$ fluctuations • Experiments with limited concentrations (0.1-1 nM) of DNA template-primer could determine whether signals correspond to base incorporations
L744E L744R	• Control measurements with poly$(dN)_{42}$ + dNTP could transduce $I(t)$ fluctuations that either enhance previously observed signals or depict an additional step • If a new signal is observed in addition to previously observed signals, experiments with limited concentrations (0.1-1 nM) of DNA template-primer could determine whether the signals correspond to single base incorporations
F762A Y766F Y766S Y766A E710A	• Control measurements with poly$(dN)_{42}$ + dNTP could transduce $I(t)$ fluctuations for variants that still retain activity. Some signals may be unique as low-fidelity mutant recognizes correct base pair as incorrect • Measurements with rNTPs, ddNTPs, and 3′-dNTP incorporation could generate unique signals corresponding to recognition of incorrect substrate and subsequent incorporation by low-fidelity mutant
E740R E752R K782E E783R K786E D789R E793R	• Control measurements with poly$(dN)_{42}$ + dNTP could transduce $I(t)$ fluctuations in the opposite direction (positive) than that observed with standard KF (negative) • Measurements with the 6-Cl-2APTP analog, which generates unique positive $\Delta I(t)$ fluctuations with KF, should generate the opposite signal and confirm contribution of the charged amino acid residue to original signal
A355D/A357E	• Control measurements with poly$(dN)_{42}$ + dNTP could transduce $I(t)$ fluctuations with altered kinetics most likely during baseline current state • Measurements with limited concentrations of DNA template-primer (0.1 nM-1 nM) could increase upper range of histogram that depicts the number of base incorporations, which would correspond to exonucleolysis and reincorporation of newly incorporated bases

Table 4-2. Experimental hypotheses for each of the engineered KF variants.

4.3 Experimental section

Preparation of Sso7d-KF and KF, functionalization of SWCNT-FETs, and measurements followed the methods outlined in Chapter 2 (2.3).[1,3,6,10,34] Experiments with KF and untested substrates used the homopolymeric DNA template $(A)_{42}$ mixed with rUTP or rTTP and PP_i. $(A)_{42}$ was fused to an M13 priming site and mixed with an M13 forward primer in a 1:1 stoichiometric ratio; the mixture was heated to 95 °C for 5 to 10 min then cooled from 65-25 °C in 5 °C increments over 40 min for hybridization. The DNA substrate to challenge Sso7d-KF processivity used the same process to anneal $(CTTT)_{26}$ fused to an FGA priming site and FGA_Fwd.

To examine processivity, Sso7d-KF nanocircuits were incubated with activity buffer containing 100 nM FGA/$(CTTT)_{26}$, 10 μM dGTP and 10 μM dATP. To test rNTP incorporation, KF nanocircuits were immersed in activity buffer with 100 nM M13/$(A)_{42}$ and 100 μM rUTP (Promega) to compensate for the enzyme's reduced affinity for rNTPs.[19] For inhibitor/pyrophorolysis experiments, sodium pyrophosphate was added to a solution of 100 nM M13/$(A)_{42}$ and 10 dTTP μM for a final PP_i concentration of 1 mM, >2-fold above the inhibitory constant.[8] Measurements testing for only pyrophosphorolysis used a sequence-randomized template (Random 45mer in Table 4-4) that was hybridized, as described above, to a complementary strand of equal length to generate double-stranded DNA. PP_i was added to 1 mM in this solution and dTTP was kept absent.

Initial measurements monitored the source-drain current, $I(t)$, through the SWCNT-FET while the attached KF or Sso7d-KF molecule interacted with its surrounding environment. The drain electrode was biased at 100 mV, and the electrolyte,

which served as a gate electrode, was held at or near 0 V. Incubation of each device with experimental substrate solutions transduced fluctuations, $\Delta I(t)$, which were amplified (Keithley 428), digitized at 100 kHz, and stored as uninterrupted, 600 s data sets for later analysis. Between measurements, the KF or Sso7d-KF nanocircuits were rinsed 5 x 1 mL activity buffer, incubated in buffer for 5 min, then rinsed 5 x 1 mL with buffer before introducing another substrate solution. Each enzyme molecule was compared to nucleotide-free buffer to confirm that $\Delta I(t)$ fluctuations corresponded to enzyme activity. For Sso7d-KF measurements with a long templates or KF measurements with rUTP and PP_i, solutions containing 100 nM M13/$(T)_{42}$ and 10 μM dATP or 100 nM M13/$(A)_{42}$ and 10 μM dTTP, respectively, were also measured to collect directly comparable data sets, confirm typical enzyme activities, and reproduce the types of $\Delta I(t)$ excursions reported previously.[3,6]

4.4 Preliminary measurements with rUTP and PP_i

A single device attempted to measure incorporation of rUTP opposite the DNA template $(A)_{42}$ and inhibition/pyrophosphorolysis with PP_i. These measurements were highly inconsistent, showing varying periods of positive $\Delta I(t)$ excursions, negative $\Delta I(t)$ excursions, mixtures of both, and multiple current states. This device also showed inconsistent behavior with the $(A)_{42}$ DNA template and the native dTTP substrate, which indicated that the device was likely not suitable for accurate measurements. Experiments are ongoing to reproduce measurements with rUTP and PP_i. Ensemble experiments are also ongoing to test incorporation, inhibition, and pyrophosphorolysis.

4.5 KF mutagenesis and KF variant activity

A total of 21 KF variants were engineered and generated (Table 4-1) using oligonucleotide-directed mutagenesis of the coding strand of template DNA with a previously described method.[35] Most of the KF variants were expressed, purified (Figure 4-4), and tested for their activity. Figure 4-5 shows the activity for each variant in a preliminary fluorescence-based activity assay. For each KF variant, cell pellets containing expressed protein were stored at -80 °C for long-term storage. Supercharged KF variants, which were designed by Dr. Tivoli Olsen (Figure 4-6A-C), were expressed, purified, and also tested for their activity using the ensemble fluorescence-based assay (Figure 4-6D,E). Atomic force microscopy confirmed attachment of superpositive KF thus far (Figure 4-6F).

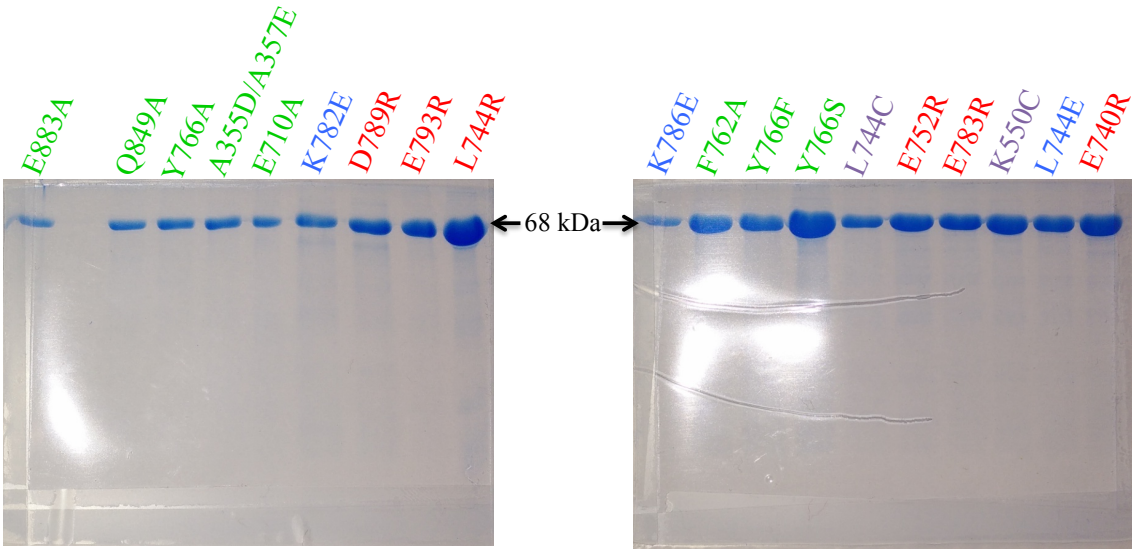


Figure 4-4. 15% SDS-PAGE gel of active site/activity variants (green), point charge variants (blue or red), and attachment site variants (purple).

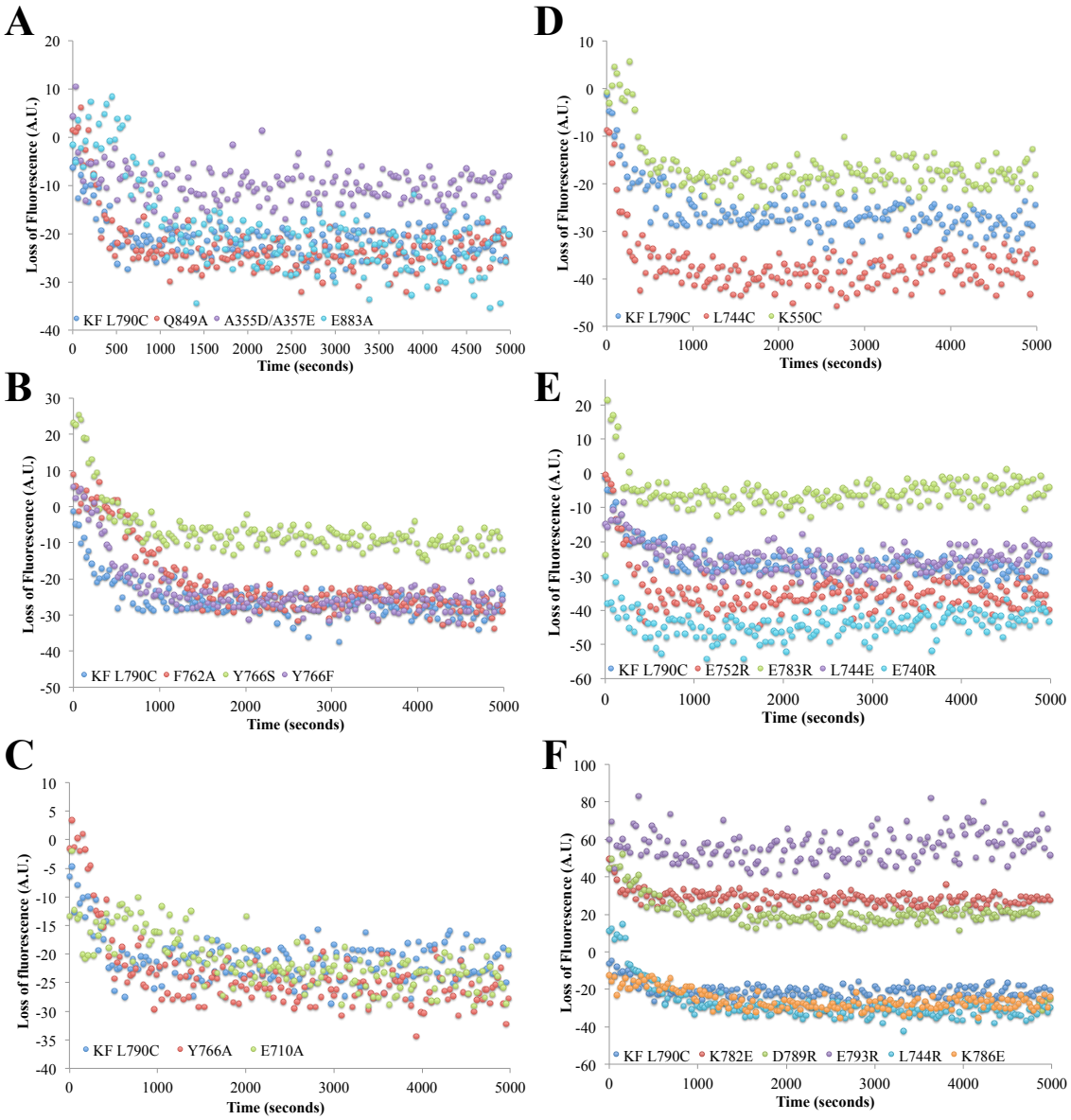


Figure 4-5. Fluorescence-based activity assays of active site/activity variants (**A, B, C**), attachment site variants (**D**), and point charge variants (**E** and **F**). Most variants appeared to show a loss in fluorescence, or increase in negative fluorescence when the background was subtracted. The decreased fluorescence correlates with enzyme activity.

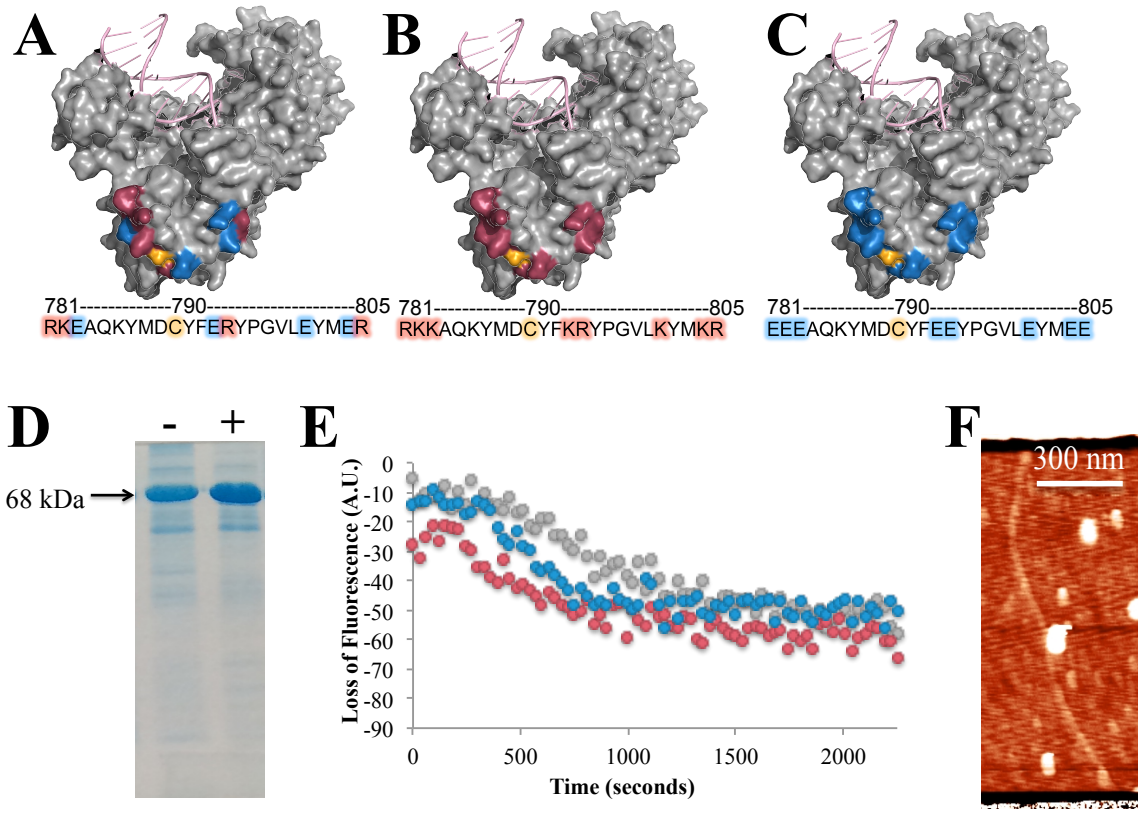


Figure 4-6. Design, generation, activity, and single-molecule attachment of supercharged KF variants. (**A**) The surface mode crystal structure of KF (gray, PDB ID: 1KFD) highlights the enzyme's cysteine attachment site in orange and nearby positively (red) and negatively (blue) charged amino acid residues. The sequence and specific targeted residues are highlighted below the structures. The (**B**) superpositive KF and (**C**) supernegative KF surface mode crystal structures also highlight the residues to represent their new charge. (**C**) Immobilized Ni^{2+} affinity chromatography followed by size exclusion chromatography purified the supernegative (−) and superpositive (+) variants and (**D**) fluorescence-based assays demonstrated their activity compared to standard KF. (**E**) AFM has only thus far confirmed single-molecule attachment of the superpositive KF.

4.6 Fusion of Sso7d to KF and electronic measurements

To increase the processivity of KF, the 7 kDa double-stranded DNA-binding protein Sso7d from *Sulfolobus solfataricus* was fused to the N-terminus of KF. Sso7d was also fused to the supernegative and superpositive KF variants in an effort to engineer highly processive polymerases with high signal to noise ratios. Using a previously described method,[36] splicing of the Sso7d and KF genes through overlap extension polymerase chain reaction (PCR) generated fusion proteins (Figure 4-7). Following expression and purification of Sso7d-KF (Figure 4-8A), ensemble fluorescence-based assays measured activity of the newly generated fusion proteins, as described for KF (Figure 4-8B).[3,34] Atomic force microscopy confirmed single-molecule attachment of Sso7d-KF to electronic nanocircuits (Figure 4-8C). Single-molecule measurements of Sso7d-KF comparing standard solutions of 100 nM M13/$(T)_{42}$ only, 10 μM dATP only, or both substrates together identified SWCNT-FETs functionalized with active Sso7d-KF (Figure 4-9A-C). A template containing 26 tetra-base repeats, or 104 template bases, which exceeded the known KF processivity of 40-50 bases (100 nM $(CTTT)_{26}$ + 10 μM dGTP + 10 μM dATP), measured the processivity of Sso7d-KF (Figure 4-9D).[31] Sso7d-KF transduced $\Delta I(t)$ fluctuations but did not appear to significantly enhance processivity of a long template compared to KF.

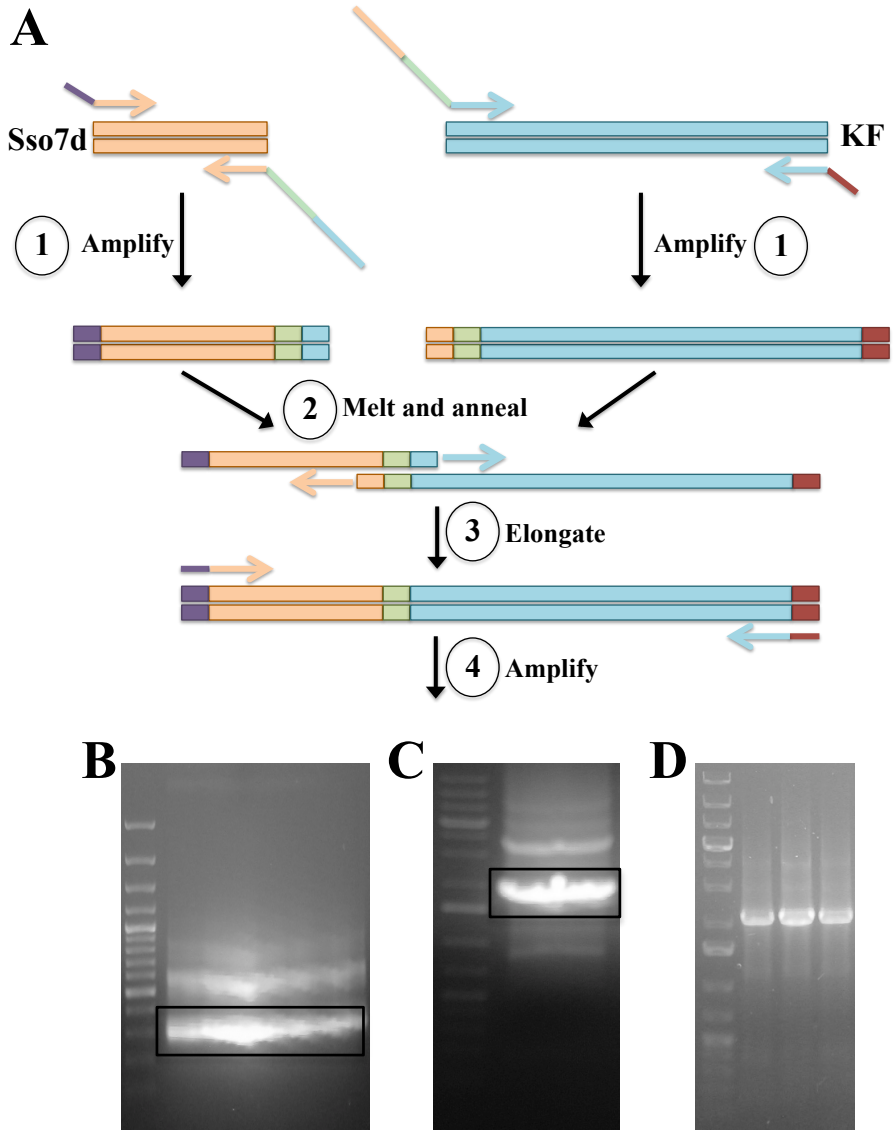


Figure 4-7. Splicing by overlap extension polymerase chain reaction for the generation of Sso7d-KF. (**A**) A schematic of overlap extension polymerase chain reaction demonstrates the steps necessary to splice together the Sso7d and KF genes with a glycine-serine linker between the two genes (green). The amplified genes (Step 1) of (**B**) Sso7d and (**C**) KF were purified by gel excision following agarose gel electrophoresis. (**D**) Elongation (Step 3) and amplification (Step 4) of the annealed genes in Step 2 produced the final gene product in preparation for subcloning into an expression vector.

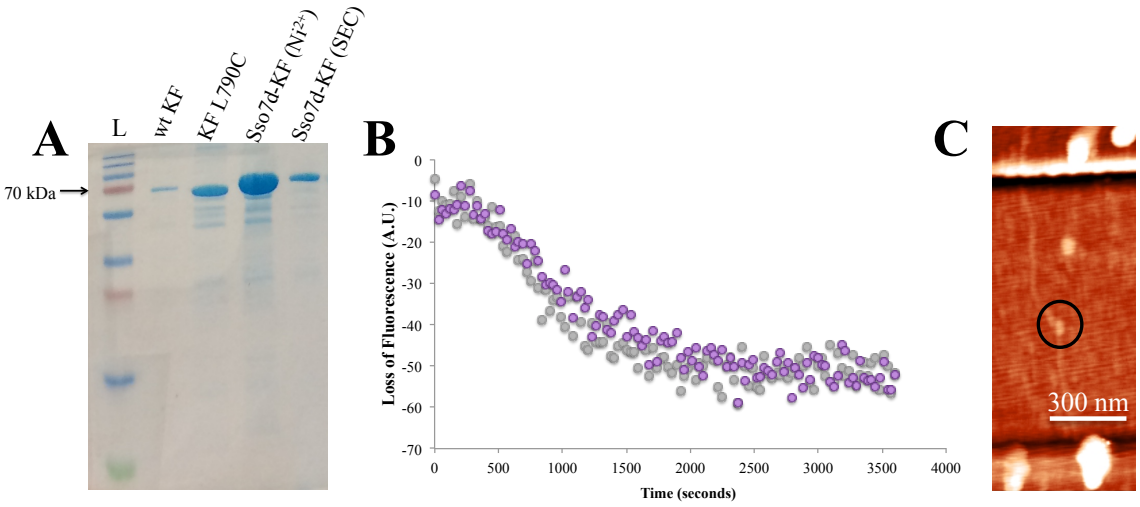


Figure 4-8. Purification, characterization, and attachment of Sso7d-KF. (**A**) A representative 15% SDS-PAGE gel shows Sso7d-KF to be larger in size compared to KF, consistent with an increase of ~7 kDa from the double-stranded DNA binding protein. Immobilized Ni^{2+} affinity chromatography followed by size exclusion chromatography purified Sso7d-KF for attachments to SWCNT-FETs. (**B**) A fluorescence-based assay established activity of Sso7d-KF (purple) consistent with KF (gray). (**C**) Atomic force microscopy confirmed single-molecule attachment of KF to a carbon nanotube.

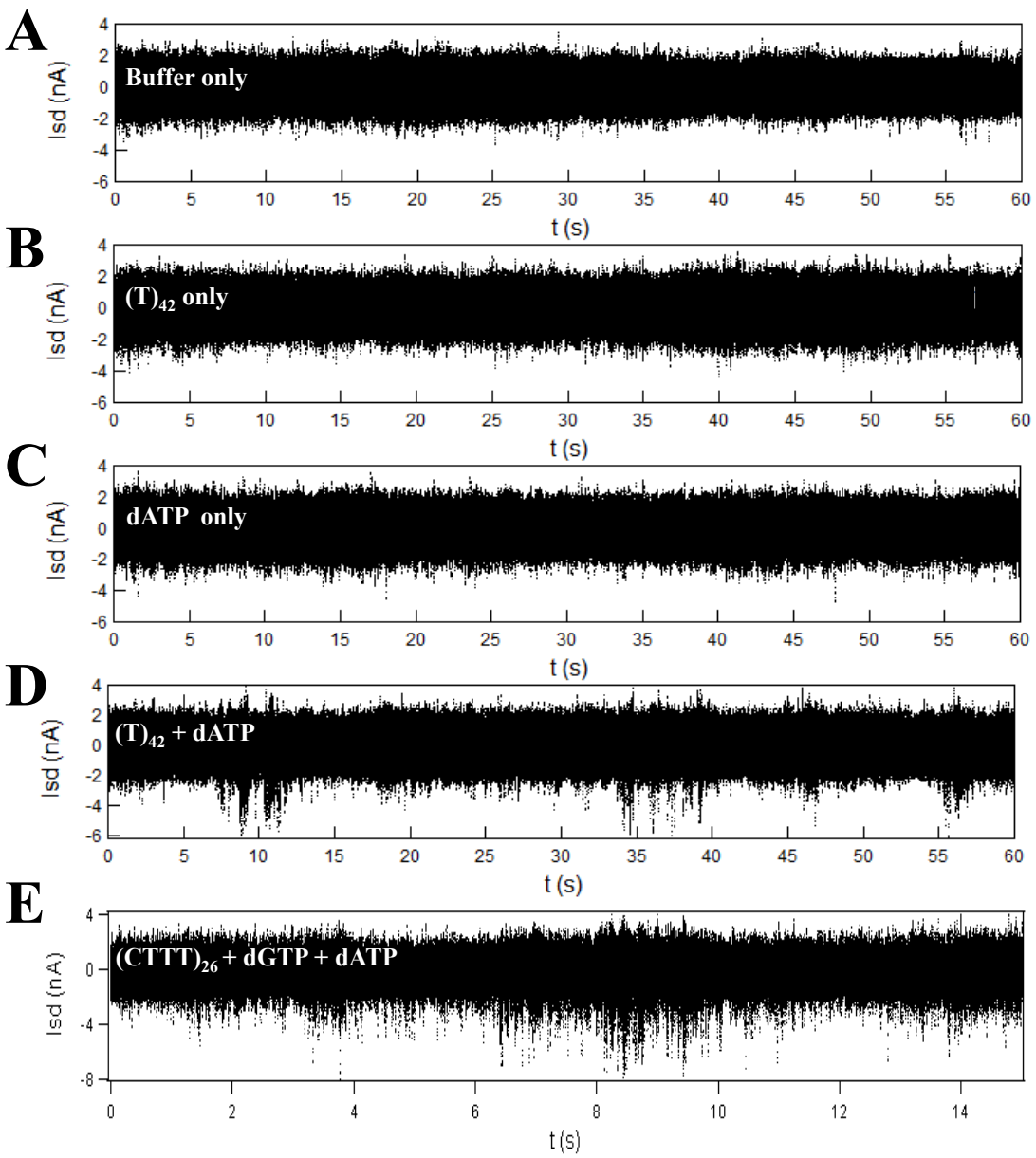


Figure 4-9. Electronic signals of Sso7d-KF. KF, when fused to a double-stranded DNA-binding domain protein, did not produce $\Delta I(t)$ fluctuations in the presence of (**A**) buffer only, (**B**) $(T)_{42}$ only, or (**C**) dATP only. Sso7d-KF did, however, produce $\Delta I(t)$ fluctuations in the presence of (**D**) both $(T)_{42}$ and dATP. This fusion protein also produced $\Delta I(t)$ fluctuations in the presence of (**E**) $(CTTT)_{26}$ + dGTP + dATP, but did not appear to significantly enhance processivity.

4.7 Conclusion

The nature of electronic signals generated by KF-functionalized SWCNT-FETs remains unclear. In order to better understand these signals and adapt these electronic biosensors for biotechnological applications, various strategies are necessary. Experiments to measure incorporation of non-cognate substrates and inhibition with pyrophosphate were designed and attempted, but must be reproduced for further analysis. Several proteins were engineered to determine the precise nature of the observed signals by alteration of the enzyme's active site/activity and charges located proximal to the nanotube. KF variants were also engineered to install new charges on domains with long-range motion or to change the enzyme's attachment site for new insight into the protein's motions. More processive enzymes and enzymes with a concentration of identical charges for higher signal to noise ratios were designed to improve data analysis and develop the devices for further application. Many of these enzymes retained bulk activity in ensemble assays, and are in various stages of progress with regard to measurements of enzyme-functionalized SWCNT-FETs. The future of measurements with the engineered KF variants was outlined here, and could potentially provide major improvements to the understanding and applications of these electronic biosensors.

4.8 Materials and methods

Materials

Reagents purchased commercially include antibiotics (Fisher Scientific), Ni-IMAC resin (Bio-Rad Laboratories), cell lines (Stratagene), deoxynucleoside triphosphates (Fisher Scientific), ribonucleoside triphosphate analogs (Promega), enzymes (New England Biolabs or Fermentas), oligonucleotides (Fisher or IDT), and 96-

well fluorescence plates (Nunc). All other chemicals were purchased commercially from Acros Organics, EMD, Fisher Scientific, or Sigma Aldrich. All reagents were used as received.

Generation of KF variants

Table 4-3 shows all of the oligonucleotides used to engineer KF for altered activity.

Table 4-3. Oligonucleotides used to engineer KF.

Oligonucleotide	**Sequence (5′-3′)**	**Use**
6H-TEV-Sso7d-pET22F	GGAGATATACATATGCATCATCATCATCATCACGAAAATCTTTATTTTCAAGGAGGCGCAACCGTAAAGTTCAAGTACAAAGGCGAAG	Forward primer for Sso7d with 6x His tag and TEV site
Sso7d-KF-OL_R	CATCAAGGATGGTGACGTAGTTGTCATAAGAAATCACGCCTGAGCCTCCCTTTTTCTGCTTCTCCAGCATCTGCAGC	Reverse overlap primer for splicing Sso7d and KF genes
Sso7d-KF-OL_F	GCTGCAGATGCTGGAGAAGCAGAAAAAGGGAGGCTCAGGCGTGATTTCTTATGACAACTACGTCACCATCCTTGATG	Forward overlap primer for splicing Sso7d and KF genes
KF-pET22R	TAGCTACGTCGACTCATTATTAGTGCGCCTGATCCCAGTTTTCGC	Reverse primer for KF
Supercharge (−)	CGGCAATTGAACATTCCAGAAGAAGAAGCGCAGAAGTACATGGACTGCTACTTCGAAGAATACCCTGGCGTGCTGGAGTATATGGAAGAAACCCGTGCTCAGGCG	To mutate R782, K782, R794, and R804 to Glu
Supercharge (+)	CAATTGAACATTCCACGTAAAAAAGCGCAGAAGTACATGGACTGCTACTTCAAACGCTACCCTGGCGTGCTGAAATATATGAAACGCACCCGTGCTCAG	To mutate E783, E793, E800, and E803 to Arg
KF_Q849A_22b	TGTCGGCGGCGGTTCCCGCCATTGGCGCGTTAATGGC	To mutate Q849 to Ala
KF_E883A_22b	CTTTATGAACTTCAAATACCAGCGCATCGTGTACCTGCATGATC	To mutate E883 to Ala
KF_L744C_22b	GGTGACGGTTTCCAGTGGGCAACCAAACACTTCTGCCGC	To mutate L744 to Cys
KF_K550C_22b	TTCAGAATGATTGTGCAGCACGCACGGATCGATCTTCACACCG	To mutate K550 to Cys
KF_L744E_22b	GGTGACGGTTTCCAGTGGTTCACCAAACACTTCTGCCGC	To mutate L744 to Glu

Oligonucleotide	Sequence (5′-3′)	Use
KF_L744R_22b	GGTGACGGTTTCCAGTGGACGACCAA ACACTTCTGCCGC	To mutate L744 to Arg
KF_F762A_22b	CGAAAGCACTCATGCCATAAATCAGA CCTGCGTTGATCGCTTTCGCGCTACGG	To mutate F762 to Ala
KF_Y766F_22b	ACCGAAAGCACTCATGCCGAAAATCA GACCAAAGTTGATCG	To mutate Y766 to Phe
KF_Y766S_22b	ACCGAAAGCACTCATGCCACTAATCA GACCAAAGTTGATCG	To mutate Y766 to Ser
KF_Y766A_22b	ACCGAAAGCACTCATGCCCGCAATCA GACCAAAGTTGATCG	To mutate Y766 to Ala
KF_E710A_22b	GATGCGCCATAATGCGCAGTGCAATC TGCGAGTAGTCCGC	To mutate E710 to Ala
KF_E740R_22b	CCAGTGGCAAACCAAACACACGTGCC GCCGTTGCCCGG	To mutate E740 to Arg
KF_E752R_22b	TTTCGCGCTACGGCGTTGACGGCTGGT GACGGTTTCCAG	To mutate E752 to Arg
KF_K782E_22b	CCATGTACTTCTGCGCTTCCTCACGTG GAATGTTCAATTGCC	To mutate K782 to Glu
KF_E783R_22b	GTCCATGTACTTCTGCGCACGTTTACG TGGAATGTTCAATTG	To mutate E783 to Arg
KF_K786E_22b	CGAAGTAGCAGTCCATGTATTCCTGC GCTTCTTTACGTGG	To mutate K786 to Glu
KF_D789R_22b	GGTAGCGTTCGAAGTAGCAACGCATG TACTTCTGCGCTTCTT	To mutate D789 to Arg
KF_E793R_22b	GCACGCCAGGGTAGCGACGGAAGTAG CAGTCCATGTAC	To mutate E793 to Arg
KF_add_exo_22b	ATGTTATCAAGGCTGTCGGTTTCGGTA TCAAATGCAAATACCGGCGCTT	To restore exonuclease activity

Splicing by overlap extension PCR to obtain Sso7d-KF

The double-stranded DNA binding protein Sso7d was genetically fused to the N-terminus of KF using previously described methods.[36] A pET15b plasmid contained a gene encoding for Pfu polymerase fused to Sso7d and was used to amplify the Sso7d gene. For the KF gene, a pET22b plasmid contained the gene encoding KF(D355A/E357A/C907S/L790C), referred to hereafter as KF L790C, and was used for amplification. The Sso7d gene was amplified with polymerase chain reaction (PCR)

using the 6HTEV-Sso7d-pET22F forward primer and the Sso7d-Klenow-OL_R reverse primer. The KF L790C gene was separately amplified using the Sso7d-Klenow-OL_F forward primer and the Klenow-pET22_R reverse primer. The 6HTEV-Sso7d-pET22F forward primer introduced a 6x histidine tag and TEV protease site to the N-terminus of Sso7d and contained the NdeI restriction site. The Sso7d-Klenow-OL_R and Sso7d-Klenow-OL_F primers contained codons for a glycine-serine linker between the C-terminus of Sso7d and the N-terminus of KF L790C. The Klenow-pET22_R primer contained a SalI restriction site.

Following initial amplification, KF L790C PCR products were first purified with the Zymo Clean and Concentrate kit. The desired DNA was further purified by extracting from a 1% agarose gel with the Zymo Gel Extraction kit following electrophoresis at 110V for 50 min. Sso7d PCR products were electrophoresed at 110V for 45 min on a 1% agarose gel, excised, and purified using the Zymo Gel Extraction kit. DNA concentrations were measured at 260/280 nm using a Thermo Scientific NanoDrop 2000 UV/Vis spectrophotometer. The amplified and purified Sso7d gene was annealed to the amplified, purified KF gene in a 1:5 or 1:10 mass ratio. The mixtures were heated to 95 °C for 5 min then cooled from 60 °C to 55 °C to 50 °C for 5 min at each temperature. The annealed genes were further amplified with the 6HTEV-Sso7d-pET22F forward primer and the Klenow-pET22_R reverse primer.

For subcloning, a pET22b plasmid was digested with NdeI and XhoI restriction enzymes at 37 °C for 2 h then heated to 65 °C for 20 min to stop the reaction. Since the KF L790C gene contains a XhoI restriction site, the amplified Sso7d-KF gene was digested with the restriction enzymes NdeI and SalI 37 °C for 2 h followed by heating to

65 °C for 20 min. Upon digestion, SalI generates a digested terminus in Sso7d-KF L790C that is complementary to the XhoI-digested site on the plasmid. Various ratios of digested gene to digested vector were mixed and heated to 45 °C for 5 min to anneal the digested termini. After addition of T4 DNA ligase in T4 DNA ligase buffer, ligation reactions were held at 16 °C for 16 h. Ligation products were then used to transform $CaCl_2$-competent Top10 *E. coli* cells by heat shock. Cells were plated and grown overnight at 37 °C on solid media supplemented with 50 μg/mL carbenicillin. Several colonies were selected for amplification of their DNA by Sso7d-KF-specific primers and sequences of PCR products with the appropriate size were confirmed.

Generation and isolation of uracil-doped single-stranded DNA templates for mutagenesis

Two separate pET22b plasmids containing a 6x histidine tag fused to a TEV protease site also encoded for one of two genes: either KF(D355A/E357A/C907S), referred to hereafter as KF C907S, or KF(D355A/E357A/C907S/L790C),[3,10] referred to hereafter as KF L790C. Mutations were introduced into KF C907S and KF L790C using previously described methods.[35] The plasmids were used to separately transform $CaCl_2$-competent CJ236 *E. coli* cells by heat shock. Cells were plated on LB/agar containing 50 μg/mL carbenicillin and 15 μg/mL chloramphenicol and grown overnight at 37 °C. Single colonies were used to inoculate 6 x 2.5 mL 2YT media containing 50 μg/mL carbenicillin and 5 μg/mL chloramphenicol. Following growth for 6-8 h at 37 °C with shaking, log phase cultures were infected with ~1.5 x 10^{10} phage/mL helper phage KO7 for 45 min at 37 °C. Cultures were then transferred to fresh 2YT media (50x volume) containing 50 μg/mL carbenicillin, 20 μg/mL kanamycin, and 0.25 μg/mL uridine for 16-18 h growth at 37 °C. PEG-NaCl precipitation (1/5 v/v) was used to isolate phage. KF

C907S and KF L790C uracil-doped single-stranded DNA templates (dU-ssDNA) were isolated with a Qiaprep Spin M13 kit and nucleic acid concentration was measured at 260/280 nm with a Thermo Scientific NanoDrop 2000 UV/Vis spectrophotometer.

Oligonucleotide-directed mutagenesis to obtain KF variants

The KF C907S dU-ssDNA template was used for mutagenesis to obtain the L744C and K550C attachment site variants. The KF L790C dU-ssDNA template was used for mutagenesis of the remaining variants. First, oligonucleotides (Table 4-3) were phosphorylated with ATP by T4 polynucleotide kinase in the presence of DTT for 1 h at 37 °C. Phosphorylated oligonucleotides were annealed to the dU-ssDNA template at a 100:1 molar ratio by heating the mixture to 95 °C for 5 min, then cooling gradually from 50 °C for 5 min, to 40 °C for 5 min, followed by 30 °C for 3 min, and finally 25 °C for 10 min. The annealed oligonucleotide-templates were then elongated by incubation at room temperature overnight with dNTPs, ATP, DTT, T7 DNA polymerase, and T4 DNA ligase. Reaction products were electrophoresed for 45 min at 90V in a 0.8% agarose gel and the bottom band was excised. Double-stranded covalently closed circular DNA (ds-cccDNA) was isolated with the Zymoclean Gel DNA Recovery kit and used to transform $CaCl_2$-competent Top10 *E. coli* cells by heat shock. After growth overnight at 37 °C on LB/agar containing 50 μg/mL carbenicillin, DNA from individual colonies was amplified by PCR, and the desired mutation was confirmed by sequencing.

Expression and purification of KF and KF variants

The following expression and purification protocol applied buffers and plasmids from a previously described procedure.[3] Separate plasmids that contained the genes encoding KF L790C,[3,10] Sso7d-KF, or the variants generated by oligonucleotide-directed

mutagenesis were used to transform $CaCl_2$-competent BL21(DE3) *E. coli* cells by heat shock. For each transformation following overnight growth on solid media, a single colony was used to inoculate 25 mL LB media supplemented with 50 μg/mL carbenicillin for growth in liquid media overnight at 37 °C with shaking. LB (1 L) supplemented with 50 μg/mL carbenicillin was inoculated with 10 mL of the overnight culture and incubated with shaking at 37 °C for several hours. Once the cells reached late log phage (OD_{600} = 0.9), expression was induced by the addition of 1 mM IPTG. After 3-4 h of protein expression at 37 °C with shaking, cells were harvested by centrifugation (6000 rpm, 20 min, 4 °C) and resuspended in lysis buffer (20 mM Tris, 50 mM NaCl, 10 mM BME, pH 8.0).

Cells were lysed by sonication and the cell debris was collected by centrifugation (15,000 rpm, 45 min, 4 °C). Following filtration through a 0.45 μm pore filter, the lysate supernatant was incubated with Ni-IMAC resin overnight at 4 °C. Protein was eluted in the lysis buffer with 250 mM imidazole, concentrated, and then treated with TEV protease for two days at 4 °C. For KF, Sso7d-KF, superpositive KF, and supernegative KF, the respective mixture was centrifuged then filtered through a 0.45 μm filter prior to size exclusion chromatography in TBS (20 mM Tris, 50 mM NaCl, 100 μM TCEP, pH 7.9) on a Bio-Rad Biologic DuoFlow FPLC. Purity was assessed by SDS-PAGE (Figure 2-1A).

Ensemble activity of KF to test native dNTP incorporation, rNTP incorporation, and PP_i inhibition/pyrophosphorolysis

Oligonucleotides used to test activity

Table 4-4 lists the oligonucleotides used to test KF activity, rNTP incorporation, PPi inhibition/pyrophosphorolysis, and for measurements with the nanocircuit. Upon receipt, HPLC-purified oligonucleotides were solubilized in water to 100 μM. Bold regions indicate the M13 priming site. [2AP] indicates 2-aminopurine.

Table 4-4. Oligonucleotides used in KF and Sso7d-KF SWCNT-FET measurements.

Oligonucleotide	**Sequence (5′-3′)**	**Use**
M13F	TGTAAAACGACGGCCAGT	M13 forward primer
ActAssay Template	TCGAGCTATCTCTAAAGC[2AP] GCTAACTATCGAGCTATCGCGA A**ACTGGCCGTCGTTTTACA**	Standard activity assay template containing 2-aminopurine
$(T)_{42}$	$(T)_{42}$**ACTGGCCGTCGTTTTACA**	Test native dATP incorporation
$(A)_{42}$	$(A)_{42}$**ACTGGCCGTCGTTTTACA**	Test native dTTP incorporation and PP_i inhibition and pyrophosphorolysis on nanocircuit
STR_FGA_Fwd	GCCCCATAGGTTTTGAACTCA	Forward primer for $(CTTT)_{26}$ template
STR_FGA_104	TTTCTTTCTTTCTTTCTTTCTTTC TTTCTTTCTTTCTTTCTTTCTTTC TTTCTTTCTTTCTTTCTTTCTTTC TTTCTTTCTTTCTTTCTTTCTTTC TTTCTTT**CTGAGTTCAAAACCT ATGGGGC**	Test processivity of Sso7d-KF
Random 45mer	CAGCCTTGTCTCGCCAGAATGC TGGTCAGCATACGGAAGAGCTC A	Test pyrophosphorolysis with PP_i after annealing to Random 45mer Rev Comp
Random 45mer Rev Comp	TGAGCTCTTCCGTATGCTGACC AGCATTCTGGCGAGACAAGGCT G	Test pyrophosphorolysis with PP_i after annealing to Random 45mer

Ensemble assay for KF and KF variant Activity

To confirm activity of KF(L790C) and all KF variants versus wild-type KF, a previously described assay was adapted as follows.[3,34] A randomized DNA template containing both 2-aminopurine (ActAssay template in Table 4-4) and an M13 priming site (bold) was annealed to the M13F primer by heating the mixture to 65 °C and slow-cooling to room temperature for 1 h. A comparable decrease in fluorescence was observed for KF(L790C) and wild-type KF (both 1 μM) upon incubation with the primer-template mixture (25 μM) and dNTPs (250 μM). The raw fluorescence data was corrected by subtraction of background, which was measured in the absence of dNTPs. The excitation and emission wavelengths employed in this experiment were 305 and 365 nm, respectively.

Ensemble activity for rNTP incorporation and PPi inhibition/pyrophosphorolysis

The following protocols have been designed for rNTP incorporation and PPi inhibition/pyrophosphorolysis for ensemble measurements. These protocols are currently undergoing further optimization by Johnathon Truong. To confirm incorporation of rNTPs, the same primer-template containing 2-aminopurine (25 μM) used in KF and KF variant activity assays will be polymerized by KF (1 μM) with native dNTPs (250 μM) in 10 mM Tris, 50 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT, pH 7.9. Reactions to test rNTP incorporation will contain 2.5 mM rNTP in place of the corresponding native dNTP, and negative control reactions will omit either the rNTP or its corresponding dNTP. Reactions will be incubated at 25 °C for 2 h with simultaneous kinetic measurement using a fluorescence plate-reader. Changes in fluorescence for rNTP experiments will be compared to the changes in fluorescence for native substrate incorporation. The

excitation and emission wavelengths employed in this experiment will be 305 and 365 nm, respectively.

4.9 References

1. Choi, Y. & Moody, I. S. *et al.* Single-molecule lysozyme dynamics monitored by an electronic circuit. *Science* **335,** 319–24 (2012).

2. Choi, Y. *et al.* Single-molecule dynamics of lysozyme processing distinguishes linear and cross-linked peptidoglycan substrates. *J. Am. Chem. Soc.* **134,** 2032–5 (2012).

3. Olsen, T. J. & Choi, Y. *et al.* Electronic measurements of single-molecule processing by DNA polymerase I (Klenow fragment). *J. Am. Chem. Soc.* **135,** 7855–60 (2013).

4. Choi, Y. & Olsen, T. J. *et al.* Dissecting single-molecule signal transduction in carbon nanotube circuits with protein engineering. *Nano Lett.* **13,** 625–31 (2013).

5. Sims, P. C. *et al.* Electronic measurements of single-molecule catalysis by cAMP-dependent protein kinase A. (2013).

6. Pugliese, K. M. & Gul, O. T. *et al.* Processive incorporation of deoxynucleoside triphosphate analogs by single-molecule DNA polymerase I (Klenow fragment) nanocircuits. *J. Am. Chem. Soc.* **137,** 9587–94 (2015).

7. Garalde, D. R. *et al.* Distinct complexes of DNA polymerase I (Klenow fragment) for base and sugar discrimination during nucleotide substrate selection. *J. Biol. Chem.* **286,** 14480–92 (2011).

8. 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,** 4835–4843 (1991).

9. Polesky, A. H., Dahlberg, M. E., Benkovic, S. J., Grindley, N. D. & Joyce, C. M. Side chains involved in catalysis of the polymerase reaction of DNA polymerase I from *Escherichia coli*. *J. Biol. Chem.* **267,** 8417–8428 (1992).

10. Derbyshire, V. *et al.* Genetic and crystallographic studies of the 3',5'-exonucleolytic site of DNA polymerase I. *Science* **240,** 199–201 (1988).

11. 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).

12. Joyce, C. How DNA travels between the separate polymerase and 3'-5'-exonuclease sites of DNA polymerase I (Klenow fragment). *J. Biol. Chem.* **264,** 10858–10866 (1989).

13. Tuske, S., Singh, K., Kaushik, N. & Modak, M. J. The J-helix of *Escherichia coli* DNA polymerase I (Klenow fragment) regulates polymerase and 3'-5'-exonuclease functions. *J. Biol. Chem.* **275,** 23759–68 (2000).

14. 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).

15. Carroll, S. S., Cowart, M. & Benkovic, S. J. A mutant of DNA polymerase I (Klenow fragment) with reduced fidelity. *Biochemistry* **30,** 804–813 (1991).

16. Tabor, S. & Richardson, C. C. A single residue in DNA polymerases of the *Escherichia coli* DNA polymerase I family is critical for distinguishing between deoxy- and dideoxyribonucleotides. *Proc. Natl. Acad. Sci. U. S. A.* **92,** 6339–43 (1995).

17. Kaushik, N., Pandey, V. N. & Modak, M. J. Significance of the O-helix residues of *Escherichia coli* DNA polymerase I in DNA synthesis: dynamics of the dNTP binding pocket. *Biochemistry* **35,** 7256–66 (1996).

18. Bell, J. B., Eckert, K. A., Joyce, C. M. & Kunkel, T. A. Base miscoding and strand misalignment errors by mutator Klenow polymerases with amino acid substitutions at tyrosine 766 in the O helix of the fingers subdomain. *J. Biol. Chem.* **272,** 7345–7351 (1997).

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

20. 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).

21. Minnick, D. T., Liu, L., Grindley, N. D. F., Kunkel, T. A. & Joyce, C. M. Discrimination against purine-pyrimidine mispairs in the polymerase active site of DNA polymerase I: a structural explanation. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 1194–9 (2002).

22. Potapova, O. *et al.* DNA polymerase catalysis in the absence of Watson-Crick hydrogen bonds: analysis by single-turnover kinetics. *Biochemistry* **45,** 890–8 (2006).

23. Hohlbein, J. *et al.* Conformational landscapes of DNA polymerase I and mutator derivatives establish fidelity checkpoints for nucleotide insertion. *Nat. Commun.* **4,**

2131 (2013).

24. Bermek, O., Grindley, N. D. F. & Joyce, C. M. Prechemistry nucleotide selection checkpoints in the reaction pathway of DNA polymerase I and roles of glu710 and tyr766. *Biochemistry* **52,** 6258–74 (2013).

25. Santoso, Y. & Joyce, C. M. *et al.* Conformational transitions in DNA polymerase I revealed by single-molecule FRET. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 715–20 (2010).

26. Bermek, O., Grindley, N. D. F. & Joyce, C. M. Distinct roles of the active-site Mg^{2+} ligands, Asp882 and Asp705, of DNA polymerase I (Klenow fragment) during the prechemistry conformational transitions. *J. Biol. Chem.* **286,** 3755–66 (2011).

27. Berezhna, S. Y., Gill, J. P., Lamichhane, R. & Millar, D. P. Single-molecule Förster resonance energy transfer reveals an innate fidelity checkpoint in DNA polymerase I. *J. Am. Chem. Soc.* **134,** 11261–8 (2012).

28. Evans, G. W., Hohlbein, J., Craggs, T., Aigrain, L. & Kapanidis, A. N. Real-time single-molecule studies of the motions of DNA polymerase fingers illuminate DNA synthesis mechanisms. *Nucleic Acids Res.* **43,** 5998–6008 (2015).

29. 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,** 14095–14101 (1993).

30. Wu, E. Y. & Beese, L. S. The structure of a high fidelity DNA polymerase bound to a mismatched nucleotide reveals an 'ajar' intermediate conformation in the nucleotide selection mechanism. *J. Biol. Chem.* **286,** 19758–67 (2011).

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

32. Wang, Y. *et al.* A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance in vitro. *Nucleic Acids Res.* **32,** 1197–207 (2004).

33. Simeonov, P., Berger-Hoffmann, R., Hoffmann, R., Sträter, N. & Zuchner, T. Surface supercharged human enteropeptidase light chain shows improved solubility and refolding yield. *Protein Eng. Des. Sel.* **24,** 261–8 (2011).

34. 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,** 9185–9192 (1995).

35. Kunkel, T. A. Rapid and efficient site-specific mutagenesis without phenotypic

selection. *Proc. Natl. Acad. Sci.* **82,** 488–492 (1985).

36. Heckman, K. L. & Pease, L. R. Gene splicing and mutagenesis by PCR-driven overlap extension. *Nat. Protoc.* **2,** 924–32 (2007).

Chapter 5: Toward the *In Vitro* Evolution of High-Affinity Binding Partners for Membrane Proteins From Phage-Displayed Affinity Reagent Libraries

5.1 Abstract

Structural elucidation of DNA polymerase I Klenow Fragment laid the groundwork for the development of techniques for understanding the biophysical nature of the enzyme's catalysis. The benefit of known crystal structures can be extended to membrane proteins, which also undergo dynamic motions and comprise over half of all drug targets. Here, phage-displayed peptide libraries in the loops of rigid scaffold proteins were used to select for high affinity binding partners to membrane proteins. Such binding partners could impart the stabilization required for crystallization of membrane proteins at various stages of molecular transport. Selections against a mock target demonstrated that high affinity binding partners could be evolved from phage-displayed libraries of the small, stable proteins human lysozyme and echistatin. Ure-I, a membrane protein involved in the molecular transport of urea for *Helicobacter pylori* survival, was immobilized by passive adsorption for solid-phase selections. However, an enzyme-linked immunosorbent assay failed to detect immobilization after week-long storage of the membrane protein, preventing successful *in vitro* evolution of high affinity binding partners. ShuA, an outer membrane protein from *Shigella dysenteriae* which could not be immobilized by passive adsorption, was expressed on the surface of *Escherichia coli* cells. Detection of ShuA on the surface of *E. coli* cells was optimized for future selections of high affinity binding partners from phage-displayed protein libraries.

5.2 Introduction

Membrane-bound proteins serve several vital functions in a variety of organisms, ranging from eukaryotic signaling pathways to bacterial virulence.[1,2] Although membrane proteins constitute 20 to 30% of all proteins encoded in any given genome,[3] only ~200 unique membrane protein structures were known as of 2009.[4] The number of unique membrane protein structures has now surpassed 500 according to an online database (http://blanco.biomol.uci.edu/mpstruc/). However, membrane protein structures remain poorly represented among the over 100,000 protein structures deposited in the RSCB Protein Data Bank (PDB).[5] Membrane proteins comprise over 50% of drug targets,[6] and, therefore, rational drug development could benefit from structural determination of membrane proteins. A lack of complete structural information for the majority of membrane-bound proteins is due to difficulty in the expression, solubilization, and crystallization of these proteins.[7]

The natural amphipathic environment of membrane proteins is necessary to stabilize the proteins' various conformational states, and can be difficult to mimic. Methodologies for accurate and complete structural determination require mimicry of a membrane protein's native environment in the form of detergents, lipid bilayers, and nanodiscs.[8–13] However, the massive size of some of these pseudo-membranes do not contribute favorably to crystallization of the membrane protein. Fusion of rigid and highly crystallizable proteins, such as T4 lysozyme, previously facilitated crystallogenesis of the dynamic membrane-bound G-protein coupled receptor (GPCR) proteins.[14,15] The ability of small, rigid proteins to influence membrane protein stability for crystallization could also be made possible through the *in vitro* evolution of high

affinity binding partners. In addition to stabilization, a specifically evolved binding protein with a rigid scaffold could also function as a probe or inhibitor for the membrane protein target.

Echistatin and human lysozyme are two proteins with potential as high affinity binding partners to membrane proteins. These proteins are small in size, easily expressed, and contain rigid scaffolds with flexible loop regions.[16–21] Peptide libraries on the loops of echistatin and human lysozyme have previously been engineered in our laboratory, and the sequence mutations are unlikely to affect the native rigidity of the proteins' scaffold. The libraries are composed of echistatin or human lysozyme displayed on the N-terminus of the M13 bacteriophage P3 minor coat protein. The use of phage display for *in vitro* evolution provides a genotype to phenotype linkage; the gene for a protein displayed on the surface of the phage is encoded in the DNA encapsulated within the phage.[22] This technique can enable the facile identification of high affinity binding partners and can be applied to membrane protein targets.

The *Helicobacter pylori* proton-gated inner membrane channel protein Ure-I and the *Shigella dysenteriae* heme/hemoglobin uptake A outer membrane receptor (ShuA) play important roles in the transport of small molecules to ensure the respective bacterium's survival. Ure-I enables the acid resistance of *H. pylori* by aiding in the transport of urea from the periplasm. The urea can then be broken down by cytoplasmic urease to neutralize the bacteria's immediate environment.[23–25] The closed channel structure of Ure-I is known,[26] but molecular dynamics simulations must be used to determine the structural changes involved in the molecular transport mechanism.[27] ShuA is a β-barrel TonB dependent transporter (TBDT) that obtains iron for survival by

sequestering and transporting heme.[28,29] ShuA sequesters heme by binding to hemoglobin after the excreted Shiga toxin from *S. dysenteriae* causes apoptosis of the host's intestinal red blood cells.[30] The crystal structure of the apo form of ShuA was previously obtained, but the conformation of heme-bound ShuA has only been proposed based on the heme-binding site of the homologous TBDT HasR.[31,32] The ShuA amino acid residues His-86 and His-420 are known to be critical for binding to heme,[29,32] but a heme- or hemoglobin-bound crystal structure is necessary to confirm this binding modality. Structural characterization of both ShuA and Ure-I at various stages of their molecular transport can provide biophysical information about these processes. Additionally, inhibition of these vital transport pathways can prevent the bacterium from obtaining a necessary resource and restore the host's health by destroying the pathogen.

Here, the feasibility of a standard method for *in vitro* evolution of a high-affinity binding partner from phage-displayed echistatin and human lysozyme libraries was first demonstrated with a mock target. Passive adsorption of Ure-I and ShuA to immunosorbent microtiter plates as a method for immobilization was then measured. This immobilization was used to evolve high-affinity binding partners to Ure-I from the echistatin and human lysozyme phage-displayed libraries, but the target protein's rapid degradation made this approach challenging. Previously, members of the Weiss lab identified an echistatin variant as a binding partner with low affinity for ShuA by using passive adsorption to immobilize ShuA. Crystals of ShuA grown in the presence of the variant diffracted at poor resolution of ~7 to 8 Å (unpublished data). Here, the same conditions for passive adsorption of ShuA did not result in immobilization. An alternative method for targeting of ShuA was optimized for future selections of high-

affinity binding partners from phage-displayed echistatin and human lysozyme libraries. This technique has the added advantage of maintaining ShuA in its native environment.

5.3 Biopanning of binding partners to a mock target to demonstrate feasibility

A monoclonal anti-FLAG antibody was first used as a mock target to demonstrate the ability to evolve high affinity binding partners from phage-displayed libraries of echistatin and human lysozyme. Phage-displayed libraries included FLAG affinity tags to monitor human lysozyme or echistatin display and, therefore, each library had an equally strong initial affinity to the anti-FLAG target. Libraries of the phage-displayed proteins echistatin and human lysozyme were used in these selections in which single loop regions were mutated individually, or in combination. All available libraries were used following molar normalization: echistatin Loop 1 and Loop 2 libraries were mixed in a 1:1 molar ratio to form the final "echistatin" library used in selections and the human lysozyme Loop 1, Loop 2, and Loop 1/2 libraries were mixed in a 1:1:1 molar ratio to form the final "human lysozyme" library.

A total of four rounds of selections were performed with each library. Each round increased the stringency for binding to the anti-FLAG antibody by increasing the number of washes following the binding step. Blocking agents (BSA, NFM, ovalbumin, casein) were alternated to prevent nonspecific binding of phage to the solid immunosorbent surface and to ensure selectivity of bound phage for the anti-FLAG antibody. The number of phage selected in each round was quantified from serial dilutions of phage-infected cells. Cells infected with carbenicillin-resistant phagemids were also compared to cell forming units after incubation with tetracycline (to select for tet-resistant XL-1 cells) and kanamycin (to select for M13 KO7 helper phage-infected cells). An enzyme-

linked immunosorbent assay (ELISA) compared the binding of repropagated phage from the first through third round of selections to the naïve library (Figure 5-1). The repropagated phage from each round of selections for human lysozyme binding partners was particularly demonstrative for the gradual increase of specificity for the target.

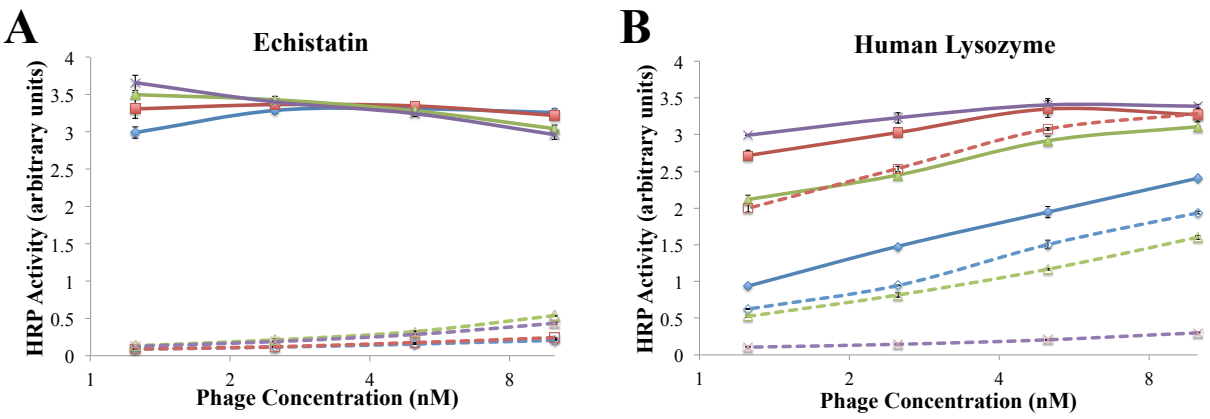


Figure 5-1. Dose-dependent ELISA of selected phage binding to anti-FLAG antibody. Both (**A**) echistatin and (**B**) human lysozyme selected phage bound more strongly to the target (solid) compared to BSA (dashed) in the naïve library (blue), the 1st round (red), the 2nd round (green), and the 3rd round (purple). The human lysozyme library showed a clear loss of binding to BSA in the 2nd and 3rd round selected phage, with the highest signal over background in the 3rd round.

5.4 ELISA to measure binding of individual selectants to anti-FLAG

After four rounds of selection, *E. coli* SS320 cells were infected with eluted phage from the third and fourth rounds of echistatin and human lysozyme selections. Selectants were grown on a medium supplemented with carbenicillin to select for cells infected with the carb-resistant echistatin or human lysozyme phagemid. A spot assay was performed in an ELISA format on cultured bacterial colonies from which phage was individually isolated to test the binding affinities of individual proteins for anti-FLAG (Figure 5-2A). Absorption of the individually propagated phage variants at 280 nm revealed phage concentrations ranging from 20-50 nM. The high phage concentration was hypothesized to be responsible for the significant background signal that was measured for each variant. This was confirmed when the ELISA was repeated after a ~6x dilution of the phage variants (Figure 5-2B). Strength of binding was based on the dimerization of *o*-phenylenediamine (OPD) by horseradish peroxidase (HRP) fused to an antibody specific for M13 bacteriophage. Proteins that bound to anti-FLAG at least three-fold over background binding to BSA were chosen for amplification, sequencing, and additional ELISAs.

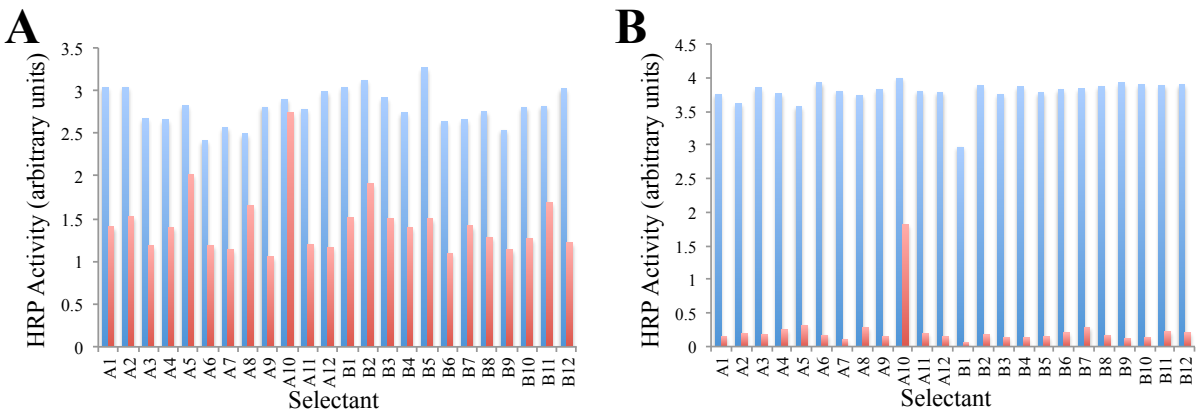
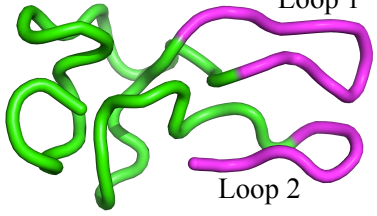


Figure 5-2. Spot assay of selected phage binding to anti-FLAG (blue) versus BSA (red). The identification of binding partners with high signal over background binding to BSA required dilution of phage from a (**A**) high concentration of 20-50 nM to a (**B**) ~6x diluted concentration.

5.5 Amplification, sequencing, and dose-dependency of selected phage-displayed proteins with high affinity for anti-FLAG

Selectants **A9** and **A12** from the echistatin library and **B5** and **B11** from the human lysozyme library were chosen for further investigation as potential high affinity binding partners for anti-FLAG. Selectant DNA was amplified from the isolated double-stranded DNA phagemids of the selectants amplified by polymerase chain reaction (PCR). Sequencing revealed that selectants **A9, A12, B5,** and **B11** were four different echistatin variants. Specifically, **A9, A12,** and **B5** were Loop 2 variants and **B11** was a Loop 1 variant (Figure 5-3). Since **B5** and **B11** were isolated from the human lysozyme selected phage, the library may have become contaminated with some of the echistatin library during selections. ELISAs showed that the four chosen variants bound to anti-FLAG in a dose-dependent manner (Figure 5-4).

A



B

	Loop 1	Loop 2
wild-type	H2N-KRARGDDMD-COOH	H2N-HKGPAT-COOH
A9	H2N-KRARGDDMD-COOH	H2N-SSWVLS-COOH
A12	H2N-KRARGDDMD-COOH	H2N-REPVVR-COOH
B5	H2N-KRARGDDMD-COOH	H2N-SDAISI-COOH
B11	H2N-YGVGGLGRWN-COOH	H2N-HKGPAT-COOH

Figure 5-3. Selected variants of echistatin. (**A**) An NMR structure of echistatin highlights the loops engineered for scaffold libraries in magenta. (**B**) Three variants had alternative sequences in Loop 2 and one variants had an alternative sequence in Loop 1 compared to wild-type echistatin.

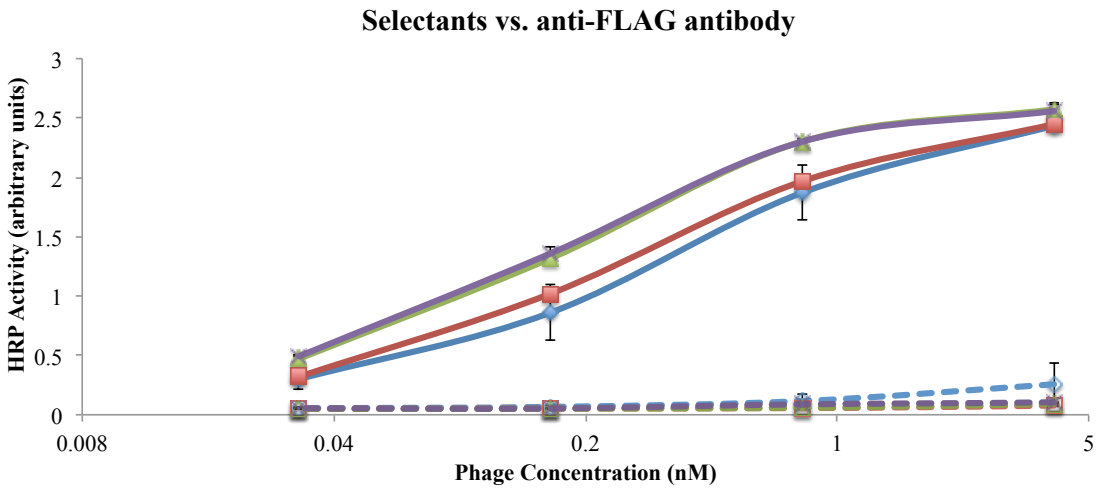


Figure 5-4. Dose-dependent ELISA of selected variants for binding to anti-FLAG target (solid) versus BSA (dashed). Variants **A9** (blue), **A12** (red), **B5** (green), and **B11** (purple) showed clear dose-dependent binding to the anti-FLAG target compared to the blocking agent BSA.

5.6 Passive adsorption of Ure-I and ShuA to immunosorbent microtiter plates

An ELISA was used to measure the immobilization of Ure-I and ShuA by passive adsorption compared to the soluble protein enhanced Green Fluorescent Protein (eGFP) as a positive control. Ure-I and ShuA were diluted in their respective storage buffers, which contained detergents for stabilization of the proteins. Each protein contained a 6x-histidine tag for purification and was detected by incubating the target first with a murine anti-His antibody followed by an HRP, anti-mouse antibody conjugate. The dimerization of OPD provided a colorimetric signal to quantify the extent of immobilization (Figure 5-5). Although it is known that detergents can prevent the passive adsorption of proteins, Ure-I clearly showed immobilization to the immunosorbent microtiter plate. The conditions for ShuA immobilization were repeated under the same conditions as used by previous lab members, but the protein did not appear to be immobilized to the microtiter plate. Therefore, only Ure-I could be targeted using this method of immobilization.

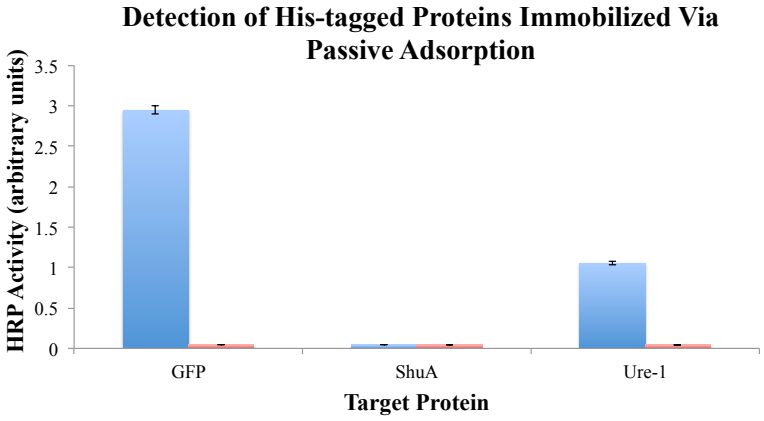


Figure 5-5. Detection of His-tagged proteins immobilized via passive adsorption. ShuA and Ure-I were tested for their immobilization to an immunosorbent microtiter plate using GFP as a positive control. Proteins were detected for binding of an anti-His antibody to the proteins' His-tag (blue) versus background binding of the anti-His antibody to BSA (red).

5.7 Biopanning of binding partners to Ure-I

In vitro evolution of high-affinity binding partners to Ure-I was attempted using the same libraries and protocols described for the mock selections. An additional negative selection step was introduced to ensure that selected phage was binding specifically to Ure-I and not to any of the buffer's components. A spot assay of phage variants from late rounds of selection showed at least one variant that appeared to bind with high affinity to Ure-I (Figure 5-6A). Unfortunately, this variant did not demonstrate the same binding activity in a follow-up dose-dependent assay (Figure 5-6B), despite the variant displaying well on the phage surface (Figure 5-6C). This observation prompted another immobilization assay to reproduce the previous results. The immobilization assay showed that Ure-I, which had been stored at 4 °C for less than 2 weeks, was no longer detectable with an anti-His antibody (Figure 5-6D). Since selections take place over the course of at least one week, the results demonstrate a major challenge for selections that is associated with a lack of Ure-I stability. Rapid Ure-I degradation either prevents passive adsorption onto the immunosorbent microtiter plate or detection with the anti-His antibody. The lack of a consistent, stable target with Ure-I prevents successful *in vitro* evolution of ligands.

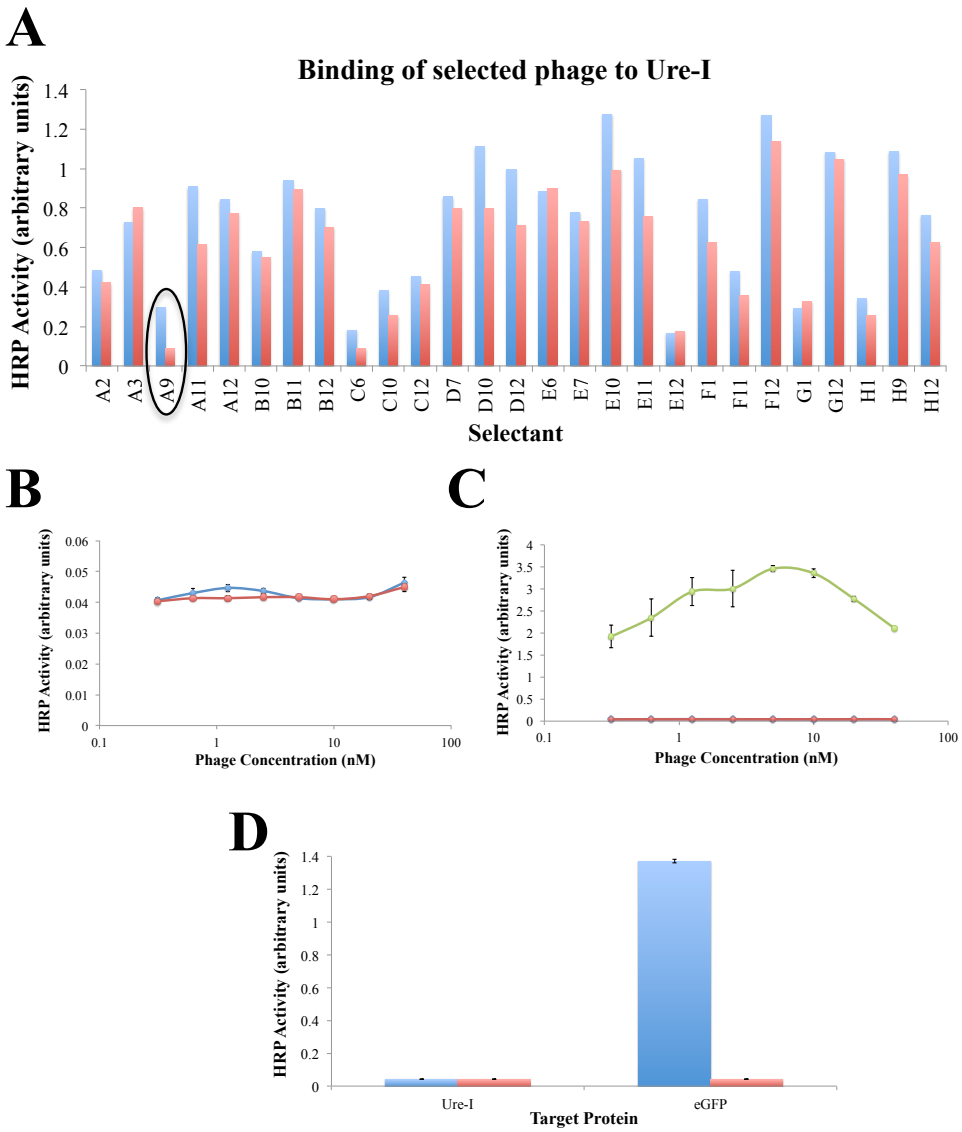


Figure 5-6. Results of *in vitro* evolution for high affinity binding partners to Ure-I. (**A**) Following selections, a spot assay showed at least one variant (A9, circled) as a possible high affinity binding partner to Ure-I (blue) compared to BSA (red). (**B**) A follow-up dose-dependent ELISA showed poor binding to Ure-I (blue) at the same level of binding to BSA (red), despite a high level of display detected with an anti-FLAG antibody as the target (green). (**D**) Measurement of Ure-I immobilization versus the positive control eGFP showed poor detection of the His-tag (blue) compared to BSA (red).

5.8 Development of immobilization strategy for ShuA

ShuA cannot be detected when immobilized using passive adsorption. Various experiments attempted to select for affinity reagents by first incubating with ShuA followed by incubated with known ShuA binding partners immobilized by passive adsorption. These efforts were unsuccessful in generating any high-affinity binding partners from phage-displayed human lysozyme and echistatin libraries (not shown). Challenges associated with using ShuA as a target prompted the development of an alternative strategy.

Homologs of ShuA can be found on the outer membrane of other bacterial organisms, including *E. coli.* This suggested that ShuA could be shuttled to the outer membrane of *E. coli* in a similar manner and potentially targeted on the surface of cells. Although an antibody specific for ShuA is known, this reagent was not initially available. Therefore, a FLAG tag was engineered into Loop 5 using oligonucleotide-directed mutagenesis in order to detect ShuA protein after expression in *E. coli* cells. Previously, this loop stably incorporated a 6x-histidine tag for purification purposes.[31,33] ShuA was expressed on the surface of cells with two expression protocols: one used a condition of slow expression at low temperature with C41(DE3) *E. coli* cells, and the other protocol was a standard expression at 37 °C. A preliminary ELISA demonstrated the highest level of ShuA detection when protein was expressed with a standard protocol and cells were immobilized in 20 mM Tris pH 8.0 (Figure 5-7A).

Testing various factors, such as cell concentration, incubation temperature, time, and centrifugation, further optimized the best conditions for immobilization of cells that expressed ShuA. Cells appeared to immobilize most efficiently to immunosorbent

microtiter plates following overnight incubation at 4 °C, subsequent incubation at 37 °C for 20 min at 180 rpm, then centrifugation for 20 min, 3,000 rpm at 4 °C. Once we received the anti-ShuA antibody, an ELISA using the optimized conditions for immobilization further confirmed detection of ShuA (Figure 5-7B). This immobilization protocol was also used to show that wild-type echistatin phage did not appear to bind non-specifically to cells that expressed ShuA (Figure 5-7C).

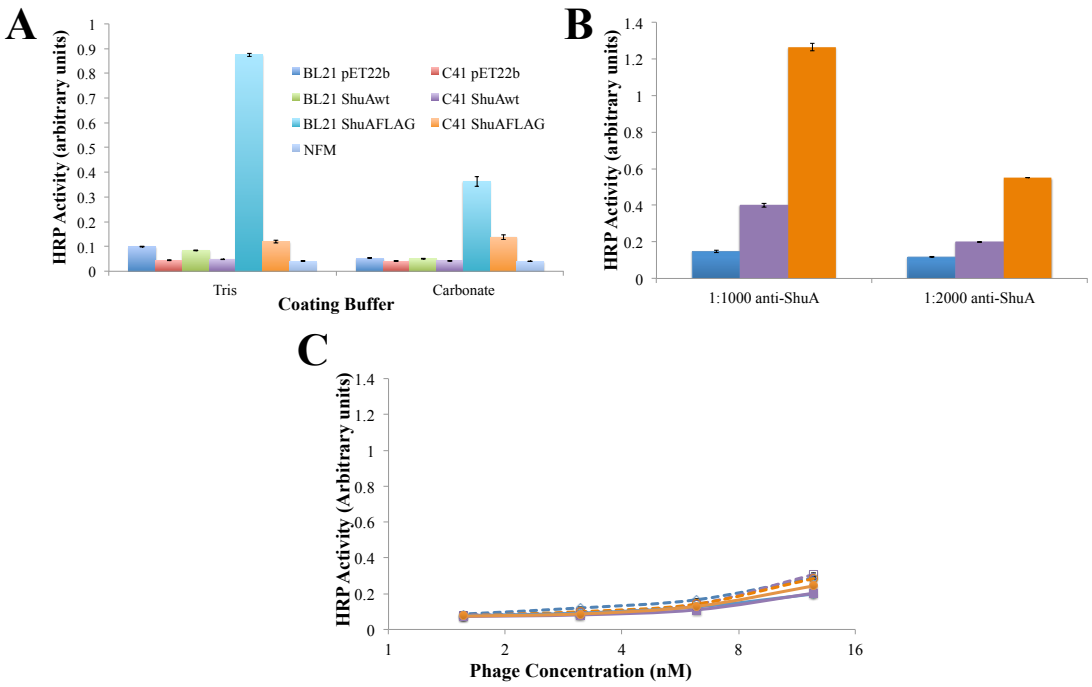


Figure 5-7. Detection of ShuA expressed in whole cells. (**A**) A preliminary immobilization assay detected the FLAG tag of ShuA, and demonstrated better immobilization when the cells were coated with Tris buffer instead of carbonate buffer. (**B**) An anti-ShuA antibody, measured at two different concentrations, confirmed immobilization of ShuA-expressed cells (orange). Cells that expressed ShuA were detected at higher levels than cells that did not express ShuA (purple) or non-fat milk (NFM, blue). (**C**) KO7 phage (solid) and phage that displayed echistatin (dashed) showed minimal non-specific binding to cells expressing ShuA (orange), null cells that did not express ShuA (purple), and NFM (blue).

The rapid degradation of Ure-I shown previously prompted a test of the extent of immobilization for cells expressing ShuA after 4 °C storage for several days. An ELISA showed that the detection of ShuA in cells stored at 4 °C tended to increase the longer they were stored (Figure 5-8A). One possibility for this observation was that the cells lysed over time and allow the cell membrane to stretch out along the bottom of the well for more surface area exposure of ShuA. This hypothesis was tested by detection with the anti-His antibody of the 6x-His tag on the N-terminus of ShuA, which faces toward the periplasm and would be detectable if cells were lysed. Such experiments, however, did not show any measureable detection of the histidine tag (Figure 5-8B). It is not clear why the FLAG tag becomes more easily detectable over a short storage period. Expression and immobilization of cells expressing ShuA are now thoroughly characterized and optimized for subsequent *in vitro* evolution of high affinity binding partners from phage-displayed libraries.

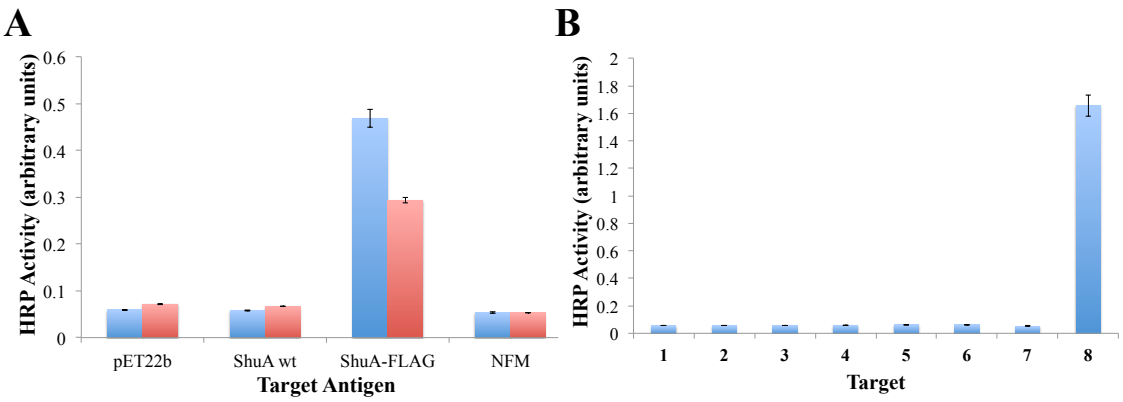


Figure 5-8. Detection of ShuA-expressing cells after storage at 4 °C for several days. (**A**) Stored ShuA-FLAG expressing cells (blue) were more readily detected by anti-FLAG than freshly resuspended ShuA-FLAG cells (red). (**B**) An ELISA attempted to determine whether enhanced anti-FLAG detection was the result of cell lysis. An anti-His antibody showed no visible detection of the 6x-His tag, located inside the cell, for stored ShuAwt (**2**), freshly resuspended ShuAwt (**5**), stored ShuAFLAG (**3**), and freshly resuspended ShuAFLAG (**6**) cells compared to negative controls stored null cells that did not express ShuA (**1**), freshly resuspended null cells (**4**), and BSA (**7**), and the positive control eGFP (**8**).

5.9 Conclusions

Highly crystallizable, soluble proteins with rigid scaffolds that bind with high affinity to membrane proteins could facilitate stabilization of various conformational states of membrane proteins. This stabilization is necessary for crystallogenesis to gather quantitative structural information. High affinity binding partners to membrane proteins may also inhibit molecular transport mechanisms for bacterial virulence. The ability to evolve high affinity binding partners from phage-displayed protein libraries was demonstrated using the anti-FLAG antibody as a target. *In vitro* evolution of binding partners to the *H. pylori* inner membrane protein Ure-I was attempted, but failed due to the apparent rapid degradation of the stored membrane protein. ShuA, an outer membrane protein from *S. dysenteriae*, could not be immobilized for *in vitro* evolution purposes by simple passive adsorption to an immunosorbent microplate. Conditions for the optimal expression and immobilization of ShuA in the *E. coli* outer membrane allowed detection of the protein in whole cells. This protocol maintains ShuA in its near-native environment and promotes the feasibility of *in vitro* evolution for high affinity binding partners using solid- or solution-phase selections.

5.10 Materials and methods

Materials

All chemical reagents and antibiotics were purchased from Fisher Scientific, Thermo Scientific, EMD, Sigma Aldrich or Acros Organics, and were used as received unless otherwise specified. Enzymes were purchased from New England Biolabs. All cell lines and the anti-FLAG M2 murine antibody were purchased from Stratagene. KO7 helper phage was purchased from Amersham Biosciences. All antibodies were purchased

from Sigma except for anti-M13/HRP conjugate (GE Healthcare) and anti-ShuA raised in rabbit (provided by Professor Angela Wilks). Maxisorp 96-well microtiter plates were purchased from Nunc and measured following ELISAs on a Bio-Tek microtiter plate reader.

Dr. Cathie Overstreet generated echistatin and human lysozyme libraries displayed on the surface of phage. Dr. Sarvind Tripathi and Professor Tom Poulos provided purified ShuA. Dr. Jack Chen and Professor Hartmut Luecke provided purified Ure-I.

General procedure for propagation and isolation of phage

Display of echistatin and human lysozyme on phage was possible by transformation of *E. coli* SS320 $CaCl_2$-competent cells with an M13 phagemid containing the protein-P3 fusion and growth on solid media overnight at 37 °C. A single colony was used to inoculate a starter liquid culture of LB media supplemented with 50 μg/mL carbenicillin and 2.5 μg/mL tetracycline. Log phase liquid cultures of transformed cells were infected with ~1.5 x 10^{10} phage/mL helper phage KO7 for 45 min at 37 °C. Infected cells were then transferred to fresh 2YT media supplemented with 50 μg/mL carbenicillin and 15 μg/mL kanamycin for phage propagation at 37 °C for 16-18 h. PEG-NaCl precipitation was used to isolate phage and the concentration of phage was measured using a microtiter plate reader (Bio-Tek) at 268 nm.

***In vitro* evolution of binding partners from phage-displayed echistatin and human lysozyme libraries**

All incubations were performed at room temperature on a 150 rpm orbital shaker. Phosphate buffered saline (PBS, pH 7.2) supplemented with 0.05% Tween-20 was used

for all wash steps unless otherwise specified. Microtiter Maxisorp plates were incubated with 100 μL of 1:10000 (Rounds 1 and 2) or 1:1000 (Rounds 3 and 4) anti-FLAG antibody in carbonate buffer (50 mM Na_2CO_3, pH 9.6) (24, 8, 8, and 8 for each library in four rounds of selection, respectively). After 2 h, 400 µL of 0.2% w/v bovine serum albumin (BSA), nonfat milk (NFM), ovalbumin, or casein in PBS blocked the wells for 30 min in Rounds 1, 2, 3, and 4, respectively. After five washes, 100 μL of phage-displayed scaffold library (echistatin or human lysozyme) in dilution buffer (0.2% blocking agent w/v and 0.05% Tween-20 in PBS) was added to the plate and incubated for 1 h in the binding step. The plate was then washed five times, increasing by three washes each round. HCl (100 µL, 0.1 M) eluted the bound phage by a 10 min sonication in a water bath. After neutralization with 1/3rd volume of 1M Tris-HCl, pH 8.0, half of the neutralized phage was used to infect a 10x volume of log phase *E. coli* SS320 cells grown in LB and incubated at 37 °C for 20 min at 225 rpm. After shaking, 10 µL was removed for titers. Helper phage KO7 was added to the infected cells at ~1.5 x 10^{10} phage/mL and incubated for 45 min at 37 °C, 225 rpm. The culture was then transferred to a 25x volume of 2YT supplemented with 50 µg/mL carbenicillin and 15 µg/mL kanamycin, then incubated 16-18 h at 37 °C, 225 rpm. Phage was isolated from the overnight culture for the subsequent selection round.

Spot assay of selected phage for anti-FLAG

Selected phage displaying proteins from the third and fourth rounds were used to infect log phase *E. coli* SS320 cells (1:10 and 1:100 v/v) and shaken at 37 °C for 20 min. After growth of infected cells on an LB-agar medium supplemented 50 µg/mL carbenicillin, selected colonies were used to inoculate 600 µL 2YT supplemented with 50

μg/mL carbenicillin and 2.5 μg/mL tetracycline in a 96-deep well plate. Log phage cultures were infected with ~1.5 x 10^{10} phage/mL helper phage KO7 for 45 min at 37 °C then transferred to 5 mL fresh 2YT media supplemented with 50 μg/mL carbenicillin and 15 μg/mL kanamycin in a 24-well block. After a 16-18 h incubation at 37 °C overnight, cells were centrifuged at 3,000 rpm, 4 °C for 30 min. Phage was precipitated with PEG-NaCl and resuspended in PBS. The phage-based ELISAs to measure binding of individual proteins to anti-FLAG were performed in a similar format to the selections described above regarding coating, incubations, washes, blocking, and binding steps with some additions. Anti-FLAG (1:1000 in coating buffer, 48 wells) and coating buffer only (48 wells) were incubated on the plate in the coating step. After blocking with BSA and four washes, 100 μL of each isolated phage variant was added to both anti-FLAG-coated wells and BSA-coated wells in the binding step. After four washes, Anti-M13/HRP conjugate (GE Healthcare) (1:5000 in dilution buffer) was incubated in the wells for 30 min. After five washes followed by two washes with PBS, absorbance was measured at 450 nm after 20-25 min with 100 μL OPD solution (2 mg/mL o-phenylenediamine dihydrochloride, 0.02% w/v H_2O_2 in citric acid buffer (50 mM citric acid, 50 mM Na_2HPO_4, pH 5.0)). A separate ELISA measured binding of selected phage to anti-FLAG in a dose-dependent manner. Phage for four variants that showed high affinity binding in the anti-FLAG spot assay were individually propagated, isolated, and resuspended in PBS. These PBS phage solutions were used for the phage-binding step (100 μL) at 0.032, 0.16, 0.8, and 4 nM concentrations.

ELISA to test immobilization of Ure-I and ShuA to immunosorbent microtiter plates

The immobilization of Ure-I and ShuA was determined by the detection of their 6x-His tag. An ELISA was performed as described previously regarding coating, incubations, washes, blocking, and binding steps. The membrane proteins were diluted in their respective buffers (Ure-I in 10 mM MOPS, 150 mM NaCl, 0.2% decylmaltoside, pH 6.5, ShuA in 50 mM Tris, 50 mM NaCl, 2 mM *N*,*N*-dimethyldodecylamine N-oxide, pH 7.8) to 10 μg/mL for coating on the immunosorbent microtiter plate. Plates were blocked with BSA and, after washing, incubated with 100 μL 1:2000 murine anti-His antibody in dilution buffer for 30 min at 150 rpm. Plates were washed four times, then incubated with 100 μL 1:1000 anti-mouse HRP conjugate for 30 min at 150 rpm. Subsequent dimerization of OPD by HRP determined the extent of membrane protein immobilization.

Oligonucleotide-directed mutagenesis to generate ShuA-FLAG

A pET22b plasmid with the gene encoding ShuA and an N-terminal 6x histidine tag was obtained from Dr. Sarvind Tripathi in Tom Poulos' laboratory. A FLAG tag was inserted into loop 5 of ShuA using a previously described method of oligonucleotide-directed mutagenesis.[34] $CaCl_2$-competent CJ236 *E. coli* cells were transformed with the ShuA-pET22b plasmid by heat shock. Cells were plated on LB/agar containing 50 μg/mL carbenicillin and 15 μg/mL chloramphenicol and grown overnight at 37 °C. Single colonies were used to inoculate 6 x 2.5 mL 2YT media supplemented with 50 μg/mL carbenicillin and 5 μg/mL chloramphenicol. Following growth for 6-8 h at 37 °C with shaking, log phase cultures were infected with ~1.5 x 10^{10} phage/mL helper phage

KO7 for 45 min at 37 °C. Cultures were then transferred to fresh 2YT media (50x volume) containing 50 μg/mL carbenicillin, 20 μg/mL kanamycin, and 0.25 μg/mL uridine for 16-18 h growth at 37 °C. PEG-NaCl precipitation (1/5 v/v) was used to isolate phage. The uracil-doped single-stranded DNA template (dU-ssDNA) was isolated with a Qiaprep Spin M13 kit and nucleic acid concentration was measured at 260/280 nm with a Thermo Scientific NanoDrop 2000 UV/Vis spectrophotometer.

An oligonucleotide encoding for the FLAG tag was designed (5′-CGCGCCGCCCGGATGTTTATCATCGTCATCTTTATAATCCAA*GGATCC*TTGTTCCTGACGATAATACTC – 3′) as the reverse complement to the coding strand template DNA. The underlined sequence represents the regions that will anneal to the template DNA and the italicized regions encodes for a glycine-serine linker. For phosphorylation, the oligonucleotude was incubated with ATP, T4 polynucleotide kinase, and DTT for 1 h at 37 °C. The phosphorylated oligonucleotide was annealed to the dU-ssDNA template at a 1:2 mass ratio by heating the mixture to 90 °C for 2 min, then cooling to 50 °C for 3 min followed by cooling to 20 °C for 5 min. The annealed oligonucleotide-templates were then elongated by incubation with dNTPs, ATP, DTT, T7 DNA polymerase, and T4 DNA ligase at room temperature overnight. Reaction products were electrophoresed for 40 min at 95V in a 0.7% agarose gel and the bottom band was excised. Double-stranded covalently closed circular DNA (ds-cccDNA) was isolated with the Zymoclean Gel DNA Recovery kit and used to transform $CaCl_2$-competent Top10 *E. coli* cells by heat shock. After growth overnight at 37 °C on LB/agar supplemented with 50 μg/mL carbenicillin, DNA from individual colonies was amplified by PCR and the desired mutation was confirmed by sequencing.

ShuA expression and cell purification

Expression of ShuA in the membrane of *E. coli* cells was tested using two different protocols. In protocol **1**, *E. coli* C41(DE3) cells were transformed with the pET22b plasmid encoding the gene for ShuA-FLAG and protocol **2** transformed *E. coli* BL21 cells. For control experiments, empty pET22b plasmids and pET22b plasmid encoding the gene for wild-type ShuA were subjected to the same transformation, expression, and purification protocols. Cells were plated and grown on solid media supplemented with 50 μg/mL carbenicillin overnight at 37 °C. For each transformation, a single colony was used to inoculate LB supplemented with 50 μg/mL carbenicillin for growth in liquid media overnight at 37 °C. In protocol **1**, 1.5 mL of the overnight culture was used to inoculate 150 mL TB media supplemented with 50 μg/mL carbenicillin and the culture was incubated at 22 °C, 80 rpm for 6-8 h until $OD_{600} = 0.6$. For protocol **2**, fresh LB media supplemented with 50 μg/mL carbenicillin and inoculated with the 1% overnight culture was grown at 37 °C, 225 rpm for ~2 h until $OD_{600} = 0.6$. For **1** and **2**, 1 mM IPTG was added to log phage cultures to induce ShuA expression. **1** expressed ShuA at 22 °C, 80 rpm for 12-16 h and **2** expressed ShuA at 37 °C, 225 rpm for 3-4 h. Cells for both cultures were harvested at 3k rpm for 30 min at 4 °C and cell pellets were flash frozen in dry ice/ethanol, then stored at -80 °C.

To prepare cells for immobilization to immunosorbent microtiter plates and remove all traces of LB media, cell pellets were thawed and resuspended in 10 mL buffer (20 mM Tris, pH 8.0), then centrifuged at 3k rpm for 30 min at 4 °C. Cells were resuspended again in 5 mL buffer and transferred to two separate 15-mL Falcon tubes before centrifugation at 10k rpm for 10 min at 4 °C. Cell pellets were then washed with

either 20 mM Tris buffer or carbonate buffer (50 mM Na_2CO_3, pH 9.6) by resuspension in 1 mL, followed by centrifugation at 10k rpm for 10 min at 4 °C. Cells were ultimately resuspended in 1 mL of the same buffer, either Tris or carbonate, and the concentration of cells was measured using a microtiter plate reader (Bio-Tek) at 600 nm. Cells were diluted to 3.0 x 10^8 cells/mL with either Tris or carbonate buffer.

Immobilization of ShuA-expressing bacterial cells was adapted from previously described protocols.[35–37] Optimal conditions for immobilization adhered cells (150 µL at 3.0 x 10^8 cells/mL) to microplate wells by incubation at 4 °C overnight followed by incubation the next morning at 37 °C with shaking at 180 rpm for 30 min. The plate was then centrifuged at 3k rpm, 4 °C for 20 min. Wells were washed three times with PBS then blocked for 30 min, 150 rpm at room temperature with 300 µL 0.2% NFM. After washing three times with PBST, 100 µL anti-FLAG M2 murine antibody (1:5000) was added to each well and incubated at room temperature for 30 min, 150 rpm. Wells were then washed three times with PBST and incubated with 100 µL HRP, anti-mouse antibody conjugate for 30 min, 150 rpm at room temperature. OPD addition was then followed as described previously. An ELISA to determine if the 6x-His tag could be detected was performed with a similar protocol, but replaced the anti-FLAG binding step with incubation of 100 µL anti-His murine antibody (1:2000 in dilution buffer). KO7 and echistatin phage, propagated and isolated as described previously, was tested for non-specific binding to the ShuA-expressing cells. For this ELISA, phage was detected with the HRP, anti-M13 antibody conjugate as described for phage-binding assays using the anti-FLAG antibody as the target. For detection of ShuA using the anti-ShuA antibody, the anti-FLAG step was replaced with incubation of anti-ShuA antibody raised in rabbit

(1:1000 or 1:2000 in dilution buffer). An HRP, anti-rabbit antibody conjugate (100 μL, 1:1000 in dilution buffer) was then used in place of anti-mouse HRP to detect the anti-ShuA antibody.

5.11 References

1. Sachs, J. N. & Engelman, D. M. Introduction to the membrane protein reviews: the interplay of structure, dynamics, and environment in membrane protein function. *Annu. Rev. Biochem.* **75,** 707–12 (2006).

2. Almén, M. S., Nordström, K. J. V, Fredriksson, R. & Schiöth, H. B. Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin. *BMC Biol.* **7,** 50 (2009).

3. Wallin, E. & von Heijne, G. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci.* **7,** 1029–38 (1998).

4. White, S. H. Biophysical dissection of membrane proteins. *Nature* **459,** 344–6 (2009).

5. Berman, H., Henrick, K. & Nakamura, H. Announcing the worldwide Protein Data Bank. *Nat. Struct. Biol.* **10,** 980 (2003).

6. Overington, J. P., Al-Lazikani, B. & Hopkins, A. L. How many drug targets are there? *Nat. Rev. Drug Discov.* **5,** 993–6 (2006).

7. Carpenter, E. P., Beis, K., Cameron, A. D. & Iwata, S. Overcoming the challenges of membrane protein crystallography. *Curr. Opin. Struct. Biol.* **18,** 581–6 (2008).

8. Shipley, G. G. Lipids: Bilayers and nonbilayers: structure, forces and protein crystallization - Editorial overview. *Curr. Opin. Struct. Biol.* **4,** 471–473 (2000).

9. Seddon, A. M., Curnow, P. & Booth, P. J. Membrane proteins, lipids and detergents: not just a soap opera. *Biochim. Biophys. Acta* **1666,** 105–17 (2004).

10. Bayburt, T. H. & Sligar, S. G. Membrane protein assembly into Nanodiscs. *FEBS Lett.* **584,** 1721–7 (2010).

11. Cross, T. A., Sharma, M., Yi, M. & Zhou, H.-X. Influence of solubilizing environments on membrane protein structures. *Trends Biochem. Sci.* **36,** 117–25 (2011).

12. Yao, Y., Ding, Y., Tian, Y., Opella, S. J. & Marassi, F. M. Membrane protein structure determination: back to the membrane. *Methods Mol. Biol.* **1063,** 145–58 (2013).

13. Ujwal, R. & Abramson, J. High-throughput crystallization of membrane proteins using the lipidic bicelle method. *J. Vis. Exp.* **59**, e3383 (2012).

14. Zou, Y., Weis, W. I. & Kobilka, B. K. N-terminal T4 lysozyme fusion facilitates crystallization of a G protein coupled receptor. *PLoS One* **7,** e46039 (2012).

15. Thorsen, T. S., Matt, R., Weis, W. I. & Kobilka, B. K. Modified T4 lysozyme fusion proteins facilitate G protein-coupled receptor crystallogenesis. *Structure* **22,** 1657–1664 (2014).

16. Blake, C. C. F. & Swan, I. D. A. X-ray analysis of structure of human lysozyme at 6 Å resolution. *Nature* **232,** 12–15 (1971).

17. Artymiuk, P. J. & Blake, C. C. F. Refinement of human lysozyme at 1.5 Å resolution analysis of non-bonded and hydrogen-bond interactions. *J. Mol. Biol.* **152,** 737–762 (1981).

18. Gan, Z., Gould, R., Jacobs, J., Friedman, P. & Polokoff, M. Echistatin. A potent platelet aggregation inhibitor from the venom of the viper, *Echis carinatus*. *J. Biol. Chem.* **263,** 19827–19832 (1988).

19. Saudek, V., Atkinson, R. A. & Pelton, J. T. Three-dimensional structure of echistatin, the smallest active RGD protein. *Biochemistry* **30,** 7369–7372 (1991).

20. Chen, Y. *et al.* Proton NMR assignments and secondary structure of the snake venom protein echistatin. *Biochemistry* **30,** 11625–11636 (1991).

21. Monleón, D., Esteve, V., Kovacs, H., Calvete, J. J. & Celda, B. Conformation and concerted dynamics of the integrin-binding site and the C-terminal region of echistatin revealed by homonuclear NMR. *Biochem. J.* **387,** 57–66 (2005).

22. Smith, G. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **228,** 1315–1317 (1985).

23. Scott, D. R. *et al.* Expression of the *Helicobacter pylori* ureI gene is required for acidic pH activation of cytoplasmic urease. *Infect. Immun.* **68,** 470–7 (2000).

24. Rektorschek, M. *et al.* Acid resistance of *Helicobacter pylori* depends on the UreI membrane protein and an inner membrane proton barrier. *Mol. Microbiol.* **36,** 141–152 (2000).

25. Bury-Mone, S., Skouloubris, S., Labigne, A. & De Reuse, H. The *Helicobacter pylori* UreI protein: role in adaptation to acidity and identification of residues essential for its activity and for acid activation. *Mol. Microbiol.* **42,** 1021–1034 (2001).

26. Strugatsky, D. & McNulty, R. *et al.* Structure of the proton-gated urea channel from the gastric pathogen *Helicobacter pylori. Nature* **493,** 255–8 (2013).

27. McNulty, R., Ulmschneider, J. P., Luecke, H. & Ulmschneider, M. B. Mechanisms of molecular transport through the urea channel of *Helicobacter pylori*. *Nat. Commun.* **4,** 2900 (2013).

28. Mills, M. & Payne, S. M. Identification of shuA, the gene encoding the heme receptor of *Shigella dysenteriae*, and analysis of invasion and intracellular multiplication of a shuA mutant. *Infect. Immun.* **65,** 5358–63 (1997).

29. Burkhard, K. A. & Wilks, A. Characterization of the outer membrane receptor ShuA from the heme uptake system of *Shigella dysenteriae*. Substrate specificity and identification of the heme protein ligands. *J. Biol. Chem.* **282,** 15126–36 (2007).

30. Cherla, R. P., Lee, S.-Y. & Tesh, V. L. Shiga toxins and apoptosis. *FEMS Microbiol. Lett.* **228,** 159–166 (2003).

31. Brillet, K., Meksem, A., Thompson, A. & Cobessi, D. Expression, purification, crystallization and preliminary X-ray diffraction analysis of the TonB-dependent haem outer membrane transporter ShuA from *Shigella dysenteriae. Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun.* **65,** 402–5 (2009).

32. Cobessi, D., Meksem, A. & Brillet, K. Structure of the heme/hemoglobin outer membrane receptor ShuA from *Shigella dysenteriae*: heme binding by an induced fit mechanism. *Proteins* **78,** 286–94 (2010).

33. Ferguson, A. D., Breed, J., Diederichs, K., Welte, W. & Coulton, J. W. An internal affinity-tag for purification and crystallization of the siderophore receptor FhuA, integral outer membrane protein from *Escherichia coli* K-12. *Protein Sci.* **7,** 1636–8 (1998).

34. Kunkel, T. A. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl. Acad. Sci.* **82,** 488–492 (1985).

35. Elder, B. L., Boraker, D. K. & Fives-Taylor, P. M. Whole-bacterial cell enzyme-linked immunosorbent assay for *Streptococcus sanguis* fimbrial antigens. *J. Clin. Microbiol.* **16,** 141–4 (1982).

36. Ley, P., Amesz, H., Tommassen, J. & Lugtenberg, B. Monoclonal antibodies

directed against the cell-surface-exposed part of PhoE pore protein of the *Escherichia coli* K-12 outer membrane. *Eur. J. Biochem.* **147,** 401–407 (1985).

37. Watanabe, K. *et al.* Development and clinical application of an immunoassay using intact *Helicobacter pylori* attached to a solid phase as an antigen. *Clin. Biochem.* **34,** 291–295 (2001).

Chapter 6: Conclusions

6.1 Conclusions

DNA polymerases are responsible for the accurate replication of DNA for the continued survival of organisms. These enzymes copy DNA with extraordinary fidelity through careful selection of deoxynucleoside triphosphate (dNTP) substrates. Kinetic analysis of DNA polymerase I Klenow Fragment (KF) determined the presence of important conformational changes before and after phosphodiester bond formation to achieve this fidelity.[1] X-ray crystal structures of related DNA polymerases visualized "open," "ajar," and "closed" conformations with different positions of a dynamic "fingers" subdomain.[2] Such structures guided placement of fluorescent probes to study the biophysics of this conformational transition in real-time.[3,4] Most recently, single-molecule fluorescence-based assays have observed transient dynamics of DNA polymerase motions in various complexes and at various stages of the catalytic cycle.[5–7]

Electronic-based methods for the study of DNA polymerases can achieve enhanced time resolution compared to fluorescence-based methods without the requirement for labeled enzymes or substrates. Sub-millisecond timescales of DNA polymerase attained through the use of nanopores[8] and electronic nanocircuits[9,10] highlight the advantages of electronic-based techniques. Single-walled carbon nanotube field-effect transistors (SWCNT-FETs) enable the measurement of KF activity at the single-molecule level through chemical modification of a single-cysteine residue on the protein's surface.[9,10] By observing enzyme motion through the mobile fingers subdomain, base incorporation events can be analyzed with SWCNT-FETs. Charged

amino acid residues on the surface of the protein proximal to the attachment site are responsible for electronically gating the SWCNT. Thus, enzyme motions result in current fluctuations when the conductivity of the SWCNT is altered.[11]

We were interested in challenging the molecular recognition by KF through the use of dNTP analogs (Chapter 2). Seven dNTP analogs were tested for their incorporation into a homopolymeric DNA template. Analogs with altered reactive centers had slower rates of incorporation compared to the corresponding native dNTPs,[10] likely due to decreased processivity after unstable helix formation. Nucleobase-modified dNTP analogs had varying effects on the incorporations rates compared to the corresponding native dNTPs. More interestingly, these base modifications caused an opposite change in the current signal during their incorporation that corresponded to alternative enzyme motions. For the analog 6-chloro-2-aminopurine dNTP (6-Cl-2APTP), in particular, this opposite current state occurred during every incorporation event opposite a template T or C base.[10] The results indicate the potential use of this dNTP analog in the discrimination of specific bases in a DNA template, which is applicable to biotechnological applications such as DNA sequencing.

For DNA sequencing and other applications, our device must accurately measure the DNA template length. We wanted to determine the ability of our device to count the number of bases in a template with a highly repetitive sequence (Chapter 3), which is notoriously error-prone during DNA replication by a DNA polymerase.[12,13] KF-functionalized SWCNT-FETs were incubated with DNA templates alone or in a mixture at limited concentrations for clearly separated clusters of activity to aid in analysis. Following strict guidelines for the identification of a single base incorporation, DNA

template lengths were measured and DNA templates of lengths differing by four bases were identified in a mixture. More precise determination of DNA template lengths benefited from the accumulation of multiple data sets. Preliminary experiments at higher bandwidths revealed short-lived conformational dynamics that occurred with high frequency and accompanied signals corresponding to base incorporations. These results underscore the importance of a limited time resolution for measuring DNA template lengths to prevent complicated analysis. Higher bandwidth measurements will be more useful for measuring transient KF motions during dNTP incorporation, but require increased signal to noise ratios.

To better understand the nature of KF-generated electronic signals and improve these signals for biotechnological applications, several experiments and variants of KF were designed and generated (Chapter 4). Measurements of KF with ribonucleoside triphosphates, for example, could provide insight into the dependence of the observed enzyme motion on ribose recognition. Such measurements, in addition to measurements with the inhibitor pyrophosphate, were attempted and must be reproduced. Variants of KF that exert low fidelity, slow the reaction rate, or restore 5′-3′ exonuclease activity were designed and generated to identify molecular recognition, phosphodiester bond formation, and exonucleolysis, respectively, in our observed electronic signals. Point-charge variants aim to determine the charged amino acid residues responsible for the generation of electronic signals. Additional KF variants were generated to concentrate the surface charge proximal to the SWCNT-FET, increase processivity through addition of a double-stranded DNA binding protein, and generate unique signals corresponding to alternative stages of catalysis through different attachment sites. These variants have the

potential to improve KF-functionalized SWCNT-FETs for use in a variety of important applications.

Similar to DNA polymerases, membrane proteins undergo a series of conformational transitions for the survival of an organism. Specifically, these membrane proteins have complex mechanisms for molecular transport, and the determination of their structures at various stages of their transport mechanism can guide experimental design and inhibition with drugs. In an effort to achieve X-ray crystal structures of the membrane proteins ShuA from *Shigella dysenteriae* and Ure-I from *Helicobacter pylori*, I attempted to select for high affinity binding partners from phage-displayed libraries of the stable, rigid proteins echistatin and human lysozyme. Such binding partners could stabilize the membrane proteins for successful crystallization. *In vitro* evolution of binding partners for Ure-I, immobilized by passive adsorption, was unsuccessful due to rapid degradation of the membrane protein. ShuA, which could not be passively adsorbed for immobilization, was expressed on the surface of bacterial cells. The immobilization of these ShuA-expressing cells and detection of ShuA on the cell surface was optimized for future use in the *in vitro* evolution of high affinity binding partners.

Various biomolecules benefit significantly from the determination of their structures and the observation of their biophysical mechanisms. SWCNT-FETs can be functionalized with DNA polymerases, for example, to monitor conformational transitions of these enzymes. The occurrence of different current stages that correspond to the DNA polymerase enzymatic mechanism can also serve biotechnological applications. This powerful technique can be applied to a range of proteins to provide a

thorough understanding of their dynamic motions. Strategies for membrane protein stabilization could permit membrane proteins to also benefit from this technique.

6.2 References

1. 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,** 4835–4843 (1991).

2. Wang, W., Wu, E. Y., Hellinga, H. W. & Beese, L. S. Structural factors that determine selectivity of a high fidelity DNA polymerase for deoxy-, dideoxy-, and ribonucleotides. *J. Biol. Chem.* **287,** 28215–26 (2012).

3. Allen, W. J., Rothwell, P. J. & Waksman, G. An intramolecular FRET system monitors fingers subdomain opening in Klentaq1. *Protein Sci.* **17,** 401–408 (2008).

4. Joyce, C. M. *et al.* Fingers-closing and other rapid conformational changes in DNA polymerase I (Klenow fragment) and their role in nucleotide selectivity. *Biochemistry* **47,** 6103–16 (2008).

5. Santoso, Y. & Joyce, C. M. *et al.* Conformational transitions in DNA polymerase I revealed by single-molecule FRET. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 715–20 (2010).

6. 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).

7. Evans, G. W., Hohlbein, J., Craggs, T., Aigrain, L. & Kapanidis, A. N. Real-time single-molecule studies of the motions of DNA polymerase fingers illuminate DNA synthesis mechanisms. *Nucleic Acids Res.* **43,** 5998–6008 (2015).

8. Garalde, D. R. *et al.* Distinct complexes of DNA polymerase I (Klenow fragment) for base and sugar discrimination during nucleotide substrate selection. *J. Biol. Chem.* **286,** 14480–92 (2011).

9. Olsen, T. J. & Choi, Y. *et al.* Electronic measurements of single-molecule processing by DNA polymerase I (Klenow fragment). *J. Am. Chem. Soc.* **135,** 7855–60 (2013).

10. Pugliese, K. M. & Gul, O. T. *et al.* Processive incorporation of deoxynucleoside triphosphate analogs by single-molecule DNA polymerase I (Klenow fragment) nanocircuits. *J. Am. Chem. Soc.* **137,** 9587–94 (2015).

11. Choi, Y. & Olsen, T. J. *et al.* Dissecting single-molecule signal transduction in carbon nanotube circuits with protein engineering. *Nano Lett.* **13,** 625–31 (2013).

12. Shendure, J. & Ji, H. Next-generation DNA sequencing. *Nat. Biotechnol.* **26,** 1135–45 (2008).

13. Budworth, H. & McMurray, C. T. A brief history of triplet repeat diseases. *Methods Mol. Biol.* **1010,** 3–17 (2013).