2019

MECHANISMS OF TRINUCLEOTIDE REPEAT INSTABILITY DURING DNA SYNTHESIS

Kara Y. Chan

University of Kentucky, kara.chan@uky.edu
Author ORCID Identifier:
https://orcid.org/0000-0003-4477-9151
Digital Object Identifier: https://doi.org/10.13023/etd.2019.394

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

<div align="right">

Kara Y. Chan, Student

Dr. Guo-Min Li, Major Professor

Dr. Isabel Mellon, Director of Graduate Studies

</div>

MECHANISMS OF TRINUCLEOTIDE REPEAT INSTABILITY DURING DNA SYNTHESIS

_____

DISSERTATION
_____

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in the
College of Medicine
at the University of Kentucky

By

Kara Yi-Wing Chan

Lexington, Kentucky

Director: Dr. Guo-Min Li, Professor of Toxicology and Cancer Biology

Lexington, Kentucky

2019

ABSTRACT OF DISSERTATION

MECHANISMS OF TRINUCLEOTIDE REPEAT INSTABILITY DURING DNA SYNTHESIS

Genomic instability, in the form of gene mutations, insertions/deletions, and gene amplifications, is one of the hallmarks in many types of cancers and other inheritable genetic disorders. Trinucleotide repeat (TNR) disorders, such as Huntington's disease (HD) and Myotonic dystrophy (DM) can be inherited and repeats may be extended through subsequent generations. However, it is not clear how the CAG repeats expand through generations in HD. Two possible repeat expansion mechanisms include: 1) polymerase mediated repeat extension; 2) persistent TNR hairpin structure formation persisting in the genome resulting in expansion after subsequent cell division. Recent *in vitro* studies suggested that a family A translesion polymerase, polymerase θ (Polθ), was able to synthesize DNA larger than the template DNA. Clinical and *in vivo* studies showed either overexpression or knock down of Polθ caused poor survival in breast cancer patients and genomic instability. However, the role of Polθ in TNR expansion remains unelucidated. Therefore, we hypothesize that Polθ can directly cause TNR expansion during DNA synthesis. The investigation of the functional properties of Polθ during DNA replication and TNR synthesis will provide insight for the mechanism of TNR expansion through generations.

Keywords:

DNA Translesion Polymerase- Polymerase θ

Base Excision Repair

Hairpin Bypass Synthesis

Mn^{2+}/Mg^{2+}

Trinucleotide Repeats Expansion

Huntington's Disease

<div align="right">

Kara Y. Chan
_____
Name of Student

09/22/2019
_____
Date

</div>

MECHANISMS OF TRINUCLEOTIDE REPEAT INSTABILITY DURING DNA
SYNTHESIS

By

Kara Y. Chan

<table>
<tr><td>Guo-Min Li, Ph.D.</td></tr>
<tr><td>Director of Dissertation</td></tr>
</table>

<table>
<tr><td>Isabel Mellon, Ph.D.</td></tr>
<tr><td>Director of Graduate Studies</td></tr>
</table>

<table>
<tr><td>09/22/2019</td></tr>
<tr><td>Date</td></tr>
</table>

Acknowledgement:

Throughout the years of my graduate studies, I learned a great number of quality skills including experimental skills, critical thinking, and communication skills from my mentor Dr. Guo-Min Li, my committee members, Dr. Isabel Mellon, Dr. Brett Spear and Dr. Yvonne Fondufe-Mittendorf, my colleagues, especially Dr. Janice Ortega, Dr. Guogen Mao, Dr. Liya Gu, and our collaborator- Dr. Wei Yang. I thank my mentor, Dr. Li, for his patience and guidance which directed me through family hardships and stumbling blocks in my progress. Dr. Li led us through a move from the University of Kentucky (UK) in Lexington, KY to the University of Southern California (USC) in Los Angeles, CA and ultimately to the University of Texas Southwestern (UTSW) in Dallas, TX. Along the way, I was exposed to various research environments and people. I met new friends and strengthened my networking skills. Leading me to finding new goals and dreams after graduation. The second person that I would like to acknowledge is Dr. Janice Ortega, Janice helped me to be a better and independent scientist. She is a great mentor and friend with whom I often discussed many scientific topics and my experimental results. Lastly, I would like to thank Dr. Guogen Mao for helping me in his free time after we moved to UTSW.

I would like to acknowledge everyone I met and my colleagues in UK, USC and UTSW, especially: Dr. Feng Li, Dr. Sanghee Lee, Dr. Michelle Pitts, Christine Kim, Nidhi Shukla and Hui Yu from UK; Dr. Hong Tao Li, Dr. Bailin Zhao, Kelly Villers and Veronica Ortiz from USC; Dr. Jinzhen Guo, Dr. Qihuang Jin, Dr. Xiao Sun, Dr.

Junhong Guan, Dr. Manisha Shrestha, Alicia Zhang, Dr. Asaithamby Aroumougame (Thambi) and Dr. Hui Ming Lu from UTSW.

Last of all, I would like to thank my family: my father, Fai Lam Chan, my mom, Po Sheung Chui, my sister, Karen Chan, and my brother, Ka Kei Chan. My boyfriend Dr. Timothy Scott who supported me emotionally and scientifically leading to the completion of my dissertation.

Thank you for everyone whom I met and became my long-term friends during these years as a Ph.D. candidate. Thank you for being a part of my life and force to finish my dissertation.

TABLE OF CONTENTS:

LIST OF TABLES:

LIST OF FIGURES:

Chapter 1: Introduction

A Novel Translesion DNA Polymerase and Trinucleotide Repeat Stability

1.1. CAG Trinucleotide Repeat Expansion in Huntington's Disease

Maintaining genomic stability is important to preserve normal cellular functions and to prevent genetic disorders. Genomic instability, including DNA nucleotide insertions/deletions, gene relocations, and DNA re-replication, is one of the hallmarks in many types of cancer and genetic disorders [1-5]. Huntington's disease (HD) is a neurodegenerative disease caused by an autosomal dominate gain-of-function mutation of the Huntingtin (mHTT) gene [1]. The mHTT protein encoded by the mutated *HTT* gene contains an expansion to more than 35 CAG repeats within exon 1 that leads to increased oxidative stress within cells [1, 7]. Huntington's Disease patient cell and tissue studies showed that mHTT protein aggregates caused transcriptional dysregulation, defective energy metabolism, increased oxidative stress, excitotoxicity, and inflammation [1, 8-16]. The mechanism by which CAG trinucleotide repeats (TNRs) undergo expansion remains unclear. However, recent studies suggested oxidative stress could be a possible cause of TNR expansion in HD through DNA repair mechanisms especially base excision repair (BER) involving MutSβ and polymerase β (Polβ) [1, 17, 18]. Previous studies showed that BER contributes to TNR expansion during oxidative damaged DNA base repair [5, 19]. During BER, apurinic/apyrimidinic endonuclease 1 (APE1) creates a nick on the damaged strand at CAG repeats, which can easily form a hairpin structure via DNA strand slippage [5]. Xu et. al. suggested that flap structure-specific endonuclease 1 (FEN1) and DNA Ligase 1

(Lig1) could promote an incomplete removal of large hairpin structures caused by BER, leading to DNA expansion [5]. During DNA synthesis, the relaxed DNA strands which contain repetitive sequences are prone to slippage and form hairpins structures [20, 21]. If these hairpin structures are not removed and repaired properly, it will cause repeat expansion. Thus, two possible mechanisms that could explain TNR expansion include: 1) error-prone DNA synthesis by a polymerase leading to expansion on the nascent strand during replication and/or repair, or 2) failure of hairpin removal. We propose to investigate which human DNA polymerase(s) is involved in error-prone DNA synthesis that causes TNR expansion.

Figure 1.1: Model of hairpin formation in CAG repeat sequence during DNA synthesis and DNA repair.

1.2. Genomic Instability and DNA Polymerases

There are two major types of DNA polymerase: 1) high fidelity polymerases-Polδ, Polε; 2) error-prone translesion polymerases- Polβ, Polζ, Polθ, Polμ, Polη, Polι, and Polκ. High-fidelity DNA polymerases exhibit a more active proofreading activity (3' → 5' exonuclease activity) and processivity than translesion polymerases. Because of this proofreading activity, high-fidelity polymerases are known to have lower error rates than error-prone translesion polymerases. During DNA replication, DNA polymerase α (Polα) synthesizes short RNA primers for high fidelity DNA polymerases- Polδ and Polε to initiate DNA synthesis. However, when Polδ or Polε encounters a DNA lesion, the affinity of Polδ or Polε to the DNA template dramatically decreases [22]. Although both Polδ or Polε can excise the secondary DNA structures on the nascent strand with their 3'→5' exonuclease activity, both Polδ and Polε would halt and disassociate from those structures allowing the recruitment of a translesion polymerase and/or a DNA repair machinery such as DNA mismatch repair to the damage site. Due to the lack of proofreading activity in translesion polymerases, they are prone to inducing errors and bypassing various DNA lesions during DNA synthesis. In addition to single nucleotide mismatches by replicative and translesion polymerases, secondary DNA structures can be formed with relaxed and unprotected DNA during replication [23]. These secondary structures can lead to insertions and deletions. Highly repetitive DNA sequences may result in hairpin structure formation because of strand slippages, especially in regions with simple repetitive DNA sequences such as CAG/CTG repeats. *In vitro* studies showed that polymerase β, together

3

with mismatch recognition protein MutSβ, promotes CAG/CTG hairpin retention, leading to repeat expansions [20, 24].

1.3. Huntington's Disease (HD) and Repeat Expansion

HD is an inheritable genetic neurodegenerative disease caused by the expansion of CAG repeats within exon 1 of the *HTT* gene. HD is an autosomal dominate disease that causes neurodegeneration when a patient carries an allele of the CAG expended *HTT* gene. The severity of HD is determined by the length of the CAG repeats: the longer the repeats, the younger that patient will experience symptoms [25]. A healthy individual has 6-35 CAG repeats while a HD patient will carry a copy longer than 35 CAG repeats and is more likely to develop HD symptoms [7, 25, 26]. People who carry an allele of 27-35 CAG repeats will have an increased lifelong risk of developing HD symptoms and having HD offspring because of the instability of the CAG repeats in the *HTT* gene [27, 28]. Interestingly, clinical studies showed that large CAG expansions were mainly from paternal inheritance as large instabilities were found in spermatogenesis more often than oogenesis [25, 29]. However, the mechanism of instability of 27 or more CAG repeats occurs is not fully understood.

1.4. Oxidative Stress and Repeat Expansion

Patients with HD and other neurodegenerative diseases were shown to have higher reactive oxygen species (ROS) generation than healthy people [1, 7, 17, 18, 30]. The increased oxidative stress caused an accumulation of oxidative damage on DNA and proteins, and led to single stranded break (SSB), double

stranded break (DSB) and the formation of oxidized protein and DNA nucleotides such as 8-oxo-guanine (8-oxo-G). In an *in vivo* study, HD fibroblasts were observed to have CAG repeat instability when treated with H_2O_2 [31]. In the same study, 8-Oxoguanine glycosylase (OGG1), one of the BER proteins, was shown to be recruited to initiate BER for the repair of 8-oxo-G, and was responsible for SSB and repeat expansion in cells [31]. HD mice were also used to study the relationship between oxidative stress and repeat expansion, genetic sequencing results showed that organs, such as brain and liver, exposed to higher oxidative stress and had high 8-oxo-G lesions experienced more repeat instability compared to their tails which were exposed to low oxidative stress with fewer 8-oxo-G lesion [31]. Since both mouse models and human tissues samples showed higher oxidative stress in HD compared to wild type control or control organs, it is suggested there is a strong correlation between oxidative stress and DNA repair mechanisms in repeat expansion.

1.5. Base Excision Repair and Repeat Extension

Multiple studies have reported that translesion Polβ, one of the components in BER, and MutSβ, shown as a partner with Polβ, could cause repeat expansion through hairpin retention *in vitro* [20, 24, 32]. In HD patient samples, increased oxidative stress was observed, suggesting that this may be one of the components to drive repeat expansion [1, 33]. If ROS attacks DNA, bases such as guanine are one of the bases that often damaged by ROS and leads to the formation of 8-oxo-G. These 8-oxo-G can be repaired by BER [34]. BER begins with a lesion-specific DNA glycosylase that removes the damaged base from the sugar backbone

5

leaving an apurinic/apyrimidinic (AP) site [35]. Apurinic/apyrimidinic endonuclease 1 (APE1) is recruited and subsequently leads to a single strand break before the AP site is replaced and the strand is re-ligated [35]. In highly repetitive regions, disruption of the double helix allows strand slippage to occur and leads to stable hairpin formation [21, 36, 37]. If these stable hairpins are not removed and repaired accurately, they will lead to repeat expansion [20, 24, 36].

Studies have also pointed to a possible mechanism of repeat expansion by which MutSβ can be recruited to further stabilize hairpin structures during the process of BER and allowed hairpin bypass synthesis by Polβ [20, 24, 32, 38]. Lokanga et. al., showed that heterozygosity of a Polβ mutant in cells, which contain a wild type Polβ and a PolβY265C, increased the rate of TNR contraction and lowered the rate of having small expansion (<5 TNR expansion) in three-month old mice [39]. Although heterozygote mutant cells had less overall expansions, a significantly larger fraction of cells had larger expansions (>10 TNR repeats) than wild type cells in older mice (11-month old mice) [39]. Lokanga et. al. suggested that an unknown pathway may be involved in the large repeat expansion in older mice [39]. Furthermore, multiple studies showed that Poly(ADP-Ribose) Polymerase 1 (PARP1) participates in BER because it can bind to the same substrate as APE1. PARP1 may also be involved in SSB repair [40-44]. Interestingly, PARP1 recruits Polθ to the damage site for strand break repair and Polθ can function as Polβ in BER [35, 45]. This led us to study the involvement of Polθ in TNR expansion as an alternative repair mechanism to cause large repeat instability.

1.6. Polymerase θ Involvements in Other DNA Repair Mechanisms

Polymerase θ was identified as the only translesion polymerase to be involved in a newly discovered double strand break (DSB) repair mechanism — alternative end joining (alt-EJ, also known as microhomology-mediated end joining (MMEJ) [46]. Polθ is proposed to dimerize and connect two free DNA ends together for DSB repair, in regions that contain microhomology [47, 48]. However, Polθ-mediated alt-EJ is error-prone and has a unique mutation signature of small insertions or deletions at the DSB sites [47, 49]. The exceptional mutation signatures of Polθ provide insight into how translesion polymerase mediates repeat expansion. Besides DSB repair, recent studies also suggested that Polθ can bypass a 6-4 double thymidine (TT) photoproduct and assist polymerase η (Polη) in bypassing UV-induced DNA damage during nucleotide excision repair (NER) [50, 51]. Because Polθ participates in multiple repair mechanisms, it is a good candidate for studying repeat expansion during DNA repair.

1.7. Characteristics of Translesion Polymerase θ

Polθ expression needs to be tightly regulated. The overexpression of Polθ is found in multiple cancer types and is associated with poor survival in breast cancer, chromatin instabilities, increased somatic mutations, and promotes DNA synthesis under replication stress [52-60]. The absence of

Figure 1.2: The structure for Polθ. **(A)** The polymerase domain of Polθ [6].

Polθ in cells results in increased formation of micronuclei, low success rate in CRISPR/Cas 9- mediated mutagenesis, unregulated replication, and sensitizes cells to DNA damaging agents [61-63]. Like other translesion polymerases, Polθ has the ability to synthesize DNA beyond the damaged sites [50, 59, 64-66]. Polθ has two isoforms: 1) a 290 kDa protein that contains a N-terminal functional helicase domain, a central domain for scaffolding Rad 51 from performing homologous recombination (HR) [67] and a polymerase domain [54, 68]; 2) a 198 kDa protein that lacks a helicase domain [69]. Polθ is a multifunctional polymerase that participates in both DNA repair and DNA replication. The N-terminal helicase domain of Polθ has ATPase activity and is able to unwind double stranded DNA or RNA-DNA hybrids [70]. Studies also showed that the functional N-terminal of Polθ is essential for interstrand cross-links repair (ICL) in *Drosophila* and alt-EJ for

double strand break (DSB) repair *in vitro* [55, 68, 71]. Its C-terminal polymerase domain contains 3 insertion loops (insert 1, insert 2, and insert 3) (Figure 1.2). Insert 2 and insert 3 are within the palm of the polymerase, and insert 2 was later found to be essential for alt-EJ [55]. Besides the helicase domain is important for DSB repair, its polymerase domain is also required for cell survival after ionizing radiation treatments, for DNA synthesis, and for causing a small expansion *in vitro* [55, 57, 71-73]. Although replication is limited in neurons, neurons still need DNA polymerases to re-synthesize and complete DNA repair. **Therefore, we are going to study the role of Polθ in TNR expansion in DNA synthesis in respect of TNR regions.**
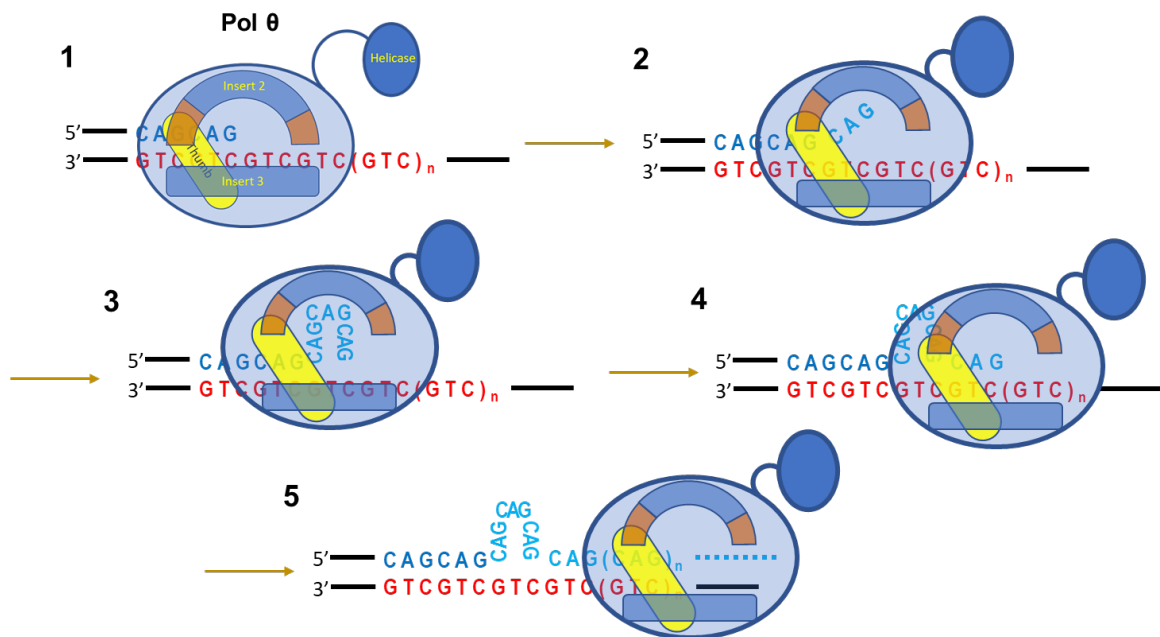
Figure 1.3: Proposed model of Polθ induced CAG repeat sequence during DNA synthesis.

1.8. Polymerase Facilitated Expansion

We will focus on the properties of translesional polymerases that allow them to synthesize through TNR regions. Previous studies demonstrated that the ability of Polβ to promote hairpin retention synthesis depends on MutSβ activity [20, 24]. Other studies highlighted the multi-functional translesion Polθ as an interesting potential candidate for an alternative translesional polymerase for various DNA repair pathways [6, 45, 49, 52-54, 61-64]. Polθ has been shown to participate in DNA repair and to regulate repair pathways, replication and lesion bypass, replication timing, genome stability maintenance, and gene editing [6, 45, 49, 52-54, 61-64]. Polθ was shown to have a larger active enzymatic site for lesion by pass and DNA synthesis during DNA replication [64, 72]; in contrast, other studies showed that Polθ can cause small expansions *in vitro* [45, 55, 56]. In the neurons of HD patients, replication-mediation expansion is uncommon because neurons do not replicate. The major DNA synthesis pathways that are required in neurons are BER, NER, transcription-coupled MMR, and DSB repair [19]. **Hence, studies showed that Polθ was involved in BER, NER, and DSB repair, Polθ can be a good target for inhibiting the progression of TNR expansion diseases through limiting its translesion synthesis properties.**

Chapter 2

Materials and Methods

2.1. Chemicals and Reagents

- Amersham: ECL Detection Reagent, ECL Select™ Detection Reagent

- Calbiochem: Hydroxylapatite (HAP)

- Fisher Biotech: n-butanol, iso-propanol, methanol, dNTPs, protein G-agarose beads, phenol

- GE Healthcare: PreScission Protease, HisTrap HP, Q Sepharose, S Sepharose, Mono Q, Mono S, Heparin, Phenyl Sepharose, S100, S200, illustra MicroSpin G-25 columns

- Gibco: FreeStyle 293 Expression Medium

- HyClone: fetal bovine serum (FBS), newborn serum (NBS), trypsin protease, RPMI 1640 Media with glutamine (RPMI), Dulbecco's High Glucose Modified Eagles Medium (DMEM), HYQ SFX-Insect

- New England Biolab (NEB): AgeI, NotI, XhoI, EcoRI, BseRI, BsmBI, BglI, PstI, Klenow fragment DNA polymerase, T7 DNA Polymerase (unmodified)

- Perkin Elmer: [γ-^{32}P]-ATP

- Phenix Research Products: GelGreen Nucleic Acid Stain

- Research Products International Corp (RPI): 2xYT, agar, myo-inositol

- Sigma: acrylamide (A9099), N,N'-methylenebis(acrylamide) (Bis-acrylamide), potassium phosphate monobasic (KH$_2$PO4), potassium phosphate dibasic (K$_2$HPO4), potassium hydroxide (KOH), sodium

hydroxide (NaOH), sodium citrate ($Na_3C_6H_5O_7 \cdot 2H_2O$), sucrose, boric acid, magnesium (II) chloride ($MgCl_2$), manganese(II) chloride, hydrochloric acid (HCl), Trizma® base (Tris), HEPES, polyehylene glycol (PEG-8000), sodium dodecyl sulfate (SDS), sodium bisulfate, Tween-20, triton X-100, NP-40, potassium acetate, Tris(2-carboxyethyl)phosphine hydrochloride (TCEP), bovine serum albumin (BSA), Dithiothreitol (DTT), Ethylenediaminetetraacetic acid (EDTA), agarose, imidazole, ethidium bromide (EtBr), polyvinylpyrrolidone (PVP), lithium chloride (Li_2Cl), β-Mercaptoethanol, 2,2,2-trichloroethanol (TCE), urea, cesium chloride (CsCl), ammonium persulfate (APS), TNM-FH, acetate acid, monoclonal ANTI-FLAG® M2 antibody (F3165)

- QuantaBio: PerfeCTa STBR Green FastMix, repliQa™ HiFi Assembly Mix (Gibson Assembly)

- Roche: Adenosine triphosphate (ATP), blocking reagent (11096176001)

- USB: Heparin, T4-PNK

- Santa Cruz: Tubulin antibody

- Novus: Polθ (1C11) antibody

- LifeSpan BioSciences, Inc: Polθ (LS-C118709-100) antibody

- Invitrogen: prolong diamond antifade mountant with DAPI

2.2. Basic Techniques

2.2.1. Buffer Preparation

All solutions and cell culture media were prepared with de-ionized distilled water (ddH$_2$O). Solutions were sterilized either by autoclaving for 20 min at 121°C or filtering through a 0.22 µm filter. All buffers for protein purification and cell culture were stored at 4°C.

2.2.2. Agarose Gel Electrophoresis

Agarose gel electrophoresis was run in 1x TAE (40 mM Tris, 20 mM acetic acid, and 2 mM EDTA) buffer. DNA samples were prepared for analysis by adding 6x DNA loading dye [100 mM Tris●HCl (pH7.6), 60% (v/v) glycerol, 60 mM EDTA, 0.03% (w/v) bromophenol blue, 0.03% (w/v) xylene cyanol FF]. Gels were either stained with 0.5 µg/mL EtBr or 0.25x GelGreen Nucleic Acid Stain. DNA fragments were visualized, and images were captured by using a ChemiDoc™ MP Imaging System (Bio-Rad).

2.3. Cell Cultures

Human lymphocytes: GM14044, GM03643, and HL60 were cultured in RPMI supplemented with 15%, 20%, and 10% fetal bovine serum (FBS), respectively. A large quantity of suspension HeLa S3 cells were cultured in RPMI 1640 with 5% FBS and 5% NBS for making nuclear extracts. Human fibroblast: GM04204 (WT), GM04210 (HD), GM04230 (HD), GM04212 (HD), GM04208 (HD), GM04220 (HD), GM21756 (HD), GM09197 (HD), and GM02153 (WT) were cultured in minimum essential medium (MEM) with 15% FBS. All human cell lines

were maintained in 5% CO_2 at 37°C. High 5 insect cells were grown in TNM-FH (Sigma) and supplemented with 10% heat inactivated FBS. SF9 cells were cultured in GE Healthcare HyClone™ SFX Insect™ Cell Culture Media. Both insect cells were maintained at 27°C.

2.4. Cell Fractionations and Whole Cell Lysates for Western Blotting

Cells were harvested by centrifugation at 180 x g for 10 min at 4°C. They were resuspended and washed twice in ice cold DPBS. Cells were then resuspended with buffer A [10 mM HEPES●KOH (pH7.9), 10 mM KCl, 1.5 mM $MgCl_2$, 0.34 M sucrose, 10% glycerol, 1 mM DTT, 0.1% Triton X-100 and protease inhibitors] and incubated on ice for 8 min. Nuclei were collected at 1,300 x g for 5 min at 4°C. The pellets were washed twice with ice cold buffer A. To recover the chromatin binding proteins, 1:1 (v/v) ice cold 0.2 M HCl were used to denature and disassociate from the chromatin by incubating for 10 min on ice. The acidic solutions were then neutralized with 1:1 (v/v) of 1 M Tris ●HCl (pH8.0) with 1 U/µL of Benzonase® Nuclease (Sigma) and incubated on ice for 0.5-1 h to allow DNA digestions. The digested genomic DNA and insoluble proteins were removed by centrifugation. Whole cell lysates were prepared by resuspending cells in 1:1 (v/v) ice cold 0.2 M HCl and incubate on ice for 30 min and neutralized with 1:1 (v/v) of 1 M Tris ●HCl (pH8.0) with 1 U/µL of Benzonase® Nuclease (Sigma) for further releasing the proteins from genomic DNA. All fractions were clarified by high-speed centrifugation at maximum speed (>21,000 x g) for 15 min at 4°C. Protein concentrations were determined with a Bradford protein assay according to manufacturer's procedures (BioRad). Extract samples were boiled for 5 min at

95°C and loaded on a home-made 6%/15% two-layers polyacrylamide gel containing final concentration of 0.5% 2,2,2-trichloroethanol (TCE). Samples were transferred to a nitrocellulose membrane and blocked with 5% non-fat milk in TBST+ 0.1% Tween 20 for 1 h at room temperature followed by a standard western blotting procedure. Membranes were incubated with 1:500 for anti-Polθ antibody (Novus/ LifeSpan/Abcam) or 1:1,000 for anti-α-tubulin (Santa Cruz) and 1:500 for anti-Lamin A/C antibodies (ProteinTech) in 5% milk solution overnight at 4°C with rotation. Membranes were washed three times with TBST+ 0.1% Tween 20 and incubated with 1:10,000 secondary antibodies with respect to primary antibodies used for 1 h at room temperature in 5% milk in TBST+ 0.1% Tween 20 after primary antibodies incubation. After washing out secondary antibodies three times with TBST, immunoreactive bands were visualized and quantified using a BioRad Imaging system and software (BioRad).

2.5. Protein Expression Vectors and Other Plasmids

2.5.1. Plasmid Construction

The pLEXm-Polθ90 (wild type Polθ polymerase domain) gene was kindly provided by our collaborator- Dr. Wei Yang (NIH). pSUMO3-PolθΔi2 (Polθ insertion 2 deleted polymerase domain mutant) was a gift from Dr. Richard T. Pomerantz (Temple University) and was cloned into pLEXm vector with AgeI and XhoI; the gene was later renamed as pLEXm-θΔi2. The polymerase Q full length gene was purchased from transomic technology (Clone ID: BC172289). Polymerase θ full length was cloned into pLEXm by Gibson assembly (SGI-DNA, GA1200) and cloned into the pEF-1a-Flag-HA gateway acceptor vector by a

Gateway LR Clonase Enzyme Mix (Invitrogen). Polymerase θ polymerase domain was also cloned into the pCW57.1-GFP vector, a gift from Dr. Asaithamby Aroumougame (UTSW), by Gibson assembling for inducible expression, and immunofluorescent studies. Polymerase δ and RFC bacterial expression vectors were a gift from Dr. Yoshihiro Matsumoto (Fox Chase Cancer Center). Polymerase δ insect cell expression vectors were a gift from Dr. Paul Modrich (Duke University). RPA bacterial expressing vector was a gift from Dr. Marc Wold (University of Iowa). pZLCv2-3xFLAG-dCas9-HA-2xNLS was a gift from Stephen Tapscott (Addgene plasmid # 106357) [74].

2.5.2. Large Scale ssDNA Extractions

M13mp18 is a phage formation plasmid; the ssDNA phage can be found in the cleared cultural medium and the dsDNA can be found in the bacterial pellet. For ssDNA substrates needed for *in vitro* DNA synthesis, either the dsDNA or the ssDNA was used to perform transformation with XL1-Blue bacteria. The transformed XL1-Blue was inoculated with 100 μL of XL1-Blue overnight culture in soft agar and could form plaques after incubation at 37°C overnight. A single plaque was isolated and amplified by incubating with fresh1:100 diluted XL1-Blue culture for 6-8 h at 37°C. The clarified supernatant was then scaled up to a 200 mL culture for ssDNA isolation. Fifty grams of polyethylene glycol 8000 per liter and 36 g of NaCl per liter were used to precipitate the phage from the medium. The phage precipitant was collected by centrifugation at 4,500 x g for 15 min at 4°C. The phage pellet was resuspended in TE buffer [100 mM Tris●HCl (pH8.0) and 10 mM EDTA (pH8.0)]. The phage was broken, and phage proteins were

removed by using TE buffer-equilibrated phenol. ssDNA was obtained by ethanol precipitation.

2.5.3. Large Scale Plasmid Extraction

Because 293T/293GnTi⁻ transfection required large quantity of pure plasmids, 1 L of pLEXm- θ90 and pLEXm- θΔi2 transformed bacteria were used. Plasmid-containing bacteria were collected and lysed with cold 20 mL of solution 1 [0.9% glucose, 10 mM EDTA, 25 mM Tris•HCl (pH8.0)] per 1 L of culture and 0.1 g of lysozyme per 1 L of culture. The mixture was incubated on ice for 10 min. Forty milliliters of freshly made solution 2 (0.2 M NaOH and 1% SDS) per 1 L of culture were added, mixed gently to avoid shearing dsDNA, and incubated on ice for 10 min. The lysed bacteria extract was neutralized with 30 mL of cold freshly made 3 M potassium acetate (pH 4.8). The white precipitate was removed by spinning at 10,000 x g for 1 h at 4°C and filtered through 4 layers of cheese cloth. 0.6 x of the supernatant volume of Isopropanol was added to the filtered supernatant and mixed well. The isopropanol-plasmid solutions were left to incubate at room temperature for 30 min. The precipitated plasmids were collected by high speed centrifugation for 30 min. The DNA pellet was washed with 70% ethanol and air dried. To increase the purity of the plasmids, the DNA pellet was resuspended in TE buffer with RNase. Cesium Chloride with ethidium bromide (CsCl -EtBr) binding was performed to remove RNase and other impurities. Briefly, for each gram of DNA solution, 1.08 g of dehydrated CsCl and 50 μL of freshly prepared 10 mg/mL EtBr were added. The density of the CsCl –EtBr -DNA solution should be about 1.55 g/mL. The plasmids were separated by ultracentrifugation

(Beckman MT65) at 45,000 rpm for 18 h at 20°C. n-Butanol was used to remove EtBr from the recovered plasmids. The EtBr-free plasmids solution was dialyzed against 1 L of TE buffer for 3-4 times and at least 4 h each time to remove butanol, CsCl, and EtBr. Trace amounts of butanol, CsCl, and EtBr would decrease transfection efficiency. The plasmid concentrations were determined, and the plasmids were stored at 4°C.

2.6. Protein Purification

Every step in the protein purification was performed at 4°C or on ice.

2.6.1. Polθ Wild Type and Insert 2 Deleted Mutant Purification

Cells (293T) were seeded in DMEM supplied with 10% FBS and poured into 60 x 150 mm^2 dishes to reach 60% confluency. pLEXm- Polθ90 or pLEXm- PolθΔi2 were transfected with PEI-max (Polyscience, Inc. Cat# 24765-1) at a 1:4 ratio. Transfected cells were incubated in a humidified environment at 37°C with 5% CO_2 for three days. Cells were collected by scraping gently and centrifuging at 180 x g for 10 min at 4°C. They were then washed with cold PBS once and frozen at -80°C for later. Cell pellets were resuspended and lysed in 10mL of buffer B_1 [20 mM Tris●HCl (pH 8.0), 1 M NaCl, 0.1 mM EDTA, 2 mM DTT, 5% glycerol] with 0.1% NP-40 and proteinase inhibitors. The cell lysate underwent 2 min-sonication-cycles (2 s on and 2 s off) with a 2-min pause between cycles until becoming less viscous or runny. The sonicated lysate was cleared by ultra-centrifugation at 142,030 x g for 1 h at 4°C. The clarified lysate was then incubated with equilibrated amylose resin (NEB, Cat# E8021) at 4°C with rotation for 1-2 h. The mixture was transferred

to a clear cooled empty column. The resin was washed with 30 mL of buffer B_1.

Then, Polθ was eluted with 10 mL of elution buffer [20 mM Tris●HCl (pH8.0), 300

mM NaCl, 0.1 mM EDTA, 2 mM DTT, 10% glycerol, 100 mM D-maltose]. Fractions

containing Polθ (170 kDa) were identified with SDS-PAGE. The Polθ containing

fractions were combined and the protein concentration was determined with a

Bradford protein assay. One unit of PreScission Protease (PP) (GE Healthcare

Life Sciences, Cat# 27084301) was mixed with every 100 µg of protein and allowed

to digest overnight at 4°C with rotation. The digested sample was then diluted to a

final salt concentration of 100 mM NaCl with buffer A_1 [20 mM Tris●HCl (pH 8.0),

0 mM NaCl, 0.1 mM EDTA, 2 mM DTT, 10% glycerol]. The diluted sample was

passed through a Mono S column (GE Healthcare Life Sciences), washed with

10% buffer B_1 for at least 15 column volumes and eluted with a gradient of 10-80%

buffer B_1. The protein containing fractions were identified by SDS-PAGE and

diluted 1:1 with storage buffer [20 mM Tris●HCl (pH 8.0), 300 mM NaCl, 0.1 mM

EDTA, 4 mM TCEP, 10% glycerol]. For long-term storage, the protein was divided

into small aliquots, snapped frozen in liquid nitrogen, and stored at -80°C.

Besides, using 293T cells, 1 L of 293GnTi⁻ cells were grown in Freestyle

293 Expression Medium (Gibco) supplied with 1% FBS in suspension. One liter of

293GnTi⁻ cells was transfected with 1 µg of plasmid with 4 mg of PEI-Max.
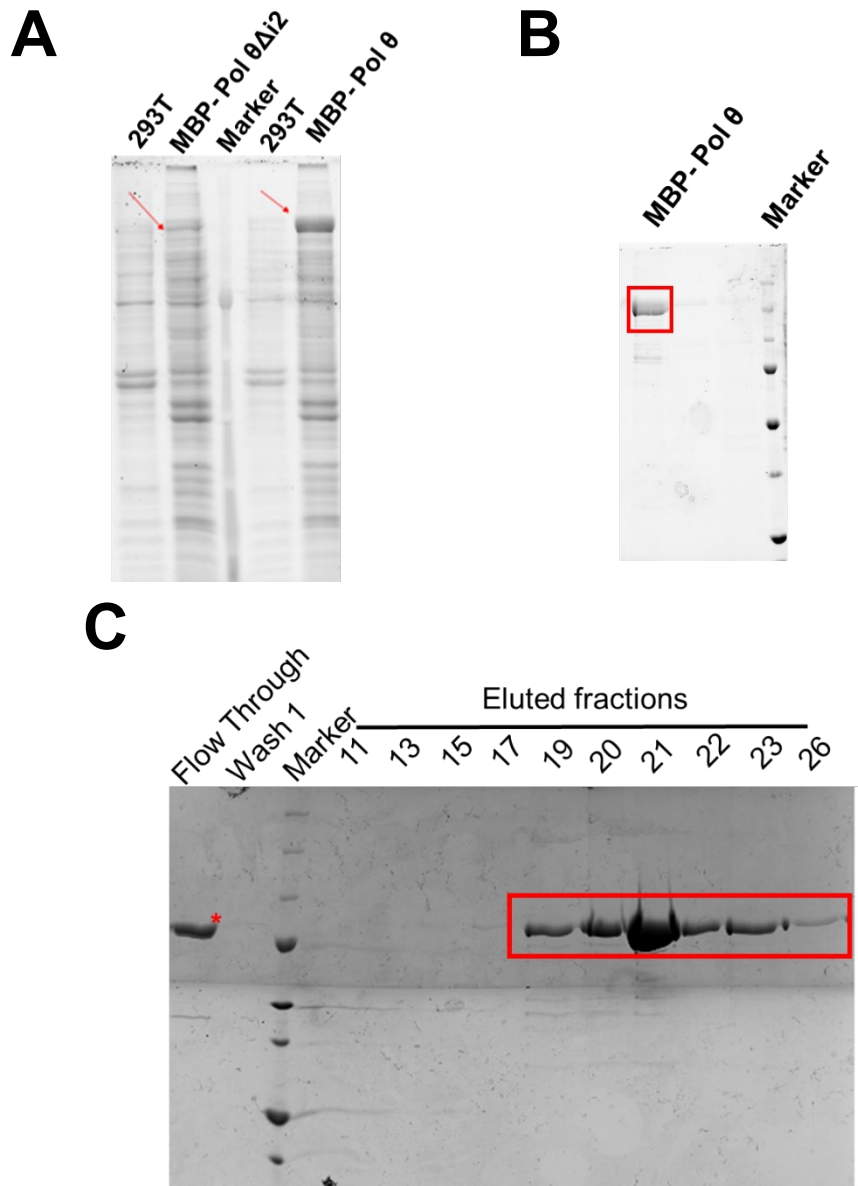
Figure 2.1: Expression and purification of Polθ WT. **(A)** Cell lysate of Polθ after sonication (Red arrows indicate MBP- Polθ). **(B)** MBP-Polθ purified via amylose resin, showed in red rectangle. **(C)** Polθ purified via Mono S to remove MBP tag (red rectangle indicates Polθ in various eluted fractions, and red asterisk showed the cleaved MBP tag). Purities of the protein were analyzed by 10% SDS-PAGE after each purification steps.

2.6.2. Polδ Purification

The four-subunit Polδ was co-expressed through a baculoviral system with either SF9 or High 5 insect cells. Cells were collected after being infected for 48 h through centrifugation at 180 x g for 15 min at 4°C. The pellet was lysed with 5 mL of extraction buffer per g of cells [20 mM Tris•HCl (pH 8.0), 160 mM NaCl, 0.1 mM EDTA, 2 mM DTT, 10% glycerol, 0.2% NP-40 and proteinase inhibitors]. Cell lysate was prepared by using a homogenizer. The lysate was cleared by centrifuging with 53,200 x g for 1 h at 4°C, then loaded to a HisTrap column (GE Healthcare Life Sciences). The HisTrap column was washed with buffer A_1 [20 mM Tris•HCl (pH 8.0), 200 mM NaCl, 5 mM 2-mercaptoethanol and 10% glycerol]. The column was washed a second time with 15% buffer B_1 [20 mM Tris•HCl (pH 8.0), 200 mM NaCl, 500 mM imidazole, 5 mM 2-mercaptoethanol and 10% glycerol]. Polδ was eluted with a 15%-80% buffer B_1. All the fractions containing protein were combined and diluted in 1:1 with buffer A_2 [20 mM Tris•HCl (pH 8.0), 0 mM NaCl, 2 mM DTT and 10% glycerol]. The diluted po δ was passed through a Mono Q anion exchange chromatography column. The Mono Q column was washed in the order of 10% buffer B_2 [20 mM Tris•HCl (pH 8.0), 1 M NaCl, 2 mM DTT and 10% glycerol], followed by 20% Buffer B_2. Polδ was eluted with a 20%-100% buffer B_2 gradient. Fractions were pooled and diluted to 150 mM NaCl with buffer B_2. The diluted fractions were passed through a Mono S cation exchange chromatography column. The Mono S was washed and eluted as Mono Q. Fractions with Polδ were identified and pooled. For long-term storage, protein stored in a final concentration of 10% sucrose, 1 mg/mL BSA, 10% glycerol, and 2 mM TCEP, then the purified

protein was divided into small aliquots, snapped frozen in liquid nitrogen, and stored in -80°C.
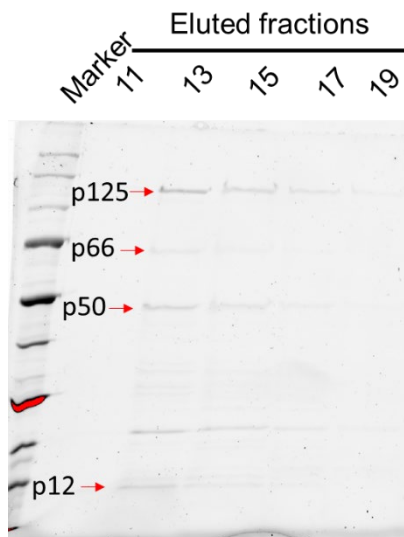
Figure 2.2: Purified human Polδ from bacteria expression system after SP Sepharose (cation exchange column).

Bacterial expression of Polδ was also expressed and purified as described [75]. In summary, the Polδ expressed bacteria at 16°C overnight and was lysed in lysate buffer [40 mM HEPES●NaOH (pH 7.5), 200 mM NaCl, 10% glycerol, 1% triton X-100, and proteinase inhibiters]. The lysate was sonicated at a power 2-3 setting for 15 s on and 15 s off. After that the lysate was clarified by ultracentrifugation at 350,000 rpm for 1 h at 4°C. The clarified lysate was adjusted to final concentration of 30 mM of imidazole with buffer His-B [40 mM HEPES●NaOH (pH 7.5), 100 mM NaCl, 10% glycerol, 500 mM of imidazole, and proteinase inhibiters]. The lysate was then passed through a HisTrap column that was equilibrated with buffer His-A [40 mM HEPES●NaOH (pH 7.5), 100 mM NaCl,

22

10% glycerol, 30 mM of imidazole, proteinase inhibiters], washed with buffer SP-B [40 mM HEPES•NaOH (pH 7.5), 1 M NaCl, 10% glycerol, 0.01% triton X-100, and proteinase inhibiters], followed by washing with 12% buffer His-B. Polδ was eluted by 60% buffer His-B. The Polδ containing fractions were loaded to the SP sepharose (GE Healthcare Life Sciences) that was equilibrated with buffer SP-A [40 mM HEPES•NaOH (pH 7.5), 100 mM NaCl, 10% glycerol, 0.01% triton X-100, and proteinase inhibiters]. The Polδ binded SP sepharose was washed by 20% of buffer SP-B and then 50% of buffer SP-B. The purified protein was eluted at 100% of buffer SP-B. Polδ was stored in a final concentration of 10% sucrose, 1 mg/mL BSA, 10% glycerol, 1 M NaCl, and 2 mM TCEP, and was divided into small aliquots, snapped frozen in liquid nitrogen, and stored at -80°C.

2.6.3. RFC Purification

The five subunits of RFC were co-expressed through a baculoviral system with either SF9 or High 5 insect cells. Cells were collected after being infected for 48 h through centrifugation at 180 x g for 15 min at 4°C. The pellet was lysed with 8 mL of RFC extraction buffer per g of cells [25 mM Hepes•KOH (pH 7.5), 350 mM NaCl, 0.1 mM EDTA, 2 mM DTT, 10% glycerol, 1.5 mM $MgCl_2$ and proteinase inhibitors]. The cell lysates were prepared by using a homogenizer. The homogenized lysate was spun at 53,200 x g for 1 h at 4°C. Supernatants were diluted with buffer A_1 [50 mM potassium phosphate (pH 7.5), 0 mM NaCl, 2 mM DTT, 0.5 mM EDTA, 0.01% NP-40 and 10% glycerol] to 120 mM NaCl and loaded onto a Q sepharose column (GE Healthcare Life Sciences). The column was washed with 15% buffer B_1 (buffer A_1 with 1 mM KCl). RFC was eluted with a 15%-

70% buffer B_1 gradient. RFC-containing fractions were determined by western blotting with the his-tag antibody (Genescript). All the RFC containing the fractions were combined and diluted with buffer A_1 to about 100 mM KCl. The diluted RFC fractions were passed through a heparin column. The heparin column was washed in the order of 10% buffer B_1, then 20% buffer B_1. Purified RFC was eluted with a 20%-80% buffer B_1 gradient. RFC fractions were pooled and diluted to 150 mM KCl again for passed through a Mono S cation exchange chromatography column. The Mono S was washed with 15% buffer B_1 and eluted with a gradient of 15%-60% Buffer B_1. Fractions with RFC were identified and pooled, and diluted to 150 mM KCl for Mono Q under the same conditions as those of Mono S. RFC fractions were concentrated with a Millipore Ultrafree concentrator to 250μL and loaded to a S200 increase column (GE Healthcare Life Sciences) equilibrated with 18% buffer B_1. RFC was stored in a final concentration of 10% sucrose, 1 mg/mL BSA, 10% glycerol, 180 mM KCL, and 2 mM TCEP, and was divided into small aliquots, snapped frozen in liquid nitrogen, and stored at -80°C.

Mono Q column could be substituted by a heparin column under the same purification condition with the Mono S.
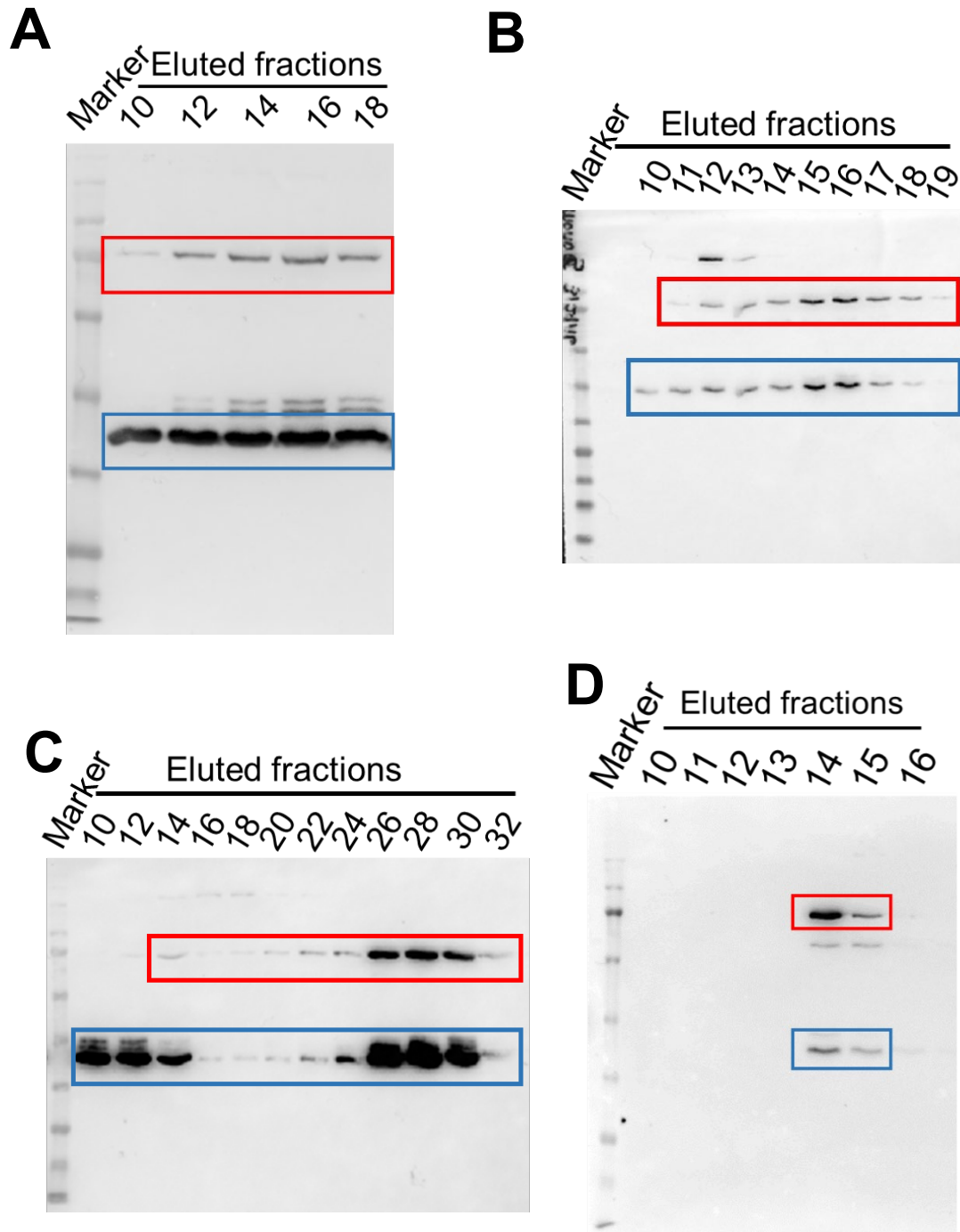
Figure 2.3: Western blot results with his-tag antibody (p140N555-His$_6$ and His$_6$-p38) for RFC purification. **(A)** RFC purification after Q sepharose. **(B)** RFC purification after Mono S. **(C)** RFC purification after heparin. **(D)** RFC purification after S200. Red rectangles indicated p140N555-His$_6$; blue rectangles showed His$_6$-p38 subunits.

2.6.4. Polβ Purification

Polβ was expressed in SF9 and collected after 48h infection. The cells were collected and washed as mentioned earlier. The cell pellet was lysed with 10mL of buffer A_1 [25 mM Hepes•NaOH (pH 7.8), 300 mM NaCl, 20 mM imidazole, 5 mM 2-mercaptoethanol and 10% glycerol]. The cell lysates were prepared and clarified as the procedures described earlier (see **2.6.2** and **2.6.3**). The supernatant was passed through a HisTrap column. The HisTrap column was washed with 30% Buffer B_1 (Buffer A_1 with 500 mM imidazole). Polβ was eluted with a gradient of 30%-100% buffer B_1 gradient. Protein containing fractions were combined and diluted to 150mM NaCl with buffer A_2 [25 mM Hepes•NaOH (pH 7.8), 0 mM NaCl, 2 mM DTT and 10% glycerol] and loaded on to a heparin column or a Mono S column. The column was washed with 15% buffer B_2 and eluted with a gradient of 15%-80% buffer B_2. Proteins were concentrated and loaded onto a S100 or a S200 with 30% buffer B_2 to remove nucleases. Polβ was stored as described in **2.6.2**.
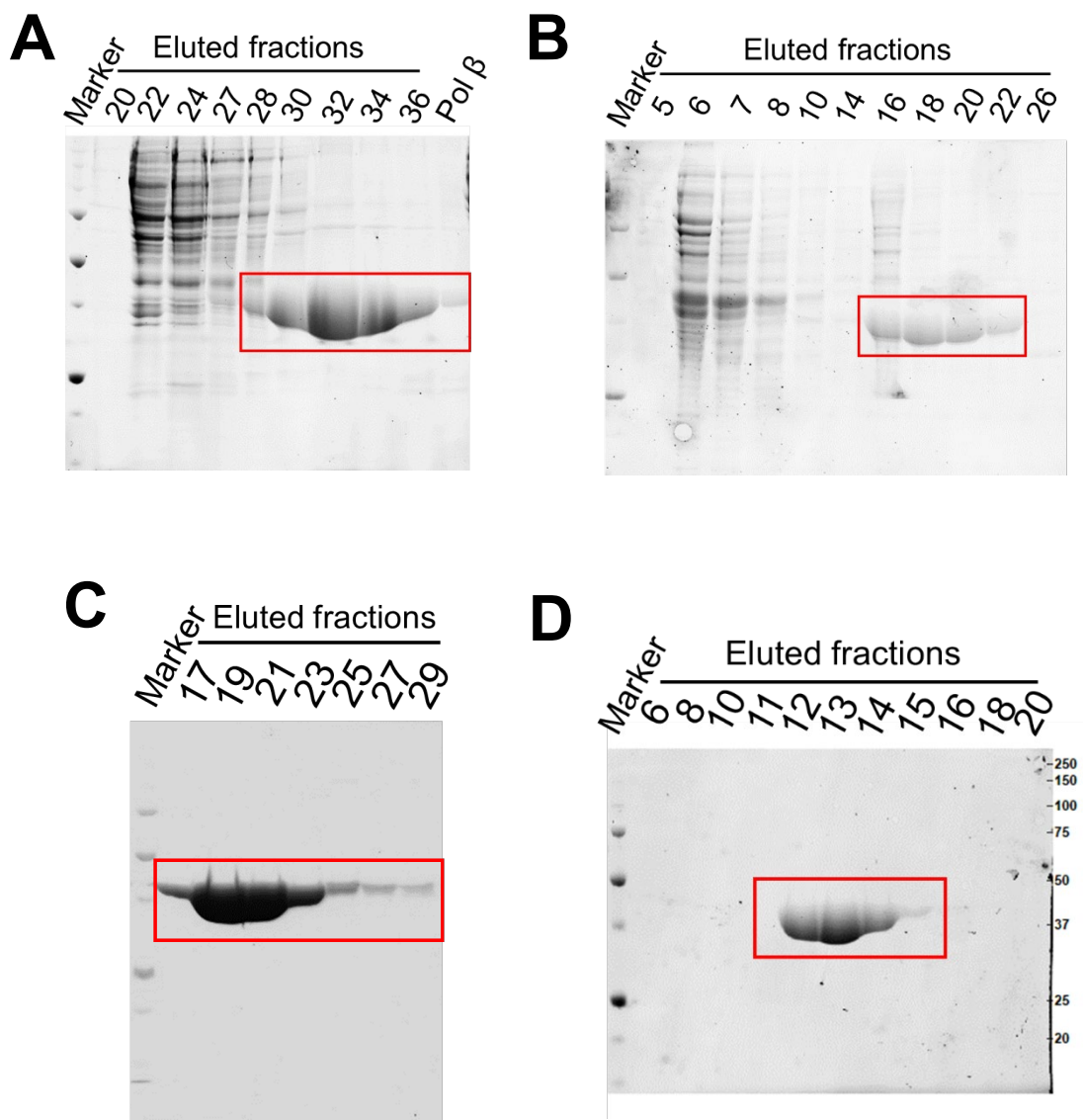
Figure 2.4: Unstained TCE gel results for Polβ purification. **(A)** Polβ purification after HisTrap column. **(B)** Polβ purification after heparin column. **(C)** Polβ purification after Mono S. **(D)** Polβ purification after S200. Red rectangles indicated Polβ protein.

2.6.5. RPA Purification

RPA was expressed in BL21(DE3) or Rossetta. hRPA should be expressed within four days of plasmid transformation. A single colony of RPA-BL21 was grown overnight and diluted 1:100 to 1 L of 2xYT (RPI) without shaking at 37°C overnight. The next morning, the cells were grown with shaking to OD_{600} = 0.8. Then, IPTG was added to a final concentration of 1 mM to allowing the bacteria to express the protein for 3 h at 37°C. The three RPA subunits were purified as previously reported [76]. Briefly, the bacteria pellet was prepared in buffer A_1 [30 mM HEPES●KOH (pH 7.8), 0.25 mM EDTA, 50 mM KCl, 0.5 M NaSCN, 0.5% (w/v) myo-inositol, 1 mM DTT, and proteinase inhibitors] with 0.1% (v/v) NP-40, followed by 2 s on and 4 s off sonication for 5 min. The lysate was cleared by at least 53,200 x g centrifugation for 1 hour at 4°C and loaded onto an Affi-Gel Blue (GE Healthcare) column per liter of bacterial culture with a flow rate of at least 1 mL/min. The column was washed with at least 10x column volumes of buffer A_1, then washed with wash buffer [30 mM HEPES●KOH (pH 7.8), 0.25 mM EDTA, 800 mM KCl, 0.5 M NaSCN, 0.5% (w/v) Myo-Inositol, 1 mM DTT and proteinase inhibitor]. RPA was eluted with 15x column volumes of buffer B_1 [30 mM HEPES●KOH (pH 7.8), 0.25 mM EDTA, 50 mM KCl,1.5 M NaSCN, 1 mM DTT and proteinase inhibitor]. The eluted RPA was loaded to a 1 mL HAP column which equilibrated with buffer A_1. After loading the sample, the HAP column was washed with 18x column volumes of buffer A_1 and eluted with elution buffer [80 mM potassium phosphate (pH7.8), 30 mM HEPES●KOH (pH 7.8), 0.25 mM EDTA, 50 mM KCl, 0.5% (w/v) myo-inositol, 1 mM DTT and proteinase inhibitor]. The RPA containing

fractions were pooled and diluted with 3x sample volumes with buffer A_2 [30 mM HEPES•KOH (pH 7.8), 0.25 mM EDTA, 0 mM KCl, 0.5% (w/v) Myo-Inositol, 1 mM DTT and proteinase inhibitor]. The diluted sample was loaded onto the Mono Q (anion exchange column). The Mono Q was washed with 15x column volumes of buffer A_2 containing 50 mM KCl and 15x column volumes of buffer A_2 containing 100 mM KCl. RPA was eluted with 15mL of 200-500 mM KCl gradient. RPA was eluted at about 300 mM KCl. The purified RPA was washed in buffer A_2 containing about 300 mM KCl and supplemented with 1 mM TCEP and proteinase inhibitors. RPA was aliquoted and stored at -80°C.
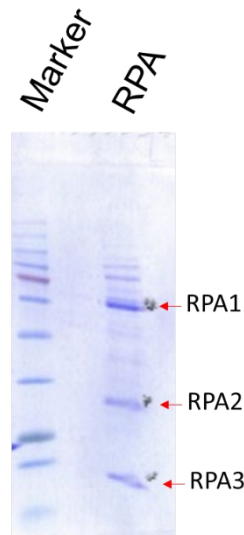
Figure 2.5: Coomassie brilliant blue staining for RPA after Mono Q column.

2.6.6. PCNA Purification

The his-tagged PCNA homotrimer was purified by previous lab members, Dr. Bailin Zhao and Dr. Janice Ortega. PCNA was expressed in BL21(DE3) or Rossetta with 1 mM IPTG for 3h at OD_{600} = 0.8. The cell pellet was lysed in lysis buffer [25 mM Tris●HCl (pH 7.4), 25 mM NaCl, 0.01% NP-40, 5 mM 2-mercaptoethanol and proteinase inhibitors] and sonicated. The lysate was clarified by ultracentrifugation as described as above and passed through a HisTrap column. PCNA was eluted with buffer A_{His} [25 mM Tris●HCl (pH 7.4), 25 mM NaCl, 0.01% NP-40, 5 mM 2-mercaptoethanol and proteinase inhibitors] contained 20 mM to 400 mM imidazole. PCNA was then further purified with a set of columns in the following order: phenyl sepharose, S sepharose, and heparin columns. Contaminants bind to the first two columns, allowing highly purified PCNA to bind to heparin column. Before washing and eluting PCNA from the heparin column, phenyl sepharose and S sepharose were removed before elution. Heparin column was washed with buffer $A_{Heparin}$ [25 mM Tris●HCl (pH7.4), 1 mM EDTA, 0.01% NP-40, 10% glycerol, 5 mM DTT, and proteinase inhibitors] with 200 mM NaCl and eluted with a gradient of 200 mM to 700 mM NaCl. The eluted PCNA was concentrated and loaded onto a S200 gel exclusion column for removing nucleases and reducing salt concentration to 200 mM NaCl. Highly purified PCNA was aliquoted and stored at -80°C.

2.7. Making HeLa S3 Nuclear Extract

A 6 L culture of HeLa S3 was grown for making nuclear extracts, according to a previous study [77]. HeLa S3 cells were washed with wash buffer [20 mM Hepes●KOH (pH7.5), 5 mM KCl, 0.5 mM MgCl$_2$, 0.2 M Sucrose, 0.5 mM DTT and proteinase inhibitors]. Cells were washed once with hypotonic buffer [20 mM Hepes●KOH (pH7.5), 5 mM KCl, 0.5 mM MgCl$_2$, 0.5 mM DTT and proteinase inhibitors]. Nuclei were gently extracted by hypotonic buffer and homogenization (Douncer), and collected by low speed centrifugation. The nuclei were resuspended with final concentration of150 mM NaCl and extraction buffer (50 mM Hepes●KOH (pH7.5), 10% sucrose, 0.5 mM DTT and proteinase inhibitors) to remove genomic DNA. Ammonium sulfate (*final concentration of*42% (w/v)) was used to precipitate out the target nuclear proteins. The precipitated nuclear proteins were dialyzed against dialysis buffer [25 mM Hepes●KOH (pH7.6), 50 mM KCl, 0.1 mM EDTA, 2 mM DTT and proteinase inhibitors] to reduce salt concentration. The dialyzed extract was clarified by centrifugation. The clarified extract was aliquoted and stored at -80°C.

2.8. General DNA Annealing

A pair of primer and template at a representative ratio of 1.5:1 was heated to 95°C for 10 min in *final concentration of*167 mM NaCl. The mixture was allowed to slowly cool down to room temperature.

2.9. DNA Substrate Preparation

The 110 bps *in vitro* DNA synthesis substrates with 20 TNR and their control template were synthesized, and gel purified by Sigma. The 35 TNR and 15 TNR substrates were driven from the ssDNA of M13mp18 and purified as *Large scale plasmid/substrate extractions* [78]. To ensure that substrates were fully digested, ssDNA was first annealed with 2 short complimentary oligos, each containing either a BglI or a BsrBI enzyme digestion site. Digested ssDNA was purified by phenol extraction and ethanol precipitation. The purified ssDNAs or synthesized 110-oligomers were then annealed with 5' [^{32}P]-labeled primer (C6277 for 35 TNR and oligo, C6291 for Random control) for in vitro DNA synthesis. For the *Hairpin Retention Assay*, a non-radioactive 5' protected synthesized 15 TNR oligo (mCTG15 or mCAG15) annealed with a ssDNA with 10 TNR that was driven from M13mp18 for hairpin substrates; another 5' protected control primer annealed with either ssDNA with 10 TNR or 15 TNR as a control.

Table 2.1: Primer Extension Substrates

Name of Substrates	Enzyme Digestion Primers		Extension Primer	Length (nt)
(CAG)35 Template	C6360 (BglI)	C6185 (BsrBI)	C6277	189
(CTG)35 Template	C6360 (BglI)	C6185 (BsrBI)	C6277	190
M13mp18/GC (Control)	C6360 (BglI)	C6185 (BsrBI)	C6291	192
(CAG)15 Template	C6360 (BglI)	C6185 (BsrBI)	C6277	125
(CTG)15 Template	C6360 (BglI)	C6185 (BsrBI)	C6277	125
M13mp18/GC (Control)	C6360 (BglI)	C6185 (BsrBI)	C6277	125
(CAG)20 Template	Synthesized by Sigma		C6277	110
(CTG)20 Template	Synthesized by Sigma		C6277	110
Random Template	Synthesized by Sigma		C6277	110

2.10. T7 Endonuclease and Mung Bean Nuclease Digestion for Hairpin Substrates

To ensure a hairpin was formed in the hairpin substrates, 110 ng of a 5' [^{32}P]-labeled hairpin substrate and their non-hairpin controls were exposed to T7 endonuclease and mung bean nuclease, independently, according to NEB protocols. The digested substrates were resolved with a 15% urea-polyacrylamide denaturing gel (8.3 M Urea, 1x TBE, 15% 38:2 Acrylamide solution, 7 µL/mL APS and 0.7 µL/mL TEMED). The gel was analyzed using an Amersham Typhoon Gel and Blot Imaging System.

2.11. *In Vitro* DNA synthesis and Urea-PAGE Electrophoresis

Unless otherwise mentioned, all reactions were performed with 9 nM of DNA substrate, 0.1 µM of purified polymerase (δ, θ, θΔi2) or 0.5 Unit of T7 polymerase (New England Biolabs), 500 nM of PCNA, 50 nM of RFC, and 0.4 fM of RPA in a 40 µL reaction containing 20 mM of Tris●HCl (pH 7.6), 200 µg/mL of BSA, 1.5 mM of ATP, 1 mM of glutathione (Reduced), 0.2 mM of each dNTPs, 50 mM of MgCl$_2$, and 110 mM with KCl incubation for 30 min at 37°C. Synthesis was terminated by 60 µL of a proteinase K solution containing 0.67% (v/v) SDS, 2.5 mM of EDTA, and 0.3 mg/mL proteinase K for 1 h at 37°C. After phenol extraction and ethanol precipitation, the DNA samples were digested with a 0.1 unit of RNase (Roche). After enzyme digestion, the reactions were terminated by adding 2x SSCP (95% formamide, 0.05% bromophnol blue, 0.05% xylene cyanol, and 20 mM EDTA) and heating at 95°C for 7 min. The synthesized samples were loaded to a 6% urea-polyacrylamide denaturing gel (final concentration of 8.3 M Urea, 1x TBE, 6% 38:2 Acrylamide solution, 7 µL/mL APS, and 0.7 µL/mL TEMED) for long DNA synthesis products and 7% urea-polyacrylamide denaturing gel for short synthesis products. The gel was pre-run in 1x TBE buffer [89 mM Tris Base (pH 8.3), 89 mM boric acid, and 2 mM EDTA] at 80 W with a 40 mL gel and 120 W with a 60 mL gel for 30 min before loading the sample. Samples were running at 42 W (1600-1800 V) with a 40 mL gel and 82 W (1800-2100 V) with a 60 mL gel for 2 h 30 min. After running, the gel was dried in a Bio-Rad Model 583 Gel Dryers for 2 h at 75°C. The products were visualized through an Amersham Typhoon Gel and Blot Imaging System.

MgCl$_2$ was substituted by MnCl$_2$ in **Chapter 4**.

2.12. Hairpin Retention Assay and Southern Blot

Unless otherwise mentioned, all reactions were performed with 2nM of DNA substrate, 0.1. μM of purified polymerase (δ, θ, θΔi2), 28 μg of HeLa Nuclear Extract in 20 uL reactions containing 2 mM of Tris●HCl (pH 7.6), 0.2 mg/mL of BSA, 1.5 mM of ATP, 1 mM of Glutathione (Reduced), 0.2 mM of each dNTPs, 50 mM of MgCl$_2$, and 120 mM with KCl incubation for 2 min at 37°C. Reactions were terminated by 30 μL of proteinase K solution containing 0.67% (v/v) SDS, 2.5 mM of EDTA, and 0.3 mg/mL proteinase K for 1 h at 37°C. After phenol extraction and ethanol precipitation, the recovered DNA samples were digested with 0.3 unit of RNase (Roche), Sau96I, HindIII, and BsrBI (New England Biolabs) for at least 2 h. The digested samples were resolved in a 6% urea-polyacrylamide denaturing gel. The gel was pre-run in 1x TBE buffer at 16 W for 30 min before loading the sample and was run at 8 W for 1 h 15 min. The resolved products were transferred to a nylon membrane (GE Healthcare Amersham Hybond™-NX) in 1x TBE at 1 A (38 V) for 1 h at 4°C. The membrane was dried and UV-crosslinked for 10 min. The cross-linked membrane was blocked with hybridizing buffer [0.002% (w/v) SDS, 0.005% (w/v) Polyvinylpyrrolidone (PVP40), 0.002% (w/v) heparin, 1 mM EDTA (pH 8.), 1 M NaCl, 5 mM Tris●HCl (pH 7.6)] for 3 min at 37°C. The membranes were probed with ^{32}P-labeled V6135 (*Table of Oligos*) in Hybridizing Buffer overnight at 37°C. The membrane was washed twice with 2x SSC Buffer (2x SSC + .1% SDS; 2x SSC [3 M NaCl and 0.3 M Sodium Citrate (pH7.6)], then washed

twice with 1x SSC Buffer (1x SSC + 0.1% SDS). The membranes were dried and visualized by an Amersham Typhoon Gel and Blot Imaging Systems

2.13. Lentivirus Generation and Increased Transfection Efficiency of Polθ Expression Virus by Double Freeze-Thaw Cycle

pCW57.1-roGFP- Polθ, PolθΔi2 and GFP control were transfected into 293T cells in 10 mL of DMEM supplied with 10% heat inactivated FBS with jetPrime (PolyPlus Transfection) separately to create lentivirus according manufacturer's instructions. After 2 days of transfection, the virus was collected and concentrated by a Lenti-X Concentrator (Clontech) according to manufacturer's instructions. The concentrated virus was resuspended in freezing medium (65% of DMEM, 30% FBS, final concentration of 8 μg/mL of hexadimethrine bromide, and 5% DMSO) and stored at 4°C for at least 1 h before use.

Transfection efficiency was improved by performing freeze-thaw cycle [79]. HD fibroblasts or mouse neuron cell lines were resuspended with virus containing freezing medium and first placed in -80°C for 6 h and transferred to liquid nitrogen for overnight [79]. The frozen virus-infected cells were thawed in a 37°C water bath for 5 min and replaced at -80°C for 6 h, followed by freezing in liquid nitrogen overnight. After the freeze-thaw cycles, the cells were incubated in fresh growing media until reaching 60% confluency. Cells that overexpressed Polθ were treated for western blotting, con-focal microscopy studies, and chromatin immunoprecipitation.

2.14. Data Quantification and Statistics

The data collected through a Bio-Rad ChemiDoc system were quantified by Bio-Rad ImageLab. All other data collected and imaged by other system were quantified by ImageJ. The ANOVA test was used to calculate the significance levels between samples (GraphPad Prism).

2.15. Table of Oligos

Name of Oligo	Sequence	Application
CTG20 (purified)	CATGATTACGAATTCCTGCTGCTGCTGCT GCTGCTGCTGCTGCTGCTGCTGCTGCTG CTGCTGCTGCTGCTGAAGCTTGGCAC TGGCCGTCGTTTTACAACGTCGTG	Primer extension
CAG20 (purified)	CATGATTACGAATTCCAGCAGCAGCAGCA GCAGCAGCAGCAGCAGCAGCAGCAGCAG CAGCAGCAGCAGCAGCAGAAGCTTGGCA CTGGCCGTCGTTTTACAACGTCGTG	Primer extension
GC (purified)	CATGATTACGAATTCTTCACACAGGAAACA GCTATGACGAGCTCGGTACCCGGGGATC ATGCACTCGAGACATGCAAGCTTGGCACT GGCCGTCGTTTTACAACGTCGTG	Primer extension
C6277 (purified)	ACGTTGTAAAACGACGGCCA	Primer extension
M13-C6291 (purified)	AGCTGGCGAAAGGGGGAT	Primer extension
C6360 (BglI)	CCATTCGCCATTCAGGCTGCGCAA	Enzyme Digestion for Primer Extension
C6185 (BsrBI)	TATCCGCTCACAATTCCACAC	Enzyme Digestion for Primer Extension
M13ctlPE	mAmCmG ACG GCC AGT GCC AAG CTT	T vector cloning

mCTG15	A[mC][mG][mA]CGGCCAGTGCCAAGCTTCTGCTGCTGCTGCTGCTGCT GCTGCTGCTGCTGCTGCTGCTGCTG	T vector sequencing from Eton bioscience Inc.
mCAG15	A[mC][mG][mA]CGGCCAGTGCCAAGCTTCAGCAGCAGCAGCAGCAGCA GCAGCAGCAGCAGCAGCAGCAGCAG	Hairpin retention assay
V6315	CTATGACCATGATTACGAATTC	Hairpin retention assay
V6221 (EcoRI)	CATGATTACGAATTC	Hairpin retention assay
CAG5	CAG CAG CAG CAG CAG	Hairpin retention assay
CAG10	CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG	Probe for Hairpin Retention Assay
CTG5	CTG CTG CTG CTG CTG	Probe for Hairpin Retention Assay
CTG10	CTG CTG CTG CTG CTG CTG CTG CTG CTG CTG	Probe for Hairpin Retention Assay
HD343-R (Purified)	CTGAGGCAGCAGCGGCTGTGCCTGCG	Hd sequence pcr
HD168-F (Purified)	ATGAAGGCCTTCGAGTCCCTCAAGTCCTTC	Hd sequence pcr
CTG (5'Cy3)	CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCATTCCCGGCTACAAGCGGCCGCGAGCAGC	Con-focal FISH
CAG (5'Cy3)	CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCATTCCCGGCTACAAGCGGCCGCGAGCAGC	Con-focal FISH
AgeI-Q90-right-F	GGACCGGTGGGGTTCAAAGACAACAGCC	Polθ Δinsert 2 cloning
PolQ-XhoI-R	GGCTCGAGTTATTACACATCAAAGTCCTTTAGC	Polθ Δinsert 2 cloning

Gibson QFL-F	GGGCTCGAAGTCCTATTCCAGGGACCGG TGATGAATCTTCTGCGTCGGAGTGGGAAA CGG	Polθ full length cloning
Gibson QFL-R	GGAGTGAATTATCGCGATACTAGTCTCGA GTTACACATCAAAGTCCTTTAGCTCTCCCC A	Polθ full length cloning
Gibson-polQFLmed-R	TGCTGCCAGGTTGGTGTCCTTTGCTCCTA GTTTGGCATTTTC	Polθ full length cloning
Gibson-PolQFLmid-F	GAAAATGCCAAACTAGGAGCAAAGGACAC CAACCTGGCAGCA	Polθ full length cloning

Chapter 3

DNA Polymerase θ-Catalyzed Error-Prone DNA Synthesis Induces Trinucleotide Repeat Expansion

3.1. Abstract

Polymerase θ (Polθ), a member of the family A translesion polymerases, has a unique structure composed of an active helicase domain and a large active site which is specific for Polθ. Polθ is involved in DNA replication timing and DNA repair [35, 54, 61]. Recent studies have showed that Polθ plays a role in base excision repair (BER) where it works in coordination with polymerase β (Polβ) [35, 40, 42-45, 80]. Several studies have shown an important role of Polθ in alt-EJ which could cause small insertion/deletion within repeat regions [54]. To investigate the role played by Polθ-mediated expansion in CAG/CTG repeats, we studied Polθ in hairpin retention and *in vitro* DNA synthesis of repeat sequences. Wild type Polθ promoted hairpin retention significantly more than the controls, while Polθ insert 2 mutant (contained a smaller active site) did not show this ability. *In vitro* DNA synthesis of CAG or CTG repeats also demonstrate that Polθ caused large expansions at trinucleotide repeat (TNR) sequences above 35 TNR without a pre-existing hairpin structure. In contrast, this was not observed with TNR sequences below 20 repeats or random sequences. These results indicate Polθ's involvement in generating large expansion of CAG repeat in TNR diseases, including Huntington's disease (HD) and myotonic dystrophy type 1 (DM1).

3.2. Introduction

Maintaining genomic stability is important to preserve normal cell functions and prevent genetic disorders. Genomic instability, including gene mutations, insertions/deletions, gene relocations and DNA re-replication, is one of the hallmarks in many types of cancer and genetic disorders [1-5]. Huntington's disease (HD), as the disease model in this study, is a neurodegenerative disease caused by an autosomal dominate gain-of-function of the mutant Huntington (mHtt) protein [1]. Clinically, mHtt protein is encoded by an expanded CAG trinucleotide repeat (TNR) of more than 40 repeats within exon 1 of the Huntington gene will cause the pathogenesis of HD [1]. Recent studies were supporting that above 35 CAG repeats should consider HD although patients carry an allele with 35-40 CAG repeats might not develop any symptoms [1, 7, 25, 81]. HD cell patient-derived lines and tissue studies in HD found mHtt protein and mRNA would aggregate, causing transcriptional dysregulation, defective energy metabolism, increased oxidative stress, excitotoxicity, and inflammation [1, 8-16]. However, the mechanism of TNR expansion through generations remains unclear. Recent studies suggest a link between HD and oxidative stress as a possible pathway leading to TNR expansion [1, 7, 17, 18]. Additionally, others found that base excision repair (BER) of oxidative damaged DNA bases may also participate in TNR expansion [5, 82]. During BER, APE1 created a nick on the damage strand allowing 3' slippage and hairpin formation [5]. Additionally, during DNA replication, genomic DNA is unprotected and relaxed at the replication fork, replicative DNA polymerase may cause a strand slippage in highly repetitive sequences and cause

the formation of secondary structures on these genomic DNA. When those secondary structures remain unrepaired, expansion will occur on the next replication cycle. Hence there are two possible mechanisms leading to TNR expansion: 1) a polymerase expands directly on to the nascent strands during replication; 2) a failure of hairpin repair results in repeat expansion.

Previous studies demonstrated that Polβ can promote the retention of hairpin structures depending on MutSβ activity and can extend DNA after the hairpin structures *in vitro* [20, 24]. Recent studies have highlighted the multi-functional translesion polymerase, Polθ, as a potential candidate. Polθ has been shown to participate in DNA repair and to regulate repair pathways (alt-EJ and BER), replication and lesion bypass, replication timing, genome stability maintenance, and gene editing [6, 45, 49, 52-54, 61-64]. Polθ has a special polymerase structure that it is the only known translesional polymerase contains a helicase domain and large active site. Polθ is a 290 kDa protein that contains a N-terminal active helicase domain and a central domain with two Rad 51-binding domains which is believed that these Rad 51-binding domains inhibit homologous recombination (HR) and allows Polθ performs alt-EJ for double strand break (DSB) repair [54, 83]. Its C-terminal polymerase domain contains three insertions (loop 1, insert 2 and loop 3). Insert 2 and loop 3 are within the palm of the polymerase, and insert 2 was later found to be essential for alt-EJ and contributes to the fidelity of Polθ [55]. Polθ was shown to have a larger active enzymatic site for lesion by pass and DNA synthesis during DNA replication [64, 72], while other studies have shown that Polθ can cause a small expansion *in vitro* [45, 55, 56]. We studied the contribution of

Polθ in TNR expansion in relation to the size of the active site in the Polθ polymerase domain.

3.3. Results

Polθ promotes (CAG)$_n$ or (CTG)$_n$ hairpin retention synthesis like Polβ.

To determine whether Polθ stimulates (CAG)•(CTG) repeat expansion by promoting hairpin retention in the nascent strand, *in vitro* DNA synthesis was performed in HeLa nuclear extracts using a (CAG)$_5$ or (CTG)$_5$ hairpin-containing DNA template with or without various purified DNA polymerases (Figure 3.1A and 3.2A). The synthesized products were resolved by urea-PAGE and analyzed by Southern blot assay. The hairpin-retained or -removal products were visualized with a ^{32}P-labeled probe specifically annealed to the newly synthesized strand (Figure 3.1A) [24]. Similar to previous studies [24], HeLa nuclear extracts along with the addition of Polδ generated a major hairpin-removal band (Figure 3.2B). Remarkably, Polθ promoted more hairpin retention compared to Polβ. Polθ enhanced hairpin retention to 61% with the (CAG)$_5$ HP template and to 57% with the (CTG)$_5$ HP template during *in vitro* DNA synthesis with a pre-existing hairpin in the nascent strand. Additionally, Polβ which only caused 38% and 26% hairpin retention with the (CAG)$_5$ HP and (CTG)$_5$ HP templates, respectively (Figure 3.2B). The enhancement of hairpin retention synthesis was observed regardless of either (CAG)$_5$ or (CTG)$_5$ hairpin in the primer strand. Interestingly, Polθ caused extension beyond hairpin retention with (CTG)$_5$ hairpin-containing *in vitro* DNA synthesis (Figure 3.2B).

43
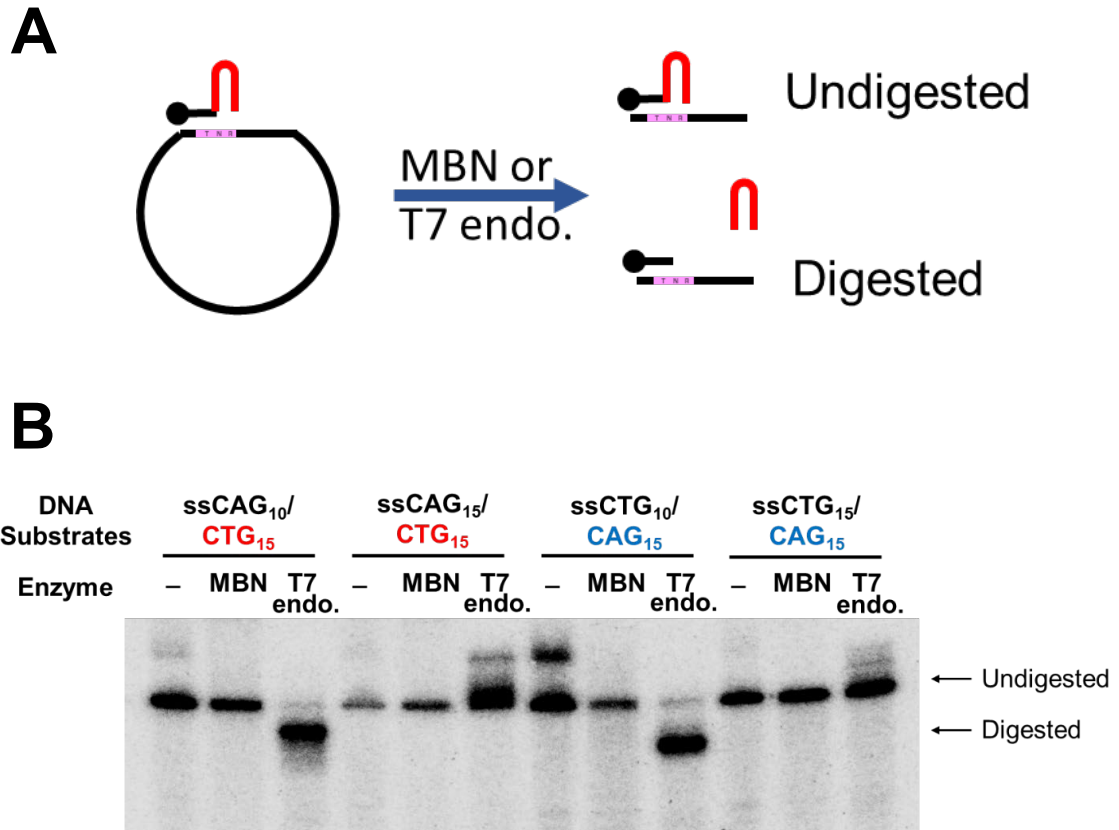
Figure 3.1: Hairpin structures were formed in the primer strand. **(A)** Experimental diagram of 5'-^{32}P labeled hairpin digestion to ensure the formation of hairpin structures. **(B)** Hairpin substrates digestion with Mung Bean Nuclease (MBN) or T7 endonuclease I (T7 endo.).
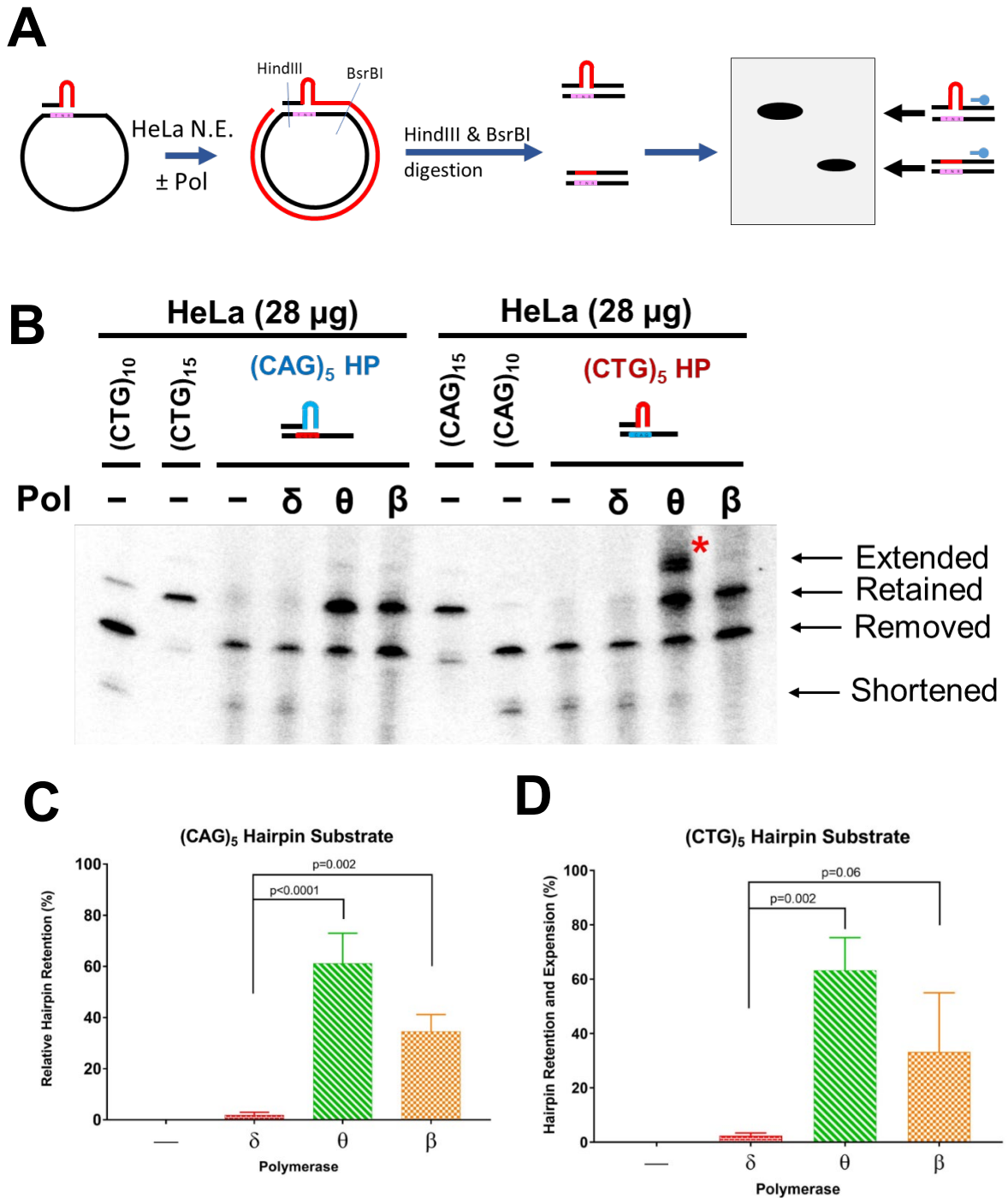
Figure 3.2: Polθ enhances either (CAG)$_5$ or (CTG)$_5$ hairpin retention synthesis in the competing HeLa nuclear extracts. **(A)** Experimental diagram of hairpin removal/retention assay by Southern blot analysis. **(B)** In vitro DNA synthesis with (CAG)$_5$ or (CTG)$_5$ hairpin in primer strand and analyzed by Southern blot assay; red asterisk indicates extended products. **(C)** Quantification for (CAG)$_5$ hairpin retention for four independence assays. **(D)** Quantification for (CTG)$_5$ hairpin retention for four independence assays.

The deletion of insert 2 in the Polθ domain does not affect hairpin bypass synthesis.

The deletion of insert 2 within the Polθ polymerase domain was previously shown to increase the fidelity of Polθ and shrink the polymerase active site [55]. To investigate the importance of insert 2 in the Polθ polymerase domain in hairpin bypass synthesis, a construct of the insert 2 deletion mutant (PolθΔi2) was cloned into the pLEXm vector with two maltose binding (MBP) protein tags for expression in 293GnTi⁻/ 293T cells and protein purification (Figure 3.3A). *In vitro* hairpin bypass synthesis was repeated with PolθΔi2 (Figure 3.3B). PolθΔi2 insignificantly induced hairpin retention synthesis by 7% with both hairpin templates compared to the addition of Polδ. Polθ wild type (WT)-mediated hairpin retention synthesis was significantly higher than PolθΔi2 synthesis with both hairpin substrates (Figure 3.3C and 3.3D). Moreover, Polθ WT polymerase induced expansion besides hairpin retention, while PolθΔi2 could not induce extra expansion. Therefore, the downsizing of the enzymatic site of the polymerase domain in PolθΔi2 is able to increase the fidelity of Polθ (Figure 3.3B).

Polθ promotes expansion without a pre-existing hairpin in the nascent strand

Polθ WT caused expansion besides promoting hairpins in retention synthesis (Figure 3.2 and Figure 3.3). To investigate the ability of Polθ-mediated expansion, *in vitro* DNA synthesis without preexisting hairpin in extension primer with purified PCNA, RFC, various polymerases with/without RPA (Figure 3.4A). Polθ WT caused an addition of one nucleotide beyond the full length of the template which was independent of the sequence in the template (Figure 3.4). Surprisingly, when the templates contained either $(CAG)_{35}$ or $(CTG)_{35}$ repeats, the threshold TNR for characterizing HD patients, only wild type Polθ caused a large expansion but not its random sequence controls (Figure 3.4C).
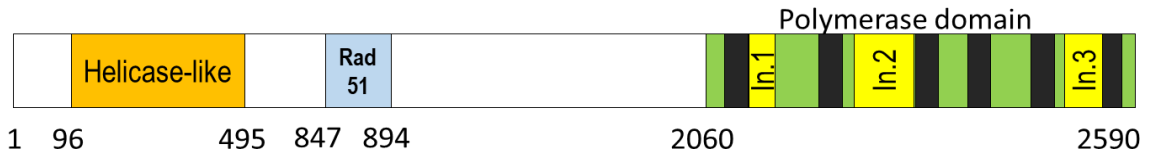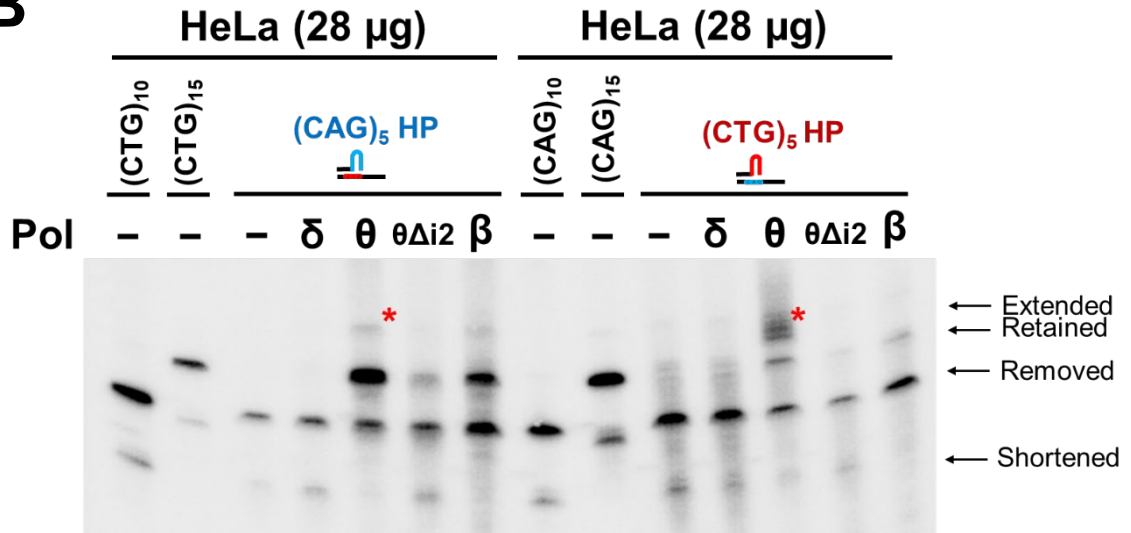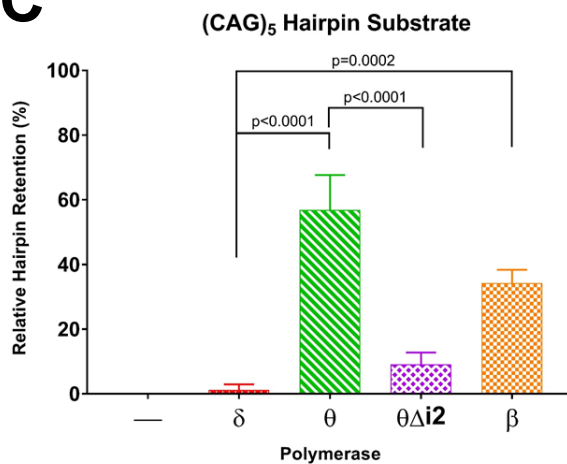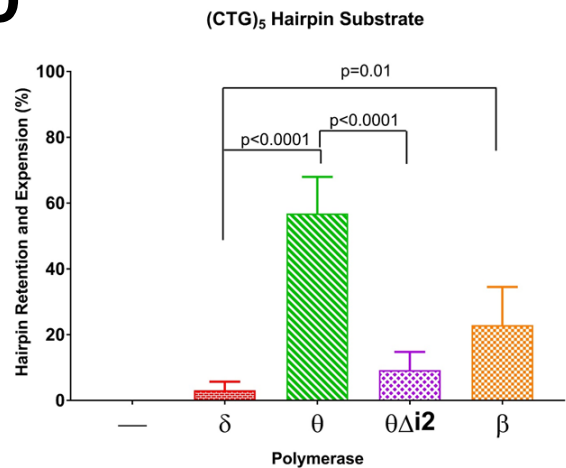
Figure 3.3: Polθ Δ insert 2 (θΔi2) mildly promotes either (CAG)$_5$ or (CTG)$_5$ hairpin retention synthesis in the competing HeLa nuclear extracts. **(A)** Model of Polθ with domains. **(B)** *In vitro* DNA synthesis with either (CAG)$_5$ or (CTG)$_5$ hairpin in the primer strand and analyzed by Southern blot assay; ; red asterisk indicates extended products. **(C)** Quantification for (CAG)$_5$ hairpin retention for three independent assays. **(D)** Quantification for (CTG)$_5$ hairpin retention for three independent assays.

Polθ-mediated large repeat expansion is sequence specific

In order to investigate if an allele with 28 or more repeats is about 10% more unstable than normal allele through generations [27, 28, 84], we used various repeat length and sequence to mimic lagging and leading strand for *in vitro* DNA synthesis without a preexisting hairpin structure to demonstrate repeat expansion during DNA replication. Our results suggested that Polθ-promoted repeat expansion depend on repeat length and sequence. With normal repeat length mimics, wild type Polθ only caused a one nucleotide addition with no significant difference between CAG repeats and CTG repeats (Figure 3.4A and 3.4B). While with a longer repeat, wild-type Polθ caused a large expansion in a sequence-dependent manner (Figure 3.4C). Wild type Polθ yielded larger expanded products with 35 CTG templates than with 35 CAG templates.

The deletion of insert 2 in Polθ's polymerase domain increases the fidelity of Polθ

PolθΔi2 had a similar processivity to that of wild-type Polθ on various substrates, except for the $(CTG)_{35}$ substrate (In Figure 3.4). Insert 2 in the polymerase domain either did not or mildly contributed to the processivity of the polymerase. However, insert 2 contributed to the fidelity of the polymerase as the insert 2 mutants were unable to cause large expansion due to the reduced active site within the polymerase domain. The reduction size of the polymerase domain in Polθ provided us an insight of a possible small molecule target site to promote error-free DNA synthesis with this multi-functional translesion polymerase.
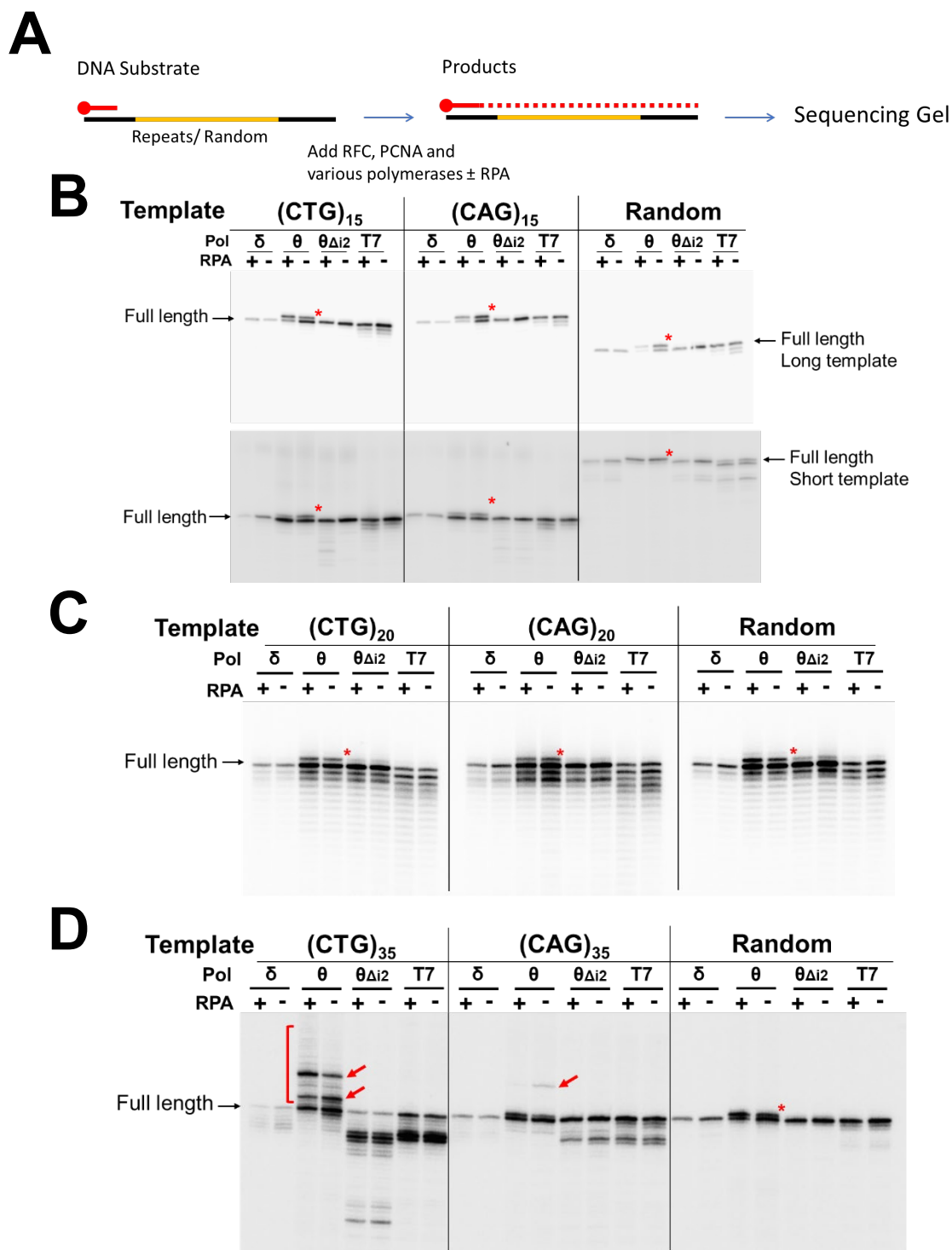
Figure 3.4: Polθ-mediated repeat expansion depends on sequence and TNR repeat length. **(A)** Experimental model for *in vitro* DNA synthesis. **(B)** Reconstituted DNA synthesis with ^{32}P-labeled primers and pre-digested purified M13mp18 templates contain either 10 CAG or 10 CTG repeats or random control. **(C)** Reconstituted DNA synthesis with ^{32}P-labeled primer and synthesized templates contain either 20 CAG or 20 CTG repeats or random control. **(D)** Reconstituted DNA synthesis with ^{32}P-labeled primers and pre-digested purified M13mp18 templates contain either 35 CAG or 35 CTG repeats or random control. Red bracket and arrows indicate extended large product; red asterisk indicates extended TdT products.

Polymerase θ mediated expansion within repeat sequence

The Polθ mediated expansion bands with the $(CAG)_{35}$ and $(CTG)_{35}$ templates were extracted and recovered from the sequencing gel. The recovered expansion bands were cloned into T-Vector pMD™20 (Takara). The cloned expanded bands v purified and sequenced. Sequencing results showed that the expended bands were located within repeat regions of the $(CTG)_{35}$ template (Figure 3.5). The major large expansion band contained $(CAG)_{40}$ repeats in the nascent strand that were 5 TNR repeats more than the original $(CTG)_{35}$ on the template strand (labeled in Figure 3.5). However, the expanded products from the $(CAG)_{35}$ template could not be recovered from the gel because the major expanded band was much weaker than with the $(CTG)_{35}$ template (red arrow in Figure 3.4D).

A

C6277

```
CTG35_PolQ_WT.txt          ACGTTGTAAAACGACGGCCAGTGCCAAGCTTCAGCAGCAG          40
CTG35_PolQ_Expansion.txt   ACGTTGTAAAACGACGGCCAGTGCCAAGCTTCAGCAGCAG          40
Consensus                  acgttgtaaaacgacggccagtgccaagcttcagcagcag
```

```
CTG35_PolQ_WT.txt          CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC          80
CTG35_PolQ_Expansion.txt   CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC          80
Consensus                  cagcagcagcagcagcagcagcagcagcagcagcagcagc
```

```
CTG35_PolQ_WT.txt          AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA          120
CTG35_PolQ_Expansion.txt   AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA          120
Consensus                  agcagcagcagcagcagcagcagcagcagcagcagcagca
```

```
CTG35_PolQ_WT.txt          GCAGCAGCAGCAGCAG...............GAATTCGTA          145
CTG35_PolQ_Expansion.txt   GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGGAATTCGTA          160
Consensus                  gcagcagcagcagcag              gaattcgta
```

Expansion   V6221

```
CTG35_PolQ_WT.txt          ATCATG          151
CTG35_PolQ_Expansion.txt   ATCATG          166
Consensus                  atcatg
```

B

C6277

```
CTG_PolQ_WT.txt            ACGTTGTAAAACGACGGCCAGTGCCAAGCTTCAGCAGCAG          40
CTG_PolQ_Deletion.txt      ACGTTGTAAAACGACGGCCAGTGCCAAGCTTCAGCAGCAG          40
Consensus                  acgttgtaaaacgacggccagtgccaagcttcagcagcag
```

```
CTG_PolQ_WT.txt            CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC          80
CTG_PolQ_Deletion.txt      CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC          80
Consensus                  cagcagcagcagcagcagcagcagcagcagcagcagcagc
```

```
CTG_PolQ_WT.txt            AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA          120
CTG_PolQ_Deletion.txt      AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA          120
Consensus                  agcagcagcagcagcagcagcagcagcagcagcagcagca
```

```
CTG_PolQ_WT.txt            GCAGCAGCAGCAGCAGGAATTCGTAATCATG          151
CTG_PolQ_Deletion.txt      GCAGCAGCAGCAG...GAATTCGTAATCATG          148
Consensus                  gcagcagcagcag   gaattcgtaatcatg
```

V6221

Deletion

Figure 3.5: Polθ promotes expansion or deletion within a repeat region. **(A)** Sequencing results from the cloned T-vectors with expanded bands in Polθ DNA synthesis with the (CTG)$_{35}$ template. **(B)** Sequencing results showed deletion by Polθ DNA synthesis with the (CTG)$_{35}$ template.

Chromatin-associated polymerase θ increased in cells that carried larger CTG/CAG repeats

The correlation between Polθ expression and HD remains unknown. To understand the relationship between Polθ protein levels and the progression of HD, protein levels of multiple patient-derived cell lines were screened by Western blotting. In lymphocytes, the total Polθ levels in HD patient-derived lymphocytes was slightly higher than HL60 but yet it was statistically significantly increase protein expression in HD lymphocytes (Figure 3.6B). However, although HL60 is from a non-HD patient, it is an acute promyelocytic leukemia cell line which replicates much faster and harbors a much higher replication stress than other HD lymphocytes (Figure 3.6A). In order to have a better control on cell conditions, better wild type cell lines and to increase the screening of HD cell lines, we used a group of nine HD patient derived fibroblast cell lines: a family of six HD patient fibroblasts including a non-HD family member, 2 unrelated HD patient and another clinically non-HD patient cell line. Since Polθ is a translesion polymerase, we focused on the comparison of Polθ protein levels in cytosol and on chromatin. To determine whether the increased Polθ is recruited directly to the chromatins, we isolated the chromatin fractions and compared to the cytoplasmic fractions with the lymphocytes. The protein level of Polθ in the cytoplasmic fractions are much lower than in the chromatin fractions (Figure 3.6C top panel). Consistent with the HD patient lymphocytes, the fibroblasts that contained more than 30 CAG repeats had higher Polθ recruitment to the chromatin than the normal control — GM04204 (Figure 3.6C). Interestingly, the clinical non-HD patient (GM02153) contains an

allele with 32 CAG repeats had a significant higher Polθ protein recruitment to the

chromatin compared to another control (GM04204) [85]. Our data showed that a

possible threshold of Polθ recruitment to chromatin happened before developing

HD symptoms. Polθ protein was stabilized and recruited to the chromatin as there

was a trace amount of Polθ in the cytoplasmic fractions which was unable to be

detected by western blot (Figure 3.6C top panel).
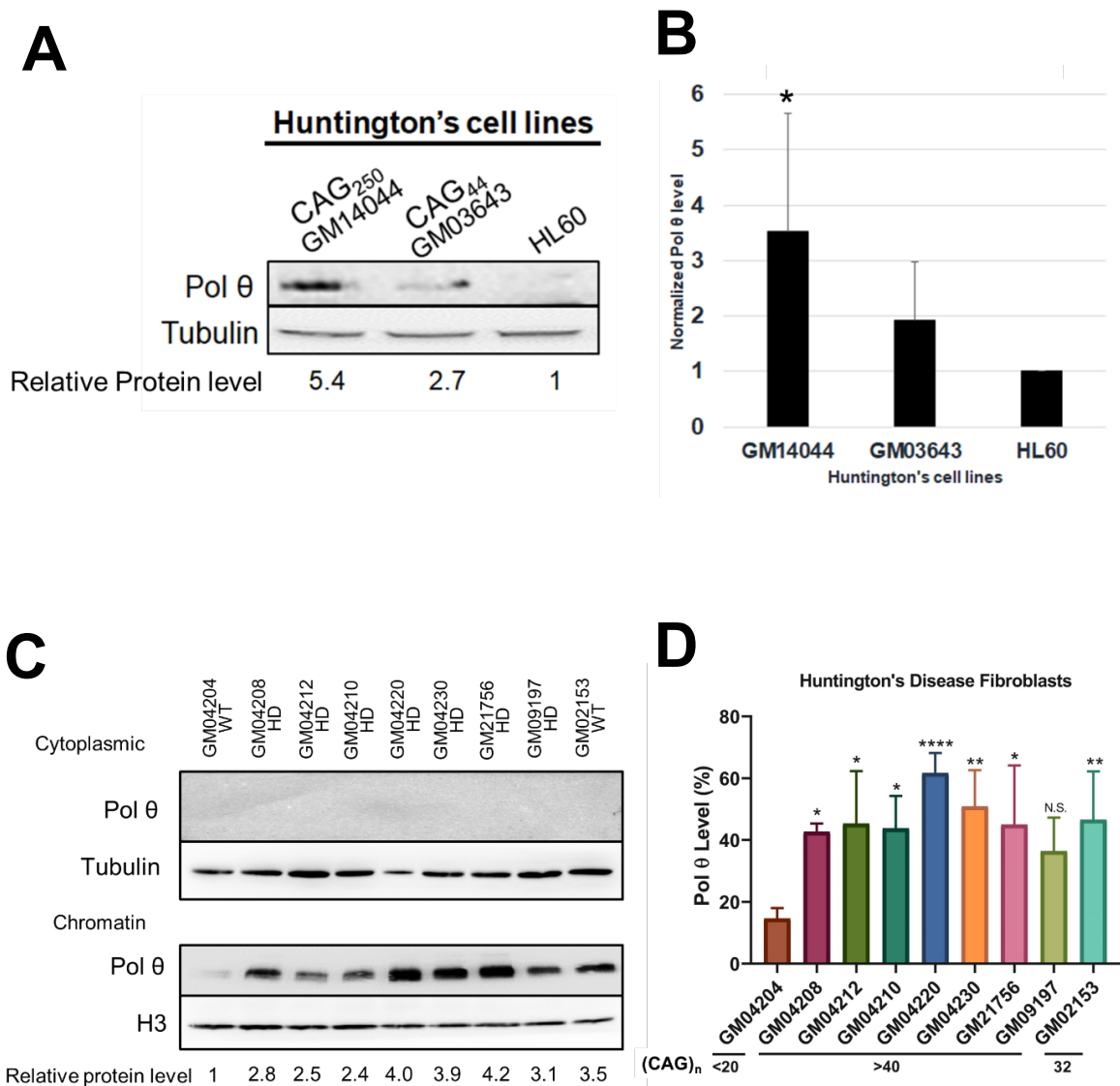
Figure 3.6: Polθ expression levels in patient cell lines. Polθ protein levels are upregulated in **(A)** western blot for HD patient lymphoblast whole cell lysates; **(B)** its quantification of three independent results is shown; **(C)** western blot for HD patient fibroblast; **(D)** quantification of Polθ recruitment on chromatin in HD fibroblasts for five independent experiments. *: $p < 0.05$, **: $p < 0.01$, ****: $p < 0.0001$, N.S.: not significant.

Estimating CAG repeat length in Huntington's disease cell lines

Most of the HD cell information on the CAG repeats' length was obtained from the Coriell Institute for Medical Research. Nevertheless, no details were provided about the CAG repeats' length of some cell lines (Figure 3.7). Therefore, we measured the distance from the bottom of the wells (top of the gel) to the middle of the detected bands by ImageJ and generated a logistic regression equation with a R^2 = 0.9728. The estimated CAG repeats in exon 1 of *HTT* gene for all the HD cell line in study is showed as follows: GM04212 has 15 and 44 CAG repeats; GM04210 has 16 and 51 CAG repeats; GM04220 has 18 and 51 CAG repeats; GM09197 has 18 and 180 CAG repeats; GM14044 has 18 and 250 CAG repeats; HL60 has 14 and 28 CAG repeats.
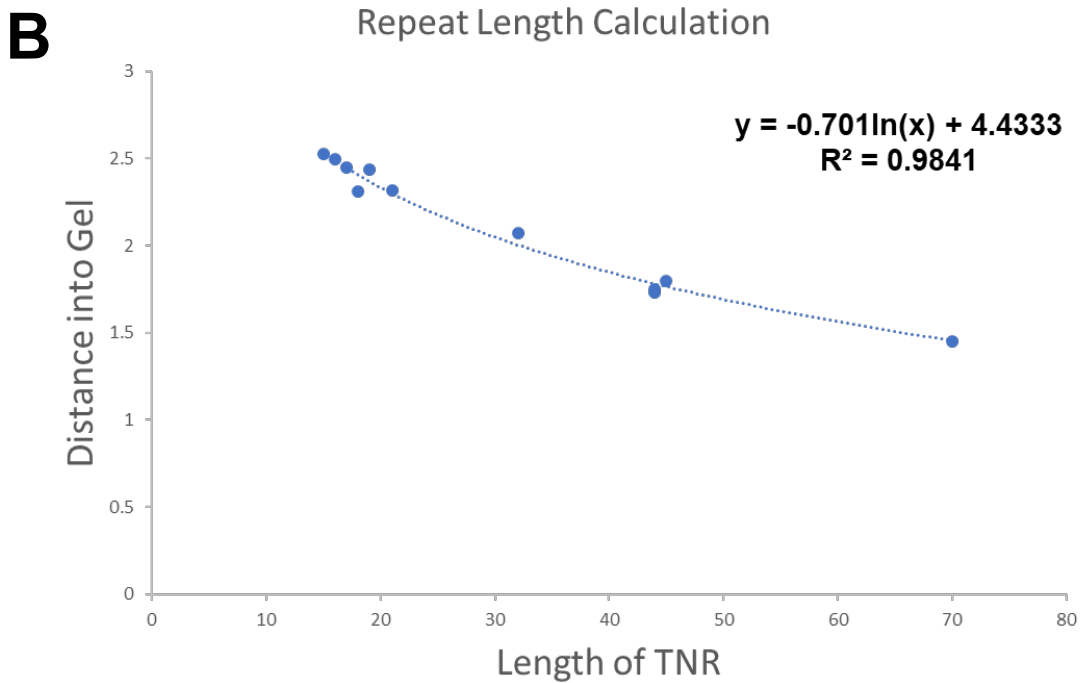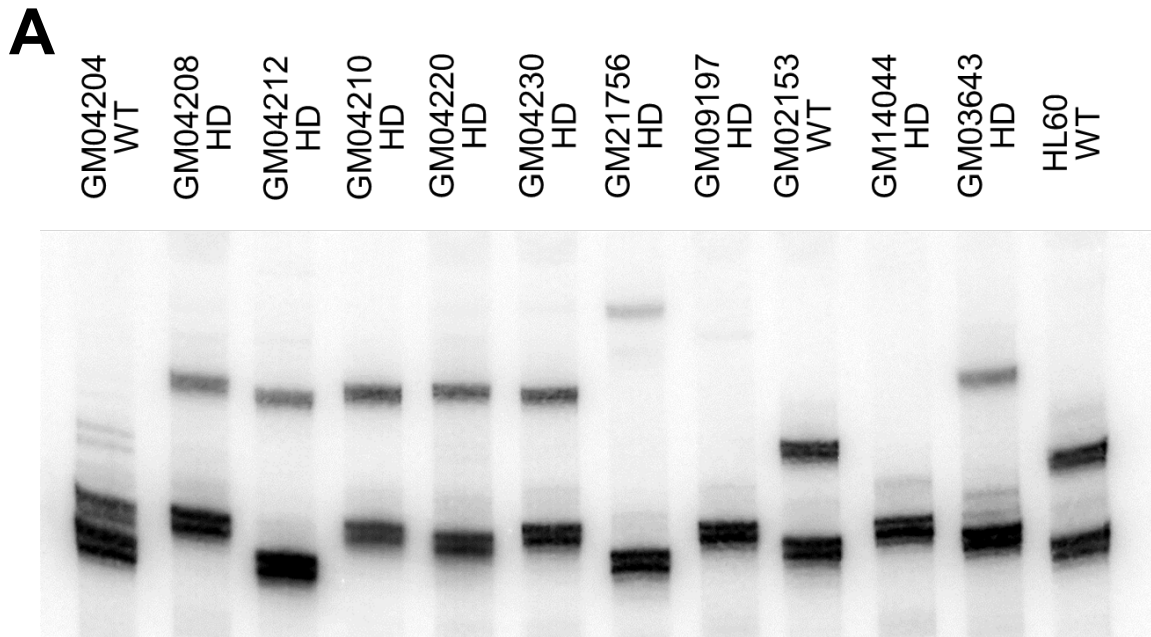
Figure 3.7: Estimating the repeat length for cells with an unknown CAG repeat length. **(A)** Southern blot analysis of the CAG repeat length in exon 1 of the *HTT* gene in various HD cell lines; **(B)** Quantification of the PCR products' distance into the gel and the logistic regression equation with a R^2 = 0.9841.

Table 3.1: Logistic regression equation for estimating CAG repeat length in *HTT* exon 1

$$Y = -0.701 ln(x) + 4.4333$$

Table 3.2: Cell line information and HD status. Most cell line information on the HD status and repeat length was obtained from the Coriell Institute for Medical Research (Coriell), unless otherwise mentioned. Cell lines with no information (highlighted in orange) about length of CAG repeats was estimated by logistic regression equation in **Table 3.1**.

Cell lines	Gender	Allele	Repeat length	Distance into Gel	HD Status	Reference
GM04204	M	1	17	2.449	No	Coriell
		2	18	2.311		Coriell
GM04208	M	1	21	2.319	Yes	Coriell
		2	44	1.733		Coriell
GM04212	F	1	16	2.504	Yes	
		2	44	1.788		
GM04210	M	1	18	2.392	Yes	
		2	43	1.794		
GM04220	F	1	17	2.435	Yes	
		2	43	1.790		
GM04230	F	1	18	2.400	Yes	
		2	45	1.795		Coriell
GM21756	F	1	15	2.524	Yes	Coriell
		2	70	1.453		Coriell
GM09197	M	1	18	2.410	Yes	
		2	180	N/A		Coriell
GM02153	F	1	16	2.495	No	[85]
		2	32	2.070		[85]
GM14044	M	1	18	2.415	Yes	Coriell
		2	250	N/A		Coriell
GM03643	F	1	19	2.437	Yes	Coriell
		2	44	1.752		Coriell
HL60	F	1	28	2.102	No	
		2	14	2.477		

3.4. Discussion

Polymerase θ is a unique translesion polymerase with an essential role in alt-EJ during DSB repair. Previously showed Polθ can cause small expansions and deletions *in vivo and* allows strand slippage when synthesized through mono- or di-nucleotide repeats *in vitro* [55, 86]. However, the mechanism through which Polθ contributes to TNR expansion-associated diseases, such as HD or DM1, is still unknown.

Polθ promotes hairpin retention DNA synthesis more than Polβ

Polθ promotes hairpin retention synthesis with CAG hairpin or CTG hairpin significantly better than HeLa nuclear extracts alone or with the addition of Polδ (Figure 3.2). Although both Polβ and Polθ can promote hairpin retention during DNA synthesis with a pre-existing DNA structure, the unique large enzymatic site in the polymerase domain of Polθ provides an advantage for accommodating and by-passing large DNA secondary structures and performing DNA synthesis with either CAG or CTG hairpin substrates. However, it is still unknown how Polθ cooperates with Polβ during BER in the brain when neurons are experiencing an increase in ROS and DNA damage in HD patient' brains [17, 18, 25, 31], and ultimately leading to the development and progression of HD. Study showed that Polθ would be recruited in various DNA repair pathways [35], it remains unclear that why Polθ is recruited to the chromatin more than non-HD patient cell lines. Since HD patient cell lines have an increased amount of ROS, we suggest Polθ plays a role in BER and against oxidative stress. In the future, HD fibroblasts can be treated with H_2O_2 to determine whether the recruitment and expression of Polθ

increases in the chromatin, and to determine if the overexpression of Polθ would increase the repeat expansion under a high ROS environment.

Polymerase θ enhances trinucleotide repeat expansion and hairpin retention with CTG hairpins

Strikingly, Polθ promotes expansion within the CTG repeats region (noted with a red asterisk in Figure 3.2B), suggesting that Polθ not only induces hairpin retention synthesis, but also promotes the growth of CTG hairpins (Figure 3.2B). Polθ may mediate additional repeat expansion (noted with red asterisks in Figure 3.2B and Figure 3.3B) during CAG hairpin bypass synthesis. However, because the formation of CAG hairpins is less stable than that of CTG hairpins [21], the phenomenon of repeat expansion with CAG hairpin substrate may be varied and could be too weak for radiative detection because of the low sensitivity of Southern blotting. The large repeat expansion caused by Polθ's error-prone synthesis without a preexisting hairpin structure suggests Polθ-mediate repeat DNA synthesis may involve in both germline expansion in HD families and somatic expansion in non-HD families. One study showed that a mutation in polq (mouse analog of human Polθ) caused spontaneous chromosome aberrations occurring leading to growth retardation [87] and indicated that Polθ is important for cells replication. Previous clinical cohort study observed a more frequent incident of a paternal inheritance of HD than maternal inheritance [25, 29]. According to the GTExPortal from The Broad Institute of MIT and Harvard, the gene expression for Polθ (ENSG00000051341.9) in testis is much higher than that in ovary, [25, 88, 89]. Although paternal inheritance occurs more often which could be explained by

a higher expression of Polθ, spermatogenesis requires many more cell replication cycles than oogenesis [90] leading to an increase the incidents of repeat expansion. Future studies on the overexpression of Polθ *in vivo* at various concentrations of Polθ *in vitro* DNA synthesis are essential to understand whether paternal inheritance is related to the higher Polθ expression.

The insert 2 domain in the polymerase domain of Polθ allows Polθ-meditated hairpin retention and repeat expansion

Polθ highly stimulates hairpin retention during DNA synthesis *in vitro*. In the polymerase domain of Polθ, three loop structures provide the large active site for DNA synthesis. Kent, et. al, (2015) suggested that insert 2 in the polymerase domain promoted DNA binding; insert 1 contributed to DNA synthesis processivity and DNA binding [6, 67]; insert 3 participated in translesion synthesis activity [6, 55, 67]. PolθΔi2 had a significantly lower ability to synthesize past hairpin structures than wild type Polθ (Figure 3.3B); no significant differences were observed in terms of promoting hairpin retention between PolθΔi2 and either the negative control or wild type Polδ (Figure 3.3C and 3.3D). Deleting insert 2 increased the fidelity of Polθ but decreased the ability of DNA lesion to bypass synthesis. The lower hairpin retention ability of PolθΔi2 could be caused by its lower activity. To understand whether the activity of PolθΔi2 has a similar DNA synthesis activity to Polθ wild type and to further study Polθ's unique properties on synthesizing through TNR sequences, we performed reconstituted *in vitro* DNA synthesis assays. The smaller catalytic domain in PolθΔi2 will provide insight into potential treatment for inhibiting repeat expansions and increasing the fidelity of

Polθ without completely disabling other essential roles of Polθ in DNA repair and replication.

Polymerase θ causes small expansions with short trinucleotides repeat template

Polymerase θ caused a small expansion in the short TNR template of 15 or 20 TNR as a mimic repeat length for a healthy person in the Huntington gene (Figure 3.4A and 3.4B). Consistent with previous studies, Polθ caused a small expansion, suggesting strong terminal deoxynucleotidyl transferase (TdT) activity showed as an addition of one extra nucleotide at the end of the newly synthesized strand. Polθ's TdT ability was observed in all short TNR substrates and with random sequence controls. Although researchers suggested that RPA could inhibit the DNA synthesis of Polθ, the presence of unmodified RPA did not significantly inhibit the fidelity, processivity, and TdT activity of Polθ (Figure 3.4). The small expansion was not restricted to the 125 nt [(CAG)$_{15}$ or (CTG)$_{15}$] and 110 nt [(CAG)$_{20}$ or (CTG)$_{20}$] that were used in Figure 3.4A and 3.4B, but also in the longer random sequence control (148 nt) for (CAG)$_{15}$ or (CTG)$_{15}$ templates (Bottom panel of Figure 3.4B). The TdT activity was suppressed by deleting insert 2 in Polθ (Figure 3.4B and 3.4C) but the DNA synthesis was mildly or unaffected with a smaller polymerase active site similar to what was reported previously with mononucleotide repeat templates [55].

Polymerase θ-mediated large repeat expansion depends on the stability of hairpin structure in the nascent strand

Wild type Polθ induced various large expansions upon synthesizing CAG repeats in the nascent strand only when repeat length is longer than 35 CAG or CTG repeats in the templates (Figure 3.4C). In contrast, CTG repeats in the nascent strand (35 CAG repeat in the template strand) only showed a single large expansion with the more sensitive *in vitro* DNA synthesis assay (Figure 3.4C). Polθ-mediated a variation of large expansions (shown as the red bracket and red arrows in Figure 3.4C*)*, however, those large expansions only happened in a low frequency. Since the CTG repeats could form a more stable secondary structure than the CAG repeats [21], Polθ-mediated repeat expansion with $(CAG)_{35}$ template has less varity of expansion than replicating through a $(CTG)_{35}$ template. Wild-type Polθ protein caused expansion beyond hairpin retention synthesis, indicated by red asterisk in $(CTG)_5$ hairpin substrates in Figure 3.2B and 3.3B. Similar to the large expansions *in vitro* DNA synthesis with 35 CAG or CTG repeats templates and considering wild-type Polθ's large active site, nascent strand slippage was induced with the $(CTG)_5$ hairpin substrate in hairpin by-pass synthesis in Figure 3.2 and 3.3. With a larger repeat number and longer template- $(CTG)_{35}$ repeats templates, the nascent strand is more flexible which it is easier to form secondary structures (Figure 3.4C). Polθ could cause more variations and larger repeat expansions with the $(CTG)_{35}$ template than with either shorter repeat templates or the $(CAG)_{35}$ template in Figure 3.4C because of the stability difference in secondary structures [21]. The CTG repeats on the template strand can also form

hairpin structures and Polθ may bypass those hairpin structures and continue DNA synthesis, leading to repeat deletions in the nascent strand which can be showed by sequencing. In the sequencing results with the Polθ replicating $(CTG)_{35}$ template, it revealed a 1 TNR repeat deletion in the product, suggesting hairpin formation in the $(CTG)_{35}$ template strand (Figure 3.5B).

Studies have shown that individuals with TNR repeats longer than 28 have an increased lifetime risk of developing HD and an increased risk that their children will be affected [27, 91] and others had showed that more than 27 CAG repeats is unstable [27]. Since Polθ is recruited to the origin of replication [55] and our data demonstrated that Polθ can increase trinucleotide repeat length, we suggest Polθ may contribute to the increased risk of developing HD in people who had $(CAG)_{35}$ repeats in the *HTT* gene during DNA repair and cell replication.

Polymerase θ-mediated large repeat expansion specifically with long repeat templates

The wild-type Polθ protein promoted strand slippage during DNA synthesis specifically in longer repeat template but not with a random sequence (Figure 3.4C). The increase in nascent strand expansion within the repeat sequence depended on the large active site in the polymerase domain of Polθ. Without insert 2 in the Polθ polymerase domain, Polθ was unable to cause expansion in any of the DNA templates. The specificity of Polθ-mediated repeat expansion depends on the length of the TNR repeats and correlates with the HD studies that showed that patients with 27-35 CAG repeat alleles were more unstable [28, 84, 91]. Previous studies showed that Polθ mediates repeat expansion with

mononucleotide repeats templates [48, 55]. The reduced size in the catalytic domain displaced as a promising target for preventing various type repeat expansion drive genomic instability by discovering small molecule that can in cooperate within the catalytic domain of wild-type Polθ.

Figure 3.9: Model of Polθ mediated error-prone synthesis in long TNR template.

Polymerase θ protein level is upregulated in cells harboring more than 30 CAG repeats in the HTT gene

Strikingly, GM02153 is considered a cell line generated from non-HD patient as she did not develop HD, however its Polθ protein level was significantly higher than in another control cell line — GM04204, but similar to that of other HD patient cell lines (Figure 3.6A and 3.6B). Later, it was shown that GM02153 has an intermediate allele with 32 repeats [85]. Others suggested that 27-35 CAG repeats are considered unstable, and showed that the TNR unstable incidents increased from 5.5-7.3% with 27-30 CAG to 13% with 32 CAG [84]. Furthermore, Polθ could be the driving force in transmitting HD to the next generation in patients with 27-

35 repeats. However, a larger screening on Polθ protein levels in patients who carry 27-35 repeats will help to show whether Polθ induces repeat instability.

Polymerase θ increased recruitment to the chromatin.

We showed an increase in Polθ recruitment in HD patient cell lines with more than 30 TNRs to chromatin (Figure 3.6C and 3.6D). The recruitment of Polθ to TNR regions is unclear. Although Polθ has higher processivity in repetitive regions than Polδ (Figure 3.4), it promotes error-prone DNA synthesis and induces TNR expansion in the offspring of high-risk HD individuals. Since we showed how Polθ mediates error prone synthesis, the increased recruitment of Polθ could rise the probability of Polθ mediates TNR instability through its role in BER and DSB repair in neurons in pre-HD patients.

The unique large active site enables Polθ to promote strand slippage in long TNR repeat templates and induce large TNR expansion during DNA synthesis. The deletion of insert 2 in the polymerase domain indicates Polθ as a possible therapeutic target for decreasing the risk of developing HD and transmitting it to next generation in intermediate people who carry 27-35 CAG repeats.

Chapter 4

**The Metal Ion Choice for DNA Synthesis Influences the Fidelity of Polymerase θ**

4.1. Introduction

Clinical studies showed that metal ion balance in the brains of HD patients was different from those of healthy individuals [92]. In structural studies, various metal ions interact slightly differently within the polymerase active site [48, 93, 94]. In translesion polymerase ι and polymerase η, the distance between Mn^{2+} ions and their interacting amino acid within the polymerase active site is farther than when Mg^{2+} ions are incorporated in those active sites [94], suggesting that the incorporation of Mn^{2+} rather than Mg^{2+} enlarged the active sites of TLS polymerases. Previous studies showed that Polθ could cause mononucleotide repeat expansion when the last nucleotides were repeats [48, 55]. Polθ, with Mn^{2+}, could cause expansion with a mononucleotide repeat template [48]. However, it is still unknown whether Polθ could cause expansion when synthesizing through longer TNR templates with Mn^{2+}. **In this chapter, we will discuss whether 1) Polθ causes larger expansions with Mn^{2+} because of increased active sites; 2) PolθΔi2's active site will be enlarged by Mn^{2+} and lead to expansions.**

4.2. Results and Discussion

Polθ-mediated expansion is induced by the presence of Mn^{2+}

Polθ caused a larger expansion with Mn^{2+} but not with Mg^{2+} in the 110 nt synthesized templates (Figure 4.1). Interestingly, in the presence of Mn^{2+}, the high fidelity polymerases, Polδ- and PolθΔi2 also showed small expansions with the 110 nt templates (the set of templates contain 20 TNR and their random sequence control), although the expansions were smaller than those for wild-type Polθ (Figure 4.1). With the 110 nt oligo templates, Polθ caused a larger expansion with Mn^{2+} than with Mg^{2+} (Figure 4.1). This observation indicates that the large catalytic site of the Polθ is necessary to promote error-prone DNA synthesis. However, with longer templates (190 nt) contained in the 35 TNR repeats and their controls, Polδ was unable to synthesize to full length in 30 min (Figure 4.2). Like Polδ, Polθ and PolθΔi2 had trouble synthesizing through longer templates (Figure 4.2). In 30 min of reaction time with longer templates, PolθΔi2 caused a small expansion when synthesizing through $(CAG)_{35}$, $(CTG)_{35}$, and random sequence (Figure 4.2). As discussed in **Chapter 3**, Polθ induced much larger expansion with $(CTG)_{35}$ template with Mg^{2+} and caused more smearing in the presence of Mn^{2+} (Figure 4.2 top panel, labelled with red bracket). Polθ caused a larger expansion with Mn^{2+} than with Mg^{2+} with the $(CAG)_{35}$ template (Figure 4.2 middle; panel labelled with red arrows). These larger expansion in the 110 nt templates and the 190 nt templates with wild-type Polθ protein could cause by the enhanced terminal deoxynucleotidyl transferase (TdT) activity in the presence of Mn^{2+} and led to more variation of large expansions in the longer TNR templates.
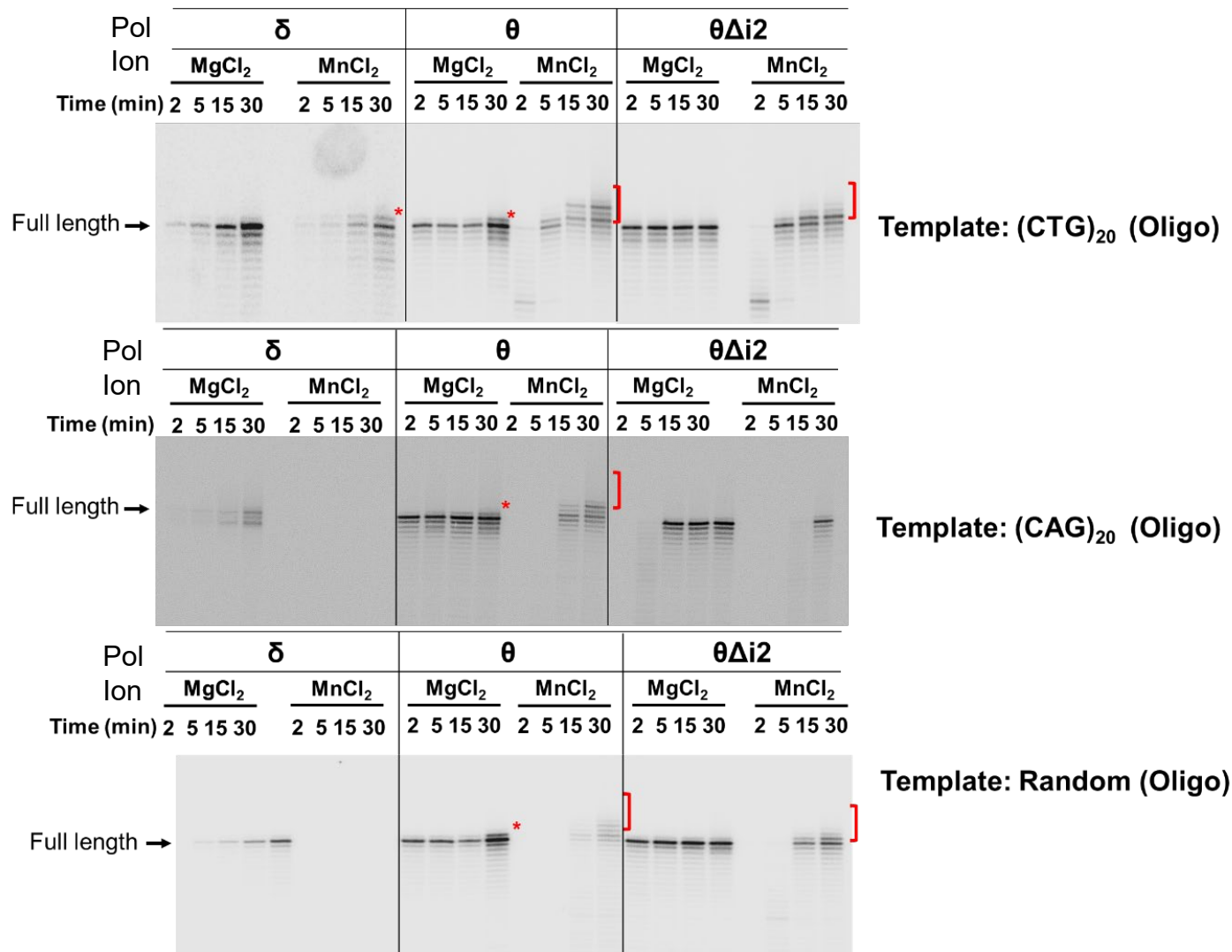
Figure 4.1: Manganese ions affects the fidelity and processivity of polymerases. Reconstituted DNA synthesis with ^{32}P-labeled primers and synthesized 110 nt oligo template contains either 20 CTG repeats, 20 CAG repeats, or random sequence at various time points.
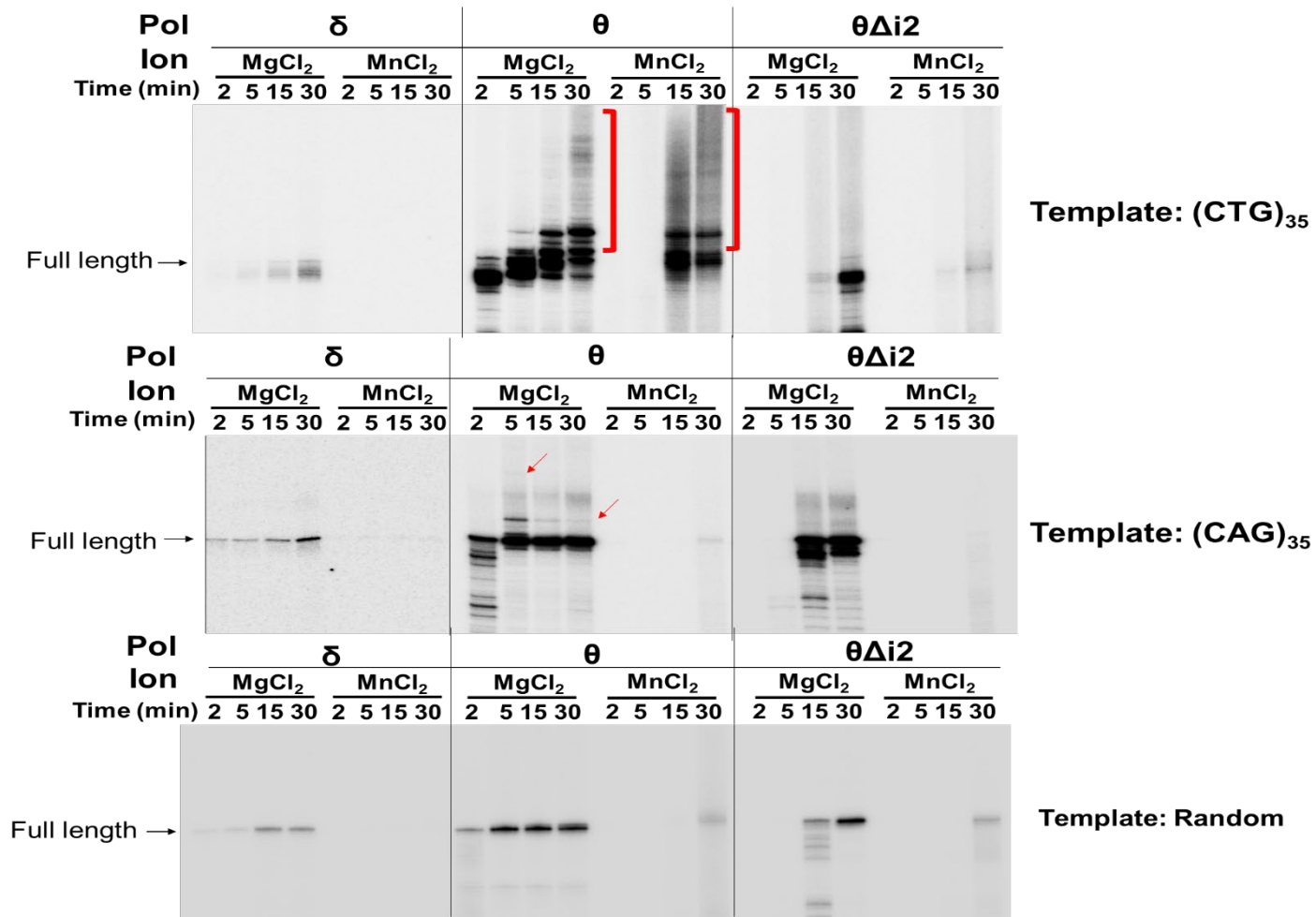
Figure 4.2: Polθ-mediated repeat expansion could be induced by the presence of Mn^{2+} ions. Reconstituted DNA synthesis with ^{32}P-labeled primers and pre-digested purified M13mp18 templates contains either 35 CAG, 35 CTG repeats, or representative random controls at various time points.

Mn^{2+} ions slowed down the processivities of Polδ, θ, and θΔi2

Polδ could synthesize to full length with the 110 nt oligo templates (Figure 4.1) but Polδ could not utilize Mn^{2+} for DNA synthesis with longer templates (Figure 4.2). PolθΔi2 could synthesize to full length in 15 min with the shorter substrates in the present of Mg^{2+} (Figure 4.1) and required 30 min or longer for replicating the longer template (Figure 4.2). Previous data showed that the processivity of wild-type Polθ with Mn^{2+} ions was slower [69], which was consistent with our observations shown in both Figures 4.1 and 4.2.

Mn^{2+}-induced expansion with Polθ did not depend on template sequences

Polδ-, θ-, and θΔi2-mediated small expansions were observed in all full-length products with Mn^{2+} and those expansions were not dependent on template length or sequences (Figure 4.1 and 4.2). With the longer templates (contain 35 TNR and their control) shown in Figure 4.2, although Polθ was the only polymerase that could synthesize to full-length in all substrates, the expansion facilitated by Polθ with Mn^{2+} was similar to that with Mg^{2+} because the larger full-length products were unable to be resolved with a 6% denaturing urea-polyacrylamide gel. Although the synthesized products from Polθ displayed a smear of larger expansion, which might be an indication of a greater variety of extension products than with Mg^{2+}, the lower resolution may be hiding the various small expansions in the 35 TNR repeats. Therefore, we were unable to observe an obvious Polθ-mediated expansion similar to that observed with the smaller oligo templates. The smearing

in the 35 TNR templates and the 20 TNR templates showed an enhanced TdT activity in the present of Mn^{2+} compared to Mg^{2+}.

4.3. Conclusion

Mn^{2+} decreased the processivity of all polymerases, and Mn^{2+} reversed the phenotype of PolθΔi2 back to wild type and facilitated small expansions with short oligos regardless of the template sequences. In the presence of Mn^{2+}, Polθ enhanced the small expansions in the short templates. This was not observed with Mg^{2+}. However, whether Mn^{2+} increases TdT activity in Polθ is still unknown. Further structural studies in Mn^{2+} incorporation in the enzymatic site of Polθ is needed to understand the importance of the unique large active site in Polθ's contribution to repeat expansion and TdT activity. Mn^{2+} is one of the trace metals that is essential for normal cell functions and presents in our food and environments [95]. Maintaining a balance between Mg^{2+} and Mn^{2+} in cell is necessary in DNA replication, our study showed that the substitution of Mg^{2+} with Mn^{2+} promoted error-prone synthesis in both Polδ and Polθ. Environmental studies showed that the increasing exposure of Mn^{2+} in fish will negatively affect fertility in male fish. Also, the exposure of increased amount of Mn^{2+} will lead to inducing ROS in cells, ROS is one of the most proposed mechanism for TNR expansion in the brain through BER as shown in this study. Kent et. al. showed that Polθ was able to cause more than 1 nt expansion in a 10:1 Mg^{2+}:Mn^{2+} [86]. With different ratio 2:1 Mg^{2+}:Mn^{2+}, Polθ was able to induce more error-prone synthesis similar to our results with Polθ and Mn^{2+} alone [86]. The competition of 2 different kinds of metal incorporation in the enzymatic pocket of Polθ can affect the fidelity and

processivity of Polθ. Ultimately, there are multiple divalent metal ions in a physiological cell condition, when the balance of metals ions is changed, it is possible that in HD patients have a different metal composition compared to non-HD patients which leaded to further development of HD.

Chapter 5

Conclusion and Future Directions

Polymerase θ has a unique large enzymatic site within the polymerase domain (Figure 5.1A). In **Chapter 3,** we showed that Polθ used its large enzymatic site for hairpin bypass synthesis (Figure 3.2 and 3.3), and the large catalytic site was required for repeat expansion without a preexisted hairpin structure (Figure 3.4). Our sequencing results also showed Polθ not only caused repeat expansion and had TdT activity, it also caused small repeat deletion (Figure 3.5). Our results suggested that the repeat expansions and TdT activity required the large enzymatic site in the polymerase domain. The deletion of insert 2 in the Polθ polymerase domain can reduce Polθ's active site and PolθΔi2 has higher fidelity on DNA synthesis. PolθΔi2 preferred error-free DNA synthesis with all the substrates used in this study. Moreover, the DNA synthesis ability of PolθΔi2 mutant was similar to that of wild-type Polθ, our results provided an insight into a potential target site for a small molecule to bind to insert 2 and to lessen the active site of Polθ (Figure 5.2). This approach will promote error-free DNA synthesis in pre-HD patients without ideally affecting the functions of Polθ in DNA repair and cell replication. Ultimately, the small molecule reduces the error-prone property of Polθ and help to reduce the life-long risk of developing Huntington's disease in patients and their offspring. The promotion of Polθ error-free synthesis can also lead to a potential direction in cancer therapy as the upregulated in protein expression of Polθ is highly correlated to the resistance of DSB inducing agents independent of homology recombination statues [73].
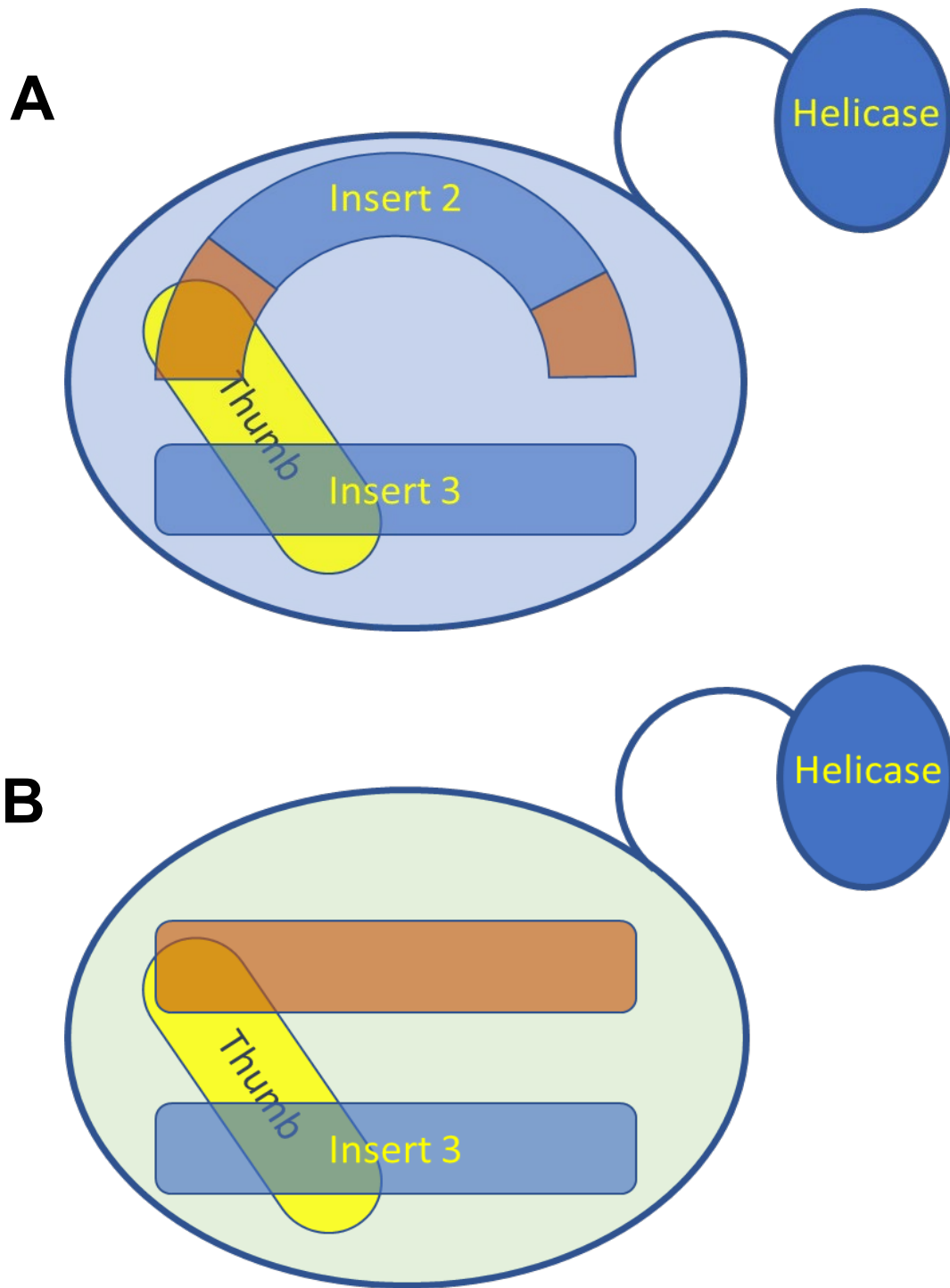
Figure 5.1: Schematic diagram of the full-length Polθ. **(A)** Full-length wild-type Polθ compared to **(B)** a proposed schematic diagram of PolθΔi2 mutant (bottom) with a smaller enzymatic site.
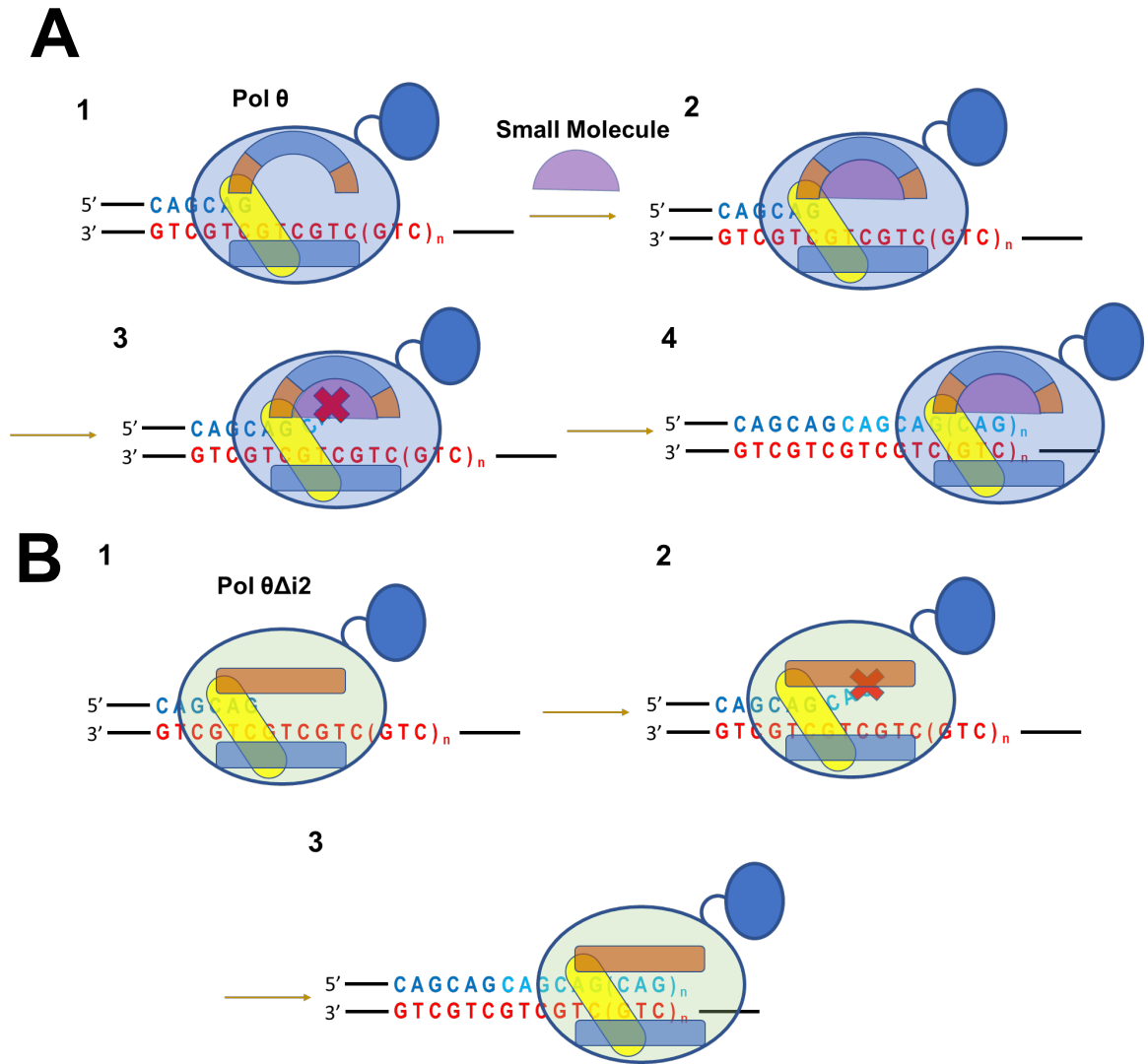
Figure 5.2: Model of a proposed therapeutic mechanism for Polθ. **(A)** Small molecule incorporates inside the enzymatic site to mimic error-free synthesis from the deletion of insertion 2. **(B)** PolθΔi2 mediates error-free DNA synthesis.

Appendix I

Author Contributions

Kara Y. Chan performed the experiments and purified all the proteins used, except PCNA; Dr. Janice Ortega and Dr. Wei Yang initiated the project; Dr. Janice Ortega and Dr. Bailin Zhao purified PCNA.

Appendix II

Acknowledgements

Appendix III

Commonly Used Abbreviation

Alt-EJ	alternative end joining
AP	apurinic/apyrmidinic
APE1	apurinic/apyrimidinic (AP) endonuclease
ATP	adenosine triphosphate
BER	base excision repair
bp	base pair
C	Complementary
CAG	cytosine-adenine-guanine
CTG	cytosine-thymine-guanine
DM1	myotonic dystrophy type 1
dNTP	deocynucleotide triphosphate
DPBS	Dulbecco's phosphate-buffered saline (PBS)
DSB	double strand break
dsDNA	double stranded DNA
DTT	dithiothreitol
EDTA	ethylediamine tetracetic acid
EtBr	ethidium bromide
FBS	fetal bovine serum
FEN1	flap endonuclease 1
g	gram
GFP	green fluorescent protein
h	hour
H_2O_2	Hydrogen peroxide
HD	Huntington's disease
HTT	human huntingtin gene
ICL	interstrand cross-links repair
IPTG	isopropyl-1-thio-β-D-galactopyranoside
Lig1	DNA ligase 1
m	mutant; mouse
Mg	Magnesium
min	minute
MMEJ	microhomology-mediated end joining
MMR	mismatch repair
Mn	Manganese
MSH	MutS homolog
MSI	microsatellite instability
MutSβ	complex of MSH2 and MSH3
NER	nucleotide excision repair

nt	nucleatide
PAGE	polyacrylamide gel electrophoresis
PARP1	poly(ADP-Ribose) polymerase 1
PCNA	proliferating cell nuclear antigen
PMSF	phenylmethylsulfonyl fluoride
Pol	polymerase
RFC	replication factor C
ROS	reactive oxygen species
RPA	replication protein A
s	second
SDS	sodium dodecyl sulfate
SSB	single strand break
ssDNA	single stranded DNA
TAE	tris/acetate acid/EDTA
TBE	tris/borate acid/EDTA
TBS	tris-buffered saline
TdT	Terminal deoxynucleotidyl transferase
TE	tris-EDTA
TNR	trinucleotide repeat
TT	6-4 double thymidine
UV	ultraviolet
V	Volts
β- ME	β-mercaptoethanol
Δi2	insert 2 deleted

Reference

1.   Kumar, A. and R.R. Ratan, *Oxidative Stress and Huntington's Disease: The Good, The Bad, and The Ugly.* J Huntingtons Dis, 2016. **5**(3): p. 217-237.

2.   Lee, H.O., J.M. Davidson, and R.J. Duronio, *Endoreplication: polyploidy with purpose.* Genes Dev, 2009. **23**(21): p. 2461-77.

3.   Sen, S., *Aneuploidy and cancer.* Curr Opin Oncol, 2000. **12**(1): p. 82-8.

4.   Storchova, Z. and D. Pellman, *From polyploidy to aneuploidy, genome instability and cancer.* Nat Rev Mol Cell Biol, 2004. **5**(1): p. 45-54.

5.   Xu, M., et al., *Base excision repair of oxidative DNA damage coupled with removal of a CAG repeat hairpin attenuates trinucleotide repeat expansion.* Nucleic Acids Res, 2014. **42**(6): p. 3675-91.

6.   Zahn, K.E., et al., *Human DNA polymerase theta grasps the primer terminus to mediate DNA repair.* Nat Struct Mol Biol, 2015. **22**(4): p. 304-11.

7.   Manoharan, S., et al., *The Role of Reactive Oxygen Species in the Pathogenesis of Alzheimer's Disease, Parkinson's Disease, and Huntington's Disease: A Mini Review.* Oxid Med Cell Longev, 2016. **2016**: p. 8590578.

8.   Gipson, T.A., et al., *Aberrantly spliced HTT, a new player in Huntington's disease pathogenesis.* RNA Biol, 2013. **10**(11): p. 1647-52.

9.   Hatters, D.M., *Putting huntingtin "aggregation" in view with windows into the cellular milieu.* Curr Top Med Chem, 2012. **12**(22): p. 2611-22.

10.   Kumar, A., M. Vaish, and R.R. Ratan, *Transcriptional dysregulation in Huntington's disease: a failure of adaptive transcriptional homeostasis.* Drug Discov Today, 2014. **19**(7): p. 956-62.

11.   Mochel, F. and R.G. Haller, *Energy deficit in Huntington disease: why it matters.* J Clin Invest, 2011. **121**(2): p. 493-9.

12.   Yano, H., et al., *Inhibition of mitochondrial protein import by mutant huntingtin.* Nat Neurosci, 2014. **17**(6): p. 822-31.

13.   Ayala-Pena, S., *Role of oxidative DNA damage in mitochondrial dysfunction and Huntington's disease pathogenesis.* Free Radic Biol Med, 2013. **62**: p. 102-110.

14.   Reijonen, S., et al., *Downregulation of NF-kappaB signaling by mutant huntingtin proteins induces oxidative stress and cell death.* Cell Mol Life Sci, 2010. **67**(11): p. 1929-41.

15. Sepers, M.D. and L.A. Raymond, *Mechanisms of synaptic dysfunction and excitotoxicity in Huntington's disease.* Drug Discov Today, 2014. **19**(7): p. 990-6.

16. Andre, R., L. Carty, and S.J. Tabrizi, *Disruption of immune cell function by mutant huntingtin in Huntington's disease pathogenesis.* Curr Opin Pharmacol, 2016. **26**: p. 33-8.

17. Machiela, E., et al., *Oxidative stress is increased in C. elegans models of Huntington's disease but does not contribute to polyglutamine toxicity phenotypes.* Neurobiol Dis, 2016. **96**: p. 1-11.

18. Salim, S., *Oxidative Stress and the Central Nervous System.* J Pharmacol Exp Ther, 2017. **360**(1): p. 201-205.

19. Talhaoui, I., et al., *Aberrant base excision repair pathway of oxidatively damaged DNA: Implications for degenerative diseases.* Free Radic Biol Med, 2017. **107**: p. 266-277.

20. Chan, N.L., et al., *Coordinated processing of 3' slipped (CAG)n/(CTG)n hairpins by DNA polymerases beta and delta preferentially induces repeat expansions.* J Biol Chem, 2013. **288**(21): p. 15015-22.

21. Petruska, J., N. Arnheim, and M.F. Goodman, *Stability of intrastrand hairpin structures formed by the CAG/CTG class of DNA triplet repeats associated with neurological diseases.* Nucleic Acids Res, 1996. **24**(11): p. 1992-8.

22. Hedglin, M., B. Pandey, and S.J. Benkovic, *Stability of the human polymerase delta holoenzyme and its implications in lagging strand DNA synthesis.* Proc Natl Acad Sci U S A, 2016. **113**(13): p. E1777-86.

23. Hosseinzadeh-Colagar, A., et al., *Microsatellite (SSR) amplification by PCR usually led to polymorphic bands: Evidence which shows replication slippage occurs in extend or nascent DNA strands.* Mol Biol Res Commun, 2016. **5**(3): p. 167-174.

24. Guo, J., et al., *MutSbeta promotes trinucleotide repeat expansion by recruiting DNA polymerase beta to nascent (CAG)n or (CTG)n hairpins for error-prone DNA synthesis.* Cell Res, 2016. **26**(7): p. 775-86.

25. Walker, F.O., *Huntington's disease.* The Lancet, 2007. **369**(9557): p. 218-228.

26. Nance, M.A., *Genetics of Huntington disease.* Handb Clin Neurol, 2017. **144**: p. 3-14.

27. Semaka, A., J.A. Collins, and M.R. Hayden, *Unstable familial transmissions of Huntington disease alleles with 27-35 CAG repeats (intermediate alleles).* Am J Med Genet B Neuropsychiatr Genet, 2010. **153B**(1): p. 314-20.

28. Semaka, A., et al., *Re: Autopsy-proven Huntington's disease with 29 trinucleotide repeats.* Mov Disord, 2008. **23**(12): p. 1794-5; author reply 1793.

29. Ridley, R.M., *<Ridley 1988 Anticipation in Huntingtons disease is inherited through the male line but may originate in the female.pdf>.* J Med Genet, 1988. **25**: p. 589-595.

30. Gil-Mohapel, J., P. Brocardo, and B. Christie, *The Role of Oxidative Stress in Huntington's Disease: Are Antioxidants Good Therapeutic Candidates?* Current Drug Targets, 2014. **15**(4): p. 454-468.

31. Kovtun, I.V., et al., *OGG1 initiates age-dependent CAG trinucleotide expansion in somatic cells.* Nature, 2007. **447**(7143): p. 447-52.

32. Nakatani, R., et al., *Large expansion of CTG*CAG repeats is exacerbated by MutSbeta in human cells.* Sci Rep, 2015. **5**: p. 11020.

33. Sorolla, M.A., et al., *Proteomic and oxidative stress analysis in human brain samples of Huntington disease.* Free Radic Biol Med, 2008. **45**(5): p. 667-78.

34. Kim, K.C., et al., *7,8-Dihydroxyflavone suppresses oxidative stress-induced base modification in DNA via induction of the repair enzyme 8-oxoguanine DNA glycosylase-1.* Biomed Res Int, 2013. **2013**: p. 863720.

35. Yoshimura, M., et al., *Vertebrate POLQ and POLbeta cooperate in base excision repair of oxidative DNA damage.* Mol Cell, 2006. **24**(1): p. 115-25.

36. Chan, N.L., et al., *The Werner syndrome protein promotes CAG/CTG repeat stability by resolving large (CAG)(n)/(CTG)(n) hairpins.* J Biol Chem, 2012. **287**(36): p. 30151-6.

37. Gacy, A.M., et al., *Trinucleotide repeats that expand in human disease form hairpin structures in vitro.* Cell, 1995. **81**(4): p. 533-40.

38. Lai, Y., et al., *Crosstalk between MSH2-MSH3 and polbeta promotes trinucleotide repeat expansion during base excision repair.* Nat Commun, 2016. **7**: p. 12465.

39. Lokanga, R.A., et al., *Heterozygosity for a hypomorphic Polbeta mutation reduces the expansion frequency in a mouse model of the Fragile X-related disorders.* PLoS Genet, 2015. **11**(4): p. e1005181.

40. Khodyreva, S.N., et al., *Apurinic/apyrimidinic (AP) site recognition by the 5'-dRP/AP lyase in poly(ADP-ribose) polymerase-1 (PARP-1).* Proc Natl Acad Sci U S A, 2010. **107**(51): p. 22090-5.

41. Gottipati, P., et al., *Poly(ADP-ribose) polymerase is hyperactivated in homologous recombination-defective cells.* Cancer Res, 2010. **70**(13): p. 5389-98.

42. Cistulli, C., et al., *AP endonuclease and poly(ADP-ribose) polymerase-1 interact with the same base excision repair intermediate.* DNA Repair (Amst), 2004. **3**(6): p. 581-91.

43. Luijsterburg, M.S., et al., *PARP1 Links CHD2-Mediated Chromatin Expansion and H3.3 Deposition to DNA Repair by Non-homologous End-Joining.* Mol Cell, 2016. **61**(4): p. 547-562.

44. Prasad, R., et al., *Mammalian Base Excision Repair: Functional Partnership between PARP-1 and APE1 in AP-Site Repair.* PLoS One, 2015. **10**(5): p. e0124269.

45. Mateos-Gomez, P.A., et al., *Mammalian polymerase theta promotes alternative NHEJ and suppresses recombination.* Nature, 2015. **518**(7538): p. 254-7.

46. Wyatt, D.W., et al., *Essential Roles for Polymerase theta-Mediated End Joining in the Repair of Chromosome Breaks.* Mol Cell, 2016. **63**(4): p. 662-673.

47. Yousefzadeh, M.J., et al., *Mechanism of suppression of chromosomal instability by DNA polymerase POLQ.* PLoS Genet, 2014. **10**(10): p. e1004654.

48. He, P. and W. Yang, *Template and primer requirements for DNA Poltheta-mediated end joining.* Proc Natl Acad Sci U S A, 2018. **115**(30): p. 7747-7752.

49. Arana, M.E., et al., *Low-fidelity DNA synthesis by human DNA polymerase theta.* Nucleic Acids Res, 2008. **36**(11): p. 3847-56.

50. Seki, M. and R.D. Wood, *DNA polymerase theta (POLQ) can extend from mismatches and from bases opposite a (6-4) photoproduct.* DNA Repair (Amst), 2008. **7**(1): p. 119-27.

51. Yoon, J.H., et al., *Error-Prone Replication through UV Lesions by DNA Polymerase theta Protects against Skin Cancers.* Cell, 2019. **176**(6): p. 1295-1309 e15.

52. Kawamura, K., et al., *DNA polymerase theta is preferentially expressed in lymphoid tissues and upregulated in human cancers.* Int J Cancer, 2004. **109**(1): p. 9-16.

53. Yoon, J.H., et al., *A role for DNA polymerase theta in promoting replication through oxidative DNA lesion, thymine glycol, in human cells.* J Biol Chem, 2014. **289**(19): p. 13177-85.

54. Yousefzadeh, M.J. and R.D. Wood, *DNA polymerase POLQ and cellular defense against DNA damage.* DNA Repair (Amst), 2013. **12**(1): p. 1-9.

55. Kent, T., et al., *Mechanism of microhomology-mediated end-joining promoted by human DNA polymerase theta.* Nat Struct Mol Biol, 2015. **22**(3): p. 230-7.

56. Masuda, K., et al., *DNA polymerases eta and theta function in the same genetic pathway to generate mutations at A/T during somatic hypermutation of Ig genes.* J Biol Chem, 2007. **282**(24): p. 17387-94.

57. Ceccaldi, R., et al., *Homologous-recombination-deficient tumours are dependent on Poltheta-mediated repair.* Nature, 2015. **518**(7538): p. 258-62.

58. Goullet de Rugy, T., et al., *Excess Poltheta functions in response to replicative stress in homologous recombination-proficient cancer cells.* Biol Open, 2016. **5**(10): p. 1485-1492.

59. Kent, T., et al., *DNA polymerase theta specializes in incorporating synthetic expanded-size (xDNA) nucleotides.* Nucleic Acids Res, 2016. **44**(19): p. 9381-9392.

60. Lemee, F., et al., *DNA polymerase theta up-regulation is associated with poor survival in breast cancer, perturbs DNA replication, and promotes genetic instability.* Proc Natl Acad Sci U S A, 2010. **107**(30): p. 13390-5.

61. Fernandez-Vidal, A., et al., *A role for DNA polymerase theta in the timing of DNA replication.* Nat Commun, 2014. **5**: p. 4285.

62. Goff, J.P., et al., *Lack of DNA polymerase theta (POLQ) radiosensitizes bone marrow stromal cells in vitro and increases reticulocyte micronuclei after total-body irradiation.* Radiat Res, 2009. **172**(2): p. 165-74.

63. van Schendel, R., et al., *Polymerase Theta is a key driver of genome evolution and of CRISPR/Cas9-mediated mutagenesis.* Nat Commun, 2015. **6**: p. 7394.

64. Hogg, M., et al., *Lesion bypass activity of DNA polymerase theta (POLQ) is an intrinsic property of the Poldomain and depends on unique sequence inserts.* J Mol Biol, 2011. **405**(3): p. 642-52.

65. Laverty, D.J., et al., *The A-Rule and Deletion Formation During Abasic and Oxidized Abasic Site Bypass by DNA Polymerase theta.* ACS Chem Biol, 2017. **12**(6): p. 1584-1592.

66. Muzzini, D.M., et al., *Caenorhabditis elegans POLQ-1 and HEL-308 function in two distinct DNA interstrand cross-link repair pathways.* DNA Repair (Amst), 2008. **7**(6): p. 941-50.

67. Black, S.J., et al., *DNA Polymerase theta: A Unique Multifunctional End-Joining Machine.* Genes (Basel), 2016. **7**(9).

68. Newman, J.A., et al., *Structure of the Helicase Domain of DNA Polymerase Theta Reveals a Possible Role in the Microhomology-Mediated End-Joining Pathway.* Structure, 2015. **23**(12): p. 2319-30.

69. Maga, G., et al., *DNA Polymerase θ Purified from Human Cells is a High-fidelity Enzyme.* Journal of Molecular Biology, 2002. **319**(2): p. 359-369.

70. Ozdemir, A.Y., et al., *Polymerase theta-helicase efficiently unwinds DNA and RNA-DNA hybrids.* J Biol Chem, 2018. **293**(14): p. 5259-5269.

71. Beagan, K., et al., *Drosophila DNA polymerase theta utilizes both helicase-like and polymerase domains during microhomology-mediated end joining and interstrand crosslink repair.* PLoS Genet, 2017. **13**(5): p. e1006813.

72. Lee, Y.S., Y. Gao, and W. Yang, *How a homolog of high-fidelity replicases conducts mutagenic DNA synthesis.* Nat Struct Mol Biol, 2015. **22**(4): p. 298-303.

73. Wang, Z., et al., *DNA polymerase theta (POLQ) is important for repair of DNA double-strand breaks caused by fork collapse.* J Biol Chem, 2019. **294**(11): p. 3909-3919.

74. Campbell, A.E., et al., *NuRD and CAF-1-mediated silencing of the D4Z4 array is modulated by DUX4-induced MBD3L proteins.* Elife, 2018. **7**.

75. Fazlieva, R., et al., *Proofreading exonuclease activity of human DNA polymerase delta and its effects on lesion-bypass DNA synthesis.* Nucleic Acids Res, 2009. **37**(9): p. 2854-66.

76. Binz, S.K., et al., *The phosphorylation domain of the 32-kDa subunit of replication protein A (RPA) modulates RPA-DNA interactions. Evidence for an intersubunit interaction.* J Biol Chem, 2003. **278**(37): p. 35584-91.

77. Li, G.M. and P. Modrich, *Restoration of mismatch repair to nuclear extracts of H6 colorectal tumor cells by a heterodimer of human MutL homologs.* Proc Natl Acad Sci U S A, 1995. **92**(6): p. 1950-4.

78. Zhang, Y., et al., *Reconstitution of 5'-directed human mismatch repair in a purified system.* Cell, 2005. **122**(5): p. 693-705.

79. Chen, S., et al., *Freeze-thaw increases adeno-associated virus transduction of cells.* Am J Physiol Cell Physiol, 2006. **291**(2): p. C386-92.

80. Gaggioli, V. and P. Zegerman, *Terminating the replication helicase.* Nat Cell Biol, 2017. **19**(5): p. 410-412.

81. Langbehn, D.R., et al., *CAG-repeat length and the age of onset in Huntington disease (HD): a review and validation study of statistical approaches.* Am J Med Genet B Neuropsychiatr Genet, 2010. **153B**(2): p. 397-408.

82. Talhaoui, I., et al., *Aberrant base excision repair pathway of oxidatively damaged DNA: Implications for degenerative diseases.* Free Radic Biol Med, 2016.

83. Newman, J.A., et al., *Structure of the Helicase Domain of DNA Polymerase Theta Reveals a Possible Role in the Microhomology-Mediated End-Joining Pathway.* Structure, 2015. **23**(12): p. 2319-2330.

84.   Semaka, A., et al., *High frequency of intermediate alleles on Huntington disease-associated haplotypes in British Columbia's general population.* Am J Med Genet B Neuropsychiatr Genet, 2013. **162B**(8): p. 864-71.

85.   Mollica, P.A., *DNA Repair Deficiency in Huntington's Disease Fibroblasts and Induced Pluripotent Stem Cells*, in *Biological Sciences*. 2015, Old Dominion University: ODU Digital Commons. p. 1-172.

86.   Kent, T., et al., *Polymerase theta is a robust terminal transferase that oscillates between three different mechanisms during end-joining.* Elife, 2016. **5**.

87.   Shima, N., R.J. Munroe, and J.C. Schimenti, *The mouse genomic instability mutation chaos1 is an allele of Polq that exhibits genetic interaction with Atm.* Mol Cell Biol, 2004. **24**(23): p. 10381-9.

88.   Trottier, Y., V. Biancalana, and J.L. Mandel, *Instability of CAG repeats in Huntington's disease: relation to parental transmission and age of onset.* J Med Genet, 1994. **31**(5): p. 377-82.

89.   Duyao, M., et al., *Trinucleotide repeat length instability and age of onset in Huntington's disease.* Nat Genet, 1993. **4**(4): p. 387-92.

90.   Gilbert, S.F., *Developmental biology.* 6th ed. 2000, Sunderland, Mass.: Sinauer Associates. xviii, 749 p.

91.   Semaka, A., et al., *CAG size-specific risk estimates for intermediate allele repeat instability in Huntington disease.* J Med Genet, 2013. **50**(10): p. 696-703.

92.   Williams, B.B., et al., *Altered manganese homeostasis and manganese toxicity in a Huntington's disease striatal cell model are not explained by defects in the iron transport system.* Toxicol Sci, 2010. **117**(1): p. 169-79.

93.   Frank, E.G. and R. Woodgate, *Increased catalytic activity and altered fidelity of human DNA polymerase iota in the presence of manganese.* J Biol Chem, 2007. **282**(34): p. 24689-96.

94.   Gao, Y. and W. Yang, *Capture of a third Mg(2)(+) is essential for catalyzing DNA synthesis.* Science, 2016. **352**(6291): p. 1334-7.

95.   Madison, J.L., et al., *Disease-toxicant interactions in manganese exposed Huntington disease mice: early changes in striatal neuron morphology and dopamine metabolism.* PLoS One, 2012. **7**(2): p. e31024.

VITA

Kara Y. Chan

Place of Birth: Kowloon, Hong Kong

Education

2008- 2012   **Bachelor of Science in Biochemistry**
Minor in Biology, Mathematics, Honors
Department of Chemistry
Northern Kentucky University**,** Highland Heights,
Kentucky, USA

Research Experience

2012-2019   **Graduate Research in Toxicology and Cancer Biology**
A Novel Translesion DNA Polymerase And Trinucleotide Repeat Stability
Department of Toxicology and Cancer Biology
College of Medicine
University of Kentucky**,** Lexington, Kentucky, USA
Principal Investigator: Dr. Guo-Min Li

2009-2012   Undergraduate Research in Chemistry
Solvent-Dependent Properties of Estrongens
Department of Chemistry
Northern Kentucky University**,** Highland Heights,
Kentucky, USA
Principal Investigator: Dr. Patrick Michael Hare

Professional Development

2019	**Poster Presentation:**
	2019 EMGS Special Symposium Genome Maintenance Systems in Cancer Etiology and Therapy: A Tribute to Paul Modrich
	Error-prone DNA synthesis by DNA polymerase θ causes CAG repeat expansions
	Washington, DC, USA
2012	**Platform Presentation:**
	The 2012 Celebration of Student Research and Creativity- Honors Capstone Project
	Effects on Estrogens
	Northern Kentucky University**,** Highland Heights, Kentucky, USA
2012	**Poster Presentation:**
	243rd American Chemical Society National Meeting & Exposition
	Solvent Effects on the Excited State Lifetimes of Natural Estrogens
	San Diego, California, USA
2011	**Platform Presentation:**
	2011 KAS Annual Meeting
	Solvent Effects on the Excited State Lifetimes of Natural Estrogens
	Murray State University, Murray, Kentucky, USA
2011	**Poster Presentation:**
	The 2011 Summer Celebration of Student Research
	The Excited-State Lifetimes of Estrogens
	Northern Kentucky University**,** Highland Heights, Kentucky, USA
2011	**Platform Presentation:**
	The 2011 Celebration of Student Research and Creativity
	Fluorescence Quantum Yields of Estrone and 17β-Estradiol
	Northern Kentucky University**,** Highland Heights, Kentucky, USA
2011	**Poster Presentation:**
	241st American Chemical Society National Meeting & Exposition
	Photophysics of the natural estrogens estrone and 17-

beta-estradiol
Anaheim, California, USA

2010	**Poster Presentation:** The 2010 Summer Celebration of Student Research Emission Yields of Estrogens in Several Solvents Northern Kentucky University**,** Highland Heights, Kentucky, USA
2010	**Poster Presentation:** The 2010 Celebration of Student Research and Creativity Solvent Dependence of Estrogen Fluorescence Northern Kentucky University**,** Highland Heights, Kentucky, USA
2009	**Poster Presentation:** 2009 KAS Annual Meeting Solvent Dependence of Estrogen Fluorescence Northern Kentucky University**,** Highland Heights, Kentucky, USA
2009	**Poster Presentation:** The 2009 Summer Celebration of Student Research Absorption and Emission Yields of Estrone and 17-beta- Estradiol Northern Kentucky University**,** Highland Heights, Kentucky, USA
2009	**Poster Presentation:** The 2009 Celebration of Student Research and Creativity Solvent effects on the steady state spectroscopy of estrone and estradiol Northern Kentucky University**,** Highland Heights, Kentucky, USA

Research Fellowships and Scholarship

2014-2016	NRSA Institutional Predoctoral Training Grants (T32)
2011	2011 Dorothy Westerman Herrmann Summer Research Fellowship

Scholarly Activities

2011-2012	Vice President of Student Affiliates of the American Chemical Society
	Northern Kentucky University, Highland Heights, Kentucky, USA
2010-2012	President of Norse Badminton Club
	Northern Kentucky University, Highland Heights, Kentucky, USA

Academic Honors and Awards

2012	Cum Lodi, B.S. in Biochemistry
2012	SAACS/ GSE Service Award
2008-2009, 2011-2012	Dean's List
2010	President's Honors List

Affiliations/Memberships

2011- Present	Gamma Sigma Epsilon Chemistry Honor Society
2010- Present	Golden Key International Honour Society
2009-2011	Student Affiliates of the American Chemical Society

Publications

Keogh, N.; **Chan, K.Y.**; Li, G.M.; Lahue, R.S. MutSβ abundance and Msh3 ATP hydrolysis activity are important drivers of CTG●CAG repeat expansions. *N.A.R.* **2017,** *45(17),* 10068-10078. DOI: 10.1093/nar/gkx650

Chan, K.Y.; Courtois, B.; Loose, K.; Hare, P.M. Solvent-Dependent Fluorescence Lifetimes of Estrone, 17β-Estradiol and 17α-Ethinylestradiol. *Photochem. & Photobio.* **2013,** *89*, 294-299. DOI: 10.1111/php.12011

Chan, K.Y.; Gavaghan, B.M.; Stoeckel, A.W.; Irizarry, K.; Hare, P.M. Solvent Effects on the Steady State Photophysics of Estrone and 17β-estradiol. *Photochem. & Photobio.* **2012,** *88*, 295-303. DOI: 10.1111/j.1751-1097.2011.01066.x