Expanding Oligonucleotide Properties through Chemical Modification: from Gene Editing and Abasic Detection to Self-Replication

by

Hansol Park

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Chemistry
University of Alberta

Abstract

DNA stores the genetic information of human beings and all living organisms. Research on oligonucleotides can lead us to find the answer to the origin of life. Self-replication of nucleic acids in the absence of enzymes such as polymerase or ligase represents an important step in the origin of life. Yet, product inhibition remains a major challenge despite efforts to develop strategies for the nonenzymatic self-replication of nucleic acids. We propose that studying a successful example of enzymatic DNA self-replication based on a simple ligation chain reaction could shed light on critical steps in self-replication. Previously our group showed an isothermal self-replication system called lesion-induced DNA amplification (LIDA). It was proposed that the destabilizing abasic group played an important role in reducing product inhibition. This thesis focuses on determining all of the factors that allow LIDA to overcome product inhibition. Thermodynamic studies were performed to measure the binding affinity of DNA complexes in the absence of enzymes, and a kinetic model was used to determine the corresponding values in the presence of T4 DNA ligase. This study suggests that T4 ligase helps overcome product inhibition by reducing the binding affinity difference between the product duplex and the intermediate complex, implying that catalysts or enzyme-like substances that stabilize the intermediate DNA complex might be a route to nonenzymatic replication.

Unnatural nucleobases have been designed to extend the role of natural bases in templated synthesis and molecular recognition. Some laboratory-designed bases that imitate the hydrogen bonding of the natural base pairs often contain pyrimidines or purine derivatives. In contrast, other unnatural base pairs rely on the geometric fit and packing force. Prior work had shown a hydrophobic pyrene nucleotide forms a strong base pair with an abasic site. As abasic lesions are the most common types of DNA damage, pyrene nucleotides are a promising candidate to detect

abasic sites using their high stability. Remarkable selectivity was observed in the ligation of 5'-phosphate-pyrene nucleotides across from an abasic lesion in a templated ligation reaction suggesting that pyrene-terminated probes could be used in abasic detection when followed by amplification of the ligated product. This result also suggested that other DNA adducts could be detected by the incorporation of DNA glycosylase in a prior step, which removes specific DNA adducts to form an abasic lesion, broadening pyrene nucleotides' application in DNA lesion or adduct detection.

The CRISPR-Cas9 system is a powerful gene-editing tool that allows researchers to alter DNA sequences and modify gene function, the identification of which was just recognized with the 2020 Nobel Prize in Chemistry. The CRISPR-Cas9 system requires Cas9 protein, crRNA, the 20 nucleotides (nt) programmable target sequence, and tracrRNA, which binds the Cas9 nuclease and complexes the crRNA through partial hybridization. The crRNA and tracrRNA sequences can also be fused into one strand resulting in a single guide RNA (sgRNA). For some applications targeting multiple genes is required to study complex biopathways and avoid the difficulty in deconvoluting overlapping and redundant roles of proteins. Yet, the synthesis of sgRNA in solid-phase synthesis and the modification of sgRNA remain challenging. Here I describe a strategy to synthesize multiple fragments that together constitute the sgRNA and connect these pieces with bioconjugation: specifically, by copper-catalyzed azide-alkyne cycloaddition (CuAAC), the most common "click" reaction. This modular bioconjugation method enables multi-color labeling with different fluorophores as well as different chemical modifications. This modular synthesis was successfully used to edit two genes in cells combined with fluorescence-activated cell sorting.

Comprehensive studies on oligonucleotides serve to further the knowledge and understanding of the oligonucleotides and their applications. As oligonucleotide plays a key role

in biotechnologies, the observation and methodologies within this thesis may prove insightful for understanding the self-replication of the oligonucleotide, developing probes for DNA damage detections, and enabling the preparation of highly modified single guide RNA for CRISPR-Cas9-based biotechnologies.

Preface

This thesis is an original work by Hansol Park. Some of the research conducted for this thesis was a collaboration with other group members in the Gibbs group (University of Alberta, AB), the Macauley group (University of Alberta, AB), and the Hubbard group (the University of Alberta, AB and the University of Toronto, ON).

Chapter 2 of the thesis is under preparation as Park, H.; Parshotam, S.; Hales, S.C.; Gibbs, J.M. Minimizing Product Inhibition in DNA Self-Replication. I was responsible for data collection, analysis, and manuscript composition. Shyam Parshotam aided in data analysis for kinetic fitting, and Sarah Hales did preliminary experiments, not part of this thesis, to initiate the study on the background triggered reaction. Dr. Julianne M. Gibbs guided this project and contributed to the writing and editing of the manuscript as a supervisory author.

Chapter 3 of the thesis is original work by me. No part of this chapter has been previously published. I was responsible for data acquisition and analysis.

Chapter 4 of the thesis has been under revision as Park, H.; Osman, E.A.; Cromwell, C.R.; St. Laurent, C.D.; Liu, Y.; Kitova, E.N.; Klassen, J.S.; Hubbard, B.P.; Macauley, M.S.; Gibbs, J.M. CRISPR-Click Enables Multi-Gene Editing with Modular Synthetic sgRNAs in *Bioconjugate Chemistry.* I am the first co-author of this paper with Dr. Eiman Osman. I was responsible for conducting experiments and analysis as well as writing the manuscript. RNA strands were synthesized and characterized by me and Dr. Eiman Osman. Most of sgRNA that have been used were synthesized by me. Yuning Liu performed a preliminary experiment, not part of this chapter, to initiate the project. Christopher Cromwell performed the *in-vitro* assays and the cell-based assays. Christopher St.Laurant conducted the cell-based assays and the flow cytometry data related to Siglec sequences. Dr. Elena N. Kitova and Dr. John S. Klassen performed the characterization

by mass spectrometry on tracrRNA and sgRNA. Dr. Matthew S. Macauley and Dr. Basil P. Hubbard helped guide this project and contribute to the writing and editing of the manuscript. Dr. Julianne M. Gibbs was the supervisory author who guided the project and contributed to the writing and editing of the manuscript.

Acknowledgments

A journey from the start of my Ph.D. to the end was a meaningful adventure for me. I learned a lot and met incredible people. I would first like to thank my supervisor, Professor Julianne M. Gibbs. Her insightful feedback pushed me to sharpen my thinking and brought my knowledge to a higher level. Without her assistance and dedicated involvement, this journey would have never been accomplished. I want to thank her very much for her outstanding support and understanding over the last five years.

I would also like to show gratitude to my committee, Professor Fredrick West and Professor Christopher Cairo, for giving me insightful feedback on my research and advice. I am grateful for the collaboration with Professor Matthew Macauley and Professor Basil Hubbard on the CRISPR project, which broadened my biochemistry knowledge. Special thanks to Professor Robert Campbell for the excellent advice. Also, I would like to thank Dr. Hayley Wan for being a great mentor. Being an organic chemistry teaching assistant for the last five years taught me teaching skills and motivated me to accomplish the graduate teaching and learning courses. I also want to express special appreciation to Gareth Lambkin in biological service. He taught me many instruments and gave me advice whenever I encountered difficulties. He was always willing to help. Special thanks to Jing Zheng and Wayne Moffat for technical advice. I would like to thank the Alberta/Technical University of Munich International Graduate School for Hybrid Functional Materials (ATUMS) programs for giving me an opportunity for the research exchange program. Special thanks to Professor Job Boekhoven for hosting me at the Technical University of Munich (TUM) and giving me a memorable research experience.

I want to thank the former and current Gibbs group members, especially the DNA group, who shared many experiences in the lab. Special thanks to Shyam Parshotam for helping me with

the kinetic simulations. I am happy to be a part of the Gibbs group. I would like to thank my friends from the Chemistry department, Yuki Kamijo, Jenny Lin, Michelle Ha, and Christina Braun, for their understanding and positive vibes. Without them, life in Edmonton would not be pleasant. I am grateful for our friendship and for sharing many memories with them.

Most importantly, I thank my family back home, South Korea, for their invaluable support and love. Studying abroad in Canada would not be possible without them. Even though I am far away by myself, I always feel their love and care. Thank you, mom, Mija Kwak and dad, Kwangseo Park, and lovely sisters, Hanbyeol Park and Hansem Park. Also, I would like to thank my future in-law family in Prince George. They are always welcoming and make me feel like having another home in Canada. Special thanks to Olga Rehl for her caring and to the little boys, Jakob, Eli, and Anthony, for cheering me up with their cute smiles. Finally, I would like to thank my fiancé, Benjamin Rehl, for giving me the most significant support and love. He always stands by me and encourages me to get through the difficulties during the Ph.D.

Table of Contents

List of Tables

List of Figures

List of Symbols/Abbreviations

APS	Ammonium persulfate
AMA	Ammonium hydroxide/40% aqueous methylamine (1:1 v/v)
AMP	Adenosine monophosphate
AP site	Apurinic/apyrimidinic sites
ARP	aldehyde-reactive probe
ATP	Adenosine triphosphate
(BimC4A)3	Tripotassium 5,5′,5″-[2,2′,2″-nitrilotris(methylene)tris(1H-benzimidazole-2,1-diyl)]tripentanoate hydrate
bp	base pair
BSA	Bovine serum albumin
Cas9	CRISPR associated protein 9
CEU	Cohesive end unit
Chi2/DoF	chi-square per degree of freedom
CMAS	Cytidine monophosphate N-Acetylneuraminic acid synthetase
CPG	Controlled pore glass
CRISPR	Clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
CuAAC	Copper-catalyzed alkyne-azide cycloaddition
Cy5	Cyanine 5
DCM	dichloromethane
DMAP	4-(dimethylamino)pyridine
DMF	Dimethylformamide
DMT	Dimethoxy trityl
DNA	Deoxyribonucleic acid
dPTP	pyrene nucleoside triphosphate
DSB	Double-strand DNA breaks
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EDC	1-ethyl-3-(3-dimethylamino propyl)-carbodiimide

EDTA	Ethylenediaminetetraacetic acid
ESI	Electrospray ionization
ESI-MS	Electrospray ionization-mass spectrometry
FACS	Fluorescence-activated cell sorting
Fln	Fluorescein
gRNA	guide RNA
HDR	Homology-directed repair
HPLC	High-performance liquid chromatography
IDT	Integrated DNA technologies
ITC	Isothermal titration calorimetry
LCR	Ligase chain reaction
LIDA	Lesion induced DNA amplification
MALDI	Matrix-assisted laser desorption/ionization
MALDI-TOF	Matrix-assisted laser desorption/ionization time-of-flight
NEB	New England Biolabs
NFW	Nuclease-free water
NHEJ	Non-homologous end-joining
NHS	N-hydroxy succinimide
NMR	Nuclear magnetic resonance spectroscopy
nt	nucleotide
ONs	Oligonucleotides
PAGE	Polyacrylamide gel electrophoresis
PAM	protospacer adjacent motif
PBCV-1 DNA ligase	*Paramecium Bursaria Chlorella* virus 1
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PPi	Pyrophosphate
RNA	Ribonucleic acid
RNaseH	Ribonuclease H
RNP	Ribonucleoprotein

RP HPLC	Reversed-phase high-performance liquid chromatography
RVD	repeat variable diresidue
sgRNA	single guide RNA
Siglecs	Sialic acid-binding immunoglobulin-type lectins
INDELs	Insertion or deletion of bases
SNP	Single nucleotide polymorphisms
SPREAD	Surface-promoted replication and exponential amplification of DNA analogs
SSB protein	Single-strand binding protein
T7EI	T7 endonuclease I
TAE	Tris-acetate-EDTA
TALEN	Transcription activator-like effector nucleases
TAMRA	Tetramethyl rhodamine
TBDMS	tert-butyldimethylsilyl ethers
TBE	Tris-borate-ethylenediaminetetraacetic acid
TBMDS	tert-butyldimethylsilyl group
TEAA	Triehylammonium acetate
TEMED	Tetramethylethylenediamine
THF	tetrahydrofuran
TLC	Thin-layer chromatography
Tm	Melting temperature
TMS	Tetramethyl silane
p-Toluoyl	4-methoxybenzoyl
TON	Turnover number
tracrRNA	trans-activating crRNA
TRIS-HCl	Tris(hydroxymethyl)aminomethane hydrochloric acid
UBP	Unnatural base pairs
UDG	uracil-DNA glycosylase
UV	Ultraviolet
ZFN	Zinc finger nucleases

Chapter 1

Introduction

1.1 Origin of Life

The question of the origin of life continues to be one of the most challenging questions in science. The formation of life on early Earth must have been preceded by complex chemical networks with emerging life-like properties. Despite the effort of numerous scientists, the question is difficult to answer due to the lack of direct evidence of early life. Geologists estimate that the Earth formed approximately 4.6 billion years ago from the measurement of the age of the oldest rocks on Earth. However, the environment was not suitable for life as it is now due to the high temperature and no or low oxygen level. Early cellular life appeared around 3.6 billion years ago based on fossil evidence (**Figure 1.1**).[1–3] Although the earliest fossils, single-cell microbes, indicate the existence of living organisms, the fundamental question of how life evolves remains unanswered.

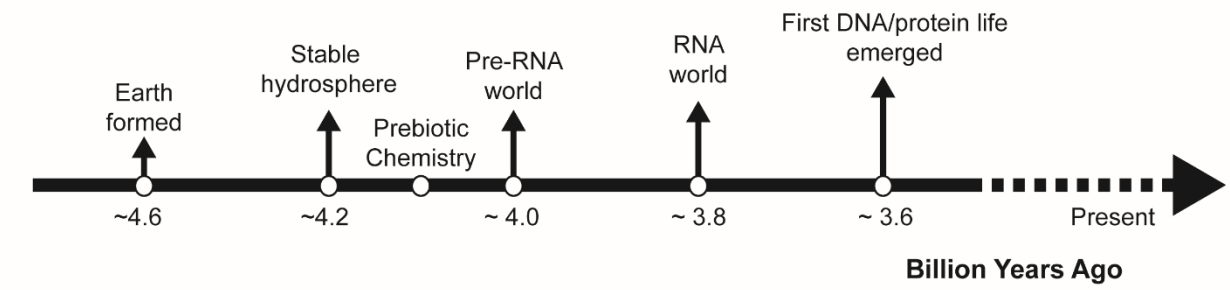


Figure 1.1 Timeline of the early history of life on Earth.[3] The figure was reconstructed from reference 3.

1.1.1 Oparin-Haldane Hypothesis

In 1924, Russian biochemist Alexander Oparin suggested that primitive Earth contained a strongly reducing or oxygen-poor atmosphere, including methane, ammonia, hydrogen, and water vapor.[4,5] Oparin proposed that this atmosphere gradually resulted in the growth and increase of complexity of molecules with new properties. Independently, an English scientist J.B.S Haldane

also proposed a similar hypothesis in 1929.[6,7] This theory is called the "Oparin-Haldane hypothesis," also known as "chemical evolution" or "heterotrophic theory." In this theory, simple inorganic molecules could have reacted to form building blocks such as amino acids and nucleotides, enabling them to accumulate in the ocean, often referred to as "primordial soup." According to Oparin, charged amphiphilic macromolecules were created through a combination of smaller organic molecules. On the other hand, Haldane proposed that complex compounds with "oily film" acting like membranes were produced and resulted in the cellular components. Although there was no substantial evidence to support the theory, this general approach significantly impacted further studies in biology. Later, this theory was supported by the "Miller-Urey experiment."

1.1.1.1 Miller-Urey Experiment

In 1953, Stanley Miller and his supervisor Harold Urey conducted the experiment to demonstrate how organic molecules could be spontaneously produced by inorganic molecules under the condition of the Oparin-Haldane theory.[8–10] They used a mixture of gases, methane (CH_4), ammonia (NH_3), water (H_2O), and hydrogen (H_2), which were believed to exist in primitive Earth's atmosphere. The gas mixture was then reacted with the continuous electrical sparks, which mimic the lightning effect that may provide the energy for chemical reactions on early Earth (**Figure 1.2**). Miller and Urey found that basic organic monomers such as amino acids were formed in this experiment. Although nucleotides were absent, proof of the formation of building blocks from inorganic molecules abiotically in early Earth conditions was a breakthrough in studying the origin of life.

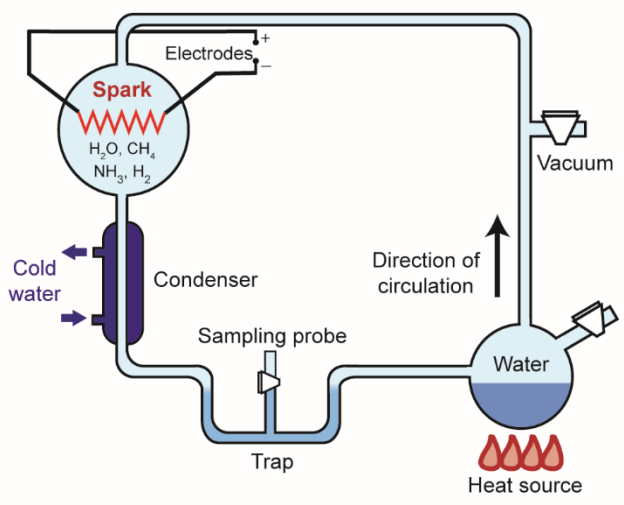


Figure 1.2 Schematic of Miller and Urey experiment. An apparatus holds a mix of gases similar to the early Earth atmosphere over a pool of water which represents the early Earth Ocean. As the water in the flask was heated to induce evaporation, the water vapor entered the larger flask. Electrodes delivered an electric current to generate a spark, similar to lightning, in a flask containing water vapor and gas mixture. The condenser was set up to cool down the atmosphere, and a U-shaped trap was set up to collect the solution. The reaction was performed over a week.[9]

1.1.2 Metabolism-First Hypothesis

To study the nature of early Earth life, it is necessary to understand how prebiotic polymers would become self-replicating or self-perpetuating materials to sustain life. There are two main hypotheses to explain the mechanism of early life, the "gene-first" hypothesis, and the "metabolism-first" hypothesis. The "metabolism-first" idea suggests that metabolic reactions could exist before the first macromolecules were produced.[11] While there are many theories underlying the "metabolism-first" view, such as prebiotic autocatalytic network,[12–15] Springsteen and Krishnamurthy showed that environmental chemical reactions could be enhanced by natural selection favoring the generation of macromolecules.[16] Springsteen and co-workers also discovered that a group of similar reactions could happen in the absence of macromolecules.[17] Their experiment was conducted in very mild conditions, producing most of the product in the reversed

citric acid cycle (or the reverse Krebs cycle) with a little waste. Although the metabolism-first hypothesis has been studied to show that metabolic reactions can produce a series of organic molecules readily in certain conditions that are relevant to biology,[18] the explanation of how the macromolecules emerged has not been answered yet.

1.1.3 RNA World Hypothesis

Another possible hypothesis is the "gene-first" hypothesis, stating that self-replicating nucleic acids are the first life form in early Earth.[2,19–21] Unlike DNA, deoxyribonucleic acids, which only store genetic information in their biological role, RNA, ribonucleic acids can store genetic information and act as catalysts (ribozymes). Accordingly, RNA should have been able to catalyze its synthesis to achieve reproduction and self-sustainability. Due to these unique properties of RNA, the "RNA world hypothesis" proposes that RNA is likely the first genetic material on early Earth (**Figure 1.3**).[1,3,12,21,2,19,20,22–26] In this theory, RNA building blocks emerged in a primordial soup of molecules. Some stable RNA would enable them to grow and make copies of themselves. Over millions of years, the RNA could reproduce and create an array of RNA machines or RNA sequences with specific functions. The RNA world theory is viewed as the most likely explanation for the origin of life.[2,3,21] However, the spontaneous generation of protein and DNA from RNA on early Earth is not straightforward to explain, and likely mechanisms are still sought. Importantly, how the first RNA replicator appeared also remains a mystery.

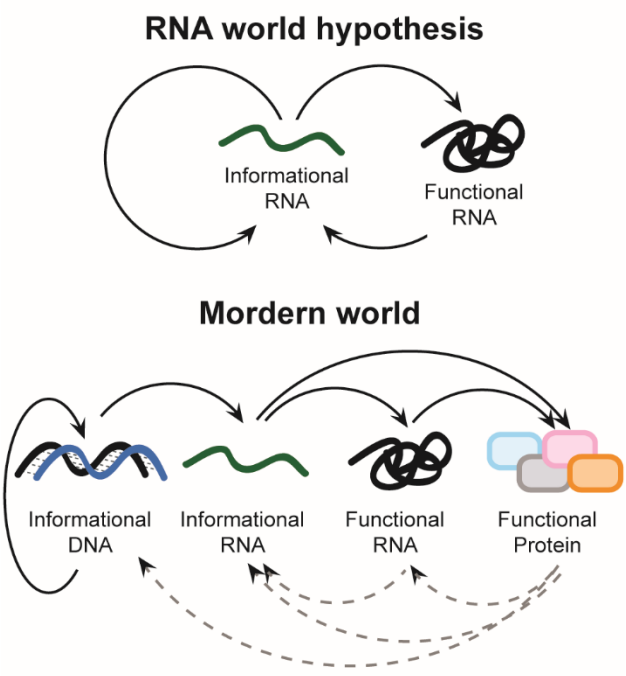


Figure 1.3 Scheme illustrates RNA world hypothesis compared to modern world molecular biology. The RNA world hypothesis suggests that RNA molecules have the ability to perform the jobs that protein and DNA do in the modern world.[27] Modified with permission from *Biology Direct* **2010**, *15*(5). Copyright © 2010 Springer Nature.

1.2 Self-Replication in the Origin of Life

An essential process in the origin of life was the evolution of the first building blocks that could replicate and transmit genetic information on the early Earth. Once the first self-replicating material appeared and underwent natural selection, favored offspring with mutations or variation would make copies of themselves. Eventually, the first simple cells would appear upon the addition of compartmentalization, metabolism, and other life-like features. According to the RNA world hypothesis, the ability of self-replication probably evolved from an RNA self-replicator.[2,3,21] However, researchers working in systems chemistry proposed that self-replicating small molecule systems were also likely to have developed from the complex molecular networks present on early Earth that may have also played a role in the origin of life.[28–30]

1.2.1 DNA Replication in Nature

Self-replication is the ability to yield the structure of an identical or a similar copy of itself. In biology, DNA replication is an essential step for genetic inheritance. During cell division, DNA is replicated and transfers the genetic information to offspring.[31,32] The first step in DNA replication is helicase enzymes unwinding the double strands of DNA by breaking the hydrogen bonds of the DNA base pairs (**Figure 1.4**).[31–33] The separation of the two single strands of DNA forms a Y shape called a replication fork. These two DNA strands, the leading strand (3' to 5' direction toward the replication fork) and the lagging strand (5' to 3' direction), act as templates for building the new strands of DNA. A short RNA strand, a primer, comes along and binds to the leading strand.[31–33] Then, DNA polymerase binds to the leading strands and allows replication by adding complementary nucleotides. Meanwhile, the primase reads the short, separated template and synthesizes a short RNA primer. Then, DNA polymerase can extend the complementary lagging strand and generate Okazaki fragments.[31–33] The fragments of the DNA strands are then linked together by DNA ligase. The key steps in DNA replication are to unzip the double helix and extend and copy the DNA template. To mimic nature's replication strategy, it is essential to induce the dehybridization of nucleic acid efficiently.

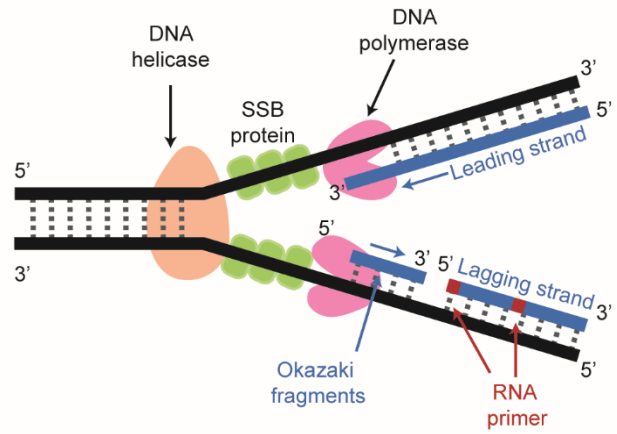


Figure 1.4 Schematic illustration shows DNA replication in nature. DNA helicase unwinds double-stranded DNA, and single-stranded binding (SSB) protein binds single-stranded regions of DNA. DNA polymerase synthesizes the complementary strand using template strands as guides.

1.2.2 Autocatalytic Self-Replication

There are two types of reactions that can replicate themselves. An autocatalytic reaction is a chemical reaction when the reaction product directly catalyzes its formation.[34–36] A simple model can be used to describe the autocatalytic process for oligonucleotide self-replication (**Figure 1.5**).[35] For example, the template (T) in the model is a self-complementary DNA strand that has the ability to amplify itself. First, the T binds to A and B fragments via Watson-Crick base pairs leading to the formation of a ternary complex. Protein machinery can ligate the A and B fragments to generate the new T template. Or in non-enzymatic conditions, chemical ligation can occur to form a covalent bond between the fragments due to proximity. The generation of the product strand after ligation is irreversible, but the dissociation of the product:template duplex is reversible, resulting in the initiation of a new replication cycle. The number of replication cycles (N) generates 2^N of the initial template. The cycle number and time for each cycle represent the catalytic activity.

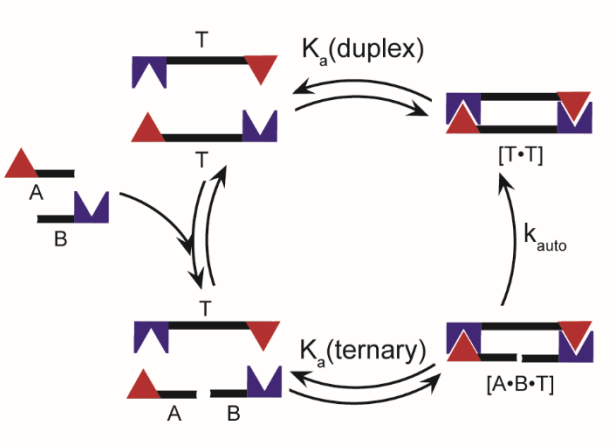


Figure 1.5 Scheme representing an autocatalytic self-replication system. Fragments A and B bind to template T, forming a ternary complex. Ternary complexes can react to generate template-product duplexes. K_A(ternary) indicates the association of the ternary complex, while K_A(duplex) indicates the template-product duplex. k_{auto} represents the rate of product formation.

1.2.3 Cross-Catalytic Self-Replication

Although the autocatalytic reaction illustrates simple self-replication, nucleic acid replication in nature primarily undergoes replication with complementary strands rather than self-complementary strands.[36] This type of replication is a cross-catalytic reaction where two templated cycles are couples, and each template catalyzes the formation of a product that is the template for the other cycle.[36,37] A basic cross-catalytic system is described below (**Figure 1.6**). In the first cycle, B and D fragments hybridize to T^{AC} template, forming the other template (T^{BD}) that acts as a template in the other cycle. When this cross-catalysis is successful, the formation of one template would show up as a sigmoidal growth when one of the fragments is limited. The key step is how fast the product duplex can dissociate and release the single-stranded product for both autocatalytic and cross-catalytic systems using oligonucleotides as the replicator.[34,38–40]

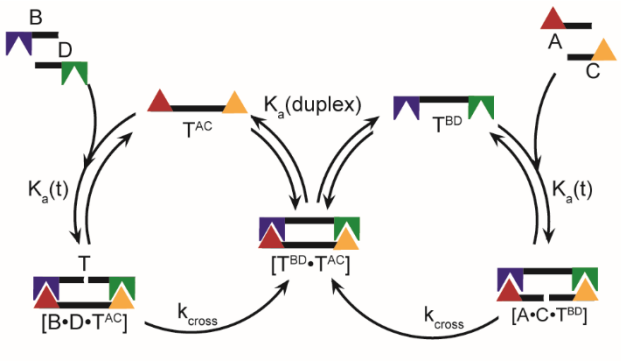


Figure 1.6 Scheme illustrates a cross-catalytic system. Fragments B and D bind to Template $T^{AC,}$ while fragments A and C bind to template T^{BD}. The ternary complexes undergo reactions, such as ligation, to generate the template-product duplexes. $K_A(t)$ represents the association affinity of the ternary complex, while K_A(duplex) is the affinity of the product duplex. k_{cross} is the rate of formation of product duplex.

1.2.4 Product Inhibition

The biggest challenge in self-replication systems based on nucleic acids is product inhibition, which prevents the release of products.[34,38–40] The ideal condition to drive spontaneous dissociation of the product is when the ternary complex is more stable than the product duplex (**Figure 1.5** and **Figure 1.6**).[34] However, DNA strands have much stability in the product duplex form due to a large number of hydrogen bond interactions between the two strands after ligation of the fragments. The stronger binding affinity of the product duplex compared with the ternary complex can be attributed to the smaller entropic penalty to hybridization that arises from forming the same number of hydrogen bonds and π-stacking interactions from two strands (the product duplex) rather than three (the ternary complex).[41] Owing to this characteristic, the template ligated reaction rarely shows catalytic activity based on spontaneous dissociation of the product:template duplex and the reintroduction of the template into another catalytic cycle.[42–44] Therefore, strategies to reduce product inhibition in templated ligation reactions of oligonucleotides have been studied, which are described in section 1.2.6.[42,43,45] Regarding the origin of life where protein machinery

(like helicases) were unlikely to exist, modulating the environmental conditions has been studied to induce the dissociation of oligonucleotide duplexes providing a possible solution to the problem of product inhibition.[46–49]

1.2.4.1 Overcoming Product Inhibition by Environmental Changes

Despite the importance of oligonucleotides replication in the absence of enzymes, the strategies to induce the separation of oligonucleotide duplex without thermal cycle in plausible prebiotic conditions remain a challenge. Theoretically, daily or seasonal cycles or geothermal activity could have caused temperature fluctuations, which could cause the water evaporation from the pool containing small organic molecules and inorganic salts.[48,50] Salt such as NaCl, KCl, or $MgCl_2$ contributes to the stability of oligonucleotides by interacting with the phosphate backbone charges.[51–53] Changing salt concentration is one of the ways to regulate the melting temperature of DNA or RNA duplex.[54,55] Braun and coworkers demonstrated that the closed system, such as a porous rock, could fluctuate temperature, leading to salt concentration changes.[46] They used the chamber, which had both a colder side of the wall and a warmer side. They observed the modest temperature difference could cause the evaporation of salt-free water from the warm side and condensation of the water droplet to the cold side, which resulted in salt concentration changes.[46] This salt fluctuation results in a decrease in melting temperature and transient dissociation of a DNA duplex (**Figure 1.8A**). This method can be one strategy to overcome product inhibition. Sutherland and colleagues demonstrated that changes in pH could also be used to modulate the melting temperature of the RNA duplex (**Figure 1.8B**).[47] This provides a possible route to overcome product inhibition as a recent study has shown that stable pH gradients can be sustained by a thermal cycle of heating and cooling in a closed system, which might have occurred on early

Earth.[56] These studies suggest that the formation and evolution of RNA in prebiotic Earth can be obtained with intrinsic temperature, salt concentration, or pH cycling and gives insights into mechanisms of nonenzymatic nucleotide replication. Additionally, Hud and coworkers demonstrated that solvent viscosity can also be used to generate longer oligonucleotide polymers (> 300 nt length) by templated ligation.[48] Increased solvent viscosity slowed down the duplex formation but allowed the assembly of short oligonucleotides on the longer template (**Figure 1.8C**).[49] This approach permits the information transfer with gene-length sequences.[49] Furthermore, with this strategy, they demonstrated the nonenzymatic replication of RNA duplex (~100-600 bp) by fluctuating the environmental conditions such as the temperature and water viscosity.[49] Although only a single round of replication was demonstrated due to the irreversible changes in the experiment, their strategy showed the potential of continuous replication of oligonucleotide using similar chemistry in the an environment with fluctuating viscosity.

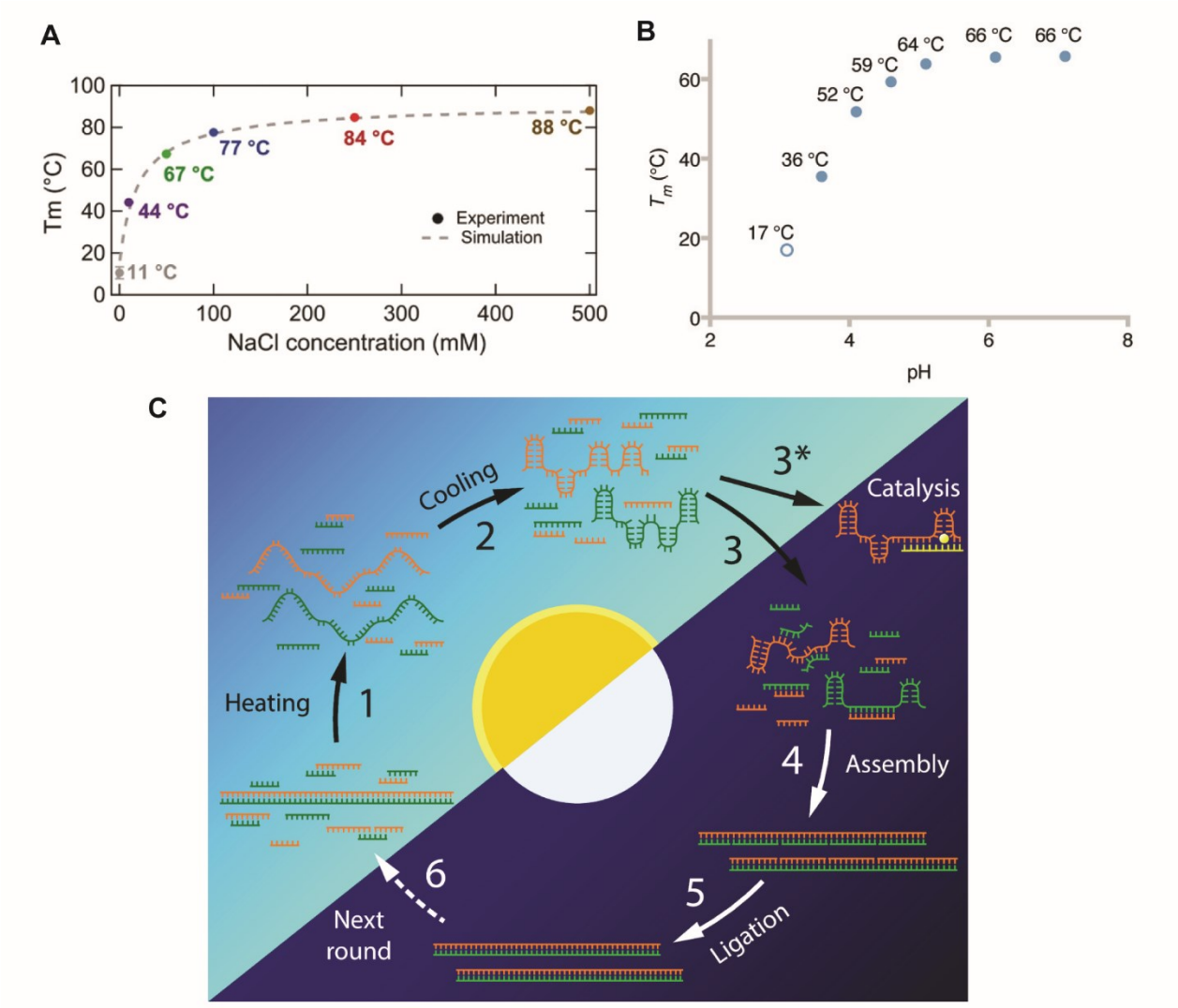


Figure 1.7 Environmental changes can reduce the stability of the oligonucleotide duplex providing a route to overcoming product inhibition. (**A**) The changes in salt concentration can reduce the melting temperature of DNA duplex.[46] Image reproduced from *Angew.Chem. Int. Ed.* **2019**, *58*, 13155 –13160. Copyright © 2019 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim. (**B**) The controlling pH can tune the melting temperature of the RNA duplex.[47] Reprinted with permission from *Biochemistry* **2018**, 57, 45, 6382-6386. https://pubs.acs.org/doi/10.1021/acs.biochem.8b01080. Copyright © 2018 American Chemical Society. (**C**) The solvent viscosity can induce the facilitation of RNA replication.[49] Reprinted with permission from *Nucleic Acids Research*, **2019**, *47*(13), 6569–6577. Copyright © 2019 Oxford University Press.

1.2.5 Self-Replication of Oligonucleotides

To begin with, the self-replication of oligonucleotides is critical in both chemical evolution and RNA world theories.[2,34,57–59] Chemical evolution supports that short oligonucleotides can be assembled by random polymerization and undergo template-directed replication.[60–62] This enzyme-free replication is at the boundary between the simple chemical evolution and the beginning of biological evolution. One condition to enable self-replication is amplification through autocatalytic reaction if the product is self-complementary and acts as a template. Von Kiedrowski[35] and Orgel[63] demonstrated the first autocatalysis in DNA self-replication.

1.2.5.1 Günter Von Kiedrowski

In 1994, Slevers and von Kiedrowski showed nonenzymatic self-replicating systems based on cross-catalytic templated reaction (**Figure 1.9**).[37] They used two self-complementary and two complementary templates, which could result in four combinations of templates (**Figure 1.9A**). For the ligation chemistry, they introduced the reaction between a 3'-phosphate and 5'-amine in the presence of water-soluble carbodiimide EDC (1-ethyl-3-(3-dimethylamino propyl)-carbodiimide), producing the 3'-5' phosphoramidate linkage. The initial template induced the amplification of the intended template (**Figure 1.9B**), although the side products were also generated. The kinetic studies indicated that this system undergoes product inhibition, which interrupts the exponential amplification of the template. Later in his research, von Kiedrowski designed surface-promoted replication and exponential amplification of DNA analogs (SPREAD) to overcome the product inhibition by immobilizing formed products (templates) on a surface.[64]

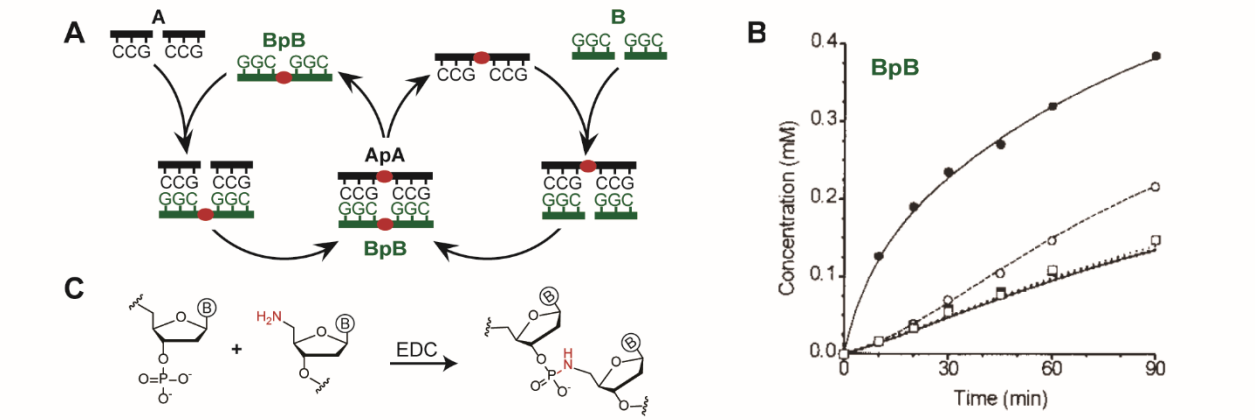


Figure 1.8 Examples of nonenzymatic self-replication of an oligonucleotide. (**A**) Cross-catalytic self-replication of von Kiedrowski.[37,65] (**B**) The chemical reaction was driven by the presence of carbodiimide (EDC). The figure is regenerated from the description of his paper. (**C**) The formation of the different products over time when the BpB template was initially added. Reconstructed with permission from *Nature* **1994,** 369, 221–224. Copyright © 1994 Springer Nature. *Chem. Eur. J.* **1998,** 4, 629-641. Copyright © 1998 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim. *ChemBioChem*, **2008**, 9(14), 2185-2192 Copyright © 2008 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

1.2.5.2 Gerald F. Joyce

Ribozymes are ribonucleic acid enzymes that catalyze biochemical reactions, including the ligation and cleavage of oligonucleotides.[66,67] Ribozymes catalyzing RNA ligation is often considered as a model for autonomous self-replication in nucleic systems.[67] Joyce and co-workers demonstrated isothermal self-replication of an RNA based on an R3C ligase ribozyme.[67] He restructured the sequence of the R3C ligase as a symmetrical dimer so that the product has the identical sequence with the template. Although RNA-catalyzed self-replication showed exponential growth, the dimer complex interrupted the ternary complex formation. In other words, product inhibition still presented a problem. Later, Lincoln and Joyce demonstrated a successful isothermal cross-replicating ribozyme (R3C ligase) by controlling the stability of substrates with the G: U mismatch pair (**Figure 1.10**).[68] Their system showed robust exponential growth of the template (R3C ligase) in a sigmoidal shape. This study is an excellent example of the self-

replication of RNA in the absence of proteins. However, this ligation reaction is a very sequence-specific reaction that requires the maintenance of the majority of the sequences for its catalytic properties. As such, it does not represent a general model for oligonucleotide replicators.

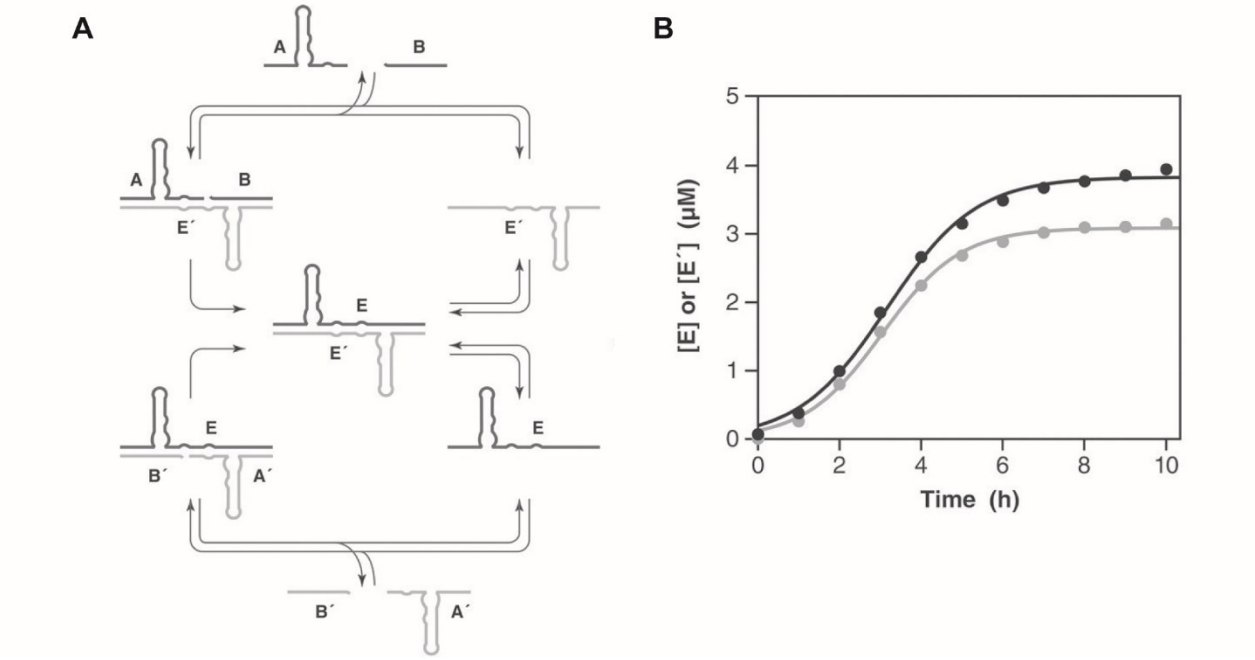


Figure 1.9 Successful example of oligonucleotide self-replication. (**A**) Cross-catalytic self-replication of R3C enzyme. The enzymes E and E' catalyze each other's formation reaction. (**B**) Amplification result of self-replicating R3C enzyme.[68] From Lincoln TA, Joyce GF. Self-sustained replication of an RNA enzyme. From *Science* **2009**, *323*(5918), 1229-1232. doi:10.1126/science.1167856. Reprinted with permission from AAAS. Copyright © 2009, The American Association for the Advancement of Science

1.2.6 Using Destabilizing Groups to Generate Turnover in Ligation Cycles

Despite Joyce's successful example of self-replication, the general strategy to overcome product inhibition remains an ongoing challenge. A typical method to solve product inhibition is the use of thermal cycling. However, natural oligonucleotide replication occurs without thermal cycles. One of the other approaches to reduce the limitation is to design destabilizing groups in

oligonucleotides. The first example of destabilizing the product duplex was proposed by Lynn in the 1990s.[42,45] Aldehyde modified DNA was reacted with amine-modified DNA via reductive amination (**Figure 1.11A**). The imine intermediate was reduced by KBH_3CN and formed the amine linkage. Due to the flexible ethylamine, the amine product duplex was much less stable (10^6-fold) than the imine duplex. Moreover, the amine product was revealed as 30-fold less stable than the ternary complex. Accordingly, they observed about 12 turnovers (13 min^{-1} for 160 min) in a single ligation cycle. Kool and coworkers also reported a simple, reagent-free chemical ligation to produce a destabilizing linker.[43] A flexible hydrocarbon linker at the ligation site was produced by phosphorothioate formation with a dabsyl leaving group (**Figure 1.11B**).[43] This linker allowed 92-fold amplification of the template at one temperature in a single ligation cycle. Seitz and coworkers utilized chemical peptide nucleic acid ligation for a selective and rapid reaction.[44] They demonstrated the enhancement of turnover by involving a flexible rearrangement step from *i*Cys-mediated ligation. The rearrangement in the main chain resulted in increased interbase distances; hence, the ligated product bound less tightly to the DNA template. They showed a turnover of 226 with 10 μM fragments. Another example is that changes in the location of phosphate linkage can be used to lower the stability of oligonucleotides. Szostak and coworkers demonstrated that 2'-5' linkage decreased the melting temperature of the RNA duplex.[69] The 3'-5' linkage can be generated naturally as a mixture of 2'-5' and 3'-5' linkage by a templated reaction due to the proximity and similar nucleophilicity of the hydroxyl group (2' and 3'). The mixture of 2'-5' phosphate linkage in ribozyme not only destabilizes the RNA duplex but also retains the activity. However, none of these strategies were utilized to generate turnover in self-replication cycles (i.e., autocatalytic, or cross-catalytic oligonucleotide replication).

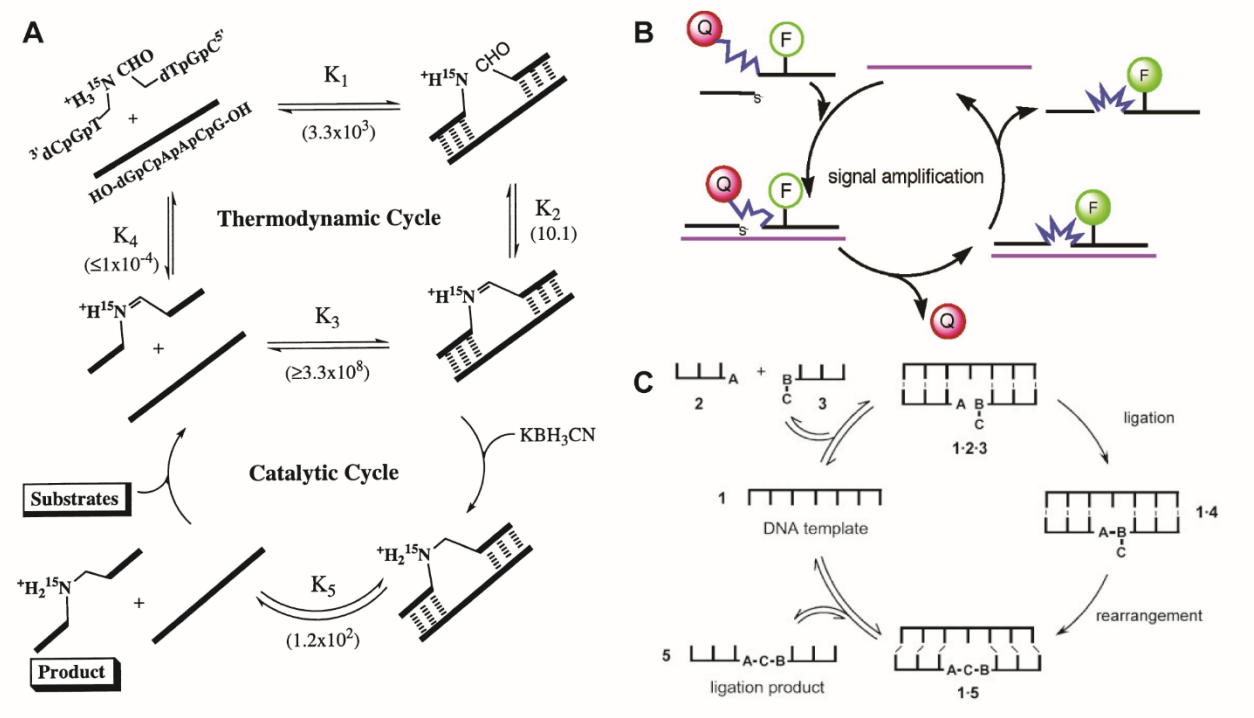


Figure 1.10 Using a destabilizing group to reduce the oligonucleotide duplex. (**A**) The hydrocarbon linker facilitates the turnover in the templated ligation reactions.[43] Reprinted with permission from *J. Am. Chem. Soc*. **2004**, *126*, 43, 13980–13986. Copyright © 2004, American Chemical Society (**B**) The chemical reaction lowers the product duplex stability.[42] Reprinted with permission from *J. Am. Chem. Soc.* **1997**, *119*, 50, 12420–12421. Copyright © 1997, American Chemical Society. (**C**) Scheme of ligation-rearrangement reactions for reducing product inhibition in DNA-templated ligation reaction.[44] Reprinted with permission from *Angew.Chem. Int. Ed*. **2006**, *45*, 5369-5373 Copyright © 2006 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.

1.3 Gene Editing

Nucleic acid chemists have utilized their ability to control the structure and sequence of oligonucleotides to introduce new or enhanced ligation properties that have been used not only in origins of life research but also in biodiagnostics.[70,71] At the same time, chemists and biologists have sought in the last decade strategies to modify the underlying genetic code in cells that also utilize our ability to readily produce RNA and DNA sequences, including with modifications.[72–74]

Gene editing is the process of changing the gene sequences by inserting, deleting, or modifying DNA bases in the genome. Since the 1970s, gene editing has been studied in various fields, such as agriculture, medicine, and biotechnologies.[75–77] In the 1990s, genetically modified foods were introduced to the market to increase resistance to insects and herbicides.[76,77]

With the advent of strategies to synthesize DNA in the lab,[78–80] DNA constructs could be designed that contained a sequence that recognized the target genes, allowing it to be cut by an endonuclease followed by repair.[81–84] Repair of DNA double-strand breaks can undergo two different mechanisms, non-homologous end joining (NHEJ) and homology-directed repair (HDR). NHEJ occurs with various enzymes to directly ligate the broken ends without a homologous template, while HDR repairs the breakpoint in accordance with the inserted template (**Figure 1.12**). Although gene-editing technologies allow researchers to study gene functions and expressions by targeting specific genes, the low efficiency of editing remains a drawback. To overcome this low efficiency and increase the specificity of target cleavage, three main nucleases have been discovered and engineered for gene-editing; zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALEN), and the clustered regularly interspaced short palindromic repeats (CRISPR/Cas) system.

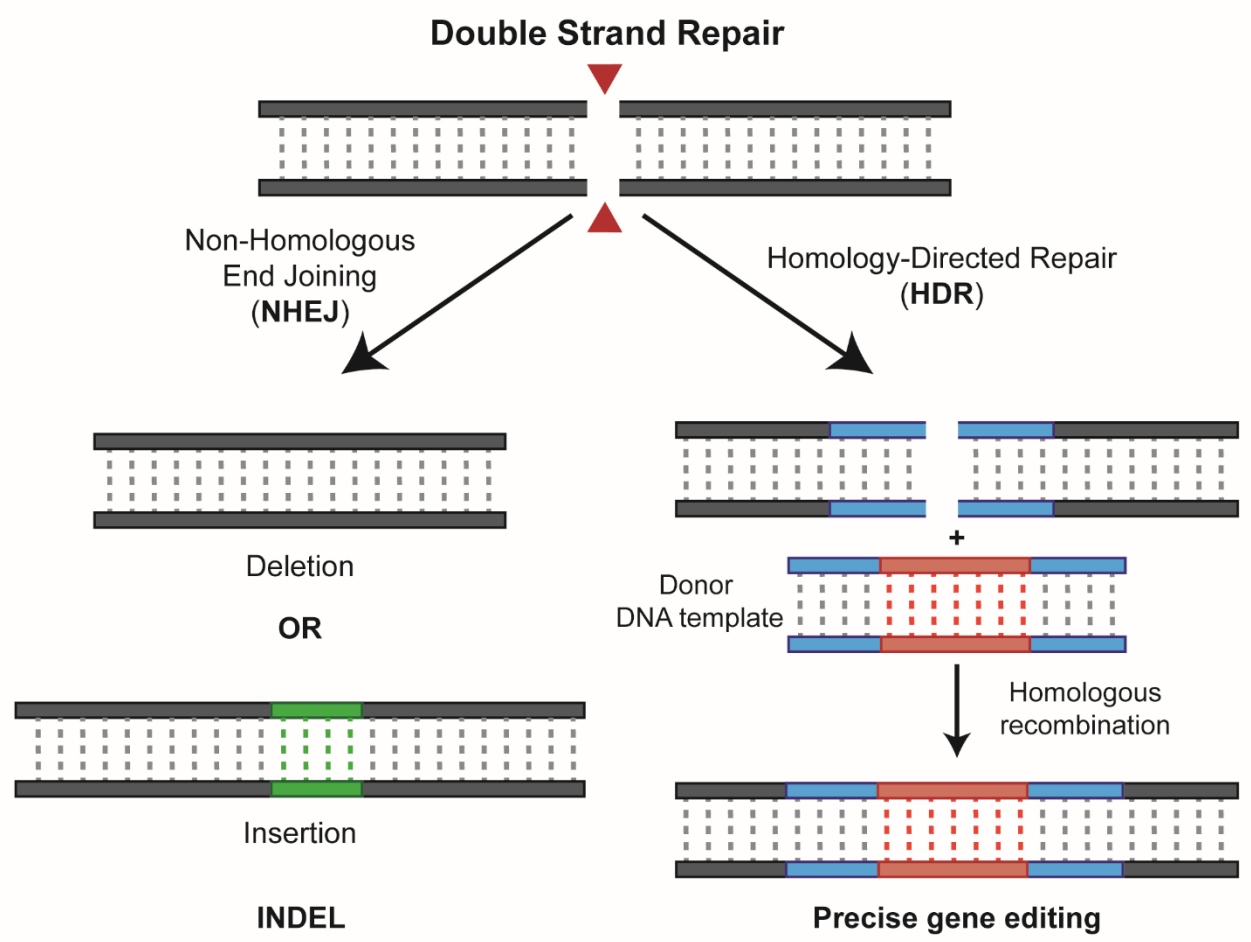


Figure 1.11 Two different methods for repairing double-strand DNA breaks (DSBs). NHEJ method ligates the breaks resulting in deletion, insertion, or knock-out. HDR pathway repairs the break using a template leading to gene replacement or knock-in.

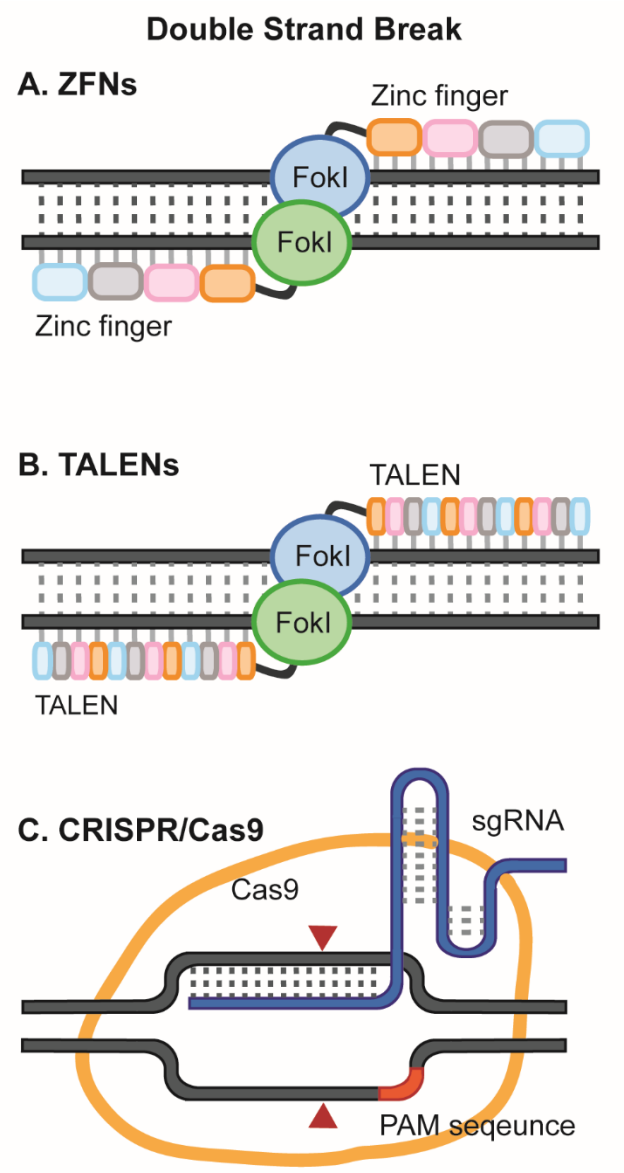


Figure 1.12 The core genome editing technologies that are most commonly used. (**A**) Zinc finger nucleases (ZFNs). (**B**) Transcription activator-like effector nuclease (TALENs). (**C**) Clustered regularly interspaced short palindromic repeats (CRISPR) - CRISPR-associated protein 9 (Cas9).

1.3.1 Zinc Finger Nucleases (ZFNs)

Zinc finger nucleases (ZFNs) are engineered DNA-binding proteins that can edit specific genes by cleaving the double-strand DNA, leading to target mutagenesis and target gene

replacement (**Figure 1.13)**.[85–87] ZFNs are combined with a zinc finger DNA binding domain and a DNA cleavage domain of Fok1 endonuclease. The most common DNA binding domain in ZFN is a set of Cys_2His_2 zinc fingers, which stabilize the protein fold with the coordination of zinc ions. Typically, the zinc finger domain contains three to six individual zinc finger repeats. The crystal study of zinc fingers showed that each finger binds to 3 base pairs of DNA.[88] DNA cleavage domains are originated from the natural type II restriction enzyme Fok1.[89] Although the cleavage domain did not have a sequence specificity, the cutting can be fused to an alternative recognition domain.[90,91] This cleavage domain requires dimerization at the opposite sites of DNA strands with a certain distance apart in order to cleave DNA.[92,93] The most common linker between the zinc finger domain and the cleavage domain is 5 or 6 bp length of the 5' edge of each bonding site.[94–96]

ZFNs recognize 9 - 18 base pairs of target DNA.[75,97] The specificity of ZFNs can be achieved by the requirement of dimerization, as the cleavage reagents only are assembled at the target sequences when the dimer is formed. 18 bp sequences are required for specific gene targeting within 68 billion bp of DNA.[75,98,99] Therefore, this method allowed for target specificity in gene editing of the human genome for the first time. However, there are some potential limitations in gene editing with ZFNs. It has been challenging to assemble the zinc finger domains to bind DNA with higher affinity.[100] Another limitation is the site selection frequency. The ZFNs can only target the binding sites of every 200 bps in DNA sequences.[101–104]

1.3.2 TALENs

TALENs, transcription activator-like effector nucleases, are engineered DNA binding proteins to cut specific sequences of DNA strands. TALENs are made of a TAL effector DNA binding domain and a DNA cleavage domain. A TAL effector originated from the *genus*

Xanthomonas, consisting of 33-34 amino acid repeats.[105,106] The amino acid sequences are mainly similar, except for two adjacent amino acids called repeat variable diresidue, RVD. The repeats with different RVD can recognize one-to-one DNA base pairs, which enables the ease of design of the targeting sequence. DNA cleavage domains are produced from the end of the FokI endonuclease, which is feasible for modification (**Figure 1.13**). In later studies, modified FokI cleavage domains are used to improve the specificity[107,108] and cleavage efficiency.[109] Same as ZFNs, the FokI domain requires the dimer to cleave the target sequence. TALENs has some advantages over ZFNs, such as ease of design of the TALENs construct, higher cleavage activity, and reduced off-target activity.[110,111]

1.3.3 CRISPR/Cas9

CRISPR, clustered regularly interspaced short palindromic repeats, are a class of DNA sequences originating from prokaryotic organisms such as bacteria and archaea.[112] The sequences are derived from the bacteriophage that had previously invaded the cells allowing the bacteria to detect and destroy similar DNA sequences, preventing viral infection. Cas, CRISPR-associated proteins, are enzymes that use CRISPR sequences to recognize the foreign DNA strands. CRISPR, along with Cas enzymes known as the CRISPR-Cas system, has been engineered to allow the modification of genes. The most commonly used CRISPR-associated enzyme, Cas9 from *Streptococcus pyogenes,* contains two domains. The Cas9 HNH nuclease domain cleaves the target strand while the Cas9 RuvC-like domain cleaves the complementary target strands.[113,114] CRISPR needs guide RNA (gRNA), which is made up of two strands: trans-activating crRNA (tracrRNA) and (CRISPR RNA) crRNA. crRNA contains 20 bp of the complement to the target sequences while tracrRNA is a longer RNA strand base pairing with crRNA, enabling crRNA to bind to the

Cas enzyme.[113,115] The tracrRNA and crRNA can be fused into a single strand called single guide RNA (sgRNA).[113] Protospacer adjacent motif, PAM is a 2 - 6 bp DNA sequence that allows the CRISPR-Cas9 system to recognize the target sequence. PAM is a critical component for targeting sequences to distinguish its own bacterial DNA from foreign sequences, avoiding the CRISPR cutting of its sequences. For Cas9, the PAM sequence is 5' NGG 3', where N represents any base followed by two guanine bases.[113,116] Gene editing only occurs when Cas9 recognizes the PAM sequences and unwinds the target DNA sequences (**Figure 1.13**).[116]

The most significant advantages of the CRISPR/Cas system are its simplicity and efficiency over other gene-editing methods. Unlike ZFNs and TALENs, which require encoding large DNA components for each target, the CRISPR/Cas system only needs to engineer gRNA by changing 20 bp of crRNA. Moreover, the CRISPR-Cas system is less expensive and requires less labor compared to ZFNs and TALENs. Accordingly, it is more straightforward to target multiple sequences simultaneously in the same cell.[115,117] However, the CRISPR-Cas9 system has a higher possibility of off-target effects due to the tolerance of multiple mismatches.[118–120]

1.3.3.1 Prospect of the CRISPR-Cas System

The CRISPR-Cas system is a breakthrough technology in genome editing.[113,114,121] In 2020, the pioneers of CRISPR technology, Jennifer Doudna and Emmanuelle Charpentier, were awarded the Nobel Prize in chemistry for the discovery and development of CRISPR. The application of genome editing has been extended in various fields, from studying gene function in plants or animals to human gene therapy. The introduction of the CRISPR technology begins a new era in gene therapy by facilitating the correction of particular gene sequences. CRISPR-Cas9 has enabled the study of diseases,[122] the treatments of multiple human diseases,[123,124] and the design of sensitive

biomarker detection platforms.[125] A recent small clinical trial of intravenous CRISPR-gene editing in humans was first reported with a promising result.[126,127]

1.4 Thesis Organization

This thesis mainly concentrates on studying kinetic behaviors of isothermal DNA self-replication called lesion-induced DNA amplification, exploring on selective ligation of pyrene nucleotide across abasic site, and developing a modular method to make single guide RNA for the CRISPR-Cas9 system with click CuAAC chemistry and fluorophore conjugation.

Chapter 2 is a study to understand the kinetic behavior of lesion-induced DNA amplification. The kinetic model was used to measure the binding affinity of the intermediate complex and product duplex and the ligation rate in the presence of T4 DNA ligase. The binding affinities of the intermediate complex and product duplex were also measured in thermodynamic measurements without T4 DNA ligase. Comparison of the changes in binding affinities between experiments with enzyme and without enzyme suggested that the enzyme reduced the gaps between intermediate complex and product duplex further with the aid of abasic group. Furthermore, the fitting to measure Michaelis constant was measured to support the effect of the abasic group. This study concludes that the critical roles in the success of LIDA are not only the abasic destabilizing group but also T4 DNA ligase by overcoming product inhibition.

Chapter 3 aims to incorporate the pyrene base in selective ligation across abasic lesions. T4 DNA ligase and PBCV-1 DNA ligase were used in various ATP concentrations and temperatures to find the most effective ligation condition. The selectivity and the efficiency of pyrene nucleotides across from abasic lesions were compared to the DNA templates containing natural bases on the opposite sites of pyrene. Furthermore, the competitive ligation confirmed the 5’-

phosphate pyrene strand has high selectivity on the opposite side of the abasic site. In addition, the effect of a bulky destabilizing group at the ligation site in the isothermal DNA self-replication system was studied and compared to the native DNA sequence without any modifications. Although further studies are needed, it shows the possibility of the usage of bulky destabilizing groups to induce successful isothermal self-replication. Moreover, it suggests that a native DNA system can replicate the 18 nt DNA sequence in optimal conditions without any destabilizing agent.

Chapter 4 focuses on designing a modular synthesis of single guide RNA for the CRISPR/Cas9 system with a fast chemical ligation method, Cu(I) catalyzed alkyne-azide cycloaddition. To reduce the burden of synthesis of a long sgRNA (~100 nt), the sgRNA was divided into three smaller fragments with azide and alkyne functional groups. For ease of synthesis of modification on RNA strands, the modified DNA bases were used and at a 5' position to avoid degradation. Simple syringe synthesis was used to functionalize 5' azide on RNA fragments, and commercially available 3' propargyl modified CPG was used in solid-phase synthesis. Fluorophore was conjugated with an NHS ester-amine reaction to label sgRNA. Two different fluorescently labeled sgRNAs were tested in an *in-vitro* cutting assay to show the activity of sgRNA. Furthermore, the sgRNAs were transfected into the Cas9 expressing cell and measured the cleavage efficiency in the cell. The synthesized sgRNAs were successfully worked both in *in-vitro* and cell-based assays, suggesting that sgRNA can be achieved by this modular synthesis.

Chapter 5 is the conclusion of this thesis. Future works and the perspectives of the projects are discussed.

Appendix I shows NMR spectra of synthesis of 1-pyrene nucleotides and ESI spectra of triazole modified tracrRNA and sgRNA.

Chapter 2

Minimizing Product Inhibition in DNA Self-Replication: Insights for Prebiotic Replication from the Role of the Enzyme

"Minimizing Product Inhibition in DNA Self-Replication: Insights for Prebiotic Replication from the Role of the Enzyme" **Park, H.**; Parshotam, S.; Hales, S.C.; Gibbs, J.M. *(in preparation)*

2.1 Introduction

The last century has yielded enormous insights into the beginnings of the universe and the fundamental properties of matter, yet the age-old question of how life emerged on Earth remains unanswered.[2,3] There has been progress in identifying likely prebiotic pathways for generating the amino acid[128–131] and nucleotide[35,37,64,132,133] building blocks of life as well as strategies for their polymerization,[131] but how these materials replicated and became integrated into self-sustaining networks remains a mystery. Recent studies have shown that nucleic acid polymerization can be coupled to vesicle growth and division, providing insight on how compartmentalization and polymerization became intertwined in protocells.[134–136] Strategies for polymerizing nucleotides are also being reported of increasing selectivity and speed based on prebiotically plausible activated phosphate and diphosphate nucleotides.[134,135] However, mechanisms for such oligonucleotides to autonomously self-replicate, which would involve templated synthesis followed by spontaneous dissociation from the template, have not been forthcoming. This dearth arises from the intrinsic thermodynamic problem inherent in the templated synthesis where the product forming on the oligonucleotide template strand binds more and more strongly the longer it grows.[43,137] This product inhibition is significant even in small oligonucleotide replication cycles like the cross-catalytic and autocatalytic systems for hexamer synthesis reported by von Kiedrowski.[35,37,64,132,133] Consequently, the field of prebiotic chemistry has mostly abandoned the idea of autonomous replication after von Kiedrowski's seminal work and instead has focused on scenarios where the local environment is modulated allowing for changes in the template:product duplex stability and therefore turnover of the replication cycle. Environmental changes that have been explored that lead to cycles of hybridization and dissociation of RNA or DNA duplexes involve varying pH,[47] salt concentrations,[46] viscosity, and temperature.[48,49] All of these scenarios can be rationalized to

some extent based on intrinsic temperature cycles and chemical gradients expected on early Earth. Yet, the question of how fragile oligonucleotides could have sustained these slow cyclical processes for many generations of replication remains unclear, making the question of oligonucleotide self-replication far from settled in any prebiotic theory.

One strategy that has also been considered to avoid product inhibition in general nonenzymatic oligonucleotide self-replication, albeit not yet demonstrated, is the introduction of oligonucleotide modifications that destabilize the interaction between the template and the growing complement. For example, it has been noted that 2'-5'- linked oligonucleotides rather than the biologically occurring 3'-5'-linkages can facilitate molecular recognition between the nucleobases in a duplex but weaken the overall duplex stability.[69,138] Such linkages might have led to strands capable of templating the formation of complements that contained the native 3'-5' linkage leading to some information transfer based on the nucleobase code as well as spontaneous turnover in templated ligation.[69,138] However, self-replicating cycles have not been demonstrated with this approach. Other destabilizing strategies have also been investigated that are not limited by prebiotic conditions but instead aim to explore how modulating the structure of the oligonucleotide backbone linkage could impact turnover. Kool and co-workers found that introducing a butyl, propyl, or ethyl linkage between the 5'-position of the terminal nucleobase and the reactive electrophile led to an enhanced turnover of the DNA template in a chemical ligation strategy.[43] Prior to this work, Lynn and co-workers also explored chemical methods to modulate the stability of a chemically ligated strand.[42] In their example, DNA-templated ligation led to an imine linkage between two DNA fragments complementary to the template. Next, the imine was reduced to an amine, which led the ligation product to spontaneously dissociate from the template. The authors proposed that such a method might be useful in cross-catalysis but did not explore it experimentally (**Figure 2.1A**).[139]

Recently, another isothermal and nonenzymatic strategy was introduced by Kool and coworkers that included one templated ligation cycle.[140] Their strategy combined two templated chemical reactions to drive fluorescent signals to detect the target template. The first reaction was a chemical ligation forming a destabilized product that dissociated from the template. This strand then acted as a template for the second reaction consisting of a Staudinger reduction between the two adjacent strands after hybridization to the ligated template (**Figure 2.1B**). The reduction released a fluorescent quencher leading to signal amplification as the template could turnover multiple times owing to the similar stability of the ternary complex before and after the Staudinger reduction upon the ligation formation. However, this catalytic network did not lead to the self-replication of the original template.

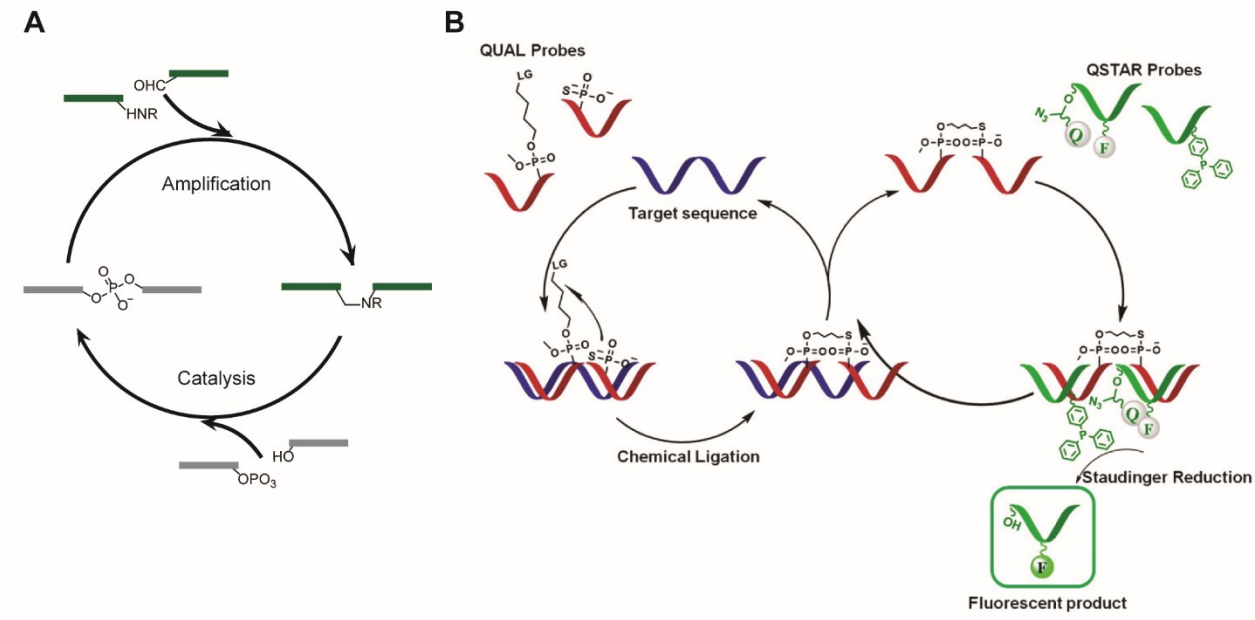


Figure 2.1 Schematic illustration of the destabilizing group-induced two-step oligonucleotide reaction. (**A**) Proposed two-stage replication cycle by Lynn and coworker.[139] Amplification to synthesize the ethylamine-containing backbone product template and the catalytic reaction to synthesize phosphate backbone of the original template. Adapted with permission from *Angew. Chem. Int. Ed.* **2000**, *39*, 3641-3643. Copyright © 2000 WILEY-VCH Verlag GmbH, Weinheim, Fed. Rep. of German. (**B**) Two-step DNA templated ligation reactions. The first reaction generated a destabilized product and the second reaction induced the release of a fluorescent quencher by a

Staudinger reduction.[140] Reprinted with permission from *J. Am. Chem. Soc.* **2017**, *139*, 15, 5405-5411. Copyright © 2017, American Chemical Society.

Thus far, the only successful self-replicating nonenzymatic oligonucleotide system that has been identified that exhibits sigmoidal growth indicative of exponential amplification is based on the ribozymes identified by the Joyce lab, which catalyze the formation of phosphodiester bonds; however, the replicating system requires a specific RNA sequence to confer this catalytic property.[68,141,142] Interestingly, the most facile replicators in this self-replicating ribozyme system did have destabilizing modifications between the template and the complementary fragments in the form of a U:G wobble pair.[68] These results suggested that destabilizing mismatches facilitated replication, although they decreased the fidelity of the replication process.

Previously our lab discovered an isothermal ligase chain reaction that exhibited rapid, exponential amplification of 18mer sequences using a common ligase enzyme and four fragment sequences, one of which contained an abasic model lesion in place of a complementary nucleotide at the 5'-phosphate terminus.[143,144] The temperature where exponential self-replication was observed was tunable by varying the stability of the fragment:template complexes through the introduction of mismatches,[145,146] revealing that self-replication was intrinsically linked to the stability of the DNA complexes formed throughout the catalytic cycle. This method of lesion-induced DNA amplification (LIDA) has proven general for amplifying a variety of 18mer sequences, usually requiring the introduction of only one abasic nucleotide into one of the four fragments.[143–146] Because of its sequence generality, LIDA might represent a model for understanding autonomous oligonucleotide self-replication. Yet all of our efforts of performing nonenzymatic self-replication using abasic destabilizing groups and rapid chemical ligation

methods yielded poor results (data not shown), suggesting that the enzyme was playing a more substantial role in our rapid self-replicating system.

To determine the role of the enzyme and identify principles for self-replication that can inform the design of autonomous, nonenzymatic oligonucleotide replication, here we utilize a kinetic model of cross-catalysis to identify the important kinetic and thermodynamic parameters in our ligase chain reaction of abasic containing strands. Remarkably, we find that the enzyme influences product inhibition by reducing the stability difference between the intermediate complex and the product duplex. This dual role of the enzyme in facilitating rapid ligation and overcoming product inhibition suggests that prebiotic replication might have required interactions with surfaces or aqueous complexes that bound the intermediate ternary complex more strongly than the product duplex, thereby overcoming product inhibition.

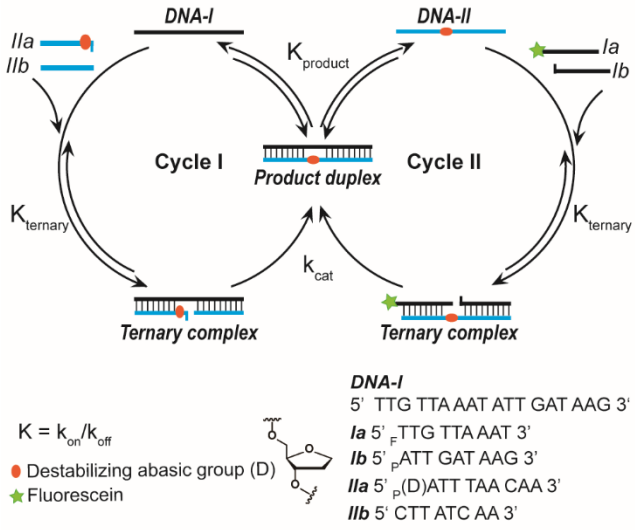


Figure 2.2 Scheme of lesion-induced DNA amplification (LIDA) describing the amplification of DNA-I strands at one temperature (30 °C). IIa$_P$ strands contain the destabilizing abasic group with 5' phosphate. The limiting reagent is labeled with fluorescein (F) on **Ia** strands. P represents a 5' phosphate group allowing the ligation reaction. T4 DNA ligase allows the ligation of ternary complexes. The replication of **DNA-I** is measured by the formation of fluorescein-labeled strands.

Our isothermal oligonucleotide self-replication scheme involves a cross-catalytic strategy where the product of one ligation reaction acts as the template of the other reaction (**Figure 2.2**). First, the native DNA template (**DNA-I**) is hybridized with two complementary fragments (**DNA-IIa$_P$** and **DNA-IIb** where p represents 5'-phosphate) through Watson-Crick base pairing to form ternary complexes where one of the fragments (**DNA-IIa$_P$**) contains the abasic group at the 5'-phosphate terminus. After ligation by T4 DNA ligase, the ternary complexes result in a product duplex consisting of the template (**DNA-I**) and the ligated product (**DNA-II**). Due to the destabilizing abasic group in the middle of the ligated product strand, the product duplex is able to dissociate at the same temperature where hybridization and ligation can occur, liberating the original template (**DNA-I**) and the new template (**DNA-II**). The two fragments (**DNA-Ia** and **DNA-Ib$_P$**) are also present in the mixture that are complementary to this newly generated template (**DNA-II**); therefore, they hybridize, are ligated, and dissociate, generating a new copy of the original target sequence (**DNA-I**), thereby triggering a cross-catalytic reaction. Our previous work showed that replacing the complementary thymine with an abasic group had a larger effect on the **DNA-I:DNA-II** product duplex than the ternary fragment complexes formed prior to ligation (**DNA-I**:**DNA-IIa**:**IIb** or **DNA-II:DNA-Ia:DNA-Ib**) based on the change in thermal dissociation or melting temperature (T_m).[143,144] However, even with the abasic group, the product duplex was much more stable based on its T_m than the ternary fragment complex, which would still result in product inhibition during replication. Herein, we identify the thermodynamic and kinetic parameters of LIDA, our rapid isothermal ligase chain reaction using a kinetic model of cross-catalysis and contrast these parameters to the stability of the DNA complexes in the absence of the ligase enzymes. Our results suggest that the presence of the abasic group brings the product and intermediate closer in stability and that the T4 DNA ligase helps to further overcome product inhibition by reducing the stability gap between the intermediate ternary complex and the product

duplex. Hence, thi work can aid the understanding of oligonucleotide self-replication, including the requirements for nonenzymatic self-replication.

2.2 Result and Discussion

2.2.1 Kinetic Profile of Lesion-Induced DNA Amplification (LIDA)

Prior to discussing our thermodynamic analysis and kinetic modeling, we will discuss typical results observed in our LIDA experiments under the conditions optimized for a given sequence. LIDA was performed using the fragment sequences shown in **Figure 2.2** in a 1:2:2:2 ratio, with the limiting fragment being the fluorescent-labeled fragment (**DNA-Ia**). The amount of fluorescent **DNA-I** formed from the **DNA-Ia** fragment was monitored versus time using polyacrylamide gel electrophoresis according to our previously reported method.[144] The resulting kinetic profile is shown in **Figure 2.5** for a reaction initiated with 0.01 equivalent **DNA-I** (14 nM) and a reaction with no initial **DNA-I**. The sigmoidal kinetic profiles are indicative of cross-catalysis: the initial exponential growth of product followed by a reduction in rate as the fragment pieces are consumed. At 30 °C, the optimal temperature for this 18mer sequence, the **DNA-I** initiated reaction consumed all of the limiting fragments within 45 minutes at these fragment and enzyme concentrations. Moreover, as shown in our earlier work,[144] we also observed a background-triggered process. This process is attributed to a pseudo-blunt end ligation of the four fragments in the absence of any **DNA-I** template. This ligation is not a true blunt end ligation as the duplex that results from hybridization of **Ia** and **IIa$_P$** has a 5'-abasic phosphate nucleotide, which should not interact with the 5'-adenosine phosphate single-base overhang from the **Ib$_P$**:**IIb** fragment duplex, hence the use of pseudo-blunt end. T4 DNA ligase is known to facilely perform blunt end ligation

of smaller DNA duplexes that contain the required 5'-phosphate modifications.[147,148] We have observed T4 DNA ligase triggers this background process in all of the sequences that we have successfully amplified by LIDA[144–146] suggesting the enzyme is also effective at ligating either at or across from abasic single-nucleotide overhangs.

2.2.2 Thermodynamic Parameters for DNA Complexes Formed During LIDA

The general idea for LIDA came from our observations of work in DNA-templated nonenzymatic ligation reactions using destabilizing linkers. These destabilizing groups, when positioned at the nick site of a ternary complex on either the nucleophilic or electrophilic termini, were shown to drive turnover in chemical (nonenzymatic) ligation suggesting that such destabilizing groups were more destabilizing when located in the middle of a duplex versus at the end (i.e. at the nick site).[38,42–44,139] The abasic group is known to destabilize DNA duplexes remarkably,[149,150] but prior to our work, it had not been reported as a means to generate turnover in the templated ligation of DNA. To determine whether the abasic group was indeed a good candidate for driving turnover in DNA-templated ligation reactions, we performed thermal denaturation experiments for some of the relevant complexes and found that the abasic site did appear to be more destabilizing for the product duplex and the ternary complexes based on the effects on the dissociation temperatures.[144] However, a rigorous thermodynamic analysis was not performed.

Thermal dissociation experiments are shown in **Table 2.1** of the relevant DNA complexes for the system explored in this chapter with the same buffer and salt concentration used in LIDA in 50 mM Tris-HCl buffer with 10 mM $MgCl_2$. For the samples prepared with a 1.2 μM concentration of each strand close to the concentrations used for our typical LIDA reactions, the

presence of the abasic group lowered the melting temperature of the product duplex notably (~12 ℃) compared to the ternary complex (~6 ℃) supporting the idea that the abasic group helped minimize product inhibition by bringing the stabilities of the ternary complex and product duplex closer together thereby facilitating turnover (**Table 2.1**).

Table 2.1 The melting temperatures of DNA complexes.

DNA Complex	T_m (°C)		
	0.6 µM	**1.2 µM**	**1.8 µM**
Product duplex (T) DNA-I: DNA-II(T)	49.84 ± 0.05	50.90 ± 0.03	51.52 ± 0.03
Product duplex (D = abasic) DNA-I: DNA-II(D)	37.2 ± 0.2	39.21 ± 0.04	39.2 ± 0.1
Ternary complex (T) Ia:Ib$_P$:DNA-II(T)	23.6 ± 0.5	26.1 ± 0.1	27.07 ± 0.06
Ternary complex (D = abasic) Ia:Ib$_P$:DNA-II(D)	19.2 ± 0.3	20.8 ± 0.1	22.2 ± 0.1
	1.2 µM	**1.8 µM**	**2.4 µM**
Intermediate complex (T) DNA-I:IIa$_P$(T)	23.11 ± 0.07	23.90 ± 0.05	24.67 ± 0.05
Intermediate complex (D = abasic) DNA-I:IIa$_P$(D)	21.3 ± 0.1	21.7 ± 0.1	22.3 ± 0.9
Intermediate complex (T) Ib$_P$:DNA-II(T)	24.5 ± 0.1	25.77 ± 0.08	27.07 ± 0.06
Intermediate complex (D = abasic) Ib:DNA-II(D)	18.5 ± 0.5	19.2 ± 0.3	19.6 ± 0.1
Background complex (D = abasic) Ia:IIa$_P$(D)	23.8 ± 0.2	24.8 ± 0.2	25.8 ± 0.1

Condition: 50 mM Tris-HCl buffer, pH 7.5 at 25 ℃, 10 mM $MgCl_2$.
The T_m was obtained from the Gaussian fit from the first derivative of absorbance at 260 nm over the increased temperatures.

To quantify the change in stability in terms of ΔH and ΔS, we followed the procedure of Marky and Breslauer that required measuring the dissociation temperature for DNA complexes formed at different concentrations.[151] The van't Hoff plot of $1/T_m$ versus the ln C_T with total DNA concentration (sum of all strands) C_T in molar units is shown in **Figure 2.3**. The change in enthalpy (ΔH) upon hybridization can be determined from the slope acquired from a linear fit to the data according to:

$$slope = \frac{(n-1)R}{\Delta H} \tag{2.1}$$

where *n* is the molecularity of the reaction, and R is the ideal gas constant. The y-intercept and ΔH determined from the slope can then be used to find the change in entropy (ΔS) according to:[151]

$$y - Intercept = \frac{(\Delta S - (n-1)R\ ln2n)}{\Delta H} \tag{2.2}$$

The hybridization equilibrium constants K_A at 30 °C, where the cross-catalytic self-replication is optimal for this sequence, were determined from the thermodynamic values using the following equation:

$$K_A = exp\left(\frac{-\Delta H}{RT} + \frac{\Delta S}{R}\right) \tag{2.3}$$

The product duplex with the abasic group in the middle showed a five order magnitude lower binding affinity ($K_A = 8 \times 10^7\ M^{-1}$) than the one with fully complementary strands containing the thymidine ($3.6 \times 10^{12}\ M^{-1}$), which agreed with the theoretical value determined using IDT OligoAnalyzer[152] ($5 \times 10^{12}\ M^{-1}$). However, the stability of the ternary complexes was not able to be calculated from this approach because this method assumed the binding of the DNA complex follows a two-state model, folded or unfolded, resulting in an underestimation of the hybridization process of three strands. Consequently, the intermediate complexes between the template hybridized with only one fragment were explored. The intermediate complex with the abasic group, either on the template or on the fragment, had only one or two orders of magnitude difference

depending on the position of the abasic group compared to the intermediate complex with the complementary T in place of the abasic site (K_A = 3 x 10^3 M^{-1} (with abasic) versus 4 x 10^5 M^{-1} or 8 x 10^4 M^{-1} (with thymine)) (**Table 2.2**). These values confirmed that the location of the abasic in the DNA duplex influenced the degree of destabilization much more significantly when located in the middle of the strand instead of at the terminus of a duplex. However, although the abasic group brought the stabilities much closer together of the product duplex and the intermediate complex compared with the strands containing the complementary thymidine, the K_A for the product and intermediate complexes still differed by five orders of magnitude (10^3 M^{-1}-intermediate with abasic versus ~10^8 M^{-1}-product with abasic), which would be expected to lead to product inhibition preventing turnover in DNA-templated ligation cycles.

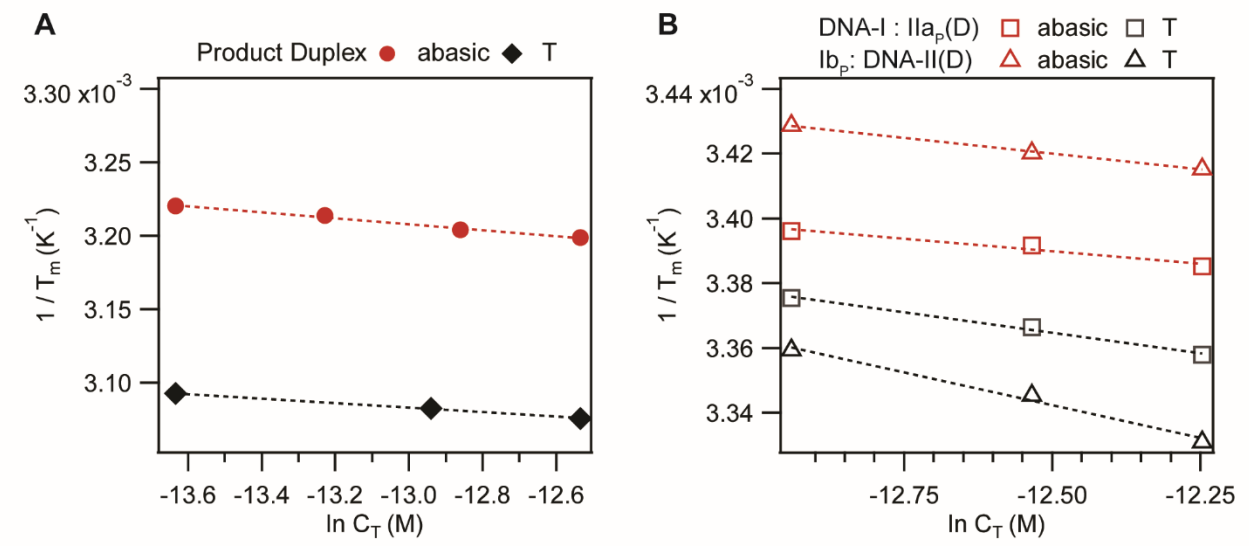


Figure 2.3 Thermodynamic data was plotted 1/Tm (K^{-1}) versus ln C_T (M). C_T indicates the total concentration of DNA strands. The plot shows the concentration dependence of the melting temperature. (**A**) Product duplex of **DNA-I:DNA-II**. The filled circle indicates the duplex containing the abasic group, and the diamond indicates the product duplex with thymidine. (**B**) Intermediate complex. Square indicates the template (**DNA-I**) hybridized with a fragment (**IIa_P**, 10 nt) with either abasic (red) or T (black). Triangle indicates a fragment (**Ib_P**) hybridized to the template (**DNA-II**) with either abasic (red) or T (black).

Table 2.2 Thermodynamic parameters of DNA complexes derived from the melting profiles

	ΔH (kJ•mol^{-1})	ΔS (kJ•mol^{-1}•K^{-1})	K_A (30 °C) (M^{-1})
Product Duplex - T	-570 ± 3	-1.646 ± 0.007	3.6 ± 0.2 x 10^{12}
Product Duplex - T (Theoretical*)	-560 ± 10	-1.62 ± 0.03	5 ± 1 x 10^{12}
Product Duplex - abasic	-440 ± 20	-1.29 ± 0.07	2 ± 1 x 10^{8}
Intermediate complex (T) DNA-I:IIa$_P$(T)	-330 ± 30	-1.00 ± 0.09	8 ± 10 x 10^{4}
Intermediate complex (T) Ib$_P$:DNA-II(T)	-210 ± 20	-0.57 ± 0.06	4 ± 3 x 10^{5}
Intermediate complex (D = abasic) DNA-I:IIa$_P$(D)	-540 ± 100	-1.7 ± 0.3	3 ± 20 x 10^{3}
Intermediate complex (D = abasic) Ib$_P$:DNA-II(D)	-430 ± 30	-1.34 ± 0.08	2 ± 4 x 10^{3}
Background complex (D = abasic)	-270 ± 20	-0.81 ± 0.07	4 ± 5 x 10^{5}

Theoretical values of T_m are obtained from the IDT analyzer. *Theoretical condition:* various concentrations of DNA complex (1.8 μM, 1.2 μM, 0.6 μM) in 10 mM NaCl and 10 mM $MgCl_2$.

2.2.3 Thermodynamic Parameters Determined from Isothermal Titration Calorimetry

To further verify these thermodynamic parameters, the DNA hybridization of the product and intermediate complexes was followed by isothermal titration calorimetry (ITC), which is widely used to measure the energy of molecular interactions based on the thermogram of heat differences during titration with one of the complementary strands.[153] Unlike the van't Hoff thermal denaturation method, a two-state is not required for this calorimetric method.[154] The ITC technique is often used to measure various binding energies in biology such as protein-DNA binding,[155–157] protein-protein interactions,[158] protein-ligand binding,[159] or oligonucleotide interactions.[160–163] ITC measures the heat difference at a single temperature when small volumes of high concentration of

the sample are injected into the solution in the cell that typically contains a low concentration of macromolecules. The heat measurements are corrected with a reference cell which contains a buffer blank. The integrated heat responses per injection are plotted with the molar ratio of the molecules to derive binding stoichiometry (N), ΔH, ΔS, and K_A. Initially, thermodynamic parameters were obtained from the ITC analysis of various DNA complexes at 25 °C. However, when the product duplex with thymidine was measured in the calorimeter at 25 °C, its binding affinity was six orders of magnitude lower than the value determined from the van't Hoff analysis from the T_m measurements. Mikulecky and Feig discussed that single strand stacking in oligonucleotides is a temperature-dependent value, affecting thermodynamic parameters determined at temperatures well below the dissociation temperature.[154,160,163] Accordingly, the binding affinity (K_A) measured by ITC at temperatures much lower than the T_m of the respective complexes did not match the values from van't Hoff approach. Therefore, the ITC experiments were performed again at about 5 °C lower than the complex's melting temperature to collect the ΔH and ΔS associated with the formation of product duplexes, ternary complexes, and intermediate complexes, neglecting the contribution of single-strand stacking at lower temperatures (**Table 2.3**). Compared to the thermodynamic parameters from those two approaches (**Table 2.2** and **Table 2.3**), both product duplexes, including that with either thymidine or the abasic group, exhibited similar binding affinities at 30 °C with the thermal dissociation or calorimetric method, (3.6×10^{12} M^{-1} vs 1.98×10^{12} M^{-1}, respectively, for thymidine, 2×10^{8} M^{-1} vs 1.03×10^{8} M^{-1}, respectively, for the abasic). This result confirmed once more that an abasic group in the DNA system lowered the stability of the product duplex considerably. Moreover, the binding affinity of the intermediate complexes from the ITC experiments was in the same order of magnitude as that evaluated for the ternary complexes, indicating that the ITC method is an adequate approach to measure the stability of the ternary complexes. The ternary and intermediate complexes with complementary bases (10^5-10^4

M^{-1}) were more stable than those possessing the abasic group (10^4 - 10^3 M^{-1}), as reflected by their different T_m values. To sum up, the thermodynamic parameters revealed that the presence of the abasic group reduced the stability gaps between the association constants of the product duplex and intermediate complexes considerably, from 10^8 M^{-1} to 10^3-10^5 M^{-1} for the abasic system compared with 10^{12} M^{-1} to 10^4-10^5 M^{-1} for that with the complementary thymidine.

Table 2.3 Isothermal titration calorimetry parameters of DNA complexes

	℃	N	ΔH ($kJ•mol^{-1}$)	ΔS ($kJ•mol^{-1}•K^{-1}$)	K_A (30 °C) (M^{-1})
Product Duplex DNA-I:DNA-II(T)	25 ℃	0.845 ± 0.002	-592 ± 3	-1.83	1.8 ± 0.1 x 10^6
	45 ℃	0.972 ± 0.002	-608 ± 3	-1.77	1.98 ± 0.07 x 10^{12}
Product Duplex (D = abasic) DNA-I:DNA-II(D)	25 ℃	1.146 ± 0.004	-494 ± 3	-1.52	3.8 ± 0.3 x 10^5
	35 ℃	1.151 ± 0.008	-444 ± 4	-1.31	1.03 ± 0.09 x 10^8
Ternary Complex Ia:Ib_P:DNA-II(T)	25 ℃	1.403 ± 0.006	-379 ± 2	-1.15	2.5 ± 0.2 x10^5
Intermediate Complex Ib_P:DNA-II(T)	25 ℃	1.01 ± 0.01	-227 ± 3	-0.65	2.0 ± 0.2 x 10^5
Intermediate Complex DNA-I:IIa_P(T)	25 ℃	0.95 ± 0.04	-320 ± 20	-0.98	4 ± 3 x 10^4
Ternary Complex (D = abasic) Ia:Ib_P:DNA-II(D)	25 ℃	0.82 ± 0.01	-288 ± 6	-0.86	7 ± 1 x 10^4
Intermediate Complex (D = abasic) Ib_P:DNA-II(D)	20 ℃	0.80 ± 0.08	-350 ± 40	-1.10	1 ± 3 x 10^3
Intermediate Complex (D = abasic) DNA-I:IIa_P(D)	20 ℃	0.99 ± 0.01	-152 ± 3	-0.39	2.7 ± 0.3 x 10^5
	25 ℃	1.05 ± 0.05	-210 ± 10	-0.60	6 ± 3 x 10^4

Condition: 50 mM Tris-HCl buffer, pH 7.5 at 25 ℃, 10 mM $MgCl_2$.
N is the number of binding sites

2.2.4 Kinetic Fitting Parameters for DNA Complexes

Rasmussen and co-workers recently performed an analysis of our published LIDA results to model autonomous self-replication and proposed that the ΔΔG between the product duplex and ternary complex had to be -8 kcal/mol or less to observe facile self-replication.[164] However, the authors did not speculate on how LIDA achieved this requirement. Our thermodynamic analysis suggests that the product duplex and the intermediate complex for the abasic system had ΔΔG values on the order of -15 kcal/mol to -28 kcal/mol at 30 °C. However, in our thermodynamic analysis, we focused only on the formation of the complexes in the absence of the enzyme. We hypothesize that the enzyme helps decrease the difference in ΔΔG or the affinity constant K_A between the product and nicked duplex, thereby overcoming inhibition. To test our hypothesis, we utilized kinetic modeling of a cross-catalytic system. Although we used a simplified kinetic model that does not explicitly include the enzyme, it is worth first reviewing the T4 DNA ligase-catalyzed ligation mechanism of a nicked duplex.

DNA ligation by T4 DNA ligase occurs in three steps.[148,165] First, the amine functional group from a lysine residue on T4 DNA ligase reacts with ATP to form a ligase-AMP (adenosine monophosphate) intermediate by releasing pyrophosphate (PPi). Next, the adenylated enzyme binds to the nicked DNA complex and transfers the AMP to the 5' phosphate strand. The ligation is completed by a 3' hydroxy group that reacts with the 5' terminus by releasing AMP. We assume the enzyme then dissociates, followed by product duplex dissociation.

Consequently, the apparent hybridization K_A values exhibited during LIDA should be a convolution of the intrinsic hybridization constant and the binding constant of the enzyme for the complex. To determine these apparent binding constants and compare them with our previous

thermodynamic analysis, a kinetic model was used to obtain the individual rate constants that constitute the affinity constants of the different complexes as well as an effective rate constant for the ligation step of our self-replication system. Unlike our previous work, two different lengths of fluorescein-labeled strands (**DNA-Ia(+2)** and **DNA-IIb**) were used to measure both *in situ* template formation **DNA-I(+2)** (20 nt) and the destabilizing product formation **DNA-II** (18 nt). The formation of both **DNA-Ia(+2)** and **DNA-II** was confirmed by polyacrylamide gel electrophoresis based on the length difference of the two products, and the formation of each product versus time was determined (**Figure 2.4**).

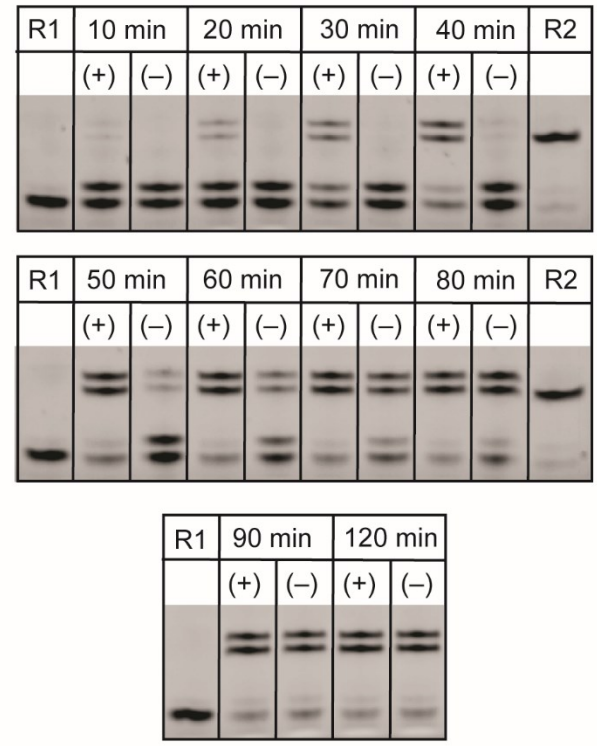


Figure 2.4 Cross-catalytic reaction of two fluorescein-labeled strands. Representing polyacrylamide gel images of cross-catalytic reaction at 30 °C. (+) represents initial 14 nM template DNA-I was added, (-) represents no initial template was added. R1 represents unreacted IIb with fluorescein labeling while R2 indicates the formation of DNA-II from the ligation of IIa_p with IIb with fluorescein labeling. Experimental conditions: 1.4 μM of DNA-Ia with extended two bases (11 nt), 2.8 μM of DNA-Ib_p, DNA-IIa_p which contains the abasic group, and DNA-IIb with

fluorescein labeling. 2000 CEU T4 DNA ligase (1 µL) was used in the reaction (total volume of 15 µL). 50 mM Tris-HCl, 10 mM $MgCl_2$, and 1 mM ATP of the ligation buffer (New England Biolab) were used.

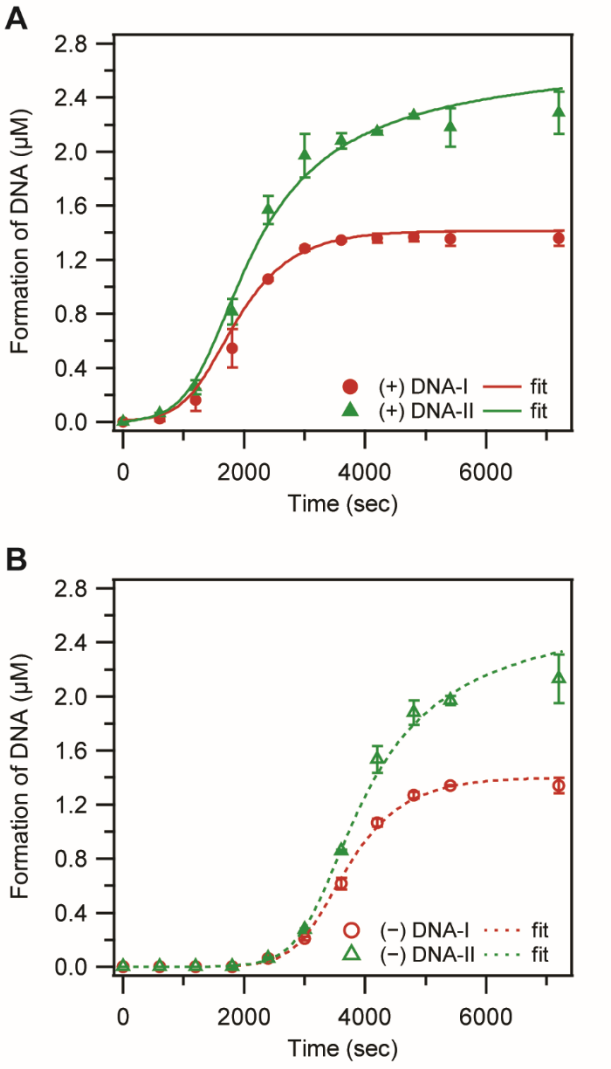


Figure 2.5 Cross-catalytic replication of F-DNA-I or F-DNA-II at 30 °C with 2000 CEU T4 DNA ligase. (**A**) initiated by 14 nM or (**B**) 0 nM DNA-I (initial template). The red circle represents the formation of **F-DNA-I** while the green triangle represents the formation of **F-DNA-II**. The data were fitted with KinTek Explorer. *Experimental conditions*: 1.4 µM of **DNA-Ia** with extended two bases (11 nt), 2.8 µM of **DNA-Ib$_p$**, **DNA-IIa$_p$** which contains the abasic group, and **DNA-IIb** with fluorescein labeling. 2000 CEU T4 DNA ligase (1 µL) was used in the reaction (total volume of 15 µL). A 50 mM Tris-HCl buffer containing 10 mM $MgCl_2$, and 1 mM ATP (New England Biolab) were used.

Next, the data were analyzed by KinTek Explorer to obtain the kinetic parameters of our self-replication system (**Figure 2.5** and **Table 2.4**). The following reaction equilibria and irreversible steps were used to generate the model in KinTek Explorer:

$$\mathrm{DNA-I} + \mathrm{IIa_P} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{DNA-I \cdot IIa_P} \quad (2.4)$$

$$\mathrm{DNA-I \cdot IIa_P} + \mathrm{IIb} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{DNA-I \cdot IIa_P \cdot IIb} \quad (2.5)$$

$$\mathrm{DNA-I \cdot IIa_P \cdot IIb} \xrightarrow{k_{cat}} \mathrm{DNA-I \cdot DNA-II} \quad (2.6)$$

$$\mathrm{DNA-I} + \mathrm{DNA-II} \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} \mathrm{DNA-I \cdot DNA-II} \quad (2.7)$$

$$\mathrm{DNA-II} + \mathrm{Ia} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{DNA-II \cdot Ia} \quad (2.8)$$

$$\mathrm{DNA-II \cdot Ia} + \mathrm{Ib_P} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{DNA-II \cdot Ia \cdot Ib_P} \quad (2.9)$$

$$\mathrm{DNA-II \cdot Ia \cdot Ib_P} \xrightarrow{k_{cat}} \mathrm{DNA-II \cdot DNA-I} \quad (2.10)$$

$$\mathrm{IIa_P} + \mathrm{IIb} \xrightarrow{k_{bi}} \mathrm{DNA-II} \quad (2.11)$$

The results from a good fit to the data are shown in **Table 2.4** (goodness of fit parameter Chi2/DoF = 13.6). As will be discussed, we performed additional experiments to determine if the k_{cat} and k_{bi} values, signifying rate constants for the templated ligation, and background pseudo-blunt end ligation observed in the fit were reasonable. As with any fitting procedure that has numerous fit parameters, multiple non-unique fits can describe our system. However, the general agreement with either measured or reported ligation rate constants suggests our fit is reasonable.[166] Additionally, we will also show in simulations how changing the most important parameters that affect the affinity constants have a drastic impact on the self-replication curve suggesting that our

modeling can at least capture the proper order of magnitude of the affinity constants for all of the intermediates and the product duplex.

To minimize the number of fit parameters, we assumed that the formation of the ternary complexes for both cycles occurred in two steps, each step with the same forward and reverse rate constant (k_{+1} and k_{-1}, respectively). The ratio of these rate constants is the affinity constant for the intermediate ($k_{+1}/k_{-1} = K_A$), which we can compare to the affinity constant of the product ($k_{+2}/k_{-2} = K_A$). Additionally, as will be described in 2.3.5, only DNA-II formation occurred via a background reaction by a non-templated ligation, not the formation of **DNA-I**. Hence, we constrained the background reaction between **Ia** and $\mathbf{Ib_P}$ to have a rate constant value of 0 in our kinetic model.

The rate constant for the ligation steps from the kinetic model fit (k_{cat}) was 0.093 ± 0.005 s^{-1}, which is of similar magnitude to the reported k_{cat} (0.57 ± 0.11 s^{-1}) for the ligation of DNA strands with T4 DNA ligase albeit without the presence of the 5'-phosphate abasic group.[166] The background reaction induced by pseudo-blunt end ligation of $\mathbf{DNA\text{-}IIa_P}$ and **IIb** yielded a rate constant value of 2 x 10^{-7} s^{-1} from the fit of the traces in the absence of initial **DNA-I** consistent with a slow process. The kinetic fit yielded an association constant for the intermediate complexes for $\mathbf{DNA\text{-}I \cdot IIa_p}$ and $\mathbf{DNA\text{-}II \cdot Ib_P}$ of 1.05 ± 0.05 x 10^5 M^{-1}. These affinity constants were slightly higher than that determined in the ITC analysis for the intermediate complexes (6 ± 3 x 10^4 M^{-1}) albeit close within error, suggesting the presence of the enzyme only slightly stabilizes these complexes. In contrast, the affinity constant determined for the product duplex from the fit with the kinetic model was 9 ± 7 x 10^5 M^{-1}, which was much lower than that determined from the ITC analysis (K_A = 1.03 ± 0.09 x 10^8 M^{-1}). The difference in affinity constants between the intermediate and produce complexes determined from the kinetic model results in ΔΔG as -5 kcal/mol from the

fit, which is lower than the value determined from thermodynamic measurements in the absence of enzyme (-15 ~ -28 kcal/mol). This result suggests the enzyme compensates for the stability gaps between the ternary complex and product duplex, helping to overcome product inhibition based on its affinity for the hybridized complexes. This argument is consistent with a study from Lohman and coworkers' where T4 DNA ligase was shown to bind more strongly to the ternary complex over the product duplex ($K_{net,nick}$ = 2.5 - 3.3 x 10^8 M^{-1} and $K_{net,duplex}$ = 2.8 - 7.25 x 10^6 M^{-1}).[166] Indeed, the difference in the ratio of the K_A values in the presence of enzyme was three orders of magnitude less than the K_A in the absence of enzyme, consistent with the two orders of magnitude difference in $K_{binding}$ for the enzyme with nicked versus produce duplex. Our fit suggests the dominant contribution is a destabilizing effect of the enzyme on the product duplex based on the lower K_A in the presence of the enzyme compared with the value for the product duplex in the absence of enzyme. Yet it is unclear whether that apparent destabilizing effect is real or just a consequence of our simplified kinetic model. Accordingly, we next explored the sensitivity of our model to the relative affinities of the intermediate and product duplex as well as to the individual binding constants associated with intermediate or product hybridization.

Table 2.4 Kinetic parameters of DNA complexes

Cycle 1	**DNA-I + IIa$_P$ ⇋ DNA-I • IIa$_P$** **DNA-I • IIa$_P$ + IIb ⇋ DNA-I • IIa$_P$ • IIb**	**k_{+1}**	5.0 ± 0.1 x 10^4 $M^{-1}s^{-1}$
		k_{-1}	0.48 ± 0.02 s^{-1}
		k_{+1}/k_{-1}	**1.05 ± 0.05 x 10^5 M^{-1}**
	DNA-I$_P$ • IIa • IIb → DNA-I • DNA-II	**k_{cat}**	**0.093 ± 0.005 s^{-1}**
Product Duplex	**DNA-I + DNA-II ⇋ DNA-I • DNA-II**	**k_{+2}**	4 ± 3 x 10^4 $M^{-1}s^{-1}$
		k_{-2}	0.05 ± 0.08 s^{-1}
		k_{+2}/k_{-2}	**9 ± 7 x 10^5 M^{-1}**
Cycle 2	**DNA-II + Ia ⇋ DNA-II • Ia** **DNA-II • Ia + Ib$_P$ ⇋ DNA-II • Ia • Ib$_P$**	**k_{+1}**	5.0 ± 0.1 x 10^4 $M^{-1}s^{-1}$
		k_{-1}	0.48 ± 0.02 s^{-1}
		k_{+1}/k_{-1}	**1.05 ± 0.05 x 10^5 M^{-1}**
	DNA-II • Ia • Ib$_P$ → DNA-II • DNA-I	**k_{cat}**	**0.093 ± 0.005 s^{-1}**
Background	**IIa$_P$ + IIb → DNA-II**	**k_{bi}**	**2 x 10^{-2} $M^{-1}s^{-1}$**
	Ia + Ib$_P$ → DNA-I	**k_{bi}**	–

DNA-I: 5'- TTG TTA AAT ATT GAT AAG - 3'
Ia 5' $_F$TA TTG TTA AAT 3'
Ib$_P$ 5' $_P$ATT GAT AAG 3'

DNA-II: 5'- CTT ATC AA(D) ATT TAA CAA - 3'
IIa$_P$ 5' $_P$(D)ATT TAA CAA 3'
IIb 5' $_F$T CTT ATC AA 3'

To evaluate how the stability of the intermediate and product duplexes affected the kinetics of self-replication in our model, simulations were performed whereby the rate constants were systematically increased or decreased by a factor of 10 from the values determined from the best fit. In **Figure 2.6A-D**, this resulted in an increase or decrease in the intermediate or product duplex K_A values by one order of magnitude. When the stability of the intermediate complex was one order of magnitude higher by changing either 10 x k_{+1} or 0.1 x k_{-1}, the formation of **DNA-I** and **DNA-II** was much faster (**Figure 2.6A, 2.6C**). However, when the stability of the intermediate complex was reduced by one order of magnitude (0.1 x k_{+1} or 10 x k_{-1}), self-replication was not observed (**Figure 2.6A 2.6C**). In contrast, when the product duplex formation had an affinity

constant of 10 x K_A (10 x k_{+2} or 0.1 x k_{-2}), the self-replication of **DNA-I** and **DNA-II** was only slightly slower (**Figure 2.6B, 2.6D**). Reducing the affinity constant of the product by 10-fold also only slightly increased the rate of **DNA-I** and **DNA-II** formation (0.1 x k_{+2} or 10 x k_{-2}). In addition, simulations were conducted with 5 x k_{cat} and 0.2 k_{cat} to evaluate the sensitivity of the model to the rate of ligation. When the k_{cat} is 5 times faster than the value determined from our best fit of the data, the self-replication was much faster than what we observed. Conversely, the ligation rate was much slower when k_{cat} was reduced by a factor of (0.2 x k_{cat}), resulting in slow amplification of the templates (**Figure 2.6E**). These simulations reveal that self-replication is much more sensitive to the intermediate stability rather than that of the product duplex, suggesting that nonenzymatic self-replicating systems should focus on means to stabilize the intermediate moreso than destabilize the product duplex to facilitate rapid self-replication. Furthermore, the effective rate constant for ligation also has a significant impact on the kinetics of self-replication, suggesting that rapid ligation reactions are required to generate appreciable self-replication.[164] For example, very little product was observed within the first 7200 seconds (2 hours) of reaction when k_{cat} was reduced to 0.1k_{cat} (simulation not shown).

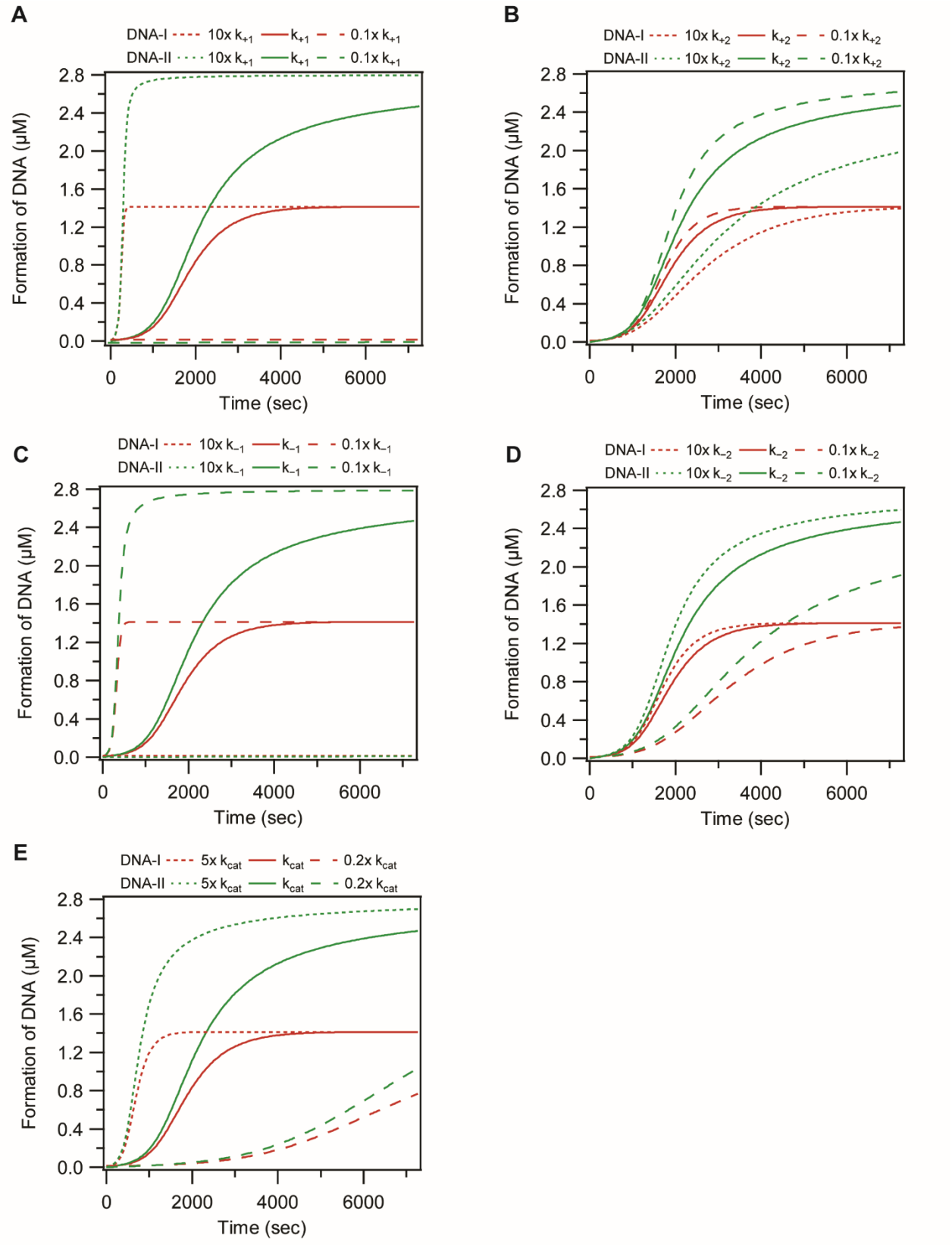


Figure 2.6 Simulation of changing the k values of the DNA complex. The red color indicates the formation of DNA-I and the green color indicates the formation of DNA-II. The solid lines

were the fit to the actual measurement. Smaller dashed lines indicate simulated data when the K_A of the DNA complex was forced to have one order higher magnitude while bigger dashed lines represent the simulated data when the K_A of the DNA complex was forced to have one order lower magnitude. (**A**) Simulation of changing the k_{+1} values. Original values: k_{+1} (5.0 x 10^4 $M^{-1}s^{-1}$) forced 10x value: k_{+1} (50 x 10^4 $M^{-1}s^{-1}$). Forced 0.1x value: k_{+1} (0.5 x 10^4 $M^{-1}s^{-1}$). (**B**) Simulation of changing the k_{+2} values. Original k_{+2} value (4 x 10^4 $M^{-1}s^{-1}$), simulated 10x k_{+2} value (40 x 10^4 $M^{-1}s^{-1}$), simulated 0.1x k_{+2} value (0.4 x 10^4 $M^{-1}s^{-1}$) (**C**) Simulation of changing the k_{-1} values. Original values: k_{-1} (0.48 s^{-1}) forced 10x value: k_{-1} (4.8 s^{-1}). Forced 0.1x value: k_{-1} (0.048 s^{-1}). (**D**) Simulation of changing the k_{-2} values. Original k_{-2} value (0.05 s^{-1}), simulated 10x k_{-2} value (0.5 s^{-1}), simulated 0.1x k_{-2} value (0.005 s^{-1}). (**E**) Simulation of changing the k_{cat} values. Original k_{cat} (0.009 s^{-1}), simulated 5x k_{cat} (0.5 s^{-1}) and simulated 0.2x k_{cat} value (0.02 s^{-1}).

2.2.5 Michaelis-Menten Parameters for DNA Complexes

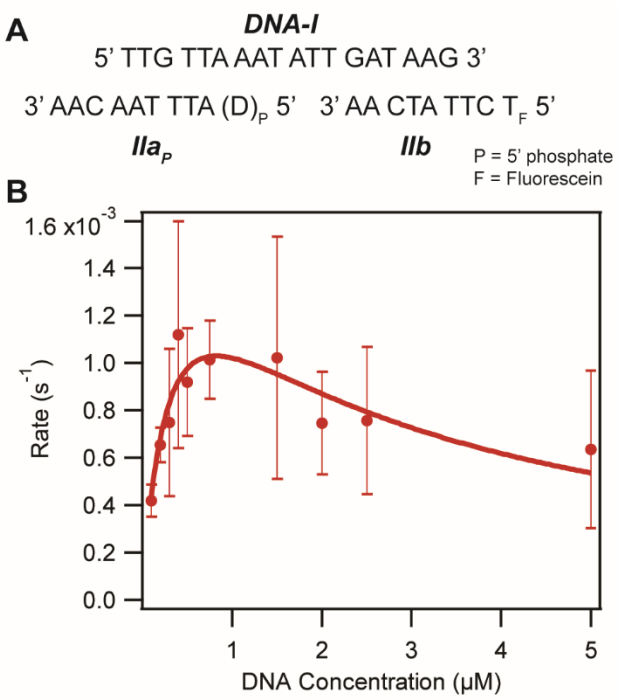
Thus far, we have rationalized the decrease in relative K_A for the intermediate versus product duplex as arising from the enzyme owing to binding interactions between the enzyme and the hybridized complexes. We attempted to measure K_{net} for the abasic-containing complexes according to the methods from Lohman and coworkers of rapid time scale analysis using stopped-flow DNA ligase single turnover and binding assay[166] to verify that the relative affinities for the nick and the product duplex were maintained or perhaps even enhanced with the presence of the abasic group, but were unsuccessful as no change in fluorescence was observed in the required stop-flow measurements upon enzyme-substrate binding. Therefore, we decided to determine the effect of the abasic group on the Michaelis-Menten parameters k_{cat} and K_m for cycle I for the system with and without the abasic group to provide some insight as to how the presence of the abasic affected the interactions with the enzyme. These parameters were determined by measuring the initial rates of the templated ligation of the two fragments. The initial ligation rate was evaluated from linear fits to the stoichiometric templated reactions over the range of fluorescein-labeled ternary (nicked) complexes **DNA-I • IIa$_p$ • IIb** from 0.1 μM to 5 μM reacting with 25 nM T4

DNA ligase (**Figure 2.7**). The series of kinetic experiments at different substrate concentrations were repeated at least three times, and the average observed rate constant (initial rate/enzyme concentration) with the corresponding standard deviation are plotted versus substrate concentration (**Figure 2.7**). Previously, studies revealed that T4 DNA ligase and other enzymes exhibited a notable reduction in the initial rate at high concentrations of the ternary DNA complex, indicating that the substrate inhibited the reaction.[167] We observed a similar decrease in the rate V_0 with increasing substrate concentration for the abasic system. Therefore, we utilized a Michael-Menten model with uncompetitive substrate inhibition to interpret our data:[167]

$$\frac{V_0}{[E]} = \frac{k_{cat} * [S]}{K_m + [S](1+\frac{[S]}{K_i})} \tag{2.12}$$

where K_m is the Michaelis constant related to the K_D of enzyme-substrate binding, [S] is the DNA substrate concentration, [E] is the enzyme concentration, V_0 is the initial rate, k_{cat} is the turnover number, and K_i is the inhibition dissociation constant for the substrate with the enzyme-substrate complex. The reported Michaelis-Menten values are those from the fit of the average V_0 data with the standard deviation returned from the fit. Although the error bars for the averaged initial rates are substantial, they and the general quality of the fits are similar to that observed by Lohman and co-workers for similar experiments done at lower enzyme and substrate concentrations.[167] In the presence of the abasic group, the K_m value was 0.3 ± 0.2 μM while with the complimentary T base K_m was 0.7 ± 0.4 μM, which implies that T4 DNA ligase has a similar affinity for the ternary complex with and without the abasic group. The substrate inhibition (K_i) constants were also similar for the abasic-containing system (2 ± 1 μM) and that with the thymidine (1.1 ± 0.7 μM). However, k_{cat} was higher for the system with the T base, suggesting that the ligation is faster with the 5'-phosphate thymidine (0.3 ± 0.1 s^{-1}) compared with the 5'-phosphate abasic (0.07 ± 0.02 s^{-1})

(**Figure 2.7C**). Previous reports showed that T4 DNA ligase could ligate 5' end abasic site albeit with reduced reaction rates[168] and reseal the abasic site that was generated by AP endonuclease.[169] These parameters suggest that the abasic-containing system has similar enzyme affinity and inhibition behavior as that of the thymidine-containing system, although it appears that the presence of the abasic group does decrease the rate constant for the ligation step.



D

	abasic	T
k_{cat} (s^{-1})	0.07 ± 0.02	0.3 ± 0.1
K_m (μM)	0.3 ± 0.2	0.7 ± 0.4
K_i (μM)	2 ± 1	1.1 ± 0.7

Figure 2.7 Michaelis-Menten parameters for abasic ternary complex versus complementary (T) base ternary complex. The data obtained was the ligation reaction of various concentrations (0.1 μM to 5 μM) of **DNA-I** • **IIa_p** • **IIb** at 16 °C. T4 DNA ligase concentration was 25 nM. Initial rates were fitted first 120-sec data. (**A**) The sequence information (**B**) The abasic complex of **DNA-**

I • IIa$_p$ • IIb. (**C**) The complementary (T) complex of **DNA-I • IIa(T)$_p$ • IIb**. (**D**) K_m, $k_{cat,}$ and K_i value comparison between the abasic and complementary base(T). The data were averaged of a minimum of three replicated data shown in **Figure 2.17** and **Figure 2.18**.

2.2.6 Background Reaction Triggered by Pseudo-blunt End Ligation

To determine what triggers the non-templated background reaction, we investigated the pseudo-blunt-end ligation reaction (a 5'-abasic and 5'-deoxyadenosine single "base" overhang) using four fragments where only one of them has the 5'-phosphate, which enables ligation (**Figure 2.8**). The pseudo-blunt-end ligation reactions were conducted at 8 ℃ and 12 ℃ with four fragments (**IIa**, **IIb**, **Ia**, and **Ib**) without any addition of the full-length template (18 nt). In cycle 1, the 5' phosphate was located on the abasic-containing strand (the 10mer **IIa$_p$**), while in cycle 2, the 5' phosphate presented on the deoxyadenosine terminated strand (the 9mer **Ia$_p$**). The result shows that only the formation of **DNA-II** in cycle 1 occurred by this pseudo-blunt end process, as the formation of **DNA-I** in cycle 2 was not observed. These results imply the background reaction is initially triggered by **DNA-II** formation based on ligation of the 5'-abasic phosphate strands. Therefore, during the fitting and simulation, the formation of **DNA-II** was only considered, as mentioned in 2.2.3.

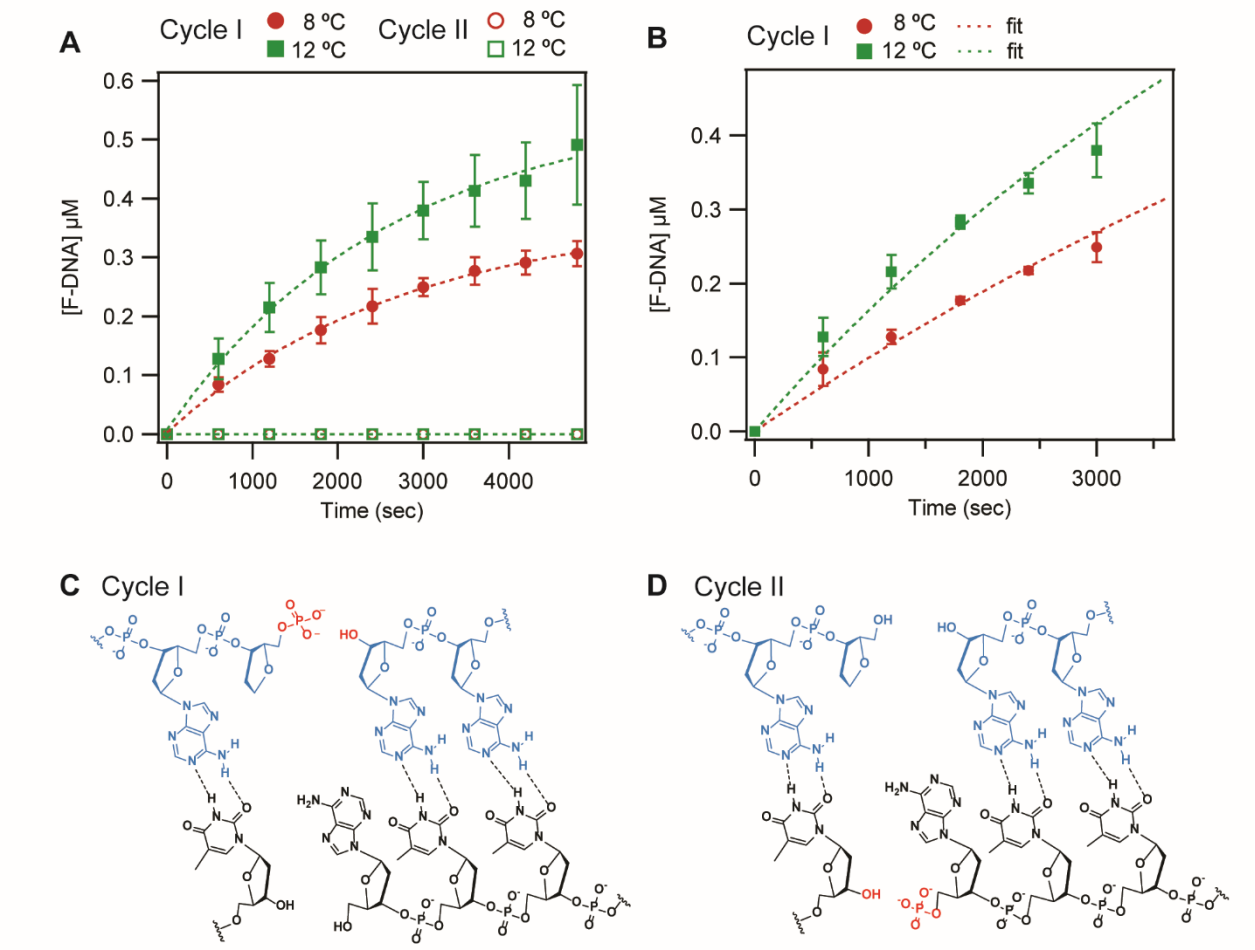


Figure 2.8 Background trigger reaction. (**A**) Pseudo-blunt-end ligation at 8 °C and 12 °C with 2000 CEU T4 DNA ligase with four fragments (1.4 µM or 2.8 µM, 1:2:2:2). The red circle indicates the reaction at 8 °C, and the green squares indicate the reaction at 12 °C. The filled circle and square represent cycle 1 while the empty circle and square represent cycle 2. (**B**) Fitted graph with KinTek Explorer with biomolecular binding until 3600 sec. (**C**) Cycle 1 consists of a 5' phosphate abasic strand (**IIa$_p$**), fluorescein-labeled strand (**IIb**), and two short template strands (**Ia, Ib**). (**D**) Cycle 2 contains 5' phosphate strand (**Ib$_p$**), fluorescein-labeled strand (**Ia**), and complementary strands (**IIb**), one including abasic group (**IIa**).

To quantify the reactivity of this background triggered (i.e., non-templated) reaction, an excess of the enzyme was used to catalyze the reaction ([Limiting substrate] = 1.4 µM, [Substrate] = 2.8 µM and [Enzyme] = 2.3 µM). The bimolecular binding (A + B → AB) was used to fit the non-templated reaction, which assumed that two hybridized fragments are one substrate since the

fragments are assumed to be hybridized at this low temperature (i.e. DNA-Ia:IIa as substrate A and DNA-Ib:IIb as substrate B). The observed rate constant (k_{bi}) was determined by fitting the data with the reaction equilibria below.

$$A + B \xrightarrow{k_{bi}} AB \qquad (2.12)$$

The k_{bi} from the fit was 2.7 ± 0.7 x 10 $M^{-1}s^{-1}$ at 8 °C and 5 ± 2 x 10 $M^{-1}s^{-1}$ at 12 °C. The k_{bi} rate constant (2 x 10^{-2} $M^{-1}s^{-1}$) for the pseudo-blunt end reaction in the kinetic model of LIDA is much smaller than that observed from the fit of the data in **Figure 2.8B**. This difference is attributed to the convolution between the ligation step and formation of the fragment duplexes in our simplified kinetic model of LIDA, which should lead to a much lower rate constant with the kinetic model owing to the low stability of the fragment duplexes at 30 °C. For example, the melting temperature of **Ia:IIa$_P$** is about 24 °C (**Table 2.1**), and we assume **Ib:IIb** has a similar melting temperature. In other words, k_{bi} from the fit took into account the hybridization of the short fragments, while the calculated rate constant from our temperature-dependent experiments corresponded to just the ligation step. In addition to its relevance to understanding the mechanism of LIDA, these results reveal that T4 DNA ligase tolerates a single 5'-phosphate abasic overhang and readily ligates it in a non-templated blunt-end like reaction. However, this enzyme will not ligate across from a single-base overhang containing an abasic group.

2.2.7 The Roles of Temperature and T4 DNA Ligase Concentration in Lesion-Induced DNA Amplification

Previous work had found that the kinetics of LIDA was highly temperature-dependent.[144] Here, we expanded the temperature range and explored both the templated and background-triggered LIDA. While previous studies were focused on finding the optimal temperature for the best difference in templated and non-templated reaction,[144] we explored higher temperatures to understand the role of the enzyme. Although 30 ℃ yielded the fastest rate of self-replication while maintaining the greatest discrimination between the DNA-I initiated and background-triggered profiles, the self-replication at 34 °C and 38 ℃ showed much faster replication (**Figure 2.9**). However, no reaction was observed at 42 °C, revealing that this temperature completely shut down replication. The melting temperature of the intermediate complex was 19 – 21 °C, which is 20°C lower than this elevated temperature. Therefore, we reason that the affinity constant for the intermediate complex is too low at this temperature (42 °C), leading to the abrupt termination of self-replication based on our simulations above. At 38 °C, the affinity constant of the intermediate should also be reduced; however, this might be compensated for by an increase in k_{cat} as well as a decrease in the affinity of the product duplex. It also appears that the rate of the background triggered process is increased at the higher temperatures, which we also attribute to faster ligation by the enzyme despite the lower stability of the probe fragment duplex at higher temperatures.

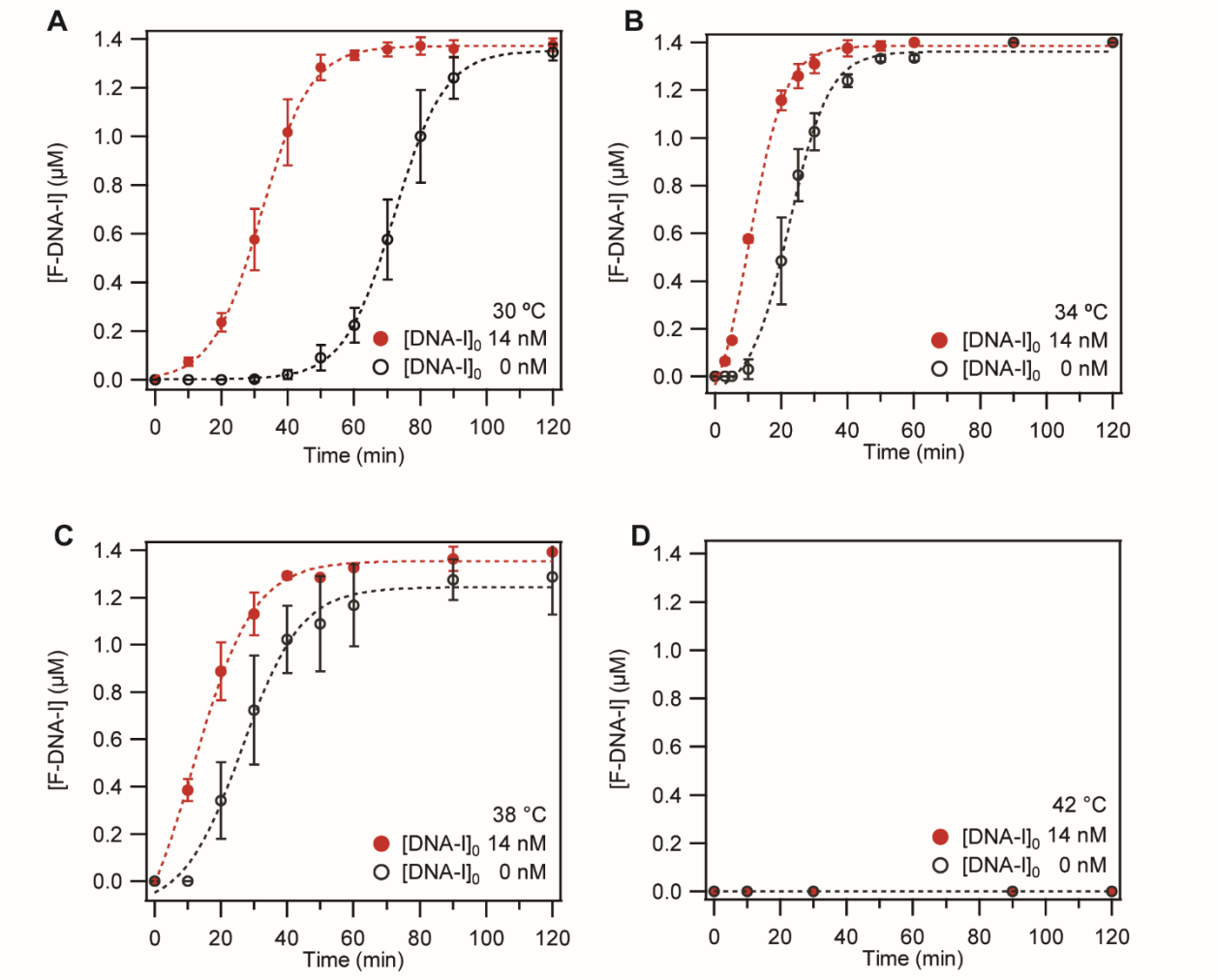


Figure 2.9 Temperature variations on the cross-catalytic reactions. (**A**) 30 °C, (**B**) 34 °C, (**C**) 38 °C and (**D**) 42 °C. The red circle indicates the experiments with an initial 14 nM **DNA-I** while the black empty circle indicates the experiments without any initial **DNA-I**. *Experimental conditions*: 1.4 μM of **DNA-Ia**, 2.8 μM of **DNA-Ib$_p$**, **DNA-IIa$_p$** which contains the abasic group, and **DNA-IIb**. 2000 CEU T4 DNA ligase (1 μL) was used in the reaction (15 μL). 50 mM Tris-HCl buffer (New England Biolab) containing 10 mM $MgCl_2$, and 1 mM ATP were used.

To test the catalytic activity of the enzyme, we varied the concentrations of the enzyme and examined the efficiencies of the self-replication with enzyme concentrations that were lower than the substrate concentrations. Previous experiments showed that the concentration of the enzyme had a major effect on the turnover in the single-cycle ligation reactions (cycle II).[143,144] More rapid

self-replication was then achieved with 2000 cohesive end unit (CEU) T4 DNA ligase, whereas lower rates of ligation were observed with five Weiss unit enzymes, albeit from a different supplier.[143,144] In addition to arising from different suppliers, these enzymes used different ligation buffers without and with the presence of PEG, respectively. Accordingly, the direct comparison of the enzyme concentration with the exact same batch and buffer had not been performed yet. Here, the cross-catalyst reactions with various concentrations of the enzyme were evaluated by dilution of 2000 CEU enzymes using the diluent buffer recommended by the company. The enzyme concentration of 2000 CEU T4 DNA ligase was considered as 35 ± 2 μM based on the reports from the company.[170] Therefore, the 2000 CEU T4 DNA ligase that was used in LIDA was determined as 2.3 μM. As shown in **Figure 2.10**, the rate of formation of the template **DNA-I** in lower concentrations 1200 CEU (1.4 μM) and 600 CEU (0.7 μM) was comparable to that at higher concentrations (2000 CEU), within the variation observed between enzyme batch to batch. However, the concentration at 400 CEU showed inconsistent results while at 200 CEU the cross-catalytic reaction did not show a sigmoidal curve and leveled off at lower percent conversions. This result implies that this concentration (600 CEU) may be the threshold of the activity of T4 DNA ligase to induce successful self-replication. These data suggest that the enzyme must be of similar concentration to the limiting reagent (typically 1.4 μM of the fluorescein-labeled strand) to facilitate LIDA.

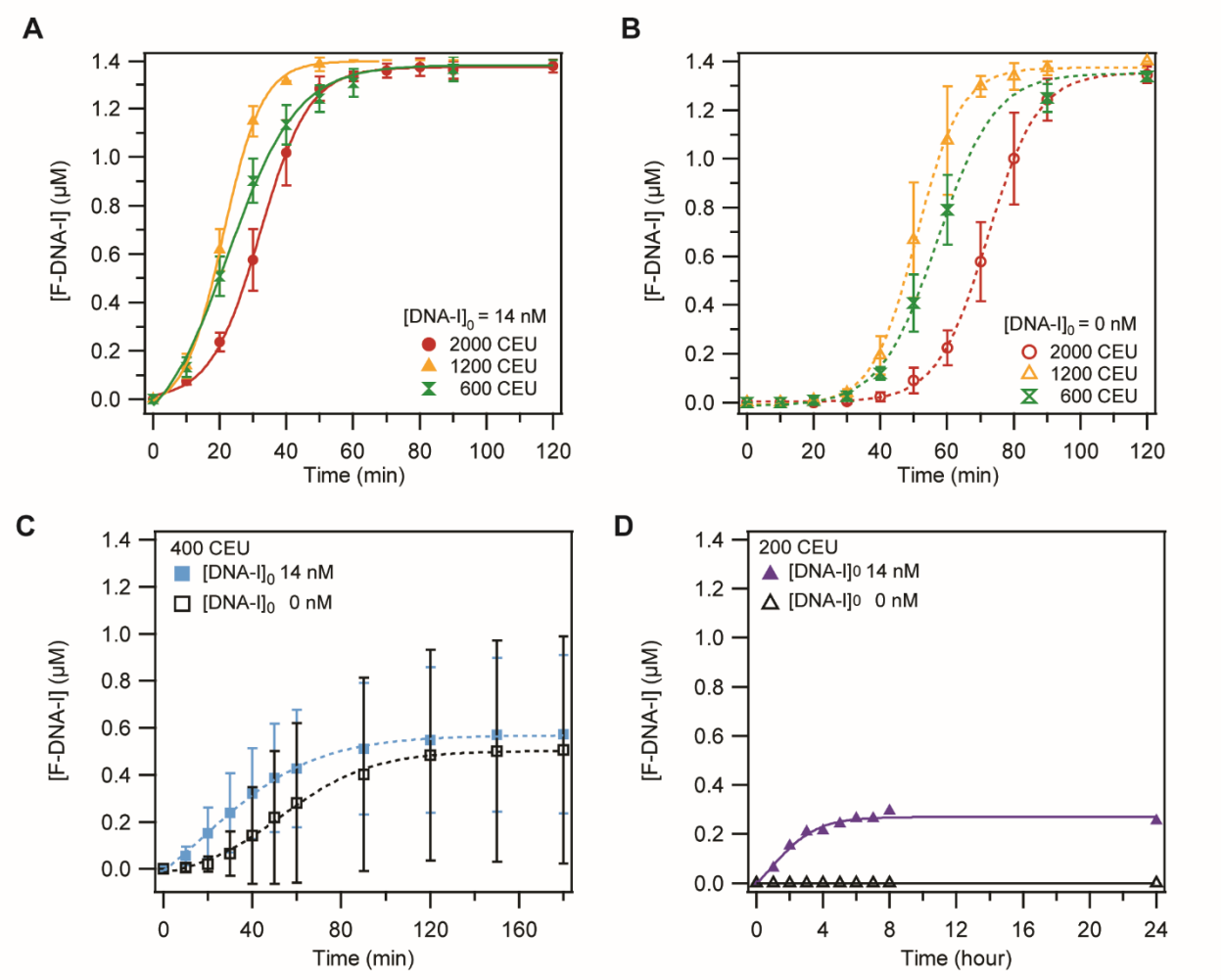


Figure 2.10 Lesion-induced DNA amplification (LIDA) with various enzyme concentrations. 2000 CEU, 1200 CEU, 600 CEU, 400 CEU and 200 CEU. The 2000 CEU enzyme was diluted to 1200 CEU, 600 CEU, 400 CEU, and 200 CEU using diluent A (B8001S) from New England Biolabs. (**A**) LIDA with 14 nM of the initial template with various concentrations of enzyme. (**B**) LIDA with 0 nM of the initial template with various concentrations of the enzyme (**C**) Cross-catalytic reaction with 400 CEU enzyme with extended time. (**D**) Cross-catalytic reaction with 200 CEU enzyme with extended time. *Experimental conditions*: 1.4 μM of **DNA-Ia**, 2.8 μM of **DNA-Ib$_p$, DNA-IIa$_p$**, which contains the abasic group, and **DNA-IIb**. Indicated T4 DNA ligase (1 μL) was used in the reaction (15 μL) at 30 °C. 50 mM Tris-HCl, 10 mM $MgCl_2$, buffer, and 1 mM ATP were manually added.

2.3 Conclusion

In conclusion, it has been shown that the addition of an abasic group instead of thymidine decreases the product duplex affinity constants for a particular 18mer duplex from 10^{12} M^{-1} to 10^{8} M^{-1} and the intermediate complex from 10^{5} M^{-1} to 10^{3}-10^{5} M^{-1} at 30 °C, the optimal replication temperature for this sequence. It was also found that T4 DNA ligase further compensates for the stability difference between the intermediate complex and the product duplex, reducing the product duplex stability by 10^{5} M^{-1}. This finding explains how spontaneous dissociation of product strands could occur at one temperature. Finally, our kinetic model reveals how rapid self-replication can be achieved even when the affinity constants are higher for the product duplex and the intermediate complex; a fast enough k_{cat} is required and a high enough affinity constant for the intermediate complex. Altogether, both the abasic group and T4 DNA ligase stabilization of the intermediate play a significant role in achieving rapid isothermal DNA self-replication. Therefore, a successful oligonucleotide self-replication system should incorporate destabilization and a catalyst that favors the intermediate complex over the product complex to overcome product inhibition further. These results can guide efforts to generate autonomous self-replicating oligonucleotides that use chemical rather than enzymatic ligation.

2.4 Experimental Section

2.4.1 General Materials

T4 DNA ligases were purchased from New England Biolabs (Ispwich, MA). T4 DNA ligase (2,000,000 cohesive end units/mL, Catalog M0202T) was used for cross-catalytic reaction

and pseudo-blunt-end ligation reaction. The corresponding ligation buffer (50 mM Tris-HCl, pH 7.5 at 25 °C, 1 mM ATP and 10 mM $MgCl_2$) as a 10x stock, and diluent A (10 mM Tris-HCl, pH 7.4 at 25 °C, 1 mM DTT, 0.1 mM EDTA, 50 % glycerol, 200 μg/mL BSA, 50 mM KCl) as 1x stock were also from New England Biolabs.

2.4.2 Oligonucleotide Synthesis

All oligonucleotides were synthesized on Applied Biosystems Model 392 DNA/RNA automated synthesizer using standard phosphoramidite reagents, CPGs (1 μmole scale), and other ancillary reagents purchased from Glen Research (Sterling, VA, USA). The abasic group was introduced with dSpacer CE phosphoramidite (10-1914-90), while Fluorescein dT phosphoramidite (10-1056-95) was introduced on oligonucleotide with extended 12 minutes coupling time. Solid chemical Phosphorylation Reagent II (10-1902-90) was used to synthesize 5' phosphate. Synthesized oligonucleotides were cleaved from the solid support by contacting ammonium hydroxide solution at room temperature overnight. Oligonucleotides were purified with Glen-pak cartridges (60-5200) by following Glen research DMT-off purification procedure. Oligonucleotides were stored in a 10 mM PBS buffer, and the concentration was determined using OligoCalc (http://www.basic.northwestern.edu/biotools/oligocalc.html).

2.4.3 Cross-Catalytic Reaction

The DNA self-replication experiments were performed as described in our earlier work.[143,144] The final concentration of three fragments (Ib, IIa, and IIb) was 2.8 μM, and the limiting fragments (Ia) was 1.4 μM. Fluorescein-labelling was on Ia or IIb. The template strand

(DNA-I, 1.4 μM) was diluted to 0.1 equivalent, as an initial concentration of 14 nM. The reaction was initiated by mixing the four fragments, with or without the template, ligation buffer, and 1μL of T4 DNA ligase in the total volume of 15 μL at 30 C. Aliquots were taken at various times and quenched with EDTA and bromophenol dye mixture. The data were analyzed in 15% PAGE gel and fit the KinTek Explorer.

2.4.4 Pseudo-Blunt End ligation

Cycle 1, two short strands (Ia, Ib) were used instead of 18 nt template. The 5' phosphate was added to the abasic strand (IIa), and fluorescein was attached to the other strand (IIb). Cycle 2, the 5' phosphate was added on the Ib strand, and a fluorescein label was added on the Ia strand. The abasic strand (IIa) and short strand (IIb) without any modification were added. In both cycles, fluorescein-labeled strands were limiting reagents as 1.4 μM, and the other strands were 2.8 μM. The T4 DNA ligase (1μL, 2,000,000 cohesive end units/mL) was added to the reaction mixture in the total volume of 15 μL and initiated the ligations at various temperatures (8 °C and 12 °C). The ligation was quenched with EDTA and bromophenol dye at each data point and analyzed by 15% PAGE gel, as described below.

2.4.5 Steady-State Ligation Assay

T4 DNA ligase (2,000,000 cohesive end units/mL, 35 μM)[170] was diluted to 1 μM using a diluent A buffer. The ligation assay contained T4 DNA ligase (final concentration of 25 nM), ligase buffer, template strand (DNA-I), fluorescein-labeled DNA (IIb), and short DNA strands (IIa) either containing abasic group or T base. The ligation reaction was performed at 16 °C. After incubating

the reaction mixture for 5 minutes, the reaction was initiated by the addition of a ternary DNA complex of various concentrations (0.1 μM to 5 μM), as shown in **Figure 2.16** and **Figure 2.17**. As described below, the data points were obtained at every minute for 10 minutes, quenched by mixing with EDTA, and analyzed by PAGE gel.

2.4.6 Isothermal Titration Calorimetry

ITC experiments were performed using Microcal VP-ITC (Microcal, Northampton, MA). The DNA strands were prepared in a 50 mM Tris-HCl buffer with 10 mM $MgCl_2$. The DNA strands in the syringe were 75 μM, and the DNA strands in the cell were 5 μM. Immediately before the experiments were performed, all the oligonucleotide solutions were degassed for 8 minutes at a temperature two degrees (°C) lower than the measurement temperature using the equipment provided with the calorimeter. The cell was washed with the buffer solution three times, then the oligonucleotide solution (5 μM) was added to the cell. A high concentration of the oligonucleotide solution (75 μM) was added to the syringe and purged twice. An initial 2 μL injection was set to counteract backlash in the auto titrator.[171] The ITC experiments consisted of ten 10 μL injections followed by eleven 15 μL injections of oligonucleotides (75 μM) to the complementary strands (5 μM). The cell temperature was set to the temperature (20 °C, 25 °C, 35 °C, or 45 °C) and the initial delay was 60 seconds. The stirring speed of the sample was 300 rpm. ITC data were analyzed with Origin software (Microcal Software Inc. Version 5.0). The raw thermogram data were analyzed and normalized in a plot of kcal/mole of injection versus molar ration. The fitting was done until Chi2 was no longer reduced. In the product template titration, the expected n=1 for DNA complex formation.

2.4.7 Polyacrylamide Gel Electrophoresis Analysis

Polyacrylamide gel (15%) was prepared with 8M urea mixed 5X TBE buffer and 40% acrylamide/bis solution 19:1 (Bio-Rad 161-0144). The gels containing the reaction mixture were analyzed using UV transillumination to visualize fluorophore strands by ImageQuant RT ECL Imager from GE Healthcare Life Science. Also, synthesized oligonucleotides were analyzed by StainsAll (Aldrich E9379) for purity. Oligonucleotide loaded gel was soaked in StainsAll solution (1:1 Formamide/water) (Formamide, Fisher F841) for 10 mins and visualized by white illumination via ImageQuant RT ECL instrument from GE Healthcare life science

2.4.8 Melting Curve Analysis

The oligonucleotide strands were prepared in three or four different concentrations as shown in **Table 2.1** in 50 mM Tris-HCl buffer (pH 7.5 at 25 °C), including 10 mM $MgCl_2$. The melting curve analysis was done using HP 8453 diode-array spectrophotometer equipped with an HP 89090A Peltier temperature controller. The buffer solution was used as blank to calibrate, then DNA solution was measured UV/vis at 260 nm with reference at 800 nm. The temperature range varies depending on the expected melting temperature of the DNA duplex (± 15 °C from theoretical melting temperature). The temperature increment was 1 °C and the holding time was 1 minute. The stirring speed was 200 rpm. The first derivative of the melting temperature was obtained from the UV-Visible ChemStation B0201 software. The data were fitted with Gaussian fit in Igor Pro (Wave Metrics. Inc. Version 7.08).

$$f(T) = f(0) + A\,exp[-\left(\frac{T - T_m}{width}\right)^2]$$

Where *f*(T) is the first derivative of absorbance at 260 nm, A is the amplitude, T is the temperature (°C), T_m is the melting temperature (°C), and *width* is the width of the Gaussian peak. The obtained T_m values were fitted into the equation below.[151]

$$\frac{1}{T_m} = \frac{(n-1)R}{\Delta H} lnC_T + \frac{[\Delta S^0 - (n-1)R\ ln2n]}{\Delta H^0}$$

(n = 2, a biomolecular process for product duplex and intermediate complex.)

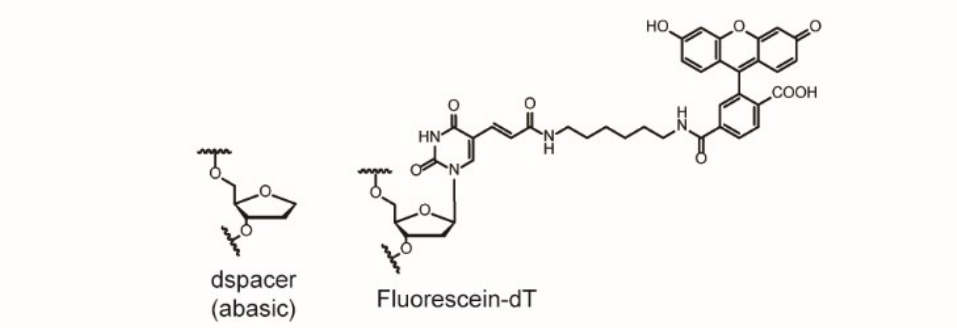
(n = 3, ternary complex)

2.4.9 Matrix-Assisted Laser Desorption Ionization (MALDI)

Matrix solutions were prepared with a 9:1 ratio of 25 mg/mL of 2,4,6-tridoxyacetophenole in 1:1 acetonitrile: water solution to 25 mg/mL ammonium citrate. Synthesized oligonucleotides were dissolved in 1:1 ratio of water and matrix (about 0.6 mM to 1 mM solution). Oligonucleotides were characterized by their mass by a Voyager Elite (Applied Biosystems, Foster City, CA) time-of-flight in linear negative mode.

Table 2.4 Oligonucleotide sequence and corresponding MALDI data.

Oligonucleotides	Sequence (5' → 3')	Expected	Founded
DNA-I	TTG TTA AAT ATT GAT AAG	5552	5555
DNA-II (abasic group)	CTT ATC AA(abasic) ATT TAA CAA	5318	5321
DNA-II (Perfect match, T)	CTT ATC AA**T** ATT TAA CAA	5441	5445
F-DNA-Ia	$T_{fluorescein}$TG TTA ATT	3239	3243
F-DNA-Ia (+2)	$T_{fluorescein}$A TTG TTA ATT	3857	3864
DNA-Ia	TTG TTA AAT	2728	2732
DNA-Ib_p	phosphate ATT GAT AAG	2842	2845
DNA-Ib	ATT GAT AAG	2762	2765
DNA-IIa_p	phosphate (abasic) ATT TAA CAA	2966	2968
DNA-$IIa_{(OH)}$	(abasic) ATT TAA CAA	2886	2888
DNA-IIb	CTT ATC AA	2842	2845
F-DNA-IIb	$T_{fluorescein}$ CTT ATC AA	3184	3188



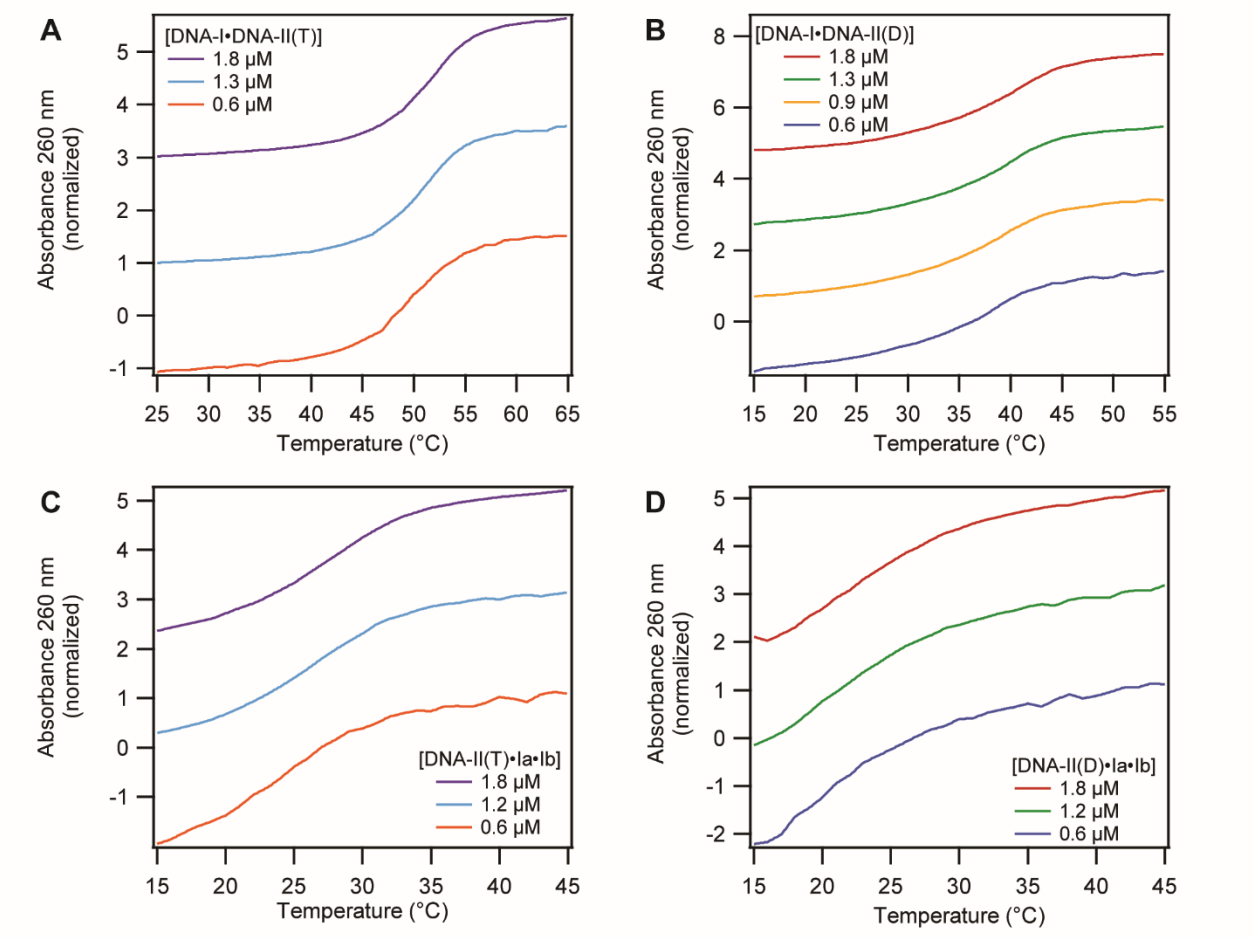


Figure 2.11 Melting profiles of DNA complexes. Data were normalized and offset for clarity. (**A**) Product complex with T base. (**B**) Product complex with the abasic group. (**C**) Ternary complex with T base. (**D**) Ternary complex with the abasic group.

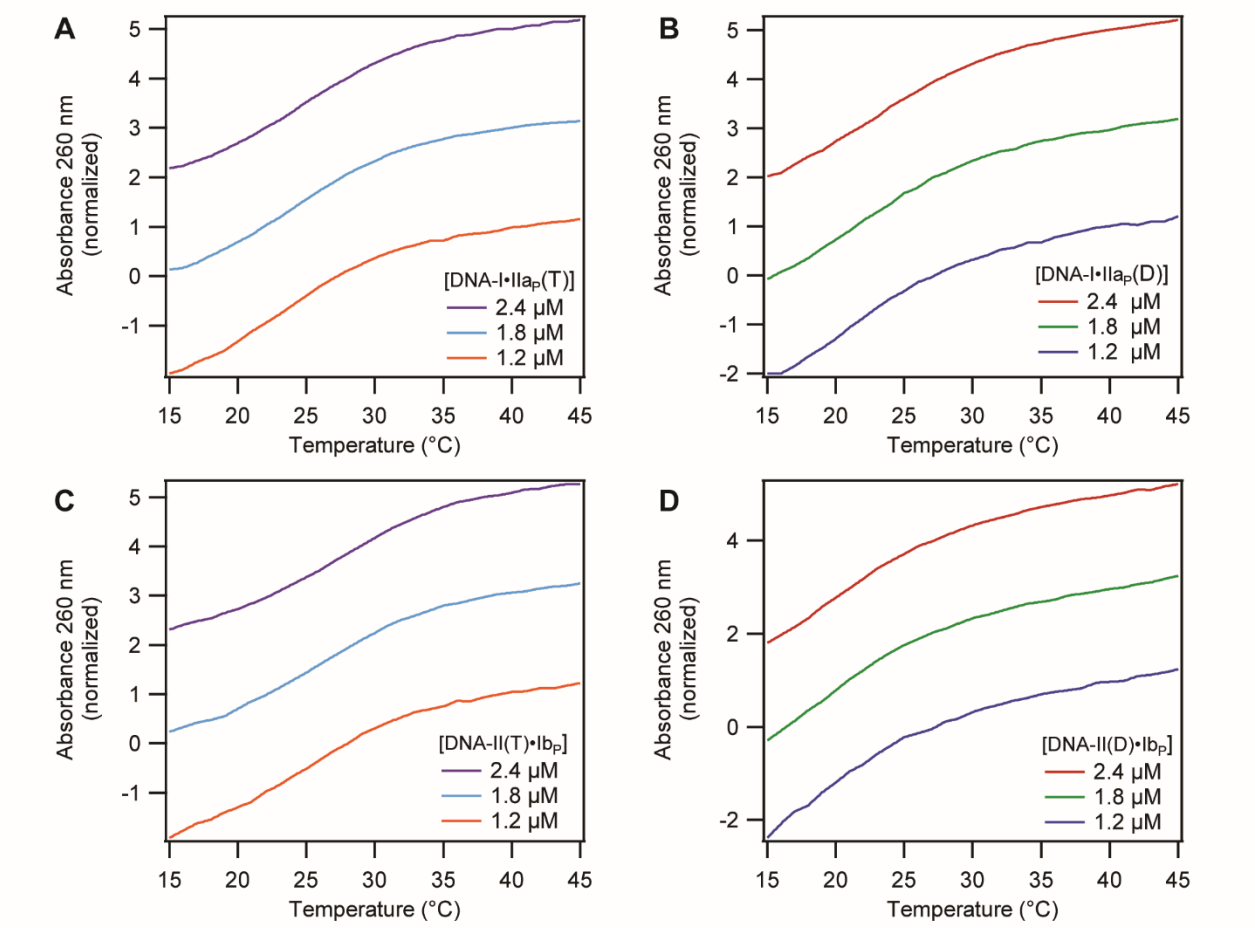


Figure 2.12 Melting profiles of DNA complexes. Data were normalized and offset for clarity. (**A**) Intermediate complex: the template hybridized with a fragment containing a T base. (**B**) Intermediate complex: the template hybridized with a fragment containing an abasic site. (**C**) Intermediate complex: the template containing T base hybridized with a fragment. (**D**) Intermediate complex: the template containing an abasic group hybridized with a fragment.

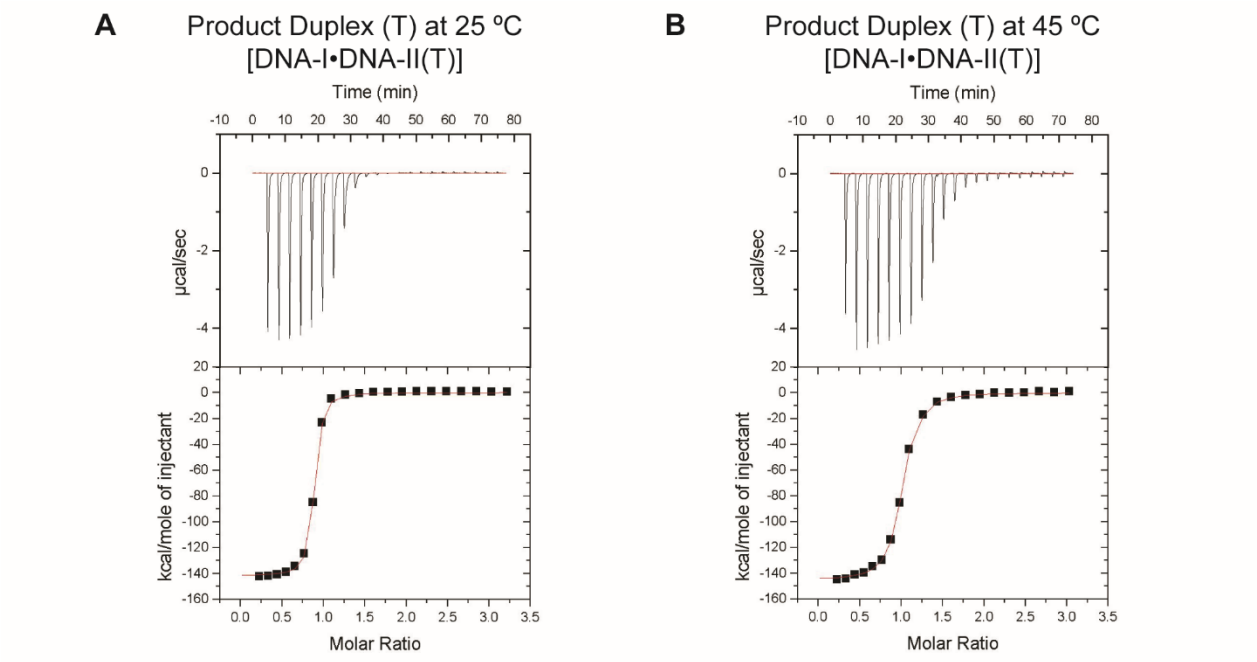


Figure 2.13 Representative ITC data of DNA complexes with a thymidine base. (**A**) Product duplex with T base measured at 25℃. (**B**) Product duplex with T base measured at 45 ℃. DNA strands in the syringe (75 µM) and the cell (5 µM). *Experimental condition:* 50 mM Tris-HCl buffer (pH 7.5 at 25 ℃) including 10 mM $MgCl_2$.

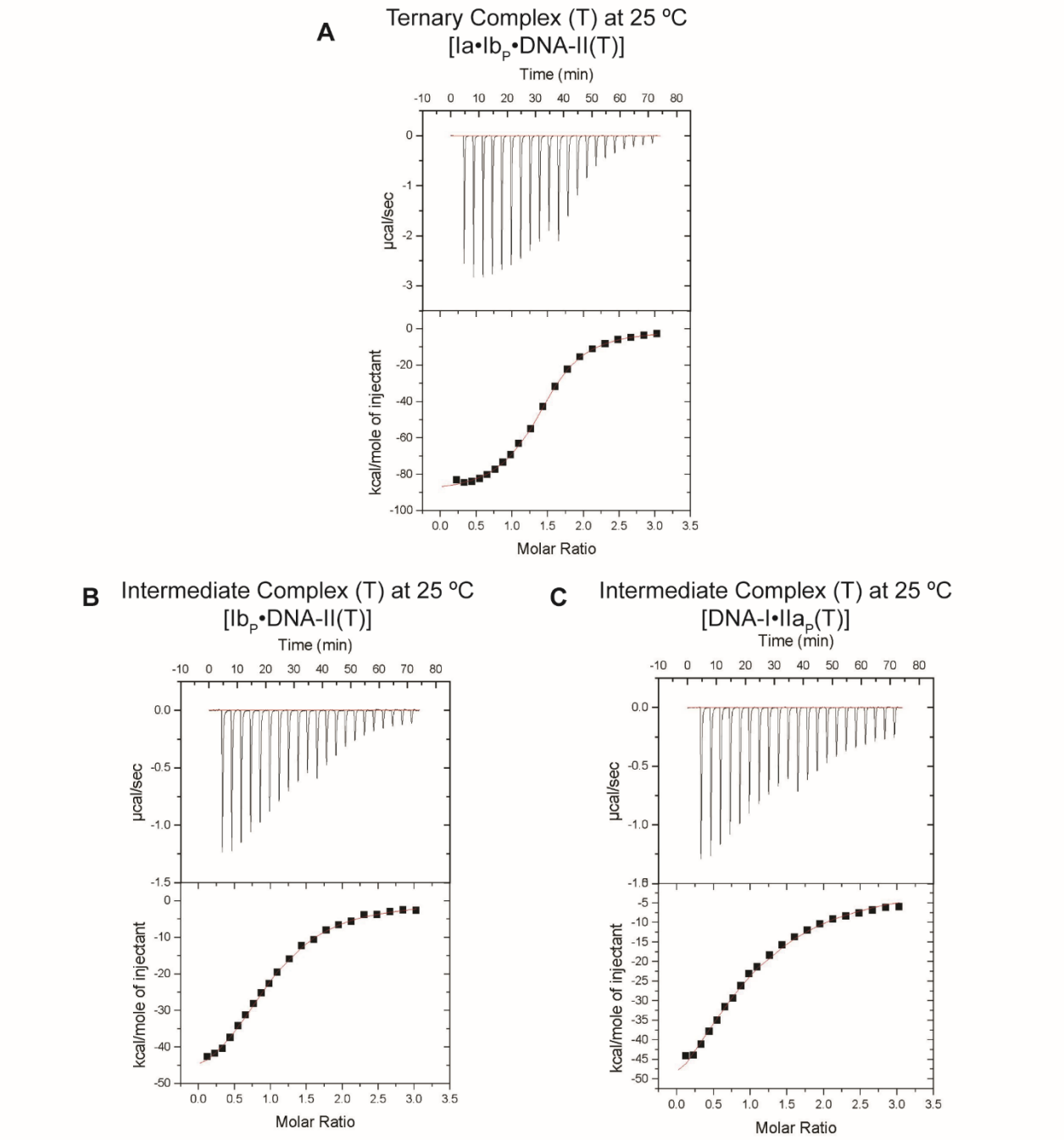


Figure 2.14 Representative ITC data of DNA complexes with a thymidine base. (**A**) Ternary complex with T base measured at 25°C. (**B**) Intermediate complex (**Ib$_P$:DNA-II(T)**) with T base measured at 25 °C. (**C**) Intermediate complex (**DNA-I:IIa$_P$(T)**) with T base measured at 25 °C. DNA strands in the syringe (75 µM) and the cell (5 µM). *Experimental condition:* 50 mM Tris-HCl buffer (pH 7.5 at 25 °C) including 10 mM $MgCl_2$.

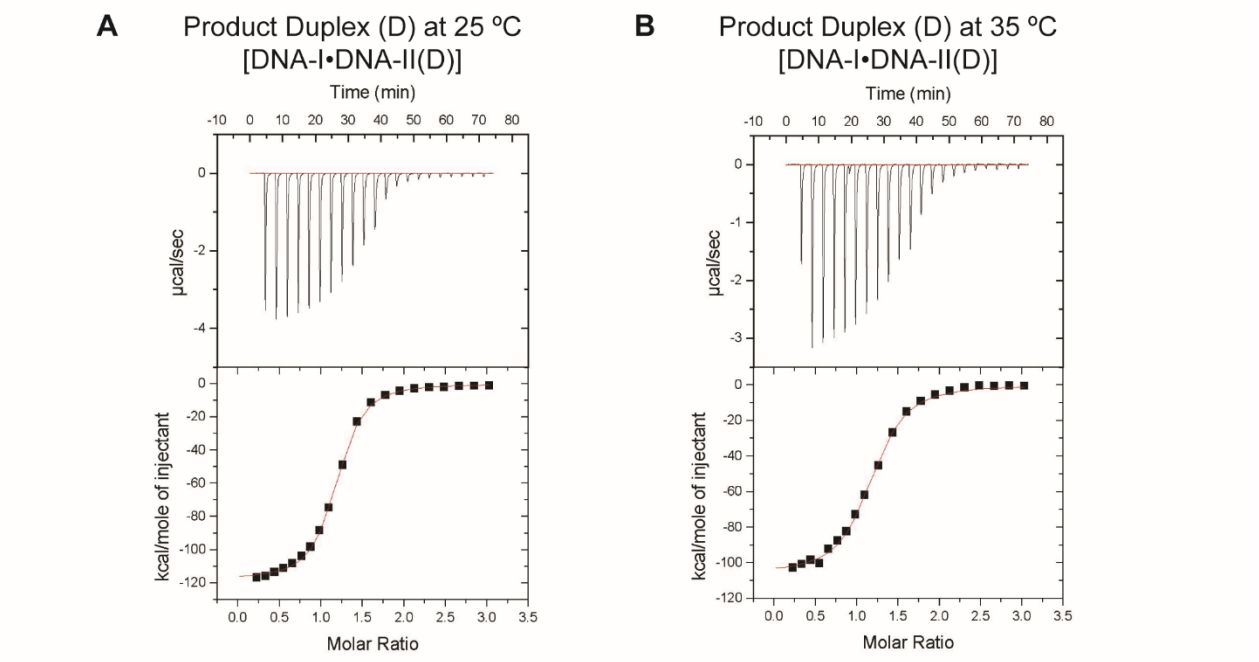


Figure 2.15 Representative ITC data of DNA complexes with the abasic group. (**A**) Product duplex with the abasic group measured at 25 ℃. (**B**) Product duplex with the abasic group obtained at 35 ℃. DNA strands in the syringe (75 µM) and the cell (5 µM). *Condition:* 50 mM Tris-HCl buffer (pH 7.5 at 25 ℃) containing 10 mM $MgCl_2$.

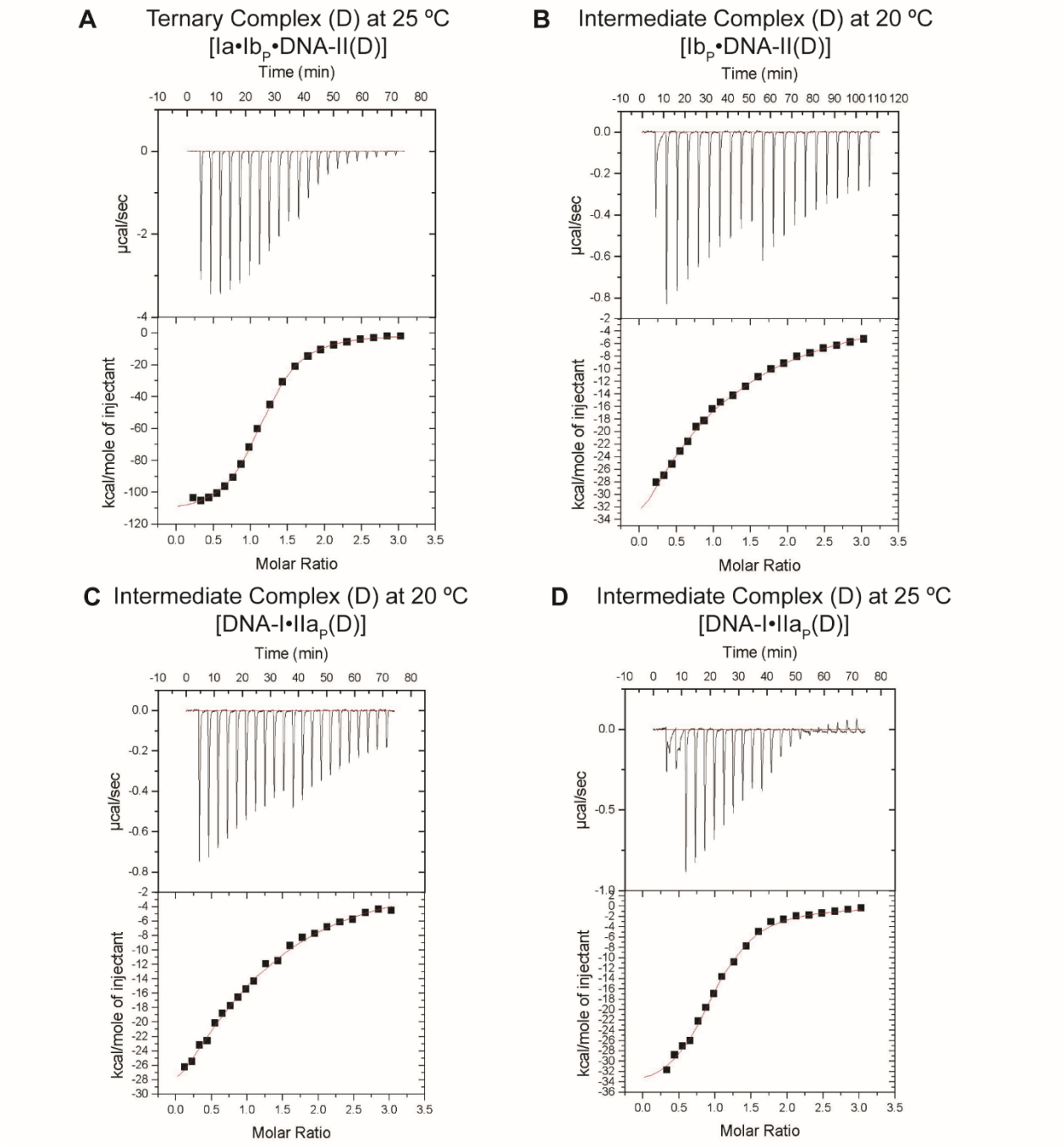


Figure 2.16 Representative ITC data of DNA complexes with the abasic group. (**A**) Ternary complex with the abasic group was measured at 25 ℃. (**B**) Intermediate complex (**Ib$_P$:DNA-II(D)**) with the abasic group was obtained at 20 ℃. (**C**) Intermediate complex (**DNA-I:IIa$_P$(D)**) with the abasic group was obtained at 20 ℃. (**D**) Intermediate complex (**DNA-I:IIa$_P$(D)**) with the abasic group was obtained at 25 ℃. *Experimental condition:* DNA strands in the syringe (75 µM) and in the cell (5 µM). 50 mM Tris-HCl buffer (pH 7.5 at 25 ℃) containing 10 mM $MgCl_2$.

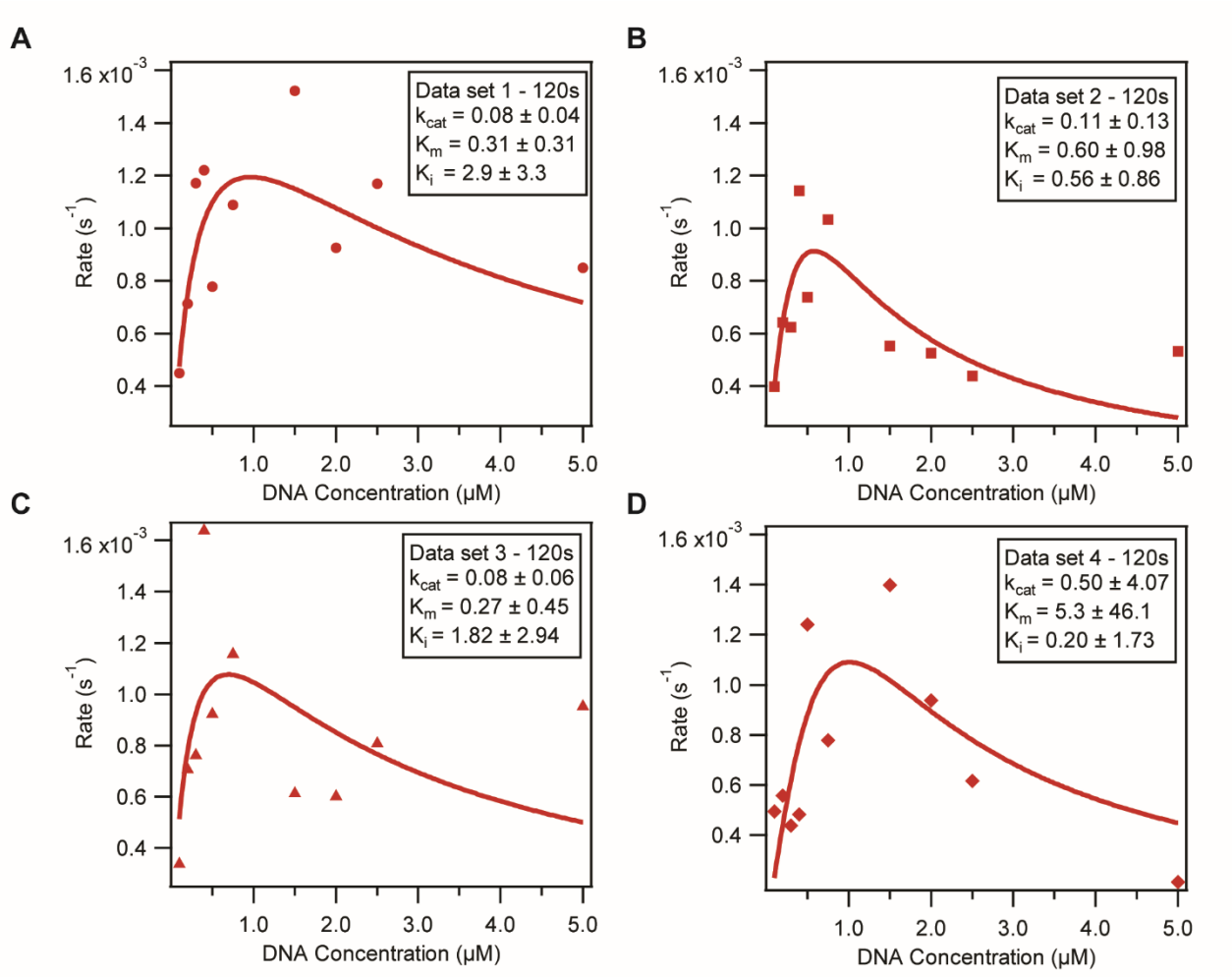


Figure 2.17 Individual data set of Michaelis-Menten parameters for abasic ternary complex. (**A**) k_{cat} = 0.08 ± 0.04 s^{-1}, K_m = 0.3 ± 0.3 μM, K_i = 2.9 ± 3.3 μM, (**B**) k_{cat} = 0.1 ± 0.1 s^{-1}, K_m = 0.6 ± 1 μM, K_i = 0.6 ± 0.9 μM, (**C**) k_{cat} = 0.08 ± 0.06 s^{-1}, K_m = 0.3 ± 0.5 μM, K_i = 2 ± 3 μM, (**D**) k_{cat} = 0.5 ± 4.1 s^{-1}, K_m = 5 ± 46 μM, K_i = 0.2 ± 1.7 μM.

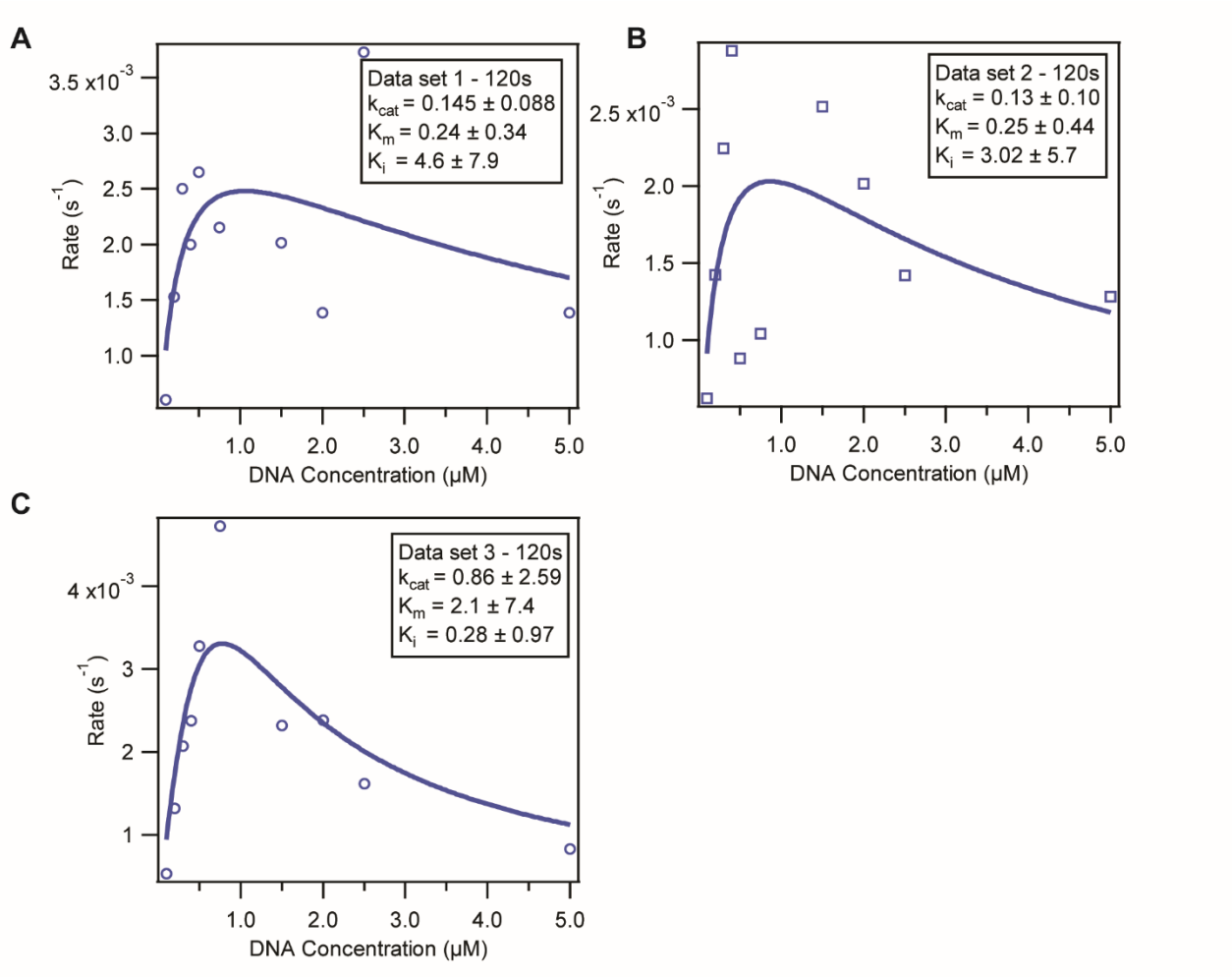


Figure 2.18 Individual data set of Michaelis-Menten parameters for T base ternary complex. (**A**) k_{cat} = 0.15 ± 0.09 s^{-1}, K_m = 0.2 ± 0.3 μM, K_i = 5 ± 8 μM, (**B**) k_{cat} = 0.1 ± 0.1 s^{-1}, K_m = 0.3 ± 0.4 μM, K_i = 3 ± 6 μM, (**C**) k_{cat} = 0.9 ± 2.6 s^{-1}, K_m = 2 ± 7 μM, K_i = 0.3 ± 1 μM

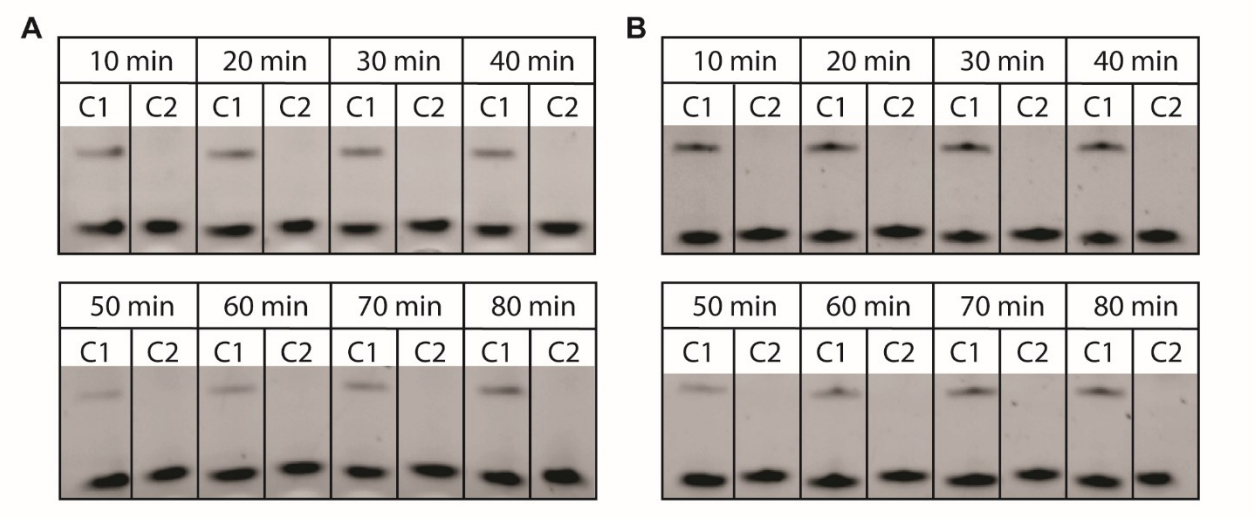


Figure 2.19 Representative gel image of pseudo-blunt-end ligation. Pseudo-blunt-end ligation (**A**) at 8 °C (**B**) at 12 °C. C1 indicates cycle 1 where Ia, Ib, 5' phosphate IIa and fluorescein-labeled IIb were used. C2 indicates cycle 2, where fluorescein-labeled Ia, 5' phosphate Ib, IIa, and IIb were used.

Chapter 3

Selectivity and Efficiency of the Ligation of the Pyrene:Abasic Base Pair by T4 DNA Ligase: Towards Abasic Lesion Detection by Ligation

3.1 Introduction

DNA damage occurs continuously as a consequence of exposure to genotoxic agents.[172,173] Carcinogens, including tobacco and alcohol, and mutagens such as UV radiation, X-rays, and oxidative materials are the types of genotoxic agents resulting in mutations that can lead to cancers.[174–178] Endogenous cellular processes also lead to DNA damage, including depurination, depyrimidination, double-strand breaks, cytosine deamination, and the formation of O^6-methylguanine and other alkylated lesions.[179–182] The most common lesion observed in DNA is an apurinic or apyrimidinic site called the abasic site, which is estimated to occur ~10,000 times in a human cell per day.[181–186] Abasic sites can occur not only by DNA damage such as DNA alkylation or oxidation,[185,187,188] but also as an intermediate of the excision repair pathway.[189–191] Although most of this damage is repaired by the DNA repair system, remaining or undetected abasic sites, which have lost their genetic information, can cause mutations or acute toxicities in cells.[192–195] Owing to the broad role of abasic sites in biology, it is important to detect their presence and locations and understand their formation within genomic DNA.

The most common method to detect abasic sites is to employ aldehyde-reactive probes (ARPs).[183,184,196–199] The aldehyde group of the ring-open form of the abasic site can react with hydroxylamines, a common functional group used in ARPs. A capture or labeling group like biotin constitutes another part of the ARP. Therefore reaction of ARPs on the abasic sites can be quantified with dot blot or enzyme-linked immunosorbent assays after binding to streptavidin by biotin.[196,199–201] However, the main shortcoming of this method is the cross-reactivity caused by other reactive aldehydes in DNA that do not arise from abasic sites.[200–203] Recently, the strategies to map the abasic site in a single-nucleotide resolution were reported.[204,205] SnAP-seq which uses ARPs to label the abasic sites, improved the specificity of the method by including an additional

chemical step, alkaline treatment, that led to a strand break only for the abasic sites allowing it to be discriminated from other ARP-reactive lesions.[204] Another method for abasic detection, Nick-seq, utilizes an AP endonuclease (Endo IV) to cleave the abasic sites resulting in strand breaks. This method achieved increased sensitivity and specificity by blocking pre-existing strand breaks and converting DNA modifications enzymatically or chemically.[205]

Abasic, or model abasic groups where the 1'-OH is replaced by a hydrogen, have also been explored as part of an unnatural base pair with 1'-pyrene nucleotides. Such unnatural base pairs (UBPs) have been developed to extend the genetic alphabet of nucleic acids.[206–208] Unnatural base pairs are of great interest in biotechnology applications because of improved hybridization or site-specific labeling abilities with different functionalities.[209–211] One typical strategy to develop unnatural base pairs is the use of nucleoside analogs with hydrogen bonding motifs that are orthogonal to the natural bases.[212,213] For example, the Benner group showed that their dP-dZ pairs exhibit high fidelity in PCR amplification (**Figure 3.1A**).[214] Another strategy is to use hydrophobic and non-hydrogen bonding nucleobases relying on the geometry fit and packing forces between the nucleobases analogs.[215–218] The imidazopyridine and pyrrole derivatives from the Hirao group were successfully amplified in PCR with high selectivity on each other (**Figure 3.2B**).[219,220] Moreover, the Romesberg group developed unnatural base pairs, dNMAP-dTPT3 pairs, with high replication fidelity demonstrating not only storing but also retrieving the expanded information (**Figure 3.2C**).[221] However, the study of enzymatic ligation of unnatural base pairs at the site of ligation remains less common.[222,223]

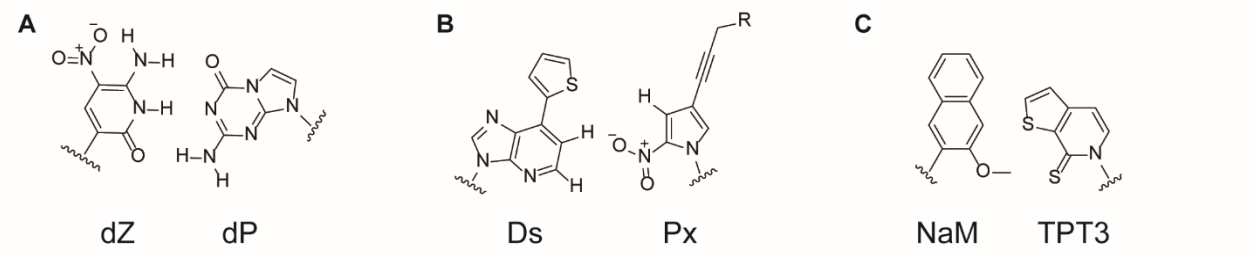


Figure 3.1 Chemical structures of unnatural base pairs system. (**A**) Hydrogen bonding UBP from the Benner group, 6-amino-5-nitro-2(*1H*)-pyridone (dZ) and 2-amino-imidazo[1,2-α]-1,3,5-triazine 4(*8H*)one (dP) (**B**) Hydrophobic UBP from the Hirao group, 7-(2-thienyl)imidazo[4,5-b]pyridine (DS), 2-nitro-4-propynyl pyrrole (Px (**C**) Hydrophobic UBP from the Romesberg group, 2-methoxyl-3-methylnaphtalene (dNAM) and thieno[2,3-c]pyridine-7(*6H*)-thione (TPT3).

Regarding non-hydrogen bonding base pairs, Matray and Kool first demonstrated that a non-hydrogen bonding pyrene nucleotide exhibited high stability when paired across from an abasic site.[215] They showed that the large size of pyrene could compensate sterically for the absence of a base when paring with the abasic site. The robust π stacking ability of pyrene as a base also aids the stability of pyrene-abasic pairs in the DNA duplex.[215,216,224] Furthermore, they demonstrated that pyrene nucleoside triphosphate (dPTP) could be incorporated specifically opposite of abasic sites by the Klenow fragment of *E.coli* DNA polymerase I.[215] Additionally, they showed the visualization of a level of 1% of abasic sites based on the incorporation of dPTP. Pyrene base was incorporated by other polymerases such as *E.coli* polymerase, DNA polymerase.[215,225] These results demonstrated that hydrogen bonding of the nucleobase is not necessary for stable DNA base pairs or enzyme recognition. Pyrene nucleotides were further studied with DNA methyltransferase MTaqI and led to enhancement of DNA-MTaqI binding that was attributed to pyrene's ability to stabilize the base flipped structure of the opposing adenosine after binding to the methyltransferase.[226]

The ligation of unnatural base pairs has also been explored to some extent, particularly with T4 DNA ligase. T4 DNA ligase is widely used to mediate the ligation of nicked sites. Also, T4 DNA ligase is known to tolerate modified nucleotides at the 5'-phosphate terminus of the ligation site.[227–229] Taking advantage of these features of T4 DNA ligase, modifications including mismatch bases and unnatural bases are typically studied at the 5'-phosphate terminus of the nicked site rather than the 3'-hydroxy terminus. For example, Hirao and coworkers reported that non-hydrogen bond base pairs (Ds:Pa where Pa has aldehyde moiety instead of nitro moiety from Px) were successfully ligated by T4 DNA ligase at the ligation junction.[230] When they located unnatural Ds base on the 5'-phosphate end, it exhibited higher % ligation across from Pa base on a template than other native nucleobases. When 3'-hydroxy Ds strand was used, it showed enhanced selectivity on Pa base containing template than thymine base containing template. However, the 3' hydroxy adenine nucleobase strand was also ligated across from Pa base containing template. They reasoned the selectivity of the imidazopyridine base (Ds) for the pyrrole derivative (Pa) was due to the similar geometry of the resulting bass pair to the canonical base pairs. Regarding the interaction of unnatural nucleotides interacting with native nucleobases, the Burrows group demonstrated the installation of unnatural bases, dNAM, or d5SICS nucleotides using T4 DNA ligase with the addition of DMSO.[222] They showed the most efficient ligation of those nucleotides was opposite the site of natural G bases with the highest yield of ~70%. Furthermore, Pack and coworkers utilized the enhanced selectivity of T4 DNA ligase on discrimination of single nucleotide polymorphisms using the unnatural bases 5-nitroindole, iso-dG, and iso-dC, at the 5' phosphate-end position.[231] The authors argued that non-hydrogen bulky bases played a role in ligation selectivity by changing the geometrical distance between the template and the ligation strands.[231]

Chlorella virus PBCV-1 DNA ligase, known as SplintR ligase, also seals nicked site on DNA substrates either on RNA or DNA template requiring co-factor ATP.[232] PBCV-1 DNA ligase is commonly used in RNA detection based on RNA-templated DNA ligation[233–236] due to its remarkable ligation efficiency for DNA substrates on an RNA template.[232] Recent studies on the end-joining of DNA substrates with different blunt-end and overhang architectures with PBCV-1 DNA ligase were reported.[237] PBCV-1 DNA ligase effectively ligates 3'-single base overhangs rather than blunt ends and 5' single-base overhang. However, the application of PBCV-1 on unnatural bases remains unexplored.

Here we report the highly selective and efficient enzymatic ligation of 5'-phosphate 1'-pyrene nucleotides across from an abasic group in the template strand with both T4 DNA ligase and PBCV1-DNA ligase. This approach uses a chemically synthesized pyrene nucleotide chosen due to its known stability and geometrical fit with an abasic lesion.[215,216] The ligation selectivity and rate of the ligation of the pyrene nucleotide across from the abasic site was as high as that observed for the natural base pairs. In contrast, the ligation of the 5'-phosphate abasic across from the pyrene nucleotide on the template strand exhibited less selectivity. Finally, we have demonstrated the selective ligation of DNA abasic sites over the canonical bases in a ligation competition experiment. We expect this strategy for detecting abasic lesions will enable us to identify other types of DNA damage sites when incorporating an additional first step of selective glycosylase generation of abasic sites.

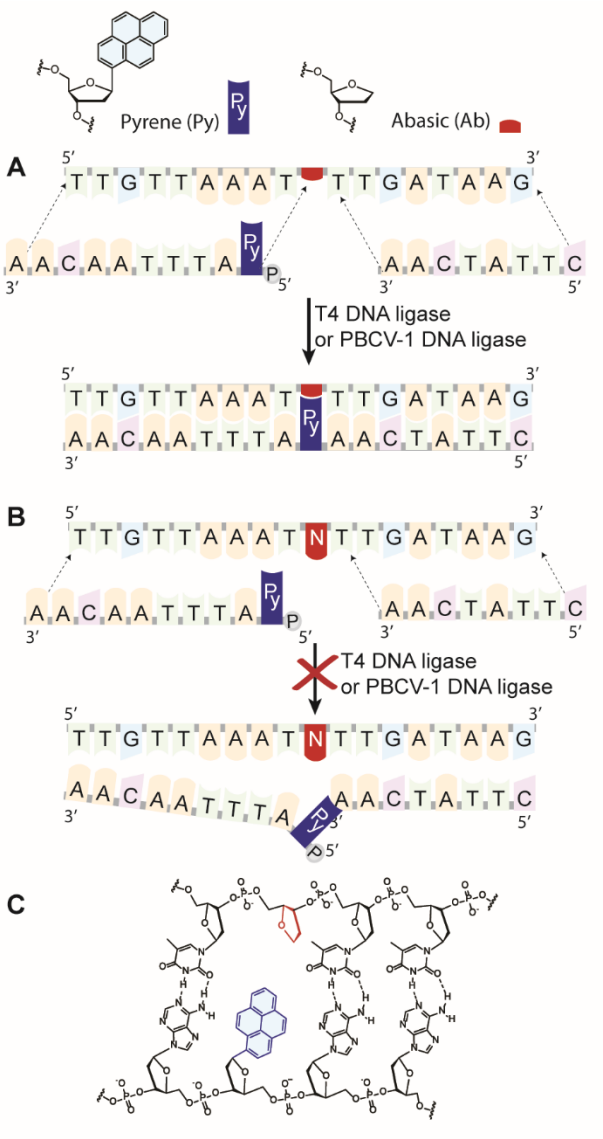


Figure 3.2 Schematic diagram of selective ligation of the pyrene strand on the abasic site. (**A**) Successful ligation of pyrene strand across from the abasic site. (**B**) Unsuccessful ligation of pyrene strand across from the natural DNA bases (N, any natural bases). (**C**) Molecular structure depicting the pyrene:abasic base pair in a duplex.

3.2 Selectivity and Efficiency of the Ligation of the Pyrene:Abasic Base Pair

3.2.1 Ligation Reaction of 5'-Phosphate Pyrene Strand with Abasic Containing Template and Adenine Base Template

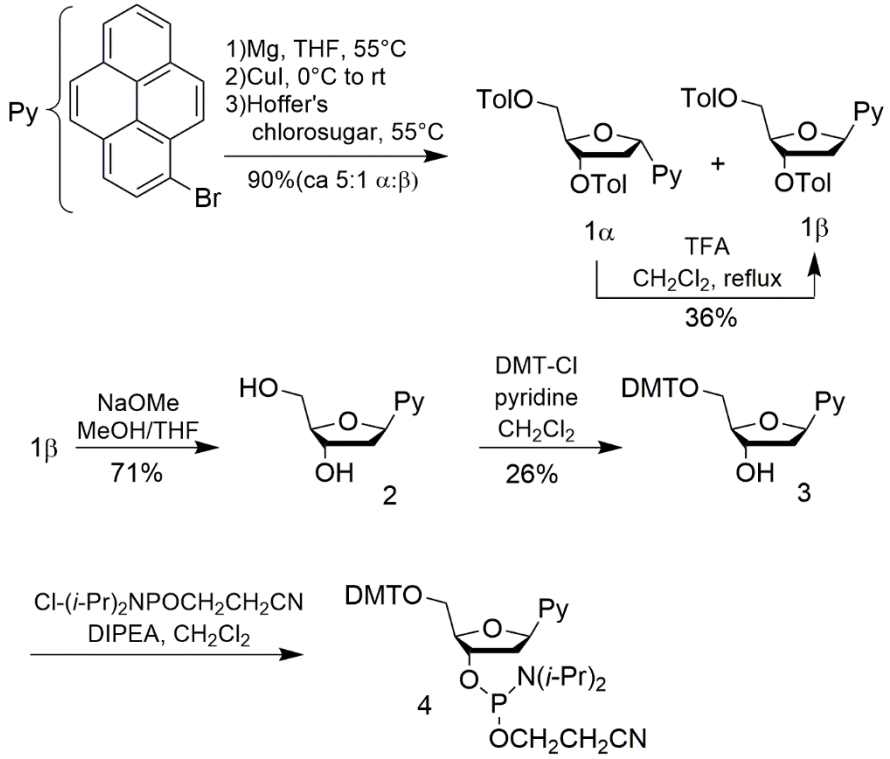


Figure 3.3 Schematic representation of the synthesis of 1-pyrene nucleotides.[238]

First, the pyrene-modified nucleotide was synthesized from 1-bromopyrene and Hoffer's chlorosugar (3,5-ditoluoyl-1-α-chloro-2-deoxy-D-ribose) following the literature procedure.[224,238,239] The β-anomor nucleoside was selectively purified and followed by removal of the toluoyl group, DMT-protection of the 5'-OH, and 3'-phosphitylation. The resulting pyrene phosphoramidite was then added into the oligonucleotides strands using a solid-phase synthesizer at the 5'-position of the CPG-bound strand. For the template strand containing pyrene, solid-phase synthesis was continued with the addition of the native nucleotides. For the pyrene-terminated

strand, after the coupling of the pyrene nucleotide, 5' phosphorylation was performed using the standard reagents on the solid-phase synthesizer. The pyrene-modified oligonucleotides were purified by RP-HPLC and characterized by polyacrylamide gel electrophoresis and mass spectrometry.

We hypothesized that T4 DNA ligase and PBCV-1 DNA ligase would catalyze the ligation of the 5'-phosphate pyrene-modified strands, selectively across from the abasic site (**Figure 3.2**), although the enzymatic ligation of unnatural nucleotides with the exception of model 1'-H-abasic groups[143,168] had been limited to nucleobases containing heteroatoms. To see whether the T4 DNA ligase and PBCV-1 DNA ligase tolerated the bulky pyrene substituent at the ligation site and whether the pyrene showed selective ligation across the abasic site, the stoichiometric ligation of a 5'-phosphate pyrene strand and a fluorescein-labeled 3'-OH terminated strand was tested in the presence of the DNA template containing the abasic group or a natural base, A, across from the pyrene (**Figure 3.4**). The ligation %yield was determined by the formation of the fluorescein-labeled product strands with polyacrylamide gel electrophoresis and plotted versus time. Both DNA ligases showed fast ligation with the 5'-phosphate pyrene nucleotide across from the abasic group on the template. However, T4 DNA ligase enabled the ligation of pyrene nucleotides across the adenine base (10% yield after 60 min), whereas the PBCV-1 DNA ligase did not show any ligation activity for the pyrene:A base pair. Further to optimize the T4 DNA ligase condition to suppress unwanted ligation across from A, the temperature was varied from 16 ℃ to 42 ℃ in the ligation reaction (**Figure 3.5A**) as the selectivity of T4 DNA ligase of the base pair against mismatched base pairs was increased at higher temperatures, well above the melting temperature of the DNA duplex as observed in our colleagues' previous report.[240] The difference in ligation between the abasic template and the adenine base-containing template was distinct in all temperatures, especially at the high temperature (36 ℃ and above) where the ligation with the

adenine base template did not occur (**Figure 3.5A**). This result agreed with our previous observations that the elevated temperature above the T_m increases the selectivity of the ligation matched base pair, in this case, the stable pyrene: abasic pair. Overall, T4 DNA ligase and PBCV-1 DNA ligase successfully facilitated the ligation of the bulky pyrene at the ligation junction and yielded the product strand (18 nt), notably when the abasic template was used.

Next, to optimize the conditions for T4 DNA ligase, which was less intrinsically discriminating than PBCV-1 DNA ligase, the effect of ATP concentration on the ligation reaction was explored since ATP is a cofactor of T4 DNA ligase enabling the nick sealing of DNA duplex.[241] Previous studies showed that the concentration of ATP affected the ligation rate and fidelity of T4 DNA ligase.[147,241,242] Since the ATP binds to the ligase very fast and irreversibly if there is a high concentration of ATP, the ligase dissociated from adenylated DNA complex would be re-adenylated fast and unable to catalyze the ligation, leading to reduced ligation efficiency.[166] Therefore, the ATP concentration was varied to control the specificity of the pyrene:abasic pair. The same types of ligation reactions were performed at 16 ℃ in different ATP concentrations. T4 DNA ligase ligated not only the pyrene:abasic pair efficiently (~100%) but also the unfavoured pyrene:A pair some amount (~20 %) with 1 mM ATP, which is the standard concentration used for ligation with this enzyme (**Figure 3.5B**). However, ligation was only observed with the pyrene:abasic pair in high ATP concentration (10 mM) without compensating the % yield, proving selective ligation of the pyrene strand across the abasic site. The ATP and temperature variation results indicate that the selectivity of the pyrene strand can be tuned by the reaction conditions, which opens up its possibility of use in an abasic lesion detection method.

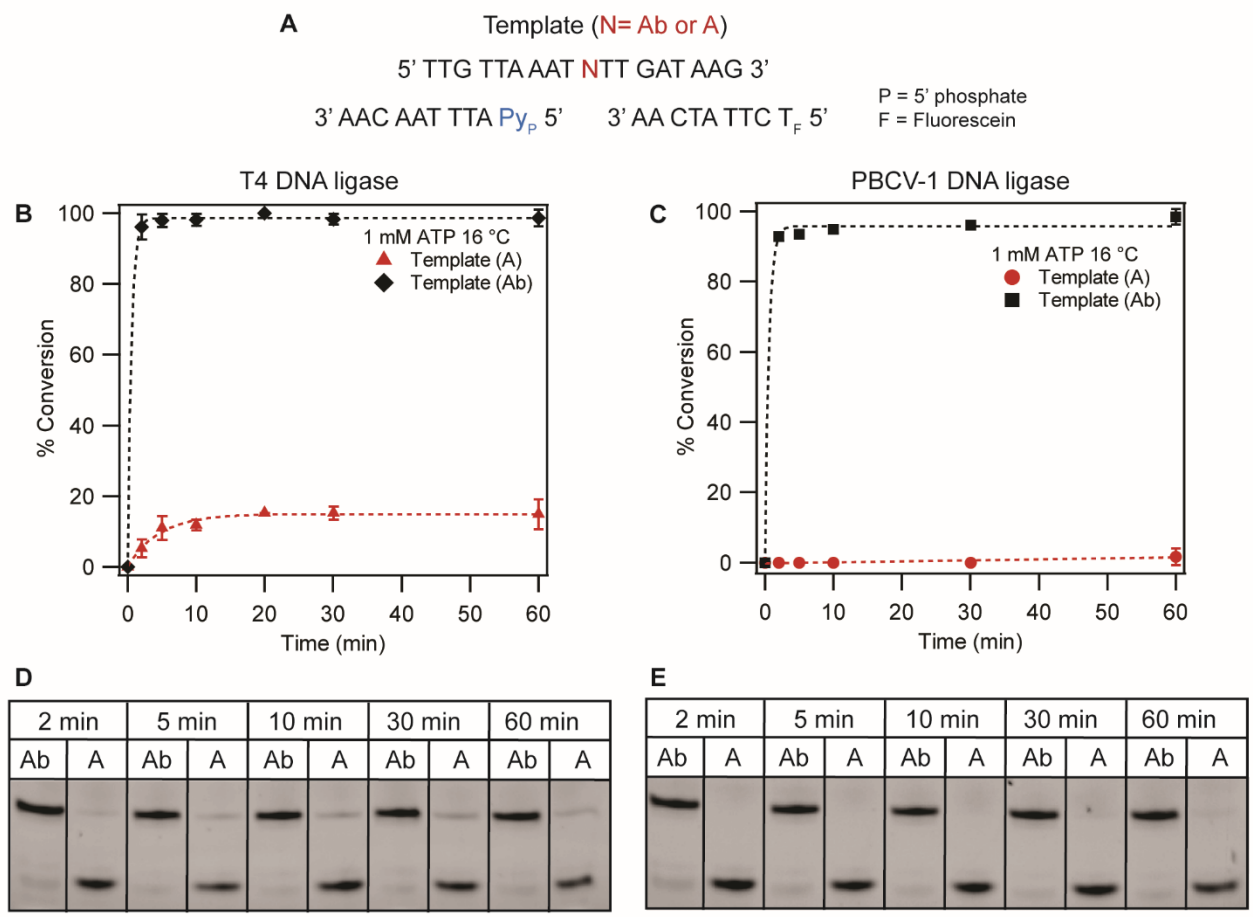


Figure 3.4 The ligation reaction of the 5'-phosphate pyrene strands. (**A**) Template and fragments sequence that was used in the experiments. (**B**) Time course of enzymatic ligation of pyrene nucleotides in the 5' position with 400 CEU T4 DNA ligase. (**C**) with PBCV-1 DNA ligase. (**D**) Representative PAGE gel image of stoichiometric ligation with 400 CEU T4 DNA ligase. (**E**) Representative PAGE gel image of stoichiometric ligation with PBCV-1 DNA ligase. *Experimental conditions*: 2.8 µM of the 5'-phosphate pyrene strand, 1.4 µM of the fluorescein-labeled strand, and 1.4 µM of the template strands. T4 DNA ligase (400 CEU, 1 µL per 15 µL total volume) or PBCV-1 DNA ligase (10.5 µM, 1.5 µL per 15 µL total volume) in 50 mM Tris-HCl, 10 mM $MgCl_2$ and 1 mM ATP concentrations.

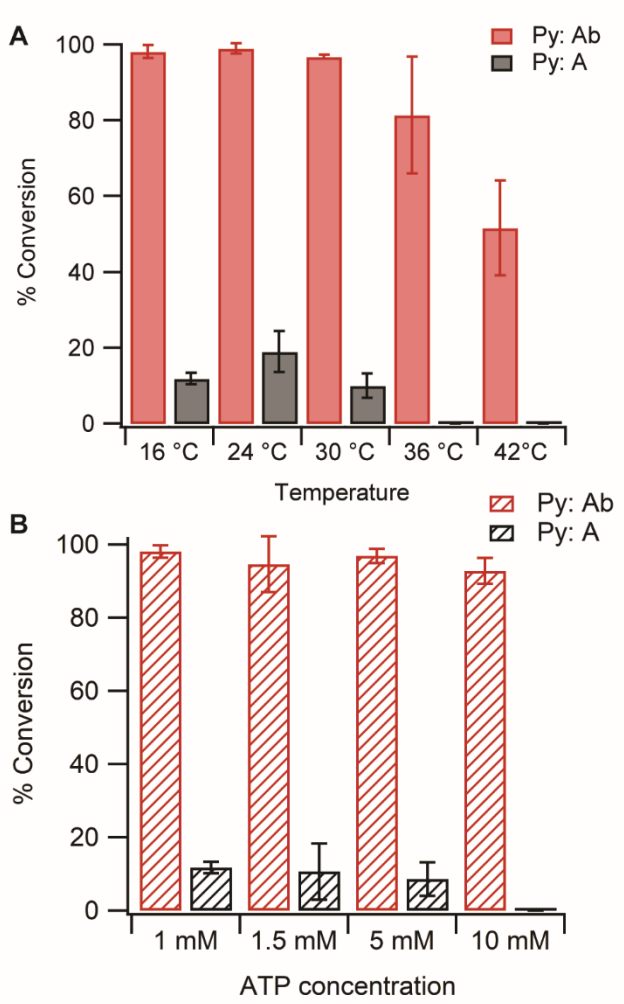


Figure 3.5 Comparison of T4 DNA ligase catalyzed ligation of 5' phosphate pyrene strand using the template with either an abasic site or an adenine base. The ligation was evaluated after 10 minutes and analyzed by PAGE gel. (**A**) Temperature variation with 1 mM ATP (**B**) ATP variation at 16 ℃. *Experimental conditions*: 2.8 µM of the 5'-phosphate pyrene strand, 1.4 µM of the fluorescein-labeled strand, and 1.4 µM of the template strands. T4 DNA ligase (400 CEU, 1 µL per 15 µL total volume) in 50 mM Tris-HCl, 10 mM $MgCl_2$ and 1 mM ATP concentrations.

3.2.2 Ligation Reactions Using a Pyrene Template and 5'-Phosphate Abasic or 5'-Phosphate Adenine Base Strands

We attribute the high efficiency and selectivity of ligation of the pyrene strand across from the abasic template to the steric fit of the pyrene:abasic pair. To determine whether the reverse scenario is also selective, the ligation of the 5'-phosphate abasic nucleotide or the 5'-phosphate

adenine nucleotide across from pyrene on the template strand, ligation reactions were performed using a template containing a pyrene across from the abasic (**Figure 3.6**). Interestingly, PBCV-1 DNA ligase selectively ligated the 5'-phosphate abasic strand on the pyrene-containing template while T4 DNA ligase did not. This result implies that the PBCV-1 DNA ligase can differentiate the mismatched bases pairs more selectively than the T4 DNA ligase.

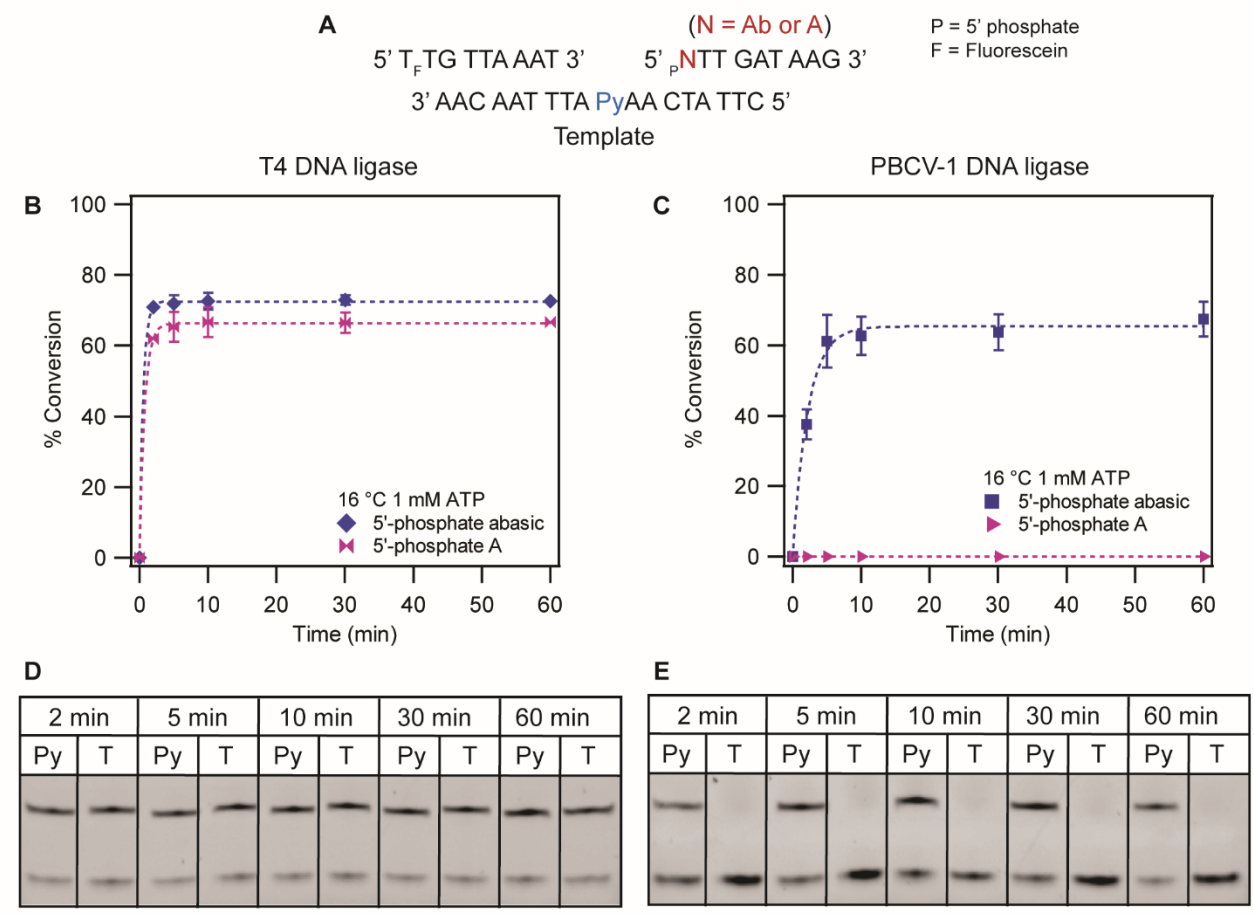


Figure 3.6 Ligation reaction of 5'-phosphate abasic strand or adenine base strand with the pyrene-containing template. (**A**) The DNA strands that were used in this experiment. (**B**) Time course of the ligation reaction by 400 CEU T4 DNA ligase (**C**) Time course of the ligation reaction by PBCV-1 DNA ligase (**D**) Representative PAGE gel of the ligation reaction by T4 DNA ligase. (**E**) Representative PAGE gel of the ligation reaction by PBCV-1 DNA ligase. *Experimental conditions*: 2.8 µM of the 5'-phosphate strand, 1.4 µM of the fluorescein-labeled strand, and 1.4 µM of the template strands. T4 DNA ligase (400 CEU, 1 µL per 15 µL total volume) or PBCV-1 DNA ligase (10.5 µM, 1.5 µL per 15 µL total volume) in 50 mM Tris-HCl, 10 mM $MgCl_2$ and 1 mM ATP concentrations.

To see whether 5'-phosphate abasic strand was selectively ligated across from the pyrene-containing template over the thymine base-containing template, similar ligation experiments were conducted. Unlike that observed in 5'-phosphate pyrene ligations, there was no selectivity observed for the 5'-phosphate abasic ligation. For both ligases, 5'-phosphate abasic nucleotides preferred to be ligated with the template with thymine nucleotide over the pyrene nucleotide, although the overall ligation yield was higher with T4 DNA ligase than PBCV-1 DNA ligase. We also altered the temperature and ATP concentration to see the effect on ligation efficiency and selectivity (**Figure 3.8**). Still, the 5'-phosphate abasic nucleotide showed a higher ligation yield on the template with the native thymine base, compared to that with the pyrene-containing template at lower temperatures (16 ℃ and 24 ℃) (**Figure 3.8A, B**). In contrast, at higher temperatures (30 ℃ to 42 ℃), the abasic nucleotide was ligated more efficiently on the pyrene template than the T-containing, which we attribute to the higher stability of the pyrene:abasic pair compared to the T:abasic pair. This trend was observed both with T4 DNA ligase and PBCV-1 DNA ligase. However, for PBCV-1 DNA ligase, which has an optimal operating temperature of 20 ℃, the ligation efficiency decreased significantly at 36 ℃, consistent with restricted activity at a higher temperature. Despite the interesting temperature variation results on selectivity, the changes in ATP concentration with T4 DNA ligase did not affect the selectivity of the abasic strand ligation on the pyrene or thymine template. At various ATP concentrations, the ligation of the 5'-phosphate abasic strand with both pyrene and thymine templates showed high % conversion (~80%, ~95% respectively) albeit reduced ligation with 10 mM ATP. These results revealed that the selectivity of the 5'-phosphate pyrene ligation was not mirrored for the 5'-phosphate abasic ligation under any conditions, indicating that the shape of the unnatural abasic-pyrene base pair was not the only reason for the selective ligation. This result could be related to the fact that T4 DNA ligase has less discrimination on mismatches or modification on the 5'-phosphate side. Sowers and coworkers

demonstrated that the base pair's geometry was more important than its relative stability, and also, the minor groove hydrogen bonding interactions had a significant impact on DNA ligase fidelity using non-hydrogen bonding pyrimidine shape analog.[228] Although there is no certain explanation or structural reason for these different selectivities between the 5'-phosphate pyrene strand and the 5'-phosphate abasic strand, it may have something to do with the adenylated DNA strand and its stability within DNA ligase.

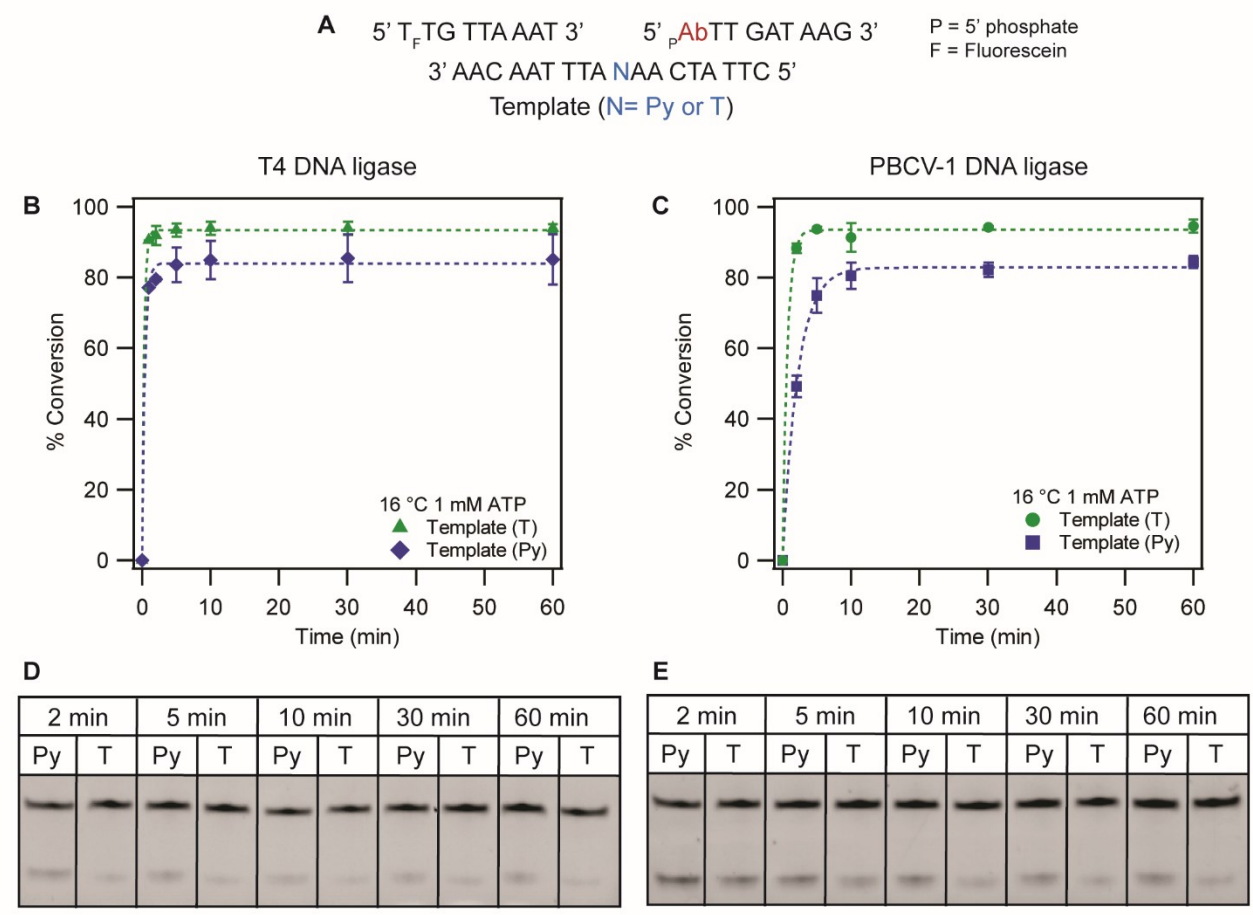


Figure 3.7 Ligation reaction of the 5'-phosphate abasic strand with the pyrene-containing template. (**A**) Templates and fragments sequence that was used in the experiments. (**B**) Time course of enzymatic ligation of pyrene nucleotides in the 5' position with 400 CEU T4 DNA ligase (**C**) with PBCV-1 DNA ligase (**D**) Representative PAGE gel image of stoichiometric ligation with 400 CEU T4 DNA ligase (**E**) Representative PAGE gel image of stoichiometric ligation with PBCV-1 DNA ligase. *Experimental conditions*: 2.8 μM of the 5'-phosphate abasic strand, 1.4 μM of the fluorescein-labeled strand, and 1.4 μM of the template strands. T4 DNA ligase (400 CEU, 1 μL per 15 μL total volume) or PBCV-1 DNA ligase (10.5 μM, 1.5 μL per 15 μL total volume) in 50 mM Tris-HCl, 10 mM $MgCl_2$ and 1 mM ATP concentrations at 16 °C.

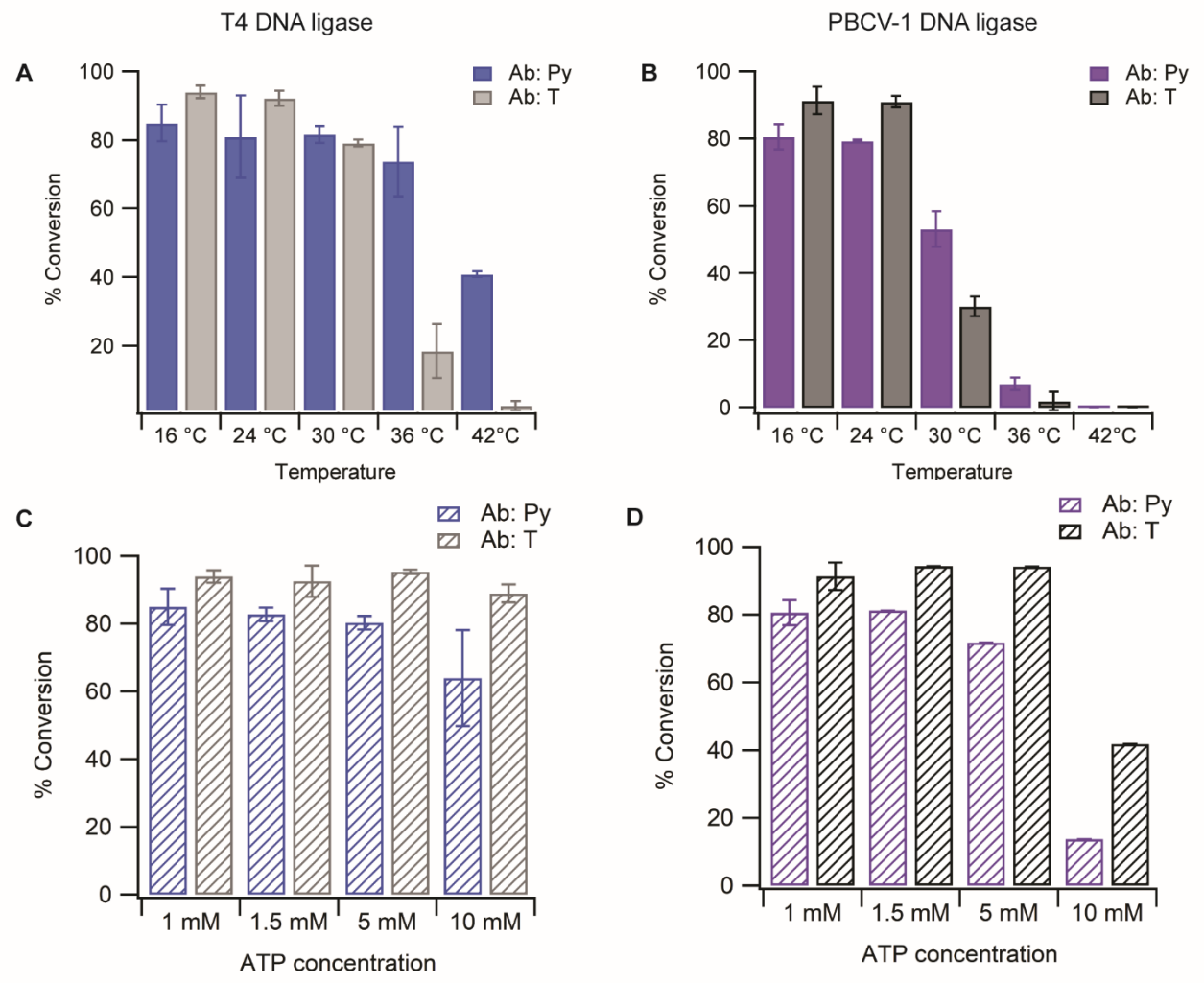


Figure 3.8 Comparison of T4 DNA ligase and PBCV-1 DNA ligase on the ligation reaction of the 5'-phosphate abasic strand to the template with either pyrene or thymine. The ligation was evaluated after 10 minutes and analyzed by PAGE gel. (**A**) Temperature variation with 1 mM ATP with T4 DNA ligase (**B**) Temperature variation with 1 mM ATP with PBCV-1 DNA ligase (**C**) ATP variation at 16 ℃ with T4 DNA ligase. (**D**) ATP variation at 16 ℃ with PBCV-1 DNA ligase. *Experimental conditions*: 2.8 µM of the 5'-phosphate abasic strand, 1.4 µM of the fluorescein-labeled strand, and 1.4 µM of the template strands. T4 DNA ligase (400 CEU, 1 µL per 15 µL total volume) or PBCV-1 DNA ligase (10.5 µM, 1.5 µL per 15 µL total volume) in 50 mM Tris-HCl, 10 mM $MgCl_2$ and 1 mM ATP concentrations.

3.2.3 Selectivity of 5'-Phosphate Pyrene Ligation on Variable Templates

To utilize pyrene ligation as a method of detecting abasic lesions, the selectivity across from all of the canonical bases must be evaluated. Therefore, to further establish the selectivity of

ligation of 5'-phosphate pyrene across from an abasic lesion, five different templates containing the natural bases (A, T, G, and C) or the abasic lesion were tested in the ligation of the 5'-phosphate pyrene terminated strand (**Figure 3.9**). When a standard condition of ATP (1 mM) was used with T4 DNA ligase, ligation was observed in all five templates. Based on the previous experiments, which showed no ligation with the adenine-containing template, the ligation was tested with T4 DNA ligase at 16 ℃ in 10 mM ATP concentration. Minor amounts of ligation were observed with thymine template (~10%) and guanine template (< 5%) in 10 mM ATP, while the pyrene nucleotide only exhibited efficient ligation with the abasic site template when the ATP concentration was increased to 15 mM (**Figure 3.9A**). Next, ligation of the pyrene strand on the native base containing templates was examined at a higher temperature of 36℃ (**Figure 3.9B**). Similar to the result with 10 mM ATP at 16 °C, a small amount of ligation with other bases (T and G) was observed at 36 ℃ and 1 mM ATP. When increased ATP concentration (10 mM) was used at 36 ℃, the overall ligation was reduced dramatically, and ligation of the pyrene strand was only for a nicked duplex with the template containing the abasic site. Next, we explored the selectivity of 5'-phosphate pyrene ligation using PBCV-1 DNA ligase under standard conditions, 1 mM ATP, and 16℃ **(Figure 3.10**). For this enzyme, excellent efficiency was observed with the abasic template, while there was negligible ligation for the reactions using templates with the other bases (less than 5 % or none). These experiments confirmed the outstanding selectivity of the pyrene nucleotide for the abasic lesion over the natural bases with two different types of DNA ligases.

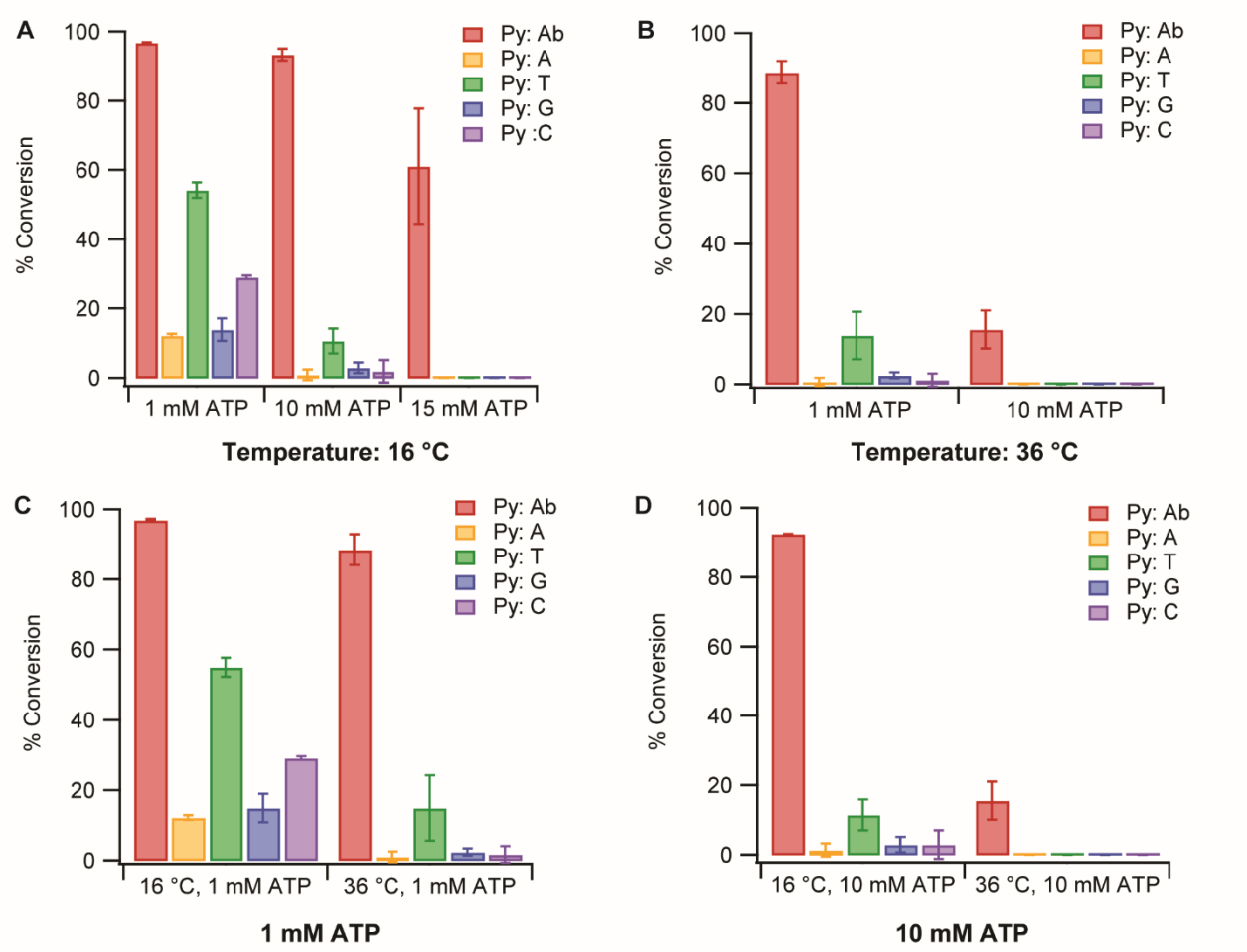


Figure 3.9 Ligation efficiencies of the unnatural pyrene base at the ligation site with abasic and natural bases with T4 DNA ligase at 10 minutes. (**A**) Comparison of ATP concentration, 1mM, 10 mM, versus 15 mM at 16 °C. (**B**) Comparison of ATP concentrations, 1 mM versus 10 mM at 36 °C. (**C**) Temperature comparison at 1 mM ATP (**D**) Template comparison at 10 mM ATP. *Experimental conditions*: 2.8 µM of the 5'-phosphate pyrene strand, 1.4 µM of the fluorescein-labeled strand, and 1.4 µM of the template strands. T4 DNA ligase (400 CEU, 1 µL per 15 µL total volume) in 50 mM Tris-HCl, 10 mM $MgCl_2$ and various ATP concentrations.

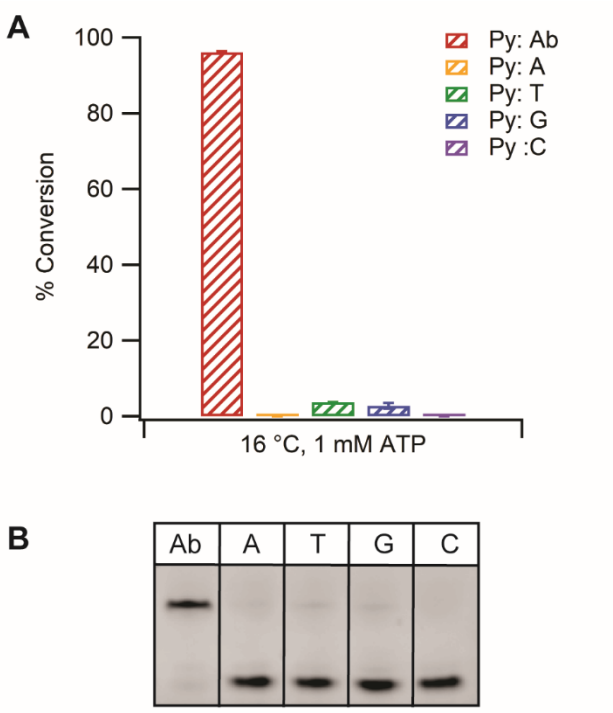


Figure 3.10 Ligation efficiencies of the unnatural pyrene base at the ligation site with abasic and natural bases with PBCV-1 DNA ligase at 10 minutes. (**A**) Standard ATP concentration (1 mM) at 36 °C. *Experimental conditions*: 2.8 µM of the 5'-phosphate pyrene strand, 1.4 µM of the fluorescein-labeled strand, and 1.4 µM of the template strands. PBCV-1 DNA ligase (10.5 µM, 1.5 µL per 15 µL total volume) in 50 mM Tris-HCl, 10 mM $MgCl_2$ and 1 mM ATP concentrations.

3.2.4 Ligation Efficiency Comparison

Next, we examined the ligation efficiency of the pyrene:abasic pair compared to the native base pairs. The time course of the ligation of 5'-phosphate thymine nucleobases with adenine base containing template was plotted together with the 5'-phosphate pyrene ligations with the abasic template. As shown in **Figure 3.11**, the rate of the 5'-phosphate pyrene ligation was as high as the one with unmodified thymine nucleobase strands. However, the reverse sequence condition, where 5'-phosphate abasic strand ligated across from the pyrene template, was not as fast as native A:T base ligation. This result may indicate that the 5'-phosphate pyrene strand was not inhibited by the bulk of the bases during the ligation mechanism, such as adenylation and AMP transfer. However,

the pyrene template seemed to affect the ligation steps due to either lack of hydrogen bonding or the bulkiness of the base since the ligation of the 5'-phosphate abasic strand across from the adenine nucleotide was as fast as native A:T base pair ligation (shown in **Figure 3.16**).

Under the same conditions, the ligation selectivity of the 5'-phosphate pyrene strand was compared to the ligation of the 5'-phosphate thymine strand across from various canonical bases, including abasic. T4 DNA ligase did not differentiate the perfect T:A base pair from the mismatched base pairs at the ligation junction at 16 °C with 1 mM ATP. Increased ATP concentration (10 mM) did improve the selectivity of T:A base pair (95 % conversion) over the ligation across from the abasic. However, there was still high percent ligation with T:T and T:G base pair (83%, 95%, respectively). Unlike PBCV-1 DNA ligase selectively ligated 5'-phosphate pyrene end strand across from the abasic, the ligase exhibited low fidelity with native bases mismatches (**Figure 3.12**). This result suggests that the geometry fit of pyrene:abasic pair plays a role in the ligation mechanisms to some extent.

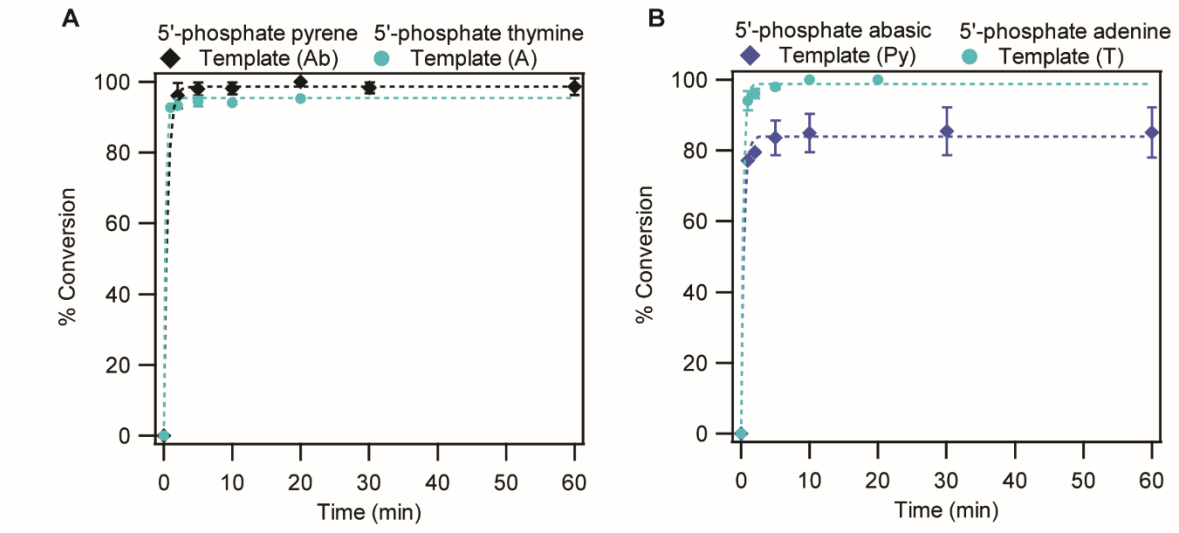


Figure 3.11 Ligation comparison with Py:Ab base pair with A:T base pair. *Experimental conditions*: 2.8 µM of the 5'-phosphate strand, 1.4 µM of the fluorescein-labeled strand, and 1.4

μM of the template strands. T4 DNA ligase (400 CEU, 1 μL per 15 μL total volume) in 50 mM Tris-HCl, 10 mM $MgCl_2$ and 1 mM ATP concentrations at 16 °C.

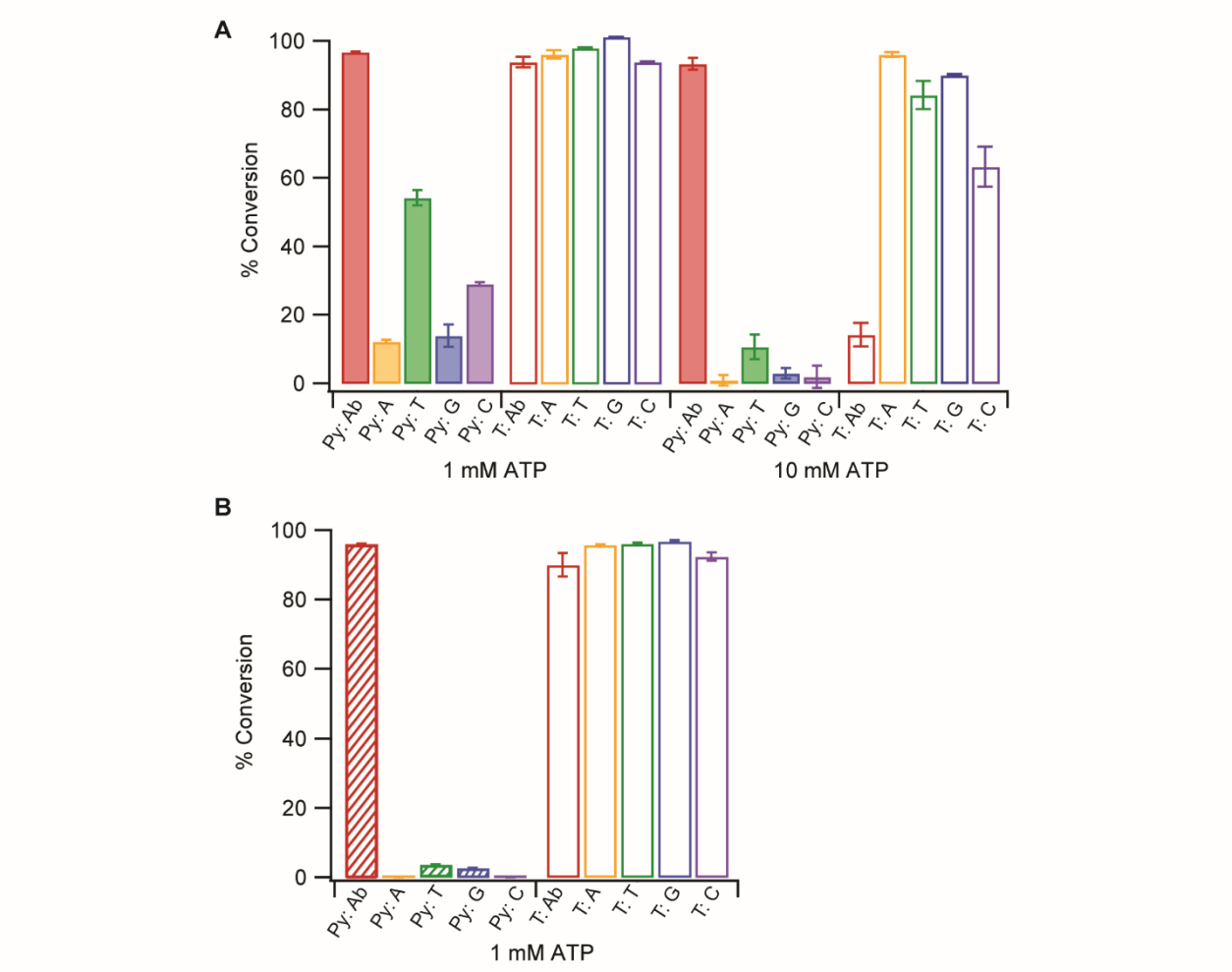


Figure 3.12 Selectivity comparison with 5'-phosphate pyrene strands with 5'-phosphate thymine strands with various templates. *Experimental conditions*: 2.8 μM of the 5'-phosphate strand, 1.4 μM of the fluorescein-labeled strand, and 1.4 μM of the template strands. T4 DNA ligase (400 CEU, 1 μL per 15 μL total volume) in 50 mM Tris-HCl, 10 mM $MgCl_2$ and 1 mM ATP concentrations at 16 °C. Data were evaluated in 10 minutes.

3.2.5 Competitive Ligation Reaction

In order to further investigate the selectivity of the ligation of a 5'-phosphate containing a pyrene-modified nucleotide for the template containing the “complementary" abasic group, a competitive ligation reaction was performed at 16 ℃ using phosphate strands containing either the terminal pyrene or a thymine nucleobase. The 5’-phosphate thymine strand that was complementary to the adenine-containing template was prepared with three extra nucleotides. As such, when both 5’-phosphate modified strands (T or pyrene) were added, the ligated product was distinguished based on the ligated product length (18 nt for the pyrene strand vs 21 nt for the thymine strand). The sum of the added templates was kept at 1.4 μM while each of the 5’-phosphate strands and the fluorescein-labeled 3’-OH strand were 2.8 μM and 1.4 μM, respectively. Each lane contained a different combination of DNA fragments and templates to see whether the selectivity of pyrene nucleotides could be achieved in a mixture of DNA substrates (**Figure 3.11**). Two different ATP concentrations (10 mM and 15 mM) were tested with T4 DNA ligase, while the standard ATP concentration (1 mM) for PBCV-1 DNA ligase was used. For the ligation of the pyrene strand in the presence of the abasic template, the ligation yield was higher in 10 mM ATP with T4 DNA ligase and in 1 mM ATP with PBCV-1 DNA ligase (lanes 1, 2, and 7). Moreover, the pyrene nucleotide was only ligated in the presence of the abasic template (lanes 1 and 7) and not in the presence of the adenine template. In the presence of both 5’-phosphate strands, the ligated product was completely determined based on the templates that were present, illustrating that ligation was very selective under competitive conditions (lanes 2 and 3). However, in the absence of competition, the fidelity of the thymine terminated strand was poor owing to templated ligation with the abasic template except when 15 mM ATP and T4 DNA ligase were used (lane 5). Interestingly, although there was only 0.7 μM of the abasic template in the experiment with both

abasic and A-containing templates (lane 7), ligation of more than 0.7 μM of the abasic strand (i.e., % conversion > 50%) was observed (for T4 DNA ligase and 15 mM ATP, 51%; for T4 DNA ligase and 10 mM ATP, 94%; for PBCV-1 ligase and 1 mM ATP, 76%). This suggests turnover occurs in the reaction using the abasic-containing template and the 5'-phosphate pyrene strand rather than some nonselective templated ligation using the A-containing templates as no ligation was observed in the presence of only the A-containing template. In summary, such a competition platform appears to be a useful strategy for abasic lesion detection based on the ligation of 5'-phosphate pyrene nucleotides. For example, SNPs, single nucleotide polymorphisms, have been detected in approaches involving the first ligation that is templated by the target sequence followed by some kind of amplification of the ligated product.[243–245] We can envision such a strategy using ligation of a pyrene-terminated probe followed by the amplification of the ligation product.

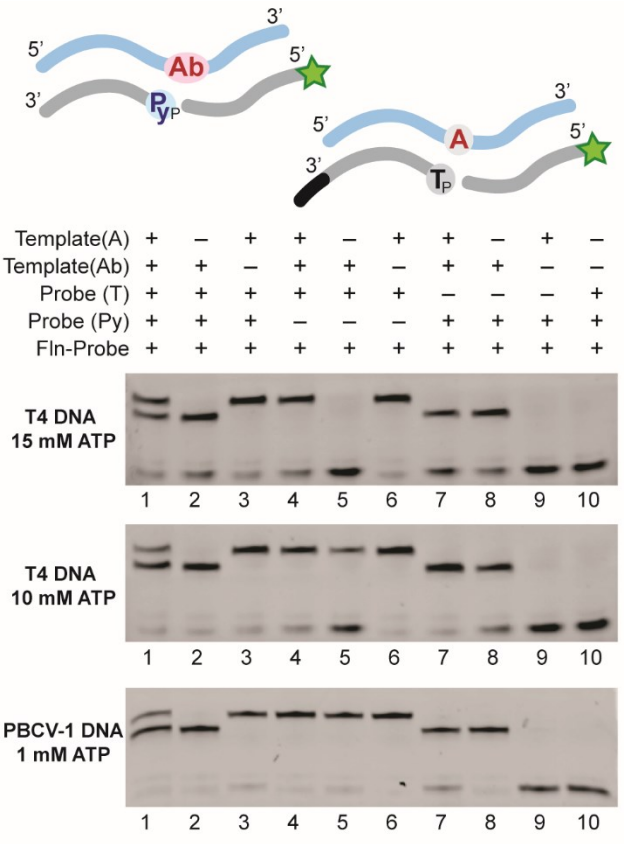


Figure 3.13 A ligation competition experiments. The reaction was evaluated after 10 min at 16 °C. *Experimental conditions*: 2.8 μM of the 5'-phosphate strand, 1.4 μM of the fluorescein-labeled strand, and 0.7 μM or 1.4 μM of the template strands. The sum of the templates added was consistent as 1.4 μM, T4 DNA ligase (400 CEU, 1 μL per 15 μL total volume) or PBCV-1 DNA ligase (10.5 μM, 1.5 μL per 15 μL total volume) in 50 mM Tris-HCl, 10 mM $MgCl_2$. ATP concentration was indicated in the figure.

3.3 Pyrene as A Bulky Destabilizing Agent in Destabilization-Induced Isothermal DNA Amplification

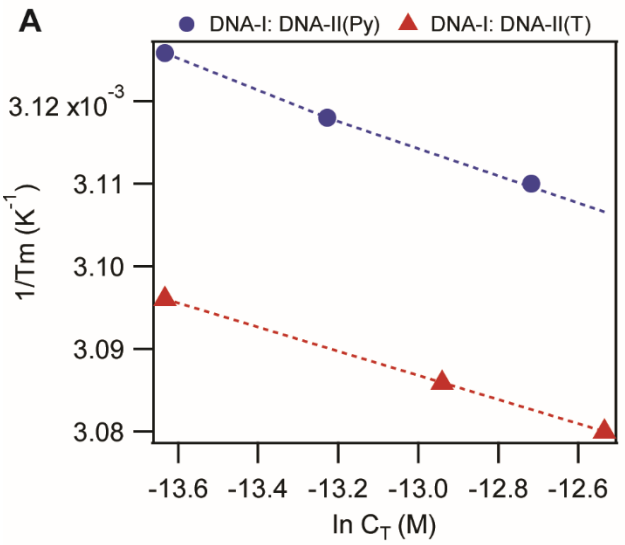
Once we confirmed that T4 DNA tolerated the bulky pyrene base at the ligation site, the question was raised whether the pyrene-modified base can act as a destabilizing group allowing for self-replication following a similar mechanism as our lesion-induced DNA amplification (LIDA), which uses an abasic group as the destabilizing lesion. As mentioned in the introduction and

Chapter 2, our group has found that incorporating a destabilizing group in a ligase chain reaction using four primers or fragments reduced product inhibition, triggering turnover in templated ligation reactions.[43] Moreover, studies on how lesion-induced DNA amplification (LIDA) overcomes product inhibition were discussed in chapter 2. It was found that the destabilizing abasic group helped to reduce product inhibition with the aid of the T4 DNA ligase enzyme. Therefore, to evaluate the pyrene nucleotide as a destabilizing agent, thermodynamic studies were first investigated in the absence of enzymes to determine how much the bulky pyrene group destabilized the product duplex. Specifically, the melting profile of the pyrene-modified product duplex was obtained and compared to the one with the native product duplex (**Table 3.1**). Unlike the abasic destabilizing group, an only minor decrease in the melting temperature (T_m) in the pyrene-containing product duplex was observed (-3 ℃ compared to the unmodified duplex). This result agreed with previously reported values of pyrene:canonical base pairs.[216] This small decrease in T_m resulted from the stabilizing effect of the π stacking of the pyrene base with other natural bases despite its bulk. The thermodynamic parameters of duplex formation were then obtained to understand the stability differences of the modified duplex using the thermal denaturation analysis and the van't Hoff method (**Figure 3.12**).[151] It was revealed that the pyrene-modified duplex exhibited a reduction by only one order of magnitude in K_A at 30 ℃ compared with the complementary duplex. Nevertheless, the effect of the pyrene nucleotide in LIDA was explored using the same cross-catalytic strategy we have previously optimized for the system with the 5'-abasic phosphate strand (**Figure 3.13**).

Table 3.1 The melting temperatures of DNA complexes.

DNA complex	T_m (°C)		
Product duplex (T)	0.6 µM	1.2 µM	1.8 µM
DNA-I: DNA-II(T)	49.84 ± 0.05	50.90 ± 0.03	51.52 ± 0.03
Product duplex (D = pyrene) DNA-I: DNA-II(D)	0.6 µM	0.9 µM	1.5 µM
	46.77 ± 0.06	47.57 ± 0.04	48.39 ± 0.04

Condition: 50 mM Tris-HCl buffer (pH 7.5 at 25 °C) with 10 mM $MgCl_2$.
The T_m was obtained from the Gaussian fit from the first derivative of absorbance at 260 nm, and the error is the standard deviation returned from the Gaussian fit.



B

	Product duplex (T)	Product duplex (Pyrene)
ΔH (kJ•mol^{-1})	- 570 ± 3	- 480 ± 30
ΔS (kJ•mol^{-1}•K^{-1})	-1.652 ± 0.007	-1.39 ± 0.09
K_A (30 °C) (M^{-1})	9.0 ± 0.4 x 10^{11}	2 ± 1 x 10^{10}

Figure 3.14 Thermodynamic analysis of the pyrene-containing product duplex compared to fully complementary duplex. (**A**) Thermodynamic data was plotted $1/T_m$ (K^{-1}) versus ln C_T (DNA complexes concentration). The plot shows the concentration dependence of the melting temperature. The filled blue circle (pyrene) or the red diamond (T base) indicates the product duplexes. (**B**) The table for thermodynamic parameters comparison.

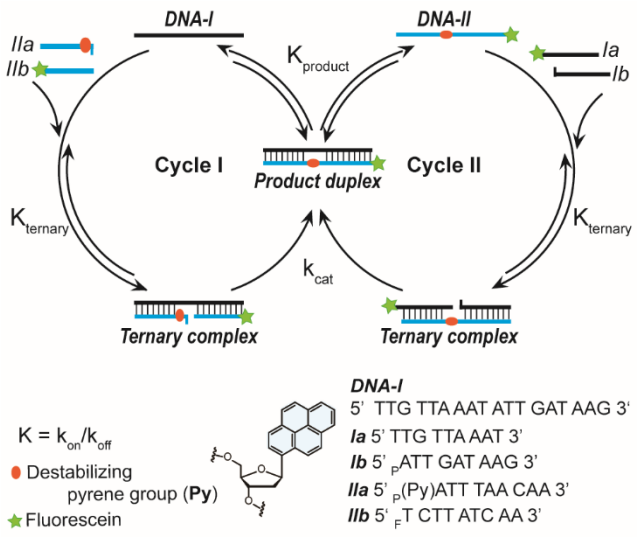


Figure 3.15 Schematic representation of isothermal lesion-induced DNA amplification (LIDA). The destabilizing pyrene group is located in 5'-phosphate of **DNA-IIa$_P$**. Cycle 1: A **DNA-I** strand templates the formation of **DNA-II** via hybridization with IIa and IIb. Cycle 2: Ligated **DNA-II** strand containing destabilizing group catalyzes the formation of **DNA-I** using the fluorescent-labeled strand (**Ia**) and the fragment **Ib**. P represents the 5'-phosphate group. F represents the fluorescein group. T4 DNA ligase allows the ligation of ternary complexes. The formation of fluorescein-labeled strands measures the replication of DNA-I

3.3.1 Single Cycle Reactions

With lesion-induced DNA amplification (LIDA), as described in Chapter 2, the strategy involves a cross-catalytic reaction where the products of each cycle act as templates for the other reaction cycle (**Figure 3.13**). This method requires the addition of two fragments that form the template (**DNA-I**), and two other fragments that compose the complementary strand and contain one or more destabilizing groups (**DNA-II**). The limiting reagent is one of the fluorescein-labeled fragments (**Ia**) to detect the amplification of **DNA-I**. In the presence of T4 DNA ligase, the continuous ligation of the intermediate nicked complex consisting of two fragments hybridized to a template occurs to generate the template of the other cycle. In the original DNA strand system involving the abasic group, the exponential replication of the **DNA-I** template was observed within 45 minutes at 30 °C when the initial template (14 nM) was added.

For cross-catalysis using pyrene as the destabilizing group, in cycle 1 **DNA-I** must hybridize with the pyrene-containing fragment (**IIa$_P$**) and the other complementary strand (**IIb**) followed by ligation and dissociation. In contrast, in cycle 2, **DNA-II** containing the pyrene nucleotide acts as the template strand and forms the nicked intermediate complex with **Ia** and **Ib$_P$** strands. Therefore, a prerequisite of LIDA is that the destabilizing modification facilitates ligation when at the 5'-phosphate position of the nick or across from the nick in the template. To see whether pyrene could be effectively ligated across from a purine nucleotide, which is predicted to lead to the greatest amount of destabilization, we compare the performance of the pyrene-containing system to a standard system that we have previously explored that uses the abasic group as a destabilizing group. The DNA-templated ligation reactions of each cycle were separately tested at 16 ℃ with a low concentration of T4 DNA ligase (1 Weiss unit) with stoichiometric amounts of the template and the fluorescein-labeled strand (1.4 μM). The ligation for cycle 1, where the 5'-phosphate pyrene strand was used, resulted in inefficient ligation (~10% conversion of the limiting strand), while cycle 2 where it used pyrene-modified templated ligation, showed a higher ligation yield (~ 80%) after 8 minutes (**Figure 3.14**). This result suggested that the pyrene modification on 5'-phosphate sites affected the ligation much more than when it was located in the template. In contrast, the abasic strand showed near-complete % conversion within five minutes in both cycles, although cycle 2 exhibited a slightly higher % yield (88% for cycle 1 and 96% for cycle 2, after 4 min). Our group's previous report showed that the enzyme's concentration had a high impact on the single-cycle turnover ligation experiment, where substoichiometric amounts of template were used.[144] Additionally, increasing the enzyme concentration above 1 Weiss unit resulted in much faster replication of the template in LIDA via cross-catalysis in the presence of all four fragments.[143,144] Therefore, the ligation was tested again with a high enzyme concentration (2000 CEU, approximately 7 Weiss units). With the 2000 CEU enzyme, templated ligations efficiencies

of each cycle were improved, especially in the pyrene templated ligation (cycle 2), which displayed near-complete ligation within five minutes. Even at higher enzyme concentration, cycle 1 ligation was slow, but a significantly improved % conversion was demonstrated with a higher concentration of enzyme (80% after 120 min). The successful ligation of the pyrene-modified strands, albeit at a slower rate than that observed for the abasic system, suggested that the pyrene-modified strand could be applied in a cross-catalytic reaction. Therefore, next, the conditions of the cross-catalytic reaction were explored with the pyrene-modified strands.

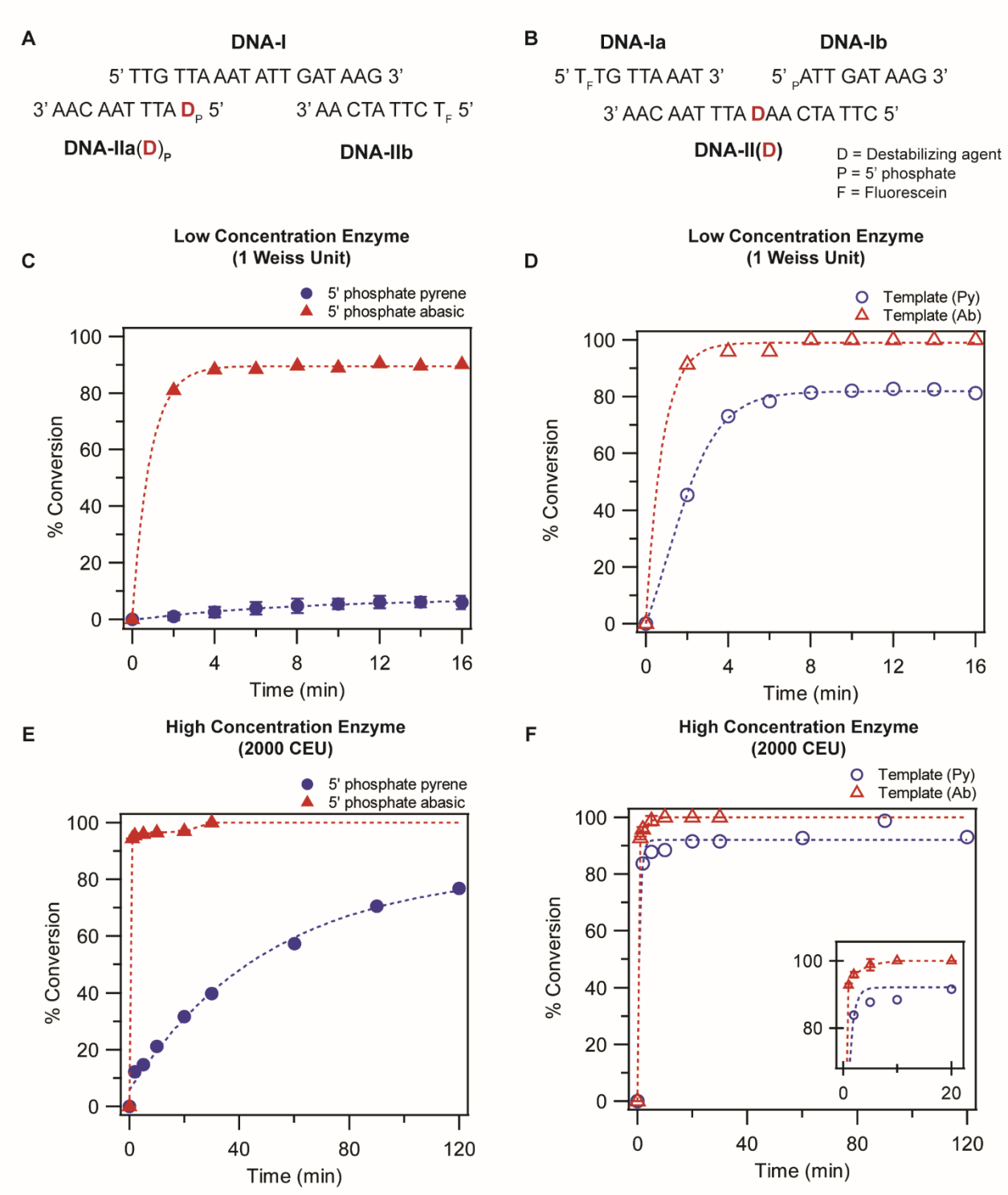


Figure 3.16 Templated ligation reaction with the abasic modified or pyrene-modified strands. (**A**) Sequence information for cycle 1. (**B**) Sequence information for cycle 2. (**C**) Ligation reaction with 5'-phosphate strand with destabilizing group (D) with 1 Weiss unit T4 DNA ligase (**D**) Ligation reaction with the destabilizing group containing template with 1 Weiss unit T4 DNA ligase. (**E**) Ligation reaction with 5'-phosphate strand with destabilizing group (D) with 2000 CEU

T4 DNA ligase. (**F**) Ligation reaction with the destabilizing group containing template with 2000 CEU T4 DNA ligase. *Experimental condition:* 2.8 μM of 5'-phosphate strands, 1.4 μM of fluorescein-labeled strands, and 1.4 μM of the templates. Low concentrations of T4 DNA ligase from Invitrogen (1 Weiss unit, 1 μL in total 15 μL volume) were used with a provided ligation buffer. High concentrations of T4 DNA ligase from New England Biolabs (2000 CEU, 1 μL in total 15 μL volume) were used with a provided ligation buffer.

3.3.2 Cross-Catalytic Reaction on Various ATP Concentrations

The pyrene-modified strands were tested in the standard condition of LIDA for this sequence with abasic as a destabilizing group, which uses 2000 CEU and 1 mM ATP at 30 ℃ with an initial 14 nM or 0 nM of the template **DNA-I**.[144] As shown in **Figure 3.17A**, the replication of the template (**DNA-I**) with the pyrene-modified strands was not efficient, implying that the conditions needed to be optimized due to the usage of different destabilizing groups. As the ATP concentration is one of the factors to optimize the reaction, three varying concentrations of ATP were tested. Interestingly, there was no difference observed between templated (14 nM) and non-templated (0 nM) reactions for all three ATP concentrations. The 100 μM ATP showed the linear increase of the ligation, whereas 10 μM ATP concentration exhibited a relatively fast increase for the first 60 minutes and leveled off. Surprisingly, extra product bands were observed in 100 μM and 10 μM ATP concentrations (**Figure 3.18B, C**). It was first assumed that the stability of the 5'-overhang pyrene base with other nucleotides might cause this polymerization-like activity. Therefore, to evaluate this unique observation, the fully complementary strand, which does not have any modification, was evaluated in the same experiments for comparison.

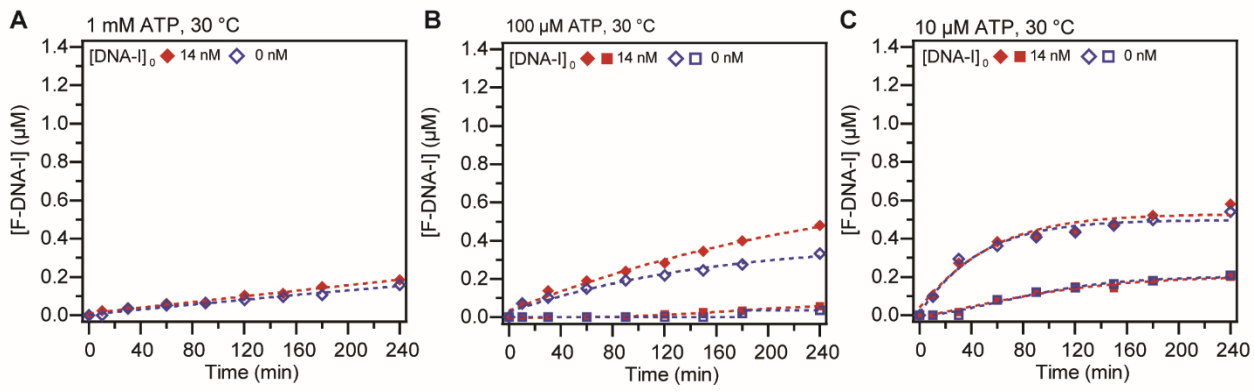


Figure 3.17 Cross-catalytic replication of DNA-I with pyrene modified strand (IIa$_P$) at 30 °C. (**A**) 1 mM ATP concentration. (**B**) 100 μM ATP concentration. The extra bands were described as square labeling. (**C**) 10 μM ATP concentration. The extra bands were described as square labeling. *Experimental conditions*: 2.8 μM of the pyrene-modified strand (**IIa$_P$**), **Ib**, and **IIb**. 1.4 μM of the fluorescein-labeled strand (**Ia**). 14 nM (filled red diamond) or 0 nM (empty blue diamond) of the initial **DNA-I**. T4 DNA ligase (2000 CEU per 15 μL total volume) in 50 mM Tris-HCl, 10 mM $MgCl_2$ and various ATP concentrations.

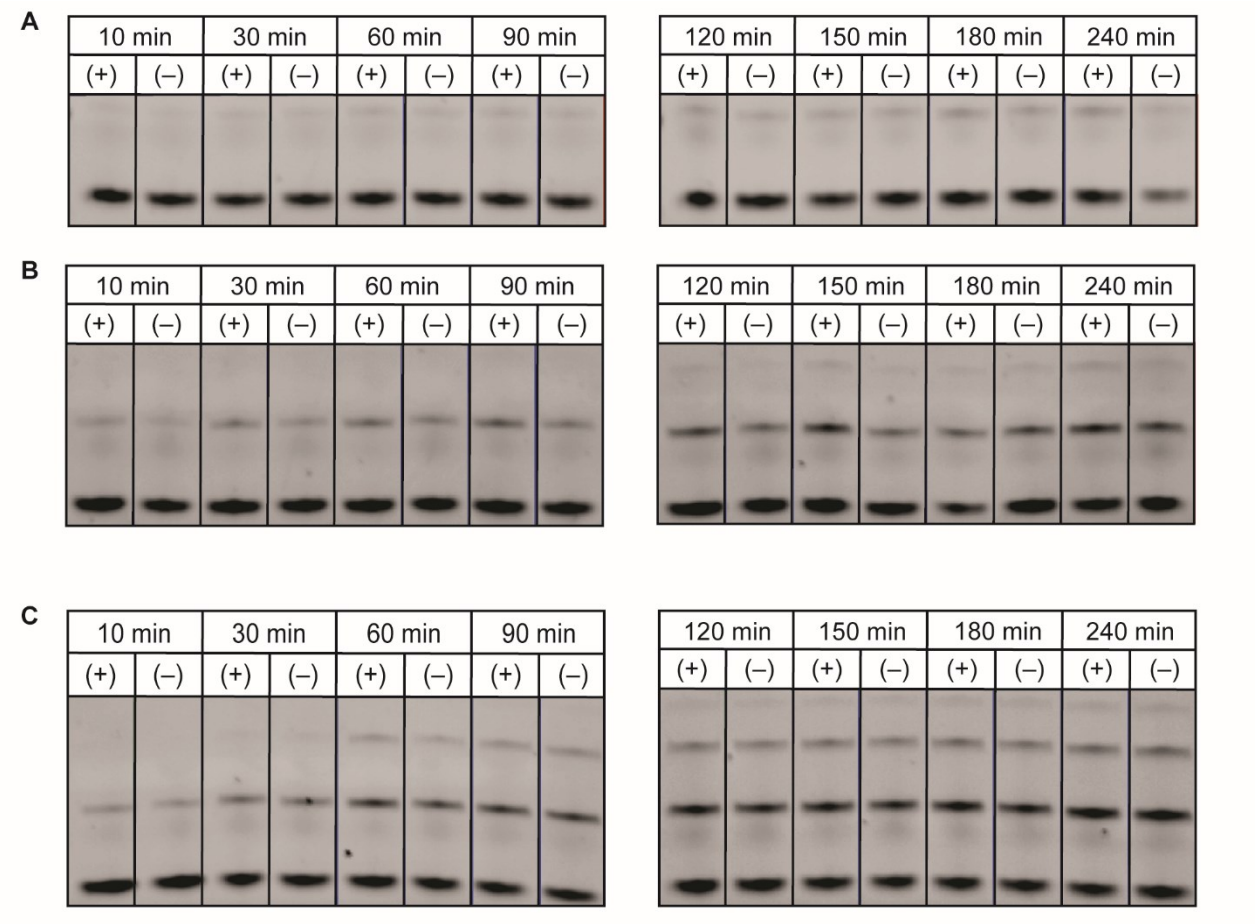


Figure 3.18 Representative PAGE image of cross-catalytic reaction with 5'-phosphate pyrene strand (DNA-IIaP (pyrene)). Figure 3.17 with (**A**) 1 mM ATP, (**B**) 100 uM ATP, and (**C**) 10 uM ATP. (+) initial template (14 nM), (-) no initial template (0 nM).

A native DNA system that contained the complementary thymine nucleobase, instead of the destabilizing group, was evaluated as the comparison tool in the same condition (**Figure 3.19**). The replication of **DNA-I** at 30 ℃ in various ATP concentrations had slow parabolic growth both in templated and non-templated reactions that are attributed to ligation of the four primers, which form a one-base overhang formed between the hybridized. This result at 30 ℃ in 1 mM ATP was in agreement with the previous report.[144] Very similar kinetic profiles between the ATP variations were observed. This outcome suggested that the ligation efficiency with the complementary strands was less influenced by the ATP variation. Moreover, the cross-catalysis controls with the fully

complementary strands lacking any destabilizing group revealed that there was no polymerization-like activity at lower ATP concentrations.

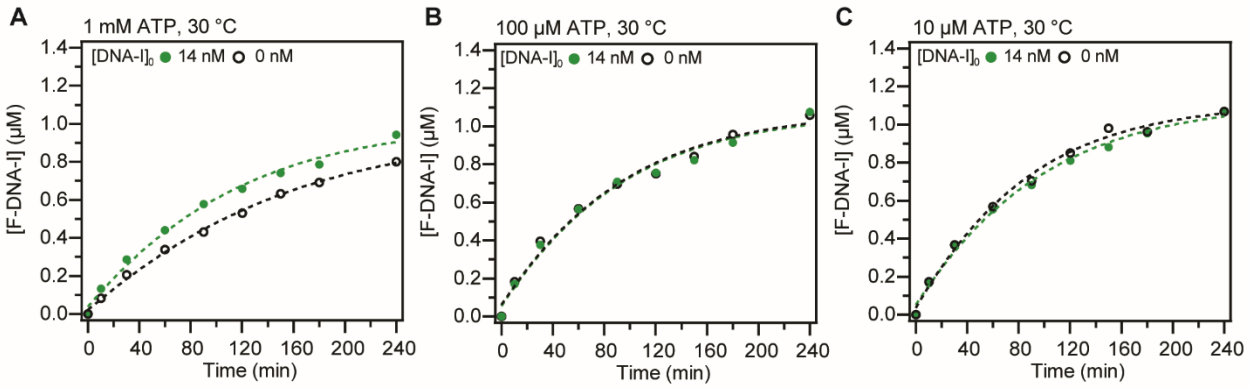


Figure 3.19 Cross-catalytic replication of DNA-I with the complementary 5'-phosphate thymine base strand at 30 °C. (**A**) 1 mM ATP concentration. (**B**) 100 μM ATP concentration. (**C**) 10 μM ATP concentration. *Experimental conditions*: 2.8 μM of the pyrene-modified strand (**IIa$_P$**), **Ib**, and **IIb**. 1.4 μM of the fluorescein-labeled strand (**Ia**). 14 nM (filled green circle) or 0 nM (empty black circle) of the initial **DNA-I**. T4 DNA ligase (2000 CEU per 15 μL total volume) in 50 mM Tris-HCl, 10 mM $MgCl_2$ and various ATP concentrations.

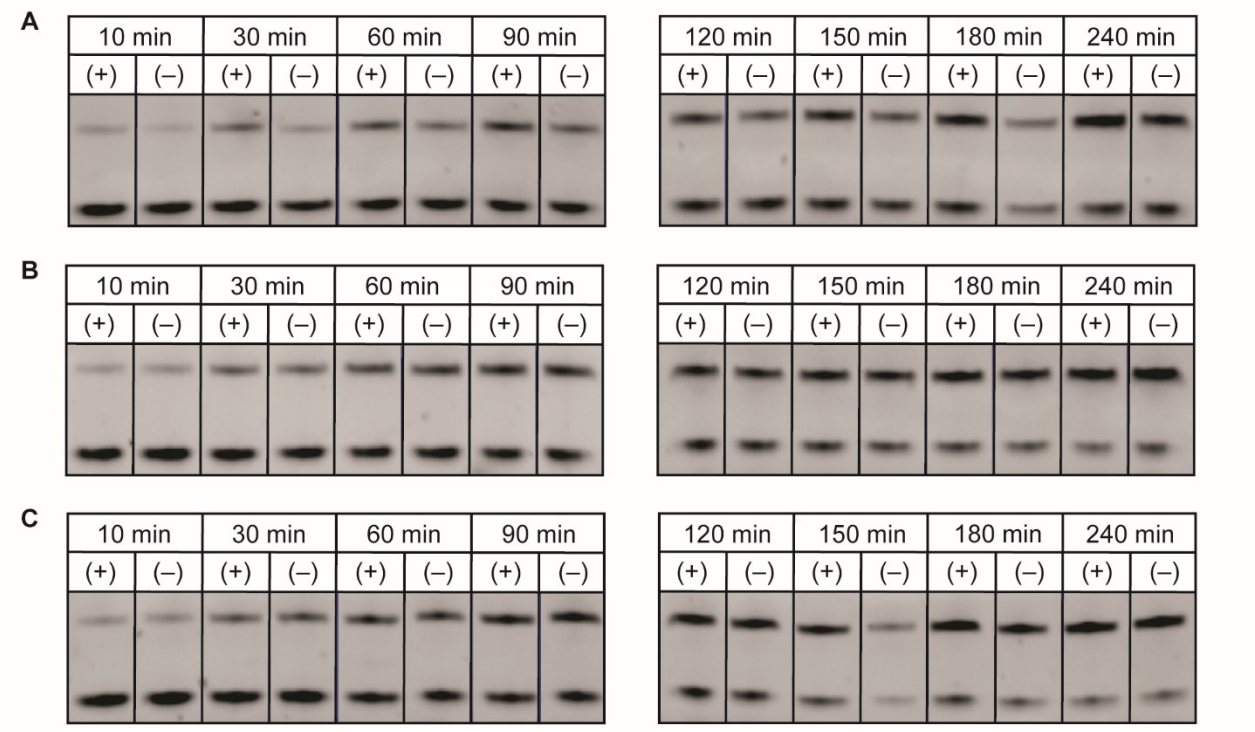


Figure 3.20 Representative PAGE image of cross-catalytic reaction with 5'-phosphate thymine strands (DNA-IIaP). Figure 3.19 with (**A**) 1 mM ATP, (**B**) 100 uM ATP, and (**C**) 10 uM ATP. (+) initial template (14 nM), (-) no initial template (0 nM).

3.3.3. Cross-Catalytic Reaction on Various Temperature

The minor decrease of the melting temperature of the pyrene-modified strands implied that the optimal temperature of the cross-catalytic reaction needed to be increased. Hence, increased temperatures were tested in LIDA with both pyrene-modified strands and non-modified strands using 100 μM ATP since it exhibited better replication than 1 mM ATP at 30 °C with pyrene-modified strands. Excitingly, as shown in **Figure 3.21**, cross-catalysis of the pyrene-modified DNA system exhibited an S-shape of growth at 34 °C and 36 °C indicative of self-replication as well as different amplification between templated and non-templated reactions. However, at higher temperatures (38 °C) the slow linear growth was observed. These experiments revealed that the optimal temperature in a pyrene-containing LIDA system is about 36 °C, which is 11 °C below the

product duplex melting temperature (47 °C). This observation agrees with finding the optimal temperature for different DNA systems that the ideal replication temperature is ~10 °C below the product duplex's melting temperature.[144] Next, the cross-catalytic reaction with the pyrene modified strands was performed with PBCV-1 DNA ligase at 30 °C and 36 °C since the PBCV-1 DNA ligase was able to ligate the pyrene modified strand across from the abasic group as shown earlier in the chapter (**Figure 3.20**). However, no ligation was observed with PBCV-1 DNA ligase. This result indicates that although PBCV-1 DNA ligase tolerates the pyrene nucleotide at the ligation site, it is not the optimal enzyme to use in this system that requires ligation of a pyrene:adenine destabilized pair.

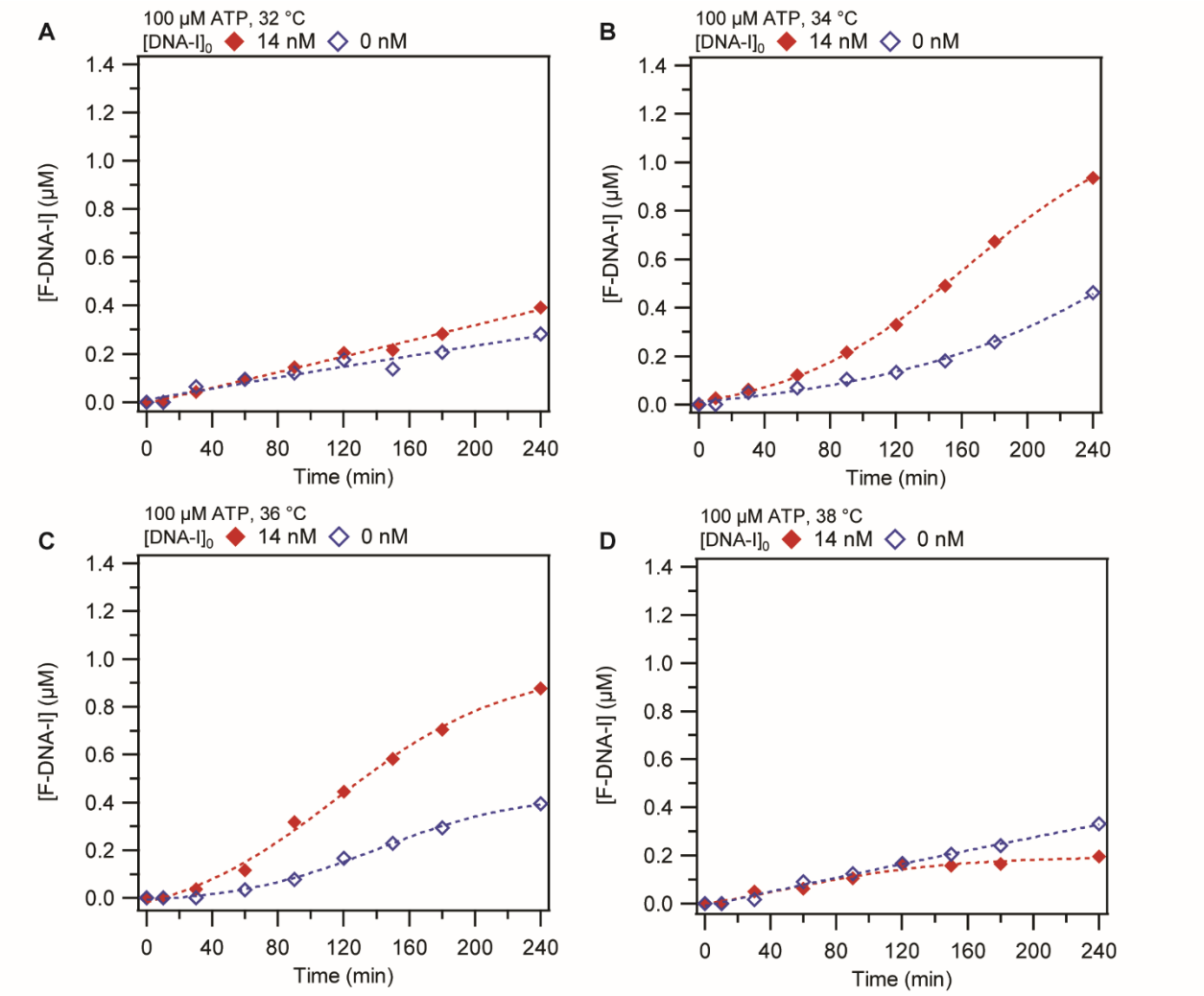


Figure 3.21 Cross-catalytic replication of DNA-I with pyrene modified strands at various temperatures. (**A**) 32 ℃ (**B**) 34 ℃ (**C**) 36 ℃ (**D**) 38 ℃. *Experimental conditions*: 2.8 µM of the pyrene-modified strand (**IIa$_P$**), **Ib**, and **IIb**. 1.4 µM of the fluorescein-labeled strand (**Ia**). 14 nM (filled red diamond) or 0 nM (empty blue diamond) of the initial **DNA-I**. T4 DNA ligase (2000 CEU per 15 µL total volume) in 50 mM Tris-HCl, 10 mM $MgCl_2$ and 100 µM ATP concentrations.

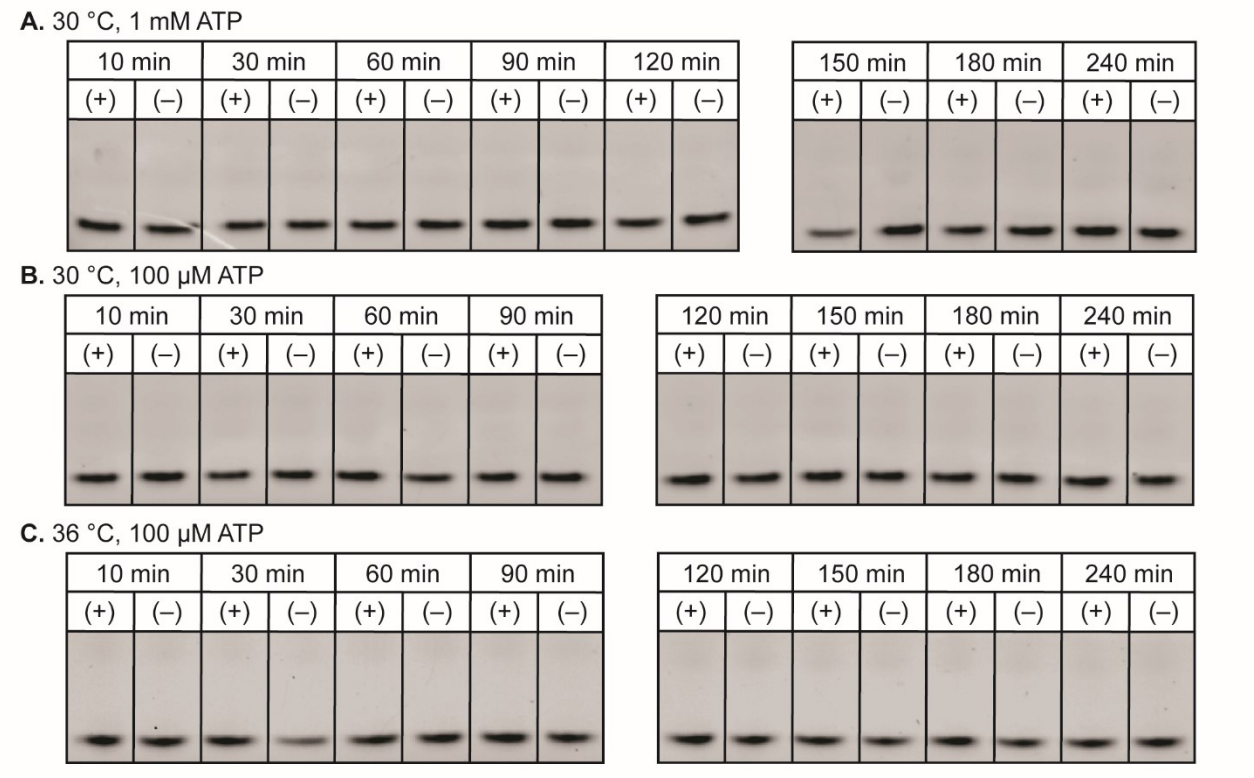


Figure 3.22 Cross-catalytic replication of DNA-I with pyrene modified strands with PBCV-1 DNA ligase. (**A**) 30 °C, 1 mM ATP (**B**) 30 °C, 100 μM ATP (**C**) 36 °C 100 μM ATP. *Experimental conditions*: 2.8 μM of the pyrene-modified strand (**IIa$_P$**), **Ib**, and **IIb**. 1.4 μM of the fluorescein-labeled strand (**Ia**). 14 nM (+) or 0 nM (-) of the initial **DNA-I**. PBCV-1 DNA ligase (10.5 μM) (1.5μL per 15 μL total volume) in 50 mM Tris-HCl, 10 mM $MgCl_2$ and various ATP concentrations.

A native DNA system was also tested in this experiment at elevated temperatures that were higher than what had been previously explored (**Figure 3.23**).[144] Generally, the higher the temperature, the faster the ligation of the primers that formed **DNA-I**. However, quite surprisingly, sigmoidal growth of **DNA-I** was observed at these higher temperatures, 34 °C and above, with a distinct difference between cross-catalysis with and without an initial template, especially at 38 °C. Cross-catalytic reaction at 38 °C with standard condition (1 mM ATP) was also conducted and exhibited similar sigmoidal growth with 100 μM ATP albeit faster non-templated reaction. This result suggested that an isothermal ligase chain reaction can occur even in a native DNA system without any destabilizing group at elevated temperatures with extended reaction time. This

behavior is likely because of the short length of the template (18 nt) and low G:C contents (17%) which affect the stability of the strand. This result brings into question whether there is a range of the stability difference between the ternary complex and the product duplex that can overcome product inhibition without destabilizing agents. **Table 3.2** shows the calculated stabilities at different temperatures for both the abasic and unmodified DNA systems. The stability gap in the natural DNA system at 38 ℃ for the product and intermediate complexes was reduced to five orders of magnitude compared with seven orders of magnitude difference in stability between the complexes at 30 ℃. The smaller stability gap at 38 ℃ is consistent with the rapid sigmoidal amplification observed at this temperature, similar to that seen for the abasic DNA system at 30 ℃, which also corresponded to a stability difference between the product and intermediate of four to five orders of magnitude (**Table 3.2**). This finding suggests LIDA can be possibly applied to any sequence of 18 nt, by tuning the temperature where it has a small enough difference in the stability of the intermediate and product DNA complexes. Future work will focus on further investigation into isothermal ligase chain reaction (LCR) and thermodynamic studies of different native DNA sequences as well as pyrene DNA systems in LIDA.

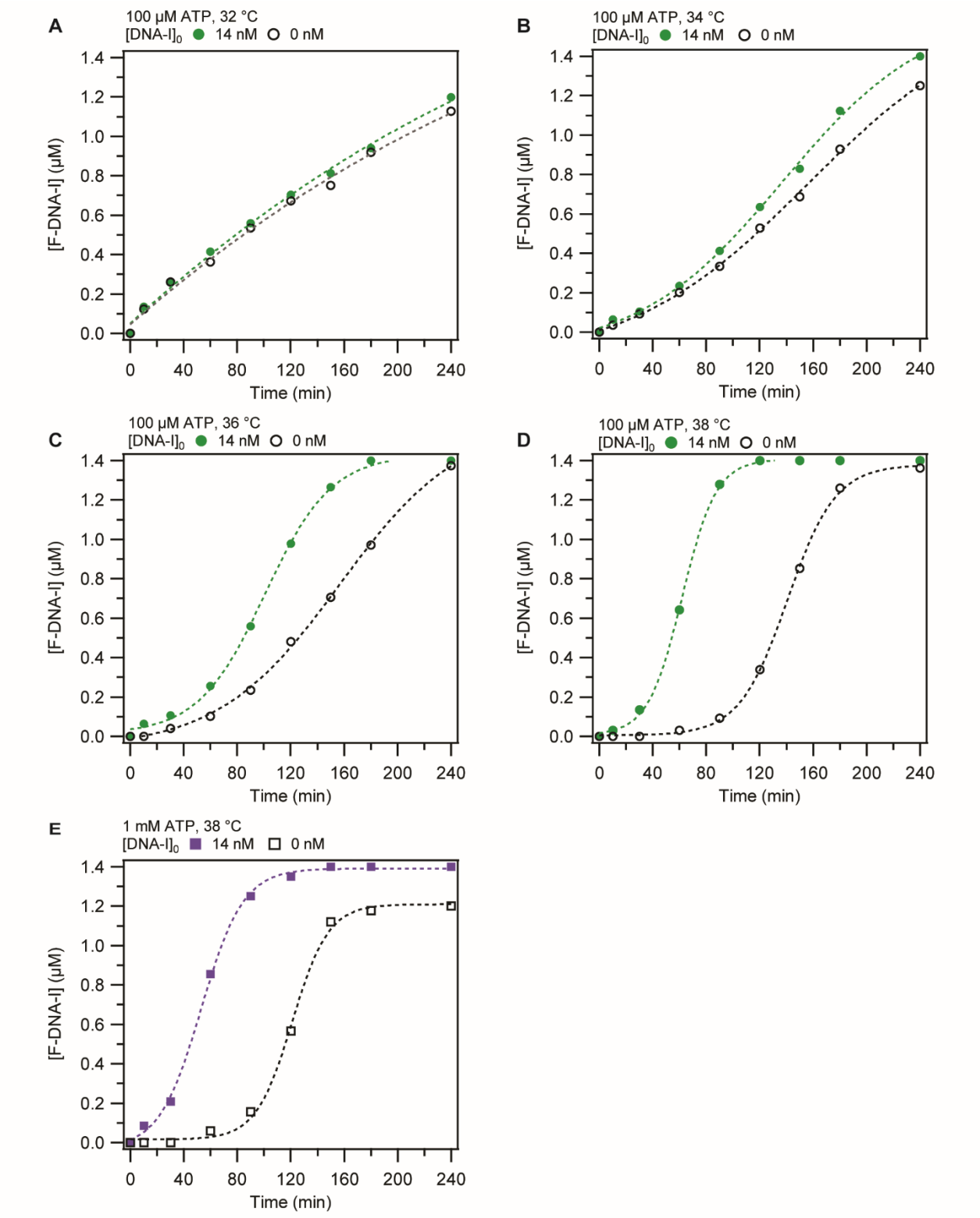


Figure 3.23 Cross-catalytic replication of DNA-I with complementary strands at various temperatures and ATP concentration. (**A**) 32 ℃ (**B**) 34 ℃ (**C**) 36 ℃ (**D**) 38 ℃ (**E**) 38 ℃ with 1 mM ATP. *Experimental conditions*: 2.8 μM of the 5'-phosphate thymine strands (**IIa$_P$**), **Ib**, and

IIb. 1.4 μM of the fluorescein-labeled strand (**Ia**). 14 nM (filled green circle) or 0 nM (empty black circle) of the initial **DNA-I**. T4 DNA ligase (2000 CEU per 15 μL total volume) in 50 mM Tris-HCl, 10 mM $MgCl_2$ and 100 μM or 1 mM ATP concentrations.

Table 3.2 Thermodynamic parameters of DNA complexes

	K_A (M^{-1})			
	30 °C	34 °C	36 °C	38 °C
Product Duplex (T) DNA-I:DNA-II(T)	$3.6 \pm 0.2 \times 10^{12}$	$1.9 \pm 0.1 \times 10^{11}$	$4.5 \pm 0.2 \times 10^{10}$	$1.1 \pm 0.06 \times 10^{8}$
Intermediate Complex (T) DNA-I:IIa_P(T)	$8 \pm 3 \times 10^{5}$	$1 \pm 2 \times 10^{4}$	$6 \pm 10 \times 10^{3}$	$3 \pm 5 \times 10^{3}$
Intermediate Complex (T) Ib_P:DNA-II(T)	$4 \pm 3 \times 10^{5}$	$1 \pm 1 \times 10^{5}$	$8 \pm 8 \times 10^{4}$	$5 \pm 5 \times 10^{4}$
Product Duplex (D = abasic) DNA-I:DNA-II(D)	$2 \pm 1 \times 10^{8}$	$2 \pm 1 \times 10^{7}$	$7 \pm 5 \times 10^{6}$	$2 \pm 2 \times 10^{6}$
Intermediate Complex (D) DNA-I:IIa_P(D)	$3 \pm 20 \times 10^{3}$	$2 \pm 20 \times 10^{2}$	50 ± 70	10 ± 30
Intermediate Complex (D) Ib_P:DNA-II(D)	$2 \pm 4 \times 10^{3}$	$2 \pm 6 \times 10^{2}$	80 ± 200	28 ± 100

Condition: 50 mM Tris-HCl buffer (pH 7.5 at 25 °C) with 10 mM $MgCl_2$.

3.4 Conclusion

Here, we have investigated ligation reactions involving an unnatural hydrophobic pyrene-abasic pair. Despite the non-hydrogen bonding interaction, ligation of a pyrene-modified end showed great selectivity when positioned across from the abasic lesion, which we attribute to the stability of the pyrene:abasic pair. Both T4 DNA ligase and PBCV-1 DNA ligase successfully ligated the pyrene-modified 5'-phosphate strand, and the ligation rate efficiency of the pyrene-abasic pair was comparable to that of A:T base pair. These results indicate that selective ligation of the pyrene base can be applied to the sequence-specific detection of abasic lesions if ligation is coupled with a subsequent amplification of the ligated products to allow for detection of a small number of abasic sites. The selectivity of 5'-phosphate pyrene ligation is consistent with previous reports showing that the pyrene-modified bases can be selectively incorporated opposite abasic lesions by the polymerase.[215] Moreover, other DNA damage sites can be detected by pyrene ligation with an extra step utilizing the appropriate DNA glycosylase, which recognizes different base modifications and produces an abasic site. Altogether, this work proves that the pyrene-modified end ligation opens the possibility of DNA damage detection.

Finally, the pyrene base was explored as a destabilizing group in our isothermal chain ligase reaction (LIDA). At increased temperature and optimized ATP concentration, the amplification of the target DNA strands showed the sigmoidal shape with the pyrene-modified strands, suggesting the pyrene-modified base can also destabilize the DNA strands. Additionally, the native DNA system without any modification was examined for comparison. Interestingly, the exhibition of sigmoidal curves for the DNA system without any destabilizing group implies that the isothermal LCR with four short primers (9, 9, 8, and 10 nt) can be extended to any sequence.

3.5 Experimental Section

3.5.1 Materials and Instrumentation

Unless noted, all chemicals and solvents for synthesizing pyrene nucleotides were purchased from Sigma-Aldrich. All glassware was oven-dried prior to use. Dry solvents (tetrahydrofuran, dichloromethane) were purified using a cartridge solvent purification system prior to use. Dry pyridine was purified by successive passage through columns of alumina and copper under argon. Silica gel (P60, 40 - 63 μm, 60 Å, #R12030B, Silicycle) was used to flash column chromatography. Thin-layer chromatography was carried out on glass-backed plates precoated with silica gel (SilicaPlate G TLC, 60 matrix, 200 μm, #Z122785 Sigma-Aldrich). TLC plates were visualized by UV light. NMR spectra (^{1}H, ^{12}C, ^{31}P) were obtained on Agilent VNMR 700 MHz, Varian VNMRS 500 MHz, or Inova 500 MHz spectrometers. ^{1}H and ^{13}C chemical shifts were reported in ppm downfield from tetramethylsilane (TMS). ^{31}P chemical shifts were reported in ppm downfield from an external phosphoric acid standard at 201.641 MHz. HPLC purification was performed using Agilent HPLC 1100 equipped with a C18 column.

3.5.2 DNA Synthesis and Purification

Oligonucleotides were synthesized on an Applied Biosystems Model 392 DNA/RNA synthesizer, as mentioned in Chapter 2. The characterization of synthesized strands, MALDI-TOF and StainsAll, followed the same procedure described in Chapter 2.

3.5.3 Synthesis of Pyrene Nucleotide

The procedure for the synthesis of pyrene nucleotide was followed and modified the literature.[143,224,238,239] Unless noted, all reactions were performed under an inert atmosphere employing the standard Schlenk technique. All the steps were monitored by TLC analysis prior to the flash column chromatography using the same eluent as mentioned on each step. The NMR spectra of each product are attached in Appendix I.

*Synthesis of 1-[2'-deoxy-3',5'-bis O-(4-methylbenzoyl)-α-D-ribofuranosyl-1]pyrene (**1α and 1β**)*[238]

Mg turnings (0.18 g, 7.40 mmol) and 1,2-bromoethane (0.09 mL, 1.04 mmol) were added to the Schlenk flask and dissolved in 1.6 mL of dry THF (flask 1). The reaction was stirred for 10 min at room temperature. 1-bromopyrene (0.38 g, 1.37 mmol) was dissolved in dry THF (4 mL) in different Schlenk flask 2 and then transferred to flask 1 by cannula. The reaction was stirred at 55 °C for 3 hours under the air-free condition, with the condenser set-up. Once the reaction was cooled down in the ice-bath, the reaction was transferred by cannula to flask 3 where Cu(I) iodide (CAS 7681-65-4, Fisher Scientific) (0.13 g, 0.68 mmol) was added. Then, the reaction was stirred at room temperature for 20 minutes and then at 45 °C for 20 minutes. In flask 4, Hoffer's chlorosugar (CAS 3601-89-6, Carbosynth) (0.25 g, 0.66 mmol) was dissolved in dry THF (2 mL). The reaction solution from flask 3 was transferred to flask 4 by cannula and stirred at 55 °C for 2 hours and then at room temperature overnight under air-free conditions. The reaction mixture was extracted with 10% NH_4Cl (10 mL) and DCM (10 mL) followed by sodium bicarbonate (10 mL) and sodium chloride (10 mL) solution. Back-extraction of the aqueous layer was carried out with DCM (10 mL). Combined the organic layer was dried with $MgSO_4$ and then dried by rotovap.

Flash column chromatography was performed with the hexane: ethyl acetate (9:1). *Yield*: 1α (0.25 g, 68%), 1β (0.024 g, 6%) and 1α + 1β mixture (0.058 g, 16%) determined from the isolated compounds and compound mixtures from flash chromatography based on the limiting reagent (0.66 mmol, Hoffer's chlorosugar). ^{1}H NMR of the 1α compound did not match completely with the reference from the literature[224] but our ^{1}H NMR spectrum of the 1β compound did match with the reported spectrum.[224]

^{1}H NMR for 1α (700 MHz, $CDCl_3$) δ 8.35 (d, J = 9.2 Hz, 1H), 8.23 – 8.17 (m, 4 H), 8.13 (d, J = 9.2 Hz, 1H), 8.10 – 7.99 (m, 5H), 7.62 (d, J = 8.2 Hz, 2H), 7.28 (d, J = 7.9 Hz, 3H), 7.09 (d, J = 7.9 Hz, 2H), 6.43 – 6.38 (m, 1H, H1'), 5.75 (m, 1H, H3'), 4.97 (d, *J* = 2.9, 1H, H4'), 4.78 – 4.68 (m, 2H, H5'), 3.32 (m, 1H, H2'β), 2.51 (m, 1H, H2'α), 2.44 (s, 3H), 2.35 (s, 3H).

^{13}C NMR for 1α (176 MHz, $CDCl_3$) δ 166.6, 166.4, 144.4, 144.0, 134.3, 130.0, 129.4, 127.9, 127.8, 127.4, 127.3, 127.2, 126.1, 125.4, 122.4, 114.8, 77.3, 77.2, 77.0, 65.0, 41.8, 21.9.

^{1}H NMR for 1β (700 MHz, $CDCl_3$) δ 8.33 (dd, J = 7.9, 0.6 Hz, 1H), 8.28 (d, J = 9.2 Hz, 1H), 8.22 – 8.14 (m, 4H), 8.12 – 8.04 (m, 5H), 8.08 – 7.99 (m, 1H), 7.99 – 7.95 (m, 2H), 7.36 – 7.32 (d, J = 7.5Hz, 2H), 7.19 (d, J = 7.5Hz, 2H), 6.30 (dd, J = 10.9, 5.3 Hz, 1H, H1'), 5.75 (m, 1H, H3'), 4.82 (m, 2H, H5'), 4.74 (m, 1H, H4'), 2.90 (dd, J = 14.1, 5.3 Hz, 1H, H2'α), 2.48 (s, 3H), 2.44 – 2.40 (m, 1H, H2'β), 2.38 (s, 3H).

^{13}C NMR for 1α (176 MHz, $CDCl_3$) δ 166.6, 166.2, 144.0, 136.2, 131.5, 130.9, 130.8, 130.0, 129.8, 129.4, 129.1, 127.8, 127.6, 127.3, 127.3, 126.9, 126.1, 125.5, 125.3, 125.0, 125.0, 122.9, 122.6, 82.8, 78.5, 76.7, 64.9, 60.5, 40.8, 21.9, 21.7, 14.4.

ESI for $C_{37}H_{30}O_5$ calculated [$M+Na^+$] 577.63; found 577.2

Epimerization of 1α and 1β mixture: **1-[2'-deoxy-3',5'-bis O-(4-methylbenzoyl)-α-D-ribofuranosyl-1]pyrene** [239]

The 1α + 1β mixture (0.125 g, 0.23 mmol) was dissolved in DCM (4 mL). 1% TFA was added to the reaction flask and stirred overnight at room temperature. Triethylamine (134 µL, 0.1 mmol) was added to the reaction flask. The reaction mixture was dried out and flash column chromatography was conducted with the eluent, hexane: ethyl acetate (9:1). The column chromatography was conducted once more. *Overall yield*: 1β (0.044 g, 36% first column and 0.029 g, 25% second column)

Deprotection of 1β compound: **1-[2'-deoxy-*β*-*D*-ribofuranosyl-1]pyrene (Compound 2)** [238]

The 1β compound (0.069 g, 0.12 mmol) and sodium methoxide (0.02 g, 0.37 mmol) were dissolved in dry methanol (4.3 mL) and dry THF (0.48 mL) and then sonicated for a minute. Then the reaction mixture was stirred overnight at room temperature. The reaction mixture dried and then extracted with 10% NH_4Cl (10 mL) and ethyl acetate (10 mL x 3). Organic layers were combined and dried with $MgSO_4$. Then the crude product was purified with flash column chromatography with 100% ethyl acetate. *Yield*: 0.028 g, 71%. Our ^{1}H NMR spectrum of the 1β compound did match with the reported spectrum[224] while the literature integration were slightly lower leading to total of 16H instead of the expected 18H.

^{1}H NMR (700 MHz, $CDCl_3$) δ 8.32 (d, J = 9.2 Hz, 1H), 8.24 – 8.17 (m, 4H), 8.14 (d, J = 9.2 Hz, 1H), 8.06 (s, 2H), 8.02 (t, J = 7.6 Hz, 1H), 6.24 (dd, J = 10.2, 5.8 Hz, 1H, H1'), 4.64 – 4.58 (m, 1H. H3'), 4.30 – 4.17 (m, 1H, H4'), 4.07 – 3.87 (m, 2H, H5'), 2.63 (ddd, J = 13.4, 5.8, 2.1 Hz, 1H, H2'α), 2.27 (ddd, J = 13.4, 10.3, 6.4 Hz, 1H, H2'β), 1.96 – 1.89 (m, 2x OH).

^{13}C NMR (126 MHz, $CDCl_3$) δ 169.0, 168.7, 134.6, 131.5, 131.0, 130.8, 128.0, 127.9, 127.6, 127.5, 126.1, 125.5, 125.3, 125.2, 125.0, 122.7, 122.5, 87.2, 74.2, 63.7, 44.2.

ESI for $C_{21}H_{18}O_3$ calculated [$M+Na^+$] 341.37; found 341.1

Protection of 5' hydroxy group on Compound 2: **1-[2'-deoxy-5'-*O*-(4',4'-dimethoxytrityl)-*β*-*D*-ribofuranosyl-1]pyrene (Compound 3)**[224]

Compound 2 (0.028 g, 0.088 mmol) was evaporated with dry pyridine (1 mL x 2) with a tertiary trap setup on the Schlenk line. Then, dry pyridine (1 mL) was added in the flask. In a separate flask, DMT-Cl (0.046 g, 0.13 mmol) and catalytic amount of DMAP (4-(dimethylamino)pyridine) (0.016g, 0.13 mmol) were prepared with dry DCM (1mL). The DMT-Cl solution was added in the reaction flask, dropwise. The addition was performed without septum, under nitrogen flush, as the gas generated. Once the addition was done, the reaction was stirred for 8 hours at room temperature. The reaction mixture was extracted with DCM (10 mL) and saturated sodium bicarbonate (10 mL x 2) followed by brine (10 mL). Aqueous layer was back extracted with DCM (10 mL x 2). Combined organic layer was dried with $MgSO_4$. The crude product was purified by flash column chromatography with eluent (hexane: ethyl acetate 8:2). *Yield*: 0.014 g, 26%. Our 1H NMR of the compound 3 were matched with the reported spectrum.[246]

^{1}H NMR (700 MHz, $CDCl_3$) δ 8.33 (d, J = 9.2 Hz, 1H), 8.29 (dd, J = 7.9, 0.6 Hz, 1H), 8.18 (ddd, J = 7.8, 4.1, 1.1 Hz, 2H), 8.15 (d, J = 7.9 Hz, 1H), 8.08 (d, J = 9.2 Hz, 1H), 8.05 (s, 2H), 8.00 (t, J = 7.6 Hz, 1H), 7.56 – 7.50 (m, 2H), 7.44 – 7.38 (m, 4H), 7.30 (m, 2H), 7.26 – 7.20 (m, 1H), 6.83 (dd, J = 9.0, 2.6 Hz, 4H), 6.21 (dd, J = 9.8, 6.0 Hz, 1H, H1'), 4.57 (dd, J = 6.3, 3.0 Hz, 1H, H3'), 4.24 (dd, J = 4.9, 3.4 Hz, 1H, H4'), 3.78 (d, J = 3.5 Hz, 6H), 3.50 (ddd, J = 43.5, 9.8, 4.8 Hz, 2H, H5'), 2.61 (ddd, J = 13.3, 6.0, 2.4 Hz, 1H, H2'α), 2.27 (ddd, J = 13.3, 9.9, 6.3 Hz, 1H, H2'β).

^{13}C NMR (176 MHz, $CDCl_3$) δ 158.7, 145.0, 136.2, 135.6, 131.5, 130.8, 130.3, 129.3, 128.4, 128.0, 127.8, 127.7, 127.2, 127.0, 126.0, 125.3, 125.2, 125.0, 124. 9, 123.0, 122.9, 113.3, 86.5, 86.3, 64.5, 55.4, 53.6, 43.9, 29.8.

ESI for $C_{42}H_{36}O_5$ calculated [$M+Na^+$] 643.26; found 643.2

Phosphorylation of Compound 3: **1-[2'-deoxy-5'-*O*-(4',4'-dimethoxytrityl)-3'-*O*-(*N,N*-diisopropyl-*β*-cyanoethylphosphoramidyl)-*β*-*D*-ribofuranosyl-1]pyrene (Compound 4)**[143]

Compound 3 (0.0144 g, 23.91 μmol) was dissolved with dry DCM (6 μL) in a Schlenk flask under the argon. *N,N*-diisopropylethylamine (CAS 7087-68-5, Fluka Analytical) (6.04 μL, 1.2 eq) was added to the reaction flask. After three times of freeze-pump-thaw degassing, cyanoethyl *N,N*-diisopropylchlorophosphoramidite (CAS 89992-70-1) (6.25 μL, 1.5 eq) was added to the reaction mixture. The reaction was stirred for 30 minutes at room temperature and was quickly concentrated to dryness and re-dissolved in $CDCl_3$ (0.8 mL) for NMR analysis. ^{31}P NMR confirmed the presence of the phosphoramidite by the distinctive signal at δ 148 ppm. Immediately after the NMR analysis, the crude compound was transferred to ABI 6.5 mL vial for installation on the solid-phase synthesizer.

^{1}H NMR of crude compound 4 (i.e. the evaporated reaction mixture) is included in Appendix I.

^{31}P NMR of crude compound 4 (202 MHz, $CDCl_3$) δ 148.4, 148.0 and we observed the oxidized phosphoramidite product or starting material at δ 14.2 and δ 2.2.

3.5.4 Synthesis of Pyrene Nucleotide on DNA Strands

The DNA strands (DNA-IIa_p(pyrene) and DNA-II(pyrene)) were synthesized from 1.0 μmole CPG using the standard DNA synthesis cycle up to the 9th base from 3' base CPG. Synthesized pyrene phosphoramidite was installed in the 5th position of the solid-phase synthesizer immediately after NMR analysis. Pyrene phosphoramidite was added with a standard DNA synthesis cycle with an extended coupling time (20 min). The synthesis was continued as designed sequences. For the 5'-phosphate pyrene strand, DNA-IIa_p(pyrene), 5'-phosphorylation was performed using solid chemical phosphorylation reagent II (10-1902-90). For the pyrene-modified

template, DNA-II(pyrene), the addition of bases was continued as designed. The standard DNA purification using a Gel-Pak cartridge was performed to purify the pyrene-modified strands. The pyrene-modified oligonucleotides were further purified by RP-HPLC. The pyrene modified strands were analyzed by MALDI-TOF and polyacrylamide gel electrophoresis with Stains-All.

3.5.5 HPLC Purification of Pyrene Modified Oligonucleotides

The crude pyrene-modified strands were analyzed by RT-HPLC. The solvent system used was described, A solvent: 0.03 M triethylammonium acetate (TEAA) buffer and B solvent: acetonitrile with 5 % of 0.03 M TEAA. The flow rate was consistent at 3 mL per minute. For the analysis, the sample (1 nmol in 20 μL H_2O, injection of 7 μL) was run through the C18 column with a gradient of B solvent, 1 % from 50 % over 55 minutes followed by 5 minutes of post-run. The analysis was performed both at 260 nm wavelength where the oligonucleotide was observed and 350 nm wavelength where pyrene moiety was observed. For the purification, the crude product (40 nmol in 180 μL H_2O) was run through the column with the same procedure. The injection of the sample was 60 μL per run. The UV detector was set at 260 nm wavelength to observe the oligonucleotide. The average purification yield was 20-23 % for both DNA-IIa_P(pyrene) and DNA-II(pyrene) strands.

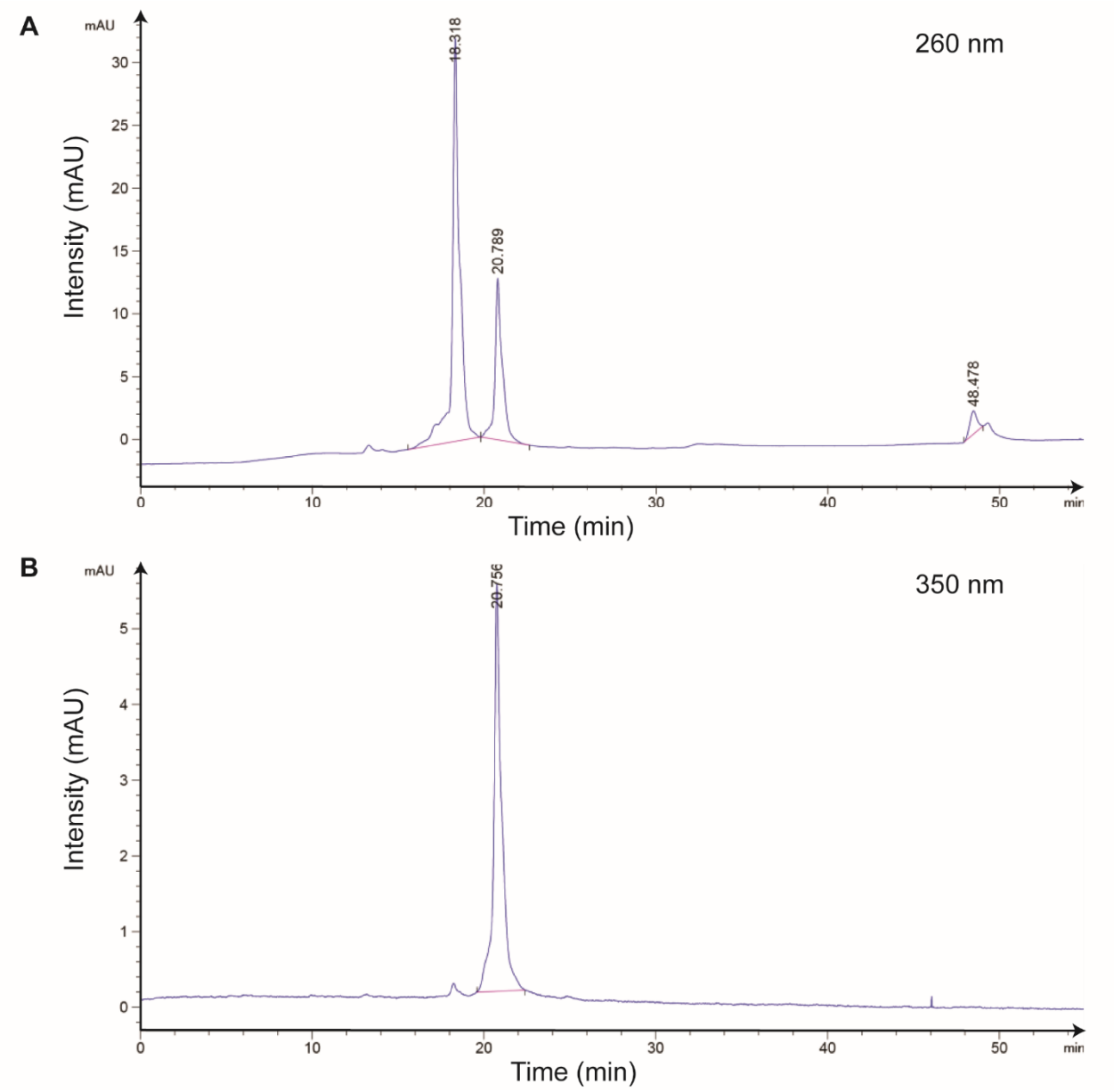


Figure 3.24 Analytical HPLC chromatogram of crude DNA-II(pyrene). The detection wavelength is at (**A**) 260 nm wavelength and (**B**) 350 nm wavelength. Based on analytical HPLC, a semi-preparatory run was performed, and the product was collected from the peak (RT 20 min from 260 nm wavelength).

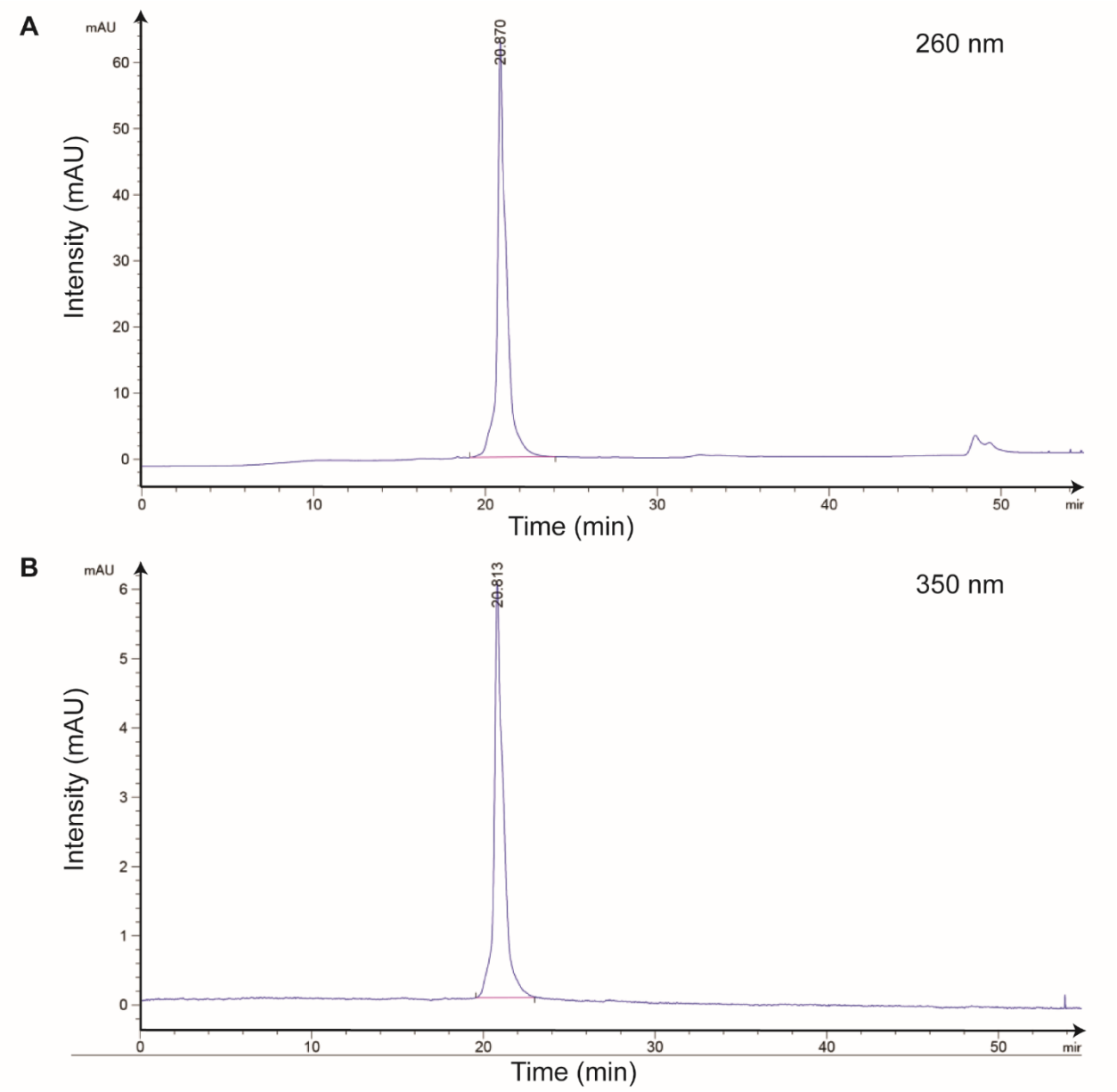


Figure 3.25 Analytical HPLC chromatogram of purified DNA-II(pyrene). The detection wavelength is at (**A**) 260 nm wavelength and (**B**) 350 nm wavelength.

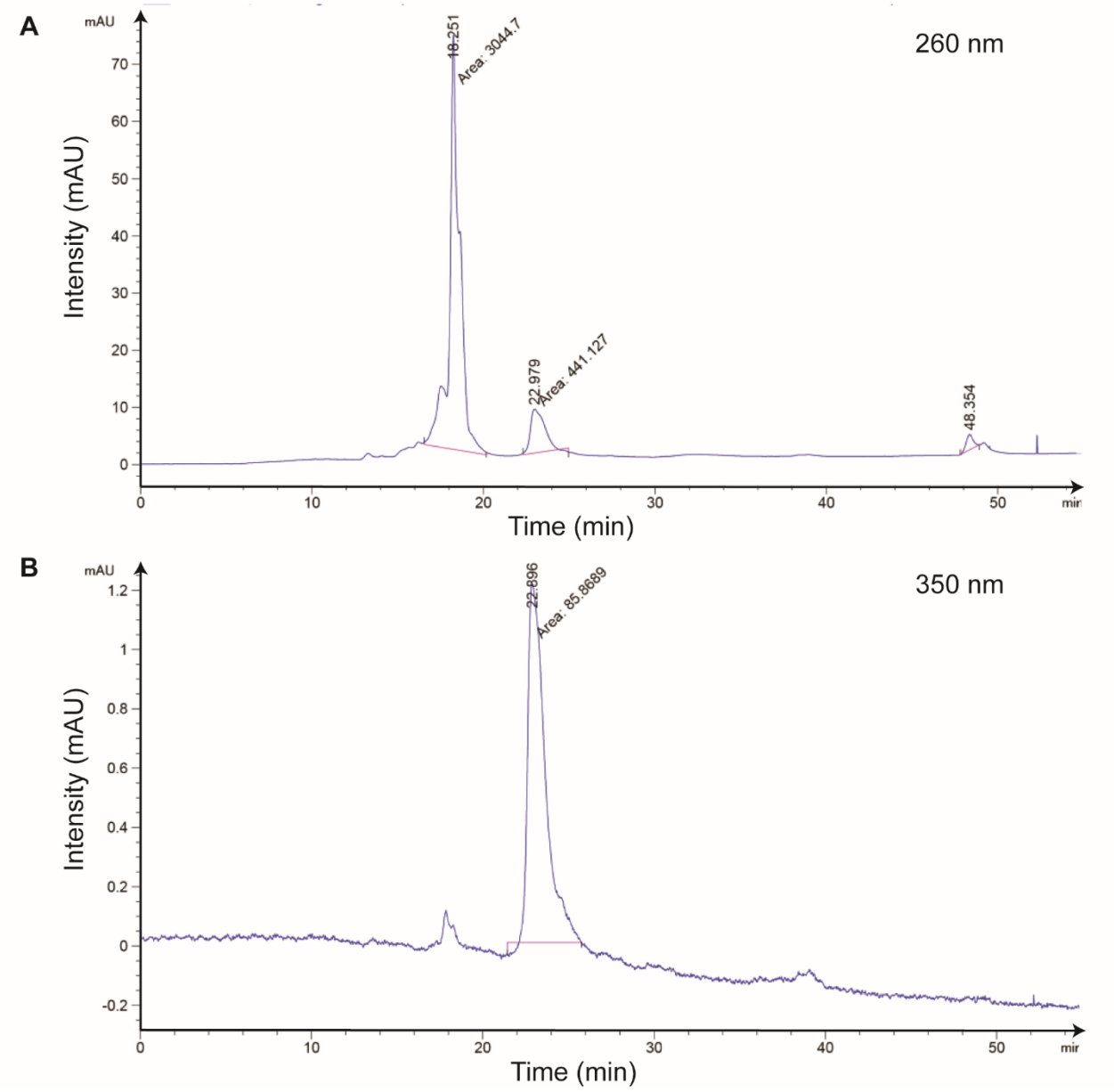


Figure 3.26 Analytical HPLC chromatogram of crude DNA-IIa$_P$(pyrene). The detection wavelength is at (**A**) 260 nm wavelength (**B**) 350 nm wavelength. Based on analytical HPLC, a semi-preparatory run was performed, and the product was collected from the peak (RT 22 min from 260 nm wavelength).

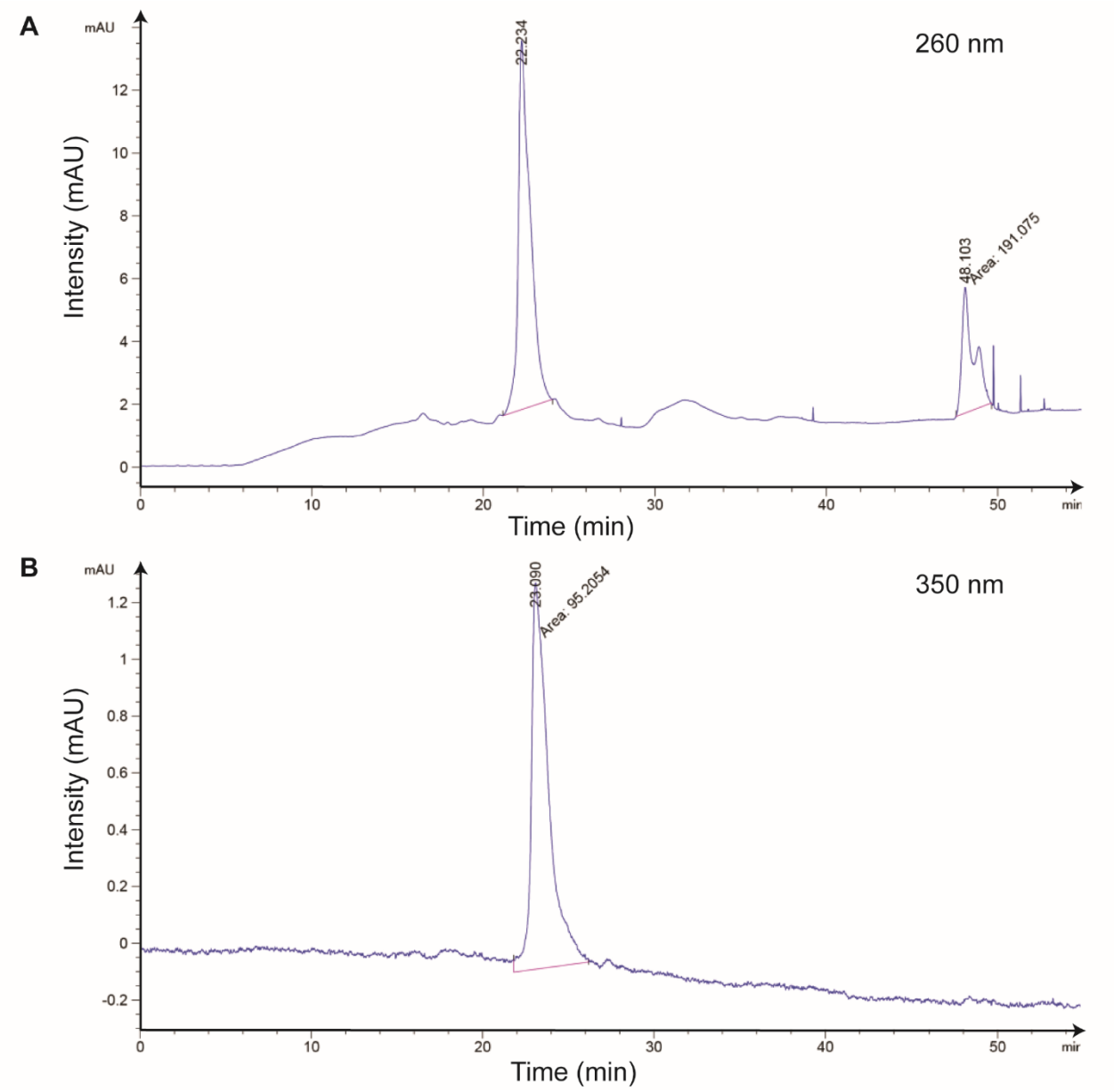


Figure 3.27 Analytical HPLC chromatogram of purified DNA-IIa$_P$(pyrene). The detection wavelength is at (**A**) 260 nm wavelength and (**B**) 350 nm wavelength.

3.5.6 Ligation Assay

Pyrene modified strand (DNA-IIa$_P$(pyrene) and DNA-II(pyrene)) was aliquoted as 0.0406 nmol per tube. All other DNA strands were prepared in a 10 mM PBS 0.1 M NaCl buffer as a stock solution. A master mix was prepared by adding 10X Tris buffer, 10X ATP, and T4 DNA ligase

(400,000 CEU/mL #M0202S from New England Biolabs, 1 μL per total 15 μL of reaction volume) or PBCV-1 DNA ligase (10.5 μM, #M0375S from New England Biolabs) (1.5 μL per total 15 μL of reaction volume) and then incubated at the reaction temperature for 10 minutes. The master mix was then added into each reaction tube, 2.5 μL for a total of 7.5 μL volume or 5 μL for a total 15 μL volume. In each reaction, the final concentration of the phosphate modified strands was 2.8 μM as excess and the final concentration of the other fluorescein-labeled strands and the templates were 1.4 μM in 50 mM Tris-HCl buffer including 10 mM $MgCl_2$, and various ATP concentrations. The reaction was collected at each data point by quenching with EDTA and 1-bromophenol mixture.

3.5.7 Competitive Ligation Assay

In competitive ligation assay, either template (abasic site) or template (adenine) or both templates were added. When either template was used, the final concentration was 1.4 μM. When both templates were added, each template was 0.7 μM, to make the sum of template concentration consistent as 1.4 μM. Both 5'-phosphate strands (pyrene or thymine) were 2.8 μM, and fluorescein-labeled strands were 1.4 μM. The DNA fragments and templates were mixed and incubated at the reaction temperature while preparing the master mix. The master mix was prepared by adding a 10X Tris buffer containing $MgCl_2$, 10X ATP, and DNA ligase. For T4 DNA ligase (400 CEU), 0.5 μL per total 7.5 μL reaction was used, while for PBCV-1 (10.5 μM), 0.75 μL per total 7.5 μL reaction volume was used. Then the master mix was incubated at the reaction temperature for 10 minutes and then added to the reaction mixture. To stop ligation after 10 minutes of reaction, the reaction mixture (3 μL) was mixed with EDTA (3 μL, 0.5 M) containing 1-bromophenol dye. The reaction was analyzed by 15% denaturing PAGE.

3.5.8 Lesion-Induced DNA Amplification with Pyrene Strand

The general procedure of LIDA was explained in Chapter 2. Instead of a stock solution, the pyrene modified strand was prepared in dryness (0.0406 nmol in each tube). Therefore, each tube was dissolved in 2.5 μL H_2O immediately before the reaction and directly added into the reaction mixture (total volume of 15 μL). For T4 DNA ligase reaction, 1 μL (2000 CEU) per total 15 μL reaction volume was used. For PBCV-1 DNA ligase reaction, 1.5 μL (10.5 μM) per total 15 μL reaction volume was used.

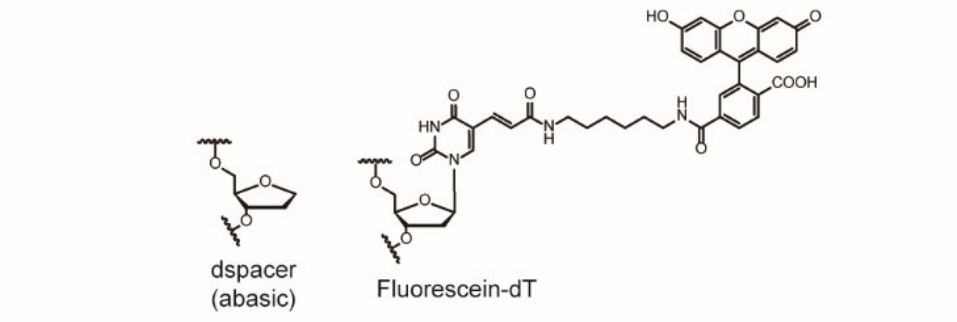
3.5.9 Thermal Dissociation Analysis of Pyrene Modified Strand

The procedure of melting curve analysis is described in Chapter 2.

Table 3.3 Oligonucleotide sequence and corresponding MALDI

Oligonucleotides	Sequence (5' → 3')	Expected	Founded
DNA- I (Ab)	TTG TTA AAT **Ab**TT GAT AAG	5419	5425
DNA- I (A)	TTG TTA AAT **A**TT GAT AAG	5552	5555
DNA- I (T)	TTG TTA AAT **T**TT GAT AAG	5543	5549
DNA- I (G)	TTG TTA AAT **G**TT GAT AAG	5568	5574
DNA- I (C)	TTG TTA AAT **C**TT GAT AAG	5528	5532
DNA- II (Pyrene)	CTT ATC AA(Pyene) ATT TAA CAA	5517	5520
DNA- II (T)	CTT ATC AA**T** ATT TAA CAA	5441	5445
DNA- Ia	$T_{fluorescein}$TG TTA ATT	3239	3243
DNA- Ib_p Abasic group	phosphate (abasic)TT GAT AAG	2708	2712
DNA- Ib_p	phosphate ATT GAT AAG	2842	2845

DNA- IIb	$T_{fluorescein}$ CTT ATC AA	3184	3188
DNA- IIa_p Pyrene group	phosphate (pyrene) ATT TAA CAA	3166	3170
DNA- IIa_p Complementary (T)	phosphate T ATT TAA CAA	3090	3094
DNA- IIa_p (+3) Complementary (T)	phosphate T ATT TAA CAA TAA	4020	4021



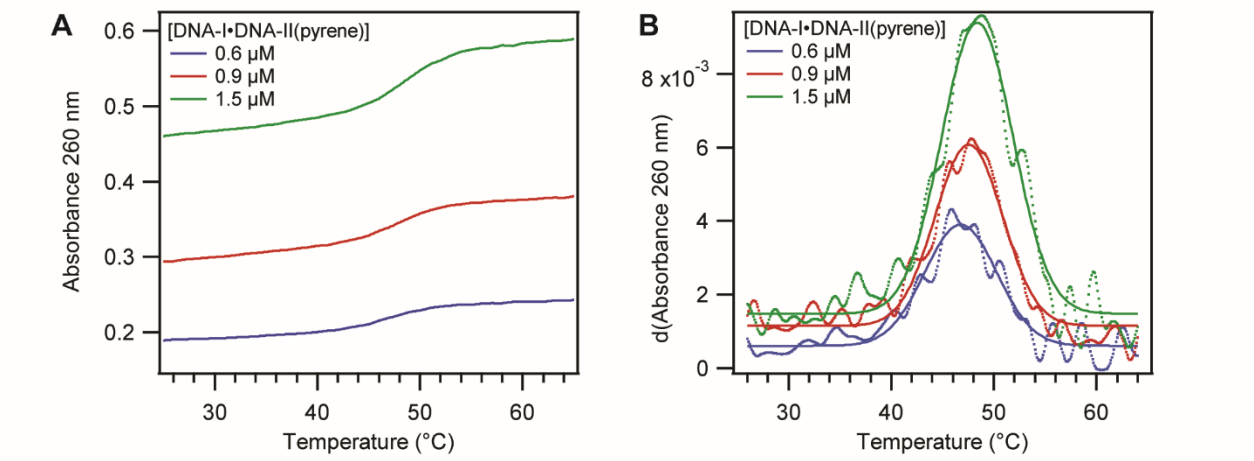


Figure 3.28 Thermal melting curves of DNA duplex (pyrene). (**A**) Monitoring the absorbance at 260 nm by the increasing temperature at various concentrations; 0.6 µM, 0.9 µM, and 1.5 µM. (**B**) The corresponding first derivative of absorbance of the melting curves. The data were fit to the Gaussian curve to determine the melting temperature.

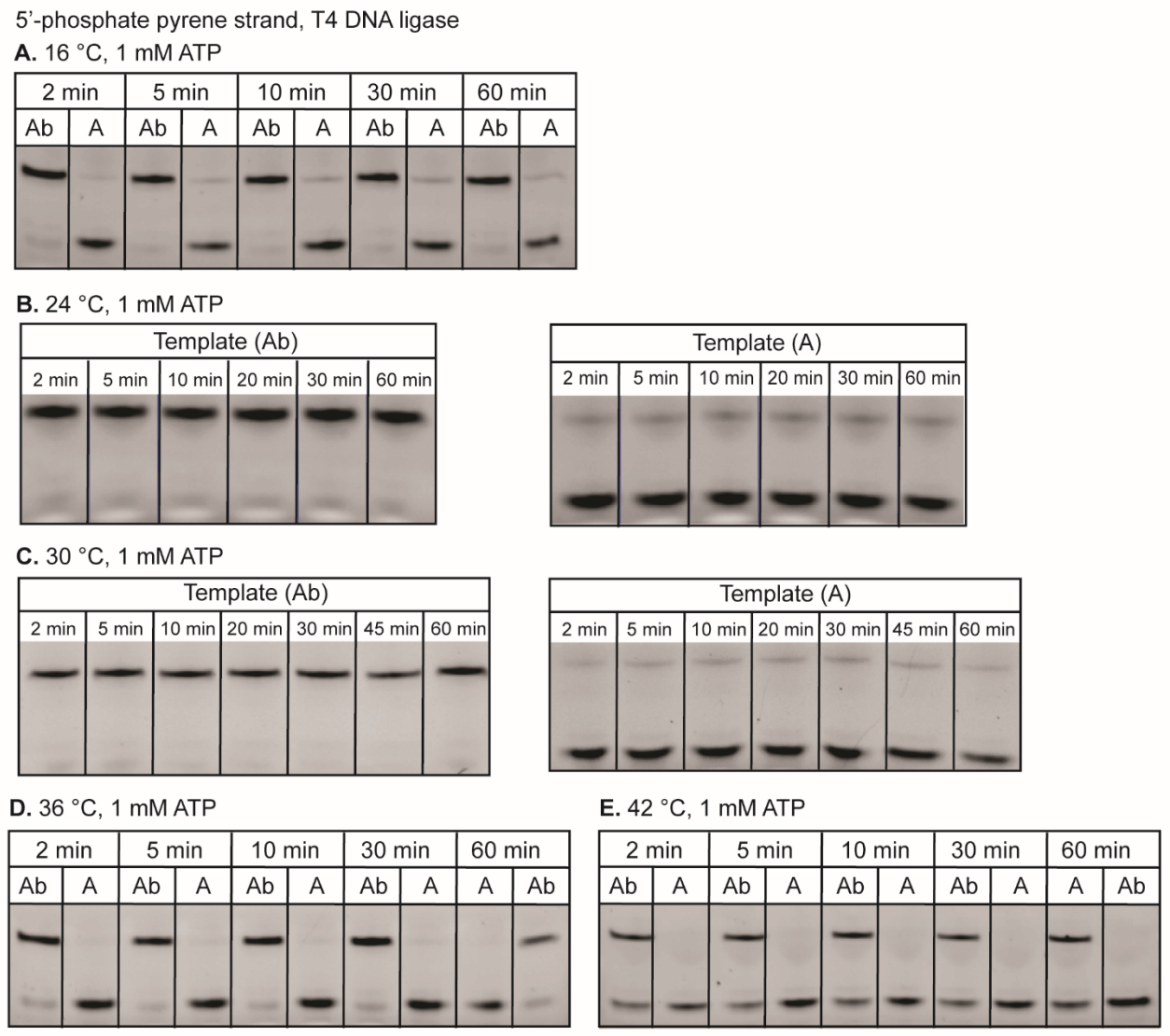


Figure 3.29 Representative PAGE image of temperature varied ligation reactions using 5'-phosphate pyrene strands. (**A**) 16 ℃, (**B**) 24 ℃, (**C**) 30 ℃, (**D**) 36 ℃ and (**E**) 42 ℃ from **Figure 3.5A.**

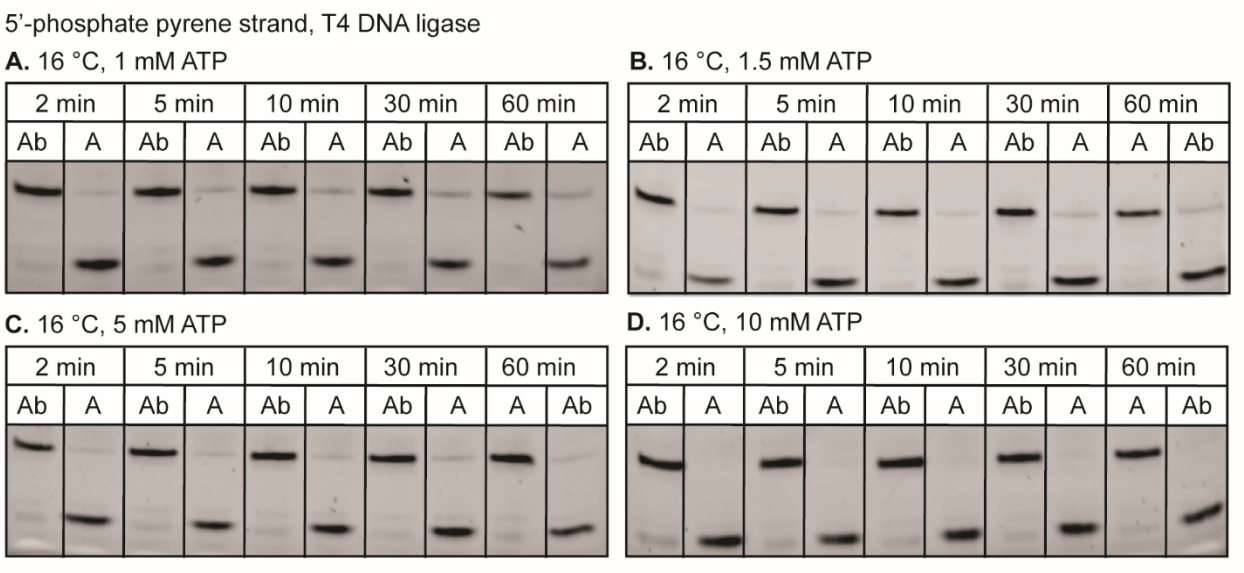


Figure 3.30 Representative PAGE image of ATP concentration variations using 5'-phosphate pyrene strands at 16 °C. (**A**) 1 mM ATP, (**B**) 1.5 mM ATP, (**C**) 5 mM ATP and (**D**) 10 mM ATP from **Figure 3.5B**.

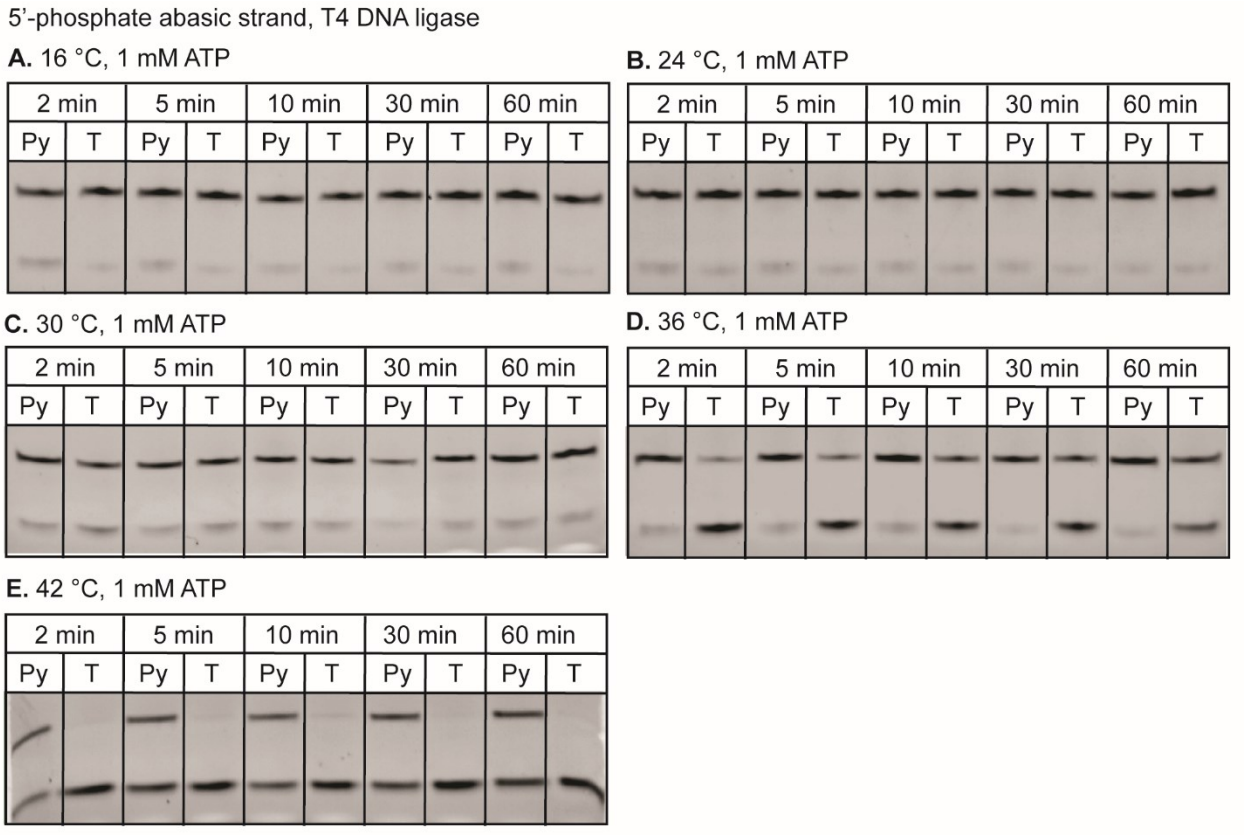


Figure 3.31 Representative PAGE image of temperature varied ligation reactions using 5'-phosphate abasic strands by T4 DNA ligase. (**A**) 16 ℃, (**B**) 24 ℃, (**C**) 30 ℃, (**D**) 36 ℃ and (**E**) 42 ℃ from **Figure 3.8A.**

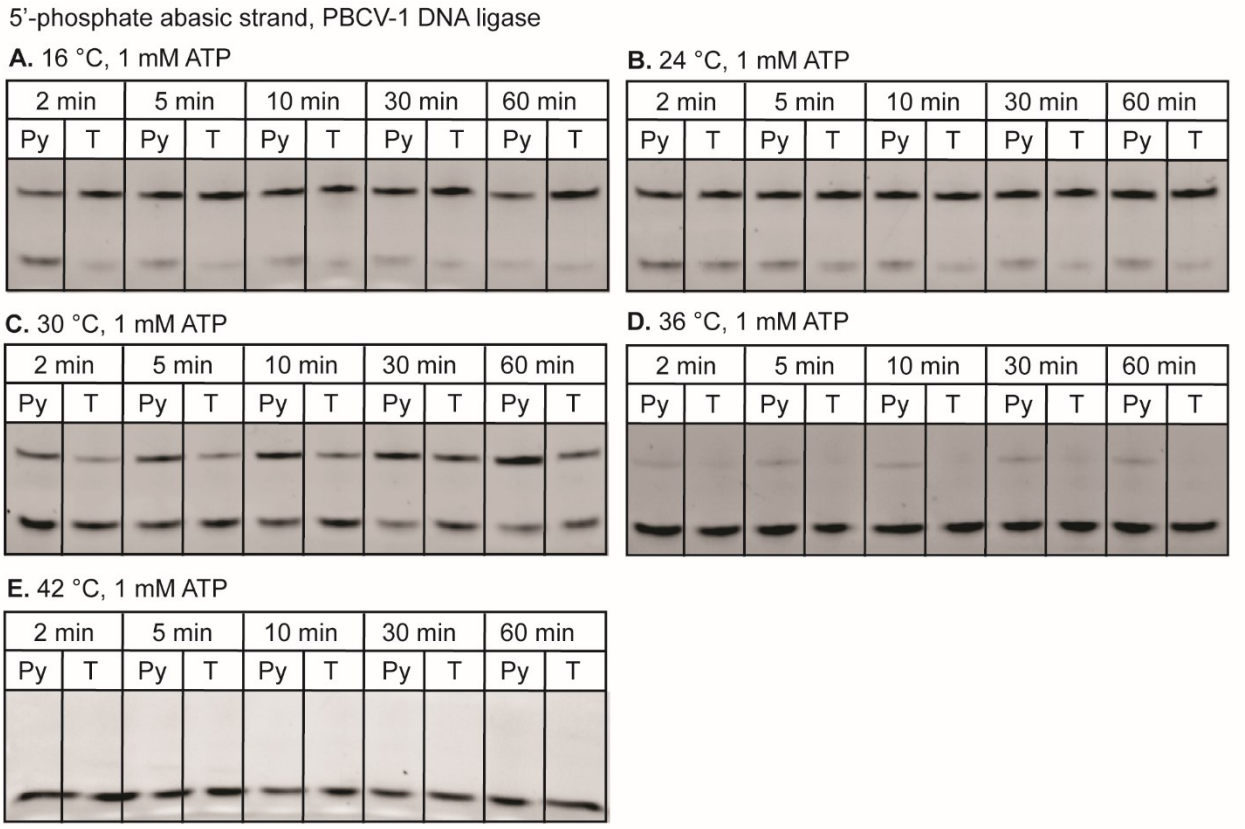


Figure 3.32 Representative PAGE image of temperature varied ligation reactions using 5'-phosphate abasic strands by PBCV-1 DNA ligase. (**A**) 16 ℃, (**B**) 24 ℃, (**C**) 30 ℃, (**D**) 36 ℃ and (**E**) 42 ℃ from **Figure 3.8B.**

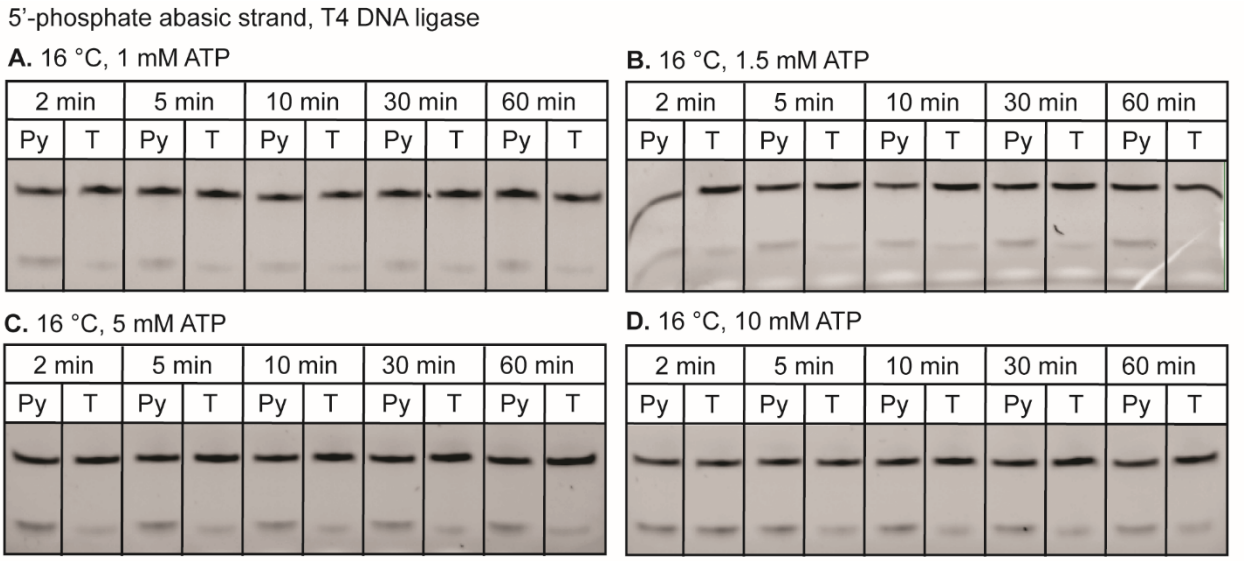


Figure 3.33 Representative PAGE image of ATP concentration variations using 5'-phosphate abasic strands by T4 DNA ligase at 16 °C. (**A**) 1 mM ATP, (**B**) 1.5 mM ATP, (**C**) 5 mM ATP and (**D**) 10 mM ATP from **Figure 3.8C**.

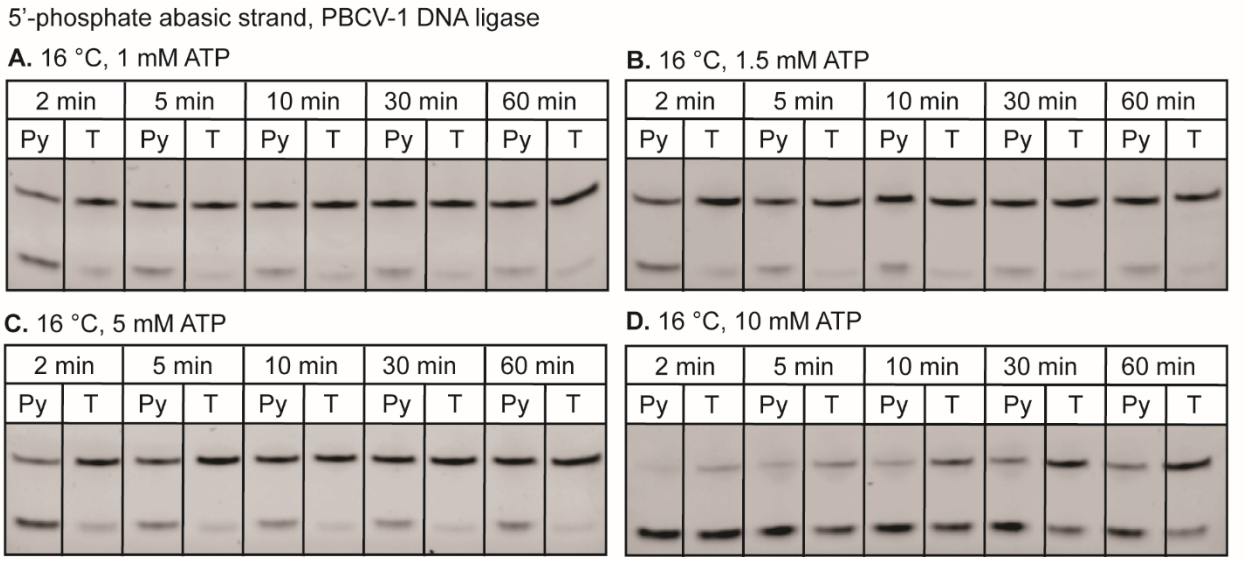


Figure 3.34 Representative PAGE image of ATP concentration variations using 5'-phosphate abasic strands by PBCV-1 DNA ligase at 16 °C. (**A**) 1 mM ATP, (**B**) 1.5 mM ATP, (**C**) 5 mM ATP and (**D**) 10 mM ATP from **Figure 3.8D**

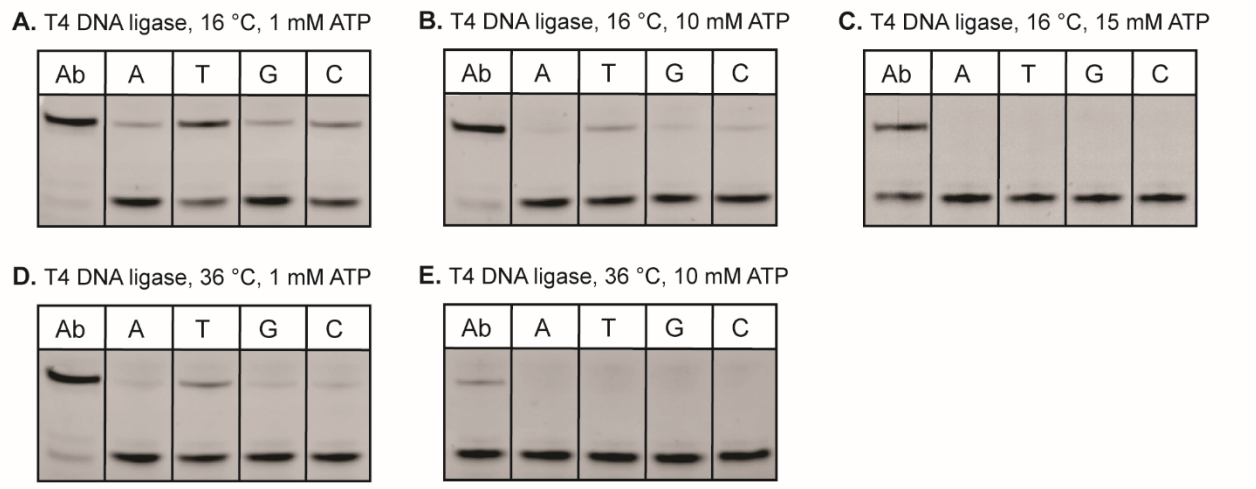


Figure 3.35 Representative PAGE image of template selectivity using 5'-phosphate pyrene strands by T4 DNA ligase. (**A**) 16 ℃, 1 mM ATP, (**B**) 16 ℃, 10 mM ATP, (**C**) 16 ℃, 15 mM ATP, (**D**) 36 ℃, 1 mM ATP and (**E**) 36 ℃, 10 mM ATP from **Figure 3.9**

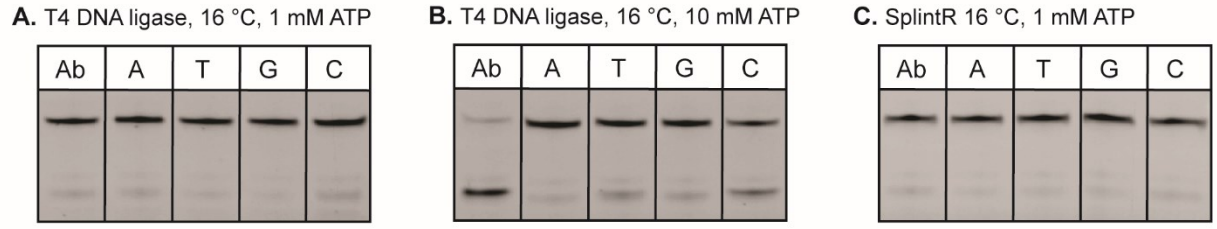


Figure 3.36 Representative PAGE image of template selectivity using 5'-phosphate pyrene strands by T4 DNA ligase or PBCV-1 DNA ligase. (**A**) T4 DNA ligase at 16 ℃, 1 mM ATP, (**B**) T4 DNA ligase at 16 ℃, 10 mM ATP, and (**C**) PBCV-1 DNA ligase at 16 ℃, 5 mM ATP from **Figure 3.12**

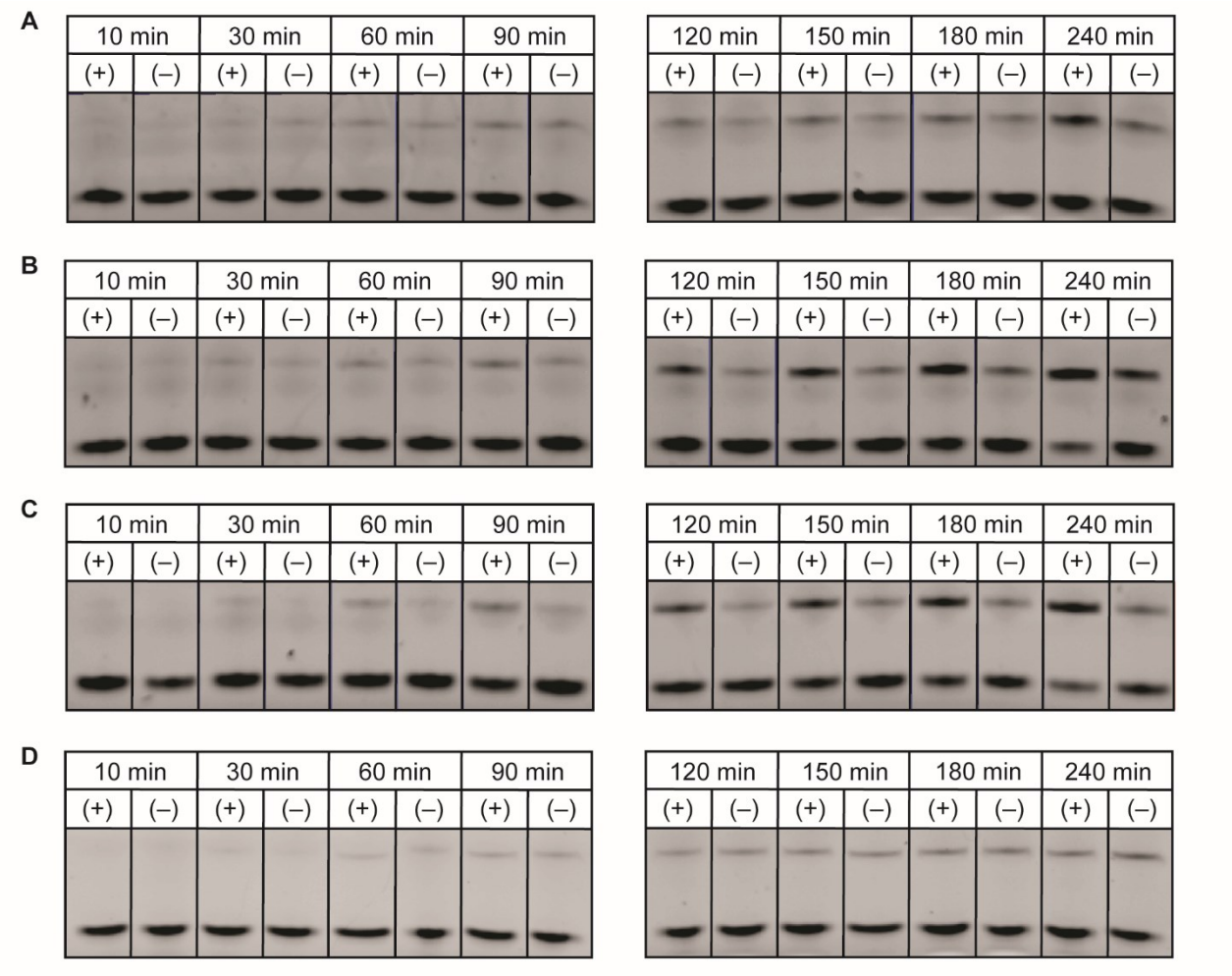


Figure 3.37 Representative PAGE image of cross-catalytic reaction with 5'-phosphate pyrene strands (DNA-IIa$_P$(pyrene)). (**A**) 32 ℃, (**B**) 34 ℃, (**C**) 36 ℃ and (**D**) 38 ℃ from **Figure 3.21.**

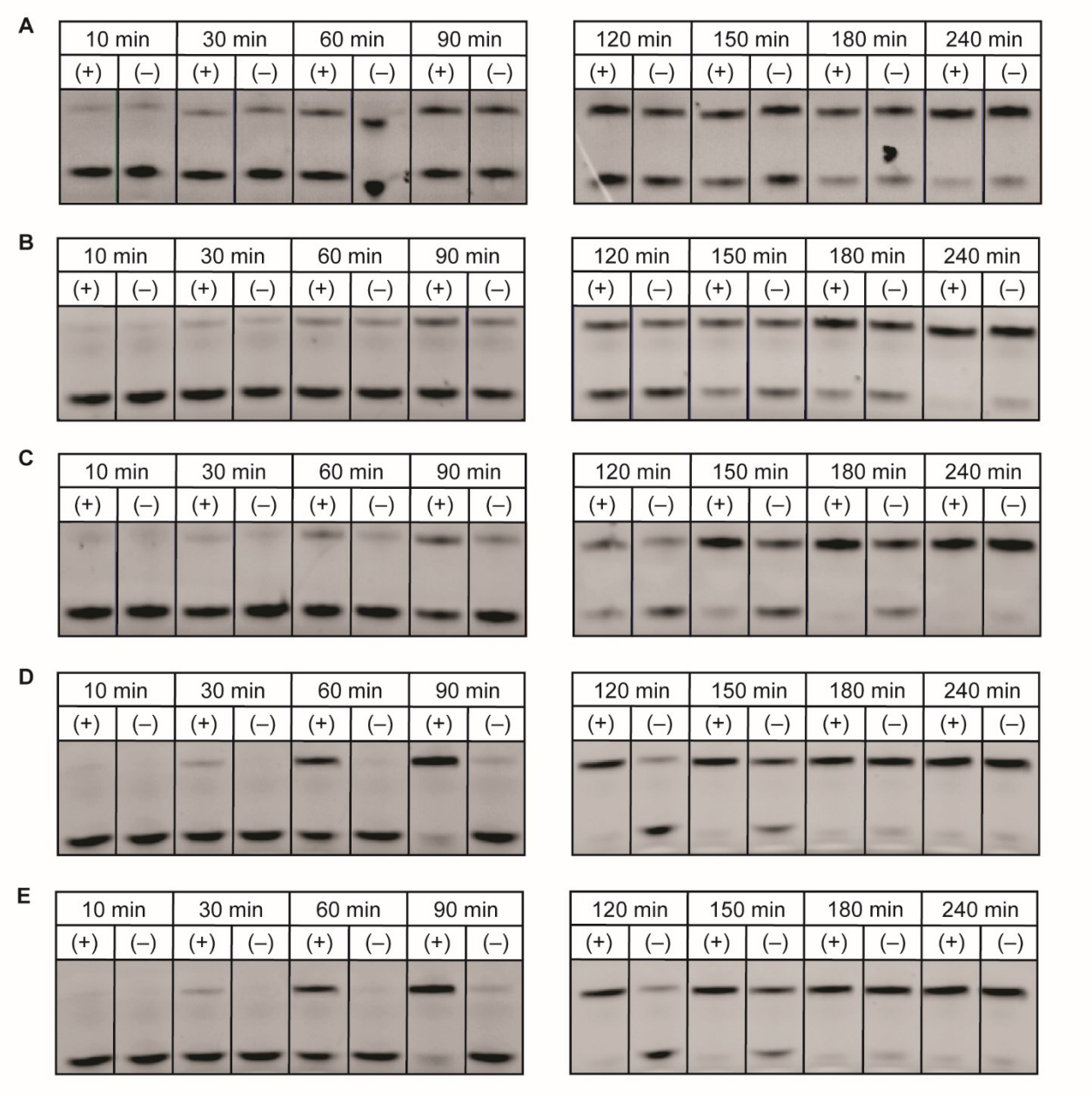


Figure 3.38 Representative PAGE image of cross-catalytic reaction with 5'-phosphate thymine nucleoside strands (DNA-IIa$_P$). (**A**) 32 °C, (**B**) 34 °C, (**C**) 36 °C, (**D**) 38 °C, and (**E**) 38 °C with 1 mM ATP from **Figure 3.23**.

Chapter 4

CRISPR-Click Enables Multi-Gene Editing with Modular Synthetic sgRNAs

"CRISPR-Click Enables Multi-Gene Editing with Modular Synthetic sgRNAs" **Park, H.**; Osman, E.A.; Cromwell, C.R.; St. Laurent, C.D.; Liu, Y.; Kitova, E.N.; Klassen, J.S.; Hubbard, B.P.; Macauley, M.S.; Gibbs, J.M. *(in revision)*

4.1 Introduction

CRISPR-Cas9 has enabled gene editing both in the lab,[113,115,117,121,247–251] allowing for the function of proteins to be readily interrogated,[252–254] and in the clinic, where it promises to offer treatments for diseases such as sickle cell anemia and cystic fibrosis.[123,124,255–258] Cas9 directs sequence specific DNA cleavage through the aid of guide RNA (gRNA) consisting of a CRISPR RNA (crRNA), which contains a 20-nucleotide (nt) programmable target sequence, and a trans-activating crRNA (tracrRNA) that binds the Cas9. The tracrRNA acts as a bridge between the crRNA and the enzyme.[113] As CRISPR-Cas9 directed cleavage can vary in efficiency depending on the target, one strategy to ensure gene knockout is to target cutting of the gene at multiple locations by introducing multiple guide RNAs.[115,117,259,260] Additionally, other applications in research, as well as clinical applications, require multiple genes to be edited.[261–263] However, targeting multiple genes through iterative rounds of gene editing requires cells to be cultured for prolonged periods of time, which can be time prohibitive and not feasible for some cell lines.[262,263] Simultaneous disruption of numerous loci in one round solves these issues.

Several solutions have been advanced with the goal of multiplexing CRISPR-Cas9 to target multiple loci simultaneously. One strategy involves encoding multiple gRNA via lentiviral transduction[264,265] or transfection of plasmids containing the appropriate sequences.[115,117,260,261,266] One particularly elegant example that has been reported is a lentiviral delivery system that uses the Cpf1 nuclease (Cas12) and a self-processing guide RNA to target four different genes.[267] While these approaches use DNA-encoded strategies, the direct introduction of a gRNA to cells has been shown to have advantages.[73,248,258,268–275] For example, direct introduction of synthetic crRNA and tracrRNA made by solid-phase, rather than enzyme-catalyzed synthesis, allows unnatural chemical modifications to be incorporated that confer benefits, such as increased stability,[269–271] reduced off-

target effects,[272,273] and fluorescent labeling.[273–275] Accordingly, an attractive approach to multiplexing genome editing that alleviates the need for in-cell processing is to introduce multiple gRNAs simultaneously into a cell. This can be achieved by complexing multiple guides using a nanoparticle delivery agent,[276,277] or introducing multiple gRNAs targeting different genes into the same cell.[261,278]

Synthetic gRNAs can be assembled using two different methods. The first is the dual strand approach, where the crRNA is hybridized to the tracrRNA; for the Cas9 system, these strands are 29-42 and 65-75 nucleotides, respectively.[113,279,280] Using this dual strand system, chemically-modified crRNAs such as those incorporating bridged nucleic acids, 2'-*O*-methyl, 2'-fluoro, phosphorothioate, 2'-*O*-methyl-thiophosphonoacetate, 2'-*O*-methyl-phosphonoacetate, or 2'-(*S*)-constrained ethyl RNA have been produced with increased activity,[269,271,279] enhanced stability,[269,271,279] or decreased off-target cutting.[273] Likewise, fluorescent modifications to the tracrRNA have been used in combination with FACS to isolate cells taking up fluorescent-labeled tracrRNA (and therefore have a higher probability of gene-target cleavage), which is helpful for cells that are difficult to transfect.[281,282] In the second synthetic approach, the full crRNA-tracrRNA sequence is combined into one strand resulting in a single-guide RNA (sgRNA) molecule of ~97 nucleotides in length.[283] It has been reported that sgRNAs synthesized by solid-phase synthesis led to higher efficiencies in gene editing compared to dual crRNA:tracrRNA systems.[271,284] However, due to challenges in synthesizing chemically-modified RNA strands of this length, it has not commonly been used to introduce chemical modifications in sgRNAs,[270–272,284,285] such as fluorescent modifications.

One of the main challenges associated with accessing chemically-modified sgRNAs stems from difficulties in synthesizing long RNA by solid-phase synthesis.[286–289] For example, the solid-phase synthesis of a 97mer sgRNA would be 5.4% if the theoretical efficiency of each step were

97% ($0.97^{96} = 0.054$). Recent progress in improving the coupling efficiency of each step in the solid-phase synthesis has allowed sgRNA to be directly produced,[270–272] but such approaches have not been expanded to introduce fluorescent modifications, which can often exhibit reduced coupling efficiencies. Moreover, the low overall yield - even with high coupling efficiencies on a per-step basis - leads to such small quantities of sgRNA material that they cannot be thoroughly purified or characterized by gel electrophoresis and mass spectrometry. This inability to produce, purify, and characterize sizable amounts of synthetic sgRNA also limits the generation of well-defined hybrid materials that facilitate cellular delivery of multiple gene-editing constructs.[290]

One strategy that has been explored to increase the amount and structural diversity of sgRNA involves synthesizing it from two smaller fragments using enzyme-catalyzed[275,291] or bioconjugation techniques.[288,289] The first reported example of the latter involved the introduction of a hexyne phosphoramidite monomer at the 5' end of the tracrRNA and a 3'-aminohexyl modified crRNA further functionalized with azido-acetic acid NHS ester.[288] The tracrRNA and crRNA were then 'clicked' together using copper-catalyzed alkyne-azide cycloaddition (CuAAC) chemistry.[288] The resulting sgRNA was capable of directing gene editing despite the large aliphatic triazole linkage that resulted from the bioconjugation reaction.[288] Tolerance of Cas9 to this linkage in the loop adjoining the crRNA and tracrRNA sequences (often referred to as the tetraloop of the sgRNA) is consistent with previous observations that various modifications at the 5'-terminus of the tracrRNA or tetraloop of the sgRNA did not negatively impact gRNA-Cas9 function.[292,293] The robust activity of sgRNA modified in this hairpin is also congruent with the crystal structure of Cas9 complexed with a gRNA, showing that the hairpin (tetraloop) connecting the crRNA and tracrRNA sequences protrudes away from the protein.[283,294] More recently, the Brown group used different alkyne-azide cycloaddition strategies to generate modified sgRNA from two smaller

fragments.[289] They explored the impact of different triazole linkages in the tetraloop as well as near the cleavage site across from the 3'-end of the protospacer adjacent motif (PAM) sequence. In both of the bioconjugation reports using CuAAC to generate the sgRNA, the fragment encompassing the tracrRNA sequence was substantial in length (65 nt, 66 nt or 79 nt), which is not optimal for solid-phase synthesis.[288,289]

Here we demonstrate that sgRNAs can be assembled from three smaller pieces of RNA, which better leverages the advantages of solid-phase synthesis, and results in a more modular synthetic strategy. This approach uses a middle fragment modified with different fluorophore labels, which is subsequently clicked to a unique crRNA sequence such that each fluorophore corresponds to a specific gene target. We demonstrate that these clicked fluorescent sgRNAs are fully functional *in vitro* and in cells, although the position of the installed triazole linkage within the tracrRNA does impact efficiency. Moreover, highlighting the utility of this approach, we have shown that two genes can be disrupted in cells with high efficiency when this strategy is used in combination with FACS. We expect this synthetic strategy for the modular assembly of sgRNAs will readily enable access to a variety of chemically-modified or functionalized RNA fragments, facilitating the production of sgRNA libraries for screening and other applications using commercially available solid-phase reagents and simple on-bead protocols.

4.2 Results and Discussion

4.2.1 Triazole Modified tracrRNA In Three Different Locations

Deconstructing the single guide RNA into three similar length fragments, unlike the two-fragment strategies previously reported,[288,289] requires that the Cas9 tolerate unnatural

oligonucleotide linkages within the portion of the gRNA corresponding to the tracrRNA sequence that binds Cas9. Therefore, we decided to first explore the impact of introducing an unnatural linkage into the tracrRNA sequence on Cas9 activity, with the motivation of being able to stitch together smaller pieces to form the longer, 63 nucleotide tracrRNA. For our unnatural linkage, we selected the triazole that resulted from the Brown lab's strategy using a copper-catalyzed alkyne-azide cycloaddition (CuAAC) of a 3'-*O*-propargyl terminus reacting with a 5'-azide terminated strand (**Figure 4.1**). This linkage is known to be remarkably well tolerated by nucleic acid-binding enzymes and can even be read through by some polymerases.[295,296] Another modification that is well tolerated by Cas9 is the replacement of ribonucleotides with deoxyribonucleotides at several locations in the gRNA sequence.[297,298] These previous reports suggested that we could use commercially available 3'-*O*-propargyl-methyldeoxycytosine-modified solid supports to introduce the propargyl functionality and 5'-iodo-modified deoxythymidine to introduce the azide.

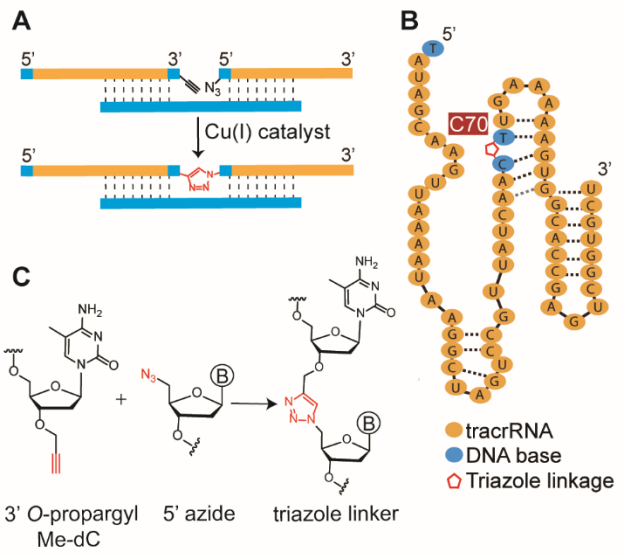


Figure 4.1 Template-assisted CuAAC assembly of the tracrRNA. (**A**) Schematic diagram illustrating the DNA-templated copper-catalyzed alkyne-azide cycloaddition (CuAAC) of RNA strands terminated with 3'-*O*-propargyl or 5'-azido modified deoxyribonucleotides. (**B**) tracrRNA sequence with the triazole linkages (C70:T71) and DNA substitutions shown. tracrRNA was terminated with a 5'-thymidine to enhance the stability against RNase degradation. (**C**) Molecular

structure depicting triazole formation from the 5'-azide and 3'-*O*-propargyl Me-dC terminated strands.

We prepared three different lengths of two pairs of RNA strands, 5' azide terminated and 3' propargyl terminated strands that would result in a triazole linkage in tracrRNA strands (**Figure 4.2 and 4.3**). These RNA strands were synthesized to test the unnatural triazole linkages in various positions in the tracrRNA, including those involved in Cas9 binding (**Figure 4.3B**). Two of the 5'-azide strands were synthesized using the commercially available 5'-iodothymidine phosphoramidite followed by nucleophilic substitution with NaN_3 prior to base deprotection and cleavage of the DNA-RNA chimera from the solid support.[299] We also synthesized a strand containing a 5'-azidodeoxyribocytodine (C61) by directly iodinating the 5'-hydroxy group while still on the bead.[299] Introducing the triazoles at these positions allowed us to explore the impact of the triazole linkage location within the tracrRNA on Cas9 activity (**Figure 4.3A, 4.3B**). Our previous work showed that the CuAAC reaction of 5'-azido and 3'-*O*-propargyl terminated DNA strands proceeded rapidly using a DNA template to bring together the reactive termini in the presence of a Cu(I) catalyst and 1.5 equivalents (with respect to copper) of a benzimidazole ligand.[300] Using similar conditions with argon degassing, we observed quantitative conversion of the limiting 3'-alkyne oligonucleotide fragment after 1 hour of reaction. The triazole-linked tracrRNA was observed in polyacrylamide gel by the appearance of the new strand (63 nt) compared to the reference DNA strand (63 nt) (**Figure 4.2**). After the reaction, the triazole-linked tracrRNAs were purified by polyacrylamide gel electrophoresis and characterized by mass spectrometry.

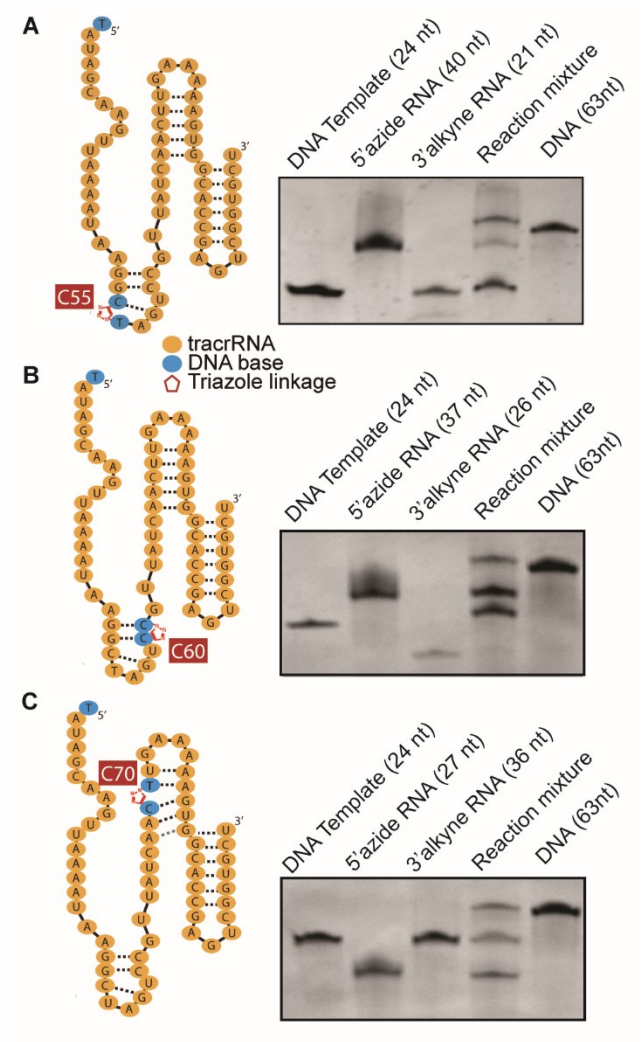


Figure 4.2. Polyacrylamide gel electrophoresis images of triazole modified tracrRNAs. Triazole linkages are in three different positions (**A**) C55:T56, (**B**) C60:C61, and (**C**) C70:T71. First lane -DNA template (24 nt), second lane - 5' azide terminated RNA strands (various length), third lane – 3' alkyne terminated RNA strands (various length), fourth lane – reaction mixture after 2 hours, and fifth lane – 63 nt DNA strand from IDT (Integrated DNA Technologies).

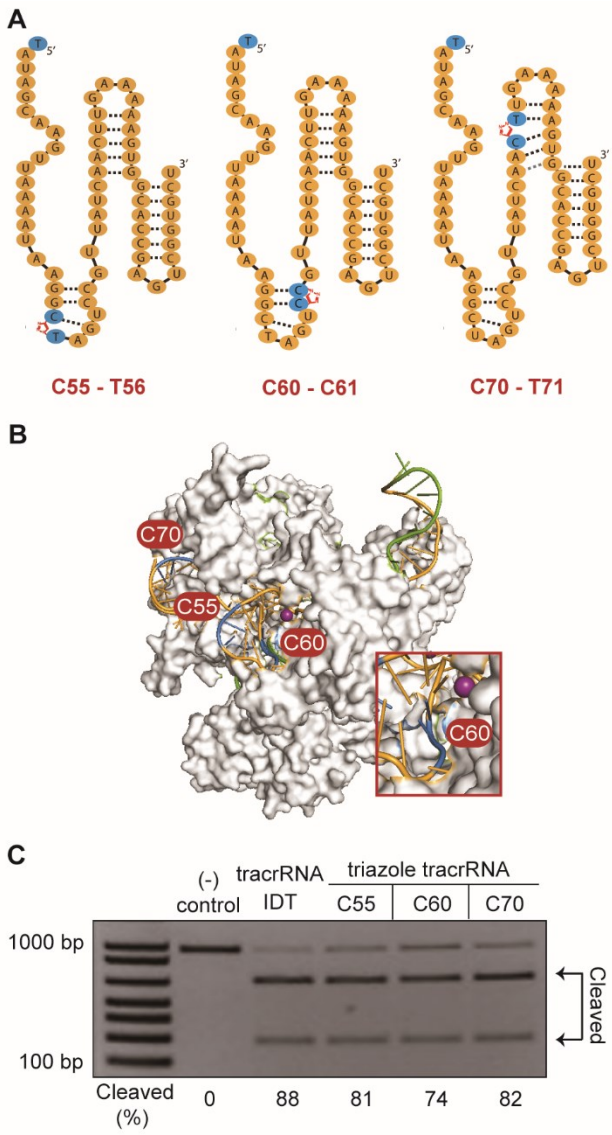


Figure 4.3 Exploring three different locations for the triazole linkage in tracrRNA. (**A**) Sequences of the three clicked tracrRNA were explored with triazole linkages resulting from 3'-*O*-propargyl modifications at the C55, C60, and C70, respectively. The tracrRNA was synthesized from two fragments that resulted in a triazole between the C55:T56, C60:C61, and C70:T71 positions, respectively. Ribonucleotides are shown in orange, and deoxyribonucleotides are shown in blue. The 5'-dT was introduced to increase the stability of the tracrRNA. The crRNA was obtained separately from a commercial source (Integrated DNA Technologies). (**B**) Structure of Cas9 in complex with a gRNA (PDB 4OO8) highlighting the position of the three locations where the tracrRNA was clicked together in our studies, shown in blue. (**C**) Gel electrophoresis image of the *in-vitro* Cas9 cutting assay with commercially available tracrRNA and triazole-containing tracrRNA at different positions.

We evaluated the activity of the three triazole-linked tracrRNAs using an *in vitro* Cas9 cutting assay that can detect DNA target strand cleavage. In this assay, the various triazole-linked tracrRNA were hybridized to a commercial crRNA (from Integrated DNA Technologies; IDT) that targets *EMX1* (**Figure 4.3C**), complexed with Cas9, then added to a reaction containing amplified DNA corresponding to the *EMX1* gene target site. We observed cutting of the DNA target for the tracrRNAs containing triazole linkages at the C55:T56 and C70:T71 positions (81% and 82% cleavage, respectively) which was comparable to the results observed for the commercial tracrRNA (88% cleavage). Remarkably, even the tracrRNA with the triazole at the C60:C61 junction in the region of the tracrRNA sequence that binds to the Cas9 nuclease[44] exhibited excellent nuclease activity, albeit slightly reduced (74% cleavage). The appreciable activity for the tracrRNA with the triazole linkage in close proximity to the bound protein provides further support that this specific linkage generated from 3'-*O*-propargyl and 5'-azide nucleotides is accepted by many different enzymes owing to its size and hydrogen-bonding ability that mimic the phosphate linkage.[301]

4.2.2 Exploring the Triazole Linkage in sgRNA

Having determined that the triazole linkage in the tracrRNA is not deleterious to Cas9 activity, particularly at locations C55:T56 and C70:T71, we next explored the synthesis of sgRNA from three strands. This strategy required the synthesis of a middle sequence that contained both a 5'-azide and a 3'-*O*-propargyl methyldeoxycytosine. When fluorescently labeled sgRNA were desired, we also introduced an additional dT nucleotide, positioned in the tetraloop region that contained or was modifiable with a fluorophore. All three strands were then clicked simultaneously using two DNA templates to selectively react the desired azide and alkyne termini and prevent unwanted by-products (**Figure 4.4**).

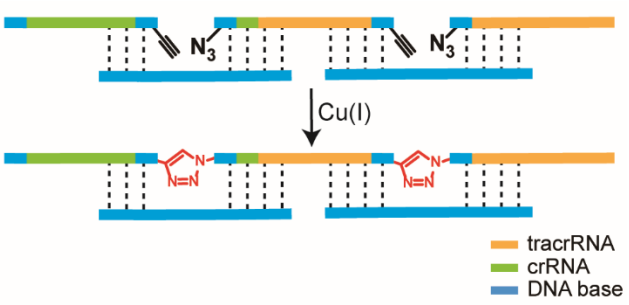


Figure 4.4 Scheme of double-click CuAAC of sgRNA. Blue represents DNA bases, green represents crRNA, and yellow indicates tracrRNA. The Cu catalyst was used to induce Cu(I) catalyzed alkyne-azide cycloaddition.

Two different lengths of pairs were examined for double-click sgRNA. Considering the high cleavage efficiency of the C55:T56 position and the C70:C71 position, we tested both locations with an additional position C30:T31 that linked the crRNA sequence which contains the sequence complementary to the target gene (**Figure 4.5**). First, we synthesized a sgRNA that targeted a particular sialic acid-binding immunoglobulin-type lectin (Siglec), *Siglec-3* or *CD33*, which has recently been implicated as an immune checkpoint. The resulting sgRNA product contained a triazole at the C55:T56 or C70:T71 position and at the C30:T31 position near the gene targeting sequence (**Figure 4.5**). The CuAAC reaction progress was monitored by denaturing polyacrylamide gel electrophoresis (PAGE), which revealed the generation of a larger strand of the expected size of the sgRNA (97 mer) (**Figure 4.5B and 4.5D**). These modified sgRNAs were purified by extraction from PAGE gels and characterized by electrospray ionization mass spectrometry. Both sgRNAs containing the two triazole linkages were tested in an *in vitro* assay to assess their activities with Cas9 protein. (**Figure 4.5**) We observed cutting of the *Siglec-3* target (C:50:T56, 88% cleavage and C70:T71, 95% cleavage) that was comparable to the results observed for the commercial dual gRNA (~86% cleavage). The triazole modifications at the C70:T71

position as well as the C30:T31 position (Figure 4.5A) henceforth were used to investigate the further studies in sgRNA due to higher cleavage activities with Cas9 protein.

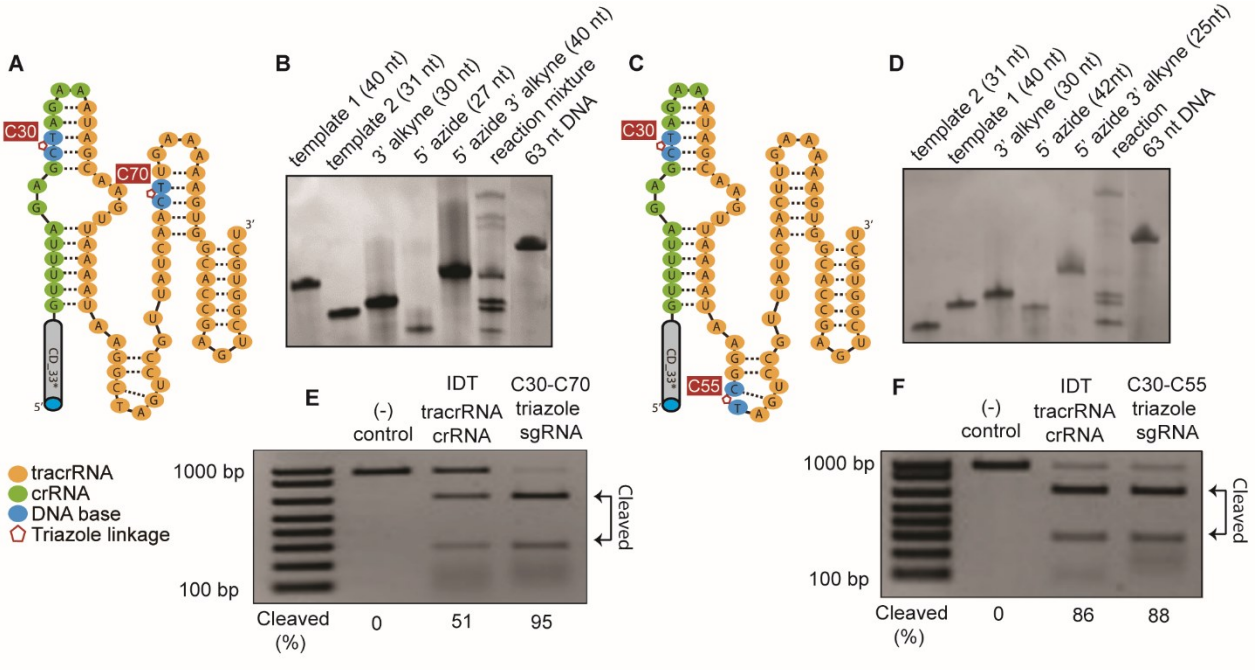


Figure 4.5. Double-clicked (CuAAC) triazole modified sgRNAs. (**A**) Schematic of sgRNA targeting *Siglec-3* containing triazole linkages at the C30:T31 and the C70:T71 positions. (**B**) Corresponding PAGE gel image of reaction mixture and starting materials for generating the sgRNA in A. (**C**) Scheme of sgRNA targeting *Siglec-3* containing triazole linkages at the C30:T31 and the C55:T56 positions. (**D**) Corresponding PAGE gel image of reaction mixture and starting materials for generating the sgRNA in C. (**E**) Agarose gel analysis of the *in vitro* cutting assay of *Siglec-3* target with double-clicked sgRNA from part A (C30-C70) compared with the commercial dual gRNA from IDT. (**F**) Agarose analysis of the *in vitro* cutting assay of *Siglec-3* target with double-clicked sgRNA from part C (C30-C55) (A) compared with the dual gRNA from IDT. *Control (-)*: only the sgRNA was added without Cas 9 protein.

4.2.3 Triazole Modified sgRNA with Fluorescence Labeling

One of the major advantages of our approach is that modifications can be incorporated into the fragments while still achieving substantial yields by solid-phase synthesis. Accordingly, we incorporated a fluorescein-modified thymidine to the tetraloop region, adding an additional

nucleotide between A35 and A36. To introduce a distinct second label, we prepared a middle fragment with a modified thymidine at the same position that contained a trifluoroacetyl-protected primary alkyl amine. After base deprotection and removal from the solid phase, the primary alkyl amine on this strand was reacted with an ATTO-550 fluorophore modified with an *N*-hydroxysuccinimidyl (NHS) ester following standard coupling and purification procedures.

4.2.3.1 Triazole Modified sgRNAs Targeting EMX1 and WAS Sequences

To test whether the fluorescent modified sgRNAs containing two unnatural triazole linkages was active against the *EMX1* target, an *in vitro* DNA cleavage assay was once again performed using the *EMX1* target sequence. We compared the *in vitro* activity of our sgRNA modified with the *EMX1* targeting sequence with a scrambled target sgRNA containing the triazole modifications at the same two positions. Excellent activity for the *EMX1* targeting sgRNA (85% cleavage) was observed with no activity for the scrambled sgRNA (**Figure 4.6D**), indicating that the sgRNA constructed from three smaller fragments maintained its selectivity in targeted cleavage. Next, we evaluated whether our double-clicked sgRNA was active in cells by transfecting Cas9-expressing HeLa cells with a ribonucleotide polymer complex composed of the synthetic sgRNA and lipofectamine. The target DNA was extracted, and the target region was amplified by PCR. The activity of the synthetically modified sgRNA in cells was measured by the presence of insertions or deletions (INDELs) in the target sequence using the T7 endonuclease I assay.[302] Despite the presence of two unnatural triazole linkages, we observed the double-clicked sgRNA was active in the unsorted cells after transfection based on the cleavage of the target by T7 endonuclease I (18% cleavage) (**Figure 4.6E**).

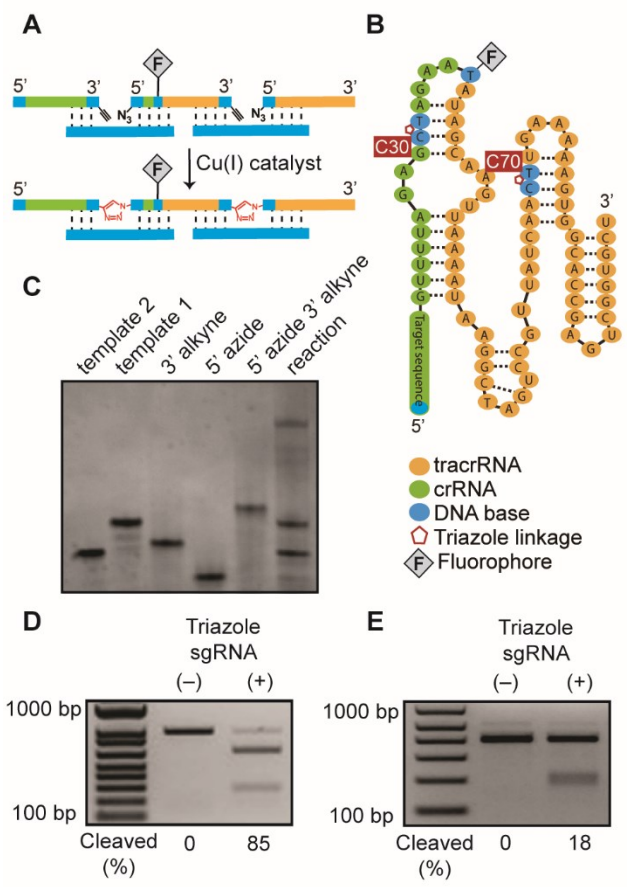


Figure 4.6 Synthesis and activity of double-clicked sgRNA. (**A**) Reaction scheme illustrating the formation of double-clicked sgRNA with fluorophore. (**B**) sgRNA sequence with the functionalized extra dT between A35 and A36. Green represents crRNA, yellow represents tracrRNA, blue represents DNA bases. Triazole linkages are located in between C30:T31 and C70:T71. (**C**) PAGE image of the double-click reaction. (**D**) *In vitro* assay illustrating the activity of the sgRNA targeting *EMX1*. The control lane (-) is the same experiment using a different sgRNA that did not target the *EMX1* gene. (**E**) The T7 endonuclease I assay illustrates cleavage of the *EMX1* target in unsorted HeLa cells expressing Cas9 after transfection with the *EMX1*-targeting sgRNA. The control lane (-) is the result of the T7 endonuclease I assay with cells not transfected with the sgRNA.

Next, we aimed to evaluate whether these fluorescent sgRNA could be used in a dual-color/dual-gene editing process. An ATTO 550-modified middle fragment was used to prepare two sgRNA targeting the *EMX1* or *WAS* genes and a fluorescein modified middle fragment was used to prepare an sgRNA targeting *WAS* (**Figure 4.8A**). We then evaluated all three sgRNAs activities

with Cas9 using the *in vitro* DNA cleavage assay (**Figure 4.7A-B**). As mentioned earlier, the *EMX1* targeting sgRNA showed excellent activities (86% cleavage). However, the *WAS* targeting sgRNA exhibited some difference in activities between ATTO-550 labeling (86% cleavage) and fluorescein labeling (42% cleavage) in the *in vitro* DNA cleavage assay. This difference in activity may come from the experimental variation although this experiment was conducted only once, the following experiment (**Figure 4.7C**) in Cas9 expressing cells did not have such a difference between fluorophores. Next the activity of both *WAS* targeted sgRNA was evaluated in the transfected unsorted Cas9-expressing HeLa cells based on the T7 endonuclease I cleavage assay (17% cleavage for ATTO 550 labeling and 18% cleavage for fluorescein labeling), similar to the efficiency of the *EMX1*-targeted sgRNA (18% cleavage) (**Figure 4.7C**). This similarity between the two WAS-targeting sgRNA suggested that the fluorophore did not impact the activity of the single guides. To take advantage of the presence of the fluorescent labels on the sgRNA and our ability to sort the cells based on their fluorescence, we transfected the Cas9-expressing cells with ATTO 550-labeled *EMX1* gRNA or fluorescein-labeled *WAS* sgRNA individually or together using lipofectamine. Using fluorescence-activated cell sorting (FACS), cells were then sorted for the single-positive ATTO 550^+ or fluorescein$^+$ cells from the individual transfections, or the double-positive ATTO 550^+ and fluorescein$^+$ population from the cells transfected with both sgRNAs (**Figure 4.8B**). Most of the cells took up the sgRNAs following transfection as seen by the shift in fluorescence in the appropriate channels (**Figure 4.8C**). The top 40% of the positive cells were sorted in each case. After cell sorting, we quantified the activity of these sgRNAs in cells by measuring cleavage of the target site using the T7 endonuclease I assay (**Figure 4.8D**).[302] We found that the ATTO 550-labelled *EMX1* sgRNA showed a very similar %cleavage in both the single sort and dual sort consistent with the high transfection (19% vs 18%, respectively). However, the fluorescein-labeled *WAS* sgRNA had higher %cleavage in the single sort versus double sort

experiment (20% vs 12%, respectively). These results demonstrate that two fluorescently modified synthetic sgRNAs produced through a double click reaction can successfully be transfected into cells to facilitate simultaneous dual-gene editing.

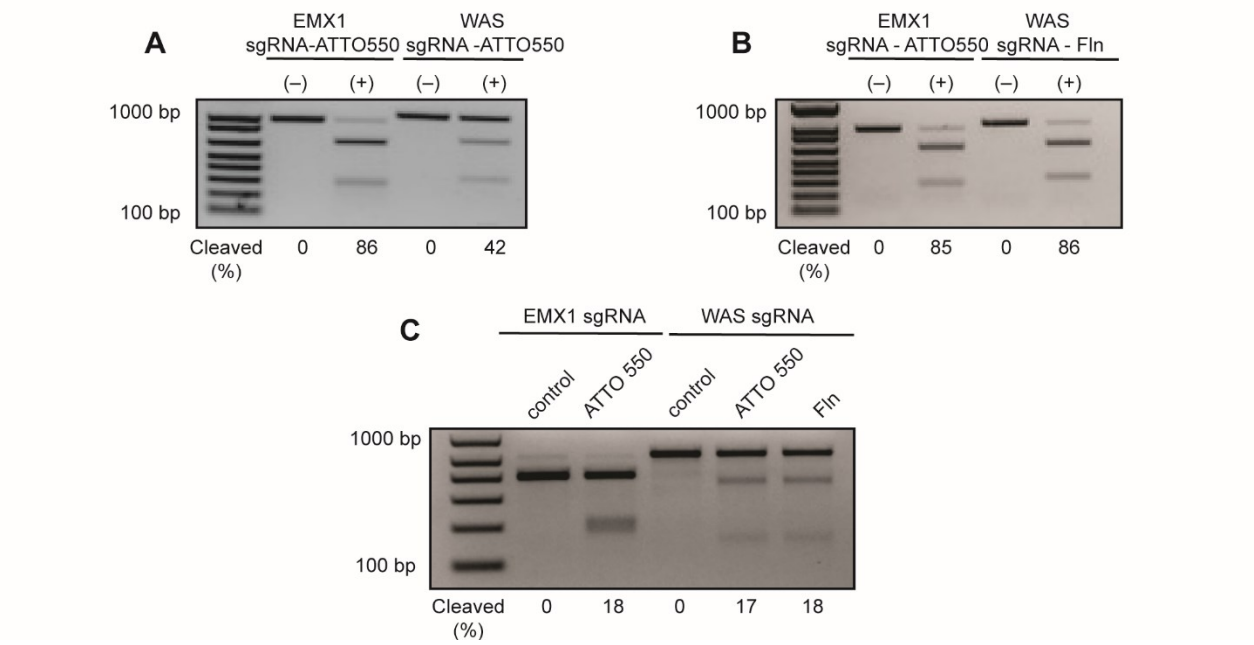


Figure 4.7 In vitro DNA cleavage assay and T7 endonuclease I assay of EMX1 and WAS targeting sgRNAs. (**A**) *In vitro a*ssay of ATTO 550-labeled *EMX1* sgRNA and ATTO550-labeled *WAS* sgRNA. *Control (-)* experimented only with sgRNA without Cas9 protein. (**B**) *In vitro* assay of ATTO 550-labeled *EMX1* sgRNA and fluorescein-labeled *WAS* sgRNA. *Control (-)* indicates the sgRNA not targeting *EMX1* or *WAS***.** (**C**) T7 endonuclease I assay in unsorted HeLa cells expressing Cas9, illustrating cleavage of ATTO 550-labeled *EMX1* sgRNA, ATTO550-labeled *WAS* sgRNA, and fluorescein-labeled *WAS* sgRNA. The *control lane (-)* is the result of the T7 endonuclease I assay with cells not transfected with the sgRNA.

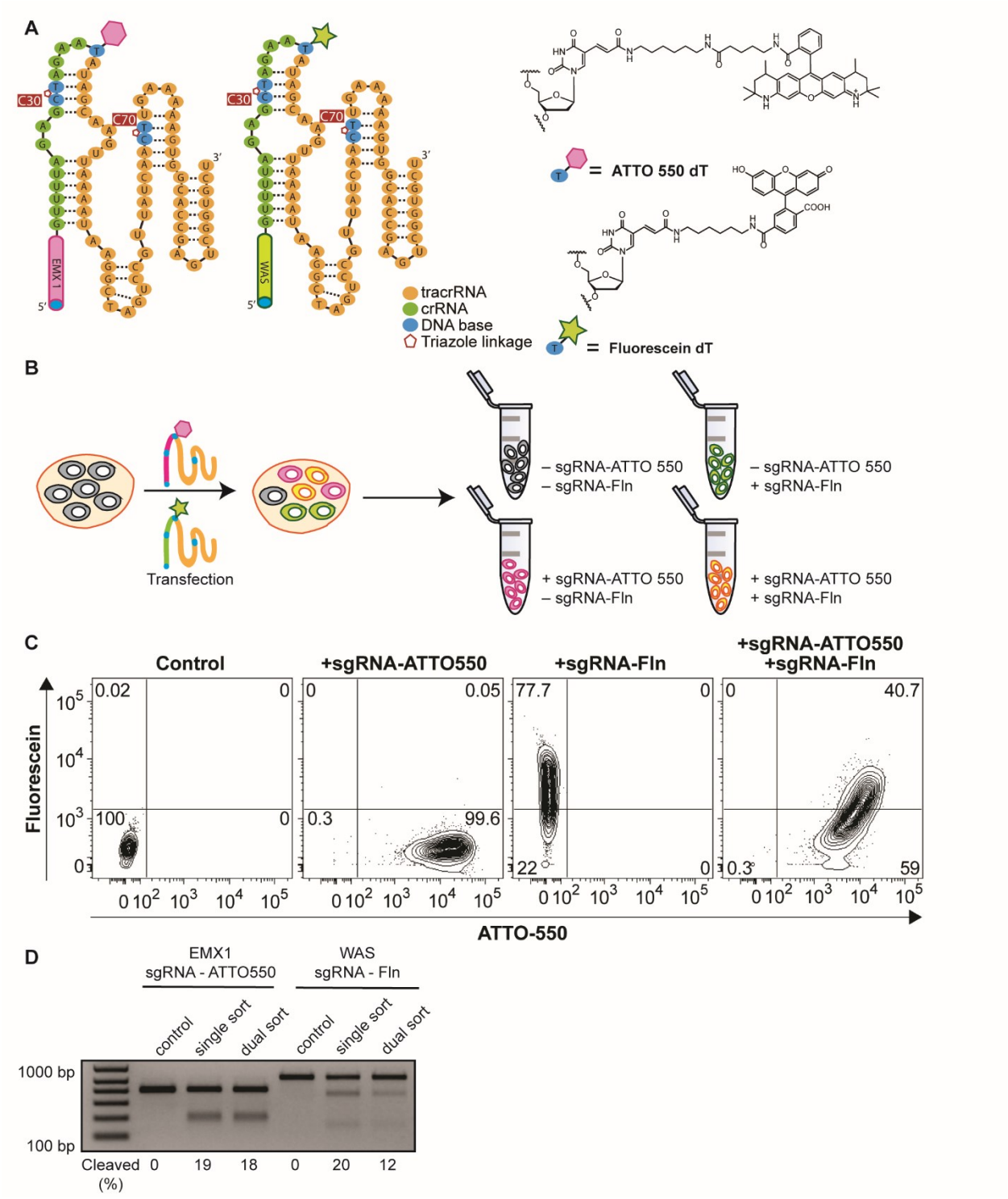


Figure 4.8 Multiple gene editing using dual fluorescent sgRNA and FACS. (**A**) sgRNA structures of *EMX1* targeting sgRNA labeled with ATTO 550 (pink hexagon) and of *WAS* targeting sgRNA labeled with fluorescein (green star). (**B**) Schematic illustrating general cell-based assay with fluorescent-labeled sgRNA. Both sgRNAs were transfected in Cas9 expressing Hela cells followed by FACS sorting based on ATTO 550 and fluorescein emission. Negative control was

not sorted. (**C**) FACS results of Hela cells transfected with ATTO 550-labeled *EMX1* sgRNA or fluorescein-labeled *WAS* sgRNA for the single-color transfections and a 50:50 mixture of both sgRNAs for the dual-color transfections. The control cells were not transfected by either sgRNA. (**D**) Results from the T7 endonuclease I cleavage assay for the *EMX1* target and *WAS* target performed on cells. *Control* - cells not transfected with any sgRNA (unsorted); *single sort* - cells transfected with either ATTO 550-labeled *EMX1* sgRNA or fluorescein-labeled *WAS* sgRNA and sorted accordingly by FACS; *dual sort* - cells transfected with both sgRNAs and sorted for both fluorescent colors by FACS.

4.2.3.2 Triazole Modified sgRNAs Targeting Siglec Sequences

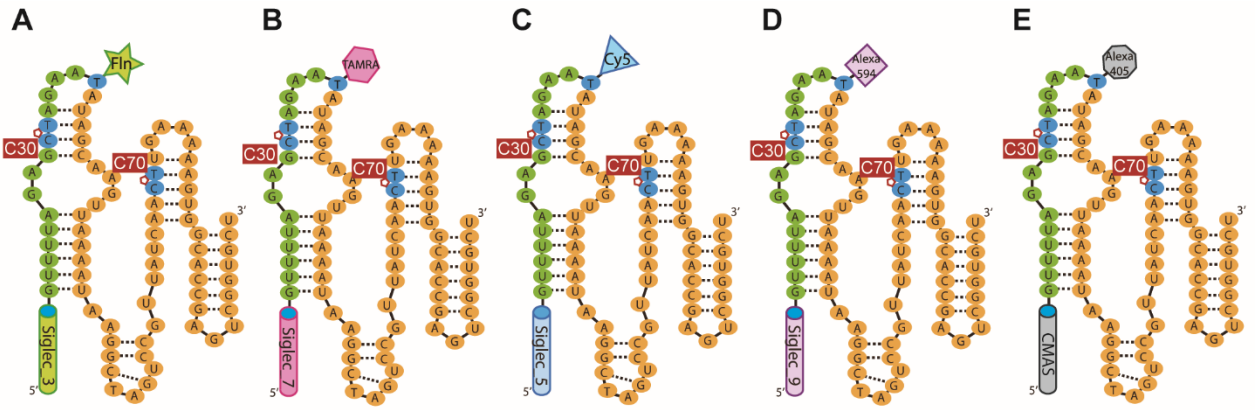


Figure 4.9 Five different fluorophores targeting five different targets. (**A**) *Siglec-3* targeting sgRNA labeled with fluorescein. (**B**) *Siglec-7* targeting sgRNA labeled with TAMRA. (**C**) *Siglec-5* targeting sgRNA labeled with Cyanine 5. (**D**) *Siglec-9* targeting sgRNA labeled with Alexa 594. (**E**) *CMAS* targeting sgRNA labeled with Alexa 405. Corresponding PAGE gel showing the click reactions are illustrated in **Figure 4.15** and **Figure 4.16**.

After establishing our general strategy for generating functional and fluorescent-labeled sgRNAs from two simultaneous click reactions, we next explored its modular nature and its amenability to FACS in a multi-color, multi-target experiment. First, we synthesized five different 5'-termini fragments of the sgRNA containing a 3'-*O*-propargyl methyl-dC that targeted *Siglec-3, Siglec-5, Siglec-7, Siglec-9,* and *CMAS* (cytidine monophosphate *N*-acetylneuraminic acid synthetase). We also conjugated the middle fragment of the sgRNA that contained an amino-modified dT with an NHS ester modified cyanine5 or Alexa 594 or Alexa 405 fluorophore as

mentioned above. TAMRA and fluorescein modifications were incorporated directly by solid-phase synthesis using the corresponding fluorophore-modified dT. To generate each sgRNA, two simultaneous DNA-templated click reactions with the three fragments were performed. Using this route, five different sequences targeting sgRNA with five different fluorophores were successfully generated (**Figure 4.9** and **Figure 4.15**).

First, three triazole-modified sgRNAs, fluorescein-labeled *Siglec-3* sgRNA, TAMRA-labeled *Siglec-7* sgRNA, and Cy5-labeled *Siglec-5* sgRNA, were combined with Cas9 and transfected into U937 cells. Cells were sorted that were positive and negative for all three fluorophores and protein expression was evaluated by flow cytometry using fluorescent-labeled antibodies for each Siglec. The results were not reproducible and inconclusive (data not shown). Further, new sgRNA for *Siglec-3*, *Siglec-5*, *Siglec-7*, and *Siglec-9* were synthesized using different targeting sequences to increase the chance for target cleavage. Using this new sequence, a fluorescein-labeled *Siglec-3* sgRNA was tested in an *in vitro* assay, and it showed very good activity (92% cleavage) (**Figure 4.10B**). Next, newly synthesized fluorescein-labeled *Siglec-3* sgRNA and TAMRA-labeled *Siglec-7* sgRNA were combined with Cas9 and lipofectamine and separately transfected in U937 cells as well as ATTO-550 labeled dual guide RNA from IDT that targeted *Siglec-3*. The top 10-30% of fluorescent cells were sorted (**Figure 4.10A**). Five days later, protein expression was assessed using flow cytometry using fluorescent labeled antibodies against Siglec-3 and Siglec-7. The flow cytometry data showed only ~7% of knockout for the cells transfected with the commercial *Siglec-3* dual guide RNA from IDT, while negligible knockout was observed for both triazole-modified sgRNAs (**Figure 4.10C**).

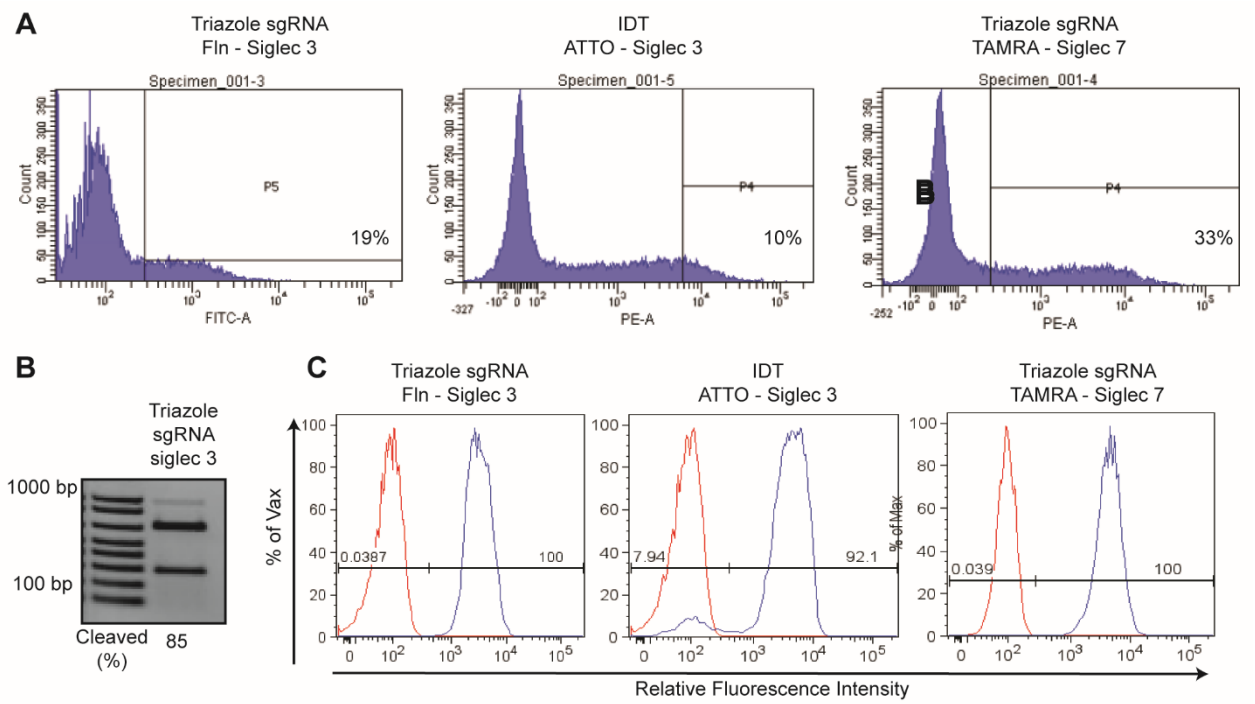


Figure 4.10 Multiple gene editing using fluorescent-labeled sgRNAs. (**A**) Raw cell sorting data of B. fluorescein-labeled *Siglec-3* sgRNA was sorted top 19 %, ATTO 550-labeled *Siglec-3* dual gRNA was sorted top 10 % and TAMRA-labeled *Siglec-7* sgRNA was sorted top 33 %. (**B**) *In vitro* assay illustrating the activity of fluorescein-labeled *Siglec-3* sgRNA. (**C**) Fluorescent cell sorting of U937 cell transfected with fluorescein-labeled *Siglec-3* sgRNA and TAMRA-labeled *Siglec-7* sgRNA or ATTO 550-labeled *Siglec-3* dual gRNA from IDT. Red indicates the unstained cell as negative control and blue is stained with anti-Siglec-3 or anti-Siglec-7 antibodies. If the knock-out was successful, a portion of the blue peak should shift towards the red peaks. Both triazole-modified sgRNAs showed ~ 0.3% knock-out while IDT *Siglec-3* dual gRNA showed 7.9% knock-out.

Our previous success with *EMX1*- and *WAS*-targeted sgRNA in HeLa cells used a Cas9 expressing cell line. To see whether the lack of gene editing of the Siglec targets stemmed from the challenges of transfection and Cas9-targeted cleavage in the cellular environment, we repeated the experiments with *EMX1*- and *WAS*-targeted sgRNA using 293T cells that did not express Cas9 (**Figure 4.11A**). As such Cas9 was introduced during sgRNA transfection with the lipofectamine. The results of the T7 endonuclease I cleavage assay are shown for these experiments with unsorted 293T cells. The cleavage efficiency decreased to some extent for the *EMX1* target in the 293T cells

when compared with that shown in the Cas9-expressing HeLa cells (14 vs 18%, respectively), while no cleavage was observed for the *WAS*-targeted sgRNA in the transfected 293T cells. The successful Cas9-targeted cleavage of *Siglec-3* and *WAS* in *in vitro*, yet the lack of effective gene editing in cells suggests that the sgRNA are not stable enough to survive intact prior to accessing the targets for targets that are relatively difficult to edit. Moreover, dual gRNA from IDT showed weaker activity in the 293T cells for *Siglec-3* compared with *EMX1* in the T7 endonuclease I assay (**Figure 4.11B**). IDT uses proprietary modifications to enhance the stability of their dual guide RNA for CRISPR-Cas9 gene editing, so the lower efficiency for the *Siglec-3* target suggests it is intrinsically more difficult to edit. The success of the IDT gRNA also indicates that extra stabilizing modifications are required to enhance the generality and stability of the double-clicked sgRNAs in cells, such as 2'-*O*-methyl, 2'-*O*-methyl-thiophosphonoacetate, or 2'-*O*-methyl-phosphonoacetate.[269,271,279] Such a strategy is being currently explored in the Gibbs and Macauley labs by my junior colleague. We speculate that the *EMX1* is an easier target to edit compared with both the *WAS* and *Siglec-3* targets, which is why we observe gene editing in the 293T cells using our double-clicked sgRNA, consistent with the common use of *EMX1* as a proof of principle in CRISPR-Cas9 tool development research.[119,273,303]

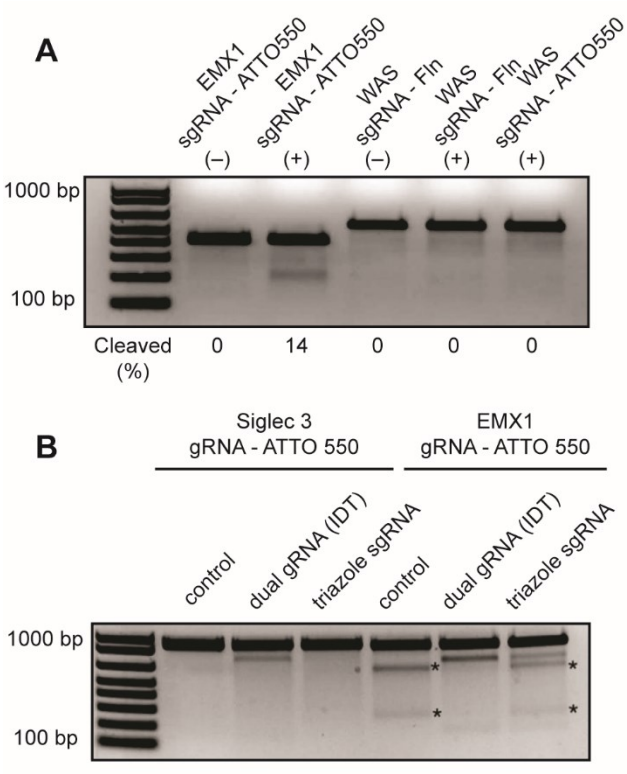


Figure 4.11 Cell based assay of triazole-modified sgRNAs. (**A**) T7 endonuclease I assay of ATTO 550-labeled *EMX1* sgRNA, fluorescein-labeled *WAS* sgRNA, and ATTO550-labeled *WAS* sgRNA for unsorted 293T cells transfected with Cas9 and sgRNA using lipofectamine. *Control (-)* represents untransfected cells. (**B**) T7 endonuclease I assay for ATTO 550-labeled *Siglec-3* sgRNA and ATTO 550-labeled *EMX1* sgRNA compared to commercially available dual gRNA (IDT). The asterisk indicates the by-product bands from PCR due to the Cas9 over digestion. All four gRNA were transfected with Cas9 ribonucleoprotein containing lipofectamine in 293T cells. *Control* - non transfected 293T cells.

4.3 Conclusions

The approach described herein enables the modular and rapid synthesis of fluorescent or other chemically modified sgRNAs and is scalable to the production of large quantities. By dividing the tracrRNA sequence into two fragments, we were able to generate sgRNA from three fragments of lengths 30, 41, and 27 nt. With these lengths, incorporating chemically modified nucleic acids that improve specificity or guide RNA stability becomes practical. Moreover, our successful implementation of deoxyribonucleic acids at the bioconjugation sites allows anyone with access to a solid-phase synthesizer to make these constructs using commercially available materials and straightforward on-bead reaction protocols.[299] Thus, this modular bioconjugation strategy of sgRNA can be assembled to achieve multi-color labeling with different fluorophores as well as different chemical modifications. Taken together, this modular strategy of synthesis of sgRNA opens up new possibilities for CRISPR/Cas biotechnological applications such as multi-gene knockout and imaging.

4.4 Experimental Section

4.4.1 Oligonucleotide Synthesis

With the exception of the 63nt DNA ruler that was purchased from Integrated DNA Technologies (IDT), oligonucleotides were synthesized with an Applied Biosystem 394 automated DNA/RNA synthesizer using reagents from Glen Research following their recommended procedures. 2'-*O*-TBDMS phosphoramidites (Bz-A, Ac-G, Ac-C, and U) and 2'-*O*-Ac-U-CPG were used for the RNA nucleotides. The DNA phosphoramidites utilized were dC, dT, Fluorescein-dT, 5'-iodo-dT, Amino-Modifier C6-dT, 3'-*O*-5-Methyl-dC Controlled Porous Glass (CPG). To

cleave the modified RNA from the solid support, a 1:1 mixture of 30% ammonium hydroxide (Aldrich 1336-21-6) and 40 % methylamine (Aldrich 74-89-5) was freshly prepared and used following the recommended procedure with an additional AMA incubation of the CPGs for 40 or more minutes. For the 5'-OH terminated strands, we utilized the DMT-on purification protocol with Glen-pak cartridges.

4.4.2 On-CPG 5'-Azide ONs Modification.

For the C60 tracrRNA strand, the 5'-hydroxy of the terminal dC nucleotide had to be iodinated following the literature procedure.[299] Specifically, after DMT deprotection of the terminal dC nucleotide, the column containing the CPGs was briefly affixed to the house vacuum to remove the additional solvent. Methyltriphenoxyphosphonium iodide (450 mg) was dissolved in dry DMF (2 mL, ≥99.9% OmniSolv®). The solution was transferred into a 1-mL plastic syringe fit. The column containing the CPGs was connected with the syringe containing the iodide solution and an empty 1 mL syringe and the solution was syringed back and forth for 5 min and then allowed to stand for 1 hour at room temperature. After an hour, the solution was removed from the CPGs and the CPGs were washed with dry DMF (2-3 mL) and dried on the house vacuum, and immediately used in the next step. The iodo-dT terminated strand made directly from the commercially available phosphoramidite or the iodo-dC prepared above was used in the next step. A solution of sodium azide (50 mg in 1 mL dry DMF) was prepared and heated for 10 min at 75 °C to facilitate dissolution. After brief cooling, the azide solution was taken into a 1 mL syringe synthesis and introduced to the iodinated strands on the CPGs also utilizing syringe-synthetic techniques. The CPGs connected to both syringes were left in the azide solution overnight at room temperature. The next morning, the azide solution was removed, the CPGs were washed with dry

acetonitrile (5 mL) and dried on the house vacuum. The 5'-azide ONs was then cleaved from the solid support and further deprotected following the recommended RNA deprotection procedure from Glen Research. To the deprotected 5'-azide ONs, aqueous sodium acetate (3M, filtered, 25 μL) and 1-butanol (1 mL) were added followed by vortexing for 30 seconds. The oligo solution was cooled down to -20 °C for 30 min. The precipitated ONs and solution were centrifuged (12,000 rpm, 10 min). The supernatant was decanted, and the pellet was washed with ethanol (95 %, 0.75 mL) twice followed by desalting of the ONs using a Glen-Pak cartridge.

4.4.3 Fluorophore-dT Modified ONs.

For the 5'-azide, 3'-alkyne modified ONs bearing the internal fluorophore label, the Amino-Modifier C6-dT was used in the solid-phase synthesis followed by azide functionalization and strand deprotection and purification as described above. This amine-functionalized oligonucleotide (50 nmol) was dissolved in a carbonate/bicarbonate buffer (200 μL, 0.1 M, pH 9.2). An NHS-ester modified ATTO 550 (Sigma Aldrich cat # 92835), Cyanine5 (Lumiprobe, cat #23020), Alexa 594 (Thermo Fisher #A200004), and Alexa 405 (Thermo Fisher #A30000) (350 nmol in 7 μL of DMSO) was added to the oligo solution, mixed well, and left to stand at room temperature for 40 min. After vortexing, the reaction mixture was stored in the refrigerator for 3 hours and then at room temperature for 1 hour. The reaction mixture was desalted with Glen Gel-pak 1.0 desalting column (cat # 61-5010).

4.4.4 DNA Templated CuAAC Ligation.

DNA-templated CuAAC was performed similarly to our previous work.[300] For triazole tracrRNA synthesis, a solution of 5' azide ONs (30 nmol), 3' alkyne ONs (15 nmol), and DNA Template-1 (15 nmol) was prepared to reach a total volume of 508 µL (0.2 M NaCl, 10 mM $MgCl_2$). For triazole sgRNA synthesis, 5' azide RNA (20 nmol), 5' azide – fluorophore – 3' alkyne ONs (15 nmol), 3' alkyne ONs (20 nmol), DNA Template-1 and -2 (15 nmol each) were dissolved in a salt solution (0.2 M NaCl, 10 mM $MgCl_2$ to reach a total volume of 508 µL. The ONs solution was heated at 80 °C for 5 min and cooled down to room temperature for 1 hour to facilitate strand annealing. The ONs solution was then flushed with argon for 2 min. Next $CuSO_4(H_2O)_5$ (24 µL, 18 mM) and tripotassium 5,5′,5″-[2,2′,2″-nitrilotris(methylene)tris(1H-benzimidazole-2,1-diyl)] tripentanoate hydrate ($BimC_4A)_3$, 36 µL, 18 mM) were combined in a small Eppendorf tube, flushed with argon for 2 min and left to sit for 3 min. Next, freshly prepared sodium ascorbate (24 µL, 340 mM) was added to the copper solution followed by flushing with argon for 2 min and left to sit for 3 min. The copper solution (68 µL) was added to the ONs solution, followed by argon flush for 2 minutes, then the reaction was incubated at 23 °C for 1 hour. At this point, the reaction was stopped with the addition of aqueous EDTA (100 µL, 0.5 M EDTA) and the solution was lyophilized. The lyophilized reaction was dissolved in 80 µL nuclease-free water and 40 µL of 0.5 M EDTA. The oligo solution was loaded with 7 µL on each lane on a 10% denaturing PAGE gel (0.75 mm, 10 wells per gel, 2 gels used for one reaction mixture). The gels were run at 200 – 250 V for 60 – 80 min. Afterward, the gels were visualized by UV or StainsAll (see below). The desired oligo band was cut, crushed, and soaked in nuclease-free water (800 µL) and stored in the refrigerator overnight. The oligo solution was filtered by using Spin-X centrifuge tube filter (costar

8160, 0.22 μm cellulose acetate) to separate from the crushed gel then desalted by Glen-pack cartridge following the Glen-pak desalting procedure.

4.4.5 Analysis by Polyacrylamide Gel Electrophoresis.

A 10% denaturing polyacrylamide gel (0.75 mm, 10 wells) was prepared. To analyze the CuAAC reactions, we used StainsAll to visualize each ONs band (Aldrich E9379). After electrophoretic separation, the gel was soaked in StainsAll solution (1:1 Formamide/water) (Formamide, Fisher F841) for 10 mins and visualized by white illumination via ImageQuant RT ECL instrument from GE Healthcare life science.

4.4.6 Cloning and Plasmid Construction.

Plasmid templates for *in vitro* cleavage assays were prepared through ligation of inserts into HindIII/XbaI double digested pUC19 (ThermoFisher). Single-stranded DNA oligonucleotides containing Cas9 targets were ordered from Integrated DNA Technologies (IDT). Forward and reverse ssDNA oligonucleotides were annealed in a thermocycler by heating to 95 °C for 5 min, then slowly cooled to 25 °C over the course of 1 hour prior to ligation.

4.4.7 In vitro Cleavage of DNA Substrates

In vitro cleavage assays were performed as previously described.[273] Briefly, primers pUC19_fwd and pUC19_rev were used to generate dsDNA substrates through PCR amplification of previously prepared plasmid templates, followed by purification with the QIAquick PCR

Purification Kit (Qiagen). DNA substrate concentration was measured by an Implen NanoPhotomoter NP80 (Implen). Prior to the addition of DNA substrate, Cas9 (New England Biolabs) and gRNA were incubated at a ratio of 1:1.5 in Cas9 cleavage buffer (100 mM NaCl, 50 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$, 100 μg mL-1 BSA) for 10 min at 25 °C to assemble complexes. For each reaction, 5 nM of DNA substrate was digested with 50 nM Cas9 RNP complex for 1 hour at 37 °C. Reactions were stopped by purifying the cleavage products using the MinElute PCR Purification Kit (Qiagen). The reaction was resolved on a 1.2 % agarose gel, followed by imaging on an Amersham Imager 600 (GE Healthcare).

4.4.8 Cationic Lipid Transfection of Stable Cell Lines

Cells stably expressing Cas9 were transfected using Lipofectamine RNAiMAX with sgRNAs according to the manufacturer's instructions to a final concentration of 30 nM.[250,251]

4.4.9 Cationic Lipid Delivery of Cas9 RNP Complexes

Cas9 (New England Biolabs) and synthesized gRNAs were assembled into ribonucleoprotein (RNP) complexes and transfected into 293T using Lipofectamine CRISPRMAX according to the manufacturer's instructions (Invitrogen).

4.4.10 Cellular Cleavage Assays.

48 hours after transfection, genomic DNA (gDNA) was isolated from transfected cells using the DNeasy Blood & Tissue Kit (Qiagen) and quantified using a NanoPhotometer NP80

(Implen). Locus-specific primers targeting *EMX1* and *WAS*, as well as 100 ng of gDNA were used to PCR amplify the target of interest, followed by purification with the QIAquick PCR Purification Kit (Qiagen). T7 endonuclease I (T7E1) digestion was performed using the purified PCR product as described in the manufacturer's protocol (New England Biolabs). Cleavage products were resolved on a 1.2 % TAE agarose gel. For experiments involving the enrichment of fluorescent-positive cells, cells were sorted on a BD FACSAria III instrument 24 hours following transfection by the Flow Cytometry Core at the University of Alberta. PCR amplification and T7E1 digestion was performed as described above.[250,251]

4.4.11 Cell Transfection and Flow Cytometry

Custom crRNA (Integrated DNA Technologies; IDT) was designed to target human *Siglec-3* (*Siglec-3* sequence = 5' CGG TGC TCA TAA TCA CCC CA 3'). U937 cells were seeded at 500,000 cells per well the day of transfection in a 6-well tissue culture plate, in 1.3 mL growth media (RPMI containing 10% FBS and 100 U/mL penicillin-streptomycin). For one well of a 6-well plate, 20 pmol of Cas9 nuclease (IDT), 20 pmol of ATTO-550 labeled crRNA:tracrRNA (IDT) duplex or triazole-modified sgRNAs, 8 µL of Cas9 Plus reagent (Thermo Fisher), and 16 µL of CRISPRMAX reagent (Thermo Fisher) in 665 µL Opti-MEM medium (Gibco), were incubated together for 15 minutes, and then added to the cells. One day after transfection, cells were removed from the plate, washed, resuspended in 300 µL of cell sorting media (PBS, 1% FBS, 1 mM EDTA), and stored on ice until sorting. Cells were sorted within the University of Alberta Flow Cytometry Core. 100K of the brightest 10-30% of cells stained with ATTO-550 (IDT CD33), fluorescein, and TAMRA were sorted into a 15 mL tube containing regular culture media, and then transferred to a 12 well plate. Cells were grown for 5 days and then analyzed by flow cytometry. Cells were

incubated with Human TruStain FcX™ Fc blocker (1:200, Biolegend) for 10 minutes at room temperature. They were then incubated with either anti-siglec-3-BV711 or anti-siglec-7-PE antibodies (1:200, Biolegend) for 30 minutes on ice. Cells were washed twice, resuspended in a flow buffer (HBSS containing 0.1% EDTA and 0.1% BSA), and analyzed by flow cytometry. Unstained wild-type U937 cells were used as a control (red graph).

4.4.12 MALDI TOF MS Characterization.

The synthesized oligonucleotide (1-3 nmol) was dissolved in 10 μL of triethylammonium acetate (0.1M TEAA, pH 7.0) and desalted using Zip tip (Millipore, ZTC18S096). The eluted oligonucleotide was lyophilized to dryness. 3-hydroxypicolinic acid (25 mg/mL) was dissolved in acetonitrile/water (1:1 v/v). Ammonium citrate (25 mg/mL) was dissolved in water. 3-hydroxypicolinic acid solution was mixed with ammonium citrate solution (9:1 v/v). The dried oligonucleotide was dissolved in 2.5 μL matrix solution and 2.5 μL nuclease-free water. The oligonucleotides were analyzed on a Voyager Elite (Applied BioSystems, Foster City, CA, USA) with linear negative mode.

4.4.13 Mass Spectrometry (Electrospray Ionization)

Electrospray ionization-mass spectrometry (ESI-MS) measurements were performed on a 15T solariX-XR Fourier-transform ion cyclotron resonance (FT-ICR) mass spectrometer (Bruker-Daltonics, Billerica, MA), which was equipped with a nanoflow ESI (nanoESI) source. NanoESI tips with ~5 μm outer diameters (o.d.) were produced from borosilicate capillaries (1.0 mm o.d., 0.68 mm inner diameter) using a P-1000 micropipette puller (Sutter Instruments, Novato, CA). To

perform nanoESI, a voltage of approximately +1.0 kV (positive mode) was applied to a platinum wire inserted inside the nanoESI tip and in contact with the solution. Ions/droplets were sampled into a vacuum using a glass capillary (exit voltage 220 V) and a source drying gas (N_2, 4 L/min, and 120 °C) to aid with desolvation. The ions were steered by a deflector (200 V) into the funnel (150 V)–skimmer (60 V) stage, and accumulated in a hexapole for 0.5 s. The ions were transferred through the quadrupole and then injected into the ParaCell ion cell for detection (sweep excitation power 45%, pressure ~5×10^{-10} mbar). Default values were applied for other instrumental parameters. Data acquisition was performed using the FtmsControl software (version 2.2). The time-domain signal, consisting of the sum of ~200 transients, containing 2M data points per transient, was subjected to one zero-fill prior to Fourier-transformation. Stock solutions were prepared from lyophilized DNA/RNA samples, dissolved in milliQ water, and stored at -20 ºC. The nanoESI-MS measurements were performed on the DNA/RNA samples (10 μM) in 200 mM aqueous ammonium acetate (pH 7).

DNA oligonucleotides are in the lower case and RNA oligonucleotides are in the upper case. Modifications are shown in bold, and the structures are shown at the bottom of the table. Unless otherwise noted, oligonucleotides were analyzed by MALDI-TOF

Table 4.1 Oligonucleotides for clicked tracrRNA and sgRNA.

Modification Positions	Oligo code	Sequence (5' → 3')	Molecular Weight	
			Calculated	**Found**
C70	5' azide	**N_3-t**UG AAA AAG UGG CAC CGA GUC GGU GCU	**8734**	8737
	3' alkyne + dT	tAU AGC AAG UUA AAA UAA GGC UAG UCC GUU AUC AA**c*-alkyne**	**11542**	11541
	5' azide + 3' alkyne	**N_3-t**AG AAA UAG CAA GUU AAA AUA AGG CUA GUC CGU UAU CAA **c*-alkyne**	**12901**	12901
C55	5' azide	**N_3-t**AG UCC GUU AUC AAC UUG AAA AAG UGG CAC CGA GUC GGU GCU	**13493**	13493
	3' alkyne + dT	tAU AGC AAG UUA AAA UAA GG**c*-alkyne**	**6784**	6785
	5' azide + 3' alkyne	**N_3-t**AG AAA UAG CAA GUU AAA AUA AGG **c*-alkyne**	**8142**	8142
C60	5' azide	**N_3-c**GU UAU CAA CUU GAA AAA GUG GCA CCG AGU CGG UGC U	**11887**	11886
	3' alkyne + dT	tAU AGC AAG UUA AAA UAA GGC UAG U**c*-alkyne**	**8376**	8378
	ATTO 550	**N_3-t**AG AA**(ATTO 550-t)** AUA GCA AGU UAA AAU AAG GCU AGU CCG UUA UCA A**c*-alkyne**	**13935**	13937 ± 1*
	Fluorescein	**N_3-t**AG AA**(Fluorescein-t)** AUA GCA AGU UAA AAU AAG GCU AGU CCG UUA UCA A**c*-alkyne**	**13716**	13719 ± 1*
	Alexa 594	**N_3-t**AG AA**(Alexa 594-t)** AUA GCA AGU UAA AAU AAG GCU AGU CCG UUA UCA A**c*-alkyne**	**14078**	14066 ± 1*
	Alexa 405	**N_3-t**AG AA**(Alexa 405-t)** AUA GCA AGU UAA AAU AAG GCU AGU CCG UUA UCA A**c*-alkyne**	**13966**	13969 ± 1*

c*-alkyne = 3'-propargyl-5-Me-dC

* Analyzed by electrospray ionization

Table 4.2 CRISPR target sequences.

Target	Sequence (5' → 3')	Molecular Weight	
		Calculated	**Found**
Siglec-3 (v1)	g AA CAC CCC CGA UCU UCU CCG UUU UAG AG***c - alkyne**	**9464**	9463
Siglec-5 (v1)	**c** AU GGC UAC UGG UUC CGG GCG UUU UAG AG***c - alkyne**	**9617**	9617
Siglec-7 (v1)	**g** AG AGG TGG TCC CGC TTC AC G UUU UAG AG***c - alkyne**	**9640**	9639
Siglec-9 (v1)	g AG AUG GAC GGU CUU GUU CG G UUU UAG AG***c-alkyne**	**9682**	9683
CMAS	c CC CGG GAC GGU CGC CCC CG G UUU UAG AG***c-alkyne**	**9589**	9589
Siglec-3	c GG UGC UCA UAA UCA CCC CAG UUU UAG AG***c - alkyne**	**9528**	9527
Siglec-5	g AG UUC CGU GAC CGU GCA AGG UUU UAG AG***c- alkyne**	**9664**	9668
Siglec-7	t CG ACG AAG CUC CAA GAU CCG UUU UAG AGC***c- alkyne**	**9605**	9604
Siglec-9	a AC CAG UAG CCA UGA ACU ACG UUU UAG AG***c- alkyne**	**9598**	9602
EMX 1	g AG UCC GAG CAG AAG AAG AAG UUU UAG AG **c*-alkyne**	**9781**	9784
WAS	t GG AUG GAG GAA UGA GGA GUG UUU UAG AG **c*-alkyne**	**9846**	9848

c*-alkyne = 3'-propargyl-5-Me-dC

v1 = version 1.

Table 4.3 DNA templates sequences.

Complementary Position	Sequence (5' → 3')	Molecular Weight	
		Calculated	**Found**
G62 – C85 (24 nt)	gtg cca ctt ttt caa gtt gat aac	**7318**	7318
A51 – C91 (40 nt)	act cgg tgc cac ttt ttc aag ttg ata acg gac tag cct t	**12237**	12240
U44 – C67 (24nt)	gat aac gga cta gcc tta ttt taa	**7351**	7353
G39 – U72 (34nt)	aag ttg ata acg gac tag cct tat ttt aac ttg c	**10431**	10430
G21 - U50 (30 nt)	att tta act tgc tat ttc tag ctc taa aac	**9105**	9106
G21–A– U50 (31nt)	att tta act tgc tat att cta gct cta aaa c	**9418**	9419

Table 4.4 Triazole modified tracrRNAs.

tracrRNA Product	Sequence (5' → 3')	Molecular Weight	
		Calculated	**Found**
C70 triazole modified	t AUA GCA AGU UAA AAU AAG GCU AGU CCG UUA UCA A **c*(triazole)t** UG AAA AAG UGG CAC CGA GUC GGU GCU	**20277**	20280 ± 1*
C55 triazole modified	t AUA GCA AGU UAA AAU AAG G **c*(triazole)t** AGU CCG UUA UCA ACU UGA AAA AGU GGC ACC GAG UCG GUG CU	**20277**	20280 ± 1*
C60 triazole modified	t AUA GCA AGU UAA AAU AAG GCU AGU **c*(triazole)c** GUU AUC AAC UUG AAA AAG UGG CAC CGA GUC GGU GCU	**20263**	20265 ± 1*

c*-alkyne = 3'-propargyl-5-Me-dC

* Analyzed by electrospray ionization

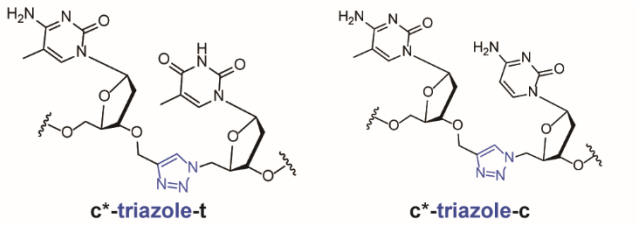


Table 4.5 Triazole modified sgRNAs.

sgRNA Product	Sequence (5' → 3')	Molecular Weight	
		Calculated	**Found**
C30 C70 modified sgRNA (Target: *Siglec-3*)	g AA CAC CCC CGA UCU UCU CCG UUU UAG AG **c*(triazole)t** AG AAA UAG CAA GUU AAA AUA AGG CUA GUC CGU UAU CAA **c*(triazole)t** UGA AAA AGU GGC ACC GAG UCG GUG CU	**31099**	30808
C30 C55 modified sgRNA (Target: *Siglec-3*)	g AA CAC CCC CGA UCU UCU CCG UUU UAG AG **c*(triazole)t** AG AAA UAG CAA GUU AAA AUA AGG CUA GUC CGU UAU CAA **c*(triazole)t** UGA AAA AGU GGC ACC GAG UCG GUG CU	**31099**	31233 ± 1*

c*-alkyne = 3'-propargyl-5-Me-dC
* Analyzed by electrospray ionization

Table 4.6 Triazole modified fluorescent-labeled sgRNAs.

sgRNA Product	Sequence (5' → 3')	Molecular Weight	
		Calculated	**Found**
Fluorescein labeled sgRNA (Target: *Siglec-3)* (version.1)	g AA CAC CCC CGA UCU UCU CCG UUU UAG AG**c*(triazole)t** AG AA**(Fluorescein-t)** AUA GCA AGU UAA AAU AAG GCU AGU CCG UUA UCA A**c*(triazole)t** UGA AAA AGU GGC ACC GAG UCG GUG CU	**31914**	31954*

TAMRA labeled sgRNA (Target: *Siglec-7*) (version.1)	c AU GGC UAC UGG UUC CGG GCG UUU UAG AG**c*(triazole)t** AG AA**(TAMRA-t)** AUA GCA AGU UAA AAU AAG GCU AGU CCG UUA UCA A**c*(triazole)t** UGA AAA AGU GGC ACC GAG UCG GUG CU	**32122**	32135*
Cy5 labeled sgRNA (Target: *Siglec-5*) (version.1)	g AG AGG UGG UCC CGC UUC ACG UUU UAG AG**c*(triazole)t** AG AA**(Cy5-T)** AUA GCA AGU UAA AAU AAG GCU AGU CCG UUA UCA A**c*(triazole)t** UGA AAA AGU GGC ACC GAG UCG GUG CU	**32198**	33177*
Alexa 594 labeled sgRNA (Target: *Siglec-9*) (version.1)	g AG AUG GAC GGU CUU GUU CG G UUU UAG AG**c*(triazole)t** AG AA**(Alexa 594-T)** AUA GCA AGU UAA AAU AAG GCU AGU CCG UUA UCA A**c*(triazole)t** UGA AAA AGU GGC ACC GAG UCG GUG CU	**32494**	32600*
Alex 405 labeled sgRNA (Target: *CMAS*)	c CC CGG GAC GGU CGC CCC CG G UUU UAG AG**c*(triazole)t** AG AA**(Alexa 405-t)** AUA GCA AGU UAA AAU AAG GCU AGU CCG UUA UCA A**c*(triazole)t** UGA AAA AGU GGC ACC GAG UCG GUG CU	**32289**	32467*
Fluorescein labeled sgRNA (Target: *Siglec-3*)	c GG UGC UCA UAA UCA CCC CAG UUU UAG AG**c*(triazole)t** AG AA**(Fluorescein-t)** AUA GCA AGU UAA AAU AAG GCU AGU CCG UUA UCA A**c*(triazole)t** UGA AAA AGU GGC ACC GAG UCG GUG CU	**31978**	31978 ± 2*
TAMRA labeled sgRNA (Target: *Siglec-7*)	g AG UUC CGU GAC CGU GCA AGG UUU UAG AG**c*(triazole)t** AG AA**(TAMRA-t)**AUA GCA AGU UAA AAU AAG GCU AGU CCG UUA UCA A**c*(triazole)t** UGA AAA AGU GGC ACC GAG UCG GUG CU	**32169**	32168 ± 2*

Cy5 labeled sgRNA (Target: *Siglec-5*)	t CG ACG AAG CUC CAA GAU CCG UUU UAG AG**c*(triazole)t** AG AA**(Cy5-t)** AUA GCA AGU UAA AAU AAG GCU AGU CCG UUA UCA A**c*(triazole)t** UGA AAA AGU GGC ACC GAG UCG GUG CU	**32163**	32163 ± 2*
Alexa 594 labeled sgRNA (Target: *Siglec-9*)	a AC CAG UAG CCA UGA ACU ACG UUU UAG AG**c*(triazole)t** AG AA**(Alexa 594-t)** AUA GCA AGU UAA AAU AAG GCU AGU CCG UUA UCA A**c*(triazole)t** UGA AAA AGU GGC ACC GAG UCG GUG CU	**32411**	32413 ± 2*
ATTO 550 labeled sgRNA (Target: *Siglec-3*)	c GG UGC UCA UAA UCA CCC CAG UUU UAG AG**c*(triazole)t** AG AA**(ATTO 550-t)** AUA GCA AGU UAA AAU AAG GCU AGU CCG UUA UCA A**c*(triazole)t** UGA AAA AGU GGC ACC GAG UCG GUG CU	**32197**	32200 ± 3*
ATTO 550 labeled sgRNA (Target: *WAS*)	t GG AUG GAG GAA UGA GGA GUG UUU UAG AG**c*(triazole)t** AG AA**(ATTO 550-t)**AUA GCA AGU UAA AAU AAG GCU AGU CCG UUA UCA A**c*(triazole)t** UGA AAA AGU GGC ACC GAG UCG GUG CU	**32516**	32517±2*
ATTO 550 labeled sgRNA (Target: *EMX 1*)	g AG UCC GAG CAG AAG AAG AAG UUU UAG AG **c*(triazole)t** AG AA**(ATTO 550-t)** AUA GCA AGU UAA AAU AAG GCU AGU CCG UUA UCA A**c*(triazole)t** UGA AAA AGU GGC ACC GAG UCG GUG CU	**32450**	32610 ± 3*
Fluorescein labeled sgRNA (Target: *WAS*)	t GG AUG GAG GAA UGA GGA GUG UUU UAG AG **c*(triazole)t** AG AA**(Fluorescein-t)** AUA GCA AGU UAA AAU AAG GCU AGU CCG UUA UCA A**c*(triazole)t** UGA AAA AGU GGC ACC GAG UCG GUG CU	**32297**	32297 ± 2*

c*-alkyne = 3'-propargyl-5-Me-dC

* Analyzed by electrospray ionization

Table 4.7 Primer sequences for PCR.

Primers	Sequence (5' → 3')
pUC19_fwd	GCG ACA CGG AAA TGT TGA ATA CTC AT
pUC19_rev	CAG CGA GTC AGT GAG CGA
EMX1_fwd	CCT CAG TCT TCC CAT CAG GC
EMX1_rev	GTG GGG CCA TGA CTC CA
WAS_fwd	GTG GCA GGG CTG TGA TAA CT
WAS_rev	TGC TTT ATC ATT CAC TGG CTC A

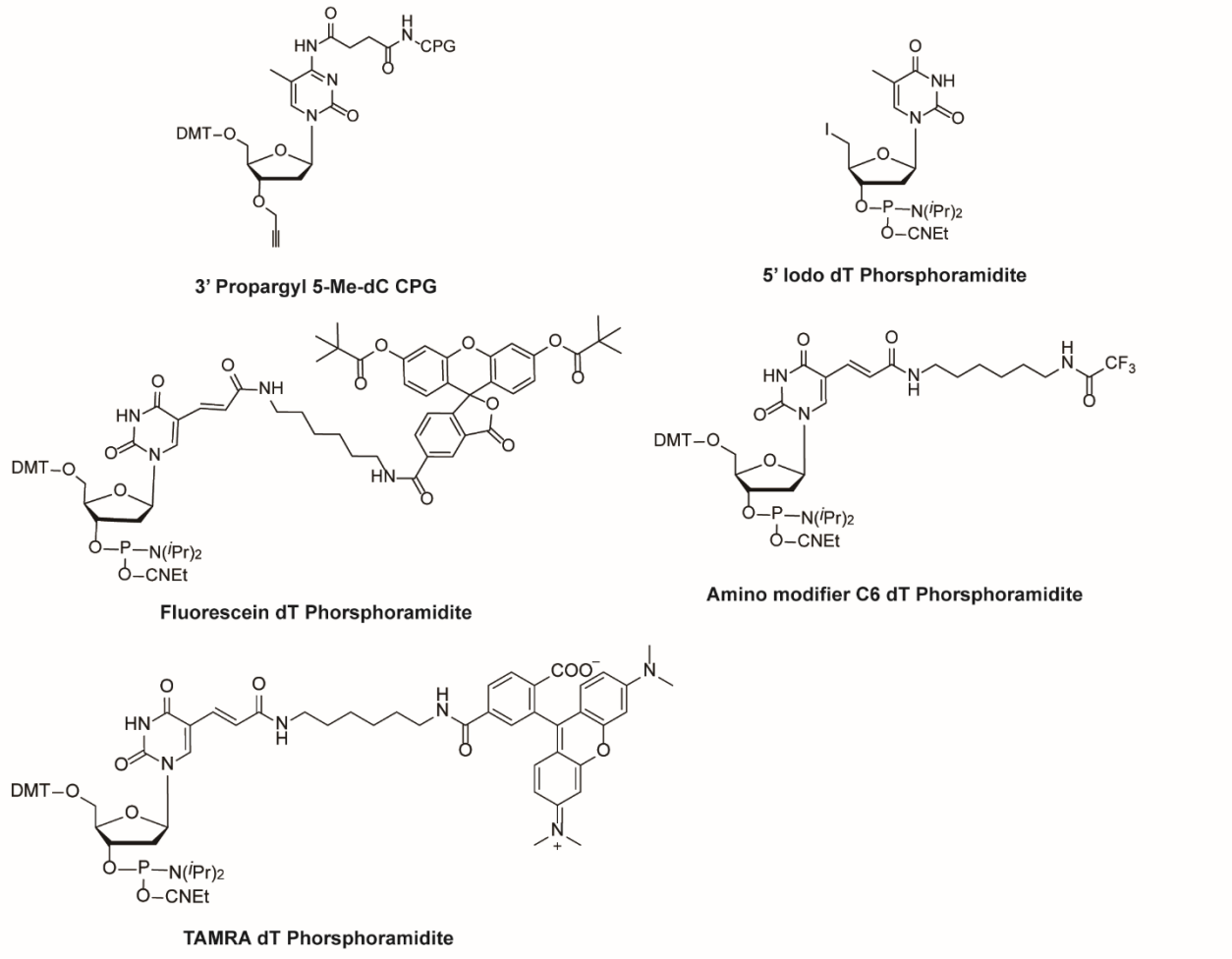


Figure 4.12 Modified nucleoside phosphoramidites and support structures.

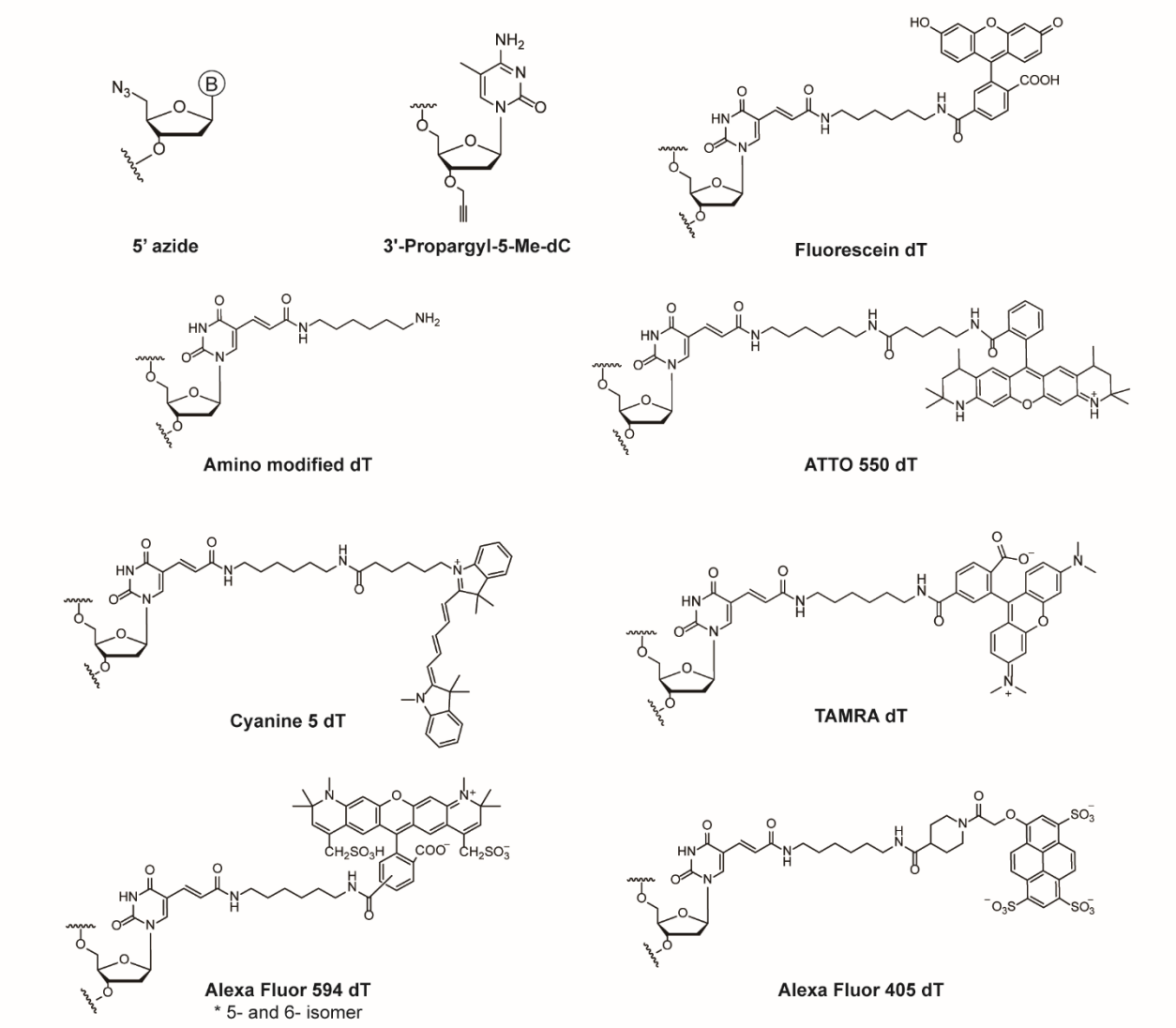


Figure 4.13 Structures of modified nucleotides.

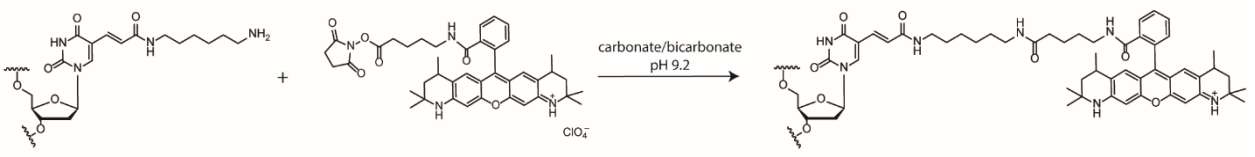


Figure 4.14 Representative Fluorophore (ATTO 550) NHS ester conjugation with NH2-dT modified RNA strand.

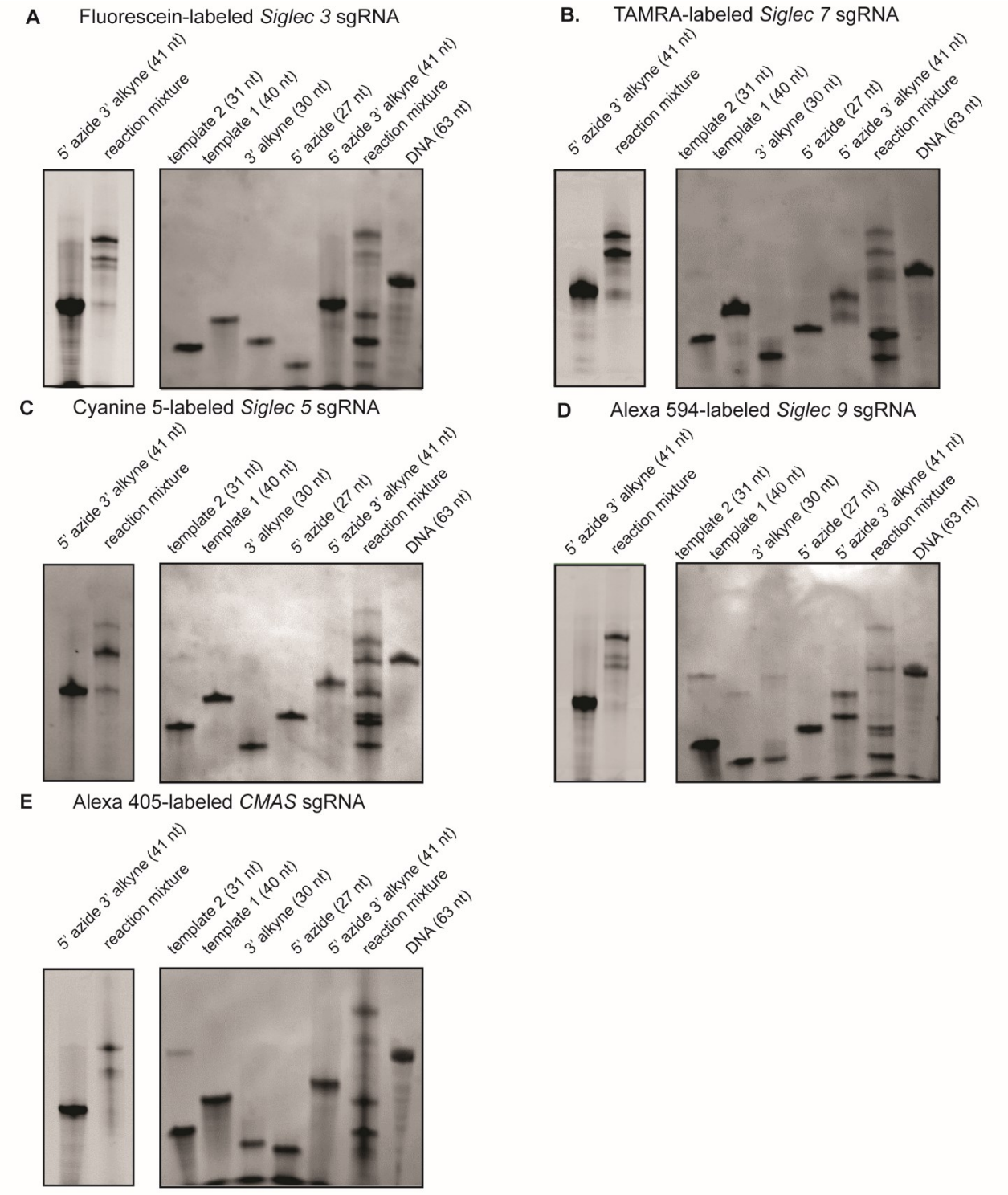


Figure 4.15 PAGE gel images of five different fluorophores with Siglec targets and CMAS sequences. (**A**) Fluorescein-labeled *Sigle-3* sgRNA. (**B**) TAMRA-labeled *Siglec-7* sgRNA. (**C**) Cyanine 5-labeled *Siglec-5* sgRNA. (**D**) Alexa 594-labeled *Siglec-9* sgRNA. and (**E**) Alexa 405-labeled *CMAS* sgRNA.Filtered images to detect fluorescence-labeled strands (left) and StainsAll images of reactants and reaction mixture of fluorescence-labeled sgRNAs. Filter (520 nm) was

used for fluorescein, Alexa 594 and Alexa 405. Filter (595 nm) was used to detect TAMRA and ATTO 550. Filter (630 nm) was used to detect Cyanine 5.

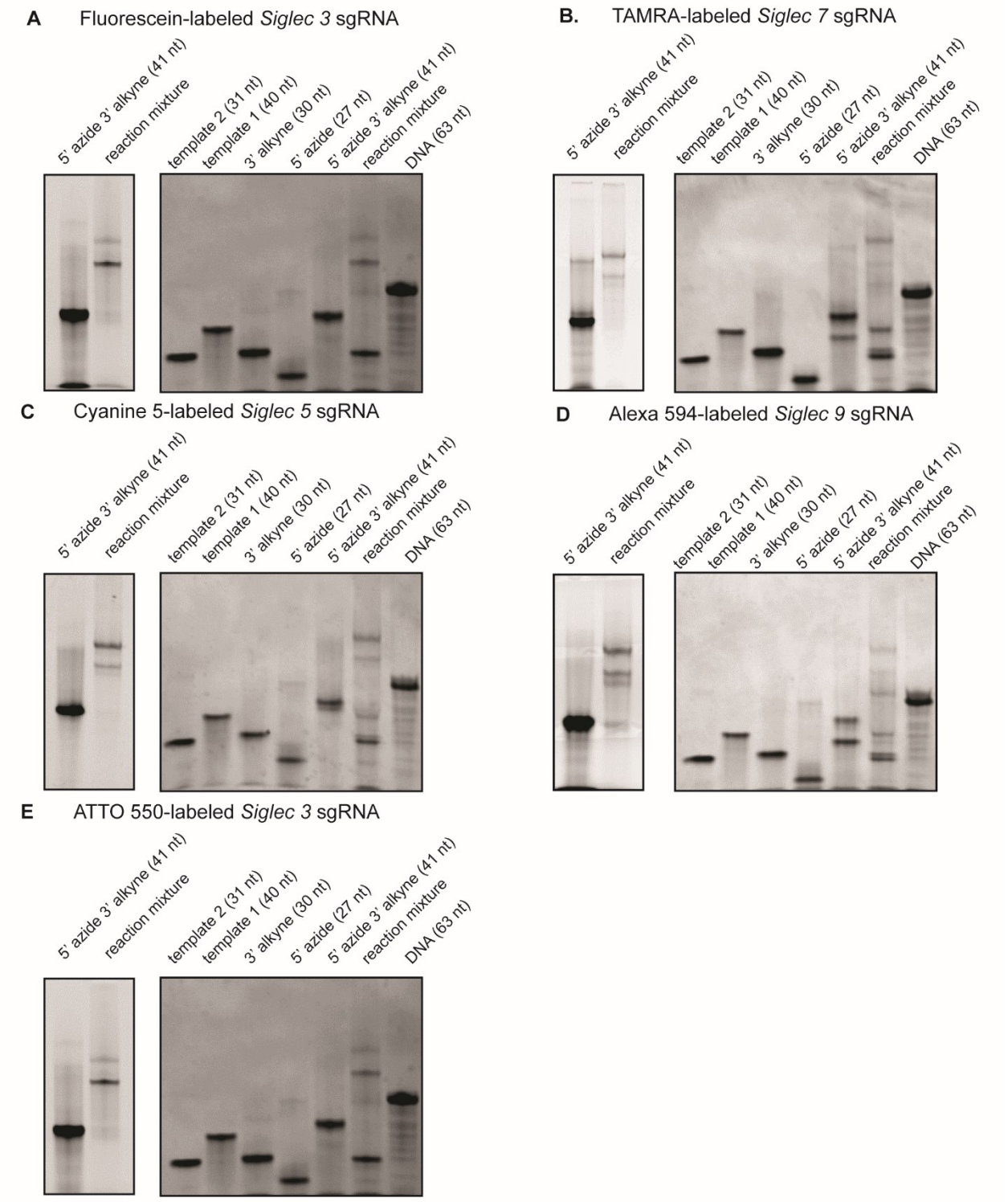


Figure 4.16 PAGE gel images of five different fluorophores with Siglec targets. (**A**) Fluorescein-labeled *Sigle-3* sgRNA. (**B**) TAMRA-labeled *Siglec-7* sgRNA. (**C**) Cyanine 5-labeled

Siglec-5 sgRNA. (**D**) Alexa 594-labeled *Siglec-9* sgRNA. and (**E**) ATTO 550-labeled *Siglec-3* sgRNA. Filtered images to detect fluorescence-labeled strands (left) and StainsAll images of reactants and reaction mixture of fluorescence-labeled sgRNAs. Filter (520 nm) was used for fluorescein, and Alexa 594. Filter (595 nm) was used to detect TAMRA and ATTO 550. Filter (630 nm) was used to detect Cyanine 5.

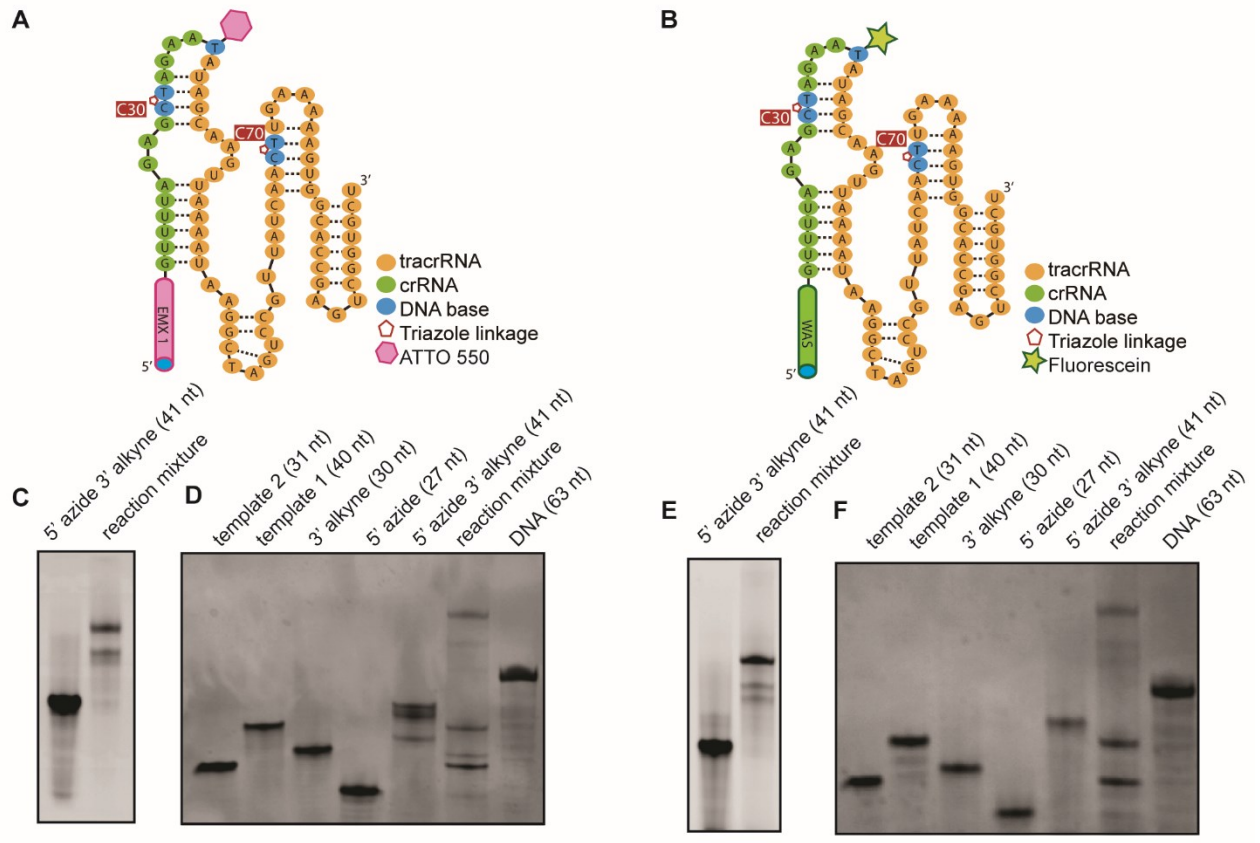


Figure 4.17 PAGE gel image of triazole modified sgRNA. (**A**) ATTO 550-labeled *EMX1* sgRNA and (**B**) fluorescein-labeled *WAS* sgRNA. (**C**) Filter (595 nm) of the imager to detect ATTO-550 labeled strands (**D**) StainsAll images of reactants and reaction mixture of ATTO 550-labelled *EMX1* sgRNA (**E**) Filter (520 nm) of the imager to detect fluorescein-labeled strands (**F**) StainsAll image of reactants and reaction mixture of fluorescein-labeled *WAS* sgRNA.

Chapter 5

Conclusion and Future Outlook

5.1 Conclusion

Since Watson, Crick, and Franklin's groundwork on understanding the structures of DNA, DNA technology has been studied widely and represents critical research in biotechnology. A common method in biotechnology is utilizing chemical syntheses and reactions with oligonucleotides, such as solid-phase synthesis, bioconjugation, and chemical modifications. This thesis conducted comprehensive research to understand oligonucleotides' role in self-replication, detection tools, and gene-editing technologies.

Finding a critical step that enables oligonucleotide self-replication is essential to study the origin of life. Our isothermal ligase chain reaction, lesion-induced DNA amplification (LIDA), is an excellent example of rapid and successful replication. While previous reports focused on finding optimal conditions for isothermal DNA amplification as a detection tool, the fundamental studies of this method have not been conducted yet.

Chapter 2 aimed to understand the kinetic behavior of lesion-induced DNA amplification by thermodynamic analysis and a kinetic fitting model. The melting profiles of DNA complexes prove that the abasic group decreased the melting temperature of DNA duplex by about 12 ℃ compared to the complementary thymidine. This lower melting temperature resulted in reduced stability of DNA duplex by five orders of magnitude at 30 ℃, the optimal self-replication temperature for this sequence system (3.6×10^{12} M^{-1} versus 8×10^{7} M^{-1}). Similarly, the isothermal calorimetry titration measured four orders of magnitude difference in stability in the presence of a destabilizing abasic group in the duplex (1.98×10^{12} M^{-1} versus 1×10^{8} M^{-1}). It was confirmed that the abasic group reduces the stability gaps between the product duplex and intermediate complexes 10^{8} M^{-1} to 10^{3}-10^{5} M^{-1} while the complementary thymidine showed the more significant difference 10^{12} M^{-1} to 10^{4}-10^{5} M^{-1}. Although the destabilizing group brought the stabilities much closer

together, the K_A differed by four orders of magnitude, which was expected to exhibit product inhibition. Kinetic modeling by KinTek Explorer fitted the observed *in situ* formation of both template and product strands in the isothermal ligase chain reaction. The fitting data showed that in the presence of T4 DNA ligase, the stability difference between the intermediate complex and the product duplex was much more reduced ($1.05 \pm 0.05 \times 10^5$ M^{-1} versus $9 \pm 7 \times 10^5$ M^{-1}), leading to continuous dissociation of product strands from the template:product duplex. Simulations on our kinetic fits indicated the stability of the intermediate complexes plays an important role in initiating self-replication. Moreover, the influence of the abasic group was presented as Michaelis-Menten parameters compared to the thymidine. The result suggests that k_{cat} was slower for the 5'-phosphate abasic complex (0.07 ± 0.02 s^{-1}) than that of the complementary bases (0.3 ± 0.1 s^{-1}), although they have similar enzyme affinity and inhibition behavior. Lastly, the background triggered reaction was explored, resulting in the pseudo-blunt ligation only occurring in one cycle with 5'-abasic overhang. In summary, the dual role of T4 DNA ligase was demonstrated by the experimental results from the enzyme concentration variation and the temperature effects on LIDA. It was revealed that T4 DNA ligase not only facilitates the fast ligation of the ternary complex but also helps to reduce the stability gaps between the DNA complexes, leading to overcoming product inhibition. This work gives insight into the design of nonenzymatic oligonucleotides self-replication by determining what is needed to achieve fast and efficient self-replication of oligonucleotides.

In Chapter 3, the pyrene nucleotide was studied as the selective ligation tool for the abasic site. T4 DNA ligase and PBCV-1 DNA ligase enabled the ligation of the 5'-phosphate pyrene nucleotide at the ligation junction selectively to the abasic-containing template. The ligation reaction was performed using a pyrene containing template in the comparison of 5'-phosphate

abasic strands and 5'-phosphate deoxyadenosine strands. The PBCV-1 DNA ligase selectively ligated the 5'-abasic strand with the pyrene template, while the T4 DNA ligase ligated them non-selectively. Additionally, the 5'-phosphate abasic nucleotide was tested for its selectivity on the pyrene-containing template versus the thymidine-containing template. Although increased temperature enables higher ligation yield on the pyrene template over the natural T template, the selectivity of the 5'-phosphate abasic strand was not obtained. The selectivity of 5'-phosphate pyrene strands was demonstrated with the template containing either the abasic group or the natural bases (A, T, G, and C). Under optimized conditions, the ligases only allowed ligation with the pyrene:abasic pair. Moreover, the competitive ligation was performed with 5'-phosphate pyrene strands and 5'-phosphate thymidine strands. This experiment demonstrated that the pyrene strands were selectively reacted with the abasic templates even the competitive strands existed. The incorporation of 5'-phosphate pyrene strand to selectively ligate to the abasic template can be utilized to detect DNA damage sites by adding a pre-step using DNA glycosylase, which generates the abasic site from the damaged bases. Furthermore, the bulky pyrene group was tested as the destabilizing group in isothermal DNA chain reaction (LIDA). The pyrene-containing DNA system demonstrated the sigmoidal shape of the self-replication reaction at the optimized condition, indicating that the pyrene group is an effective candidate for the destabilizing group to induce turnover.

Chapter 4 aims to develop the modular and fast synthesis of chemically modified sgRNA. The strategy was to chemically ligate smaller sgRNA fragments to produce a full length of sgRNA. Chemical ligation was chosen over enzymatic ligation due to the tolerance of the modification, low cost, and scalability of the reaction. Copper-catalyzed alkyne-azide cycloaddition (CuAAC) was selected because of previously demonstrated compatibility between the triazole linker and the polymerase, geometrical similarity with a phosphate group, and the hydrogen-bonding ability. First,

the triazole linker was explored in three different locations in tracrRNA and exhibited excellent activity with Cas9. Interestingly, the triazole linker, when buried in Cas9, also showed comparable activity, suggesting that triazole linker was indeed tolerated by the enzyme as a substitution candidate for the phosphate backbone. Next, the double click reaction with three smaller RNA:DNA fragments was performed to generate sgRNA. The incorporation of amine-modified thymidine in the middle fragment enabled the conjugation of fluorophores. By simply changing the crRNA sequence depending on the target, multiple sgRNA strands with different fluorophores were generated. ATTO 550-labeled *EMX1* gRNA and fluorescein-labeled *WAS* sgRNA were tested in the *in-vitro* assay for cleavage activity of Cas9. Moreover, these sgRNAs were successfully transfected in Cas9-expressing HeLa cells and sorted based on their fluorophore emission. After cell sorting, the activity of these sgRNAs in cells was quantified by measuring the cleavage of the target site using the T7 endonuclease I assay. The modular bioconjugation of sgRNA from the assembly of smaller RNA strands was achieved with the addition of fluorophores and other chemical modifications.

5.2 Future Work

5.2.1 Exploration of Different Types of Destabilizing Group in LIDA

Chapter 2 discusses the thermodynamic parameters of the destabilizing abasic group in the LIDA system. Chapter 3 showed those parameters in the destabilizing pyrene group and a native DNA system in LIDA. Based on chapter 2, T4 DNA ligase helps to overcome product inhibition by reducing the stability gap between the intermediate complex and the product duplex. T4 DNA ligase also can tolerate some modifications on the ligation site, such as the abasic group and the

pyrene nucleotides. Taking advantage of this tolerance, exploring other types of destabilizing groups in the LIDA system would be interesting and helpful to understand the characteristics of LIDA. For example, the destabilizing groups that were shown in the previous study would be a great place to start.[143] The thermal dissociation curves including ethyl, butyl, *cis*-butenyl, and xylyl group in intermediate complex and product duplex were compared with that of complementary thymidine base along with the abasic group. It showed that these destabilizing groups decreased the melting temperature by about 12 ℃ in product duplex compared to thymidine. The destabilizing templates were capable of turn-over reaction initiated by 0.01 equivalent template with fragments. These experiments suggest that those destabilizing groups have the potential to be applied in the LIDA system as 5'-phosphate end strands (**Figure 5.1**). Synthesizing the destabilizing group followed by phosphorylation can generate 5'-phosphate modification strands. T4 DNA ligase may tolerate these modifications and ligate the strands. Once the ligation is successfully achieved, the cross-catalytic reaction can be conducted to analyze how much T4 DNA ligase can reduce the stability gaps between the DNA complexes with these modifications. Moreover, the thermodynamic parameters of the destabilizing product duplexes, and their intermediate complexes can be obtained by either measuring the melting temperatures or using isothermal calorimetry techniques.

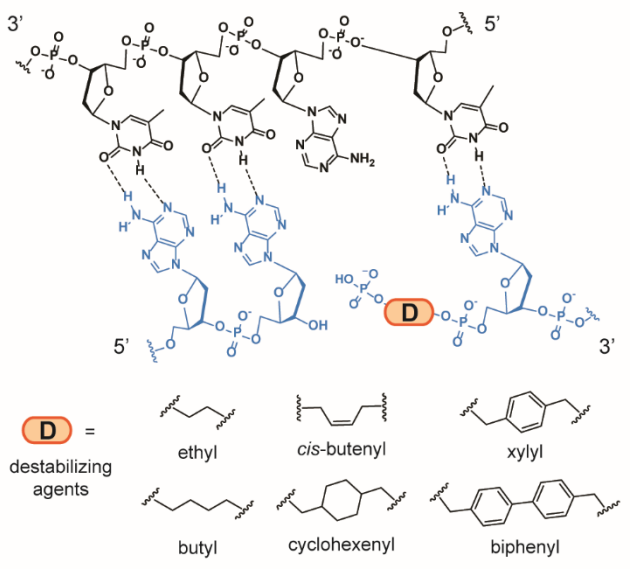


Figure 5.1 Destabilizing agent candidates for cross-catalytic reaction. The destabilizing reagents are located at 5'-phosphate nicked site, in place of thymidine.

5.2.2 Detection of DNA Damages

There are various ways to detect DNA damage sites. For example, Sturla and coworkers used non-natural bases (ExBenzi and ExBIM) that form stable base pairs with O^6-methylguanine.[304] The oligonucleotides containing the unnatural bases were conjugated to gold-nanoparticles and the hybridization with oligonucleotides that contain O^6-methylguanine sites induced the colorimetric changes of the nanoparticle system for the quantification of O^6-MeG.[304] The Burrow group utilized uracil-DNA glycosylase (UDG) to find the damaged sites and AP endonuclease to generate the nick sites.[222] Then unnatural base pairs (dNaM/d5SICS) were inserted opposite to natural bases by the polymerase. The nicked site was sealed by T4 DNA ligase with the addition of dimethylsulphoxide, and then PCR was conducted to amplify the target strands.

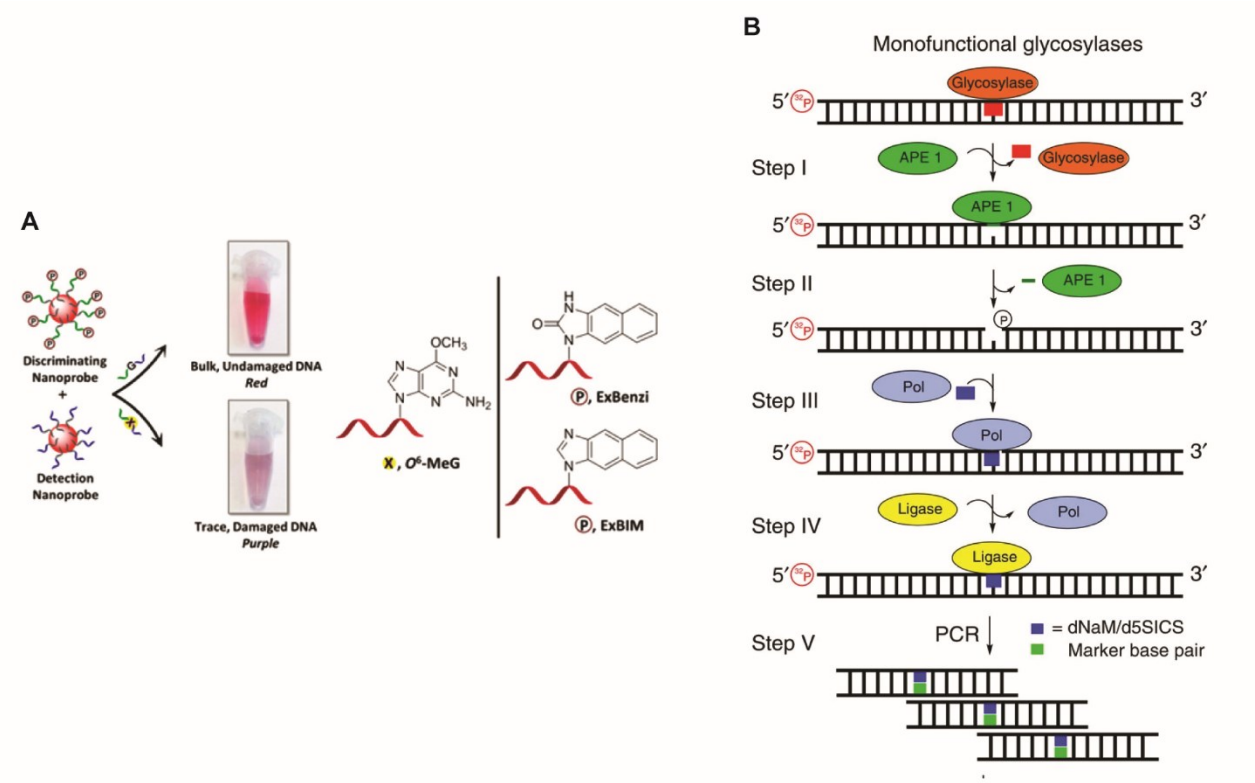


Figure 5.3 Examples of DNA damage detection. (**A**) Schematic illustration of the gold nanoparticle colorimetric detection of the mutation O^6-MeG.[304] Reprinted with permission from *J. Am. Chem. Soc.* **2016**, *138*, 27, 8497-8504. Copyright © 2016, American Chemical Society. (**B**) Scheme of labeling lesions using monofunctional base excision repair enzyme.[222] Reprinted with permission from *Nat Commun* **2015**, *6,* 8807 Copyright © 2015, Springer Nature.

As Kool and coworkers showed that pyrene nucleoside triphosphate can be inserted opposite to abasic sites by DNA polymerase,[215] it may be possible to design the amplification method by incorporating polymerase after the selective ligation of the pyrene nucleotide. However, the problem may be encountered that the canonical base is incorporated at the opposite side of the pyrene base non-selectively. Based on measurements by Kool and coworkers of the thermal dissociation temperature of pyrene:canocial base pairs, the adenine nucleobase showed a higher melting temperature (38.7 ℃) than other bases (36.4 ℃ with T, 37.6 ℃ with C and 38.2 ℃ with G).[216] Another method can be a ligation detection reaction that determines directly the amplification of the ligated product (**Figure 5.4**). The abasic-containing template can be generated

by DNA glycosylase from the damaged or mutated DNA strand. Then the 5'-phosphate pyrene strand and the labeled 3'-hydroxy strand can be hybridized to the template and ligated by T4 DNA ligase which has a wider range of temperature activity than PBCV-1 DNA ligase. The incorporation of thermal cycling can induce the dissociation of the target template and the ligated product. To avoid the re-hybridization of two long strands, biotin or a magnetic bead can be added on a 5' end to trap the ligated product. This detection method will result in the linear amplification of the ligated strand.

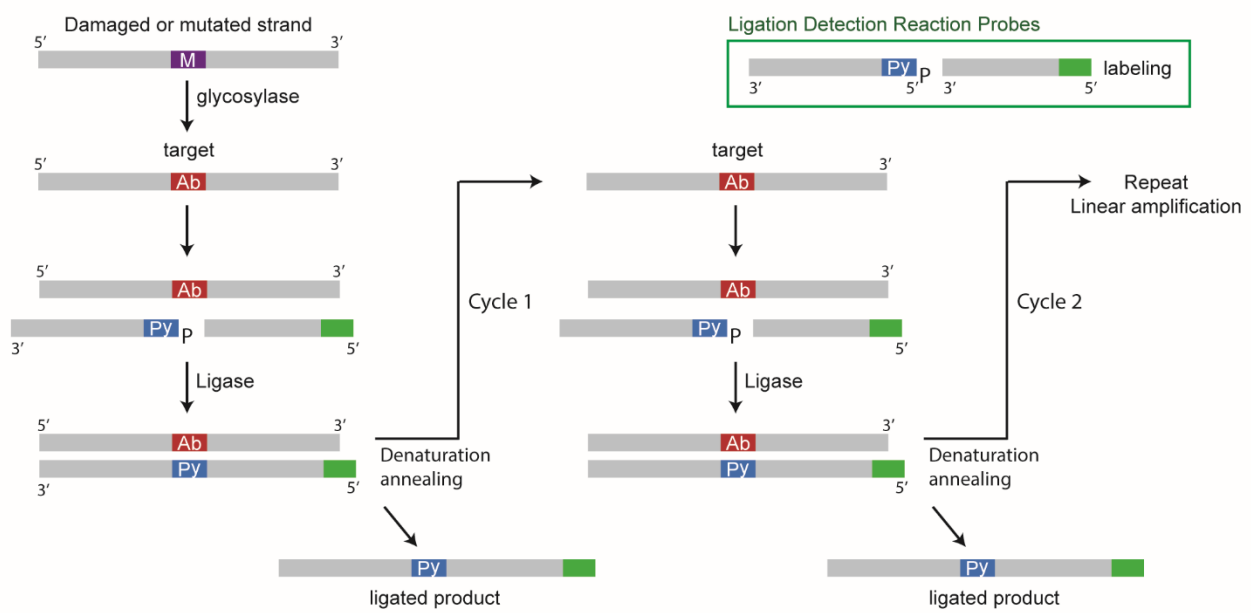


Figure 5.4 Proposed scheme of ligation detection reaction. This ligation detection reaction will result in the linear amplification of the product strand.

5.2.3 Enhancing the Stability of Triazole-Modified sgRNA

Although the *EMX1* targeting sgRNA and *WAS* targeting sgRNA were synthesized and ligated to sgRNA with different fluorophores and had good activity in the *in-vitro* assay, Siglec targeting sgRNAs showed very low or no activity in the cells. The possible reason for this poor activity is due to the lack of stabilizing agents. The deoxyribonucleotide was used on the 5' position

of sgRNA to enhance the resistance against RNase; however, this replacement of one base to deoxynucleotide was not sufficient. In the direct comparison of commercial dual gRNA on the same target, a difference in activity was observed likely due to the extra stabilizing agents on commercial gRNA. Recent work utilized modifications of sgRNA such as 2'-*O*-methyl, 2'-*O*-methyl-phosphorothioate, or 2'-*O*-methyl-phosphorothioate to increase the stability of sgRNA.[269,271,279] Therefore, further investigation of adding modification such as 2'-*O*-methyl or 2'-*O*-methyl-thiophosphate on either 3' end position or 5' end position in sgRNA is currently underway by my junior colleague, Santiago Tijaro Bulla. The amount of modification on either side or both may show different activities in the cell. The current goal is to find the best modification combinations in sgRNA to enhance the stability in cells. The next challenge for this project is to enable multiple gene-editing, more than three sgRNA, in the same cell. This modular synthesis of sgRNA can be incorporated in various ways. For example, the incorporation of the donor DNA template instead of the fluorophore on sgRNA can introduce a sequence of interest to induce homology-directed DNA repair. Finally, another example involves the incorporation of nanoparticles, which are a great candidate for gRNA delivery into cells.[276,277] The introduction of multiple sgRNAs simultaneously into a cell can be achieved by complexing multiple guides using a nanoparticle delivery agent.

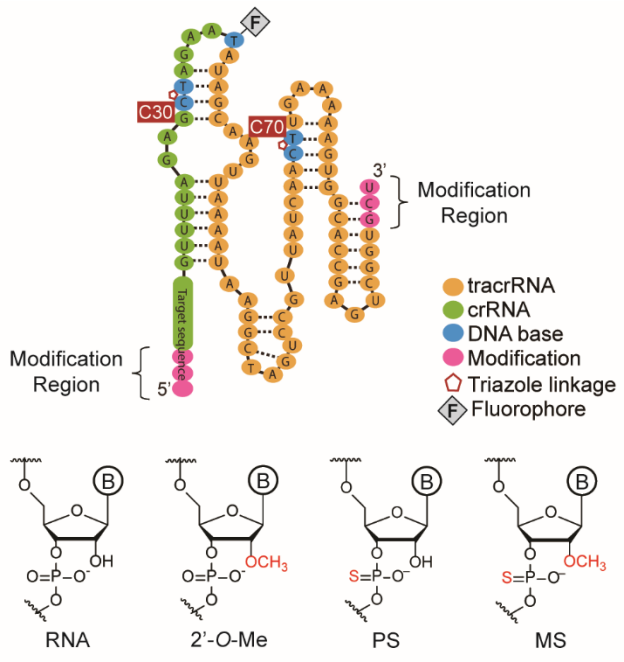


Figure 5.3 Modification candidates for stabilizing agents. Schematic illustration of triazole modified sgRNA. tracrRNA (yellow), crRNA (green), DNA bases (blue), and modification (pink). Modification candidates: 2'-*O*-methyl (2'-*O*-Me), phosphorothioate (PS), and 2'-*O*-methyl 3' phosphorothioate (MS).

Reference

(1) Orgel, L. E. The Origin of Life—a Review of Facts and Speculations. *Trends Biochem. Sci.* **1998**, *23* (12), 491–495. https://doi.org/10.1016/S0968-0004(98)01300-0.

(2) Joyce, G. F. RNA Evolution and the Origins of Life. *Nature* **1989**, *338* (6212), 217–224. https://doi.org/10.1038/338217a0.

(3) Joyce, G. F. The Antiquity of RNA-Based Evolution. *Nature* **2002**, *418* (6894), 214–221. https://doi.org/10.1038/418214a.

(4) Oparin, A. I. *The Origin of Life*; Dover Publications, 1965.

(5) Lazcano, A. Alexandr I. Oparin and the Origin of Life: A Historical Reassessment of the Heterotrophic Theory. *J. Mol. Evol.* **2016**, *83* (5–6), 214–222. https://doi.org/10.1007/s00239-016-9773-5.

(6) HALDANE, J. B. S. The Origin of Life. *Ration. Annu.* **1929**, *148*, 3–10.

(7) Tirard, S. J. B. S. Haldane and the Origin of Life. *J. Genet.* **2017**, *96* (5), 735–739. https://doi.org/10.1007/s12041-017-0831-6.

(8) Miller, S. L. A Production of Amino Acids Under Possible Primitive Earth Conditions. *Science* **1953**, *117* (3046), 528–529. https://doi.org/10.1126/science.117.3046.528.

(9) Miller, S. L.; Urey, H. C. Organic Compound Synthes on the Primitive Eart: Several Questions about the Origin of Life Have Been Answered, but Much Remains to Be Studied. *Science* **1959**, *130* (3370), 245–251. https://doi.org/10.1126/science.130.3370.245.

(10) Lazcano, A.; Bada, J. L. The 1953 Stanley L. Miller Experiment: Fifty Years of Prebiotic Organic Chemistry. 8.

(11) Stewart, J. E. The Origins of Life: The Managed-Metabolism Hypothesis. *Found. Sci.* **2019**, *24* (1), 171–195. https://doi.org/10.1007/s10699-018-9563-1.

(12) Hordijk, W.; Hein, J.; Steel, M. Autocatalytic Sets and the Origin of Life. *Entropy* **2010**, *12* (7), 1733–1742. https://doi.org/10.3390/e12071733.

(13) Hordijk, W.; Steel, M. Autocatalytic Networks at the Basis of Life's Origin and Organization. *Life* **2018**, *8* (4), 62. https://doi.org/10.3390/life8040062.

(14) Steel, M.; Hordijk, W.; Xavier, J. C. Autocatalytic Networks in Biology: Structural Theory and Algorithms. *J. R. Soc. Interface* **2019**, *16* (151), 20180808. https://doi.org/10.1098/rsif.2018.0808.

(15) Xavier, J. C.; Hordijk, W.; Kauffman, S.; Steel, M.; Martin, W. F. Autocatalytic Chemical Networks at the Origin of Metabolism. *Proc. R. Soc. B Biol. Sci.* **2020**, *287* (1922), 20192377. https://doi.org/10.1098/rspb.2019.2377.

(16) Springsteen, G.; Yerabolu, J. R.; Nelson, J.; Rhea, C. J.; Krishnamurthy, R. Linked Cycles of Oxidative Decarboxylation of Glyoxylate as Protometabolic Analogs of the Citric Acid Cycle. *Nat. Commun.* **2018**, *9* (1), 91. https://doi.org/10.1038/s41467-017-02591-0.

(17) Stubbs, R. T.; Yadav, M.; Krishnamurthy, R.; Springsteen, G. A Plausible Metal-Free Ancestral Analogue of the Krebs Cycle Composed Entirely of α-Ketoacids. *Nat. Chem.* **2020**, *12* (11), 1016–1022. https://doi.org/10.1038/s41557-020-00560-7.

(18) Kauffman, S. A.; Kauffman, M. S. F. I. P. B. S. A. *The Origins of Order: Self-Organization and Selection in Evolution*; Oxford University Press, 1993.

(19) Woese, C. R.; Dugre, D. H.; Saxinger, W. C.; Dugre, S. A. The Molecular Basis for the Genetic Code. *Proc. Natl. Acad. Sci.* **1966**, *55* (4), 966–974. https://doi.org/10.1073/pnas.55.4.966.

(20) Orgel, L. E. Evolution of the Genetic Apparatus. *J. Mol. Biol.* **1968**, *38* (3), 381–393. https://doi.org/10.1016/0022-2836(68)90393-8.

(21) Gilbert, W. Origin of Life: The RNA World. *Nature* **1986**, *319* (6055), 618–618. https://doi.org/10.1038/319618a0.

(22) Crick, F. H. C. The Origin of the Genetic Code. *J. Mol. Biol.* **1968**, *38* (3), 367–379. https://doi.org/10.1016/0022-2836(68)90392-6.

(23) McGIVAN, J. D. Horizons in Biochemistry and Biophysics (Volume 1). *Biochem. Soc. Trans.* **1976**, *4* (1), 179–180. https://doi.org/10.1042/bst0040179.

(24) Cech, T. R. The RNA Worlds in Context. *Cold Spring Harb. Perspect. Biol.* **2012**, *4* (7), a006742–a006742. https://doi.org/10.1101/cshperspect.a006742.

(25) Bernhardt, H. S. The RNA World Hypothesis: The Worst Theory of the Early Evolution of Life (except for All the Others)a. *Biol. Direct* **2012**, *7* (1), 23. https://doi.org/10.1186/1745-6150-7-23.

(26) Higgs, P. G.; Lehman, N. The RNA World: Molecular Cooperation at the Origins of Life. *Nat. Rev. Genet.* **2015**, *16* (1), 7–17. https://doi.org/10.1038/nrg3841.

(27) Goldman, A. D.; Samudrala, R.; Baross, J. A. The Evolution and Functional Repertoire of Translation Proteins Following the Origin of Life. **2010**, 12.

(28) Ashkenasy, G.; Jagasia, R.; Yadav, M.; Ghadiri, M. R. Design of a Directed Molecular Network. *Proc. Natl. Acad. Sci.* **2004**, *101* (30), 10872–10877. https://doi.org/10.1073/pnas.0402674101.

(29) Mattia, E.; Otto, S. Supramolecular Systems Chemistry. *Nat. Nanotechnol.* **2015**, *10* (2), 111–119. https://doi.org/10.1038/nnano.2014.337.

(30) Ashkenasy, G.; Hermans, T. M.; Otto, S.; Taylor, A. F. Systems Chemistry. *Chem. Soc. Rev.* **2017**, *46* (9), 2543–2554. https://doi.org/10.1039/C7CS00117G.

(31) Berg, J. M.; Tymoczko, J. L.; Stryer, L.; Stryer, L. *Biochemistry*, 5th ed.; W.H. Freeman: New York, 2002.

(32) Rye, C.; Wise, R.; Jurukovski, V.; DeSaix, J.; Choi, J.; Avissar, Y. DNA Replication in Eukaryotes. In *Biology*; OpenStax: Houston, Texas, 10212016.

(33) Lodish, H.; Berk, A.; Zipursky, L.; Matsudaira, P.; Baltimore, D.; Darnell, J. *Molecular Cell Biology*, Fourth ed.; Media connected W. H. Freeman and company: New York, 2000.

(34) Robertson, A.; Sinclair, A. J.; Philp, D. Minimal Self-Replicating Systems. *Chem. Soc. Rev.* **2000**, *29* (2), 141–152. https://doi.org/10.1039/a803602k.

(35) von Kiedrowski, G. A Self-Replicating Hexadeoxynucleotide. *Angew Chem Int Ed Engl* **1986**, *25* (10), 932–935. https://doi.org/10.1002/anie.198609322.

(36) Patzke, V.; Kiedrowski, G. von. Self Replicating Systems. *Arkivoc* **2007**, *2007* (5), 293–310. https://doi.org/10.3998/ark.5550190.0008.522.

(37) Sievers, D.; von Kiedrowski, G. Self-Replication of Complementary Nucleotide-Based Oligomers. *Nature* **1994**, *369* (6477), 221–224. https://doi.org/10.1038/369221a0.

(38) Grossmann, T. N.; Strohbach, A.; Seitz, O. Achieving Turnover in DNA-Templated Reactions. *ChemBioChem* **2008**, *9* (14), 2185–2192. https://doi.org/10.1002/cbic.200800290.

(39) Clixby, G.; Twyman, L. Self-Replicating Systems. *Org. Biomol. Chem.* **2016**, *14* (18), 4170–4184. https://doi.org/10.1039/C6OB00280C.

(40) Duim, H.; Otto, S. Towards Open-Ended Evolution in Self-Replicating Molecular Systems. *Beilstein J. Org. Chem.* **2017**, *13*, 1189–1203. https://doi.org/10.3762/bjoc.13.118.

(41) Yeldell, S. B.; Seitz, O. Nucleic Acid Constructs for the Interrogation of Multivalent Protein Interactions. *Chem. Soc. Rev.* **2020**, *49* (19), 6848–6865. https://doi.org/10.1039/D0CS00518E.

(42) Zhan, Z.-Y. J.; Lynn, D. G. Chemical Amplification through Template-Directed Synthesis. *J. Am. Chem. Soc.* **1997**, *119* (50), 12420–12421. https://doi.org/10.1021/ja972870v.

(43) Abe, H.; Kool, E. T. Destabilizing Universal Linkers for Signal Amplification in Self-Ligating Probes for RNA. *J. Am. Chem. Soc.* **2004**, *126* (43), 13980–13986. https://doi.org/10.1021/ja046791c.

(44) Dose, C.; Ficht, S.; Seitz, O. Reducing Product Inhibition in DNA-Template-Controlled Ligation Reactions. *Angew. Chem. Int. Ed.* **2006**, *45* (32), 5369–5373. https://doi.org/10.1002/anie.200600464.

(45) Goodwin, J. T.; Lynn, D. G. Template-Directed Synthesis: Use of a Reversible Reaction. *J. Am. Chem. Soc.* **1992**, *114* (23), 9197–9198. https://doi.org/10.1021/ja00049a067.

(46) Ianeselli, A.; Mast, C. B.; Braun, D. Periodic Melting of Oligonucleotides by Oscillating Salt Concentrations Triggered by Microscale Water Cycles Inside Heated Rock Pores. *Angew. Chem. Int. Ed.* **2019**, *58* (37), 13155–13160. https://doi.org/10.1002/anie.201907909.

(47) Mariani, A.; Bonfio, C.; Johnson, C. M.; Sutherland, J. D. PH-Driven RNA Strand Separation under Prebiotically Plausible Conditions. *Biochemistry* **2018**, *57* (45), 6382–6386. https://doi.org/10.1021/acs.biochem.8b01080.

(48) He, C.; Gállego, I.; Laughlin, B.; Grover, M. A.; Hud, N. V. A Viscous Solvent Enables Information Transfer from Gene-Length Nucleic Acids in a Model Prebiotic Replication Cycle. *Nat. Chem.* **2017**, *9* (4), 318–324. https://doi.org/10.1038/nchem.2628.

(49) He, C.; Lozoya-Colinas, A.; Gállego, I.; Grover, M. A.; Hud, N. V. Solvent Viscosity Facilitates Replication and Ribozyme Catalysis from an RNA Duplex in a Model Prebiotic Process. *Nucleic Acids Res.* **2019**, *47* (13), 6569–6577. https://doi.org/10.1093/nar/gkz496.

(50) Szostak, J. W. ON THE ORIGIN OF LIFE. **2016**, 5.

(51) Nakano, S. Nucleic Acid Duplex Stability: Influence of Base Composition on Cation Effects. *Nucleic Acids Res.* **1999**, *27* (14), 2957–2965. https://doi.org/10.1093/nar/27.14.2957.

(52) Tan, Z.-J.; Chen, S.-J. Nucleic Acid Helix Stability: Effects of Salt Concentration, Cation Valence and Size, and Chain Length. *Biophys. J.* **2006**, *90* (4), 1175–1190. https://doi.org/10.1529/biophysj.105.070904.

(53) Singh, A.; Singh, N. Effect of Salt Concentration on the Stability of Heterogeneous DNA. *Phys. Stat. Mech. Its Appl.* **2015**, *419*, 328–334. https://doi.org/10.1016/j.physa.2014.10.029.

(54) Schildkraut, C.; Lifson, S. Dependence of the melting temperature of DNA on salt concentration. *Biopolymers* **1965**, *3* (2), 195–208. https://doi.org/10.1002/bip.360030207.

(55) Gruenwedel, D. W.; Hsu, C.-H. Salt Effects on the Denaturation of DNA. *Biopolymers* **1969**, *7* (4), 557–570. https://doi.org/10.1002/bip.1969.360070412.

(56) Keil, L. M. R.; Möller, F. M.; Kieß, M.; Kudella, P. W.; Mast, C. B. Proton Gradients and PH Oscillations Emerge from Heat Flow at the Microscale. *Nat. Commun.* **2017**, *8* (1), 1897. https://doi.org/10.1038/s41467-017-02065-3.

(57) Bag, B. G.; von Kiedrowski, G. Templates, Autocatalysis and Molecular Replication. *Pure Appl. Chem.* **1996**, *68* (11), 2145–2152. https://doi.org/10.1351/pac199668112145.

(58) Kosikova, T.; Philp, D. Exploring the Emergence of Complexity Using Synthetic Replicators. *Chem. Soc. Rev.* **2017**, *46* (23), 7274–7305. https://doi.org/10.1039/C7CS00123A.

(59) Ellington, A. D.; Szostak, J. W. In Vitro Selection of RNA Molecules That Bind Specific Ligands. *Nature* **1990**, *346* (6287), 818–822. https://doi.org/10.1038/346818a0.

(60) Dickerson, R. E. Chemical Evolution and the Origin of Life. *Sci. Am.* **1978**, *239* (3), 70–87.

(61) Higgs, P. G. Chemical Evolution and the Evolutionary Definition of Life. *J. Mol. Evol.* **2017**, *84* (5–6), 225–235. https://doi.org/10.1007/s00239-017-9799-3.

(62) Schrum, J. P.; Zhu, T. F.; Szostak, J. W. The Origins of Cellular Life. *Cold Spring Harb. Perspect. Biol.* **2010**, *2* (9), a002212–a002212. https://doi.org/10.1101/cshperspect.a002212.

(63) Zielinski, W. S.; Orgel, L. E. Autocatalytic Synthesis of a Tetranucleotide Analogue. *Nature* **1987**, *327* (6120), 346–347. https://doi.org/10.1038/327346a0.

(64) Luther, A.; Brandsch, R.; von Kiedrowski, G. Surface-Promoted Replication and Exponential Amplification of DNA Analogues. *Nature* **1998**, *396* (6708), 245–248. https://doi.org/10.1038/24343.

(65) Sievers, D.; von Kiedrowski, G. Self-Replication of Hexadeoxynucleotide Analogues: Autocatalysis versus Cross-Catalysis. *Chem. - Eur. J.* **1998**, *4* (4), 629–641. https://doi.org/10.1002/(SICI)1521-3765(19980416)4:4<629::AID-CHEM629>3.0.CO;2-0.

(66) Ferré-D'Amaré, A. R.; Scott, W. G. Small Self-Cleaving Ribozymes. *Cold Spring Harb. Perspect. Biol.* **2010**, *2* (10), a003574. https://doi.org/10.1101/cshperspect.a003574.

(67) Paul, N.; Joyce, G. F. A Self-Replicating Ligase Ribozyme. *Proc. Natl. Acad. Sci.* **2002**, *99* (20), 12733–12740. https://doi.org/10.1073/pnas.202471099.

(68) Lincoln, T. A.; Joyce, G. F. Self-Sustained Replication of an RNA Enzyme. *Science* **2009**, *323* (5918), 1229–1232. https://doi.org/10.1126/science.1167856.

(69) Engelhart, A. E.; Powner, M. W.; Szostak, J. W. Functional RNAs Exhibit Tolerance for Non-Heritable 2′–5′ versus 3′–5′ Backbone Heterogeneity. *Nat. Chem.* **2013**, *5* (5), 390–394. https://doi.org/10.1038/nchem.1623.

(70) Di Pisa, M.; Seitz, O. Nucleic Acid Templated Reactions for Chemical Biology. *ChemMedChem* **2017**, *12* (12), 872–882. https://doi.org/10.1002/cmdc.201700266.

(71) Silverman, A. P.; Kool, E. T. Detecting RNA and DNA with Templated Chemical Reactions. *Chem. Rev.* **2006**, *106* (9), 3775–3789. https://doi.org/10.1021/cr050057+.

(72) Allen, D.; Rosenberg, M.; Hendel, A. Using Synthetically Engineered Guide RNAs to Enhance CRISPR Genome Editing Systems in Mammalian Cells. *Front. Genome Ed.* **2021**, *2*, 617910. https://doi.org/10.3389/fgeed.2020.617910.

(73) Filippova, J.; Matveeva, A.; Zhuravlev, E.; Stepanov, G. Guide RNA Modification as a Way to Improve CRISPR/Cas9-Based Genome-Editing Systems. *Biochimie* **2019**, *167*, 49–60. https://doi.org/10.1016/j.biochi.2019.09.003.

(74) Kelley, M. L.; Strezoska, Ž.; He, K.; Vermeulen, A.; Smith, A. van B. Versatility of Chemically Synthesized Guide RNAs for CRISPR-Cas9 Genome Editing. *J. Biotechnol.* **2016**, *233*, 74–83. https://doi.org/10.1016/j.jbiotec.2016.06.011.

(75) Li, H.; Yang, Y.; Hong, W.; Huang, M.; Wu, M.; Zhao, X. Applications of Genome Editing Technology in the Targeted Therapy of Human Diseases: Mechanisms, Advances and Prospects. *Signal Transduct. Target. Ther.* **2020**, *5* (1), 1. https://doi.org/10.1038/s41392-019-0089-y.

(76) Chen, K.; Wang, Y.; Zhang, R.; Zhang, H.; Gao, C. CRISPR/Cas Genome Editing and Precision Plant Breeding in Agriculture. *Annu. Rev. Plant Biol.* **2019**, *70* (1), 667–697. https://doi.org/10.1146/annurev-arplant-050718-100049.

(77) Ricroch, A. Global Developments of Genome Editing in Agriculture. *Transgenic Res.* **2019**, *28* (S2), 45–52. https://doi.org/10.1007/s11248-019-00133-6.

(78) Letsinger, R. L.; Ogilvie, K. K. Nucleotide Chemistry. XIII. Synthesis of Oligothymidylates via Phosphotriester Intermediates. *J. Am. Chem. Soc.* **1969**, *91* (12), 3350–3355. https://doi.org/10.1021/ja01040a042.

(79) Ogilvie, K. K.; Theriault, N.; Sadana, K. L. Synthesis of Oligoribonucleotides. *J. Am. Chem. Soc.* **1977**, *99* (23), 7741–7743. https://doi.org/10.1021/ja00465a073.

(80) Usman, N.; Ogilvie, K. K.; Jiang, M. Y.; Cedergren, R. J. The Automated Chemical Synthesis of Long Oligoribuncleotides Using 2'-O-Silylated Ribonucleoside 3'-O-Phosphoramidites on a Controlled-Pore Glass Support: Synthesis of a 43-Nucleotide Sequence Similar to the 3'-Half Molecule of an Escherichia Coli Formylmethionine TRNA. *J. Am. Chem. Soc.* **1987**, *109* (25), 7845–7854. https://doi.org/10.1021/ja00259a037.

(81) Evans, M. J.; Kaufman, M. H. Establishment in Culture of Pluripotential Cells from Mouse Embryos. *Nature* **1981**, *292* (5819), 154–156. https://doi.org/10.1038/292154a0.

(82) Devereux, J.; Haeberli, P.; Smithies, O. A Comprehensive Set of Sequence Analysis Programs for the VAX. *Nucleic Acids Res.* **1984**, *12* (1Part1), 387–395. https://doi.org/10.1093/nar/12.1Part1.387.

(83) Thomas, K. R.; Capecchi, M. R. Site-Directed Mutagenesis by Gene Targeting in Mouse Embryo-Derived Stem Cells. *Cell* **1987**, *51* (3), 503–512. https://doi.org/10.1016/0092-8674(87)90646-5.

(84) Capecchi, M. R. The New Mouse Genetics: Altering the Genome by Gene Targeting. *Trends Genet.* **1989**, *5*, 70–76. https://doi.org/10.1016/0168-9525(89)90029-2.

(85) Urnov, F. D.; Rebar, E. J.; Holmes, M. C.; Zhang, H. S.; Gregory, P. D. Genome Editing with Engineered Zinc Finger Nucleases. *Nat. Rev. Genet.* **2010**, *11* (9), 636–646. https://doi.org/10.1038/nrg2842.

(86) Durai, S. Zinc Finger Nucleases: Custom-Designed Molecular Scissors for Genome Engineering of Plant and Mammalian Cells. *Nucleic Acids Res.* **2005**, *33* (18), 5978–5990. https://doi.org/10.1093/nar/gki912.

(87) Carroll, D. Genome Engineering With Zinc-Finger Nucleases. *Genetics* **2011**, *188* (4), 773–782. https://doi.org/10.1534/genetics.111.131433.

(88) Pavletich, N.; Pabo, C. Zinc Finger-DNA Recognition: Crystal Structure of a Zif268-DNA Complex at 2.1 A. *Science* **1991**, *252* (5007), 809–817. https://doi.org/10.1126/science.2028256.

(89) Li, L.; Wu, L. P.; Chandrasegaran, S. Functional Domains in Fok I Restriction Endonuclease. *Proc. Natl. Acad. Sci.* **1992**, *89* (10), 4275–4279. https://doi.org/10.1073/pnas.89.10.4275.

(90) Kim, Y. G.; Chandrasegaran, S. Chimeric Restriction Endonuclease. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91* (3), 883–887.

(91) Kim, Y. G.; Cha, J.; Chandrasegaran, S. Hybrid Restriction Enzymes: Zinc Finger Fusions to Fok I Cleavage Domain. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93* (3), 1156–1160.

(92) Wah, D. A.; Bitinaite, J.; Schildkraut, I.; Aggarwal, A. K. Structure of FokI Has Implications for DNA Cleavage. *Proc. Natl. Acad. Sci.* **1998**, *95* (18), 10564–10569. https://doi.org/10.1073/pnas.95.18.10564.

(93) Smith, J. Requirements for Double-Strand Cleavage by Chimeric Restriction Enzymes with Zinc Finger DNA-Recognition Domains. *Nucleic Acids Res.* **2000**, *28* (17), 3361–3369. https://doi.org/10.1093/nar/28.17.3361.

(94) Bibikova, M.; Carroll, D.; Segal, D. J.; Trautman, J. K.; Smith, J.; Kim, Y.-G.; Chandrasegaran, S. Stimulation of Homologous Recombination through Targeted Cleavage by Chimeric Nucleases. *Mol. Cell. Biol.* **2001**, *21* (1), 289–297. https://doi.org/10.1128/MCB.21.1.289-297.2001.

(95) Shimizu, Y.; Bhakta, M. S.; Segal, D. J. Restricted Spacer Tolerance of a Zinc Finger Nuclease with a Six Amino Acid Linker. *Bioorg. Med. Chem. Lett.* **2009**, *19* (14), 3970–3972. https://doi.org/10.1016/j.bmcl.2009.02.109.

(96) Händel, E.-M.; Alwin, S.; Cathomen, T. Expanding or Restricting the Target Site Repertoire of Zinc-Finger Nucleases: The Inter-Domain Linker as a Major Determinant of Target Site Selectivity. *Mol. Ther.* **2009**, *17* (1), 104–111. https://doi.org/10.1038/mt.2008.233.

(97) Beumer, K. J.; Trautman, J. K.; Bozas, A.; Liu, J.-L.; Rutter, J.; Gall, J. G.; Carroll, D. Efficient Gene Targeting in Drosophila by Direct Embryo Injection with Zinc-Finger Nucleases. *Proc. Natl. Acad. Sci.* **2008**, *105* (50), 19821–19826. https://doi.org/10.1073/pnas.0810475105.

(98) Miller, J. C.; Holmes, M. C.; Wang, J.; Guschin, D. Y.; Lee, Y.-L.; Rupniewski, I.; Beausejour, C. M.; Waite, A. J.; Wang, N. S.; Kim, K. A.; et al. An Improved Zinc-Finger

Nuclease Architecture for Highly Specific Genome Editing. *Nat. Biotechnol.* **2007**, *25* (7), 778–785. https://doi.org/10.1038/nbt1319.

(99) Paschon, D. E.; Lussier, S.; Wangzor, T.; Xia, D. F.; Li, P. W.; Hinkley, S. J.; Scarlott, N. A.; Lam, S. C.; Waite, A. J.; Truong, L. N.; et al. Diversifying the Structure of Zinc Finger Nucleases for High-Precision Genome Editing. *Nat. Commun.* **2019**, *10* (1), 1133. https://doi.org/10.1038/s41467-019-08867-x.

(100) Ramirez, C. L.; Foley, J. E.; Wright, D. A.; Müller-Lerch, F.; Rahman, S. H.; Cornu, T. I.; Winfrey, R. J.; Sander, J. D.; Fu, F.; Townsend, J. A.; et al. Unexpected Failure Rates for Modular Assembly of Engineered Zinc Fingers. *Nat. Methods* **2008**, *5* (5), 374–375. https://doi.org/10.1038/nmeth0508-374.

(101) Sander, J. D.; Dahlborg, E. J.; Goodwin, M. J.; Cade, L.; Zhang, F.; Cifuentes, D.; Curtin, S. J.; Blackburn, J. S.; Thibodeau-Beganny, S.; Qi, Y.; et al. Selection-Free Zinc-Finger-Nuclease Engineering by Context-Dependent Assembly (CoDA). *Nat. Methods* **2011**, *8* (1), 67–69. https://doi.org/10.1038/nmeth.1542.

(102) Gupta, A.; Christensen, R. G.; Rayla, A. L.; Lakshmanan, A.; Stormo, G. D.; Wolfe, S. A. An Optimized Two-Finger Archive for ZFN-Mediated Gene Targeting. *Nat. Methods* **2012**, *9* (6), 588–590. https://doi.org/10.1038/nmeth.1994.

(103) Gupta, R. M.; Musunuru, K. Expanding the Genetic Editing Tool Kit: ZFNs, TALENs, and CRISPR-Cas9. *J. Clin. Invest.* **2014**, *124* (10), 4154–4161. https://doi.org/10.1172/JCI72992.

(104) Hillary, V. E.; Ceasar, S. A.; Ignacimuthu, S. Genome Engineering in Insects: Focus on the CRISPR/Cas9 System. In *Genome Engineering via CRISPR-Cas9 System*; Elsevier, 2020; pp 219–249. https://doi.org/10.1016/B978-0-12-818140-9.00018-0.

(105) Christian, M.; Cermak, T.; Doyle, E. L.; Schmidt, C.; Zhang, F.; Hummel, A.; Bogdanove, A. J.; Voytas, D. F. Targeting DNA Double-Strand Breaks with TAL Effector Nucleases. *Genetics* **2010**, *186* (2), 757–761. https://doi.org/10.1534/genetics.110.120717.

(106) Zhang, F.; Cong, L.; Lodato, S.; Kosuri, S.; Church, G. M.; Arlotta, P. Efficient Construction of Sequence-Specific TAL Effectors for Modulating Mammalian Transcription. *Nat. Biotechnol.* **2011**, *29* (2), 149–153. https://doi.org/10.1038/nbt.1775.

(107) Szczepek, M.; Brondani, V.; Büchel, J.; Serrano, L.; Segal, D. J.; Cathomen, T. Structure-Based Redesign of the Dimerization Interface Reduces the Toxicity of Zinc-Finger Nucleases. *Nat. Biotechnol.* **2007**, *25* (7), 786–793. https://doi.org/10.1038/nbt1317.

(108) Doyon, Y.; Vo, T. D.; Mendel, M. C.; Greenberg, S. G.; Wang, J.; Xia, D. F.; Miller, J. C.; Urnov, F. D.; Gregory, P. D.; Holmes, M. C. Enhancing Zinc-Finger-Nuclease Activity with Improved Obligate Heterodimeric Architectures. *Nat. Methods* **2011**, *8* (1), 74–79. https://doi.org/10.1038/nmeth.1539.

(109) Guo, J.; Gaj, T.; Barbas, C. F. Directed Evolution of an Enhanced and Highly Efficient FokI Cleavage Domain for Zinc Finger Nucleases. *J. Mol. Biol.* **2010**, *400* (1), 96–107. https://doi.org/10.1016/j.jmb.2010.04.060.

(110) Miller, J. C.; Tan, S.; Qiao, G.; Barlow, K. A.; Wang, J.; Xia, D. F.; Meng, X.; Paschon, D. E.; Leung, E.; Hinkley, S. J.; et al. A TALE Nuclease Architecture for Efficient Genome Editing. *Nat. Biotechnol.* **2011**, *29* (2), 143–148. https://doi.org/10.1038/nbt.1755.

(111) Mussolino, C.; Morbitzer, R.; Lütge, F.; Dannemann, N.; Lahaye, T.; Cathomen, T. A Novel TALE Nuclease Scaffold Enables High Genome Editing Activity in Combination with Low Toxicity. *Nucleic Acids Res.* **2011**, *39* (21), 9283–9293. https://doi.org/10.1093/nar/gkr597.

(112) Horvath, P.; Barrangou, R. CRISPR/Cas, the Immune System of Bacteria and Archaea. *Science* **2010**, *327* (5962), 167–170. https://doi.org/10.1126/science.1179555.

(113) Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J. A.; Charpentier, E. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **2012**, *337* (6096), 816–821. https://doi.org/10.1126/science.1225829.

(114) Jiang, F.; Zhou, K.; Ma, L.; Gressel, S.; Doudna, J. A. A Cas9–Guide RNA Complex Preorganized for Target DNA Recognition. *Science* **2015**, *348* (6242), 1477. https://doi.org/10.1126/science.aab1452.

(115) Cong, L.; Ran, F. A.; Cox, D.; Lin, S.; Barretto, R.; Habib, N.; Hsu, P. D.; Wu, X.; Jiang, W.; Marraffini, L. A.; et al. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **2013**, *339* (6121), 819–823. https://doi.org/10.1126/science.1231143.

(116) Anders, C.; Niewoehner, O.; Duerst, A.; Jinek, M. Structural Basis of PAM-Dependent Target DNA Recognition by the Cas9 Endonuclease. *Nature* **2014**, *513* (7519), 569–573. https://doi.org/10.1038/nature13579.

(117) Mali, P.; Yang, L.; Esvelt, K. M.; Aach, J.; Guell, M.; DiCarlo, J. E.; Norville, J. E.; Church, G. M. RNA-Guided Human Genome Engineering via Cas9. *Science* **2013**, *339* (6121), 823–826. https://doi.org/10.1126/science.1232033.

(118) Fu, Y.; Foden, J. A.; Khayter, C.; Maeder, M. L.; Reyon, D.; Joung, J. K.; Sander, J. D. High-Frequency off-Target Mutagenesis Induced by CRISPR-Cas Nucleases in Human Cells. *Nat. Biotechnol.* **2013**, *31* (9), 822–826. https://doi.org/10.1038/nbt.2623.

(119) Hsu, P. D.; Scott, D. A.; Weinstein, J. A.; Ran, F. A.; Konermann, S.; Agarwala, V.; Li, Y.; Fine, E. J.; Wu, X.; Shalem, O.; et al. DNA Targeting Specificity of RNA-Guided Cas9 Nucleases. *Nat. Biotechnol.* **2013**, *31* (9), 827–832. https://doi.org/10.1038/nbt.2647.

(120) Jain, S.; Shukla, S.; Yang, C.; Zhang, M.; Fatma, Z.; Lingamaneni, M.; Abesteh, S.; Lane, S. T.; Xiong, X.; Wang, Y.; et al. TALEN Outperforms Cas9 in Editing Heterochromatin Target Sites. *Nat. Commun.* **2021**, *12* (1), 606. https://doi.org/10.1038/s41467-020-20672-5.

(121) Doudna, J. A.; Charpentier, E. The New Frontier of Genome Engineering with CRISPR-Cas9. *Science* **2014**, *346* (6213), 1258096. https://doi.org/10.1126/science.1258096.

(122) Dow, L. E. Modeling Disease In Vivo With CRISPR/Cas9. *Trends Mol. Med.* **2015**, *21* (10), 609–621. https://doi.org/10.1016/j.molmed.2015.07.006.

(123) Dever, D. P.; Bak, R. O.; Reinisch, A.; Camarena, J.; Washington, G.; Nicolas, C. E.; Pavel-Dinu, M.; Saxena, N.; Wilkens, A. B.; Mantri, S.; et al. CRISPR/Cas9 β-Globin Gene Targeting in Human Haematopoietic Stem Cells. *Nature* **2016**, *539* (7629), 384–389. https://doi.org/10.1038/nature20134.

(124) Frangoul, H.; Altshuler, D.; Cappellini, M. D.; Chen, Y.-S.; Domm, J.; Eustace, B. K.; Foell, J.; de la Fuente, J.; Grupp, S.; Handgretinger, R.; et al. CRISPR-Cas9 Gene Editing for Sickle Cell Disease and β-Thalassemia. *N. Engl. J. Med.* **2021**, *384* (3), 252–260. https://doi.org/10.1056/NEJMoa2031054.

(125) Kaminski, M. M.; Abudayyeh, O. O.; Gootenberg, J. S.; Zhang, F.; Collins, J. J. CRISPR-Based Diagnostics. *Nat. Biomed. Eng.* **2021**, *5* (7), 643–656. https://doi.org/10.1038/s41551-021-00760-7.

(126) Ledford, H. CRISPR Treatment Inserted Directly into the Body for First Time. *Nature* **2020**, *579* (7798), 185–185. https://doi.org/10.1038/d41586-020-00655-8.

(127) Gillmore, J. D.; Gane, E.; Taubel, J.; Kao, J.; Fontana, M.; Maitland, M. L.; Seitzer, J.; O'Connell, D.; Walsh, K. R.; Wood, K.; et al. CRISPR-Cas9 In Vivo Gene Editing for Transthyretin Amyloidosis. *N. Engl. J. Med.* **2021**, *385* (6), 493–502. https://doi.org/10.1056/NEJMoa2107454.

(128) Lee, D. H.; Granja, J. R.; Martinez, J. A.; Severin, K.; Ghadiri, M. R. A Self-Replicating Peptide. *Nature* **1996**, *382* (6591), 525–528. https://doi.org/10.1038/382525a0.

(129) Lee, D. H.; Severin, K.; Yokobayashi, Y.; Ghadiri, M. R. Emergence of Symbiosis in Peptide Self-Replication through a Hypercyclic Network. *Nature* **1997**, *390* (6660), 591–594. https://doi.org/10.1038/37569.

(130) Severin, K.; Lee, D. H.; Martinez, J. A.; Vieth, M.; Ghadiri, M. R. Dynamic Error Correction in Autocatalytic Peptide Networks. *Angew. Chem. Int. Ed.* **1998**, *37* (1-2), 126–128. https://doi.org/10.1002/(SICI)1521-3773(19980202)37:1/2<126::AID-ANIE126>3.0.CO;2-4.

(131) Saghatelian, A.; Yokobayashi, Y.; Soltani, K.; Ghadiri, M. R. A Chiroselective Peptide Replicator. *Nature* **2001**, *409* (6822), 797–801. https://doi.org/10.1038/35057238.

(132) von Kiedrowski, G.; Wlotzka, B.; Helbing, J.; Matzen, M.; Jordan, S. Parabolic Growth of a Self-Replicating Hexadeoxynucleotide Bearing a 3'-5'-Phosphoamidate Linkage. *Angew. Chem. Int. Ed. Engl.* **1991**, *30* (4), 423–426. https://doi.org/10.1002/anie.199104231.

(133) Achilles, T.; von Kiedrowski, G. A Self-Replicating System FromThree Starting Materials. *Angew. Chem. Int. Ed. Engl.* **1993**, *32* (8), 1198–1201. https://doi.org/10.1002/anie.199311981.

(134) Mansy, S. S.; Schrum, J. P.; Krishnamurthy, M.; Tobé, S.; Treco, D. A.; Szostak, J. W. Template-Directed Synthesis of a Genetic Polymer in a Model Protocell. *Nature* **2008**, *454* (7200), 122–125. https://doi.org/10.1038/nature07018.

(135) Mansy, S. S.; Szostak, J. W. Thermostability of Model Protocell Membranes. *Proc. Natl. Acad. Sci.* **2008**, *105* (36), 13351–13355. https://doi.org/10.1073/pnas.0805086105.

(136) Toparlak, O. D.; Mansy, S. S. Progress in Synthesizing Protocells. *Exp. Biol. Med.* **2019**, *244* (4), 304–313. https://doi.org/10.1177/1535370218816657.

(137) SantaLucia, J. A Unified View of Polymer, Dumbbell, and Oligonucleotide DNA Nearest-Neighbor Thermodynamics. *Proc. Natl. Acad. Sci.* **1998**, *95* (4), 1460–1465. https://doi.org/10.1073/pnas.95.4.1460.

(138) Sheng, J.; Li, L.; Engelhart, A. E.; Gan, J.; Wang, J.; Szostak, J. W. Structural Insights into the Effects of 2'-5' Linkages on the RNA Duplex. *Proc. Natl. Acad. Sci.* **2014**, *111* (8), 3050–3055. https://doi.org/10.1073/pnas.1317799111.

(139) Ye, J.; Gat, Y.; Lynn, D. G. Catalyst for DNA Ligation: Towards a Two-Stage Replication Cycle. *Angew. Chem. Int. Ed.* **2000**, *39* (20), 3641–3643. https://doi.org/10.1002/1521-3773(20001016)39:20<3641::AID-ANIE3641>3.0.CO;2-J.

(140) Velema, W. A.; Kool, E. T. Fluorogenic Templated Reaction Cascades for RNA Detection. *J. Am. Chem. Soc.* **2017**, *139* (15), 5405–5411. https://doi.org/10.1021/jacs.7b00466.

(141) Paul, N.; Joyce, G. F. Minimal Self-Replicating Systems. *Curr. Opin. Chem. Biol.* **2004**, *8* (6), 634–639. https://doi.org/10.1016/j.cbpa.2004.09.005.

(142) Kim, D.-E.; Joyce, G. F. Cross-Catalytic Replication of an RNA Ligase Ribozyme. *Chem. Biol.* **2004**, *11* (11), 1505–1512. https://doi.org/10.1016/j.chembiol.2004.08.021.

(143) Kausar, A.; McKay, R. D.; Lam, J.; Bhogal, R. S.; Tang, A. Y.; Gibbs-Davis, J. M. Tuning DNA Stability To Achieve Turnover in Template for an Enzymatic Ligation Reaction. *Angew. Chem. Int. Ed.* **2011**, *50* (38), 8922–8926. https://doi.org/10.1002/anie.201102579.

(144) Kausar, A.; Mitran, C. J.; Li, Y.; Gibbs-Davis, J. M. Rapid, Isothermal DNA Self-Replication Induced by a Destabilizing Lesion. *Angew. Chem. Int. Ed.* **2013**, *52* (40), 10577–10581. https://doi.org/10.1002/anie.201303225.

(145) Alladin-Mustan, B. S.; Mitran, C. J.; Gibbs-Davis, J. M. Achieving Room Temperature DNA Amplification by Dialling in Destabilization. *Chem. Commun.* **2015**, *51* (44), 9101–9104. https://doi.org/10.1039/C5CC01548K.

(146) Alladin-Mustan, B. S.; Liu, Y.; Li, Y.; de Almeida, D. R. Q.; Yuzik, J.; Mendes, C. F.; Gibbs, J. M. Reverse Transcription Lesion-Induced DNA Amplification: An Instrument-Free Isothermal Method to Detect RNA. *Anal. Chim. Acta* **2021**, *1149*, 238130. https://doi.org/10.1016/j.aca.2020.12.005.

(147) Ferretti, L.; Sgaramella, V. Specific and Reversible Inhibition of the Blunt End Joining Activity of the T4 DNA Ligase. *Nucleic Acids Res.* **1981**, *9* (15), 3695–3705. https://doi.org/10.1093/nar/9.15.3695.

(148) Rossi, R.; Montecucco, A; Ciarrocchi, G; Biamonti, G. Functional Characterization of the T4 DNA Ligase: A New Insight into the Mechanism of Action. *Nucleic Acids Res.* **1997**, *25* (11), 2106–2113. https://doi.org/10.1093/nar/25.11.2106.

(149) Kalnik, M. W.; Chang, C. N.; Johnson, F.; Grollman, A. P.; Patel, D. J. NMR Studies of Abasic Sites in DNA Duplexes: Deoxyadenosine Stacks into the Helix Opposite Acyclic Lesions. *Biochemistry* **1989**, *28* (8), 3373–3383. https://doi.org/10.1021/bi00434a037.

(150) Vesnaver, G.; Chang, C. N.; Eisenberg, M.; Grollman, A. P.; Breslauer, K. J. Influence of Abasic and Anucleosidic Sites on the Stability, Conformation, and Melting Behavior of a DNA Duplex: Correlations of Thermodynamic and Structural Data. *Proc. Natl. Acad. Sci.* **1989**, *86* (10), 3614–3618. https://doi.org/10.1073/pnas.86.10.3614.

(151) Marky, L. A.; Breslauer, K. J. Calculating Thermodynamic Data for Transitions of Any Molecularity from Equilibrium Melting Curves. *Biopolymers* **1987**, *26* (9), 1601–1620. https://doi.org/10.1002/bip.360260911.

(152) Oligo Analyzer https://www.idtdna.com/calc/analyzer (accessed 2021 -11 -16).

(153) Ghai, R.; Falconer, R. J.; Collins, B. M. Applications of Isothermal Titration Calorimetry in Pure and Applied Research—Survey of the Literature from 2010. *J. Mol. Recognit.* **2012**, *25* (1), 32–52. https://doi.org/10.1002/jmr.1167.

(154) Mikulecky, P. J.; Feig, A. L. Heat Capacity Changes Associated with Nucleic Acid Folding. *Biopolymers* **2006**, *82* (1), 38–58. https://doi.org/10.1002/bip.20457.

(155) Kozlov, A. G.; Lohman, T. M. Calorimetric Studies of E. Coli SSB Protein-Single-Stranded DNA Interactions. Effects of Monovalent Salts on Binding Enthalpy 1 1Edited by D. Draper. *J. Mol. Biol.* **1998**, *278* (5), 999–1014. https://doi.org/10.1006/jmbi.1998.1738.

(156) Loregian, A.; Sinigalia, E.; Mercorelli, B.; Palù, G.; Coen, D. M. Binding Parameters and Thermodynamics of the Interaction of the Human Cytomegalovirus DNA Polymerase Accessory Protein, UL44, with DNA: Implications for the Processivity Mechanism. *Nucleic Acids Res.* **2007**, *35* (14), 4779–4791. https://doi.org/10.1093/nar/gkm506.

(157) Kozlov, A. G.; Lohman, T. M. SSB Binding to SsDNA Using Isothermal Titration Calorimetry. In *Single-Stranded DNA Binding Proteins: Methods and Protocols*; Keck, J. L., Ed.; Methods in Molecular Biology; Humana Press: Totowa, NJ, 2012; pp 37–54. https://doi.org/10.1007/978-1-62703-032-8_3.

(158) Pierce, M. M.; Raman, C. S.; Nall, B. T. Isothermal Titration Calorimetry of Protein–Protein Interactions. *Methods* **1999**, *19* (2), 213–221. https://doi.org/10.1006/meth.1999.0852.

(159) Dam, T. K.; Brewer, C. F. Thermodynamic Studies of Lectin−Carbohydrate Interactions by Isothermal Titration Calorimetry. *Chem. Rev.* **2002**, *102* (2), 387–430. https://doi.org/10.1021/cr000401x.

(160) Takach, J. C.; Mikulecky, P. J.; Feig, A. L. Salt-Dependent Heat Capacity Changes for RNA Duplex Formation. *J. Am. Chem. Soc.* **2004**, *126* (21), 6530–6531. https://doi.org/10.1021/ja0316263.

(161) Tikhomirova, A.; Taulier, N.; Chalikian, T. V. Energetics of Nucleic Acid Stability: The Effect of ΔC_p. *J. Am. Chem. Soc.* **2004**, *126* (50), 16387–16394. https://doi.org/10.1021/ja046387d.

(162) Gourishankar, A.; Shukla, S.; Ganesh, K. N.; Sastry, M. Isothermal Titration Calorimetry Studies on the Binding of DNA Bases and PNA Base Monomers to Gold Nanoparticles. *J. Am. Chem. Soc.* **2004**, *126* (41), 13186–13187. https://doi.org/10.1021/ja046785g.

(163) Mikulecky, P. J.; Feig, A. L. Heat Capacity Changes Associated with DNA Duplex Formation: Salt- and Sequence-Dependent Effects. *Biochemistry* **2006**, *45* (2), 604–616. https://doi.org/10.1021/bi0517178.

(164) Engelhardt, J.; Andresen, B.; Rasmussen, S. Thermodynamics and Kinetics of Lesion Induced DNA Amplification (LIDA); 2020; pp 269–272. https://doi.org/10.1162/isal_a_00312.

(165) Lehman, I. R. DNA Ligase: Structure, Mechanism, and Function: The Joining of DNA Chains by DNA Ligase Is an Essential Component of DNA Repair. Replication, and Recombination. *Science* **1974**, *186* (4166), 790–797. https://doi.org/10.1126/science.186.4166.790.

(166) Bauer, R. J.; Jurkiw, T. J.; Evans, T. C.; Lohman, G. J. S. Rapid Time Scale Analysis of T4 DNA Ligase–DNA Binding. *Biochemistry* **2017**, *56* (8), 1117–1129. https://doi.org/10.1021/acs.biochem.6b01261.

(167) Bauer, R. J.; Evans, T. C.; Lohman, G. J. S. The Inhibitory Effect of Non-Substrate and Substrate DNA on the Ligation and Self-Adenylylation Reactions Catalyzed by T4 DNA Ligase. *PLOS ONE* **2016**, *11* (3), e0150802. https://doi.org/10.1371/journal.pone.0150802.

(168) Goffin, C.; Bailly, V.; Verly, W. G. Nicks 3′ or 5′ to AP Sites or to Mispaired Bases, and One-Nucleotide Gaps Can Be Sealed by T4 DNA Ligase. *Nucleic Acids Res.* **1987**, *15* (21), 8755–8771. https://doi.org/10.1093/nar/15.21.8755.

(169) Bogenhagen, D. F.; Pinz, K. G. The Action of DNA Ligase at Abasic Sites in DNA. *J. Biol. Chem.* **1998**, *273* (14), 7888–7893. https://doi.org/10.1074/jbc.273.14.7888.

(170) Lohman, G. J. S.; Chen, L.; Evans, T. C. Kinetic Characterization of Single Strand Break Ligation in Duplex DNA by T4 DNA Ligase. *J. Biol. Chem.* **2011**, *286* (51), 44187–44196. https://doi.org/10.1074/jbc.M111.284992.

(171) Mizoue, L. S.; Tellinghuisen, J. The Role of Backlash in the "First Injection Anomaly" in Isothermal Titration Calorimetry. *Anal Biochem* **2004**, *326* (1), 125–127. https://doi.org/10.1016/j.ab.2003.10.048.

(172) Zamenhof, S.; Alexander, H. E.; Leidy, G. STUDIES ON THE CHEMISTRY OF THE TRANSFORMING ACTIVITY. *J. Exp. Med.* **1953**, *98* (4), 373–397. https://doi.org/10.1084/jem.98.4.373.

(173) Dahlmann, H. A.; Vaidyanathan, V. G.; Sturla, S. J. Investigating the Biochemical Impact of DNA Damage with Structure-Based Probes: Abasic Sites, Photodimers, Alkylation Adducts, and Oxidative Lesions. *Biochemistry* **2009**, *48* (40), 9347–9359. https://doi.org/10.1021/bi901059k.

(174) Hecht, S. S. Tobacco Smoke Carcinogens and Lung Cancer. *J. Natl. Cancer Inst.* **1999**, *91* (14), 17.

(175) Hecht, S. S. Lung Carcinogenesis by Tobacco Smoke. *Int. J. Cancer* **2012**, *131* (12), 2724–2732. https://doi.org/10.1002/ijc.27816.

(176) Pfeifer, G. P.; Denissenko, M. F.; Olivier, M.; Tretyakova, N.; Hecht, S. S.; Hainaut, P. Tobacco Smoke Carcinogens, DNA Damage and P53 Mutations in Smoking-Associated Cancers. *Oncogene* **2002**, *21* (48), 7435–7451. https://doi.org/10.1038/sj.onc.1205803.

(177) Friedberg, E. C. DNA Damage and Repair. *Nature* **2003**, *421* (6921), 436–440. https://doi.org/10.1038/nature01408.

(178) Friedberg, E. C. A Brief History of the DNA Repair Field. *Cell Res.* **2008**, *18* (1), 3–7. https://doi.org/10.1038/cr.2007.113.

(179) Gillet, L. C. J.; Schärer, O. D. Molecular Mechanisms of Mammalian Global Genome Nucleotide Excision Repair. *Chem. Rev.* **2006**, *106* (2), 253–276. https://doi.org/10.1021/cr040483f.

(180) David, S. S.; O'Shea, V. L.; Kundu, S. Base-Excision Repair of Oxidative DNA Damage. *Nature* **2007**, *447* (7147), 941–950. https://doi.org/10.1038/nature05978.

(181) Lindahl, T.; Nyberg, B. Rate of Depurination of Native Deoxyribonucleic Acid. *Biochemistry* **1972**, *11* (19), 3610–3618. https://doi.org/10.1021/bi00769a018.

(182) Lindahl, T. Instability and Decay of the Primary Structure of DNA. *Nature* **1993**, *362* (6422), 709–715. https://doi.org/10.1038/362709a0.

(183) Nakamura, J.; Walker, V. E.; Upton, P. B.; Chiang, S.-Y.; Kow, Y. W.; Swenberg, J. A. Highly Sensitive Apurinic/Apyrimidinic Site Assay Can Detect Spontaneous and

Chemically Induced Depurination under Physiological Conditions. *Cancer Res.* **1998**, *58* (2), 222.

(184) Nakamura, J.; Swenberg, J. A. Endogenous Apurinic/Apyrimidinic Sites in Genomic DNA of Mammalian Tissues. *Cancer Res.* **1999**, *59* (11), 2522.

(185) Nakamura, J.; La, D. K.; Swenberg, J. A. 5′-Nicked Apurinic/Apyrimidinic Sites Are Resistant to β-Elimination by β-Polymerase and Are Persistent in Human Cultured Cells after Oxidative Stress. *J. Biol. Chem.* **2000**, *275* (8), 5323–5328. https://doi.org/10.1074/jbc.275.8.5323.

(186) Tubbs, A.; Nussenzweig, A. Endogenous DNA Damage as a Source of Genomic Instability in Cancer. *Cell* **2017**, *168* (4), 644–656. https://doi.org/10.1016/j.cell.2017.01.002.

(187) Wang, Y.; Liu, L.; Wu, C.; Bulgar, A.; Somoza, E.; Zhu, W.; Gerson, S. L. Direct Detection and Quantification of Abasic Sites for in Vivo Studies of DNA Damage and Repair. *Nucl. Med. Biol.* **2009**, *36* (8), 975–983. https://doi.org/10.1016/j.nucmedbio.2009.07.007.

(188) Kidane, D.; Murphy, D. L.; Sweasy, J. B. Accumulation of Abasic Sites Induces Genomic Instability in Normal Human Gastric Epithelial Cells during Helicobacter Pylori Infection. *Oncogenesis* **2014**, *3* (11), e128–e128. https://doi.org/10.1038/oncsis.2014.42.

(189) McCullough, A. K.; Dodson, M. L.; Lloyd, R. S. Initiation of Base Excision Repair: Glycosylase Mechanisms and Structures. *Annu. Rev. Biochem.* **1999**, *68* (1), 255–285. https://doi.org/10.1146/annurev.biochem.68.1.255.

(190) Jacobs, A. L.; Schär, P. DNA Glycosylases: In DNA Repair and Beyond. *Chromosoma* **2012**, *121* (1), 1–20. https://doi.org/10.1007/s00412-011-0347-4.

(191) Krokan, H. E.; Bjoras, M. Base Excision Repair. *Cold Spring Harb. Perspect. Biol.* **2013**, *5* (4), a012583–a012583. https://doi.org/10.1101/cshperspect.a012583.

(192) Schaaper, R. M.; Kunkel, T. A.; Loeb, L. A. Infidelity of DNA Synthesis Associated with Bypass of Apurinic Sites. *Proc. Natl. Acad. Sci.* **1983**, *80* (2), 487–491. https://doi.org/10.1073/pnas.80.2.487.

(193) Kunkel, T. A. Mutational Specificity of Depurination. *Proc. Natl. Acad. Sci.* **1984**, *81* (5), 1494–1498. https://doi.org/10.1073/pnas.81.5.1494.

(194) Loeb, L. A.; Preston, B. D. Mutagenesis by Apurinic/Apyrimidinic Sites. *Annu. Rev. Genet.* **1986**, *20* (1), 201–230. https://doi.org/10.1146/annurev.ge.20.120186.001221.

(195) Boiteux, S.; Guillet, M. Abasic Sites in DNA: Repair and Biological Consequences in Saccharomyces Cerevisiae. *DNA Repair* **2004**, *3* (1), 1–12. https://doi.org/10.1016/j.dnarep.2003.10.002.

(196) Kubo, K.; Ide, H.; Wallace, S. S.; Kow, Y. W. A Novel Sensitive and Specific Assay for Abasic Sites, the Most Commonly Produced DNA Lesion. *Biochemistry* **1992**, *31* (14), 3703–3708. https://doi.org/10.1021/bi00129a020.

(197) Atamna, H.; Cheung, I.; Ames, B. N. A Method for Detecting Abasic Sites in Living Cells: Age-Dependent Changes in Base Excision Repair. *Proc. Natl. Acad. Sci.* **2000**, *97* (2), 686–691. https://doi.org/10.1073/pnas.97.2.686.

(198) Asaeda, A.; Ide, H.; Terato, H.; Takamori, Y.; Kubo, K. Highly Sensitive Assay of DNA Abasic Sites in Mammalian Cells-Optimization of the Aldehyde Reactive Probe Method. *Anal. Chim. Acta* **1998**, *365* (1–3), 35–41. https://doi.org/10.1016/S0003-2670(97)00648-X.

(199) Kurisu, S.; Miya, T.; Terato, H.; Masaoka, A.; Ohyama, Y.; Kubo, K.; Ide, H. Quantitation of DNA Damage by an Aldehyde Reactive Probe (ARP). *Nucleic Acids Symp. Ser.* **2001**, *1* (1), 45–46. https://doi.org/10.1093/nass/1.1.45.

(200) Ide, H.; Akamatsu, K.; Kimura, Y.; Michiue, K.; Makino, K.; Asaeda, A.; Takamori, Y.; Kubo, K. Synthesis and Damage Specificity of a Novel Probe for the Detection of Abasic Sites in DNA. *Biochemistry* **1993**, *32* (32), 8276–8283. https://doi.org/10.1021/bi00083a031.

(201) Poetsch, A. R.; Boulton, S. J.; Luscombe, N. M. Genomic Landscape of Oxidative DNA Damage and Repair Reveals Regioselective Protection from Mutagenesis. *Genome Biol.* **2018**, *19* (1), 215. https://doi.org/10.1186/s13059-018-1582-2.

(202) Raiber, E.-A.; Beraldi, D.; Ficz, G.; Burgess, H. E.; Branco, M. R.; Murat, P.; Oxley, D.; Booth, M. J.; Reik, W.; Balasubramanian, S. Genome-Wide Distribution of 5-Formylcytosine in Embryonic Stem Cells Is Associated with Transcription and Depends on Thymine DNA Glycosylase. *Genome Biol.* **2012**, *13* (8), R69. https://doi.org/10.1186/gb-2012-13-8-r69.

(203) Hardisty, R. E.; Kawasaki, F.; Sahakyan, A. B.; Balasubramanian, S. Selective Chemical Labeling of Natural T Modifications in DNA. *J. Am. Chem. Soc.* **2015**, *137* (29), 9270–9272. https://doi.org/10.1021/jacs.5b03730.

(204) Liu, Z. J.; Martínez Cuesta, S.; van Delft, P.; Balasubramanian, S. Sequencing Abasic Sites in DNA at Single-Nucleotide Resolution. *Nat. Chem.* **2019**, *11* (7), 629–637. https://doi.org/10.1038/s41557-019-0279-9.

(205) Cao, B.; Wu, X.; Zhou, J.; Wu, H.; Liu, L.; Zhang, Q.; DeMott, M. S.; Gu, C.; Wang, L.; You, D.; et al. Nick-Seq for Single-Nucleotide Resolution Genomic Maps of DNA Modifications and Damage. *Nucleic Acids Res.* **2020**, *48* (12), 6715–6725. https://doi.org/10.1093/nar/gkaa473.

(206) Benner, S. A.; Sismour, A. M. Synthetic Biology. *Nat. Rev. Genet.* **2005**, *6* (7), 533–543. https://doi.org/10.1038/nrg1637.

(207) Henry, A. Beyond A, C, G and T: Augmenting Nature's Alphabet. *Curr. Opin. Chem. Biol.* **2003**, *7* (6), 727–733. https://doi.org/10.1016/j.cbpa.2003.10.011.

(208) Malyshev, D. A.; Dhami, K.; Quach, H. T.; Lavergne, T.; Ordoukhanian, P.; Torkamani, A.; Romesberg, F. E. Efficient and Sequence-Independent Replication of DNA Containing a Third Base Pair Establishes a Functional Six-Letter Genetic Alphabet. *Proc. Natl. Acad. Sci.* **2012**, *109* (30), 12005–12010. https://doi.org/10.1073/pnas.1205176109.

(209) Kawai, R.; Kimoto, M.; Ikeda, S.; Mitsui, T.; Endo, M.; Yokoyama, S.; Hirao, I. Site-Specific Fluorescent Labeling of RNA Molecules by Specific Transcription Using Unnatural Base Pairs. *J. Am. Chem. Soc.* **2005**, *127* (49), 17286–17295. https://doi.org/10.1021/ja0542946.

(210) Seo, Y. J.; Malyshev, D. A.; Lavergne, T.; Ordoukhanian, P.; Romesberg, F. E. Site-Specific Labeling of DNA and RNA Using an Efficiently Replicated and Transcribed Class of Unnatural Base Pairs. *J. Am. Chem. Soc.* **2011**, *133* (49), 19878–19888. https://doi.org/10.1021/ja207907d.

(211) Kimoto, M.; Cox, R. S.; Hirao, I. Unnatural Base Pair Systems for Sensing and Diagnostic Applications. *Expert Rev. Mol. Diagn.* **2011**, *11* (3), 321–331. https://doi.org/10.1586/erm.11.5.

(212) Piccirilli, J. A.; Benner, S. A.; Krauch, T.; Moroney, S. E.; Benner, S. A. Enzymatic Incorporation of a New Base Pair into DNA and RNA Extends the Genetic Alphabet. *Nature* **1990**, *343* (6253), 33–37. https://doi.org/10.1038/343033a0.

(213) Yang, Z.; Chen, F.; Chamberlin, S. G.; Benner, S. A. Expanded Genetic Alphabets in the Polymerase Chain Reaction. *Angew. Chem. Int. Ed.* **2010**, *49* (1), 177–180. https://doi.org/10.1002/anie.200905173.

(214) Yang, Z.; Chen, F.; Alvarado, J. B.; Benner, S. A. Amplification, Mutation, and Sequencing of a Six-Letter Synthetic Genetic System. *J. Am. Chem. Soc.* **2011**, *133* (38), 15105–15112. https://doi.org/10.1021/ja204910n.

(215) Matray, T. J.; Kool, E. T. A Specific Partner for Abasic Damage in DNA. *Nature* **1999**, *399* (6737), 704–708. https://doi.org/10.1038/21453.

(216) Matray, T. J.; Kool, E. T. Selective and Stable DNA Base Pairing without Hydrogen Bonds. *J. Am. Chem. Soc.* **1998**, *120* (24), 6191–6192. https://doi.org/10.1021/ja9803310.

(217) Moran, S.; Ren, R. X.-F.; Rumney; Kool, E. T. Difluorotoluene, a Nonpolar Isostere for Thymine, Codes Specifically and Efficiently for Adenine in DNA Replication. *J. Am. Chem. Soc.* **1997**, *119* (8), 2056–2057. https://doi.org/10.1021/ja963718g.

(218) McMinn, D. L.; Ogawa, A. K.; Wu, Y.; Liu, J.; Schultz, P. G.; Romesberg, F. E. Efforts toward Expansion of the Genetic Alphabet: DNA Polymerase Recognition of a Highly Stable, Self-Pairing Hydrophobic Base. *J. Am. Chem. Soc.* **1999**, *121* (49), 11585–11586. https://doi.org/10.1021/ja9925150.

(219) Hirao, I.; Mitsui, T.; Kimoto, M.; Yokoyama, S. An Efficient Unnatural Base Pair for PCR Amplification. *J. Am. Chem. Soc.* **2007**, *129* (50), 15549–15555. https://doi.org/10.1021/ja073830m.

(220) Yamashige, R.; Kimoto, M.; Takezawa, Y.; Sato, A.; Mitsui, T.; Yokoyama, S.; Hirao, I. Highly Specific Unnatural Base Pair Systems as a Third Base Pair for PCR Amplification. *Nucleic Acids Res.* **2012**, *40* (6), 2793–2806. https://doi.org/10.1093/nar/gkr1068.

(221) Zhang, Y.; Ptacin, J. L.; Fischer, E. C.; Aerni, H. R.; Caffaro, C. E.; San Jose, K.; Feldman, A. W.; Turner, C. R.; Romesberg, F. E. A Semi-Synthetic Organism That Stores and Retrieves Increased Genetic Information. *Nature* **2017**, *551* (7682), 644–647. https://doi.org/10.1038/nature24659.

(222) Riedl, J.; Ding, Y.; Fleming, A. M.; Burrows, C. J. Identification of DNA Lesions Using a Third Base Pair for Amplification and Nanopore Sequencing. *Nat. Commun.* **2015**, *6* (1), 8807. https://doi.org/10.1038/ncomms9807.

(223) Moser, M. J. Enzymatic Repair of an Expanded Genetic Information System. *Nucleic Acids Res.* **2003**, *31* (17), 5048–5053. https://doi.org/10.1093/nar/gkg709.

(224) Ren, R. X.-F.; Chaudhuri, N. C.; Paris, P. L.; Rumney; Kool, E. T. Naphthalene, Phenanthrene, and Pyrene as DNA Base Analogues: Synthesis, Structure, and Fluorescence in DNA. *J. Am. Chem. Soc.* **1996**, *118* (33), 7671–7678. https://doi.org/10.1021/ja9612763.

(225) Dzantiev, L.; Alekseyev, Y. O.; Morales, J. C.; Kool, E. T.; Romano, L. J. Significance of Nucleobase Shape Complementarity and Hydrogen Bonding in the Formation and Stability of the Closed Polymerase−DNA Complex. *Biochemistry* **2001**, *40* (10), 3215–3221. https://doi.org/10.1021/bi002569i.

(226) Singh, I.; Beuck, C.; Bhattacharya, A.; Hecker, W.; Parmar, V. S.; Weinhold, E.; Seitz, O. Abasic site stabilization by aromatic DNA base surrogates: High-affinity binding to a base-flipping DNA-methyltransferase. *Pure Appl. Chem.* **2004**, *76* (7–8), 1563–1570. https://doi.org/10.1351/pac200476071563.

(227) Luo, J. Improving the Fidelity of Thermus Thermophilus DNA Ligase. *Nucleic Acids Res.* **1996**, *24* (15), 3071–3078. https://doi.org/10.1093/nar/24.15.3071.

(228) Liu, P. DNA Ligases Ensure Fidelity by Interrogating Minor Groove Contacts. *Nucleic Acids Res.* **2004**, *32* (15), 4503–4511. https://doi.org/10.1093/nar/gkh781.

(229) Zhao, X.; Muller, J. G.; Halasyam, M.; David, S. S.; Burrows, C. J. In Vitro Ligation of Oligodeoxynucleotides Containing C8-Oxidized Purine Lesions Using Bacteriophage T4 DNA Ligase. *Biochemistry* **2007**, *46* (12), 3734–3744. https://doi.org/10.1021/bi062214k.

(230) Kimoto, M.; Soh, S. H. G.; Tan, H. P.; Okamoto, I.; Hirao, I. Cognate Base-Pair Selectivity of Hydrophobic Unnatural Bases in DNA Ligation by T4 DNA Ligase. *Biopolymers* **2021**, *112* (1), e23407. https://doi.org/10.1002/bip.23407.

(231) Jang, E. K.; Yang, M.; Pack, S. P. Highly-Efficient T4 DNA Ligase-Based SNP Analysis Using a Ligation Fragment Containing a Modified Nucleobase at the End. *Chem. Commun.* **2015**, *51* (66), 13090–13093. https://doi.org/10.1039/C5CC03761A.

(232) Lohman, G. J. S.; Zhang, Y.; Zhelkovsky, A. M.; Cantor, E. J.; Evans, T. C. Efficient DNA Ligation in DNA–RNA Hybrid Helices by Chlorella Virus DNA Ligase. *Nucleic Acids Res.* **2014**, *42* (3), 1831–1844. https://doi.org/10.1093/nar/gkt1032.

(233) Jin, J.; Vaud, S.; Zhelkovsky, A. M.; Posfai, J.; McReynolds, L. A. Sensitive and Specific MiRNA Detection Method Using SplintR Ligase. *Nucleic Acids Res.* **2016**, *44* (13), e116–e116. https://doi.org/10.1093/nar/gkw399.

(234) Wee, E. J. H.; Trau, M. Simple Isothermal Strategy for Multiplexed, Rapid, Sensitive, and Accurate MiRNA Detection. *ACS Sens.* **2016**, *1* (6), 670–675. https://doi.org/10.1021/acssensors.6b00105.

(235) Krzywkowski, T.; Nilsson, M. Fidelity of RNA Templated End-Joining by Chlorella Virus DNA Ligase and a Novel ILock Assay with Improved Direct RNA Detection Accuracy. *Nucleic Acids Res.* **2017**, *45* (18), e161–e161. https://doi.org/10.1093/nar/gkx708.

(236) Woo, C. H.; Jang, S.; Shin, G.; Jung, G. Y.; Lee, J. W. Sensitive Fluorescence Detection of SARS-CoV-2 RNA in Clinical Samples via One-Pot Isothermal Ligation and Transcription. *Nat. Biomed. Eng.* **2020**, *4* (12), 1168–1179. https://doi.org/10.1038/s41551-020-00617-5.

(237) Bauer, R. J.; Zhelkovsky, A.; Bilotti, K.; Crowell, L. E.; Evans, T. C.; McReynolds, L. A.; Lohman, G. J. S. Comparative Analysis of the End-Joining Activity of Several DNA Ligases. *PLOS ONE* **2017**, *12* (12), e0190062. https://doi.org/10.1371/journal.pone.0190062.

(238) Wojciechowski, F.; Lietard, J.; Leumann, C. J. 2-Pyrenyl-DNA: Synthesis, Pairing, and Fluorescence Properties. *Org. Lett.* **2012**, *14* (20), 5176–5179. https://doi.org/10.1021/ol302150a.

(239) Jiang, Y. L.; Stivers, J. T. Efficient Epimerization of Pyrene and Other Aromatic C-Nucleosides with Trifluoroacetic Acid in Dichloromethane. *Tetrahedron Lett.* **2003**, *44* (1), 85–88. https://doi.org/10.1016/S0040-4039(02)02481-4.

(240) Osman, E. A.; Alladin-Mustan, B. S.; Hales, S. C.; Matharu, G. K.; Gibbs, J. M. Enhanced Mismatch Selectivity of T4 DNA Ligase Far above the Probe: Target Duplex Dissociation Temperature. *Biopolymers* **2021**, *112* (1). https://doi.org/10.1002/bip.23393.

(241) Cherepanov, A. V.; De Vries, S. Kinetics and Thermodynamics of Nick Sealing by T4 DNA Ligase: Kinetics of Nick Sealing by T4 DNA Ligase. *Eur. J. Biochem.* **2003**, *270* (21), 4315–4325. https://doi.org/10.1046/j.1432-1033.2003.03824.x.

(242) Lei, Y.; Hili, R. Structure–Activity Relationships of the ATP Cofactor in Ligase-Catalysed Oligonucleotide Polymerisations. *Org. Biomol. Chem.* **2017**, *15* (11), 2349–2352. https://doi.org/10.1039/C6OB02792J.

(243) Xu, W.; Xie, X.; Li, D.; Yang, Z.; Li, T.; Liu, X. Ultrasensitive Colorimetric DNA Detection Using a Combination of Rolling Circle Amplification and Nicking Endonuclease-Assisted Nanoparticle Amplification (NEANA). *Small* **2012**, *8* (12), 1846–1850. https://doi.org/10.1002/smll.201200263.

(244) Cheng, Y.; Zhao, J.; Jia, H.; Yuan, Z.; Li, Z. Ligase Chain Reaction Coupled with Rolling Circle Amplification for High Sensitivity Detection of Single Nucleotide Polymorphisms. *The Analyst* **2013**, *138* (10), 2958. https://doi.org/10.1039/c3an36920j.

(245) Zhang, Y.; Guo, Y.; Quirke, P.; Zhou, D. Ultrasensitive Single-Nucleotide Polymorphism Detection Using Target-Recycled Ligation, Strand Displacement and Enzymatic Amplification. *Nanoscale* **2013**, *5* (11), 5027. https://doi.org/10.1039/c3nr01010d.

(246) Lee, S. H.; Wang, S.; Kool, E. T. Templated Chemistry for Monitoring Damage and Repair Directly in Duplex DNA. *Chem. Commun.* **2012**, *48* (65), 8069. https://doi.org/10.1039/c2cc34060g.

(247) Jinek, M.; East, A.; Cheng, A.; Lin, S.; Ma, E.; Doudna, J. RNA-Programmed Genome Editing in Human Cells. *eLife* **2013**, *2*, e00471. https://doi.org/10.7554/eLife.00471.

(248) Lin, S.; Staahl, B. T.; Alla, R. K.; Doudna, J. A. Enhanced Homology-Directed Human Genome Engineering by Controlled Timing of CRISPR/Cas9 Delivery. *eLife* **2014**, *3*, e04766. https://doi.org/10.7554/eLife.04766.

(249) Tuladhar, R.; Yeu, Y.; Tyler Piazza, J.; Tan, Z.; Rene Clemenceau, J.; Wu, X.; Barrett, Q.; Herbert, J.; Mathews, D. H.; Kim, J.; et al. CRISPR-Cas9-Based Mutagenesis Frequently Provokes on-Target MRNA Misregulation. *Nat. Commun.* **2019**, *10* (1), 4056. https://doi.org/10.1038/s41467-019-12028-5.

(250) Cromwell, C. R.; Jovel, J.; Hubbard, B. P. Methods for Measuring CRISPR/Cas9 DNA Cleavage in Cells. In *CRISPR Guide RNA Design*; Fulga, T. A., Knapp, D. J. H. F., Ferry, Q. R. V., Eds.; Methods in Molecular Biology; Springer US: New York, NY, 2021; Vol. 2162, pp 197–213. https://doi.org/10.1007/978-1-0716-0687-2_11.

(251) Cromwell, C. R.; Hubbard, B. P. In Vitro Assays for Comparing the Specificity of First- and Next-Generation CRISPR/Cas9 Systems. In *CRISPR Guide RNA Design*; Fulga, T. A., Knapp, D. J. H. F., Ferry, Q. R. V., Eds.; Methods in Molecular Biology; Springer US: New York, NY, 2021; Vol. 2162, pp 215–232. https://doi.org/10.1007/978-1-0716-0687-2_12.

(252) Wang, T.; Wei, J. J.; Sabatini, D. M.; Lander, E. S. Genetic Screens in Human Cells Using the CRISPR-Cas9 System. *Science* **2014**, *343* (6166), 80–84. https://doi.org/10.1126/science.1246981.

(253) Shi, J.; Wang, E.; Milazzo, J. P.; Wang, Z.; Kinney, J. B.; Vakoc, C. R. Discovery of Cancer Drug Targets by CRISPR-Cas9 Screening of Protein Domains. *Nat. Biotechnol.* **2015**, *33* (6), 661–667. https://doi.org/10.1038/nbt.3235.

(254) He, W.; Zhang, L.; Villarreal, O. D.; Fu, R.; Bedford, E.; Dou, J.; Patel, A. Y.; Bedford, M. T.; Shi, X.; Chen, T.; et al. De Novo Identification of Essential Protein Domains from CRISPR-Cas9 Tiling-SgRNA Knockout Screens. *Nat. Commun.* **2019**, *10* (1), 4541. https://doi.org/10.1038/s41467-019-12489-8.

(255) Schwank, G.; Koo, B.-K.; Sasselli, V.; Dekkers, J. F.; Heo, I.; Demircan, T.; Sasaki, N.; Boymans, S.; Cuppen, E.; van der Ent, C. K.; et al. Functional Repair of CFTR by CRISPR/Cas9 in Intestinal Stem Cell Organoids of Cystic Fibrosis Patients. *Cell Stem Cell* **2013**, *13* (6), 653–658. https://doi.org/10.1016/j.stem.2013.11.002.

(256) Hsu, P. D.; Lander, E. S.; Zhang, F. Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell* **2014**, *157* (6), 1262–1278. https://doi.org/10.1016/j.cell.2014.05.010.

(257) Ye, L.; Wang, J.; Tan, Y.; Beyer, A. I.; Xie, F.; Muench, M. O.; Kan, Y. W. Genome Editing Using CRISPR-Cas9 to Create the HPFH Genotype in HSPCs: An Approach for Treating Sickle Cell Disease and β-Thalassemia. *Proc. Natl. Acad. Sci.* **2016**, *113* (38), 10661–10665. https://doi.org/10.1073/pnas.1612075113.

(258) DeWitt, M. A.; Magis, W.; Bray, N. L.; Wang, T.; Berman, J. R.; Urbinati, F.; Heo, S.-J.; Mitros, T.; Munoz, D. P.; Boffelli, D.; et al. Selection-Free Genome Editing of the Sickle Mutation in Human Adult Hematopoietic Stem/Progenitor Cells. *Sci. Transl. Med.* **2016**, *8* (360), 360ra134-360ra134. https://doi.org/10.1126/scitranslmed.aaf9336.

(259) Leonetti, M. D.; Sekine, S.; Kamiyama, D.; Weissman, J. S.; Huang, B. A Scalable Strategy for High-Throughput GFP Tagging of Endogenous Human Proteins. *Proc. Natl. Acad. Sci.* **2016**, *113* (25), E3501–E3508. https://doi.org/10.1073/pnas.1606731113.

(260) McCarty, N. S.; Graham, A. E.; Studená, L.; Ledesma-Amaro, R. Multiplexed CRISPR Technologies for Gene Editing and Transcriptional Regulation. *Nat. Commun.* **2020**, *11* (1), 1281. https://doi.org/10.1038/s41467-020-15053-x.

(261) Wang, H.; Yang, H.; Shivalila, C. S.; Dawlaty, M. M.; Cheng, A. W.; Zhang, F.; Jaenisch, R. One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering. *Cell* **2013**, *153* (4), 910–918. https://doi.org/10.1016/j.cell.2013.04.025.

(262) Jiang, Y.; Chen, B.; Duan, C.; Sun, B.; Yang, J.; Yang, S. Multigene Editing in the Escherichia Coli Genome via the CRISPR-Cas9 System. *Appl. Environ. Microbiol.* **2015**, *81* (7), 2506–2514. https://doi.org/10.1128/AEM.04023-14.

(263) Sekine, R.; Kawata, T.; Muramoto, T. CRISPR/Cas9 Mediated Targeting of Multiple Genes in Dictyostelium. *Sci. Rep.* **2018**, *8* (1), 8471. https://doi.org/10.1038/s41598-018-26756-z.

(264) Kabadi, A. M.; Ousterout, D. G.; Hilton, I. B.; Gersbach, C. A. Multiplex CRISPR/Cas9-Based Genome Engineering from a Single Lentiviral Vector. *Nucleic Acids Res.* **2014**, *42* (19), e147–e147. https://doi.org/10.1093/nar/gku749.

(265) Fu, Y.; Rocha, P. P.; Luo, V. M.; Raviram, R.; Deng, Y.; Mazzoni, E. O.; Skok, J. A. CRISPR-DCas9 and SgRNA Scaffolds Enable Dual-Colour Live Imaging of Satellite Sequences and Repeat-Enriched Individual Loci. *Nat. Commun.* **2016**, *7* (1), 11707. https://doi.org/10.1038/ncomms11707.

(266) Lim, H.; Choi, S.-K. Programmed GRNA Removal System for CRISPR-Cas9-Mediated Multi-Round Genome Editing in Bacillus Subtilis. *Front. Microbiol.* **2019**, *10*, 1140. https://doi.org/10.3389/fmicb.2019.01140.

(267) Zetsche, B.; Heidenreich, M.; Mohanraju, P.; Fedorova, I.; Kneppers, J.; DeGennaro, E. M.; Winblad, N.; Choudhury, S. R.; Abudayyeh, O. O.; Gootenberg, J. S.; et al. Multiplex Gene Editing by CRISPR–Cpf1 Using a Single CrRNA Array. *Nat. Biotechnol.* **2017**, *35* (1), 31–34. https://doi.org/10.1038/nbt.3737.

(268) Kim, S.; Kim, D.; Cho, S. W.; Kim, J.; Kim, J.-S. Highly Efficient RNA-Guided Genome Editing in Human Cells via Delivery of Purified Cas9 Ribonucleoproteins. *Genome Res.* **2014**, *24* (6), 1012–1019. https://doi.org/10.1101/gr.171322.113.

(269) Mir, A.; Alterman, J. F.; Hassler, M. R.; Debacker, A. J.; Hudgens, E.; Echeverria, D.; Brodsky, M. H.; Khvorova, A.; Watts, J. K.; Sontheimer, E. J. Heavily and Fully Modified RNAs Guide Efficient SpyCas9-Mediated Genome Editing. *Nat. Commun.* **2018**, *9* (1), 2641. https://doi.org/10.1038/s41467-018-05073-z.

(270) Hendel, A.; Bak, R. O.; Clark, J. T.; Kennedy, A. B.; Ryan, D. E.; Roy, S.; Steinfeld, I.; Lunstad, B. D.; Kaiser, R. J.; Wilkens, A. B.; et al. Chemically Modified Guide RNAs Enhance CRISPR-Cas Genome Editing in Human Primary Cells. *Nat. Biotechnol.* **2015**, *33* (9), 985–989. https://doi.org/10.1038/nbt.3290.

(271) Basila, M.; Kelley, M. L.; Smith, A. van B. Minimal 2'-O-Methyl Phosphorothioate Linkage Modification Pattern of Synthetic Guide RNAs for Increased Stability and Efficient CRISPR-Cas9 Gene Editing Avoiding Cellular Toxicity. *PLOS ONE* **2017**, *12* (11), e0188593. https://doi.org/10.1371/journal.pone.0188593.

(272) Ryan, D. E.; Taussig, D.; Steinfeld, I.; Phadnis, S. M.; Lunstad, B. D.; Singh, M.; Vuong, X.; Okochi, K. D.; McCaffrey, R.; Olesiak, M.; et al. Improving CRISPR–Cas Specificity with Chemical Modifications in Single-Guide RNAs. *Nucleic Acids Res.* **2017**, *46* (2), 792–803. https://doi.org/10.1093/nar/gkx1199.

(273) Cromwell, C. R.; Sung, K.; Park, J.; Krysler, A. R.; Jovel, J.; Kim, S. K.; Hubbard, B. P. Incorporation of Bridged Nucleic Acids into CRISPR RNAs Improves Cas9 Endonuclease Specificity. *Nat. Commun.* **2018**, *9* (1), 1448. https://doi.org/10.1038/s41467-018-03927-0.

(274) Lee, K.; Mackley, V. A.; Rao, A.; Chong, A. T.; Dewitt, M. A.; Corn, J. E.; Murthy, N. Synthetically Modified Guide RNA and Donor DNA Are a Versatile Platform for CRISPR-Cas9 Engineering. *eLife* **2017**, *6*, e25312. https://doi.org/10.7554/eLife.25312.

(275) Palumbo, C. M.; Gutierrez-Bujari, J. M.; O'Geen, H.; Segal, D. J.; Beal, P. A. Versatile 3′ Functionalization of CRISPR Single Guide RNA. *ChemBioChem* **2020**, *21* (11), 1633–1640. https://doi.org/10.1002/cbic.201900736.

(276) Wang, H.-X.; Song, Z.; Lao, Y.-H.; Xu, X.; Gong, J.; Cheng, D.; Chakraborty, S.; Park, J. S.; Li, M.; Huang, D.; et al. Nonviral Gene Editing via CRISPR/Cas9 Delivery by Membrane-Disruptive and Endosomolytic Helical Polypeptide. *Proc. Natl. Acad. Sci.* **2018**, *115* (19), 4903–4908. https://doi.org/10.1073/pnas.1712963115.

(277) Wei, T.; Cheng, Q.; Min, Y.-L.; Olson, E. N.; Siegwart, D. J. Systemic Nanoparticle Delivery of CRISPR-Cas9 Ribonucleoproteins for Effective Tissue Specific Genome Editing. *Nat. Commun.* **2020**, *11* (1), 3232. https://doi.org/10.1038/s41467-020-17029-3.

(278) Khan, F. J.; Yuen, G.; Luo, J. Multiplexed CRISPR/Cas9 Gene Knockout with Simple CrRNA:TracrRNA Co-Transfection. *Cell Biosci.* **2019**, *9* (1), 41. https://doi.org/10.1186/s13578-019-0304-0.

(279) Rahdar, M.; McMahon, M. A.; Prakash, T. P.; Swayze, E. E.; Bennett, C. F.; Cleveland, D. W. Synthetic CRISPR RNA-Cas9–Guided Genome Editing in Human Cells. *Proc. Natl. Acad. Sci.* **2015**, *112* (51), E7110–E7117. https://doi.org/10.1073/pnas.1520883112.

(280) Karvelis, T.; Gasiunas, G.; Miksys, A.; Barrangou, R.; Horvath, P.; Siksnys, V. CrRNA and TracrRNA Guide Cas9-Mediated DNA Interference in Streptococcus Thermophilus. *RNA Biol.* **2013**, *10* (5), 841–851. https://doi.org/10.4161/rna.24203.

(281) Seki, A.; Rutz, S. Optimized RNP Transfection for Highly Efficient CRISPR/Cas9-Mediated Gene Knockout in Primary T Cells. *J. Exp. Med.* **2018**, *215* (3), 985–997. https://doi.org/10.1084/jem.20171626.

(282) Nasri, M.; Mir, P.; Dannenmann, B.; Amend, D.; Skroblyn, T.; Xu, Y.; Schulze-Osthoff, K.; Klimiankou, M.; Welte, K.; Skokowa, J. Fluorescent Labeling of CRISPR/Cas9 RNP for Gene Knockout in HSPCs and IPSCs Reveals an Essential Role for GADD45b in Stress Response. *Blood Adv.* **2019**, *3* (1), 63–71. https://doi.org/10.1182/bloodadvances.2017015511.

(283) Nishimasu, H.; Ran, F. A.; Hsu, P. D.; Konermann, S.; Shehata, S. I.; Dohmae, N.; Ishitani, R.; Zhang, F.; Nureki, O. Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell* **2014**, *156* (5), 935–949. https://doi.org/10.1016/j.cell.2014.02.001.

(284) Finn, J. D.; Smith, A. R.; Patel, M. C.; Shaw, L.; Youniss, M. R.; van Heteren, J.; Dirstine, T.; Ciullo, C.; Lescarbeau, R.; Seitzer, J.; et al. A Single Administration of CRISPR/Cas9 Lipid Nanoparticles Achieves Robust and Persistent In Vivo Genome Editing. *Cell Rep.* **2018**, *22* (9), 2227–2235. https://doi.org/10.1016/j.celrep.2018.02.014.

(285) Yin, H.; Song, C.-Q.; Suresh, S.; Wu, Q.; Walsh, S.; Rhym, L. H.; Mintzer, E.; Bolukbasi, M. F.; Zhu, L. J.; Kauffman, K.; et al. Structure-Guided Chemical Modification of Guide RNA Enables Potent Non-Viral in Vivo Genome Editing. *Nat. Biotechnol.* **2017**, *35* (12), 1179–1187. https://doi.org/10.1038/nbt.4005.

(286) Dellinger, D. J.; Timár, Z.; Myerson, J.; Sierzchala, A. B.; Turner, J.; Ferreira, F.; Kupihár, Z.; Dellinger, G.; Hill, K. W.; Powell, J. A.; et al. Streamlined Process for the Chemical Synthesis of RNA Using 2′- *O* -Thionocarbamate-Protected Nucleoside Phosphoramidites in the Solid Phase. *J. Am. Chem. Soc.* **2011**, *133* (30), 11540–11556. https://doi.org/10.1021/ja201561z.

(287) El-Sagheer, A. H.; Brown, T. New Strategy for the Synthesis of Chemically Modified RNA Constructs Exemplified by Hairpin and Hammerhead Ribozymes. *Proc. Natl. Acad. Sci.* **2010**, *107* (35), 15329–15334. https://doi.org/10.1073/pnas.1006447107.

(288) He, K.; Chou, E. T.; Begay, S.; Anderson, E. M.; van Brabant Smith, A. Conjugation and Evaluation of Triazole-Linked Single Guide RNA for CRISPR-Cas9 Gene Editing. *ChemBioChem* **2016**, *17* (19), 1809–1812. https://doi.org/10.1002/cbic.201600320.

(289) Taemaitree, L.; Shivalingam, A.; El-Sagheer, A. H.; Brown, T. An Artificial Triazole Backbone Linkage Provides a Split-and-Click Strategy to Bioactive Chemically Modified CRISPR SgRNA. *Nat. Commun.* **2019**, *10* (1), 1610. https://doi.org/10.1038/s41467-019-09600-4.

(290) Lee, K.; Conboy, M.; Park, H. M.; Jiang, F.; Kim, H. J.; Dewitt, M. A.; Mackley, V. A.; Chang, K.; Rao, A.; Skinner, C.; et al. Nanoparticle Delivery of Cas9 Ribonucleoprotein and Donor DNA in Vivo Induces Homology-Directed DNA Repair. *Nat. Biomed. Eng.* **2017**, *1* (11), 889–901. https://doi.org/10.1038/s41551-017-0137-2.

(291) Deng, W.; Shi, X.; Tjian, R.; Lionnet, T.; Singer, R. H. CASFISH: CRISPR/Cas9-Mediated in Situ Labeling of Genomic Loci in Fixed Cells. *Proc. Natl. Acad. Sci.* **2015**, *112* (38), 11870–11875. https://doi.org/10.1073/pnas.1515692112.

(292) Briner, A. E.; Donohoue, P. D.; Gomaa, A. A.; Selle, K.; Slorach, E. M.; Nye, C. H.; Haurwitz, R. E.; Beisel, C. L.; May, A. P.; Barrangou, R. Guide RNA Functional Modules Direct Cas9 Activity and Orthogonality. *Mol. Cell* **2014**, *56* (2), 333–339. https://doi.org/10.1016/j.molcel.2014.09.019.

(293) Shechner, D. M.; Hacisuleyman, E.; Younger, S. T.; Rinn, J. L. Multiplexable, Locus-Specific Targeting of Long RNAs with CRISPR-Display. *Nat. Methods* **2015**, *12* (7), 664–670. https://doi.org/10.1038/nmeth.3433.

(294) Nishimasu, H.; Cong, L.; Yan, W. X.; Ran, F. A.; Zetsche, B.; Li, Y.; Kurabayashi, A.; Ishitani, R.; Zhang, F.; Nureki, O. Crystal Structure of Staphylococcus Aureus Cas9. *Cell* **2015**, *162* (5), 1113–1126. https://doi.org/10.1016/j.cell.2015.08.007.

(295) El-Sagheer, A. H.; Sanzone, A. P.; Gao, R.; Tavassoli, A.; Brown, T. Biocompatible Artificial DNA Linker That Is Read through by DNA Polymerases and Is Functional in Escherichia Coli. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (28), 11338–11343. https://doi.org/10.1073/pnas.1101519108.

(296) El-Sagheer, A. H.; Brown, T. Efficient RNA Synthesis by in Vitro Transcription of a Triazole-Modified DNA Template. *Chem. Commun.* **2011**, *47* (44), 12057. https://doi.org/10.1039/c1cc14316f.

(297) Rueda, F. O.; Bista, M.; Newton, M. D.; Goeppert, A. U.; Cuomo, M. E.; Gordon, E.; Kröner, F.; Read, J. A.; Wrigley, J. D.; Rueda, D.; et al. Mapping the Sugar Dependency for Rational Generation of a DNA-RNA Hybrid-Guided Cas9 Endonuclease. *Nat. Commun.* **2017**, *8* (1), 1610. https://doi.org/10.1038/s41467-017-01732-9.

(298) Yin, H.; Song, C.-Q.; Suresh, S.; Kwan, S.-Y.; Wu, Q.; Walsh, S.; Ding, J.; Bogorad, R. L.; Zhu, L. J.; Wolfe, S. A.; et al. Partial DNA-Guided Cas9 Enables Genome Editing with Reduced off-Target Activity. *Nat. Chem. Biol.* **2018**, *14* (3), 311–316. https://doi.org/10.1038/nchembio.2559.

(299) Sanzone, A. P.; El-Sagheer, A. H.; Brown, T.; Tavassoli, A. Assessing the Biocompatibility of Click-Linked DNA in Escherichia Coli. *Nucleic Acids Res.* **2012**, *40* (20), 10567–10575. https://doi.org/10.1093/nar/gks756.

(300) Osman, E. A.; Gadzikwa, T.; Gibbs, J. M. Quick Click: The DNA-Templated Ligation of 3′-*O* -Propargyl- and 5′-Azide-Modified Strands Is as Rapid as and More Selective than Ligase. *ChemBioChem* **2018**, *19* (19), 2081–2087. https://doi.org/10.1002/cbic.201800305.

(301) El-Sagheer, A. H.; Brown, T. Click Nucleic Acid Ligation: Applications in Biology and Nanotechnology. *Acc. Chem. Res.* **2012**, *45* (8), 1258–1267. https://doi.org/10.1021/ar200321n.

(302) Guschin, D. Y.; Waite, A. J.; Katibah, G. E.; Miller, J. C.; Holmes, M. C.; Rebar, E. J. A Rapid and General Assay for Monitoring Endogenous Gene Modification. In *Engineered Zinc Finger Proteins*; Mackay, J. P., Segal, D. J., Eds.; Methods in Molecular Biology; Humana Press: Totowa, NJ, 2010; Vol. 649, pp 247–256. https://doi.org/10.1007/978-1-60761-753-2_15.

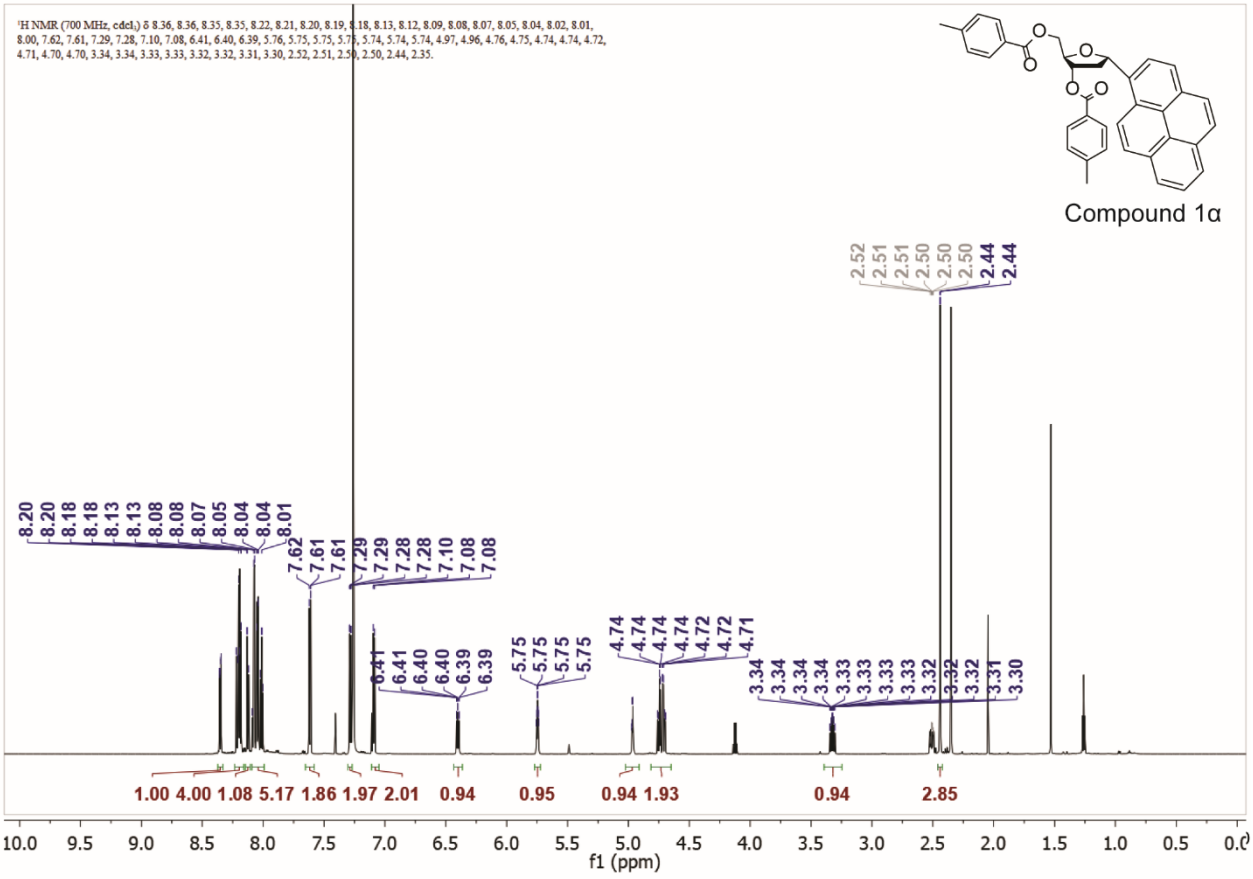
(303) Kim, D.; Kim, S.; Kim, S.; Park, J.; Kim, J.-S. Genome-Wide Target Specificities of CRISPR-Cas9 Nucleases Revealed by Multiplex Digenome-Seq. *Genome Res.* **2016**, *26* (3), 406–415. https://doi.org/10.1101/gr.199588.115.

(304) Trantakis, I. A.; Nilforoushan, A.; Dahlmann, H. A.; Stäuble, C. K.; Sturla, S. J. In-Gene Quantification of O^6 -Methylguanine with Elongated Nucleoside Analogues on Gold Nanoprobes. *J. Am. Chem. Soc.* **2016**, *138* (27), 8497–8504. https://doi.org/10.1021/jacs.6b03599.

Appendix I

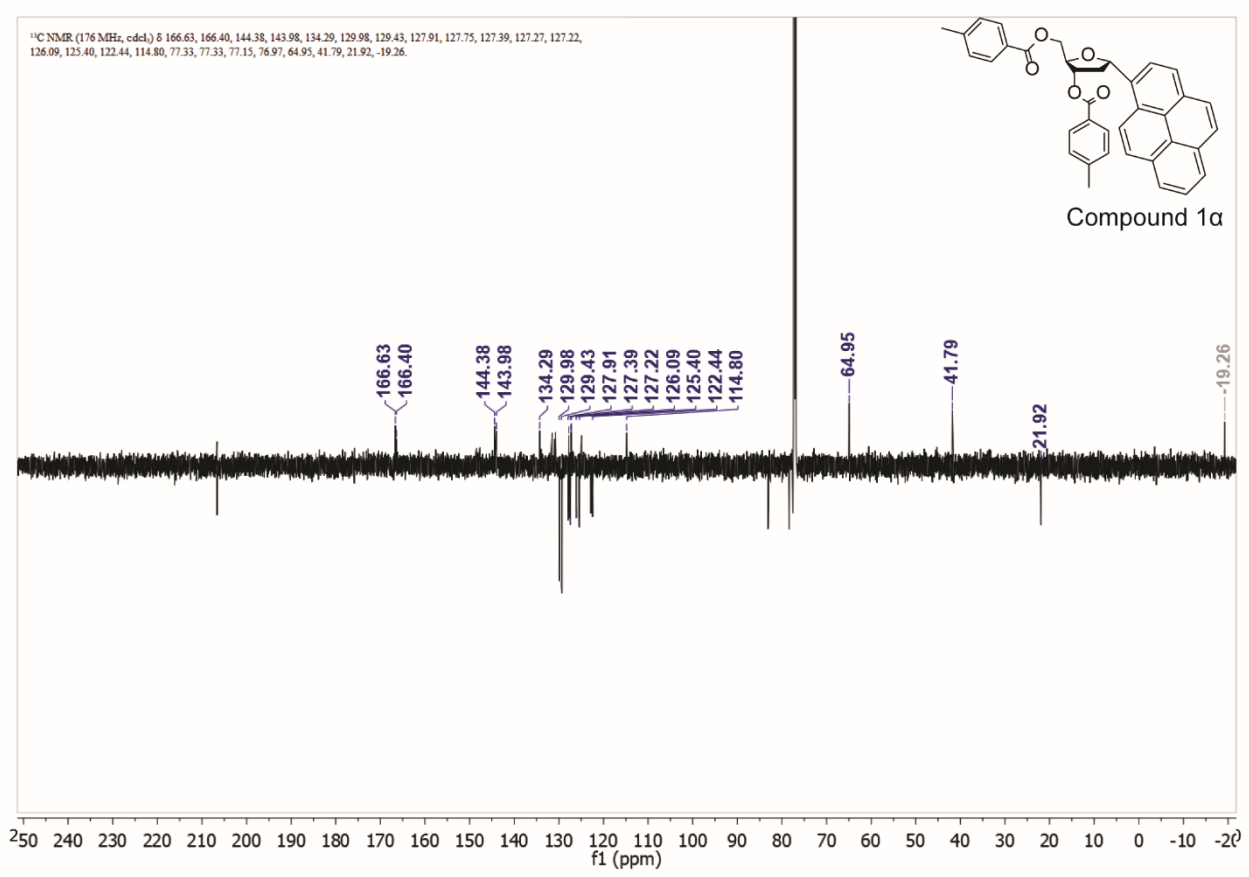
Supporting Information of Chapter 3 and

Chapter 4



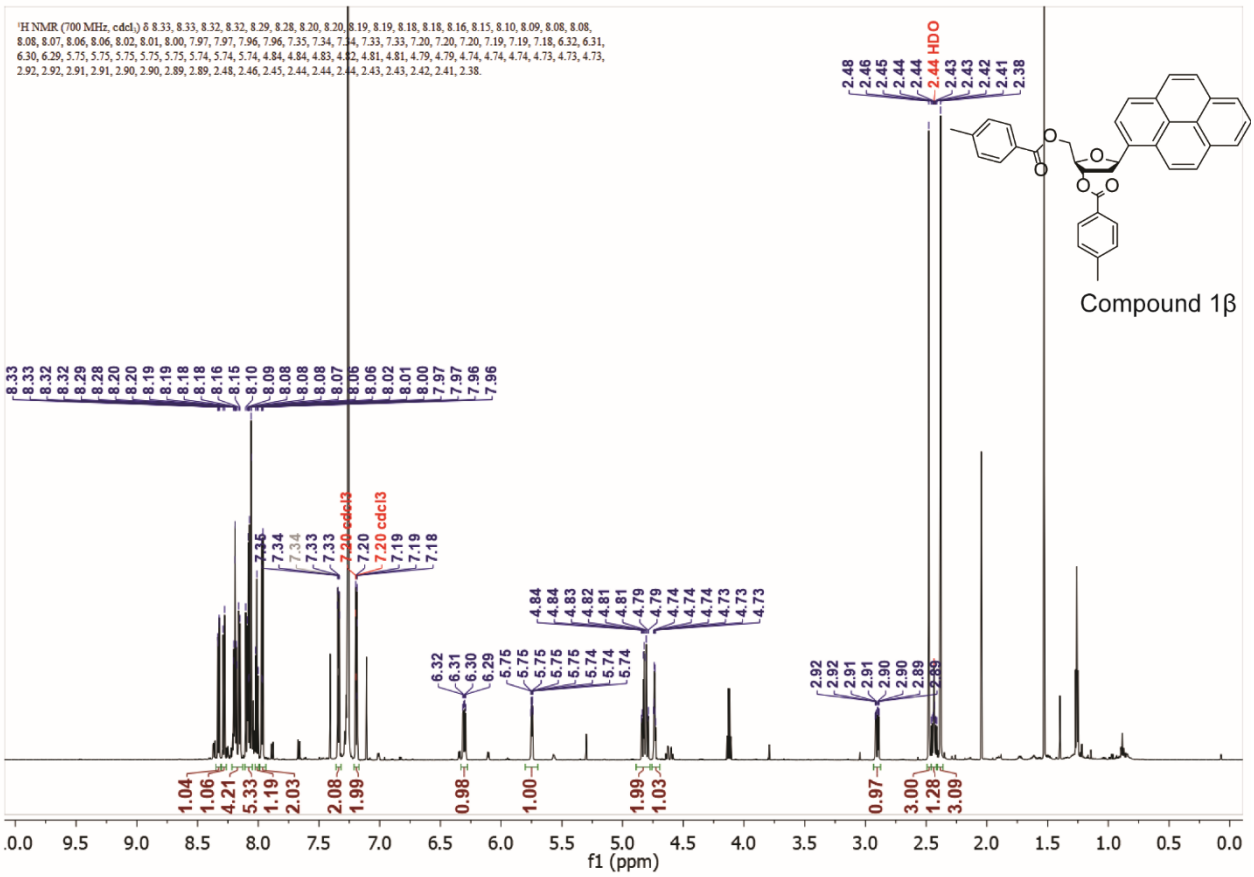
AI. 1 ^{1}H NMR spectrum of compound 1α.

^{1}H NMR for 1α (700 MHz, $CDCl_3$) δ 8.35 (d, J = 9.2 Hz, 1H), 8.23 – 8.17 (m, 4 H), 8.13 (d, J = 9.2 Hz, 1H), 8.10 – 7.99 (m, 5H), 7.62 (d, J = 8.2 Hz, 2H), 7.28 (d, J = 7.9 Hz, 3H), 7.09 (d, J = 7.9 Hz, 2H), 6.43 – 6.38 (m, 1H, H1'), 5.75 (m, 1H, H3'), 4.97 (d, *J* = 2.9, 1H, H4'), 4.78 – 4.68 (m, 2H, H5'), 3.32 (m, 1H, H2'β), 2.51 (m, 1H, H2'α), 2.44 (s, 3H), 2.35 (s, 3H).



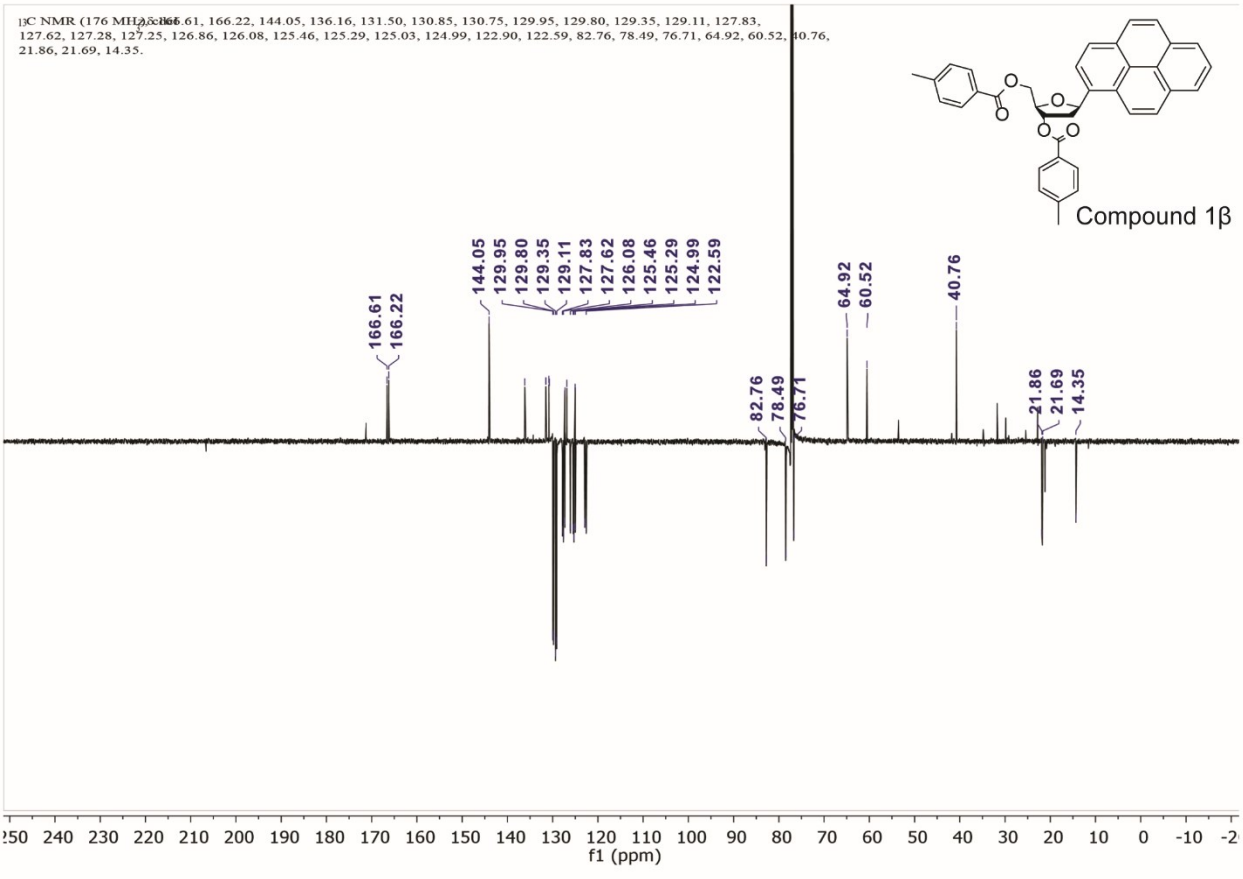
AI. 2 ^{13}C NMR spectrum of compound 1α.

^{13}C NMR for 1α (176 MHz, $CDCl_3$) δ 166.6, 166.4, 144.4, 144.0, 134.3, 130.0, 129.4, 127.9, 127.8, 127.4, 127.3, 127.2, 126.1, 125.4, 122.4, 114.8, 77.3, 77.2, 77.0, 65.0, 41.8, 21.9.



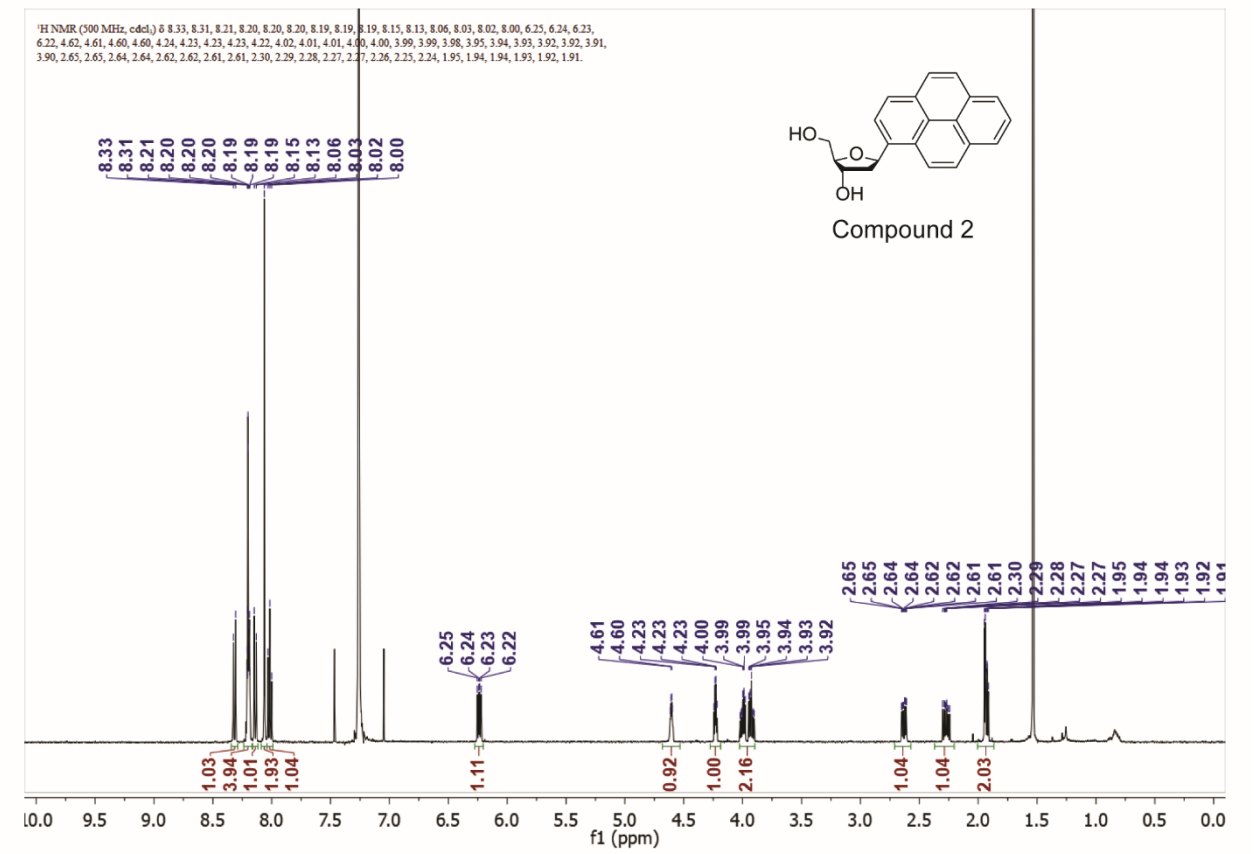
AI. 3 ^{1}H NMR spectrum of compound 1β.

^{1}H NMR for 1β (700 MHz, $CDCl_3$) δ 8.33 (dd, J = 7.9, 0.6 Hz, 1H), 8.28 (d, J = 9.2 Hz, 1H), 8.22 – 8.14 (m, 4H), 8.12 – 8.04 (m, 5H), 8.08 – 7.99 (m, 1H), 7.99 – 7.95 (m, 2H), 7.36 – 7.32 (d, J = 7.5Hz, 2H), 7.19 (d, J = 7.5Hz, 2H), 6.30 (dd, J = 10.9, 5.3 Hz, 1H, H1'), 5.75 (m, 1H, H3'), 4.82 (m, 2H, H5'), 4.74 (m, 1H, H4'), 2.90 (dd, J = 14.1, 5.3 Hz, 1H, H2'α), 2.48 (s, 3H), 2.44 – 2.40 (m, 1H, H2'β), 2.38 (s, 3H).



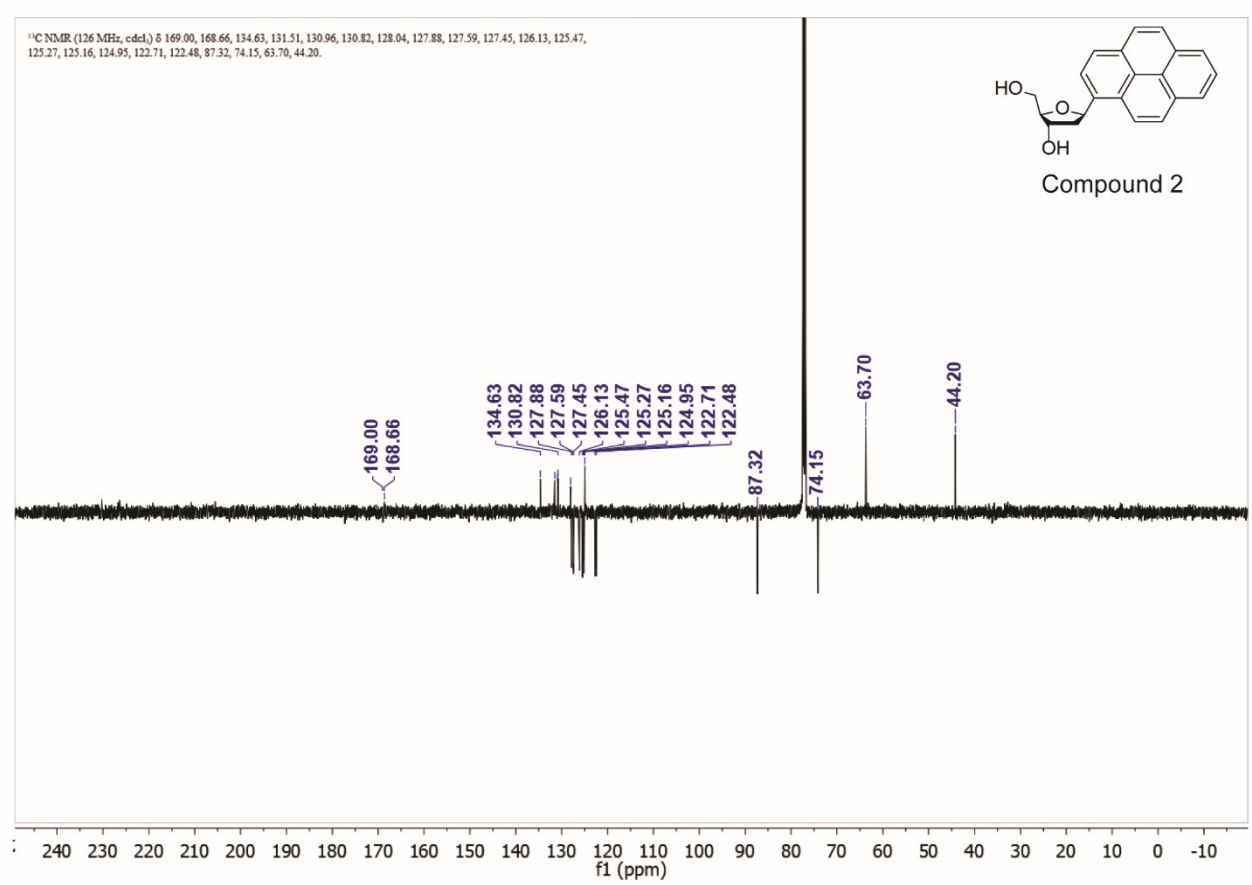
AI. 4 ^{13}C NMR spectrum of compound 1β.

^{13}C NMR for 1α (176 MHz, $CDCl_3$) δ 166.6, 166.2, 144.0, 136.2, 131.5, 130.9, 130.8, 130.0, 129.8, 129.4, 129.1, 127.8, 127.6, 127.3, 127.3, 126.9, 126.1, 125.5, 125.3, 125.0, 125.0, 122.9, 122.6, 82.8, 78.5, 76.7, 64.9, 60.5, 40.8, 21.9, 21.7, 14.4.



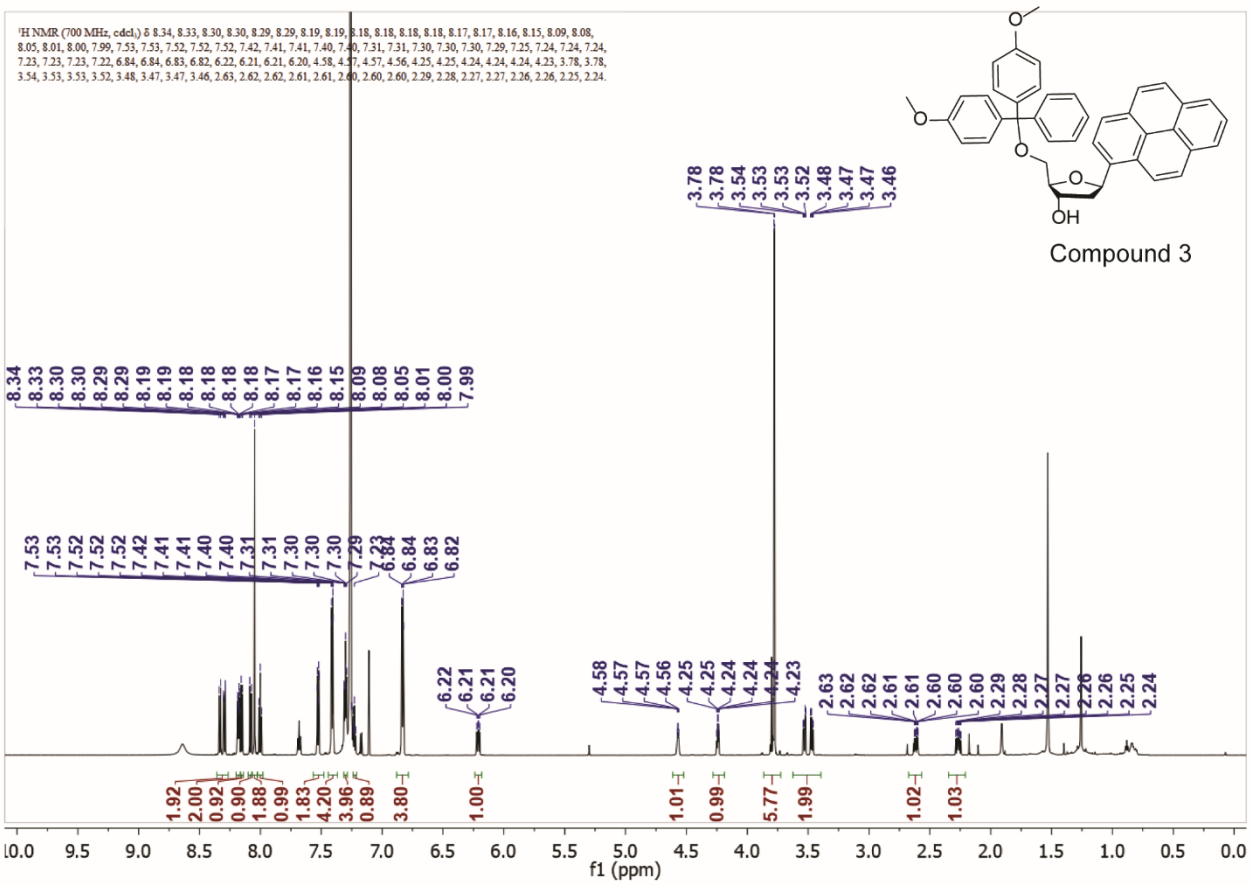
AI. 5 ^{1}H NMR spectrum of compound 2.

^{1}H NMR (700 MHz, $CDCl_3$) δ 8.32 (d, J = 9.2 Hz, 1H), 8.24 – 8.17 (m, 4H), 8.14 (d, J = 9.2 Hz, 1H), 8.06 (s, 2H), 8.02 (t, J = 7.6 Hz, 1H), 6.24 (dd, J = 10.2, 5.8 Hz, 1H, H1'), 4.64 – 4.58 (m, 1H. H3'), 4.30 – 4.17 (m, 1H, H4'), 4.07 – 3.87 (m, 2H, H5'2), 2.63 (ddd, J = 13.4, 5.8, 2.1 Hz, 1H, H2'α), 2.27 (ddd, J = 13.4, 10.3, 6.4 Hz, 1H, H2'β), 1.96 – 1.89 (m, 2x OH).



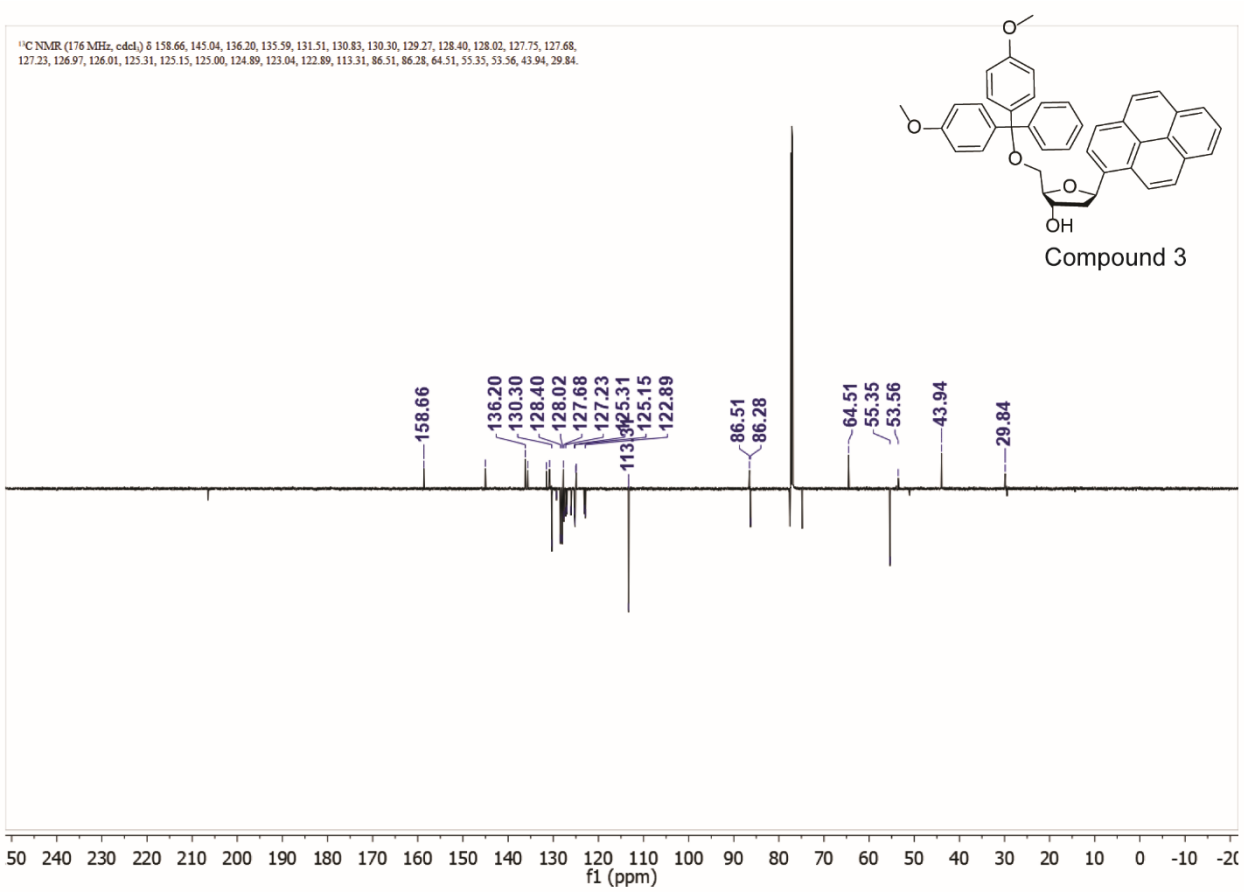
AI. 6 ^{13}C NMR spectrum of compound 2.

^{13}C NMR (126 MHz, $CDCl_3$) δ 169.0, 168.7, 134.6, 131.5, 131.0, 130.8, 128.0, 127.9, 127.6, 127.5, 126.1, 125.5, 125.3, 125.2, 125.0, 122.7, 122.5, 87.2, 74.2, 63.7, 44.2.



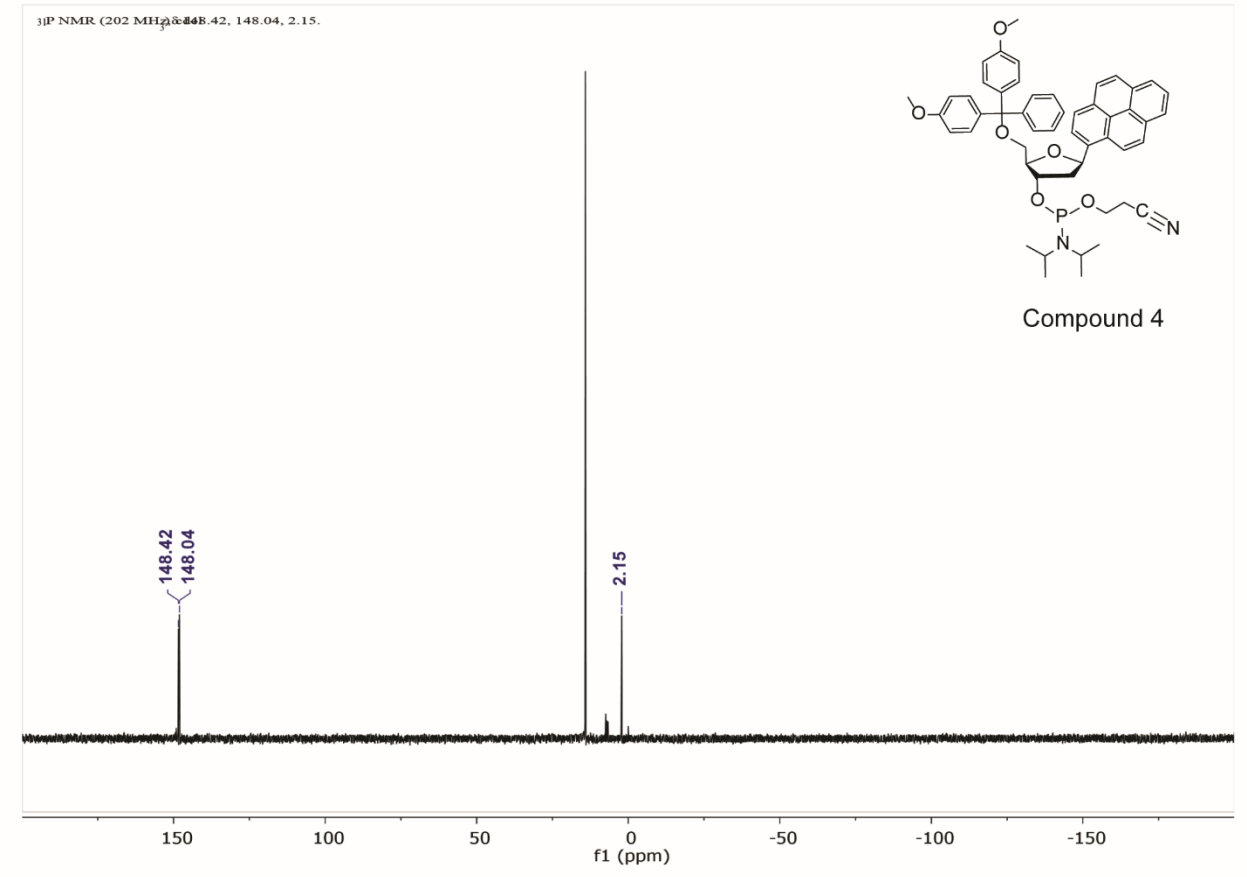
AI. 7 ^{1}H NMR spectrum of compound 3.

^{1}H NMR (700 MHz, $CDCl_3$) δ 8.33 (d, J = 9.2 Hz, 1H), 8.29 (dd, J = 7.9, 0.6 Hz, 1H), 8.18 (ddd, J = 7.8, 4.1, 1.1 Hz, 2H), 8.15 (d, J = 7.9 Hz, 1H), 8.08 (d, J = 9.2 Hz, 1H), 8.05 (s, 2H), 8.00 (t, J = 7.6 Hz, 1H), 7.56 – 7.50 (m, 2H), 7.44 – 7.38 (m, 4H), 7.30 (m, 2H), 7.26 – 7.20 (m, 1H), 6.83 (dd, J = 9.0, 2.6 Hz, 4H), 6.21 (dd, J = 9.8, 6.0 Hz, 1H, H1'), 4.57 (dd, J = 6.3, 3.0 Hz, 1H, H3'), 4.24 (dd, J = 4.9, 3.4 Hz, 1H, H4'), 3.78 (d, J = 3.5 Hz, 6H), 3.50 (ddd, J = 43.5, 9.8, 4.8 Hz, 2H, H5'), 2.61 (ddd, J = 13.3, 6.0, 2.4 Hz, 1H, H2'α), 2.27 (ddd, J = 13.3, 9.9, 6.3 Hz, 1H, H2'β).



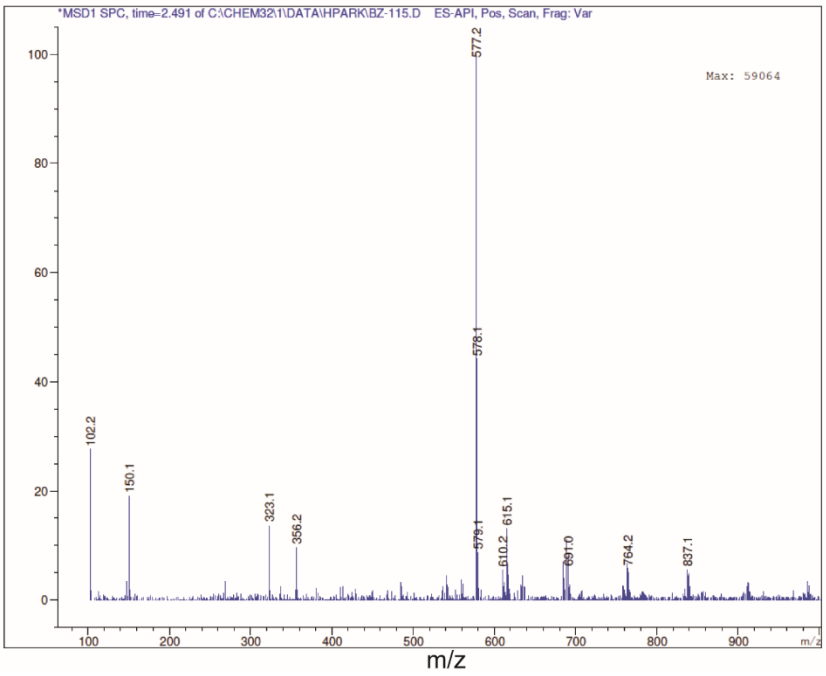
AI. 8 ^{13}C NMR spectrum of compound 3.

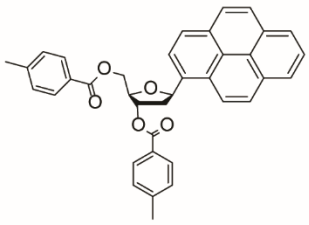
^{13}C NMR (176 MHz, $CDCl_3$) δ 158.7, 145.0, 136.2, 135.6, 131.5, 130.8, 130.3, 129.3, 128.4, 128.0, 127.8, 127.7, 127.2, 127.0, 126.0, 125.3, 125.2, 125.0, 124. 9, 123.0, 122.9, 113.3, 86.5, 86.3, 64.5, 55.4, 53.6, 43.9, 29.8.



AI. 9 ^{31}P NMR spectrum of crude compound 4.

^{31}P NMR of crude compound 4 (202 MHz, $CDCl_3$) δ 148.4, 148.0 and observed the oxidized phosphoramidite product or starting material at δ 14.2 and δ 2.2.

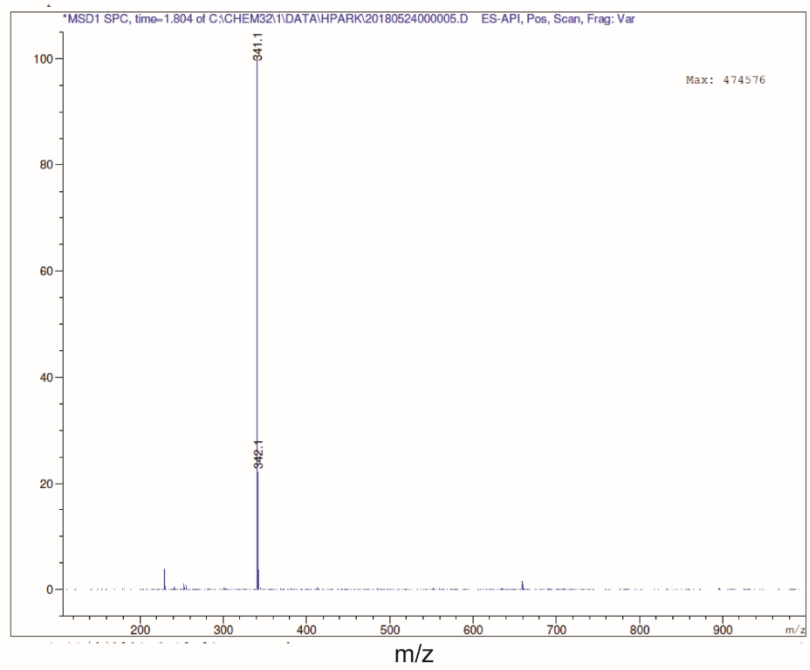


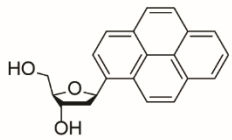


Compound 1 mixture ($C_{37}H_{30}O_5$)

Calculated [M+Na] 577.63
Found 577.2

AI. 10 ESI mass spectrum of compound 1.

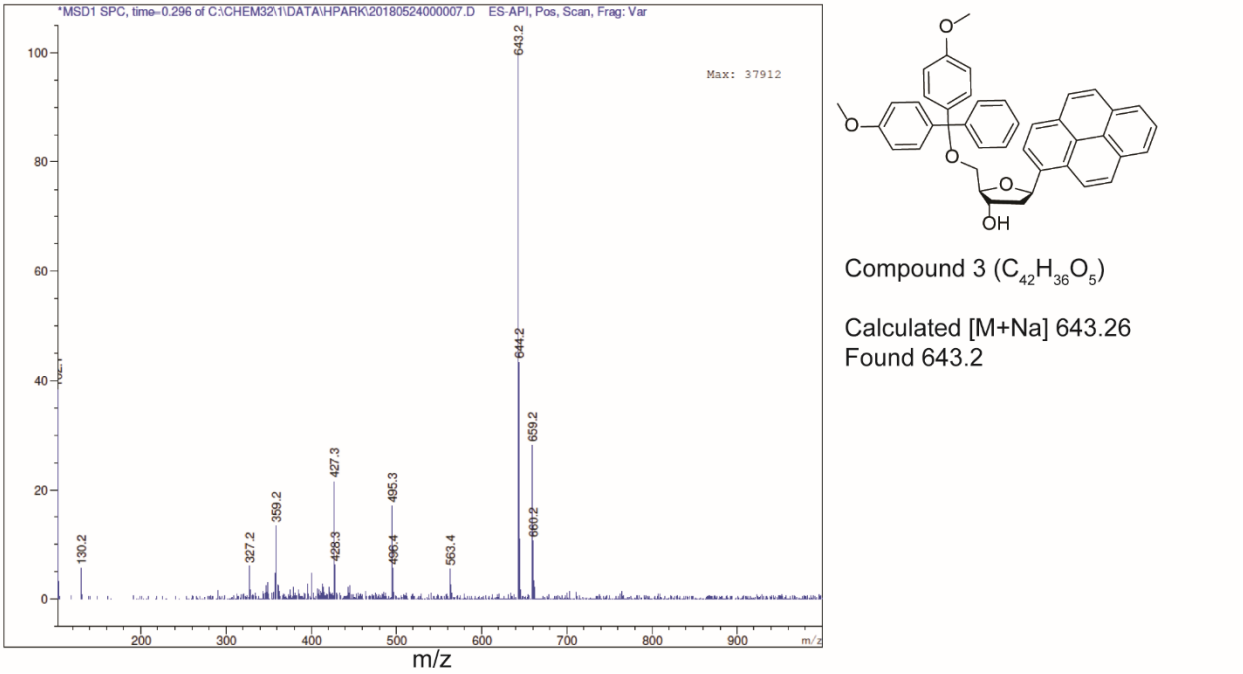




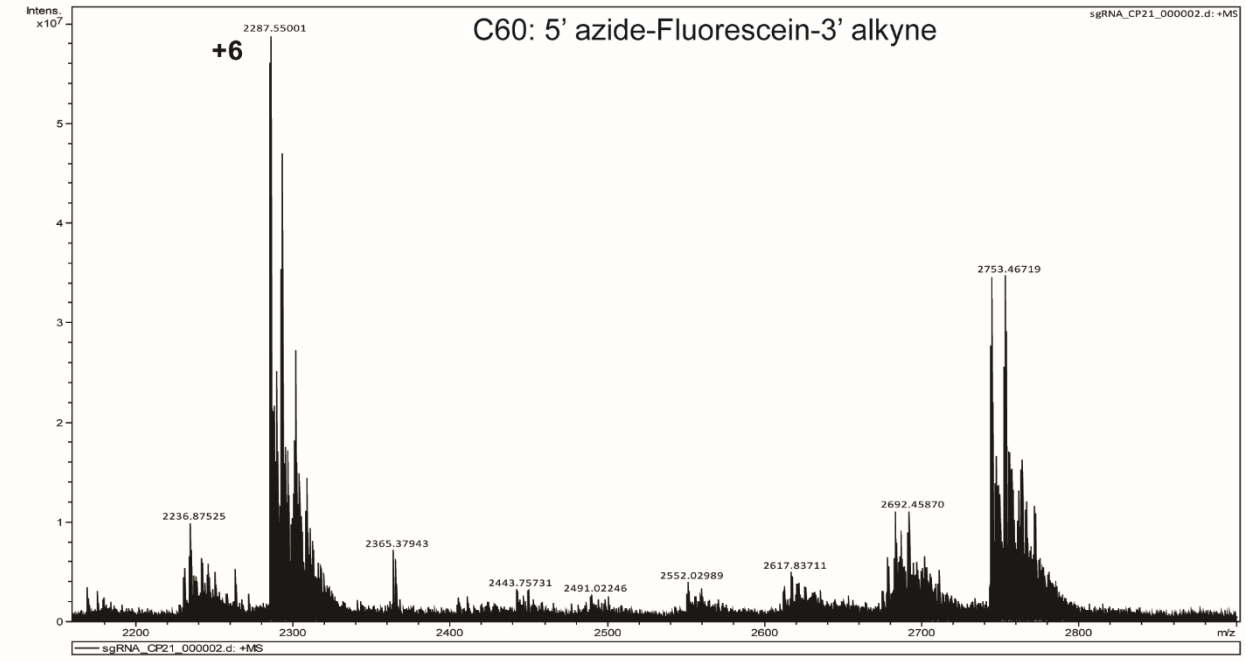
Compound 2 ($C_{21}H_{18}O_3$)

Calculated [M+Na] 341.37
Found 341.1

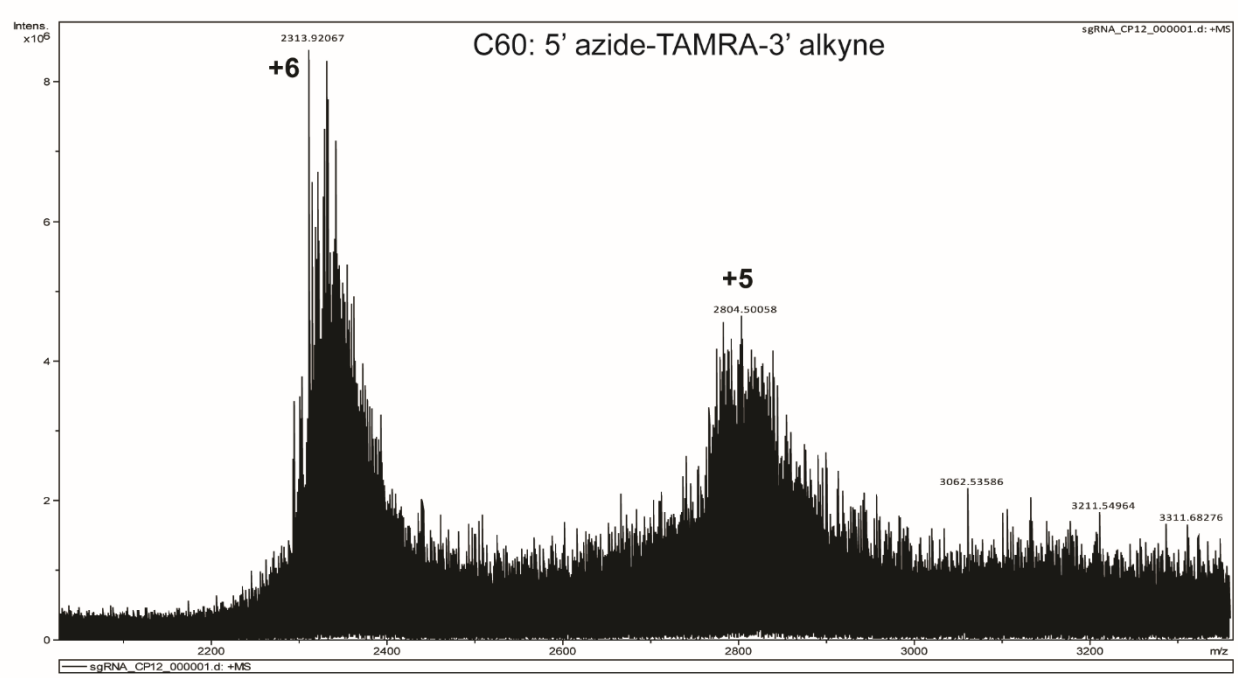
AI. 11 ESI mass spectrum of compound 2.



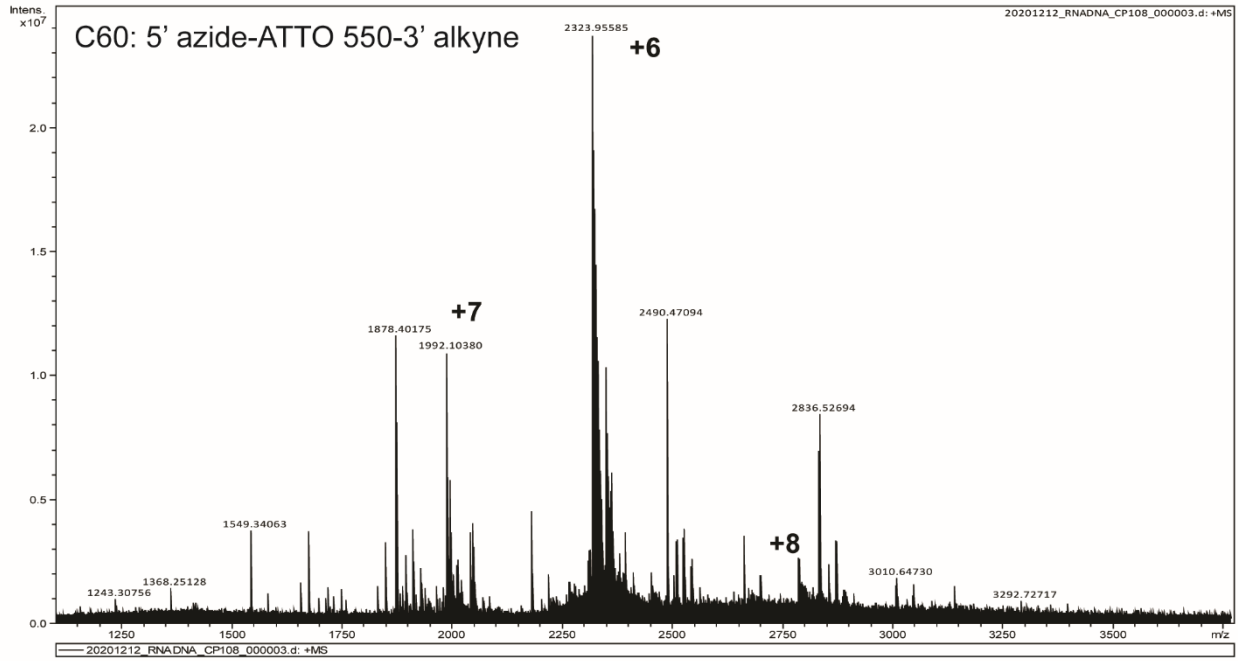
AI. 12 ESI mass spectrum of compound 3.



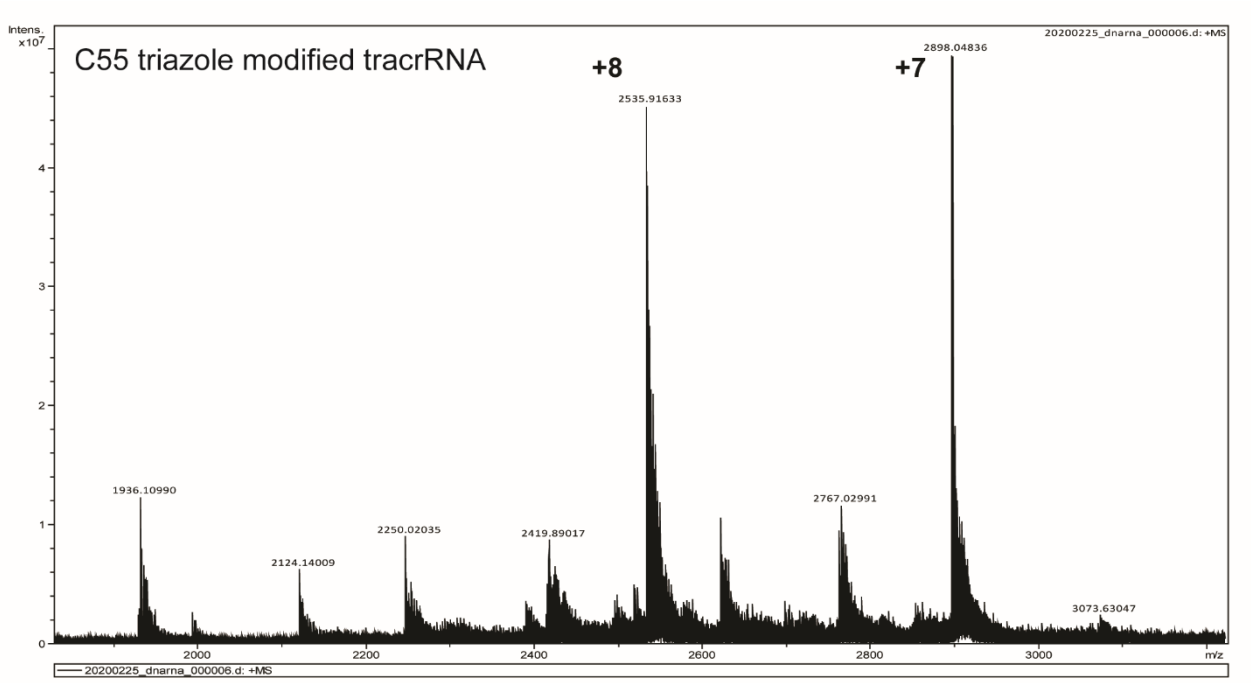
AI. 13 ESI mass spectrum of fluorescein modified 5' azide-3' alkyne strand in positive mode for aqueous ammonium acetate solution.



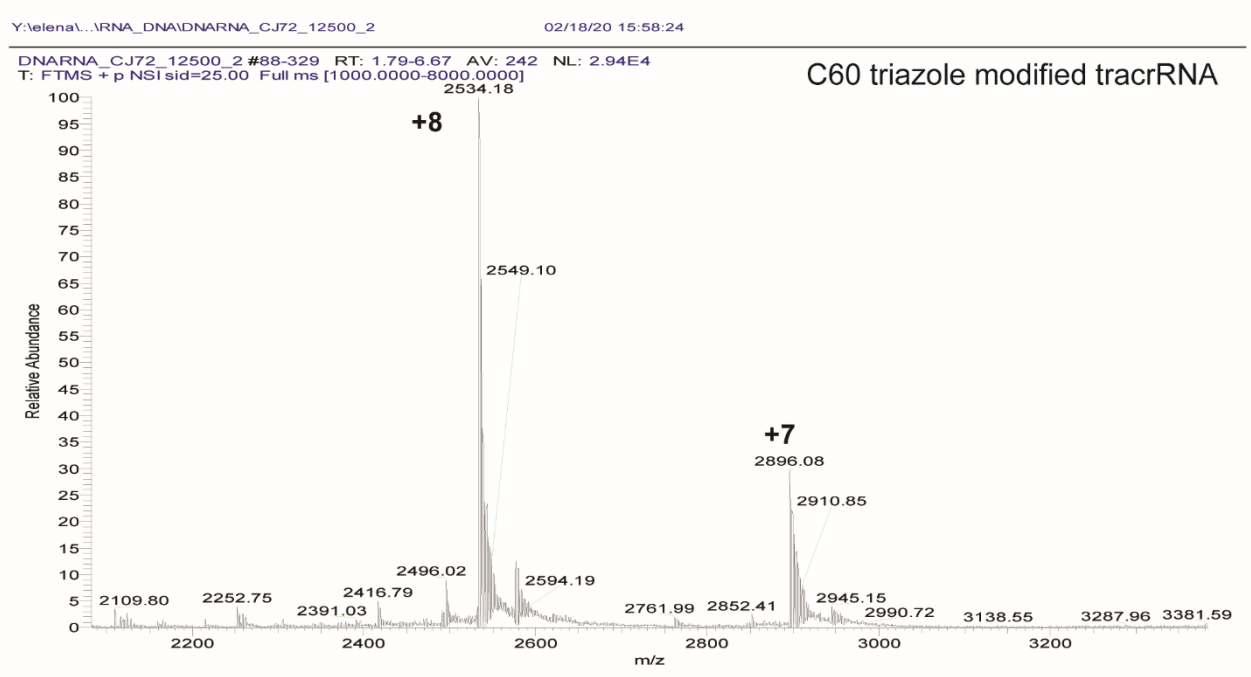
AI. 14 ESI mass spectrum of TAMRA modified 5' azide-3' alkyne strand in positive mode for aqueous ammonium acetate solution.



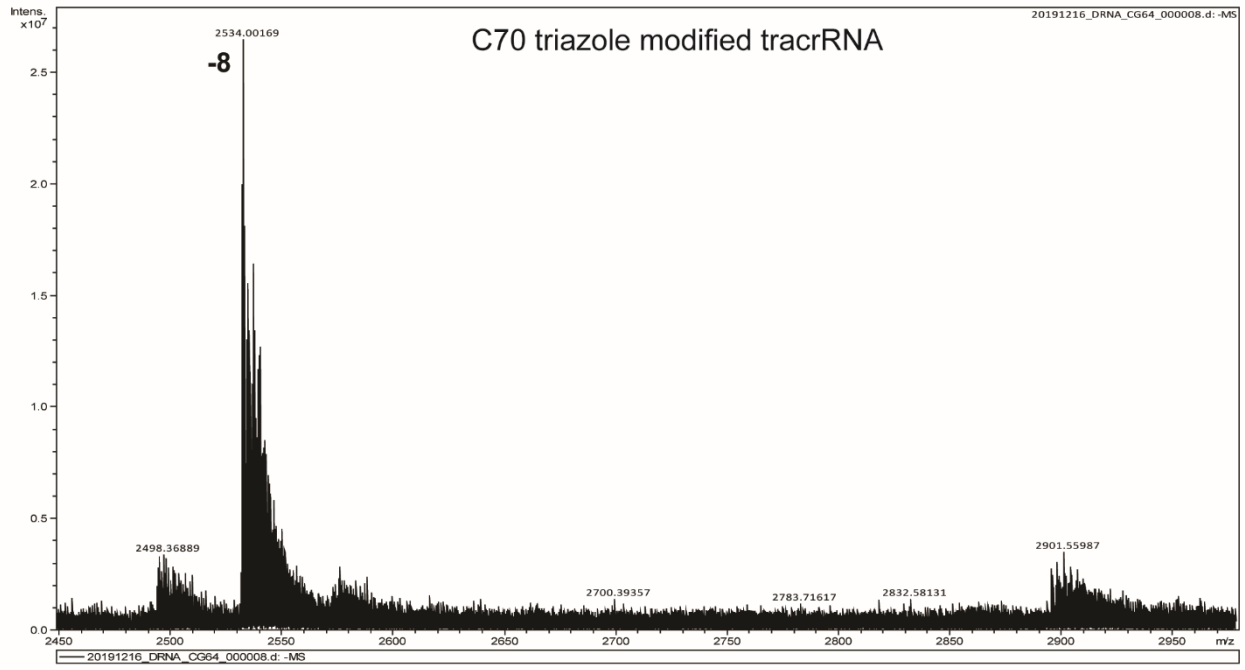
AI. 15 ESI mass spectrum of ATTO 550 modified 5' azide-3' alkyne strand in positive mode for aqueous ammonium acetate solution.



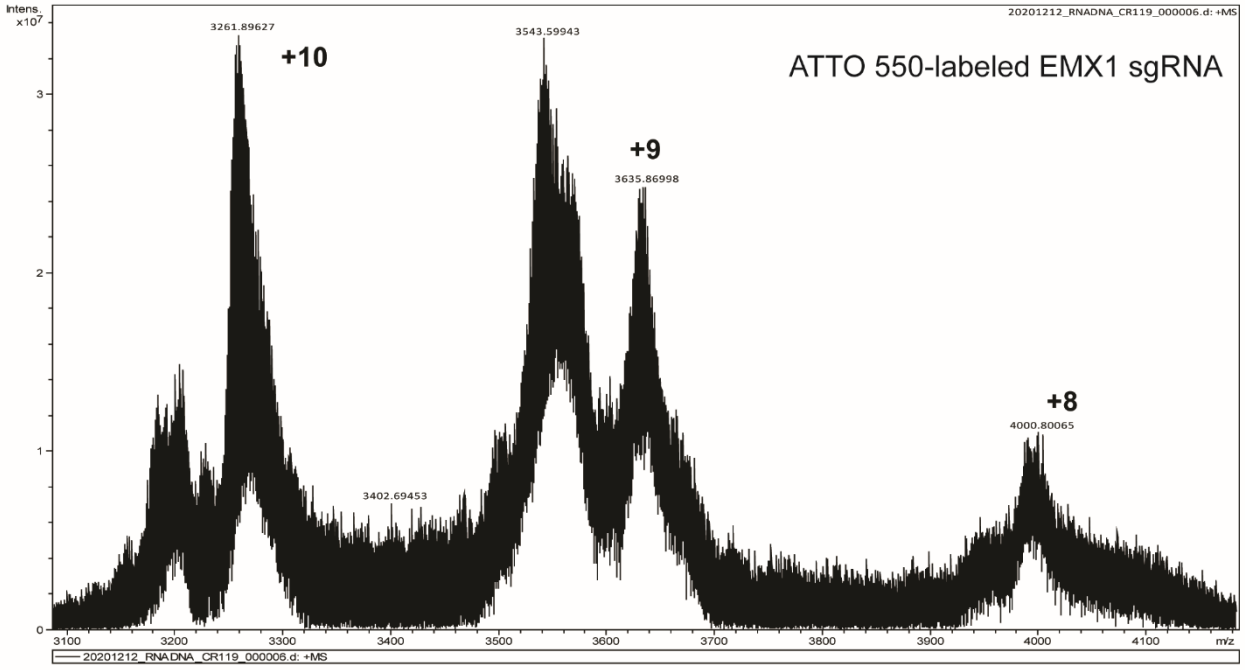
AI. 16 ESI mass spectrum of C55 triazole modified tracrRNA strand in positive mode for aqueous ammonium acetate solution.



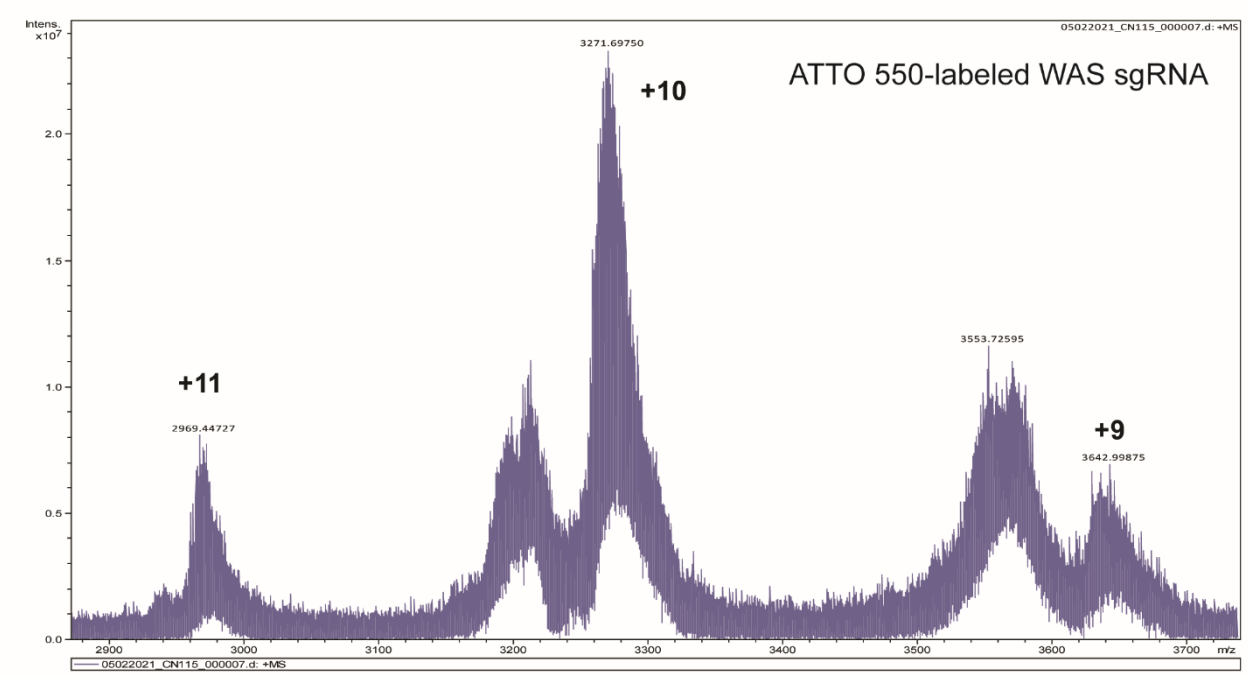
AI. 17 ESI mass spectrum of C60 triazole modified tracrRNA strand in positive mode for aqueous ammonium acetate solution.



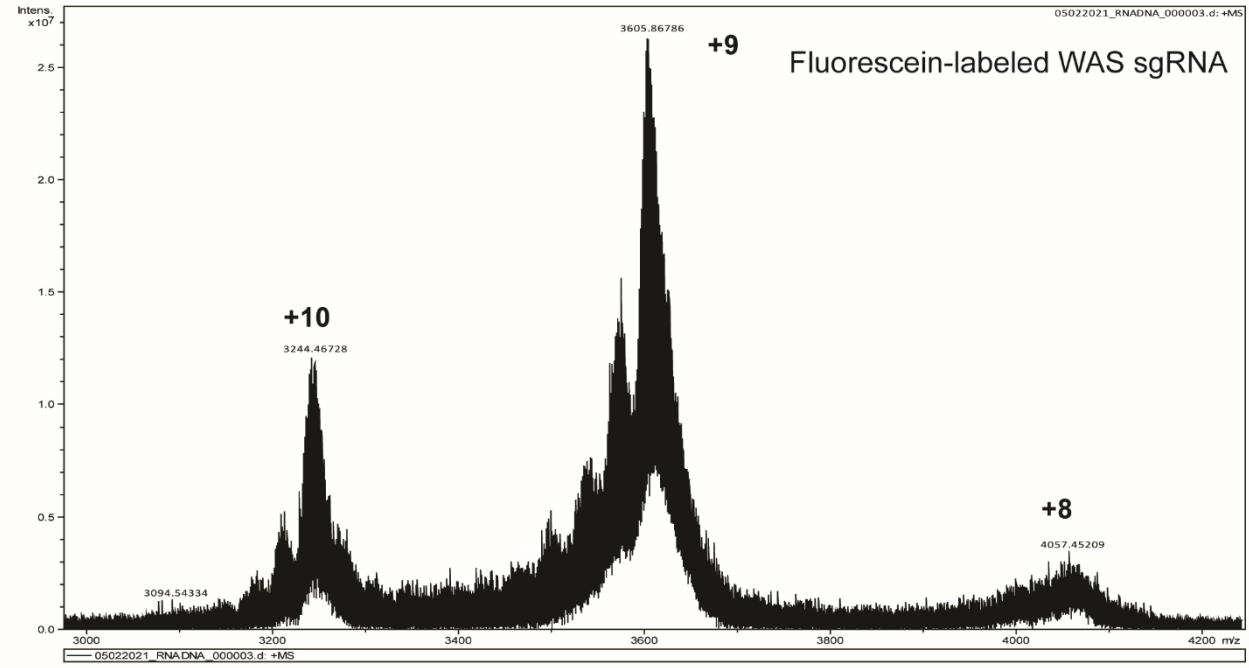
AI. 18 ESI mass spectrum of C70 triazole modified tracrRNA strand in negative mode for aqueous ammonium acetate solution.



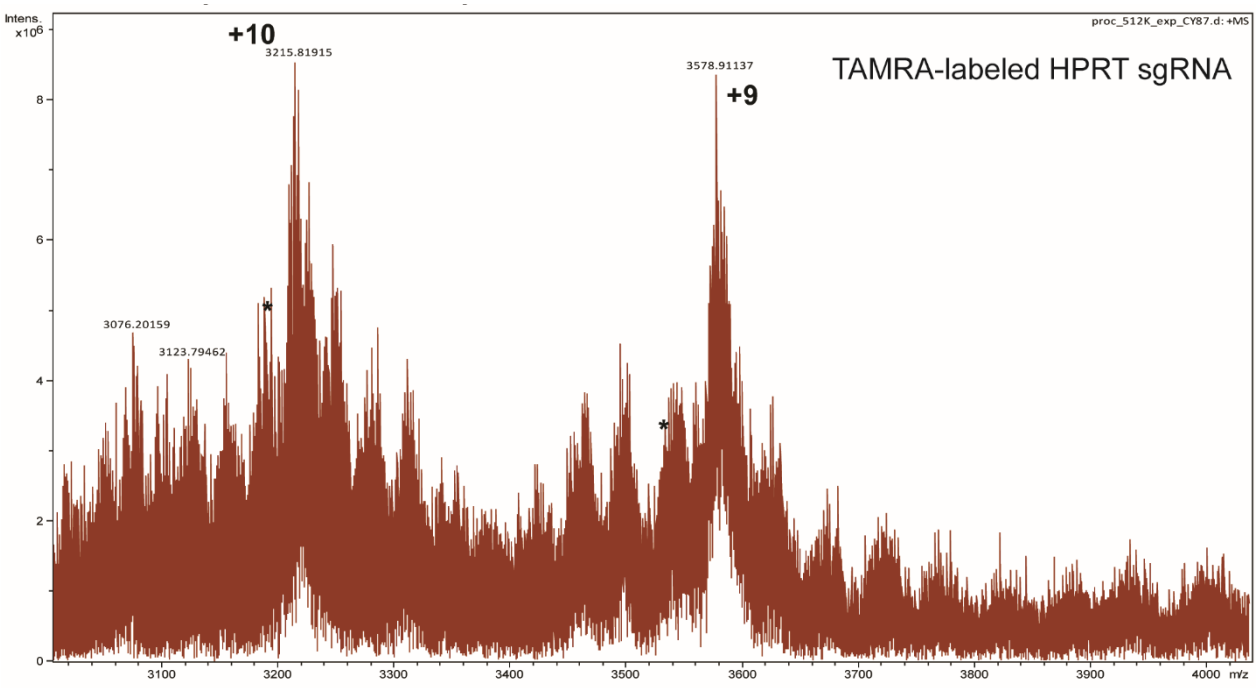
AI. 19 ESI mass spectrum of ATTO 550-labeled *EMX1* sgRNA in positive mode for aqueous ammonium acetate solution.



AI. 20 ESI mass spectrum of ATTO 550-labeled *WAS* sgRNA in positive mode for aqueous ammonium acetate solution.



AI. 21 ESI mass spectrum of fluorescein-labeled *WAS* sgRNA in positive mode for aqueous ammonium acetate solution.



AI. 22 ESI mass spectrum of TAMRA-labeled *HPRT* sgRNA in positive mode for aqueous ammonium acetate solution.