THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF BIOCHEMISTRY AND MOLECULAR BIOLOGY

CHARACTERIZING REPLICATION PROTEIN RECRUITMENT AT GAA MICROSATELLITE REPEAT IN A MAMMALIAN PRIMARY REPLICATION CYCLE

REBECCA ANNE DEGIOSIO
SPRING 2017

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Biochemistry and Molecular Biology
with honors in Biochemistry and Molecular Biology

Reviewed and approved* by the following:

Maria Krasilnikova
Research Assistant Professor of Biochemistry and Molecular Biology
Thesis Supervisor

Lorraine Santy
Associate Professor of Biochemistry and Molecular Biology
Honors Adviser

* Signatures are on file in the Schreyer Honors College.

ABSTRACT

The expansion of microsatellites – stretches of DNA consisting of repeating 2-5 nucleotide motifs – has been implicated as the key precipitant to a host of genetic disorders as well as general genomic instability. The mechanism of such expansion, however, remains unclear. It has been previously demonstrated that mammalian cells undergo a distinct primary cycle of DNA replication which is particularly susceptible to replication stalling by non-canonical DNA structures formed within repetitive genomic regions. It is unknown if this primary replication cycle plays any role in facilitating microsatellite repeat expansion. Herein, we investigate patterns of replication protein recruitment during primary and subsequent replication cycles in regions surrounding a trinucleotide repeat. SV40-origin based plasmids containing GAA_{57} repeat in two orientations or no repeat transfected into mammalian cells were used as models of DNA replication. Chromatin immunoprecipitation (ChIP) was performed during both primary and subsequent cycles using antibodies to polymerase alpha (pol α)/primase, MCM4, and polymerase zeta (pol ζ) proteins. Results were analyzed by regular PCR and quantitative PCR (qPCR). We demonstrate that pol α and MCM4 are more strongly recruited in repeat-containing plasmids than in control plasmid during the primary replication cycle specifically. We propose that these proteins' recruitment in this early stage results from elevated non-canonical DNA structure formation in the presence of repeat, permitted by underdeveloped chromatin structure. The polymerase may then initiate the unstable primary cycle of replication, leading to possible repeat expansion and increased genomic instability.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Primarily I would like to thank Dr. Maria Krasilnikova for mentoring me throughout this research project, among others. In addition to being an extremely caring and helpful supervisor who has treated me and her other mentees like family, Dr. Krasilnikova taught me to think critically and independently about my research as well as the work of others. She has been the single most influential person to my development as a professional scientist at Penn State, and I feel extremely fortunate to have been a part of her laboratory group.

I'd also like to thank Dr. Gong Chen of Penn State's Biology department, who has provided our group with several reagents for use in the ChIP and qPCR procedures presented herein. He has also offered valuable advice towards the writing of this thesis and the presentation of scientific data, which will no doubt continue to prove useful throughout my career. Jiuchao Yin has been particularly helpful to my secondary research project in Dr. Chen's group, so I would like to thank her for her consistent support and guidance.

I'd also like to acknowledge the support of current Krasilnikova lab members - Caitlin Komm, Ari Scechtman and Hansa Nawani – as well as past members. They have been nothing but helpful and encouraging to me, and I wish them all the best after their time at Penn State!

Chapter 1

Introduction

Microsatellites – also called simple sequence repeats (SSRs) or short tandem repeats (STRs) - are defined as stretches of DNA consisting of a repeating 2-5 nucleotide motif. Such motifs typically repeat 5-50 times, but can expand to over 200 iterations. Repetitive elements such as microsatellites are found in every organism's genome. They are found in varying proportions across genomes, accounting for approximately 3% of the human genome specifically (30). Their distribution is non-random, concentrating mainly in noncoding DNA (23). For example, in the Japanese pufferfish *Fugu rubripes*, only 11.6% of 6042 microsatellite regions were located in protein-coding regions (15). This distribution pattern is most likely due to the greater tolerance toward mutations found in noncoding genomic regions.

The repetitive nature of microsatellite regions leads to an unusually high rate of mutation: approximately 10^{-2}-10^{-6} events per generation (23). Unlike the rarer point mutations typical of non-repetitive DNA sequences, mutations within microsatellite regions often represent expansions or contractions of repeat length – that is, the addition or deletion of repeating units. Such mutations are mostly attributable to DNA strand slippage (also called replication slippage), in which a repeating region of ssDNA (on either the template or nascent strand) loops out from the double helix structure (often adopting its own "hairpin" structure, **Figure 1.2B**) and replication machinery "skips" over the extra sequence. A second cycle of replication - or in some cases, DNA repair mechanisms - then add or delete repeating units, leading to an overall insertion or deletion (**Figure 1.1**).

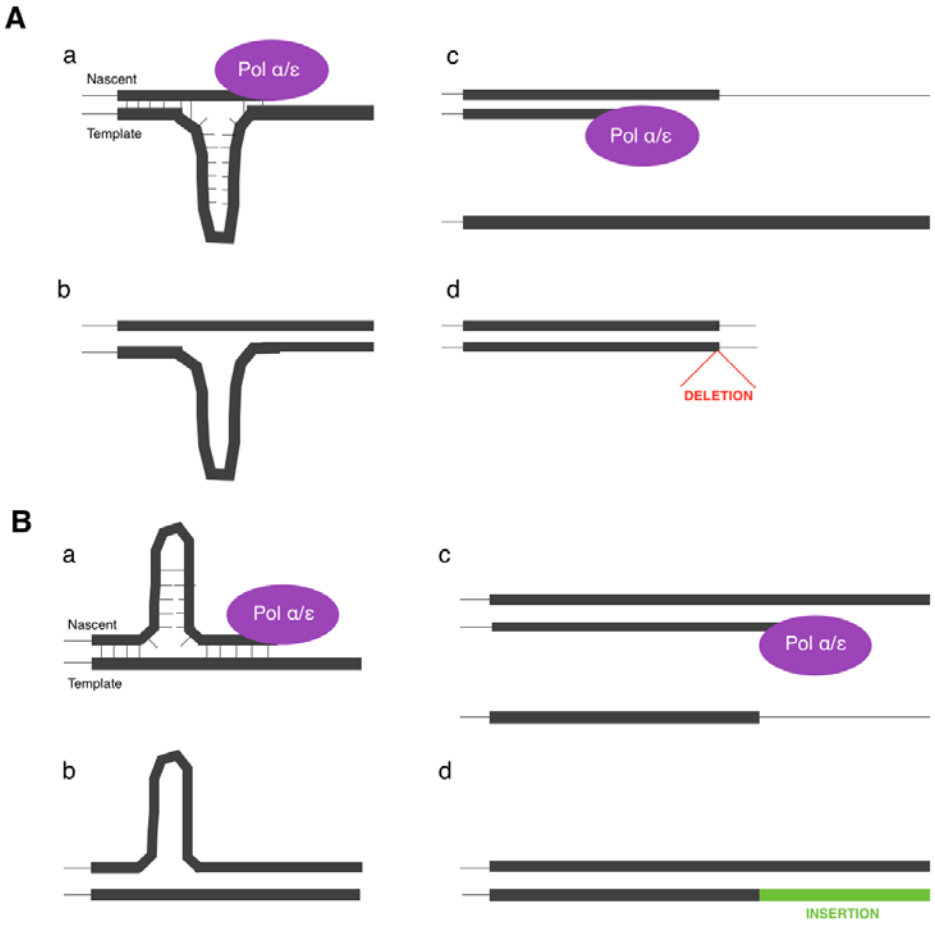


Figure 1.1. Schematic of replication slippage resulting in microsatellite (thick line) contraction (A) or expansion (B). Replication proceeds through an obstacle, in this case a hairpin structure on either strand (a). Completion of replication results in one strand being longer than the other (b). A second cycle of replication uses the mutated nascent strand as a template (c), resulting in a total deletion or insertion (d).

The hairpin structure is just one of many non-canonical DNA structures (also called non-B structures or secondary structures) that have been observed in repetitive regions such as microsatellites. These structures are so named as they vary from the "canonical" double-helical structure of DNA. In total, more than 10 distinct DNA conformations have been observed and

characterized (12). Such conformations include the G-quadruplex structure, in which evenly-spaced guanine nucleotides hydrogen-bond to each other in squares, creating structures called G-tetrads. The repeating trinucleotide motif GAA is particularly susceptible to triplex formation, in which a single-stranded region of DNA dissociates from its double helix and forms a triple helix with an upstream region of double-stranded DNA (18). The aforementioned hairpin structure represents a single repetitive strand of DNA which has hydrogen-bonded to itself. Finally, the cruciform structure is similar to a hairpin structure forming on both DNA strands simultaneously. See **Figure 1.2** for a review of these structures.

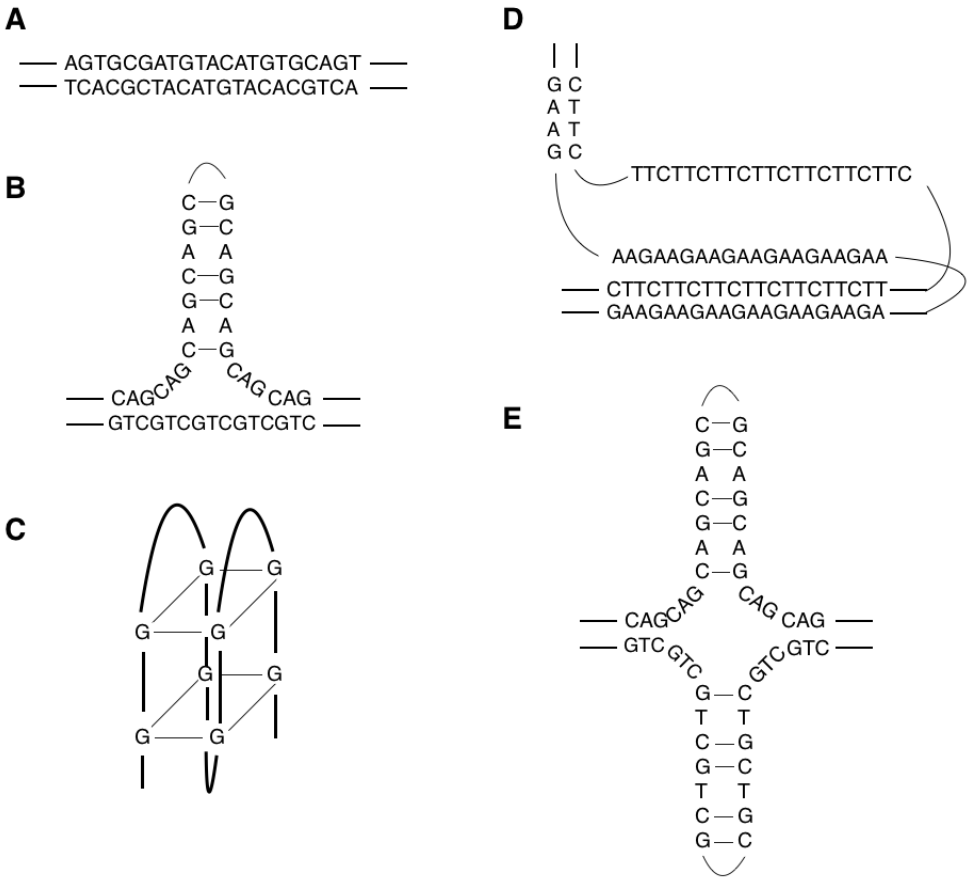


Figure 1.2. Illustrations of several non-canonical DNA structures that can form in microsatellite regions. Canonical double-helix DNA (A) is compared to a hairpin (B), G-quartet (C), triplex (D), and cruciform structure (E).

It has been suggested that the non-canonical structures formed by microsatellite sequences may help mediate several of their various functions. For instance, it has been suggested that repetitive elements like microsatellites may have an influence over DNA recombination. Dinucleotide repeats appear to be particularly apt recombination sites, due to their high affinity for recombination enzymes such as Rad51 and RecA. The degree of secondary structure formation within such dinucleotide repeats has been shown to influence the efficiency of these proteins' binding (6). Microsatellite regions are also known to influence DNA replication processes. For instance, previous work in our laboratory used 2D-PAGE gels of replication intermediates to demonstrate that $(GAA)_n$ repeat can stall replication of an SV40 origin-based plasmid. Efficacy of the stalling was dependent on the length of the repeat region. Moreover, plasmid sample that was subjected to alkali conditions and Nt.BstNBI nicking enzyme to prevent supercoiling, resulted in a significantly less pronounced replication arc on 2D-PAGE gel, suggesting reduced stalling in the absence of secondary structures (10). Finally, repetitive regions are understood to have profound impacts on gene expression. They can serve as purposeful promoter elements - as in the case of a critical GAA_{12} repeat in the promoter of *E. Coli lacZ* gene – or can lead to stalled transcription through the formation of various non-canonical structures which block RNA polymerase progression (3, 23).

Beyond these, still more possible functions of microsatellite regions have been proposed. For instance, they appear to exhibit significantly conserved distribution patterns within taxi, suggesting a possible role in taxon-specific chromosomal organization. Repeating mono-, di-, tri- and tetranucleotide motifs also tend to be highly concentrated at centromeric regions, where it has been suggested they may play roles in sister chromatid cohesion and/or kinetochore formation (23, 27). **Figure 1.3** summarizes the various roles of microsatellites in the genome.

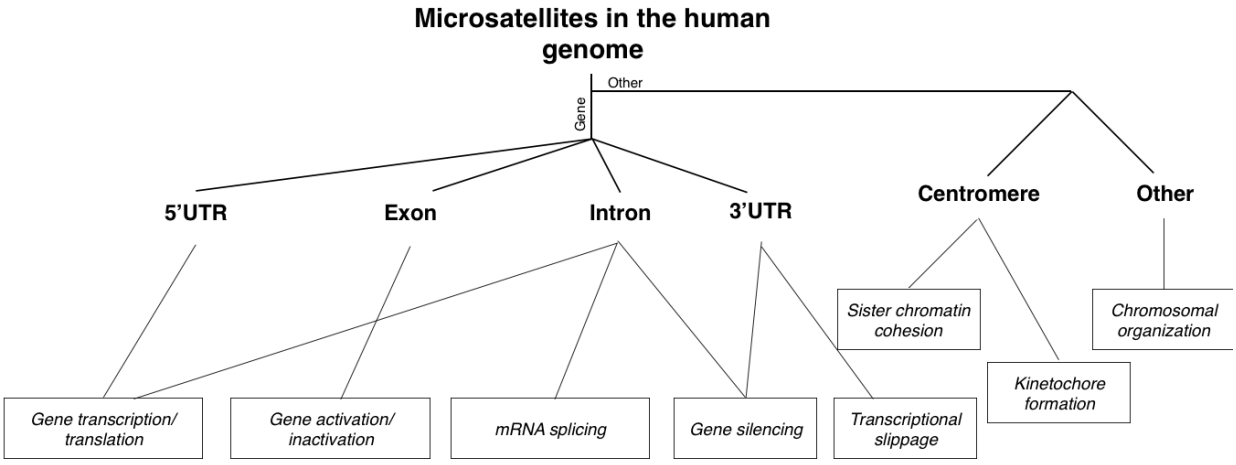


Figure 1.3. Functions of microsatellites in the genome, categorized by location.

One significant type of microsatellite is the trinucleotide repeat: a stretch of repeating three-nucleotide units. Expansion of trinucleotide repeats (in any portion of a gene, coding or noncoding) can lead to a variety of disorders called trinucleotide repeat disorders. These disorders often have an impact on neurological development and function. For example, Fragile X syndrome – a neurodevelopmental disorder with similar symptoms to autism – is caused by the expansion of a $(CGG)_n$ repeat in the 5'UTR of Fragile X mental retardation 1 (*FMR1*) gene. 14 disorders are known to be caused by trinucleotide repeat expansion (23). More examples of trinucleotide repeat disorders are shown in **Table 1.1**.

Disease	Repeating unit	*Gene* [location]	Normal length (units)	Pre-mutation length (units)	Disease length (units)
Dentatorubral-Pallidoluysian Atrophy	CAG	*ATN1* [exon 5]	6-35	35-48	49-88
Fragile X Syndrome	CGG	*FMR1* [5'UTR]	6-50	55-200	200-4000
Friedreich's Ataxia	GAA	*FXN* [intron 1]	5-30	31-100	70-1000
Huntington's Disease	CAG	*HTT* [exon 1]	6-29	29-37	38-180
Myotonic Dystrophy (type I)	CTG	*DMPK* [3'UTR]	5-37	37-50	>50
Spinocerebellar Ataxia (type I)	CAG	*ATXN1* [exon 8]	6-39	40	41-83

Table 1.1. Trinucleotide repeat disorders and their causative repeat expansions. "Pre-mutation" lengths indicate a high risk of progeny becoming affected by the disorder.

A critical feature of a trinucleotide repeat is its so-called "threshold" length, often approximately 100-200 units. After achieving this length, the repeat can expand or contract much more rapidly – in some cases, accumulating several thousand new copies through only a few generations (25). The rapidly increasing length of the repeat often leads to more severe disease symptoms and earlier disease outset over successive generations - a phenomenon known as anticipation.

It is known that fewer secondary structures form in those trinucleotide repeats which have not been observed to expand. Therefore, it has been suggested that the threshold length of trinucleotide repeats may merely correspond to a stability threshold for secondary structures to form, at which point these structures destabilize the region and lead to expansion. However, further studies have established that secondary structures can form at much smaller repeat lengths than the general threshold value of 100-200 units; for example, stable hairpin structures

have been observed in synthetic 10-unit stretches of CTG repeat (25, 29). The true mechanism of instability thus remains unclear.

Previous work in our lab has demonstrated the existence of a unique primary mode of mammalian DNA replication which is distinguished from canonical subsequent replication cycles by randomized initiation sites and underdeveloped chromatin structure. As studied in SV40-origin based plasmid pUCneo, this primary replication cycle occurs by 6 hours post-transfection into mammalian cells, and is not intiated at the SV40 origin. It is also not stimulated by sites of DNA damage, nor does it depend on homologous recombination or A/T rich regions (9). However, the mechanism of this novel, randomly-initiated form of replication - and which components of the canonical replication fork mediate it - is not yet known.

This primary cycle of replication is significantly more susceptible to fork stalling by $(GAA)_n$ repeats than subsequent cycles. This is presumably due to the repeat's secondary structure, which may form more easily in the absence of mature chromatin (9, 10). It has been reported, for instance, that the formation of triple helices within $d(GA•TC)_n$ sequences is prevented by nucleosomes (16). Indeed, the timing of the primary replication cycle matches well with nucleosome assembly on pUCneo as analyzed by micrococcal nuclease digestion of intact chromatin (9). Moreover, these events also appear to coincide with timepoints of heightened microsatellite repeat instability in the genome; trinucleotide repeat expansions have generally been understood to originate in early stages of embryogenesis, during which the unusual primary replication cycle would occur (26).

From this we theorize that the non-canonical DNA structures formed by microsatellites during the primary replication cycle generate single-stranded regions of DNA that can serve as loading docks for replication proteins. This recruitment of replication proteins may then initiate

the unstable primary cycle of DNA replication, which - due to its increased vulnerability to fork stalling - could significantly contribute to microsatellite instability as well as repeat expansions leading to genetic disorder. Multiple proteins which specifically recognize triplex structures have already been identified and characterized from HeLa cell extracts (28).

Herein, we analyze the recruitment of several replication proteins at various locations on a $(GAA)_{57}$-containing plasmid transfected into mammalian cells, during both primary and subsequent replication cycles. Proteins studied include pol α/primase, mini-chromosome maintenance complex protein 4 (MCM4), and polymerase zeta (pol ζ).

ORC is the first protein complex to load onto DNA in the process of replication. It is, in most eukaryotes, followed by Cdc6, Cdt1, and a hexamer of MCM proteins (which act as a DNA helicase complex). This total assembly of proteins is known as the pre-replication (pre-RC) complex. To initiate DNA replication, pol α and primase enzymes combine in a separate complex. The former enzyme intiates replication of DNA close to origins of replication, as well as Okazaki fragments throughout the replication process. The latter enzyme generates RNA primers from which pol α extends.

In contrast, pol ζ is a translesion synthesis (TLS) polymerase, mediating DNA replication across obstacles formed by DNA-damaging elements such as radiation. Indeed, though this enzyme is non-essential to yeast, pol ζ homolog deficiency in mice resulted in embryonic death, demonstrating its critical role in protecting against DNA damage (4). The depletion of similar TLS polymerases (specifically eta [η] and kappa [κ]) in HeLa cells containing many copies of noncanonical structure-forming sequences resulted in more double-stranded breaks (DSBs) than wild-type cells, suggesting the additional importance of TLS polymerases in transversing secondary structures which may stall other polymerases (5). Of particular relevance to our

studies, the loss of pol ζ in mammalian cells has been shown to lead to severe chromosomal instability, with frequent translocation and increased extrachromosomal material (also called double minute chromosomes) found in the nucleus (32). Pol ζ is also known to recognize and bind cruciform structures such as those found in microsatellites (33). Therefore, we saught to evaluate its patterns of recruitment to microsatellites during the primary mammalian replication cycle, speculating that it may have a preventative effect against unstable replication initiation.

As studied in the yeast *Saccharomyces cerevisiae,* pol ζ works closely with protein Rev1. In the event that a normal, elongating polymerase such as α or eta [ε] is stalled during replication, the pol ζ/Rev1 complex is activated. Rev1 of this complex ubiquitinates proliferating cell nuclear antigen (PCNA), an essential replication protein which typically acts as a DNA clamp. This ubiquitination signal then mediates the lesion bypass as shown in **Figure 1.4** (19).

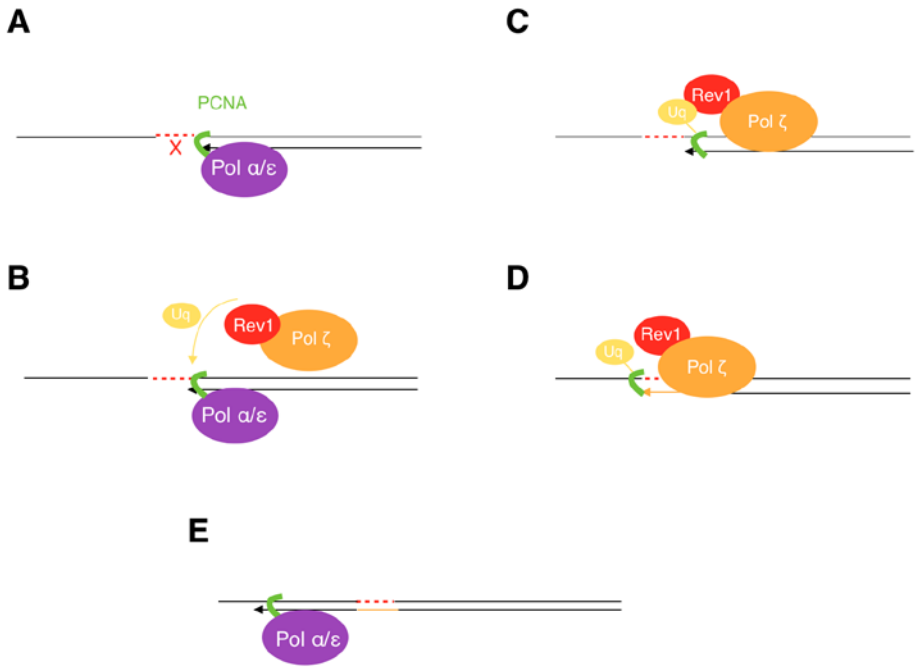


Figure 1.4. A model of lesion bypass by pol ζ/Rev1 complex. Regular, processive polymerase encounters a replication block, such as that caused by radiation damage (dotted red region) (A). This stalling activates pol ζ/Rev1 complex, which ubiquitinates PCNA (B). This results in dissociation of regular polymerase and docking of pol ζ/Rev1 complex (C). The complex completes replication through the blocked region (D). Through an unknown process, the complex then dissociates, the ubiquitin tag is removed and regular polymerase redocks to continue replication (E).

As previously mentioned, our studies focus on the trinucleotide repeat $(GAA)_n$. Expansion of this repeat in the first intron of frataxin (*FXN*) gene causes Friedreich's Ataxia (FDRA), a disease in which nervous tissue of the spinal cord degenerates, leading to a loss of control over muscle function in all four limbs. It can also lead to vision and hearing impairments, slurred speech, and heart problems such as atrial fibrillation. It is the most common inherited ataxia, with a prevalence of 1 in 50,000 people. The expanded $(GAA)_n$ repeat disrupts *FXN* gene expression, resulting in frataxin deficiency. The biological function of frataxin remains

unknown, though studies of yeast frataxin homolog knockouts suggest a role in regulating mitochondrial iron levels (7). As in most trinucleotide repeat disorders, FDRA demonstrates genetic anticipation, where the length of the repeat is positively associated with disease severity.

Understanding the pattern and mechanism of replication protein recruitment at $(GAA)_n$ repeats in a mammalian system can provide insight into mechanisms of microsatellite genomic instability in humans, and may lead to treatments for sufferers of FDRA and similar trinucleotide repeat disorders. It is unclear whether this knowledge can be transferred to other types of unstable microsatellites in the genome, such as mono- or dinucleotide repeats. However, previous work by our laboratory has shown that the degree of plasmid dinucleotide repeat instability is correlated with strength of replication stalling in a mammalian system, implying that secondary structures in the dinucleotide repeat play a similar role in microsatellite destabilization (2).

Chromatin immunoprecipitation (ChIP) and PCR were used to evaluate degrees of protein recruitment for MCM4, pol α/primase, and pol ζ in a plasmid model of FDRA transfected into mammalian cells. Quantitative PCR (qPCR) was also employed to quantify this protein enrichment. Our results suggest preferential recruitment of pol α and MCM to repeat-containing plasmids during the primary replication cycle.

Chapter 2

Materials & Methods

Plasmid preparation

Cloning of repeat-containing plasmids was carried out in *E. Coli* XL1 Blue strain (Stratagene).

Blunt-ended EcoRI-HindIII fragment from pBluescript-GAA_{57} was inserted into the blunt-ended Eco81I-site of pYES-Bsg plasmid. The repeat (underlined) was flanked by two BsgI sites in an inverted orientation (highlighted) and two EcoRI sites (bolded).

GAATTCTGCAGATATCCATCACACTGGCGGCCGCTCGAGTGCAGACCTCAAATTCG
AT(GAA)$_{57}$GATCAAGCTTCAGGTTCTGCACATCGAGCATGCATCTA**GAATTC**.

pUC19neoGAA_{57} and pUC19neoCTT_{57} plasmids were generated by inserting repeat-containing EcoRI-EcoRI fragments of pYES-CTT_{57} and pYES-GAA_{57} into the blunt-ended AatII site of pUC19neo plasmid.

Chromatin Immunoprecipitation (ChIP)

Cell cultures and transfections

Cos-1 cells were grown in Dulbecco's Modified Eagle Medium (DMEM) and 10% fetal bovine serum (FBS) to 50% confluence. For the analysis of protein recruitment during the

primary replication cycle, Cos-1 cells were transfected with pUC19neo, pUC19neoGAA_{57}, or pUC19neoCTT_{57} plasmid using Lipofecamine 2000 (Thermo Scientific) for 5 hours. After transfections, media was removed and 10 mL of DMEM supplemented with 10% FBS was added to cells. Cells were incubated for 1h prior to cross-linking. For the analysis of the protein recruitment in subsequent replication cycles, Cos-1 cells were lysed at 48 hours after transfection.

Cross-linking and chromatin preparation

Cells were incubated with 354 µL of 37% formaldehyde at room temperature for 10 min to cross-link proteins to DNA. 885 µL of 2M glycine was added to the media for 5 min at room temperature. Medium was removed and cells washed three times with cold 1X phosphate-buffered saline (PBS). The cells were placed on ice, then suspended in 1mL of Nuclei Washing Buffer (NWB) (5 mM $MgCl_2$, 10 mM Tris-Cl [pH 8.0], 0.32 M sucrose, 1mM PMSF) supplemented with 0.5% Triton X-100. Sample was pelleted at 4°C at 3000 rpm for 5 min. This pellet was resuspended in 1 mL of NWB and repelleted at 4°C at 3000 rpm for 5 min. The pellet was resuspended in 2 mL of IP buffer (0.1 M NaCl, 66.7 mM Tris-Cl [pH 8.0], 5 mM EDTA, 0.33% SDS, 1.67% Triton X-100, 1 mM PMSF). This sample was fragmented using ten 15-sec pulses of 50% amplitude with 30 s rest intervals (Fisher Scientific Sonic Dismembrator Model 500).

Chromatin pre-clearing and immunoprecipitation

The sample was centrifuged at 4°C at 14000 rpm for 5 min. Supernatant was pre-cleared using 60 µL of 50% slurry Protein A agarose/ Salmon Sperm DNA (Millipore) by rotating for 1h

at 4°C. Beads were pelleted at 4°C at 4000 rpm for 1 min and the supernatants were centrifuged at 4°C at 14000 rpm for 15 min.

1 µg of each antibody (DNA pol α rabbit polyclonal IgG [Santa Cruz Biotechnology], MCM 4 rabbit polyclonal IgG [Santa Cruz Biotechnology] and ORC2 rabbit polyclonal IgG [Santa Cruz Biotechnology]) was added to each 500 µL of pre-cleared supernatant. Another 500 µL of supernatant was preserved without the addition of antibodies. All samples were incubated overnight with rotation at 4 °C. The samples were then centrifuged at 4°C at 14000 rpm for 5 min. 30 µL of no-antibody sample was collected as "Total Input" control. 15 µL of 50% slurry Protein A agarose/salmon sperm DNA (Millipore) were added to each 500 µL sample except "Total Input." These samples were rotated at 4°C for 1h, then pelleted at 4000 rpm at 4°C for 1 min. Supernatants were discarded.

Washes

Pellets were washed with three 2 min washes in mixed micelle buffer (0.15 M NaCl, 20 mM Tris-Cl pH 8.0, 5 mM EDTA, 5.2% sucrose, 0.2% SDS, 1% Triton X-100, 1 mM PMSF), two 2 min washes in Buffer 500 (10 mM Tris-Cl pH 8.0, 50 mM HEPES, 0.5 M NaCl, 1% Triton X-100, 0.1% sodium deoxycholate, 1 mM EDTA, 1mM PMSF), two 2 min washes in LiCl/detergent buffer (10 mM Tris-Cl pH 8.0, 250 mM LiCl, 0.5% NP40, 0.5% sodium deoxycholate, 10 mM EDTA, 1mM PMSF,) and two 2 min washes in TE (10 mM Tris-Cl, pH 8.0, 1 mM EDTA, 1 mM PMSF). All washes were performed using 1 mL solution with gentle rotation at room temperature. Samples were repelleted at 4000 rpm at 4°C for 1 min between washes.

DNA elution and purification

After washes, all no-antibody controls and experimental samples were resuspended in 300 µL of SDS/bicarbonate buffer (0.1 M $NaHCO_3$, 1% SDS) plus 1 µL GADPH-2. Total Input controls were resuspended in 270 µL of SDS/bicarbonate buffer plus 1 µL GAPDH-2. All samples were vortexed at setting 3 (Vortex-Genie 2) for 15 min. Samples were spun at 4000 rpm for 1 min and the supernatants were collected. All samples were reverse cross-linked by incubating at 65°C overnight.

Each sample was then incubated with 3 µL RNase A at 37°C for 30 min, followed by incubation with 5 µL of Proteinase K for 1h at 55°C. DNA was then purified by two phenol/chloroform extractions and one chloroform extraction. DNA was precipitated with 1/10 volume 3 M NaOAc, 1 µL polyacryl carrier (MRC) and 3 volumes of 100% ethanol. Sample was pelleted at 14000 rpm for 5 min at 4°C, and supernatant was discarded. Pellets were dissolved in 300 µL of TE buffer (10 mM Tris-HCl ph 8.0, 1 mM EDTA). 30 µL 3M NaOAc was then added and the sample was vortexed. 900 µL of 100% ethanol was added and the sample was inverted 10 times to mix. The sample was incubated for 2 min at room temperature, then pelleted at 14000 rpm for 5 min at room temperature. The pellet was washed twice with 70% ethanol, then dried for 5 min before being dissolved in 300 µL of TE buffer.

Polymerase chain reaction (PCR) analysis

Immunoprecipitated DNA was amplified in a total volume of 25 µL with Maxima Hot Start TaqDNA Polymerase (Fermentas) based on the manufacturer's instructions. Primer

sequences are listed in **Table 2.1.** PCR was performed in a DeltaCycler II system (Ericomp). Samples were denatured by incubation at 95°C for 4 min, then subjected to 30 cycles of denaturation at 95°C for 30 s, annealing at 55°C for 30 s, and extension at 72°C for 30 s. Final extension was performed at 72°C for 5 min. Amplified samples were resolved by gel electrophoresis using 1.5% agarose gels (Sigma-Aldrich). All samples were run alongside FastRuler Low Range DNA Ladder (Thermo Scientific).

Quantitative PCR (qPCR) analysis

Standard curve generation

Primer sets 3, 4, 5, 5' and 3' (**Table 2.1**) were assessed for qPCR efficiency through the generation of standard curves. 1 ug each of purified pUC19neo, pUC19neoGAA$_{57}$ and pUC19neoCTT$_{57}$ plasmid DNA was linearized by digestion with BamHI. 1 uL polyacryl carrier, 2 uL 3M sodium acetate and 3 volumes of 100% ethanol were added to each digestion mixture, with vortexing between each addition. The samples were centrifuged at 14000 rpm for 5 minutes and supernatant removed. 300 uL 70% ethanol was then added to each pellet with vortexing. The samples were centrifuged at 14000 rpm for 5 minutes and supernatant removed. The pellets were dried with a SpeedVac Plus system (Thermo Savant) and resuspended in 30 uL TE buffer. The samples were heated to 65°C for 10 minutes to reverse any triplex formation.

These samples were then serial diluted in a 10-fold series, and amplified using PerfeCTa SYBR Green Supermix (Quantabio) by manufacturer's instruction on a Microamp Optical 96-well reaction plate (Applied Biosystems) using the StepOnePlus real-time PCR system (Applied Biosystems). Samples were denatured by incubation at 95°C for 10 min, then subjected to 40

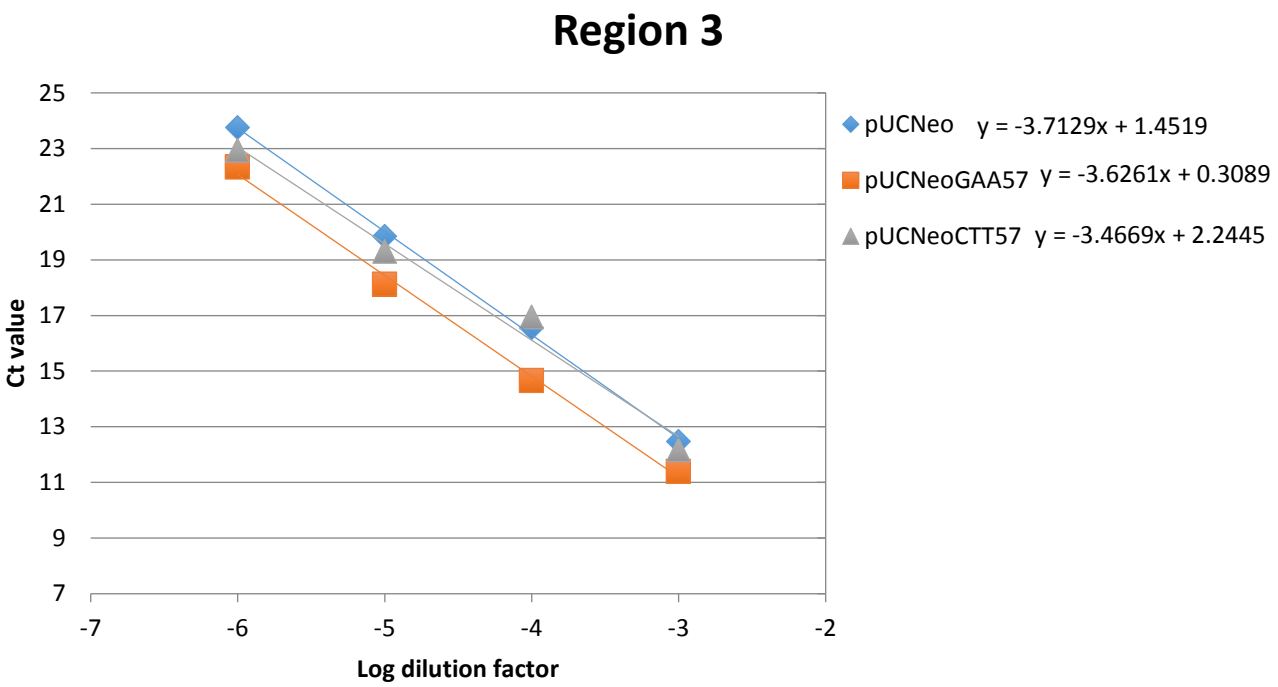
cycles of denaturation at 95°C for 30 s, annealing at 55°C for 30 s, and extension at 72°C for 30 s. Cycling was followed by automated melt curve generation. Resulting C_T values were plotted against log dilution factor (**Figure 2.1**). The data was given a linear fit. Efficiencies and amplification factors were calculated from trendline slopes using the online GE Dharmacon qPCR efficiency calculator (**Table 2.2**).

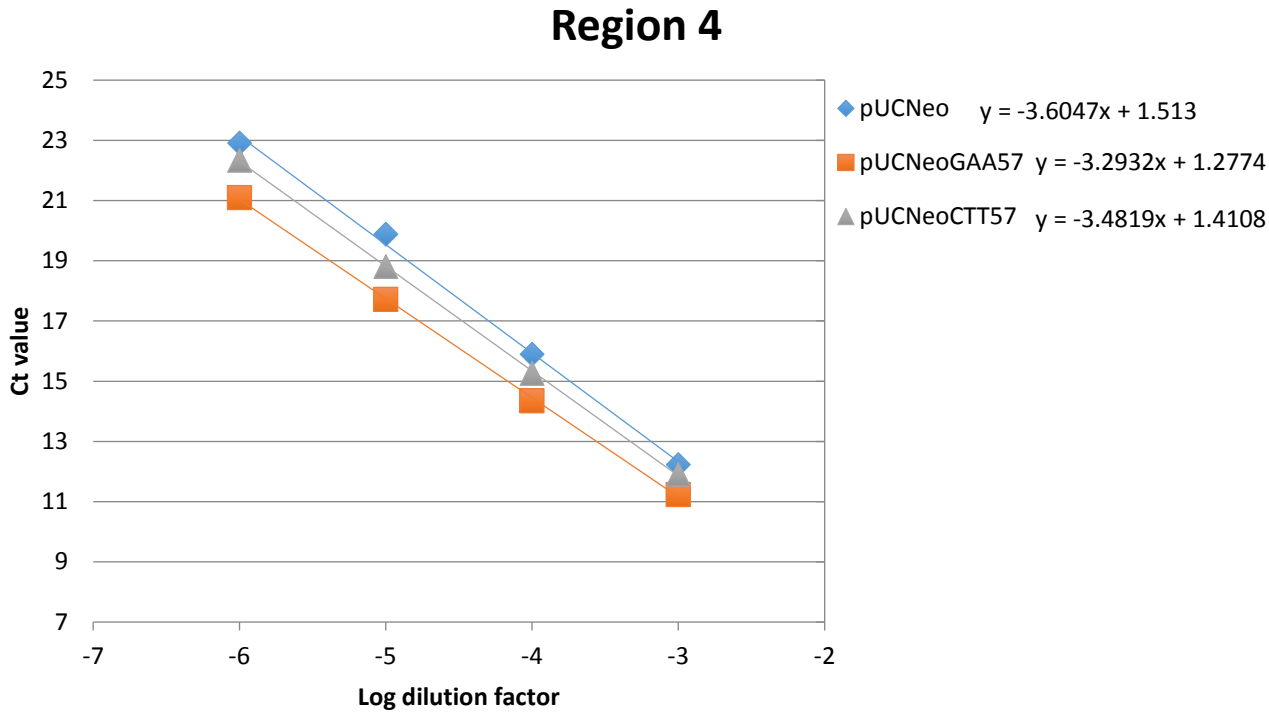
Experimental assays

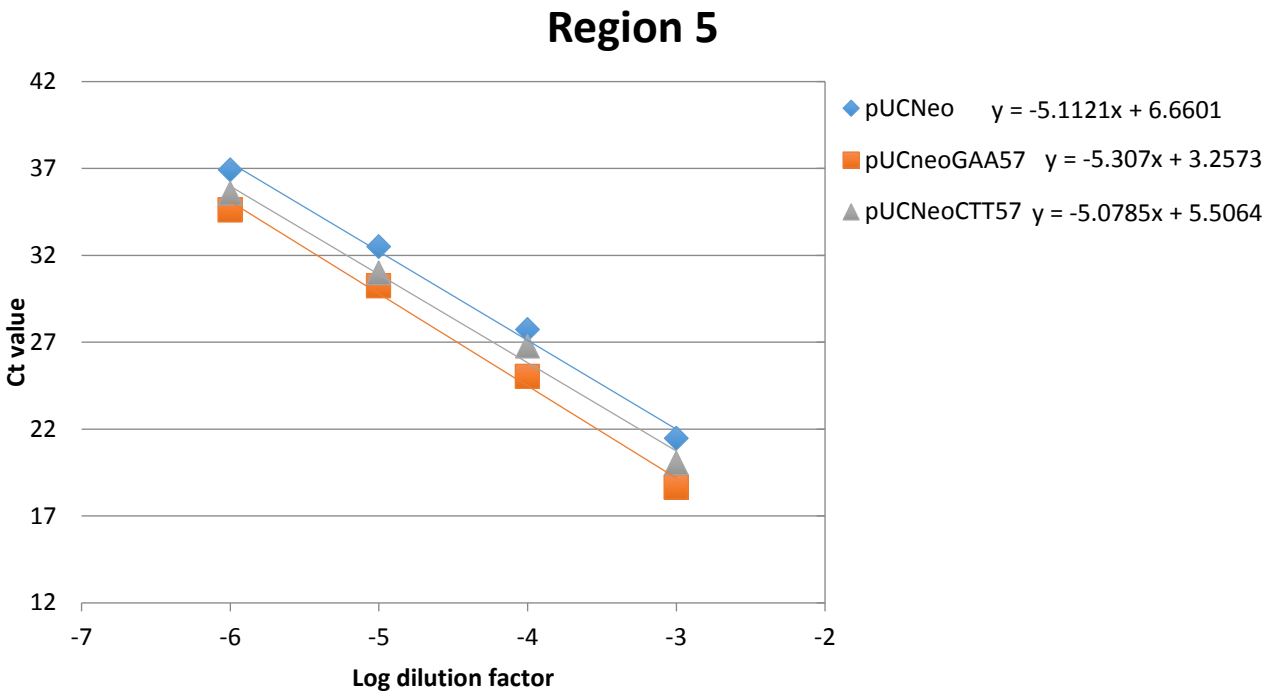
Immunoprecipitated DNA was amplified using PerfeCTa SYBR Green Supermix (Quantabio) by manufacturer's instruction on a Microamp Optical 96-well reaction plate (Applied Biosystems) using the StepOnePlus real-time PCR system (Applied Biosystems). Samples were denatured by incubation at 95°C denaturation for 10 min, then subjected to 40 cycles of denaturation at 95°C for 30 s, annealing at 55°C for 30 s, and extension at 72°C for 30 s. Cycling was followed by automated melt curve generation. C_T values were normalized by percent input method.

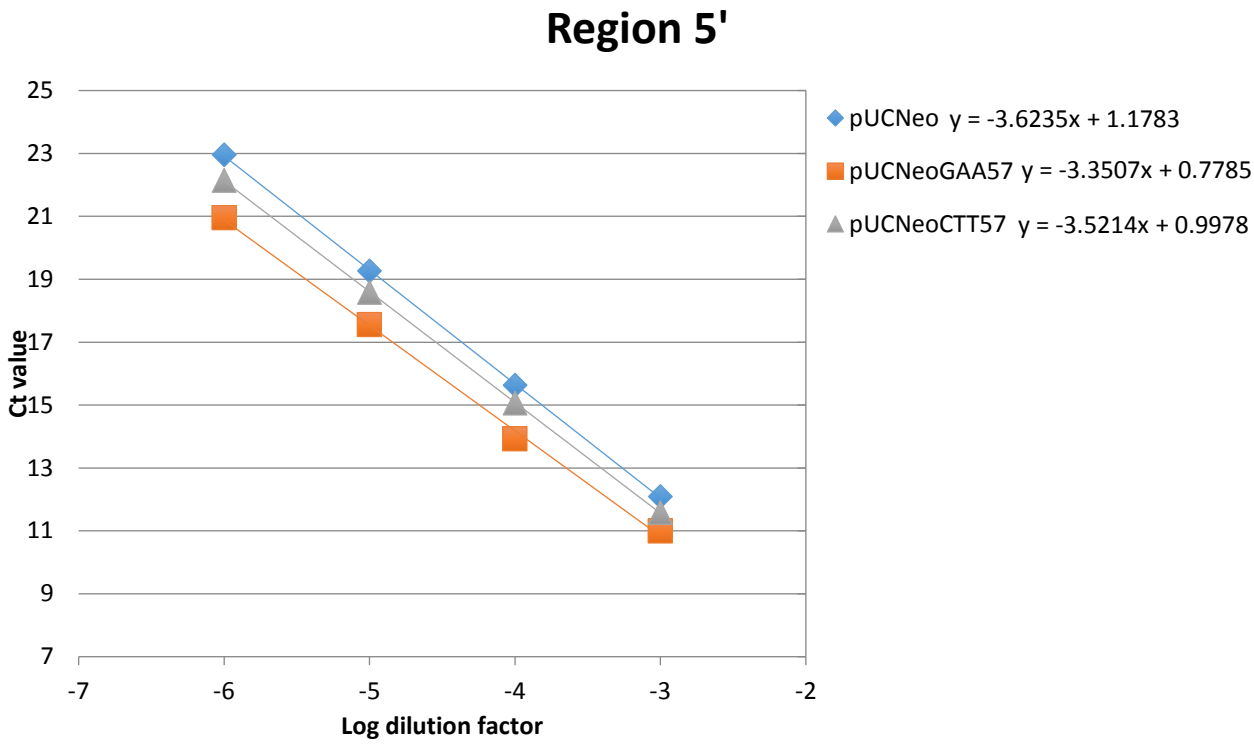
Region	Primer	Sequence	Melting Temperature (°C)
3	Forward	5’-CAC CGC CGC CTT CTA TGA-3’	57.6
	Reverse	5’-GGC GAA GAA CTC CAG CAT GA-3’	57.6
4	Forward	5’-AAT CAT AGG CTG CCC ATC CA-3’	56.4
	Reverse	5’-CCC TTA GAA AGC GGT CTG TGA-3’	56.7
5	Forward	5’-GTA GCT CTT GAT CCG GCA AA-3’	55.0
	Reverse	5’-CGT CAG ACC CCG TAG AAA AG-3’	55.1
5'	Forward	5’-AAG TTG GCC GCA GTG TTA TC-3’	55.8
	Reverse	5’-GCT ATG TGG CGC GGT ATT AT-3’	55.2
3'	Forward	5’-GTC GGG GCT GGC TTA ACT A-3’	56.9
	Reverse	5’-CTG GGG TAA TAG CGA AGA GG-3’	55.3
Repeat Region	Forward	5’-CAA ATA GGG GTT CCG CGC AC-3’	59.2
	Reverse	5’-CTC GTG ATA CGC CTA TTT TTA TAG G-3’	53.2

Table 2.1. Primers used for PCR/qPCR reactions. Repeat Region primers were used for regular PCR only.


Region 3
25
23
21
19
17
15
13
11
9
7
Ct value
-7
-6
-5
-4
-3
-2
Log dilution factor
pUCNeo y = -3.7129x + 1.4519
pUCNeoGAA57 y = -3.6261x + 0.3089
pUCNeoCTT57 y = -3.4669x + 2.2445


Region 4
25
23
21
19
17
15
13
11
9
7
Ct value
-7
-6
-5
-4
-3
-2
Log dilution factor
pUCNeo y = -3.6047x + 1.513
pUCNeoGAA57 y = -3.2932x + 1.2774
pUCNeoCTT57 y = -3.4819x + 1.4108


Region 5
42
37
32
27
22
17
12
Ct value
-7
-6
-5
-4
-3
-2
Log dilution factor
pUCNeo y = -5.1121x + 6.6601
pUCneoGAA57 y = -5.307x + 3.2573
pUCNeoCTT57 y = -5.0785x + 5.5064


Region 5'
25
23
21
19
17
15
13
11
9
7
Ct value
-7
-6
-5
-4
-3
-2
Log dilution factor
pUCNeo y = -3.6235x + 1.1783
pUCNeoGAA57 y = -3.3507x + 0.7785
pUCNeoCTT57 y = -3.5214x + 0.9978

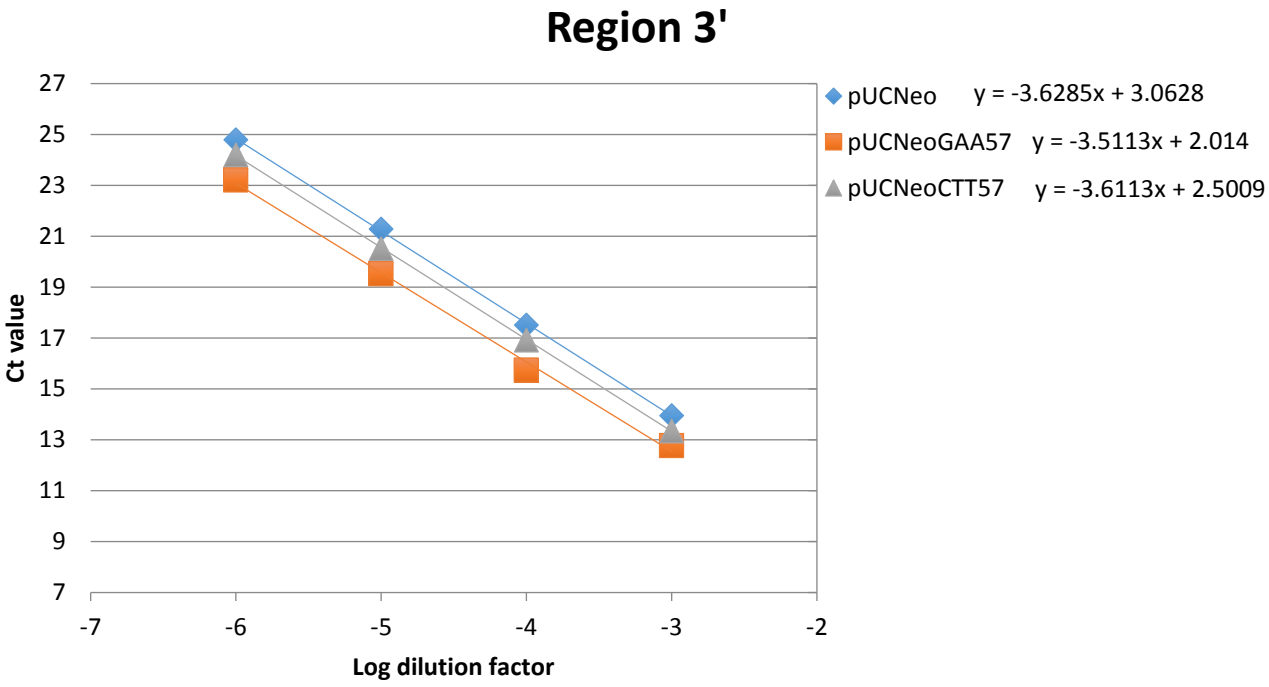


Figure 2.1. Standard curves for determination of primer set qPCR efficiencies.

Region	Plasmid	Slope	Efficiency (%)	Amplification Factor
3	pUC19neo	-3.7129	85.92	1.86
	pUC19neoGAA$_{57}$	-3.6261	88.7	1.89
	pUC19neoCTT$_{57}$	-3.4469	95.04	1.95
4	pUC19neo	-3.6047	89.42	1.89
	pUC19neoGAA$_{57}$	-3.2932	101.21	2.01
	pUC19neoCTT$_{57}$	-3.4819	93.73	1.94
5	pUC19neo	-5.1121	56.9	1.57
	pUC19neoGAA$_{57}$	-5.307	56.64	1.57
	pUC19neoCTT$_{57}$	-5.0785	57.37	1.57
5'	pUC19neo	-3.6235	88.79	1.89
	pUC19neoGAA$_{57}$	-3.3507	98.81	1.99
	pUC19neoCTT$_{57}$	-3.5214	92.3	1.92
3'	pUC19neo	-3.6285	88.62	1.89
	pUC19neoGAA$_{57}$	-3.5113	92.66	1.93
	pUC19neoCTT$_{57}$	-3.6113	89.19	1.89

Table 2.2. Calculated PCR efficiencies and amplification factors.

Chapter 3

Results

Qualitative assessment of replication protein recruitment by ChIP-PCR

Chromatin immunoprecipitation (ChIP) and PCR were together utilized to assess patterns of replication protein recruitment on plasmids containing a GAA_{57} repeat in either of two orientations (called pUC19neoGAA_{57} and pUC19neoCTT_{57}), as well as on an identical plasmid lacking this repeat (pUC19neo). Plasmids were transfected into mammalian cells and lysed after 6 and 48 hours to observe protein recruitment during primary and subsequent cycles of replication, respectively. Plasmids isolated after the shorter transfection time do not have fully-matured chromatin structure.

Significant PCR amplification at a specific region of immunoprecipitated DNA indicates a greater concentration of antibody-protein-DNA complex, thereby implying significant protein recruitment to that region. Results were compared to no-antibody (-) and Total Input (T) controls for each plasmid/region pair. The former accounts for the occasional non-specific binding of protein-DNA complexes to agarose beads, while the latter acts as a positive control, allowing one to confirm sufficient isolation of chromatin. **Figure 3.1** visualizes the ChIP workflow - a summary of the procedure described in *Materials & Methods*.

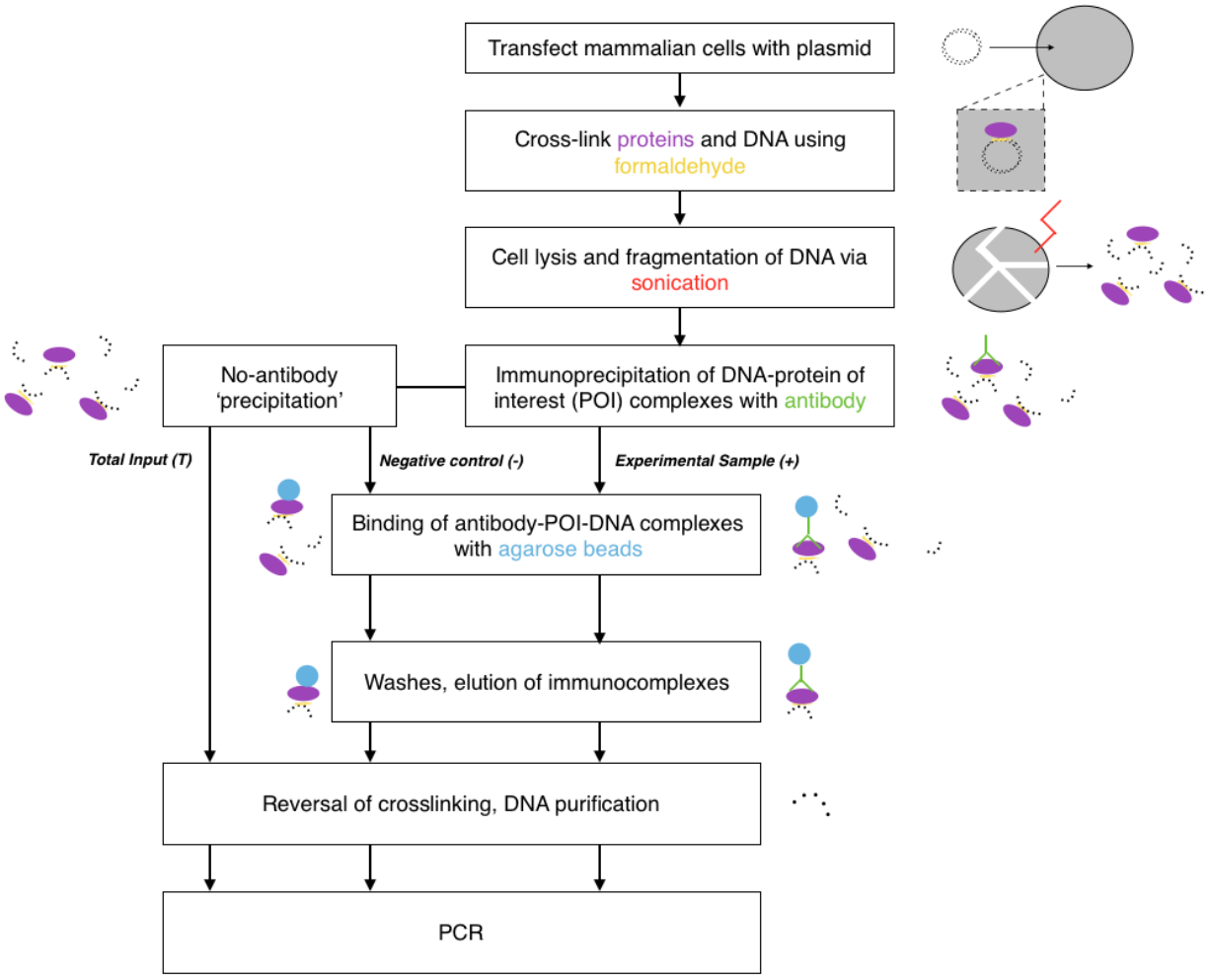


Figure 3.1. ChIP procedure workflow. Methods of generating Total Input (T) and negative (-) controls are also shown.

Protein recruitment was assessed at 6 distinct regions of each pUC19neo plasmid. Regions 3 (primers P3F, P3R) and 4 (P4F, P4R) represent generic regions of the plasmid. Region 5 (P5F, P5R) represents a segment of the bacterial pUC19neo origin of replication (at which no recruitment is to be expected after transfection into a mammalian host system). Regions 5' (5'F, 5'R) and 3' (3'F, 3'R) are so named as they represent the 5' and 3' ends of the GAA_{57}/CTT_{57}

repeat. Repeat Region (REPEAT-F, REPEAT-R) lies at the more immediate 3' end of the repeat.

Figure 3.2 maps all primer binding sites on each plasmid.

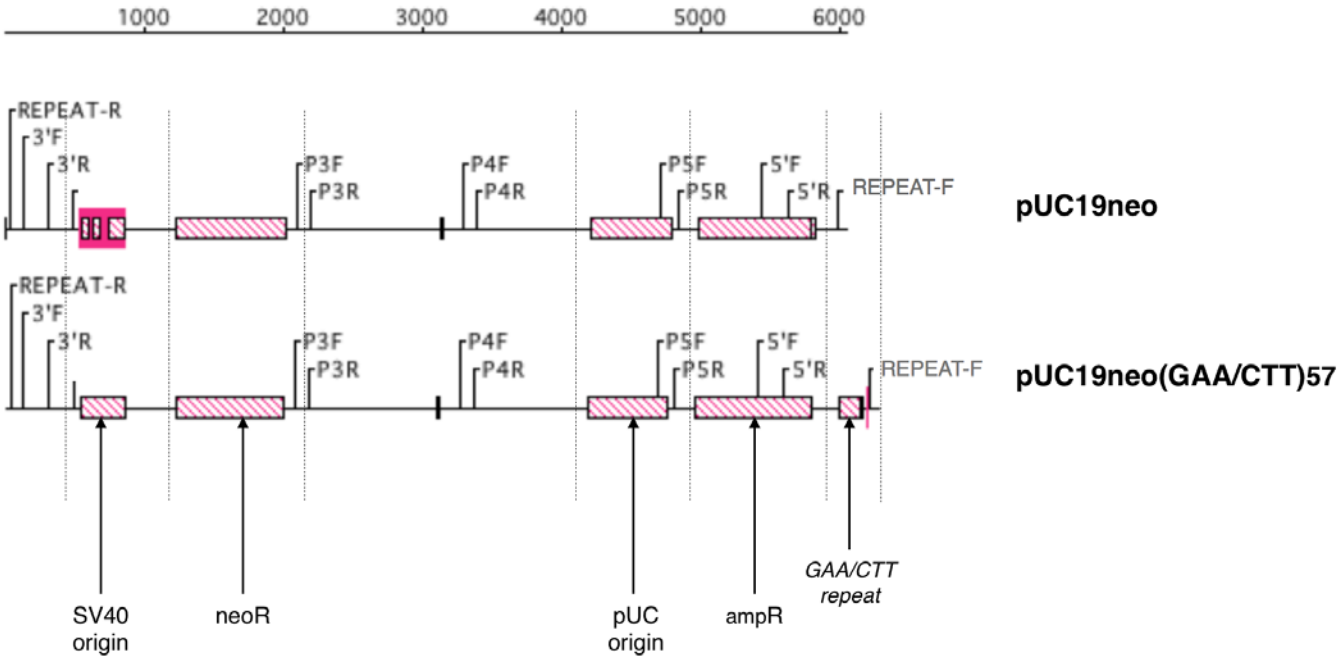


Figure 3.2. Map of primer binding sites on pUC19neo, pUC19neoGAA$_{57}$ and pUC19neoCTT$_{57}$ plasmids. Several key regions common among the plasmids are also identified. "ampR" and "neoR" indicate ampicillin and neomycin resistance genes, respectively.

First, protein recruitment during the primary replication cycle of pUC19neoCTT$_{57}$ was assessed.

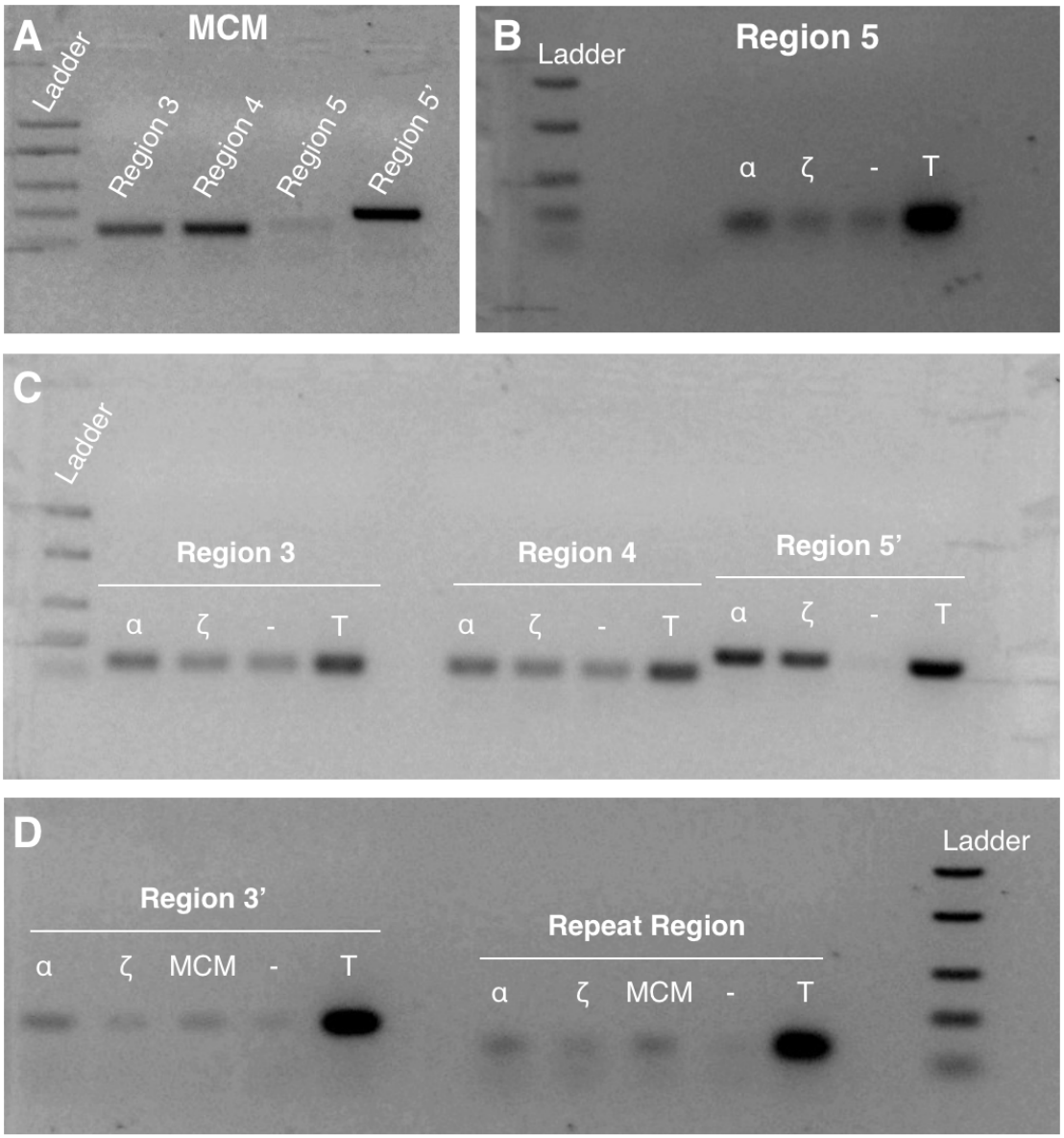


Figure 3.3. Resolved PCR products from ChIP of pUC19neoCTT$_{57}$ plasmid during the primary replication cycle. Samples from various regions using antibodies to MCM4 are shown separately (A) from those with antibodies to pols α and ζ, as well as controls (C). Recruitment of these proteins at Region 5 is shown in the top right (B). Recruitment of all proteins at Region 3' and the Repeat Region are shown at the bottom (D). Ladder bands represent, from top to bottom, 1500, 850, 400, 200, and 50 bp fragments.

All resulting bands of **Figure 3.3** show expected lengths between 200 and 50 bp and indicate sufficient primer specificity. The gels show an increase in pol α across the entire plasmid, particularly in region 5', which lies in the vicinity of $(CTT)_{57}$ repeat **(Figure 3.2, Figure 3.3**). MCM4 is also recruited at all regions of the plasmid except Region 5 **(Figure 3.3A, Figure 3.3D**). There appears to also be some recruitment of pol ζ at Region 5' (**Figure 3.3C**).

Analysis was repeated for products of immunoprecipitated pUC19neo and pUC19neoGAA_{57} plasmid DNA during the primary replication cycle for the regions of greatest interest: 5' and 3'. Assuming our ChIP protocol lead to sufficient chromatin isolation judging from the strong Total Input signals in **Figure 3.3**, Total Input controls for these additional ChIPs were not included on gels.

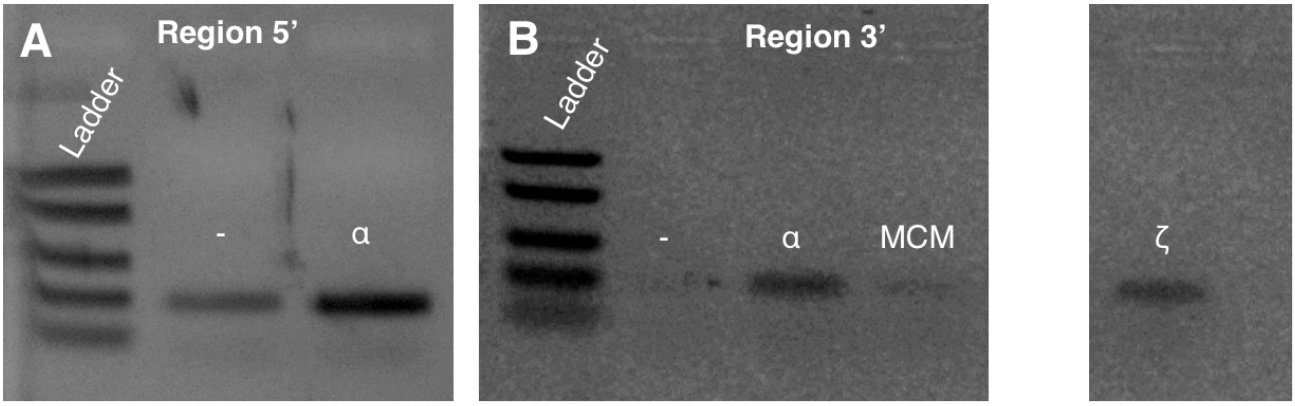


Figure 3.4. Resolved PCR products from ChIP of pUC19neoGAA_{57} plasmid during the primary replication cycle. Pol α recruitment at Region 5' is on the left (A), while total protein recruitment at Region 3' is on the right (B).

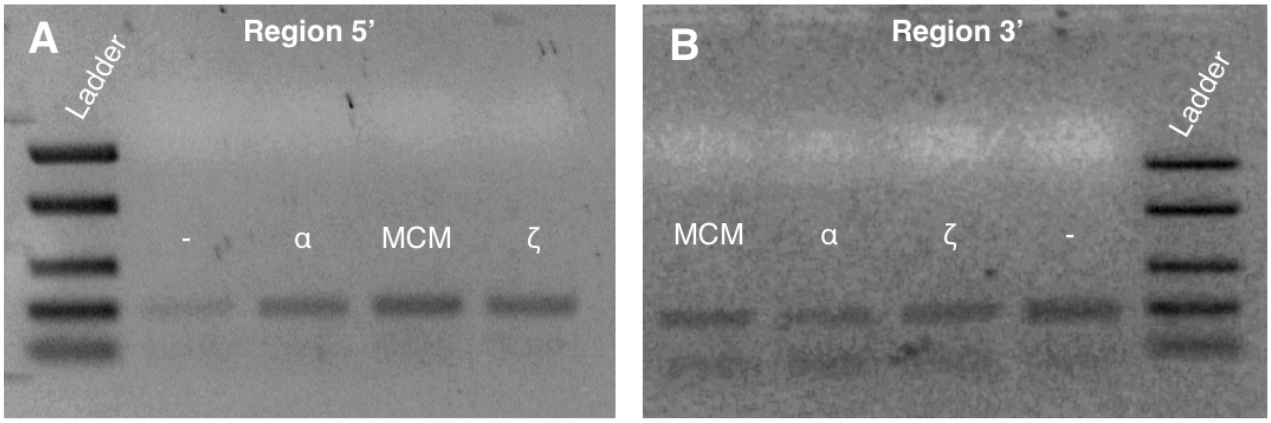


Figure 3.5. Resolved PCR products from ChIP of pUC19neo plasmid during the primary replication cycle. Region 5' is shown on the left (A) and Region 3' on the right (B).

Figure 3.5 shows recruitment of pol α and MCM at Region 5' in pUC19neo; however, the contrasts between no-antibody control (-) and sample are lesser than those seen for the repeat-containing plasmids (**Figure 3.3C, Figure 3.4A**). In addition, there is no observed Region 3' recruitment of MCM4 or α for pUC19neo, in contrast to the pUC19neoCTT$_{57}$ results (**Figure 3.3D**).

Finally, subsequent cycle ChIP (48 hr transfection time) results were briefly assessed for pUC19neo and pUC19neoGAA$_{57}$ plasmids at Region 3' only.

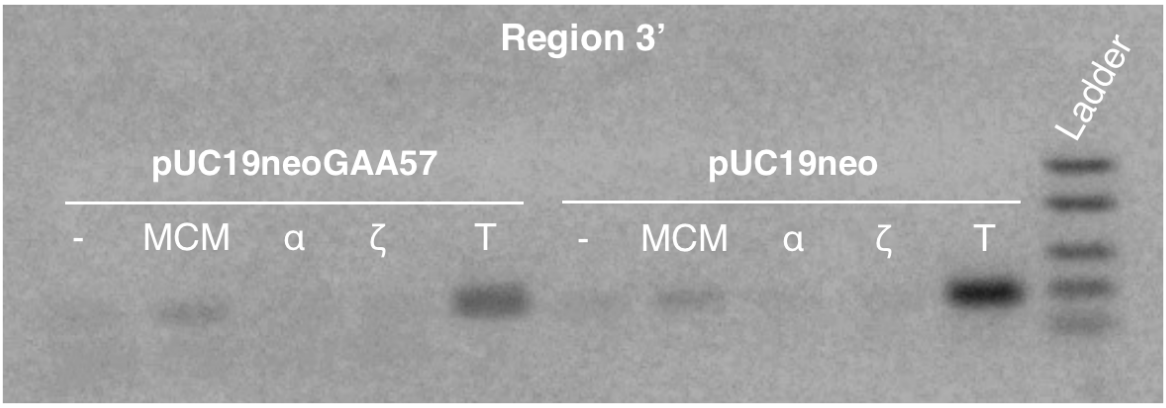


Figure 3.6. Resolved PCR products from ChIP of pUC19neo and pUC19neoGAA_{57} plasmids during subsequent replication cycles.

Figure 3.6 clearly shows an absence of the pol α recruitment on GAA_{57}-containing plasmid at Region 3' that was observed during the primary replication cycle **(Figure 3.4B**). However, mild MCM4 recruitment is observed in both plasmids during subsequent replication.

In summary, the qualitative data presented here show some recruitment of replication proteins across the majority of all three plasmids during the primary replication cycle, but the relative strength of recruitment for each protein and plasmid is difficult to assess due to varying levels of background signal and indirect comparisons across multiple gels. Some of this recruitment - particularly that of pol α - appears to dissipate during subsequent replication.

Quantitative assessment of replication protein recruitment by ChIP-qPCR

Following this PCR analysis of primary and subsequent cycle ChIPs, we saught to quantify the observed protein enrichment through qPCR. Data was normalized via the percent input (%IP) method, which uses comparison with Total Input controls (**Figure 3.1**) to determine

the percentage of starting chromatin isolated in each experimental sample. First, Total Input control C_T values are adjusted to 100% input by subtracting an appropriate number of cycles (for example, $\log_2[10] = 3.322$ cycles for controls using 10% starting chromatin). Then, %IP of any experimental sample (including no-antibody controls) is defined as 100 * [Amplification Factor]$^{(\text{adjusted Total Ct} - \text{experimental Ct})}$ (refer to **Table 2.2** for amplification factors).

The ChIP assay for each replication cycle was performed in duplicate (with few exceptions indicated by flat error bars). %IP values were averaged for each sample, and normalized to no-antibody controls for each respective region. Results for all three plasmids are presented, followed by graphs in which the two repeat-containing plasmids results are additionally normalized to control pUC19neo plasmid. The analysis was repeated for subsequent replication cycle data. Error bars of all graphs represent the range between the biological duplicates.

Repeat Region primers were excluded from all quantitative analyses due to the consistent appearance of multiple T_M peaks, leading to unreliable data. These peaks may be due to polymerase stalling at the repeat region, generating qPCR products of varying length.

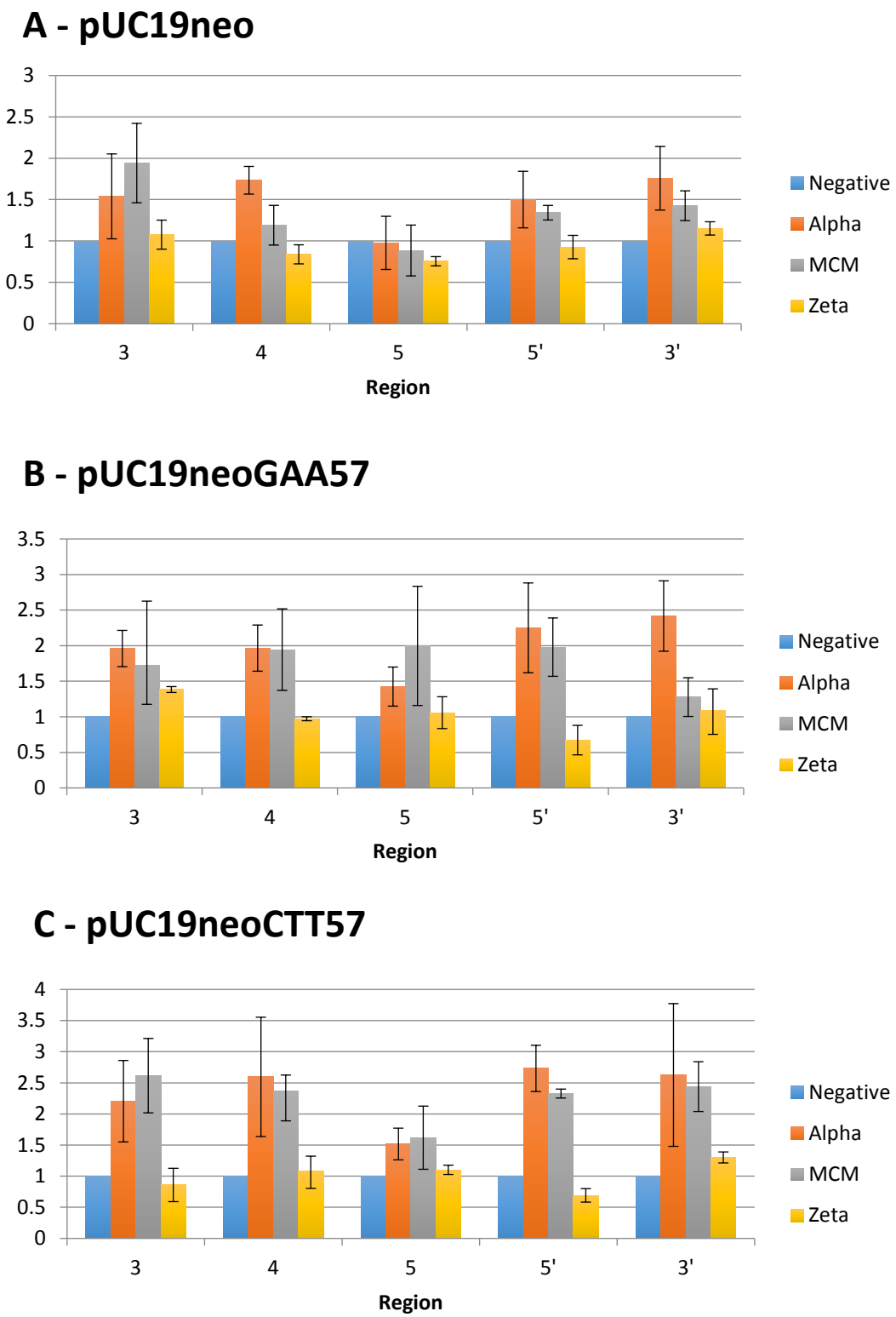


Figure 3.7. ChIP-qPCR results for pUC19neo (A), pUC19neoGAA$_{57}$ (B), and pUC19neoCTT$_{57}$ (C) plasmids during the primary replication cycle.

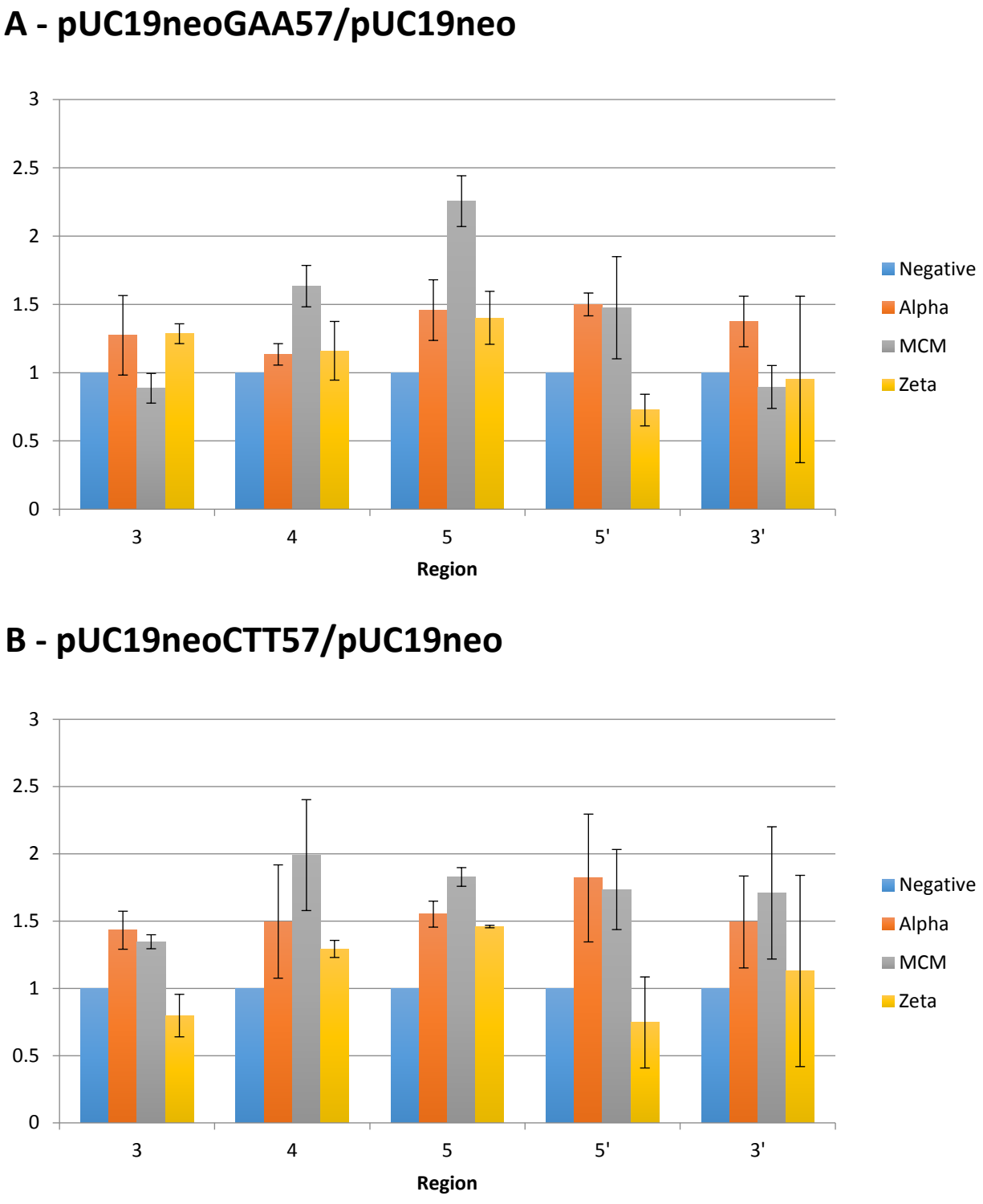


Figure 3.8. ChIP-qPCR results of pUC19neoGAA$_{57}$ (A) and pUC19neoCTT$_{57}$ (B) plasmids relative to pUC19neo during the primary replication cycle.

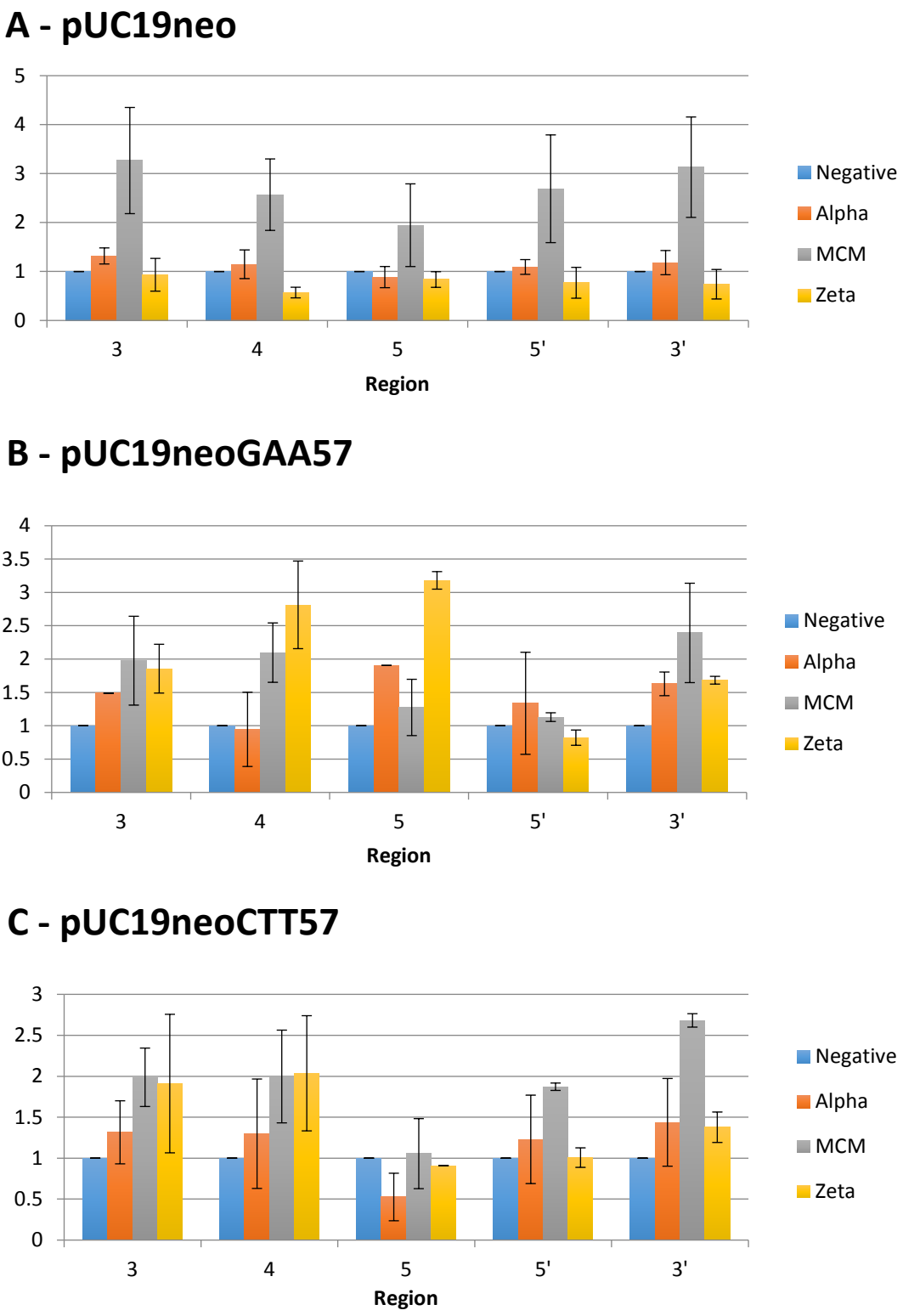


Figure 3.9. ChIP-qPCR results for pUC19neo (A), pUC19neoGAA$_{57}$ (B), and pUC19neoCTT$_{57}$ (C) plasmids during subsequent replication cycles.

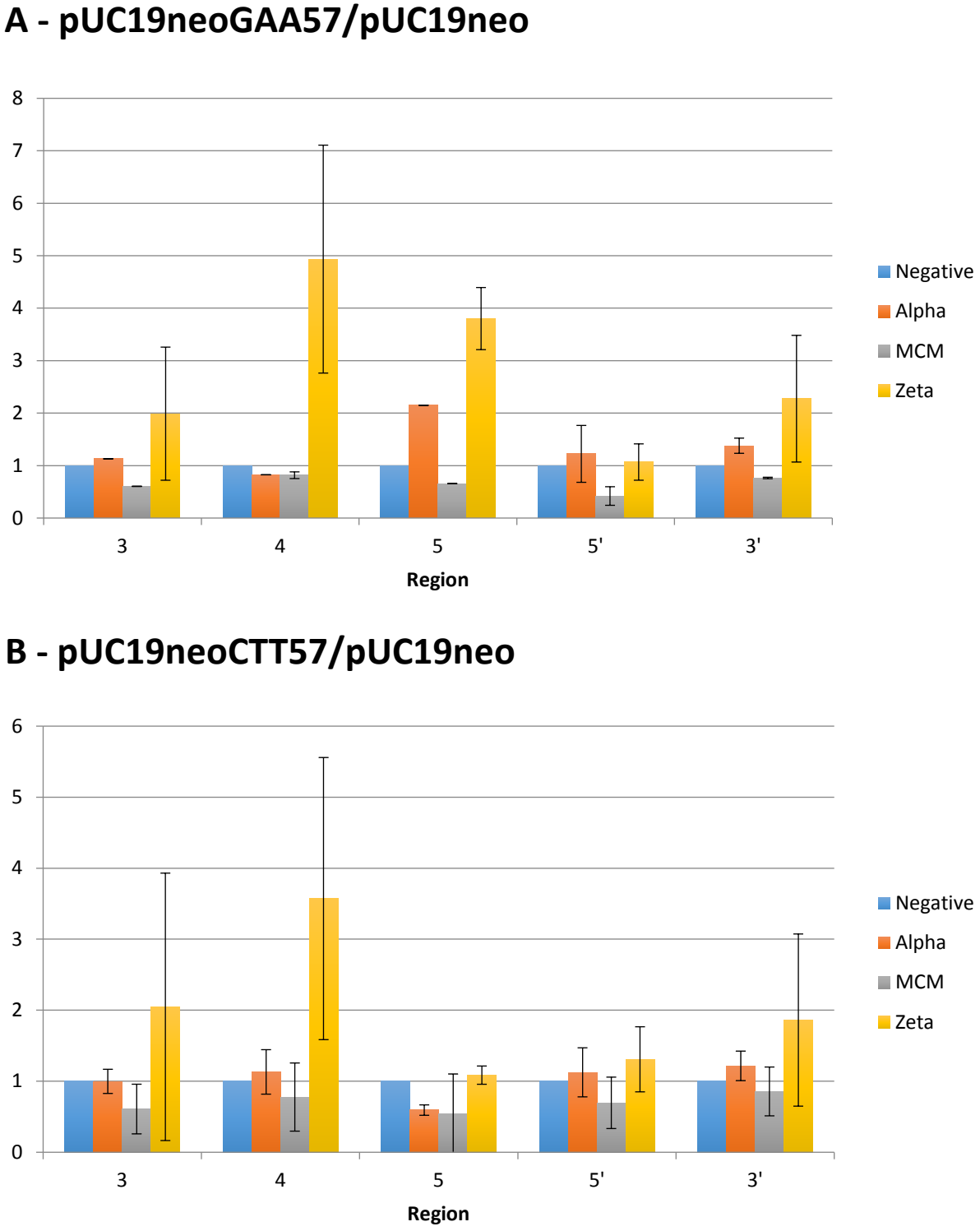


Figure 3.10. ChIP-qPCR results of pUC19neoGAA$_{57}$ (A) and pUC19neoCTT$_{57}$ (B) plasmids relative to pUC19neo during subsequent replication cycles.

Using this quantitative data we can more precisely analyze pol α and MCM4 enrichment across the pUC19neo(GAA/CTT)$_{57}$ and pUC19neo plasmids. **Figure 3.7** reveals consistent recruitment of polymerase α and MCM4 proteins (1.5 to 2.8-fold increases) over no-antibody control across all regions of all plasmids, including control plasmid (excluding Region 5). In contrast, pol ζ shows low to moderate recruitment (<1.4-fold increase) and is inconsistent across the three plasmids.

The enrichment of protein recruitment specifically associated with microsatellite repeat can be seen directly by dividing pUC19neo(GAA/CTT)$_{57}$ %IP results by those of pUC19neo control plasmid. For (GAA)$_{57}$-containing plasmid, pol α shows an average 1.35-fold enrichment over pUC19neo and no-antibody controls across the regions, and MCM4 shows 1.5 to 2.25 fold increases in Regions 4, 5 and 5' (**Figure 3.8A**). Pol ζ recruitment is enhanced minimally through the plasmid (~1.25-fold in Regions 3, 4 and 5). For the opposite orientation of repeat, pol α and MCM4 are recruited significantly over pUC19neo and no-antibody controls in all regions (average fold changes of 1.5 and 1.7 for pol α and MCM, respectively), with some additional pol ζ recruitment at Regions 4 and 5 (~1.25-fold increase) (**Figure 3.8B**). It should be noted that the pol ζ data shows larger error bars (that is, variation between biological replicates) than other analyzed proteins, particularly in Region 3'.

These data verify that the presence of (GAA/CTT)$_{57}$ repeat is correlated with increased pol α and MCM4 recruitment across plasmid, to an average of 1.5-fold over plasmid without the repeat. Moreover, these proteins appear to have a preference for repeat that is in the (CTT)$_{57}$ orientation, evinced by higher and more consistent fold changes in pUC19neoCTT$_{57}$ than in pUC19neoGAA$_{57}$. Pol ζ recruitment is present in both repeat-containing plasmids, but to a lesser degree and less consistently.

Quantitative data for subsequent cycle protein recruitment reveals similar general recruitment of pol α and MCM4 at all regions of all plasmids (with the exception of pol α on pUC19neo control plasmid, which falls near background levels) (**Figure 3.9**). Normalization of these results to control pUC19neo plasmid, however, shows that pol α and MCM4 hold little-to-no preference for repeat-containing plasmids during subsequent DNA replication (**Figure 3.10**). Pol α is recruited an average 1.1-fold over controls, excluding a 2-fold change in Region 5. However, this outlying data point was obtained from only a single biological replicate. MCM signals consistently fall below background levels in both repeat-containing plasmids. Interestingly, these plasmids tend to draw much more pol ζ than pUC19neo control during subsequent cycles of replication, though large error between biological replicates is again observed for this protein. These data suggest that the pol α and MCM recruitment shown in the primary replication cycle is specific to that cycle and is not maintained as the plasmid develops a more mature chromatin structure.

Chapter 4

Discussion

In summary, both qualitative and quantitative ChIP analysis of primary and subsequent replication cycles in our plasmids suggests enrichment of several replication proteins to plasmids containing $(GAA/CTT)_{57}$ repeat in the primary replication mode. For this mode, ChIP-PCR revealed signals in pol α and MCM immunoprecipitated samples that exceeded the no-antibody control at nearly all analyzed regions of all three plasmids (**Figure 3.3**, **Figure 3.4**, **Figure 3.5**). This included our control plasmid not containing microsatellite repeat. However, qPCR analysis revealed that the fold changes of these samples over no-antibody controls were higher in repeat-containing plasmids than in control plasmid (**Figure 3.8**). Pol α recruitment diminished on gel for both pUC19neo and pUC19neoGAA_{57} in Region 3' during subsequent replication (**Figure 3.6**). Similar, plasmid-wide losses of pol α enrichment - well as an equalization of MCM enrichment between repeat-containing plasmids and control – were largely confirmed by our qPCR data for subsequent replication (**Figure 3.10**).

Previous work on the topic of replication protein recruitment in the vicinity of microsatellite regions has already been presented by another Schreyer Scholar within our group. The results of this previous study showed higher enrichment of pol α and MCM4 (over no-antibody control) in Region 3 for pUC19neo and pUC19neoGAA_{57} plasmids, and the same in Region 4 for all 3 plasmids. pUC19neoGAA_{57} plasmid showed enriched pol α and MCM4 in Region 5', with pUC19neoCTT_{57} and pUC19neo plasmids showing enrichment in only MCM4 or α, respectively (31). Analyses of Region 3' and Repeat Region - as well as assessment of pol ζ recruitment - were not included.

In the current work, heightened recruitment of pol α and MCM4 at Regions 3 and 4 was again observed in pUC19neoCTT$_{57}$ on gel. Similar recruitment could be expected in the other two plasmids, though this gel data is not available. Additionally, pol α recruitment was again seen in Region 5' for pUC19neoGAA$_{57}$ on gel (**Figure 3.4A**). However, it should be noted that this same Region 5' pol α recruitment was also observed in pUC19neoCTT$_{57}$, at odds with the older results (**Figure 3.3C**). Repetition of the experiment using qPCR analysis supports the newer result, showing a 2.75-fold increase of pol α at Region 5' of (CTT)$_{57}$-containing plasmid over no-antibody control (**Figure 3.7C**).

Such a discrepancy highlights the deeply subjective nature of ChIP-PCR gel interpretation. The method of regular PCR analysis of immunoprecipitated samples is also limited by indirect comparison between experimental and control conditions – in this case, the comparison between pUC19neo(GAA/CTT)$_{57}$ plasmids and the control pUC19neo plasmid. The current work expands on the analysis of microsatellite replication protein recruitment through its use of quantitative PCR (qPCR), which allows for the calculation of exact fold-change differences over control conditions, normalizing to both pUC19neo plasmid and no-antibody samples, thereby yielding more reliable, definative data.

The quantitative data obtained for both repeat-containing plasmids in the primary replication cycle are generally in keeping with available qualitative results, with some exceptions. On gel, the no-antibody background of pUC19neo at Region 3' appeared higher than any immunoprecipitated sample, yet qPCR shows significant increases in pol α and MCM over control (**Figure 3.5, Figure 3.7A**). It is known that background signal can sometimes vary between no-antibody and immunoprecipitated samples due to differences in handling, and that this effect can also vary across biological replicates (20). Such minor error is a likely

explanation, as two biological replicates agree on these large increases in opposition to the single replicate used for qualitative analysis.

In addition, no large quantitative increases in pol ζ signal over background were observed for Region 5' of pUC19neo and pUC19neoCTT_{57} plasmids, as might have been suggested by the strong bands seen for these regions on gel (**Figure 3.3**, **Figure 3.5**, **Figure 3.7A/C**). The inconsistancy in pol ζ recruitment specifically is sensible, since pol ζ is the only protein analyzed which is not absolutely required for replication initation at all sites of plasmid. Instead, pol ζ is canonically recruited more-or-less randomly at sites of DNA lesion (including mismatches, single-stranded breaks, among other forms of damage) (19). Such sites could easily vary between biological replicates. Since two separate biological sets of ChIP samples were used for qualitative versus quantitative analysis, separate patterns of recruitment may therefore be expected for this polymerase. This variance between biological replicates may also explain the large error bars observed for many quantitative pol ζ data points, particularly in the graphs of **Figure 3.10**. Due to such inconsistencies, we conclude that pol ζ is likely not enriched on repeat-containing plasmids during primary replication in any biologically-relevant manner.

In contrast, any variance seen in pol α and MCM recruitment data between biological duplicates is insufficient to affect their interpretation. The fold changes of repeat-containing plasmids' recruitment of pol α and MCM over pUC19neo and no-antibody controls show an ~1.3-fold and 1.6-fold increase in pol α for GAA_{57} and CTT_{57}-containing plasmids, respectively, and ~1.4-fold and 1.7-fold increases for MCM4 (**Figure 3.8**). Though advanced statistical tests may be necessary to prove the significance of these changes, we demonstrate that the effects are robust and reproducible across biological replicates. The heightened presence of pol α and MCM4 is observed in nearly all regions of plasmid, not limited to regions nearby the

microsatellite repeat or SV40 origin. This is to be expected, as after their initial recruitment, MCM4 is constantly helping unwind DNA at progressing replication forks as part of a helicase complex, and pol α is continually extending from new RNA primers (adding ~20 nucleotides before replacement by a more processive polymerase) in lagging strand synthesis throughout the plasmid.

Our results demonstrate that the presence of microsatellite repeat is correlated with enhanced recruitment of replication proteins, particularly pol α and MCM, supporting our hypothesis that non-canonical DNA structures of microsatellite regions could lead to unscheduled replication initiation. However, additional study will be needed to transform this observed correlation into a casuality. One future experiment of interest may be to observe protein recruitment on pUC19neo$(GAA/CTT)_{57}$ plasmids that have been treated to minimize triplex formation - possibly using the same alkali conditioning and Nt.BstNBI nicking protocol used previously by our group to establish reduced replication fork stalling in the absence of secondary structure (10).

The difference in results between pUC19neoGAA_{57} and pUC19neoCTT_{57} plasmid is also noteworthy. Pol α and MCM proteins are generally recruited to a somewhat higher degree on the plasmid with repeat in the $(CTT)_{57}$ orientation – that is, with the homopyrimidine sequence in the leading strand position and the homopurine GAA sequence in the lagging position. It has previously been shown that replication stalling within FRDA $(GAA/CTT)_n$ repeat is much more pronounced in this orientation, due to more stable triplex formation (22). Such orientation-dependant stability has also been observed for other trinucleotide repeats, such as CTG/CAG and CGG/CCG (1, 17). This further supports our hypothesis that stable triplex formation leads to replication protein recruitment during the primary replication cycle.

Moreover, the fact that the preferential recruitment is mostly limited to the primary replication cycle discredits mechanisms of recruitment which are not reliant upon secondary structure formation permitted by underdeveloped chromatin. It makes sense for proteins such as histones to sterically hinder secondary structure formation; indeed, single-stranded binding proteins (SSBPs) are designed to inhibit such structures after parental DNA strands are separated in the process of DNA replication. Interestingly, trinucleotide repeat expansions have long been speculated to have an epigenetic regulatory component; in mice, deficiency of DNA methyltransferase Dnmt1 caused by siRNA knockdown promotes expansion of CAG repeat at the spinocerebellar ataxia type 1 (*Sca1*) locus (13).

It has also been established that mature episomal plasmids like pUC19neo are covered with histones in mammalian cells, making our plasmid model sufficient to simulate typical genomic chromatin development (11). However, a dependence of non-canonical DNA structure formation upon chromatin development throughout stages of DNA replication has not yet been directly observed. This may present another interesting new line of investigation which can complement this work. Other future directions will include expanding our ChIP-qPCR assays to include other replication proteins, including more processive polymerases like delta (δ) and epsilon (ε). Establishing the presence or absence of these proteins at repeat-containing plasmids will further illuminate the mechanisms of alternative primary cycle DNA replication at microsatellite regions.

Our model of genomic instability generated by the unstable primary replication mode aligns well with what is known about the timing of trinucleotide repeat expansions. Evidence suggests that such expansions tend to occur in the earliest stages of embryogenesis (26). Some individuals with Fragile X Syndrome, for instance, have two major *FMR1* CGG repeat lengths in

all cells throughout all tissues, suggesting that instability originated in the first zygotic cell division (14). In addition, we have previously demonstrated that the primary replication mode is particularly prone to fork stalling at repetitive regions, showing that it cannot proceed through the A/T-rich microsatellite region FRA16B in pUCneo plasmid (9). This makes microsatellites (such as trinucleotide repeats) the main target of instability caused by the primary replication cycle (though it is also possible for misregulation of primary replication to contribute to genome-wide destabilization through re-replication of DNA, similarly to promiscuous origin firing in G1 phase of replication as it contributes to tumorigenesis) (21). Therefore, a better understanding of the primary replication mode and why it shows particular susceptibility to fork stalling may well become critical to the treatment of genomic disorders like Fragile X Syndrome or FDRA.

The presented work demonstrates that certain replication proteins - particularly pol α and MCM4 - are recruited more strongly to microsatellite-containing plasmids during the alternative primary cycle of mammalian DNA replication previously described by our group. However, it remains unknown whether or not this recruitment is truly mediated by non-canonical DNA structures formed within microsatellites. Our limited list of replication proteins assessed also does not yet allow for a complete understanding of this how this primary mode of replication proceeds when stimulated by repetitive genomic elements. Ongoing work will answer such questions, and may have keen relevance to the development of therapies intended to treat genetic disorders like FDRA.

BIBLIOGRAPHY

1. Balakumaran, B. S., Freudenreich, C. H., & Zakian, V. A. (2000). CGG/CCG repeats exhibit orientation-dependent instability and orientation-independent fragility in Saccharomyces cerevisiae. *Human molecular genetics*, *9*(1), 93-100.
2. Baptiste, B. A., Ananda, G., Strubczewski, N., Lutzkanin, A., Khoo, S. J., Srikanth, A., … Eckert, K. A. (2013). Mature Microsatellites: Mechanisms Underlying Dinucleotide Microsatellite Mutational Biases in Human Cells. *G3: Genes|Genomes|Genetics*, *3*(3), 451–463.
3. Belotserkovskii, B. P., De Silva, E., Tornaletti, S., Wang, G., Vasquez, K. M., & Hanawalt, P. C. (2007). A triplex-forming sequence from the human c-MYC promoter interferes with DNA transcription. *Journal of Biological Chemistry*, *282*(44), 32433-32441.
4. Bemark, M., Khamlichi, A. A., Davies, S. L., & Neuberger, M. S. (2000). Disruption of mouse polymerase zeta (Rev3) leads to embryonic lethality and impairs blastocyst development in vitro. *Current Biology*, *10*(19), 1213–1216.
5. Bétous, R., Rey, L., Wang, G., Pillaire, M. J., Puget, N., Selves, J., ... & Hoffmann, J. S. (2009). Role of TLS DNA polymerases eta and kappa in processing naturally occurring structured DNA in human cells. *Molecular carcinogenesis*, *48*(4), 369-378.
6. Biet, E., Sun, J. S., & Dutreix, M. (1999). Conserved sequence preference in DNA binding among recombination proteins: an effect of ssDNA secondary structure. *Nucleic Acids Research*, *27*(2), 596-600.

7. Bradley, J. L. (2000). Clinical, biochemical and molecular genetic correlations in Friedreich's ataxia. *Human Molecular Genetics*, *9*(2), 275–282.
8. Brinkmann, B., Klintschar, M., Neuhuber, F., Hühne, J., & Rolf, B. (1998). Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *American Journal of Human Genetics*, *62*, 1408–1415.
9. Chandok, G. S., Kapoor, K. K., Brick, R. M., Sidorova, J. M., & Krasilnikova, M. M. (2011). A distinct first replication cycle of DNA introduced in mammalian cells. *Nucleic Acids Research*, *39*(6), 2103–2115.
10. Chandok, G. S., Patel, M. P., Mirkin, S. M., & Krasilnikova, M. M. (2012). Effects of Friedreich's ataxia GAA repeats on DNA replication in mammalian cells. *Nucleic Acids Research*, *40*(9), 3964–3974.
11. Chen, Z. Y., Riu, E., He, C. Y., Xu, H., & Kay, M. A. (2008). Silencing of episomal transgene expression in liver by plasmid bacterial backbone DNA is independent of CpG methylation. *Molecular Therapy*, *16*(3), 548-556.
12. Choi, J., & Majima, T. (2011). Conformational changes of non-B DNA. *Chemical Society Reviews*, *40*(12), 5893-5909.
13. Dion, V., Lin, Y., Hubert, L., Waterland, R. A., & Wilson, J. H. (2008). Dnmt1 deficiency promotes CAG repeat expansion in the mouse germline. *Human Molecular Genetics*, *17*(9), 1306–1317.
14. Dion, V., & Wilson, J. H. (2009). Instability and chromatin structure of expanded trinucleotide repeats. *Trends in Genetics*, *25*(7), 288–297.
15. Edwards, Y. J., Elgar, G., Clark, M. S., & Bishop, M. J. (1998). The identification and characterization of microsatellites in the compact genome of the Japanese pufferfish,

Fugu rubripes: perspectives in functional and comparative genomic analyses. *Journal of molecular biology*, *278*(4), 843-854.

16. Espinás, M. L., Jiménez-garcía, E., Chem, J. B., Jime, E., & Azorı, F. (1996). Molecular Genetics: Formation of Triple-stranded DNA at d(GA · TC)n Sequences Prevents Nucleosome Assembly and Is Hindered by Nucleosomes, *271*(50), 31807–31812.
17. Freudenreich, C. H., Stavenhagen, J. B., & Zakian, V. A. (1997). Stability of a CTG/CAG trinucleotide repeat in yeast is dependent on its orientation in the genome. *Molecular and cellular biology*, *17*(4), 2090-2098.
18. Gacy, A. M., Goellner, G. M., Spiro, C., Chen, X., Gupta, G., Bradbury, E. M., … McMurray, C. T. (1998). GAA instability in Friedreich's Ataxia shares a common, DNA-directed and intraallelic mechanism with other trinucleotide diseases. *Molecular Cell*, *1*(4), 583–93.
19. Gan, G. N., Wittschieben, J. P., Wittschieben, B. Ø., & Wood, R. D. (2008). DNA polymerase zeta (pol zeta) in higher eukaryotes. *Cell Research*, *18*(1), 174–183.
20. Haring, M., Offermann, S., Danker, T., Horst, I., Peterhansel, C., & Stam, M. (2007). Chromatin immunoprecipitation: optimization, quantitative analysis and data normalization. *Plant Methods*, *3*, 11.
21. Hook, S., Lin, J. and Dutta, A. (2007) Mechanisms to control rereplication and implications for cancer. Curr. Opin. Cell Biol., 19, 663–671.
22. Krasilnikova, M. M., & Mirkin, S. M. (2004). Replication Stalling at Friedreich's Ataxia (GAA)n Repeats In Vivo. *Mol. Cell. Biol.*, *24*(6), 2286–2295.

23. Li, Y., Korol, A., Fahima, T., Beiles, A., & Nevo, E. (2002). Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Molecular Ecology*, (11), 2453–2465.

24. Marquis Gacy, A., Goellner, G., Juranić, N., Macura, S., & McMurray, C. T. (1995). Trinucleotide repeats that expand in human disease form hairpin structures in vitro. *Cell*, *81*(4), 533–540.

25. Mirkin, S. M. (2006). DNA structures, repeat expansions and human hereditary disorders. *Current Opinion in Structural Biology*, *16*(3), 351–358.

26. Mirkin, S. M. (2007). Expandable DNA repeats and human disease. *Nature*, *447*(7147), 932-940.

27. Murphy, T. D., & Karpen, G. H. (1995). Localization of centromere function in a drosophila minichromosome. *Cell*, *82*(4), 599–609.

28. Musso, M., Nelson, L. D., & Van Dyke, M. W. (1998). Characterization of purine-motif triplex DNA-binding proteins in HeLa extracts. *Biochemistry*, *37*(9), 3086–3095.

29. Petruska, J., Arnheim, N., & Goodman, M. F. (1996). Stability of intrastrand hairpin structures formed by the CAG/CTG class of DNA triplet repeats associated with neurological diseases. *Nucleic Acids Research*, *24*(11), 1992–1998.

30. Subramanian, S., Mishra, R. K., & Singh, L. (2003). Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions. *Genome Biology*, *4*(2), R13.

31. Teo, Y. V. (2014). Unscheduled Replication at Microsatellite Repeats. Retrieved from ETDA.

32. Wittschieben, J. P., Reshmi, S. C., Gollin, S. M., & Wood, R. D. (2006). Loss of DNA polymerase ζ causes chromosomal instability in mammalian cells. *Cancer research*, *66*(1), 134-142.

33. Zannis-Hadjopoulos, M., Yahyaoui, W., & Callejo, M. (2008). 14-3-3 cruciform-binding proteins as regulators of eukaryotic DNA replication. *Trends in biochemical sciences*, *33*(1), 44-50.

Academic Vita

REBECCA DEGIOISO

425 Waupelani Drive, Apt 25• State College, PA 16801
Phone: 570-847-0902 • E-Mail: beccaanne95@gmail.com

EDUCATION

B.S. **Pennsylvania State University,** State College, PA 2017

Biochemistry

Schreyer Honors Scholar

EXPERIENCE

- **Pennsylvania State University, Department of Biochemistry and Molecular Biology**, State College, PA 2014-2017

  Research Assistant

  PI: Maria Krasilnikova
  - Investigated replication protein recruitment at microsatellite repeats in the human genome, contributions to genomic instability

  PI: Gong Chen
  - Tested small molecule reprogramming methods to convert human glial cells into GABAergic neurons, using immunostaining, RT-PCR

- **Pennsylvania State University, Department of Biochemistry and Molecular Biology**, State College, PA 2016

  Teaching Assistant, *BISC004 "Human Body: Form and Function"*
  - Managed gradebooks, scored exams, and answered student questions through one-on-one appointments, office hours and review sessions

- **Sanofi Pasteur (pay-rolled through PRO Unlimited),** Swiftwater, PA 2015

  Undergraduate Summer Intern
  - Collaborated with the Analytical Processes and Technology (AP&T) department to conduct pre-qualification method development studies for a new sterility testing method

KEY SKILLS

PCR, qPCR, RT-PCR

Cell culture

Chromatin immunoprecipitation (ChIP)

Immunostaining

DNA cloning

Familiarity with NCBI databases, BLAST

PAPERS & PRESENTATIONS

- **"Characterization of Replication Protein Recruitment at GAA Microsatellite Repeat in a Mammalian Primary Replication Cycle"** 2017
  - Thesis paper for Honors in Biochemistry
- **"Replication at Non-Canonical DNA Structures"** 2017
  - Presented at Penn State 2017 Undergraduate Poster Exhibition