UNIVERSITY OF CALIFORNIA,
IRVINE

Advancing DNA Sequencing Technologies, Antibody Therapeutics, and Viral Diagnostics

DISSERTATION

submitted in partial satisfaction of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in Biological Sciences

by

Kristin Nichelle Gabriel

Dissertation Committee:
Professor Gregory A. Weiss, Chair
Professor Jennifer A. Prescher
Professor Melanie J. Cocco

2021

DEDICATION

To

My mentors,

For giving me inspiration, guidance, and confidence to climb new heights

My friends and family,

For your endless love, support, and encouragement every step of the way

"Look deep into nature, and then you will understand everything better."
– Albert Einstein

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

Page

ACKNOWLEDGEMENTS

I would like to thank all the people who have supported me during the past five years of my doctoral journey.

Every journey begins with a first step. In this regard, I am forever thankful to my Ph.D. advisor, Professor Gregory Weiss, for accepting me into his chemical biology laboratory and letting me embark on a PhD under his direction and guidance. Greg is someone you will instantly be drawn to and never forget once you meet him. I hope that I could be as lively, enthusiastic, and energic as Greg and someday be able to captivate an audience as well as he can. Greg not only provides me with mentorship to emulate, but also challenges me to think creatively about science. Although I had extensive research previously, I have been exposed to the idea of pushing the boundary of science. I appreciate his scientific advice and knowledge, and the confidence he has ensued in me to succeed in my research career.

I would like to express my deepest appreciation to my committee members. My advancement to candidacy committee comprised of phenomenal women in science – Professor Celia Goulding, Professor Lan Huang, Professor Melanie Cocco, and Professor Jennifer Prescher. The latter two professors also served on my dissertation committee. Celia encouraged and expected me to think more independently about my experiments and results. Lan is a great mass spectrometrist and has taught me a lot about experimental design and data analysis for my bioconjugation research. Melanie served as my first-year advisor in the Cellular and Molecular Biosciences (CMB) program, and I am thankful for the help selecting my courses and rotations. Throughout the years, she has provided constructive criticism, invaluable insights, and enthusiasm for my work in

structure-function relationship of DNA polymerases. Jenn has provided unparallel support and profound belief in my abilities. I am thankful for her enthusiasm in my work and her practical chemical biology suggestions to elevate my research. Jenn also extended a great amount of assistance by allowing me and my team to use her laboratory's instrumentation for experiments. Over the course of my graduate career, my committee members have shared helpful career advice that has helped me become a confident and successful woman in science.

My dissertation would not have been possible without the collaborative support and guidance of Professor Philip Collins. He has challenged me to think critically about science, from experimental design to data interpretation. Phil has provided feedback on several of my practice oral presentations and suggestions on how to improve my science communication skills. In the last six months of my graduate career, he has provided an excellent sounding board to discuss knowledge gaps in DNA polymerase literature and single-molecule studies, which helped with framing the significance of our single-molecule studies using nanocircuits.

I am extremely grateful to the past and present Collins laboratory members for collaboration over the past five years. The two post-docs, Dr. Narenda Kumar and Dr. Denys Marushchak, were both helpful physicists who supported me during my first and second years in graduate school. I am thankful for Narenda and Denys sharing their expertise in bioelectronics and teaching me the fundamentals of our nanocircuit approach. I am grateful to have worked alongside former graduate students, Dr. Mackenzie Turvey and Dr. Calvin Lau. Mackenzie broke down complex physics data into digestible results and provided intellectual discussions about DNA polymerase dynamics.

He continually conveyed a spirit of adventure regarding research and scholarship, and an excitement revolving around bioelectronics. Mackenzie was relentless in helping me write an interesting story about single-molecule Taq DNA polymerase dynamics at elevated temperatures (Chapter 2). Calvin explained the physics and computational concepts behind our single-molecule studies. His friendly, easy-going, and all around kind and bright personality have made working together an absolute joy. Calvin was instrumental in teaching me to organize my references and write a cohesive story about Φ29 DNA polymerase pausing (Chapter 3). I am thankful to have worked with Jeffrey Taulbee on single-molecule studies of Taq DNA polymerase during the second half of my graduate career. He is diligent, organized, and a creative thinker in terms of experimental design. Jeff was supportive in helping me prepare for multiple conference presentations in my fourth and fifth years of graduate school.

I would like to thank the current members of the Weiss laboratory. I am grateful for Sanjana Sen, who has served as my role model from my first year rotation in the Weiss laboratory. She is a resilient, hard-working individual who has inspired me to think creatively when it comes to translating basic science research into useful tools. Thank you to Emily Sanders for your support and comic relief as we both went through this graduate school journey together – we officially joined the Weiss laboratory a few days apart and now will complete our PhDs back to back. It was such a great opportunity to have collaborated with both Sanjana and Emily during our final year and share a co-first author journal article on antibody responses to COVID-19 severity (Chapter 5). I am grateful to Michael Spano for fixing any broken laboratory equipment and providing outside-of-the-box ideas on future directions for my research projects. He has also been

a great source of optimism and provided me with good company while hiking the mountains or snorkeling along the coast. I want to thank Brian Miller, Wyatt Ramirez, and Jason Garrido for the many conversations about exciting science as well as encouragement and support during the second half of my graduate career. My appreciation also goes to the current undergraduate students, including Keertna Bhuvan, Aidan Gelston, Brandon Lai, Assa Magassa, Arjun Pamidi, Gabriela Salcedo, Brandon Tram, and Elizabeth Wiriadinata, for inspiring me to be a better scientist. Best wishes on future research projects in the Wiess laboratory!

I would like to thank all the undergraduate, masters, or first year PhD students whom I have mentored over the years. Many thanks to Chanelle Hunter, Sloan Lewis, Alicia Santos, Heather Romero Mercieca, Joseph Garcia, Joshua Kim, Rebecca Vargas, David Narel, Jenifer Pham, Vy Pham, Cynthia Chen, Kim Hoang, Sean Tang, and Elizabeth Wiriadinata. I have learned so much while working alongside each of you, and I cannot thank you enough for the impact you have made on my personal and professional growth. I've learned how to better manage my time, projects, and mentees and how to communicate the big picture behind my research. Thank you for inspiring me each day. I wish you all the best in your future endeavors!

I also want to thank past members of the Weiss laboratory with whom I overlapped in my doctoral journey. Dr. Sudipta Majumdar taught me everything I know about cloning, protein expression and purification, and enzyme activity characterization. His experimental guidance and troubleshooting helped me overcome many obstacles in my research. I am so grateful for Sudipta reviewing my writing including grant proposals, advancement reports, manuscript drafts, and presentation materials over the course of

my graduate career. Dr. Mark Richardson guided me through the scientific process from start to finish. He has served as my mentor in many ways, such as, explaining chemical concepts, teaching experiment design using proper controls, and analyzing complex data. I am much indebted to Mark for the opportunity to share a co-first author journal article on bioconjugation (Chapter 4). Cheers mate! Sincere thank you to Dr. Rebekah Dyer for also being a source of positivity and a brilliant researcher. She has served as a mentor and role model, training me on how to conduct meticulous experiments and how to engage and inspire women in science. I treasure working on the same team as Rebekah and the opportunity to co-author journal articles together. I am thankful for her tips and tricks on science visualization as well as her delicious baked goods. I would also like to thank Dr. Kaitlin Pugliese, who set a great foundation from which to build upon this dissertation. Specifically, her impeccable organization of laboratory notebooks and inventory set me up for great success in graduate school. She also challenged me to find answers to questions previously unanswered about DNA polymerase at the single-molecule level. Dr. Mariam Iftikhar introduced me to the world of mass spectrometry, which was a key technique that made a significant contribution to charactering bioconjugates (Chapter 4). She also showed me how to break down complex concepts to a general audience. Dr. Josh Smith taught me early on about scaling up protein expression and how to view mistakes as an opportunity for learning. Thank you to Dr. Gaetano Speciale for being my bay-mate and role model. He was a creative post-doc who showed me the value of taking a multidisciplinary approach while pushing the limits of science. I am thankful for his advice on how to manage undergraduates, dress to impress, and navigate the business of science. Thanks to Dr. Joshua Britton for also acquainting me with how scientists can

thrive as entrepreneurs. Dr. Madison Fletcher made science engaging and exciting in the laboratory not only with her STEM themed attire but also her outgoing, radiant personality. I am grateful for Madison reviewing my writing, especially my NSF grant materials, and recommending science outreach and professional development activities. I appreciated Dr. Aisha Attar for her kindness and positive, can-do attitude. Her warm smile brightened my face every day we worked together. It was also a pleasure working alongside Ryan Le Tourneau, Jessica Pham, Jeffrey Briggs, Shea Patterson, Jessica Phillips, Emily Crawley, John Hong, Christian Totoiu, Jessica Fong, Hariny Isoda, Katrin Safronuk, Thu Nguyen, Johnathon Truong, Ashby Shareen, Kohi Nguyen, Andrew Chu, Carlos Vasquez, Marcus Woodworth, and Kevin Johnston.

I would like to thank the laboratories who shared reagents and laboratory instrumentation, especially the Martin, Prescher, Heyduk, Goulding, Edinger, Wang, and Poulos laboratories. My neighbors were a breath of fresh air amongst any chaos that I was facing, especially Marc Sprague-Piercy, Brenna Norton-Baker, Jessica Kelz, Anastasia Ionkina, Anna Love, Sierra Williams, Sean Nguyen, Zi Yao, Dr. Drew Ferreira, Brendan Zhang, Bronte Charette, Dr. Bonnie Cuthbert, Rodger de Miranda, Matt Lewis, and Alec Follmer.

Thank you to Professor Chang Liu and Professor Christopher Hughes for the valuable rotation experiences during my first year. I appreciated learning transferable skills and the excellent project guidance and mentorship from Dr. Rishi Jajoo, Arjun Ravikumar, Garri Arzymanyan, and Hugh Bender.

I would like to thank several funding organizations for financial support of my dissertation research. This work described here has been supported in part by the CMB

Program, National Human Genome Research Institute of the National Institute of Health (1R01HG009188-01), and National Science Foundation Graduate Research Fellowship Program (NSF GRFP) (DGE-1839285). I want to also thank the journals (ACS Bioconjugate Chemistry, mSphere, and ACS Synthetic Biology) for permission to include my research publications in my dissertation.

Thank you to the University of California, Irvine for cultivating a cutting-edge research environment with facilities and programs to help me achieve success. The Mass Spectrometry Facility was instrumental in helping me characterize bioconjugates (Chapter 4), and I greatly appreciated the advice and creative suggestions from Dr. Felix Grun and Ben Katz. The Graduate Division extended a great amount of assistance with my academic writing, especially for my NSF GFRP grant and parts of my dissertation. Likewise, Professor Michael Mulligan and Dr. Lori Greene provided practical suggestions on how to elevate my NSF GRFP grant. I would also like to recognize the career guidance that I received from Lauren Lyon through the Division of Career Pathways.

I would like to gratefully acknowledge the CMB program and Department of Molecular Biology and Biochemistry (MB&B) for accepting me into their program and department, respectively. The wonderful staff was helpful and friendly, including Renee Frigo, Gary Roman, Morgan Oldham, Bessy Varela, and Jackie Ramirez. The transition from college to graduation school proved challenging at times. Many thanks to my CMB cohort, especially Andrew Oliver, Andrew Wheat, Jasper Liedtke, Joann Phan, Karissa Munoz, and Rachel Waymack for helpful support during my first year of graduate school. My colleagues in the MB&B department also provided valuable peer discussion during

research in progress talks. Additionally, several professors opened my eyes to new experimental approaches during classes and journal clubs.

Thanks should also go to Graduate Professional Success for PhD students and Post-docs in STEM (GPS-STEM) for training me how to become a polished professional, build my scientific network, and become think tank leader. I learned an immense about of knowledge outside my main research focus, such as data analysis, industry insights, and clinical trials through GPS-STEM events and sponsored courses at UCI's Continuing Division of Education. Special thanks to Dr. Emma Flores-Kim, Dr. Harinder Singh, and Professor David Fruman for being my role models and cheerleaders throughout my PhD. Bri McWhorter was instrumental in helping me find new ways of describing my research and coaching me through GPS-STEM's elevator pitch and business pitch competitions. It has been an absolute pleasure to connect and build meaningful relationships with several GPS-STEM think tank members, including Caitlin Suire, Megan Newcomb, Lee-or Herzog, Duc Phan, Bri Caver, Hamsa Gowda, and Elizabeth Selwan.

I would like to thank the Association for Women in Science for bringing together women in STEM disciplines and advocating for the advancement of woman in science. I am extremely grateful to have served as the Professional Development Chair. I would also like to thank DeviceAlliance for welcoming me into a tight-knit group of professionals and the opportunity to be a founding member and Career Development Chair for the Young Professionals Committee. Through both organizations, I enjoyed implementing activities, such as a strengths finder workshop, career navigation event, and mentorship programs. Managing time and planning activities played an important role in becoming a more productive and effective researcher.

I am very grateful for my undergraduate role models who were instrumental in encouraging me to pursue a PhD and preparing for success in graduate school. During my first semester of college at Chapman University, Professor Melissa Rowland-Goldsmith and Professor Elaine Schwartz shared their research experiences and inspired me to pursue a career in biochemistry. I am much indebted to my undergraduate advisors, Professor Jason Keller and Professor Marco Bisoffi, for giving me the opportunity to do research. My laboratory experiences spanning fields of wetland biogeochemistry and prostate cancer laid a solid foundation for graduate school. I am grateful to have had variety of learning opportunities from setting up lab equipment to communicating my science through presentations and co-first author publications in peer-reviewed journals. Moreover, I would like to thank my former undergraduate research advisors for being encouraging, supportive, and enthusiastic throughout my graduate career.

I am grateful to my friends from college for encouraging me to push my boundaries, and inspiring me to enjoy life's beautiful moments. Jenny Bowen opened my eyes to graduate school as a next step after college, taught me the importance of building and maintaining a network, and encouraged me to think critically about my science. I share an adventurous spirit with Alana Garibaldi, who supported me while I faced new challenges and cheered me on while I reached new heights. She taught me how to persevere and reminded me to celebrate every one of my triumphs of graduate school. Lindsey Abbott has brought laughter and joy into my life. She always encouraged me to dream big, reevaluate my priorities, and reminded me to enjoy life and laugh more often.

I want to thank the people that have welcomed me into their family in such an accepting and loving way – Linda Ambrosini, Len Ambrosini, Monica Lee, Craig Lee, Paul

Ambrosini, Amy Ambrosini, Carol Ayers, and Dolly Ambrosini. I cherish your loving support and kind words of encouragement. I cannot begin to express my thanks to you for making so many ordinary moments extraordinary throughout my doctoral journey.

A warm thank you goes to my family for their continuous and unparallel love and support. I am grateful to my aunts and uncles for cherished times spent together during visits. I very much appreciated the moments I shared with Timmy, Sue, Janice, Tom, Pat, Sue, Peggy, Susie, Rick, and Anna. I would also like to say thank you to my cousins, especially Brittany Gregory. Brittany has been an inspiration for me to pursue higher education and provided support throughout my education. I admire her ambition and courage and am so thankful to have her as a role model. Special thanks to my grandma, Anna Kapala, for teaching me how to be resourceful and always encouraging me to achieve my absolute best.

Through the peaks and valleys, I am fortunate to have a support system that cheers me on and lifts me up. I would like to thank my brother, Ryan Gabriel, and sister, Brenna Gabriel, for always being there for me when I need you. Both of my siblings showed ensued that I did not forget to enjoy the beauty life has to offer. I truly appreciate the unwavering love, support, and encouragement as well as positive influence you have on my life.

The life lessons that I learned from my parents have allowed me to achieve tremendous success throughout my academic career. My parents instilled the value of education from a young age. This notion transformed into a hunger for knowledge that motivated me to become the first in my family to attend college, then pursue a terminal degree. My parents raised me to follow my dreams and prepared me to succeed in my

most difficult challenge – earning a doctoral degree. My father, Jeffery Gabriel, taught me how to be a hard worker, perfectionist, and problem solver. He also taught me the importance of persistence, patience, and passion and how to think outside of the box. My mother, Debra Gabriel, taught me how to be spontaneous, dream big, take risks, and live in the moment. She also taught me the importance of being selfless, thoughtful, generous, and kind. Thank you, Mom and Dad, for inspiring me to work hard and reach for the stars!

My five year journey in graduate school has been an absolute pleasure with my fiancé, Mitchell Ambrosini, by my side. His humor is contagious, supplying me with loads of laughter and smiles at many occasions. He has been a pillar of my strength. His work ethic has been an inspiration to me, fueling my work ethic as a graduate student. His comfort, support, and encouragement has helped me through challenging times. Whenever I faced hurdles, he helped me refocus and recharge by joining me on an adventure in the great outdoors or organizing game nights with friends. Graduate school was still sweet, despite difficult times. He has been there to celebrate my major milestones in graduate school, whether it's with an ice cream cone or a champagne toast. His deep affection, especially towards me and our puppy dog, Lyla, has been a cornerstone in our family. I'm so grateful to have found a loving and supportive partner in life.

CURRICULUM VITAE

Kristin N. Gabriel

kngabrie@uci.edu • (708)-870-4769 • www.linkedin.com/in/kristingabriel/

EDUCATION

Ph.D. in Molecular Biology and Biochemistry present
University of California, Irvine (Irvine, CA)

B.S. in Biochemistry and Molecular Biology, cum laude Dec 2015
Chapman University (Orange, CA)

HANDS-ON RESEARCH EXPERIENCE

Graduate Student Researcher Sep 2016 – present
UCI - Department of Molecular Biology and Biochemistry (Irvine, CA)
Thesis Advisor: Prof. Gregory A. Weiss

- Engineer DNA polymerase variants for attachment on electronic nanotransistors, optimize activity assays to study polymerization, and collaborate with Illumina Inc. to conduct single-molecule studies
- Planned, executed, and analyzed mass spectrometry bioconjugation reactions to assess pyrocinchonimide reaction for improved development of antibody-drug conjugates to improve cancer therapies
- Identified specific Nucleocapsid antibodies using phage display plus developed patient disease risk factor score for predicting COVID-19 severity and development of COVID-19 disease prognostic
- Communicated results and challenges weekly with colleagues and senior management and presented comprehensive technical project reports and formal presentations at yearly professional conferences
- Translated ideas and conceptualized experimental findings through authored scientific publications, patent, and a funded grant proposal from the National Science Foundation Graduate Research Fellowship Program
- Lectured undergraduate biochemistry discussion and microbiology laboratory sessions as teaching assistant

Junior Specialist Mar 2016 – May 2016
UCI - Department of Chemistry (Irvine, CA)
Advisor: Prof. Gregory A. Weiss

- Advanced single molecule bioelectronic studies through enzyme expression and purification
- Optimized enzyme purification using size-exclusion chromatography
- Training in Laboratory Safety Fundamentals, Hazardous Materials, and Cyber Security Awareness Fundamentals

Undergraduate Researcher Jan 2013 – Dec 2015
CU – Schmid College of Science and Tech. (Orange, CA)
Advisors: Prof. Marco Bisoffi and Prof. Jason Keller

- Explored molecular insights into prostate field cancerization using human cell models by measuring gene regulation with qRT-PCR while in the Bisoffi Cancer Biology Laboratory
- Studied soil microbial processes in wetland ecosystems by measuring methane and carbon dioxide production using gas chromatograph while in the Keller Wetland Biogeochemistry Laboratory

SKILLS

Laboratory: Enzyme expression (*E. coli* and yeast expression systems) and purification (ion and size-exclusion chromatography), optical spectroscopy (UV-Vis, fluorescence, fluorescence polarization, circular dichroism), mass spectrometry (MALDI, ESI-LCMS), mutagenesis (site-directed PCR, Gibson assembly, site-directed Q5), mammalian cell culture, immunostaining (immunohistochemistry, flow cytometry, western blotting, ELISA), phage display, microscopy

Interpersonal: Exceptional verbal and written communication, organizational, public speaking, teaching, and mentoring skills as well as ability to work independently in challenging and fast paced environment

Computer: Data analysis (Knime, Python, MATLAB, ImageJ, GraphPad Prism), Chimera, Protein sequence analysis (SnapGene, Benchling, ExPASy), Adobe Illustrator, Microsoft Office (Word, PowerPoint, Excel)

PUBLICATIONS AND PATENTS

Sen S.*, Sanders E.*, **Gabriel K.*,** Miller B., Isoda H., Salcedo G., Garrido J., Dyer R., Nakajimaa R., Jain A., Santos A., Bhuvan K., Trifea D., Ricks-Oddie J., Felgner P., Edwards R., Majumdar S., and Weiss G. (2021) “Antibody response to a nucleocapsid epitope as a marker for COVID-19 disease severity.” mSphere, 6(2): e00203-21.

Richardson M.*, **Gabriel K.***, Garcia J., Ashby S., Dyer R., Kim J., Lau C., Hong J., Le Tourneau R., Sen S., Narel D., Katz B., Ziller J., Majumdar S., and Weiss G. (2020) “Pyrocinchonimides Conjugate to Amine Groups on Proteins via Imide Transfer.” Bioconjugate Chemistry, 31(5):1149-1462.

Richardson M., **Gabriel K.,** Dyer R., and Weiss G. (2020) “Improved bioconjugation reaction for antibody-drug conjugates.” U.S. Provisional Application filed (UC Case 2019-077-0).

Arzmanyan G., **Gabriel K**., Ravikumar A., Javanpour A., and Liu C. (2018). “Mutually Orthogonal DNA Replication Systems In Vivo.” ACS Synthetic Biology 7(7): 1722-1729.

Gabriel K.*, Jones C.*, Nguyen J., Antillon K., Janos S., Overton H., Jenkins S., Frisch E., Trujillo A., and Bisoffi M. (2016) “Association of regulation of protein factors of field effect in prostate tissues” International Journal of Oncology. 49(4):1541-1552.

Gabriel K., Medvedeff C.A., and Keller J.K. (2016) “Microbial organic matter reduction in a peatland soil: the importance of water-table level” BIOS. 88(1):39–45.

Keller J.K., Anthony T., Clark D., **Gabriel K**., Gamalath D., Kabala R., King J., Medina L., and Nguyen M. (2015) "Soil Organic Carbon and Nitrogen Storage in Two Southern California Salt Marshes: the Role of Pre-restoration Vegetation." Bulletin of the Southern California Academy of Sciences. 114(1): 22–32.

*Authors contributed equally

ADDITIONAL LEADERSHIP AND COMMUNICATION EXPERIENCE

Founding Member and Team Lead Sep 2020 - present
STEM Alumni PhD Network (CU)

- Established a community of Chapman alumni PhD students to share career advice with undergraduates
- Contributed ideas and orchestrated opportunities to help undergraduates apply and choose a graduate school

Laboratory Management July 2017 – present
Weiss Laboratory (UCI)

- Managed relationships with vendors and laboratory personnel, including hiring, orientation, and training
- Supervised work-study students by setting goals, objectives, and tactical plan to organize freezer inventory
- Managed budget for all supplies and implemented online lab management platform to track order requests
- Negotiated service contracts and served as liaison between Weiss lab and purchasing for equipment

Graduate Student Mentor Jun 2017 – present
Weiss Laboratory (UCI)

- Mentored +13 undergraduate, masters, and graduate students in performing independent research projects and lead a research team in managing projects, defining research strategies, and developing goals

Student Program Lead and Moderator May 2019, May 2020
Adv. Genomic Tech. Development Meeting

- Spearheaded inaugural grantee session on career transitions for >50 student and post-docs
- Moderated the Sequencing Tech Awards session, Grantee Talks, and Trainee collaboration discussion

Founding Member and Career Development Chair Jan 2018 – Nov 2019
DeviceAlliance Young Professionals

- Initiated 6-month inaugural mentorship program for >30 young professionals in the medical device industry
- Consulted with President and worked with Membership Chair to define mentee needs, recruit mentors, identify venues, negotiate event expenses, create marketing materials, and conduct follow-up survey
- Served as invited panelist for Career Navigation Event in Life Science & Medical Devices (>100 attendees)

Professional Development Chair Sep 2018 – Sep 2019
Association for Women in Science (UCI)

- Partnered with other departments to recruit graduate students and post-docs in STEM and orchestrated 3-month inaugural mentoring circle that boasted a career navigation and strengthen finder events
- Worked with Vice President and Membership Chair to define student needs, create interactive learning workshops, recruit speakers, design event flyers, publicize events, and collect participant feedback

Member and Trainee-led Think Tank Sep 2017 – Oct 2020
Graduate Professional Success in STEM (UCI)

- Raised awareness of GPS-STEM and sparked collaboration with community and campus-wide groups
- Evaluated current efforts and overhauled interactive workshops to fill gaps in professional development

INDUSTRY EXPERIENCE

Lab Technician Jan 2016 – Mar 2016
DiaSorin Molecular LLC (Cypress, CA)

- Manufactured ELISA diagnostic kits through methods of formulation, adjustment, filtration, packaging, and antigen production under design control process (GDP/GMP/QSR)

Cancer Biology Intern May 2015 – Aug 2015
CytomX Therapeutics (South San Francisco, CA)

- Advanced probody drug conjugates toward clinical investigation through ELISA and Immunohistochemistry assay development, optimization, and validation for contract research organization activities
- Trainings in Employee Health and Safety, FDA regulations, Pre-IND and IND Phase, IPO Process

AWARDS AND HONORS

1st Place – Applied Innovation & GPS-STEM's Industry Insights Pitch Competition, UCI 2021
Graphical Abstract Winner – GPS-STEM's Visualizing Science Competition, UCI 2020
Anti-Cancer Challenge research grant - Chao Family Comprehensive Cancer Center, UCI 2020
1st Place Blitz Talk – Dept. Molecular Biology and Biochemistry Annual Retreat, UCI 2019
Audience's Award – GPS-STEM's Elevator Pitch Competition, UCI 2018
Robert Warner Award for Outstanding Achievement in Nucleic Acid Biochemistry, UCI 2018
1st Place TED-style Talk - Associated Graduate Students Symposium, UCI 2018
National Science Foundation Graduate Research Fellowship Program (NSF GFRP), UCI 2018
UCI Biological Sciences Graduate Fellowship, UCI 2016
Honors B.S. in Biochemistry and Molecular Biology, *cum laude*, CU 2015
Schmid College of Science and Technology Program Honors, CU 2015
2nd Place Oral Presentation - Tri Beta Biological Honor Society Conference, CU 2015
2nd Place Poster Presentation - Tri Beta Biological Honor Society Conference, CU 2014
Travel Grants – Student Government Association, CU 2014 – 2015
Undergraduate Research Grant – Office of Undergraduate Research, CU 2013, 2015
Dean's Scholarship – Office of Admissions, CU 2012 – 2016
Chapman Grant – Office of Financial Aid, CU 2012 – 2016

GRADUATE COURSEWORK

Medical Product Life-Cycle Management (BME X408)
Fundamentals of Clinical Trials (BIOSCI X450)
Regulatory Requirements for Medical Devices (EECS X445.2)
Regulation of Gene Expression (M&MG 206)
Intro to Proteomics (PHYSIO 252)

Protein Structure and Function (MB 204)
Adv. Molecular Genomics (M&MG 250)
High-Res Structures: NMR & X-ray (MB 211
Effective Comm. for Scientists (Activate to Captivate, 12 hr)
Epigenetics in Health and Disease (BIOCHEM 225)
Scientific Writing and Publishing (Nature masterclass, 11 hr)
Basics of Medical Writing (GPS-STEM, 6 hr)
Intro to Data Science (I&C SCI X426.6)
Visualizing Science (GPS-STEM, 8 hr)
Business of Science (SciPhD, 30 hr)
Craft of Science (CHEM 251)
Micro-MBA (UCSD Rady School, 30 hr)
Oncology Drug Development (Pfizer, 6 hr)
Conduct of Research (MMS 250)
Industry Insights (GPS-STEM,12 hr)

RELEVANT PRESENTATIONS

Oral presentations

Gabriel K. (2021) "Monitoring Taq DNA Polymerase Dynamics at the Single-Molecule Level with Native Temperatures." St. Jude Children's Research Hospital National Graduate Student Symposium.

Gabriel K. (2020) "Pausing dynamics of single-molecule Φ29 DNA polymerase observed by carbon nanotube transistors." NIH National Human Genome Research Institute annual meeting.

Gabriel K. (2019) "Gimme a break or DNA polymerase pausing at Single-Molecule Level." UCI MB&B RIP talk.

Gabriel K. (2019) "Dissecting DNA polymerases at the Single-Molecule Level." ACS national meeting.

Gabriel K. (2019) "Dawn of Clinical DNA Sequencing." UCI Grad Slam semi-finals.

Gabriel K. (2018) "DNA polymerases at the Single-Molecule Level." UCI MB&B RIP talk.

Gabriel K. (2018) "Electronically Listening to Single Enzymes." UCI Associated Graduate Student Symposium.

Gabriel K. (2018) "Electronically Listening to Single Enzymes." UCI Grad Slam semi-finals.

Invited presentations

Invited Panelist (2021) CU Careers Seminar in Biotechnology (BCHM 100)

Invited Panelist (2020) CU Graduate student Panel and Networking for First-Generation undergraduates

Invited Speaker (2020) "DNA Sequencing Technologies." CU Biotech and Bioengineering (BCHM 320)

Invited Panelist (2020) CU Careers Seminar in Molecular Biology and Medicine (BCHM 100)

Invited Panelist (2019) CU Graduate student Panel and Networking for First-Generation undergraduates

Invited Panelist (2019) CU Careers Seminar in Biochemistry and Molecular Biology (BCHM 100)

Invited Panelist (2019) CU-DeviceAlliance Career Navigation Event in Life Science & Medical Devices

Poster presentations

Gabriel K., Turvey M., Lau C., Vargas R., Taulbee J., Kim J., Chen S., Wonbae L., Pugliese K., Majumdar S., Collins P., and G. Weiss. (2021) "Monitoring Taq DNA Polymerase Dynamics

at the Single-Molecule Level with Native Temperatures." St. Jude Children's Research Hospital National Graduate Student Symposium.

Gabriel K., Lewis S., Lau C., Turvey M., Kumar N., Pugliese K., Boyanov B., Majumdar S., Collins P., and G. Weiss. (2019) "DNA sequencing with single-molecule nano-electronics." Frontiers in Protein Mass Spectrometry Symposium in Irvine, CA.

Gabriel K., Lau C., Turvey M., Kumar N., Pugliese K., Boyanov B., Majumdar S., Collins P., and G. Weiss. (2019) "Single-molecule electronics reveal pausing dynamics of DNA polymerase." NIH National Human Genome Research Institute annual meeting in Boston, MA.

Gabriel K., Richardson M., Garcia J., Dyer R., Majumdar S., and G. Weiss. (2019) "Dimethylmaleimide: a new reagent for protein bioconjugation." UCI MB&B dept. meeting in Carlsbad, CA.

Gabriel K., Lewis S., Lau C., Turvey M., Kumar N., Pugliese K., Boyanov B., Majumdar S., Collins P., and G. Weiss. (2018) "DNA sequencing with single-molecule nano-electronics." NIH National Human Genome Research Institute annual meeting in Boston, MA.

Gabriel K., Lewis S., Lau C., Kumar N., Majumdar S., Collins P., and G. Weiss. (2018) "Enzyme dynamic of single-molecule DNA Polymerases in bio-nanocircuits." UCI MB&B dept. meeting in Lake Arrowhead, CA.

Dyer R., **Gabriel K.**, Iftikhar M., Majumdar S., Pugliese K., Truong J., and G. Weiss. (2016) "Single molecule protein nanocircuits with DNA Synthesis Enzymes." UCI Cancer Center Retreat in Long Beach, CA.

Gabriel K., Iftikhar M., Majumdar S., Pugliese K., Truong J., and G. Weiss. (2016) "Single Molecule Protein Nanocircuits with DNA Polymerase KF." CUSP-UCI Research Mixer in Irvine, CA.

COMMUNITY INVOLVEMENT

Volunteer May 2017 – present

UCI anti-cancer challenge fundraiser

- Support, mobilize, and encourage bike riders as a fundraiser member on Team OncoSlayers.

Volunteer Scientist Oct 2016 – present

Irvine Unified School District

- Translate the scientific method into a digestible concept for students in the Irvine Unified School District
- Captivated students on the latest DNA sequencing technologies and motivated first-generation students to pursue graduate school as guest lecturer (BCHM 100/320) at Chapman University

ABSTRACT OF THE DISSERTATION

Advancing DNA Sequencing Technologies, Antibody Therapeutics, and Viral Diagnostics

by

Kristin Nichelle Gabriel

Doctor of Philosophy in Biological Sciences

University of California, Irvine, 2021

Professor Gregory A. Weiss, Chair

DNA polymerase catalyzes the correct replication and repair of DNA, an essential step in the life cycle of all organisms. The enzyme incorporates incoming deoxynucleotide triphosphates (dNTPs) into a nascent DNA strand that is complementary to a single-stranded DNA (ssDNA) template. This capability makes DNA polymerases workhorses for molecular biology and biotechnology. Single-molecule studies can identify the dynamics of DNA polymerase structural conformations, which are otherwise lost through averaging in ensemble populations. Therefore, the Weiss and Collins labs (UCI) collaboratively developed a single-walled carbon nanotube field-effect transistor (nanocircuit) to translate enzyme motions into electronic signals. These nanocircuits may illuminate hidden conformational transitions of DNA polymerase activity and reveal new mechanisms and dynamics during dNTP incorporation. The goal of this work is to apply this approach to understand the conformational dynamics of the thermostable DNA

polymerase from *Thermus aquaticus* (Taq) and processive DNA polymerase from the *Bacillus subtilis* bacteriophage phi29 (Φ29). The work here resolves subtle dynamics and transient intermediate states of DNA polymerase that modulate catalytic speed and molecular recognition during DNA synthesis. Additional efforts are focused on evaluating the performance of nanocircuits for direct electrical monitoring of polymerase in future DNA sequencing.

Advances in bioconjugation, the ability to link biomolecules to each other, small molecules, surfaces, and more, can spur the development of advanced materials and therapeutics. In work reported here, pyrocinchonimide (Pci) undergoes a surprising transformation with biomolecules. The reaction targets amines and involves an imide transfer, which has not been previously reported for bioconjugation purposes. The Pci motif can reduce combinatorial diversity when many available reactive amines are available, such as in the formation of antibody-drug conjugates. The reaction offers a thermodynamically controlled route to single or multiple modifications of proteins for a wide range of applications.

In response to the ongoing Coronavirus disease (COVID-19) pandemic, my laboratory members and I investigated the antibody response of patients infected by SARS-CoV-2 for predicting disease trajectories. Using methods of enzyme-linked immunosorbent assay (ELISA) and coronavirus antigen microarray (COVAM) analysis, antibody epitopes were mapped in the plasma of COVID-19 patients (n = 186) experiencing a wide range of disease states. This work identified antibodies to a 21-residue epitope from nucleocapsid (termed Ep9) associated with severe disease, including admission to the intensive care unit (ICU), requirement for ventilators, or death.

By combining a disease risk factor score with a test for anti-Ep9 antibodies, severe COVID-19 outcomes could be predicted with 13.4 likelihood ratio (96.7% specificity). The results lay the groundwork for a new type of COVID-19 prognostic that could guide more effective therapeutic intervention.

CHAPTER 1: Listening to enzymes move using carbon nanocircuits

1.1 Abstract

Single-molecule studies record the dynamic activities of individual molecules that are otherwise lost through averaging in ensemble populations. Over the past several decades, a variety of experimental techniques have emerged to explore the single-molecule world and revolutionized multiple fields – enzymology, structural biology, biochemistry, biotechnology, and DNA sequencing. The three categories of single-molecule experimental techniques include force-based (magnetic and optical tweezers), optical (fluorescence resonance energy transfer), and electrical (nanopore and nano-transistors). Despite their versatility and precision, forced-based and optical techniques have been largely limited to time scales poorly matched with physiological scenarios. The latter category of electrical techniques can overcome this problem by offering an observation window of a single biomolecule from microseconds to weeks.

Our laboratory in collaboration with the Collins laboratory (UCI) has developed single-walled carbon nanotube (SWCNT) field-effect transistors, termed hereafter nanocircuits. This electrical device employs carbon nanotubes as a sensitive microphone to which a single enzyme molecule is attached through a pyrene linker π-stacked on the SWCNT. Specifically, enzyme-functionalized nanocircuits generate electrical signals by the movement of charged residues on the enzyme's surface during catalysis. With microsecond time resolution over long-durations (e.g., up to 2-weeks), this solid-state approach can uncover phenomena unobservable by other ensemble and single-molecule

enzymology techniques. To date, our laboratory has investigated over 500 nanocircuit devices with 25 different enzymes. Key enzymology insights include reaction kinetics, mechanisms, memory effects, dynamic disorder, and processive variability. This chapter will describe nanocircuit biofunctionalization and measurement principles, review the enzymology uncovered by nanocircuits, and survey nanocircuit improvements for fundamental and biotechnology applications.

1.2 Introduction

Monitoring the conformational changes of single molecules can reveal their otherwise hidden dynamic processes required for function. Ensemble populations hide such information due to averaging. Thus, single-molecule experiments have revolutionized multiple fields – enzymology,[1,2] biochemistry,[3] biotechnology,[4] and DNA sequencing.[5,6] Over the past several decades, a variety of experimental techniques have emerged to explore the single-molecule world, including force-based (magnetic and optical tweezers),[7,8] fluorescence (smFRET and FCS),[9,10] optical (ultrafast spectroscopy and whispering gallery mode),[11,12] and electrical (atomic force microscopy, nanopore, nano-transistor).[13,14] Recent reviews have surveyed the state-of-the-art for these single-molecule tools.[15–17] This chapter reviews the enzymology uncovered by single-walled carbon nanotube field-effect transistors (nanocircuits), which have been the focus of a long running collaboration between the Weiss and Collins laboratories.

Nanocircuits apply an electrical-based sensing to monitor the conformational motions of enzymes.[18] A single enzyme tethered to the nanocircuit can be examined in real-time at ultrafast timescales (microsecond resolution) over long durations (e.g., up to

2-weeks). This label-free, electronic technique can overcome some of the limitations associated with labels (e.g., low temporal resolution and inevitable photobleaching of fluorescent tags).[19] Additionally, nanocircuits are participially attractive due to their versatile architecture – carbon nanotubes,[18,20] silicon nanowires,[21] molecular junctions,[13] nanopores,[22] and nanotransistors.[23]

This joint project has applied this approach to investigate the conformational dynamics of a number of enzymes during protein unfolding, folding, binding, and catalysis. Such findings have uncovered previously "inaccessible" information about enzymes' catalytic mechanism[24] such as inefficient substrate processing.[25] Additionally, nanocircuits have measured the incorporation of deoxnucleotide triphosphates (dNTPs) by a DNA polymerase for the development of future technologies for fast and low-cost DNA sequencing.[26]

In this chapter, I will discuss efforts over the last decade that showcase the technique's versatile ability to uncover new information about enzyme conformational dynamics and activities. First, I will describe biofunctionalization and measurement principles of nanocircuits. Next, I will review the enzymology uncovered by nanocircuits over the decade. Namely, our single-molecule experiments have revealed reaction kinetics of complex biochemical events, such as, conformational transitions,[27] enzyme memory effects,[24] dynamic disorder,[25] and substrate processive variability.[26] Finally, I will survey nanocircuits for applications in biology and biotechnology.

1.3 Biofunctionalization of nanocircuits

Enzymes can be attached to the carbon nanotube surface either covalently[28] or noncovalently.[29] Since covalent attachments can disrupt the electrical-based sensing of nanocircuits, noncovalent attachment strategies are commonly employed to preserve the carbon nanotube and lifetime of nanocircuits.[30,31] For example, pyrene-based linkers bearing maleimide[32] or pyrocinchonimide[33] (Pci) functional groups can provide stable anchors to the sidewall of the carbon nanotube (**Figure 1-1**). Solvent-exposed cysteine or lysine residues on the surfaces of enzymes can react with these functional groups, providing a relatively straightforward method for carbon nanotube attachment. The aromatic pyrene of this linker can strongly adhere to the carbon nanotube sidewall via π-π interactions.[34]

The thiol-maleimide reaction is predominantly used the to attach individual enzyme molecules to nanocircuits (**Figure 1-1A**). In this approach, a single cysteine variant of the enzyme is engineered to ensure a consistent orientation of enzyme on the SWCNT after bioconjugation to pyrene maleimide.[25,35] However, in cases where mutations to native cysteines render the enzyme insoluble and inactive (e.g., Phi29 DNA polymerase), I discovered the Pci imide-transfer reaction can attach the enzyme through selective lysine functionalization (**Figure 1-1B**).[33] Chapter 2 details how the Pci reaction modality is distinct from maleimides. Following biofunctionalization using either pyrene-based linker, atomic-force microscopy (AFM) confirms successful enzyme-carbon nanotube attachment (**Figure 1-1,** white arrows).[14] These resultant hybrid complexes are referred to as an enzyme-nanocircuit (e.g., lysozyme-nanocircuit).

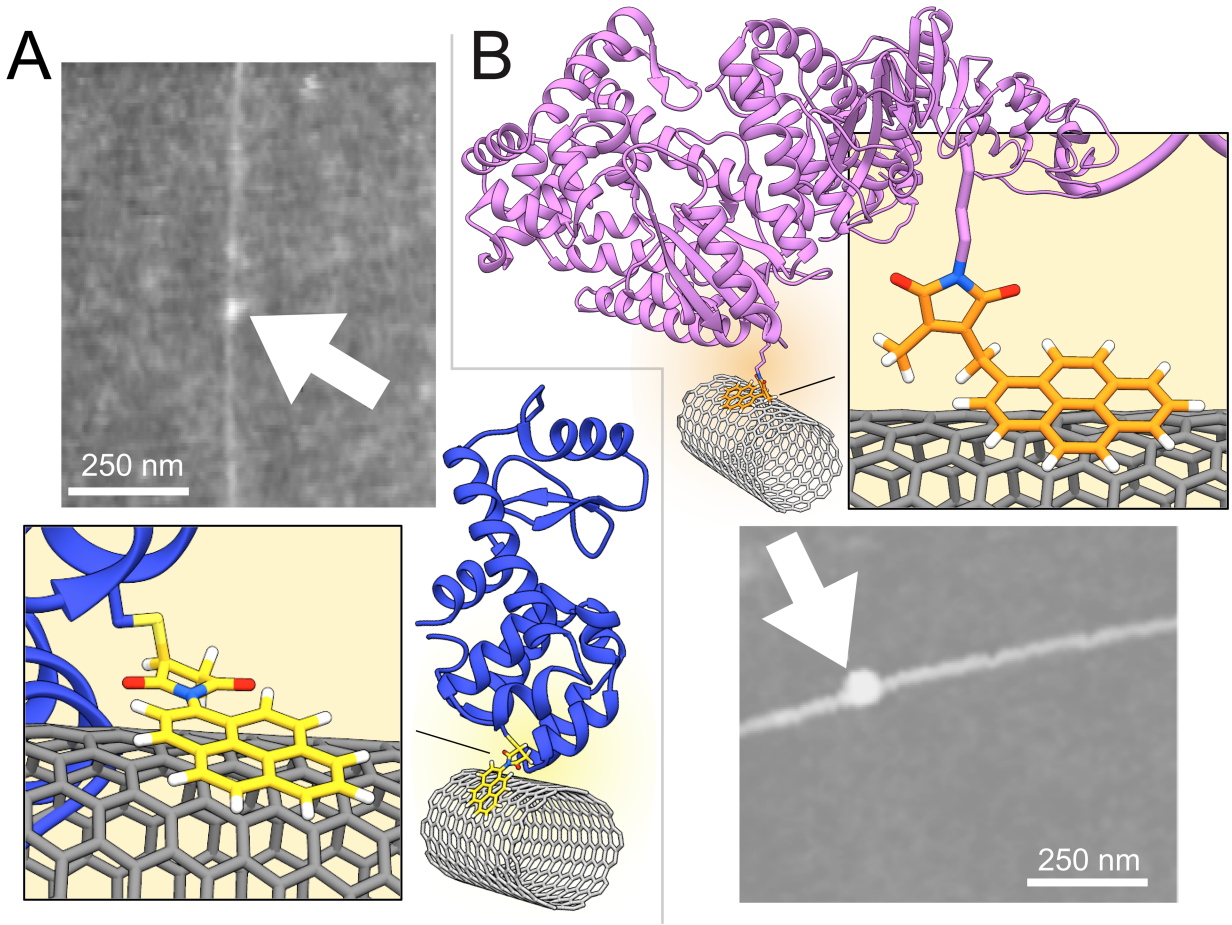


Figure 1-1. Schematic diagram of single molecules bioconjugated to nanocircuit. **A)** Pyrene maleimide (yellow) or **B)** pyrene pyrocinchonimide (orange) is adhered to the nanocircuit through π−π interactions. AFM shows the expected 1−2 nm diameter of the nanocircuit with a single enzyme (lysozyme in blue or Taq polymerase in pink) attachment obtained using the pyrene maleimide or pyrene pyrocinchonimide linkers, respectively (1 nm, white arrow). The AFM images were reproduced from ref[14] and ref[33]. Copyright 2015 Cambridge University Press and Copyright 2020 American Chemical Society. PDB: 148L and 1TAQ.[36,37]

Successful nanocircuit biofunctionalization has required several optimization procedures. For example, an effective strategy using the thiol-maleimide reaction is to eliminate tris(2-carboxyethyl)phosphine (TCEP). This reducing agent is commonly added to preserve protein activity over time, however TCEP can readily react with maleimides and reverse the covalent modification of cysteine residues in a proteins.[38] To control the final number of individual enzyme molecules on a nanocircuit, the concentration of pyrene

linkers and pyrene concentrations can be adjusted to yield only one enzyme per 1 μm of carbon nanotube length (e.g., 1:5,000 ratio of pyrene maleimide linker to pyrene). Despite improvements during the biofunctionalization process, challenges with nanocircuit biofouling slow work towards dissecting enzyme dynamics at the single-molecule level. Rigorous protein purification helps to reduce proteinaceous debris deposited on the device surface. Overcoming surface biofouling, a consequent of long measurements (e.g., up to 2-weeks), is currently under investigation.

1.4 Principles of nanocircuit measurement

Following biofunctionalization, a voltage is applied across the carbon nanotube between the source and drain electrodes to generate a current over time $I(t)$ (**Figure 1-2A**). The $\Delta I(t)$ signal is subjected to a highpass filter to account for wandering current and low frequency fluctuations from the raw $I(t)$ data. In activity buffer, the $\Delta I(t)$ signal presents as a "baseline" with featureless band of 1/f noise. In the presence of the enzyme's catalytic substrate, conformational changes during catalysis shift charged residues nearby the carbon nanotube, causing $\Delta I(t)$ excursions.[39] Negative controls in the absence of substrate or with catalytically inactive enzyme did not produce $\Delta I(t)$ excursions.

The simplest example of $\Delta I(t)$ excursions has the enzyme cycling between open and closed conformations. An illustration of a single-cysteine variant of lysozyme (C54T/C97A/S90C referred to as lysozyme) tethered to the carbon nanotube is presented (**Figure 1-2B**). As lysozyme opens and closes upon its peptidoglycan substrate during catalysis, charged residues (K83 and R119) move 1 to 2 Å away from the carbon nanotube (**Figure 1-2B**, red arrows). Such motions within the Debye screening radius of

1 nm of the carbon nanotube (**Figure 1-2B**, **dashed**) give rise to $\Delta I(t)$ excursions.[40] For quantitative statistical analysis, each $\Delta I(t)$ excursion is assigned a duration, τ_{closed}, and waiting time before the $\Delta I(t)$ excursion, τ_{open}, from the baseline (**Figure 1-2C**, **left**). The distributions between the open (τ_{open}) and closed (τ_{closed}) conformations are presented in **Figure 1-2C** (**right**).

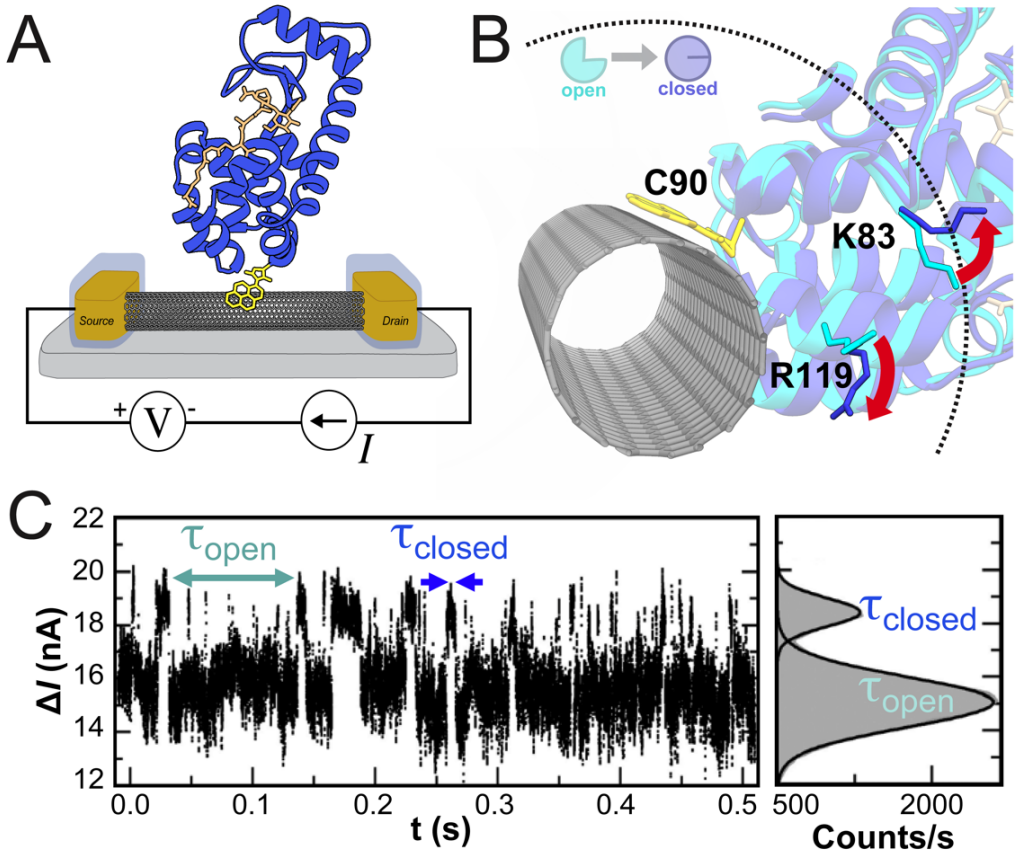


Figure 1-2. Principles of nanocircuit measurement. A) Schematic of a nanocircuit functionalized with a single enzyme (lysozyme, blue) tethered to a carbon nanotube via a pyrene maleimide linker (yellow). When a voltage is applied, the current travels from the source to the drain electrodes (gold). **B)** As lysosome transitions from open (cyan) to closed (blue) conformations, charged residues (K83 and R119) near to the attachment site (C90) and within the 1.2 nm Debye screening radius (dashed) of the carbon nanotube can electrostatically modulate the current over time, $\Delta I(t)$. **C)** Representative $\Delta I(t)$ signal during lysozyme catalysis of peptidoglycan (left) with $\Delta I(t)$ excursions (blue arrows) and waiting time before each $\Delta I(t)$ excursion (teal arrow) from the baseline. Histogram showing distributions of the open (τ_{open}) and closed (τ_{closed}) conformations. The image in (C) was adapted from ref[41]. Copyright 2013 American Chemical Society. PDB: 1QTV and 148L.[36,42]

Excursions in $\Delta I(t)$ are attributed to moving charged residues close to the SWCNT during catalysis.[25] Our laboratory provided supporting evidence for this mechanism using two strategies. First, seven lysozyme variants were engineered to have positive (blue), neutral (yellow), or negative (red) charged residues at positions 83 and 119 (**Figure 1-3A**, **top**). Comparisons of each lysozyme variant's ability to transduce $\Delta I(t)$ excursions demonstrated how the charge and positioning of nearby sidechain residues substantially altered the effective voltage gating (ΔV_G) on the nanocircuit (**Figure 1-3A**, **bottom**). Second, the enzyme's environment (e.g., salt concentration) was altered. Decreasing the salt concentration improved the signal-to-noise ratio of the ΔI signal (**Figure 1-3B**, **top**); this finding built upon previous work that showed ionic screening affected nanocircuits.[43] However, the ΔI signal dropped unexpectedly at low salt concentrations (10 mM NaCl); this decreased activity was attributed to enzyme activity (**Figure 1-3B**, **bottom**. The reduced activity likely resulted from partial protein denaturation, a known lysozyme sensitivity to low salt.[44] Both strategies characterized key parameters of the nanocircuit's electrostatic transduction and suggested ways to optimize $\Delta I(t)$ signal with any molecule of interest.

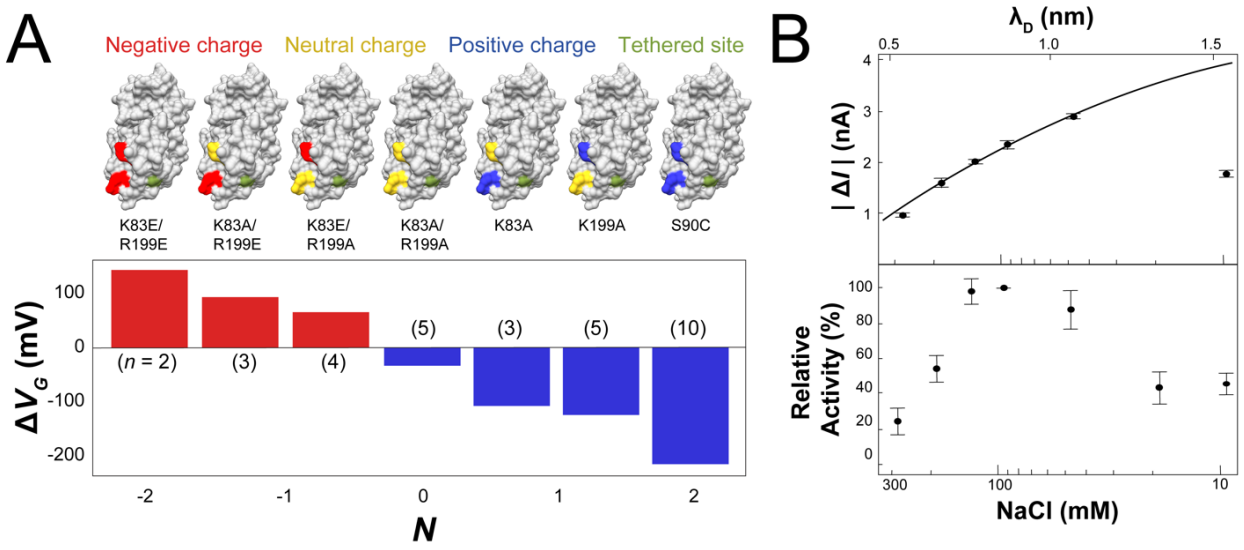


Figure 1-3. Signal transduction mechanism of nanocircuits. A) Lysozyme variants engineered to have positive (blue), neutral (yellow), or negative (red) sidechain residues near the attachment site (green). Among the *n* different lysozyme-nanocircuits, the average effective voltage gating of ΔV_G was proportional to the enzyme's net charge (*N,* x-axis). **B)** The signal amplitude ΔI of a representative lysozyme-nanocircuit was screened by buffer salt concentration (bottom axis), and had a corresponding Debye screening radius (top axis) that followed the Debye-Hückel model (solid line). The decrease in ΔI signal at a low salt concentration correlated to a decrease in lysozyme's catalytic activity. Error bars indicate ± standard deviation. The images in (A) and (B) were adapted from ref[41] and ref[45], respectively. Copyright 2013 Physical Chemistry Chemical Physics and 2013 American Chemical Society.

1.5 Overview of enzyme dynamics observed by nanocircuits

To date, over 500 single-molecule electronic devices with 25 different enzymes have been investigated using nanocircuits.[14] As examples, single-molecule measurements are described for the following three enzymes: lysozyme, cAMP-dependent protein kinase A (PKA), and the Klenow Fragment (KF) of DNA Polymerase I from *E. coli*. Electronic recordings revealed new information of single molecule's activity and motions for each of three enzymes, two of which had been previously characterized using smFRET. In situations where it is not possible to easily achieve 10-50 μs of

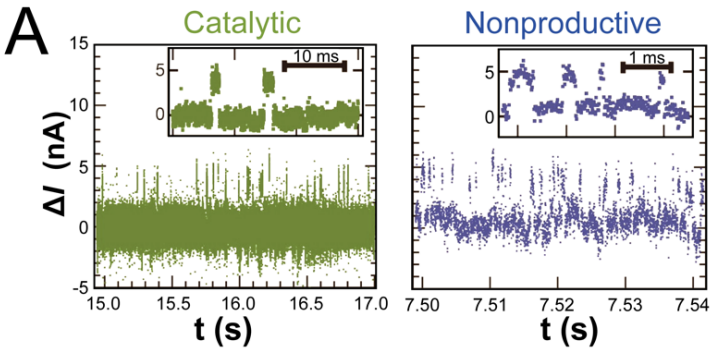
resolution during measurements,[46,47] nanocircuits complimented smFRET, providing a more complete single-molecule technique.

pH dependence of lysozyme's catalytic activity

Nanocircuits permit long-term monitoring of the same molecule. Such long time scales (e.g., up to 2-weeks) are beyond the capacity of smFRET due to limitations of fluorophore quenching and bleaching.[19] Initial efforts using nanocircuits focused on lysozyme, an enzyme that catalyzes the destruction of the peptidoglycan layer of bacterial cells walls, occurring especially in tears and egg white.[48] Our laboratory was interested in characterizing unexplained lysozyme behaviors, such as the enzyme's closing "hinge" motion during hydrolysis of its peptidoglycan substrate.[49] smFRET work revealed this closure occurs at two different rates – a slow, "catalytic" closure (20-50 s^{-1}) associated with processing its peptidoglycan substrate or a fast, "nonproductive" closure (200-400 s^{-1}).[50,51] However, the pH dependence of the enzyme's hinge closures remained ill-defined.

The long-duration ΔI signal from lysozyme-nanocircuits captured the enzyme undergoing slow, catalytic closures during hydrolysis of its peptidoglycan substrate or fast, nonproductive closures (**Figure 1-4A**). This single-molecule recording of catalytic activity at pH 7 agreed with previous ensemble measurements over the enzyme's active pH range over 6-9.[52] Likewise, the average catalytic rate of a single-molecule over a 10-minute period at three different pH values agreed with smFRET measurements (**Figure 1-4B**).[50,51] Notably, the lysozyme-nanocircuits gave insight into the proportion of time the enzyme spends hydrolyzing its peptidoglycan substrate as a function of pH (**Figure 1-4C**). However, lysozyme-nanocircuits suggest the increase in nonproductive closures and

time the enzyme spends in an inactive, closed conformation contributes to the enzyme's pH dependence. This insight illustrates how nanocircuits can provide a better understanding of catalytic activity with long-duration measurements, which otherwise remain a challenge to other state-of-the-art single-molecule tools.



B

	Catalytic rate, *k* (Hz)			
	Nanocircuits			smFRET
	pH 5	pH 7	pH 11	pH 7
Processing	11.4	15.4	11.5	20 to 90
Nonproductive	198	316	235	200 to 400

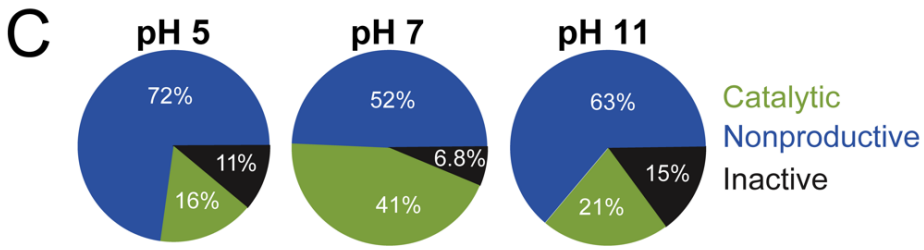


Figure 1-4. Long-duration recording from lysozyme-nanocircuit uncovers relationship between pH and the enzyme's hinge closure motions. A) Representative Δ*I* signals of lysozyme-nanocircuit showing slow, catalytic (green) or fast, nonproductive (blue) closures. The inset illustrates individual Δ*I* excursions for the two types of hinge closures. **B)** Catalytic rate (Hz) of lysozyme for catalytic or nonproductive closures at three different pH values measured by nanocircuits is comparable to range measured by smFRET observations.[50,51] **C)** pH dependence of lysozyme's catalytic activity. Percentage of time the enzyme underwent catalytic (green) or nonproductive (blue) closures or remained in an inactive (black) state. The data in (A-C) were adapted from ref[18]. Copyright 2012 American Association for the Advancement of Science.

Enzymatic turnover of lysozyme with linear and crosslinked peptidoglycan substrates

Another important and challenging problem is correlating lysozyme's processivity with the hinge closure dynamics during hydrolysis of glycosidic bonds in its peptidoglycan substrate. As a processive enzyme, lysozyme catalyzes an average of 100 glycosidic bonds at a rate of 15 per second before dissociating from its substrate.[24] The crosslinks within a substrate can affect the enzymatic degradation rates,[53] and may compromise lysozyme's processivity.[54] To test this hypothesis, lysozyme-nanocircuits were used to monitor the enzymatic hydrolysis of a peptidoglycan substrate without (linear in **Figure 1-5A**) or with (cross-linked in **Figure 1-5B**) crosslinks. The proportion of time the enzyme spent processing the two substrates varied (**Figure 1-5C**). Specifically, lysozyme processed the linear substrate 2-fold faster than crosslinked substrate due to fewer nonproductive motions. For both substrates, the catalytic rate depended largely on the type of hinge closure (**Figure 1-5D**). For example, the average catalytic rate was 9-fold higher for nonproductive (329 s^{-1}) than catalytic (36 s^{-1}) closures with the linear substrate. Compared to the cross-linked substrate, the linear substrate had a slightly higher rate for catalytic and nonproductive closures – an increase of 16% and 13% was observed, respectively. Taken together, the long-duration observations of individual lysozyme molecules provide insights into the enzyme's strategy for enzymatic hydrolysis and processivity. One strategy used by lysozyme during hydrolysis of linear substrates is to reduce nonproductive closures. For cross-linked substrates, a different strategy could involve the enzyme sidestepping around the crosslinks to reach a neighboring peptidoglycan.

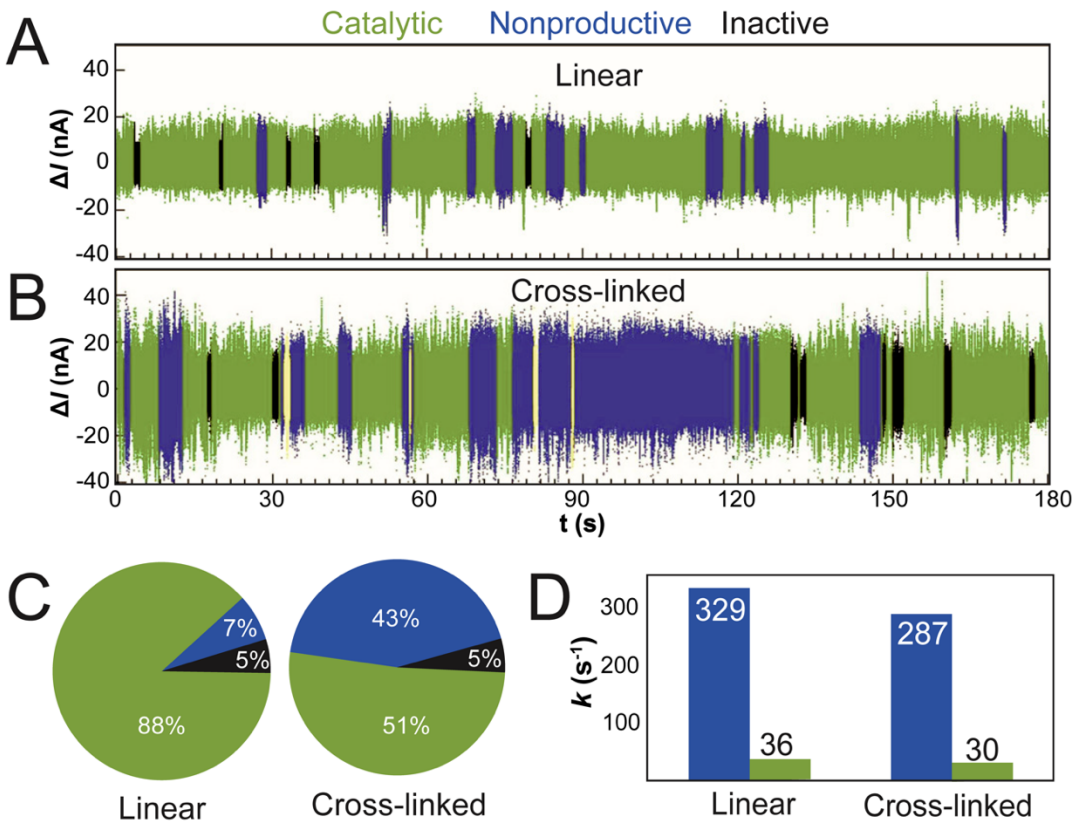


Figure 1-5. Fluctuations in enzymatic turnover of lysozyme processing linear and crosslinked peptidoglycan substrates. Lysozyme processing **A**) cross-linked or **B)** synthetic, linear peptidoglycan reveals slow catalytic motions (green) and fast nonproductive motions (blue). **C)** Compared to processing of cross-linked peptidoglycan (green), processing of synthetic, linear substrate (blue) has a substantially higher percentage of catalytic motions. The percentage of time spent inactive (black) was independent of the substrate. **D)** The average rates of k_{fast} and k_{slow} shows cross-linked decreases the rate for both catalytic (green) and nonproductive (blue) motions compared to the presence of synthetic, linear substrates. The images in (A-D) were adapted from ref[41]. Copyright 2013 Owner Societies.

Lysozyme's opening and closing has a finite transition and can pause in an intermediate conformation

Nanocircuits can track single-molecule trajectories indefinitely at microsecond resolution. Observations with such timescales have gone largely unreported due to the time resolution of smFRET measurements (hundreds of microseconds to 1 ms).[46,47] With an experimental resolution of 2 μs, lysozyme-nanocircuits have directly observed the enzyme's underlying transition pathway and intermediate conformational dynamics

(**Figure 1-6A**).[27] Most of the time enzymatic closures followed a concerted mechanism, with a continuous and smooth transition occuring between the open to closed conformations (**Figure 1-6A**, **left).** Lysosome's closing had a finite duration (average of 37 μs) that was independent of the type of closure motion (e.g., catalytic or nonproductive closures) (**Figure 1-6B**). Furthermore, lysosome's closing has symmetric properties to opening with similar transition timing and features. The results demonstrate that lysozyme operates as a Brownian motor during hydrolysis of its peptidoglycan substrate.[55,56] The remaining 10% of enzymatic closures were interrupted and require the enzyme to transition to a non-concerted mechanism, where the enzyme paused in an intermediate, partly closed conformation during closing or opening (**Figure 1-6A**, **right**). A possible explanation for the intermediate conformation may be the enzyme checking alignment of catalytic residues and substrate functionalities during the peptidoglycan hydrolysis reaction (**Figure 1-6C**). Once the necessary molecular orbitals align, the enzyme can proceed to catalyze the bond breaking and forming transformation. Taken together, these single-molecule recordings provide a more complete picture of the intrinsic, dynamic motions of enzymes.

Recent breakthroughs in single-molecule techniques have further characterized the intermediate conformational dynamics of lysosome. Advances in smFRET with molecular dynamics simulations can now resolve conformations over nano- to millisecond timescales.[57] Consequently, lysosome has been shown to remain in dynamic equilibrium between the open, intermediate, and closed conformations.[58] Through mutagenesis of nearby active site residues, Sananbria *et al.* 2020 found this intermediate may be actively involved in the release of the hydrolyzed peptidoglycan substrate in catalysis.[58] This

intermediate, however, appears to be distinct from the previously reported crystallized "ajar" structure captured using mutagenic modifications that stabilized the conformation.[59] The characterization of conformational dynamics and intermediates at the single-molecule level are often necessary for an in-depth understanding of complex and dynamic enzymatic systems.

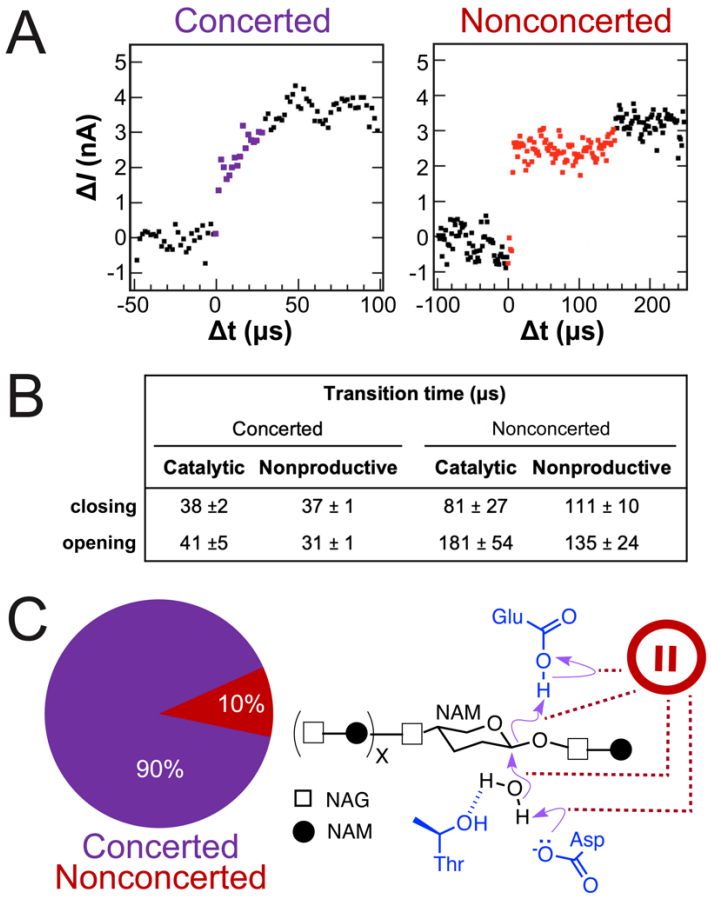


Figure 1-6. Lysozyme's open-to-closed transition. A) Representative Δ*I* signals of lysozyme-nanocircuit showing the enzyme closing via the concerted (left) or nonconcerted (right) transition pathways. The colored data points show a continuous transition (purple) or pause in the middle of transition (red). **B**) Lysozyme's closing or opening transition times ± standard deviation. **C**) For 90% of enzyme closures, the concerted transition pathway plays a role in lysozyme's hydrolysis of glycosidic bonds (purple arrows). The remaining 10% of enzyme closures "pause" (red dashed line) from an interruption during hydrolysis, leading to an intermediate within the nonconcerted transition pathway. The glycosidic bond between N-acetylmuramic acid (NAM) and N-acetylglucosamine (NAG) subunits within the peptidoglycan substrate are depicted with one water molecule (black) surrounded by lysozyme's key active-site residues (blue) of

Glu11, Asp20, and Thr26. The images in (A) and data in (B) was adapted from ref[27]. Copyright 2015 American Chemical Society.

Enzymatic activity of PKA in the presence of its substrate and cofactors

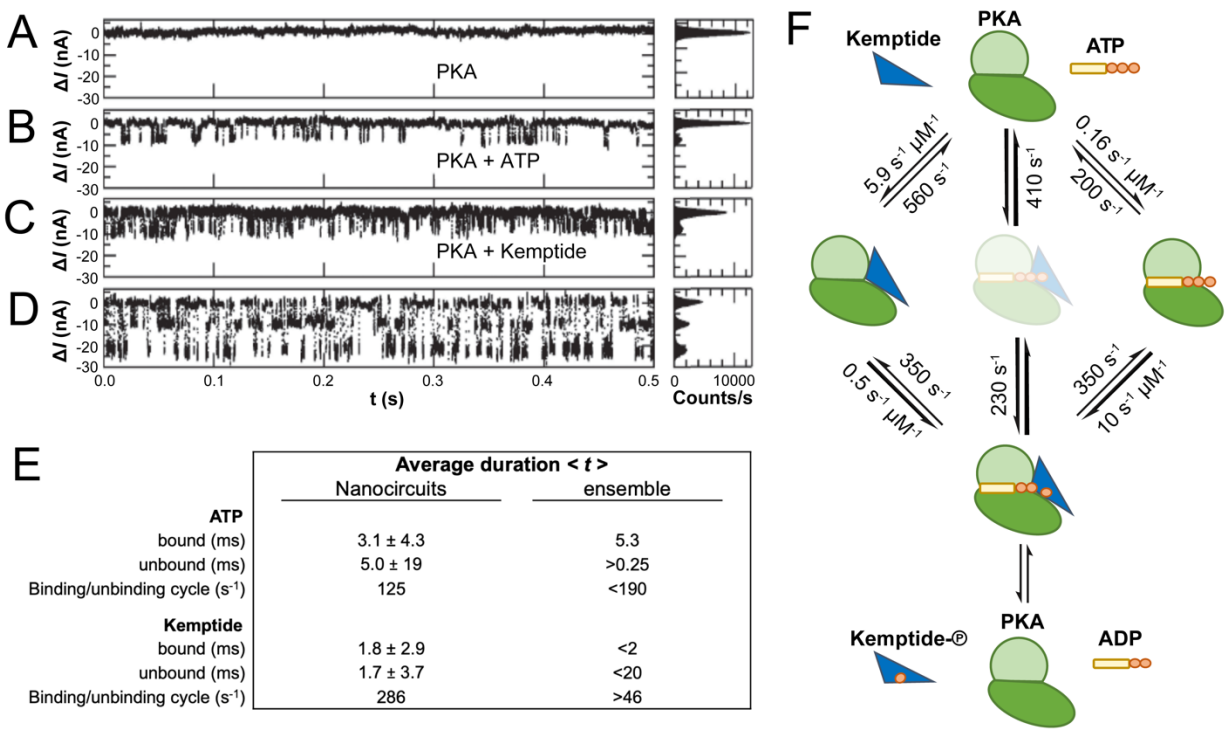
As a second enzyme example with PKA, a dynamically controlled signaling enzyme, offers insight into kinase regulation. PKA function regulates many cellular process including energy metabolism, membrane transport, muscle contraction, and signaling pathways.[60–64] A wide range of substrates exist for PKA, including enzymes, ion channels, chromosomal proteins, and transcription factors.[65] As a molecular switch, PKA can turn the activity of its substrate "on" and "off" by phosphorylation.[66–68] Binding two cofactors, adenosine-5'-triphosphate (ATP) and Mg^{2+}, induces conformational changes in the regulatory subunits required for PKA's catalytic subunits to transfer the gamma phosphate of ATP to its substrate.[66,69,70] Characterizing individual molecules of PKA can elucidate the basis of its regulated, variable behavior.

Nanocircuits were used to examine the catalytic subunit of PKA. A single cysteine (T32C) variant, called PKA hereafter, was engineered for site specific attachment on the nanocircuits. As described previously, charged residues (K28 and E31) near the attachment site (T32C) within the Debye screening radius of 1 nm perturb the $\Delta I(t)$ signal. In contrast to the previously described single molecule recordings with lysozyme-nanocircuits, the PKA-nanocircuits transduced $\Delta I(t)$ excursions in the negative $\Delta I(t)$; this reversal reflects the charge and movement diration of the protein residues close to the SWCNT. In the absence of binding partners, $\Delta I(t)$ excursions were nonexistent in activity buffer alone (100 mM MOPS, 9 mM $MgCl_2$, 100 μM TCEP, pH 7.2) (**Figure 1-7A**). Two-level $\Delta I(t)$ excursions were transduced in the presence of either the ATP cofactor (**Figure**

1-7B) or Kemptide substrate (**Figure 1-7C**). Three level Δ*I*(*t*) excursions were transduced in the presence of both binding partners (ATP and Kemptide) by the PKA-nanocircuits transduced (**Figure 1-7D**). Such Δ*I*(*t*) signals clearly resolve the enzyme in open, intermediate, and closed conformations as the enzyme forms its ternary complex. Histograms showed the distributions of the open, intermediate, or closed conformations relative to the baseline Δ*I*(*t*) signal (**Figure 1-7A to Figure 1-7D**, **right panels**).

PKA-nanocircuits determined the binding rates for the ATP cofactor or Kemptide substrate by measuring the closed (bound) or open (unbound) conformations (**Figure 1-7E**). Consistent with molecular dynamic simulations and rate constant analyses,[71] the enzyme's binding and unbinding to its peptide substate is faster than to ATP. Furthermore, results reported by PKA-nanocircuits support conventional ensemble measurements showing PKA has a shared active site that is capable of binding independently to both Kemptide and ATP.[72,73]

Over the past few decades, it has become clear that enzymes, including kinases, have diverse motion over a vast range of timescales. With an experimental resolution in the microsecond timescale, nanocircuits can uniquely capture the highly variable on- and off- rates of PKA-catalyzed phosphorylation (**Figure 1-7F**). Given the complexity and dynamic nature of this enzymatic system, nanocircuits provides insights into the titanic struggle between on- and off- rates of PKA's complex transitions.



	Average duration < *t* >	
	Nanocircuits	ensemble
ATP		
bound (ms)	3.1 ± 4.3	5.3
unbound (ms)	5.0 ± 19	>0.25
Binding/unbinding cycle (s^{-1})	125	<190
Kemptide		
bound (ms)	1.8 ± 2.9	<2
unbound (ms)	1.7 ± 3.7	<20
Binding/unbinding cycle (s^{-1})	286	>46

Figure 1-7. Nanocircuit measurements of single molecule PKA. Representative Δ*I*(*t*) signals (left) and histograms (right) from PKA-nanocircuits in the presence of **A**) activity buffer alone, **B**) ATP (2mM), **C**) Kemptide (100 μM), or **D**) both ATP and Kemptide. The baseline defines the enzyme's open conformation (0 nA), and Δ*I*(*t*) excursions correspond to the intermediate (-10 nA) and closed conformations (-20 nA). **E**) Average duration < *t* > of substrate binding ± standard deviation measured by nanocircuits is comparable to ensemble kinetic parameters.[74–76] **F**) Schematic of PKA-catalyzed phosphorylation of Kemptide, with rates measured by PKA-nanocircuits. Kinetic second order rates are represented for the indicated substrate concentration. Higher transition probabilities are reflected by increased line thickness. The intermediate conformation (transparent) is unobserved by nanocircuits. The images in (A-D) and data in (E-F) were adapted from ref[32]. Copyright 2013 American Chemical Society.

Understanding the dynamic disorder of PKA activity

To build upon the mechanistic information of PKA catalysis,[66] additional single-molecule studies focused on understanding the dynamic disorder of PKA. In contrast to static disorder, which results from heterogeneity of turnover rates among individual molecules, dynamic disorder results from slow transitions between open and closed conformations.[77] Such slow fluctuations in reaction rates are beyond the scope of

conventional timescales (e.g., nanoseconds to milliseconds). Dynamic disorder is an intrinsic property of enzymes. Memory or the ability of the enzymatic cycle to be affected by the previous cycle of that enzyme, leads to a non-Markovian system of dynamic disorder.[78,79] PKA-nanocircuits indirectly observed the enzymatic cycling and revealed a multi-second memory effort that had been previously unobserved.[32]

Specifically, PKA-nanocircuits uncovered divergent responses to ATP and peptide binding. Over the span of many thousands of cofactor ATP binding and unbinding cycles, the on- and off- rates are synced, experiencing increases and decreases together (**Figure 1-8A**). Since ATP is present in excess concentrations in the cell, PKA has on- and off- rates resulting from Brownian randomness; in other words, the enzyme cannot respond to small fluctuations in ATP concentrations and therefore its ATP binding dynamics reflect thermal sensitivity. In contrast, the second-by-second variations for the on- and off- rates diverged for the substrate Kemptide (**Figure 1-8B**). The divergence in direction of these rates suggests the evolution of an anti-correlated relationship for the peptide. The highly regulated activity of PKA reflected in the enzyme's conformational dynamics to its substrates. The anti-corelated on- and off- rates for substrate binding reflect a sensitive response to this highly regulated event in the cell. In other words, the enzyme has evolved to increase on- rates together with simultaneous decreased off- rates. Thus, the enzyme has evolved a simple mechanism to enhance its responsiveness. The correlated conformational changes observed for ATP on- and off- rates is likely due to random thermal fluctuations in the local environment.[80]

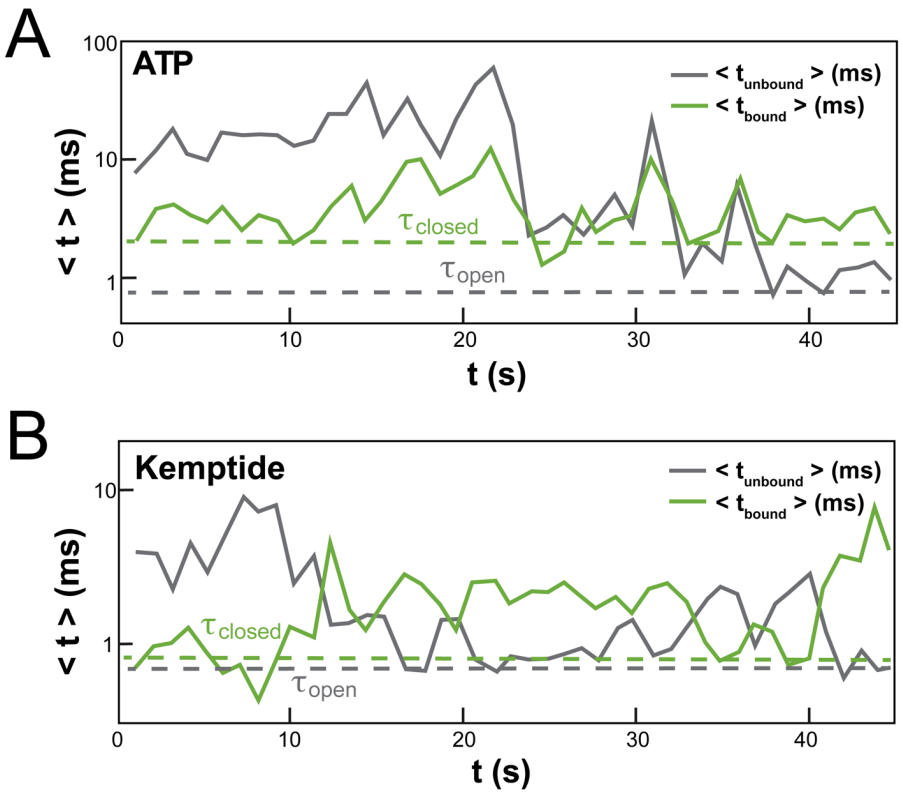


Figure 1-8. PKA's dynamic disorder with ATP or Kemptide observed by nanocircuits. Variation in waiting times for PKA binding to **A)** ATP or **B)** Kemptide. Average durations (dashed lines) were calculated for $<t_{bound}>$ and $<t_{unbound}>$ across one second of data. The images in (A-B) were adapted from ref[32]. Copyright 2013 American Chemical Society.

PKA-nanocircuits revealed that the enzyme may cycle between the intermediate and closed conformations multiple times before phosphorylation occurs (**Figure 1-9**).[32] The majority of enzyme closures (77%) involved a simple catalytic event, where the enzyme closed once in the intermediate conformation between the open and closed conformations (**Figure 1-9A**). The remaining closures (23% of the time) represented inefficient catalysis; the enzyme required multiple closures between the open and closed conformations (**Figure 1-9B**). Consequently, the catalytic rate was higher for simple (155 s^{-1}) compared to inefficient (91 s^{-1}) catalysis. This single-molecule recording of PKA's variable catalytic activity provides rich insight into the enzyme's catalytic mechanism.

Other studies have described PKA's conformational dynamics and provided insights into its catalytic mechanism.[60,67,70] Nuclear magnetic resonance spectroscopy (NMR) of the enzyme showed transition through three conformations, including the open, intermediate, and closed conformations.[75,80] Our laboratory showed using NMR and biochemical studies on another kinase family member show a similar catalytic inefficiency; such cycling behavior may provide a molecular basis for dynamic phosphorylation of highly regulated enzymes.[81] Recent X-ray crystallography and single turnover enzyme activity measurements reinforce findings that describe the role of PKA's substrate and metal cofactors modulating the catalytic cycle.[82] Molecular dynamics simulations on PKA in the presence of its preferred cofactor (Mg^{2+}) or another metal ion (Ca^{2+}) suggested a possible regulatory mechanism through structural stability.[83] Elucidating the conformational dynamics can provide and explain the complexity of the enzyme's interaction with its cofactors and substrate.

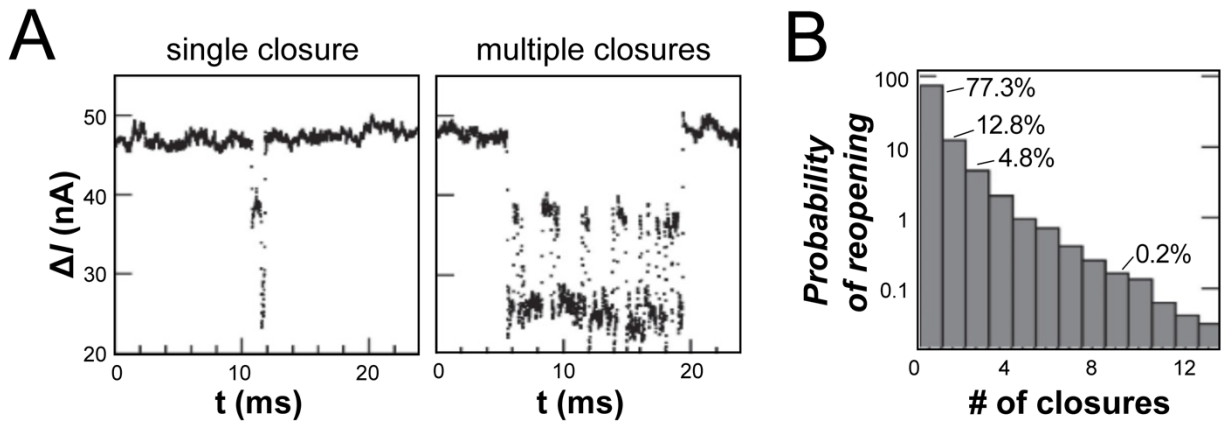


Figure 1-9. PKA-nanocircuits measure catalysis efficiently. A) Representative electronic signals of single (left) or multiple (right) closures during PKA catalysis. The recording of the single closures portrays the enzyme transition from open to closed conformations, pausing briefly at an intermediate conformation. The recording of multiple closures captures the enzyme undergoing nine repeated closures from the intermediate conformation before returning to the open conformation. **B**) Probability distribution of over 10,000 catalytic cycles. Simple catalysis, or events that occurred after a single closure, occurred 77% of the time. Approximately 5% of catalytic activity was inefficient activity

involving four or more repeated closures. The images in (A) and data in (B) were adapted from ref[14]. Copyright 2015 Cambridge University Press.

DNA polymerase substrate recognition

DNA polymerase catalyzes the replication and repair of DNA, two essential functions in all living organisms. The high fidelity of this process ensures incorporation of the correct deoxynucleotide triphosphates (dNTPs) into the nascent single-stranded DNA (ssDNA) template through extension of a DNA primer to form a complementary double-stranded DNA (dsDNA) strand. In *E. coli* and other prokaryotic replication systems, the enzyme relies on three fidelity mechanisms: dNTP selection (5'-3' polymerase), editing (3'-5' exonuclease), and mismatch repair (5'-3' endonuclease).[81] The Klenow Fragment of *E. coli* DNA polymerase I (KF) retains both the dNTP selection and the editing activities, which act in concert to synthesize Okazaki fragments during lagging-strand DNA replication with high fidelity.[82,83]

Structural characterization of KF defines a core polymerase architecture as a right-handed structure, having Palm, Fingers, and Thumb domains.[84,85] In addition to the polymerase core, the enzyme also contains an Intervening non-functional 3'-5' exonuclease domain.[86] Analysis of KF's open and closed conformations provides key insights into how the enzyme functions.[87] The Palm-like active site contains the catalytic residues. As the nascent ssDNA template is held in place by the Thumb, the Fingers region closes the active site around the appropriate dNTP to facilitate phosphoryl transfer, termed the chemistry step.[88] During the Fingers-closing transition, a part of the Fingers domain, the O-helix or dNTP-binding site, rotates ~40° upon dNTP recognition and binding.[89]

Kinetic studies provide the basic mechanism by which KF performs its dNTP selection and incorporation.[90,91] The enzyme's catalytic speeds range from 13-15 dNTPs/second[92,93] with an average error rate of 1.8×10^{-8} errors/dNTP[94] and processivity of 45 dNTPs/binding event.[91] Ensemble studies have outlined a minimal reaction pathway of KF's catalytic polymerization (**Figure 1-10A**).[90,91,95–97] Though steps of a minimal reaction pathway for the Klenow Fragment of DNA polymerase I from *E. coli* (KF) have been characterized, the subtle conformational changes and dynamic activity (e.g., processivity and kinetics) remains unclear.

Single molecule techniques can elucidate enzyme conformational dynamics required for dNTP discrimination. For example, smFRET has been widely used to study KF. Early smFRET studies on KF have monitored the conformational transitions that precede the chemistry step or bond breaking and forming event and reported that the unliganded polymerase fluctuates between the open and closed conformations.[78] Additional work examining KF's Fingers-closing motion revealed new insights into fidelity checkpoints and dNTP discrimination.[98–100] However, electrically-based techniques are required to study KF's opening and closing of the Fingers domain with sub-millisecond resolution to elucidate how conformational dynamics facilitate structure-specific dNTP discrimination by DNA polymerase. Such electronic techniques have characterized individual molecules of KF,[26,101,102] and illuminated the dynamics of the open and closed conformations in KF's catalytic cycle and correlated conformational change dynamics with dNTP discrimination.

To conduct single-molecule studies using nanocircuits, a single cysteine variant of KF was created with mutations remove the native cysteine (specifically the mutation

C907S), abrogate the 3-5' exonuclease activity (D335A/E357D), and allow for single point attachment through a cysteine residue (L790C) to the nanocircuit. This variant, hereafter termed KF, features a L790C anchor in the middle of an α-helix nearby to the Palm-like active site. In activity buffer (10 mM Tris, 50 mM NaCl, 10 mM $MgCl_2$, 10 mM DTT, pH 7.8) or with mismatched dNTPs, a baseline with no $\Delta I(t)$ excursions was observed for the rate-limiting open conformation (**Figure 1-10B, top** and **middle**). In the presence of activity buffer containing a template and complementary dNTPs, two-level $\Delta I(t)$ excursions appeared. This signal revealed the enzyme in the closed (τ_{closed}) and open (τ_{open}) conformations (**Figure 1-10B, bottom**). The $\Delta I(t)$ signal was characterized by the duration of $\Delta I(t)$ excursions, τ_{closed}, waiting times between $\Delta I(t)$ excursions, τ_{open}, and the average height H of the $\Delta I(t)$ excursions (**Figure 1-10C**). The catalytic rate for the homopolymeric DNA templates represents a complete catalytic cycle in **Figure 1-10A**, and agrees with previous ensemble-measured catalytic rates of KF (**Figure 1-10D**).[91,93] Thus, each $\Delta I(t)$ excursion captured by KF-nanocircuits correlates with the enzyme closing during a single dNTP incorporation event.

This electronic system proved quite capable at enumerating the lengths of DNA templates. Low concentrations of DNA templates (1 nM) were required to impose inactive periods (average of 3 s) between $\Delta I(t)$ excursions (**Figure 1-10E**). Such bursts of $\Delta I(t)$ excursions correspond to template arrival and its processing by KF. The experiments were reminders the poor processivity of KF, likely reflecting its evolution as a DNA polymerase specializing in the synthesis of the lagging strand by Okazaki fragments.[82,83] However, a histogram of number of $\Delta I(t)$ excursions observed for PolyC shows a strong peak at 42, matching the 42 basepair length of the DNA template and suggesting that KF

efficiently processes to the end of the template molecule with one Δ*I*(*t*) excursion per base (**Figure 1-10F**). The burst sequences rarely extended beyond 42; <4% of Δ*I*(*t*) excursions were beyond 42, implying the enzyme closing-motion has a high success rate (99.8%) for dNTP incorporation. Thus, the nanocircuit technique provides a new, direct method for measuring processivity to resolve the precise value from the wide range of previously reported ensemble-measurements (1 to 50 dNTP/binding event).[91,103,104]

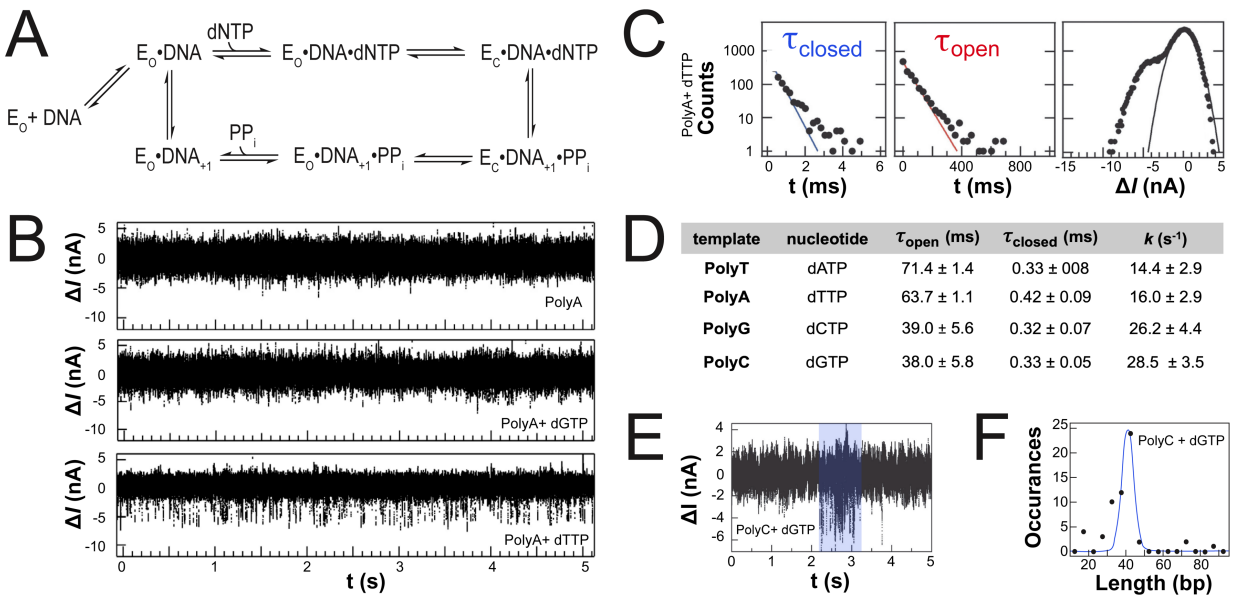
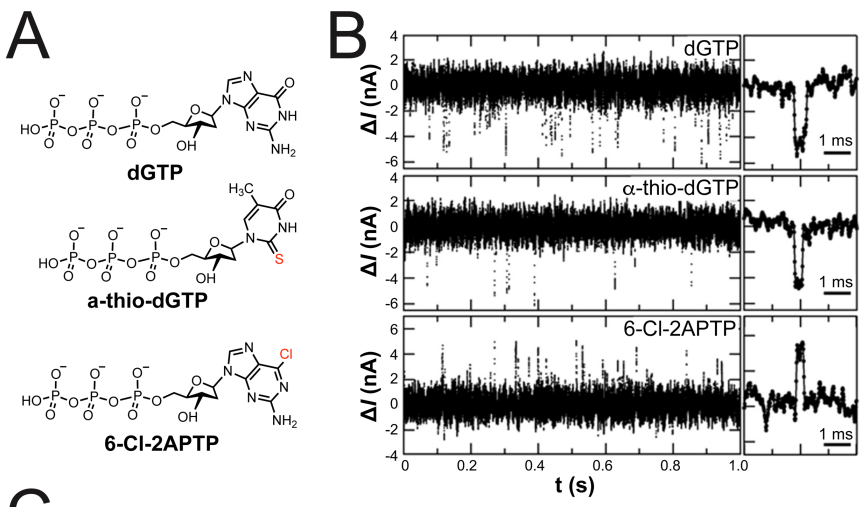


Figure 1-10. Enzymatic turnover by KF. A) Minimal reaction pathway of KF's catalytic polymerization of template-complementary DNA with mechanical motions. **B)** Representative ΔI(t) signals generated by KF in the presence of containing PolyT only, mismatched (PolyT + dGTP), or matched (PolyT + dATP) solutions. **C)** Histograms of closed (left) or open (middle) conformation durations and Δ*I* recordings (right) as measured by KF with matched substrates (PolyT + dATP). Solid lines indicated exponential fits or Gaussian fits for durations or Δ*I* recordings, respectively. Enzyme closure caused a shoulder to the left of the main peak (right). **D)** Representative signal by KF-nanocircuits in the presence of matched substrates (PolyC + dGTP) shows a burst of Δ*I* excursions. **E)** Histogram of the number of excursions (i.e., dNTP incorporations) in a burst could be counted and unambiguously associated with KF processing a single template (PolyC). The peak at 42 bases is in excellent agreement with the template length, and occurrences of less than 42 bases indicate polymerase dissociation from the template. **F**) Single-molecule kinetic parameters for KF processing homopolymeric templates data represents average values ± standard deviation. The images in (A and D-E) and (B and F) were adapted from ref[14] and ref[101], respectively. Copyright the 2015 Cambridge University Press and 2016 MDPI.

This system could provide a future generation DNA sequencing approach. Initial experiments aimed at differentiating the $\Delta I(t)$ excursions associated with each dNTP demonstrated the challenges inherent to this goal. The incorporation rates for dTTP or dATP could be readily distinguished from dGTP or dCTP incorporation rates during processing of complementary, homopolymeric DNA templates.[102] However, differentiating purines from pyrimidines is difficult as their excursion features overlap substantially. Experiments with dNTP analogs demonstrate one approach to differentiate between the two of bases. Measurements of KF-nanocircuits with 2'-deoxynucleoside-5'-O-(1-thiotriphosphate) (α-thio-dNTP) or 2-amino-6-Cl-purine-2'-deoxyriboside-triphosphate (6-Cl-2APTP) analogs (**Figure 1-11A**) produced $\Delta I(t)$ signals that appeared similar to the native dNTPs (**Figure 1-11B).** However, histograms of the $\Delta I(t)$ excursion durations and waiting times between $\Delta I(t)$ excursions (**Figure 1-11C**) showed differences in the incorporation rates for dNTP analogs (**Figure 1-11D**). 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; this is the opposite direction from signals with native dNTPs (**Figure 1-11D**). The approach, like other electronic sensing techniques,[6,105] remains highly promising for DNA sequencing, but requires further efforts focused on the pyrimidine-purine differentiation problem.



D

template	nucleotide	τ_{open} (ms)	τ_{closed} (ms)	k (s^{-1})
PolyC	dGTP	40.8 ± 6.0	0.40 ± 0.20	24.3 ± 4.3
PolyC	α-thio-dGTP	63.6 ± 2.8	0.21 ± 0.15	15.7 ± 0.7
PolyC	6-Cl-2ATPP	50.5 ± 1.4	0.20 ± 0.12	19.7 ± 0.6

Figure 1-11. KF processing native and unnatural substrates. A) Chemical structures of representative dNTP analogs with chemical modifications from the native dNTPs highlighted in red. **B)** $\Delta I(t)$ excursions during native and analog dNTP incorporation with a magnified view of a typical $\Delta I(t)$ excursion (right). **C**) Direct comparison of the probability distributions of $\langle\tau_{closed}\rangle$ (top) and $\langle\tau_{open}\rangle$ (bottom) durations during incorporation of the indicated dNTPs into PolyC. **D)** Table of kinetic native and analog dNTP incorporations by KF into PolyC. Data represents average values ± standard deviation. The images in (A-D) were adapted from ref[102]. Copyright the 2015 American Chemical Society.

1.6 Conclusion

Recent advances in single-molecule experiments have overcome challenges that hinder employing such investigations for biology, biochemistry, and biophysics. A lack of

sensitivity, bandwidth, and mechanical stability has largely constrained the early single-molecule electronics.[79,106] To expand nanofabrication techniques, a promising electronic architecture based on field-effect transistors (FETs) has been explored using carbon nanotubes,[20] graphene,[107,108] silicon nanowires,[21] nanopores,[22] and nanotransistors.[23] Our laboratory has paved the way for fabricating nanocircuits with individual enzymes to probe their catalytic functions. Furthermore, this collaboration has developed a stable and high bandwidth capable of sensing over extended periods of time (e.g., up to 2-weeks) for monitoring single-molecule dynamics.

Improvements to nanocircuits will benefit many aspects of the studies illustrated here. Robust and scalable bioelectronic sensing methods are required to monitor an even larger number of diverse individual enzymes. More site-directed attachment strategies are also required to accelerate the study of enzymes with many cysteines. Additionally, the attachment site should be designed in such a manner that conformational changes cause observable $\Delta I(t)$ excursions in the $\Delta I(t)$ signal recording by the nanocircuit while not restricting the protein movement.[109] Current investigations are underway to lay a foundation for a reproducible and scalable bioelectronic sensing technology to develop a tool for single-molecule enzymology on a large scale. Furthermore, developments in key aspects and data processing are necessary to better discriminate the enzyme's activity from noisy background, simplifying data analysis. Current and future work on nanocircuits will fine-tune this technology with orders-of-magnitude improvements for single-molecule enzymology.

The continuous recording of real-time complex biological processes with molecule-by-molecule precision is crucial for revealing complex, hidden enzyme dynamics.

Bioelectronic sensing has been a widely employed technique to monitor chemical reactions, molecular motions, conformational dynamics, and enzymatic activities. This list of individual molecules tested will continue to grow as electronic experimental techniques are improved. The past two decades has witnessed an explosion in single-molecule bioelectronic sensing for numerous applications. In particular, the development of nanocircuit to study enzyme motions has extended single-molecule enzymology unexplored timescales (**Figure 1-12**).

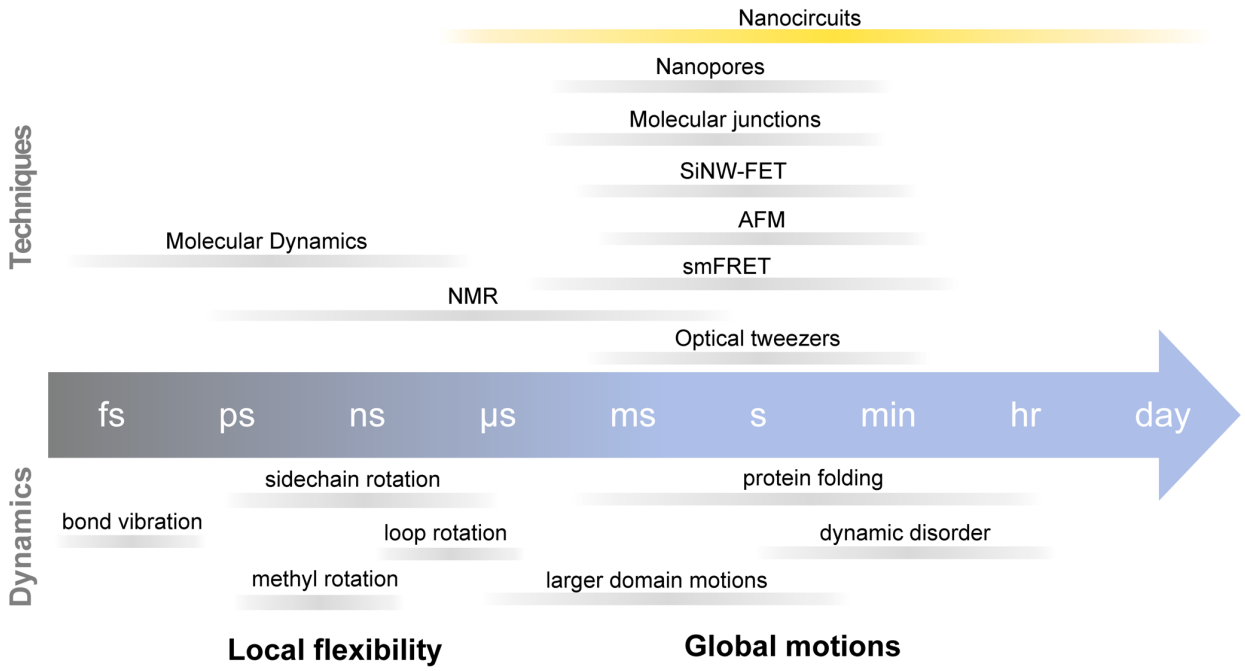


Figure 1-12. Nanocircuits extend single-molecule enzymology to relatively unexplored timescales. Single-molecule sensing from our laboratory by nanocircuits (yellow) and other laboratories using similar electronic methods or other techniques (molecular dynamics, fluorescence, or force-based). Image adapted from ref.[110]

Mechanistic questions in enzymology that were previously out of reach because of ensemble averaging are accessible with this and related bioelectronic sensing techniques. Lessons learned from nanocircuit development should be broadly applicable to other single-molecule approaches. In cases where fluorescent labeling is beyond the

bounds of possibility, nanocircuits can establish detailed recordings of individual enzymes. The ability to collect long-duration recordings up to microsecond resolution using nanocircuits opens the possibility of observing new effects in unexplored regimes. For example, some single molecule systems show memory effects, fluctuating, flickering, or stochastic behavior. This behavior can provide a glimpse into the underlying physical mechanism that is often obscured by averaging in ensemble populations. The ability to probe catalytic functions of individual molecules provides promise for the growing field of single-molecule studies and applications in biology and biotechnology.

1.7 References

(1) Voith von Voithenberg, L.; Lamb, D. C. Single Pair Förster Resonance Energy Transfer: A Versatile Tool To Investigate Protein Conformational Dynamics. *BioEssays* **2018**, *40*, 1–14.

(2) Lu, H. P. Single-Molecule Enzymatic Dynamics. *Science* **1998**, *282*, 1877–1882.

(3) Xie, X. S.; Choi, P. J.; Li, G.-W.; Lee, N. K.; Lia, G. Single-Molecule Approach to Molecular Biology in Living Bacterial Cells. *Annu. Rev. Biophys.* **2008**, *37*, 417–444.

(4) Uphoff, S.; Sherratt, D. J. Single-Molecule Analysis of Bacterial DNA Repair and Mutagenesis. *Annu. Rev. Biophys.* **2017**, *46*, 411–432.

(5) van Dijk, E. L.; Jaszczyszyn, Y.; Naquin, D.; Thermes, C. The Third Revolution in Sequencing Technology. *Trends Genet.* **2018**, *34*, 666–681.

(6) Agah, S.; Zheng, M.; Pasquali, M.; Kolomeisky, A. B. DNA Sequencing by Nanopores: Advances and Challenges. *J. Phys. D. Appl. Phys.* **2016**, *49*, 1–23.

(7) Moffitt, J. R.; Chemla, Y. R.; Smith, S. B.; Bustamante, C. Recent Advances in Optical Tweezers. *Annu. Rev. Biochem.* **2008**, *77*, 205–228.

(8) De Vlaminck, I.; Dekker, C. Recent Advances in Magnetic Tweezers. *Annu. Rev. Biophys.* **2012**, *41*, 453–472.

(9) Roy, R.; Hohng, S.; Ha, T. A Practical Guide to Single Molecule FRET. *Nat.*

Methods **2008**, *5*, 507–516.

(10) Moerner, W. E.; Kador, L. Optical Detection and Spectroscopy of Single Molecules in a Solid. *Phys. Rev. Lett.* **1989**, *62*, 2535–2538.

(11) Armani, A. M.; Kulkarni, R. P.; Fraser, S. E.; Flagan, R. C.; Vahala, K. J. Label-Free, Single-Molecule Detection with Optical Microcavities. *Science* **2007**, *317*, 783–787.

(12) Liebel, M.; Toninelli, C.; Van Hulst, N. F. Room-Temperature Ultrafast Nonlinear Spectroscopy of a Single Molecule. *Nat. Photonics* **2018**, *12*, 45–49.

(13) Jia, C.; Ma, B.; Xin, N.; Guo, X. Carbon Electrode-Molecule Junctions: A Reliable Platform for Molecular Electronics. *Acc. Chem. Res.* **2015**, *48*, 2565–2575.

(14) Choi, Y.; Weiss, G. A.; Collins, P. G. Single-Molecule Bioelectronics. In *Handbook of Bioelectronics*; 2015; pp 66–85.

(15) Walter, N. G.; Huang, C.-Y.; Manzo, A. J.; Sobhy, M. A. Do-It-Yourself Guide: How to Use the Modern Single Molecule Toolkit. *Nat. Methods* **2010**, *5*, 254–257.

(16) Li, Y.; Yang, C.; Guo, X. Single-Molecule Electrical Detection: A Promising Route toward the Fundamental Limits of Chemistry and Life Science. *Acc. Chem. Res.* **2020**, *53*, 159–169.

(17) Li, Y.; Zhao, L.; Yao, Y.; Guo, X. Single-Molecule Nanotechnologies: An Evolution in Biological Dynamics Detection. *ACS Appl. Bio Mater.* **2020**, *3*, 68–85.

(18) Choi, Y.; Moody, I. S.; Sims, P. C.; Hunt, S. R.; Corso, B. L.; Perez, I.; Weiss, G. A.; Collins, P. G. Single-Molecule Lysozyme Dynamics Monitored by an Electronic Circuit. *Science* **2012**, *335*, 319–324.

(19) Gauer, J. W. Single-Molecule FRET to Measure Conformational Dynamics of DNA Mismatch Repair Proteins. *Methods Enzymol.* **2016**, *581*, 285–315.

(20) Guo, X.; Small, J. P.; Klare, J. E.; Wang, Y.; Purewal, M. S.; Tam, I. W.; Hong, B. H.; Caldwell, R.; Huang, L.; O'Brien, S.; Yan, J.; Breslow, R.; Wind, S. J.; Hone, J.; Kim, P.; Nuckolls, C. Covalently Bridging-Gaps in Single-Walled Carbon Nanotubes with Conducting Molecules. *Science* **2006**, *311*, 356–359.

(21) He, G.; Li, J.; Ci, H.; Qi, C.; Guo, X. Direct Measurement of Single-Molecule DNA Hybridization Dynamics with Single-Base Resolution. *Angew. Chemie - Int. Ed.* **2016**, *55*, 9036–9040.

(22) Howorka, S.; Siwy, Z. Nanopore Analytics: Sensing of Single Molecules. *Chem. Soc. Rev.* **2009**, *38*, 2360–2384.

(23) Macchia, E.; Manoli, K.; Holzer, B.; Di Franco, C.; Ghittorelli, M.; Torricelli, F.; Alberga, D.; Mangiatordi, G. F.; Palazzo, G.; Scamarcio, G.; Torsi, L. Single-Molecule Detection with a Millimetre-Sized Transistor. *Nat. Commun.* **2018**, *9*, 1–10.

(24) Moody, I. S.; Choi, Y.; Olsen, T. J.; Sims, P. C.; Collins, P. G.; Weiss, G. A. Dissecting Lysozyme by Single-Molecule Techniques. *Lysozymes Sources, Funct. Role Dis.* **2013**, 193–213.

(25) Choi, Y.; Olsen, T. J.; Sims, P. C.; Moody, I. S.; Corso, B. L.; Dang, M. N.; Weiss, G. A.; Collins, P. G. Dissecting Single-Molecle Signal Transduction in Carbon Nanotube Circuits with Protein Engineering. *Nano Lett.* **2013**, *13*, 625–631.

(26) Olsen, T. J.; Choi, Y.; Sims, P. C.; Gul, O. T.; Corso, B. L.; Dong, C.; Brown, W. A.; Collins, P. G.; Weiss, G. A. Electronic Measurements of Single-Molecule Processing by DNA Polymerase I (Klenow Fragment). *J. Am. Chem. Soc.* **2013**, *135*, 7855–7860.

(27) Akhterov, M. V.; Choi, Y.; Olsen, T. J.; Sims, P. C.; Iftikhar, M.; Gul, O. T.; Corso, B. L.; Weiss, G. A.; Collins, P. G. Observing Lysozyme's Closing and Opening Motions by High-Resolution Single-Molecule Enzymology. *ACS Chem. Biol.* **2015**, *10*, 1495–1501.

(28) Cheung, C. L.; Lieber, C. M. Covalently Functionalized Nanotubesasnanometre-Sizedprobes in Chemistryandbiology. *Nature* **1998**, *394*, 52–55.

(29) Chen, R. J.; Zhang, Y.; Wang, D.; Dai, H. Noncovalent Sidewall Functionalization of Single-Walled Carbon Nanotubes for Protein Immobilization. *J. Am. Chem. Soc.* **2001**, *123*, 3838–3839.

(30) Goldsmith, B. R.; Coroneus, J. G.; Khalap, V. R.; Kane, A. A.; Weiss, G. A.; Collins, P. G. Conductance-Controlled Point Functionalization of Single-Walled Carbon Nanotubes. *Science* **2007**, *315*, 77–81.

(31) Goldsmith, B. R.; Coroneus, J. G.; Weiss, G. A.; Collins, P. G. Scaffolding Carbon Nanotubes into Single-Molecule Circuitry. *Mater. Res. Soc. Symp. Proc.* **2007**, *1018*, 99–104.

(32) Sims, P. C.; Moody, I. S.; Choi, Y.; Dong, C.; Corso, B. L.; Gul, O. T.; Collins, P. G.; Weiss, G. A. Electronic Measurements of Single-Molecule Catalysis by CAMP-Dependent Protein Kinase A Electronic Measurements of Single-Molecule Catalysis by CAMP-Dependent Protein Kinase A. *J. Am. Chem. Soc.* **2013**, *135*, 7861−7868.

(33) Richardson, M. B.; Gabriel, K. N.; Garcia, J.; Ashby, S.; Dyer, R.; Kim, J.; Lau, C.; Hong, J.; Le Tourneau, R. J.; Sen, S.; Narel, D.; Katz, B. B.; Ziller, J. W.; Majumdar,

S.; Collins, P. G.; Weiss, G. A. Pyrocinchonimides Conjugate to Amine Groups on Proteins via Imide Transfer. *Bioconjug. Chem.* **2020**, *31*, 1449–1462.

(34) Tomonari, Y.; Murakami, H.; Nakashima, N. Solubilization of Single-Walled Carbon Nanotubes by Using Polycyclic Aromatic Ammonium Amphiphiles in Water—Strategy for the Design of High-Performance Solubilizers. *Chem. - A Eur. J.* **2006**, *12*, 4027–4034.

(35) Choi, Y.; Moody, I. S.; Sims, P. C.; Hunt, S. R.; Corso, B. L.; Seitz, D. E.; Blaszcazk, L. C.; Collins, P. G.; Weiss, G. A. Single-Molecule Dynamics of Lysozyme Processing Distinguishes Linear and Cross-Linked Peptidoglycan Substrates. *J. Am. Chem. Soc.* **2012**, *134*, 2032–2035.

(36) Kuroki, R.; Weaver, L. H.; Matthews, B. W.; Kuroki, R.; Weaver, L. H.; Matthewst, B. W. A Covalent Enzyme-Substrate Intermediate with Saccharide Distortion in a Mutant T4 Lysozyme. *Science* **1993**, *262*, 2030–2033.

(37) Kim, Y.; Eom, S. H.; Wang, J.; Lee, D.-S.; Suh, S. W.; Steitz, T. A. Crystal Structure of Thermus Aquaticus DNA Polymerase. *Nature* **1995**, *376*, 612–616.

(38) Smith, M. E. B.; Schumacher, F. F.; Ryan, C. P.; Tedaldi, L. M.; Papaioannou, D.; Waksman, G.; Caddick, S.; Baker, J. R. Protein Modification, Bioconjugation, and Disulfide Bridging Using Bromomaleimides. *J. Am. Chem. Soc.* **2010**, *132*, 1960–1965.

(39) Lerner, M. B.; Resczenski, J. M.; Amin, A.; Johnson, R. R.; Goldsmith, J. I.; Johnson, A. T. C. Toward Quantifying the Electrostatic Transduction Mechanism in Carbon Nanotube Molecular Sensors. *J. Am. Chem. Soc.* **2012**, *134*, 14318–14321.

(40) Hunt, S. R.; Wan, D.; Khalap, V. R.; Corso, B. L.; Collins, P. G. Scanning Gate Spectroscopy and Its Application to Carbon Nanotube Defects. *Nano Lett.* **2011**, *11*, 1055–1060.

(41) Choi, Y.; Weiss, G. A.; Collins, P. G. Single Molecule Recordings of Lysozyme Activity. *Phys. Chem. Chem. Phys.* **2013**, *15*, 14879–14895.

(42) Kuroki, R.; Weaver, L. H.; Matthews, B. W. Structural Basis of the Conversion of T4 Lysozyme into a Transglycosidase by Reengineering the Active Site. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 8949–8954.

(43) Sorgenfrei, S.; Chiu, C.; Johnston, M.; Nuckolls, C.; Shepard, K. L. Debye Screening in Single-Molecule Carbon Nanotube Field-Effect Transistors. *Nano Lett.* **2011**, *11*, 13739–3743.

(44) Choi, Y.; Olsen, T. J.; Sims, P. C.; Moody, I. S.; Corso, B. L.; Dang, M. N.; Weiss, G. A.; Collins, P. G. Supporting Information: Dissecting Single-Molecule Signal

Transduction in Carbon Nanotube Circuits with Protein Engineering. *Nano Lett.* **2013**, *13*, 625–631.

(45) Lerner, E.; Barth, A.; Hendrix, J.; Ambrose, B.; Birkedal, V.; Blanchard, S. C.; Börner, R.; Chung, H. S.; Cordes, T.; Craggs, T. D.; Deniz, A. A.; Diao, J.; Fei, J.; Gonzalez, R. L.; Gopich, I. V.; Ha, T.; Hanke, C. A.; Haran, G.; Hatzakis, N. S.; Hohng, S.; Hong, S. C.; Hugel, T.; Ingargiola, A.; Joo, C.; Kapanidis, A. N.; Kim, H. D.; Laurence, T.; Lee, N. K.; Lee, T. H.; Lemke, E. A.; Margeat, E.; Michaelis, J.; Michalet, X.; Myong, S.; Nettels, D.; Peulen, T. O.; Ploetz, E.; Razvag, Y.; Robb, N. C.; Schuler, B.; Soleimaninejad, H.; Tang, C.; Vafabakhsh, R.; Lamb, D. C.; Seidel, C. A. M.; Weiss, S.; Boudker, O. FRET-Based Dynamic Structural Biology: Challenges, Perspectives and an Appeal for Open-Science Practices. *Elife* **2021**, *10*, 1–69.

(46) Juette, M. F.; Terry, D. S.; Wasserman, M. R.; Altman, R. B.; Zhou, Z.; Zhao, H.; Blanchard, S. C. Single-Molecule Imaging of Non-Equilibrium Molecular Ensembles on the Millisecond Timescale. *Nat. Methods* **2016**, *13*, 341–344.

(47) Swaminathan, R.; Ravi, V. K.; Kumar, S.; Kumar, M.; Satish, V.; Chandra, N. Lysozyme: A Model Protein for Amyloid Research. *Adv. Protein Chem. Struct. Biol.* **2011**, *84*, 63–111.

(48) Yirdaw, R. B.; McHaourab, H. S. Direct Observation of T4 Lysozyme Hinge-Bending Motion by Fluorescence Correlation Spectroscopy. *Biophys. J.* **2012**, *103*, 1525–1536.

(49) Chen, Y.; Hu, D.; Vorpagel, E. R.; Lu, H. P. Probing Single-Molecule T4 Lysozyme Conformational Dynamics by Intramolecular Fluorescence Energy Transfer. *J. Phys. Chem. B* **2003**, *107*, 7947–7956.

(50) Wang, Y.; Lu, H. P. Bunching Effect in Single-Molecule T4 Lysozyme Nonequilibrium Conformational Dynamics under Enzymatic Reactions. *J. Phys. Chem. B* **2010**, *114*, 6669–6674.

(51) Matsunaga, T.; Hishiya, T.; Takeuchi, T. Surface Plasmon Resonance Sensor for Lysozyme Based on Molecularly Imprinted Thin Films. *Anal. Chim. Acta* **2007**, *591*, 63–67.

(52) Mi, F.-L.; Tan, Y.-C.; Liang, H.-C.; Huang, R.-N.; Sung, H.-W. In Vitro Evaluation of a Chitosan Membrane Cross-Linked with Genipin. *J. Biomater. Sci. Polym. Ed.* **2001**, *12*, 835–850.

(53) Azevedo, H. S.; Reis, R. L. Understanding the Enzymatic Degradation of Biodegradable Polymers and Strategies to Control Their Degradation Rate. In *Biodegradable Systems in Tissue Engineering and Regenerative Medicine*; CRC Press: Boca Raton, FL, 2005; p 183.

(54) Hwang, W.; Karplus, M. Structural Basis for Power Stroke vs. Brownian Ratchet Mechanisms of Motor Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 19777–19785.

(55) Saffarian, S.; Collier, I. E.; Marmer, B. L.; Elson, E. L.; Goldberg, G. Interstitial Collagenase Is a Brownian Ratchet Driven by Proteolysis of Collagen. *Science* **2004**, *306*, 108–111.

(56) Dimura, M.; Peulen, T. O.; Hanke, C. A.; Prakash, A.; Gohlke, H.; Seidel, C. A. Quantitative FRET Studies and Integrative Modeling Unravel the Structure and Dynamics of Biomolecular Systems. *Curr. Opin. Struct. Biol.* **2016**, *40*, 163–185.

(57) Sanabria, H.; Rodnin, D.; Hemmen, K.; Peulen, T. O.; Felekyan, S.; Fleissner, M. R.; Dimura, M.; Koberling, F.; Kühnemuth, R.; Hubbell, W.; Gohlke, H.; Seidel, C. A. M. Resolving Dynamics and Function of Transient States in Single Enzyme Molecules. *Nat. Commun.* **2020**, *11*.

(58) Sagermann, M.; Mårtensson, L.-G.; Baase, W. A.; Matthews, B. W. A Test of Proposed Rules for Helix Capping: Implications for Protein Design. *Protein Sci.* **2009**, *11*, 516–521.

(59) Bauman, A. L.; Scott, J. D. Kinase- and Phosphatase-Anchoring Proteins: Harnessing the Dynamic Duo. *Nat. Cell Biol.* **2002**, *1*, 1–4.

(60) Chin, K. V.; Yang, W. L.; Ravatn, R.; Kita, T.; Reitman, E.; Vettori, D.; Cvijic, M. E.; Shin, M.; Iacono, L. Reinventing the Wheel of Cyclic AMP Novel Mechanisms of CAMP Signaling. *Ann. N. Y. Acad. Sci.* **2002**, *968*, 49–64.

(61) Johnson, D. A.; Akamine, P.; Radzio-Andzelm, E.; Madhusudan; Taylor, S. S. Dynamics of CAMP-Dependent Protein Kinase. *Chem. Rev.* **2001**, *101*, 2243–2270.

(62) Robinson-White, A.; Stratakis, C. A. Protein Kinase a Signaling: “Cross-Talk” with Other Pathways in Endocrine Cells. *Ann. N. Y. Acad. Sci.* **2002**, *968*, 256–270.

(63) Shabb, J. B. Physiological Substrates of CAMP-Dependent Protein Kinase. *Chem. Rev.* **2001**, *101*, 2381–2411.

(64) Isobe, K.; Jung, H. J.; Yang, C. R.; Claxton, J.; Sandoval, P.; Burg, M. B.; Raghuram, V.; Knepper, M. A. Systems-Level Identification of PKA-Dependent Signaling in Epithelial Cells. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 8875–8884.

(65) Adams, J. A. Kinetic and Catalytic Mechanisms of Protein Kinases. *Chem. Rev.* **2001**, *101*, 2271–2290.

(66) Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S. The Protein

Kinase Complement of the Human Genome. *Science* **2002**, *298*, 1912–1934.

(67) Taylor, S. S.; Yang, J.; Wu, J.; Haste, N. M.; Radzio-Andzelm, E.; Anand, G. PKA: A Portrait of Protein Kinase Dynamics. *Biochim. Biophys. Acta - Proteins Proteomics* **2004**, *1697*, 259–269.

(68) Kemp, B. E.; Graves, D. J.; Benjamini, E.; Krebs, E. G. Role of Multiple Basic Residues in Determining the Substrate Specificity of Cyclic AMP Dependent Protein Kinase. *J. Biol. Chem.* **1977**, *252*, 4888–4894.

(69) Adams, J. A.; Taylor, S. S. Energetic Limits of Phosphotransfer in the Catalytic Subunit of CAMP-Dependent Protein Kinase As Measured by Viscosity Experiments. *Biochemistry* **1992**, *31*, 8516–8522.

(70) Chu, W. T.; Chu, X.; Wang, J. Binding Mechanism and Dynamic Conformational Change of C Subunit of PKA with Different Pathways. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, E7959–E7968.

(71) Whitehouse, S.; Feramisco, J. R.; Casnellie, J. E.; Krebs, E. G.; Walsh, D. A. Studies on the Kinetic Mechanism of the Catalytic Subunit of the CAMP-Dependent Protein Kinase. *J. Biol. Chem.* **1983**, *258*, 3693–3701.

(72) Kong, C. te; Cook, P. F. Isotope Partitioning In The Adenosine 3',5'-Monophosphate Dependent Protein Kinase Reaction Indicates A Steady-State Random Kinetic Mechanism. *Biochemistry* **1988**, *27*, 4795–4799.

(73) Ni, D. Q.; Shaffer, J.; Adams, J. A. Insights into Nucleotide Binding in Protein Kinase A Using Fluorescent Adenosine Derivatives. *Protein Sci.* **2000**, *9*, 1818–1827.

(74) Grant, B. D.; Adams, J. A. Pre-Steady-State Kinetic Analysis of CAMP-Dependent Protein Kinase Using Rapid Quench Flow Techniques. *Biochemistry* **1996**, *35*, 2022–2029.

(75) Masterson, L. R.; Mascioni, A.; Traaseth, N. J.; Taylor, S. S.; Veglia, G. Allosteric Cooperativity in Protein Kinase A. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 506–511.

(76) Chen, Q.; Groote, R.; Schönherr, H.; Vancso, G. J. Probing Single Enzyme Kinetics in Real-Time. *Chem. Soc. Rev.* **2009**, *38*, 2671–2683.

(77) Santoso, Y.; Joyce, C. M.; Potapova, O.; Le Reste, L.; Hohlbein, J.; Torella, J. P.; Grindley, N. D. F.; Kapanidis, A. N. Conformational Transitions in DNA Polymerase I Revealed by Single-Molecule FRET. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 715–720.

(78) Xie, S. Single-Molecule Approach to Enzymology. *Single Mol.* **2001**, *2*, 229–236.

(79) Lu, H. P. Single-Molecule Protein Interaction Conformational Dynamics. *Curr. Pharm. Biotechnol.* **2009**, *10*, 522–531.

(80) Masterson, L. R.; Cheng, C.; Yu, T.; Tonelli, M.; Kornev, A.; Taylor, S. S.; Veglia, G. Dynamics Connect Substrate Recognition to Catalysis in Protein Kinase A. *Nat. Cell Biol.* **2010**, *6*, 821–828.

(81) Piserchio, A.; Warthaka, M.; Kaoud, T. S.; Callaway, K.; Dalby, K. N.; Ghose, R. Local Destabilization, Rigid Body, and Fuzzy Docking Facilitate the Phosphorylation of the Transcription Factor Ets-1 by the Mitogen-Activated Protein Kinase ERK2. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 6287–6296.

(82) Gerlits, O.; Das, A.; Keshwani, M. M.; Taylor, S.; Waltman, M. J.; Langan, P.; Heller, W. T.; Kovalevsky, A. Metal-Free CAMP-Dependent Protein Kinase Can Catalyze Phosphoryl Transfer. *Biochemistry* **2014**, *53*, 3179–3186.

(83) Solorza, J.; Recabarren, R.; Alzate-Morales, J. Molecular Insights into the Trapping Effect of Ca2+ in Protein Kinase A: A Molecular Dynamics Study. *J. Chem. Inf. Model.* **2020**, *60*, 898–914.

(84) Echols, H.; Goodman, M. F. Fidelity Mechanisms in DNA Replication. *Annu. Rev. Biochem.* **1991**, *60*, 477–511.

(85) Joyce, C. M.; Grindley, N. D. F. Construction of a Plasmid That Overproduces the Large Proteolytic Fragment (Klenow Fragment) of DNA Polymerase I of Escherichia Coli. *Proc. Natl. Acad. Sci. U. S. A.* **1983**, *80*, 1830–1834.

(86) Joyce, C. M.; Potapova, O.; DeLucia, A. M.; Huang, X.; Basu, V. P.; Grindley, N. D. F. Fingers-Closing and Other Rapid Conformational Changes in DNA Polymerase I (Klenow Fragment) and Their Role in Nucleotide Selectivity. *Biochemistry* **2008**, *47*, 6103–6116.

(87) Ollis, D. L.; Brick, P.; Hamlio, R.; Xuong, N. G.; Steitz, T. A. Structure of Large Fragment of Escherichia Coli DNA Polymerase I Complexed with DTMP. *Struct. Insights into Gene Expr. Protein Synth.* **1985**, *4*, 235–239.

(88) Steitz, T. A. DNA Polymerases: Structural Diversity and Common Mechanisms. *J. Biol. Chem.* **1999**, *274*, 17395–17398.

(89) Freemont, P. S.; Friedman, J. M.; Beese, L. S.; Sandersont, M. R.; Steitz, T. A. Cocrystal Structure of an Editing Complex of Klenow Fragment with DNA (3'-5' Exonuclease/DNA Polymerase/Protein-DNA Interaction/x-Ray Crystallography/Metal Ion Catalysis). *Biochemistry* **1988**, *85*, 8924–8928.

(90) Patel, P. H.; Suzuki, M.; Adman, E.; Shinkai, A.; Loeb, L. A. Prokaryotic DNA Polymerase I: Evolution, Structure, and "Base Flipping" Mechanism for Nucleotide

Selection. *J. Mol. Biol.* **2001**, *308*, 823–837.

(91) Joyce, C. M.; Steitz, T. A. Polymerase Structures and Function: Variations on a Theme? *J. Bacteriol.* **1995**, *177*, 6321–6329.

(92) Miller, B. R.; Beese, L. S.; Parish, C. A.; Wu, E. Y. The Closing Mechanism of DNA Polymerase I at Atomic Resolution. *Struct. Des.* **2015**, *23*, 1609–1620.

(93) Kuchta, R. D.; Benkovic, P.; Benkovic, S. J. Kinetic Mechanism Whereby DNA Polymerase I (Klenow) Replicates DNA with High Fidelity. *Biochemistry* **1988**, *27*, 6716–6725.

(94) Kuchta, R. D.; Mizrahi, V.; Benkovic, P. A.; Johnson, K. A.; Benkovic, S. J. Kinetic Mechanism of DNA Polymerase I (Klenow). *Biochemistry* **1987**, *26*, 8410–8417.

(95) Singh, K.; Srivastava, A.; Patel, S. S.; Modak, M. J. Participation of the Fingers Subdomain of Escherichia Coli DNA Polymerase I in the Strand Displacement Synthesis of DNA. *J. Biol. Chem.* **2007**, *282*, 10594–10604.

(96) 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* **1991**, *30*, 4835–4843.

(97) Bebenek, K.; Joyce, C. M.; Fitzgerald, M. P.; Kunkel, T. A. The Fidelity of DNA Synthesis Catalyzed by Derivatives of Escherichia Coli DNA Polymerase I*. *J. Biol. Chem.* **1990**, *265*, 13878–13887.

(98) 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* **1991**, *30*, 511–525.

(99) Rothwell, P. J.; Waksman, G. Structure and Mechanism of DNA Polymerases. *Adv. Protein Chem.* **2005**, *71*, 401–440.

(100) Eger, B. T.; Carroll, S. S.; Benkovic, P. A.; Dahlberg, M. E.; Benkovic, S. J.; Kuchta, R. D.; Joyce, C. M. Mechanism of DNA Replication Fidelity for Three Mutants of DNA Polymerase I: Klenow Fragment KF(Exo+), KF(PolA5), and KF(Exo−). *Biochemistry* **1991**, *30*, 1441–1448.

(101) 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.* **2012**, *40*, 7975–7984.

(102) 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.* **2012**, *134*, 11261–11268.

(103) Hohlbein, J.; Kapanidis, A. N. *Probing the Conformational Landscape of DNA Polymerases Using Diffusion-Based Single-Molecule FRET*, 1st ed.; Elsevier Inc., 2016; Vol. 581.

(104) Gül, O. T.; Pugliese, K. M.; Choi, Y.; Sims, P. C.; Pan, D.; Rajapakse, A. J.; Weiss, G. A.; Collins, P. G. Single Molecule Bioelectronics and Their Application to Amplification-Free Measurement of DNA Lengths. *Biosensors* **2016**, *6*, 1–19.

(105) Pugliese, K. M.; Gül, O. T.; Choi, Y.; Olsen, T. J.; Sims, P. C.; Collins, P. G.; Weiss, G. A. Processive Incorporation of Deoxynucleoside Triphosphate Analogs by Single-Molecule DNA Polymerase i (Klenow Fragment) Nanocircuits. *J. Am. Chem. Soc.* **2015**, *137*, 9587–9594.

(106) 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* **1996**, *35*, 7256–7266.

(107) Bryant, F. R.; Johnson, K. A.; Benkovic, S. J. Elementary Steps in the DNA Polymerase I Reaction Pathway. *Biochemistry* **1983**, *22*, 3537–3546.

(108) Shuo Huang; He, J.; Chang, S.; Zhang, P.; Liang, F.; Li, S.; Tuchb, M.; Fuhrman, A.; Ros, R.; Lindsay, S. Single Base Resolution in Tunneling Reads of DNA Composition. *Nat. Nanotechnol.* **2010**, *5*, 868–873.

(109) Lu, H. P. Probing Single-Molecule Protein Conformational Dynamics. *Acc. Chem. Res.* **2005**, *38*, 557–565.

(110) Schedin, F.; Geim, A. K.; Morozov, S. V.; Hill, E. W.; Blake, P.; Katsnelson, M. I.; Novoselov, K. S. Detection of Individual Gas Molecules Adsorbed on Graphene. *Nat. Mater.* **2007**, *6*, 652–655.

(111) Lu, Y.; Goldsmith, B. R.; Kybert, N. J.; Johnson, A. T. C. DNA-Decorated Graphene Chemical Sensors. *Appl. Phys. Lett.* **2010**, *97*, 2–5.

(112) Tian, B.; Xu, S.; Rogers, J. A.; Cestellos-blanco, S.; Yang, P.; Bezanilla, F.; Liu, J.; Bao, Z.; Hjort, M.; Cao, Y.; Melosh, N.; Lanzani, G.; Benfenati, F.; Galli, G.; Gygi, F.; Kautz, R.; Gorodetsky, A. A.; Kim, S. S.; Lu, T. K.; Anikeeva, P.; Cifra, M.; Havelka, D.; States, U.; Division, S.; Berkeley, L. Roadmap on Semiconductor–Cell Biointerfaces Bozhi. *Phys. Biol.* **2019**, *15*, 1–65.

(113) Henzler-Wildman, K.; Kern, D. Dynamic Personalities of Proteins. *Nature* **2007**, *450*, 964–972.

CHAPTER 2: Single-Molecule Taq DNA Polymerase Dynamics at Multiple Domains and Native Temperatures

2.1 Abstract

Taq DNA polymerase functions at elevated temperatures with fast conformational dynamics – regimes previously inaccessible to mechanistic, single molecule studies. Here, single-walled carbon nanotube transistors recorded the motions of Taq molecules processing matched or mismatched template-deoxynucleotide triphosphate pairs from 22 to 85 °C. By using four enzyme orientations, the whole-enzyme closures of nucleotide incorporations were distinguished from more rapid, 20-μs closures of Taq's fingers domain testing complementarity and orientation. On average, one transient closure was observed for every nucleotide binding event; even complementary substrate pairs averaged five transient closures between each catalytic incorporation at 72 °C. All of Taq's temperature sensitivity occurred in its rate-determining open state; the rate and duration of the transient closures and the catalytic events had almost no temperature dependence.

2.2 Introduction

DNA polymerases catalyze the reactions essential to life – DNA replication and repair.[1] They incorporate incoming deoxynucleotide triphosphates (dNTPs) into a nascent DNA strand that is complementary to a single-stranded DNA (ssDNA) template. This capability makes DNA polymerases workhorses for molecular biology and biotechnology.[2] In particular, the thermostable DNA polymerase from *Thermus aquaticus* (Taq) is widely used in the polymerase chain reaction (PCR).[3]

The simplest model for DNA polymerase function is a catalytic cycle between open and closed conformations.[1,4] In the open conformation, dNTP and ssDNA template bind to designated sites on the enzyme's "fingers" and "thumb" domains, respectively.[5] After binding, the two domains come together towards a closed conformation that either rejects the dNTP or catalytically incorporates it into the nascent ssDNA.[4,6] While this simple model is broadly accepted, fundamental questions about the fidelity-checking mechanism have motivated searches for additional intermediate conformations.[7,8] How does fidelity checking occur in the open conformation before the dNTP and DNA template can coordinate? Alternatively, if fidelity-checking is postponed until the catalytically active, closed conformation, how are misincorporation rates of 10^{-6} errors/dNTP achieved? Early research hypothesized the necessity of partly-closed, intermediate conformations where dNTP identification and mismatch rejection could occur before the closed conformation.[6,9]

Over the past decade, multiple breakthroughs began to identify and characterize these intermediate conformations in Taq and homologous A-family DNA polymerases. In 2011, Wu and Beese crystallized an "ajar" structure using a noncomplementary pairing

and mutagenic modifications that stabilized the conformation.[10] Independently, advances in single-molecule Förster resonance energy transfer (smFRET) inferred partly-closed conformations from fluorescence emission populations[8,11,12] and developed a dynamic model of rapidly interconverting conformational populations influenced by complementarity.[13,14] On the theoretical side, all-atom molecular dynamics (MD) advanced to produce the first detailed trajectories of closing and reopening transitions, [15,16] followed by observations of interconversion dynamics among a diverse ensemble of atomic configurations.[17,18] [17,18]. Taken together, this progress has revealed partially closed intermediates and dynamics between the open and closed conformations, but it has not directly observed the fidelity-checking mechanism nor resolved the debate over whether "conformational selection," "induced-fitting," or a combination of the two is the best reductionist model.[8,19–26] The diversity of trajectories merely indicates that fidelity-checking is more complex than a binary choice between two deterministic, concerted pathways.

Here, we contribute long-duration, single-molecule trajectories of Taq opening and closing upon either matched or mismatched dNTP-template pairs. Using a solid-state electronic technique with microsecond resolution, we monitored individual Taq molecules over minutes of activity, directly recording processive catalytic dNTP incorporations interspersed among 20 μs, non-catalytic closures. The results agree with the glimpses provided by smFRET measurements[11,12,14] or MD trajectories,[17] detailing the timing with which Taq tests for complementarity by closing upon matched or mismatched dNTPs. A further advantage of the electronic technique is its compatibility with a wide range of

temperatures; here, Taq's single-molecule activity was studied from 22 to 85 °C to cover the range relevant for PCR.

2.3 Results and Discussion

To record these trajectories, we attached Taq molecules to field-effect transistors (FETs) comprised of individual, single-walled carbon nanotubes (SWNTs, **Figure 2-1A**). The technique has proven useful with other enzymes,[27,28] including the Klenow Fragment (KF) of DNA polymerase.[29–31] Following methods similar to our previous work[29] and adopting the Taq variants studied by Xu *et al*.,[32] we used a pyrene-maleimide linker to bind individual Taq molecules to SWNT devices in four distinct orientations designed to transduce signals from the backside of Taq's intervening (R411C), palm (A814C), fingers (R695C), or thumb (E524C) domains (**Figure 2-2** and **Figure 2-3**). In this chapter, the resultant hybrid devices are each referred to as Taq-Domain (e.g., Taq-Intervening). Atomic force microscopy (AFM) of an example Taq-Intervening device is shown in **Figure 2-1B** and **Figure 2-19**. In each orientation, the movement of charged amino acid sidechains closest to the SWNT electrostatically modulated the source-drain current $I(t)$, whereas more distant domains were screened by the conductivity of the activity buffer (40 mM HEPES, 50 mM KCl, 5 mM MgCl2, pH 8.5), which had a Debye screening radius of 1.2 nm (**Figure 2-3**). The electrical current as a function of time, $I(t)$, was recorded while Taq interacted with a solution of homopolymeric "polyT" DNA primer-template (4 nM poly(dT)$_{42}$ fused to an M13 priming site and annealed with M13 primer, Table S1) and/or different dNTPs (typ. 10 μM). The $\Delta I(t)$ signals presented here were highpass filtered at 15 Hz to remove DC components of 10 to 500 nA and the lowest-frequency

fluctuations from the raw $I(t)$ data. The Supplementary Materials provide full descriptions of the experimental methods for production and purification of each full-length Taq variant and SWNT FET fabrication, bioconjugation, and measurement.

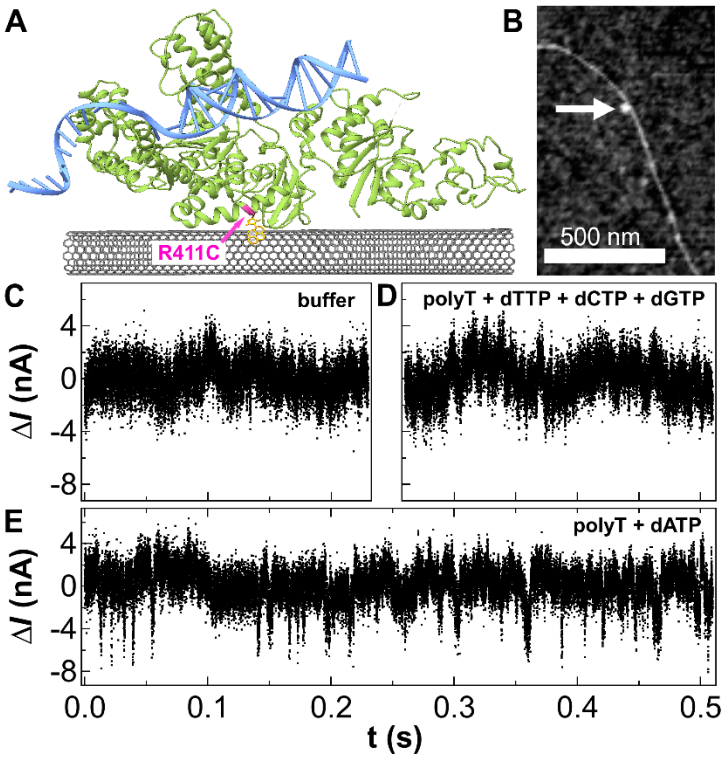


Figure 2-1. A Taq-Intervening device shown by schematic, AFM characterization, and signals generated at 72 °C. A) The Taq-Intervening device featured an individual Taq DNA polymerase (green) site-specifically bioconjugated to a SWNT through a cysteine in the Intervening domain (residue R411C, pink) to a pyrene-maleimide linker molecule (yellow). **B)** AFM image of a sample device with a single Taq attachment (white arrow). Representative $\Delta I(t)$ signals generated in solutions of **C)** activity buffer, and polyT with **D)** mismatched (dTTP + dCTP + dGTP) or **E)** matched (dATP) dNTPs. Two-level excursions only occurred when polyT and complementary dNTPs were both present. PDB: 1TAQ.[33]

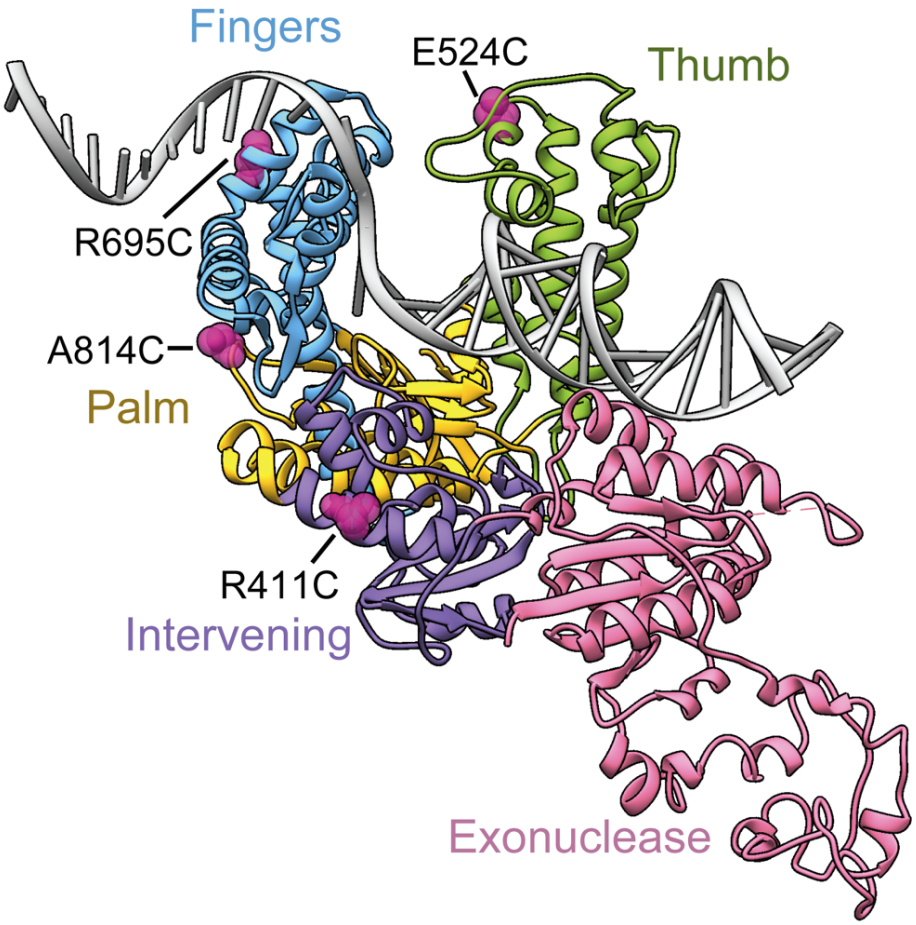


Figure 2-2. Locations of the four attachment sites on Taq. The position of the single cysteine attachment sites (pink spheres) probed Taq's motions in the Intervening (purple), Fingers (blue), Thumb (green), and Palm (yellow) domains during replication of DNA (gray). PDB: 1TAQ.[33]

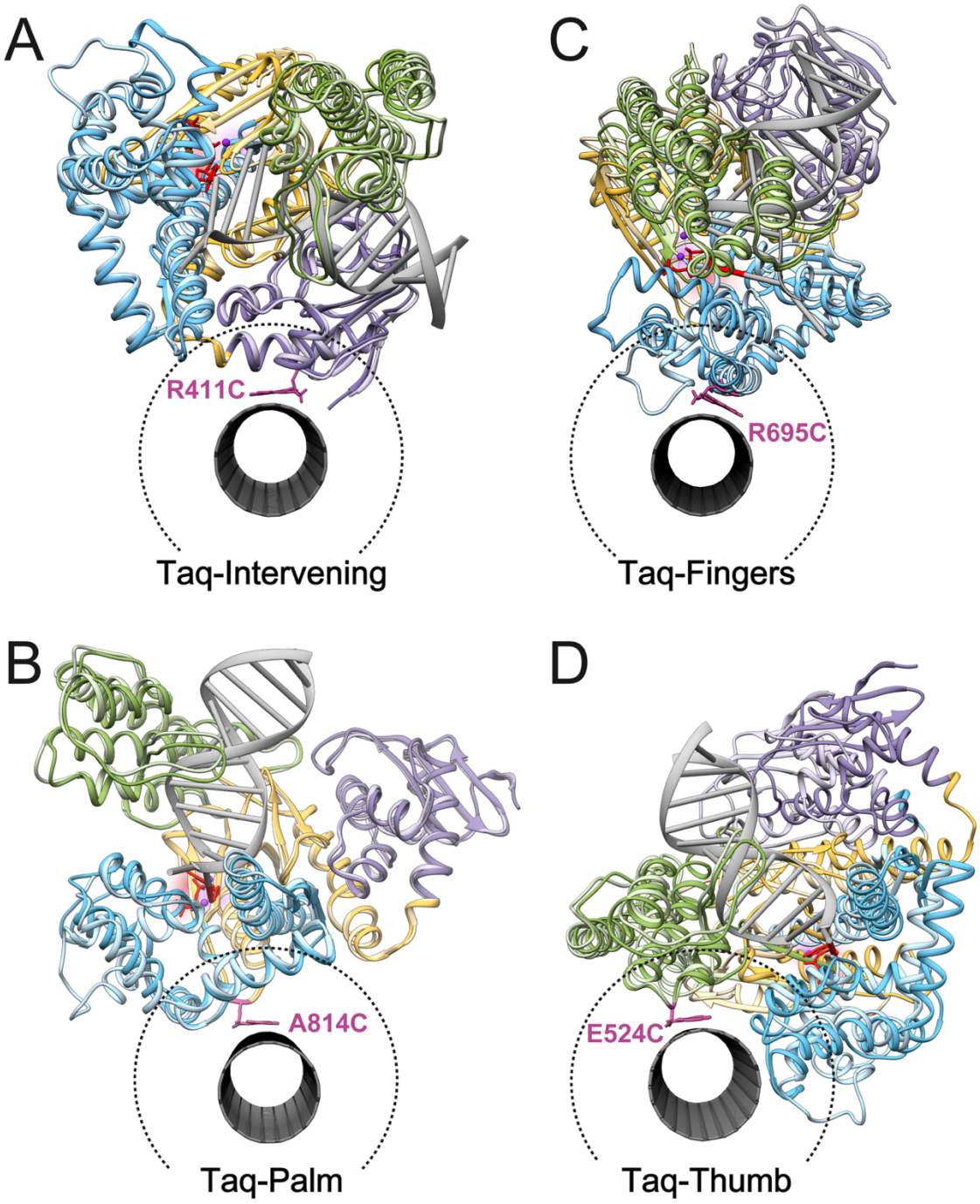


Figure 2-3. Comparing Taq open and closed motions. Here, the overlay includes atoms from both the active site DNA (light gray), incoming dNTP (red), and metal ions (purple) for the indicated cysteine attachment site (pink) via the pyrene maleimide linker on the SWNT (dark gray). The different hybrid devices probe enzyme motions from Taq's **A**) Intervening (purple), **B**) Palm (yellow), **C**) Fingers (blue), or **D**) Thumb (green) domains. The open (dark) and closed (light) structures show large secondary structure motions adjacent to the SWNT (dark gray) within a 1.2 nm radius (dotted circle). Such motions can perturb the local gating of the SWNT and generate the observed signals. PDB: 3KTQ and 4KTQ for open and closed structures of KlenTaq, respectively.[4]

Aside from being acquired at 72 °C, the signals generated by Taq-Intervening devices were essentially identical to previous electrical measurements of KF.[34] In activity buffer (**Figure 1C**) or with mismatched dNTPs (**Figure 1D**), $\Delta I(t)$ was a featureless band of 1/f noise. In the presence of matched dNTPs, additional two-level switching appeared (**Figure 1E**). At 72 °C, the average rate of Taq's two-level excursions below the baseline was 61 s^{-1}, effectively equal to Taq's catalytic rate k_{cat} at that temperature.[35] Led by this agreement and the similarity to previous KF results,[29] where event counting correctly enumerated DNA primer-template lengths,[31] we assigned $\Delta I(t)$ to a two-state model for dNTP incorporations. The $\Delta I(t) = 0$ baseline corresponded to Taq's open conformation, and each excursion recorded one dNTP incorporation occurring in Taq's closed conformation. For brevity, we henceforth refer to these excursions as "catalytic closures."

Temperature dependence provided new confirmation of the assignment of $\Delta I(t)$ excursions to Taq's catalytic closures. From 22 to 85 °C, the average rate of catalytic closures increased exponentially from 4 to 96 s^{-1} (**Figure 2**). While exact values varied from one molecule to another, the range of rates and the inferred activation energy of 11 kcal/mol was comparable to previous reports of $k_{cat.}$[36–39] As in ensemble measurements,[36,38] we observed a softening of the Arrhenius rate law above 50 °C, where thermal fluctuations decrease Taq's dNTP and primer-template binding stabilities. At 95 °C, $\Delta I(t)$ became disordered and not enumerable, but organized activity returned upon cooling devices below 85 °C.

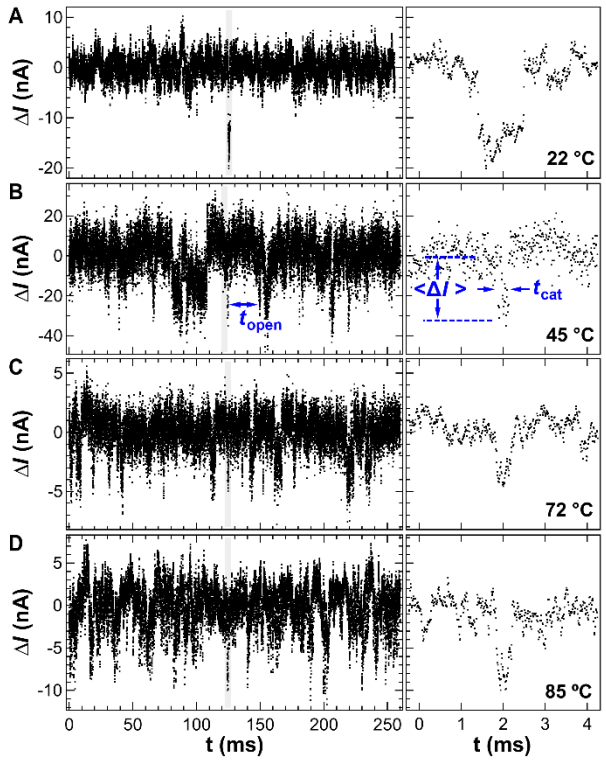


Figure 2-4. Temperature dependence of ΔI(t) and individual catalytic closures. Taq-Intervening with matched dATP and polyT generated different rates of catalytic closures at **A)** 22 °C, **B)** 45 °C, **C)** 72 °C and **D)** 85 °C. While the rate of catalytic closures increased with temperature (left), the duration of individual catalytic closures stayed constant (gray highlights and right). At right, individual data points represent 1 μs intervals, except where decimated for clarity in (A).

Figure 2-1 and **Figure 2-4** are short samples taken from continuous Δ*I*(*t*) records lasting no less than 5 minutes and, in some cases, as long as 45 minutes. At PCR temperatures, the full recordings contained thousands of catalytic closures per minute, each of which could be resolved and individually characterized in the manner depicted in **Figure 2-4B**. For quantitative statistical analysis, each catalytic closure was assigned a duration τ_{cat}, a waiting time before the closure τ_{open}, and an average amplitude <Δ*I*> measured from the event's instantaneous baseline. Transition times at the front and back step edges of each catalytic closure were also inspected but determined to be limited by the amplifier rise time (0.5 μs). Unlike previous work with T4 lysozyme, which directly

observed finite transition times averaging 37 μs,[40] most Taq open-to-closed transitions had sharp edges and rise times faster than 1 μs.

Figure 2-5 shows probability distribution functions for large populations of τ_{open} and τ_{cat} from Taq-Intervening activity at four temperatures. τ_{open} exhibited strong temperature dependence, consistent with previous ensemble work,[6] whereas τ_{cat} was largely temperature independent. k_{cat} was essentially determined by the former, since τ_{open} was 100 to 1000 times longer than τ_{cat}. Consequently, **Figure 2-5A** is a single-molecule representation of Taq's k_{cat} and its temperature dependence, illustrating the single-molecule, event-by-event composition of k_{cat} as well as its instantaneous variability. Solid lines in **Figure 2-5A** depict single-exponential fits for a Poisson process with a single, rate-limiting step. The slope of each line equals the mean value of k_{cat} at each temperature, so that steepening τ_{open} distributions graphically illustrate Taq's increasing activity up to 85 °C.

The τ_{open} distributions exceeded the exponential fits at longer durations, but this deviation occurred over many minutes of acquisition; shorter subpopulations of τ_{open} fit single exponentials with minute-by-minute variations in slope. Similar variability was observed in τ_{cat}, causing its cumulative distributions to resemble double-exponential curves (**Figure 2-5B**). This artificial stretching of single-exponential distributions can indicate external factors like temperature variation or dynamic disorder intrinsic to an enzyme,[41–43] but the latter explanation seems more likely since τ_{cat} had no other temperature sensitivity. **Table 2-1** summarizes numeric values for the fits in **Figure 2-5** and the range of single-exponential τ_{cat} fits observed at each temperature.

Table 2-1. Taq-Intervening open and closed durations processing matched substrate (polyT + dATP).

T (°C)	τ_{open} ± σ (ms)	k (s^{-1})	τ_{cat} ± σ (ms)
22	236 ± 15	4	0.67 ± 0.41
42	62 ± 2	16	0.57 ± 0.33
72	16.4 ± 0.6	61	0.45 ± 0.15
85	10.4 ± 0.2	96	0.33 ± 0.09

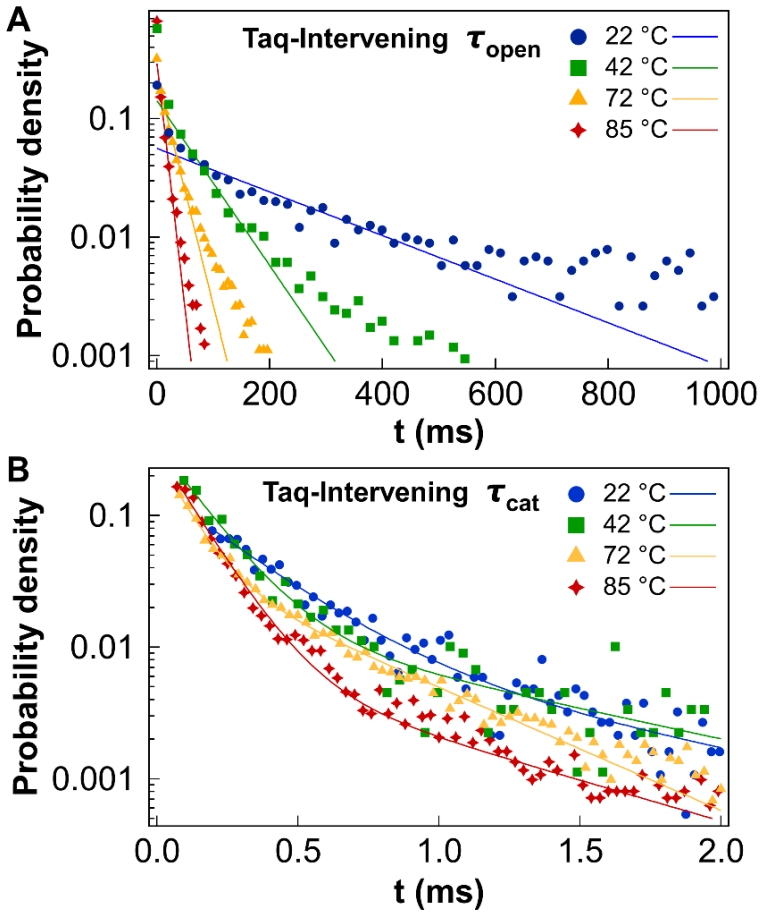


Figure 2-5. Taq-Intervening open and closed durations processing matched substrate (polyT + dATP). A) Probability distributions of the rate-determining waiting time, τ_{open}, with single-exponential fits. **B)** Probability distributions of the catalytic closure durations, τ_{cat}, with double-exponential fits.

Taq-Palm devices corroborated the findings from Taq-Intervening. The Taq-Palm orientation suffered from sparse activity, perhaps inhibited by its attachment to the SWNT

FET, but otherwise minute-long bursts of $\Delta I(t)$ excursions had the same characteristics as Taq-Intervening (**Figure 2-6**). Even studied in detail, though, the single-molecule trajectories from Taq-Palm and Taq-Intervening devices only observed k_{cat} activity already well established by ensemble and smFRET studies. The palm and intervening domains participated in the enzyme's global transformations during catalysis,[32] but they were too far from the incoming dNTPs to transduce the dynamic activity associated with complementarity checking (**Figure 2-3**). Studying those signals required new orientations that attached the SWNT FET directly to the fingers or thumb domains.

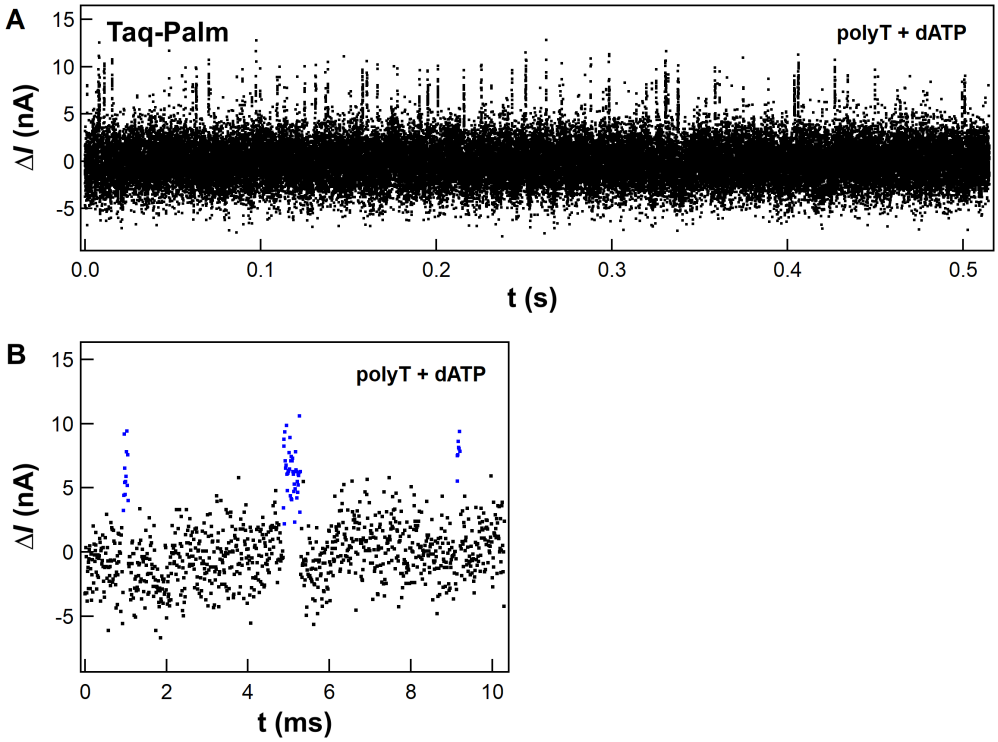


Figure 2-6. Representative signals generated by Taq-Palm at 72 °C. A) In matched polyT + dATP, Taq-Palm generated upward excursions that were otherwise similar in duration and rate to the catalytic closures transduced by Taq-Intervening. **B)** A 10-ms sample of data shows three catalytic closures with τ_{cat} = 120, 70, and 420 µs (blue). All figures depict raw data flattened with a 15-Hz highpass filter. Individual points in (B) represent 10 µs.

Taq-Fingers devices produced $\Delta I(t)$ signals with two distinct populations of events. The first population was identical to the Taq-Intervening catalytic closures, having the same τ_{cat} probability distribution and thermally-activated rates k_{cat}, and only occurring in polyT solutions with matched dNTPs (e.g. dATP). The second population of events, on the other hand, appeared in $\Delta I(t)$ signals collected in both matched (**Figure 4A**) and mismatched dNTPs (**Figure 4B**). This population of events, which we term "transient closures," was also distinguished by an average duration $<\tau_{transient}>$ = 18 ± 6 μs, approximately 8 times shorter than $<\tau_{cat}>$. In mismatched dNTPs (where transient closures were easiest to analyze because catalytic closures were absent), transient closures occurred at a mean rate $k_{transient}$ = 10 ± 8 s^{-1} with a single-exponential distribution of durations (**Figure 4C**). Data sets acquired at 45, 60, and 75 °C indicated no significant temperature dependence in either $k_{transient}$ or $<\tau_{transient}>$ (**Figure 2-7**). As with catalytic closures, observing transient closures required the presence of both DNA primer-template and dNTP; these events disappeared in control measurements using buffer, polyT alone, or dNTPs alone (**Figure 2-8**). Similar signals were generated by Taq-Thumb (**Figure 2-9**).

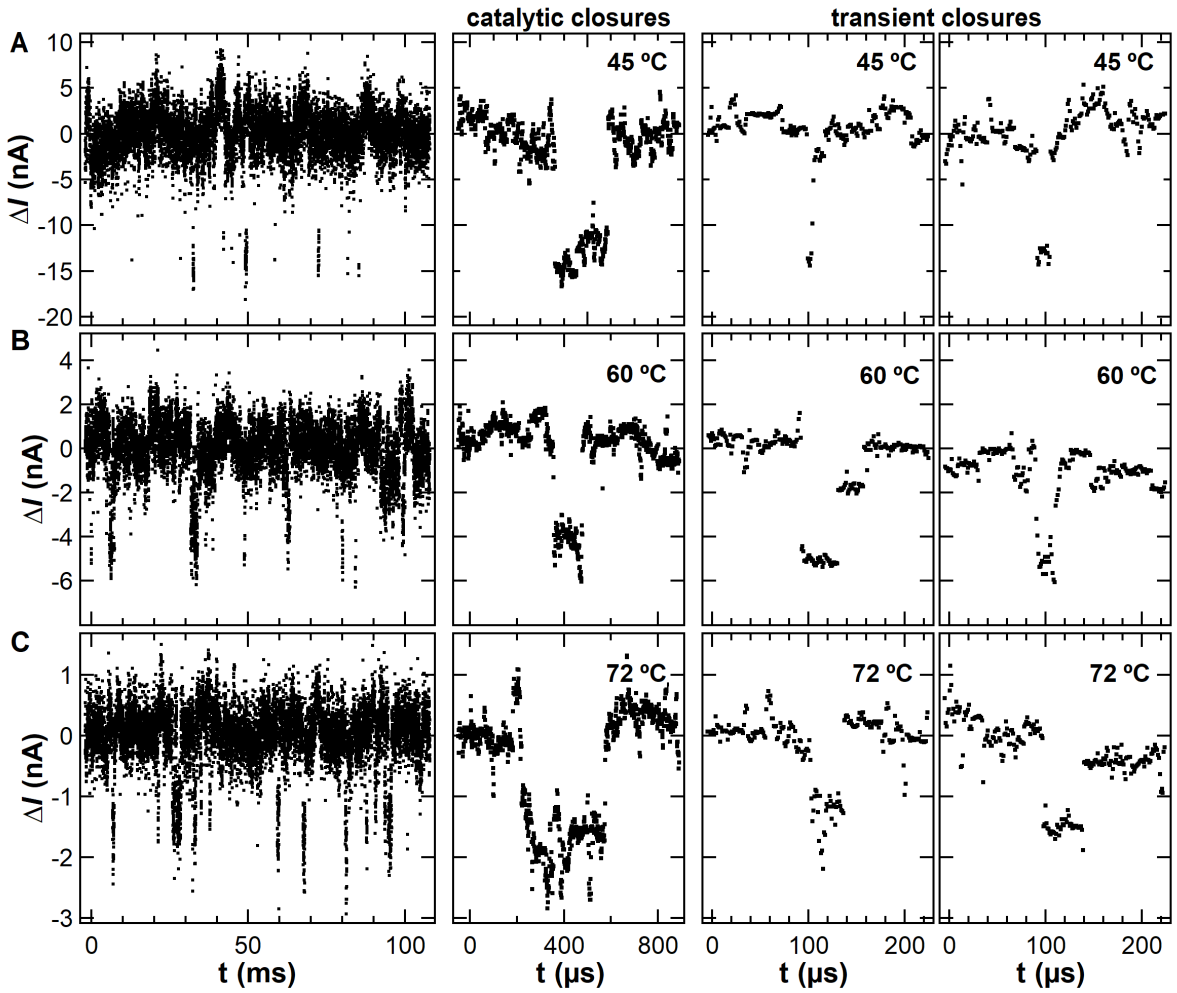


Figure 2-7. Example signals and events from Taq-Fingers in polyT + dATP. Representative Δ*I*(*t*) signals and events generated by three Taq-Fingers devices measured at **A)** 45 °C, **B)** 60 °C, and **C)** 72 °C in solutions of polyT + dATP. Data in the first column are decimated for clarity, but the columns of example catalytic closures and transient closures depict raw flattened data at 1 MHz sampling rates.

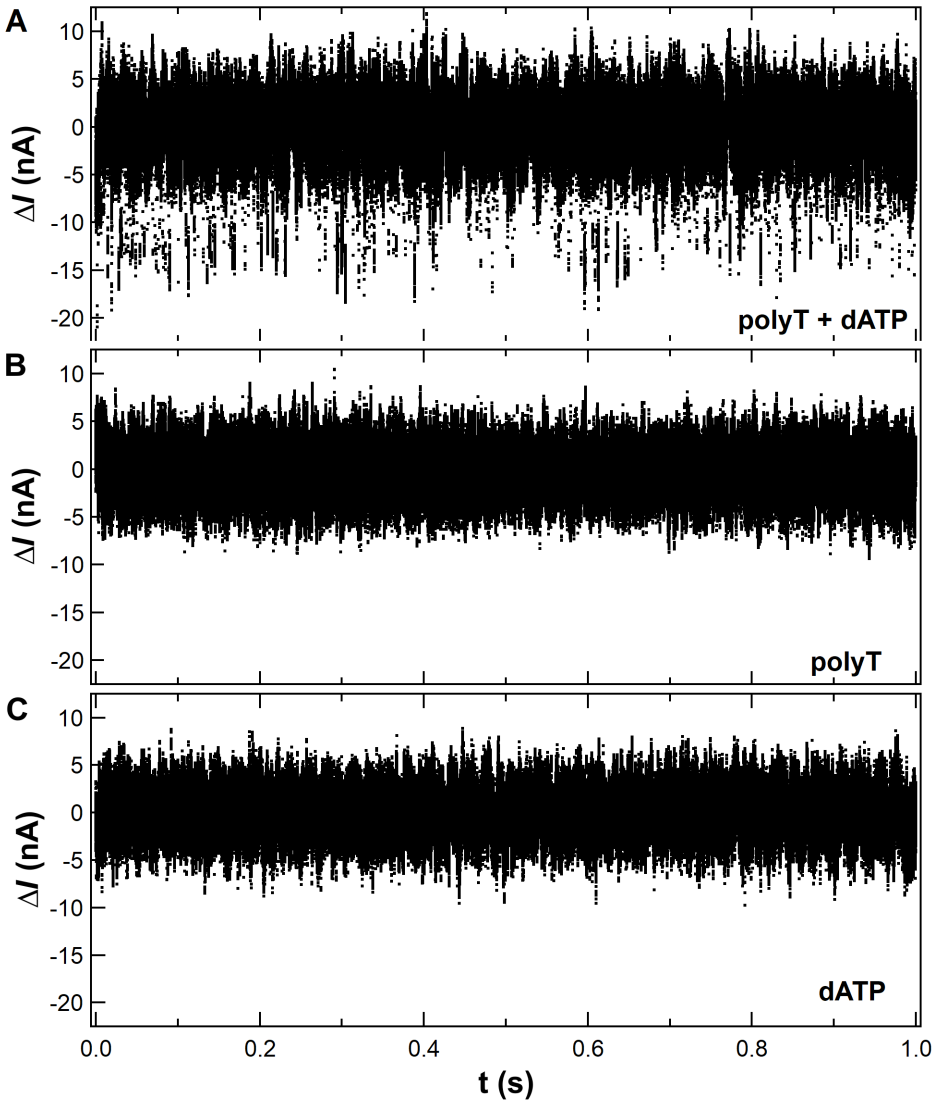


Figure 2-8. Control measurements from Taq-Fingers at 45 °C. Representative Δ*I*(*t*) signals generated by Taq-Fingers in buffer solutions containing **A)** matched dNTPs, polyT + dATP, **B)** ssDNA template only, or **C)** dNTPs only. Step-like current excursions are only observed when templates and dNTPs are both present. Figures depict raw data flattened with a 15-Hz highpass filter.

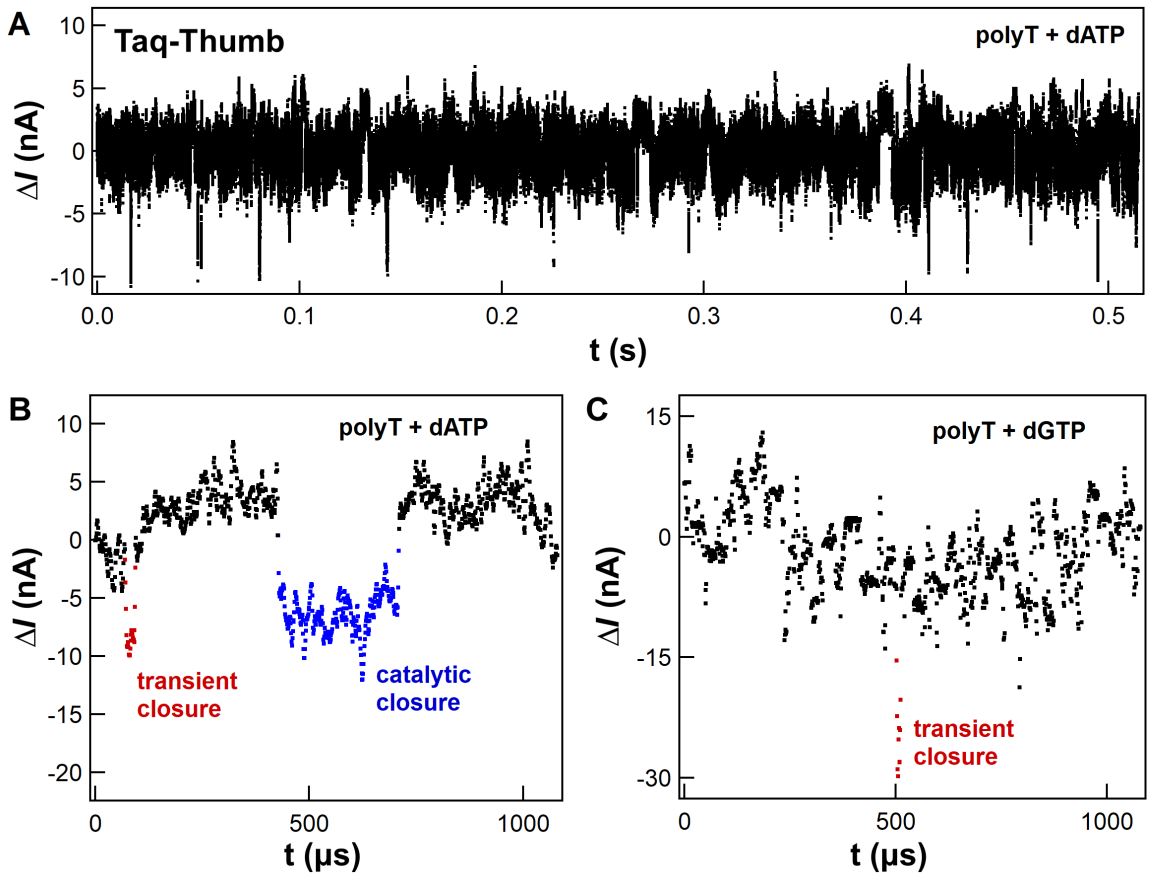


Figure 2-9. Representative signals generated by Taq-Thumb at 72 °C. A) In matched polyT + dATP, Taq-Thumb transduced a combination of transient closures and catalytic closures, similar to Taq-Fingers. **B)** A 1-ms sample of data shows a transient closure with $\tau_{transient}$ = 22 μs (red) followed by a catalytic closure τ_{cat} = 282 μs (blue). **C)** In mismatched polyT + dGTP, only transient closures were seen, like the one shown with $\tau_{transient}$ = 9 μs (red). All figures depict raw data flattened with a 15-Hz highpass filter. Individual points in (B) and (C) represent 1 μs.

In matched dNTPs, Taq-Fingers signals contained a mixture of the new transient closures along with the catalytic closures observed with Taq-Intervening. The two types of events overlapped in a bi-exponential distribution (**Figure 2-10C**), making classification of individual events impossible in the 75- to 125-μs region where the two subpopulations had comparable probabilities. Overall, though, the subpopulations were easily identified by their overlap with the single-population data sets: the transient closures in mismatched dNTPs or the Taq-Intervening catalytic closures (**Figure 2-10C inset**). The sole

difference among these comparisons was that the rate $k_{transient}$ rose from 10 s^{-1} in mismatched dNTPs up to 280 ± 150 s^{-1} in matched dNTPs. To depict this difference, **Figure 2-10C** shows the matched and mismatched population distributions in units of events per second.

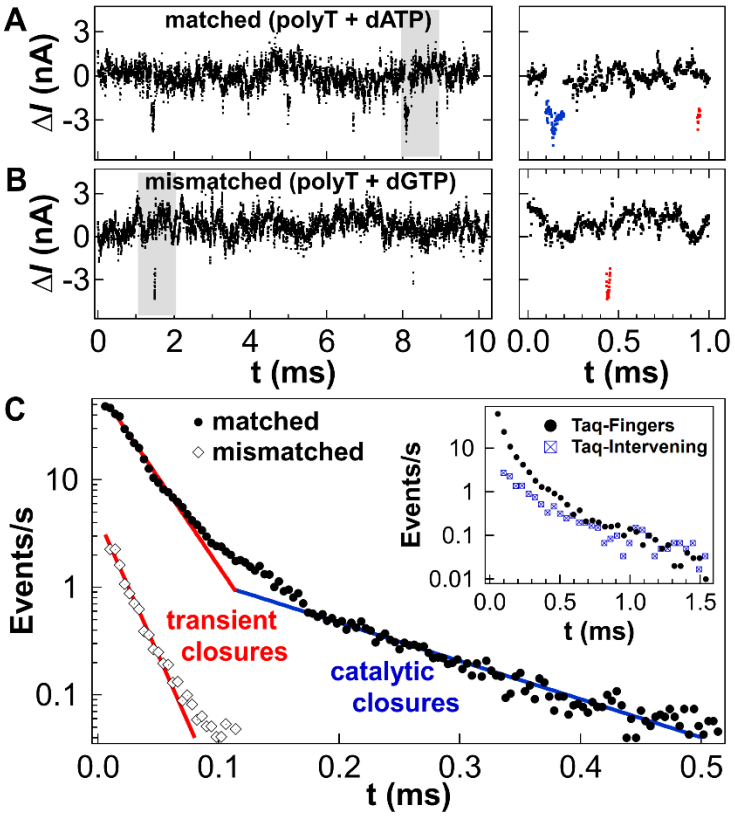


Figure 2-10. Taq-Fingers activity in matched and mismatched dNTPs at 45 °C. Taq-Fingers closures upon polyT with **A)** matched dATP or **B)** mismatched dGTP. Gray highlights the regions expanded at right with examples of individual catalytic closures (blue) and transient closures (red) at 1 μs resolution. Transient closures occurred in both matched dATP and mismatched dGTP, while catalytic closures only occurred with matched dATP. **C)** Statistical distributions of closures per second in matched dATP (circles) and mismatched dGTP (diamonds) with 4-μs bin widths, and (inset) comparison to Taq-Intervening with 20-μs bins.

As noted above, single-exponential fits to the τ_{cat} and $\tau_{transient}$ distributions disguise the actual dynamics and variability in k_{cat} and $k_{transient}$. When calculated on 1-second intervals from single-molecule trajectories, both rates varied tenfold. **Figure 2-11** shows

instantaneous $k_{transient}$ values (**Figure 2-11A**) and the moment-to-moment accumulation of those values into log-normal probability distributions, with well-separated distributions for matched and mismatched dNTPs (**Figure 2-11B**). The log-normal fitting indicates that $k_{transient}$ obeyed a thermodynamic rate equation $\ln(k_{transient}/k_0) = -\Delta G^{\ddagger}/k_B T$; where $\Delta G^{\ddagger}$ is the energy barrier for transient-closing motions, k_0 is an attempt rate, and k_B is Boltzmann's constant. Plotting $k_{transient}$ on a relative energy axis $\Delta G^{\ddagger}$ illustrates two important features. First, the distribution widths FWHM = 0.65 ± 0.05 kcal/mol equaled 1 $k_B T$, showing that thermal fluctuations in $\Delta G^{\ddagger}$ governed the $k_{transient}$ variability and exceeded any other temperature dependence (**Figure 2-12**). Second, the peak-to-peak separation of 2.1 ± 0.2 kcal/mol thermodynamically summarizes the 30-fold difference between matched and mismatched $k_{transient}$ values, and it provides a convenient point of comparison to conventional, ensemble measurements.[6,9,44,45]

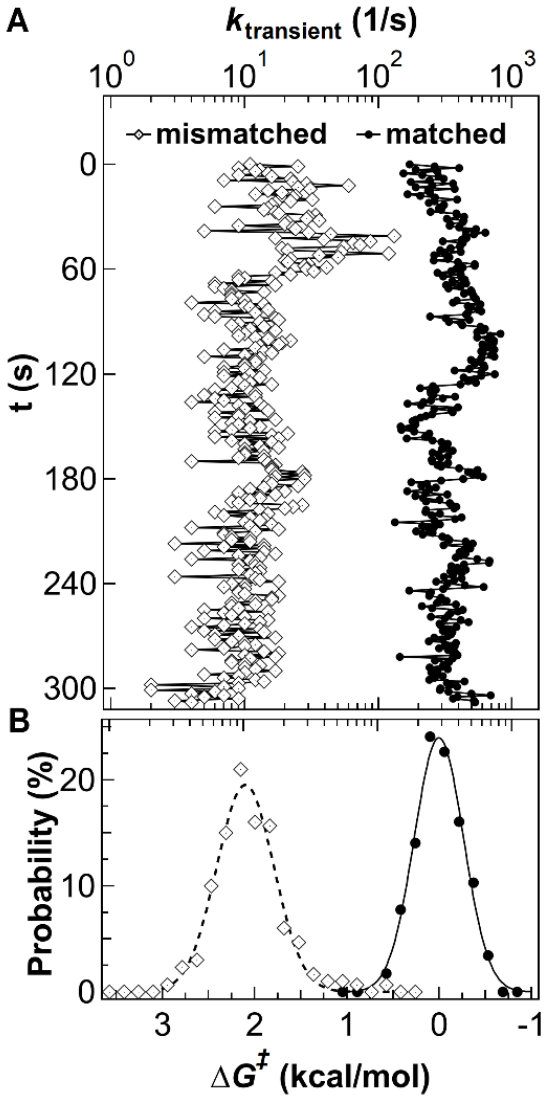


Figure 2-11. Variable kinetics of Taq-Fingers transient closures. A) Instantaneous rates calculated on 1-s intervals depict $k_{transient}$'s mean and range of variation with matched dATP (circles) or mismatched dGTP (diamonds). **B)** Cumulative probability distributions on dual horizontal axes of rate (1/s, top) and relative energy (kcal/mol, bottom), overlaid with log-normal fits.

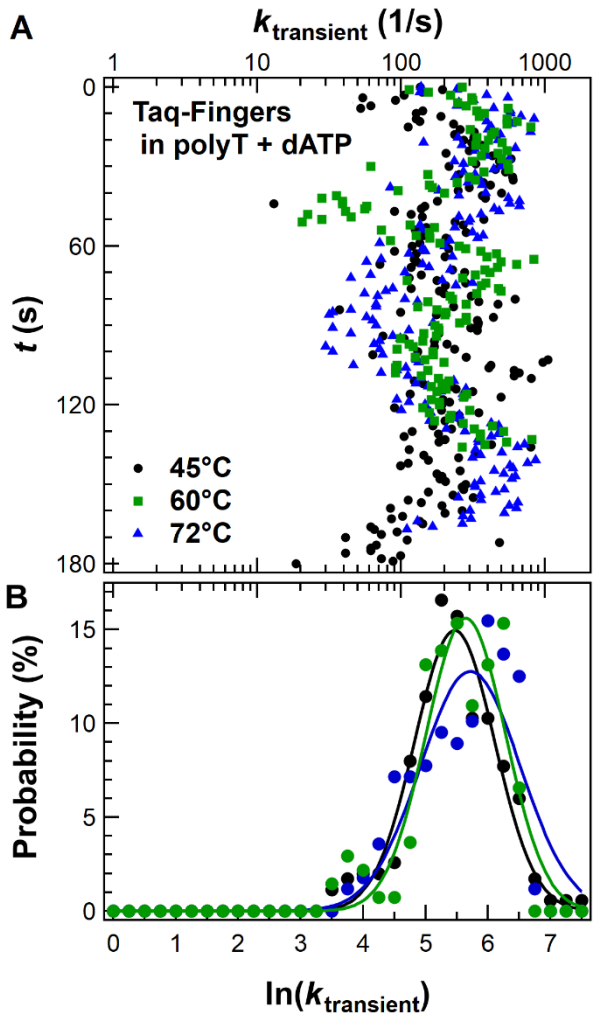


Figure 2-12. Variance of $k_{transient}$ at three temperatures. A) Instantaneous rates calculated for 1-s intervals depict $k_{transient}$'s mean and range of variation with matched dATP at 45, 60, and 72 °C. **B**) Cumulative probability distributions on dual horizontal axes of rate (1/s, top) and ln($k_{transient}$) (bottom), with the latter representing the relative energy scale ln($k_{transient}$) = -$\Delta G^{\ddagger}/k_B T$. Solid lines are log-normal fits to each distribution. The shift of the peak positions <$k_{transient}$> amounted to ≤ 0.25 $k_B T$ over the temperature range, even including contributions from molecule-to-molecule differences (i.e., static disorder).

Specifically, past ensemble research has reported identical differences between matched and mismatched dNTPs. Over two decades ago, matched dNTPs were observed to have binding affinities and rates 2 kcal/mol higher than mismatched dNTPs, leading authors to question how that difference arises in polymerases' open conformations, where the dNTP binds the fingers domain far from the templating

base.[6,9,44] Ever since, hidden fingers-closing transients have been proposed as a possible key to the fidelity testing mechanism and a widely anticipated feature of Taq's dynamics. Modern MD simulations have reproduced the 2 kcal/mol difference,[15,17,18] but only in closed or partly-closed conformations with dynamical time scales too brief for direct observation by smFRET.[13,14,46]

Despite agreement on this energy difference, the electronic technique was not directly sensing the dNTP binding events of ensemble affinity experiments. For instance, when dNTPs bound to Taq in the absence of template, no $\Delta I(t)$ excursions were produced. Instead, dNTP binding to the Taq-template complex and subsequent transient closures of the fingers domain appeared to have a one-to-one correspondence, with dNTP binding *inducing* the closures observed here. This correspondence suggests that mismatched dNTPs dissociated after a single transient closure, this efficient mismatch-rejection mechanism avoids multiple closure attempts.

In addition to sensing transient closures upon mismatched dGTP, the Taq-Fingers trajectories also resolved combinations of transient closures and catalytic closures upon matched dATP. Other than the differences already noted, no other characteristics distinguished the three populations (**Figure 2-13**). Firstly, the $\tau_{transient}$ and τ_{cat} events were uniformly intermixed in dATP without evidence of bunching or sequencing, as if both were independent, memoryless Poisson processes. Secondly, the $\tau_{transient}$ durations were the same for dATP and dGTP, making them unlikely direct measurements of the fidelity-checking step. Thirdly, the transient and catalytic closures had identical distributions of signal amplitude $\langle \Delta I \rangle$ (**Figure 2-14**), indicating that the SWNT FET was transducing the same degree of Taq motion for both. All three observations led us to conclude that $\Delta I(t)$

was not sensitive to Taq's partly-closed conformations, in which fidelity-checking is believed to occur.[12] [12]. Close inspection of Taq's sequence in the Taq-Fingers orientation suggests that one charged amino acid (D655) generated the $\Delta I(t)$ excursions, but only when Taq closed fully and not in the partially closed conformation reported by Wu and Beese.[10] Distinguishing partially closed conformations may require carefully designed mutations in future experiments, as well as higher bandwidths if the event durations are less than 1 μs.

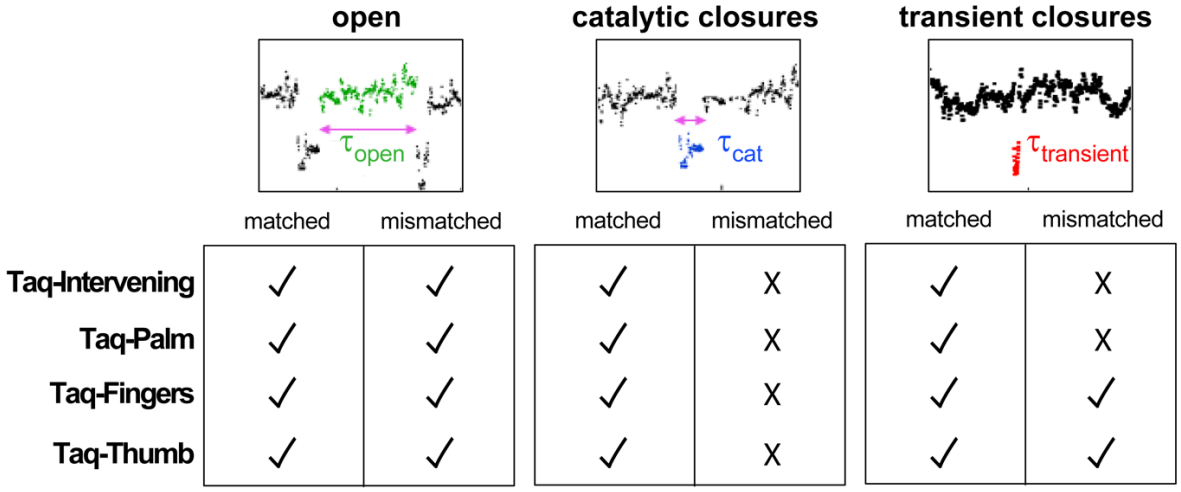


Figure 2-13. Summary of local motions recorded by SWNT FET for Taq's four different orientations in the presence of matched or mismatched dNTPs. The signals reported here capture three enzyme states – open, closed-long, closed-brief – that capture the enzyme during the open state, catalytic closure, or transient closure, respectively. The observation (✓) or absence (X) of each event state was indicated for matched dATP or mismatched dGTP.

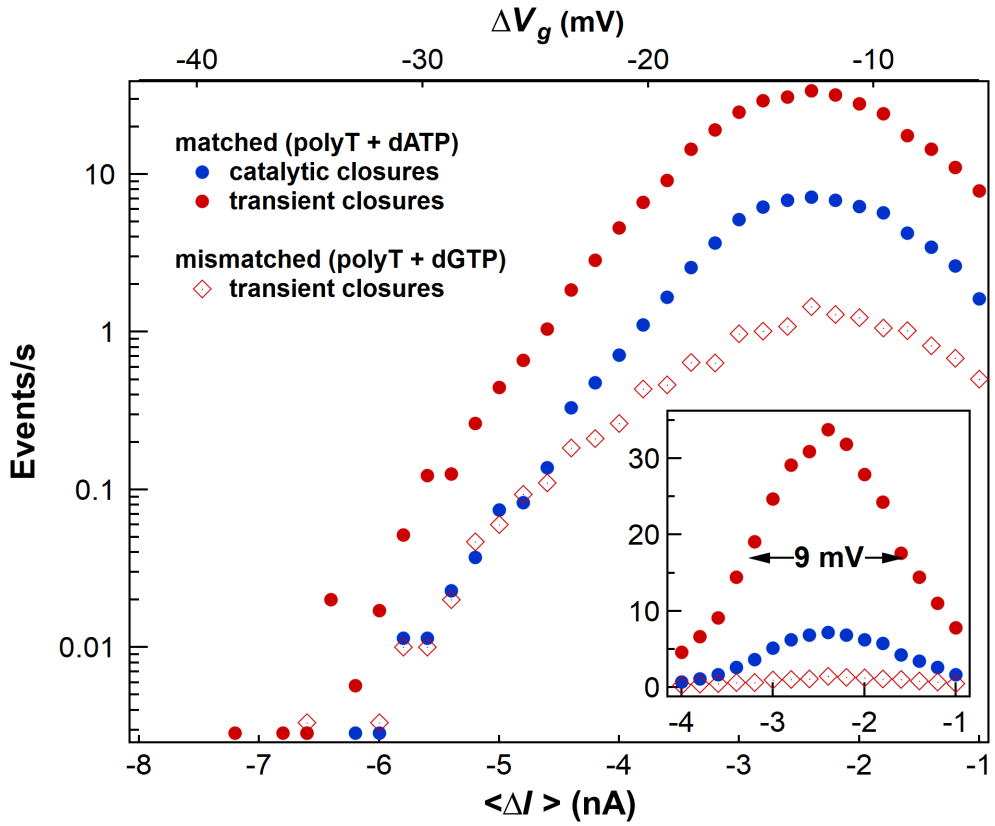


Figure 2-14. Event amplitudes generated by Taq-Fingers. The population distribution of event amplitudes <ΔI> generated by Taq-Fingers. Events were separated into transient closures (red) or catalytic closures (blue) based on event duration. All three distributions had similar widths and peak positions, suggesting that the charges responsible for electrostatic gating underwent the same degree of movement in mismatched transient closures, matched transient closures, and catalytic closures. Top x-axis provides a conversion of <Δ*I*> to its equivalent electrostatic gating ΔVg, as determined from the device transconductance dI/dVg. Inset shows the distributions on a linear scale with full-width half maximum equal to 9 mV.

Even without resolving intermediate conformations, the detailed observation of transient closures is important in two ways. First, transient closures occur with matched dNTPs. At 72 °C, there were approximately five transient closures for every catalytic closure ($k_{transient}/k_{cat}$ = 5), and no model has suggested that successful fidelity checks upon matched dNTPs should proceed to so many non-catalytic closures and rejections. At lower temperatures, the number of transient closures per cycle grew exponentially with $1/k_{cat}$. The entire τ_{open} duration filled with transient closures, suggesting that a continuous

cycle of dNTP binding, fidelity-checking, and dissociation repeated while awaiting the completion of some other thermally activated, rate-limiting step. This free-running process may require revisions to the linear, multistep diagrams used to represent Taq's catalytic cycle.[9,47]

Second, transient closures occur with mismatched dNTPs. Again, the premise of a fidelity checkpoint is to reduce misincorporations by reducing closures upon mismatched dNTPs. Here, Taq appears to have closed once upon every dNTP, whether matched or mismatched. Analysis of smFRET transition probabilities[14] and MD trajectories[17] have previously suggested that the mismatch rejection mechanism might need to follow a trajectory in which Taq briefly accesses the closed conformation after the fidelity check and on the way to re-opening and dNTP dissociation. Transient closures may be the observation of this process.

When matched dNTPs are rejected, the transient closures may reflect misalignment of the molecular orbitals required for bond formation and breakage (**Figure 2-15**). Multiple transient closures per catalytic closure might allow previewing the closed conformation and cycling back to open or partially open conformations to achieve correct template positioning, dNTP orientation, or alignment of active site and substrate functionalities, including Mg^{2+}. On the other hand, the $\tau_{transient}$ distributions for matched and mismatched dNTPs were identical, so any repositioning with its consequent change in binding energy must occur during either τ_{open} or τ_{cat}. Closures lasting less than 5 μs were not counted here, leaving many opportunities for the sub-microsecond interconversions predicted by MD simulations[16,18] Once molecular orbital alignment is correct, a catalytic closure can proceed[7] with the global conformational changes that

extend to the palm and intervening domains.[32] **Figure 2-14** depicts a likely timing of transient closures and catalytic closures among these steps of the catalytic sequence.

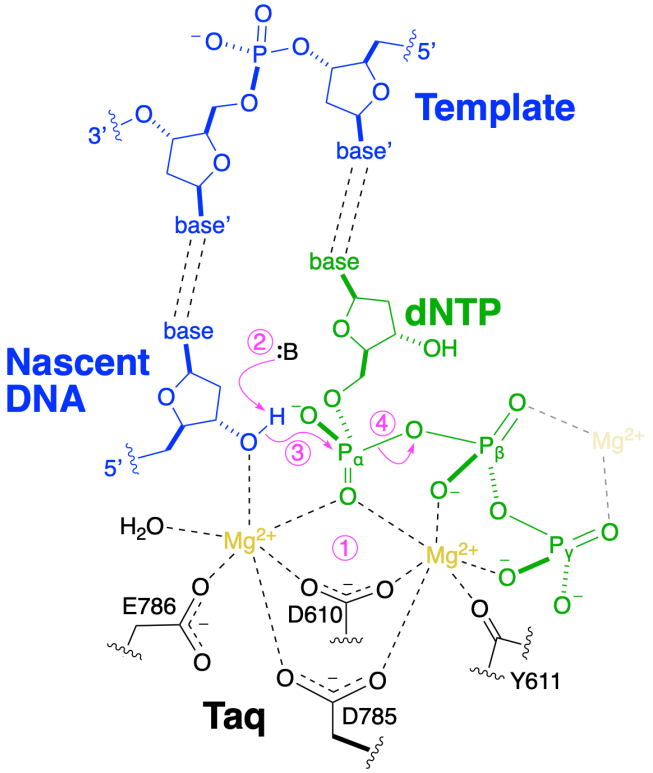


Figure 2-15. Putative role of transient closures in Taq's catalytic cycle. The data reported here suggest transient closures might be a successful process for previewing the closed conformation and cycling back to the fidelity checkpoint to achieve correct template positioning, dNTP orientation, or alignment of active site and dNTP molecular orbitals, including Mg^{2+}. This stage includes bond-formation and -breakage steps (pink arrows) like binding by Mg^{2+} (step 1), deprotonation of the 3'-hydroxyl (step 2), nucleophilic attack on Pα by the 3'-oxygen (step 3), and displacement of PP_i (step 4). Figure adapted from previous work.[7,48,49] This figure includes the templating dNTP (green), nascent DNA (blue), metal ions (yellow), and key active-site residues for matched dNTP incorporation. This schematic did not rule out the possibility of a third metal ion (transparent yellow) coordinating the templating dNTP.

2.4 Conclusion

This work opens frontiers in protein engineering and biophysics. The conceptual separation of Taq's catalytic cycle into a thermally activated step and temperature-independent motions offers a successful paradigm for the evolution of high-temperature

enzymes. Directed evolution of industrially important enzymes could benefit from exploiting this paradigm to achieve Taq's catalytic rate acceleration at elevated temperatures. Such capabilities are notoriously hard to engineer by standard experimental and computational approaches.[49,50] Furthermore, the dynamics of DNA replication is merely one case of an important but complex chemical process challenging state-of-the-art biophysical techniques. More broadly, enzyme evolution, optimization, and engineering could all benefit from the ability to track single-molecule chemical trajectories indefinitely with microsecond resolution. The added benefits of operating over a wide temperature range or with solvent compatibility make the solid-state technique of single-molecule electronics a promising approach to reveal timing, sequence, and dynamic kinetics in a wide range of systems.

2.5 Materials and Methods

Materials

Reagents purchased commercially include the following: cell lines (Stratagene), kanamycin (Carbosynth), IPTG (Carbosynth), glycine (Biobasic), and Ni-IMAC resin (Bio-Rad Laboratories). All other chemicals were supplied by Fisher Scientific or ThermoFisher Scientific. Supplies purchased commercially include: 0.22 μm top vacuum filter (Genesee Scientific), 3.5K MWCO 0.1 mL Slide-A-Lyzer mini dialysis device (ThermoFisher), and Amicon Ultra-0.5 mL centrifuge filter (Fisher Scientific). All reagents and supplies were used as received. All solutions were sterile-filtered or autoclaved before use. Kits and services purchased commercially include: Q5 Site-directed mutagenesis (New England

Biolabs), QIAprep spin miniprep kit (Qiagen), Pierce BCA assay kit (ThermoFisher), and DNA sequencing (Genewiz).

Mutagenesis of Taq

A pET28c plasmid for ligation-independent cloning containing an ORF encoding Taq fused to an N-terminal His_6 peptide epitope was mutated to abolish the polymerase's exonuclease activity (G46D) and introduce a single cysteine (R411C, E524C, R695C, or A814C) (**Table 2-2**). Q5 Site-directed mutagenesis was applied according to the manufacturer's instructions using Q5 Hot Start High-Fidelity DNA Polymerase and the appropriate oligos, **Oligo1** to **Oligo10** (**Table 2-3**). The KLD mix (5 µL) was transformed into TOP10 *E. coli* competent cells, and transformants were plated on a kanamycin-supplemented (50 µg/mL) agar plate before incubation at 37 °C overnight. Five single colonies were selected to inoculate 4 mL of 2YT media in a 15 mL culture tube supplemented with kanamycin (50 µg/mL). The seed cultures were incubated at 37 °C with shaking at 225 rpm for 8-12 h. Plasmid DNA was isolated using the QIAprep spin miniprep kit according to the manufacturer's instructions. The successful subcloning of the ORF of the Taq variants was checked via DNA sequencing using the appropriate oligos, **Oligo11** to **Oligo14** (**Table 2-3**).

Table 2-2. Oligos used for cloning of Taq variants or activity assays.

Oligo#	Sequence (5' to 3') with substitutions denoted in lowercase	Purpose
Oligo1	GGCGGTCTACgacTTCGCCAAGAGCC	To generate D46G mutation in Taq exo$^-$
Oligo2	TGCACCGGCTCCCCCCGG	
Oligo3	CCTTTCCGAGtgcCTCTTCGCCAACCTGTGGGGGAGGCTTG	To generate R411C mutation in Taq-Intervening
Oligo4	GCGGCCCGCTCCCCCGCC	
Oligo5	GGCCCTCCGCtgcGCCCACCCCATCG	To generate E524C mutation in Taq-Thumb
Oligo6	TCCAGGACGGCGGCGCTG	
Oligo7	CTTCATTGAGtgcTACTTTCAGAGCTTC	

Oligo8	GCCTGGGCCTCCTCGTAA	To generate R695C mutation in Taq-Fingers
Oligo9	GTATCCCCTGtgcGTGCCCCTGG	To generate A814C mutation in Taq-Palm
Oligo10	ACCCCCTCCATGACCTCC	
Oligo11	TAATACGACTCACTATAGGG	To sequence ends of ORF encoding Taq gene
Oligo12	GCTAGTTATTGCTCAGCGG	
Oligo13	TTGGGAAAAGTACGGCCTGA	To sequence middle of ORF encoding Taq gene
Oligo14	CTTCTCGCCAAAGACCTGAG	
Oligo15	TGTAAAACGACGGCCAGT	M13 primer
Oligo16	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTACTGG CCGTCGTTTTACA	poly(dT)$_{42}$

Table 2-3. Taq variants used in this study.

Variant	Sequence (N-terminal His_6 epitope indicated in blue and mutations in red)	Theoretical MW (g/mol)
Taq-Intervening (G46D/R411C)	MHHHHHHENLYFQGIFRAPMRGMLPLFEPKGRVLLVDGHHLAYRTFHALKGLTTSRGEPV QAVYDFAKSLLKALKEDGDAVIVVFDAKAPSFRHEAYGGYKAGRAPTPEDFPRQLALIKE LVDLLGLARLEVPGYEADDVLASLAKKAEKEGYEVRILTADKDLYQLLSDRIHVLHPEGY LITPAWLWEKYGLRPDQWADYRALTGDESDNLPGVKGIGEKTARKLLEEWGSLEALLKNL DRLKPAIREKILAHMDDLKLSWDLAKVRTDLPLEVDFAKRREPDRERLRAFLERLEFGSL LHEFGLLESPKALEEAPWPPPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKA LRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARRYGGEWTEE AGERAALSECLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVRLDVAYLRALSL EVAEEIARLEAEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTEKTGKRSTSAAVLEA LREAHPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSSDPNL QNIPVRTPLGQRIRRAFIAEEGWLLVALDYSQIELRVLAHLSGDENLIRVFQEGRDIHTE TASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHRLSQELAIPYEEAQAFIERYFQSFP KVRAWIEKTLEEGRRRGYVETLFGRRRYVPDLEARVKSVREAAERMAFNMPVQGTAADLM KLAMVKLFPRLEEMGARMLLQVHDELVLEAPKERAEAVARLAKEVMEGVYPLAVPLEVEV GIGEDWLSAKE	96305.82
Taq-Thumb (G46D/E524C)	MHHHHHHENLYFQGIFRAPMRGMLPLFEPKGRVLLVDGHHLAYRTFHALKGLTTSRGEPV QAVYDFAKSLLKALKEDGDAVIVVFDAKAPSFRHEAYGGYKAGRAPTPEDFPRQLALIKE LVDLLGLARLEVPGYEADDVLASLAKKAEKEGYEVRILTADKDLYQLLSDRIHVLHPEGY LITPAWLWEKYGLRPDQWADYRALTGDESDNLPGVKGIGEKTARKLLEEWGSLEALLKNL DRLKPAIREKILAHMDDLKLSWDLAKVRTDLPLEVDFAKRREPDRERLRAFLERLEFGSL LHEFGLLESPKALEEAPWPPPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKA LRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARRYGGEWTEE AGERAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVRLDVAYLRALSL EVAEEIARLEAEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTEKTGKRSTSAAVLEA LRCAHPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSSDPNL QNIPVRTPLGQRIRRAFIAEEGWLLVALDYSQIELRVLAHLSGDENLIRVFQEGRDIHTE TASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHRLSQELAIPYEEAQAFIERYFQSFP KVRAWIEKTLEEGRRRGYVETLFGRRRYVPDLEARVKSVREAAERMAFNMPVQGTAADLM KLAMVKLFPRLEEMGARMLLQVHDELVLEAPKERAEAVARLAKEVMEGVYPLAVPLEVEV GIGEDWLSAKE	96332.89
Taq-Fingers (G46D/R695C)	MHHHHHHENLYFQGIFRAPMRGMLPLFEPKGRVLLVDGHHLAYRTFHALKGLTTSRGEPV QAVYDFAKSLLKALKEDGDAVIVVFDAKAPSFRHEAYGGYKAGRAPTPEDFPRQLALIKE LVDLLGLARLEVPGYEADDVLASLAKKAEKEGYEVRILTADKDLYQLLSDRIHVLHPEGY LITPAWLWEKYGLRPDQWADYRALTGDESDNLPGVKGIGEKTARKLLEEWGSLEALLKNL DRLKPAIREKILAHMDDLKLSWDLAKVRTDLPLEVDFAKRREPDRERLRAFLERLEFGSL LHEFGLLESPKALEEAPWPPPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKA LRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARRYGGEWTEE AGERAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVRLDVAYLRALSL EVAEEIARLEAEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTEKTGKRSTSAAVLEA	96305.82

	LREAHPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSSDPNL QNIPVRTPLGQRIRRAFIAEEGWLLVALDYSQIELRVLAHLSGDENLIRVFQEGRDIHTE TASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHRLSQELAIPYEEAQAFIE**C**YFQSFP KVRAWIEKTLEEGRRRGYVETLFGRRRYVPDLEARVKSVREAAERMAFNMPVQGTAADLM KLAMVKLFPRLEEMGARMLLQVHDELVLEAPKERAEAVARLAKEVMEGVYPLAVPLEVEV GIGEDWLSAKE	
Taq-Palm (G46D/A814C)	MHHHHHHENLYFQGIFRAPMRGMLPLFEPKGRVLLVDGHHLAYRTFHALKGLTTSRGEPV QAVY**D**FAKSLLKALKEDGDAVIVVFDAKAPSFRHEAYGGYKAGRAPTPEDFPRQLALIKE LVDLLGLARLEVPGYEADDVLASLAKKAEKEGYEVRILTADKDLYQLLSDRIHVLHPEGY LITPAWLWEKYGLRPDQWADYRALTGDESDNLPGVKGIGEKTARKLLEEWGSLEALLKNL DRLKPAIREKILAHMDDLKLSWDLAKVRTDLPLEVDFAKRREPDRERLRAFLERLEFGSL LHEFGLLESPKALEEAPWPPPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKA LRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARRYGGEWTEE AGERAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVRLDVAYLRALSL EVAEEIARLEAEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTEKTGKRSTSAAVLEA LREAHPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSSDPNL QNIPVRTPLGQRIRRAFIAEEGWLLVALDYSQIELRVLAHLSGDENLIRVFQEGRDIHTE TASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHRLSQELAIPYEEAQAFIERYFQSFP KVRAWIEKTLEEGRRRGYVETLFGRRRYVPDLEARVKSVREAAERMAFNMPVQGTAADLM KLAMVKLFPRLEEMGARMLLQVHDELVLEAPKERAEAVARLAKEVMEGVYPL**C**VPLEVEV GIGEDWLSAKE	96390.92

Protein expression and purification

Protein expression and purification was conducted using procedures previously described (*28, 40*) with modifications as follows. The plasmid encoding each Taq variant was transformed via heat shock into competent BL21 Star (DE3) *E. coli*. A single colony was transferred to LB media (25 mL) supplemented with kanamycin (50 µg/mL), and incubated at 37 °C for 10 to 16 h. An aliquot of the starter culture (10 mL) was transferred to LB media (1 L LB in a 4 L baffled flask). After reaching an OD_{600} of ~0.4, the culture was induced through addition of IPTG (1 mM final concentration) before incubation, with shaking (225 rpm) at 37 °C for 2 h. The cells were centrifuged and the cell pellet and resuspended in lysis buffer (50 mM Tris-HCl, 25 mM KCl, 50 mM glucose, 0.25% Tween, 10 mM BME, 1 mM PMSF, pH 7.8, and protease inhibitor cocktail). Following sonication for 12 min (1 s pulse on, 3 s pulse off), DNase was added to the cell lysate, and the solutions was incubated at 37 °C for 30 min to eliminate DNA bound to the polymerase. Heating at 70 °C for 30 min deactivated the DNase. The DNase-treated lysate was

subjected to centrifugation (26,892 rcf, 45 min, 4 °C). The supernatant was incubated with charged Ni-NTA resin and bound overnight on a shaker (150 rpm at 4 °C). The resin was loaded onto a column, and subject to washes (50 mM Tris-HCl, 25 mM KCl, 50 mM glucose, 0.25% Tween, 10 mM BME, pH 7.9) with a concentration gradient of imidazole (wash buffer supplemented with 25 to 250 mM imidazole). After pooling and concentrating (by microconcentration with a 50 kDa MWCO) the eluted protein was further purified by size exclusion chromatography (Superdex 200 pg, 16/600 at a flow rate of 1.0 mL/min) in activity buffer (40 mM HEPES, 50 mM KCl, 5 mM $MgCl_2$, pH 8.5). The eluted fractions were collected and visualized by 12% SDS-PAGE (**Figure 2-16**). For single-molecule measurements on SWNT devices, the purified recombinant protein was dialyzed using 3.5K MWCO 0.1 mL Slide-A-Lyzer mini dialysis device into attachment buffer (40 mM HEPES, 50 mM KCl, 5 mM $MgCl_2$, pH 6.5). The protein concentration was determined by a bicinchoninic acid (BCA) assay using the enzyme's estimated MW (http://www.expasy.org). For storage, the purified recombinant protein was dialyzed into storage buffer (20 mM Tris-HCl, 100 mM KCl, 0.1 mM EDTA, 1 mM DTT, 0.5% Tween20, 10% Glycerol, pH 7.4) and maintained to -80 °C.

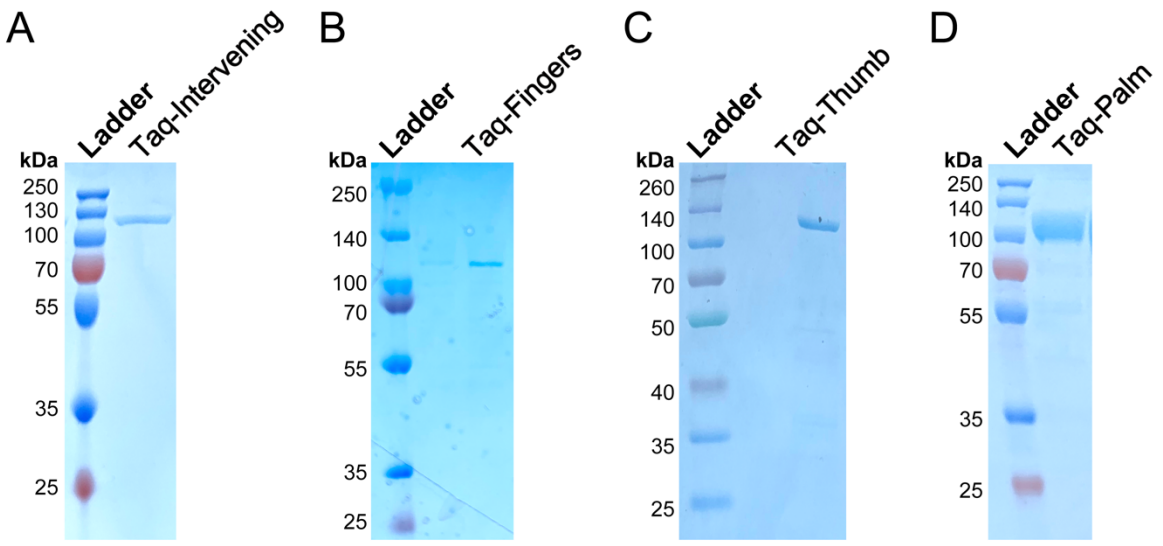


Figure 2-16. Taq variant homogeneity by 12% SDS-PAGE. These representative SDS-PAGEs visualize **A)** Taq-Intervening, **B)** Taq-Fingers, **C)** Taq-Thumb, or **D)** Taq-Palm after size exclusion chromatography. Taq variants were purified to >95% homogeneity, and migrated at their expected mass of ≈96 kDa. The PageRuler plus prestained protein ladder in panels A, B, and D or Spectra multicolor broad range protein ladder in panel C were used as reference MWs.

Ensemble activity assay

Before bioconjugation to SWNTs, Taq activity was examined using previously described ensemble activity assays,[30,50] using **Oligo15** and polyT or **Oligo16** (**Table 2-2**). The HPLC-purified oligos were solubilized in water to 100 μM and hybridized to the template through annealing from 45 to 5 °C with a 5 °C temperature decrease every 5 min. The positive control reactions contained the Taq variant (2 μM), dATP (100 μM), and DNA template-primer (5 μM) in 1x Taq Reaction Buffer (20 mM Tris-HCl, 50 mM KCl, 5 mM $MgCl_2$, pH 8.3). The negative control reaction omitted Taq. Reactions were incubated at 72 °C for 1 h, before the addition of SYBR Green Nucleic Acid I stain, and then visualized by electrophoresis using 5% high resolution agarose gel. Gels were

imaged using a Typhoon scanner at 256 nm. A representative ensemble activity assay is shown in **Figure 2-17** and **Figure 2-18**.

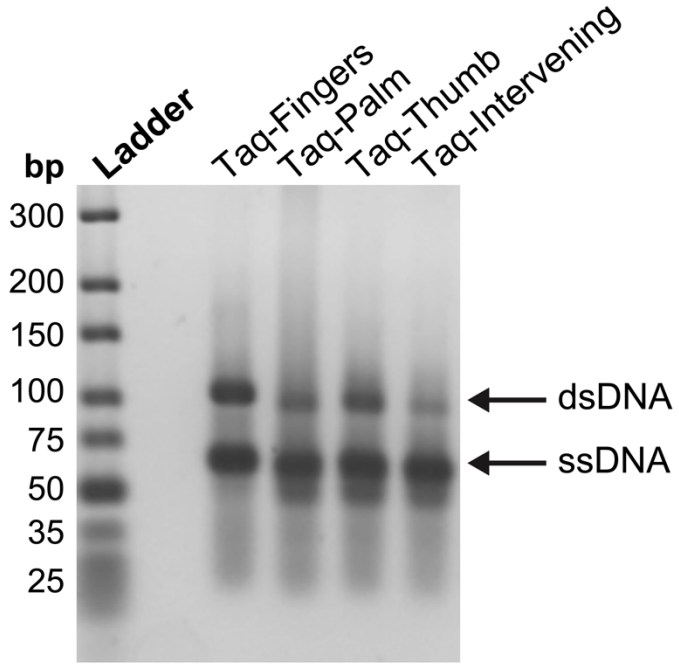


Figure 2-17. Ensemble activity assays of Taq variants. This representative 5% agarose gel visualizes the activity of the indicated Taq variant (1 μM) after incubation at 72 °C for 1 h with dATP (10 μM) and polyT template (100 nM) in Taq Reaction Buffer (20 mM Tris-HCl, 50 mM KCl, 5 mM $MgCl_2$, pH 8.3). SYBR green nucleic acid one prestained reactions prior to running on 5% agarose gels. The template was pre-annealed to the M13 primer.

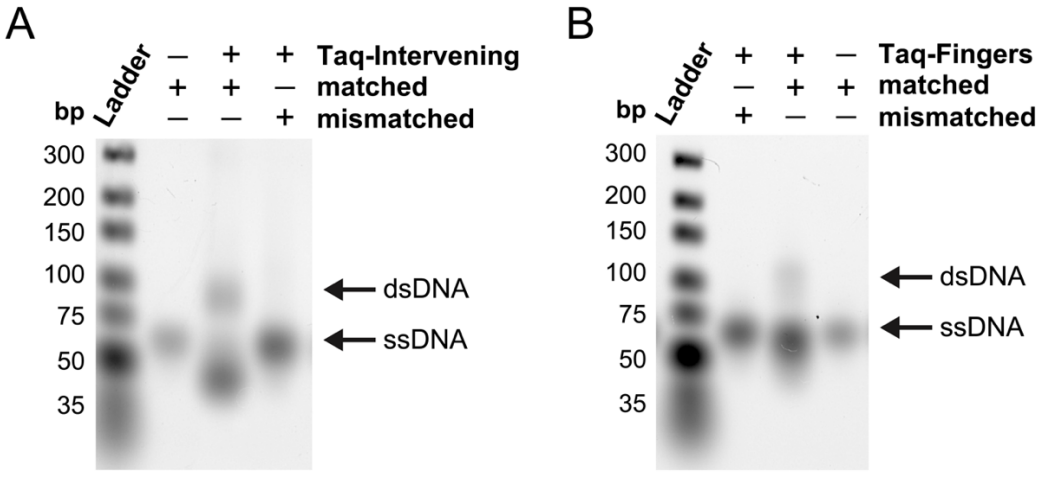


Figure 2-18. Ensemble activity assay for Taq-Intervening and Taq-Fingers in the presence of matched or mismatched dNTPs. These representative 5% agarose gels visualize the

activity of **A)** Taq-Intervening (1 μM) or **B)** Taq-Fingers (1 μM) after incubation at 72 °C for 1 h with polyT template (100 nM) in Taq Reaction Buffer (20 mM Tris-HCl, 50 mM KCl, 5 mM $MgCl_2$, pH 8.3) in the presence of matched (10 μM dATP) or mismatched (10 μM dGTP) dNTPs. SYBR green nucleic acid I prestained the reaction solutions prior to their visualization on 5% agarose gels. The template was pre-annealed to the M13 primer. Taq variants produced dsDNA only in the presence of matched dNTPs.

Fabrication and Measurement of SWNT FETs

SWNT FET devices were fabricated using techniques described previously.[29] In brief, SWNTs were grown via chemical vapor deposition (CVD) from $Fe_{30}Mo_{84}$ catalyst seeds.[51] The $Fe_{30}Mo_{84}$ catalyst nanoparticles were confined to Al_2O_3 islands photolithographically defined on a 4" p++ silicon wafer with a 250 nm thermal oxide (Silicon Quest International). A fast-heating CVD recipe facilitated growth of approximately one straight, 100 μm long SWNT per island. Immediately after growth, the wafer was baked in air (315 °C, 30 min) to remove excess amorphous carbon, annealed (95%/5% flowing Ar/H_2 at 600 °C, 60 min), and given a 15-nm protective Al_2O_3 coating by atomic layer deposition (Savannah S200, Cambridge Nanotech). Electrodes were defined by lithographically patterning a bilayer resist (S1808 on LOR-A1), selectively etching the Al_2O_3 (Transetch-N, Transene), e-beam evaporating 10 nm Pt and 40 nm Ni, and lifting off the resist (Remover PG, MicroChemicals). Devices were then passivated with PMMA, with a small opening defined over each 2-μm SWNT channel by e-beam lithography. In the preferred protocol, devices were individually etched in Transetch-N to remove the protective Al_2O_3 in this windowed region immediately before biofunctionalization.

Biofunctionalization of the newly-exposed channel proceeded in two stages. First, the device was soaked in a solution of 100 pM pyrene-maleimide diluted in 1 μM of pyrene

in ethanol for 2 min to coat the SWNT surface with a monolayer of pyrene and pyrene-maleimide linkers. Excess pyrene was then gently rinsed away with ethanol for 2 min in a drop-by-drop fashion. Next, the device was submerged in flowing attachment buffer (40 mM HEPES, 50 mM KCl, 5 mM $MgCl_2$, pH 6.5) for 10 min to equilibrate the surface. For bioconjugation, the device was incubated in a solution of the same buffer with 4 nM Taq for 5 min, followed by 10 min of continuous rinsing with buffer to remove nonspecific protein. This protocol was optimized to yield approximately 1 Taq attachment per 2 μm of SWNT, which in our geometry equaled one active, single-molecule SWNT FET device for every pair of devices prepared (**Figure 2-19**).

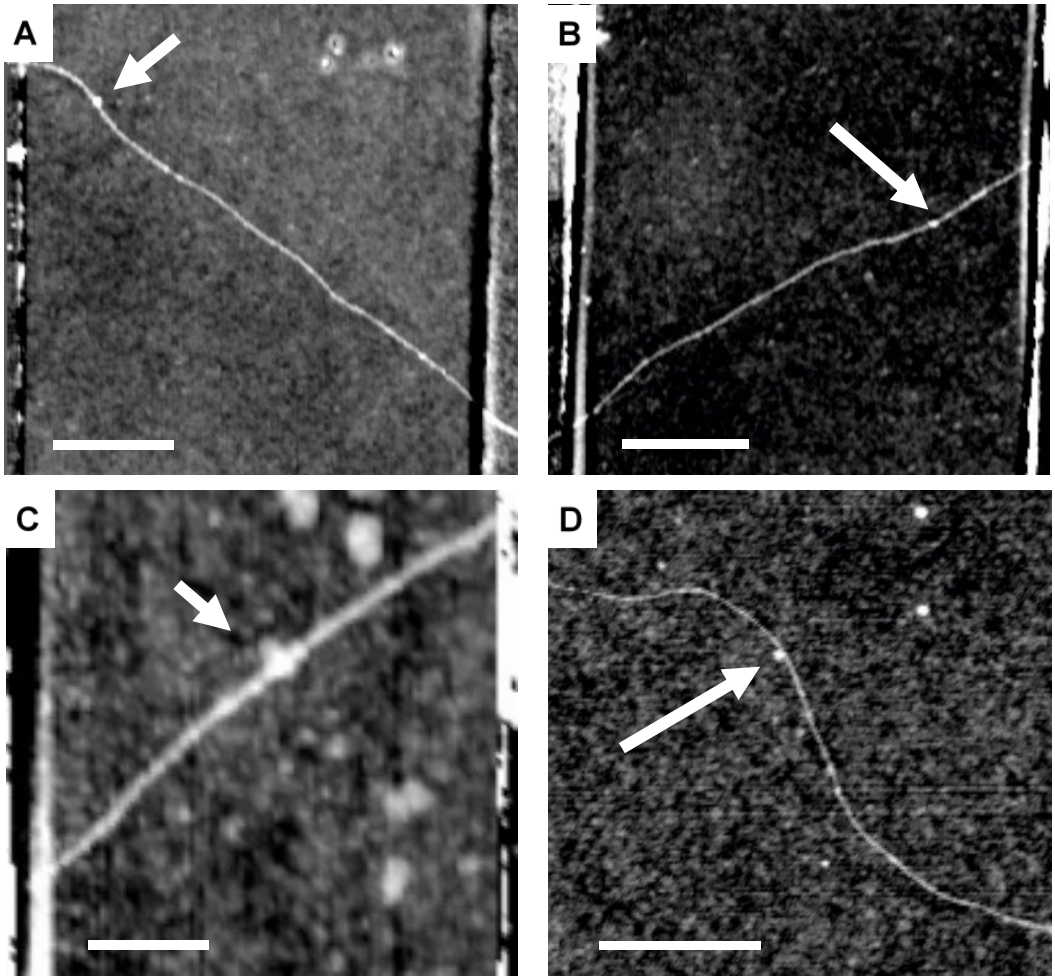


Figure 2-19. Additional AFM topography images of sample devices with single-molecule Taq. After electrical measurements, devices were rinsed, dried, and imaged in air with

tapping mode AFM. Vertical lines along the sides of images **A)**, **B)**, and **C)** mark the edges of the Pt source and drain lithography connecting each SWNT. The edges of the PMMA/Al_2O_3 passivation layer are outside of the range of each image; AFM imaging was confined to the windows where this passivation had been removed. Apparent topographic heights of SWNTs measured 1.2 to 1.7 nm; attached Taq molecules appeared as 0.8 to 1.8 nm protrusions above the SWNT height. **D)** A wider view of the image in **Figure 1A**. All scale bars measure 500 nm.

Before measurements, devices were mounted in a holder, wirebonded, and compressed with a PDMS gasket. To achieve a proper seal between the PDMS and PMMA surface of the device, the PDMS gasket was sonicated in 80% ethanol + 20% isopropanol for 10 min before being clamped in place. A 100-μm wide microfluidic channel in the PDMS confined the liquid solutions to flow over the exposed SWNT FET device. The potential of the liquid was controlled with Pt counter electrodes and Ag/AgCl reference electrodes mounted in the fluid inlets and outlets.

During electrical measurements, electrodes were controlled and monitored using National Instruments data acquisition hardware (PCI-6281) and LabVIEW 2018 software. A transimpedance current preamplifier with 10^7 V/A gain (DHPCA-100 with 0.55 μs rise time, or DLPCA-200 with 2.5 μs rise time, FEMTO) was used to bias the drain electrode (20 to 100 mV) and measure the source-drain current $I(t)$. The Pt counter electrode was set to electrostatically gate the SWNT near the middle of its linear operating range (typically -200 mV with respect to Pt), and then signals were continuously digitized and stored at 2 MHz.

Statistical analysis

To isolate and analyze enzyme-induced $I(t)$ signals, a denoising filter was implemented with the LabVIEW 2018 wavelet analysis library. Two parallel, non-adaptive

bandpass filters used Haar wavelets with garrote thresholding from 66 ms to 256 μs for a low-noise output channel or from 66 ms to 8 μs for a sharper-edged, high-frequency output channel. After denoising, an adaptive thresholding filter was applied to both channels to identify individual events, excluding fluctuations with durations ≤ 4 μs to avoid counting external noise sources.

After identification, each event was individually characterized. In addition to the event duration and mean event amplitude reported here, tabulated parameters included total variation, kurtosis, and frequency components calculated from both the raw signal and the denoised outputs. No features distinguished closed-long and closed-brief events except for their durations and, in fact, the duration distribution was smooth and continuous when closed-long and closed-brief events were both present in a data set. Categorization into closed-long and closed-brief subsets was therefore guided by bi-exponential fitting and comparison against single-exponential distributions, as shown in **Figure 2-4C**. For quantitative statistical analysis, non-overlapping subsets were artificially divided into events lasting < 75 μs (brief events) or > 125 μs (long events), ignoring the indeterminate events in between. This definition of subsets introduced systematic undercounting of event frequencies. It also limited our pursuit of quantitative comparisons among the different Taq mutants.

We note that the 4 μs noise cutoff caused the analysis to ignore very short signals. The cutoff truncated no less than 20% of events in a Poisson process having a mean duration $<\tau_{transient}>$ = 18 μs (but only 3% of events with $<\tau_{cat}>$ = 150 μs). All $k_{transient}$ and k_{cat} rates in the main text are based on events enumerated by the analysis, without any adjustment to account for this systematic undercounting.

2.6 References

(1) Maga, G. DNA Polymerases. In *Reference Module in Biomedical Sciences*; Elsevier Inc.: Amsterdam, Netherlands, 2019; pp 376–378.

(2) Aschenbrenner, J.; Marx, A. DNA Polymerases and Biotechnological Applications. *Curr. Opin. Biotechnol.* **2017**, *48*, 187–195.

(3) Saiki, R. K.; Gelfand, D. H.; Stoffel, S. Primer-Directed Enzymatic Amplification of DNA with a Thermostable DNA Polymerase. *Science* **1988**, *239*, 487–491.

(4) 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.* **1998**, *17*, 7514–7525.

(5) Eom, S. H.; Wang, J.; Steitz, T. A. Structure of Taq Polymerase with DNA at the Polymerase Active Site. *Nature* **1996**, *382*, 278–281.

(6) 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* **2005**, *19*, 345–355.

(7) 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* **2012**, *13*, 635–639.

(8) Rothwell, P. J.; Allen, W. J.; Sisamakis, E.; Kalinin, S.; Felekyan, S.; Widengren, J.; Waksman, G.; Seidel, C. A. M. DNTP-Dependent Conformational Transitions in the Fingers Subdomain of Klentaq1 DNA Polymerase: Insights into the Role of the "Nucleotide- Binding" State. *J. Biol. Chem.* **2013**, *288*, 13575–13591.

(9) Joyce, C. M.; Benkovic, S. J. DNA Polymerase Fidelity: Kinetics, Structure, and Checkpoints. *Biochemistry* **2004**, *43*, 14317–14324.

(10) 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.* **2011**, *286*, 19758–19767.

(11) Santoso, Y.; Joyce, C. M.; Potapova, O.; Le Reste, L.; Hohlbein, J.; Torella, J. P.; Grindley, N. D. F.; Kapanidis, A. N. Conformational Transitions in DNA Polymerase I Revealed by Single-Molecule FRET. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 715–720.

(12) 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.* **2012**, *134*, 11261–11268.

(13) 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.* **2015**, *43*, 5998–6008.

(14) Hohlbein, J.; Kapanidis, A. N. Probing the Conformational Landscape of DNA Polymerases Using Diffusion-Based Single-Molecule FRET. In *Methods in Enzymology*; Elsevier Inc.: Amsterdam, Netherlands, 2016; Vol. 581, pp 353–378.

(15) Miller, B. R.; Parish, C. A.; Wu, E. Y. Molecular Dynamics Study of the Opening Mechanism for DNA Polymerase I. *PLoS Comput. Biol.* **2014**, *10*.

(16) Miller, B. R.; Beese, L. S.; Parish, C. A.; Wu, E. Y. The Closing Mechanism of DNA Polymerase i at Atomic Resolution. *Structure* **2015**, *23*, 1609–1620.

(17) Meli, M.; Sustarsic, M.; Craggs, T. D.; Kapanidis, A. N.; Colombo, G. DNA Polymerase Conformational Dynamics and the Role of Fidelity-Conferring Residues: Insights from Computational Simulations. *Front. Mol. Biosci.* **2016**, *3*, 1–16.

(18) Yeager, A. V.; Humphries, K.; Farmer, E.; Cline, G.; Miller, B.; Yeager, A.; Miller Iii, B. R. Investigation of Nascent Base Pair and Polymerase Behavior in the Presence of Mismatches in DNA Polymerase I Using Molecular Dynamics. *J. Chem. Inf. Model.* **2017**, *58*, 338–349.

(19) Radhakrishnan, R.; Arora, K.; Wang, Y.; Beard, W. A.; Wilson, S. H.; Schlick, T. Regulation of DNA Repair Fidelity by Molecular Checkpoints: “Gates” in DNA Polymerase β’s Substrate Selection. *Biochemistry* **2006**, *45*, 15142–15156.

(20) Prasad, B. R.; Kamerlin, S. C. L.; Florián, J.; Warshel, A. Prechemistry Barriers and Checkpoints Do Not Contribute to Fidelity and Catalysis as Long as They Are Not Rate Limiting. *Theor. Chem. Acc.* **2012**, *131*, 1–15.

(21) Wong, I.; Patel, S. S.; Johnson, K. A. An Induced-Fit Kinetic Mechanism for DNA Replication Fidelity: Direct Measurement by Single-Turnover Kinetics. *Biochemistry* **1991**, *30*, 526–537.

(22) Datta, K.; Wowor, A. J.; Richard, A. J.; Licata, V. J. Temperature Dependence and Thermodynamics of Klenow Polymerase Binding to Primed-Template DNA. *Biophys. J.* **2006**, *90*, 1739–1751.

(23) Kirmizialtin, S.; Nguyen, V.; Johnson, K. A.; Elber, R. How Conformational Dynamics of DNA Dolymerase Select Correct Substrates: Experiments and

Simulations. *Structure* **2012**, *20*, 618–627.

(24) Prasad, B. R.; Warshel, A. Prechemistry Versus Preorganization in DNA Replication Fidelity. *Proteins* **2011**, *79*, 2900–2919.

(25) Changeux, J. P.; Edelstein, S. Conformational Selection or Induced Fit? 50 Years of Debate Resolved. *F1000 Biol. Rep.* **2011**, *3*, 1–15.

(26) Mulholland, A. J.; Roitberg, A. E.; Tuñón, I. Enzyme Dynamics and Catalysis in the Mechanism of DNA Polymerase. *Theor. Chem. Acc.* **2012**, *131*, 1–4.

(27) Choi, Y.; Moody, I. S.; Sims, P. C.; Hunt, S. R.; Corso, B. L.; Perez, I.; Weiss, G. A.; Collins, P. G. Single-Molecule Lysozyme Dynamics Monitored by an Electronic Circuit. *Science* **2012**, *335*, 319–324.

(28) Sims, P. C.; Moody, I. S.; Choi, Y.; Dong, C.; Corso, B. L.; Gul, O. T.; Collins, P. G.; Weiss, G. A. Electronic Measurements of Single-Molecule Catalysis by CAMP-Dependent Protein Kinase A Electronic Measurements of Single-Molecule Catalysis by CAMP-Dependent Protein Kinase A. *J. Am. Chem. Soc.* **2013**, *135*, 7861–7868.

(29) Olsen, T. J.; Choi, Y.; Sims, P. C.; Gul, O. T.; Corso, B. L.; Dong, C.; Brown, W. A.; Collins, P. G.; Weiss, G. A. Electronic Measurements of Single-Molecule Processing by DNA Polymerase I (Klenow Fragment). *J. Am. Chem. Soc.* **2013**, *135*, 7855–7860.

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

(31) Gül, O. T.; Pugliese, K. M.; Choi, Y.; Sims, P. C.; Pan, D.; Rajapakse, A. J.; Weiss, G. A.; Collins, P. G. Single Molecule Bioelectronics and Their Application to Amplification-Free Measurement of DNA Lengths. *Biosensors* **2016**, *6*, 1–19.

(32) Xu, C.; Maxwell, B. A.; Suo, Z. Conformational Dynamics of Thermus Aquaticus DNA Polymerase I during Catalysis. *J Mol Biol.* **2014**, *426*, 2901–2917.

(33) Soo Hyun Eom; Wang, J.; Steitz, T. A. Structure of Taq Polymerase with DNA at the Polymerase Active Site. *Nature* **1996**, *382*, 278–281.

(34) Choi, Y.; Olsen, T. J.; Sims, P. C.; Moody, I. S.; Corso, B. L.; Dang, M. N.; Weiss, G. A.; Collins, P. G. Dissecting Single-Molecle Signal Transduction in Carbon Nanotube Circuits with Protein Engineering. *Nano Lett.* **2013**, *13*, 625–631.

(35) Gelfand, D. H. Taq DNA Polymerase. In *PCR Technology*; Erlich H.A., Ed.;

Stockton Press: Palgrave Macmillan, London, 1989; pp 17–22.

(36) Chien, A.; Edgar, D. B.; Trela, J. M. Deoxyribonucleic Acid Polymerase from the Extreme Thermophile Thermus Aquaticus. *J. Bacteriol.* **1976**, *127*, 1550–1557.

(37) Innis, M. A.; Myambo, K. B.; Gelfand, D. H.; Brow, M. A. D. DNA Sequencing with Thermus Aquaticus DNA Polymerase and Direct Sequencing of Polymerase Chain Reaction-Amplified DNA. *Proc. Natl. Acad. Sci.* **1988**, *85*, 9436–9440.

(38) Langer, A.; Schräml, M.; Strasser, R.; Daub, H.; Myers, T.; Heindl, D.; Rant, U. Polymerase/DNA Interactions and Enzymatic Activity: Multi-Parameter Analysis with Electro-Switchable Biosurfaces. *Sci. Rep.* **2015**, *5*, 1–15.

(39) Kong, H.; Kucera, R. B.; Jack, W. E. Characterization of a DNA Polymerase from the Hyperthermophile Archaea Thermococcus Litoralis. Vent DNA Polymerase, Steady State Kinetics, Thermal Stability, Processivity, Strand Displacement, and Exonuclease Activities. *J. Biol. Chem.* **1993**, *268*, 1965–1975.

(40) Akhterov, M. V.; Choi, Y.; Olsen, T. J.; Sims, P. C.; Iftikhar, M.; Gul, O. T.; Corso, B. L.; Weiss, G. A.; Collins, P. G. Observing Lysozymes Closing and Opening Motions by High-Resolution Single-Molecule Enzymology. *ACS Chem. Biol.* **2015**, *10*, 1495–1501.

(41) English, B. P.; Min, W.; Van Oijen, A. M.; Kang, T. L.; Luo, G.; Sun, H.; Cherayil, B. J.; Kou, S. C.; Xie, X. S. Ever-Fluctuating Single Enzyme Molecules: Michaelis-Menten Equation Revisited. *Nat. Chem. Biol.* **2006**, *2*, 87–94.

(42) Dan, N. Understanding Dynamic Disorder Fluctuations in Single-Molecule Enzymatic Reactions. *Curr. Opin. Colloid Interface Sci.* **2007**, *12*, 314–321.

(43) Costescu, B. I.; Sturm, S.; Gräter, F. Dynamic Disorder Can Explain Non-Exponential Kinetics of Fast Protein Mechanical Unfolding. *J. Struct. Biol.* **2017**, *197*, 43–49.

(44) Moran, S.; Ren, R. X. F.; Kool, E. T. A Thymidine Triphosphate Shape Analog Lacking Watson-Crick Pairing Ability Is Replicated with High Sequence Selectivity. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 10506–10511.

(45) Kunkel, T. A. DNA Replication Fidelity. *J. Biol. Chem.* **2004**, *279*, 16895–16898.

(46) Juette, M. F.; Terry, D. S.; Wasserman, M. R.; Altman, R. B.; Zhou, Z.; Zhao, H.; Blanchard, S. C. Single-Molecule Imaging of Non-Equilibrium Molecular Ensembles on the Millisecond Timescale. *Nat. Methods* **2016**, *13*, 341–344.

(47) Joyce, C. M.; Potapova, O.; DeLucia, A. M.; Huang, X.; Basu, V. P.; Grindley, N. D. F. Fingers-Closing and Other Rapid Conformational Changes in DNA Polymerase

I (Klenow Fragment) and Their Role in Nucleotide Selectivity. *Biochemistry* **2008**, *47*, 6103–6116.

(48) Raper, A. T.; Reed, A. J.; Suo, Z. Kinetic Mechanism of DNA Polymerases: Contributions of Conformational Dynamics and a Third Divalent Metal Ion. *Chem. Rev.* **2018**, *118*, 6000–6025.

(49) Kropp, H. M.; Dürr, S. L.; Peter, C.; Diederichs, K.; Marx, A. Snapshots of a Modified Nucleotide Moving through the Confines of a DNA Polymerase. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 9992–9997.

(50) Richardson, M. B.; Gabriel, K. N.; Garcia, J.; Ashby, S.; Dyer, R.; Kim, J.; Lau, C.; Hong, J.; Le Tourneau, R. J.; Sen, S.; Narel, D.; Katz, B. B.; Ziller, J. W.; Majumdar, S.; Collins, P. G.; Weiss, G. A. Pyrocinchonimides Conjugate to Amine Groups on Proteins via Imide Transfer. *Bioconjug. Chem.* **2020**, *31*, 1449–1462.

(51) An, L.; Owens, J. M.; McNeil, L. E.; Liu, J. Synthesis of Nearly Uniform Single-Walled Carbon Nanotubes Using Identical Metal-Containing Molecular Nanoclusters as Catalysts. *J. Am. Chem. Soc.* **2002**, *124*, 13688–13689.

CHAPTER 3: Dynamics of Single-Molecule Φ29 DNA Polymerase Observed by Carbon Nanotube Transistors

3.1 Abstract

DNA polymerases catalyze processes essential to life, including DNA replication and repair. In the *Bacillus subtilis* bacteriophage phi29 replication system, the DNA polymerase (Φ29) has high processivity, polymerizing DNA templates greater than 70 kilobase pairs. Here, individual exonuclease-deficient Φ29 molecules were tethered by a pyrene maleimide linker to single-walled carbon nanotube field-effect transistors (nanocircuits) to examine DNA polymerase variability. Φ29-nanocricuits produced template- and dNTP- dependent signals, resulting from catalytic activity, that revealed the highly variable nature of the enzyme's catalytic rate. Homopolymeric DNA templates with 42 unpaired cytosines (**Poly C**), thymidines (**Poly T)**, guanosines (**Poly G**), or adenosines (**Poly A**) showed that the enzyme's catalytic rate is template-dependent. We interpret the enzyme's difficulty to process some homopolymeric templates as a result of DNA secondary structure (e.g., G-quadruplexes in **Poly G** and rigidity of **Poly A**). We conclude that sequence sensitivity can contribute an important functional role during bacteriophage DNA and viral replication.

3.2 Introduction

DNA polymerases catalyze processes essential to life, including DNA replication and repair. In the *Bacillus subtilis* (*B. subtilis*) bacteriophage phi29 replication system, the DNA polymerase (Φ29) directs the correct incoming deoxnucleotide triphosphate (dNTP) into a deep "palm"-like active site through "fingers" moving towards the primer-template DNA held in place by the enzyme's "thumb".[1] The Φ29 enzyme is known for its high processivity, polymerizing DNA templates greater than 70 kilobase pairs.[2] Such activities occur while the enzyme maintains an average catalytic speed of 50 dNTPs/sec[3] with an error-rate of 1×10^{-6} errors/dNTP.[4] Φ29 also engages in multiple associated activities during DNA polymerization, including 3'-5' exonucleolysis, pyrophosphorolysis, and terminal protein deoxynucleotidylation.[5] Such functions impact the enzyme's catalytic properties including its processivity, speed, and fidelity.[6]

Physical and chemical barriers can adversely affect catalysis by DNA polymerases. DNA hairpins or G-quadruplexes in the DNA template, for example, can block polymerization catalyzed by the enzyme.[7–9] To overcome this physical barrier, additives like betaine (N,N,N-trimethyl glycine) can reduce the formation of DNA secondary structure in vitro.[10] The unbalance of dNTPs poses another physical barrier; in fact, depletion of dNTPs lowers polymerization rates[8] and base substitution fidelity,[9] adversely affecting mutation rates in vivo.[11] Furthermore, the enzyme's stability and consequent catalytic activity is dependent on non-covalent driving forces that are strongly influenced by the chemical environment, including pH, salt concentration, divalent metal ion, and temperature. For example, a low pH can be rate limiting for the chemical step, but a high pH can be rate limiting for the conformational change by the enzyme following

formation of the new phophodiester bond, termed “the chemisty step.”[12] The required salt concentration[13] and divalent metal sensitivity can also determine catalytic rates; for example, a two-fold reduction in catalytic activity is observed when Ca^{2+} replaces the preferred co-factor, Mg^{2+}, for the archaeal *Pyrococcus abyssi* family B DNA Polymerase).[14] Despite such evident sensitivity to environmental conditions, conventional studies at the ensemble-level fail to capture hidden dynamics of DNA polymerase variablity and the relationship between chemical environments and their impact on the enzyme’s mechanism.[15]

Recent single-molecule studies describe Φ29’s complex pausing behavior during catalysis. Experiments using optical tweezers detected pausing during the DNA unwinding activity[16,17] and showed the pause event frequency (pauses/second) increases linearly with the G/C density of the DNA template.[18] Other studies with Φ29 monitored by nanoscale pores identified short and long pauses during catalysis.[19–21] Previous work also demonstrated the polymerase pauses longer on average to incorporate deoxythymidine triphosphate when it encounters a methylated versus unmethylated adenosine in the DNA template.[22] However, these methods lack the ability to distinguish many pause events – short fluctuations are hidden at the time resolutions of 0.4-0.5 ms using optical tweezers[16,18] and 0.2-0.3 ms using the nanopore-based technique.[19–22]

The single-molecule technique applied here relies on voltage-gating of a single-walled carbon nanotube field-effect transistor, termed a nanocircuit. Electronic signals are generated by charged amino acids on the enzyme’s surface that move near the nanotube during catalysis. The single enzyme is bioconjugated to the nanocircuit using the thiol-maleimide reaction between a single cysteine from the enzyme and a pyrene-maleimide

linker, which non-covalently π-stacks on the carbon nanotube. This setup can detect conformational transitions during catalysis by all enzymes with moving charged residues on their surfaces (i.e., all proteins), including the Klenow Fragment of DNA polymerase I from *E. coli* (KF).[23,24] The method provides sensitive detection of enzyme conformation at ultrafast time scales (e.g., 40 μs resolution reported here) and over long durations; a two week period is our longest observation of an individual KF molecule. Through its fast time resolution and long duration observations, the approach can capture hidden dynamics in catalytic activity, which may have been previously unobserved by other ensemble and single-molecule techniques.

In this study, we examine fluctuations in the catalytic activity of Φ29, and shed light on kinetics of DNA polymerization hidden from ensemble studies (**Figure 3-1A**). These devices, termed Φ29-nanocircuits can detect transient non-functional conformations that are hidden in ensemble studies (**Figure 3-1B**). The highly variable nature of the enzyme's catalytic rate is template-dependent, with enzyme activity on **Poly A** or **Poly G** templates having substantial longer pauses than activity on **Poly T** or **Poly C** templates. The effect of repetitive DNA sequence context on the average extension rate and pause kinetics suggests the enzyme has difficulty overcoming molecular barriers. These findings provide new insights into the unexpected Φ29 dynamic sequence sensitivity and lay the groundwork to understand how pausing coordinates DNA replication by bacteriophages.

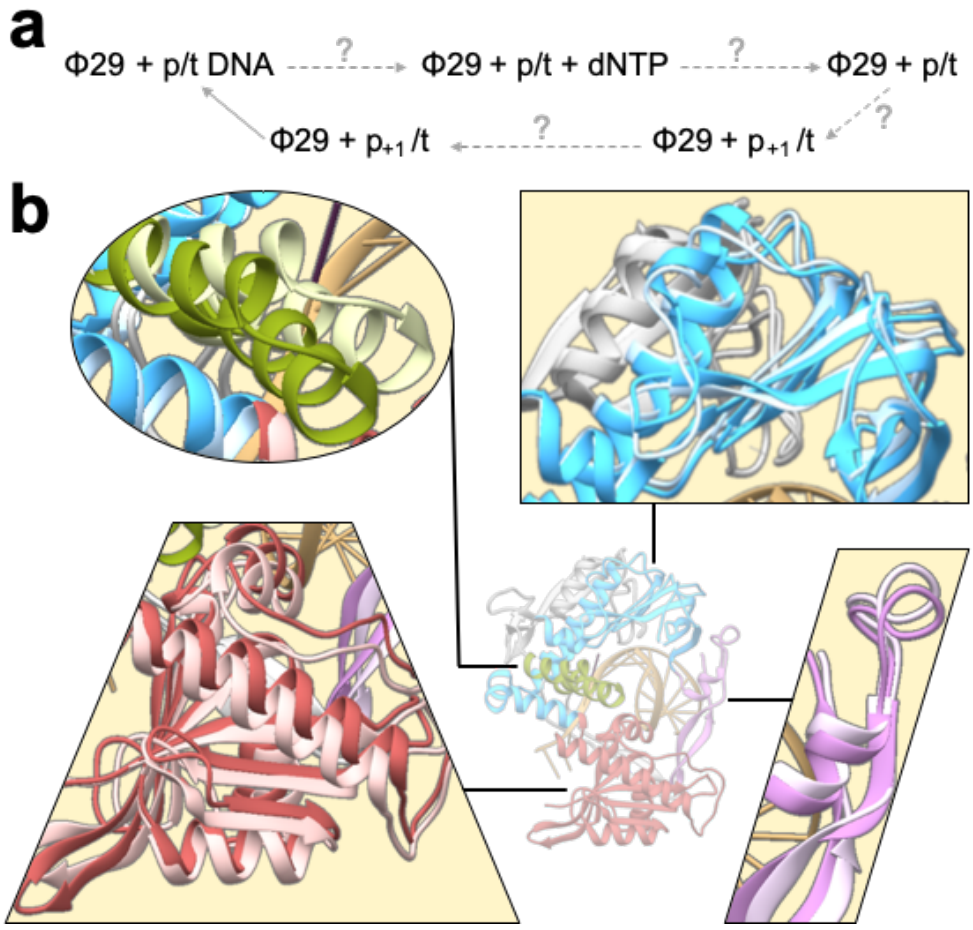


Figure 3-1. Φ29's conformational changes during DNA polymerization. A) This chapter describes the dynamic aspects of these conformational changes that are unresolved by structural studies (dashed arrows). **b**) A schematic diagram of the fingers (green), palm (blue), thumb (pink), and exonuclease (red) domains between open and closed conformations upon Φ29 binding a primer-template substrate (tan) with the incoming dNTP (purple). No direct evidence for a transient non-functional conformation change during pausing has been obtained.

3.3 Results and Discussion

Φ29-nanocricuits generated three categories of signals (**Figure 3-2**). As demonstrated previously, such signals result from dynamics and composition of charged functionalities near the attachment site.[25] Φ29 has seven native cysteine residues, and proved stubbornly resistant to substitution mutagenesis (**Figure 3-3**). Thus, experiments

reported here applied an exonuclease-deficient Φ29 DNA polymerase variant, referred to as Φ29, with five, surface-exposed thiols. Mass spectrometry of the maleimide-modified protein (intact and post-digestion) identified the following four cysteines as susceptible to maleimide modification: C106, C290, C448, and C530 (**Figure 3-4**). We interpret the three categories of signals generated by Φ29-nanocircuits as resulting from attachment to the four available cysteine residues on Φ29. For two positions, the signals appeared identical within error, and cannot be distinguished from each other. The vast majority of collected data produced either no signals or template-independent signals (**Figure 3-2b** or **3-2c,** respectively). For this second category of data, signals were observed in the absence of DNA template and complementary dNTPs.

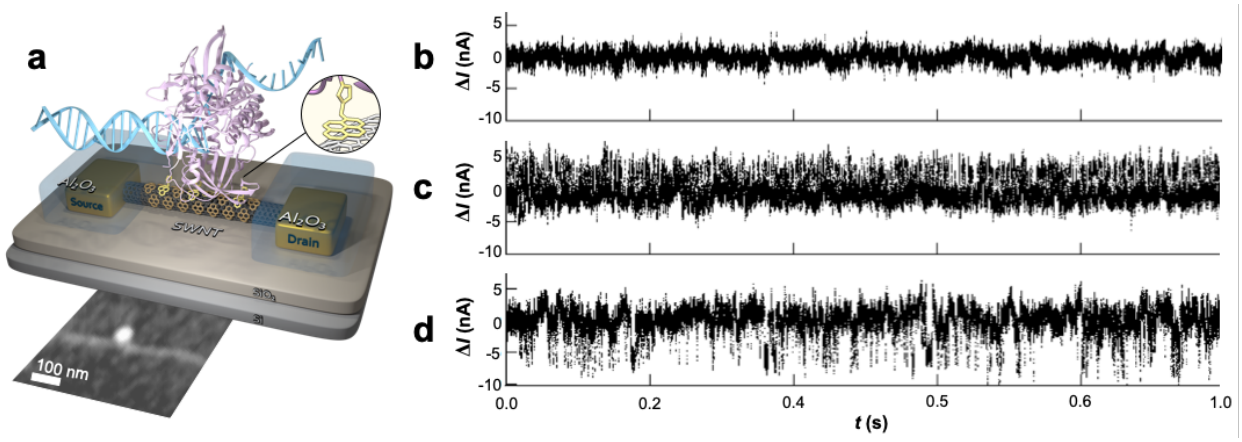


Figure 3-2. Φ29-nanocricuits generated three categories of signals. A) Schematic diagram of a Φ29-nanocircuit complex. The nanocircuit is non-covalently bioconjugated to a single molecule of Φ29. The nanocircuit is grown on SiO_2, and is connected to source and drain metal electrodes. Aluminum oxide (Al_2O_3, blue) passivates the electrodes, and leaves only a small window of SWCNT accessible. A pyrene-maleimide linker (yellow) adheres to the nanocircuit through π-π stacking and covalently attaches to a surface-exposed cysteine to immobilize the Φ29. Atomic force microscopy shows the expected 1−2 nm diameter of the nanocircuit with a single Φ29 attachment (inset). **B)** Some complexes produced no $\Delta I(t)$ signal above baseline under any conditions, **C)** others produced constant fluctuations in $\Delta I(t)$ despite the absence of DNA template and complementary dNTPs, and D) others produced template-dependent excursions in $\Delta I(t)$. This report exclusively focuses on the latter category of signals.

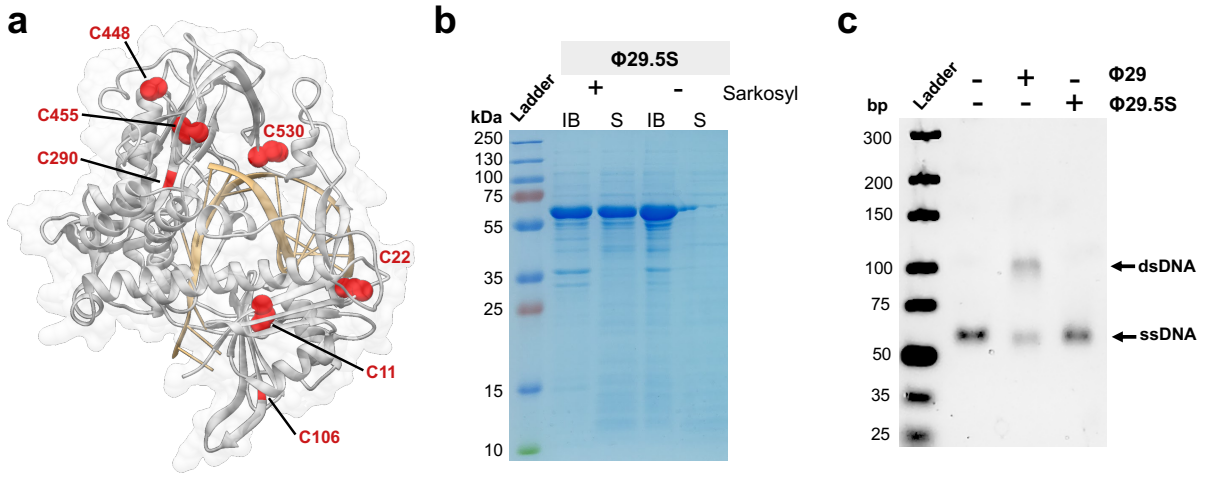


Figure 3-3. Mutations to five of Φ29's cysteines render the polymerase insoluble and inactive. A) Schematic of the five native cysteines were targeted for mutagenesis (red, spheres) to generate the Φ29 variant Φ29.5S (with mutations at D12A, D66A, C11S, C22S, C448S, C445S, and C530S). **b**) The detergent sarkosyl (1%) can solubilize this protein variant (12% SDS-PAGE, IB = inclusion bodies were insoluble protein is concentrated, S = supernatant fraction for the soluble protein). As expected, the detergent-solublized protein migrated at its expected MW of ≈66 kDa. **c**) This variant proved inactive in an ensemble DNA polymerase activity assay (5% high resolution agarose gel of samples pre-stained with SYBR Green I, where dsDNA = double-stranded DNA, resulting from polymerization with the **Poly T** template and dATP). The stained dsDNA exhibits slower electrophoresis than the single-stranded DNA (ssDNA). The negative control reaction omitted the enzyme, and the positive control reaction applies Φ29. The Ultra Low Range DNA ladder (ThermoFisher) was used to estimate DNA sizes.

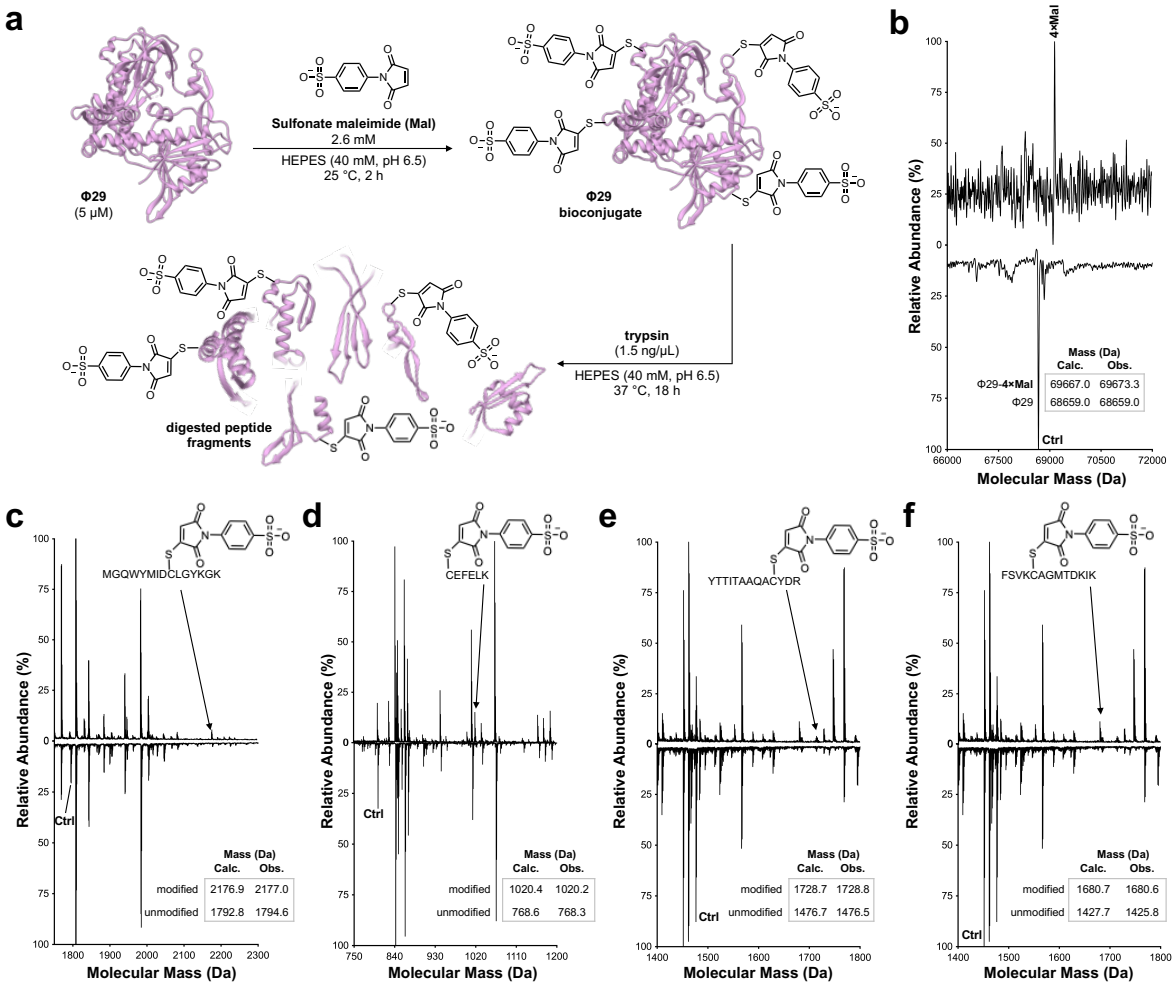


Figure 3-4. Mass spectrometry of intact and digested Φ29 identified four cysteines susceptible to maleimide modification. A) Schematic depiction of Φ29 reacted with sulfonate maleimide (Mal) by intact protein ESI LC-MS and digest MALDI-TOF of Φ29 bioconjugate. Mal is a derivative of pyrene maleimide designed to be more compatible for mass spectrometry analysis. b) Deconvoluted mass spectra of Φ29 bioconjugate after treatment with 2.6 mM Mal for 2 h at 25 °C. Traces on the top-half of the graph were obtained for the Mal-modified samples; traces on the bottom-half of the graph were obtained from the unmodified control samples of Φ29. b-e) Peptide fragments with (above) or without (below) the Mal modification appeared in a MALDI-relevant MW range. Specifically, the spectra capture peptide fragments with the b) C106, c) C290, d) C448, and e) C530. Four Mal-modified sites were identified on Φ29. Controls (Ctrl) depict spectra obtained from identical reaction conditions in the absence of Mal.

This study thus focuses exclusively on the third category of signal, which were template- and dNTP- dependent (**Figure 3-2d).** $\Delta I(t)$ excursions of this type disappeared when the DNA template and complementary dNTPs were removed from the solution

(**Figure 3-5a**). Based on this negative control, we interpret the third category of signals as resulting from catalytic activity. Similar binary $\Delta I(t)$ excursions were observed during substrate-dependent conditions in previous experiments with KF.[26,27] Specifically, the binary $\Delta I(t)$ excursions correlate with the enzyme's open conformation (τ_{hi}) and its closed conformation (τ_{lo}) during DNA polymerization. **Figure 3-5** shows the typical $\Delta I(t)$ excursion during the extended recording of Φ29-nanocricuits in DNA templates and complementary dNTPs. In the presence of a 59 basepair (bp) template (**DNA1**) composed of the four DNA bases, the signal exhibited a succession of stochastic pulses at 10 to 400 s^{-1} below the mean baseline currents (**Figure 3-5b**). Similar results were observed for a longer heteropolymer template (**DNA2**) of 1476 bp (**Figure 3-6**). We interpret the short and long sporadic excursions, occasionally in short bursts, to be associated with DNA polymerization and incorporation of dNTPs. Similar, sporadic signals have been observed during DNA polymerization by Φ29 using nanopore and other DNA sequencing technologies.[28,29]

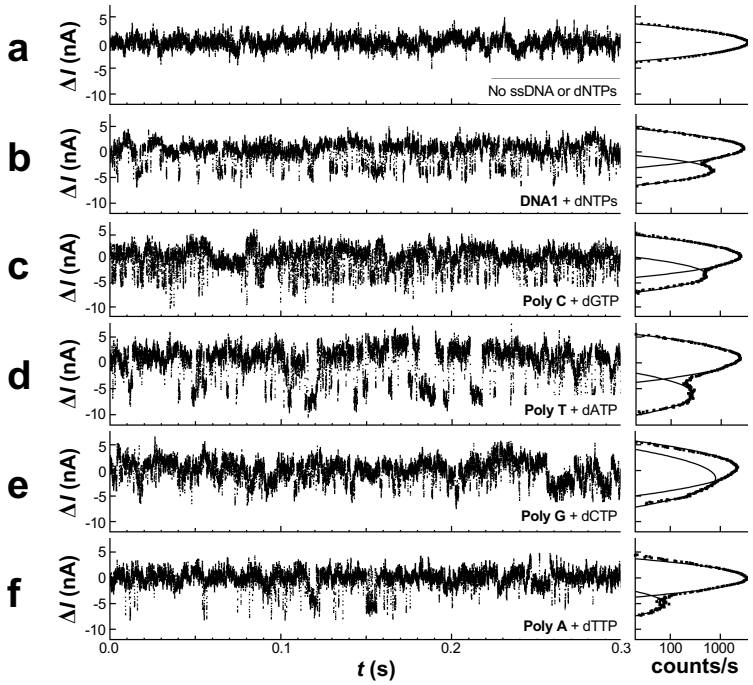


Figure 3-5. Representative Δ*I*(t) signals (left) and the corresponding histograms (right) produced by a single Φ29-nanocircuit complex. A) Omitting the DNA template and dNTPs (negative control) results in Φ29-nanocricuits incapable of generating Δ*I*(*t*) signals. **b**) Δ*I*(*t*) signals are observed in the presence of **DNA1** and dNTPs and also **c-f) Poly C**, **Poly T**, **Poly G**, and **Poly A** in the presence of a complementary dNTP.

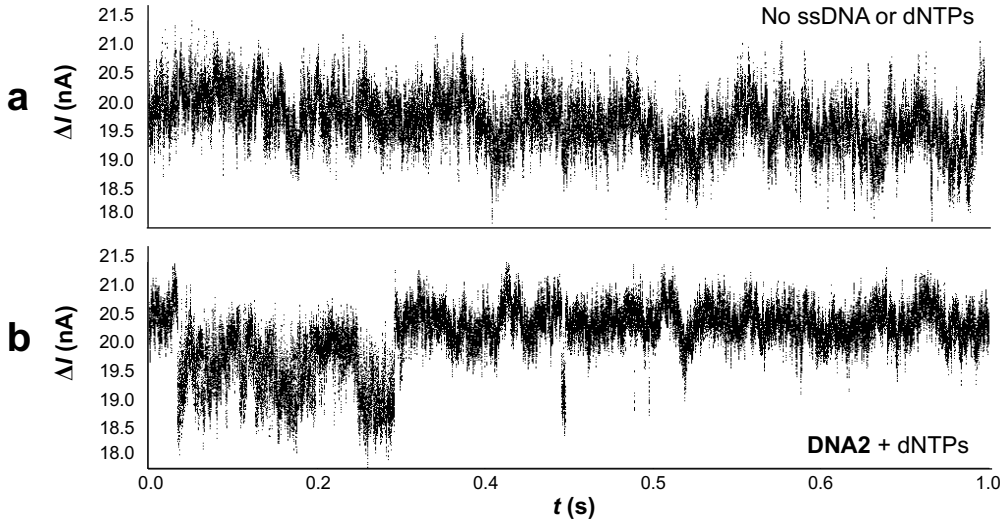


Figure 3-6. Representative Δ*I*(t) signals produced by a single Φ29-nanocircuit complex from a long heterotemplate (DNA2) and dNTPs. A) No excursions of Δ*I*(*t*) from the mean baseline currents were observed for the Φ29-nanocricuits in the absence of the DNA template and dNTPs. **b)** Δ*I*(*t*) signals generated in the presence of a **DNA2** and dNTPs.

To examine the sequence dependence of Φ29 kinetics, homopolymeric DNA templates with 42 unpaired cytosines (**Poly C**), thymidines (**Poly T)**, guanosines (**Poly G**), or adenosines (**Poly A**) were annealed to the M13 primer. Prior to single-molecule studies, an ensemble-level polymerase extension assay was performed with these templates in the presence of their complementary dNTP. A 5% high resolution agarose gel stained with the dsDNA-specific SYBR Green dye shows **DNA1** is converted to dsDNA in the presence of dNTPs (**Figure 3-7**). Similarly, the homopolymeric DNA templates **Poly C** and **Poly T** are converted to dsDNA in the presence of dGTP and dATP, respectively (**Figure 3-8a** and **Figure 3-8b**, lanes 9 and 15). In contrast, no clear polymerization by Φ29 was observed after 30 min reactions using **Poly G** or **Poly A** in the presence of their complementary dNTP (**Figure 3-8a**, lanes 1 and 8; **Figure 3-8b**, lanes 1 and 7). Despite the absence of ensemble observable polymerization for **Poly G** and **Poly A**, the characterization and kinetics were enabled by the single-molecule experiments described here.

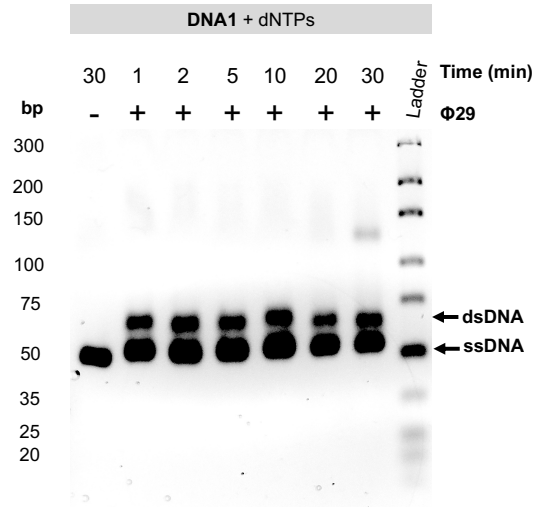


Figure 3-7. Φ29 polymerization of heterotemplate, DNA1, in the presence of complementary dNTPs after 1 minute of incubation. Φ29 demonstrated polymerization in an ensemble DNA polymerase activity assay (5% high resolution agarose gel of samples

pre-stained with SYBR Green I, where dsDNA = double-stranded DNA, resulting from polymerization with the **DNA1** and dNTPs). The stained dsDNA exhibits slower electrophoresis than the single-stranded DNA (ssDNA). The negative control reaction omitted the enzyme. The Ultra Low Range DNA ladder (ThermoFisher) was used to estimate DNA sizes.

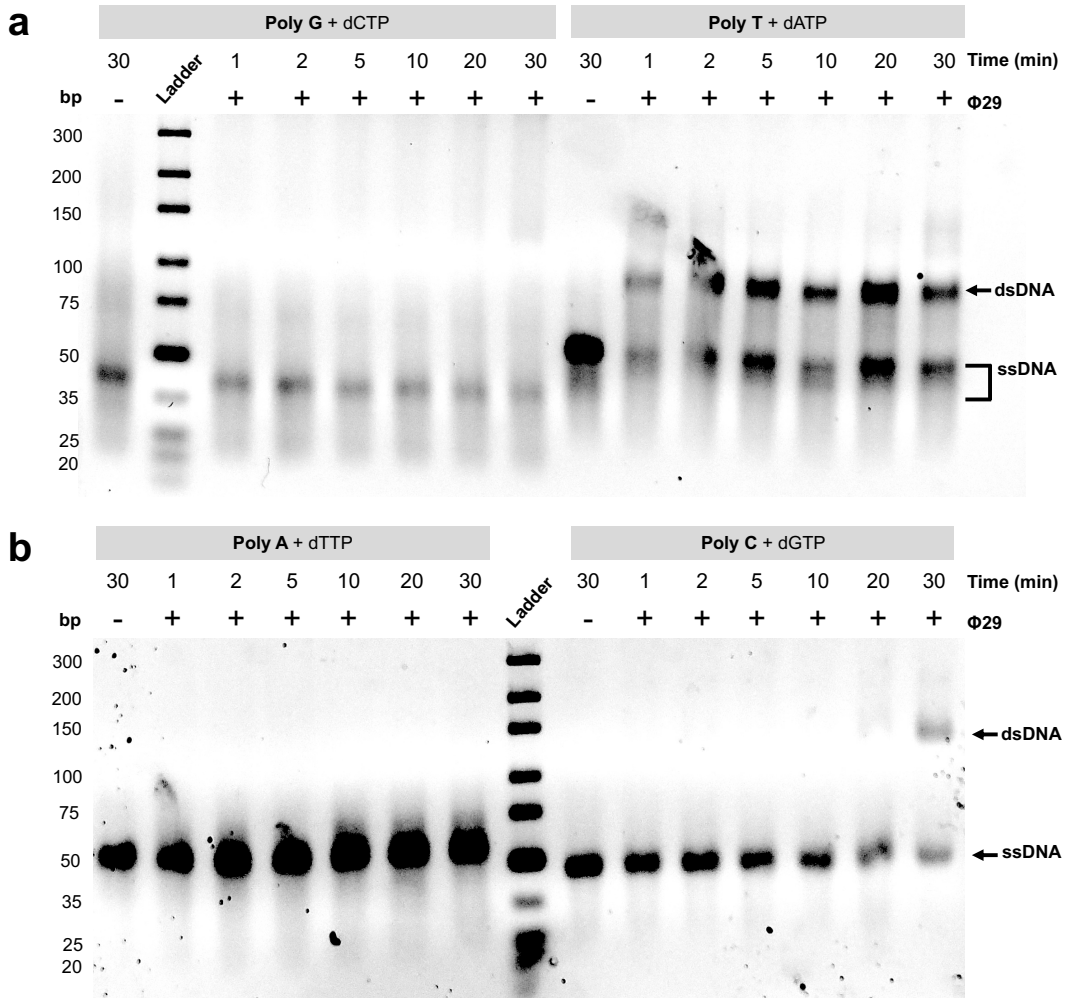


Figure 3-8. Φ29 polymerization of Poly T and Poly C, however Poly A and Poly G remain replication obstacles. Φ29 demonstrated polymerization in an ensemble DNA polymerase activity assay (5% high resolution agarose gel of samples pre-stained with SYBR Green I, where dsDNA = double-stranded DNA and ssDNA = single-stranded DNA), resulting from polymerization with the **a) Poly G** or **Poly T** and **b) Poly A** or **Poly C** in the presence of the complementary dNTP). The stained dsDNA exhibits a slower electrophoretic migration rate than the single-stranded DNA (ssDNA). The negative control reaction omitted the enzyme (-). The Ultra Low Range DNA ladder (ThermoFisher) was used to estimate DNA sizes.

Single-molecule catalysis examined polymerization with the homopolymeric templates. The complementary dNTPs were present in excess, and the resultant average catalytic rates of Φ29 appear in **Table 3-1**. The catalytic rate of Φ29 was fastest for dGTP incorporation into **Poly C**; fed this template, the enzyme exhibits rapid, short excursions in $\Delta I(t)$ excursions, which were interspersed with periods of decreased activity or even inactivity (**Figure 3-5c** and described further below). This rate for **Poly C** is in excellent agreement with ensemble-measured replication rates reported previously for this polymerase operating on heteropolymeric templates (~80 nt/s).[18,30,31] Similarly, the average rate of Φ29 with **Poly T** and **Poly G** are within error of the reported ensemble rates (40-60 nt/s).[32,33] The catalytic rates obtained from our single-molecule measurements were highly heterogeneous. For example, **Poly T** displayed a sporadic mix of short and long excursions of $\Delta I(t)$ below the mean baseline currents (**Figure 3-5d**), and **Poly G** exhibited periods of silence after 10 to 50 surrounding low-amplitude excursions below the baseline (**Figure 3-5e**). The slowest rate was observed for dTTP incorporation into **Poly A** (~3 nt/s). This low average rate results from sparse bursts of catalysis interspersed with long, silent periods of 5-20 s (**Figure 3-5f)**. In summary, the single-molecule experiments clarify the basis for observations made at the ensemble level. First, the enzyme exhibits a maximum rate of speed with either **Poly C** or a heteropolymeric template. Second, the enzyme can process problematic templates (as explained further below), but experiences long stalling periods.

Table 3-1. Kinetics of catalysis by Φ29. The average rate of incorporation k was calculated as the sum of the kinetic parameters (τ_{hi} and τ_{lo} ± standard deviation) for the indicated templates in the presence of the complementary dNTP(s). The ΔIh-l is the amplitude of the $\Delta I(t)$ between the high and low states according to the fitted Gaussian, and is calculated for each template with the indicated uncertainty (σ_{prim}).

Template	τ_{lo} (ms)	τ_{hi} (ms)	k (1/s)	$\Delta I_{h\text{-}1}$ (nA)	σ_{prim} (nA)
Poly C	130 ± 70	5 ± 3	80	-3.8	2.0
Poly T	170 ± 90	13 ± 7	21	-6.9	2.1
Poly G	130 ± 70	7 ± 4	21	-1.9	2.5
Poly A	220 ± 110	30 ± 15	3	-5.0	1.7

The enzyme varies considerably in the durations of its open (τ_{hi}) and closed (τ_{lo}) conformations, when processing homopolymeric templates. Individual histograms for these states are normalized to form probability distributions based on this data (**Figure 3-9**). The distributions were used to calculate the pausing kinetics of Φ29. Φ29 spends 1, 3, 3, or 8 min paused when processing **Poly C**, **Poly T, Poly G,** or **Poly A,** respectively. We interpret the enzyme's difficulty to process some homopolymeric templates, as a result of DNA secondary structure. Φ29 has endogenous helicase activity, but cannot displace DNA structures that do not maintain bending of the template and non-Watson-Crick base pairing.[18]

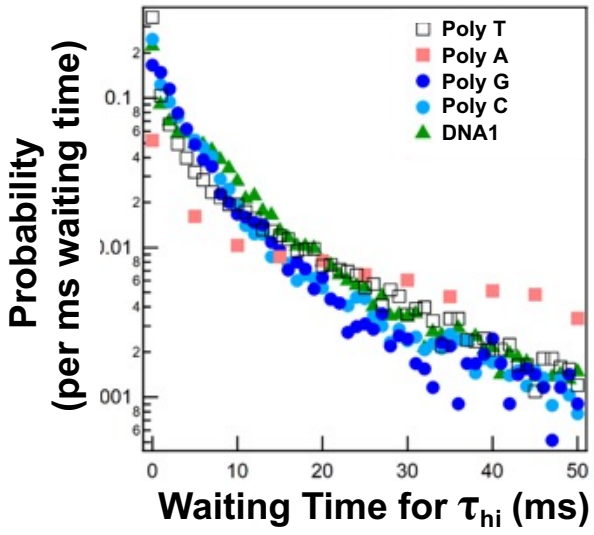


Figure 3-9. Representative distributions for open state (τ_{hi}) waiting times between excursions for each of the measured DNA templates, calculated as the probability. The distributions for the heterotemplate, DNA1, overlaps with the homotemplates with the expectation of PolyT.

Poly G, for example, can form the well-characterized G-quadruplexes to induce the observed pausing. Previous experiment with T4 DNA polymerase demonstrated pauses at kinetic barriers formed by G/C-rich hairpins with stems greater than eight base pairs (e.g., a 7.5-fold increase in the rate of progression through a pause site with the addition of a "helix-destabilizing protein" called 32 protein).[34] Φ29 has a single-stranded binding protein called protein gene 5 that increases the DNA replication,[35] which may play a role in reducing the enzyme's pausing behavior. Other studies have shown the addition of other factors known to disrupt DNA base pairing and secondary structure, such as betaine and increasing temperatures, have decreased the enzyme's pausing at G/C rich sequences during primer extension.[10,36–39] Thus, our studies uncover the fundamental basis for the ensemble-level observation of stalling at G-quadruplexes; specifically, the enzyme gets stuck in its open state, and cannot accept the template into the active site.

Poly A does not form any stable secondary structure,[40] but exhibits greater rigidity due to stronger π–π stacking between bases.[41] For example, *E. coli* DNA polymerase III slows while processing templates with unusual base compositions, such as high levels of adenosine (48% slower) and low levels of cytidine and guanosine (9% and 10%, slower rates respectively).[42] Notably, the requirement for flexible ssDNA to fit through the constrained Φ29 is not mirrored for its dsDNA product. **Poly T**, for example, and its attendant π-stack of dsDNA Poly A's has no problem being processed by this polymerase. Evidently, the ssDNA template must exhibit flexibility, but the exit tunnel imposes no such constraints. Inspection of X-ray crystal structures for this polymerase reveals a kink in the template strand to avoid steric hinderance with the enzyme's exonuclease C-terminal α-helix and fingers. Such conserved residues have been previously shown to affect polymerization.[43,44] The downstream template is large enough to accommodate a purine and pyrimidine bases; yet, from structural modeling, before the downstream tunnel template, the dsDNA must bend 90°. Thus, the enzyme stalls upon encountering the bend-resistant **Poly A** sequence.

To characterize the length dependence of Φ29's sequence-specific pausing, we next generated DNA templates with different numbers of adenosine repeats within a backbone of the thymidines. This design takes advantage of the enzyme's proven capabilities to quickly process **Poly T**, yet stall on **Poly A**. In the ensemble assay, five or more adenosines dramatically decreases the DNA polymerization catalyzed by Φ29 (**Figure 3-10a**, lanes 9 and 10). This observation is consistent with recent work that showed the signal for DNA templates with ten repeated adenosines was typically more irregular at the single-molecule level; further investigation at the ensemble-level

determined polymerization of such templates is only partially successful.[45] As observed for **Poly A** at the single molecule level, the enzymes stalls likely due to decreased flexibility to accommodate necessary conformational changes during DNA synthesis.

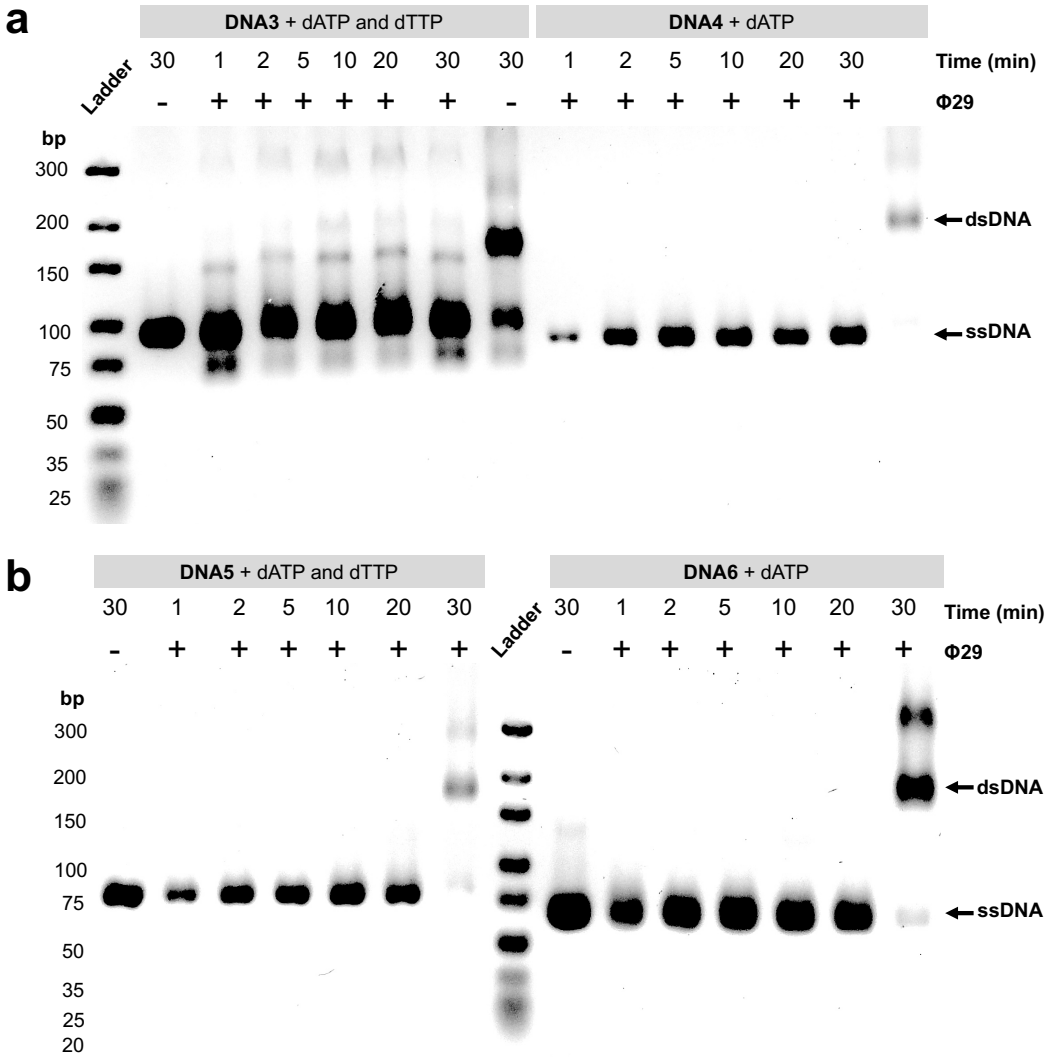


Figure 3-10. Five consecutive adenines pose an obstacle for Φ29 during polymerization of Poly T. Assays were carried out as described in Methods in the presence of 1 μM Φ29, 5 μM DNA template, and 10 μM dATP at the indicated incubation times. Templates used were **Poly T** with consecutive adenines (A_n) in the middle of the sequence and are indicated at the top of the panel (**DNA3** = Poly T A_4, **DNA4** = Poly T A_5, **DNA5** = Poly T A_6, **DNA6** = Poly T A_{12}). The negative controls had identical conditions, but omitted the enzyme and incubated the dATP and dTTP for 30 min.

The Φ29 bacteriophage genome includes stretches of adenosines, which could impact the virus's replication. Specifically, the Φ29 genome contains 26 adenosine repeats (A_n, where n ≥ 5 **Table 3-2**). Severe pausing by Φ29 was observed during the replication of A_6 containing genome sequences. However, substituting the A_6 for T_6 alleviated enzymatic pausing (**Figure 3-11**). Thus, we conclude that sequence sensitivity could contribute a functional role during bacteriophage DNA and viral replication. Notably, the bacterial host for this bacteriophage and other bacteriophages include a large number of such adenosine repeats, including higher frequencies of such repeats and also longer lengths (**Figure 3-12**). Thus, viruses could benefit from such stalling by avoiding the synthesis of the host genome during viral infection. Such pausing could provide a regulatory role to focus the virus' machinery on viral expansion.

Table 3-2. Genome composition analysis for Φ29 bacteriophage and its host, *B. subtilis*. The frequency of consecutive adenosine (A), cytidine (C), guanosine (G), and thymidine (T) sections in the Φ29 bacteriophage (top) and *B. subtilis 168* (bottom) genomes are given for repeat lengths (bp). There highest number of adenosines repeated in the Φ29 bacteriophage and *B. subtilis* genomes is six and nine, respectively.

Φ29 bacteriophage

Repeat length (bp):	1	2	3	4	5	6
A	3936	1297	407	128	28	2
C	3120	540	118	21	2	0
G	3122	636	128	24	0	0
T	4011	1232	395	107	24	3

Bacillus subtilis 168

Repeat length (bp):	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
A	778690	254162	96529	37531	14451	5031	1476	193	9	1	0	0	0	0	0	0	0
C	728254	157785	26862	5229	1000	141	10	3	0	0	0	0	0	0	0	0	0
G	725557	156998	26429	5063	945	112	7	1	0	0	0	0	0	0	0	0	0
T	782230	254410	97022	37852	14668	5259	1484	184	19	2	1	1	1	1	1	1	1

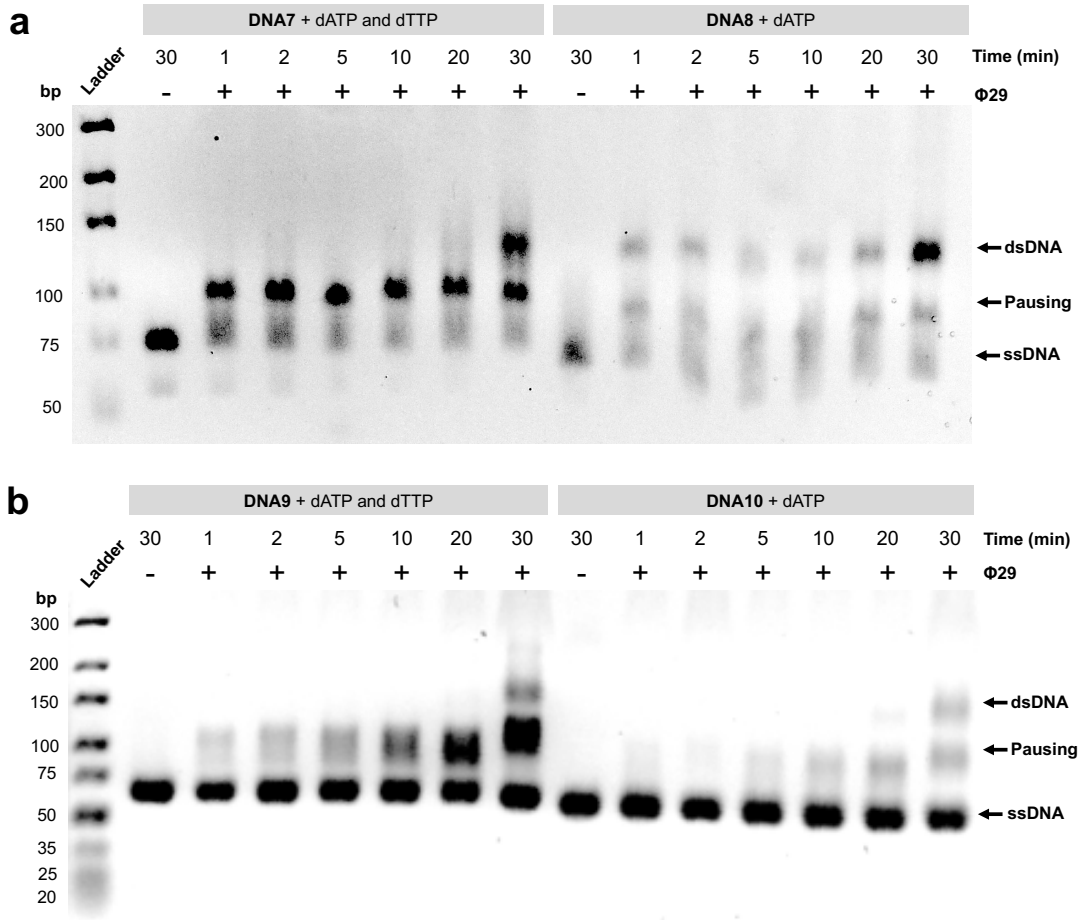


Figure 3-11. Six or ten consecutive adenines in heteropolymeric DNA templates pose an obstacle for Φ29 polymerization. Assays were carried out as described in Methods in the presence of 1 μM Φ29, 5 μM DNA template, and 10 μM dATP and/or dTTP and increasing incubation times. Templates used were based on the consecutive adenosine sections in the **a**) Φ29 bacteriophage or **b**) *B. subtilis* genomes. The negative controls omitted the enzyme and were incubated for 30 min in the presence of template and complementary dNTP.

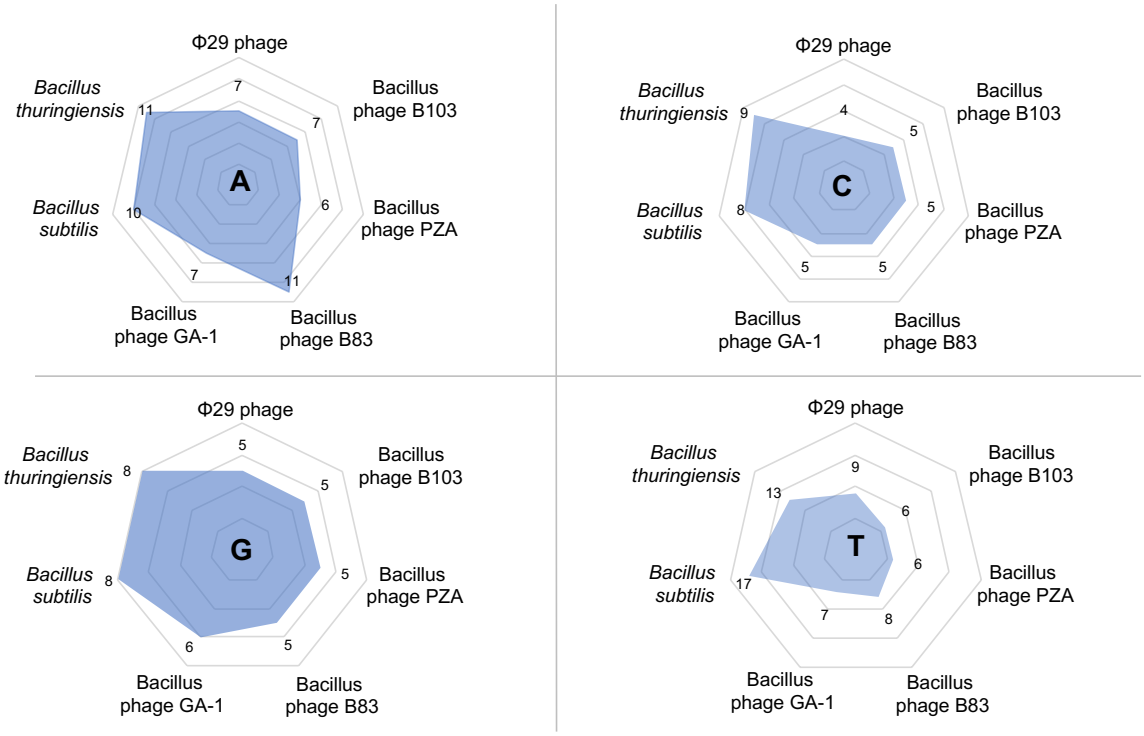


Figure 3-12. Genome composition analysis for Φ29 bacteriophage and its host, *B. subtilis*, and similar bacteriophages and bacteria. The maximum number of repeated adenosine (A), cytosine (C), guanosine (G), and thymine (T) with genomes. Φ29 bacteriophage's genome has fewer repeated adenosines then *B. subtilis*' genome. In general, similar bacteriophages to Φ29 bacteriophage share a low number of repeated adenosines. A similar bacteria, *Bacillus thuringiesis*, has a higher number of repeated adenosines.

Figure 3-13 shows a histogram of the average rate during an active period revealed two distinct Gaussian peaks corresponding to two distinct rates at ~40s^{-1} and ~190s^{-1} with Φ29 switching between these two rates and also the inactive state. The long excursions were infrequent, however, and exhibited a similar duration to the waiting time between excursions. Histograms for **Poly T** showed slight bumps, each with two maxima at ~20 s^{-1} and ~100 s^{-1} (**Figure 3-13)**. Compared to the baseline generated in the presence of buffer along, the increased background noise of the nanocircuit precluded estimates of rates below 10 s^{-1}. However, the $\Delta I(t)$ from **Poly G** exhibited larger fluctuations in the baseline comparison to the other homopolymeric DNA templates. For

example, there were longer periods of scarce activity or inactivity (up to 500 s) and short bursts in excursions (< 2 s), which were correlated with the template length and often separated by gaps (5-60 s).

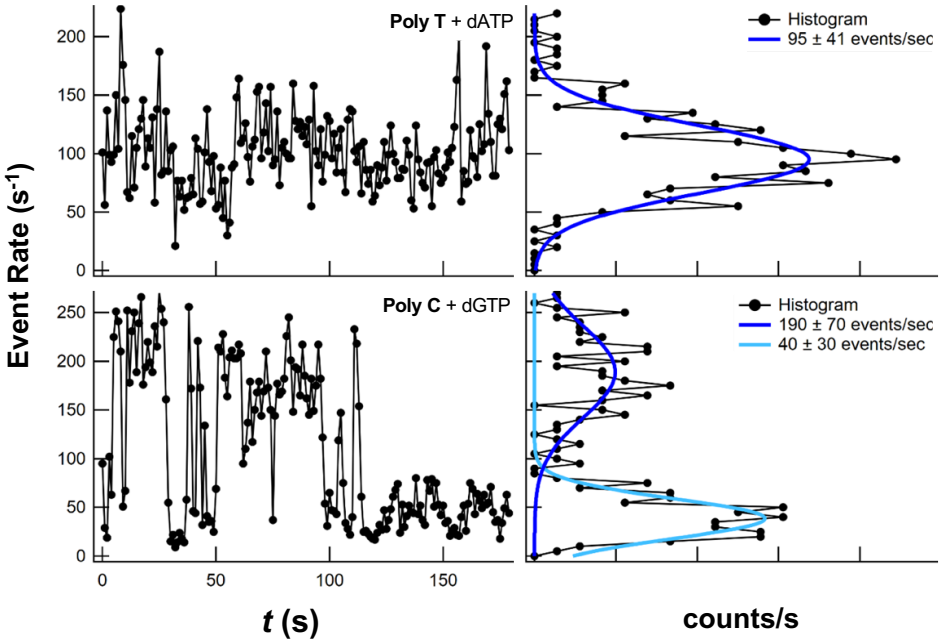


Figure 3-13. Histograms of Φ29 catalytic rate. Φ29 changes catalytic rates over time, as shown in these plots of the rate per second for catalysis with two different DNA templates (left) and the histograms of the rate plotted vertically on a log-linear scale (right) for the **Poly T** and **Poly C**. Each plot shows the rate over one complete continuous measurement, with the time stamp of each averaged measurement (in seconds) shown on the bottom axis.

3.4 Conclusion

The DNA polymerase from the *B. subtilis* bacteriophage (Φ29) is highly processive (70,000 bp is its maximum recorded processivity),[46] yet exhibits a complex pausing behavior. Variants defective for exonuclease activity (D12A, D66A) are more prone to pausing. Here, individual exonuclease-deficient Φ29 molecules were tethered by a pyrene maleimide linker to nanocircuits to examine DNA polymerase variability of pause events. Φ29-nanocricuits generated three categories of signals, one of which was

template- and dNTP- dependent category resulting from catalytic activity. Specifically, the binary $\Delta I(t)$ signals correlate with the enzyme's open conformation (τ_{hi}) and its closed conformation (τ_{lo}) during DNA polymerization. The average rates of catalysis were largely determined by the durations, and revealed a highly variable nature of the enzyme's catalytic rate. Homopolymeric DNA templates (e.g., **Poly A, Poly C, Poly G,** and **Poly T**) were used to demonstrate that the enzyme's catalytic rate is template-dependent. The average rates of dNTP incorporation for **Poly C**, **Poly T, Poly G,** and **Poly A** were 81, 21, 21, and 2 dNTP/sec, respectively. During polymerization, Φ29 spends 1, 3, 3, or 8 min paused when processing **Poly C**, **Poly T, Poly G,** or **Poly A,** respectively. We interpret the enzyme's difficulty to process some homopolymeric templates, as a result of DNA secondary structure (e.g., G-quadruplexes in **Poly G** and rigidity of **Poly A**). We conclude that sequence sensitivity can contribute an important functional role during bacteriophage DNA and viral replication.

3.5 Materials and Methods

Materials

Reagents purchased commercially include the following: cell lines (Stratagene), IPTG (Carbosynth), glycine (Biobasic), and Ni-IMAC resin (Bio-Rad Laboratories). All other chemicals were supplied by Fisher Scientific or ThermoFisher Scientific. Supplies purchased commercially include: 0.45 μm pore filter (Genesee Scientific), 0.22 μm pore filter (Denville Scientific), 0.22 μm top vacuum filter (Genesee Scientific), 3.5K MWCO 0.1 mL Slide-A-Lyzer mini dialysis device (ThermoFisher), and Amicon Ultra-0.5 mL centrifuge filter (Fisher Scientific). All reagents and supplies were used as received. All solutions were sterile-filtered or autoclaved before use.

Proteins used in this study

Table 3-3. The amino acid sequence and theoretical MW of the Φ29 variants used in this study.

Protein	Sequence (N-terminal His_6 epitope indicated in blue and mutations indicated in red)	Description
Φ29	MGSSHHHHHHSSGLVPRGSHMKHMPRKMYS**SA**FETTTKVED**S**RVWAY GYMNIEDHSEYKIGNSLDEFMAWVLKVQADLYFHNLKF**A**GAFIINWL ERNGFKWSADGLPNTYNTIISRMGQWYMIDICLGYKGKRKIHTVIYD SLKKLPFPVKKIAKDFKLTVLKGDIDYHKERPVGYKITPEEYAYIKN DIQIIAEALLIQFKQGLDRMTAGSDSLKGFKDIITTKKFKKVFPTLS LGLDKEVRYAYRGGFTWLNDRFKEKEIGEGMVFDVNSLYPAQMYSRL LPYGEPIVFEGKYVWDEDYPLHIQHIRCEFELKEGYIPTIQIKRSRF YKGNEYLKSSGGEIADLWLSNVDLELMKEHYDLYNVEYISGLKFKAT TGLFKDFIDKWTYIKTTSEGAIKQLAKLMLNSLYGKFASNPDVTGKV PYLKENGALGFRLGEEETKDPVYTPMGVFITAWARYTTITAAQA**S**YD RIIY**S**DTDSIHLTGTEIPDVIKDIVDPKKLGYWAHESTFKRAKYLRQ KTYIQDIYMKEVDGKLVEGSPDDYTDIKFSVK**S**AGMTDKIKKEVTFE NFKVGFSRKMKPKPVQVPGGVVLVDDTFTIK	A pET15b plasmid having an ORF encoding the *B. subtilis* bacteriophage phi29 DNA polymerase I (a kind gift from Illumina) was mutated (D12A/D66A) to remove the polymerase's exonuclease activity and fused to the N-terminal His_6 peptide epitope; this gene product is referred to as Φ29.
Φ29.5S	MGSSHHHHHHSSGLVPRGSHMKHMPRKMYS**SA**FETTTKVED**S**RVWAY GYMNIEDHSEYKIGNSLDEFMAWVLKVQADLYFHNLKF**A**GAFIINWL ERNGFKWSADGLPNTYNTIISRMGQWYMIDICLGYKGKRKIHTVIYD SLKKLPFPVKKIAKDFKLTVLKGDIDYHKERPVGYKITPEEYAYIKN DIQIIAEALLIQFKQGLDRMTAGSDSLKGFKDIITTKKFKKVFPTLS LGLDKEVRYAYRGGFTWLNDRFKEKEIGEGMVFDVNSLYPAQMYSRL LPYGEPIVFEGKYVWDEDYPLHIQHIRCEFELKEGYIPTIQIKRSRF YKGNEYLKSSGGEIADLWLSNVDLELMKEHYDLYNVEYISGLKFKAT TGLFKDFIDKWTYIKTTSEGAIKQLAKLMLNSLYGKFASNPDVTGKV PYLKENGALGFRLGEEETKDPVYTPMGVFITAWARYTTITAAQA**S**YD RIIY**S**DTDSIHLTGTEIPDVIKDIVDPKKLGYWAHESTFKRAKYLRQ KTYIQDIYMKEVDGKLVEGSPDDYTDIKFSVK**S**AGMTDKIKKEVTFE NFKVGFSRKMKPKPVQVPGGVVLVDDTFTIK	Five native cysteines were mutated to serine in Φ29 to generate the Φ29 variant D12A/D66A/C11S/C22S/C448S/C445S/C530S, referred to as **Φ29.5S**. The ORF encoding **Φ29.5S** was ligated into a LIC-pET28 plasmid.
Φ29 exo$^+$	MGSSHHHHHHSSGLVPRGSHMKHMPRKMYSC**D**FETTTKVEDCRVWAY GYMNIEDHSEYKIGNSLDEFMAWVLKVQADLYFHNLKF**D**GAFIINWL ERNGFKWSADGLPNTYNTIISRMGQWYMIDICLGYKGKRKIHTVIYD SLKKLPFPVKKIAKDFKLTVLKGDIDYHKERPVGYKITPEEYAYIKN DIQIIAEALLIQFKQGLDRMTAGSDSLKGFKDIITTKKFKKVFPTLS LGLDKEVRYAYRGGFTWLNDRFKEKEIGEGMVFDVNSLYPAQMYSRL LPYGEPIVFEGKYVWDEDYPLHIQHIRCEFELKEGYIPTIQIKRSRF YKGNEYLKSSGGEIADLWLSNVDLELMKEHYDLYNVEYISGLKFKAT TGLFKDFIDKWTYIKTTSEGAIKQLAKLMLNSLYGKFASNPDVTGKV PYLKENGALGFRLGEEETKDPVYTPMGVFITAWARYTTITAAQACYD RIIYCDTDSIHLTGTEIPDVIKDIVDPKKLGYWAHESTFKRAKYLRQ KTYIQDIYMKEVDGKLVEGSPDDYTDIKFSVKCAGMTDKIKKEVTFE NFKVGFSRKMKPKPVQVPGGVVLVDDTFTIK	The native 3' to 5' exonuclease activity was restored in Φ29 to generate wild-type Φ29, referred to as **Φ29 exo$^+$**.
Φ29.P	MGSSHHHHHHSSGLVPRGSHMKHMPRKMYSC**A**FETTTKVEDCRVWAY GYMNIEDHSEYKIGNSLDEFMAWVLKVQADLYFHNLKF**A**GAFIINWL ERNGFKWSADGLPNTYNTIISRMGQWYMIDICLGYKGKRKIHTVIYD SLKKLPFPVKKIAKDFKLTVLKGDIDYHKERPVGYKITPEEYAYIKN	To reduce polymerase pausing, a Φ29 variant referred from here as **Φ29.P**, was created with

DIQIIAEALLIQFKQGLDRMTAGSDSLKGFKDIITTKKFKKVFPTLS
LGLDKEVRYAYRGGFTWLNDRFKEKEIGEGMVFDVNSHYPAQMYSRL
LPYGEPIVFEGKYVWDEDYPLHIQHIRCEFELKEGYIPTIQIKRSRF
YKGNEYLKSSGGEIADLWLSNVDLELMKEHYDLYNVEYISGLKFKAT
TGLFKDFIDKWTYIKTTSEGAIKQLAKLMLNSLYGKFASNPDVTGKV
PYLKENGALGFRLGEEETKDPVYTPMGVFITAWGRYTTITAAQACYD
RIIYCDTDSIHLTGTEIPDVIKDIVDPKKLGYWAHESTFKRAKYLRQ
KTYIQDIYMKEVDGKLVEGSPDDYTDIKFSVKCAGMTDKIKKEVTFE
NFKVGFSRKMKPKPVQVPGGVVLVDDTFTIK

mutations
D12A/D66A/L253H/L437G
based on patent literature.[47]

Mutagenesis of Φ29

Overlap extension mutagenesis was employed to generate **Φ29.5S.** In this method, six PCR steps using Herculase II DNA Polymerase (Invitrogen) with **Oligo1** to **Oligo12** were performed to obtain the full length DNA (**Table 3-4**). The ORF of **Φ29.5S** cloned into a pet28 LIC vector (in-house) and was checked via DNA sequencing (Genewiz).

Q5 Site-directed mutagenesis (New England Biolabs) was applied to generate **Φ29 exo$^+$** and **Φ29.P** variants. For each variant, a two-step approach was applied to obtain the full length DNA using Q5 Hot Start High-Fidelity DNA Polymerase and **Oligo13** to **Oligo20** (**Table 3-4**). The ORF of **Φ29 exo$^+$** and **Φ29.P** variants were confirmed via DNA sequencing (Genewiz).

Table. 3-4. Oligos used for mutagenesis of Φ29 variants.

Entry	Sequence (5' to 3')	Purpose
Oligo1	GATGTACAGCAGCGCCTTCGAAACCACCACCAAGGTT GAAGATAGCCGTGTGTG	To generate C11S, C22S, and C106S mutations in **Φ29.5S** during PCR Step 1
Oligo2	CGTTTACCCTTGTAACCCAGACTGATATCAATCATG	
Oligo3	CTGGGTTACAAGGGTAAACG	PCR Step 2 for **Φ29.5S**
Oligo4	CTCTTTCAGTTCGAACTCACAACGGATATGCTG	
Oligo5	GAGTTCGAACTGAAAGAGGGC	To generate C448S and C455S mutations in **Φ29.5S** during PCR Step 3
Oligo6	GAGTCAGTGTCACTGTAGATAATGCGGTCGTAGCTCG CTTGCGCTG	
Oligo7	GACACTGACTCCATCCAC	To generate C530S mutation in **Φ29.5S** during PCR Step 4
Oligo8	TCATACCCGCGCTCTTAACCGAG	
Oligo9	CTCGGTTAAGAGCGCGGGTATGACGGACAAG	PCR Step 5 for **Φ29.5S**

Oligo10	TTACTTAATGGTGAAAGTATCATCTACC	
Oligo11	CAGGGCGCCAATGAAACACATGCCACGCAAGATGTAC	To ligate **Φ29.5S** in LIC-pET28 plasmid during PCR Step 6
Oligo12	GACCCGACGCGGTTATTACTTAATGGTGAAAGTATC	
Oligo13	GTACAGCTGCGACTTCGAAACCA	To generate A12D mutation in **Φ29 exo$^+$** during Q5 mutagenesis Step 1
Oligo14	ATCTTGCGTGGCATGTGTTTC	
Oligo15	TCTGAAATTTGACGGTGCGTTTATC	To generate A66D mutation in **Φ29 exo$^+$** during Q5 mutagenesis Step 2
Oligo16	TTATGGAAGTACAGGTCTG	
Oligo17	TGTCAACAGCCACTATCCGGCTCAAATGTAC	To generate L253H mutation in **Φ29.P** during Q5 mutagenesis Step 1
Oligo18	TCAAAGACCATGCCCTCG	
Oligo19	TCACCGCGTGGGGCCGTTATACCA	To generate A437G mutation in **Φ29.P** during Q5 mutagenesis Step 2
Oligo20	TGAAAACACCCATCGGGGTGTACAC	

Protein expression and purification of Φ29 variants

The following experimental details were used to express and purify **Φ29**, **Φ29.5S**, **Φ29 exo$^+$**, and **Φ29.P** variants. The DNA plasmid for each variant was transformed into BL21 (DE3) *E. coli* by the heat shock method. The transformed cells were transferred to an LB agar plate supplemented with an appropriate antibiotic - carbenicillin (50 µg/mL) for **Φ29**, **Φ29 exo$^+$,** and **Φ29.P** and kanamycin (40 µg/mL) for **Φ29.5S**. Following incubation at 37 °C for 14 to 16 h, a single colony from the transformation plate was deposited into 25 mL of 2YT media in a 125 mL baffled flask with the appropriate antibiotic. The primary seed culture was incubated at 37 °C with shaking at 225 rpm for 14 to 16 h. The secondary expression culture was prepared by inoculating 10 mL of the primary culture into LB media (1.0 L in a 2.0 L baffled flask) supplemented with the appropriate antibiotic. After shaking the secondary expression culture at 225 rpm at 37 °C, the density of the culture was monitored until the culture reached an OD_{600} value of 0.6 to 0.7. Then, isopropyl β-D thiogalactopyranoside (IPTG, 0.25 mM) was added to the secondary expression culture to induce overexpression of Φ29 before further incubation for 24 h at 16 °C with shaking at 225 rpm. The cells were harvested by centrifugation (6080 rcf, 20 min, 4 °C), and resuspended in lysis buffer (50 mM Tris HCl, 300 mM NaCl, 10 mM $MgCl_2$, 0.8 M trehalose, 10 mM $CaCl_2$, 1 mM phenylmethanesulfonyl fluoride –

PMSF, 10 mM β-mercaptoethanol – BME, and 1X Halt Protease and Phosphatase Inhibitor Cocktail (ThermoFisher), pH 7.5).

Cell lysis was performed by sonication using a Q500 Sonicator with the Standard Probe 4220 (QSonica, USA), temperature <8 °C, and 4 pulses (40% amplitude with 1 s per pulse followed by 3 s rest for 1 min). DNAse was added to the cell lystate and incubated at 37 °C for 1 h to eliminate DNA bound to the polymerase, then heated to 70 °C for 30 minutes to deactivate the DNAse. To improve protein solubility of **Φ29.5S**, 0.1% sarkosyl (Sigma) was added to the cell lysate and incubated overnight on a stirring plate at 4 °C. The cell lysis was subjected to centrifugation (26,892 rcf, 45 min, 4 °C) to remove cell debris. The lysate supernatant was filtered through a 0.45 µm pore filter, and then incubated with a Ni^{2+}-NTA column pre-equilibrated with lysis buffer. The supernatant mixed with Ni^{2+}-NTA resin was incubated overnight on a slow rotatory shaker at 4 °C. The column was washed with wash buffer (50 mM Tris HCl, 300 mM NaCl, 10 mM $MgCl_2$, 20 mM Imidazole, 10 mM BME, pH 7.5) and the purified protein was eluted using elution buffer (50 mM Tris HCl, 300 mM NaCl, 10 mM $MgCl_2$, 250 mM imidazole, 10 mM BME, pH 7.5).

The samples from the protein purification steps, flow-through, washes, and elutions were visualized using 12% SDS-PAGE (Bio-rad Mini-PROTEAN Tetra electrophoresis system) stained with Coomassie brilliant blue stain. Elutions containing the purified protein were pooled and concentrated using a 10 kDa cutoff microconcentrator (Vivaspin, Fisher Scientific, USA). The mixture was centrifuged, and then filtered through a 0.22 µm pore filter prior to size exclusion FPLC (Bio-Rad NGC Chromatography System, Superdex 75 pg, 26/600 at a flow rate of 1.0 mL. min^{-1}) in

activity buffer (40 mM HEPES, 300 mM NaCl, 10 mM $MgCl_2$, 100 µM Tris(2-carboxyethyl)phosphine) (TCEP) at pH 6.5). Following size exclusion FPLC, fractions were analyzed for purity using 12% SDS PAGE via ImageJ (**Figure 3-14**). For Φ29 ensemble activity assays the purified recombinant protein was dialyzed in activity buffer (40 mM HEPES, 300 mM NaCl, 10 mM $MgCl_2$, 0.1 mM TCEP, pH 6.5). For single-molecule measurements on nanocircuits, the recombinant protein was dialyzed in high salt buffer (40 mM HEPES, 300 mM NaCl, 10 mM $MgCl_2$, pH 6.5). The protein concentration was determined by a bicinchoninic acid (BCA) assay or Bradford assay using 68,558.15 g/mol as the estimated MW of the protein (http://www.expasy.org).

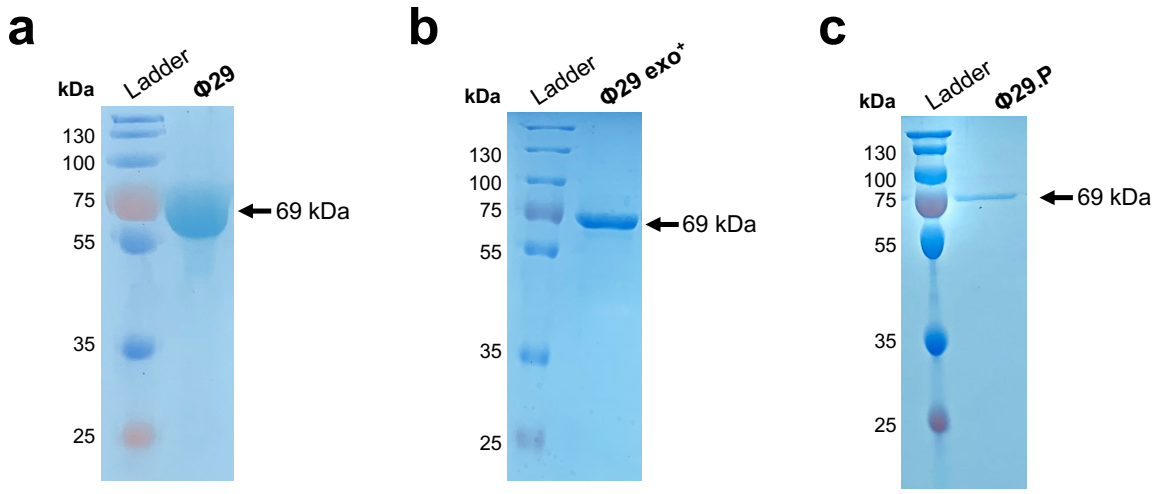


Figure 3-14. Φ29 variants purity assessed by 12% SDS-PAGE. These representative SDS-PAGEs visualize **Φ29**, **Φ29 exo$^+$,** and **Φ29.P** variants after size exclusion chromatography. Φ29 variants were purified to >90% homogeneity and migrated at their expected mass of ≈69 kDa.

Ensemble activity of Φ29 and dNTP incorporation

A. Oligonucleotides used to test activity

The following table lists the oligonucleotides used to test Φ29 activity and for measurements with the nanocircuit. Upon receipt, HPLC- purified oligonucleotides were solubilized in water to 100 μM.

Table 3-5. Oligos used for activity and electronic measurements. Bold regions indicate the M13 (5' ACTGGCCGTCGTTTTACA 3') or P8 (5'AGAATATTTGCCAGAACCGTTATGATGTCGGC GCAAAA 3') priming sites.

Entry	Sequence (5' to 3')	Purpose
DNA1	CCTAACGCAGATAGACGTTGTTTAGAGATTTAAATTCGGCC **ACTGGCCGTCGTTTTACA**	Test dNTP incorporation on DNA template < 50 bp
Poly A	$(A)_{42}$**ACTGGCCGTCGTTTTACA**	Test dTTP incorporation
Poly G	$(G)_{42}$**ACTGGCCGTCGTTTTACA**	Test dCTP incorporation
Poly C	$(C)_{42}$**ACTGGCCGTCGTTTTACA**	Test dGTP incorporation
Poly T	$(T)_{42}$**ACTGGCCGTCGTTTTACA**	Test dATP incorporation
DNA2	TTTGTAGAGCTCATCCATGCCATGTGTAATCCCAGCA GCAGTTACAAACTCAAGAAGGACCATGTGGTCACGCT TTTCGTTGGGATCTTTCGAAAGGGCAGATTGTGTCGA CAGGTAATGGTTGTCTGGTAAAAGGACAGGGCCATCG CCAATTGGAGTATTTTGTTGATAATGGTCTGCTAGTT GAACGGAACCATCTTCAATGTTGTGGCGAATTTTGAA GTTAGCTTTGATTCCATTCTTTTGTTTGTCTGCCGTG ATGTATACATTGTGTGAGTTATAGTTGTACTCGAGTT TGTGTCCGAGAATGTTTCCATCTTCTTTAAAATCAAT ACCTTTTAACTCGATACGATTAACAAGGGTATCACCT TCAAACTTGACTTCAGCACGCGTCTTGTAGTTCCCGT CATCTTTGAAAGATATAGTGCGTTCCTGTACATAACC TTCGGGCATGGCACTCTTGAAAAAGTCATGCCGTTTC ATGTGATCCGGATAACGGGAAAAGCATTGAACACCAT AGGTCAGAGTAGTGACAAGTGTTGGCCATGGAACAGG TAGTTTTCCAGTAGTGCAAATAAATTTAAGGGTGAGT TTTCCGTATGTAGCATCACCTTCACCCTCTCCACTGA CAGAAAATTTGTGCCCATTAACATCACCATCTAATTC AACAAGAATTGGGACAACTCCAGTGAAAAGTTCTTCT CCTTTGCTCGGGATCCCCAGCAGTTGTTTTGCAGGTG CGACGAGCTGTTGCTTGCGTCTATCGCCCGTTGGCTG AATCGGAAGAATCGGCCTAGGACAATAGAGTTCAGCA CGCTTCATGCGCACGAGCAACTCAGTGATTGTTTCTG CTTTAACTGCACCAAAGTTGAACGATGTCGGCAAATG CCGAGCATGCGTCGTCGTTAGGTGAGCCGAATCCCCG CGTGCACTGGCGGTGTAGGTCGTAGTGCCAGTGTACG CCGTAGCCAACACACGGTGTGGCGCGGTGTATGGGAG AGCCAGTCGAGTCACCGGGCCCTTGTGGTAGGCAGTG GGATTGGTTGTGTTGTTAAGTGCAGAAACTGGTGCAC CGTTGGGCACCCATGTGAGCTTCCCAGTGTGGGTCAC TGCTATTTCCAAATCAGAAAAATAGTACGTGGCTGCG CGAAGAAGCGCGCCCACGATATTGTCTTTGTGTGCCT GCATCACGTCGAGTGTGTGTTGGTTATCCGACACTGT GACCTCCACAAACCGGTCAAGAACGAAGGCAACGTCG	Test dNTP incorporation on DNA template > 500 bp

	GTGTGGTGGCGACGTTGGACCTGAGTCTCTCCTCCGT AGTTCTCAACGGTAGTGGTGACGGGGTCAGCAGATTC ACCAGTGGTCGTAGTTTCCATGGTCTGTTTCCTGTGT GAAATTGTTATCCGCTCACAATTCCACACATTATACG AGCCGGATGATTAATTGTCAACAGCTCATTTC**AGAAT ATTTGCCAGAACCGTTATGATGTCGGCGCAAAA**	
DNA3	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAAAATTTTT TTTTTTTTTTTTTTTTTTTTTTTTTTTTTT**ACTGGCCGTCG TTTTACA**	Test incorporation on Poly T template containing four consecutive adenines
DNA4	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAAAAATTTTT TTTTTTTTTTTTTTTTTTTTTTTTTTTTTT**ACTGGCCGTCG TTTTACA**	Test incorporation on Poly T template containing five consecutive adenines
DNA5	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAAAAAATTTTT TTTTTTTTTTTTTTTTTTTTTTTTTTTTTT**ACTGGCCGTCG TTTTACA**	Test incorporation on Poly T template containing six consecutive adenines
DNA6	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAAAAAAAAA AAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT **ACTGGCCGTCGTTTTACA**	Test incorporation on Poly T template containing twelve consecutive adenines
DNA7	AAACTTTAACAGTAGACAATATCTCGGCGATGGCGATGAAA AAAGCAAGGGACAACAGAATGAGCACAGCAAGACAAACCGC **ACTGGCCGTCGTTTTACA**	Test incorporation on DNA template based on a five consecutive adenosine section in the Φ29 bacteriophage genome
DNA8	AAACTTTAACAGTAGACAATATCTCGGCGATGGCGATGCTC TCTGCAAGGGACAACAGAATGAGCACAGCAAGACAAACCGC **ACTGGCCGTCGTTTTACA**	Test incorporation on DNA template based on the Φ29 bacteriophage genome containing consecutive adenosines replaced with $(TC)_3$
DNA9	AGTAGTTTGTAACAAAAATTCGATGACATAAAAAAAAAACA ATTAAGGACGTTTTGTCTGTTGAGGAAATT**ACTGGCCGTCG TTTTACA**	Test incorporation on DNA template based on a ten consecutive adenosine section in the *B. subtillis* genome
DNA10	AGTAGTTTGTAACAAAAATTCGATGACATTTTTTTTTTTCA ATTAAGGACGTTTTGTCTGTTGAGGAAATT**ACTGGCCGTCG TTTTACA**	Test incorporation on DNA template based on the *B. subtillis* genome containing consecutive adenosines replaced with $(T)_{10}$

B. Ensemble assay for Φ29 Activity

To confirm activity of Φ29, a previously described assay was adapted as follows.[27] DNA templates (**Table 3-5**) were polymerized by Φ29 after hybridization to an M13F primer (5' TGTAAAACGACGGCCAGT 3'). Positive control reactions contained Φ29 (1 μM), dNTPs (100 μM), and DNA template-primer (5 μM) in 1x Φ29 Reaction Buffer (330 mM Tris-CH_3COOH, 100 mM $MgCH_3COOH$, 660 mM KCH_3COOH, 1% (v/v) Tween 20, 10 mM DTT, pH 7.9). Negative control reactions omitted either Φ29 or the native dNTPs. Reactions were stored at 30 °C for 1 h in a thermal cycler then quenched with a 2X stop solution (8 M Urea, 0.5 M EDTA, 9 mM Tris-HCl, 9 mM Boric acid) at the indicated times. The reactions (4.5 μL) were mixed with DNA Loading Dye (5x, 1 μL) and prestained with

SYBR Green I Nucleic Acid Gel Stain (Thermo Scientific) prior to visualization on 5% high-resolution agarose gel. The GeneRuler Ultra Low Range DNA Ladder (Thermo Scientific) was used to estimate nucleic acid sizes. Following electrophoresis (120 V for 35 min) in 1X TAE Buffer (40 mM Tris base, 20 mM Acetic acid, 1 mM EDTA), the gel was imaged using UV-light (Digital camera E-M5 Mark II, Olympus, USA and Multi Doc-it Digital Imaging System, Analytik Jena, USA) or fluorescent light (Typhoon Trio Variable Mode Imager System, GE Healthcare, USA; fluorescence acquisition mode, 526 nm, 10 microns). Representative 5% high resolution agarose gel of an ensemble assay time-course is portrayed in **Figure 3-15**.

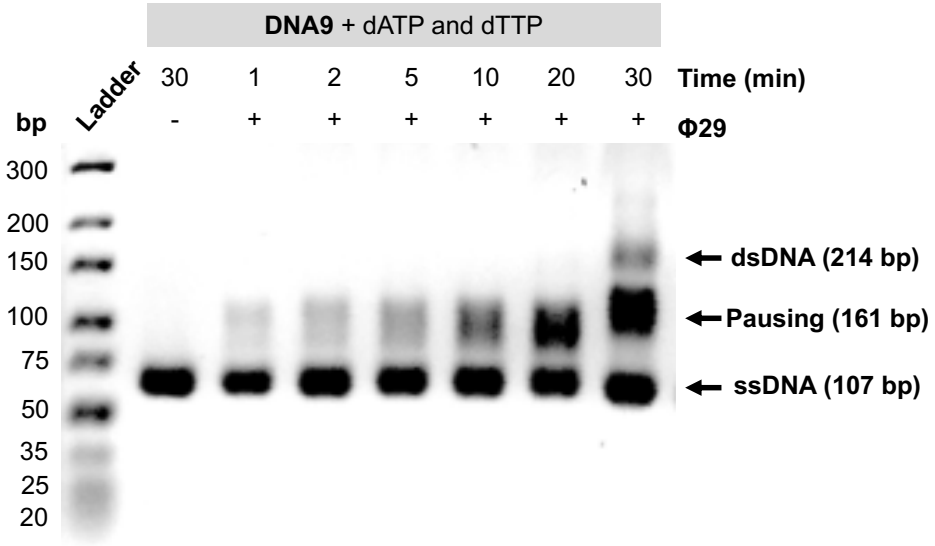


Figure 3-15. Representative 5% high resolution agarose gel of an ensemble assay time-course showing incorporation of dATP and dTTP into DNA9 template described in Table 3-5. Amplified double-stranded DNA (dsDNA) migrates slower than single-stranded DNA (ssDNA) on a 5% high resolution agarose gel stained with SYBR green I nucleic acid. The negative control reaction omitted Φ29. The positive control reaction with dATP and dTTP showed conversion to dsDNA with **DNA9** template.

Attachment of bioconjugate to nanocircuits

A nanocircuit was fabricated using techniques described previously.[46] In brief, the nanocircuit was grown on SiO_2, connected to source and drain metal electrodes and passivated with a polymer (PMMA). The nanocircuit was incubated at 275 °C in air for 30 min, soaked in PG remover at 65 °C for 10 min, rinsed with IPA and DI water, then placed on a hot plate at 135 °C for 5 min. Attachment of Φ29 variants to nanocircuits was accomplished by soaking the devices in a solution of **46** (100 pM) and 1 μM of pyrene in ethanol for 5 min, then rinsing with ethanol, IPA, DI water, and activity buffer (40 mM HEPES, 300 mM NaCl, 10 mM $MgCl_2$, 0.1 mM TCEP, pH 6.5) for 5 min. Following a rinse with DI water, a solution of 4 nM of Φ29 in activity buffer was incubated at RT for 5 min. Such devices are referred to simply as Φ29-SWCNTs.

Protein gels

The protein sample (16 μL) was mixed with SDS-PAGE loading dye (5x, 4 μL). The mixture was heated at 95 °C for 5 min, then 10 μL was visualized by SDS-PAGE with 12% polyacrylamide concentration. The PageRuler Plus Prestained Protein Ladder (ThermoFisher) were used to estimate protein sizes. After electrophoresis (80 V for 15 min, then 180 V for 30 min) in 1x SDS Running Buffer (25 mM Tris-base, 192 mM glycine, 0.1% SDS, pH 8.3), the gel was stained with Coomassie blue (3.2 mM brilliant blue, 45% ethanol, 10% glacial acetic acid in deionized water) for 30 min and destained using destaining buffer (40% methanol, 10% acetic acid in deionized water) at room temperature overnight.

3.6 References

(1) Patel, P. H.; Suzuki, M.; Adman, E.; Shinkai, A.; Loeb, L. A. Prokaryotic DNA Polymerase I: Evolution, Structure, and "Base Flipping" Mechanism for Nucleotide Selection. *J. Mol. Biol.* **2001**, *308*, 823–837.

(2) Blanco, L.; Bernad, A.; Lázaro, J. M.; Martín, G.; Garmendia, C.; Salas, M. Highly Efficient DNA Synthesis by the Phage Phi 29 DNA Polymerase. Symmetrical Mode of DNA Replication. *J. Biol. Chem.* **1989**, *264*, 8935–8940.

(3) Schwartz, J. J.; Quake, S. R. Single Molecule Measurement of the "speed Limit" of DNA Polymerase. *Proc. Natl. Acad. Sci.* **2010**, *107*, 1254–1254.

(4) Esteban, J. A.; Salas, M.; Blanco, L. Fidelity of Φ29 DNA Polymerase. Comparison between Protein-Primed Initiation and DNA Polymerization. *J. Biol. Chem.* **1993**, *268*, 2719-26.

(5) Blanco, L.; Salas, M. Relating Structure to Function in Phi29 DNA Polymerase. *Biochemistry* **1996**.

(6) Tindall, K. R.; Kunkel, T. A. Fidelity of DNA Synthesis by the Thermus Aquaticus DNA Polymerase. *Biochemistry* **1988**, *27*, 6008–6013.

(7) Gyawali, P.; GC, K.; Ma, Y.; Abeysirigunawardena, S.; Nagasawa, K.; Balci, H. Impact of Small Molecules on Intermolecular G-Quadruplex Formation. *Molecules* **2019**, *24*, 1570.

(8) Lipps, H. J.; Rhodes, D. G-Quadruplex Structures: In Vivo Evidence and Function. *Trends Cell Biol.* **2009**, *19*, 414–422.

(9) Bochman, M. L.; Paeschke, K.; Zakian, V. A. DNA Secondary Structures: Stability and Function of G- Quadruplex Structures. *Nat. Rev. Genet.* **2012**, *13*, 770–780.

(10) Mytelka, D. S.; Chamberlin, M. J. Analysis and Suppression of DNA Polymerase Pauses Associated with a Trinucleotide Consensus. *Nucleic Acids Res.* **1996**, *24*, 2774–2781.

(11) Chabes, A.; Georgieva, B.; Domkin, V.; Zhao, X.; Rothstein, R.; Thelander, L. Survival of DNA Damage in Yeast Directly Depends on Increased DNTP Levels Allowed by Relaxed Feedback Inhibition of Ribonucleotide Reductase. *Cell* **2003**, *112*, 391–401.

(12) Balbo, P. B.; Wang, E. C. W.; Tsai, M. D. Kinetic Mechanism of Active Site Assembly and Chemical Catalysis of DNA Polymerase β. *Biochemistry* **2011**, *50*, 9865–9875.

(13) Datta, K.; Wowor, A. J.; Richard, A. J.; Licata, V. J. Temperature Dependence and Thermodynamics of Klenow Polymerase Binding to Primed-Template DNA. *Biophys. J.* **2006**, *90*, 1739–1751.

(14) Ralec, C.; Henry, E.; Lemor, M.; Killelea, T.; Henneke, G. Calcium-Driven DNA Synthesis by a High-Fidelity DNA Polymerase. *Nucleic Acids Res.* **2017**, *45*, 12425–12440.

(15) Min, W.; English, B. P.; Luo, G.; Cherayil, B. J.; Kou, S. C.; Xie, X. S. Fluctuating Enzymes: Lessons from Single-Molecule Studies. *Acc. Chem. Res.* **2005**, *38*, 923–931.

(16) Ibarra, B.; Chemla, Y. R.; Plyasunov, S.; Smith, S. B.; Lá Zaro, J. M.; Salas, M.; Bustamante, C. Proofreading Dynamics of a Processive DNA Polymerase. *EMBO J.* **2009**, *28*, 2794–2802.

(17) Morin, J. A.; Cao, F. J.; Valpuesta, J. M.; Carrascosa, J. L.; Salas, M.; Ibarra, B. Manipulation of Single Polymerase-DNA Complexes: A Mechanical View of DNA Unwinding during Replication. *Cell Cycle* **2012**, *11*, 2967–2968.

(18) Morin, J. A.; Cao, F. J.; Lázaro, J. M.; Arias-Gonzalez, J. R.; Valpuesta, J. M.; Carrascosa, J. L.; Salas, M.; Ibarra, B. Active DNA Unwinding Dynamics during Processive DNA Replication. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 8115–8120.

(19) Cherf, G. M.; Lieberman, K. R.; Rashid, H.; Lam, C. E.; Karplus, K.; Akeson, M. Automated Forward and Reverse Ratcheting of DNA in a Nanopore at Five Angstrom Precision. *Nat Biotechnol.* **2012**, *30*, 130.

(20) Lieberman, K. R.; Dahl, J. M.; Mai, A. H.; Cox, A.; Akeson, M.; Wang, H. Kinetic Mechanism of Translocation and DNTP Binding in Individual DNA Polymerase Complexes. *J. Am. Chem. Soc.* **2013**, *135*, 9149–9155.

(21) Manrao, E. A.; Derrington, I. M.; Laszlo, A. H.; Langford, K. W.; Hopper, M. K.; Gillgren, N.; Pavlenok, M.; Niederweis, M.; Gundlach, J. H. Reading DNA at Single-Nucleotide Resolution with a Mutant MspA Nanopore and Phi29 DNA Polymerase. *Nat. Biotechnol.* **2013**, *10*, 54–56.

(22) Eid, J.; Fehr, A.; Gray, J.; Luong, K.; Lyle, J.; Otto, G.; Peluso, P.; Rank, D.; Baybayan, P.; Bettman, B.; Bibillo, A.; Bjornson, K.; Chaudhuri, B.; Christians, F.; Cicero, R.; Clark, S.; Dalal, R.; Dixon, J.; Foquet, M.; Gaertner, A.; Hardenbol, P.; Heiner, C.; Hester, K.; Holden, D.; Kearns, G.; Kong, X.; Kuse, R.; Lacroix, Y.; Lin, S.; Lundquist, P.; Ma, C.; Marks, P.; Maxham, M.; Murphy, D.; Park, I.; Pham, T.; Phillips, M.; Roy, J.; Sebra, R.; Shen, G.; Sorenson, J.; Tomaney, A.; Travers, K.; Trulson, M.; Vieceli, J.; Wegener, J.; Wu, D.; Yang, A.; Zaccarin, D.; Zhao, P.; Zhong, F.; Korlach, J.; Turner, S. Real-Time DNA Sequencing from Single Polymerase Molecules. *Sci. Rep.* **2009**, *323*, 133–139.

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

(24) Choi, Y.; Moody, I. S.; Sims, P. C.; Hunt, S. R.; Corso, B. L.; Perez, I.; Weiss, G. A.; Collins, P. G. Single-Molecule Lysozyme Dynamics Monitored by an Electronic Circuit. *Science (80-.).* **2012**, *335*, 319–324.

(25) Choi, Y.; Olsen, T. J.; Sims, P. C.; Moody, I. S.; Corso, B. L.; Dang, M. N.; Weiss, G. A.; Collins, P. G. Dissecting Single-Molecle Signal Transduction in Carbon Nanotube Circuits with Protein Engineering. *Nano Lett.* **2013**, *13*, 625–631.

(26) Olsen, T. J.; Choi, Y.; Sims, P. C.; Gul, O. T.; Corso, B. L.; Dong, C.; Brown, W. A.; Collins, P. G.; Weiss, G. A. Electronic Measurements of Single-Molecule Processing by DNA Polymerase i (Klenow Fragment). *J. Am. Chem. Soc.* **2013**, *135*, 7855–7860.

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

(28) Lieberman, K. R.; Cherf, G. M.; Doody, M. J.; Olasagasti, F.; Kolodji, Y.; Akeson, M. Processive Replication of Single DNA Molecules in a Nanopore Catalyzed by Phi29 DNA Polymerase. *J. Am. Chem. Soc.* **2010**.

(29) Quail, M. A.; Smith, M.; Coupland, P.; Otto, T. D.; Harris, S. R.; Connor, T. R.; Bertoni, A.; Swerdlow, H. P.; Gu, Y. A Tale of Three next Generation Sequencing Platforms: Comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq Sequencers. *BMC Genomics* **2012**, *13*, 1.

(30) Ibarra, B.; Chemla, Y. R.; Plyasunov, S.; Smith, S. B.; Lázaro, J. M.; Salas, M.; Bustamante, C. Proofreading Dynamics of a Processive DNA Polymerase. *EMBO J.* **2009**, *28*, 2794–2802.

(31) Soengas, M. S.; Gutiérrez, C.; Salas, M. Helix-Destabilizing Activity of Φ29 Single-Stranded DNA Binding Protein: Effect on the Elongation Rate during Strand Displacement DNA Replication. *J. Mol. Biol.* **1995**, *253*, 517–529.

(32) Soengas, M. S.; Esteban, J. A.; Lazaro, J. M.; Bernad, A.; Blasco, M. A.; Salas1, M.; Blanco, L. Site-Directed Mutagenesis at the Exo III Motif of Phi29 DNA Polymerase; Overlapping Structural Domains for the 3'-5' Exonuclease and Strand-Displacement Activities. *EMBO J.* **1992**, *1*, 4227–4237.

(33) de Vega, M.; Lazaro, J. M.; Salas, M.; Blanco, L. Primer-Terminus Stabilization at

the 3′-5′ Exonuclease Active Site of Phi29 DNA Polymerase. Involvement of Two Amino Acid Residues Highly Conserved in Proofreading DNA Polymerases. *EMBO J.* **1996**, *15*, 1182–1192.

(34) Huang, C.-C.; Hearst, J. E. Pauses at Positions of Single-Stranded of Secondary Structure during in Vitro Replication Fd Bacteriophage DNA by T4 DNA Polymerase. *Anal. Biochem.* **1980**, *103*, 127–139.

(35) Kang, S.; Ohshima, K.; Shimizu, M.; Amirhaeri, S.; Wells, R. D. Pausing of DNA Synthesis in Vitro at Specific Loci in CTG and CGG Triplet Repeats from Human Hereditary Disease Genes. *J. Biol. Chem.* **1995**, *270*, 27014–27021.

(36) Samadashwily, G. M.; Raca, G.; Mirkin, S. M. Trinucleotide Repeats Affect DNA Replication in Vivo. *Nat. Genet.* **1997**, *17*, 298–304.

(37) Schwartz, J. J.; Quake, S. R. Single Molecule Measurement of the “Speed Limit” of DNA Polymerase. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 20294–20299.

(38) Schroeder, C. M.; Xie, X. S.; Kim, S. Single-Molecule Study of DNA Polymerization Activity of HIV-1 Reverse Transcriptase on DNA. *J. Mol. Biol.* **2010**, *395*.

(39) Snyder, T. M.; Tse, B. N.; Liu, D. R. Effects of Template Sequence and Secondary Structure on DNA- Templated Reactivity Thomas. *J. Am. Chem. Soc.* **2008**, *130*, 1392–1401.

(40) Goddard, N. L.; Bonnet, G.; Krichevsky, O.; Libchaber, A. Sequence Dependent Rigidity of Single Stranded DNA. *Phys. Rev. Lett.* **2000**, *85*, 2400–2403.

(41) Laduca, R. J.; Fay, P. J.; Charles, S.; Bambara, R. A. Site-Specific Pausing of Deoxyribonucleic Acid Synthesis Catalyzed by Four Forms of Escherichia Coli DNA Polymerase III. **1983**, 5177–5188.

(42) Berman, A. J.; Kamtekar, S.; Goodman, J. L.; Lázaro, J. M.; de Vega, M.; Blanco, L.; Salas, M.; Steitz, T. A. Structures of Phi29 DNA Polymerase Complexed with Substrate: The Mechanism of Translocation in B-Family Polymerases. *EMBO J.* **2007**, *26*, 3494–3505.

(43) Blasco, M. A.; Blanco, L.; Pares, E.; Salas+, M.; Bernad, A. Structural and Functional Analysis of Temperature-Sensitive Mutants of the Phage F29 DNA Polymerase. *Nucleic Acids Res. 18*.

(44) Zhang, B.; DENG, H.; Mukherjee, S.; Song, W.; Wang, X.; Lindsay, S. Engineering an Enzyme for Direct Electrical Monitoring of Activity. *ACS Nano* **2019**.

(45) Kamtekar, S.; Bibillo, A.; Bjornson, K.; Christians, F.; Cutcliffe, C.; Hanes, J.; Jia, L.; Lee, W.; Miller, E.; Patel, P. Recombinant Polymerases with Increased

Phototolerance. US8906660B2, 2014.

(46) Olsen, T. J.; Choi, Y.; Sims, P. C.; Gul, O. T.; Corso, B. L.; Dong, C.; Brown, W. A.; Collins, P. G.; Weiss, G. A. Electronic Measurements of Single-Molecule Processing by DNA Polymerase I (Klenow Fragment). *J. Am. Chem. Soc.* **2013**, *135*, 7855–7860.

CHAPTER 4: Pyrocinchonimides Conjugate to Amine Groups on Proteins via Imide Transfer

4.1 Abstract

Advances in bioconjugation, the ability to link biomolecules to each other, small molecules, surfaces, and more, can spur the development of advanced materials and therapeutics. We have discovered that pyrocinchonimide, the dimethylated analog of maleimide, undergoes a surprising transformation with biomolecules. The reaction targets amines and involves an imide transfer, which has not been previously reported for bioconjugation purposes. Despite their similarity to maleimides, pyrocinchonimides do not react with free thiols. Though both lysine residues and the N-termini of proteins can receive the transferred imide, the reaction also exhibits a marked preference for certain amines that cannot solely be ascribed to solvent accessibility. This property is peculiar among amine-targeting reactions and can reduce combinatorial diversity when many available reactive amines are available, such as in the formation of antibody-drug conjugates. Unlike amides, the modification undergoes very slow reversion under high pH conditions. The reaction offers a thermodynamically controlled route to single or multiple modifications of proteins for a wide range of applications.

Key reference: Richardson, M. B.*; Gabriel, K. N.*; Garcia, J.; Ashby, S.; Dyer, R.; Kim, J.; Lau, C.; Hong, J.; Le Tourneau, R. J.; Sen, S.; Narel, D.; Katz, B. B.; Ziller, J. W.; Majumdar, S.; Collins, P. G.; Weiss, G. A. Pyrocinchonimides Conjugate to Amine Groups on Proteins via Imide Transfer. *Bioconjug. Chem.* **2020**, *31*, 1449–1462.

*Authors contrubuted equally. My contribution was designing experiments, performing experiments, analyzing data, and writing the manuscript for the protein bioconjugation research.

4.2 Introduction

The thiol-maleimide click reaction is an important mainstay in bioconjugation chemistry. Typically, the maleimide nitrogen atom carries a payload molecule chosen for experimental or therapeutic purposes. The reaction proceeds via nucleophilic attack by a sulfur atom at the maleimide alkene in a classical 1,4-conjugate addition (or "thio-Michael") mechanism. This reaction covalently links the payload to the target under mild conditions in aqueous buffers.[1,2] The product of the thiol-maleimide reaction is an unstable thiosuccinimide, which can undergo reversion via a retro-Michael process. In situ, the regenerated maleimide can then react with other thiols, shuffling the payload to off-target sites, introducing a potential source of toxicity, and decreasing selectivity for therapeutic bioconjugates.[3–5] Thiosuccinimides are also susceptible to irreversible hydrolytic ring-opening reactions in aqueous conditions, forming long-lived succinamic acid thioethers. [3,4,6–10] These and other impish behaviors have inspired detailed investigations,[11,12] especially in the context of maleimides used in preparing antibody-drug conjugates (ADCs).[13,14]

Next-generation maleimides (NGMs) are maleimide derivatives designed to attenuate the issues described above. The archetypal NGMs, halomaleimides, feature a halogen substituent at one or both alkene carbons of the maleimide.[15–17] Halomaleimides undergo an overall addition-elimination reaction; after the canonical thio-Michael step, the intermediate halosuccinimide rapidly eliminates HX, reforms the hydrolytically stable maleimide heterocycle, and generates a vinylic S-C covalent bond. Dihalo-, dithio-, and diaryloxymaleimides have also been developed for cross-linking reactions.[18–21] This technology has been successfully employed for re-bridging reduced interchain antibody

disulfides, and arming antibodies with the anticancer drug, monomethyl auristatin E[22] or the photodynamic, theranostic fluorophore IRDye 800 CW.[23]

The mono-methylated analog of maleimide ("citraconimide") has also been reported for the assembly of doxorubicin-Fab′ ADCs,[24] and may be considered a member of the NGM family. The citraconimide thio-Michael addition reaction exhibits excellent regioselectivity for nucleophilic attack at the methylated alkene carbon,[25,26] and the product methylsuccinimides exhibit improved hydrolytic stability compared to standard succinimides.[24] 5-Methylene pyrrolones are maleimide =O → $=CH_2$ isosteres that exhibit soft electrophilic character at the exocyclic γ/δ-alkene, and can participate in 1,6-conjugate addition reactions with thiols. The adduct gains no additional stereocenters and can tracelessly revert to starting materials at elevated pH, thus providing a pH-controlled catch-and-release process.[27] Another class of maleimide isosteres, 1,2-dihydropyridazine-3,6-diones undergo a similar retro-Michael reaction for cargo release.[28]

Examined herein is an ostensibly overlooked NGM reagent, pyrocinchonimide (Pci), the dimethylated analog of maleimide. The Pci motif (also referred to as dimethylmaleimide or N-dimethylmaleoyl) was previously established as an N-protecting group for aminosugar synthesis by Richard R. Schmidt.[29] Prominent examples of Pci compounds in the patent literature include antimicrobials,[30–33] water soluble[34] and photochemically cross-linked polymers,[35] silicon-based adhesives,[36] herbicides and other plant growth regulators and their synthetic intermediates,[37–44] pharmaceuticals and prophylactics for treatment of obesity[45] and urinary incontinence,[46] liquid crystal thermoset monomers,[47] synthetic intermediates of allosteric glucokinase modulators[48] and alpha-2C adrenergic receptor antagonists,[49] internal ribosome entry site-mediated

protein synthesis inhibitors,[50] and in oligosaccharide manufacturing.[51] Despite this considerable precedent, the reactivity of Pci compounds with proteins has never been specifically addressed.

The investigation revealed that the reactivity of Pci towards thiols is profoundly diminished compared to maleimide and citraconimide. Fluorescence and mass spectrometry (MS) labeling experiments with proteins revealed only trace quantities of the expected thio-Michael adduct. Instead, a previously unreported imide transfer reaction was observed (i.e., a transpyrocinchonimidation); both acyl groups are moved from the electron-poor nitrogen atoms of select Pci derivatives, to the solvent-exposed amines of lysine side chains and *N*-termini on target proteins. This imide transfer was then leveraged for direct bioconjugation of a fluorescent label to a target protein under mild conditions using a modified pyrocinchonimide derivative. Additionally, we have explored the possibility for protein-bound Pci groups to participate in UV-activated intermolecular cycloadditions with terminal alkenes. The results establish Pci compounds as a new species of amine-targeting bioconjugation reagents, with qualities that make it distinct from NHS esters, maleimides, and other NGMs (**Figure 4-1**).

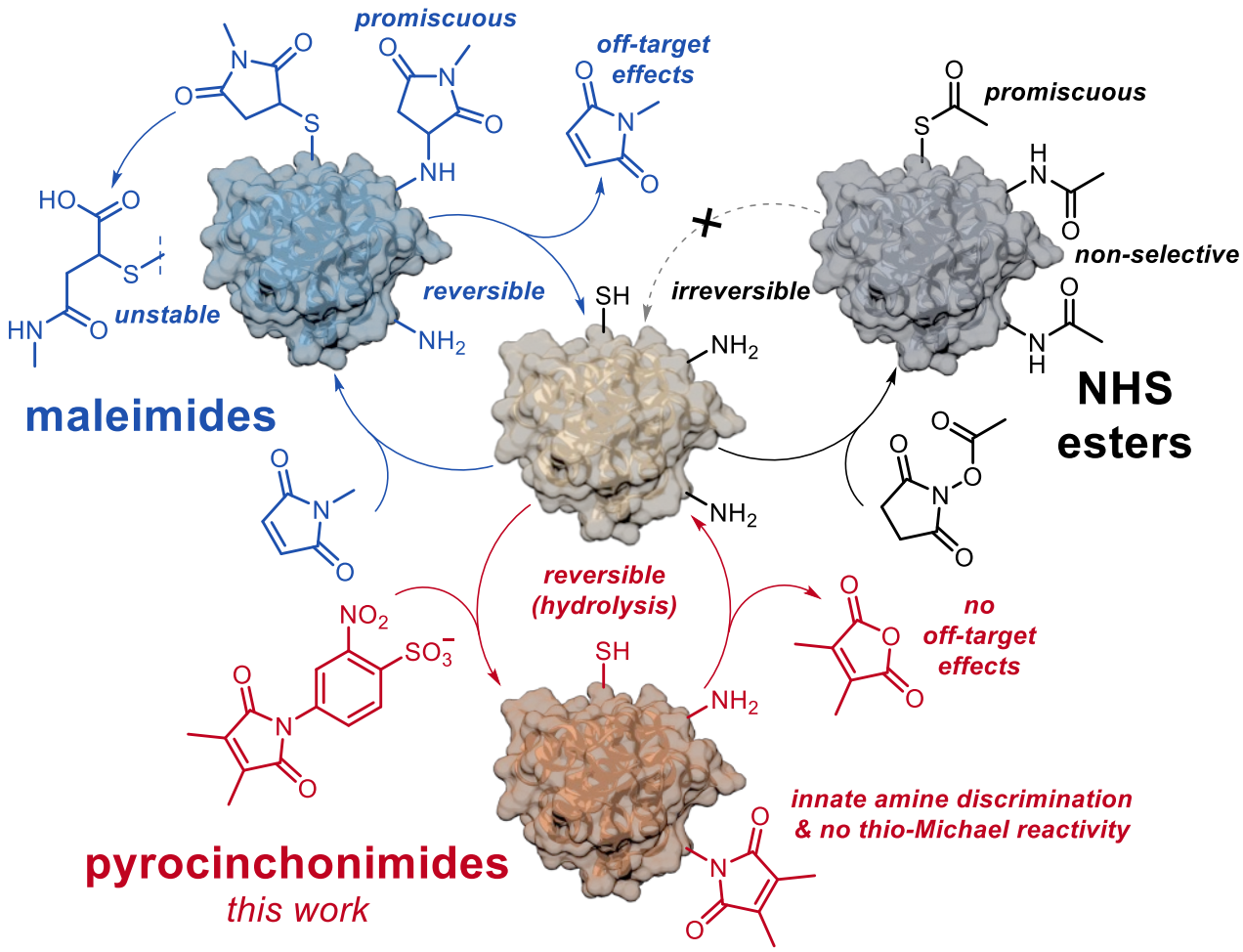


Figure 4-1. The unique modality of pyrocinchonimides. The behavior of the Pci bioconjugation is compared with its two closest analogs, maleimides and NHS esters.

4.3 Results and Discussion

Synthesis of Pyrocinchonimides

The most common method for preparing N-substituted maleimides is via the reaction of 1° amines with maleic anhydride. Ring-opening acylation of the amine results in an isolable maleamic acid intermediate, which can be cyclized under dehydrating conditions. This two-step sequence is equally appropriate for citraconimides, but not Pci compounds. At least two independent sources have reported that the condensation of 1° amines with pyrocinchonic anhydride furnishes the cyclic imide without observation of the

intermediate pyrocinchonamic acid (**Figure 4-2**).[52,53] However, the existence of the pyrocinchonamic acid intermediate is well-established spectroscopically, and the species can be trapped as the carboxylate triethylammonium salt.[40,54]

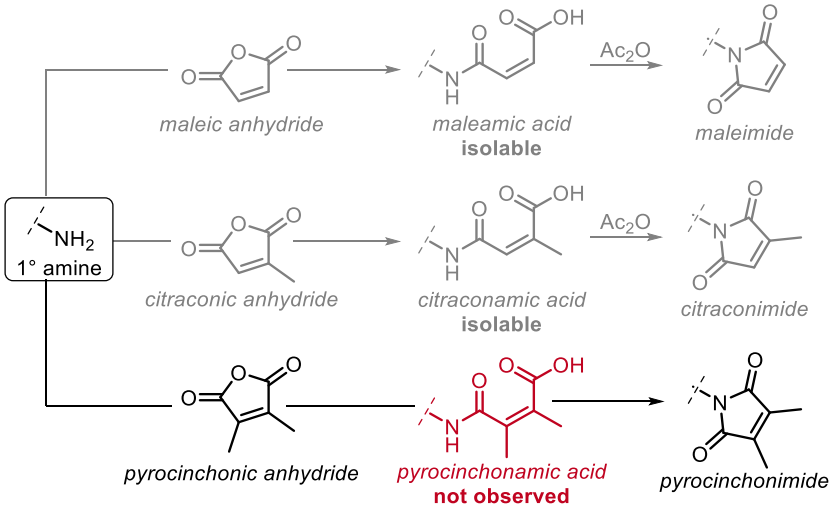
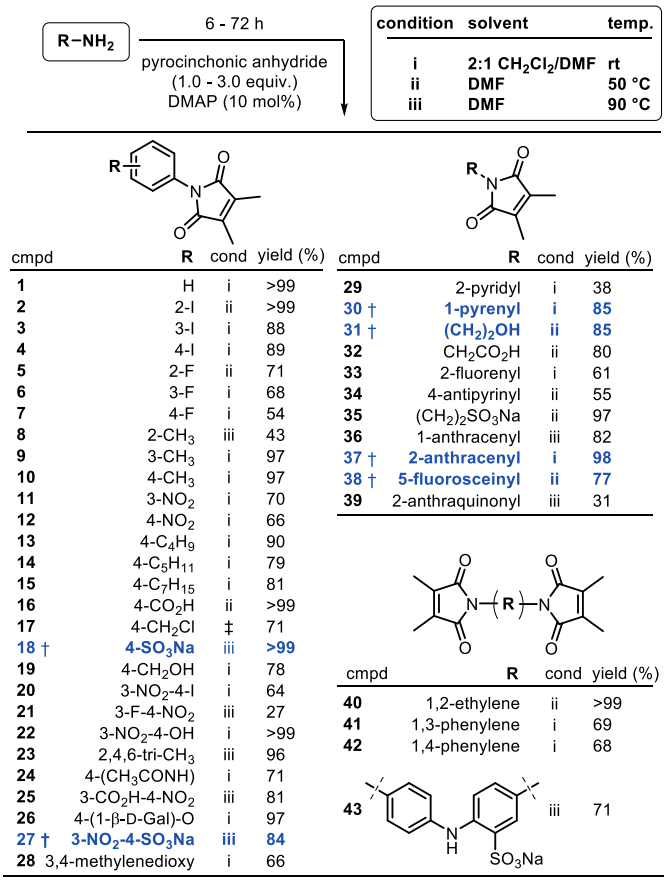


Figure 4-2. Facile cyclization of pyrocinchonimides. In contrast to the reactions of 1° amines with maleic and citraconic anhydrides, the ring-opened pyrocinchonamic acid intermediate spontaneously cyclizes to afford the imide at ambient temperatures without requiring dehydrating conditions.

The disparity in cyclization efficiency between citraconimide and pyrocinchonimide is consistent with the behaviors of their parent diacids. Both maleic (CAS: 110-16-7) and citraconic (CAS: 498-23-7) acids have been isolated and characterized as pure compounds. However, pyrocinchonic acid undergoes rapid cyclization to its anhydride during attempted isolation, and a sample of the purified acid has never been described.[55] The formation of pyrocinchonic anhydride exhibits immeasurably fast ring-closing kinetics in organic solvents, reflecting a profound entropic preference for the cyclized state.[56,57] This behavior has been ascribed to angle compression (the Thorpe-Ingold effect); the mutual repulsion of the vinylic methyl groups drives the carboxylic acids close together, and accelerates the ring closing step.[58,59] The same stereoelectronic and steric effects

could operate in the amic acid, resulting in the apparent absence of this intermediate during the synthesis of Pci derivatives. Thus, mild conditions should suffice for Pci synthesis, despite many reports applying the strongly dehydrating conditions typical for citraconimide and maleimide formation.

Guided by the reports described above, a general methodology provided relatively effortless access to a wide variety of Pci derivatives in satisfactory-to-excellent yields with respect to the amine starting material (**Figure 4-3**). Briefly, a 1° aryl- or alkylamine was combined with catalytic 4-dimethylaminopyridine (DMAP) and excess pyrocinchonic anhydride in a mixture of CH_2Cl_2 and DMF. The reaction was typically judged complete by thin layer chromatography (TLC) after stirring at rt for 6-72 h (condition i). Anhydrous DMF is not required for these syntheses. In general, the components of the product mixture were easily separated by flash chromatography, and unreacted pyrocinchonic anhydride can be recovered for use in future reactions. Alcohol substituents are tolerated (i.e., ester formation does not occur), as are the sodium salts of carboxylic and sulfonic acids. Electron poor and sterically demanding amines can result in very slow reactions; however, Pci derivatives for challenging starting materials could still be achieved in reasonable time periods by omitting the CH_2Cl_2 and increasing the reaction temperature (condition ii at 50 °C and condition iii at 90 °C).



cmpd	R	cond	yield (%)
1	H	i	>99
2	2-I	ii	>99
3	3-I	i	88
4	4-I	i	89
5	2-F	ii	71
6	3-F	i	68
7	4-F	i	54
8	2-CH_3	iii	43
9	3-CH_3	i	97
10	4-CH_3	i	97
11	3-NO_2	i	70
12	4-NO_2	i	66
13	4-C_4H_9	i	90
14	4-C_5H_{11}	i	79
15	4-C_7H_{15}	i	81
16	4-CO_2H	ii	>99
17	4-CH_2Cl	‡	71
18 †	**4-SO_3Na**	iii	**>99**
19	4-CH_2OH	i	78
20	3-NO_2-4-I	i	64
21	3-F-4-NO_2	iii	27
22	3-NO_2-4-OH	i	>99
23	2,4,6-tri-CH_3	iii	96
24	4-(CH_3CONH)	i	71
25	3-CO_2H-4-NO_2	iii	81
26	4-(1-β-D-Gal)-O	i	97
27 †	**3-NO_2-4-SO_3Na**	**iii**	**84**
28	3,4-methylenedioxy	i	66

cmpd	R	cond	yield (%)
29	2-pyridyl	i	38
30 †	**1-pyrenyl**	**i**	**85**
31 †	**(CH_2)$_2$OH**	**ii**	**85**
32	CH_2CO_2H	ii	80
33	2-fluorenyl	i	61
34	4-antipyrinyl	ii	55
35	(CH_2)$_2SO_3Na$	ii	97
36	1-anthracenyl	iii	82
37 †	**2-anthracenyl**	**i**	**98**
38 †	**5-fluorosceinyl**	**ii**	**77**
39	2-anthraquinonyl	iii	31

cmpd	R	cond	yield (%)
40	1,2-ethylene	ii	>99
41	1,3-phenylene	i	69
42	1,4-phenylene	i	68

43 iii 71

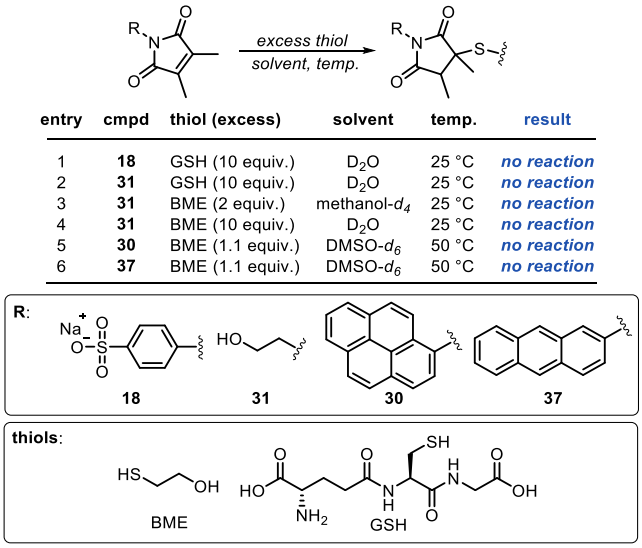
Figure 4-3. Synthesis of N-substituted Pci compounds. †Compounds **18**, **27**, **30**, **31**, **37**, and **38** (in blue) are the subject of bioconjugation experiments reported in the following sections. ‡Synthesized from 19 via reaction with MsCl/Et_3N. The identities of **18** and **27** were unambiguously confirmed by single crystal X-ray diffraction analysis (Supporting Information).

In the following sections, the ethanolamine derivative **31**, polyaromatic derivatives **30** and **37**, and the fluorescein derivative **38** were used as probes to study the surprisingly reluctant thio-Michael addition of pyrocinchonimides. Then, experiments involving the sulfanilic acid derivatives **18** and **27** establish a previously unreported transpyrocinchonimidation reaction with proteins.

Pyrocinchonimides Do Not Bioconjugate to Free Thiols

Pci participation in thio-Michael reactions appears sporadically in the prior literature. The N_ε-Pci derivative of N_α-acetyl lysine was reported to undergo a thio-Michael reaction with methyl 3-mercaptopropionate and N_α-acetylcysteine methyl ester.[60] Reactions of glutathione (GSH) with *N*-(2-hydroxyethyl)pyrocinchonimide (**32**) and *N*-phenylpyrocinchonimide (**1**) have also been reported.[26,61] However, in at least one counter example, no reaction was observed after treatment of a Pci derivative with GSH.[62] No spectra are provided for the thio-Michael Pci adducts in these prior reports.

In our hands Pci compounds are *not* reactive towards thiols. NMR experiments failed to detect any transformation of the Pci groups of **30** and **37** after prolonged exposure (3 h) to 10 mol% excess ß-mercaptoethanol (BME) at elevated temperatures (50 °C) in DMSO-d_6. In contrast, conventional maleimides were consumed within minutes under identical conditions (**Figure 4-4** and **Figure 4-5**). Further NMR experiments with GSH and water-soluble Pci derivatives **17** and **31** established the general lack of reactivity to be independent of the Pci *N*-substituent or choice of solvent. The rate constant for pyrocinchonimide reactions with thiols has been previously shown to be slowed by a factor of 3.15×10^5 compared to maleimides.[26] However, we have failed to observe any appreciable formation of thiol-Pci adducts, even in the presence of large excess of free thiol for extended periods. It is possible that the conjugate addition could occur at extreme temperatures or pH, outside of acceptable ranges for bioconjugation.



entry	cmpd	thiol (excess)	solvent	temp.	result
1	**18**	GSH (10 equiv.)	D_2O	25 °C	*no reaction*
2	**31**	GSH (10 equiv.)	D_2O	25 °C	*no reaction*
3	**31**	BME (2 equiv.)	methanol-d_4	25 °C	*no reaction*
4	**31**	BME (10 equiv.)	D_2O	25 °C	*no reaction*
5	**30**	BME (1.1 equiv.)	DMSO-d_6	50 °C	*no reaction*
6	**37**	BME (1.1 equiv.)	DMSO-d_6	50 °C	*no reaction*

Figure 4-4. Attempted small molecule thio-Michael reactions. Reactions of Pci compounds **18**, **31**, **30** and **37** with GSH and BME. Additional reactions and positive controls appear in **Figure 4-5**. No reaction indicates undetectable thio- or aza- Michael addition products in an NMR experiment with a cryoprobe-equipped 600 MHz spectrometer.

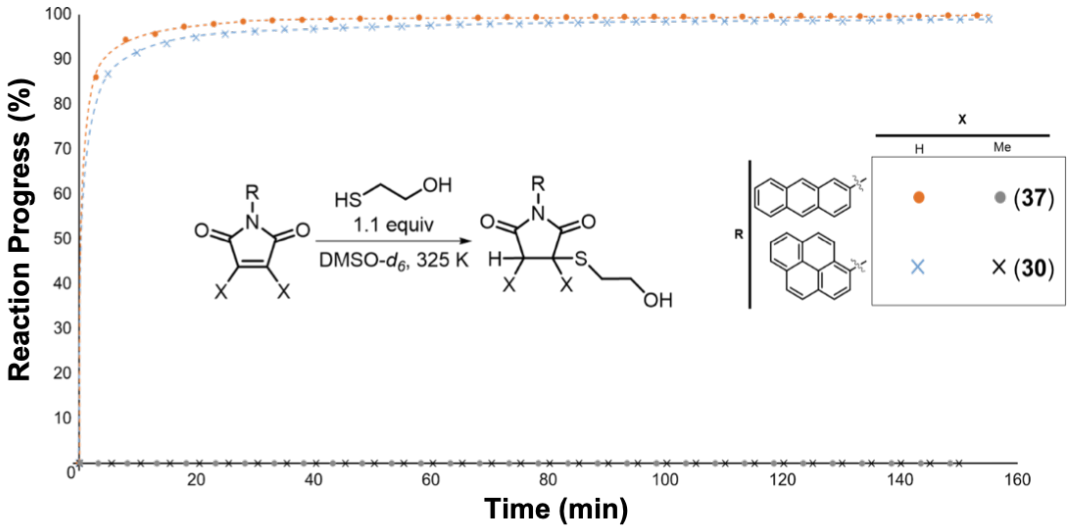


Figure 4-5. Time-course NMR for thio-Michael reactions of β-mercaptoethanol (BME) with N-aryl maleimides and pyrocinchonimides. This time-course NMR illustrates the rapid reaction of BME with conventional maleimides; however, BME exhibits essentially no observable reaction with pyrocinchonimides. The data was collected using a Bruker Avance 600 MHz CryoProbe spectrometer, as follows. A solution of the maleimide, or pyrocinchonimide (**30** or **37**) (400 µL, 150 mM in DMSO-d_6) was combined with a solution of the NMR standard ethylene carbonate (150 µL, 400 mM in DMSO-d_6), in a 0.5 mm OD NMR tube. The sample was heated to 325 K, the receiver coils were tuned, the sample lock and shim were optimized (sample spinning at 20 Hz), and a spectrum at t_0 was collected. Then, the sample was ejected, and a solution of BME (150 µL, 400 mM in DMSO-d_6) was added. The final reaction volume was 700 µL, and final concentrations of

maleimide, ethylene carbonate, and BME were 90.0, 100, and 100 μM, respectively. The sample tube was inverted several times to mix the solution thoroughly, and then returned to the spectrometer. The sample shimming was then quickly reoptimized, and the first spectrum was collected. Single-scan ^{1}H NMR spectra were collected in 5 min intervals. The reaction progress was determined through comparison of the integration values for the maleimide protons (δ 7.3 ppm) or pyrocinchonimide methyl signals (δ 2.0 ppm) relative to ethylene carbonate (δ 4.5 ppm). The concentration of ethylene carbonate in the reaction mixture is assumed to be constant throughout the experiment. No changes in the NMR spectra were observed during this time period for pyrocinchonimide samples **30** and **37**.

The reaction of Pci groups with cysteine residues on proteins was also investigated using single cysteine and cysteine-free variants of dihydrofolate reductase (DHFR-Cys and DHFR) and an exonuclease deficient Klenow fragment of DNA polymerase I (KF-Cys and KF). Fluoroscein derivatives 5-F-Pci (**38**) and its standard maleimide analog 5-F-Mal were chosen as fluorescent reporters, and the bioconjugation result was evaluated by SDS-PAGE (**Figure 4-6** and **Figure 4-7**). The two KF variants readily reacted with 5-F-Mal, and no difference in fluorescence intensity for the Cys and Cys-free variants was observed. The apparent unimportance of a free cysteine residue in the bioconjugation of KF with 5-F-Mal is striking, but not entirely unexpected. Despite their reputation for exclusive thiol reactivity, maleimides are known to react with other nucleophiles, and especially amines.[63–67] 5-F-Pci (**38**) also reacted to a small extent with both KF variants, suggesting a Cys-independent addition occurs at a low level for KF. These 5-F-Pci (**38**) conjugates cannot be detected or identified by mass spectrometry (MS), indicative of their vanishingly small quantities. The low fluorescence of the reaction with 5-F-Pci (**38**), compared to its maleimide analog, is commensurate with its greatly attenuated thiol-maleimide bioconjugation efficiency, and corroborates the results obtained from the small molecule studies.

The reactions of 5-F-Pci (**38**) and 5-F-Mal with DHFR are equally informative. As with KF, the presence of a free cysteine is unnecessary for the formation of bioconjugates with 5-F-Mal, although DHFR presents less non-specific labeling than KF. Remarkably, bioconjugation of DHFR with 5-F-Pci (**38**) appears cysteine-dependent, but the dimness of the band permits only a conservative estimate for this reaction's efficiency, which again failed to yield products detectable by MS. Taken together, the results with both proteins demonstrate Pci compounds, unlike conventional maleimides, are not good candidates for bioconjugation via the canonical thio-Michael reaction. This may not be a surprising result; the two methyl substituents on the Pci heterocycle are expected to contribute approximately -12 kcal·mol^{-1} of stability to the alkene, relative to the unsubstituted maleimide, through hyperconjugative effects alone.[68] This inference can explain the lack of both thio- and aza-Michael addition reactions by Pci compounds.

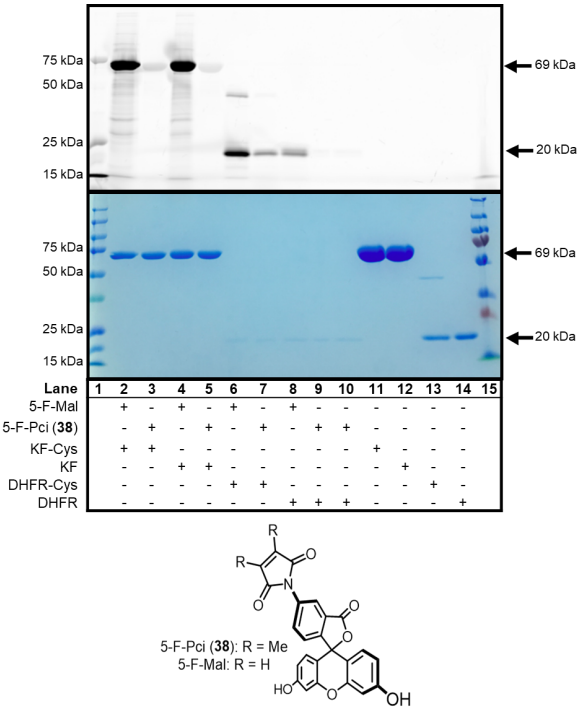


Figure 4-6. Probing the elusive Pci thio-Michael bioconjugation. In this representative 15% SDS-PAGE, KF and DHFR are analyzed after treatment with 5-F-Pci (**38**) or 5-F-Mal for 2 h at pH 7.2 and 25 °C. The protein gels were visualized by a Typhoon scanner with

fluorescence excitation at 256 nm (top gel), then stained with Coomassie brilliant blue as a control for loading and protein purity (lower gel). The reaction of single cysteine proteins with 5-F-Pci (38) uncovered Cys-dependent fluorescence for DHFR-Cys (lanes 7 and 9-10), but not KF-Cys (lanes 3 and 5). Conventional 5-F-Mal yielded efficient fluorescent labeling of single cysteine proteins as expected (lanes 2 and 6), but also extensive Cys independent labeling (lanes 4 and 8). Negative control experiments featured identical conditions without 5-F-Pci (**38**) or 5-F-Mal treatment (lanes 11-14). The precision plus Kaleidoscope standards fluorescent ladder (lane 1) and PageRuler plus prestained protein ladder (lane 15) were used as references.

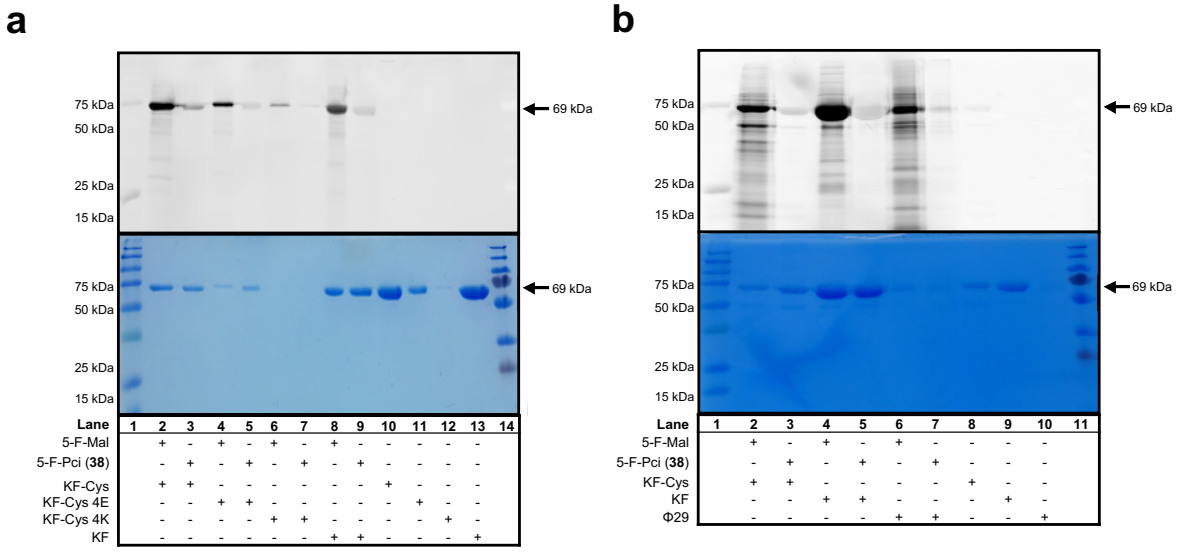


Figure 4-7. Additional examples of the elusive Pci thio-Michael bioconjugation. Two 15% SDS-PAGE show KF analyzed after treatment with thio-Michael reagents 5-F-Pci (**38**) or 5-F-Mal for 2 h at 25 °C. The protein gels were visualized by a Typhoon scanner at 256 nm (top gel), then stained with Coomassie brilliant blue as a control for loading and protein purity (lower gel). Positive control conditions featured **A)** single-cysteine KF variants (KF-Cys 4E and KF-Cys 4K) or B**)** Phi29 DNA polymerase (Φ29) from Phi29 bacteriophage with seven native cysteines. The reaction of proteins with 5-F-Pci (**38**) uncovered Cys-independent fluorescence for KF-Cys (lanes 3 and 9 in panel a, lanes 3 and 5 in panel B). Conventional 5-F-Mal demonstrated efficient fluorescent labeling of single cysteine proteins as expected (lanes 2 and 8 in panel a, lanes 2 and 4 in panel B), but also extensive Cys independent labeling (lane 4). Negative control experiments featured identical conditions without 5-F-Pci (**38**) or 5-F-Mal treatment (lanes 10-13 in panel a, lanes 8-10 in panel B). The precision plus Kaleidoscope standards fluorescent ladder in lane 1 and PageRuler plus prestained protein ladder in **A)** lane 14 and **B)** lane 11 were used as references.

The Transpyrocinchonimidation Reaction

In aqueous conditions, pyrocinchonimides equilibrate between their imide and non-isolable amic acid forms.[54–57] The mechanism likely proceeds via nucleophilic attack of a water molecule or hydroxide ion at the imide carbonyl groups, followed by elimination of the amide, resulting in the ring-opened structure. The product amic acid then rapidly recyclizes via the reverse process, reaching an equilibrium position that is strongly influenced by pH.[54] This reactivity forms the mechanistic rationale for the following observations.

In addition to oxygen-based nucleophiles, nitrogen atoms can attack Pci carbonyls. In the presence of free amines, and presumably operating by the same mechanism, transient pyrocinchonic diamides can be expected to form. In this situation, the ring-closing step associated with Pci re-formation must occur with the ejection of a constituent nitrogen atom. Thermodynamic considerations dictate that the amine with the lowest pK_a offers the best potential leaving group during this step. Therefore, leveraging the established mechanism for pyrocinchonimide equilibration, it should be possible to achieve an overall imide-transfer reaction from an electron-depleted nitrogen atom to an electron-rich one (**Figure 4-8**).

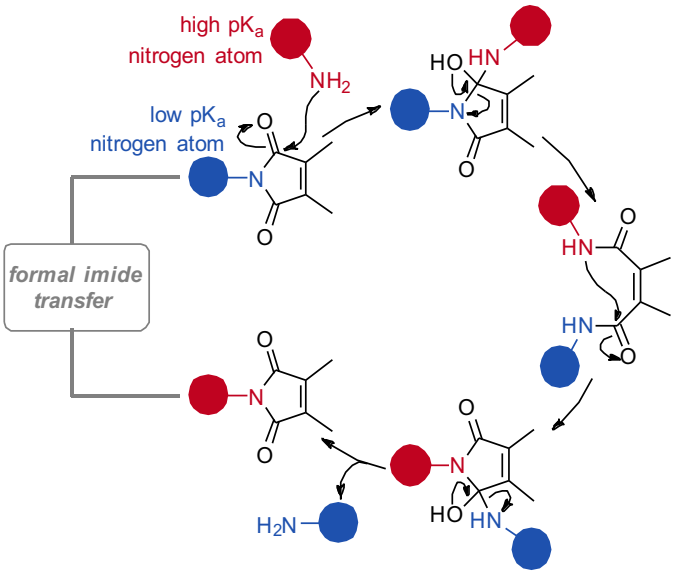


Figure 4-8. Mechanism for transpyrocinchonimidation. Attack of the pyrocinchonimide carbonyl groups by an electron-rich 1° amine can result in an overall imide transfer reaction (proton transfers omitted for simplicity). The depicted process is reversible and thermodynamically controlled.

The operability of this concept is now demonstrated. Compound **18** was originally synthesized for its ease of use in bioconjugation experiments owing to its excellent stability and solubility in water. Compound **18** is derived from sulfanilic acid, which features an ammonium ion >10-fold more acidic than acetic acid (pK_a 3.23 vs 4.75).[69] The electron deficiency of the nitrogen atom is further frustrated by the withdrawing effects of the twin acyl substituents of the Pci motif in **18**. In the presence of the considerably more basic ε-NH_2 of lysine residues ($pK_a \approx 10.5$) on target biomolecules, a formal imide transfer (transpyrocinchonimidation) was consistently observed. The reaction is a close analogue of a reported phthaloyl (Phth) protecting group installation method, where Phth-transfer also occurs from an acidic (carbamate) nitrogen atom.[70] Compound **27**, which carries an additional electron-withdrawing nitro substituent on the sulfanilic acid portion of the

molecule, was designed to further lower the pK_a for the pyrocinchonimide donor nitrogen atom and accelerate the imide transfer.

To further investigate the imide transfer reaction, compounds **18** and **27** were reacted with human insulin which provides a small, readily characterized peptide hormone to receive the imide. First, 20 μM insulin solutions were incubated with 100 mM **18** or **27** at 25 °C for 2 h; imide-transfer products were quantified by whole-protein ESI LC-MS, and the modified residues were subsequently identified by digest analysis (**Figure 4-9A**). The deconvoluted mass spectra were integrated over the regions corresponding to the calculated masses of the starting material and bioconjugates within a ± 0.5 Da range, and the relative intensities were calculated as a proportion of the sum total. Consistent with the proposed mechanism, insulin samples treated with **27** exhibited the highest proportions of Pci modification (**Figure 4-9B**). Additionally, insulin modified with **18** resulted exclusively in a single modification (**1×Pci**) of the peptide (within noise levels), whereas **27** delivered high proportions of the doubly-modified species (**2×Pci**).

Next, compound **27** (20 to 100 mM) was incubated with insulin (40 μM) in pH 8.5 solution at 25 and 37 °C for 2 and 18 h, and the full matrix of these conditions was explored (**Figure 4-9C**). At the most aggressive conditions examined (37 °C, 18 h, 100 mM **27**), the unmodified insulin levels were undetectable above the noise, and 69% of the bioconjugates carried two Pci groups. Notably, only two of the three available amines in insulin were modified; the ε-NH_2 nitrogen atom of insulin's one lysine side chain, and one of the two *N*-termini participated in the imide-transfer with either reagent, despite all three atoms being solvent-exposed. For comparison, bioconjugation with an NHS ester modified all three amines. (**Figure 4-9D**, **Figure 4-9E**, and **Figure 4-10**). Thus, imide

transfer by compound **27** to insulin reaction exhibits an innate selectivity for certain amines.

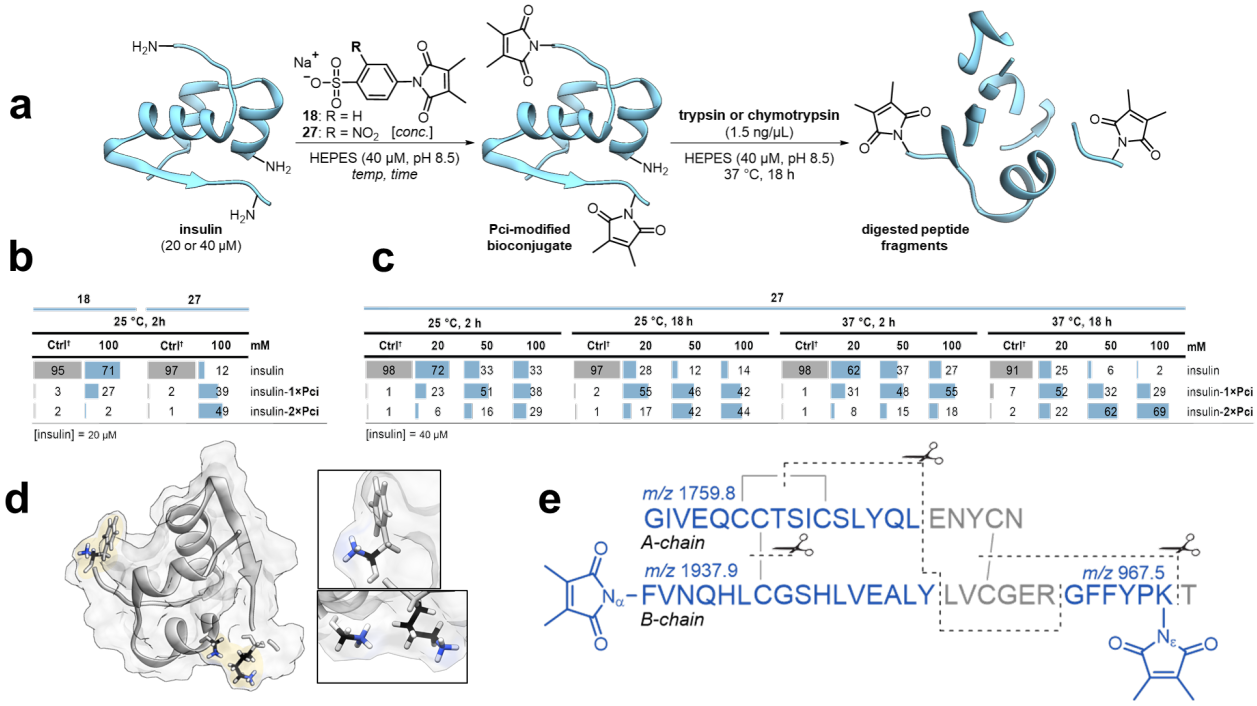


Figure 4-9. Transpyrocinchonimidation of insulin. A) Schematic for Pci transfer followed by intact protein ESI LC-MS and tandem digest MALDI-TOF of insulin-**Pci** bioconjugates. **B-C)** Estimated proportions (%) of Pci bioconjugates detected in reaction mixtures, calculated as the quotient of the signal intensities ascribed to the starting human, monomeric insulin and its bioconjugates. Peak integration values obtained from the deconvoluted spectra were used to perform this analysis, which assumes equal ionizability for the insulin starting material and the Pci-modified products. **D)** This view of insulin's structure highlights the two *N*-terminal α-NH_2 groups and the ε-NH_2 of B-chain K29 (PDB: 1MSO). **E)** Two modification sites were identified on the insulin B-chain by trypsin or chymotrypsin digest analysis: the *N*-terminal α-NH_2, and the K29 ε-NH_2. No modification was detected on the A chain's α-NH_2. The detected fragment masses, and corresponding cleavage sites are indicated. †Control samples were not exposed to the Pci-transfer reagents; this measurement can assess background noise in the mass ranges of the detected bioconjugates and helps to benchmark confidence in product detection.

a

A-chain *N*-terminal α-NH_2	B-chain *N*-terminal α-NH_2	B-chain K29 ε-NH_2	Peptide fragments
2 h, 25 °C in HEPES (40 mM, pH 6.5)	same	same	*Reaction conditions*
20 μM	40 μM	40 μM	*Insulin concentration*
10 mM	100 mM	100 mM	***27** concentration*
none	5 mM, 1 h, RT,	none	*DTT reduction*
chymotrypsin	chymotrypsin	trypsin	*Protease*

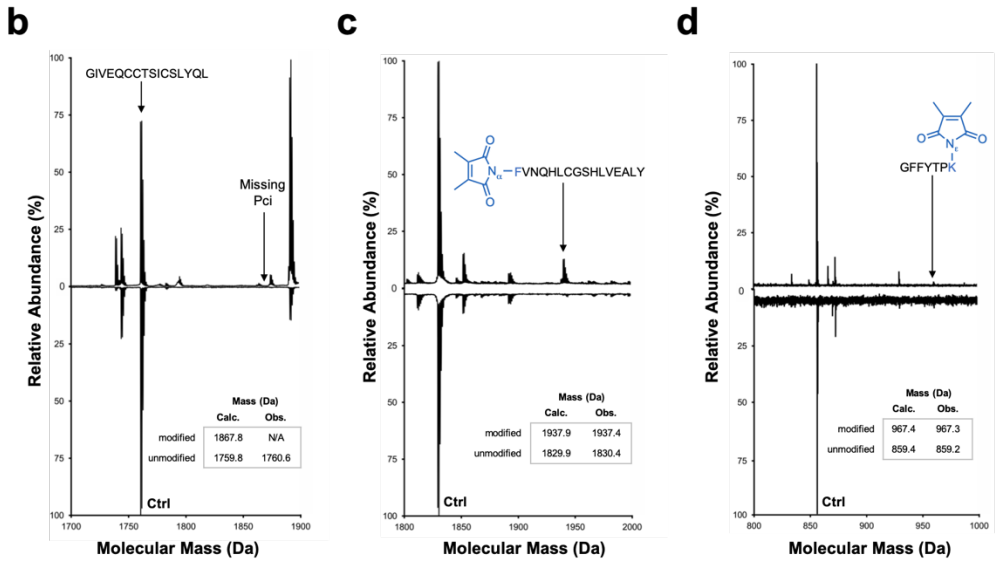


Figure 4-10. Characterization of insulin-Pci bioconjugates. **A**) Different proteases, addition of DTT, and other conditions were required to ensure each depicted peptide fragment with or without the Pci modification appeared in a MALDI-relevant MW range. Specifically, the spectra capture peptide fragments with the **B**) A-chain N-terminal α-NH_2, **C)** the B-chain *N*-terminal α-NH_2, or **D)** B-chain K29 ε-NH_2. Two **Pci-**modified sites were identified on the insulin B-chain, and none were detected on the A chain. Controls (Ctrl) depict spectra obtained from identical reaction conditions in the absence of **27**.

The imide transfer reaction was then evaluated in a more challenging context. Trastuzumab, also known as Herceptin (Genentech), is an antibody that binds to human epithelial growth factor receptor 2 (HER2), and is a biotherapeutic treatment for HER2-positive breast cancer.[71] Methods for preparing well-defined trastuzumab bioconjugates are in high demand for antibody-mediated delivery of chemo- and radiotherapeutic agents to tumor sites.[72,73] Two 3.5 μM trastuzumab solutions were incubated with 3.5 and 8.75 mM **27** at 25 °C for 2 and 18 h (conditions **I** and **II**, respectively) in buffered 40 mM HEPES

(pH 8.5). Prior to LC-MS analysis, the product mixtures were stripped of their *N*-linked surface glycans via digestion with PNGase F at 37 °C for 18 h.[74] ESI-MS analysis of the intact antibodies revealed average drug-to-antibody ratios (DAR, with Pci representing the 'drug')[75,76] of 0.9 and 2.3 for conditions **I** and **II**, respectively. In the product mixture obtained from **I**, the trastuzumab-**1×Pci** bioconjugate was found to be the most dominant signal, save for signal from the unmodified antibody. The trastuzumab-**4×Pci** bioconjugate was the highest level of modification observed for this condition (**Figure 4-11**). In contrast, condition **II** yielded a more extensively modified antibody product mixture, with the lowest ion count observed for the unmodified antibody. The modification density was also increased, with at most eight Pci-moieties detected per antibody, and the most dominant signal was assigned to the trastuzumab-**2×Pci** bioconjugate.

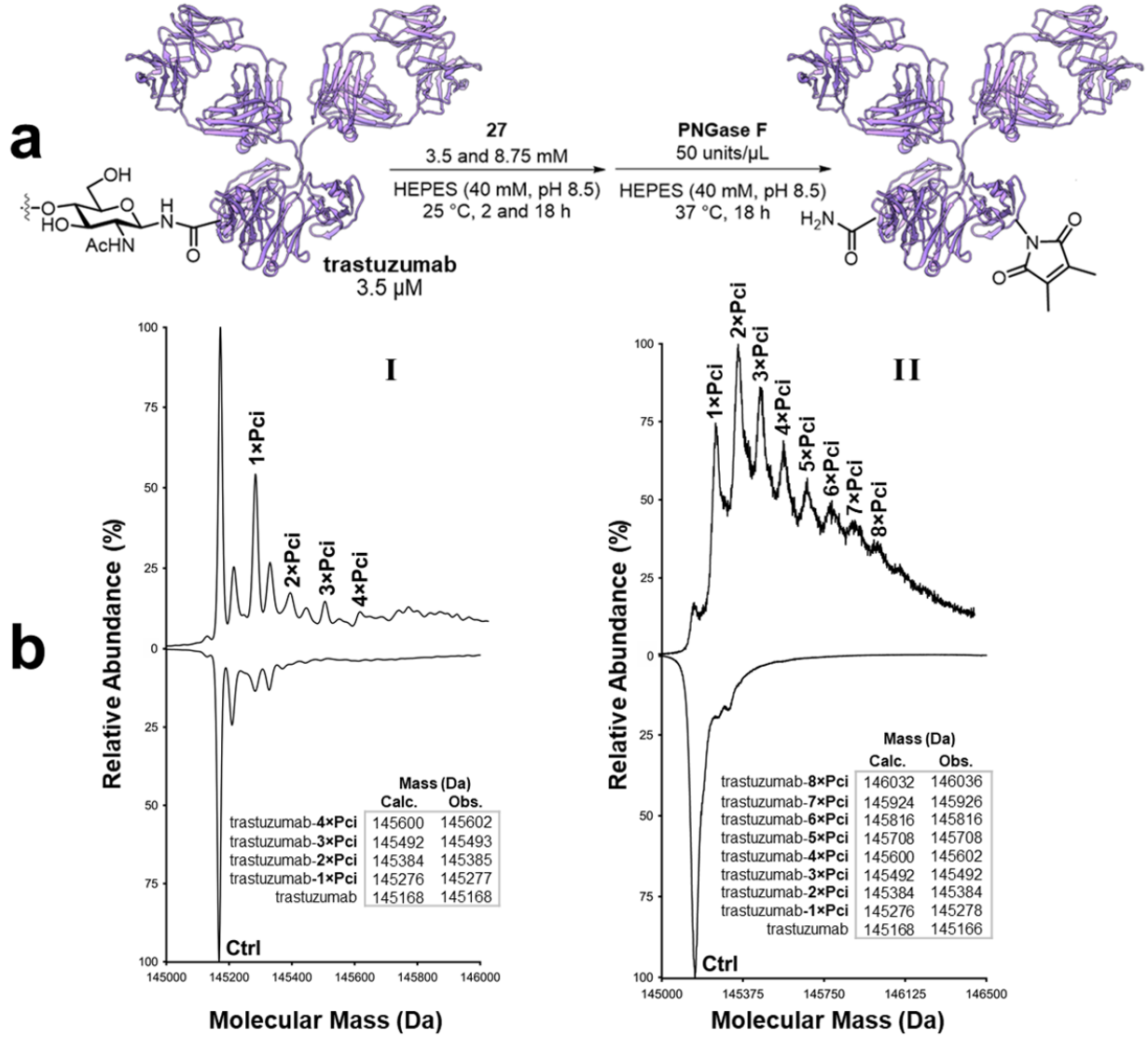


Figure 4-11. Transpyrocinchonimidation of trastuzumab. A) Schematic depiction of Pci bioconjugation of trastuzumab, followed by deglycosylation prior to ESI LC-MS analysis. **B**) Deconvoluted mass spectra of deglycosylated trastuzumab Pci-bioconjugates after treatment with (**I**) 3.5 mM **27** for 2 h, or (**II**) 8.75 mM **27** for 18 h at 25 °C. Traces on the top-half of the graph were obtained for the Pci-modified samples; traces on the bottom-half of the graph were obtained from the unmodified control samples of trastuzumab.

The cytotoxic trastuzumab ADC, 'emtansine' (Kadcyla), exhibits an average DAR of 3.5, also with as many as eight modifications per antibody.[77] Even small increases in DAR values can dramatically increase product heterogeneity. Lysine-modified ADCs with

DARs in the range 0–6, with approximately 40-45 lysine residues available for modification, can generate a combinatorial library with over four million unique members.[73,78,79] Such heterogeneity can result in unwanted behavior, including toxicity and loss of efficacy.[80,81] The challenge of addressing the combinatorial diversity of lysine-targeted bioconjugates has only recently started to receive a reply.[82–85] The low DAR values obtained with the Pci reagent **27** suggest that transpyrocinchonimidation may be a viable addition to the NHS alternatives toolkit, particularly in situations where bioconjugate heterogeneity is a persistent problem.

Protein-bound Pci groups could feasibly be expected to migrate from the bioconjugate to off-target amines. To examine this possibility, Pci-modified insulin bioconjugates were incubated with a series of potential small molecule-, peptide-, and protein-based Pci-scavengers for periods of up to 22 days (**Figure 4-12** and **Figure 4-13**). The insulin bioconjugate was observed to undergo a slow, pH-dependent reversion to its unmodified form; however, no evidence for Pci-exchange to amine groups on scavenger molecules was found. This behavior is typical of a class of bioconjugates intended for extended release of therapeutic payloads from targeted delivery systems, such as hydrogels and nanoparticles.[86–88]

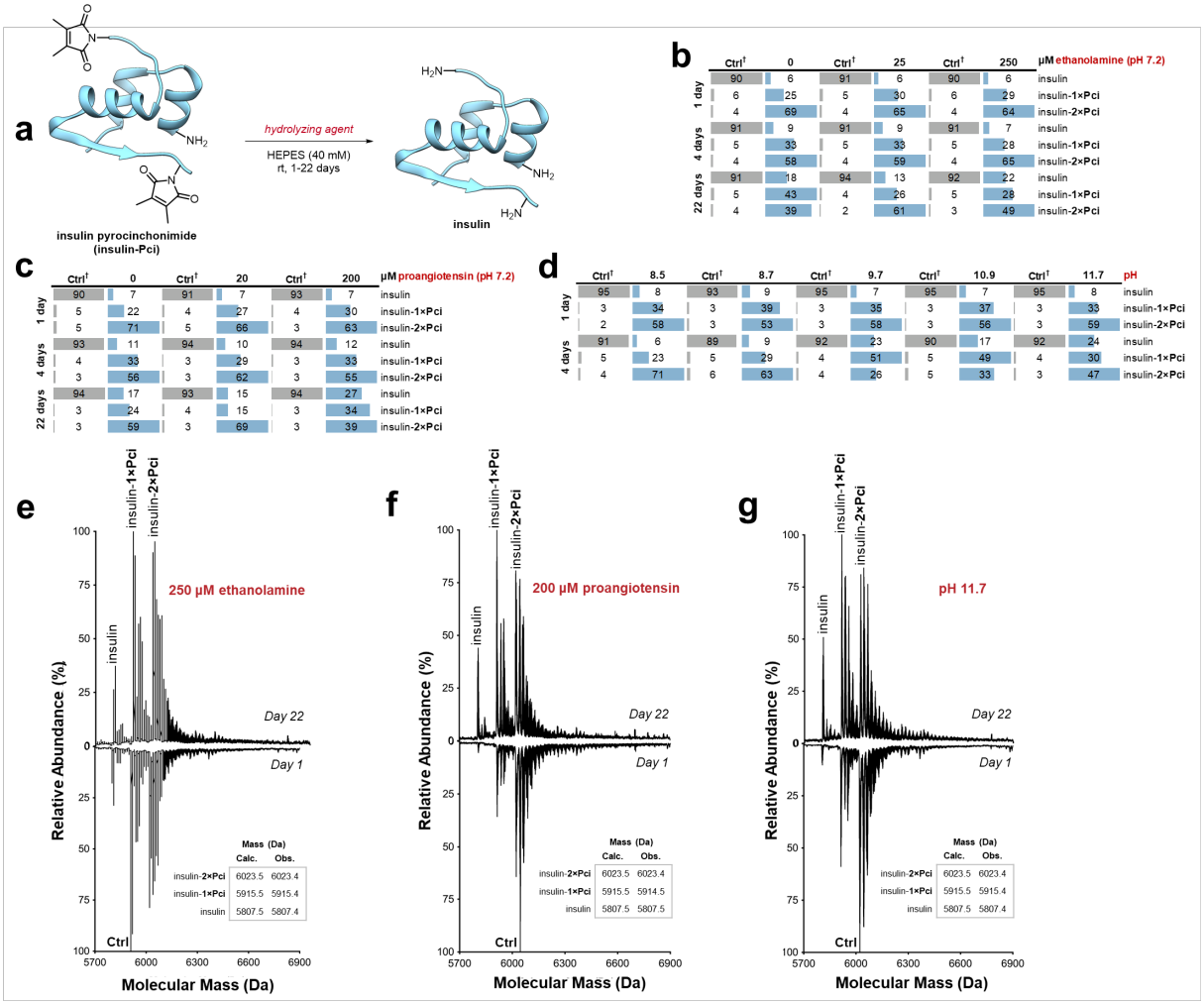


Figure 4-12. Analysis of insulin-Pci hydrolysis. A) Schematic of Pci hydrolysis in the presence of ethanolamine, proangiotensin, and elevated pH. Proportions (%) of insulin-Pci in the presence of **B**) ethanolamine (25 or 250 μM), **C**) proangiotensin (20 or 200 μM), or **D**) elevated pH (pH 8.5, 8.7, 9.7, 10.9, or 11.7) were estimated by ESI-LCMS. The relative differences in hydrolysis between ethanolamine and proangiotensin are unremarkable. The hydrolysis of the Pci groups appears to be pH-dependent over the range 8.5 -9.7; pH values outside of this range have little impact on the observed rates of hydrolysis. These semi-quantitative assessments of the product mixtures are obtained from integration of the peak area over a ±0.05 Da range of the calculated masses, and expressed as the proportion (%) of the total relative abundance. The deconvoluted mass spectrum of the reaction mixtures in the presence of **E**) ethanolamine (250 μM), **F**) proangiotensin (200 μM), and **G**) pH 11.7 after 1 and 22 days at rt.

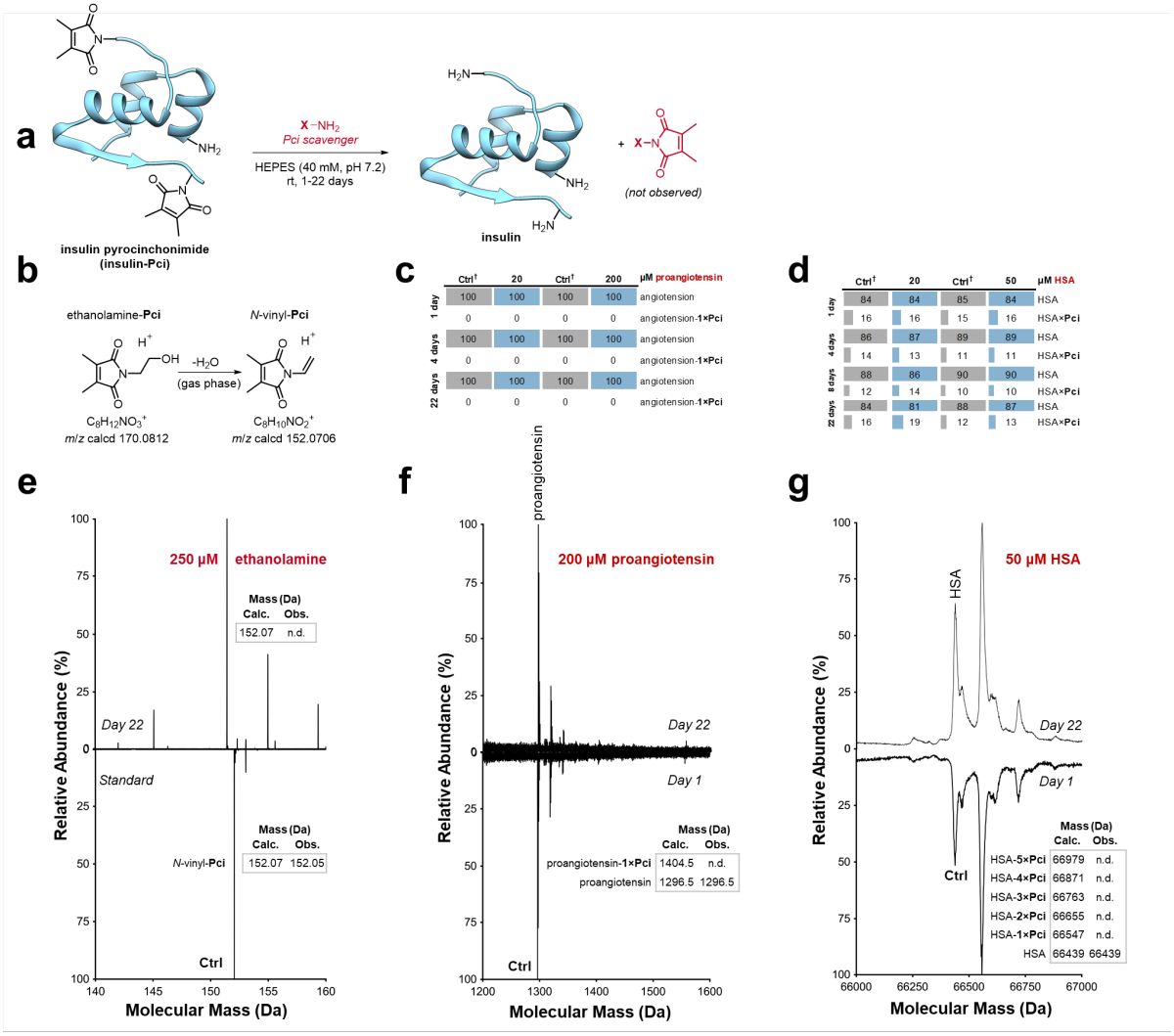


Figure 4-13. Analysis of Pci-scavenging reactions. A) Schematic of Pci-modified insulin reacting with the free amines of Pci scavenging agents: ethanolamine (25 or 250 µM, unencumbered primary amine), proangiotensin‡ (20 or 200 µM, one N-terminal amine only), or HSA† (20 or 50 µM, 66 lysines and one N-terminal amine). **B**) The reaction of ethanolamine (400 nM) with **27** (150 nM) yielded the expected product (**31**); this compound is unstable under the conditions of the analysis, and was detected as the anhydro-derivative, *N*-vinyl pyrocinchonimide, exclusively. The *N*-vinyl-Pci therefore serves the purpose of an authentic standard for assessment of ethanolamine Pci-scavenging by ESI-LCMS. **C-D**) HSA-Pci and proangiotensin-Pci were undetected at all concentrations and time periods examined. Semi-quantitative assessments of the product mixtures were obtained from integration of the peak areas over a ±0.05 Da range, corresponding to the calculated masses of the exchange products. **E-G**) The deconvoluted mass spectra of ethanolamine (250 µM), HSA (50 µM) and proangiotensin (200 µM) following incubation with insulin-Pci for 22 days in HEPES (40 mM, pH 7.2), showing no apparent Pci-exchange between amines. †The HSA starting material contained varying degrees of post-translational modifications (e.g., cysteinylation, glycosylation) with molecular weight components mainly in the range of 66,437 to 66,700

Da. ‡The minor peaks most are most likely due to proangiotensin hydrates (e.g., proangiotensin-H_2O with calculated mass, 1,312.5 Da and observed mass, 1,312.5 Da).

Our attention then turned to developing a Pci reagent with a covalently attached payload. Our laboratory uses bioconjugation for construction of single molecule devices based on single-walled carbon nanotube field-effect transistors (SWCNT-FETs).[89–91] These devices rely on the quasi-irreversible binding of aromatic hydrocarbon-based anchors to the SWCNT side-wall.[92] Our previously reported strategies for achieving protein-nanotube bioconjugates exclusively applied N-(1-pyrenyl)maleimide, reacting with a single-cysteine mutant of the target protein. Here, a Pci reagent was designed to target amines, not Cys thiols, and the pyrene substituent is carried by the Pci β-carbon, rather than its heterocyclic *N*-atom. The key step in the synthesis of the target is a frustratingly stubborn Wittig condensation, involving the phosphorus ylide derived from citraconimide. (**Figure 4-14**). The poor performance of this process is likely the result of a combination of ylide stabilization via its betaine resonance, and the poor reactivity of the aryl aldehyde (compared to the alkyl aldehyde precedents).[93–96] Improved procedures for introducing cargo to the Pci motif are currently in active development.

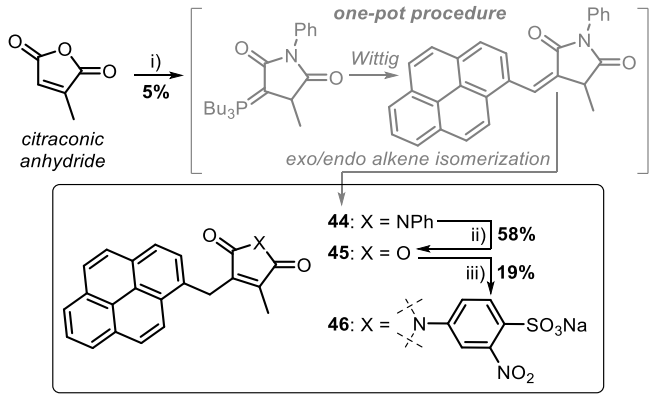


Figure 4-14. Synthesis of Pyrene-modified Pci Reagent 46. Reagents and conditions (i) $PhNH_2$, AcOH, 160 °C, 1 h, then Ac_2O, 160 °C, 0.5 h, then PBu_3, pyrene-1-

carboxaldehyde, NaOAc 160 °C, 48 h, then aq. HCl, 180 °C, 16 h. ii) NaOH, MeOH, CH_2Cl_2, rt, 30 min;[97] iii) sodium 4-sulfonato-3-nitraniline, DMF, 90 °C, 48 h. Compound **44**, the product of the multistep one-pot procedure, was unambiguously identified by single-crystal X-ray diffraction analysis (Supporting Information).

The reactivity of the pyrene-modified Pci reagent **46** was first investigated with insulin. Unsurprisingly, the cumbersome steric character and greatly reduced solubility of compound **46** compared to **27** resulted in lower transformation efficiency (**Figure 4-15**). Furthermore, the more sterically crowded molecule **46** also affects the intrinsic selectivity of the bioconjugation; exactly one modification of insulin was observed in this example. Compound **46** was then employed for modification of Taq polymerase (Taq). Taq was incubated with compound **46** at 37 °C for 18 h (pH 8.5). After removal of the excess Pci reagent by dialysis, the activity of the modified enzyme was verified by high resolution agarose gel assays to resolve dsDNA from ssDNA.[91] The Taq bioconjugate mixture retained 95% of the activity observed for the unmodified enzyme (**Figure 4-16**). The bioconjugate mixture was incubated with pristine SWCNTs using standard procedures and single-molecule attachments were investigated by atomic-force microscopy (AFM). AFM imaging revealed a single 1 nm feature, consistent with previous observations of similar proteins (**Figure 4-17**).

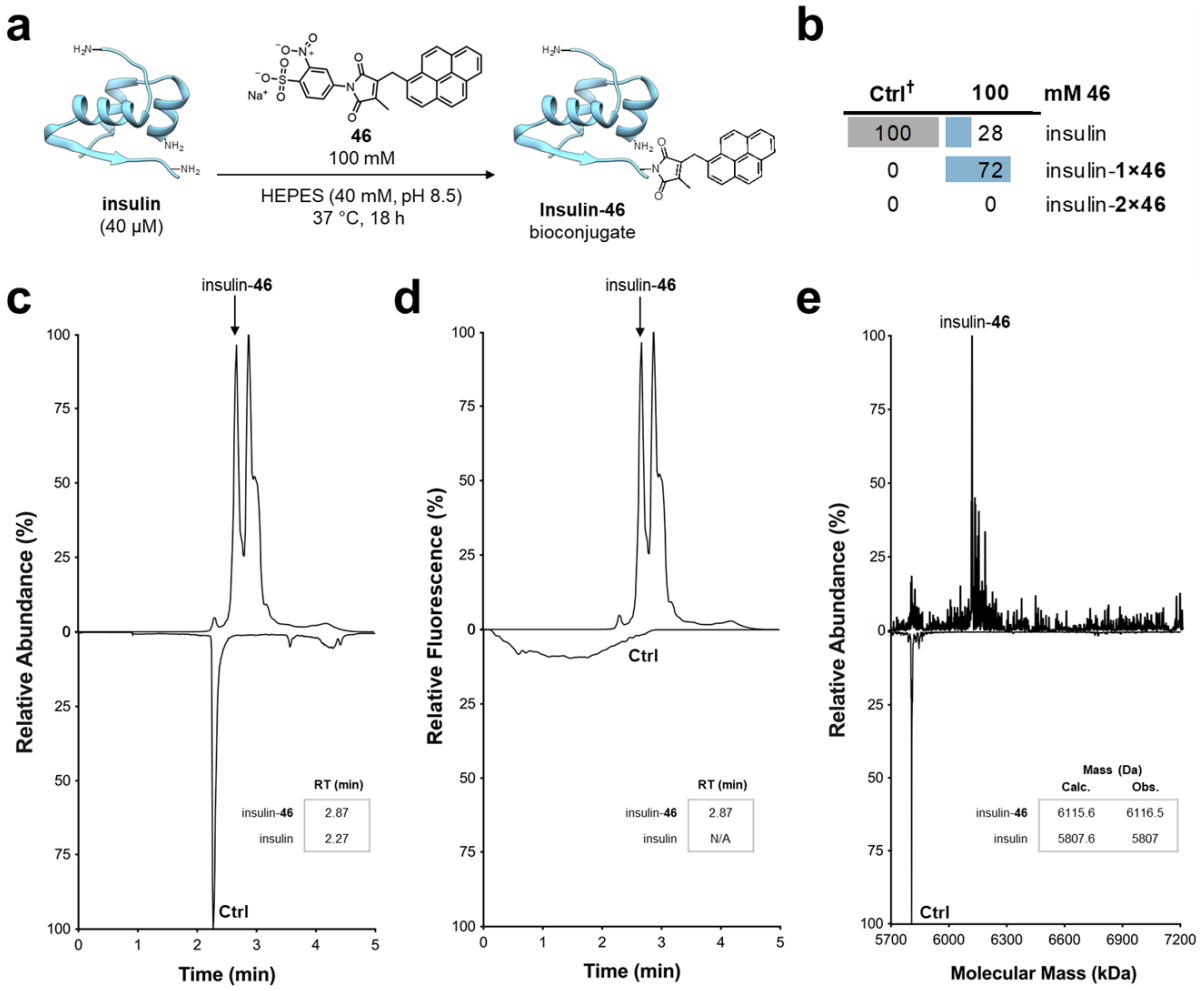


Figure 4-15. Characterization of the insulin-46 bioconjugate by ESI LC-MS. A) Schematic depiction of the insulin reaction with **46** and product characterization. A hypothetical **Pci** attachment position is modelled here (B-chain K29 ε-NH_2). **B)** The estimated proportions (%) of **Pci** bioconjugates detected in reaction mixtures were calculated as the quotient of the signal intensities ascribed to the starting human, monomeric insulin and its bioconjugates (**Figure 4-25**). †Throughout this report, the control samples (Ctrl) were not exposed to the Pci-transfer reagents; this measurement can assess background noise in the mass ranges of the detected bioconjugates, and helps confirm product detection. Total ion chromatograms (TICs) **C)** from UPLC-MS and D) UPLC-PDA (fluorescence) for insulin before (bottom) and after (top) its reaction with **46** (100 mM) for 18 h at 25 °C. The insulin-**46** bioconjugate eluted 0.6 min after the starting material. Similar profiles were obtained by both MS and fluorescence detection. Fluorescence parameters were λ_{ex} 320 nm and λ_{em} 350-500 nm. The second peak that elutes after the insulin-**46** bioconjugate is most likely the pyrene π-π complex heterodimer. E) The deconvoluted mass spectrum of insulin before (bottom) and after (top) the reaction with **46** (100 mM) for 18 h at 25 °C showed ~72% conversion to the expected product compared to the starting material.

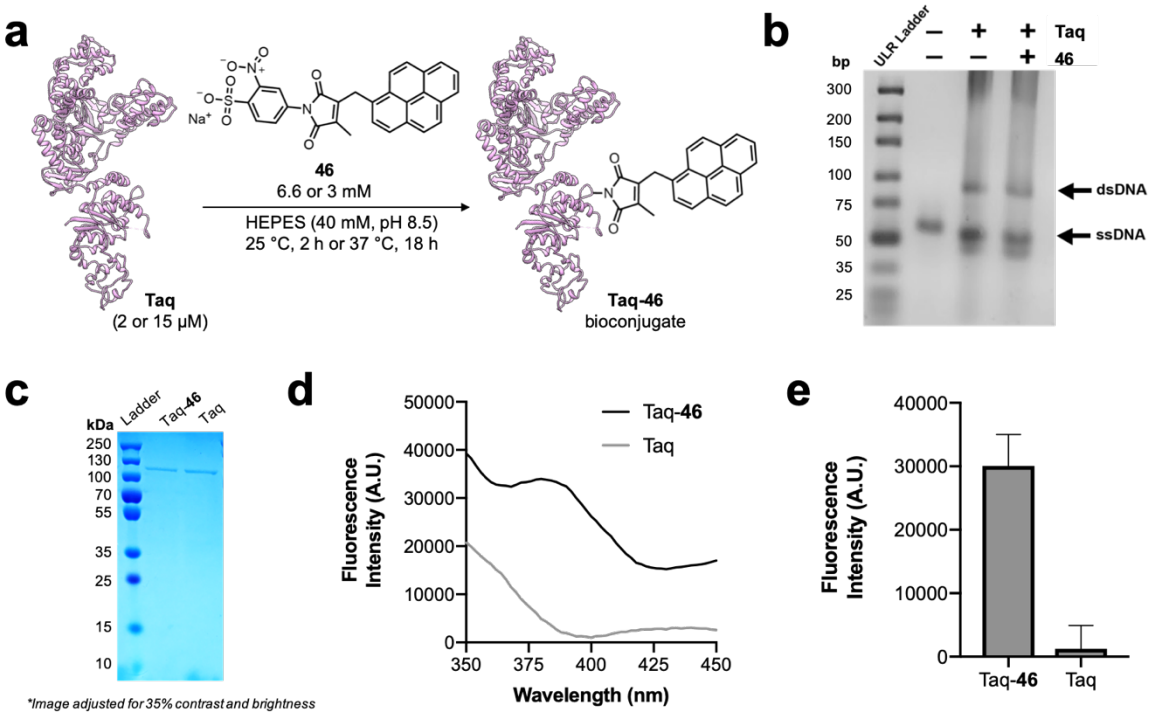


Figure 4-16. Characterization of the Taq-46 bioconjugate. A) Schematic depiction of the reaction of Taq with **46** (18 h at 37 °C in reaction buffer of 40 mM HEPES, pH 8.5) before characterization by ensemble activity assay and ESI LC-MS analysis. The precise Pci attachment position was not investigated; the position of the modification in the cartoon is arbitrary. **B)** In this DNA polymerase assay with Taq (G46D/R695C), a 5% high resolution agarose gel demonstrates formation of double-stranded DNA (dsDNA) from the PolyT42 template and dATP by Taq-**46**; before electrophoresis, the samples were pre-stained with the dsDNA stain SYBR Green I. The stained dsDNA exhibits slower electrophoresis than single-stranded DNA (ssDNA). The negative control reaction omitted Taq, and the positive control reaction applies Taq not subjected to modification by **46**. The Ultra Low Range (ULR) DNA ladder (ThermoFisher) was used to estimate DNA sizing. Band intensities were quantified using ImageJ[98] to determine a 5% decrease in the activity of the Taq-**46** bioconjugate compared to the Taq positive control. Demonstrating the generality of Pci labeling, panels c to e feature a variant of Taq lacking a surface-exposed Cys (D46G/C695R) that was reacted with **46** for 2 h at 25 °C. **C)** A 12% SDS PAGE demonstrates that the Taq-**46** bioconjugate migrated similarly to the unreacted protein. D) The fluorescence emission scan with λ_{ex} 240 nm and λ_{em} 350-450 nm and E) the relative emission at λ_{em} 394 nm quantified used a Tecan Spark 10M multimode microplate reader (Tecan Group Ltd., France) of Taq-**46** demonstrate incorporation of the pyrene fluorophore. Scans were normalized with the reaction buffer. Error bars indicate standard deviations (n = 4).

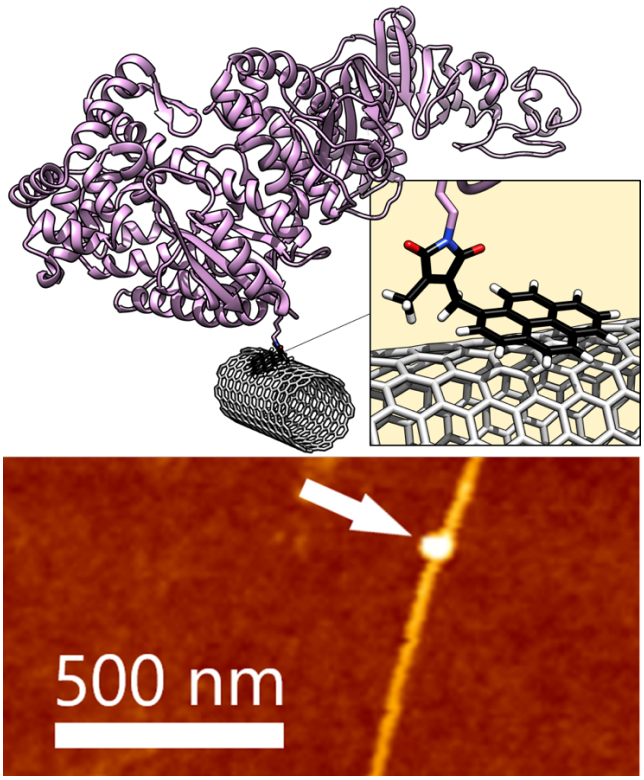


Figure 4-17. A schematic diagram of a SWCNT-FET noncovalently bioconjugated to a single molecule of Taq. A pyrene-Pci molecule, **46**, (black) is adhered to the SWCNT-FET through π−π stacking. Atomic force microscopy shows the expected 1−2 nm diameter of the SWCNT-FET with a single Taq attachment obtained using linker **46** (1 nm, white arrow).

UV-Activated Cycloadditions

Motivated by the efficiency of **27** to introduce Pci into proteins, we sought reactions capable of further elaborating the Pci bioconjugates. Cycloadditions of alkenes with maleimide derivatives have been extensively examined in the literature. The synthesis of cyclobutane-containing fused di-, tri- and tetracyclic scaffolds via UV-initiated [2+2] cycloadditions of maleimide-like compounds are especially well-documented. In 2001, an unfamiliar [5+2] cycloaddition of Pci was reported,[99] and was then applied to a challenging total synthesis.[100] The reaction involves a formal insertion of two alkene carbons into the Pci ring between the nitrogen atom and one of its carbonyls, resulting in an overall two-

carbon ring-expansion, and the formation of a seven-membered dihydroazepinedione (Dhzd). Initial mechanistic proposals for this unusual transformation have since been succeeded by robust, experimentally validated models.[101–103]

Two hypothesized photochemical processes compete to support the [2+2] or the [5+2] cycloaddition pathways. Sensitized irradiation in the presence of an appropriate chromophore populates the C=C triplet state and facilitates the [2+2] pathway. Direct irradiation, which operates in the absence of a photosensitizer, populates the C-N singlet state, causing homolytic cleavage, and results in an amide/acyl diradical intermediate. While energy transfer from the singlet to the triplet state can occur through intersystem crossing, the presence of electron-donating methyl groups at the alkene weakens the C-N bond and supports diradical formation.[103] The diradical intermediate undergoes a formal [5+2] cycloaddition in the presence of alkenes, or else spontaneously recombines to regenerate the starting Pci molecule.

These UV-activated processes offer a bioorthogonal route to further functionalization of Pci-modified proteins (**Figure 4-18A)**. Two alkene-based secondary modifiers were chosen to investigate this approach: a commercially available PEG derivative and an EDTA-derived metal ligand (**47**), which was synthesized using a one-pot method inspired by existing procedures[104] (**Figure 4-18B)**. The PEG derivative demonstrates a typical solubilization scheme for biotherapeutics. The introduction of metal complexes to proteins has a wide scope of utility, including the chelation of radioisotopes for targeted radiotherapy, as contrast and imaging agents in nuclear medicine, and as luminescent probes, serving as alternatives to traditional organic fluorophores. EDTA is a general-purpose ligand that can form stable complexes with a

variety of different metal atoms with many potential applications. Solutions of Pci-modified insulin (~40 μM) were incubated with mPEG-alkene-1K (40 mM, MW ~1000 Da) or **47**, under direct irradiation from a 450 W medium-pressure Hg lamp (228-420 nm), and temperature regulation below 8 °C was achieved with a high-velocity flow of chilled water (**Figure 4-19**).

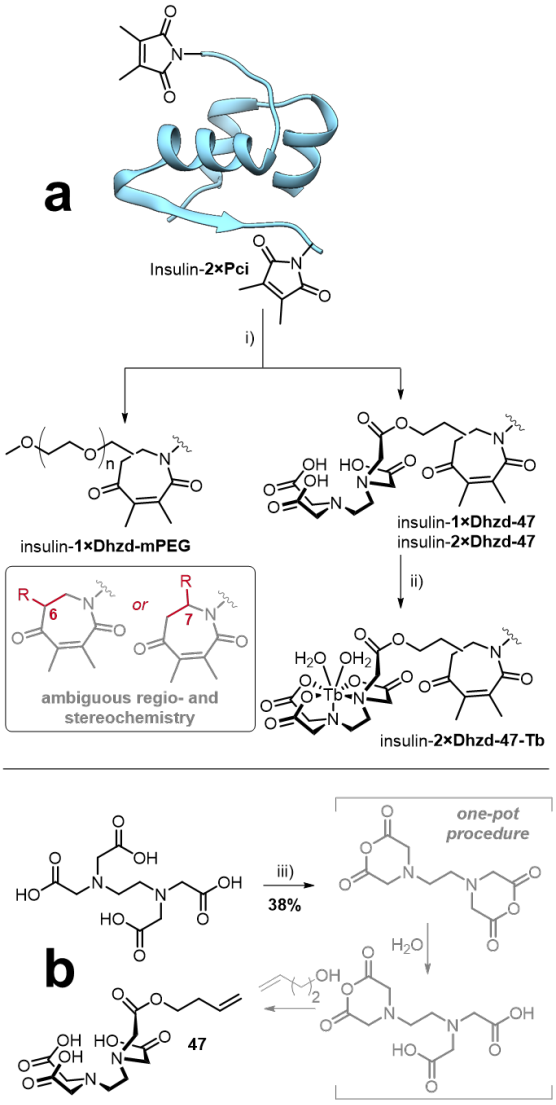


Figure 4-18. ***Reagents and conditions*** i) 40 mM mPEG-alkene 1K or 40 mM **47**, hv 228-420 nm, 40 mM HEPES (pH 8.5), 5-10 °C, 4 h; ii) 10 mM terbium(III) chloride, 40 mM HEPES (pH 8.5), 25 °C, 18 h; iii) Ac_2O, DMF, 60 °C, 24 h, then H_2O (1 equiv.), 75 °C, 2 h, then 3-buten-1-ol, DMAP, 80 °C, 24 h.

Analysis of the product mixtures by ESI LC-MS revealed masses consistent with the target cycloadducts (**Figure 4-19**). The insulin-**1×Dhzd-mPEG** cycloadduct presented a range of *m*/*z* signals separated by ~44 Da, a typical polymeric PEG MS fingerprint. Only the **1×Dhzd-mPEG** attachment was detected, with a second Pci group remaining unmodified. In the cycloaddition reaction with **47**, both the **1×** and **2×Dhzd-47** insulin adducts were observed. After incubation with terbium(III) chloride prior to further analysis, only the insulin-**2×Dhzd-47-Tb** complex was observed (as the dihydrate). The increased background noise of the mass spectra precluded estimation of product yields; photolytic decomposition of the protein likely occurred, resulting in additional mass fragments. Furthermore, the structures in **Figure 4-18** are drawn for the [5+2] product Dhzd compounds, which is concordant with the published mechanism. However, the [2+2] cyclobutane product with the same mass could also form, and thus the exact identity of the cycloadducts remains unclear. Resolving this uncertainty will likely require a comprehensive X-ray crystallography experiment beyond the scope of the present study.

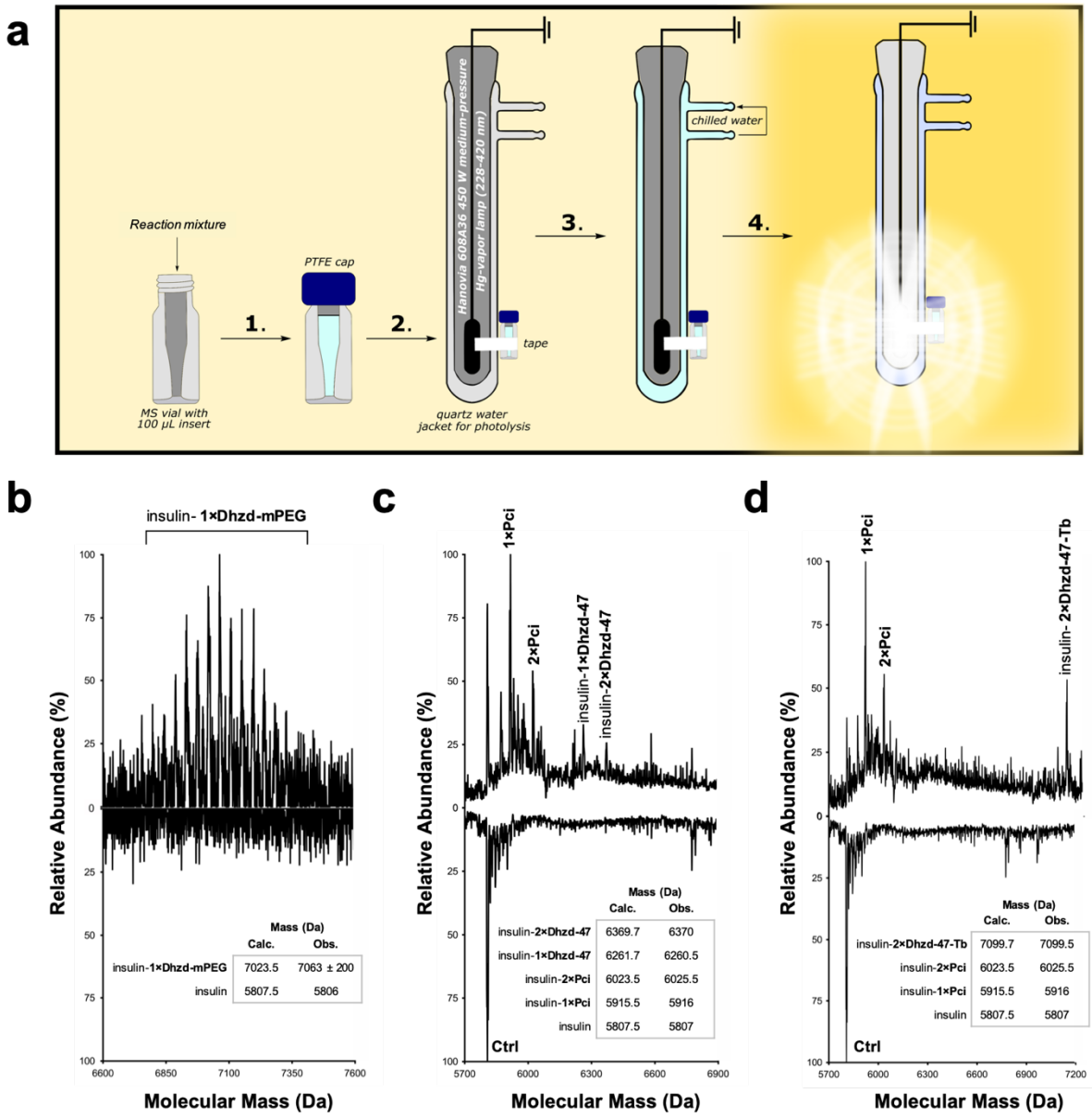


Figure 4-19. Cycloaddition reactions of the insulin-Pci. A) Schematic of the UV-catalyzed processes used to further functionalize Pci-modified insulin: **1)** A mass spectrometry vial was charged with an aqueous solution of the insulin-**Pci** bioconjugate and mPEG-alkene-1K or **47** (100 µL volume), then capped. **2)** The vial was secured to the outside surface of a photolysis quartz water jacket with tape, then a Hanovia 608A36 450 W medium-pressure mercury vapor lamp was placed inside the apparatus. **3)** The water jacket was connected to a recirculating pump, and filled with chilled water (~4 °C) from an ice bath. **4)** The apparatus was enclosed inside a steel box, the lamp was activated, and irradiation of the sample was applied continuously for 4 h. The temperature of the recirculating water was monitored at the exit, and did not exceed 8 °C during the experiment. **B)** MALDI mass spectrum of insulin before (bottom) and after (top) reaction with **27** (100 mM) in HEPES (40 mM, pH 8.5) for 18 h at 37 °C, then mPEG-alkene-1K (100 mM) after 4 h under UVC light at 266 nm and temperatures ≤8 °C using a water-cooled quartz jacket. The mass

spectra exhibit a distribution of masses commensurate with the expected insulin-**1×Dhzd-mPEG** bioconjugate. We did not observe the insulin-**2×Dhzd-mPEG** bioconjugate. **C)** Deconvoluted mass spectrum of the insulin-**Pci** bioconjugate following reaction with **47** (40 mM) under UV irradiation conditions described above. The observed products included insulin-1×Dhzd-**47** and insulin-2×Dhzd-**47**, resulting from one and two cycloadditions per insulin molecule, respectively. D) Deconvoluted mass spectrum of insulin-**Dhzd-47** following a 18 h reaction with $TbCl_3$ at 25 °C showed conversion of the starting material to the expected product, insulin-**2×Dhzd-47-Tb** (detected exclusively as the dihydrate complex). Other major peaks represent the unreacted insulin-**Pci**.

4.4 Conclusion

The Pci bioconjugation reported here offers useful capabilities and a reaction modality that is distinctive from maleimides and NHS esters. The investigation demonstrates that Pci compounds react very poorly with thiols via the conventional thio-Michael pathway, despite scattered reports to the contrary. Therefore, Pci reagents are likely to be compatible with thiols, including Cys residues on protein targets and bystanders, and the reducing agents BME and GSH. The imide transfer reaction may also be helpful in addressing frontier challenges in bioconjugation chemistry. We demonstrate effective imide transfer can result in low conjugation loadings and DARs compared to NHS methods, even when very large excesses of Pci reagent are employed (**Figure 4-20** and **Figure 4-21**). The Pci modification of proteins is slowly lost to hydrolysis at neutral-to-basic pH; however, no evidence for protein-protein imide exchange was found (**Figure 4-12** and **Figure 4-13**). The fact that the hydrolyzed anhydride/diacid is unreactive towards amine groups on neighboring proteins also suggests that prior activation (e.g. **18** and **27**) is essential. In contrast to maleimide reversion, the terminal nature of Pci hydrolysis may protect against unwanted off-target modification in situ.[105] We expect the Pci group may be of value in situations where ADC development is confounded by product heterogeneity issues, or where a slow release of the payload is

desirable. Lastly, evidence has been collected that suggests tandem Pci-transfer/UV-cycloaddition procedures may be a viable approach to the future development of Pci capabilities.

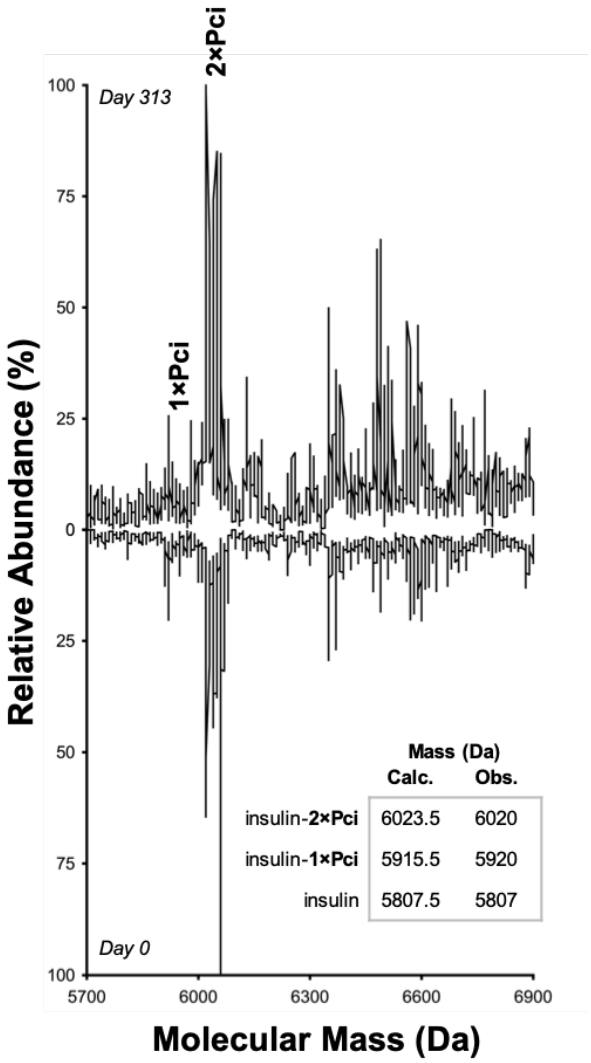


Figure 4-20. Insulin-Pci bioconjugate stability investigated over a 10-month period by ESI LC-MS. The deconvoluted mass spectrum of insulin (20 μM) after reaction with **27** (100 mM, 18 h at 37 °C), comparing 10 months of storage at 4 °C (top) with the initial analysis of the product mixture (bottom). As expected, the sample after 10 months features evidence of some protein decomposition, but comparable integrity of the sample and its Pci bioconjugate. Specifically, the proportion of insulin-**2×Pci** after 10-months of storage (86%) is similar to the amount in the initial product analysis (84%).

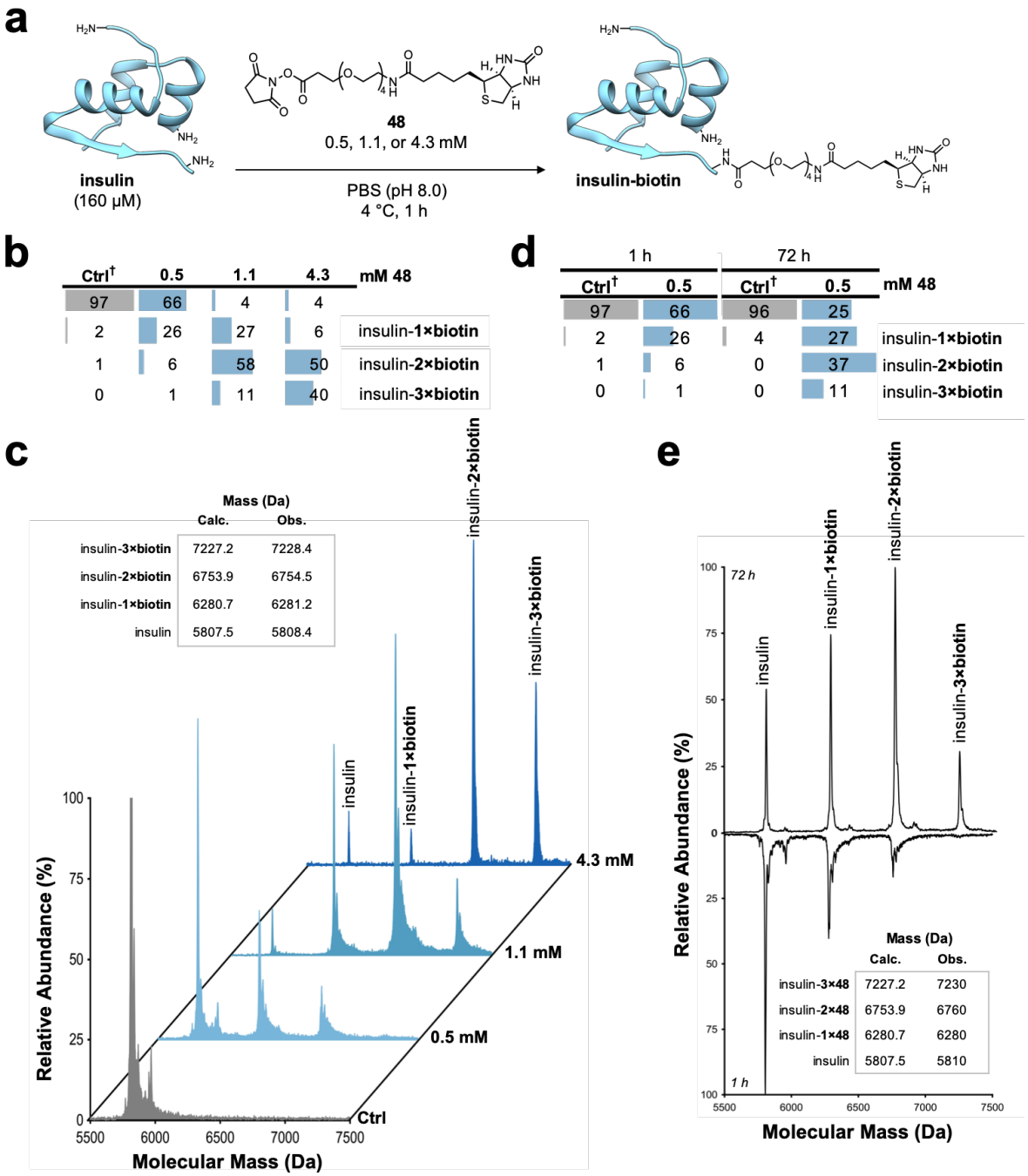


Figure 4-21. Characterization of a conventional NHS (48) reacting with insulin. A) Schematic depiction of insulin reacted with **48** for characterization by MALDI analysis. **B)** The estimated proportions (%) and **C)** deconvoluted mass spectrum of insulin reacted with **48** (0.5, 1.1, or 4.3 mM) for 1 h at 4 °C illustrate an increasing number of insulin-biotin products with higher concentrations of **48**. **D)** The reaction of insulin with 0.5 mM **48** at 4 °C in PBS (pH 8.0) was analyzed after 1 and 72 h. The estimated proportions (%) and **E)** deconvoluted mass spectrum of this reaction mixture depict the starting material (insulin) after the 3-day reaction. This semi-quantitative assessment of the product

mixtures integrated the peak area over a ±1.0 Da range of the theoretical masses for each peak, and expressed the proportion (%) as total relative abundances.

Although Pci is a structural analogue of maleimide, the Pci-transfer is more functionally aligned with an emerging class of alternative amine-targeting methods and reagents,[106] including the 6 π -azaelectrocyclization reagents,[107–111] squarimides,[112–114] iminoboronates,[115] TAK-242 derivatives,[116,117] diazonium terephthalate esters,[118] two-component formaldehyde/indole couplings,[119] phthalaldehydes,[120] acylfluorides,[121] palladium-catalyzed arylations,[122] sulfonyl acrylates,[84,123] fluorophenyl esters,[85] DSH-mediated acylations,[124] heteroaryl methylsulfones,[125] dinitroimidazoles,[126] and linchpin-directed modification strategies.[127,128] The Pci-transfer reagents **18** and **27** were used in unusually high concentrations in the present study, compared with typical concentration ranges employed in bioconjugate chemistry. However, these are only the first generation of these reagents. Exploration of other water-soluble variants with alternative substitution patterns around the aniline portion of the Pci transfer reagent is fertile ground for further research. We anticipate that new iterations of these reagents with improved reaction kinetics and defined site-selectivity profiles will be the subjects of future studies.

4.5 Materials and Methods

Materials

Reagents purchased commercially include the following: cell lines (Stratagene), IPTG (Carbosynth), glycine (Biobasic), Ni-IMAC resin (Bio-Rad Laboratories), and mPEG-Alkene (Biochempeg). All other chemicals were supplied by Fisher Scientific or ThermoFisher Scientific. Supplies purchased commercially include: 0.45 μm pore filter (Genesee Scientific), 0.22 μm pore filter (Denville Scientific), 0.22 μm top vacuum filter

(Genesee Scientific), 3.5K MWCO 0.1 mL Slide-A-Lyzer mini dialysis device (ThermoFisher), and Amicon Ultra-0.5 mL centrifuge filter (Fisher Scientific). All reagents and supplies were used as received. All solutions were sterile-filtered or autoclaved before use.

Small molecule characterization

Small molecule high-resolution mass spectra (HRMS) were obtained by electrospray ionization (ESI) on a Waters (Micromass) LCT Premier equipped with a time-of-flight (TOF) mass analyzer. Proton (^{1}H, 400 and 600 MHz), carbon (^{13}C, 100 and 150 MHz), and fluorine (^{19}F, 564 MHz) nuclear magnetic resonance (NMR) spectra were obtained on Bruker instruments equipped with a switchable QNP or BBFO probe. NMR samples were prepared in $CDCl_3$ and DMSO-d_6, and residual protonated solvent was used as an internal chemical shift standard. ^{1}H and ^{13}C assignments were determined using HSQC and 10 Hz optimized HMBC 2D-NMR analyses. Optical rotatory powers were measured in the sodium D band (589 nm) on a Jasco P-1020 polarimeter. Fourier transform infrared (FTIR) spectra were obtained as neat samples on a Jasco 4700 attenuated total reflectance (ATR) instrument using a diamond-coated zinc selenide sample accessory. X-ray data was collected on a Bruker SMART APEX II diffractometer. Flash chromatography was carried out on silica gel 60 according to the method established by Still et al.,[129] or using a Teledyne Isco CombiFlash preparative chromatography system. Analytical thin layer chromatography (tlc) was conducted on aluminum-backed 2 mm thick silica gel 60 GF_{254}, and chromatograms were visualized

under a UV lamp (254 and 365 nm) or by chemical staining with ceric ammonium molybdate (Hanessian's stain) or potassium permanganate ($KMnO_4$).

Proteins and antibodies used in this study

Table 4-1. The amino acid sequence and theoretical MW of the proteins and antibodies used in this study.

Protein	Sequence (N-terminal His_6 epitope indicated in blue and mutations indicated in red)	Theoretical MW (Da)
Proangiotensin	DRVYIHPFHL	1,296.48
Insulin	A Chain - GIVEQCCTSICSLYQLENYCN B Chain - FVNQHLCGSHLVEALYLVCGERGFFYTPKT	5,807.6
Dihydrofolate reductase (DHFR) from *E. coli*	MAHHHHHHVDDDDKMISLIAALAVDRVIGMENAMPWNLPADLAWFKRNTLNKPV IMGRHTWESIGRPLPGRKNIILSSQPGTDDRVTWVKSVDEAIAAAGDVPEIMVI GGGRVYEQFLPKAQKLYLTHIDAEVEGDTHFPDYEPDDWESVFSEFHDADAQNS HSYSFEILERR	19,664.03
Human serum albumin (HSA)	MKWVTFISLLFLFSSAYSRGVFRRDAHKSEVAHRFKDLGEENFKALVLIAFAQY LQQCPFEDHVKLVNEVTEFAKTCVADESAENCDKSLHTLFGDKLCTVATLRETY GEMADCCAKQEPERNECFLQHKDDNPNLPRLVRPEVDVMCTAFHDNEETFLKKY LYEIARRHPYFYAPELLFFAKRYKAAFTECCQAADKAACLLPKLDELRDEGKAS SAKQGLKCASLQKFGERAFKAWAV	66,437.02
Klenow fragment of DNA polymerase I (KF) from E. coli	HHHHHHENLYFQGVISYDNYVTILAEATLKAWIAKLEKAPVFAFATATDSLDNI SANLVGLSFAIEPGVAAYIPVAHDYLDAPDQISRERALELLKPLLEDEKALKVG QNLKYDRGILANYGIELRGIAFDTMLESYILNSVAGRHDMDSLAERWLKHKTIT FEEIAGKGKNQLTFNQIALEEAGRYAAEDADVTLQLHLKMWPDLQKHKGPLNVF ENIEMPLVPVLSRIERNGVKIDPKVLHNHSEELTLRLAELEKKAHEIAGEEFNL SSTKQLQTILFEKQGIKPLKKTPGGAPSTSEEVLELALDYPLPKVILEYRGLAK LKSTYTDKLPLMINPKTGRVHTSYHQAVTATGRLSSTDPNLQNIPVRNEEGRRI RQAFIAPEDYVIVSADYSQIELRIMAHLSRDKGLLTAFAEGKDIHRATAAEVFG LPLETVTSEQRRSAKAINFGLIYGMSAFGLARQLNIPRKEAQKYMDCYFERYPG VLEYMERTRAQAKEQGYVETLDGRRLYLPDIKSSNGARRAAAERAAINAPMQGT AADIIKRAMIAVDAWLQAEQPRVRMIMQVHDELVFEVHKDDVDAVAKQIHQLME NSTRLDVPLLVEVGSGENWDQAH	69,387.29
Phi29 DNA polymerase (Φ29) from Phi29 bacteriophage	MGSSHHHHHHSSGLVPRGSHMKHMPRKMYSCAFETTTKVEDCRVWAYGYMNIED HSEYKIGNSLDEFMAWVLKVQADLYFHNLKFAGAFIINWLERNGFKWSADGLPN TYNTIISRMGQWYMIDICLGYKGKRKIHTVIYDSLKKLPFPVKKIAKDFKLTVL KGDIDYHKERPVGYKITPEEYAYIKNDIQIIAEALLIQFKQGLDRMTAGSDSLK GFKDIITTKKFKKVFPTLSLGLDKEVRYAYRGGFTWLNDRFKEKEIGEGMVFDV NSLYPAQMYSRLLPYGEPIVFEGKYVWDEDYPLHIQHIRCEFELKEGYIPTIQI KRSRFYKGNEYLKSSGGEIADLWLSNVDLELMKEHYDLYNVEYISGLKFKATTG LFKDFIDKWTYIKTTSEGAIKQLAKLMLNSLYGKFASNPDVTGKVPYLKENGAL GFRLGEEETKDPVYTPMGVFITAWARYTTITAAQACYDRIIYCDTDSIHLTGTE IPDVIKDIVDPKKLGYWAHESTFKRAKYLRQKTYIQDIYMKEVDGKLVEGSPDD YTDIKFSVKCAGMTDKIKKEVTFENFKVGFSRKMKPKPVQVPGGVVLVDDTFTI K	68,558.15

DNA polymerase I (Taq) from *Thermus aquaticus*	MHHHHHHENLYFQGIFRAPMRGMLPLFEPKGRVLLVDGHHLAYRTFHALKGLTT SRGEPVQAVYGFAKSLLKALKEDGDAVIVVFDAKAPSFRHEAYGGYKAGRAPTP EDFPRQLALIKELVDLLGLARLEVPGYEADDVLASLAKKAEKEGYEVRILTADK DLYQLLSDRIHVLHPEGYLITPAWLWEKYGLRPDQWADYRALTGDESDNLPGVK GIGEKTARKLLEEWGSLEALLKNLDRLKPAIREKILAHMDDLKLSWDLAKVRTD LPLEVDFAKRREPDRERLRAFLERLEFGSLLHEFGLLESPKALEEAPWPPPEGA FVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVL ALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARRYGGEWTEEAGERAALSERLF ANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVRLDVAYLRALSLEVAEEI ARLEAEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTEKTGKRSTSAAVLEA LREAHPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLS SSDPNLQNIPVRTPLGQRIRRAFIAEEGWLLVALDYSQIELRVLAHLSGDENLI RVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHRLSQELA IPYEEAQAFIECYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRRRYVPDLEARV KSVREAAERMAFNMPVQGTAADLMKLAMVKLFPRLEEMGARMLLQVHDELVLEA PKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGEDWLSAKE	96,247.78
Trastuzumab	*Light chain protein sequence:* DIQMTQSPSSLSASVGDRVTITCRASQDVNTAVAWYQQKPGKAPKLLIYSASFL YSGVPSRFSGSRSGTDFTLTISSLQPEDFATYYCQQHYTTPPTFGQGTKVEIKR TVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQES VTEQDSKDSTYSLSSTLTLSKADYEKHKVYACVTHQGLSSPVTKSFNRGEC *Heavy chain protein sequence:* EVQLVESGGGLVQPGGSLRLSCAASGFNIKDTYIHWVRQAPGKGLEWVARIYPT NGYTRYADSVKGRFTISADTSKNTAYLQMNSLRAEDTAVYYCSRWGGDGFYAMD YWGQGTLVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWN SGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDK KVEPPKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDV SHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYK CKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYP SDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVM HEALHNHYTQKSLSLSPGK	145,549.56

Protein expression and purification

Proangiotensin

Proangiotensin, as known as angiotensin I human acetate salt hydrate, was purchased from a commercial supplier (Sigma). Prior to bioconjugation reactions, proangiotensin sample purity was examined by ESI LC-MS analysis (**Figure 4-22**).

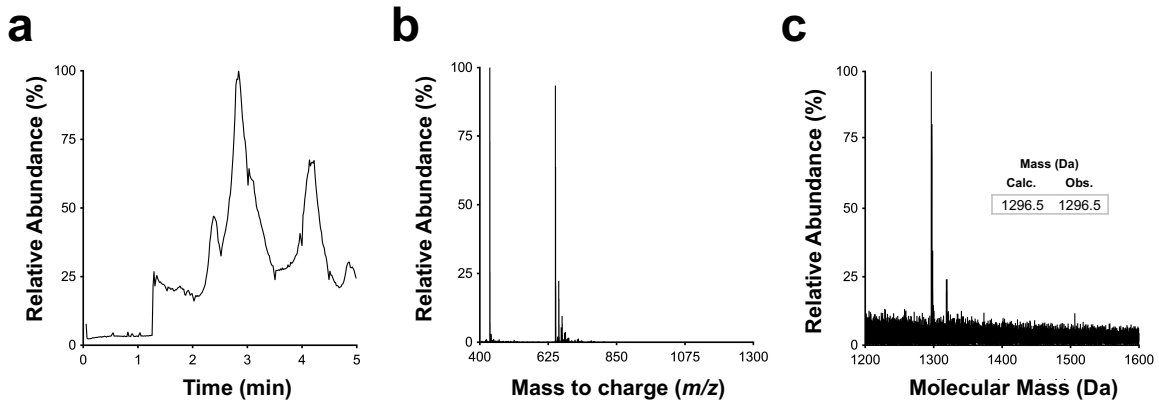


Figure 4-22. Proangiotensin purity assessed by ESI LC-MS analysis. Representative **A)** TIC, **B)** combined *m/z* spectrum, and **C)** combined *m/z* spectrum in 1200-1600 Da range of proangiotensin. The minor peak most likely represents a proangiotensin hydrate (e.g., proangiotensin-H_2O with calculated mass, 1312.5 Da and observed mass, 1312.5 Da).

Insulin

Recombinant human, monomeric insulin was purchased from a commercial supplier (MP Biomedicals). Prior to bioconjugation reactions, sample purity was examined by ESI LC-MS analysis (**Figure 4-23**).

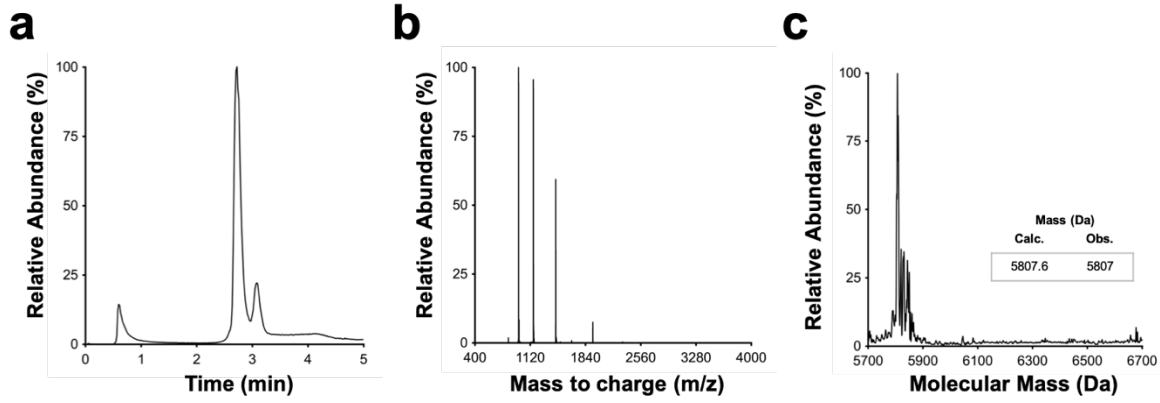


Figure 4-23. Insulin purity assessed by ESI LC-MS analysis. Representative **A)** TIC, **B)** combined *m/z* spectrum, and **C)** deconvoluted mass spectrum of insulin. The minor peaks most likely represent insulin hydrates (e.g., insulin-H_2O with calculated mass, 5823.5 Da and observed mass, 5824.5 Da).

DHFR

A pET46 Ek/LIC (Novagen) plasmid containing the gene encoding *E. coli* DHFR was sub-cloned after mutations to remove the native cysteines (C85A/C152S) and fuse an N-terminal His_6 peptide epitope; this gene product is referred to hereafter as DHFR. The experimental details to express and purify DHFR and the single-cysteine variant, DHFR-Cys (N37C), have been previously described.[130] Following size exclusion FPLC, fractions were analyzed for purity using 15% SDS PAGE via ImageJ (**Figure 4-24**). The concentration of DHFR was determined by a bicinchoninic acid (BCA) assay using 19664.03 g/mol as the expected MW of the protein (http://www.expasy.org) and the protein was either used immediately or mixed with glycerol (20% final concentration) for storage at -80 °C.

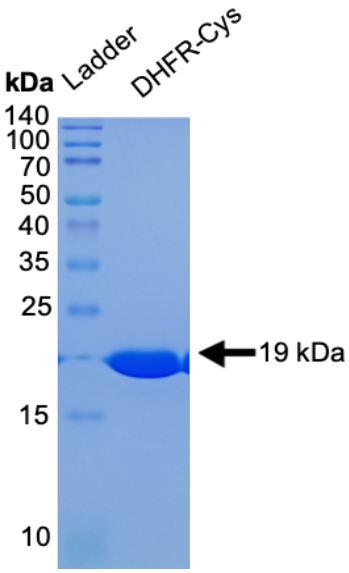


Figure 4-24. DHFR-Cys purity assessed by 15% SDS-PAGE. This representative SDS-PAGE visualizes DHFR-Cys after size exclusion chromatography. The protein was purified to >95% homogeneity, and migrated at its expected mass of ≈19 kDa.

HSA

Albumin from human serum was purchased from a commercial supplier (Sigma). Prior to bioconjugation reactions, sample purity was examined by ESI LC-MS analysis (**Figure 4-25**).

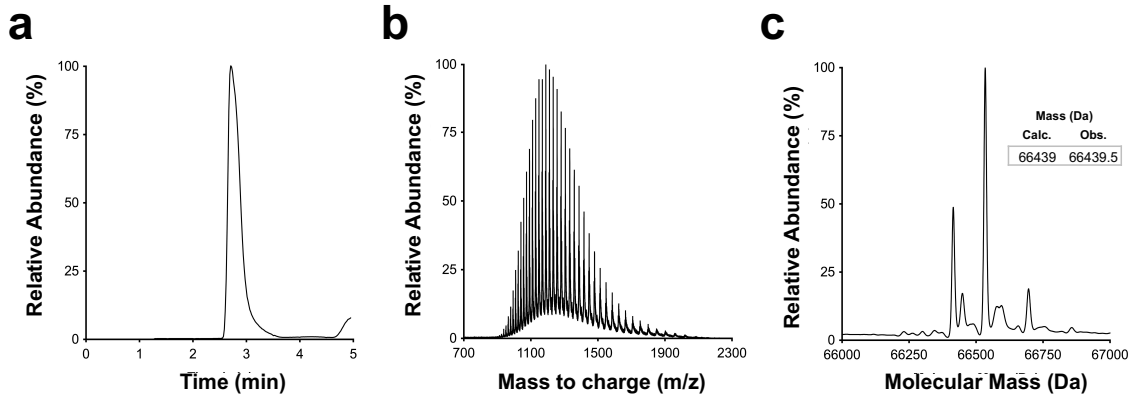


Figure 4-25. HSA purity assessed by ESI LC-MS analysis. Representative **A)** TIC, **B)** combined m/z spectrum, and **C)** deconvoluted mass spectrum of HSA. The other major peaks most likely represent HSA with varying degrees of post-translational modifications (e.g., HSA-cysteinylation with calculated mass, 66,557 Da and observed mass, 66,559 Da; HSA-glycosylated with calculated mass, 66,600 Da and observed mass, 66600.5 Da).

KF

A pET28c plasmid containing the gene encoding *E. coli* DNA polymerase I Klenow fragment (KF) was mutated to remove the polymerase's exonuclease activity (D335A/E357A) and a native cysteine (C907S); DNA encoding an *N*-terminal fusion of His_6 peptide epitope was added. A previous report described the experimental details to express and purify KF and the single-cysteine variant, KF-Cys (L790C).[91] Following size exclusion FPLC, protein purity was assessed using 15% SDS PAGE via ImageJ (**Figure 4-26**). The protein concentration was determined by a bicinchoninic acid (BCA) assay using 69387.29 g/mol as the estimated MW of the protein (http://www.expasy.org). KF was either used immediately, or mixed with glycerol (20% final concentration) for storage at -80 °C.

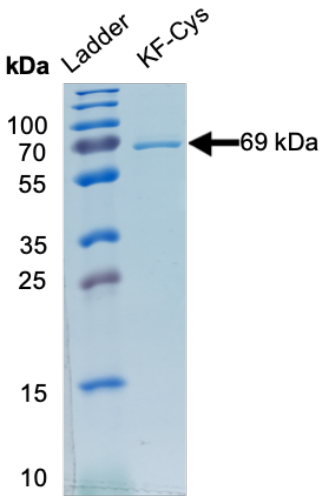


Figure 4-26. KF purity assessed by 15% SDS-PAGE. This representative SDS-PAGE visualizes KF-Cys after size exclusion FPLC. KF was purified to >95% homogeneity, and migrated at its expected mass of ≈69 kDa.

Φ29

A pET15b plasmid having an ORF encoding the *B. subtillis* bacteriophage phi29 DNA polymerase I (a kind gift from Illumina) was mutated to remove the polymerase's exonuclease activity (D12A/D66A) and to fuse an N-terminal His_6 peptide epitope; this gene product is hereafter referred to as Φ29. Φ29 was transformed into BL21 (DE3) *E. coli* by the heat shock method. The transformed cells were transferred to an LB agar plate supplemented with carbenicillin (50 µg/mL), and incubated at 37 °C for 14 to 16 h. A seed culture was prepared by inoculating a single colony from the transformation plate into 25 mL of 2YT media in a 125 mL baffled flask with carbenicillin (50 mg/mL) and incubating the culture at 37 °C with shaking at 225 rpm for 14-16 h. The expression culture was then prepared by inoculating 10 mL of the seed culture into LB media (1.0 L in a 2.0 L baffled flask) supplemented with carbenicillin (50 mg/mL) and shaking the culture at 225 rpm at 37 °C. When the density of the culture reached an OD_{600} value of 0.6 to 0.7, the overexpression of Φ29 was induced through the addition of isopropyl β-D

thiogalactopyranoside (IPTG, 0.25 mM) before further incubation for 24 h at 16 °C with shaking at 225 rpm. The cells were harvested by centrifugation (6080 rcf, 20 min, 4 °C), and resuspended in lysis buffer (50 mM Tris HCl, 300 mM NaCl, 10 mM $MgCl_2$, 0.8 M trehalose, 10 mM $CaCl_2$, 1 mM phenylmethanesulfonyl fluoride – PMSF, and 10 mM β-mercaptoethanol – BME, pH 7.5).

The cell lysate was prepared by sonication of the cells using a Q500 Sonicator with the Standard Probe 4220 (QSonica, USA), temperature <8 °C, and 4 pulses (40% amplitude with 1 s per pulse followed by 3 s rest for 1 min), then incubated with DNase I (0.5 U/μL) for 30 min at 37 °C. Following incubation for 10 min at 65 to 70 °C to inactivate the DNase, the cell lysate was subjected to centrifugation (26,892 rcf, 45 min, 4 °C) to remove cell debris. The lysate supernatant was filtered through a 0.45 μm pore filter, and then incubated with a Ni^{2+}-NTA column pre-equilibrated with lysis buffer. The supernatant mixed with Ni^{2+}-NTA resin was incubated overnight on a slow rotatory shaker at 4 °C. The column was washed with wash buffer (50 mM Tris HCl, 300 mM NaCl, 10 mM $MgCl_2$, 20 mM Imidazole, 10 mM BME, pH 7.5) and the purified Φ29 protein was eluted using elution buffer (50 mM Tris HCl, 300 mM NaCl, 10 mM $MgCl_2$, 250 mM imidazole, 10 mM BME, pH 7.5). Elutions containing the purified protein were visualized using 12% SDS-PAGE (Bio-rad Mini-PROTEAN Tetra electrophoresis system) stained with Coomassie brilliant blue stain. The eluted fractions containing the purified Φ29 were pooled and concentrated using a 10 kDa cutoff microconcentrator (Vivaspin, Fisher Scientific, USA). The mixture was centrifuged, and then filtered through a 0.22 μm pore filter prior to size exclusion FPLC (Bio-Rad NGC Chromatography System, Superdex 75 pg, 26/600 at a flow rate of 1.0 mL. min^{-1}) in activity buffer (40 mM HEPES, 300 mM NaCl, 10 mM $MgCl_2$, 100 μM

Tris(2-carboxyethyl)phosphine) (TCEP) at pH 6.5. The purity (≥95%) of the protein was assessed by 12% SDS-PAGE stained with Coomassie brilliant blue stain (**Figure 4-27)** followed by ImageJ quantification. For Φ29 assays the purified recombinant protein was dialyzed in high salt buffer (40 mM HEPES, 300 mM NaCl, 10 mM $MgCl_2$, pH 6.5). The protein concentration was determined by a bicinchoninic acid (BCA) assay using 68,558.15 g/mol as the estimated MW of the protein (http://www.expasy.org).

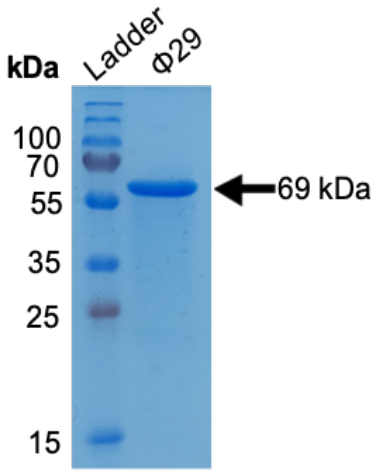


Figure 4-27. Φ29 purity assessed by 12% SDS-PAGE. This representative SDS-PAGE visualizes Φ29 after size exclusion chromatography. Φ29 was purified to >95% homogeneity, and migrated at its expected mass of ≈69 kDa.

Taq

A pET28c plasmid for ligation-independent cloning containing an ORF encoding *Thermus aquaticus* DNA polymerase fused to an N-terminal His_6 peptide epitope was mutated to remove the polymerase's exonuclease activity (G46D) and introduce a single cysteine (R965C); hereafter, this DNA polymerase is referred to as Taq. The plasmid encoding Taq was transformed into BL21 Star (DE3) *E. coli* heat shock, competent cells. A single colony was transferred to LB media (25 mL) supplemented with kanamycin (50 µg/mL), and incubated at 37 °C for 6 to 8 h. An aliquot of the starter culture (10 mL) was

transferred to LB media (1 L LB in a 4 L baffled flask). After reaching an OD_{600} of ~0.6, the culture was induced through addition of IPTG (1 mM) before incubation at 16 °C for 18 h. The cells were collected and resuspended in Taq resuspension buffer (50 mM Tris-HCl, 300 mM NaCl, 25 mM KCl, and 10 mM BME, pH 7.8). Cells were lysed with Taq lysis buffer (50 mM Tris-HCl, 25 mM KCl, 50 mM glucose, 0.25% Tween, 10 mM BME, 1 mM PMSF, and protease inhibitor cocktail – EDTA-free 100X, ThermoFisher) followed by sonication. DNAse (ThermoFisher) was added to the cell lysate, and the solutions was incubated at 37 °C for 1 h to eliminate DNA bound to the polymerase. Heating at 70 °C for 30 min deactivated the DNase. The DNase-treated lysate was subjected to centrifugation (26,892 rcf, 45 min, 4 °C). Supernatant was incubated with charged Ni-NTA resin and bound overnight on a shaker (150 rpm at 4 °C). Resin was packed into a column, where the flow-through, and washes were collected followed by a gradient imidazole solution (25 to 250 mM imidazole). After pooling and concentrating (by microconcentration with a 30 kDa MWCO) the eluted protein, the sample was further purified by size exclusion chromatography (Bio-Rad NGC Chromatography System, Superdex 200 pg, 16/600 at a flow rate of 1.0 mL. min^{-1}) in Taq activity buffer (20 mM Tris-HCl, 100 mM KCl, 0.1 mM EDTA, 1 mM DTT, 0.50% Tween20, pH 7.4). The eluted fractions were collected and visualized by 12% SDS-PAGE (**Figure 4-28**).

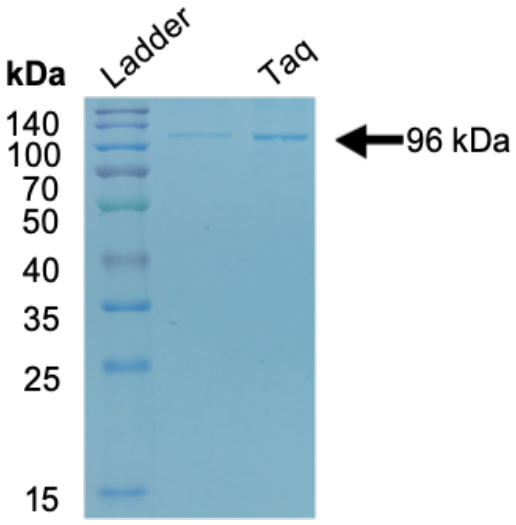


Figure 4-28. Taq purity assessed by 12% SDS-PAGE. This representative SDS-PAGE visualizes Taq after size exclusion chromatography. Taq was purified to >85% homogeneity, and migrated at its expected mass of ≈96 kDa.

Trastuzumab

The Trastuzumab antibody (Genentech) was a generous gift from Dr. Edward Nelson (UCI). Prior to bioconjugation reactions, sample homogeneity was evaluated by ESI LC-MS analysis (**Figure 4-29**).

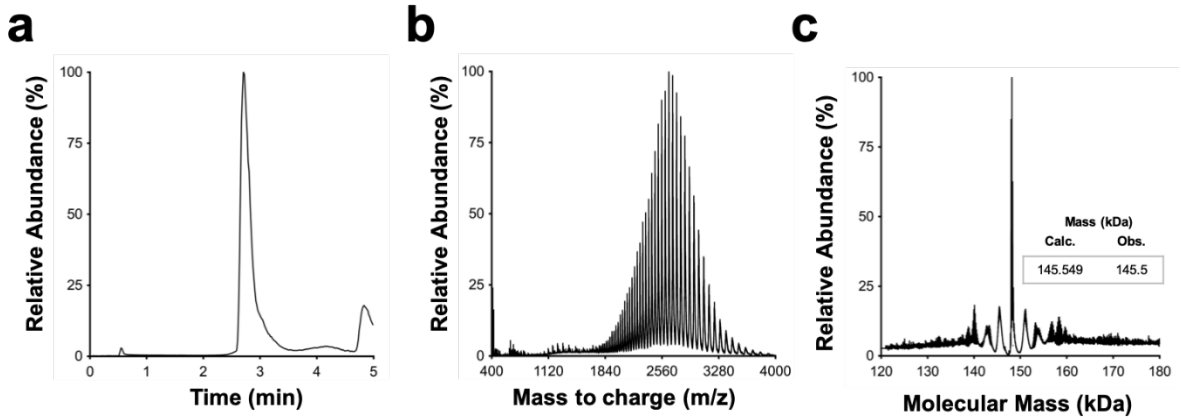


Figure 4-29. Trastuzumab purity assessed by ESI LC-MS analysis. Representative **A)** TIC, **B)** combined *m/z* spectrum, and **C)** the deconvoluted mass spectrum of Trastuzumab. The observed product may reflect hydration (three H_2O, calculated mass addition +54 Da, observed mass addition +56.6 Da).

ESI LC-MS analysis of protein-Pci bioconjugates

Transpyrocinchonimidation products were quantified by ESI LC-MS (ACQUITY UPLC H-class system, Xevo G2-XS QTof, Waters Corporation). The protein was separated from reaction buffer using a phenyl guard column at 45 °C (ACQUITY UPLC BEH Phenyl VanGuard Pre-column, 130 Å, 1.7 µm, 2.1 mm X 5 mm, Waters). The 5 min analytical method used 0.2 mL/min flow rate of a gradient of Buffer A consisting of 0.1% formic acid in water and Buffer B, 100% acetonitrile (**Table 4-2**). The Xevo Z-spray source was operated in positive MS resolution mode, 400-4000 Da, with a capillary voltage of 3000 V and a cone voltage of 40 V (NaI / CsI and Leu-enkephalin lock-mass calibration).[131] Nitrogen was used as the desolvation gas at 350 °C at a flow-rate of 800 L/h. Total average mass spectra were reconstructed from the charge state ion series using the MaxEnt1 algorithm from Waters MassLynx software V4.1 SCN949 according to the manufacturer's instructions. The parameters for the deconvolution of insulin and insulin-**Pci** were 0.05 Da peak separation, and the combined ion series for 650-2300 *m*/*z* was processed under MaxEnt1 using an acquisition range of 5700-7300 Da, 0.5 Da/channel resolution, Uniform Gaussian with 0.5 Da width at half height, minimum intensity ratios of 33% for left and right, and 12 maximum number of iterations. The deconvolution of trastuzumab and trastuzumab-**Pci** applied the following modifications to the parameters mentioned above: combined ion series for *m*/*z* 1800-4000 processed under MaxEnt1 using 145,000-150,000 Da range and 1.0 Da/channel resolution. A representative total ion chromatogram (TIC), combined *m/z* spectrum, and deconvoluted spectrum for insulin and trastuzumab can be found in **Figure 4-23** and **Figure 4-29**. Data files were exported in .txt format for further processing (GraphPad Prism, v. 8.0) and

quantification (MS-Excel). A typical semi-quantitative assessment of the relative abundances (%) as measured by the ESI LC-MS analysis is illustrated in **Figure 4-30**. This method was applied to the analysis of Pci bioconjugates with both insulin and trastuzumab.

Table 4-2. Programmed gradient for Buffer A and Buffer B used during ESI LC-MS.

Time (mins)	Flow (mL/min)	Buffer A (%)	Buffer B (%)	Curve
Initial	0.2	100	0	6
0.5	0.2	100	0	6
2.0	0.2	10	90	6
2.5	0.2	10	90	6
4.0	0.2	100	0	6
5.0	0.2	100	0	6

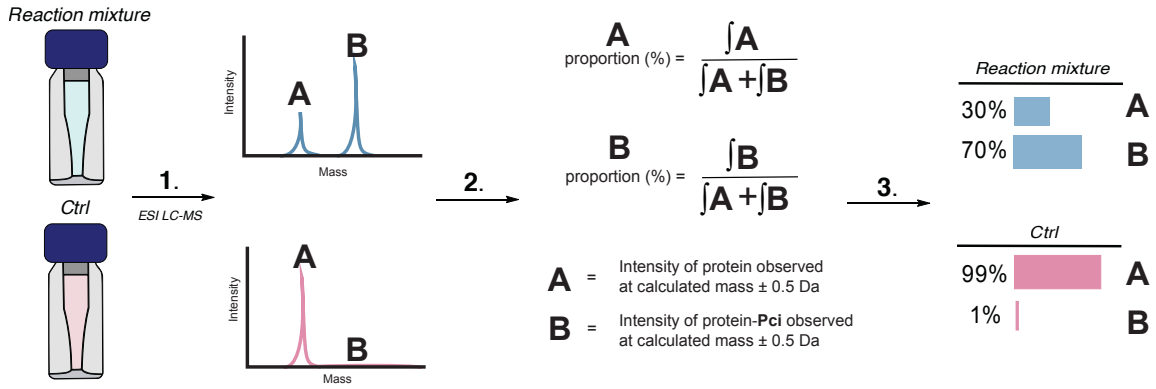


Figure 4-30. A typical semi-quantitative estimation of the product mixture from the transpyrocinchonimidation transfer reaction. 1) Mass spectrometry vials with the protein-**Pci** bioconjugate (reaction mixture) or protein (Ctrl) were analyzed by ESI LC-MS. **2)** The intensities of **A)** starting material and **B)** bioconjugate peaks from the deconvoluted mass spectra were integrated over a ±0.5 Da range (ESI LC-MS) or ±1.0 Da range (MALDI) of their theoretical masses, and expressed as total relative abundances. **3)** The reported values are the proportion (%) of the total measured peak abundances from this analysis. The control samples (Ctrl) were not exposed to the **Pci** transfer reagents; this measurement can assess background noise in the mass ranges of the detected bioconjugates, and helps confirm product detection. This analysis assumed equal ionizability for starting insulin and **Pci-**modified products, and was repeated in triplicate (n = 3) for insulin and one time (n = 1) for Trastuzumab.

MALDI analysis of insulin-Pci bioconjugate digests, insulin-Dhzd-mPEG bioconjugates, and insulin-biotin bioconjugates

An ABI SCIEX TOF/TOF 5800 (Applied Biosystems/MDS Analytics Tech) was used to perform matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF). A saturated solution (20 mg/mL) of α-cyano-4-hydroxycinnamic acid (CHCA, Sigma Aldrich) or 2,5-Dihydroxybenzoic acid (DHB, Sigma Aldrich) dissolved in a 1:1 solution of 0.1% trifluoroacetic acid (TFA) and acetonitrile (CH_3CN) in water was used as the matrix for the MALDI. Different MALDI analysis conditions were required to insure the resultant bioconjugates appeared in a MALDI-relevant MW range (Table 4-3).

Each sample of 0.5 µL was spotted, mixed with the CHCA or DHB matrix on the MALDI plate, and analyzed using a solid-state 355 nm YAG laser operated at 400 Hz with a variety of laser powers as shown. The instrument was operated in positive reflector or positive linear mode across various mass ranges (700 to 20,000 Da). The data was acquired from 3 or 10 summed spectra (see **Table 4-3**) in an automated random walk pattern across the spot to generate 2000 shots of data for each spectra. The data was acquired and processed by the TOF/TOF series explorer software (V4.0, build 8, ABI Sciex Corp). A negative control that omitted modifiers (**27,** mPEG-alkene, **47,** or **48)** in HEPES (40 mM, pH 6.5 or pH 8.5) showed no conversion to the expected product (yielding insulin's observed mass, 5807.5 Da; calculated mass 5808 Da).

Table 4-3. MALDI conditions used for insulin and insulin conjugate characterization. Different conditions were required to ensure the insulin-Pci digests (**Figure 4-10**), insulin-

Dhzd-mPEG (**Figure 4-19**), and insulin-biotin (**Figure 4-21**) appeared in a MALDI-relevant MW range.

Bioconjugates	insulin-Pci digests	insulin-Dhzd-mPEG	insulin-Dhzd-47-Tb	insulin-biotin
Matrix	CHCA	DHB	DHB	DHB
0.1% TFA	½ dilution	½ dilution	½ dilution	1/10 dilution
Laser	4300 or 5500	6300	6300	5000
Acquisition mode	Reflector Pos.	Linear Mid.	Linear Mid.	Linear Mid.
Range (Da)	700-10,000	2,000-8,000	2,000-8,000	2,000-20,000
Spectra summed	10	10	10	3

Insulin-Pci hydrolysis and Pci-exchange reactions

The dialyzed reaction mixture containing insulin-**Pci** bioconjugate (30 µL at a final concentration of ~20 µM) in HEPES (40 mM, pH 7.2) was added to a 1.7 mL Eppendorf tube. To study the insulin-Pci hydrolysis at pH 7.2, a stock solution of ethanolamine (30 µL, 25 or 250 µM), proangiotensin (30 µL, 20 or 200 µM), or HSA (30 µL, 20 or 500 µM), in HEPES (40 mM, pH 7.2) was added. To study the pH-dependent hydrolysis of the insulin Pci groups, a stock solution of HEPES (40 mM, 30 µL) at elevated pH (pH 8.5, 8.7, 9.7, 10.9, or 11.7) was added. The resulting mixtures were mixed via pipette for 3 s and incubated for 1-22 days at rt in the dark. After the reaction time elapsed, the reaction mixture (30 µL) was transferred to 1 mL glass mass spectrometry vial and analyzed by ESI-LCMS.

Different instruments and parameters were required to ensure the detection of insulin-**Pci** hydrolysis and **Pci**-exchange reactions in a relevant MW range. The insulin-**Pci** hydrolysis products were quantified using ESI LC-MS (ACQUITY UPLC H-class system, Xevo G2-XS QTOF, Waters Corporation) as described in the *General Experimental Methods section.* The **Pci**-exchange reactions with ethanolamine were

analyzed by ESI LC-MS (ACQUITY UPLC H-class system, Synapt G2 QTOF, Waters Corporation) using direct injection delivered at a flow rate of 0.1 mL/min in positive ion mode. A stand mixture of ethanolamine (400 nM) and **27** (150 nM) provided an accurate quantification of the anhydro-derivative, *N*-vinyl pyrocinchonimide. The N-vinyl-**Pci** products were quantified in the combined ion series for *m*/*z* 50-1000. The **Pci**-exchange reactions with proangiotensin and HSA were quantified by ESI LC-MS (ACQUITY UPLC H-class system, Xevo G2-XS QTOF, Waters Corporation) as described in the *General Experimental Methods section* with the following modifications: deconvolution of proangiotensin and proangiotensin-**Pci** applied combined ion series for *m*/*z* 400-1000 processed under MaxEnt1 using 1,000-2,000 Da range; deconvolution of HSA and HSA-**Pci** applied combined ion series for *m*/*z* 650-2300 processed under MaxEnt1 using 55,000-75,000 Da range. A representative total ion chromatogram (TIC), combined *m/z* spectrum, and deconvoluted spectrum for proangiotensin and HSA can be found in the **Figure 4-22** and **Figure 4-25***.*

*Fluorescence characterization of insulin-**46***

For the fluorescence characterization by UPLC-MS (**Figure 4-15**), the insulin-**46** bioconjugate (50 µL) was placed in glass, 2 mL mass spectrometry vials. Next, an injection (10 µL) was analyzed by UPLC-PDA on an ACQUITY UPLC (FLR Detector, Waters corporation). A semi-quantitative assessment of the product mixture of the insulin-**46** bioconjugate followed the description presented above (**Figure 4-30**). Fluorescence parameters were λ_{ex} 320 nm and λ_{em} 350-500 nm.

*Activity Assay of Taq-**46** bioconjugate*

To confirm activity of Taq-**46** bioconjugate, a previously described ensemble activity assay[90] was adapted as follows. The following oligonucleotides were used for the M13F primer and DNA template for Taq activity:

M13F primer: 5'-TGTAAAACGACGGCCAGT-3'

Poly T_{42}: 5' – TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTACTGGCCGT CGTTTTACA – 3'

The HPLC-purified oligonucleotides were solubilized in water to 100 μM and hybridized prior to Taq polymerization. The positive control reactions contained Taq (2 μM), dATP (100 μM), and Poly T_{42} DNA template-primer (5 μM) in 1x Taq Reaction Buffer (20 mM Tris-HCl, 50 mM KCl, 5 mM $MgCl_2$, pH 8.3). The negative control reaction omitted Taq. Reactions were incubated at 72 °C for 1 h, prestained with SYBR Green Nucleic Acid I stain, then electrophoresed on a 5% high resolution agarose gel. Gels were imaged using UV-light (Digital camera E-M5 Mark II, Olympus, USA and Multi Doc-it Digital Imaging System, Analytik Jena, USA).

Attachment of Taq-46 bioconjugate to SWCNT-FET

A SWCNT-FET was fabricated using techniques described previously.[90] In brief, the SWCNT-FET was grown on SiO_2, connected to source and drain metal electrodes and passivated with a polymer (PMMA). The SWCNT-FET was incubated at 275 °C in air for 30 min, soaked in PG remover at 65 °C for 10 min, rinsed with IPA and DI water, then

placed on a hot plate at 135 °C for 5 min. Attachment of Taq to SWCNT-FETs was accomplished by soaking the devices in a solution of **46** (100 pM) and 1 μM of pyrene in ethanol for 5 min, then rinsing with ethanol, IPA, DI water, and Taq activity buffer (20 mM Tris-HCl, 50 mM KCl, 5 mM $MgCl_2$, pH 8.3) for 5 min. Following a rinse with DI water, a solution of 4 nM of Taq in Taq activity buffer was incubated at RT for 5 min. Such devices are referred to simply as Taq nanocircuits.

Cycloaddition reactions of insulin-Pci bioconjugates

The dialyzed reaction mixture containing insulin-**Pci** bioconjugate (100 μL at a final concentration of ~10 μM) was added to a sealed, 2 mL glass mass spectrometry vial. Afterwards, stock solution (100 μL, 10 mM) of mPEG-alkene-1K or **47** in HEPES (40 mM, pH 8.5) was added, and the resulting mixture was mixed via pipette for 3 s. The reaction was incubated for 4 h under UVC light at ≤8 °C using a cooling water jacket. After the reaction time elapsed, the reaction mixture (100 μL) was dialyzed using a Slide-A-Lyzer mini dialysis device (3.5K MWCO) with stirring overnight at 4 °C to remove unreactive, excess mPEG-alkene or **47**. The dialyzed reaction mixture was transferred to a 1.7 mL Eppendorf tube and a 1 μL aliquot was analyzed by MALDI.

Protein gels

SDS-PAGE loading dye (5x, 4 μL) was mixed with the protein sample (16 μL). The solution was heated at 95 °C for 5 min, then 10 μL was visualized by SDS-PAGE with 12 or 15% polyacrylamide concentration. The Precision Plus Kaleidoscope Standards Fluorescent Ladder (Biorad) and PageRuler Plus Prestained Protein Ladder (ThermoFisher) were used to estimate protein sizes. After electrophoresis (80 V for 15

min, then 180 V for 30 min) in 1x SDS Running Buffer (25 mM Tris-base, 192 mM glycine, 0.1% SDS, pH 8.3), the gel was stained with Coomassie blue (3.2 mM brilliant blue, 45% ethanol, 10% glacial acetic acid in deionized water) for 30 min and destained using destaining buffer (40% methanol, 10% acetic acid in deionized water) at room temperature overnight.

4.6 Supporting Information

Synthesis and characterization of compounds **1-47**, NMR and IR spectra for compounds **1**−**47**, and crystallographic data for **18**, **27** and **44** (CIF) is available at https://pubs.acs.org/doi/10.1021/acs.bioconjchem.0c00143.

4.7 References

(1) Ravasco, J. M. J. M.; Faustino, H.; Trindade, A.; Gois, P. M. P. Bioconjugation with Maleimides: A Useful Tool for Chemical Biology. *Chem. - Eur. J.* **2019**, *25*, 43–59.

(2) Nair, D. P.; Podgórski, M.; Chatani, S.; Gong, T.; Xi, W.; Fenoli, C. R.; Bowman, C. N. The Thiol-Michael Addition Click Reaction: A Powerful and Widely Used Tool in Materials Chemistry. *Chem. Mater.* **2014**, *26*, 724–744.

(3) Baldwin, A. D.; Kiick, K. L. Tunable Degradation of Maleimide-Thiol Adducts in Reducing Environments. *Bioconjug. Chem.* **2011**, *22*, 1946–1953.

(4) Fontaine, S. D.; Reid, R.; Robinson, L.; Ashley, G. W.; Santi, D. V. Long-Term Stabilization of Maleimide-Thiol Conjugates. *Bioconjug. Chem.* **2015**, *26*, 145–152.

(5) Szijj, P. A.; Bahou, C.; Chudasama, V. Minireview: Addressing the Retro-Michael Instability of Maleimide Bioconjugates. *Drug Discov. Today Technol.* **2018**, *30*, 27–34.

(6) Gregory, J. D. The Stability of N-Ethylmaleimide and Its Reaction Wth Sulfhydryl Groups. *J. Am. Chem. Soc.* **1955**, *77*, 3922–3923.

(7) Finnegan, R. A.; Mueller, W. H. Base-Catalyzed Addition and Solvolysis Reactions of N-Phenylmaleimide in Methanol. *J. Pharm. Sci.* **1964**, *54*, 1257–1260.

(8) Machida, M.; Machida, M. I.; Kanaoka, Y. Hydrolysis of N-Subsituted Maleimides: Stability of Fluorescence Thiol Reagents in Aqueous Media. *Chem. Pharm. Bull.* **1977**, *25*, 2739–2743.

(9) Kalia, J.; Raines, R. T. Catalysis of Imido Group Hydrolysis in a Maleimide Conjugate. *Bioorganic Med. Chem. Lett.* **2007**, *17*, 6286–6289.

(10) Tan, X. J.; Cheng, S. S.; Shi, Y.; Xing, D. X.; Liu, Y.; Li, H.; Feng, W. Q.; Yang, J. B. Hydrolytic Degradation of N,N′-Ethylenedimaleimide: Crystal Structures of Key Intermediates and Proposed Mechanisms. *J. Mol. Struct.* **2016**, *1125*, 514–521.

(11) Shafer, D. E.; Inman, J. K.; Lee, A. Reaction of Tris(2-Carboxyethyl)Phosphine (Tcep) with Maleimide and A- Haloacyl Groups: Anomalous Elution of Tcep by Gel Filtration. *Anal. Biochem.* **2000**, *282*, 161–164.

(12) Kantner, T.; Watts, A. G. Characterization of Reactions between Water-Soluble Trialkylphosphines and Thiol Alkylating Reagents: Implications for Protein-Conjugation Reactions. *Bioconjug. Chem.* **2016**, *27*, 2400–2406.

(13) Wei, C.; Zhang, G.; Clark, T.; Barletta, F.; Tumey, L. N.; Rago, B.; Hansel, S.; Han, X. Where Did the Linker-Payload Go? A Quantitative Investigation on the Destination of the Released Linker-Payload from an Antibody-Drug Conjugate with a Maleimide Linker in Plasma. *Anal. Chem.* **2016**, *88*, 4979–4986.

(14) Ohri, R.; Bhakta, S.; Fourie-O'Donohue, A.; Dela Cruz-Chuh, J.; Tsai, S. P.; Cook, R.; Wei, B.; Ng, C.; Wong, A. W.; Bos, A. B.; Farahi, F.; Bhakta, J.; Pillow, T. H.; Raab, H.; Vandlen, R.; Polakis, P.; Liu, Y.; Erickson, H.; Junutula, J. R.; Kozak, K. R.; et al. High-Throughput Cysteine Scanning to Identify Stable Antibody Conjugation Sites for Maleimide- and Disulfide-Based Linkers. *Bioconjug. Chem.* **2018**, *29*, 473–485.

(15) Nunes, J. P. M.; Vassileva, V.; Robinson, E.; Morais, M.; Smith, M. E. B.; Pedley, R. B.; Caddick, S.; Baker, J. R.; Chudasama, V. Use of a Next Generation Maleimide in Combination with Thiomab Antibody Technology Delivers a Highly Stable, Potent and near Homogeneous Thiomab Antibody-Drug Conjugate (Tdc). *RSC Adv.* **2017**, *7*, 24828–24832.

(16) Schumacher, F. F.; Nunes, J. P. M.; Maruani, A.; Chudasama, V.; Smith, M. E. B.; Chester, K. A.; Baker, J. R.; Caddick, S. Next Generation Maleimides Enable the Controlled Assembly of Antibody-Drug Conjugates via Native Disulfide Bond Bridging. *Org. Biomol. Chem.* **2014**, *12*, 7261–7269.

(17) Forte, N.; Livanos, M.; Miranda, E.; Morais, M.; Yang, X.; Rajkumar, V. S.; Chester, K. A.; Chudasama, V.; Baker, J. R. Tuning the Hydrolytic Stability of Next Generation Maleimide Cross-Linkers Enables Access to Albumin-Antibody Fragment Conjugates and Tri-ScFvs. *Bioconjug. Chem.* **2018**, *29*, 486–492.

(18) Hull, E. A.; Livanos, M.; Miranda, E.; Smith, M. E. B.; Chester, K. A.; Baker, J. R. Homogeneous Bispecifics by Disulfide Bridging. *Bioconjug. Chem.* **2014**, *25*, 1395–1401.

(19) Lee, M. T. W.; Maruani, A.; Baker, J. R.; Caddick, S.; Chudasama, V. Next-Generation Disulfide Stapling: Reduction and Functional Re-Bridging All in One. *Chem. Sci.* **2016**, *7*, 799–802.

(20) Marculescu, C.; Kossen, H.; Morgan, R. E.; Mayer, P.; Fletcher, S. A.; Tolner, B.; Chester, K. A.; Jones, L. H.; Baker, J. R. Aryloxymaleimides for Cysteine Modification, Disulfide Bridging and the Dual Functionalization of Disulfide Bonds. *Chem. Commun.* **2014**, *50*, 7139–7142.

(21) Morais, M.; Nunes, J. P. M.; Karu, K.; Forte, N.; Benni, I.; Smith, M. E. B.; Caddick, S.; Chudasama, V.; Baker, J. R. Optimisation of the Dibromomaleimide (Dbm) Platform for Native Antibody Conjugation by Accelerated Post-Conjugation Hydrolysis. *Org. Biomol. Chem.* **2017**, *15*, 2947–2952.

(22) Nunes, J. P. M.; Morais, M.; Vassileva, V.; Robinson, E.; Rajkumar, V. S.; Smith, M. E. B.; Pedley, R. B.; Caddick, S.; Baker, J. R.; Chudasama, V. Functional Native Disulfide Bridging Enables Delivery of a Potent, Stable and Targeted Antibody-Drug Conjugate (ADC). *Chem. Commun.* **2015**, *51*, 10624–10627.

(23) Pye, H.; Butt, M. A.; Reinert, H. W.; Maruani, A.; Nunes, J. P. M.; Marklew, J. S.; Qurashi, M.; Funnell, L.; May, A.; Stamati, I.; Hamoudi, R.; Baker, J. R.; Smith, M. E. B.; Caddick, S.; Deonarain, M. P.; Yahioglu, G.; Chudasama, V.; Lovat, L. B. A HER2 Selective Theranostic Agent for Surgical Resection Guidance and Photodynamic Therapy. *Photochem. Photobiol. Sci.* **2016**, *15*, 1227–1238.

(24) Hong, L. P. T.; Scoble, J. A.; Doughty, L.; Coia, G.; Williams, C. C. Cancer-Targeting Antibody-Drug Conjugates: Site-Specific Conjugation of Doxorubicin to Anti-EGFR 528 Fab′ through a Polyethylene Glycol Linker. *Aust. J. Chem.* **2011**, *64*, 779–789.

(25) Earl, R. A.; Clough, F. W.; Townsend, L. B. Chemical Investigations of Citraconimide. *J. Heterocylcic Chem.* **1977**, *15*, 1479–1483.

(26) Miyaderaf, T.; Kosower, E. M. Receptor Site Labeling Through Functional Groups. 2. Reactivity of Maleimide Groups. *J. Med. Chem.* **1972**, *15*, 534–537.

(27) Zhang, Y.; Zhou, X.; Xie, Y.; Greenberg, M. M.; Xi, Z.; Zhou, C. Thiol Specific and Tracelessly Removable Bioconjugation via Michael Addition to 5-Methylene Pyrrolones. *J. Am. Chem. Soc.* **2017**, *139*, 6146–6151.

(28) Bahou, C.; Spears, R. J.; Aliev, A. E.; Maruani, A.; Fernandez, M.; Javaid, F.; Szijj, P. A.; Baker, J. R.; Chudasama, V. Use of Pyridazinediones as Extracellular

Cleavable Linkers through Reversible Cysteine Conjugation. *Chem. Commun.* **2019**, *55*, 14829–14832.

(29) Aly, M. R. E.; Castro-Palomino, J. C.; Ibrahim, E.-S. I.; El-Ashry, E.-S. H.; Schmidt, R. R. The Dimethylmaleoyl Group as Amino Protective Group. Application to the Synthesis of Glucosamine-Containing Oligosaccharides. *European J. Org. Chem.* **1998**, 2305–2316.

(30) Chen, W.; Zhu, Y.; Geng, X. Synthesis Method of 3-Amino Pyrrolidine Dihydrochloride. CN109349493A, 2018.

(31) Acharya, S.; Panda, S. K.; Das, P.; Agarwal, B. Novel Compositions and Therapeutic Methods. WO2017173024A1, 2017.

(32) N-Substituted Cis-Butene Imidodicarbonic Diamide Com- Pound, Preparation Method Thereof and Antibacterial Application. CN107033060B, 2020.

(33) Chen, X.; Lu, Y.; Li, Z.; Fan, Y.; Zhang, L.; Li, Y.; Shen, Y. N-Substitution Maleimide Compounds and Its Preparation Based on Pharmaceutical Intermediate Are Studied with Antibacterial Activity. CN107162950A, 2017.

(34) Roth, M. 2,3-Dimethylmaleimido-Alkyl Haloacetates. US4656292A, 1987.

(35) Baumann, M.; Kvita, V.; Roth, M.; Waterhouse, J. S. Imidyl Compounds. US4107174A, 1978.

(36) Lohmann, D.; Wyler, S. Silanes Containing Imide Groups. US4271074A, 1981.

(37) Kume, T.; Goto, T.; Kamochi, A.; Yamaguchi, N.; Yanagi, A.; Hayakawa, H.; Yagi, S. 1-(1,4-Benzoxazin-3-on-6-Yl)-Dialkylma- Leimides and Use as Herbicides. US4729784A, 1988.

(38) Kume, T.; Goto, T.; Kamochi, A.; Yanagi, A.; Yagi, S.; Miyauchi, H. 2,5-Dihydropyrroles. US4828604A, 1989.

(39) Moser, H.; Pissiotas, G.; Brunner, H.-G.; Bohner, B.; Baumann, M. N-Phenyl-Maleimides and Herbicidal and Plant Growth Regulating Methods of Use Thereof. US4804400A, 1989.

(40) Kohsaka, H.; Takase, M. Butenoic Acid Derivatives and Use as Herbicides. US5059237A, 1991.

(41) Arabori, H.; Yamazaki, S.; Arahira, M.; Murakami, A. N. Substituted-3-(Nitrogen-Containing 5-Membered Ring)-Benzenesul- Fonamide Derivatives, and Herbicidal Compositions. US5127937A, 1992.

(42) Wepplo, P. J. Substituted Benzisoxazole and Benzisothiazole Herbicidal Agents. US5484763A, 1996.

(43) Lachia, M. D.; Screpanti, C.; De Mesmaeker, A.; Lumbroso, A. F. J. C.; Rendine, S. Plant Growth Regulating Compounds. US9751867B2, 2017.

(44) Morris, J. A.; Hennessy, A. J.; Boehmer, J. E. Herbicidal Pyrrolone Derivatives. WO2016071360A1, 2016.

(45) Suzuki, N.; Nihei, Y.; Ichinose, H.; Tanaka, H.; Yasa, N.; Hatanaka, T.; Masuzawa, Y.; Nakanishi, E.; Kondo, N. Benzene Compounds. US20070105899A1, 2017.

(46) Izumimoto, N.; Kawai, K.; Kawamura, K.; Fujimura, M.; Komagata, T. Remedies or Preventives for Urinary Frequency or Urinary Incontinence and Morphinan Derivatives Having Nitrogen-Containing Heterocyclic Group. US7320984B2, 2008.

(47) Jung, M. S.; Cho, C. K.; Kim, M. J.; Park, C. W.; Ahn, Y. H.; Jang, S. W.; Chang, J. H. Liquid Crystal Thermoset Monomer or Oligomer, Thermosetting Liquid Crystal Polymer Composition Comprising the Same and Printed Circuit Board Using the Same. US7655155B2, 2010.

(48) Urbanski, M.; Xiang, A.; Zeck, R. Substituted Pyrrolones as Allosteric Modulators of Glucokinase. US8680122B2, 2014.

(49) Haikarainen, A.; Kumpulainen, E.; Pohjakallio, A.; Pystynen, J.; Wang, S. Benzodioxane Derivatives and Their Pharmaceutical Use. WO2018002437A1, 2018.

(50) Geraa, J. F.; Lichtenstein, L.; Jung, M. E.; Lee, J.; Holmes, B.; Benavides-Serrato, A. Inhibitors of Ires-Mediated Protein Synthesis. WO2017192665A1, 2017.

(51) Fumitaka, G.; Tomoyasu, T.; Hiroki, O. S.; Hiroaki, Y.; Hiroshi, A.; Grand, S.; Koichi, S. Oligosaccharide Compound, Manufacturing Method Therefor, and Intermediate Thereof. JPWO2013141350A1, 2018.

(52) Fleš, D.; Vuković, R.; Kuzmić, A. E.; Bogdanić, G.; Piližota, V.; Karlović, D.; Markuš, K.; Wolsperger, K.; Vikić-Topić, D. Synthesis and Spectroscopic Evidences of N-Arylmaleimides and N-Aryl-2,3-Dimethylmaleimides. *Croat. Chem. Acta* **2003**, *76*, 69–74.

(53) Kshirsagar, U. A.; Argade, N. P. Facile Approach to Diverse Range of 1,3-Diaza-Heterocycles: Angular/Linear Selectivity Paradigm and a Remarkable Intramolecular Methyl Migration. *Tetrahedron* **2009**, *65*, 5244–5250.

(54) Su, S.; Du, F. S.; Li, Z. C. Synthesis and PH-Dependent Hydrolysis Profiles of Mono- and Dialkyl Substituted Maleamic Acids. *Org. Biomol. Chem.* **2017**, *15*,

8384–8392.

(55) Bartlett, P. D. The Chemical Properties of the Methyl Group. *J. Chem. Educ.* **1953**, *30*, 22–31.

(56) Koskikallio, J. Dimethylmaleic Acid-Dimethylmaleic Anthydride Equilibrium. *Suom. Kemistil.* **1956**, *29B*, 5–7.

(57) Koskikallio, J.; Olsson, S. Kinetics of the Hydrolysis and Formation of Dimethylmaleic Anhydride in Solvent Mixtures. *Acta Chem. Scandinvica* **1956**, *10*, 822–830.

(58) Allinger, N. L.; Zalkow, V. Conformational Analysis. IX. The Gem-Dimethyl Effect. *J. Org. Chem.* **1960**, *25*, 701–704.

(59) Wheeler, O. H.; Rodríguez, E. E. G. Solvolysis of Methylmaleic Anhydrides. *J. Org. Chem.* **1961**, *26*, 4763–4764.

(60) Tressl, R.; Wondrak, G.; Kersten, E.; Rewicki, D. Structure and Potential Cross-linking Reactivity of a New Pentose-Specific Maillard Product. *J. Agric. Food Chem.* **1994**, *42*, 2692–2697.

(61) Palanki, M. S. S.; Bhat, A.; Lappe, R. W.; Liu, B.; Oates, B.; Rizzo, J.; Stankovic, N.; Bradshaw, C. Development of Novel Linkers to Conjugate Pharmacophores to a Carrier Antibody. *Bioorganic Med. Chem. Lett.* **2012**, *22*, 4249–4253.

(62) Koechel, D. A.; Tarloff, J. B.; Rankin, G. O. Acute Effects of Alkylating Agents on Canine Renal Function. *J. Med. Chem.* **1983**, *26*, 85–90.

(63) Uno, B. E.; Dicken, R. D.; Redfern, L. R.; Stern, C. M.; Krzywicki, G. G.; Scheidt, K. A. Calcium(II)-Catalyzed Enantioselective Conjugate Additions of Amines. *Chem. Sci.* **2018**, *9*, 1634–1639.

(64) Uno, B. E.; Deibler, K. K.; Villa, C.; Raghuraman, A.; Scheidt, K. A. Conjugate Additions of Amines to Maleimides via Cooperative Catalysis. *Adv. Synth. Catal.* **2018**, *360*, 1719–1725.

(65) Zubkov, F. I.; Kvyatkovskaya, E. A.; Nikitina, E. V.; Amoyaw, P. N. A.; Kouznetsov, V. V.; Lazarenko, V. A.; Khrustalev, V. N. Comment on “an Unexpected Formation of the Novel 7-Oxa-2-Azabicyclo[2.2.1]Hept-5-Ene Skeleton during the Reaction of Furfurylamine with Maleimides and Their Bioprospection Using a Zebrafish Embryo Model” by C. E. Puerto Galvis and V. V. Kouznetsov,: Org. . *Org. Biomol. Chem.* **2017**, *15*, 6447–6450.

(66) Alonso, C. M. A.; Palumbo, A.; Bullous, A. J.; Pretto, F.; Neri, D.; Boyle, R. W. Site-Specific and Stoichiometric Conjugation of Cationic Porphyrins to Antiangiogenic

Monoclonal Antibodies. *Bioconjug. Chem.* **2010**, *21*, 302–312.

(67) Mustafa, A.; Asker, W.; Khattab, S.; Zayed, S. M. A. D. On the Reactivity of the Unstaturated System in N-Arylmaleimides. *J. Chem. Inf. Model.* **1960**, *26*, 787–789.

(68) Braida, B.; Prana, V.; Hiberty, P. C. The Physical Origin of Saytzeff's Rule. *Angew. Chemie - Int. Ed.* **2009**, *48*, 5724–5728.

(69) Serjeant, E. P.; Dempsey, B. Ionisation Constants of Organic Acids in Aqueous Solution. In *IUPAC Chemical Data Series No. 23*; Pergamon Press, Inc: New York, 1979; p 251.

(70) Nefkens, G. H. L. Synthesis of Phthaloyl Amino Acids under Mild Conditions. *Nature* **1960**, *185*, 309.

(71) Wilson, F. R.; Coombes, M. E.; Brezden-Masley, C.; Yurchenko, M.; Wylie, Q.; Douma, R.; Varu, A.; Hutton, B.; Skidmore, B.; Cameron, C. Herceptin (Trastuzumab) in Her2-Positive Early Breast Cancer: A Systematic Review and Cumulative Network Meta- Analysis. *Syst. Rev.* **2018**, *7*, 191.

(72) Scott, A. M.; Wolchok, J. D.; Old, L. J. Antibody Therapy of Cancer. *Nat. Rev. Chem.* **2012**, *12*, 278–287.

(73) Agarwal, P.; Bertozzi, C. R. Site-SpecificAntibody−DrugConjugates:TheNexusofBioorthogonalChemistry,Protein Engineering, and Drug Development. *Bioconjug. Chem.* **2015**, *26*, 176–192.

(74) Tarentino, A. L.; Plummer, T. H. Methods Enzymology; Academic Press, 1994; Vol. 230, pp 44–57.

(75) Wakankar, A.; Chen, Y.; Gokarn, Y.; Jacobson, F. S.; Wakankar, A.; Chen, Y.; Gokarn, Y.; Jacobson, F. S. Analytical Methods for Physicochemical Characterization of Antibody Drug Conjugates. *MAbs* **2011**, *3*, 161–172.

(76) Basa, L. In. In *Antibody-Drug Conjugates*; Ducry, L., Ed.; Humana Press: Totowa, NJ, 2013; pp 285–293.

(77) Kim, M. T.; Chen, Y.; Marhoul, J.; Jacobson, F. Statistical Modeling of the Drug Load Distribution on Trastuzumab Emtansine (Kadcyla), a Lysine-Linked Antibody Drug Conjugate. *Bioconjug. Chem.* **2014**, *25*, 1223–1232.

(78) Wang, L.; Amphlett, G.; Blättler, W. A.; Lambert, J. M.; Zhang, W. Structural Characterization of the Maytansinoid-Monoclonal Antibody Immunoconjugate, HuN901-DM1, by Mass Spectrometry. *Protein Sci.* **2005**, *14*, 2436–2446.

(79) Ponte, J. F.; Sun, X.; Yode, N. C.; Fishkin, N.; Laleau, R.; Coccia, J.; Lanieri, L.;

Bogalhas, M.; Wang, L.; Wilhelm, S.; Widdison, W.; Pinkas, J.; Keating, T. A.; Chari, R.; Erickson, H. K.; Lambert, J. M. Understanding How the Stability of the Thiol-Maleimide Linkage Impacts the Pharmacokinetics of Lysine-Linked Antibody–Maytansinoid Conjugates. *Bioconjug. Chem.* **2016**, *27*, 1588–1598.

(80) Hamblett, K. J.; Senter, P. D.; Chace, D. F.; Sun, M. M. C.; Lenox, J.; Cerveny, C. G.; Kissler, K. M.; Bernhardt, S. X.; Kopcha, A. K.; Zabinski, R. F.; Meyer, D. L.; Francisco, J. A. Effects of Drug Loading on the Antitumor Activity of a Monoclonal Antibody Drug Conjugate. **2004**, *10*, 7063–7070.

(81) Tsuchikama, K.; An, Z. Antibody-Drug Conjugates: Recent Advances in Conjugation and Linker Chemistries. *Protein Cell* **2018**, *9*, 33–46.

(82) Chen, X.; Muthoosamy, K.; Pfisterer, A.; Neumann, B.; Weil, T. Site-Selective Lysine Modification of Native Proteins and Peptides via Kinetically Controlled Labeling. *Bioconjug. Chem.* **2012**, *23*, 500–508.

(83) Forte, N.; Benni, I.; Karu, K.; Chudasama, V.; Baker, J. R. Cysteine-To-Lysine Transfer Antibody Fragment Conjugation. *Chem. Sci.* **2019**, *10*, 10919–10924.

(84) Matos, M. J.; Oliveira, B. L.; Martínez-Sáez, N.; Guerreiro, A.; Cal, P. M. S. D.; Bertoldo, J.; Maneiro, M.; Perkins, E.; Howard, J.; Deery, M. J.; Chalker, J. M.; Corzana, F.; Jiménez-Osés, G.; Bernardes, G. J. L. Chemo- and Regioselective Lysine Modification on Native Proteins. *J. Am. Chem. Soc.* **2018**, *140*, 4004–4017.

(85) Pham, G. H.; Ou, W.; Bursulaya, B.; DiDonato, M.; Herath, A.; Jin, Y.; Hao, X.; Loren, J.; Spraggon, G.; Brock, A.; Uno, T.; Geierstanger, B. H.; Cellitti, S. E. Tuning a Protein-Labeling Reaction to Achieve Highly Site Selective Lysine Conjugation. *ChemBioChem* **2018**, *19*, 799–804.

(86) Baldwin, A. D.; Kiick, K. L. Reversible Maleimide-Thiol Adducts Yield Glutathione-Sensitive Poly(Ethylene Glycol)-Heparin Hydrogels. *Polym. Chem.* **2013**, *4*, 133–143.

(87) Bandyopadhyay, A.; Gao, J. Targeting Biomolecules with Reversible Covalent Chemistry. *Curr. Opin. Chem. Biol.* **2016**, *34*, 110–116.

(88) Cao, Z.; Li, W.; Liu, R.; Li, X.; Li, H.; Liu, L.; Chen, Y.; Lv, C.; Liu, Y. PH- and Enzyme-Triggered Drug Release as an Important Process in the Design of Anti-Tumor Drug Delivery Systems. *Biomed. Pharmacother.* **2019**, *118*, 109340.

(89) Choi, Y.; Moody, I. S.; Sims, P. C.; Hunt, S. R.; Corso, B. L.; Perez, I.; Weiss, G. A.; Collins, P. G. Single-Molecule Lysozyme Dynamics Monitored by an Electronic Circuit. *Science* **2012**, *335*, 319–324.

(90) Olsen, T. J.; Choi, Y.; Sims, P. C.; Gul, O. T.; Corso, B. L.; Dong, C.; Brown, W. A.;

Collins, P. G.; Weiss, G. A. Electronic Measurements of Single-Molecule Processing by DNA Polymerase i (Klenow Fragment). *J. Am. Chem. Soc.* **2013**, *135*, 7855–7860.

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

(92) Chen, R. J.; Zhang, Y.; Wang, D.; Dai, H. Noncovalent Sidewall Functionalization of Single-Walled Carbon Nanotubes for Protein Immobilization. *J. Am. Chem. Soc.* **2001**, *123*, 3838–3839.

(93) Hedaya, E.; Theodoropulos, S. The Preparation and Reactions of Stable Phosphorus Ylides Derived from Maleic Anhydrides, Maleimides or Isomaleimides. *Tetrahedron* **1968**, *24*, 2241–2254.

(94) Mangaleswaran, S.; Argade, N. P. An Efficient Synthesis of Dimethylmaleic Anhydride. *Synthesis (Stuttg).* **2002**, *7*, 865–868.

(95) Easwar, S.; Argade, N. P. A Facile Synthesis and Enzymatic Resolution of Naturally Occurring Remotely Functionalized Alkylmethylmaleic Anhydrides from Aspergillus Wentii: Aspergillus Acids A-D. *Synthesis (Stuttg).* **2006**, *5*, 831–838.

(96) Desai, S. B.; Argade, N. P. A Facile Synthesis of Ras Farnesyl-Protein Transferase Inhibitor Chaetomellic. *J. Org. Chem.* **1997**, *3263*, 4862–4863.

(97) Theodorou, V.; Skobridis, K.; Tzakos, G.; Ragoussis, V. A Simple Method for the Alkaline Hydrolysis of Esters. *Tetrahedron Lett.* **2007**, *48*, 8230–8233.

(98) Lieberman, K. R.; Cherf, G. M.; Doody, M. J.; Olasagasti, F.; Kolodji, Y.; Akeson, M. Processive Replication of Single DNA Molecules in a Nanopore Catalyzed by Phi29 DNA Polymerase. *J. Am. Chem. Soc.* **2010**, *132*, 17961–17972.

(99) Booker-milburn, K. I.; Anson, C. E.; Clissold, C.; Costin, N. J.; Dainty, R. F.; Murray, M.; Patel, D.; Sharpe, A. Intramolecular Photocycloaddition of N-Alkenyl Substituted Maleimides: A Potential Tool for the Rapid Construction of Perhydroazaazulene Alkaloids. *European J. Org. Chem.* **2001**, 1473–1482.

(100) Booker-milburn, K. I.; Dudin, L. F.; Anson, C. E.; Guile, S. D. Formal Intramolecular [5 + 2] Photocycloaddition Reactions of Maleimides : A Novel Approach to the CDE Ring Skeleton of (−)-Cephalotaxine. *Org. Lett.* **2001**, *7*, 3005–3008.

(101) Roscini, C.; Cubbage, K. L.; Berry, M.; Orr-ewing, A. J.; Booker-milburn, K. I. Reaction Control in Synthetic Organic Photochemistry: Switching between [5 + 2] and [2 + 2]Modes of Cycloaddition. *Angew. Chemie - Int. Ed.* **2009**, *48*, 8716–8720.

(102) Kumarasamy, E.; Raghunathan, R.; Ugrinov, A.; Sivaguru, J. Tailoring Atropisomeric Maleimides for Stereospecific [2 + 2] Photocycloaddition−Photochemical and Photophysical Investigations Leading to Visible-Light Photocatalysis. *J. Am. Chem. Soc.* **2014**, *136*, 8729–8737.

(103) Raghunathan, R.; Kumarasamy, E.; Jockusch, S.; Ugrinov, A.; Sivaguru, J. Engaging Electronic Effects for Atropselective [5+2]-Photocycloaddition of Maleimides. *Chem. Commun.* **2016**, *52*, 8305–8308.

(104) Takeshita, T.; Shimohara, T.; Maeda, S. Synthesis of EDTA-monoalkylamide Chelates and Evaluation of the Surface-active Properties. *J. Am. Oil Chem. Soc.* **1982**, *59*, 104–107.

(105) Shen, B. Q.; Xu, K.; Liu, L.; Raab, H.; Bhakta, S.; Kenrick, M.; Parsons-Reponte, K. L.; Tien, J.; Yu, S. F.; Mai, E.; Li, D.; Tibbitts, J.; Baudys, J.; Saad, O. M.; Scales, S. J.; McDonald, P. J.; Hass, P. E.; Eigenbrot, C.; Nguyen, T.; Solis, W. A.; Fuji, R. N.; Flagella, K. M.; Patel, D.; Spencer, S. D.; Khawli, L. A.; Ebens, A.; Wong, W. L.; Vandlen, R.; Kaur, S.; Sliwkowski, M. X.; Scheller, R. H.; Polakis, P.; Junutula, J. R. Conjugation Site Modulates the in Vivo Stability and Therapeutic Activity of Antibody-Drug Conjugates. *Nat. Biotechnol.* **2012**, *30*, 184–189.

(106) Chaubet, G.; Thoreau, F.; Wagner, A. Recent, Non-Classical, Approaches to Antibody Lysine Modification. *Drug Discov. Today Technol.* **2018**, *30*, 21–26.

(107) DeGruyter, J. N.; Malins, L. R.; Baran, P. S. Residue-Specific Peptide Modification: A Chemist's Guide. *Biochemistry* **2017**, *56*, 3863–3873.

(108) Koniev, O.; Wagner, A. Developments and Recent Advancements in the Field of Endogenous Amino Acid Selective Bond Forming Reactions for Bioconjugation. *Chem. Soc. Rev.* **2015**, *44*, 5495–5551.

(109) Tanaka, K.; Kitadani, M.; Fukase, K. Target-Selective Fluorescent "Switch-on" Protein Labeling by 6π-Azaelectrocyclization. *Org. Biomol. Chem.* **2011**, *9*, 5346–5349.

(110) Tanaka, K.; Masuyama, T.; Hasegawa, K.; Tahara, T.; Mizuma, H.; Wada, Y.; Watanabe, Y.; Fukase, K. A Submicrogram-Scale Protocol for Biomolecule-Based PET Imaging by Rapid 6π-Azaelectrocyclization: Visualization of Sialic Acid Dependent Circulatory Residence of Glycoproteins. *Angew. Chemie - Int. Ed.* **2008**, *47*, 102–105.

(111) Tanaka, K.; Fujii, Y.; Fukase, K. Site-Selective and Nondestructive Protein Labeling through Azaelectrocyclization-Induced Cascade Reactions. *ChemBioChem* **2008**, *9*, 2392–2397.

(112) Rudd, S. E.; Roselt, P.; Cullinane, C.; Hicks, R. J.; Donnelly, P. S. A

Desferrioxamine B Squaramide Ester for the Incorporation of Zirconium-89 into Antibodies. *Chem. Commun.* **2016**, *52*, 11889–11892.

(113) Storer, R. I.; Aciro, C.; Jones, L. H. Squaramides: Physical Properties, Synthesis and Applications. *Chem. Soc. Rev.* **2011**, *40*, 2330–2346.

(114) Marchetti, L. A.; Kumawat, L. K.; Mao, N.; Stephens, J. C.; Elmes, R. B. P. The Versatility of Squaramides: From Supramolecular Chemistry to Chemical Biology. *Chem* **2019**, *5*, 1398–1485.

(115) Cal, P. M. S. D.; Vicente, J. B.; Pires, E.; Coelho, A. V.; Veiros, L. F.; Cordeiro, C.; Gois, P. M. P. Iminoboronates: A New Strategy for Reversible Protein Modification. *J. Am. Chem. Soc.* **2012**, *134*, 10299–10305.

(116) Patterson, J. T.; Wilson, H. D.; Asano, S.; Nilchan, N.; Fuller, R. P.; Roush, W. R.; Rader, C.; Barbas, C. F. Human Serum Albumin Domain i Fusion Protein for Antibody Conjugation. *Bioconjug. Chem.* **2016**, *27*, 2271–2275.

(117) Asano, S.; Patterson, J. T.; Gaj, T.; Barbas, C. F. Site-Selective Labeling of a Lysine Residue in Human Serum Albumin. *Angew. Chemie - Int. Ed.* **2014**, *53*, 11783–11786.

(118) Diethelm, S.; Schafroth, M. A.; Carreira, E. M. Amine-Selective Bioconjugation Using Arene Diazonium Salts. *Org. Lett.* **2014**, *16*, 3908–3911.

(119) Larda, S. T.; Pichugin, D.; Prosser, R. S. Site-Specific Labeling of Protein Lysine Residues and N-Terminal Amino Groups with Indoles and Indole-Derivatives. *Bioconjug. Chem.* **2015**, *26*, 2376–2383.

(120) Tung, C. L.; Wong, C. T. T.; Fung, E. Y. M.; Li, X. Traceless and Chemoselective Amine Bioconjugation via Phthalimidine Formation in Native Protein Modification. *Org. Lett.* **2016**, *18*, 2600–2603.

(121) Dovgan, I.; Ursuegui, S.; Erb, S.; Michel, C.; Kolodych, S.; Cianférani, S.; Wagner, A. Acyl Fluorides: Fast, Efficient, and Versatile Lysine-Based Protein Conjugation via Plug-and-Play Strategy. *Bioconjug. Chem.* **2017**, *28*, 1452–1457.

(122) Lee, H. G.; Lautrette, G.; Pentelute, B. L.; Buchwald, S. L. Palladium-Mediated Arylation of Lysine in Unprotected Peptides. *Angew. Chemie - Int. Ed.* **2017**, *56*, 3177–3181.

(123) Matos, M. J.; Jimeńez-Oseś, G.; Bernardes, G. J. L. Bioconjugation: Methods and Protocols; Massa, S., Devoogdt, N., Eds.; Springer New York: New York, NY, 2019; pp 25–37.

(124) Yamatsugu, K.; Furuta, M.; Xi, S.; Amamoto, Y.; Liu, J.; Kawashima, S. A.; Kanai,

M. Kinetic Analyses and Structure-Activity Relationship Studies of Synthetic Lysine Acetylation Catalysts. *Bioorganic Med. Chem.* **2018**, *26*, 5359–5367.

(125) Hwang, D.; Tsuji, K.; Park, H.; Burke, T. R.; Rader, C. Site-Specific Lysine Arylation as an Alternative Bioconjugation Strategy for Chemically Programmed Antibodies and Antibody-Drug Conjugates. *Bioconjug. Chem.* **2019**, *30*, 2889–2896.

(126) Luo, Q.; Tao, Y.; Sheng, W.; Lu, J.; Wang, H. Dinitroimidazoles as Bifunctional Bioconjugation Reagents for Protein Functionalization and Peptide Macrocyclization. *Nat. Commun.* **2019**, *10*, 142.

(127) Adusumalli, S. R.; Rawale, D. G.; Thakur, K.; Purushottam, L.; Reddy, N. C.; Kalra, N.; Shukla, S.; Rai, V. Chemoselective and Site-Selective Lysine-Directed Lysine Modification Enables Single-Site Labeling of Native Proteins. *Angew. Chemie - Int. Ed.* **2020**, *59*, 10332–10336.

(128) Adusumalli, S. R.; Rawale, D. G.; Singh, U.; Tripathi, P.; Paul, R.; Kalra, N.; Mishra, R. K.; Shukla, S.; Rai, V. Single-Site Labeling of Native Proteins Enabled by a Chemoselective and Site-Selective Chemical Technology. *J. Am. Chem. Soc.* **2018**, *140*, 15114–15123.

(129) Still, W. C.; Kahn, M.; Mitra, A. Rapid Chromatographic Technique for Preparative Separations with Moderate Resolution. *J. Org. Chem.* **1978**, *43*, 2923–2925.

(130) Attar, A. M.; Richardson, M. B.; Speciale, G.; Majumdar, S.; Dyer, R. P.; Sanders, E. C.; Penner, R. M.; Weiss, G. A. Electrochemical Quantification of Glycated and Non-Glycated Human Serum Albumin in Synthetic Urine. *ACS Appl. Mater. Interfaces* **2019**, *11*, 4757–4765.

(131) Sztaray, J.; Memboeuf, A.; Drahos, L.; Vekey, K. Leucine Enkephalin−A Mass Spectrometry Standard. *Mass Spectrom. Rev.* **2011**, *30*, 298–320.

CHAPTER 5: Predicting COVID-19 Severity with a Specific Nucleocapsid Antibody plus Disease Risk Factor Score

5.1 Abstract

Effective methods for predicting COVID-19 disease trajectories are urgently needed. Here, enzyme-linked immunosorbent assay (ELISA) and coronavirus antigen microarray (COVAM) analysis mapped antibody epitopes in the plasma of COVID-19 patients (n=86) experiencing a wide range of disease states. The experiments identified antibodies to a 21-residue epitope from nucleocapsid (termed Ep9) associated with severe disease, including admission to the intensive care unit (ICU), requirement for ventilators, or death. Importantly, anti-Ep9 antibodies can be detected within 6 days post-symptom onset and sometimes within 1 day. Furthermore, anti-Ep9 antibodies correlate with various comorbidities and hallmarks of immune hyperactivity. We introduce a simple-to-calculate, disease risk factor score to quantitate each patient's comorbidities and age. For patients with anti-Ep9 antibodies, scores above 3.0 predict more severe disease outcomes with a 13.42 likelihood ratio (96.7% specificity). The results lay the groundwork for a new type of COVID-19 prognostic to allow early identification and triage of high-risk patients. Such information could guide more effective therapeutic intervention.

Key reference: Sen, S. R.*; Sanders, E. C.*; Gabriel, K. N.*; Miller, B. M.; Isoda, H. M.; Salcedo, G. S.; Garrido, J. E.; Dyer, R. P.; Nakajima, R.; Jain, A.; Caldaruse, A.-M.; Santos, A. M.; Bhuvan, K.; Tifrea, D. F.; Ricks-Oddie, J. L.; Felgner, P. L.; Edwards, R. A.; Majumdar, S.; Weiss, G. A. Predicting COVID-19 Severity with a Specific Nucleocapsid Antibody plus Disease Risk Factor Score. *mSphere* **2021**, *6*, 1–13.

*Authors contrubuted equally. My contribution was designing epitopes, performing cloning and phage display experiments, analyzing data, and writing the manuscript.

5.2 Introduction

The COVID-19 pandemic has triggered an ongoing global health crisis. More than 119.8 million confirmed cases and 2.7 million deaths have been reported worldwide as of March 16, 2021.[1] The virus that causes COVID-19, severe acute respiratory syndrome coronavirus (SARS-CoV-2), belongs to the same family of viruses responsible for respiratory illness linked to recent epidemics – severe acute respiratory syndrome (SARS-CoV-1 termed SARS here) in 2002-2003 and Middle East respiratory syndrome (MERS) in 2012.[2] The current and previous outbreaks suggest coronaviruses will remain viruses of concern for global health.

Many risk factors and comorbidities, including age, sex, hypertension, diabetes, and obesity, can influence COVID-19 patient outcomes.[3] Analysis of patient immune parameters has linked disease severity to elevated levels of biomarkers for inflammation (c-reactive protein and cardiac troponin I), organ damage (aspartate aminotransferase, abbreviated AST, and hypoalbuminemia), immune hyperactivity (IL-6 and IL-10), and clotting (D-dimer).[4] Mortality in COVID-19 is often caused by multi-organ injury and severe pneumonia attributed to an excessive immune response, termed a cytokine storm.[5] Given the rapid and wide spectrum of COVID-19 disease progression, a more precise prognostic linking disease risk factors and specific immune responses can potentially predict disease trajectories and guide interventions.

One hypothesis to explain differences in severity of COVID-19 implicates weakly binding, non-neutralizing antibodies (Abs) to SARS-CoV-2 proteins.[6] However, the potential harm of these suboptimal Abs in COVID-19 patient outcomes remains ill-defined. Furthermore, a recent review on antibody-dependent enhancement of SARS-

CoV-2 stated, “At present, there are no known clinical findings, immunological assays or biomarkers that can differentiate any severe infection from immune-enhanced disease, whether by measuring antibodies, T cells or intrinsic host responses.[7]" This conclusion inspired our study.

SARS-CoV-2 encodes four major structural proteins – spike (S), nucleocapsid (N), membrane (M), and envelope (E). The S, N, and M proteins from SARS elicit an Ab-based immune response.[8,9] The Ab response and its effects on disease progression in SARS-CoV-2 remain under investigation.[10,11] Bioinformatics has predicted >55 Ab binding epitope regions from SARS-CoV-2.[12–17] The epitopes for N, M or E proteins are less well-characterized than for S protein. Several studies have reported comprehensive epitope mapping of the antibody response to SARS-CoV-2.[18–21] Here, we sought to characterize epitopes from SARS-CoV-2 and their correlations with disease severity. ELISAs with phage-displayed epitopes (phage ELISAs) and coronavirus antigen microarray (COVAM) analysis[22] examined plasma samples from COVID-19 patients (n = 86). The results demonstrate that Abs to a specific epitope from N protein plus disease risk factors strongly correlate with COVID-19 disease severity.

5.3 Results

Design and production of candidate epitopes

Twenty-one putative SARS-CoV-2 epitopes were predicted through bioinformatics[12–14] and structure-based analysis. The candidate epitopes spanned the S, N, M, or E proteins and were on average 34 amino acids in length (**Figure 5-1** and **Table 5-1**). These epitopes were phage-displayed as fragments of the full-length protein and

were likely unstructured. Here, epitope refers to the predicted region of the antigenic protein recognized by the antibody's paratope. The structure of S protein bound to a neutralizing antibody[23,24] provided the starting point for 12 of these antibody epitopes. Epitopes were designed to potentially isolate even suboptimal Abs binding to small portions of these structural proteins; such suboptimal Abs were hypothesized to provide insight into disease severity. After display of each potential epitope on the surface of phage, the quality of the epitopes was evaluated by PCR, DNA sequencing, and QC ELISA (**Figure 5-2**). A total of 18 phage-displayed, putative epitopes passed quality control PCR, and were selected for further study.

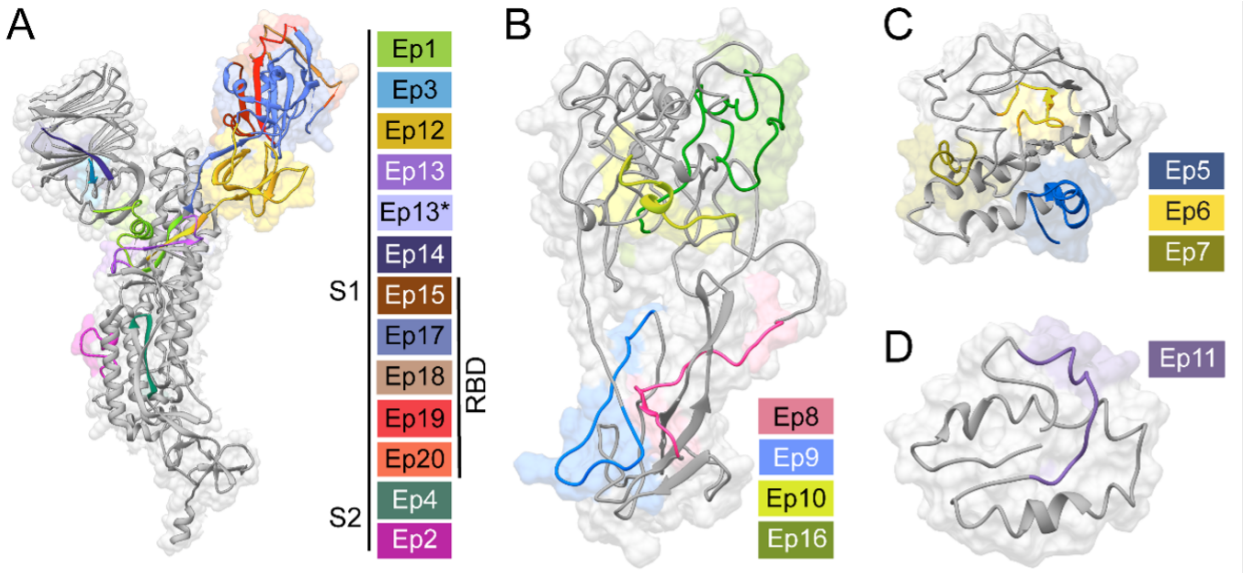


Figure 5-1. Predicted SARS-CoV-2 epitopes examined by phage ELISA. Structural models (gray) of the SARS-CoV-2 **A**) S, **B**) N, **C**) M, or **D**) E proteins illustrate our epitope design (colored). Sequence Ep13* has the mutation D614G, which increases the fitness of SARS-CoV-2 [25–27]. The depicted structural models were derived from an S protein X-ray structure (PDB: 6VXX) [23] or computation modeling of N, M, and E proteins (Protein Gene Bank: QHD43423, QHD43419, and QHD43418, respectively) [28]. **Table S1** provides sequences, sources, and rationale for selections.

Table 5-1. Phage-displayed putative epitopes of SARS-CoV-2 and Ep9 orthologous sequences from SARS, MERS, HKU-1, and NL63.

Epitope	Virus	Protein	Residues*	Amino Acid Sequence	Sequence Length	Rationale for selection	Ref. for rationale
Ep1	SARS-CoV-2	S	287-317	DAVDCALDPLSETKCTLKSFTVEKGIYQTSN	31	Predicted B-cell epitope in SARS-Cov-2 identified by bioinformatics to have immunodominant region	(*13*)
Ep2	SARS-CoV-2	S	802-819	FSQILPDPSKPSKRSFIE	18	Predicted B-cell epitope in SARS-Cov-2 identified by bioinformatics to have immunodominant region	(*13*)
Ep3	SARS-CoV-2	S	15-30	CVNLTTRTQLPPAYTN	16	Predicted B-cell epitope in SARS-CoV-2 identified by immunoinformatics	(*14*)
Ep4	SARS-CoV-2	S	1056-1070	APHGVVFLHVTYVPA	15	Predicted B-cell epitope in SARS-CoV-2 identified by structural biology and machine learning, SARS homolog of this peptide is a known B-cell epitope	(*12*)
Ep5	SARS-CoV-2	M	1-24	MADSNGTITVEELKKLLEQWNLVI	24	Predicted B-cell epitope in SARS-Cov-2 identified by bioinformatics to have immunodominant region	(*13*)
Ep6	SARS-CoV-2	M	132-151	PLLESELVIGAVILRGHLRI	20	Predicted B-cell epitope in SARS-Cov-2 identified by bioinformatics to have immunodominant region	(*13*)
Ep7	SARS-CoV-2	M	97-111	IASFRLFARTRSMWS	15	Predicted B-cell epitope in SARS-CoV-2 identified by structural biology and machine learning, SARS homolog of this peptide is a known B-cell epitope	(*12*)
Ep8	SARS-CoV-2	N	41-61	RPQGLPNNTASWFTALTQHGK	21	Predicted B-cell epitope in SARS-Cov-2 identified by bioinformatics to have immunodominant region n	(*13*)
Ep9	SARS-CoV-2	N	152-172	ANNAAIVLQLPQGTTLPKGFY	21	Predicted B-cell epitope in SARS-Cov-2 identified by bioinformatics to have immunodominant region	(*13*)
Ep10	SARS-CoV-2	N	264-278	ATKAYNVTQAFGRRG	15	Predicted B-cell epitope in SARS-CoV-2 identified by structural biology and machine learning, SARS homolog of this peptide is a known B-cell epitope	(*12*)
Ep11	SARS-CoV-2	E	52-66	VKPSFYVYSRVKNLN	15	Predicted B-cell epitope in SARS-CoV-2 identified by structural biology and machine learning, SARS homolog of this peptide is a known B-cell epitope	(*12*)
Ep12	SARS-CoV-2	S	524-598	VCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVI	75	Predicted B-cell epitope in SARS-CoV-2 identified by bioinformatics to have immunodominant region	(*13*)
Ep13	SARS-CoV-2	S	601-640	GTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGS	40	Predicted B-cell epitope in SARS-Cov-2 identified by bioinformatics to have immunodominant region. Based on the D614G variant.	(*13*)
Ep13*	SARS-CoV-2	S	601-640	GTNTSNQVAVLYQGVNCTEVPVAIHADQLTPTWRVYSTGS	40	Sequence has the mutation D614G, which increases the fitness of SARS-CoV-2	(25,26,*53*)
Ep14	SARS-CoV-2	S	61-76	NVTWFHAIHVSGTNGT	16	Predicted B-cell epitope in SARS-CoV-2 identified by immunoinformatics	(*14*)
Ep15	SARS-CoV-2	S	373-390	SFSTFKCYGVSPTKLNDL	18	Predicted B-cell epitope in SARS-CoV-2 identified by immunoinformatics	(*14*)
Ep16	SARS-CoV-2	N	354-400	NKHIDAYKTFPPTEPKKDKKKKADETQALPQRQKKQQTVTLLPAADL	47	Predicted B-cell epitope in SARS-Cov-2 identified by bioinformatics to have immunodominant region	(*13*)
Ep17	SARS-CoV-2	S	319-529	RVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKK	211	Based on structure of S protein X-ray structure (PBD: 6VXX)	(*23*)
Ep18	SARS-CoV-2	S	488-507	CYFPLQSYGFQPTNGVGYQP	20	Based on X-ray structure of S protein (PBD: 6VXX)	(*23*)
Ep19	SARS-CoV-2	S	429-448	FTGCVIAWNSNNLDSKVGGN	20	Based on X-ray structure of S protein (PBD: 6VXX)	(*23*)
Ep20	SARS-CoV-2	S	448-466	NYNYLYRLFRKSNLKPFER	19	Based on X-ray structure of S protein (PBD: 6VXX)	(*23*)
Ep21	SARS-CoV-2	S	467-487	DISTEIYQAGSTPCNGVEGFN	21	Based on X-ray structure of S protein (PBD: 6VXX)	(*23*)
sEp9	SARS	N	153-173	NNNAATVLQLPQGTTLPKGFY	21	SARS homolog of Ep9	
mEp9	MERS	N	141-161	NNDSAIVTQFAPGTKLPKNFH	21	MERS homolog of Ep9	
hEp9	HKU-1	N	166-186	TTQEAIPTRFPPGTILPQGYY	21	HKU-1 homolog of Ep9	
nEp9	NL63	N	119-136	NQKPLEPKFSIALPPELS	18	NL63 homolog of Ep9	

*Residue numbering from protein sequences deposited in GenBank. Specifically, the accession numbers were as follows: S protein (YP_009724390.1), M protein (YP_009724393.1), N protein (YP_009724397.2) and E protein (YP_009724392.1) from SARS-CoV-2 and N protein from SARS (NP_828855.1), MERS (YP_009047211.1), HKU-1 (YP_173242.1), or NL63 (YP_003771.1).

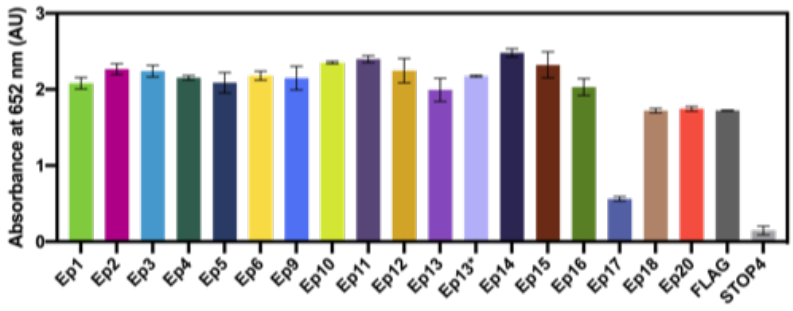


Figure 5-2. Quality control (QC) ELISA for phage-displayed, epitope candidates. Anti-FLAG antibodies (1:1000 in coating buffer) were immobilized on a microtiter plate. Subsequent steps followed the ELISA protocol provided here. Error bars represent SEM (n = 3). Ep8 was apparently toxic to *E. coli*, and Ep7 and Ep19 repeatedly failed sequencing quality controls after phage propagation.

Mapping epitope binding to anti-SARS-CoV-2 Abs

Plasma from COVID-19 patients was subjected to ELISAs with the phage-displayed SARS-CoV-2 epitopes (**Figure 5-3A**). Unless otherwise indicated (e.g., healthy controls), plasma refers to samples from PCR-verified, COVID-19 patients. In this initial assay, plasma was pooled, diluted 100-fold, and coated on a microtiter plate as the antigen of interest (3 pools of n = 5 patients per pool). Nonspecific interactions were blocked (ChonBlock), and phage-displayed epitopes were added for ELISA. The resultant data were normalized by signal from the corresponding negative control (phage without a displayed epitope). Seven candidate epitopes from the pooled patients were further investigated with a larger number of individual patient samples (n = 28) (**Figure 5-3B**). The strongest, reproducible binding was observed for three epitopes from M (Ep6), N (Ep9), and S (Ep20) proteins. Additional COVID-19 plasma samples were profiled for binding to these three epitopes (n = 86 total) (**Figure 5-3B**).

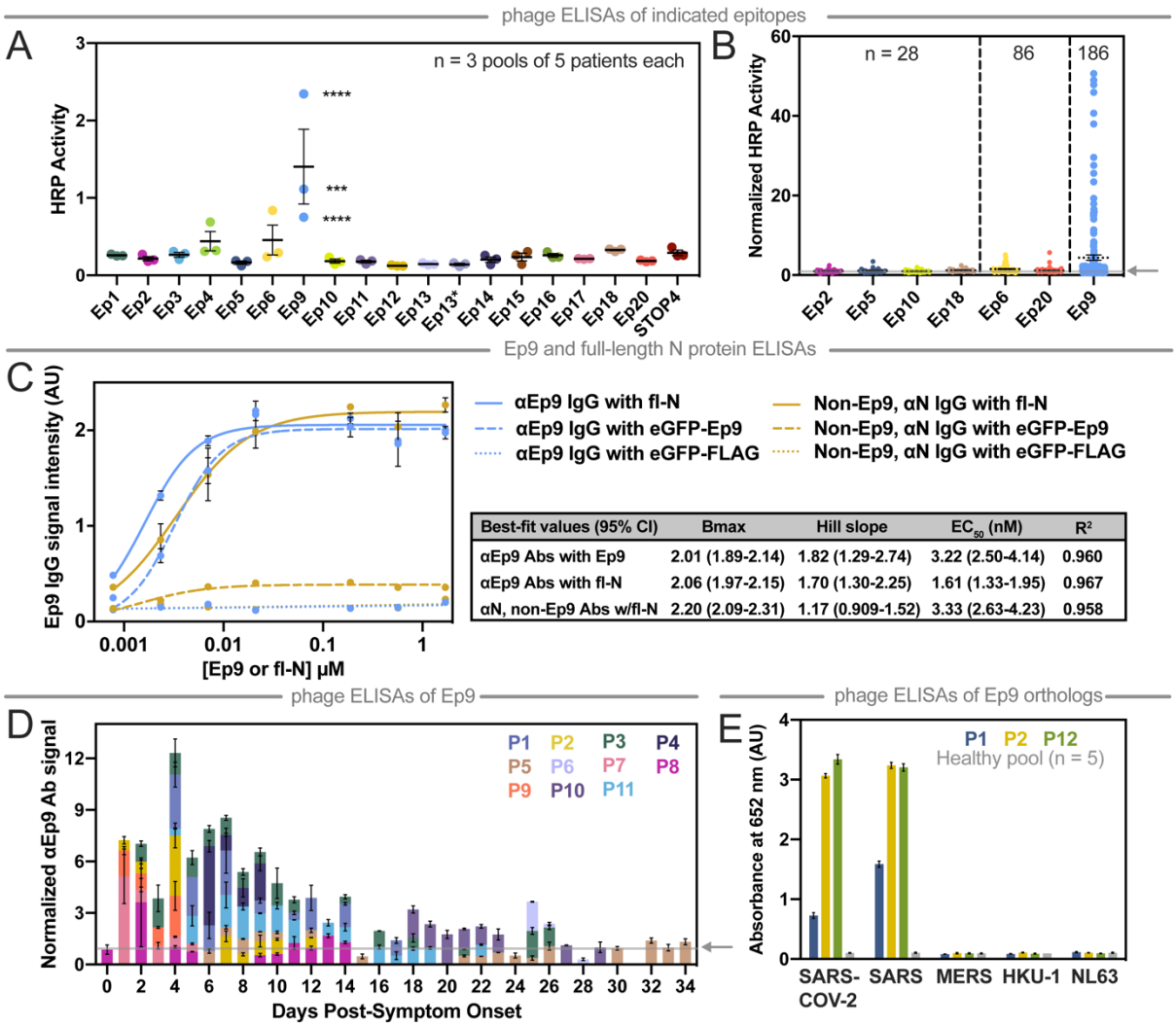


Figure 5-3. Mapping COVID-19 patient antibody responses with phage-displayed SARS-CoV-2 epitopes. A) This phage ELISA with the indicated epitopes (x-axis) examined plasma pooled from patients (n = 3 pools of 5 patients each, 2 technical replicates). STOP4 is the phage negative control. **B)** The epitopes with the highest signals were then further examined by ELISA with plasma from individual patients (n as indicated). **C)** This ELISA measures dose-dependent binding of αN IgGs from plasma pooled from five αEp9 positive patients and five non-Ep9, αN positive patients to eGFP-Ep9 (dashed line), eGFP negative control (eGFP-FLAG, dotted line) or full-length N protein (fl-N, solid line). Varying doses of Ep9 or fl-N were immobilized on microtiter plates, and binding of pooled patient plasma (1:100) was detected using α-Fc IgG-HRP Abs (1:10,000). Pooled patients were matched by similar αN IgG binding signal in COVAM analysis (inset). Non-linear lines of best-fit for binding saturation are represented. Statistical comparisons of Bmax, Hill slope and EC50 between groups, determines that binding of αEp9 IgGs to fl-N or eGFP-Ep9, and non-Ep9, αN IgGs to fl-N are significantly different ($p<0.0001$). Error bars represent ± SD. The data demonstrates that the EC50 value of αEp9 Abs is equal to the cumulative EC50 of all other αN Abs in patients lacking the αEp9 Abs. In the presence of the αEp9 Abs, the apparent binding levels of αN Abs against fl-N approximately doubles. **D)** With

samples from individual patients (designated as P# and by color) collected at the indicated times, αEp9 Abs were measured. The subset of patients shown here comprise all samples for which longitudinal data was available. **E)** Phage ELISA with samples from patients with strong αEp9 Ab response (two from the longitudinal study and one from the patient population) examines cross-reactive binding to Ep9 or Ep9 orthologs from the indicated coronaviruses (x-axis, 3 technical replicates). The arrow on the y-axis and gray line (panels B and C) represents the negative control used for normalizing the data. Error bars represent SEM (panels A, B, and D) or range of two measurements (panel C).

Only the Ep9 epitope from N protein demonstrated robust, statistically significant antibody binding in 27% of patients (n = 186) (**Figure 5-3B**). Of these patients, 100 did not have corresponding health information and were not analyzed further in this report. To test non-phage displayed epitopes, dose-dependent binding of antibodies to Ep9 fused to eGFP (eGFP-Ep9) or to full-length N protein demonstrated that αEp9 IgGs bound its antigen with EC_{50} = 3.22 nM (95% CI = 2.49 to 4.14 nM). This experiment examined plasma samples with the highest IgG response against the N protein in the COVAM assay. Patients without αEp9 Abs had roughly the same level of binding to N protein as observed for αEp9 Abs binding to Ep9. However, such αEp9 Abs appeared to add to N protein binding by antibodies; approximately two-fold increase in apparent antibody binding levels for N protein was observed, if the patient also has αEp9 Abs (**Figure 5-3C**). In patients for whom longitudinal samples were available, the highest levels of αEp9 Abs were observed at days 1 to 14 post-symptom onset (n = 11) and were detectable within 6 days (**Figure 5-3D**). In four of these patients, αEp9 Abs persisted after day 14.

Cross-reactivity of αEp9 Abs against orthologous epitopes from other coronaviruses

Next, the cross-reactivity of αEp9 Abs was examined with Ep9-orthologs from four phylogenetically related coronaviruses known to infect humans (**Figure 5-4A**). Specifically, plasma with αEp9 Abs (n = 3) and pooled plasma from healthy individuals (n = 5) were assayed. The Ep9 epitopes from SARS-CoV-2 and SARS have 90% amino acid sequence homology. Unsurprisingly, this high degree of similarity resulted in a cross-reactive Ep9 epitope, and a strong antibody response was observed to Ep9 epitopes from both viruses (**Figure 5-3E**). The coronaviruses, MERS, HKU-1, and NL63 have 52%, 43%, and 8% sequence homology to SARS-CoV-2 Ep9, respectively (**Figure 5-4B**). These more distantly related orthologs exhibited no cross-reactivity with the αEp9 Abs. Furthermore, no response was observed to Ep9 in pooled plasma from healthy individuals.

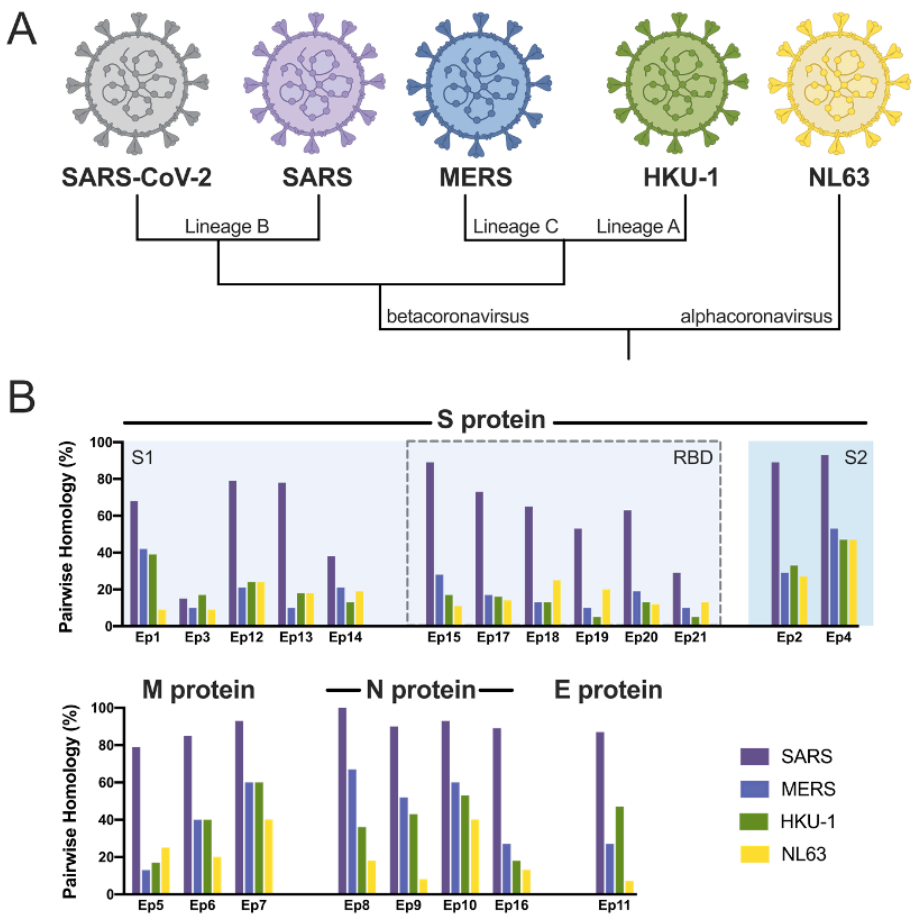


Figure 5-4. Epitope homology of SARS-CoV-2 with four phylogenetically related coronaviruses known to infect humans. A) Evolutionary lineages of the human coronaviruses investigated here, including the highly pathogenic (SARS-CoV-2, SARS, and MERS) and the less virulent (HKU-1 and NL63). **B)** The pairwise homology (% amino acid identity) between SARS-CoV-2 and the indicated coronavirus. Labels (top) indicate the proteins and domains (e.g., S1) from which the epitopes are derived.

The protein microarray COVAM analysis is a high-throughput serological test for SARS-CoV-2 Ab cross-reactivity with a panel of 61 antigens from 23 strains of 10 respiratory tract infection-causing viruses.[22] In this assay, each antigen was printed onto microarrays, probed with human plasma, and analyzed with an ArrayCam imager. COVAM distinguishes between IgG and IgM Abs binding to the full-length N protein (**Figure 5-5** and **Figure 5-6**, respectively). Thus, the COVAM analysis complemented the phage ELISA by expanding the scope of antigens surveyed and adding Ab serotype

information. The ELISA and COVAM data both demonstrated that αEp9 Abs were highly specific for lineage B betacoronaviruses, and were unlikely to be found in patients before their infection with SARS-CoV-2.

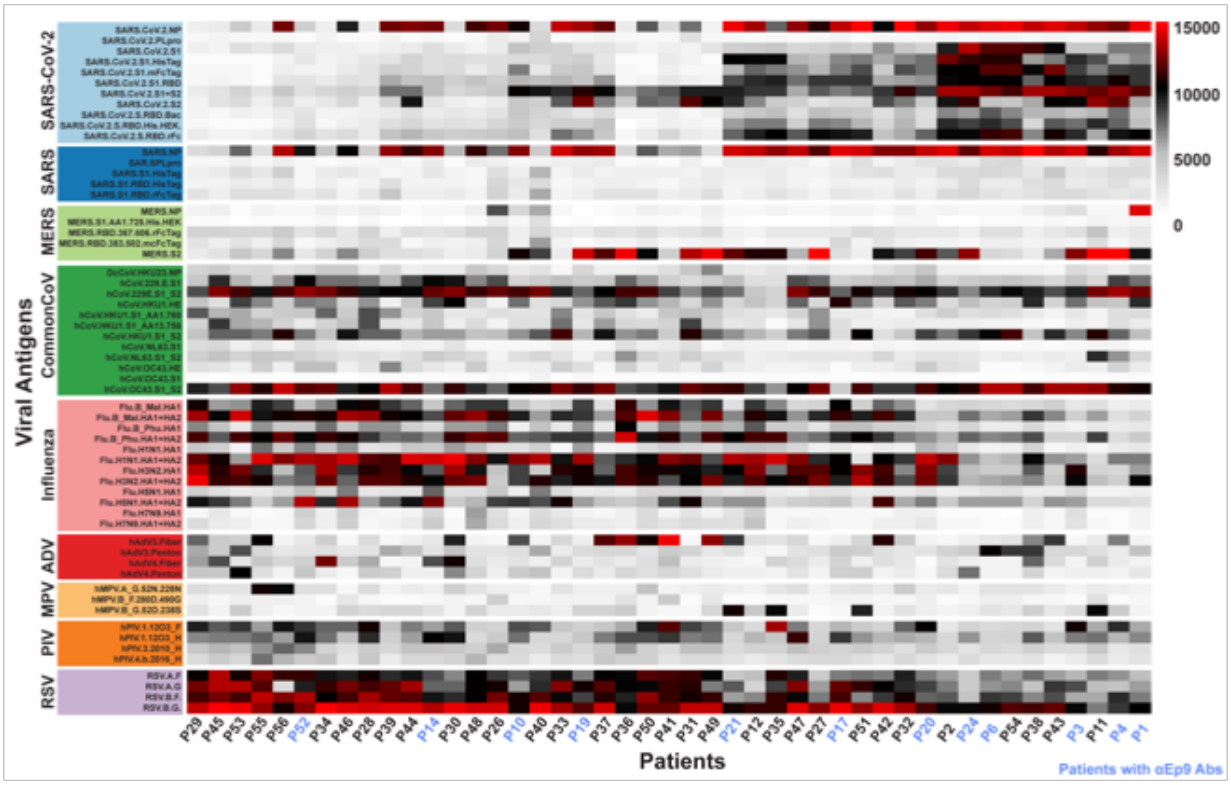


Figure 5-5. Variation in IgG seroreactivity of patient plasma. The heatmap from COVAM data shows normalized signal intensity from plasma samples (n = 45). Plasma samples are in columns and sorted left to right by increasing average intensity to differentially reactive IgG, and viruses are in rows sorted by decreasing average seroreactivity.

Figure 5-6. Variation in IgM seroreactivity of patient plasma. Heatmap from COVAM data showing normalized signal intensity from plasma samples (n = 45). Plasma samples are in columns and sorted left to right by increasing average intensity to differentially reactive IgM, and viruses are in rows sorted by decreasing average seroreactivity.

More severe disease and poorer outcomes for αEp9 patients

Direct comparison of data with full-length N protein from COVAM and Ep9 phage ELISA (n = 40 patients assayed with both techniques) revealed five unique categories of patients (**Figure 5-7A**). To enable this comparison, raw data from each assay was normalized as a percentage of the negative control. Category 1 consists of patients without Abs to the N protein. The next categories included patients with IgMs (Category 2) or IgGs (Category 3) binding to N protein, but not Ep9, termed non-Ep9 αN Abs. Category 4 included patients with αEp9 Abs (both IgMs and IgGs). Category 5 patients had exclusively IgG αEp9 Abs. The αEp9 Abs were only found in patients with IgMs or

IgGs against full-length N protein from the COVAM assay; the COVAM analysis thus independently corroborated the phage ELISAs (**Figure 5-7A**).

Interestingly, the patients with αEp9 Abs suffered more prolonged illness and worse clinical outcomes compared to patients with non-Ep9 αN Abs or no αN Abs. In this study, severe COVID-19 cases were defined as resulting in death or requiring admission to the ICU or intubation. The fraction of severe COVID-19 cases was 2.5 times higher in αEp9 Abs patients than non-Ep9 αN Abs patients (**Figure 5-7, yellow panel**); the differences in proportions of severe and non-severe αN-positive patients with or without αEp9 Abs were statistically significant ($p<0.030$, Fisher's exact test). Patients without αN Abs (Category 1) had less severe symptoms. The αEp9 Abs patients also had longer durations of symptoms and hospital stays relative to non-Ep9 αN Abs and no αN Abs patients (**Figure 5-7C** and **Figure 5-7D**). A larger data set of patient plasma analyzed by phage ELISA confirmed this conclusion ($p<0.0013$, Fisher's exact test) (**Figure 5-7B, blue panel**). Our data further demonstrated that asymptomatic COVID-19 patients (n = 3) also tested negative for αEp9 Abs (**Table 5-2**). The data also revealed early seroconversion of αEp9 IgGs (**Figure 5-7E**), but not αEp9 IgMs (**Figure 5-7F**).

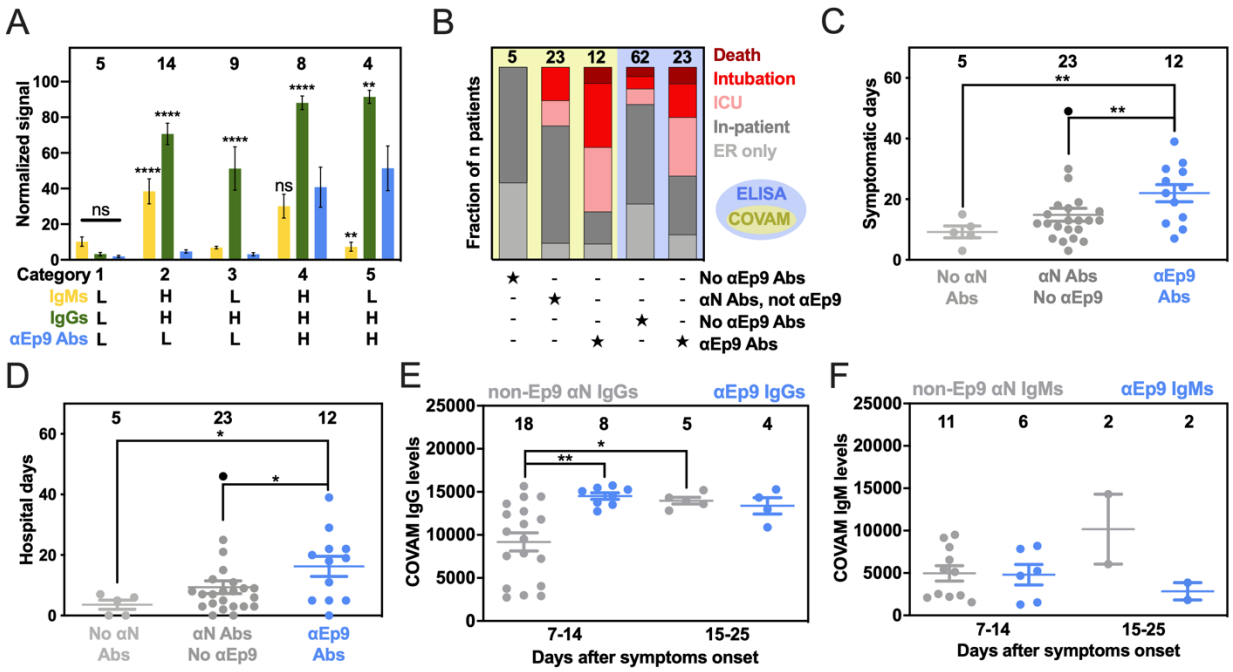


Figure 5-7. Patients with αEp9 Abs have more severe disease. A) Normalized and categorized data from measurements by COVAM (IgMs in yellow, IgGs in green) and Ep9 phage ELISA (blue). ANOVA comparing COVAM to ELISA with Dunnett's multiple comparisons yields p-values of **<0.01, ****<0.0001, or ns: not significant. **B)** Disease severity (color) binned by antibody response (★) for COVAM (yellow) or ELISA (blue). Statistical analysis reveals significant differences between distributions of severe and non-severe disease comparing patient categories, p<0.01 (χ^2) and p<0.001 (Fisher's exact test) for COVAM and ELISA, respectively. Patients with αEp9 Abs are **C)** symptomatic for longer durations and **D)** spend more days in the hospital than those with other αN Abs or no αN Abs. ANOVA with Tukey's multiple comparisons yields p-values of *<0.05 and **<0.01. One outlier (black) (ROUT = 0.1%) was omitted from statistical calculations for panels C and D. **E)** The αN IgG appear at high levels early in the course of disease only for αEp9-positive patients, but are lower in non-Ep9, αN-positive patients. After >15 days post symptom onset, αN IgG levels increase for both groups of patients. **F)** However, IgM levels do not change significantly. Error bars depict SEM with the indicated number of patients (n, numbers above columns).

Table 5-2. Demographics and clinical characteristics of COVID-19 patients categorized by αEp9 Abs response.

Characteristics	No αEp9 Abs (n=63)	αEp9 Abs (n=23)	p-value
Demographics			
Age (± SD)	49.75 (± 18.45)	47.26 (± 18.45)	0.5668
Gender F: M (%)	21:42 (44.4/66.7)	10:13 (43.5/56.5)	0.4502
Ethnicity n, (%) (Hispanic: Asian: Caucasian: Black: Other)	15 (65.2): 4 (17.4): 3 (13.0): 1 (4.3): 0 (0)	39 (61.9): 8 (12.7): 9 (14.3) :3 (4.8) :4 (6.3)	0.7760
BMI (± SD)	28.89 (± 6.445)	32.06 (± 7.896)	0.0642
Preconditions, n (%)			
Hypertension	23 (36.5)	10 (43.5)	0.6203
Diabetes	21 (33.3)	6 (26.1)	0.6065
CVD	6 (9.5)	2 (8.7)	1.0000
CAD	6 (9.5)	2 (8.7)	1.0000
CKD/ESRD	6 (9.5)	2 (8.7)	1.0000
Asthma/COPD	8 (12.7)	3 (13.0)	1.0000
Obesity	24 (38.1)	13 (56.5)	0.1461
Cancer	2 (3.17)	3 (13.0)	0.1163
Symptoms, n (%)			
Total Days of Symptoms	9.8 (± 8.98)	17 (± 10.13)	0.0059**
Cough	43 (68.3)	15 (65.2)	0.7997
Dyspnea/SOB	28 (44.4)	11 (47.8)	0.8108
Myalgia/Fatigue	17 (27.0)	8 (34.8)	0.5926
Headache	12 (19.0)	2 (8.7)	0.3349
Chest pain	7 (11.1)	3 (13.0)	1.0000
Anosmia	4 (6.3)	2 (8.7)	0.6561
Stroke-like Symptoms	0	2 (8.7)	0.0692
Abdominal pain	3 (4.8)	0	0.5611
Pulmonary symptoms^ (Pneumonia: Other: None)	16 (25.4): 36 (52.38): 8 (12.7)	13 (56.5): 7 (30.4): 1 (4.3)	0.0142*
Severity, n (%)			
Asymptomatic	3	0	0.5611
Non-severe: Severe^^	51:12 (n, severity 19.0%)	10:13 (n, severity 56.5%)	0.0013**
Days in Hospital	5.79 (± 8.01)	10.95 (± 10.74)	0.0183*
Days in ICU	12.63 (± 13.19) n=11	12.50 (± 6.93), n=12	0.8004
Days on ventilator	14.00(± 3.96), n=6	12.86 (± 5.40), n=7	0.7934

Results are presented as mean ± standard deviation (SD) or patient number (n) and percentage of population (%). P-values for continuous variables are calculated using unpaired, two-tailed T-tests. P-values for categorical variables use Fisher's exact test for single value parameters, and Chi-squared test for multi-group variables. *, ** p-values < 0.05, 0.01, respectively.

^ Pulmonary symptoms are based descriptive reports of X-ray and CT scans. "Other" pulmonary symptoms include, but are not limited to, atelectasis, pleural scarring, pleural effusion, pulmonary edema, mild peribronchial thickening.
^^ non-severe include ER and In-patients only, severe includes patients in the ICU, on the ventilator or death.

BMI = body mass index, CVD = cardiovascular disease, CAD = coronary artery disease, CKD = chronic kidney disease, ESRD = end-stage renal disease, SOB = shortness of breath, COPD = chronic obstructive pulmonary disease

Strong correlation between disease severity and comorbidities in patients with αEp9 Abs

We compared risk factors, clinical parameters, and disease outcomes among patients with αEp9 Abs (n = 23) (**Figure 5-8A** and **Figure 5-9**). A *disease risk factor score* (DRFS) was developed to evaluate the relationship between clinical preconditions and disease severity in patients with αEp9 Abs. The DRFS quantified a patient's age, sex, and pre-existing health conditions associated with COVID-19 disease severity and mortality. Risk factors include hypertension, diabetes, obesity, cancer, and chronic conditions of the following: cardiac, cerebrovascular, kidney, and pulmonary.[29–32] Using the *age score* from the Charlson Comorbidity Index[33] yields a patient's DRFS as:

$$DRFS = \Sigma \text{ (\# of risk factors)} + \text{(age score)}$$

where each risk factor was valued as either 0 or 1 if absent or present, respectively. The DRFS of patients with αEp9 Abs strongly correlated with COVID-19 disease severity (Pearson's r = 0.72, p-value <0.0001, and R^2= 0.52) (**Figure 5-8A**). The correlation in patients without αEp9 Abs was weak (r = 0.30, p-value = 0.089, R^2= 0.018) (**Figure 5-8A**). Amongst patients with αEp9 Abs (n = 23), a DRFS ≥3 determined disease severity with 92.3% sensitivity (1/13 false negatives) and 80% specificity (2/10 false positives) (**Figure 5-8B**). In the entire study cohort (n = 86), patients with αEp9 Abs and DRFS ≥3 (n = 11) have severe disease with a high degree of specificity (96.7%) and a sensitivity of 44%. Notably, DRFS predicted disease severity only for patients with αEp9 Abs (n = 23), and patients without such Abs (n = 63) had no correlation with disease outcomes.

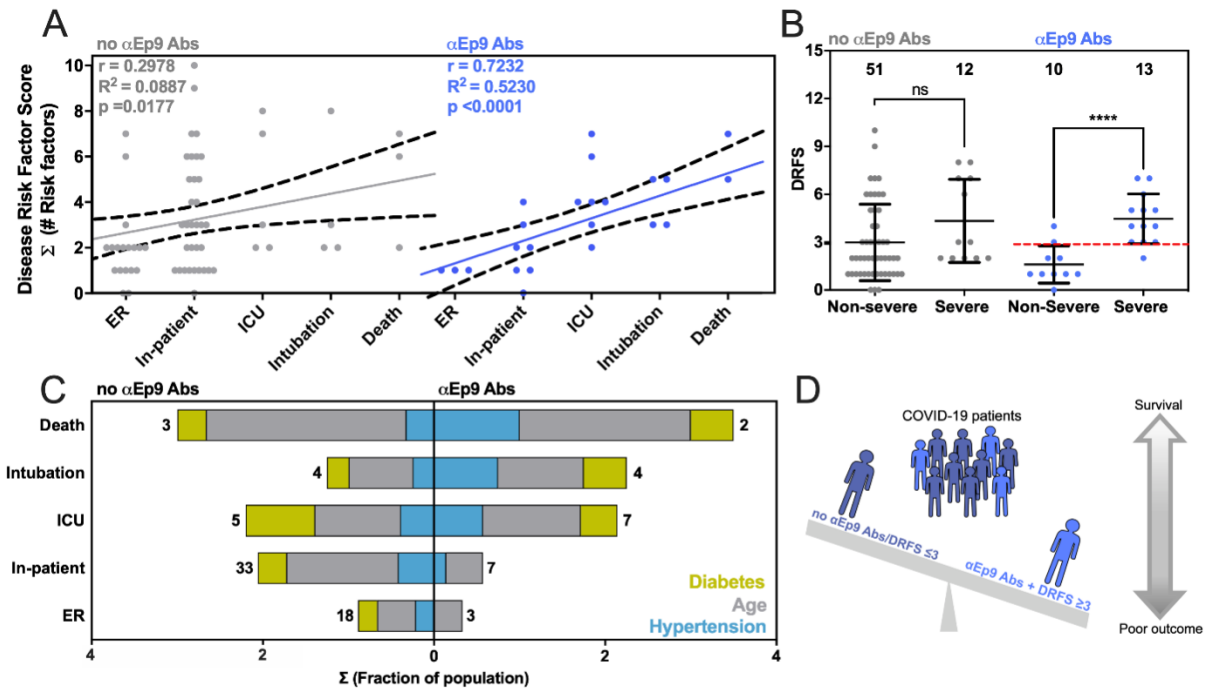


Figure 5-8. Correlation between disease severity and risk factors in patients with αEp9 Abs. A) The relationship between DRFS and disease severity of COVID-19 patients with αEp9 Abs (blue) or no αEp9 Abs (gray). Each data point represents one patient. The solid lines indicate linear regression fits with 95% confidence intervals (dotted lines), and Pearson's r-value as noted. **B)** Correlation of disease severity with DRFS in patients with αEp9 Abs. The data depicts a significant correlation between DRFS and disease severity in patients with αEp9 Abs (blue), but not in patients lacking αEp Abs (gray). In αEp9 patients, a DRFS threshold of 3.0 can predict severe disease (red). Two-tailed, parametric t-tests were conducted to compare non-severe and severe disease outcomes of patients with and without αEp9 Abs, where ****p<0.0001. The error bars represent SD with the indicated n. **C)** The color-indicated risk factors (diabetes, hypertension, and age score) are depicted on the x-axis as the fractions of patients in each disease severity category (y-axis). Numbers indicate total patients (n) without αEp9 Abs (left) or with αEp9 Abs (right). The prevalence of risk factors (colors) increases with disease severity in patients with αEp9 Abs, but not in patients without these Abs. **D)** Patients with αEp9 Abs and DRFS ≥3 are predisposed to increased COVID-19 severity and poorer outcomes.

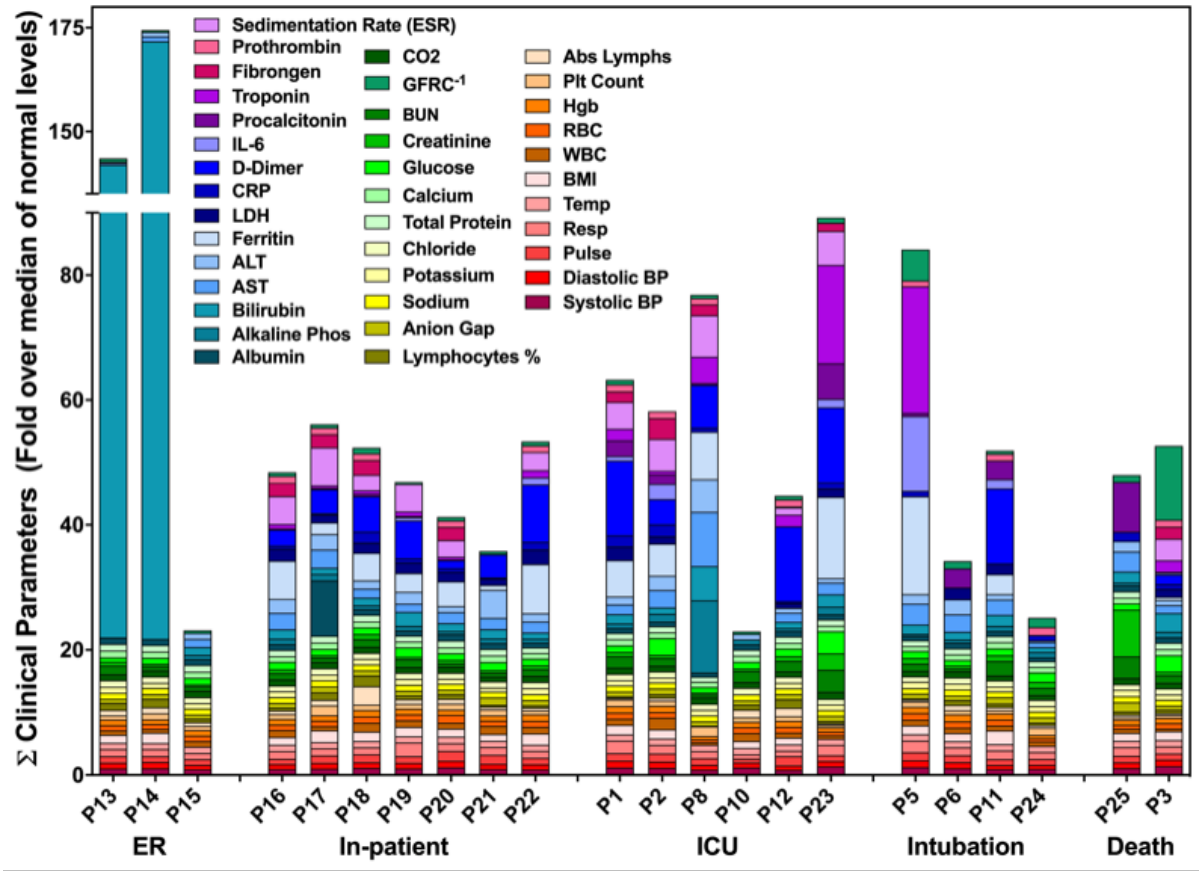


Figure 5-9. Comparison of disease severity and clinical parameters of patients with αEp9 Abs. The data shown represents the fold change of each clinical parameter over the mean of the normal range. The sum of all the fold changes of the clinical parameters for each Ep9-responsive patient is binned according to COVID-19 disease severity. For facile visualization and comparison of clinical biomarkers between Ep9-reponsive patients, the values of each parameter were normalized to fold over the mean of healthy values. No significant trends in clinical parameters (color indicated) were observed with increased disease severity or relative to patients lacking αEp9 Abs.

Examining key contributors to high DRFS, the presence of αEp9 Abs correlated with more severe disease in patients who have hypertension, diabetes, or age >50 years. Such correlation was not observed for patients lacking αEp9 Abs (**Figure 5-8C**). Such risk factors were prevalent at roughly the same percentages in both populations of patients (**Table 5-2**). Thus, these risk factors were particularly acute for patients with αEp9 Abs.

High levels of inflammatory cytokine and tissue damage markers in patients with αEp9 Abs

COVID-19 patients can have elevated serum concentrations of >20 inflammatory cytokines and chemokines.[34] However, information on the cytokine levels and the association with tissue damage and worse COVID-19 outcomes have been inconsistent.[34–36] For patients with IL-6 concentrations measured in plasma, patients with (n = 8) or without (n = 11) αEp9 Abs were compared. Interestingly, the comparison uncovered a strong positive sigmoidal association between IL-6 and AST unique to patients with αEp9 Abs (R^2 = 0.968, Spearman's r = 1.0, p-value <0.0001, n = 8) (Red line, **Figure 5-10A**); correlation of IL-6 and AST in patients with αEp9 Abs remained strong even after removal of the data point at the highest IL-6 concentration. Conversely, a slight negative trend was observed in patients lacking αEp9 Abs (Spearman's r = -0.575, p-value= 0.0612, n = 13). Thus, the presence of αEp9 Abs could disambiguate the sometimes contradictory association of IL-6 with disease severity.

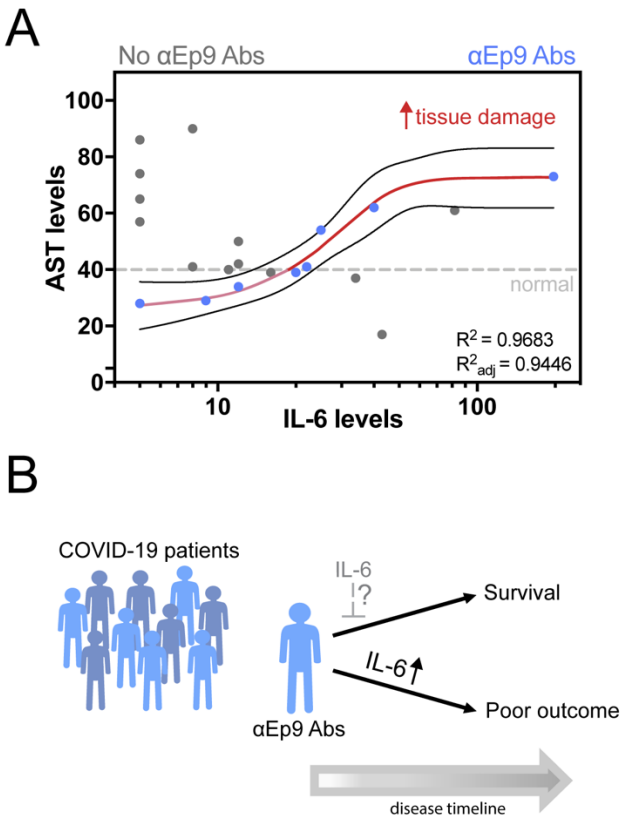


Figure 5-10. Association of inflammatory cytokine and tissue damage markers in patients with αEp9 Abs. A) Association between the inflammatory cytokine, IL-6, and the tissue damage marker, aspartate transaminase (AST), shows a sigmoidal curve fit for patients with αEp9 Abs, R2 = 0.9683, Spearman's correlation coefficient = 1.0, p< 0.0001. **B)** Schematic of patients with αEp9 Abs with increasing IL-6 levels leading to poor outcomes. We hypothesize patients with αEp9 Abs could benefit from IL-6 inhibition early in the disease, such as monoclonal antibody drugs targeting IL-6 or its receptor (IL6R), to disrupt a cytokine storm and reduce severe outcomes.

5.4 Discussion

This study introduces a two-step test as a prognostic for predicting COVID-19 disease severity and its worst outcomes. Specifically, αEp9 Abs can effectively predict severe disease (specificity 83.6%). However, combining presence of αEp9 Abs with DRFS ≥3 provides much higher specificity (96.7%) for predicting severe disease. Previously, αN IgGs have been recognized as a focal site for an antibody response [18,19,21,37] and associated with disease severity and poor outcomes.[11,37,38]

The present investigation expands on previous reports that recognize various regions of the RNA binding domain of N protein as focal sites for anti-SARS-CoV-2 antibody response. For example, the phage display-based VirScan identified an epitope region spanning residues 141-196 and microarrays further isolated peptides including residues 134-171, 155-171, 153-190, and 153-171.[18,19,37] The above investigations, however, do not find correlations between any these epitopes and disease severity. Our results are confirmed by observations from a patient cohort in Singapore, which identify an epitope (residues 153-170) very similar to Ep9 (residues 152-172) and shows a correlation between antibody response against the epitope and pneumonia and the tissue damage markers (CRP and LDH).[20] In our investigation, we examine in-depth patient clinical histories, test results, disease outcomes ranging from asymptomatic to fatal, and longer longitudinal profiling post-symptom onset, to determine the association of a larger subset of markers and risk factors. Such data allows calculation of the DRFS. Together with the presence of αEp9 Abs, patient DRFS allows early discrimination of severe from non-severe disease outcomes. Additionally, fine epitope mapping demonstrates that αEp9 Abs strongly and uniquely correlate with COVID-19 disease severity relative to other αN Abs.

We hypothesize that the underlying mechanism relating αEp9 Abs to increased disease severity involves an overzealous immune response. Specifically, we observe early seroconversion and strong early upregulation of αEp9 IgGs (**Figure 5-7E**). Similar IgG observations have been correlated with poor viral neutralization and clearance, resulting in increased COVID-19 severity.[10,38,39] Also, high levels of IL-6 are observed for αEp9-positive patients with increased levels of the tissue damage marker AST; this

correlation does not exist for patients lacking αEp9 Abs (**Figure 5-10A**). The sensitivity to IL-6 concentration before AST-monitored organ damage suggests anti-IL-6 therapeutics could be an effective management in the early and rapidly progressive stages of respiratory distress for αEp9-positive patients.[34,40–44] Since binding to N protein by αEp9 antibodies is unlikely to enhance uptake of SARS-CoV-2, an antibody-dependent enhancement mechanism could invoke antigen uptake by macrophages. This mechanism could stimulate complement activation and the cytokine storm observed here as elevated IL-6 response. Further investigation is required to determine the basis for increased disease severity in αEp9 patients.

The data demonstrate that αEp9 positive patients with DRFS ≥3 are 13.42 times (Likelihood Ratio) more likely to have severe COVID-19 disease symptoms within the study cohort (n = 86). The presence of αEp9 without DRFS is less effective as a prognostic (Likelihood Ratio of 3.17). Despite its high specificity (96.7%), the sensitivity of this two-step test is 44% (n = 86). However, this test could predict a subset of patients with a specific immune response (i.e., early IgG response and IL-6 dependent immune hyperactivity), and could suggest targeted treatment options (e.g., targeting IL-6 and its pathways).

Importantly, αEp9 Abs appear early in the course of disease. Thus, such a prognostic could outperform traditional markers for the cytokine storm such as IL-6, which appears 6-8 days after symptom onset[34,42]; all plasma collected from αEp9 positive patients (n = 7, **Figure 5-3D**) between 1 to 6 days post-symptoms onset demonstrate detectable levels of αEp9 IgG (≥ 2 fold over negative control). Early detection of αEp9 Abs in patients could be used to triage and treat COVID-19 prior to the onset of its most

severe symptoms; delayed treatments of IL-6 targeting drugs can decrease their efficacy or be counterproductive[34,40–45] (**Figure 5-10B**). The αEp9 Ab biomarker could identify patients most likely to benefit from anti-IL-6 therapeutics and avoid ineffective treatments.

This study demonstrates the usefulness of fine epitope mapping, but the following limitations should be noted. Short linear epitopes, unlike conformational epitopes in larger domains, might not resemble the tertiary structure of an antigen. Post-translational modifications, such as glycosylation were omitted for the phage-displayed S protein epitopes; the COVAM antigens, however, are produced in baculovirus or HEK-293 cells, which could glycosylate the antigens. Our analysis is largely based upon a population of 86 COVID-19 patients and 5 healthy individuals, with the majority of Hispanic descent. The conclusions could be further strengthened with follow-up investigations in a larger population. Additionally, the population examined here only included three asymptomatic individuals, and additional testing is required to verify absence of αEp9 Abs in such patients. The sample size of patients with multiple antibody targets was too limited to allow correlation analysis; future investigations could examine associations between αEp9 and other Abs. Abs recognizing other SARS-CoV-2 structural proteins could also exhibit similar characteristics to αEp9 Abs.

5.5 Conclusion

Existing diagnostic platforms could readily be adapted to test for αEp9 Abs, and the DRFS calculation is quite simple to implement (e.g., assay with eGFP-Ep9 fusion demonstrated here). As shown here, αEp9 Abs do not recognize orthologous sequences from closely related coronaviruses, providing good specificity for αEp9 as a prognostic.

Previous studies have shown that the high homology of N protein among related coronaviruses can lead to high false positive rates in serodiagnostics with full-length N antigen.[46] Thus, the two-step prognostic reported here could mitigate the worst outcomes of COVID-19, particularly for patients at high risk.

5.6 Materials and Methods

Cloning

For phage display of epitopes, the pm1165a phagemid vector as previously described[47] as engineered to encode an N-terminal FLAG-tag and a C-terminal fusion to the P8 coat protein of M13-phage. This template, termed FlagTemplate, was used for subcloning of SARS-CoV-2, SARS, MERS, HKU-1 and NL63 epitopes. A vector map of the FlagTemplate (**Figure 5-11A)**, cloning procedures, and a list of oligos (**Table 5-3**) for Q5 site-directed mutagenesis and Gibson Assembly are provided here.

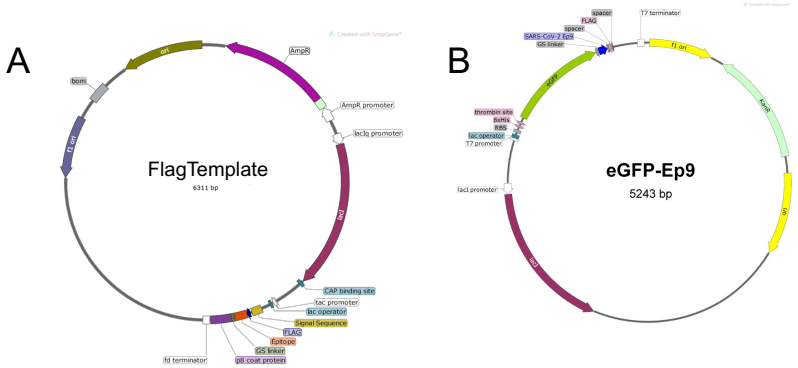


Figure 5-11. Schematic of plasmid map for phage-displayed or recombinant expression epitopes. A) Schematic of the FlagTemplate phagemid used for cloning phage-displayed epitopes. The phagemid, termed FlagTemplate, for the subcloning of SARS-CoV-2, SARS, HKU-1 and NL63 epitopes encodes an N-terminal FLAG tag, followed by a GSG linker to the epitope before a C-terminal GGGSGSSS linker to the P8 coat protein of M13-phage. **B)** The plasmid map for recombinant expression of the eGFP-Ep9 fusion. Ep9 was subcloned into a pET28-eGFP fusion vector with a C-terminal FLAG tag. The fusion protein is connected through a linker (GGGSGSS), and two spacers flank the N- and C-termini of the FLAG-tag (SGSG and GSG, respectively). The plasmid backbone lacking the Ep9 sequence was used to express the eGFP-FLAG negative control.

Table 5-3. Oligos used for cloning of phage-displayed putative epitope for SARS-CoV-2 and ortholog Ep9 epitope from SARS, MERS, HKU-1, and NL63.

Oligo#	Sequence (5' to 3') with insertions denoted in lowercase and substitutions in bold	Product	Mutagenesis	Rounds
Oligo1	aggaagtggaggtggaggatccgggagctccagcCCGAGGGTGACGATCCCG	FlagTemplate	Q5	1
Oligo2	ttatcatcgtcatctttataatcaaccaatgcataGCCGAGGCGGAAAACATC			
Oligo3	attaaagtcgttcaccgtcgaaaaaggaatctatcagacctctaacGGTGGAGGATCCGGGAGC	Ep1	Q5	1
Oligo4	gtacactttgtttcactcagtggatctaatgcacaatcgaccgcatcTTTATCATCGTCATCTTTATAATCAAC CAATGC			
Oligo5	gagcaagccttctaagcgctctttcattgaaGGTGGAGGATCCGGGAGC	Ep2	Q5	1
Oligo6	gggtcaggcaggatctgcgagaatccacttccTTTATCATCGTCATCTTTATAATCAACCAATGC			
Oligo7	tacacagttacctcccgcgtatacaaatGGTGGAGGATCCGGGAGC	Ep3	Q5	1
Oligo8	cgagttgtcaagttcacacatccacttccTTTATCATCGTCATCTTTATAATCAACCAATGC			
Oligo9	ttcttgcacgtgacgtatgtgcctgctGGTGGAGGATCCGGGAGC	Ep4	Q5	1
Oligo10	caccactccatggggcgctccacttccTTTATCATCGTCATCTTTATAATCAACCAATGC			
Oligo11	ggagctgaaaaaactgttggaacaatggaaccttgtaatcGGTGGAGGATCCGGGAGC	Ep5	Q5	1
Oligo12	tctacggtaatcgtaccgttcgagtccgccattccacttccTTTATCATCGTCATCTTTATAATCAACCAATGC			
Oligo13	tggagctgtgatcttacgcggacacctgcgtatcGGTGGAGGATCCGGGAGC	Ep6	Q5	1
Oligo14	atcactaattctgattccaacaggggtccacttccTTTATCATCGTCATCTTTATAATCAACCAATGC			
Oligo15	ttcgcacgcactcgctccatgtggtctGGTGGAGGATCCGGGAGC	Ep7	Q5	1
Oligo16	caagcggaagctcgcaattccacttccTTTATCATCGTCATCTTTATAATCAACCAATGC			
Oligo17	gcttcgtggttcactgcgcttacccagcacggaaagGGTGGAGGATCCGGGAGC	Ep8	Q5	1
Oligo18	tgtattattaggcagcccttgagggcgtccacttccTTTATCATCGTCATCTTTATAATCAACCAATGC			
Oligo19	tcaagggactaccttgcccaaggggttctatGGTGGAGGATCCGGGAGC	Ep9	Q5	1
Oligo20	ggtaattgtaacacgattgcagcgttattagcTCCACTTCCTTTATCATCGTCATCTTTATAATC			
Oligo21	gtaacccaagcgttcggtcgccgcgggGGTGGAGGATCCGGGAGC	Ep10	Q5	1
Oligo22	attatacgcttttgtagctccacttccTTTATCATCGTCATCTTTATAATCAACCAATGC			
Oligo23	gtctactctcgtgtaaaaaacttgaatGGTGGAGGATCCGGGAGC	Ep11	Q5	1
Oligo24	gtaaaaggaaggcttcactccacttccTTTATCATCGTCATCTTTATAATCAACCAATGC			
Oligo25	tttcaactttaatggcctgacggggaccggagtcctgactgaatccaatGGTGGAGGATCCGGGAGC	Ep12	Q5	3
Oligo26	ttgacacacttatttttaaccaggtttgttgacttcttcggcccgcatacTTTATCATCGTCATCTTTATAATC AACCAATGC			
Oligo27	agacgctgttcgtgacccacagactctggagattttggacattacacctGGTGGAGGATCCGGGAGC			
Oligo28	gttgtatcggcgatgtcgcgtccaaattgctggaacggcagaaatttttttATTGGATTCAGTCAGGACTCCG			
Oligo29	tgtctccgtcatcGGTGGAGGATCCGGGAGC			
Oligo30	cctccgaatgaacaAGGTGTAATGTCCAAAATCTCCAGAG			
Oligo31	ctgtatcaggatgtcaattgcacagaagtcGGTGGAGGATCCGGGAGC	Ep13	Q5	2
Oligo32	tacagcaacttggttacttgtgtttgtcccTTTATCATCGTCATCTTTATAATCAACCAATGC			
Oligo33	cccacttggcgcgtctacagcacaggcagtGGTGGAGGATCCGGGAGC			
Oligo34	cgtcagttggtcggcgtggatggctaccggGACTTCTGTGCAATTGACATCCTGATAC			
Oligo35	GCTGTACCAG**ggc**GTGAACTGTA	Ep13*	Q5	1
Oligo36	ACTGCCACCTGATTGCTG			
Oligo37	tatccacgtatcgggtacgaatggaacgGGTGGAGGATCCGGGAGC	Ep14	Q5	1
Oligo38	gcgtggaaccacgtcacatttccacttccTTTATCATCGTCATCTTTATAATCAACCAATGC			
Oligo39	tggagtgtctcccactaaattgaacgaccttGGTGGAGGATCCGGGAGC	Ep15	Q5	1
Oligo40	tagcatttgaaggttgagaacgatccacttccTTTATCATCGTCATCTTTATAATCAACCAATGC			
Oligo41	tcccccgacggagccgaagaaggataagaaaaaaaaaGGTGGAGGATCCGGGAGC	Ep16	Q5	2
Oligo42	aaggttttatacgcatcaatgtgtttatttccacttccTTTATCATCGTCATCTTTATAATCAACCAATGC			
Oligo43	acttccacagggaacgacactgccaaagggatttGGTGGAGGATCCGGGAGC			
Oligo44	tgcagtacagtggcagcattgttatttccacttccTTTATCATCGTCATCTTTATAATCAACCAATGC			
Oligo45	GGTGGAGGATCCGGGAGCTCCAGC	GA Vector	-	-
Oligo46	TTTATCATCGTCATCTTTATAATCAACCAATGCATAAGCCGAGGC			
Oligo47	CTGTATGTGGGCCCAAAAAGGGTGGAGGATCCGGGAGCTC	Ep17	GA	
Oligo48	GGCTGTACGCGTCCACTTCCTTTATCATCGTCATCTTTATAATCAACCAATGCATAAG			
Oligo49	ACGATGATAAAGGAAGTGGACGCGTACAGCCCACTGAAAG			
Oligo50	GAGCTCCCGGATCCTCCACCCTTTTTGGGCCCACATACAGTCG			
Oligo51	caaccgacaaatggtgtgggttaccagccgGGTGGAGGATCCGGGAGC	Ep18	Q5	1
Oligo52	gaaaccatatgactgcaagggaaaataacaTCCACTTCCTTTATCATCGTCATCTTTATAATC			
Oligo53	taataacttggattccaaagtaggaggcGGTGGAGGATCCGGGAGC	Ep19	Q5	1
Oligo54	gaattccaggcaataacacacccagtgaaTCCACTTCCTTTATCATCGTCATCTTTATAATC			
Oligo55	gaaatctaacttgaaaccgtttgaaaggGGTGGAGGATCCGGGAGC	Ep20	Q5	1
Oligo56	cggaaaagcctgtaaagatagttgtaattTCCACTTCCTTTATCATCGTCATCTTTATAATC			
Oligo57	cacgccttgcaatggggtagagggttttaatGGTGGAGGATCCGGGAGC	Ep21	Q5	1
Oligo58	gaccctgcctggtagatctccgttgagatgtcTCCACTTCCTTTATCATCGTCATCTTTATAATC			
Oligo59	caaccgacaaatggtgtgggttaccagccgGGTGGAGGATCCGGGAGC	Ep22	Q5	1
Oligo60	gaaaccatatgactgcaagggaaaataacaTCCACTTCCTTTATCATCGTCATCTTTATAATC			
Oligo61	acttccacagggaacgacactgccaaagggatttGGTGGAGGATCCGGGAGC	sEp9	Q5	1
Oligo62	tgcagtacagtggcagcattgttatttccacttccTTTATCATCGTCATCTTTATAATCAACCAATGC			
Oligo63	tccgggtacaaagttaccaaagaacttccacGGTGGAGGATCCGGGAGC	mEp9	Q5	1
Oligo64	gcaaattgagtaactatcgctgaatcattgttTCCACTTCCTTTATCATCGTCATCTTTATAATC			
Oligo65	tcccggaactattttaccccaaggatactatGGTGGAGGATCCGGGAGC	hEp9	Q5	1
Oligo66	gggaatctagtgggaatcgcctcctgagtagtTCCACTTCCTTTATCATCGTCATCTTTATAATC			
Oligo67	agtattgccttgccacctgagttatctGGTGGAGGATCCGGGAGC	nEp9	Q5	1
Oligo68	aaatttcggttcaagcggcttttgattTCCACTTCCTTTATCATCGTCATCTTTATAATC			
Oligo69	TTTTGCGCCGACATCATAACGGTTCT	pm1165a P8	-	-
Oligo70	TATGGGGTTTTGCTAAACAACTTTCAACAG			
Oligo71	agatgacgatgataaaggaagtggatataaGGCGGTTCGGCAGGCGAA	eGFP-FLAG	Q5	1
Oligo72	ttataatcgctggagctcccggatcctccaccTTTGTATAGTTCATCCATGCCATGTGTAATCCC			
Oligo73	tcaagggactaccttgcccaaggggttctatAGCGGAAGTGGAGATTATAAAGATGAC	eGFP-Ep9	Q5	1
Oligo74	ggtaattgtaacacgattgcagcgttattagcGGAGCTCCCGGATCCTCC			
Oligo75	TAATACGACTCACTATAGGG	T7	-	-
Oligo76	GCTAGTTATTGCTCAGCGG	T7 term	-	-

List of oligos used in Q5 site-directed mutagenesis (Q5) or Gibson Assembly (GA) for the cloning of putative epitopes Ep1-22 for phage display or eGFP recombinant protein expression. For epitope constructs between 100-210 bp in length, multiple rounds of Q5 were denoted.

Short (approximately 30 amino acids) putative epitopes for phage display and *E. coli* expression as eGFP fusion peptides in the pet28 vector were cloned via Q5 site-directed mutagenesis according to the manufacturer's instructions. A vector map of Ep9 fused peptide to eGFP, termed eGFP-Ep9, is shown below (**Figure 5-11B**). For large epitopes (>500 bp), such as Ep17, Gibson Assembly (New England Biolabs) was conducted in two PCR steps with the **FlagTemplate** or pCAGGS containing the SARS-CoV-2 S protein gene (BEI Resources) to generate the vectors and inserts, respectively. The Gibson Assembly (2 µL) or KLD (Kinase, Ligase, DpnI) mix (5 µL) was transformed into Nova Blue *E. coli* competent cells, and transformants were plated on a carbenicillin-supplemented (50 µg/mL) agar plate before incubation at 37 °C overnight. Five single colonies were selected to inoculate 4 mL of SOC media in a 15 mL culture tube supplemented with carbenicillin (50 µg/mL). The seed cultures were incubated at 37 °C with shaking at 225 rpm for 8-12 h. Phagemid DNA was isolated using the QIAprep spin miniprep kit according to the manufacturer's instructions. The successful subcloning of the ORF encoding each epitope was verified via DNA sequencing (Genewiz). The full-length N protein in a pLVX-EF1α-IRES-Puro plasmid was a generous gift from Prof. Rachel Martin (UCI).

Purification and preparation of phage

Phage were propagated and purified using procedures previously described[48] with the following changes. A single colony was selected to inoculate 15 mL of 2YT and shaken at 37 °C until the OD_{600} reached 0.6. After incubation at 37 °C for 45 min, 8 mL of the primary culture was used to inoculate 300 mL of 2YT supplemented with carbenicillin

(50 μg/mL), kanamycin (20 μg/mL), and isopropyl β-D thiogalactopyranoside (IPTG, 30 μM).

To precipitate the phage, the cultures were centrifuged at 10 krpm (15300 x g) for 10 min at 4 °C. The supernatant was decanted into a centrifuge tube containing 60 mL PEG-8000 (20%, w/v) and NaCl (2.5 M). The tube was inverted 10 times and stored on ice for 30 min followed by an additional centrifugation at 10 krpm (15300 x g) for 20 min at 4 °C. The supernatant was decanted, and tubes were centrifuged for an additional 4 min at 4 krpm (2429 x g) at 4 °C. The pellets were resuspended in PBS (10 mM phosphate, 137 mM NaCl, pH 7.2) with TWEEN 20 (0.05%, v/v) and glycerol (10%, v/v), separated into 1 mL aliquots, flash frozen with liquid nitrogen, and stored at -80 °C. For binding assays via ELISA, the purified phage was thawed on ice, precipitated a second time as before. The quality of each phage preparation was routinely checked by quality control ELISA, termed QC ELISA, to a FLAG peptide fused to the N-terminus of each epitope (**Figure 5-2**); additionally, PCR using Oligo69 and Oligo70 followed by DNA sequencing (Genewiz) was performed for every phage preparation. Such quality control allowed for identification of toxic clones; for example, C8, was apparently toxic to *E. coli*, and three protein epitopes failed to express in *E. coli* for unknown reasons. The phage concentration was determined by absorbance at 260 nm using a coefficient of molar absorptivity of 0.003 nM^{-1} cm^{-1} and diluted to 40 nM in PBS.

Expression and Purification of eGFP-Ep9 and N protein

A pET28c plasmid containing Ep9 fused to an N-terminal eGFP (**Figure 5-11B**) was transformed into BL21 DE3* *E. coli* heat shock, competent cells. A single colony was

transferred to LB media (20 mL) supplemented with kanamycin (40 µg/mL) and incubated at 37 °C for 18 h. An aliquot of the starter culture (2.5 mL) was transferred to LB media with 1% glucose (250 mL LB in a 1 L baffled flask). After reaching an OD_{600} between 0.4-0.6, the culture was induced through addition of IPTG (0.5 mM) before incubation at 25 °C for 18 h. The cells were centrifuged (15,300 x g) for 20 min at 4 °C, and the cell pellet was resuspended in lysis buffer (25 mM Tris-HCl and 200 mM NaCl, pH 8.0 and supplemented with protease inhibitor cocktail) followed by sonication. The lysate was subjected to centrifugation (26,892 rcf, 45 min, 4 °C). The supernatant was incubated with charged Ni-IMAC resin overnight on a rotary shaker (150 rpm at 4 °C). The resin was equilibrated in a column, washed with wash buffer (20 mM imidazole in lysis buffer), and the purified protein was eluted using elution buffer (250 mM imidazole in lysis buffer). Elutions containing the purified protein were visualized using 10% or 12% SDS-PAGE (Bio-rad Mini-PROTEAN Tetra electrophoresis system) stained with Coomassie brilliant blue stain (**Figure 5-12**). The eluted fractions containing the purified eGFP-Ep9 were pooled and buffer exchanged for 3 column volumes (20 mL) with lysis buffer without imidazole using a 10 kDa cutoff microconcentrator (Vivaspin, Fisher Scientific). The protein concentration was determined by a bicinchoninic acid (BCA) assay or Bradford assay using the estimated MW (http://www.expasy.org). Similar to eGFP-Ep9, the full-length N protein was expressed in 250 ml LB with 1% glucose and induced with 0.25 mM IPTG at OD_{600} = 0.8. Protein overexpression cultures were incubated at 16 °C for 22 h. Lysis and purification were conducted as described above, using N protein lysis buffer (20 mM Tris-HCl, 300 mM NaCl, 5 mM $MgCl_2$, 5 mM BME, 10% glycerol pH 8.0). The purified full-length N protein was analyzed using 10% SDS-PAGE (**Figure 5-12**).

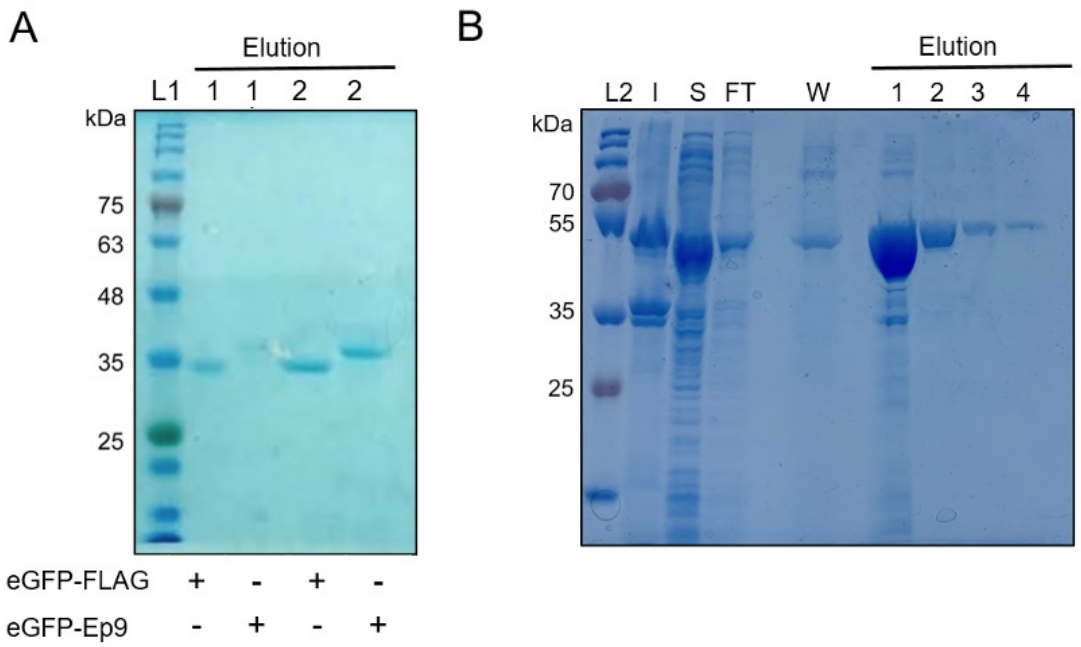


Figure 5-12. eGFP-FLAG, eGFP-Ep9, and full-length N protein purity assessed by 10% SDS-PAGE. Representative SDS-PAGE to visualize **A)** eGFP-FLAG and eGFP-Ep9 or **B)** full-length N protein after immobilized metal (Ni2+) affinity chromatography purification. Elution fraction #2 of eGFP-FLAG, eGPF-Ep9, and elution fraction #3 were purified to >95% homogeneity, and migrated at the expected mass of ≈32, ≈34, and ≈48 kDa, respectively. Elution fraction #1 of eGFP-FLAG and eGFP-Ep9 was diluted 20-fold. L1, BLUE stain (Goldbio) and L2, Prestain PAGE-Ruler Plus (Thermo Fisher Scientific) protein ladder was used as reference. I = insoluble fraction, S = soluble fractions, FT = flow-through, W = wash.

Patient sample collection

The UC Irvine Experimental Tissue Resource operates under a blanket IRB protocol (UCI #2012-8716) that gives ETR personnel 'Honest Broker' status and enables the collection of any fluid or tissue remnant in excess to that needed for clinical diagnosis and distribution to investigators under the conditions of their own IRB approval. Patients undergoing COVID testing in the Emergency Department or on the inpatient service with confirmed COVID (+) pharyngeal swabs, were followed for their blood collections daily. Specimens collected originally for diagnostic purposes were processed and stored by the hospital laboratory in a manner compliant with College of American Pathologists (CAP)

standards. EDTA-anticoagulated whole blood was stored for 2 days at 4 °C after clinical diagnosis and released for research purposes. Plasma from heparin-anticoagulated blood was centrifuged immediately after collection and preserved at 4 °C for 3-4 days before being released for research use. All COVID (+) specimens were handled under BSL-2 conditions, aliquoted into screw cap cryovials, and stored at -80 °C long term with constant temperature monitoring. Specimens were coded by the ETR with unique de-identifiers, and accompanying clinical information was stripped of PHI, such that investigators could receive specimens under a Non Human Subjects Determination exemption from the UC Irvine IRB. All samples from SARS-CoV-2 infected patients were inactivated by incubation in a water bath at 56 °C for 30 min[49], aliquoted (40 µL each), and stored at -80 °C.

Phage ELISA with plasma

The phage-displayed SARS-CoV-2 epitopes were used in phage ELISAs with patient plasma samples diluted 100-fold in coating buffer (50 mM Na_2CO_3, pH 9.6). After incubation in a 96-well Nunc MaxiSorp flat-bottom microtiter plate with shaking at 150 rpm at 4 °C for 12-18 h, plasma was aspirated by a plate washer (BioTek). Next, the plate was treated with 100 µL per well of ChonBlock Blocking/Sample Dilution Buffer (Chondrex, Inc.) for 1 h with shaking at 150 rpm at room temperature and washed three times with wash buffer (0.05% v/v Tween-20 in PBS). The epitope displaying phage and controls were diluted to 1 nM in ChonBlock Blocking/Sample Dilution Buffer and 100 µL were added to each well before incubating for 2 h with shaking (150 rpm) at room temperature. The plate was then washed three times with wash buffer. The primary antibody, anti-M13-

HRP (Creative Diagnostics), was diluted 1:5000 in ChonBlock Secondary Antibody Buffer and 100 µL was added per well; the plate was incubated for 1 h at 150 rpm and room temperature. Following three washes with wash buffer, 1-Step Ultra TMB-ELISA Substrate Solution (100 µL per well, ThermoScientific) was added. Absorbance of TMB substrate was measured twice at 652 nm by UV-Vis plate reader (BioTek) after 5 and 15 min of incubation.

ELISA of eGFP-Ep9 and full-length N protein with plasma

Varying doses, with a maximum concentration of 1.7 µM, of eGFP-Ep9, eGFP-FLAG or full-length N protein (fl-N) were diluted in PBS pH 8.0, and then immobilized on a 96-well Nunc MaxiSorp flat-bottom microtiter plate before incubatation on a shaker (150 rpm) at 4 °C for 12 to 18 h. After incubation, unattached proteins were removed through aspiratation using a plate washer (BioTek) and wells were blocked with 100 µL ChonBlock Blocking/Sample Dilution Buffer (Chondrex, Inc.) for 30 min with shaking (150 rpm) at room temperature. The plate was then washed three times with wash buffer (0.05% v/v Tween-20 in PBS). Pooled plasma from five patients within each experimental group was diluted 100-fold in ChonBlock Blocking/Sample Dilution Buffer and 100 µL was added to each well before incubating for 1 h with shaking (150 rpm) at room temperature. The plate was then washed three times with wash buffer. The detection antibody, IgG Fc Goat anti-Human, HRP (Invitrogen), was diluted 1:5000 in ChonBlock Secondary Antibody Buffer and 100 µL was added per well; the plate was incubated for 30 min at 150 rpm and room temperature. Following six washes with wash buffer, 1-Step Ultra TMB-ELISA Substrate Solution (100 µL per well, ThermoScientific) was added. Absorbance of TMB substrate

was measured twice at 652 nm by UV-Vis plate reader (BioTek) after 5 and 15 min of incubation.

Serum coronavirus antigen microarray (COVAM)

COVAM included 61 antigens across respiratory virus subtypes including 11 antigens from SARS-CoV-2 expressed in either baculovirus or HEK-293 cells as previously detailed.[22] These antigens were provided by Sino Biological U.S. Inc. as either catalog products or custom synthesis service products. The antigens were printed onto microarrays, probed with human sera, and analyzed as previously described [50–52]. Briefly, lyophilized antigens were reconstituted with sterile water to a concentration of 0.1 mg/mL protein in PBS, and printing buffer was added. Antigens were then printed onto ONCYTE AVID nitrocellulose-coated slides (Grace Bio-Labs) using an OmniGrid 100 microarray printer (GeneMachines). The microarray slides were probed with human sera diluted 1:100 in 1X Protein Array Blocking Buffer (GVS Life Sciences, Sanford, ME) overnight at 4°C and washed with TTBS buffer (20 mM Tris-HCl, 150 mM NaCl, 0.05% Tween-20 in ddH2O adjusted to pH 7.5 and filtered) three times for 5 min each. A mixture of human IgG and IgM secondary antibodies conjugated to quantum dot fluorophores Q800 and Q585 respectively was applied to each of the microarray pads and incubated for 2 h at room temperature, and pads were then washed with TTBS three times for 5 min each and dried. The slides were imaged using an ArrayCam imager (Grace Bio-Labs) to measure background-subtracted median spot fluorescence. Non-specific binding of secondary antibodies was subtracted using a saline control. The mean fluorescence of the 4 replicate spots for each antigen was used for analysis.

Statistical analysis

The ELISA data were analyzed in GraphPad Prism 8. Since the total antibody content differs from person to person, the raw absorbance values for every patient sample were normalized and represented as the ratio as compared to a negative control. Analysis of variance (ANOVA) with Dunnett's multiple comparisons was performed to determine if values were statistically significant. Correlations between COVAM IgG/IgM and ELISA were determined by plotting normalized values on an XY graph and performing a non-parametric correlation analysis using a Spearman's rank correlation coefficient test.

For data visualization of clinical patient data, trends in data were evaluated using Knime Analytics Platform software. GraphPad Prism was used to calculate column statistics including mean, standard deviation, SEM, p-values, Odds Ratios, and Likelihood Ratios defined as sensitivity / (1 - specificity). ANOVA with Tukey's multiple comparisons test was used to evaluate antibody response and disease severity between patients with αEp9 Abs, non-Ep9, αN Abs, or non αN Abs. Comparisons of patients with αEp9 Abs and non-αEp9 Abs were conducted using unpaired, two-tailed, parametric t-tests. Contingency graphs were statistically evaluated using Fisher's exact test, for groups with binary categorization, and Chi-squared test for groups with multiple categories. Different datasets were fitted with linear or non-linear regression methods, the fit with the higher R^2 value was chosen. Correlations between two clinical parameters (e.g., IL-6 and AST) were evaluated using the Pearson coefficient or Spearman coefficients (r) for linear or non-linear regressions, respectively; r-values between 1.0-0.7 were considered strong correlations, r-values between 0.7 and 0.5 were considered a moderate correlation, and

values below 0.5 were considered a weak correlation.[53] The significance of the correlation was evaluated based on p-value <0.05.

5.7 References

(1) World Health Organization. *Coronavirus Disease (COVID-19)*; Geneva, Switzerland, 2021.

(2) Llanes, A.; Restrepo, C. M.; Caballero, Z.; Rajeev, S.; Kennedy, M. A.; Lleonart, R. Betacoronavirus Genomes: How Genomic Information Has Been Used to Deal with Past Outbreaks and the Covid-19 Pandemic. *Int. J. Mol. Sci.* **2020**, *21*, 1–28.

(3) Richardson, S.; Hirsch, J. S.; Narasimhan, M.; Crawford, J. M.; McGinn, T.; Davidson, K. W.; Barnaby, D. P.; Becker, L. B.; Chelico, J. D.; Cohen, S. L.; Cookingham, J.; Coppa, K.; Diefenbach, M. A.; Dominello, A. J.; Duer-Hefele, J.; Falzon, L.; Gitlin, J.; Hajizadeh, N.; Harvin, T. G.; Hirschwerk, D. A.; Kim, E. J.; Kozel, Z. M.; Marrast, L. M.; Mogavero, J. N.; Osorio, G. A.; Qiu, M.; Zanos, T. P. Presenting Characteristics, Comorbidities, and Outcomes among 5700 Patients Hospitalized with COVID-19 in the New York City Area. *JAMA - J. Am. Med. Assoc.* **2020**, *323*, 2052–2059.

(4) Gallo Marin, B.; Aghagoli, G.; Lavine, K.; Yang, L.; Siff, E. J.; Chiang, S. S.; Salazar-Mather, T. P.; Dumenco, L.; Savaria, M. C.; Aung, S. N.; Flanigan, T.; Michelow, I. C. Predictors of COVID-19 Severity: A Literature Review. *Rev. Med. Virol.* **2020**, *31*, 1–10.

(5) Song, P.; Li, W.; Xie, J.; Hou, Y.; You, C. Cytokine Storm Induced by SARS-CoV-2. *Clin. Chim. Acta* **2020**, *509*, 280–287.

(6) Iwasaki, A.; Yang, Y. The Potential Danger of Suboptimal Antibody Responses in COVID-19. *Nat. Rev. Immunol.* **2020**, *20*, 339–341.

(7) Arvin, A. M.; Fink, K.; Schmid, M. A.; Cathcart, A.; Spreafico, R.; Havenar-Daughton, C.; Lanzavecchia, A.; Corti, D.; Virgin, H. W. A Perspective on Potential Antibody-Dependent Enhancement of SARS-CoV-2. *Nature* **2020**, *584*, 353–363.

(8) Huang, L. R.; Chiu, C. M.; Yeh, S. H.; Huang, W. H.; Hsueh, P. R.; Yang, W. Z.; Yang, J. Y.; Su, I. J.; Chang, S. C.; Chen, P. J. Evaluation of Antibody Responses against SARS Coronaviral Nucleocapsid or Spike Proteins by Immunoblotting or ELISA. *J. Med. Virol.* **2004**, *73*, 338–346.

(9) He, Y.; Zhou, Y.; Siddiqui, P.; Niu, J.; Jiang, S. Identification of Immunodominant Epitopes on the Membrane Protein of the Severe Acute Respiratory Syndrome-

Associated Coronavirus. *J. Clin. Microbiol.* **2005**, *43*, 3718–3726.

(10) Long, Q. X.; Liu, B. Z.; Deng, H. J.; Wu, G. C.; Deng, K.; Chen, Y. K.; Liao, P.; Qiu, J. F.; Lin, Y.; Cai, X. F.; Wang, D. Q.; Hu, Y.; Ren, J. H.; Tang, N.; Xu, Y. Y.; Yu, L. H.; Mo, Z.; Gong, F.; Zhang, X. L.; Tian, W. G.; Hu, L.; Zhang, X. X.; Xiang, J. L.; Du, H. X.; Liu, H. W.; Lang, C. H.; Luo, X. H.; Wu, S. B.; Cui, X. P.; Zhou, Z.; Zhu, M. M.; Wang, J.; Xue, C. J.; Li, X. F.; Wang, L.; Li, Z. J.; Wang, K.; Niu, C. C.; Yang, Q. J.; Tang, X. J.; Zhang, Y.; Liu, X. M.; Li, J. J.; Zhang, D. C.; Zhang, F.; Liu, P.; Yuan, J.; Li, Q.; Hu, J. L.; Chen, J.; Huang, A. L. Antibody Responses to SARS-CoV-2 in Patients with COVID-19. *Nat. Med.* **2020**, *26*, 845–848.

(11) Batra, M.; Tian, R.; Zhang, C.; Clarence, E.; Sacher, C. S.; Miranda, J. N.; De La Fuente, J. R. O.; Mathew, M.; Green, D.; Patel, S.; Bastidas, M. V. P.; Haddadi, S.; Murthi, M.; Gonzalez, M. S.; Kambali, S.; Santos, K. H. M.; Asif, H.; Modarresi, F.; Faghihi, M.; Mirsaeidi, M. Role of IgG against N-Protein of SARS-CoV2 in COVID19 Clinical Outcomes. *Sci. Rep.* **2021**, *11*, 3455–3464.

(12) Fast, E.; Altman, R. B.; Chen, B. Potential T-Cell and B-Cell Epitopes of 2019-NCoV. *bioRxiv* **2020**, 1–9.

(13) Grifoni, A.; Sidney, J.; Zhang, Y.; Scheuermann, R. H.; Peters, B.; Sette, A. A Sequence Homology and Bioinformatic Approach Can Predict Candidate Targets for Immune Responses to SARS-CoV-2. *Cell Host Microbe* **2020**, *27*, 671–680.

(14) Baruah, V.; Bose, S. Immunoinformatics-Aided Identification of T Cell and B Cell Epitopes in the Surface Glycoprotein of 2019-NCoV. *J. Med. Virol.* **2020**, *92*, 495–500.

(15) Zheng, M.; Song, L. Novel Antibody Epitopes Dominate the Antigenicity of Spike Glycoprotein in SARS-CoV-2 Compared to SARS-CoV. *Cell. Mol. Immunol.* **2020**, *17*, 536–538.

(16) Tilocca, B.; Soggiu, A.; Sanguinetti, M.; Musella, V.; Britti, D.; Bonizzi, L.; Urbani, A.; Roncada, P. Comparative Computational Analysis of SARS-CoV-2 Nucleocapsid Protein Epitopes in Taxonomically Related Coronaviruses. *Microbes Infect.* **2020**, *22*, 188–194.

(17) Rakib, A.; Sami, S. A.; Mimi, N. J.; Chowdhury, M. M.; Eva, T. A.; Nainu, F.; Paul, A.; Shahriar, A.; Tareq, A. M.; Emon, N. U.; Chakraborty, S.; Shil, S.; Mily, S. J.; Ben Hadda, T.; Almalki, F. A.; Emran, T. Bin. Immunoinformatics-Guided Design of an Epitope-Based Vaccine against Severe Acute Respiratory Syndrome Coronavirus 2 Spike Glycoprotein. *Comput. Biol. Med.* **2020**, *124*, 103967–103983.

(18) Wang, L.; Candia, J.; Ma, L.; Zhao, Y.; Imberti, L.; Sottini, A.; Dobbs, K.; NIAID-NCI COVID Consortium; Lisco, A.; Sereti, I.; Su, H. C.; Notarangelo, L. D.; Wang, X. W. Serological Responses to Human Virome Define Clinical Outcomes of Italian

Patients Infected with SARS-CoV-2. *medRxiv* **2020**, 1–35.

(19) Zamecnik, C. R.; Rajan, J. V.; Yamauchi, K. A.; Mann, S. A.; Loudermilk, R. P.; Sowa, G. M.; Zorn, K. C.; Alvarenga, B. D.; Gaebler, C.; Caskey, M.; Stone, M.; Norris, P. J.; Gu, W.; Chiu, C. Y.; Ng, D.; Byrnes, J. R.; Zhou, X. X.; Wells, J. A.; Robbiani, D. F.; Nussenzweig, M. C.; DeRisi, J. L.; Wilson, M. R. ReScan, a Multiplex Diagnostic Pipeline, Pans Human Sera for SARS-CoV-2 Antigens. *Cell Reports Med.* **2020**, *1*, 100123–100142.

(20) Naqiah Amrun, S.; Yi-Pin Lee, C.; Lee, B.; Fong, S.-W.; Edward Young, B.; Sin-Ling Chee, R.; Kim-Wah Yeo, N.; Torres-Ruesta, A.; Carissimo, G.; Meng Poh, C.; Wei Chang, Z.; Zirui Tay, M.; Chan, Y.-H.; I-Cheng Chen, M.; Guek-Hong Low, J.; Tambyah, P. A.; Kalimuddin, S.; Pada, S.; Tan, S.-Y.; Jin Sun, L.; Leo, Y.-S.; Lye, D. C.; Renia, L.; Ng, L. F. Linear B-Cell Epitopes in the Spike and Nucleocapsid Proteins as Markers of SARS-CoV-2 Exposure and Disease Severity. *EBioMedicine* **2020**, *58*, 102911–102919.

(21) Musicò, A.; Frigerio, R.; Mussida, A.; Barzon, L.; Sinigaglia, A.; Riccetti, S.; Gobbi, F.; Piubelli, C.; Bergamaschi, G.; Chiari, M.; Gori, A.; Cretich, M. SARS-CoV-2 Epitope Mapping on Microarrays Highlights Strong Immune-Response to n Protein Region. *Vaccines* **2021**, *9*, 35–46.

(22) Hedde, P. N.; Abram, T. J.; Jain, A.; Nakajima, R.; Ramiro de Assis, R.; Pearce, T.; Jasinskas, A.; Toosky, M. N.; Khan, S.; Felgner, P. L.; Gratton, E.; Zhao, W. A Modular Microarray Imaging System for Highly Specific COVID-19 Antibody Testing. *Lab Chip* **2020**, *20*, 3302–3309.

(23) Walls, A. C.; Park, Y. J.; Tortorici, M. A.; Wall, A.; McGuire, A. T.; Veesler, D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **2020**, *181*, 281–292.

(24) Lan, J.; Ge, J.; Yu, J.; Shan, S.; Zhou, H.; Fan, S.; Zhang, Q.; Shi, X.; Wang, Q.; Zhang, L.; Wang, X. Structure of the SARS-CoV-2 Spike Receptor-Binding Domain Bound to the ACE2 Receptor. *Nature* **2020**, *581*, 215–220.

(25) Baric, R. S. Emergence of a Highly Fit SARS-CoV-2 Variant. *N. Engl. J. Med.* **2020**, *383*, 2684–2686.

(26) Plante, J. A.; Liu, Y.; Liu, J.; Xia, H.; Johnson, B. A.; Lokugamage, K. G.; Zhang, X.; Muruato, A. E.; Zou, J.; Fontes-Garfias, C. R.; Mirchandani, D.; Scharton, D.; Bilello, J. P.; Ku, Z.; An, Z.; Kalveram, B.; Freiberg, A. N.; Menachery, V. D.; Xie, X.; Plante, K. S.; Weaver, S. C.; Shi, P. Y. Spike Mutation D614G Alters SARS-CoV-2 Fitness. *Nature* **2020**, *575*, 1–19.

(27) Korber, B.; Fischer, W. M.; Gnanakaran, S.; Yoon, H.; Theiler, J.; Abfalterer, W.; Hengartner, N.; Giorgi, E. E.; Bhattacharya, T.; Foley, B.; Hastie, K. M.; Parker, M.

D.; Partridge, D. G.; Evans, C. M.; Freeman, T. M.; de Silva, T. I.; Angyal, A.; Brown, R. L.; Carrilero, L.; Green, L. R.; Groves, D. C.; Johnson, K. J.; Keeley, A. J.; Lindsey, B. B.; Parsons, P. J.; Raza, M.; Rowland-Jones, S.; Smith, N.; Tucker, R. M.; Wang, D.; Wyles, M. D.; McDanal, C.; Perez, L. G.; Tang, H.; Moon-Walker, A.; Whelan, S. P.; LaBranche, C. C.; Saphire, E. O.; Montefiori, D. C. Tracking Changes in SARS-CoV-2 Spike: Evidence That D614G Increases Infectivity of the COVID-19 Virus. *Cell* **2020**, *182*, 812–827.

(28) Zheng, W.; Li, Y.; Zhang, C.; Pearce, R.; Mortuza, S. M.; Zhang, Y. Deep-Learning Contact-Map Guided Protein Structure Prediction in CASP13. *Proteins Struct. Funct. Bioinforma.* **2019**, *87*, 1149–1164.

(29) Tian, W.; Jiang, W.; Yao, J.; Nicholson, C. J.; Li, R. H.; Sigurslid, H. H.; Wooster, L.; Rotter, J. I.; Guo, X.; Malhotra, R. Predictors of Mortality in Hospitalized COVID-19 Patients: A Systematic Review and Meta-Analysis. *J. Med. Virol.* **2020**, *92*, 1875–1883.

(30) Simonnet, A.; Chetboun, M.; Poissy, J.; Raverdy, V.; Noulette, J.; Duhamel, A.; Labreuche, J.; Mathieu, D.; Pattou, F.; Jourdain, M.; Caizzo, R.; Caplan, M.; Cousin, N.; Duburcq, T.; Durand, A.; El kalioubie, A.; Favory, R.; Garcia, B.; Girardie, P.; Goutay, J.; Houard, M.; Jaillette, E.; Kostuj, N.; Ledoux, G.; Mathieu, D.; Moreau, A. S.; Niles, C.; Nseir, S.; Onimus, T.; Parmentier, E.; Préau, S.; Robriquet, L.; Rouze, A.; Six, S.; Verkindt, H. High Prevalence of Obesity in Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) Requiring Invasive Mechanical Ventilation. *Obesity* **2020**, *28*, 1195–1199.

(31) Lee, L. Y. W.; Cazier, J. B.; Starkey, T.; Briggs, S. E. W.; Arnold, R.; Bisht, V.; Booth, S.; Campton, N. A.; Cheng, V. W. T.; Collins, G.; Curley, H. M.; Earwaker, P.; Fittall, M. W.; Gennatas, S.; Goel, A.; Hartley, S.; Hughes, D. J.; Kerr, D.; Lee, A. J. X.; Lee, R. J.; Lee, S. M.; Mckenzie, H.; Middleton, C. P.; Murugaesu, N.; Newsom-Davis, T.; Olsson-Brown, A. C.; Palles, C.; Powles, T.; Protheroe, E. A.; Purshouse, K.; Sharma-Oates, A.; Sivakumar, S.; Smith, A. J.; Topping, O.; Turnbull, C. D.; Várnai, C.; Briggs, A. D. M.; Middleton, G.; Kerr, R.; Gault, A.; Agnieszka, M.; Bedair, A.; Ghaus, A.; Akingboye, A.; Maynard, A.; Pawsey, A.; Mohamed, A. A.; Okines, A.; Massey, A.; Kwan, A.; Ferreira, A.; Angelakas, A.; Wu, A.; Tivey, A.; Armstrong, A.; Madhan, A.; Pillai, A.; Poon-King, A.; Kurec, B.; Usborne, C.; Dobeson, C.; Thirlwell, C.; Mitchell, C.; Sng, C.; Scrase, C.; Jingree, C.; Brunner, C.; Fuller, C.; Griffin, C.; Barrington, C.; Muller, D.; Ottaviani, D.; Gilbert, D.; Tacconi, E.; Copson, E.; Renninson, E.; Cattell, E.; Burke, E.; Smith, F.; Holt, F.; Soosaipillai, G.; Boyce, H.; Shaw, H.; Hollis, H.; Bowyer, H.; Anil, I.; Illingworth Gibson, J. J.; Bhosle, J.; Best, J.; Barrett, J.; Noble, J.; Sacco, J.; Chacko, J.; Chackathayil, J.; Banfill, K.; Feeney, L.; Horsley, L.; Cammaert, L.; Mukherjee, L.; Eastlake, L.; Devereaux, L.; Melcher, L.; Cook, L.; Teng, M.; Hewish, M.; Bhattacharyya, M.; Choudhury, M.; Baxter, M.; Scott-Brown, M.; Fittall, M.; Tilby, M.; Rowe, M.; Alihilali, M.; Galazi, M.; Yousaf, N.; Chopra, N.; Cox, N.; Chan, O.; Sheikh, O.; Ramage, P.; Greaves, P.; Leonard, P.; Hall, P. S.; Naksukpaiboon,

P.; Corrie Peck, P. R.; Sharkey, R.; Bolton, R.; Sargent, R.; Jyothirmayi, R.; Goldstein, R.; Oakes, R.; Shotton Kanani, R. R.; Board, R.; Pettengell, R.; Claydon, R.; Moody, S.; Massalha, S.; Kathirgamakarthigeyan, S.; Dolly, S.; Derby, S.; Lowndes, S.; Benafif, S.; Kingdon, S.; Ayers, S.; Brown, S.; Ellis, S.; Parikh, S.; Pugh, S.; Shamas, S.; Wyatt, S.; Grumett, S.; Lau, S.; Wong, Y. N. S.; McGrath, S.; Cornthwaite, S.; Eeckelaers, S.; Hibbs, S.; Tillet, T.; Rabbi, T.; Robinson, T.; Roques, T.; Angelis, V.; Woodcock, V.; Brown, V.; Peng, Y. Y.; Drew, Y.; Hudson, Z. COVID-19 Prevalence and Mortality in Patients with Cancer and the Effect of Primary Tumour Subtype and Patient Demographics: A Prospective Cohort Study. *Lancet Oncol.* **2020**, *21*, 1309–1316.

(32) Zhang, L.; Sun, W.; Wang, Y.; Wang, X.; Liu, Y.; Zhao, S.; Long, D.; Chen, L.; Yu, L. Clinical Course and Mortality of Stroke Patients with Coronavirus Disease 2019 in Wuhan, China. *Stroke* **2020**, No. September, 2674–2682.

(33) Charlson, M. E.; Pompei, P.; Ales, K. L.; MacKenzie, C. R. A New Method of Classifying Prognostic Comorbidity in Longitudinal Studies: Development and Validation. *J. Chronic Dis.* **1987**, *40*, 373–383.

(34) Lu, L.; Zhang, H.; Zhan, M.; Jiang, J.; Yin, H.; Dauphars, D. J.; Li, S. Y.; Li, Y.; He, Y. W. Preventing Mortality in COVID-19 Patients: Which Cytokine to Target in a Raging Storm? *Front. Cell Dev. Biol.* **2020**, *8*, 677–687.

(35) Cummings, M. J.; Baldwin, M. R.; Abrams, D.; Jacobson, S. D.; Meyer, B. J.; Balough, E. M.; Aaron, J. G.; Claassen, J.; Rabbani, L. R. E.; Hastie, J.; Hochman, B. R.; Salazar-Schicchi, J.; Yip, N. H.; Brodie, D.; O'Donnell, M. R. Epidemiology, Clinical Course, and Outcomes of Critically Ill Adults with COVID-19 in New York City: A Prospective Cohort Study. *Lancet* **2020**, *395*, 1763–1770.

(36) Yang, Y.; Shen, C.; Li, J.; Yuan, J.; Wei, J.; Huang, F.; Wang, F.; Li, G.; Li, Y.; Xing, L.; Peng, L.; Yang, M.; Cao, M.; Zheng, H.; Wu, W.; Zou, R.; Li, D.; Xu, Z.; Wang, H.; Zhang, M.; Zhang, Z.; Gao, G. F.; Jiang, C.; Liu, L.; Liu, Y. Plasma IP-10 and MCP-3 Levels Are Highly Associated with Disease Severity and Predict the Progression of COVID-19. *J. Allergy Clin. Immunol.* **2020**, *146*, 119–127.

(37) Shrock, E.; Fujimura, E.; Kula, T.; Timms, R. T.; Lee, I.-H.; Leng, Y.; Robinson, M. L.; Sie, B. M.; Li, M. Z.; Chen, Y.; Logue, J.; Zuiani, A.; McCulloch, D.; Lelis, F. J. N.; Henson, S.; Monaco, D. R.; Travers, M.; Habibi, S.; Clarke, W. A.; Caturegli, P.; Laeyendecker, O.; Piechocka-Trocha, A.; Li, J.; Khatri, A.; Chu, H. Y.; Villani, A.-C.; Kays, K.; Goldberg, M. B.; Hacohen, N.; Filbin, M. R.; Yu, X. G.; Walker, B. D.; Wesemann, D. R.; Larman, H. B.; Lederer, J. A.; Elledge, S. J. Viral Epitope Profiling of COVID-19 Patients Reveals Cross-Reactivity and Correlates of Severity. *Science* **2020**, *370*, 1–22.

(38) Tan, W.; Lu, Y.; Zhang, J.; Wang, J.; Dan, Y.; Tan, Z.; He, X.; Qian, C.; Sun, Q.; Hu, Q.; Liu, H.; Ye, S.; Xiang, X.; Zhou, Y.; Zhang, W.; Guo, Y.; Wang, X. H.; He,

W.; Wan, X.; Sun, F.; Wei, Q.; Chen, C.; Pan, G.; Xia, J.; Mao, Q.; Chen, Y.; Deng, G. Viral Kinetics and Antibody Responses in Patients with COVID-19. *medRxiv* **2020**.

(39) Jiang, H. wei; Li, Y.; Zhang, H. nan; Wang, W.; Yang, X.; Qi, H.; Li, H.; Men, D.; Zhou, J.; Tao, S. ce. SARS-CoV-2 Proteome Microarray for Global Profiling of COVID-19 Specific IgG and IgM Responses. *Nat. Commun.* **2020**, *11*, 3581–3592.

(40) Benucci, M.; Giannasi, G.; Cecchini, P.; Gobbi, F. L.; Damiani, A.; Grossi, V.; Infantino, M.; Manfredi, M. COVID-19 Pneumonia Treated with Sarilumab: A Clinical Series of Eight Patients. *J. Med. Virol.* **2020**, *92*, 2368–2370.

(41) Kooistra, E. J.; Waalders, N. J. B.; Grondman, I.; Janssen, N. A. F.; de Nooijer, A. H.; Netea, M. G.; van de Veerdonk, F. L.; Ewalds, E.; van der Hoeven, J. G.; Kox, M.; Pickkers, P.; Kooistra, E. J.; Waalders, N. J. B.; Grondman, I.; Janssen, N. A. F.; de Nooijer, A. H.; Netea, M. G.; van de Veerdonk, F. L.; Ewalds, E.; van der Hoeven, J. G.; Kox, M.; Pickkers, P.; Hemelaar, P.; Beunders, R.; Bruse, N.; Frenzel, T.; Schouten, J.; Touw, H.; van der Velde, S.; van der Eng, H.; Roovers, N.; Klop-Riehl, M.; Gerretsen, J.; Claassen, W.; Heesakkers, H.; van Schaik, T.; Buijsse, L.; Joosten, L.; de Mast, Q.; Jaeger, M.; Kouijzer, I.; Dijkstra, H.; Lemmers, H.; van Crevel, R.; van de Maat, J.; Nijman, G.; Moorlag, S.; Taks, E.; Debisarun, P.; Wertheim, H.; Hopman, J.; Rahamat-Langendoen, J.; Bleeker-Rovers, C.; Koenen, H.; Fasse, E.; van Rijssen, E.; Kolkman, M.; van Cranenbroek, B.; Smeets, R.; Joosten, I. Anakinra Treatment in Critically Ill COVID-19 Patients: A Prospective Cohort Study. *Crit. Care* **2020**, *24*, 688–700.

(42) Langer-Gould, A.; Smith, J. B.; Gonzales, E. G.; Castillo, R. D.; Garza Figueroa, J.; Ramanathan, A.; Li, B. H.; Gould, M. K. Early Identification of COVID-19 Cytokine Storm and Treatment with Anakinra or Tocilizumab. *Int. J. Infect. Dis.* **2020**, *99*, 291–297.

(43) Castelnovo, L.; Tamburello, A.; Lurati, A.; Zaccara, E.; Marrazza, M. G.; Olivetti, M.; Mumoli, N.; Mastroiacovo, D.; Colombo, D.; Ricchiuti, E.; Paola, F.; Mazzone, A. Anti-IL6 Treatment of Serious COVID-19 Disease. *Med. (Baltimore, MD)* **2021**, *100*, 1–6.

(44) Salama, C.; Han, J.; Yau, L.; Reiss, W. G.; Kramer, B.; Neidhart, J. D.; Criner, G. J.; Kaplan-Lewis, E.; Baden, R.; Pandit, L.; Cameron, M. L.; Garcia-Diaz, J.; Chávez, V.; Mekebeb-Reuter, M.; Lima de Menezes, F.; Shah, R.; González-Lara, M. F.; Assman, B.; Freedman, J.; Mohan, S. V. Tocilizumab in Patients Hospitalized with Covid-19 Pneumonia. *N. Engl. J. Med.* **2021**, *384*, 20–30.

(45) Veiga, V. C.; Prats, J. A. G. G.; Farias, D. L. C.; Rosa, R. G.; Dourado, L. K.; Zampieri, F. G.; MacHado, F. R.; Lopes, R. D.; Berwanger, O.; Azevedo, L. C. P.; Avezum, Á.; Lisboa, T. C.; Rojas, S. S. O.; Coelho, J. C.; Leite, R. T.; Carvalho, J. C.; Andrade, L. E. C.; Sandes, A. F.; Pintão, M. C. T.; Castro, C. G.; Santos, S. V.;

De Almeida, T. M. L.; Costa, A. N.; Gebara, O. C. E.; De Freitas, F. G. R.; Pacheco, E. S.; MacHado, D. J. B.; Martin, J.; Conceição, F. G.; Siqueira, S. R. R.; Damiani, L. P.; Ishihara, L. M.; Schneider, D.; De Souza, D.; Cavalcanti, A. B.; Scheinberg, P. Effect of Tocilizumab on Clinical Outcomes at 15 Days in Patients with Severe or Critical Coronavirus Disease 2019: Randomised Controlled Trial. *BMJ* **2021**, *372*, n84.

(46) Yamaoka, Y.; Jeremiah, S. S.; Miyakawa, K.; Saji, R.; Nishii, M.; Takeuchi, I.; Ryo, A. Whole Nucleocapsid Protein of Severe Acute Respiratory Syndrome Coronavirus 2 May Cause False-Positive Results in Serological Assays. *Clin. Infect. Dis.* **2021**, *72*, 1291–1292.

(47) Murase, K.; Morrison, K. L.; Tam, P. Y.; Stafford, R. L.; Jurnak, F.; Weiss, G. A. EF-Tu Binding Peptides Identified, Dissected, and Affinity Optimized by Phage Display. *Chem. Biol.* **2003**, *10*, 161–168.

(48) Bhasin, A.; Sanders, E. C.; Ziegler, J. M.; Briggs, J. S.; Drago, N. P.; Attar, A. M.; Santos, A. M.; True, M. Y.; Ogata, A. F.; Yoon, D. V.; Majumdar, S.; Wheat, A. J.; Patterson, S. V.; Weiss, G. A.; Penner, R. M. Virus Bioresistor (VBR) for Detection of Bladder Cancer Marker DJ-1 in Urine at 10 PM in One Minute. *Anal. Chem.* **2020**, *92*, 6654–6666.

(49) Pastorino, B.; Touret, F.; Gilles, M.; Lamballerie, X. De; Charrel, R. N. Heat Inactivation of Different Types of SARS-CoV-2 Samples: What Protocols for Biosafety, Molecular Detection and Serological Diagnostics? *Viruses* **2020**, *12*, 735–738.

(50) Khan, S.; Jain, A.; Taghavian, O.; Nakajima, R.; Jasinskas, A.; Supnet, M.; Felgner, J.; Davies, J.; de Assis, R. R.; Jan, S.; Obiero, J.; Strahsburger, E.; Pone, E. J.; Liang, L.; Davies, D. H.; Felgner, P. L. Use of an Influenza Antigen Microarray to Measure the Breadth of Serum Antibodies across Virus Subtypes. *J. Vis. Exp.* **2019**, *(149)*, 1–8.

(51) Jain, A.; Taghavian, O.; Vallejo, D.; Dotsey, E.; Schwartz, D.; Bell, F. G.; Greef, C.; Davies, D. H.; Grudzien, J.; Lee, A. P.; Felgner, P.; Liang, L. Evaluation of Quantum Dot Immunofluorescence and a Digital CMOS Imaging System as an Alternative to Conventional Organic Fluorescence Dyes and Laser Scanning for Quantifying Protein Microarrays. *Proteomics* **2016**, *16*, 1271–1279.

(52) Nakajima, R.; Supnet, M.; Jasinskas, A.; Jain, A.; Taghavian, O.; Obiero, J.; Milton, D. K.; Chen, W. H.; Grantham, M.; Webby, R.; Krammer, F.; Carter, D.; Felgner, P. L.; Davies, D. H. Protein Microarray Analysis of the Specificity and Cross-Reactivity of Influenza Virus Hemagglutinin-Specific Antibodies. *mSphere* **2018**, *3*, 1–15.

(53) Mukaka, M. M. Statistics Corner: A Guide to Appropriate Use of Correlation Coefficient in Medical Research. *Malawi Med. J.* **2012**, *24*, 69–71.

CHAPTER 6: Conclusion and future directions

The overarching theme and long-term goal of my dissertation has been the development of chemical biology tools for advancing DNA sequencing, antibody-drug conjugates, and viral diagnostics.

6.1 DNA sequencing

DNA polymerase catalyzes the replication of DNA, an essential step in the life cycle of all organisms.[1] This enzyme requires an intricate mechanism to ensure incorporation of the correct deoxynucleotide triphosphate (dNTP) substrate. The high fidelity and efficiency allows for accurate replication of genomes, and has value for many biotechnology applications.[2] One of the earliest and most impactful examples of DNA polymerase was DNA sequencing.[3]

DNA sequencing has revolutionized biology.[4] In less than 20 years, sequencing one human genome declined in price from $2.7 billion in 2003 to less than $1000 today.[5] Illumina's sequencing by synthesis chemistry allows for rapid, cheap and accurate whole genome analysis, but its short read lengths (< 300 bases) typically do not span repetitive regions.[6] One solution is to compliment Illumina's short reads with long reads generated by the Pacific Biosciences (PacBio) SMRT light-based technique or the Oxford Nanopore Technologies (ONT) sequencing platform.[7] PacBio's SMRT sequencing uses fluorescently labelled dNTP, which limits data collection to 4-6 hours. Additionally, an extensive workflow – sample prep, sequencing prep, sequencing – and a sufficient amount of DNA is required to obtain complete coverage. ONT's platform uses a protein nanopore fixed in an electrically resistant polymer membrane which results in a high error

rate (around 10%) using the newest pores and software. Using these single-molecule approaches, ultra-long reads (reads in excess of 100,000 bases) have been sequenced.[4] Technology with increased sequencing capability, lower costs, improved throughput, and reduced errors would further revolutionize biomedicine and research.[8]

The Weiss laboratory in collaboration with the Collins laboratory (UCI) developed a single-walled carbon nanotube field-effect transistor (nanocircuit), which translates enzyme motions into electronic signals.[9–13] This nanocircuit device may illuminate hidden conformational transitions of DNA polymerase activity and reveal new mechanisms and dynamics during dNTP incorporation.[9–11] This technology provides sensitive detection of enzyme conformation at microsecond time scales and over longer durations (e.g., up to 2-weeks). In theory, DNA polymerase nanocircuits could be fine-tuned for direct electrical monitoring in DNA sequencing. Additionally, nanocircuits can capture hidden structural information about DNA beyond the sequence.

To better characterize the temperature range and time scale of nanocircuits, several single-molecule experiments on the thermostable DNA polymerase from *Thermus aquaticus* (Taq) were conducted (Chapter 2). Nanocircuits recorded motions of individual Taq molecules, discriminating between matched and mismatched dNTPs from 22 to 85 °C. Such measurements were attempted in a solution of homopolymeric “polyT” DNA primer-template and/or dNTPs. Future studies incorporating dNTPs into other DNA primer-templates, such as heteropolymeric DNA primer-templates, may confirm the nanocircuit’s temperature range with a variety of substrates. Local motions of four variants recorded Taq dynamically moving between an open conformation and both catalytic and transient closed conformations. Using an experimental resolution of 1 μs, nanocircuits

effectively resolved the previously unobserved, transient closures. The findings suggest transient closures travel the same pathway between the open and closed conformations, as evidenced by similar signal amplitudes and temperature dependence to catalytic closures. Nanocircuits are only able to observe Taq in the closed conformation or open conformation, which includes the apo, and the partially closed or ajar conformations. Therefore, Taq variants with mutations E466C or E712C have been designed to record Taq's partially closed conformations in future experiments. My work on Taq not only resolves hidden dynamics of DNA polymerase, but also illuminates the polymerase chain reaction (PCR) and lays the groundwork for a post-next generation DNA sequencing at elevated temperatures with microsecond resolution.

As a key enzyme of many DNA sequencing technologies, the processive DNA polymerase from the *Bacillus subtilis* bacteriophage phi29 (Φ29) was studied using nanocircuits (Chapter 3). This single-molecule technique produced template- and dNTP-dependent signals from catalytic activity of individual Φ29 molecules. During analysis of the signals obtained, the highly variable nature of the enzyme's catalytic rate was observed. Measurements of Φ29 with homopolymeric DNA templates with 42 unpaired cytosines, thymidines, guanosines, or adenosines demonstrated the enzyme's sensitivity to the template in determining catalytic rate. The enzyme has difficulty processing some homopolymeric templates and pauses during DNA replication. Additional Φ29 variants were designed based on patent literature to putatively reduce polymerase pausing.[14] Ensemble activity assays suggest extrinsic factors may play a role in polymerase pausing of Φ29. Future work may provide insight on how other factors act to stall or reverse pausing of Φ29. Additionally, future work may optimize this technology to offer a faster

and simpler profiling of pausing information, allowing for direct electrical monitoring of polymerase activity to be exploited for DNA sequencing.

6.2 Antibody-drug conjugates

Cancer is the leading cause of death worldwide.[15] Over the last two decades, targeted cancer therapy has been revolutionized by antibody-drug conjugates to deliver highly potent drugs directly to cancer cells. There are currently over 65 antibody-drug conjugates in clinical evaluation, however, many face challenges gaining regulatory approval due to product heterogeniety and instability.[16] Such heterogeneity can result in unwanted behavior, including toxicity and loss of efficacy.[17,18] Methods for preparing antibody-drug conjugates are in high demand for antibody-mediated delivery of chemo- and radiotherapeutic agents to tumor sites.[19,20] New and improved bioconjugation technologies would accelerate the use of antibody-drug conjugates for targeted cancer therapy and eventually benefit cancer patients globally.

The work described here builds a foundation for improving bioconjugation technologies and providing more effective antibody-drug conjugates (Chapter 4). My work discovered that pyrocinchonimide (Pci), the dimethylated analog of maleimide, undergoes a surprising transformation with biomolecules.[21] The reaction targets amines and involves an imide transfer, which has not been previously reported for bioconjugation purposes. Using mass spectrometry methods, the purity and stabilty of antibodies with the Pci motif was chacacterized. The results demonstated that the Pci reaction's selectivity reduces overmodification of the antibody compared to conventional methods. Thus, the Pci motif can reduce combinatorial diversity when many available reactive amines are available.

This finding suggests the Pci reaction may be of value in situations where antibody-drug conjugate development is confounded by product heterogeneity issues,[21] or where a slow release of the drug payload is desirable.

One potential pitfall of the Pci reaction chemistry is the slow reactivity (h) compared to maleimides (min), which would make preparation of antibody conjugate inefficient. An alternative approach would be to collaborate with a synthetic chemist to introduce a more electronegative constituent in the leaving group, which would improve reaction kinetics. Such work can dial in a predictable number of drugs on each antibody, which may increase the effectiviness of the cancer therapy. Future work on new iterations of the Pci bioconjugation reagent may define site selectivity profiles. Additionally, derivatives of the linker can be designed to deliver radiotherapeutic agents to tumor sites to expand the scope of the Pci bioconjugation reaction and eventually be applied to both cancer diagnostics and therapies.

6.3 Viral diagnostics

The coronavirus disease 2019 (COVID-19) pandemic has triggered an ongoing global health crisis. The virus that causes COVID-19, SARS-CoV-2, belongs to the same family viruses linked to recent epidemics - severe acute respiratory syndrome (SARS-CoV-1 termed SARS here) in 2002-2003 and Middle East respiratory syndrome (MERS) in 2012.[22] Patients infected with SARS-CoV-2 generate antibodies, or disease fighting components of the immune system, to mark specific pieces of the virus called antigens for destruction. Antibody detection can provide information on the viral infection time course. A desperate need exists for diagnosing a patient's disease status (e.g., infected

with symptoms, infected but asymptomatic, or fully virus cleared). Such information could potentially predict disease trajectories and guide inventions.

The work described here investigated the antibody response of patients infected by SARS-CoV-2 and developed new type of COVID-19 diagnostic for predicting disease severity (Chapter 5).[23] Using methods of enzyme-linked immunosorbent assay (ELISA) and coronavirus antigen microarray (COVAM) analysis, antibody epitopes were mapped in the plasma of COVID-19 patients (n = 86) experiencing a wide range of disease states. This work identified antibodies to a 21-residue epitope from nucleocapsid (termed Ep9) associated with severe disease, including admission to the intensive care unit, requirement for ventilators, or death. By combining a disease risk factor score with a test for αEp9 antibodies, severe COVID-19 outcomes could be predicted with 13.4 likelihood ratio (96.7% specificity). This two-step prognostic tool could mitigate the worst outcomes of COVID-19, particularly for patients at high risk.

Follow-up investigations in a larger and more diverse population of COVID-19 patients could be conducted to confirm αEp9 antibodies' diagnostic accuracy and the correlation with disease severity. Additionally, since the population examined here only included three asymptomatic individuals, and additional testing is required to confirm absence of αEp9 antibodies in such patients. Future work can adapt existing diagnostic to test for αEp9 antibodies. For example, the Wiess laboratory in collaboration with the Penner laboratory (UCI), has been developing a diagnostic Virus BioResistor technology for rapid (<60 sec) and accurate sensing of αEp9 antibodies in blood. In summary, a better understanding of the antibody response against SARS-CoV-2 could provide key insight for the eventual development of diagnostics, vaccines, and anti-viral therapeutics.

6.4 References

(1) Maga, G. DNA Polymerases. In *Reference Module in Biomedical Sciences*; Elsevier Inc.: Amsterdam, Netherlands, 2019; pp 376–378.

(2) Aschenbrenner, J.; Marx, A. DNA Polymerases and Biotechnological Applications. *Curr. Opin. Biotechnol.* **2017**, *48*, 187–195.

(3) Carrilho, E.; Kist, T. B. L.; Franc, L. T. C. A Review of DNA Sequencing Techniques. **2002**, *2*, 169–200.

(4) Shendure, J.; Balasubramanian, S.; Church, G. M.; Gilbert, W.; Rogers, J.; Schloss, J. A.; Waterston, R. H. DNA Sequencing at 40: Past, Present and Future. *Nature* **2017**, *550*, 345–353.

(5) NHGRI. The cost of sequencing a human genome https://www.genome.gov/ about-genomics/fact-sheets/Sequencing-Human-Genome-cost.

(6) Mardis, E. R. A Decade's Perspective on DNA Sequencing Technology. *Nature* **2011**, *470*, 198–203.

(7) Quail, M. A.; Smith, M.; Coupland, P.; Otto, T. D.; Harris, S. R.; Connor, T. R.; Bertoni, A.; Swerdlow, H. P.; Gu, Y. A Tale of Three next Generation Sequencing Platforms: Comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq Sequencers. *BMC Genomics* **2012**, *13*, 1–13.

(8) NHGRI. DNA Sequencing Fact Sheet https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Fact-Sheet.

(9) Choi, Y.; Moody, I. S.; Sims, P. C.; Hunt, S. R.; Corso, B. L.; Perez, I.; Weiss, G. A.; Collins, P. G. Single-Molecule Lysozyme Dynamics Monitored by an Electronic Circuit. *Science* **2012**, *335*, 319–324.

(10) Gül, O. T.; Pugliese, K. M.; Choi, Y.; Sims, P. C.; Pan, D.; Rajapakse, A. J.; Weiss, G. A.; Collins, P. G. Single Molecule Bioelectronics and Their Application to Amplification-Free Measurement of DNA Lengths. *Biosensors* **2016**, *6*, 1–19.

(11) Olsen, T. J.; Choi, Y.; Sims, P. C.; Gul, O. T.; Corso, B. L.; Dong, C.; Brown, W. A.; Collins, P. G.; Weiss, G. A. Electronic Measurements of Single-Molecule Processing by DNA Polymerase I (Klenow Fragment). *J. Am. Chem. Soc.* **2013**, *135*, 7855–7860.

(12) Akhterov, M. V.; Choi, Y.; Olsen, T. J.; Sims, P. C.; Iftikhar, M.; Gul, O. T.; Corso, B. L.; Weiss, G. A.; Collins, P. G. Observing Lysozyme's Closing and Opening

Motions by High-Resolution Single-Molecule Enzymology. *ACS Chem. Biol.* **2015**, *10*, 1495–1501.

(13) Sims, P. C.; Moody, I. S.; Choi, Y.; Dong, C.; Corso, B. L.; Gul, O. T.; Collins, P. G.; Weiss, G. A. Electronic Measurements of Single-Molecule Catalysis by CAMP-Dependent Protein Kinase A Electronic Measurements of Single-Molecule Catalysis by CAMP-Dependent Protein Kinase A. *J. Am. Chem. Soc.* **2013**, *135*, 7861−7868.

(14) Fuller, C.; Kumar, S.; Sood, A.; Nelson, J. Terminal Phosphate-Labeled Nucleotides with Methods for Use. US0162213A, 2003.

(15) Ferlay, J.; Laversanne, M.; Ervik, M.; Lam, F.; Colombet, M.; Mery, L.; Piñeros, M.; Znaor, A.; Soerjomataram, I.; Bray, F. *Global Cancer Observatory: Cancer Today*; Lyon, France, 2020.

(16) Beck, A.; Goetsch, L.; Dumontet, C.; Corvaïa, N. Strategies and Challenges for the next Generation of Antibody-Drug Conjugates. *Nat. Rev. Drug Discov.* **2017**, *16*, 315–337.

(17) Hamblett, K. J.; Senter, P. D.; Chace, D. F.; Sun, M. M. C.; Lenox, J.; Cerveny, C. G.; Kissler, K. M.; Bernhardt, S. X.; Kopcha, A. K.; Zabinski, R. F.; Meyer, D. L.; Francisco, J. A. Effects of Drug Loading on the Antitumor Activity of a Monoclonal Antibody Drug Conjugate. **2004**, *10*, 7063–7070.

(18) Tsuchikama, K.; An, Z. Antibody-Drug Conjugates: Recent Advances in Conjugation and Linker Chemistries. *Protein Cell* **2018**, *9*, 33–46.

(19) Scott, A. M.; Wolchok, J. D.; Old, L. J. Antibody Therapy of Cancer. *Nat. Publ. Gr.* **2012**, *12*, 278–287.

(20) Agarwal, P.; Bertozzi, C. R. Site-Specific Antibody − Drug Conjugates: The Nexus of Bioorthogonal Chemistry, Protein Engineering, and Drug Development. *Bioconjug. Chem.* **2015**, *26*, 176–192.

(21) Richardson, M. B.; Gabriel, K. N.; Garcia, J.; Ashby, S.; Dyer, R.; Kim, J.; Lau, C.; Hong, J.; Le Tourneau, R. J.; Sen, S.; Narel, D.; Katz, B. B.; Ziller, J. W.; Majumdar, S.; Collins, P. G.; Weiss, G. A. Pyrocinchonimides Conjugate to Amine Groups on Proteins via Imide Transfer. *Bioconjug. Chem.* **2020**, *31*, 1449–1462.

(22) Llanes, A.; Restrepo, C. M.; Caballero, Z.; Rajeev, S.; Kennedy, M. A.; Lleonart, R. Betacoronavirus Genomes: How Genomic Information Has Been Used to Deal with Past Outbreaks and the Covid-19 Pandemic. *Int. J. Mol. Sci.* **2020**, *21*, 1–28.

(23) Sen, S. R.; Sanders, E. C.; Gabriel, K. N.; Miller, B. M.; Isoda, H. M.; Salcedo, G. S.; Garrido, J. E.; Dyer, R. P.; Nakajima, R.; Jain, A.; Caldaruse, A.-M.; Santos, A.

M.; Bhuvan, K.; Tifrea, D. F.; Ricks-Oddie, J. L.; Felgner, P. L.; Edwards, R. A.; Majumdar, S.; Weiss, G. A. Predicting COVID-19 Severity with a Specific Nucleocapsid Antibody plus Disease Risk Factor Score. *mSphere* **2021**, *6*, 1–13.

APPENDIX I: Mutually Orthogonal DNA Replication Systems In Vivo

I.1 Abstract

The yeast cytoplasmically localized pGKL1/TP-DNAP1 plasmid/DNA polymerase pair forms an orthogonal DNA replication system whose mutation rate can be drastically increased without influencing genomic replication, thereby supporting *in vivo* continuous evolution. Here, we report that the pGKL2/TP-DNAP2 plasmid/DNA polymerase pair forms a second orthogonal replication system. We show that custom genes can be encoded and expressed from pGKL2, that error-prone TP-DNAP2s can be engineered, and that pGKL2 replication by TP-DNAP2 is both orthogonal to genomic replication in *Saccharomyces cerevisiae* and mutually orthogonal with pGKL1 replication by TP-DNAP1. This demonstration of two mutually orthogonal DNA replication systems with tunable error rates and properties should enable new applications in cell-based continuous evolution, genetic recording, and synthetic biology at large.

Key reference: Arzumanyan, G. A.; Gabriel, K. N.; Ravikumar, A.; Javanpour, A. A.; Liu, C. C. Mutually Orthogonal DNA Replication Systems in Vivo. *ACS Synth. Biol.* **2018**, *7*, 1722–1729.

My contribution was engineering error-prone TP-DNAP2s, generating library of TP-DNAP2s, and testing p2 fluctuations with leu2* or *URA3* to show that pGKL2 replication by TP-DNAP2 is both orthogonal to genomic replication in *Saccharomyces cerevisiae* and mutually orthogonal with pGKL1 replication by TP-DNAP1. This work was completed during one of my first-year rotations.

I.2 Introduction

Expanding the capabilities of DNA polymerases (DNAPs) to change how they copy DNA *in vitro* has resulted in staple biotechnologies including PCR mutagenesis, DNA sequencing, nucleic acid diagnostics, and the evolution of synthetic genetic polymers.[1–5] Engineering DNAPs with new properties *in vivo* could lead to similarly important advances, including cells that continuously evolve target genes, record lineage or exposure to custom stimuli through mutating barcodes, or use new genetic alphabets for biocontainment.[6–10] However, the operation of highly engineered DNAPs in a living cell is challenging because they can easily harm genomic replication.[11] We recently developed a special DNA replication system in *Saccharomyces cerevisiae* consisting of a DNAP/plasmid pair that is orthogonal to genomic replication and showed that engineered orthogonal DNAPs only replicate the orthogonal plasmid.[6,7] Here, we demonstrate that an additional DNAP/plasmid pair is orthogonal to genomic replication in *S. cerevisiae* and that the two DNAP/plasmid pairs are mutually orthogonal to each other. This solidifies two platforms for independently expanding the properties of DNA replication *in vivo*.

Our two orthogonal replication systems are based on the cytoplasmically localizedpGKL1/pGKL2 (p1/p2) plasmids originating from *Kluyveromyces lactis*.[11,12] Both p1 and p2 encode their own DNAPs, TP-DNAP1 and TP-DNAP2, respectively. We previously showed that engineered error-prone variants of TP-DNAP1 increase the mutation rate of p1 to ~10−5 substitutions per base (spb) without affecting the genomic mutation rate of ~10−10 spb in *S. cerevisiae*.[6] We also found that p1 replication strictly requires TP-DNAP1.[7] This allowed us to conclude that TP-DNAP1 and p1 constitute an orthogonal DNAP/plasmid pair such that engineered changes to TP-DNAP1 only act on

p1 but not on the genome. However, it was not known whether orthogonality to genomic replication holds true for TP-DNAP2 and p2 nor whether the TP-DNAP1/p1 and TP-DNAP2/p2 pairs are mutually orthogonal. Here, we develop error-prone TP-DNAP2s, report the associated genetic techniques necessary to engineer the p2 plasmid and TP-DNAP2, and ultimately prove that the TP-DNAP2/p2 pair is both orthogonal to genomic replication and to p1 replication.

The existence of mutually orthogonal genetically tractable replication systems is significant for three main reasons. First, our finding of two mutually orthogonal DNA replication systems should lay the foundation for novel applications in synthetic biology. For example, *in vivo* accelerated evolution6 of different genes or sets of genes can now be carried out at two distinct custom mutation rates, which could be useful for evolving components in hierarchically organized signaling pathways. Another possibility includes using inducible error-prone orthogonal DNAPs to record multiple cellular events or external stimuli, where the number of mutational events in p1 and p2 would be independent readouts of the amount of exposure to two signals experienced by cells. In addition, the freedom to engineer two orthogonal DNAPs *in vivo* may enable propagation of different XNA's in living cells, whereas current efforts are limited to either using novel base-pairs recognized by host DNAPs or engineering DNAPs to synthesize XNA with novel backbones *in vitro*.[4,5,9] Second is a practical consideration. Our main motivation for developing an orthogonal replication (OrthoRep) system was to achieve continuous rapid evolution of target genes *in vivo* at extreme mutation rates that the genome cannot withstand.[6] This was achieved by making TP-DNAP1 highly error-prone so that it rapidly mutates the p1 plasmid but spares the genome. The mutual orthogonality demonstrated

here ensures that the essential accessory genes encoded on p2 are also spared by error-prone TP-DNAP1s during directed evolution experiments of genes on p1 in OrthoRep. Third, our result provides *in vivo* evidence that DNA initiation of p1 and p2 use independent components. p1 and p2 both contain terminal proteins (TPs) linked to their 5′ termini, which act as origins of replication, akin to other protein-primed DNA replication systems like those found in bacteriophage Φ29 and adenovirus.[13,14] The lack of homology between the TPs of p1 and p2 suggested that TP-DNAP1 and TP-DNAP2 may use distinct molecular interactions for plasmid initiation. In addition, preliminary *in vitro* biochemical data from our lab shows that TP-DNAP1 can initiate replication from p1's inverted terminal repeat (ITR), hypothesized to act in concert with p1's TP to form an original of replication, but cannot initiate replication from p2's ITR (unpublished). Our observation of mutual orthogonality between p1 and p2 replication all but proves that highly specific TP-DNAP interactions with cognate TPs and ITRs govern plasmid initiation, encouraging future studies on the mechanisms of protein-primed DNA replication and suggesting a potential approach for engineering additional orthogonal replication systems that operate concurrently in the same cell.

I.3 Results and Discussion

Strategy for Characterizing Orthogonality

Our strategy for probing orthogonality of the TP-DNAP2/p2 DNAP/ plasmid pair is based on engineering and using error-prone TP-DNAP1 and TP-DNAP2 variants to measure whether they increase mutation rates of genes on p1, p2, and/or the host genome. If error-prone TP-DNAP1s only increase the mutation rate of p1 (but not of p2

and the host genome) and error-prone TP-DNAP2s only increase the mutation rate of p2 (but not of p1 and the host genome), then we may conclude that TP-DNAP1/p1 and TP-DNAP2/p2 are mutually orthogonal DNA replication systems that are both orthogonal to genomic replication. This is indeed what we found.

Integration of Heterologous Genes onto p2 Using CRISPR/Cas9

We first developed a reliable method for encoding and expressing user-defined genes on p2. To measure mutation rates of p2 replication, we needed to encode *mKate2*, *URA3*, and *leu2** (mUL*) on p2. *mKate2* would serve as a fluorescent reporter for copy number, *URA3* as a selection marker, and *leu2**, which contains a stop codon at a permissive site in *LEU2* (Q180*), would serve as a reporter for substitution mutation rates in fluctuation tests that measure reversion to functional *LEU2*. We expected that mUL* could be integrated onto p2 via *in vivo* homologous recombination, following our procedures for manipulating p1.[6,7] We constructed a DNA cassette encoding mUL* flanked by regions homologous to p2 (**Figure I-1**) such that successful recombination would result in the replacement of the nonessential ORF1 found on wildtype (wt) p2.[15] After transformation of this cassette into *S. cerevisiae* strain AR-Y292 containing wt p1 and p2 (**Figure I-2A**, **top**), we isolated several clones exhibiting uracil prototrophy and detectable fluorescence from *mKate2*. Extraction of cytoplasmic plasmids from these clones confirmed presence of the recombinant p2-delORF1-mUL*, but only at low copy as confirmed by DNA gel electrophoresis and PCR with primers specific to p2-delORF1-mUL*. In contrast to similar cassette integrations into p1, passaging under selection for *URA3* expression failed to cure the parental wt p2 plasmid and increase the copy number

of p2-delORF1-mUL* to levels easily detectable by gel electrophoresis. Although p2-delORF1-mUL* encodes all the necessary genes for its own replication and was selected for through *URA3*, it is likely that the shorter size of wt p2 provided it with enough of a replicative advantage to be maintained.

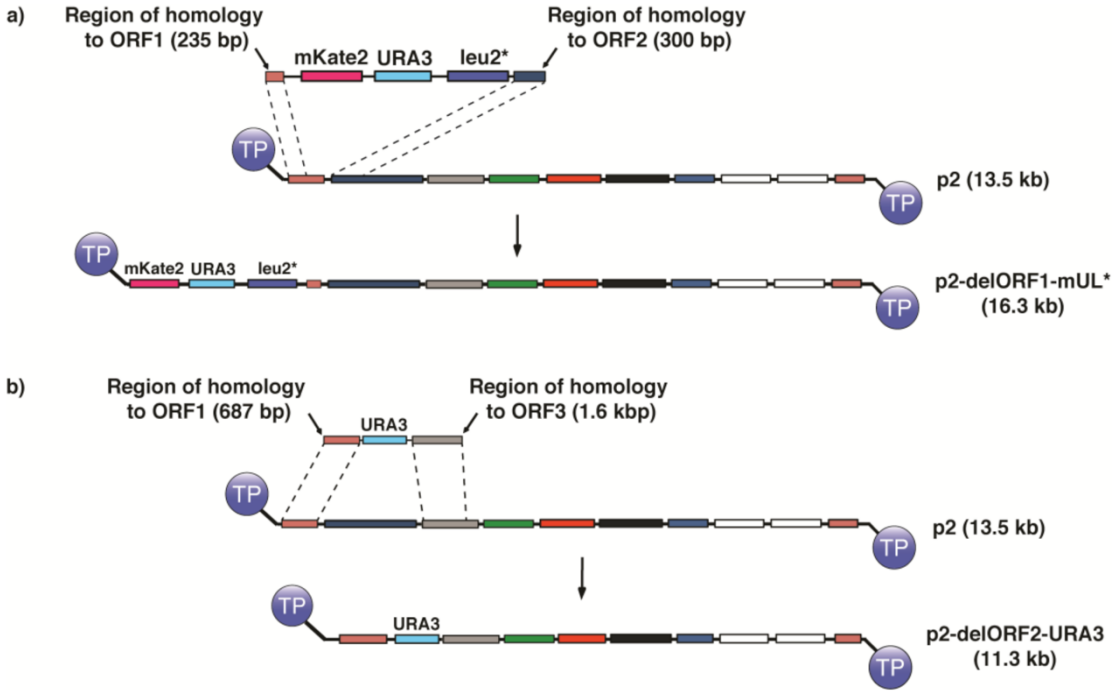


Figure I-1. Homologous recombination cassettes for p2 integration. A) Homologous recombination cassette encoding *mKate2*, *URA3*, and *leu2**, and appropriate flanking regions for integration over ORF1 of p2. This linearized cassette is generated by digestion of pGA1 with ScaI. **B)** homologous recombination cassette encoding *URA3* with flanking regions for integration over *ORF2* of p2, resulting in full deletion of TP-DNAP2. This linearized cassette is generated by digestion of pGA59 with ScaI.

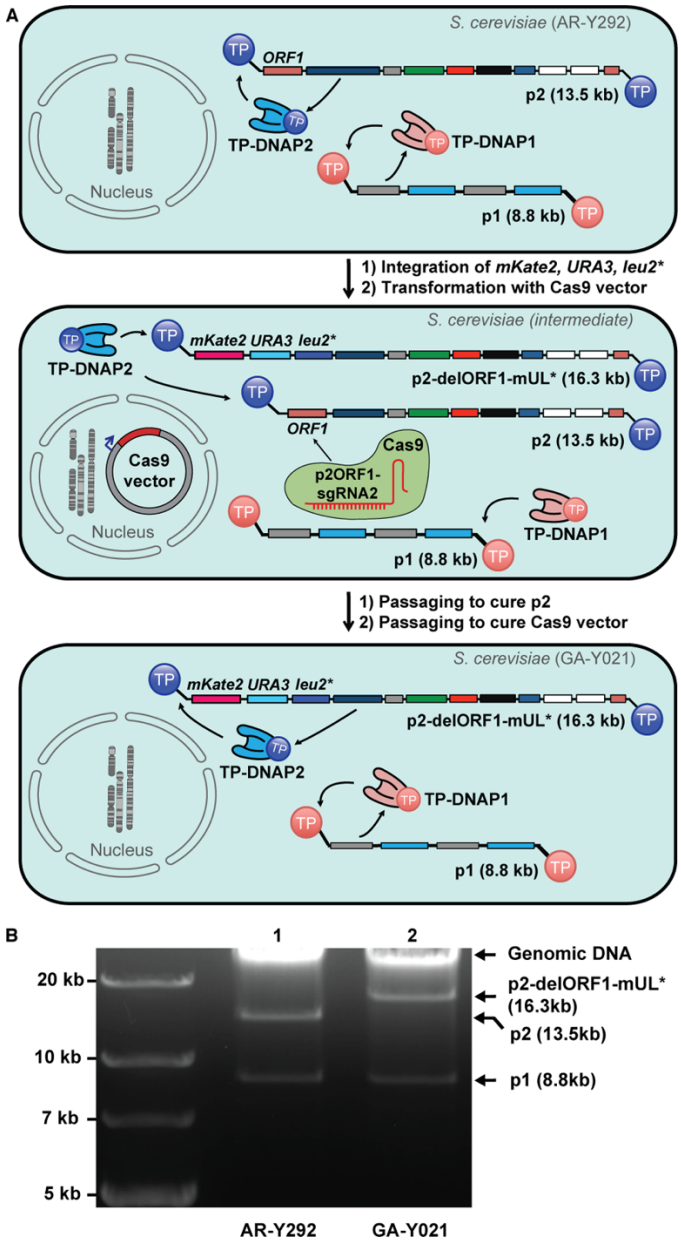


Figure I-2. Engineering recombinant p2. A) Top: AR-Y292 strain harboring p1 and p2, replicated by self-encoded TP-DNAP1 and TP-DNAP2, respectively. Middle: Intermediate strain derived from AR-Y292 by integration of *mKate2*, *URA3*, and *leu2** over ORF1 of p2 and expression of Cas9 targeting ORF1 of p2, which is missing from the recombinant p2-delORF1-mUL*. Bottom: Complete curing of the parental p2 plasmid induced by Cas9 results in GA-Y021, which contains only p1 and recombinant p2-delORF1-mUL*. GA-Y021 is used as the base strain for subsequent measurements of p2 mutation rate. **B)** Gel analysis (0.9% agarose) verifying integration of *mKate2*, *URA3*, *leu2** and loss of parental p2 after Cas9 treatment. Lane 1: DNA extracted from AR-Y292, containing linear plasmids p1 and p2. Lane 2: DNA extracted from GA-Y021, confirming loss of p2 and its replacement by p2-delORF1-mUL* at high copy.

To cure the parental wt p2 plasmid, we attempted more active methods. We used a yeast CRISPR/Cas9 vector25 and three candidate sgRNAs to target ORF1 (**Table I-1**), which is present in wt p2 but not in p2-delORF1-mUL* (**Figure I-2A**, **middle**). One of the three sgRNA's expressed in conjunction with a cytoplasmically localized Cas9 achieved complete curing of p2 and a concomitant increase in the copy number of p2-delORF1-mUL*. This was evidenced by an increase in *mKate2* fluorescent signal and a brighter p2-delORF1-mUL* DNA gel electrophoresis band (**Figure I-2B**). Curing of parental p2 to undetectable levels was confirmed by lack of PCR amplification with primers specific to p2, yielding strain GA-Y021 (**Figure I-2A**, **bottom**).

Table I-1. Target sequences used with Cas9 to cure p2. Cutting efficiency shown for six sgRNA's used for curing linear plasmids with Cas9, based on qualitative observations for number of generations needed for Cas9 for complete curing. (-): no curing, (+): curing required passaging under selection for the Cas9 plasmid (+G418), (+++): curing was complete in colonies picked off Cas9 transformation plates, with no passaging required. ORF1 guide RNA's were used to generate GA-Y021, and *ORF2* guide RNA's were used to cure p2 specifically in the TP-DNAP2 complementation experiment described in **Figure I-3**.

Entry	sgRNA	Target sequence + PAM	Curing efficiency
1	p2ORF1-sgRNA1	ttatgccataaatagaatggAGG	-
2	p2ORF1-sgRNA2	gcaacagtataatttaatttGGG	+
3	p2ORF1-sgRNA3	tgatatttacaaaatgttaaTGG	-
4	p2ORF2-sgRNA1	tcatgtaaggcagtacatgtTGG	+++
5	p2ORF2-sgRNA2	atacatttacttttcatgtaAGG	+++
6	p2ORF2-sgRNA3	ctatatgcatggtatggttcTGG	+++

Using GA-Y021, we measured the mutation rate of p2 replication by wt TP-DNAP2 still encoded on p2-delORF1-mUL*, with a previously described fluctuation test where the number distribution of functional *LEU2* mutants is used to calculate mutation rate by the MSS method.[16–18] The mutation rate of p2-delORF1-mUL* replication was 5.96 × 10−10

spb (95% CI: 3.57 × 10−10 to 8.77 × 10−10) with a copy number of 128 per cell (**Table I-2**, **Entry 1**; **Table I-3**). This is similar to the wild type p1 mutation rate and copy number, which are 1.39 × 10−9 and 124, respectively.[7]

Table I-2. TP-DNAP2 Variants with Elevated Mutation Rates. Entry 1 is the native mutation rate of p2 replication measured in GA-Y021, where p2-delORF1-mUL* is replicated only by the self-encoded wt TP-DNAP2. Entries 2−18 are mutation rate measurements of p2 replication in the presence of 17 error-prone TP-DNAP2 hits from library screening. All per-base substitution rates were measured using *leu2** fluctuation analyses in GA-Y021, where p2-delORF1-mUL* is replicated in tandem by wt TP-DNAP2 and a TP-DNAP2 variant encoded in trans on a nuclear CEN6/ARS4 plasmid. Ranges in parentheses correspond to 95% confidence intervals, determined according to the MSS method.[16–18]

entry	TP-DNAP2 variant *in trans*	per-base substitution rate of p2 ($\times 10^{-10}$)	fold increase
1	none	5.96 (3.57−8.77)	–
2	S370Q	83.0 (62.8−105)	13.9
3	Y424Q	77.5 (56.9−100)	13.0
4	Y424E	73.3 (55.2−93.3)	12.3
5	S370P	50.1 (32.4−70.4)	8.4
6	Y424K	44.4 (32.2−58.0)	7.5
7	S370R	31.9 (22.4−42.5)	5.3
8	Y424G	28.8 (20.0−38.7)	4.8
9	S370E	27.9 (19.7−37.2)	4.7
10	S370K	26.2 (18.0−35.4)	4.4
11	Y424R	19.6 (13.0−27.1)	3.3
12	S370L	14.8 (9.32−21.1)	2.5
13	L474D	12.2 (7.33−17.9)	2.0
14	F882A	11.7 (4.92−20.4)	2.0
15	F882V	10.8 (6.30−16.2)	1.8
16	L474A	8.37 (4.80−12.6)	1.4
17	F882R	7.32 (3.91−11.5)	1.2
18	L474V	5.76 (2.88−9.33)	1.0

Table I-3. Preparation of calibration curve relating p2 copy number to *mKate2* fluorescence. p2 copy number was determined by performing qPCR on DNA extracts from various strains with varying p2 copy number, accompanied with measurements of *mKate2* fluorescence by flow cytometry. Values in parentheses are 95% confidence intervals calculated from triplicate quantitative PCR's. The calibration curve was determined as described in *Materials and Methods*. The linear regression equation relating p2 copy number (y) to mean *mKate2* fluorescence (x) is: y = 0.0562x – 43.639, with R2 = 0.86.

Base strain	Polymerase *in trans*	p2 copy number by absolute quantification of *leu2**	Mean mKate2 fluorescence (a.u.)
GA-Y021	None	248 (219-277)	5093.6
	TP-DNAP1^{WT}	188 (183-193)	4188.9
	TP-DNAP1$^{I777K, L900S}$	130 (117-143)	3445.9
	TP-DNAP1$^{L477V, L640Y, I777K, W814N}$	100 (93-107)	2973.9
	TP-DNAP2^{WT}	94 (90-98)	2388.3
	TP-DNAP2^{Y424Q}	167 (158-176)	3073.3
	TP-DNAP2^{S370Q}	299 (247-350)	2808.6

Replication of p2 by TP-DNAP2 *In Trans*

To facilitate the straightforward testing of error-prone TP-DNAP2s, we showed that p2 replication could be fully sustained by TP-DNAP2 encoded *in trans*, on a standard yeast nuclear plasmid, rather than in cis, on p2. First, TP-DNAP2, which is *ORF2* of p2, was deleted by homologous recombination of a synthetic cassette encoding *URA3* (**Figure I-1**). Since p2 is a multicopy plasmid, the resulting strain (GA-Y069) harbored a mixture of the parental wt p2 and the recombinant p2 with *ORF2* deleted (p2-delORF2-*URA3*), along with unaltered p1 (**Figure I-3A**). In this strain, both the wt p2 and recombinant p2-delORF2-*URA3* plasmids rely on TP-DNAP2 encoded on the parental wt p2 plasmid for replication. Thus, loss of wt p2 should disable replication of p2-delORF2-*URA3*. Indeed, when we cured the parental p2 plasmid by targeting *ORF2* of wt p2 with Cas9 (**Table I-1**), we found that all p2-delORF2-*URA3* was also lost and that the strain could no longer grow in the absence of uracil (**Figure I-3B**). Next, we repeated the same experiment but now in the presence of a codon-optimized TP-DNAP2 expressed *in trans*

from a standard yeast *CEN6/ARS4* nuclear plasmid (pGA55-reTP-DNAP2). After p2 was fully cured by Cas9, we found that p2 was not present, p2-delORF2-*URA3* remained, and this strain could grow in the absence of uracil (**Figure I-3B**). In addition, p1 was also maintained, indicating that the accessory genes encoded on p2-delORF2-*URA3* necessary for replication of both p1 and p2 and transcription of TP-DNAP1 on p1 were still functional. Therefore, p2-derived plasmids can be replicated by TP-DNAP2 encoded on a standard nuclear plasmid, simplifying the characterization of p2 replication by error-prone TP-DNAP2 variants.

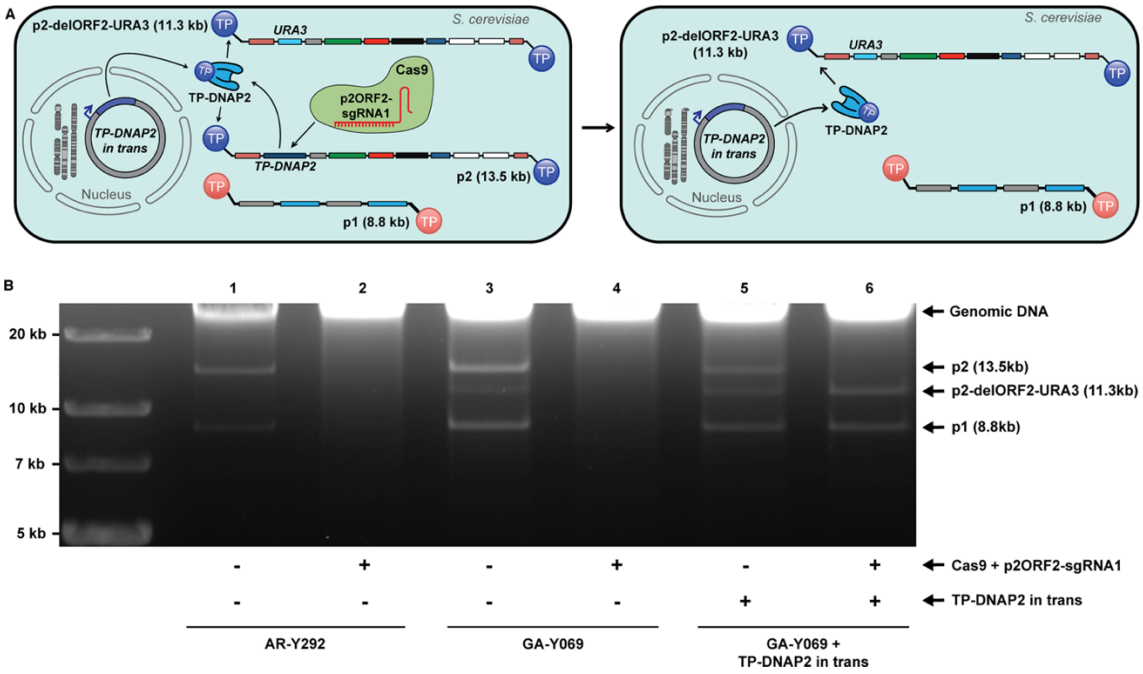


Figure I-3. Functional complementation by recoded TP-DNAP2 *in trans*. A) Left: Linear plasmid system consisting of p1, p2, and recombinant p2-delORF2-*URA3*. Cas9 coexpressed with p2ORF2-sgRNA1 specifically targets double strand breaks to *ORF2* of p2. Right: Complete curing of p2 by Cas9 results in loss of p2's self-encoded TP-DNAP2. This deletion is functionally complemented by TP-DNAP2 encoded *in trans*, which can stably replicate p2. **B)** Gel analysis (0.9% agarose) testing for stability of linear plasmid combinations in the absence (odd-numbered lanes) or the presence (even-numbered lanes) of Cas9 activity targeted to *ORF2* of p2. Lanes 1 and 2: DNA extracted from AR-Y292, containing only wt p1 and p2, which are both lost in the presence of Cas9. Lanes

3 and 4: DNA extracted from GA-Y069 containing p1, p2, and recombinant p2-delORF-*URA3*. Elimination of wt p2 by Cas9 curing eliminates the sole source of TP-DNAP2, causing subsequent loss of p2-delORF1-*URA3* and p1. Following loss of the linear plasmids, the strain in lane 4 was no longer prototrophic for uracil. Lanes 5 and 6: DNA extracted from a GA-Y069 derivative, which contains the linear plasmids p1, p2 and p2-delORF2-*URA3* and encodes TP-DNAP2 on a nuclear *CEN6/ARS4* plasmid *in trans*. Treatment with Cas9 results in a stable linear plasmid system lacking wt p2, instead consisting only of p1 and p2-delORF2-*URA3* replicated entirely by TP-DNAP2 encoded *in trans*.

Discovery of Error-Prone TP-DNAP2 Variants through Library Screening

To discover error-prone TP-DNAP2s, we screened a small library of TP-DNAP2s diversified at locations we hypothesized would be responsible for DNAP fidelity. An alignment between TP-DNAP1 and TP-DNAP2 revealed that S370, Y424, L474, and F882 in TP-DNAP2 were homologous to residues in TP-DNAP1 that we previously found could be mutated to yield error-prone TP-DNAP1s (**Figure I-4**).[6,7] We therefore generated four distinct site-saturation mutagenesis libraries, each diversifying S370, Y424, L474, or F882 in TP-DNAP2 encoded on pGA55-reTP-DNAP2. Due to unsuccessful attempts to generate a strain with a full deletion of p2-encoded TP-DNAP2 and simultaneous integration of *mKate*, *URA3*, and *leu2**, each library was transformed into GA-Y021 for convenient screening. However, since GA-Y021 still encodes wt TP-DNAP2 on p2-delORF1-mUL*, p2 mutation rates measured in this format are the result of in tandem replication of p2 by wt TP-DNAP2 and each TP-DNAP2 variant encoded *in trans*. We sampled each NNK library (theoretical diversity = 32 codon variants) at 6-fold coverage by picking 190 yeast colonies and passaging under selection for *URA3* to stabilize the copy number of p2-delORF1-mUL* in the presence of the newly introduced TP-DNAP2 variants. To screen for TP-DNAP2 variants with increased p2 mutation rate, each library member was subjected to a preliminary, small scale *leu2** fluctuation test with

six replicates. Seventeen candidate mutators with the highest expected number of mutants m calculated by the p0 method were then chosen for reconfirmation. Reconfirmation consisted of extracting the *CEN6/ARS4* plasmids encoding TP-DNAP2 variants, retransforming into a fresh GA-Y021 background and repeating *leu2** fluctuation tests with 36 replicates to determine p2 mutation rate with higher precision (**Table I-2**). Two error-prone TP-DNAP2 variants, S370Q and Y424Q, increased p2's substitution mutation rate by ~14- and ~13-fold, respectively (**Table I-2**, **Entry 2** and **Entry 3**). These variants were not active enough to fully complement a deletion of the native TP-DNAP2 and sustain p2-delORF2-*URA3* replication on their own, making our measured mutation rates an underestimate of their true per-base substitution rate. We suspect that even slight decreases in p2 copy number when replicated by mutant polymerases may cause instability in the system due to disruption of expression levels of the nine essential genes encoded on p2. Despite this, these two error-prone TP-DNAP2 variants elevate p2 mutation rate to high enough levels for measuring orthogonality

Domain (residues in TP-DNAP2)	Exo I (366-376)	Exo II (res. 416-430)	Pre-(S/T)Lx2h (464-481)	Motif C (874-883)
	S370 ↓	**Y424** ↓	**L474** ↓	**F882** ↓
TP-DNAP2	FDIESFSDETK	LYAWYGSGYDYQHVL	KDPYLFILTSLDKASKAF	IYSDTDSIFV
TP-DNAP1	FDIESYFDPEK	LIAHNGGGYDFHYIL	KDSYSFLLCSLANASKAF	IYSDTDSIFV
PHI29 DNAP	CDFETTT	YFHN-LKFDGAF	N/A	IYCDTDSIHL

Figure I-4. Alignment between TP-DNAP2, TP-DNAP1 and Phi29 DNAP at chosen library sites. Selected regions from the alignment of TP-DNAP2, TP-DNAP1, and Phi29 DNA polymerase are shown, which highlights the homology between conserved domains at the chosen library sites.[18] While these conserved domains from TP-DNAP1 and TP-DNAP2 show high homology, the homology to Phi29 is limited. Due to this reason, library sites were chosen based on previously discovered residues that govern TP-DNAP1 fidelity, rather than introducing mutations from the well-studied Phi29 DNAP which increase mutation rate.[6,7] N/A indicates poor alignment.

Orthogonality between p2 Replication and Genomic Replication

To show that p2 replication is orthogonal to genomic replication, we measured the mutation rate of the host genome in the presence of error-prone TP-DNAP2 variants. Like p1 replication by TP-DNAP1, p2 replication by TP-DNAP2 occurs in the cytoplasm *via* a protein-primed mechanism, making it likely that p2 replication is orthogonal to host genome replication. To test this, we transformed CEN6/ ARS4 vectors lacking TP-DNAP2, or encoding codon optimized versions of TP-DNAP2WT, TP-DNAP2S370Q, or TP-DNAP2Y424Q into AR-Y436 (**Figure I-5**). AR-Y436 contains wt p1 and p2, as well as an intact *URA3* locus in the host genome. Genomic per-base substitution rates were determined *via* fluctuation tests based on the frequency of 5-FOA resistant clones arising from mutations in the genomic *URA3* locus, as previously described,[7,19] and are reported in **Table I-4**. The substitution-rich spectrum of *URA3*-disabling mutations makes this assay ideal for detecting whether the elevated substitution rate of TP-DNAP2S370Q or TP-DNAP2Y424Q contributes to genomic mutation. We observed no increase in the host genomic mutation rate when error-prone variants of TP-DNAP2S370Q and TP-DNAP2Y424Q were present.

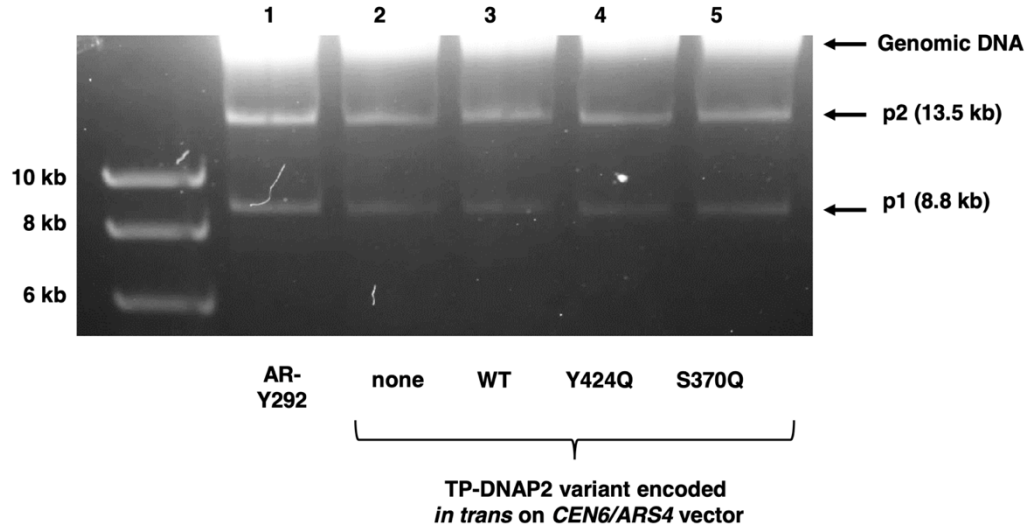


Figure I-5. Cytoplasmic plasmids from strains used to test orthogonality between TP-DNAP2 variants and genomic replication. Agarose gel (0.9%) analysis of DNA extracted after measuring genomic mutation rate with *URA3* fluctuation tests, in presence of mutagenic TP-DNAP2 variants. Lane 1: DNA extracted from parent strain AR-Y292, containing wild type p1 and p2. Lanes 2-5: DNA extracted from strains corresponding to Entries 1-4 in **Table I-4**. Replication of p1 and p2 is sustained in all strains throughout the fluctuation test procedure, ensuring that TP-DNAP2 variants had the opportunity to replicate p2 and lose their terminal proteins, which may play a role in determining specificity for p2 versus the host genome.

Table I-4. Genomic Mutation Rate Is Unaltered by Error-Prone TP-DNAP2 Variants. Entries 1−4 are per-base substitution rates of genomic replication in the presence of TP-DNAP2 variants, measured using the genomic *URA3* locus. Mutagenic TP-DNAP2 variants Y424Q and S370Q cause no statistically significant change in genomic mutation rate, relative to no TP-DNAP2 variant *in trans* (84% CI overlap method).[20] Ranges in parentheses correspond to 95% confidence intervals, determined according to the MSS method.[16–18]

entry	TP-DNAP2 *in trans*	per-base substitution rate at genomic *URA3* locus ($\times 10^{-10}$)
1	none	1.6 (0.8−2.4)
2	WT	2.2 (1.2−3.4)
3	Y424Q	1.8 (1.0−2.9)
4	S370Q	1.8 (1.0−2.8)

Orthogonality between p2 Replication and Genomic Replication

Next, we wanted to test whether the replication mechanisms of p1 and p2 are mutually orthogonal *in vivo* by measuring p1 and p2 mutation rates in the presence of a panel of TP-DNAP1 and TP-DNAP2 variants with varying mutation rates. Changes in p1 or p2 mutation rate induced by TP-DNAP variants would therefore signal a degree of cross-replication between TP-DNAP1/p1 and TP-DNAP2/p2, if any. A panel of six polymerases was introduced on *CEN6/ARS4* vectors into two separate base strains, AR-Y304 (p1 mutation rate reporter strain) and GA-Y021 (p2 mutation rate reporter strain), encoding *mKate2*, *URA3*, and *leu2** on either p1 or p2, respectively (**Figure I-6A**). Included in this panel were TP-DNAP1WT and two error-prone TP-DNAP1 variants found in previous screens: TP-DNAP1I777K,L900S and TP-DNAP1L477V,L640Y,I777K, W814N (**Figure I-7**). Also included were TP-DNAP2WT and the error-prone TP-DNAP2S370Q and TP-DNAP2Y424Q variants found in this work. In AR-Y304 and GA-Y021, both p1 and p2 still encode their native wt TP-DNAP's. Any contribution to replication by a third TP-DNAP encoded *in trans* is monitored by detecting changes in linear plasmid mutation rates. Importantly, these experiments afford each DNAP an opportunity to replicate its native plasmid and lose its attached TP, becoming "spent" and perhaps more likely to replicate its noncognate linear plasmid through DNAP exchange.[20] The presence of TP-DNAP1WT or TP-DNAP2WT *in trans* had no effect on their noncognate linear plasmids' mutation rate (**Figure 3B**, **Figure 3C**, and **Table I-5**). Mutagenic variants TP-DNAP1I777K,L900S and TP-DNAP1L477V,L640Y,I777K,W814N increased p1's mutation rate by 380- and 870-fold, respectively, but caused no statistically significant change in p2's mutation rate (**Figure I-6B**, 84% CI overlap method).[20]

Likewise, p1 mutation rate was unaltered in the presence of mutagenic TP-DNAP2S370Q and TP-DNAP2Y424Q, which increased p2 mutation rate by 29- and 16-fold, respectively (**Figure I-6C**). Thus, TP-DNAP1 replicates p1 with at least 870- fold specificity over p2, while TP-DNAP2 targets p2 with at least 29-fold specificity over p1. The level of mutual orthogonality measured here is limited by the error-rates of DNAPs used, especially that of TP-DNAP2 variants. Future discovery of more mutagenic TP-DNAP1 or TP-DNAP2 variants may prove an even greater orthogonality between these two replication systems, which we expect to be the case.

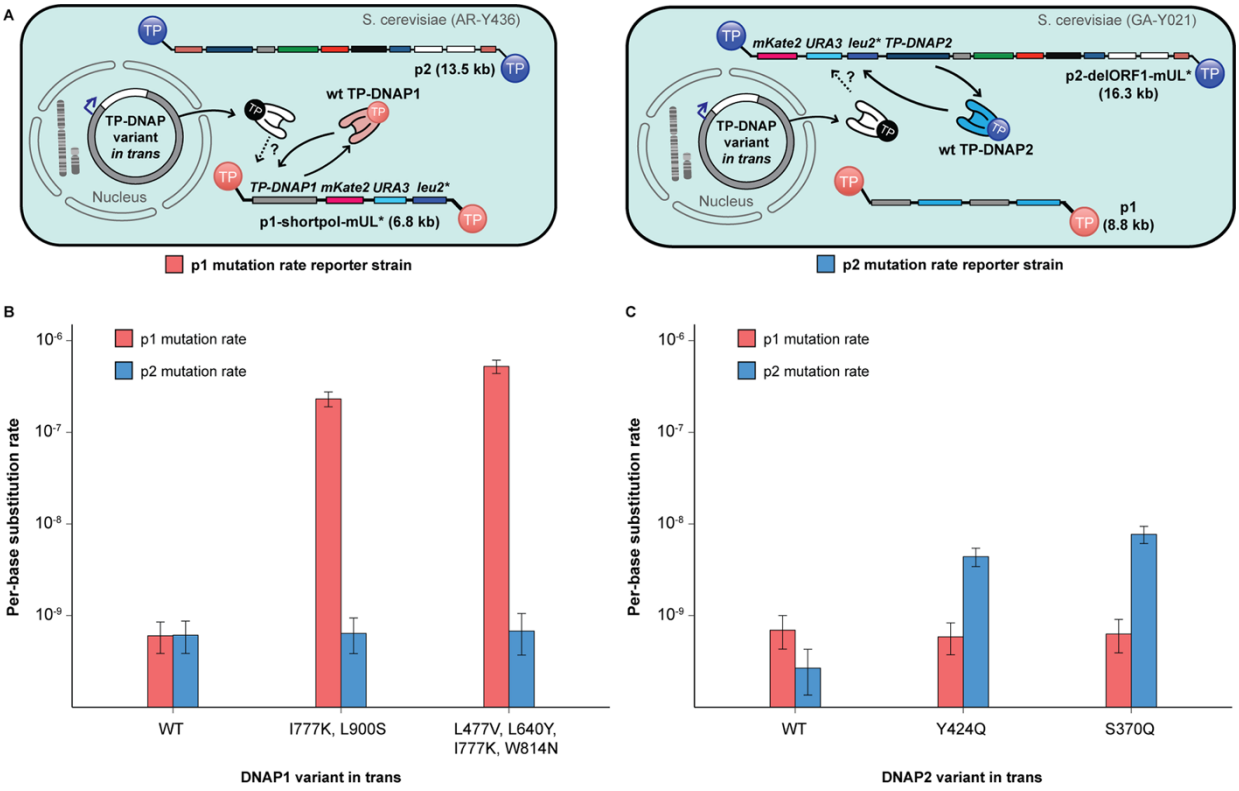


Figure I-6. Mutual orthogonality between p1 and p2 replication. A) Schematics detailing strains for measuring p1 mutation rate (left, red) and measuring p2 mutation rate (right, blue). Left: p1 mutation rate reporter strains derived from AR-Y304 encode *mKate2*, *URA3* and *leu2** on p1, enabling measurement of p1 copy number and mutation rate in the presence of TP-DNAP variants *in trans*. Right: p2 mutation rate reporter strains derived from GA-Y021 encode *mKate2*, *URA3* and *leu2** on p2, enabling measurement of p2 copy number and mutation rate in the presence of TP-DNAP variants *in trans*. **B)**

Mutagenic TP-DNAP1 variants increase p1 mutation rate (red) by 380- and 870-fold, without any increase in p2 mutation rate (blue). **C)** Mutagenic TP-DNAP2 variants increase p2 mutation rate by 16- and 29-fold, without a similar increase in p1 mutation rate. Error bars correspond to 95% confidence intervals calculated according to the MSS method.[16–18] See **Table I-5** for exact mutation rate values.

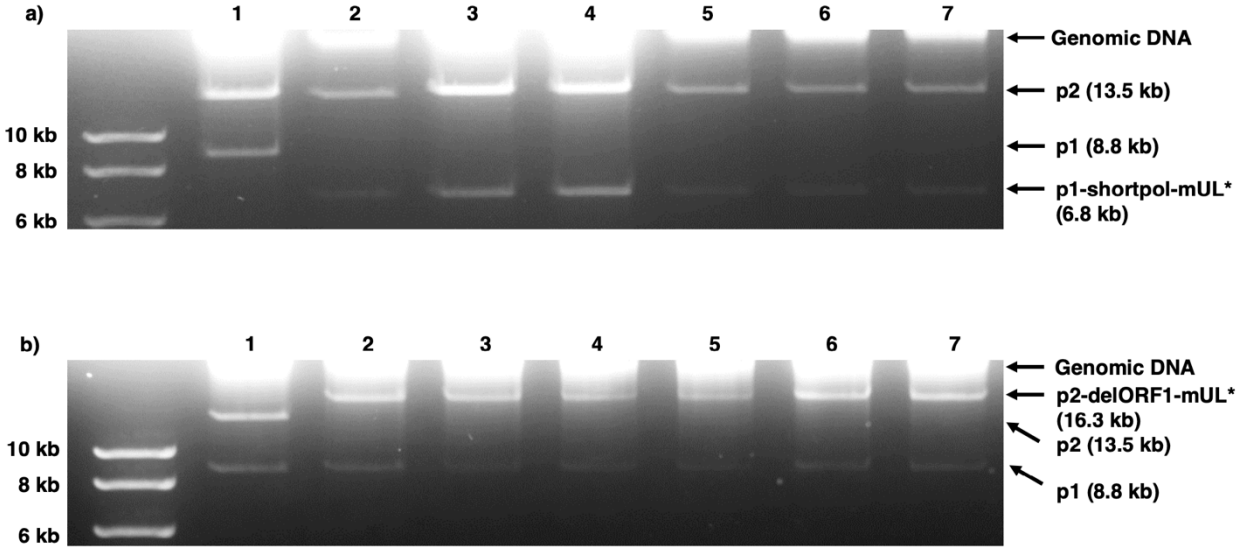


Figure I-7. Cytoplasmic plasmids from strains used to test mutual orthogonality. A) Gel analysis (0.9% agarose) of DNA extracted from p1 mutation rate reporter strains used to measure p1 mutation rate in presence of 6 TP-DNAP variants. Lane 1: parental strain AR-Y292. Lanes 2-7: In order, TP-DNAP1WT, TP-DNAP1I777K, L900S, TP-DNAP1L477V, L640Y, I777K, W814N, TP-DNAP2WT, TP-DNAP2Y424Q, TP-DNAP2S370Q. Wild type p2 is present in all strains, along with the recombinant p1-shortpol-mUL* which encodes *mKate2*, *URA3*, and *leu2**. **B)** Gel analysis (0.9% agarose) of DNA extracted from strains used to measure p2 mutation rate in presence of 6 TP-DNAP variants. These strains correspond to p2 mutation rate reporter strains described in Fig. 3a. Wild type p1 is present in all strains, along with the recombinant p2-delORF1-mUL* which encodes *mKate2*, *URA3*, and *leu2**. Lanes 2-7: In order, TP-DNAP1WT, TP-DNAP1I777K, L900S, TP-DNAP1L477V, L640Y, I777K, W814N, TP-DNAP2WT, TP-DNAP2Y424Q, TP-DNAP2S370Q.

Table I-5. Mutation rate data from mutual orthogonality experiment. The exact mutation rates represented in **Figure I-6B** and **Figure I-3C** are listed here. *leu2** fluctuation analyses were performed with 48 replicates, in Formats A and B as described in **Figure I-6A**. Values in parentheses are 95% confidence intervals calculated with the MSS method.[16–18]

Entry	Polymerase variant *in trans*	p1 s.p.b. mutation rate ($x10^{-10}$)	p2 s.p.b. mutation rate ($x10^{-10}$)
1	TP-DNAP1^{WT}	6.0 (3.9-8.5)	6.1 (3.9-8.7)
2	TP-DNAP1$^{I777K,\ L900S}$	2309 (1894-2752)	6.4 (3.8-9.4)
3	TP-DNAP1$^{L477V,\ L640Y,\ I777K,\ W814N}$	5225 (4372-6131)	6.8 (3.7-11)
4	TP-DNAP2^{WT}	6.9 (4.3-10)	2.7 (1.4-4.3)
5	TP-DNAP2^{Y424Q}	5.9 (3.7-8.3)	43.9 (34.2-54.5)
6	TP-DNAP2^{S370Q}	6.3 (3.9-9.1)	76.9 (61.1-94.2)

I.4 Conclusion

In summary, p1 replication by TP-DNAP1 and p2 replication by TP-DNAP2 are both orthogonal to genomic replication and to each other, resulting in two mutually orthogonal DNA replication systems in the same cell. We envision that this pair of orthogonal replication systems will enable the *in vivo* evolution of multiple genes at different elevated mutation rates, molecular recording of biological signals in two distinct DNA channels, and the establishment of additional mutually orthogonal DNAP/ plasmid pairs by engineering new TPs.

I.5 Materials and Methods

DNA Cloning

Plasmids cloned and used in this study are listed in **Table I-6**. All oligonucleotide primers and synthesized gene fragments (gBlocks) were purchased from IDT. Enzymes for PCR and cloning were obtained from NEB. All plasmids, including TP-DNAP2 libraries were cloned using Gibson assembly with overlap regions of 20−30 bp.[21] Vectors

harboring homologous recombination cassettes for p2 integrations were cloned as previously described for p1 integration cassettes.[6,7]

To clone pGA55, three gene fragments constituting recoded TP-DNAP2 were assembled with the vector backbone of pAR318, a yeast shuttle vector containing *CEN6/ARS4* and *HIS3* for propagation in yeast, and *ColE1* and *KanR* for propagation in *E. coli*. The resulting vector was used for TP-DNAP2 complementation and generation of TP-DNAP2 libraries (see TP-DNAP2 Library Generation section).

Table I-6. List of plasmids used in this study. Plasmids used and cloned in this work are listed here. Plasmids encoding individual TP-DNAP2 variants derived from pGA55 (other than Y424Q and S370Q) are not individually listed here.

#	Name	Source	Parent Plasmid	Origin of replication (yeast, bacterial)	Selection markers (yeast, bacterial)	Notes
1	pAR505	Previous work[7]	See previous work	*N/A, ColE1*	*URA3 (p1), AmpR*	p1 recombination cassette that integrates *mKate2, URA3,* and *leu2* (538 T>C)* in place of ORFs 2, 3, and 4 of p1
2	pGA1	This work	pAR505	*N/A, ColE1*	*URA3 (p2), AmpR*	p2 recombination cassette that integrates *mKate2, URA3,* and *leu2* (538 T>C)* ove
3	pGA59	This work	pGA1	*N/A, ColE1*	*URA3 (p2), AmpR*	p2 recombination cassette that integrates *URA3* over *ORF2* of p2
4	pGA109	This work	pGA1	*N/A, ColE1*	*URA3 (p2), AmpR*	p2 recombination cassette that integrates *URA3* and *leu2* (538 T>C)* over *ORF2* of p2
5	pAJ137	Unpublished work	Addgene (#60847)	*2μ*, *ColE1*	*KanMX, KanR*	*tRNA TYR* promoter > *HDV* ribozyme + sgRNA> *SNR52* terminator and *RNR2* promoter > *SpCAS9*> *CYC1* terminator
6	pGA3	This work	pAJ137	*2μ*, *ColE1*	*KanMX, KanR*	*tRNA TYR* promoter > *HDV* ribozyme + *p2-ORF1-sgRNA1* > *SNR52* terminator and *RNR2* promoter > *SpCAS9*> *CYC1* terminator

7	pGA4	This work	pAJ137	*2μ*, *ColE1*	*KanMX, KanR*	*tRNA TYR* promoter > *HDV* ribozyme + p2-ORF1-sgRNA2 > *SNR52* terminator and *RNR2* promoter > *SpCAS9*> *CYC1* terminator
8	pGA5	This work	pAJ137	*2μ*, *ColE1*	*KanMX, KanR*	*tRNA TYR* promoter > *HDV* ribozyme + p2-ORF1-sgRNA3 > *SNR52* terminator and *RNR2* promoter > *SpCAS9* > *CYC1 terminator*
9	pGA72	This work	pAJ137	*2μ*, *ColE1*	*KanMX, KanR*	*tRNA TYR* promoter > *HDV* ribozyme + p2-ORF2-sgRNA1 > *SNR52* terminator and *RNR2* promoter > *SpCAS9*> *CYC1* terminator
10	pGA73	This work	pAJ137	*2μ*, *ColE1*	*KanMX, KanR*	*RNA TYR* promoter > *HDV* ribozyme + p2-ORF2-sgRNA2 > *SNR52* terminator and *RNR2* promoter > *SpCAS9*> *CYC1* terminator
11	pGA74	This work	pAJ137	*2μ*, *ColE1*	*KanMX, KanR*	*RNA TYR* promoter > *HDV* ribozyme + p2-ORF2-sgRNA3 > *SNR52* terminator and *RNR2* promoter > *SpCAS9*> *CYC1* terminator
12	pAR317	Previous work[7]	See previous work	*CEN6/ARS4, ColE1*	*URA3, KanR*	Shuttle vector
13	pAR318	Previous work[7]	See previous work	*CEN6/ARS4, ColE1*	*URA3, KanR*	REV1 promoter > recoded wildtype TP-DNAP1
14	pAR611	Previous work[6]	See previous work	*CEN6/ARS4, ColE1*	*URA3, KanR*	REV1 promoter > recoded TP-DNAP1I777K, L900S
15	pAR633	Previous work[6]	See previous work	*CEN6/ARS4, ColE1*	*URA3, KanR*	REV1 promoter > TP-DNAP1L477V, L640Y, I777K, W814N
16	pGA55	This work	pAR318	*CEN6/ARS4, ColE1*	*URA3, KanR*	REV1 promoter > recoded wildtype TP-DNAP2. All other TP-DNAP2 variants were derived from this plasmid
17	pGA84	This work	pGA55	*CEN6/ARS4, ColE1*	*URA3, KanR*	REV1 promoter > recoded TP-DNAP2 Y424Q
18	pGA87	This work	pGA55	*CEN6/ARS4, ColE1*	*URA3, KanR*	REV1 promoter > recoded TP-DNAP2 S370Q

Yeast Strains

All yeast strains used in this work are listed in **Table I-7**. *S. cerevisiae* strain AR-Y292 served as the parent for all strains used in this study and contains the wild type pGKL1 and pGKL2 (or p1 and p2) linear plasmids. GA-Y021 and GA-Y069 were created from AR-Y292 by p2 integration methods described below. AR-Y436 is a derivative of AR-Y292 encoding a functional copy of *URA3* at the endogenous genomic locus, for 5-FOA-based fluctuation tests of genomic mutation rates in the presence of mutagenic TP-DNAP2 variants.

Strains for testing mutual orthogonality were generated by transforming a panel of *CEN6/ARS4* vectors encoding TP-DNAP1 or TP-DNAP2 variants into two base strains, AR-Y304 and GA-Y021. AR-Y304 (described previously[7]) contains recombinant p1 encoding *mKate2*, *URA3* and *leu2** without disturbing the native TP-DNAP1 ORF. Similarly, GA-Y021 encodes recombinant p2 that replaces ORF1 with *mKate2*, *URA3* and *leu2** without disturbing the native TP-DNAP2 ORF.

Table I-7. List of yeast strains used in this study. All yeast strains used and created in this work are listed here. Not included are derivatives of GA-Y021 and AR-Y304 containing TP-DNAP2 or TP-DNAP1 variants used for mutual orthogonality experiments, as well as derivatives of GA-Y069 containing combinations of CRISPR/Cas9 and TP-DNAP2 used to show TP-DNAP2 complementation.

Strain	Genotype	Parent Strain	Source	Notes
AR-Y292	*MAT***a** *can1 URA3 his3 leu2* TRP1 ρ0 + p1 + p2	F102-2 (ATCC (Catalog #200585)	Previous work	F102-2 derived strain with full deletions of genomic *URA3*, *URA3*, *URA3* loci
AR-Y436	*MAT***a** *can1 URA3 his3 leu2 trp1* ρ0 + p1 + p2	F102-2 (ATCC (Catalog #200585)	Previous work	F102-2 derived strain with a genomic *URA3* locus amenable for fluctuation analyses
GA-Y021	*MAT***a** *can1 URA3 his3 leu2* TRP1 ρ0 + p1 + p2-delORF1-*mKate2-URA3-leu2**	AR-Y292	This work	F102-2 derived strain with *mKate2-URA3-leu2** integrated over ORF1 of p2, and with p2 fully removed by CRISPR/Cas9
GA-Y069	*MAT***a** *can1 URA3 his3 leu2* TRP1 ρ0 + p1 +	AR-Y292	This work	F102-2 derived strain with a metastable mixture of p2 and p2 with *ORF2* deleted with *URA3*

	p2 + p2-delORF2-*URA3*			
GA-Y256	*MATa can1 URA3 his3 leu2 trp1 ρ0 + p1 + p2 + GA-Ec32*	AR-Y436	This work	F102-2 derived strain with genomic *URA3* copy, and TP-DNAP2 variants *in trans* for
GA-Y257	*MATa can1 URA3 his3 leu2 trp1 ρ0 + p1 + p2 + GA-Ec55*	AR-Y436	This work	F102-2 derived strain with genomic *URA3* copy, and TP-DNAP2 variants *in trans* for
GA-Y258	*MATa can1 URA3 his3 leu2 trp1 ρ0 + p1 + p2 + GA-Ec84*	AR-Y436	This work	F102-2 derived strain with genomic *URA3*copy, and TP-DNAP2 variants *in trans* for
GA-Y259	*MATa can1 URA3 his3 leu2 trp1 ρ0 + p1 + p2 + GA-Ec87*	AR-Y436	This work	F102-2 derived strain with genomic *URA3* copy, and TP-DNAP2 variants *in trans* for
GA-Y262	*MAT***a** *can1 URA3 his3 leu2* TRP1 ρ0 + p1 + p2 + p2-delORF2-URA3 + GA-Ec55	AR-Y292	This work	F102-2 derived strain with p1, p2, p2-delORF2-*URA3* and recoded wild type TP-DNAP2 *in trans*
GA-Y263	*MAT***a** *can1 URA3 his3 leu2* TRP1 ρ0 + p1 + p2 + p2-delORF2-URA3 + GA-Ec55 + GA-Ec72	AR-Y292	This work	F102-2 derived strain with p1, p2, p2-delORF2-*URA3* and recoded wild type TP-DNAP2 *in trans,* with a vector encoding Cas9 and p2ORF2-sgRNA1
AR-Y304	*MAT***a** *can1 ura3 his3 leu2* TRP1 ρ0 + p1 + p2 + p1-shortpol-*mKate2-URA3-leu2**	AR-Y292	Previous work	Previously made F102-2 derivative with *mKate2-URA3-leu2** integrated over ORFs 2, 3, and 4 of p1, while keeping the TP-DNAP1 encoded on ORF1

Yeast Transformation

All yeast transformations were performed using the high-efficiency LiAc/SS carrier DNA/PEG method.[22] For integrations into p2, 2−5 μg of plasmid containing the appropriate integration cassette was digested with ScaI, yielding a linearized cassette with blunt ends containing the genes of interest flanked by regions of homology to p2. The products of the digestion reaction were directly transformed into appropriate AR-Y292-derived strains harboring wild type p1 and p2, and plated on selective solid SC medium. Colonies appeared after 4−5 d of growth at 30 °C. *CEN6/ARS4* plasmids were also transformed with the LiAc/SS carrier DNA/ PEG method, but with only 500−3000 ng of DNA for individual vectors and with at least 10 μg of plasmid DNA for TP-DNAP2 library transformations, to maintain 6-fold coverage.

Linear Plasmid DNA Extraction

p1, p2, and all derived linear plasmids were extracted using a modified version of the yeast DNA extraction protocol detailed by Amberg et al.[23] The modifications were as follows: (i) cells spun down from 40 mL of saturated culture were washed in 0.9% NaCl before treatment with Zymolyase (US Biological) to break up flocculated cells; (ii) 200 µg/mL proteinase K (Sigma) was supplemented during SDS treatment for degradation of TP; (iii) rotation at ~10 rpm was used during Zymolyase and proteinase K treatments. This large-scale extraction protocol was used for preparing DNA for absolute quantification by qPCR. For qualitative analysis by agarose gel electrophoresis, this extraction protocol was scaled down to extract DNA from only 1.5 mL of saturated yeast culture.

Curing of Cytoplasmic Plasmids Using Cas9

To achieve active curing of the cytoplasmic p2 plasmid, the yeast Cas9 genomic modification system developed by Cate and co-workers was repurposed for cytoplasmic targeting.[24] The SV-40 nuclear localization signal and 8× HIS tag were removed from the pCAS plasmid (Addgene plasmid # 60847) to localize Cas9 to the cytoplasm where p2 (and p1) plasmids propagate. Appropriate 20 nt spacers were cloned into this vector to target different sites in p2 (**Table I-1**). These modified pCAS vectors were transformed into the strains harboring p2 plasmids to be cut, and plated on solid selective SC medium containing 1 g/L monosodium glutamate (MSG) as the nitrogen source and supplemented with G418 (400 µg/mL). Colonies that appeared after incubation at 30 °C for 2 d were

inoculated into liquid selective SC medium with 1 g/L MSG and G418 (200 μg/mL) and passaged once at a 1:1000 dilution to cure the targeted p2 plasmid. The resulting cultures were then subjected to DNA extraction and analysis by gel electrophoresis to verify loss of the targeted p2 plasmid. To minimize potential toxicity due to Cas9 expression, final strains lacking the pCAS vector were isolated by passaging without G418 selection, and replica plating clones on solid medium with and without G418 to screen for loss of the pCAS vector.

TP-DNAP2 Library Generation

TP-DNAP1 and TP-DNAP2 peptide sequences were aligned using protein BLAST and four candidates residues for library generation were chosen by two criteria.[25] First, candidate TP-DNAP2 residues must match a residue in TP-DNAP1 known to affect fidelity, based on prior studies.[6,7] Second, at least 25% of the 20 neighboring residues must align. This analysis yielded positions S370, Y424, L474 and F882.

To clone the expression vector for wild type TP-DNAP2 (pGA55), TP-DNAP2 was codon optimized for expression in *S. cerevisiae* with GenScript's OptimumGene tool, and the recoded ORF was synthesized as three gene fragments, which were assembled downstream of the REV1 promoter in a CEN6/ ARS4 vector containing selection markers *HIS3* and *KanR*. The four TP-DNAP2 NNK libraries were cloned from pGA55 *via* Gibson assembly. Gibson fragments containing the NNK codon at each library site were prepared with two sequential PCR's, in order to limit bias in the NNK incorporation that may result from annealing between the degenerate codon and the plasmid template. The first PCR generated a linear fragment from pGA55 with a 5′ end that terminates immediately 3′ of

the library codons, and a 3′ end in *KanR*. These linear amplicons were purified, diluted to 40 ng/μL, and reamplified in a second PCR reaction with a forward primer containing Gibson overlap regions and NNK overhangs at the corresponding library site in TP-DNAP2, and the same reverse primer used in the initial PCR. These PCR products were then purified and treated with DpnI for 6 h at 37 °C to digest any pGA55 plasmid carry-through. The second Gibson fragment was PCR amplified from pGA55 to include the vector backbone starting 5′ in *KanR* and 3′ leading up to, but not including the library codon. For each library, 100 ng of corresponding PCR amplicons were combined in a 20 μL Gibson assembly reaction, and incubated at 50 °C for 1 h. The assemblies were purified and concentrated in 12 μL of ddH2O. Five μL of the purified assembly products were then transformed into electrocompetent Top10 cells and recovered at 37 °C for 1 h. Each transformation was plated at 1×, 10× and 100× dilutions on solid LB medium supplemented with kanamycin (50 μg/mL). After overnight incubation at 37 °C, colony counts on the 10× an 100x plates were used to calculate the transformation efficiency. All transformations yielded more than 3200 transformants, corresponding to >100-fold coverage of each library in *E. coli* (each NNK library has a theoretical diversity of 32). The transformants from the 1× plates were then harvested by resuspension in 5 mL of sterile ddH2O. Library plasmid DNA was extracted from *E. coli* using the Zyppy Plasmid Miniprep Kit. Library quality was determined by verifying plasmid library sizes by agarose gel and Sanger sequencing of bulk library populations as well as 8 individual clones from each library.

p2 Fluctuation Tests Using leu2*

For fluctuation tests used to measure p2 mutation rates, *leu2** served as a marker for detecting mutational events. *leu2** is a disabled version of *LEU2*, where Q180 is replaced with a TAA stop codon. Q180 is a permissive site where mutation to any other codon other than TAG and TGA results in functional reversion to *LEU2*. These mutational events can be detected by plating scores of parallel cultures on medium lacking leucine and counting the number distribution of functional *LEU2* mutants.

For TP-DNAP2 library screening, each library member was subjected to small-scale *leu2** fluctuation tests with six replicates. 190 library members from each yeast library transformation were arrayed and inoculated into liquid SC medium lacking uracil and histidine, and passaged three times at 1:10 000 dilutions. *mKate2* fluorescence was measured at every passage on a microplate reader (TECAN Infinite 200 PRO, settings: λex = 588, λem = 633) to track p2 copy number stabilization. To perform fluctuation tests, each library member was diluted 1:10 000 into liquid SC medium buffered to pH 5.8 and lacking uracil, histidine, tryptophan, and dilutions were split into six 100 µL replicates. Cultures were grown for 48 h at 30 °C to reach saturation. Saturated cultures were washed with 200 µL 0.9% NaCl to remove residual leucine and resuspended in 35 µL of 0.9% NaCl. Ten µL of this resuspension was spot plated onto solid SC medium buffered to pH 5.8 and lacking uracil, histidine, tryptophan and leucine. Spot plates were allowed to dry and incubated at 30 °C. Colonies were counted after 5 d. Colony counts were used to calculate the expected number of *LEU2* functional mutants (m), using the p0 method.[16]

To precisely measure p1 and p2 per-base substitution mutation rates, several modifications were made to the small-scale protocol. First, 36 replicates were performed for reconfirmation of candidate TP-DNAP2 mutators, and 48 replicates for genomic and

mutual orthogonality experiments. Titers were also determined for each strain after spot plating by pooling the residual volume from 4 replicates and plating dilutions on YPD. The expected number of *LEU2* functional mutants (m) was determined by the Ma-Sandri-Sarkar maximum likelihood estimator (as calculated by the FALCOR tool and corrected for partial plating.[16–18] The mean *mKate2* fluorescence was determined from 50,000 event counts on Attune NxT Flow Cytometer and converted to p2 copy number by using a calibration curve (see below). To determine per-base substitution rates, the corrected m was normalized to the average cell titer, the p2 copy number, and the target size for functional *leu2** reversion (2.33 bp). 95% confidence intervals were similarly scaled.

Genomic Fluctuation Tests Using URA3

Fluctuation tests using genomically encoded *URA3* were performed in the presence of TP-DNAP2 variants to determine the genomic per-base substitution rates, similarly to previously described protocols.[19] AR-Y292 derived strains harboring the appropriate TP-DNAP2 variant encoded on a *CEN6/ARS4* vector (with a *HIS3* marker) were grown in liquid SC medium lacking uracil and histidine until saturation. Each strain was diluted 1:5000 into SC medium lacking histidine and aliquoted into 48 replicates of 200 µL each. Cultures were grown for 48 h at 30 °C to reach saturation. Saturated cultures were washed with 400 µL 0.9% NaCl and resuspended in 420 µL 0.9% NaCl. 400 µL from each replicate was spot plated on predried solid SC medium lacking histidine and supplemented with 5-FOA (1 g/L). The residual 20 µL from six replicates were pooled, diluted, and plated on solid YPD medium to determine cell titers. Plates were allowed to dry before incubation at 30 °C. Colonies were counted on titer plates after 2 days, and on

spot plates after 5 days of growth. The expected number of mutants (m) was calculated using the MSS maximum likelihood estimator method *via* the FALCOR tool, and corrected for partial plating, as described above. To determine per-base substitution rates, the corrected m was normalized to the average cell titer, the *URA3* copy number (1 in haploid yeast), and the target size for 5-FOA resistance *via* substitutions in *URA3* (104 bp). 95% confidence intervals were similarly scaled.

Calibration Curve Relating p2 Copy Number to mKate2 *Fluorescence*

A standard curve relating p2 copy number to *mKate2* fluorescence was prepared by combining quantitative PCR with flow cytometry. During the 1:10 000 back dilution step of the *leu2** fluctuation tests for the mutual orthogonality experiment, six strains with *mKate2* encoded on p2 were diluted into liquid SC medium buffered to pH 5.8 and lacking uracil, histidine and tryptophan to yield 50 mL of saturated culture. After 48 h of growth at 30 °C, a small portion of each culture was diluted 1:100 in 0.9% NaCl and analyzed on a flow cytometer (Attune NxT Flow Cytometer, settings: λex = 561, λem = 620; gain = 550) to determine the mean red fluorescence from 50 000 counts. Genomic DNA and linear plasmids were extracted from the remaining 40 mL of each culture using the large-scale DNA extraction protocol detailed in Linear Plasmid DNA Extraction to ensure complete and unbiased extraction of linear plasmids relative to genomic DNA. All extracts were diluted 4000-fold for use in two distinct qPCR reactions, one to quantify p2-encoded *leu2** and the other to quantify the genomic copy of LEU3 (see **Table I-8** for primers and **Table I-9** for qPCR cycling parameters). Each 20 µL qPCR reaction consisted of 5 µL of

template DNA, 2 µL forward primer (5 µM), 2 µL reverse primer (5 µM), 1 µL ddH2O, and 10 µL of Thermo Scientific Maxima SYBR Green/Fluorescein qPCR Master Mix (2×).

A standard curve for each primer set was prepared by performing qPCR on a dilution series of DNA extracted from F102−2 (25×, 125×, 625×, 3125×). Nontemplate controls with only ddH2O were included for each primer set to detect contamination. All qPCR's were performed in triplicate on the Roche LightCycler 480 System using the following protocol.

qPCR

Table I-8. Primers used for qPCR.

Entry	Sequence (5' to 3')	Purpose
1	GCTAATGTTTTGGCCTCTTC	Forward primer used for *LEU2* qPCR
2	ATTTAGGTGGGTTGGGTTCT	Reverse primer used for *LEU2* qPCR
3	CAGCAACTAAGGACAAGG	Forward primer used for *LEU3* qPCR
4	GGTCGTTAATGAGCTTCC	Reverse primer used for *LEU3* qPCR

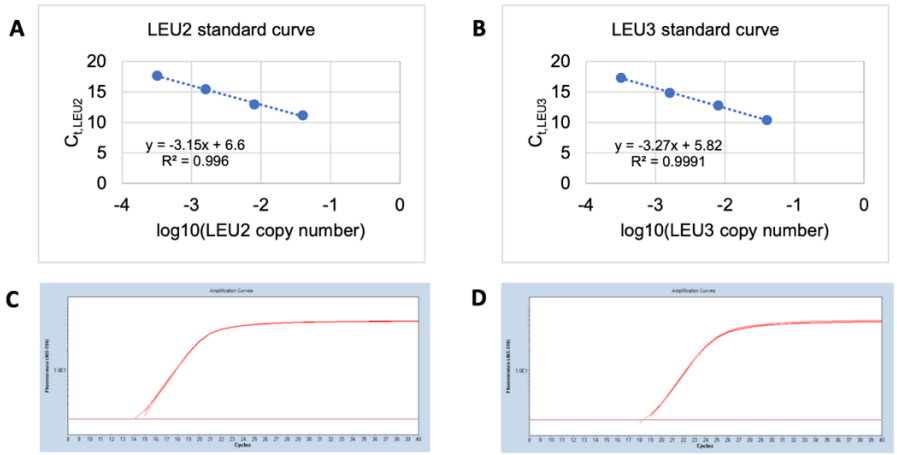
Table I-9. qPCR cycling parameters.

Step	Temperature (°C)	Time	Notes	Cycles
1	95	10 minutes		1
2.1 2.2 2.3	95 60	15 seconds 1 minute	 Measurement	40
3	95	1 minute		1
4	55	1 minute		1

Primer Melting Curve

Ramp up to 95 °C at 0.11 °C/s, with 5 measurements per °C. Cycle threshold (Ct) values were determined by the LightCycler 480 software (fit-points method, threshold = 1.75). Ct values from both standard curves were plotted against log ([DNA]). The slope and y-intercept were calculated using linear regression (**Table I-3**). Each sample's

average Ct values were converted into copy number values by using the following equation: copy number = 10((sample Ct−y-intercept)/slope). The calculated *leu2** copy number was divided by the LEU3 copy number to normalize to genomes extracted and account for variance in DNA extraction efficiency across samples. For detailed information on qPCR data analysis, see **Figure I-8**.



E

	LEU2 C_t	LEU3 C_t
Replicate 1	13.74	17.96
Replicate 2	13.45	18.19
Replicate 3	13.14	17.85
Mean	13.44	18.00

LEU2 copy number calculation:	**LEU3 copy number calculation:**
y = -3.15x + 6.6; where y = $C_{t,LEU2}$; x = log10(LEU2 copy number, 4000x diluted) 13.44 = -3.15x + 6.6 x = -2.17 LEU2 copy number (4000x diluted) = 10^x = $10^{-2.17}$ = 0.0068 LEU2 copy number (1x) = 4000*0.0068 – 1 genomic copy of LEU2 = **26.2**	y = -3.27x + 5.82; where y = $C_{t,LEU3}$; x = log10(LEU3 copy number, 4000x diluted) 18.00 = -3.27x + 5.82 x = -3.72 LEU3 copy number (4000x diluted) = 10^x = $10^{-3.72}$ = 0.000191 LEU3 copy number (1x) = 4000*0.000195 = **0.76**
Normalized LEU2 copy number = LEU2 copy number/LEU3 copy number = 26.2/0.76 = **34.5 copies**	

Figure I-8. qPCR data analysis. Standard curves prepared by performing triplicate qPCR reactions with primers specific to *LEU2* (A) or LEU3 (B), where the template DNA is a dilution series of DNA extracted from AR-Y003. AR-Y003 contains a single genomic copy of *LEU2* and LEU3, allowing a standard curve of Ct vs. copy number to be constructed. Triplicate qPCR reactions performed on a sample strain, with primers specific to *LEU2* **C)** or LEU3 **D)**, with Ct threshold shown at Fluorescence (465-510) = 1.75. **E)** Each sample's mean Ct,*LEU2* and Ct,LEU3 value is first converted to copy numbers, respectively, using the linear fits from standard curves shown in C and D. *LEU2* copy number is then normalized by LEU3 copy number to correct for potential differences in DNA extraction efficiency between samples.

I.6 References

(1) Cadwell, R. C.; Joyce, G. F. Randomization of Genes by PCR Mutagenesis. *Genome Res.* **1992**, *2*, 28–33.

(2) Metzker, M. L. Sequencing Technologies the next Generation. *Nat. Rev. Genet.* **2010**, *11*, 31–46.

(3) Kranaster, R.; Marx, A. Engineered DNA Polymerases in Biotechnology. *ChemBioChem* **2010**, *11*, 2077–2084.

(4) Taylor, A. I.; Pinheiro, V. B.; Smola, M. J.; Morgunov, A. S.; Peak-Chew, S.; Cozens, C.; Weeks, K. M.; Herdewijn, P.; Holliger, P. Catalysts from Synthetic Genetic Polymers Europe PMC Funders Group. *Nature* **2015**, *518*, 427–430.

(5) Pinheiro, V. B.; Taylor, A. I.; Cozens, C.; Abramov, M.; Renders, M.; Zhang, S.; Chaput, J. C.; Wengel, J.; Peak-Chew, S.-Y.; McLaughlin, S. H.; Herdewijn, P.; Holliger, P. Synthetic Genetic Polymers Capable of Heredity and Evolution. *Science* **2012**, *336*, 341–344.

(6) Ravikumar, A.; Arzumanyan, G. A.; Obadi, M. K. A.; Javanpour, A. A.; Liu, C. C. Scalable, Continuous Evolution of Genes at Mutation Rates above Genomic Error Thresholds. *Cell* **2018**, *175*, 1946–1957.

(7) Ravikumar, A.; Arrieta, A.; Liu, C. C. An Orthogonal DNA Replication System in Yeast. *Nat. Chem. Biol.* **2014**, *10*, 175–177.

(8) Perli, S. D.; Cui, C. H.; Lu, T. K. Continuous Genetic Recording with Self-Targeting CRISPR-Cas in Human Cells. *Science* **2016**, *353*, 1115–1125.

(9) Malyshev, D. A.; Dhami, K.; Lavergne, T.; Chen, T.; Dai, N.; Foster, J. M.; Correâ, I. R.; Romesberg, F. E. A Semi-Synthetic Organism with an Expanded Genetic Alphabet. *Nature* **2014**, *509*, 385–388.

(10) Ravikumar, A.; Liu, C. C. Biocontainment through Reengineered Genetic Codes. *ChemBioChem* **2015**, *16*, 1149–1151.

(11) Herr, A. J.; Ogawa, M.; Lawrence, N. A.; Williams, L. N.; Eggington, J. M.; Singh, M.; Smith, R. A.; Preston, B. D. Mutator Suppression and Escape from Replication Error-Induced Extinction in Yeast. *PLoS Genet.* **2011**, *7*, 1–16.

(12) Stark, M. J. R.; Boyd, A.; Mileham, A. J.; Romanos, M. A. The Plasmid-encoded Killer System of Kluyveromyces Lactis: A Review. *Yeast* **1990**, *6*, 1–29.

(13) Rodríguez, I.; Lázaro, J. M.; Salas, M.; De Vega, M. Φ29 DNA Polymerase-Terminal Protein Interaction. Involvement of Residues Specifically Conserved

among Protein-Primed DNA Polymerases. *J. Mol. Biol.* **2004**, *337*, 829–841.

(14) Mysiak, M. E.; Holthuizen, P. E.; van der Vliet, P. The Adenovirus Priming Protein PTP Contributes to the Kinetics of Initiation of DNA Replication. *Nucleic Acids Res.* **2004**, *32*, 3913–3920.

(15) Schaffrath, R.; Stark, M. J. R.; Gunge, N.; Meinhardt, F. Kluyveromyces Lactis Killer System: ORF1 of PGKL2 Has No Function in Immunity Expression and Is Dispensable for Killer Plasmid Replication and Maintenance. *Curr. Genet.* **1992**, *21*, 357–363.

(16) Foster, P. L. Methods for Determining Spontaneous Mutation Rates Patricia. In *Methods in Enzymology*; Academic Press, 2006; Vol. 409, pp 195–213.

(17) Sarkar, S.; Ma, W. T.; Sandri, G. V. H. On Fluctuation Analysis: A New, Simple, and Efficient Method for Computing the Expected Number of Mutants. *Genetica* **1992**, *85*, 173–179.

(18) Hall, B. M.; Ma, C. X.; Liang, P.; Singh, K. K. Fluctuation AnaLysis Calculator: A Web Tool for the Determination of Mutation Rate Using Luria-Delbück Fluctuation Analysis. *Bioinformatics* **2009**, *25*, 1564–1565.

(19) Lang, G. I.; Murray, A. W. Estimating the Per-Base-Pair Mutation Rate in the Yeast Saccharomyces Cerevisiae. *Genetics* **2008**, *178*, 67–82.

(20) Zheng, Q. Comparing Mutation Rates under the Luria–Delbrück Protocol. *Genetica* **2016**, *144*, 351–359.

(21) Gibson, D. G. Enzymatic Assembly of Overlapping DNA Fragments. *Methods Enzymol.* **2011**, *498*, 349–361.

(22) Gietz, R. D.; Schiestl, R. H. High-Efficiency Yeast Transformation Using the LiAc/SS Carrier DNA/PEG Method. *Nat. Protoc.* **2007**, *2*, 31–34.

(23) Amberg, D. C. *Methods in Yeast Genetics: A Cold Spring Harbor Laboratory Course Manual*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2005.

(24) Ryan, O. W.; Skerker, J. M.; Maurer, M. J.; Li, X.; Tsai, J. C.; Poddar, S.; Lee, M. E.; DeLoache, W.; Dueber, J. E.; Arkin, A. P.; Cate, J. H. D. Selection of Chromosomal DNA Libraries Using a Multiplex CRISPR System. *Elife* **2014**, *3*, 1–15.

(25) Altschul, S. F.; Madden, T. L.; Schäffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. Gapped BLAST and PSI- BLAST: A New Generation of Protein Database Search Programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402.