Determining the mechanism of large-scale CAG repeat expansion

An honors thesis for the Department of Biology

Teresa Dinter

Tufts University, 2016

Acknowledgements

I would like to thank Dr. Sergei Mirkin for giving me the exciting opportunity to perform research in his lab. His belief in my capabilities and support has been very influential in my development as a scientist. I would also like to thank Dr. Jane Kim, whose mentorship has been defining in my decision to pursue science. Her meticulous methodology has taught me to perform careful research and the importance of always asking questions. I'd also like to thank Dr. Mitch McVey not only for advising my thesis, but also providing guidance on academics and careers.

I am also grateful for the Mirkin lab members – Artem Kononenko, Elina Radchenko, Jessica Wilks, Alex Neil, Ekaterina Spivakovsky-Gonzalez, Ryan McGinty, Tri Hoang, Miranda Liang, and Rachel Rubenstein – for their constant encouragement and feedback throughout the year.

Lastly, I am thankful for the funding I received through Summer Scholars, which allowed me to pursue this research over summer 2015 and transform it into a senior thesis.

Table of Contents

Abstract

Nearly 30 genetic diseases are caused by the repetition of a short sequence of 2-9 base pairs, known as microsatellites. These microsatellite repeats occur in humans normally; however disease symptoms become apparent once a certain repeat threshold is reached. Usually, these diseases follow a phenomenon known as genetic anticipation, in which each subsequent generation experiences increased disease severity and earlier onset, which was found to correlate with microsatellite repeat length. To study these expansions, budding yeast (*S. cerevisiae*) strains have been designed to carry repeats within or near selectable genes. On selective media, cells with expansions will form colonies, allowing for their detection. Here, proteins were investigated by two approaches to determine their effect on large-scale CAG repeat expansion. First, a candidate gene approach was used to test two Holliday junction resolvases, Mus81 and Yen1, because homologous recombination had previously been shown to be involved in large-scale CAG repeat expansion. Second, in an unbiased genetic screen, Los1 had two independent hits as a gene protecting against CAG expansion. It was of particular interest because of its role at the nuclear pore in tRNA export and a recent study had found that expanded CAG repeats were repaired at the nuclear pore. To characterize the roles of these proteins, knockout strains were created and fluctuation assays were performed to calculate CAG expansion rates. The double *mus81Δyen1Δ* strain showed a decrease in the expansion rate, suggesting that these proteins are required for CAG expansion to occur. The Los1 results suggest that *los1Δ* increased canavanine resistance, but does not affect repeat expansion directly.

Introduction

Microsatellite Repeats and Human Disease

Microsatellites, repeated sequences of 2-9 nucleotides, are found in various regions of the human genome. These microsatellites can be comprised of a variety of base pair combinations and are found in all parts of a gene: coding sequences, the 3' and 5'-untranslated regions, introns, and untranscribed regions. The length of these microsatellites can vary between individuals and still present a normal phenotype. In certain genes, if the microsatellite is expanded beyond a threshold limit, disease symptoms will appear. Nearly 30 genetic diseases have been identified that are caused by expanded microsatellites.[1] Examples include Huntington's disease, myotonic dystrophy, Friedrich's Ataxia, fragile X syndrome, and familial ALS. These diseases are often neurodegenerative and are severely debilitating.

CAG•CTG repeats, which are studied in this project, are known to cause 12 of these diseases.[1] One example is Huntington's disease, which is caused by expanded CAG repeats within the coding region of the Huntingtin gene.[2] The wild type gene has between 6 and 34 CAG repeats. At 36 repeats, the expanded CAG repeats translate into polyglutamine tracts that cause the proteins to aggregate. This gain-of-function mutation manifests itself as chorea, cognitive disorders, and psychiatric problems. Usually within 10-15 years of onset, death occurs. Myotonic dystrophy type I is also caused by expanded CTG•CAG repeats and is the most common muscular dystrophy, affecting about 1 in 8000 people.[2] In this disease, the expanded CTG•CAG repeat is found in the 3'-UTR, and expanded repeat lengths can range from 50 to 10,000 repeats. The gain-of-function mutation that occurs here is at the RNA level. It is believed that RNA transcripts accumulate in the nucleus and prevent the

regulation of alternative splicing. Disease symptoms include myotonia, cardiac problems, and insulin resistance.

Since these diseases don't appear until after reproductive age, they are often passed from one generation to the next. Generally, each successive generation has earlier onset and increased severity of the disease, a phenomenon known as genetic anticipation. This was found to be correlated with repeat length.[3] Thus, longer repeats indicate earlier onset and more acute symptoms. Additionally, it has been shown that the rate of expansion increases with the length of the repetitive run.[4] Therefore, the length of the repeat is crucial because it indicates the severity of the disease and the likelihood of expansion.

Currently, it is known that expansions occur to transform the wild type allele to a disease allele and that expansions occur between generations to increase disease severity. However, the mechanisms of these expansions and how they differ depending on various factors is still not well understood. Understanding the mechanisms behind expansions would be beneficial in creating a deeper understanding of how cellular machinery interacts with repetitive DNA sequences, and it would lead to a better understanding of microsatellite diseases and their progression.

Current Models of CAG Repeat Expansion, With a Focus on Holliday Junction Resolvases

Current models of repeat expansion focus on DNA metabolic processes such as replication, transcription, and recombination.[1] It is important to note that different models of expansion exist for different types of microsatellites, due to microsatellites' ability to form secondary structures. For example, CAG repeats form hairpin loops or cruciform structures, and GAA repeats can form G quadruplexes or H-DNA. These secondary

structures are thought to be determining factors in microsatellite expansion, and thus it is expected that GAA and CAG repeats expand through different mechanisms. Current models for CAG expansion are explored further here.

Replication based models for small-scale CAG expansion focus on the formation of 5' flaps or 3' slippage.[5] 5' flaps can be formed when DNA polymerase displaces the end of an Okazaki fragment, causing it to fold into a hairpin (Figure 1A). In the 3' slippage model, DNA polymerase "slips" on the repetitive tract so that a hairpin is formed, and then it resynthesizes part of the repetitive run (Figure 1B). If these hairpins are not correctly repaired, they would become expansions in the next replication cycle, adding a hairpin-length of DNA to the repetitive tract. Large-scale expansions may be caused by fork reversal, in which DNA polymerase stalls due to some barrier, such as a secondary structure (Figure 1C). The replication fork may then reverse so that a chicken foot structure is formed. In order to restart replication, the repetitive tract can fold back on itself, causing a longer repeat fragment to be synthesized in the nascent strand. In another replication-based model, the leading strand may switch its template to use the newly synthesized lagging strand as a template, thus increasing the number of repetitive runs created (Figure 1D). Once the end of the fragment has been reached, the repetitive DNA can fold back on itself and DNA polymerase can continue synthesis of the leading strand.

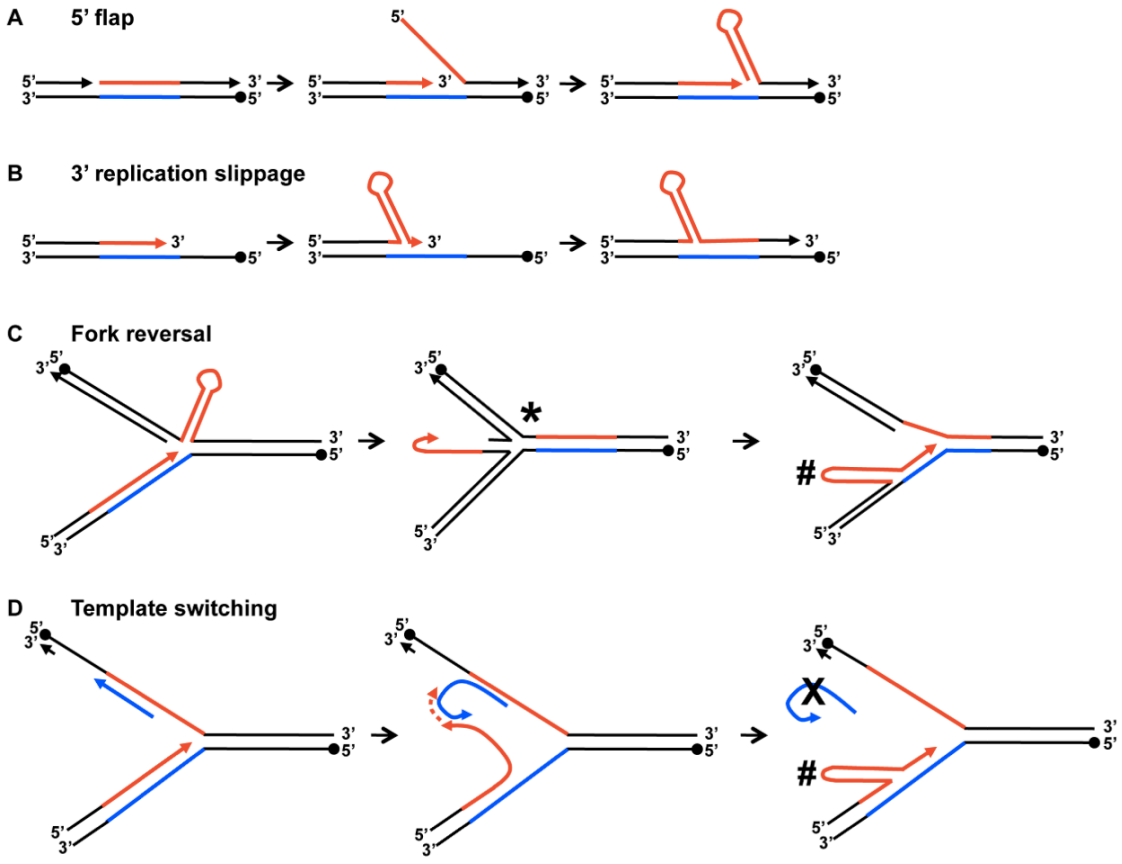

Figure 1. Replication-based models of repeat expansion, adapted from Dr. Kim.[5] 5' flap formation, 3' replication slippage, fork reversal, and template switching are shown (a-d).

However, these replication-based models do not explain all expansions; even cells that are non-dividing, and therefore not duplicating their genomes, have been shown to experience repeat expansions.[6] Repair mechanisms, such as homologous recombination (HR), could explain expansions in non-dividing cells. HR-based models are especially applicable for CAG expansion because CAG repeats are known to be fragile sites;[7] thus it is reasonable to expect that HR may be acting on double strand breaks near or within CAG repeats. When a double strand break occurs, broken strands are first resected to create a 3' overhang which can invade a homologous region on a homologous chromosome, sister chromatid, or ectopic region (Figure 2A). To initiate HR in yeast, Replication Protein A (RPA) coats the single-stranded regions of the break site.[8] Rad52 then displaces the RPA

and aids in Rad51 binding to the single-stranded DNA (ssDNA). Rad51 then leads the ssDNA invasion of a region of homology, causing displacement of a DNA strand, which Rad52 binds[9]. Recently, Dr. Jane Kim found that Rad51 and Rad52 are involved in causing large-scale expansions (Kim, unpublished), providing convincing evidence that HR may be promoting large-scale CAG repeat expansion. After synthesizing new DNA, double Holliday junctions can be formed. Resolvases are responsible for cutting DNA strands and reannealing them to create repaired, separated double helixes (Figure 2G). The resolution of the Holliday junction is a key step of HR.

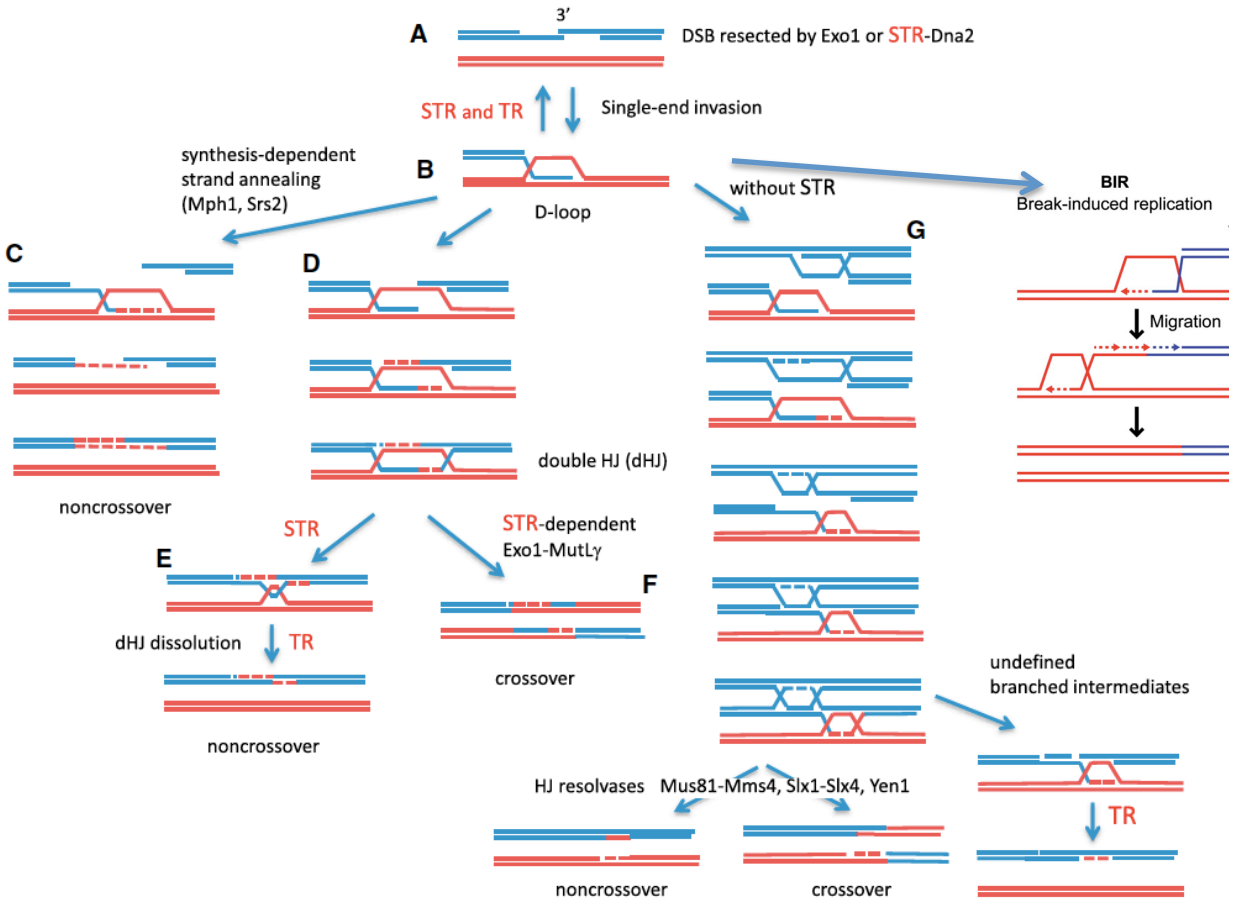

Figure 2. Models for double strand break repair, adapted from Dr. Haber[10] and Dr. Symington.[11]

Mus81 and Yen1 are resolvases in *S. cerevisiae*, budding yeast, with homologues in human cells.[11] They repair Holliday junctions through exchange of DNA fragments, resulting in both noncrossover and crossover products.[12] Since both Mus81 and Yen1 serve the same function, Matos and West have shown evidence that Mus81 and Yen1 are activated during different stages of the cell cycle.[12] According to their data, Yen1 is most active during anaphase and G1, while Mus81 is most active in prometaphase and metaphase. Though these two proteins play an important role in repair of Holliday junctions, other repair mechanisms, such as Holliday junction dissolution (Figure 1E) and synthesis-dependent strand annealing (Figure 1C), may be favorable in mitotic cells in order to prevent formation of crossover products. By determining which resolvase, if any, is involved in expansion of microsatellites, a mechanism of expansion could be described.

Genetic Screen

Though many expansion models have been proposed, it is possible that the current models do not encompass all ways in which repeat expansion occurs or all genes involved in the process. To determine which processes may be involved and which have been overlooked, genetic screens are used. In an unbiased genetic screen, mutations are made randomly in the genome and mutated cells are screened for a particular phenotype. Mutants that show the desired phenotype can be sequenced to determine which gene was involved in creating that phenotype.

An unbiased genetic screen for CAG repeat expansion was previously performed by undergraduates Samantha Harris, Emma Marshal, and Tanner Byer. In this screen, the *CAN1* cassette was used, which has $(CAG)_{140}$ repeats between the *CAN1* promoter and its

upstream activating sequence (Figure 3). In this cassette, colonies without an expansion are not able to grow on the selective galactose-canavanine media. However, colonies with expansions are able to grow on galactose-canavanine media. Thus, canavanine resistance acts as an indicator of overall expansion rate for strains with this cassette.

The goal of this screen was to select for mutants that had increased canavanine resistance, in hopes of finding genes that protect from repeat expansion. To perform the screen, Leu$^-$ cells containing the *CAN1* cassette were mutated using transposons carrying the *LEU2* gene. Transposons are mobile DNA elements that can integrate themselves into DNA and inactivate genes if integrated into the gene's open reading frame. Cells were plated on synthetic complete minus leucine (SC-Leu) media to determine if the transposon had been integrated into the genome. Colonies with integration were plated on galactose-canavanine plates to select for the desired phenotype, increased growth on canavanine. Of colonies that showed increased growth on canavanine, an inverse PCR was performed and sequenced to determine into which gene the transposon had been integrated and was responsible for the desired phenotype.

This transposon screen resulted in many hits that showed increased resistance to canavanine and presumably higher CAG expansion rate, one of which was Los1. Los1 was interesting for three reasons. Firstly, it had two independent hits in the genetic screen. Secondly, Los1 is a protein active at the nuclear pore, involved in tRNA export from the nucleus. This was interesting because it was recently discovered that repair of expanded CAG repeats was localized to the nuclear pore.[13] Lastly, Los1 is able to recognize the looped secondary structure of tRNAs, and thus it was hypothesized that it may be able to recognize other nucleic acid secondary structures, like the hairpins formed by CAG repeats.

Goals of this Project

The goal of this project was to explore the mechanism of CAG repeat expansion and to determine the roles of specific proteins in this process. Mus81 and Yen1 are Holliday junction resolvases, identified by the candidate gene approach. Los1 was identified through an unbiased genetic screen. Knockouts of these proteins were generated and their effect on CAG repeat expansion was characterized using fluctuation assays to calculate the rate of CAG repeat expansion. The expansion rate could then be compared to the wild type strain to indicate whether the protein was required for repeat expansion or prevented repeat expansion.

Methods

Yeast Experimental System

Two gene cassettes in an experimental *S. cerevisiae* (budding yeast) system were used in this project.

The *CAN1* cassette is an experimental yeast system designed to detect large-scale CAG repeat expansions when grown on galactose and canavanine-containing media. In this cassette, $(CAG)_{140}$ repeats are located between a galactose-induced upstream activating sequence (UAS_{gal}) and the promoter of the *CAN1* gene (Figure 3). The distance created by the $(CAG)_{140}$ repetitive tract, between the promoter and UAS_{gal}, allows for transcription and translation of the gene. Can1 is an arginine permease, which allows canavanine, an arginine analog, into the cell. Canavanine is toxic to the cell. It becomes integrated in proteins in place of arginine, thereby causing protein misfolding and cell death. Therefore, all cells containing this cassette and $(CAG)_{140}$ repeats will be unable to grow on galactose and canavanine containing media. If an expansion of the CAG repeats occurs, the distance between the UAS_{gal} and promoter becomes too great and transcription can no longer occur. Thus, cells containing an expansion will be able to form colonies on the selective galactose-canavanine media. This system allows for the selection of cells that have experienced a CAG expansion event.

Another analogue of this cassette was used, in which $(GAA)_{100}$ repeats, instead of $(CAG)_{140}$ repeats, were place between the UAS_{gal} and the promoter. This cassette acts in the same manner, where cells without expansions are sensitive to the galactose and canavanine containing media, but cells with expansions are able to grow.

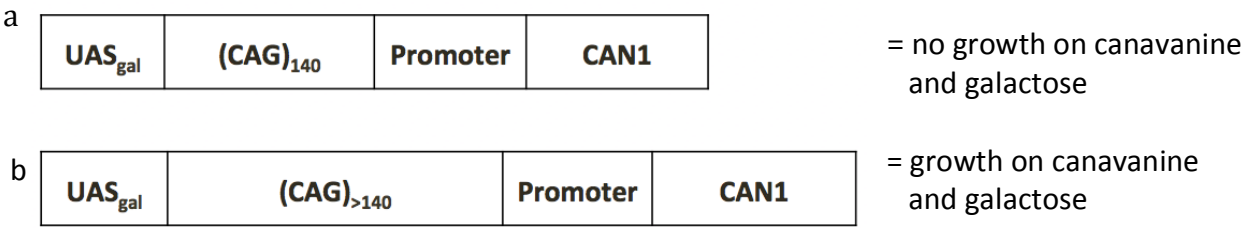

Figure 3. Schematic of the *CAN1* cassette. CAG repeats are shown between the UASgal and promoter. a) No growth will be detected on canavanine and galactose media. b) If an expansion occurs, growth will be detected on canavanine and galactose containing media.

The *URA3* cassette was used to detect large-scale GAA repeat expansions. In this cassette $(GAA)_{100}$ repeats are located within the intron of the *URA3* gene (Figure 4). This repeat length allows for successful splicing of the mRNA to occur. Ura3 can turn 5-fluoroorotic acid (5-FOA) into 5-fluorouracil, a substance toxic to the cell. Thus, any cells with the *URA3* cassette will not be able to grow on media containing 5-FOA. If an expansion occurs, the intron becomes too large to be spliced. Thus, cells with expansions are able to form colonies on 5-FOA media, allowing for the selection of GAA expansion events.

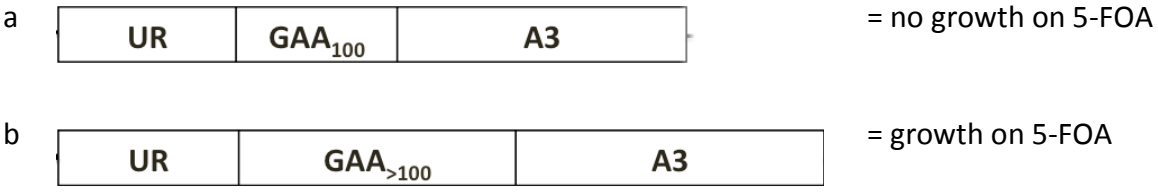

Figure 4. Schematic of the *URA3* cassette. GAA repeats are shown in the intron of the *URA3* gene. *a)* No growth will be detected on 5-FOA media. *b)* If an expansion occurs growth will be detected on 5-FOA media.

Generation of Knock Outs

All knockouts were created from strains containing either the *URA3* or *CAN1* cassettes with the correct repeat length, $(GAA)_{100}$ and $(CAG)_{140}$. *los1Δ* and *mus81Δ* were generated by using a plasmid containing a hygromycin resistance gene as a template and

primers with 60bp long tails homologous to the gene of interest. This PCR product was then transformed into yeast cells using a modified Gietz protocol from Anna Askenova. Integration of the PCR product was checked by growing transformed cells on hygromycin containing plates. Colonies from these plates were checked for specificity of integration using PCR with primers homologous to regions within the hygromycin gene and flanking sequences. Colonies were streaked for singles and the repeat length and knockout were checked again. Finally, frozen stocks were made from isolates with the correct repeat length and knockout. The *yen1Δ* was made using the same protocol, except a nourseothricin (NAT) resistance gene was incorporated instead of hygromycin. This was done to allow for the creation of a *yen1Δ* strain and a double *mus81Δyen1Δ* strain.

Spot Assay

A spot assay was performed for *los1Δ*, *mus81Δ*, *yen1Δ*, and *mus81Δyen1Δ* to approximate the effect of the knockout strains on canvanine resistance and to determine if the strains had any growth defects. 1:5 dilutions of the knockout strains, *can1Δ* , and wild type $(CAG)_{140}$ were plated on complete media (YPD), galactose-canavanine 60µg/mL plates, and galactose-canavanine 200µg/mL plates. *can1Δ* strains do not contain the *CAN1* cassette and should have equal growth on all media. Wild type strains contain the *CAN1* cassette with $(CAG)_{140}$ repeats and serve as a comparison for the knockout strains. YPD plates were grown 3 days and galactose-canavanine plates were grown 5 days.

Fluctuation Assay

Fluctuation assays, based off of the Luria-Delbrook experiment, were performed to calculate the rate of expansion. For strains containing the $(CAG)_{140}$ *CAN1* cassette, colonies were grown three days on YP-galactose and CAG repeat length was checked through PCR. Colonies with the correct starting length were chosen and dilutions were plated on YPD and galactose-canavanine media. The number of colonies was counted on Day 3, Day 4, and Day 5 for YPD, galactose-canavanine 60μg/mL plates, and galactose-canavanine 200μg/mL plates, respectively. PCR of colonies on the galactose-canavanine plates was done to determine the frequency of expansions. 1.5% agarose gels were used to visualize expansions.

For strains containing the *URA3* cassette, colonies were grown 40 hours on YPD+uracil (50 μg/mL). Starting GAA repeat length was checked using PCR. Colonies with the correct starting length were chosen and dilutions plated on YPD and 5-FOA (0.09%). After three days, colonies were counted on YPD and on 5-FOA. PCR of colonies on 5-FOA were done to determine the frequency of expansion. 1.5% agarose gels were used to visualize expansions. All raw values are included in Appendix I.

The number of colonies on each plate and expansion frequency was then used to calculate the rate of canavanine/5-FOA resistance and expansion rate using MSS-maximum likelihood method.[14] The median rate and 95% confidence intervals were calculated using the formulas $p_0 = e^{-m}. p_r = \frac{m}{r}\sum_{i=0}^{r-1}\frac{p_i}{(r-i+1)}$ and $f(r|m) = \prod_{i=1}^{C} f(r_i|m)$. This method is preferred over other methods as it corrects for error that may have occurred during dilutions and pipetting and it is valid over all ranges of mutation rates.

These expansion rates could then be compared to the wild type expansion rate. If an increase in expansion rate was seen in the mutant, then it suggested that the protein normally prevented expansions from occurring. If a decrease in expansion rate was seen in the mutant, then it suggested that the protein normally was required for expansions to occur.

A full fluctuation assay for expansion rate was not done for the strains with the $(GAA)_{100}$ *CAN1* cassette. These were grown 60 hours on YP-Galactose. Then colonies with the correct starting length, confirmed by PCR, were plated on YPD and galactose-canavanine 60µg/mL plates. YPD colonies were counted on Day 3. Galactose-canavanine colonies were counted on Day 3, 4 and 5. PCRs were done to confirm that expansion frequency was approximately the same between the wild type and knockout strains. Canavanine resistance was used as a proxy for expansion rate.

Small Scale Expansion Assay

The small-scale expansion protocol was adapted from the Freudenreich laboratory of Tufts University and was used to determine if Los1 affected small-scale as well as large-scale CAG repeat expansion. A colony of correct starting repeat length was inoculated in liquid culture and grown until it experienced 6-7 cell divisions, measured by changes in optical density (OD). Liquid cultures of YPD and YP-Galactose were done in parallel. This was done because the galactose sugar source ensures that there are high rates of transcription through *CAN1*, due to the UAS_{gal}. If glucose is the sugar source, there will be little transcription through *CAN1*, which may affect CAG expansion differently in the *los1Δ* background. Dilutions were plated onto YPD and grown for 2 days. PCR of approximately

96 colonies per treatment was performed to check CAG repeat length. 1.5% agarose gel images were scored with a program called Total Lab Quant. Changes of 5 repeats (15bp) or more was deemed an expansion or contraction. Significance of these events was determined with a Fisher's Exact Test.

Results

Large-Scale CAG Expansion in *los1Δ*

The knockout strain of *los1Δ* was successfully created and confirmed with PCR amplification.

First a spot assay was done. The YPD plate shows equal growth between strains, indicating that the *los1Δ* mutation does not affect cell growth (Figure 5). As seen in the genetic screen, the *los1Δ* strain showed more growth than the wild type strains on both concentrations of the galactose-canavanine plates, indicative of increased canavanine resistance.

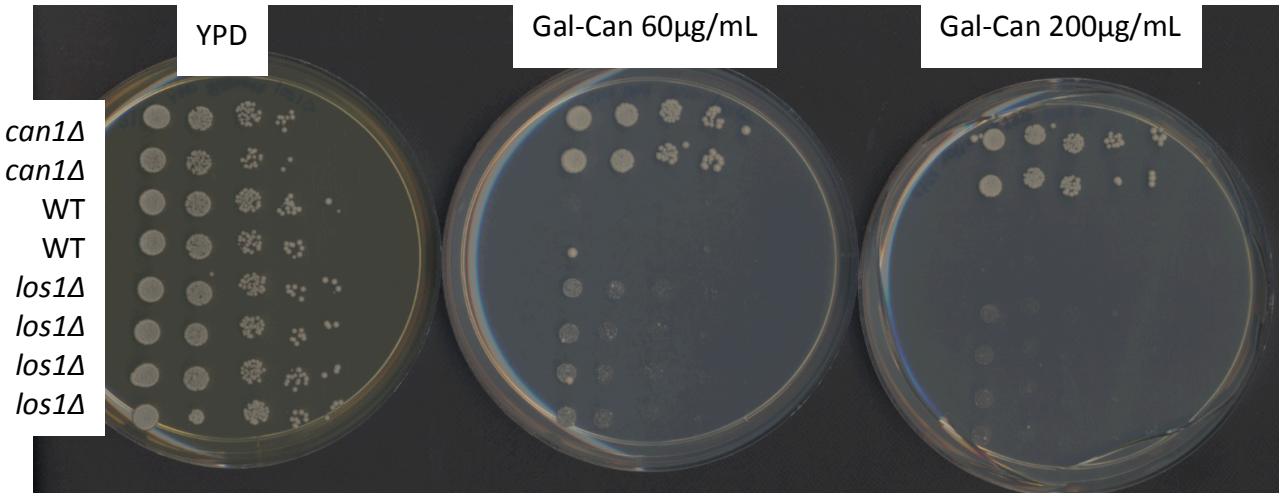

Figure 5. Spot assay of 1:5 dilutions of *los1Δ*, WT, and *can1Δ*. *can1Δ* do not contain the *CAN1* cassette. WT strains contain the *CAN1* cassette with $(CAG)_{140}$ repeats. *los1Δ* contain the *CAN1* cassette and do not have a functional *LOS1* gene. YPD media contain yeast extract, peptone, and dextrose. Gal-Can media contain yeast extract, peptone, galactose, amino acid dropout mix –arg, and different concentrations of canavanine (60µg/mL or 200µg/mL).

Next, a fluctuation assay was performed to calculate canavanine resistance rate and expansion rate (Table 1). The first *los1Δ* trial only tested the galactose-canavanine 60µg/mL concentration and showed a significant increase in canavanine resistance rate.

The expansion rate was also significantly increased over wild type. This prompted further investigation of the role of Los1 in repeat expansion.

Further fluctuation assay trials of *los1Δ* were performed, but no expansion rates could be calculated due to malfunctioning of the PCR where expansions could not be visualized, most likely caused by technical errors with a specific Taq polymerase lot number. These later trials show a significant decrease in canavanine resistance rates as compared to the rates from Trial 1 (Table 1, Figure 6). The rates from Trial 2 are still increased over wild type for both canavanine concentrations. The Trial 3 galactose-canavanine 60µg/mL rate is significantly lower than wild type, while the 200µg/mL rate is not significantly different from wild type.

Table 1. Fluctuation assay results from *los1Δ* trials. Rates of canavanine resistance and expansion rate were calculated using the MSS method. Top number representations the median rate; numbers below indicate 95% confidence intervals.

	Canavanine Concentration: 60µg/mL plates counted on day 4		200µg/mL plates counted on day 5	
	Canavanine Resistance Rate (10^{-5})	**Expansion Rate (10^{-5})**	**Canavanine Resistance Rate (10^{-5})**	**Expansion Rate (10^{-5})**
Wild Type	1.1 0.939-1.28	0.274 .212-.342	0.501 0.425-0.581	0.14 0.109-174
los1Δ Trial 1	12.8 12.0-13.6	4.51 4.91-5.13		
los1Δ Trial 2	4.3 3.97-4.64		1.58 1.42-1.76	
los1Δ Trial 3	0.541 0.459-0.628		0.48 0.404-0.561	

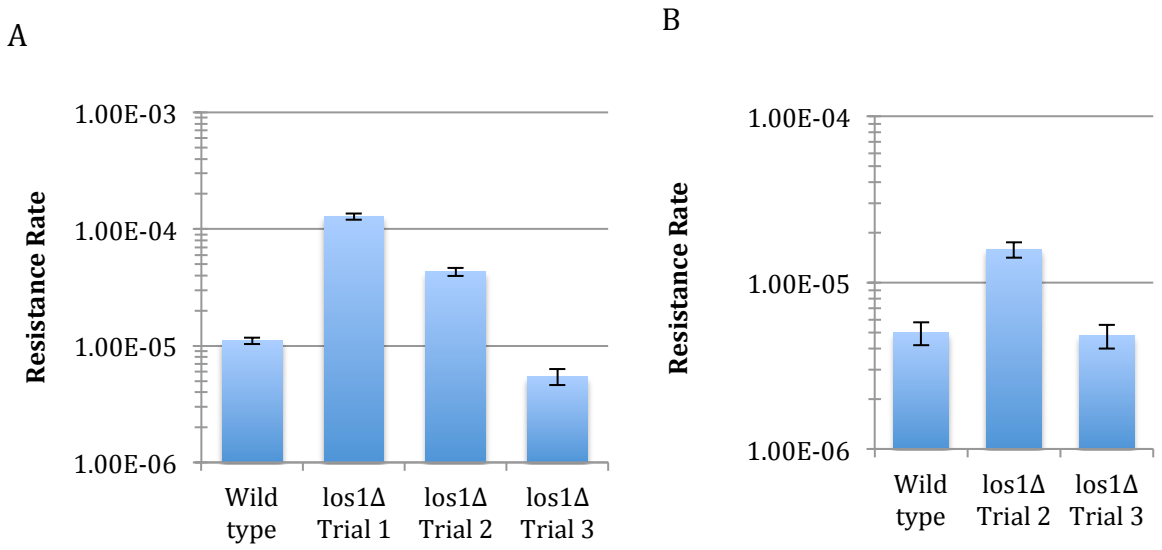

Figure 6. Canavanine resistance rates for gal-can 60µg/mL (A) and gal-can 200µg/mL (B).

Small-Scale CAG Expansion in *los1Δ*

To further characterize *los1Δ* , its effect on small-scale expansion was analyzed. Liquid cultures of glucose and galactose were grown in parallel and then plated on nonselective YPD media. Both glucose and galactose were tested since the *CAN1* cassette, which contains the $(CAG)_{140}$, relies on a UAS_{gal} (Figure 3). In galactose media, there should be high rates of transcription through the *CAN1* cassette. In glucose media, the transcription rate should be much lower. Since this could affect frequency of expansion, both media were tested.

Mixed results were seen for Los1's effect on small-scale expansion and contraction (Table 2). In media with galactose as the sugar source, there was no difference in expansion frequency, but there was a significant difference in contraction frequency, with a higher frequency in *los1Δ* than in the wild type. When glucose was the sugar source, there was a significant difference in the expansion frequency with a higher frequency in wild type than

los1Δ, and there was a significant difference in contraction frequency with a higher frequency in *los1Δ* than wild type. However it is important to note that the total number of colonies checked for repeat length in glucose media was lower for the wild type, which may have had an impact on the calculated frequencies.

Table 2. Small-scale expansion and contraction frequency of CAG repeats. Differences of 5 repeats (15bp) were considered expansions and contractions. Significance was determined using Fisher's Exact Test.

Galactose	**Expansions**	**Total**	**Frequency**	
Wild type	5	96	0.053	p=1
los1Δ	4	95	0.042	
	Contractions	**Total**	**Frequency**	
Wild type	13	96	0.137	p=0.0007
los1Δ	33	95	0.347	
Glucose	**Expansions**	**Total**	**Frequency**	
Wild type	11	47	0.234	p=0.0032
los1Δ	5	95	0.053	
	Contractions	**Total**	**Frequency**	
Wild type	5	47	0.106	p= 0.0300
los1Δ	26	95	0.274	

The length distribution of the small-scale expansions for glucose and galactose is shown in Figure 7. The majority of colonies did not experience an expansion or contraction event. Though *los1Δ* has a higher frequency of contraction on galactose (Table 2), there is no markedly different length distribution between the wild type and *los1Δ* (Figure 7A). On glucose there was a significant difference in both the frequency of expansion and the frequency of contraction. This is clearly seen in the length distribution, with a much higher percentage of wild type colonies having experienced a small-scale expansion and a much higher percentage of *los1Δ* colonies having experienced a small-scale contraction (Figure 7B). Though few wild type colonies experienced a contraction in glucose, the length distribution of the contractions differs from *los1Δ*. The *los1Δ* strain experienced more small

contractions with a change of 5-20 repeats, while the wild type had more contractions ranging between a change of 20-60 repeats.

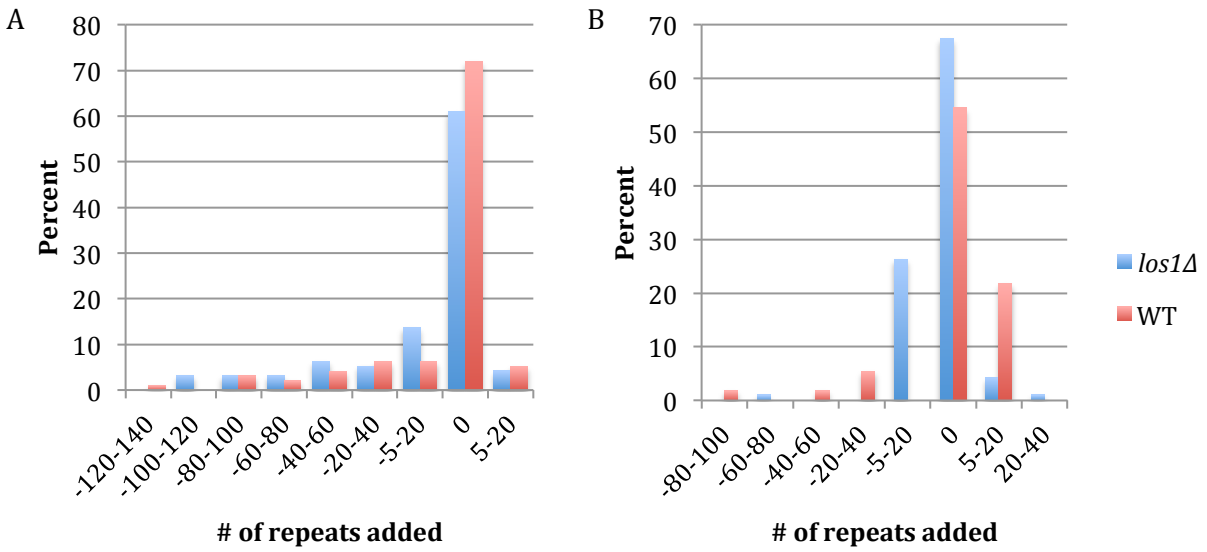

Figure 7. Length distribution based off of starting $(CAG)_{140}$ length for *los1Δ* and wild type strains on galactose (A) and glucose (B). A) No significant difference in the frequency of expansion was seen on galactose, but there was significantly more contraction in the *los1Δ* strain. Data collected from one trial on galactose with n=95,96 for *los1Δ* and wild type, respectively. B) There was a significant decrease in the frequency of expansions and a significant increase in the frequency of contraction for the *los1Δ* strain as compared to wild type. Data collected from one trial on glucose with n=95,47 for *los1Δ* and wild type, respectively.

Large-Scale GAA expansion in *los1Δ*

To determine whether Los1 affected GAA repeat expansion, a successful knockout of *los1Δ* in the *URA3* cassette was created and confirmed with PCR.

Then a fluctuation assay was performed and expansion rates were calculated (Figure 8). Both the 5-FOA resistance rate and expansion rate were slightly decreased in *los1Δ*. It is important to note that the wild type rates in this trial are significantly higher than the wild type rates reported by Zhang et al.[15] Part of this difference may be due to the rates being calculated using different methods. Zhang et al. used Drake's method,[16] whereas these rates were calculated using MSS-maximum likelihood. Even so, the

difference in calculation method cannot entirely account for the large difference between the wild type rates. But because this *los1Δ* and wild type trial were done in parallel, these values can still be compared.

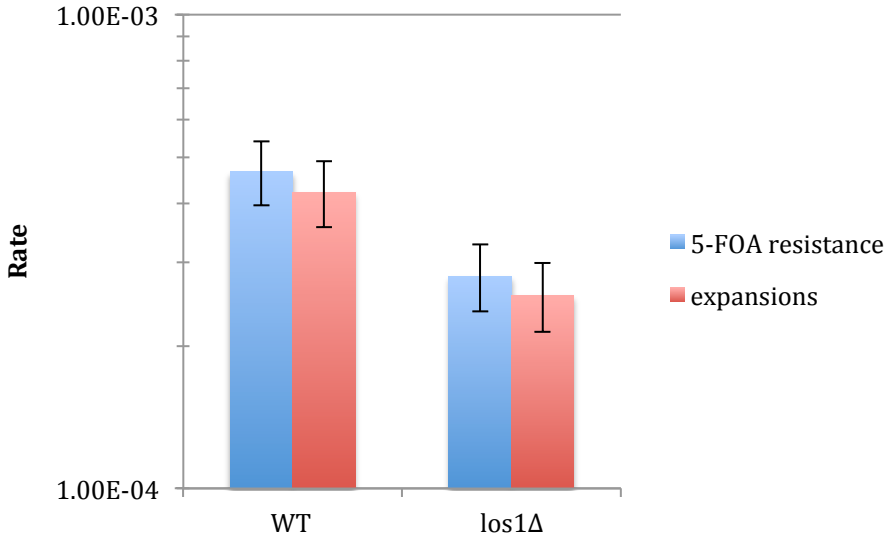

Figure 8. Fluctuation assay results for *los1Δ* and wild type of the $(GAA)_{100}$ *URA3* cassette. 5-FOA resistance and expansion rates were calculated using the MSS method. 95% confidence intervals are shown.

GAA repeats in a *los1Δ CAN1* Cassette

A recent paper by McCormik et al. found that the Los1 knockout increased arginine biosynthesis within yeast cells.[17] This is potentially problematic as the *CAN1* cassette relies on canavanine integration into proteins to have its effect. If there is increased arginine levels present in the cell, it is possible that there is competition between arginine and canavanine for integration into proteins, causing canavanine resistance to increase. This increased canavanine resistance would affect the expansion rate as the expansion rate is based off of the number of colonies growing on the selective media.

To test whether the effects seen in the spot assay and in the $(CAG)_{140}$ *CAN1* cassette fluctuation assays were affected by increased canavanine resistance, a *los1Δ* strain was

made of the $(GAA)_{100}$ *CAN1* cassette. It was hypothesized that if no increase in expansion rate is seen in this cassette, then the expansion rates in the *los1Δ* $(CAG)_{140}$ *CAN1* cassette were truly a result of CAG repeat expansion, not just increased canvanine resistance, and that Los1 affects CAG repeats.

First, a knockout of Los1 was made in the $(GAA)_{100}$ *CAN1* cassette and confirmed through PCR analysis. Then a fluctuation assay was done (Figure 9). Expansion rates were not calculated. However, PCRs for repeat length were done for the wild type and *los1Δ* strains and the frequency of expansion between the two strains was approximately equal, so the canavanine resistance rates were used as proxies for the expansion rates. *los1Δ* has a slight, but significant, decrease in the canavanine resistance rate on day 3 and day 4. However, on day 5 there is no longer a significant difference between the two strains.

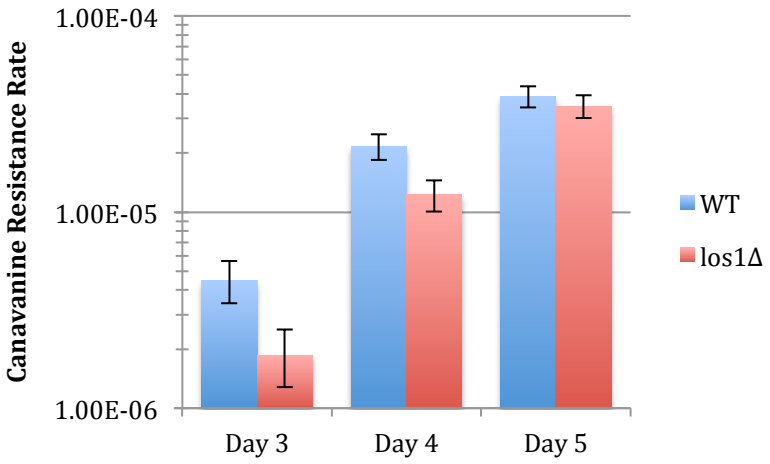

Figure 9. Canavanine resistance rates from fluctuation assay of *los1Δ* in the $(GAA)_{100}$ *CAN1* cassette done with galactose-canavanine 60µg/mL. Rates of canavanine resistance were calculated using the MSS method. 95% confidence intervals are shown.

During this fluctuation assay, differences in growth were observed between the wild type and *los1Δ* on the galactose-canavanine 60µg/mL plates. The wild type had more large, distinct colonies (Figure 10). The *los1Δ* strain had fewer distinct colonies, but many more background colonies.

Wild Type *los1Δ*

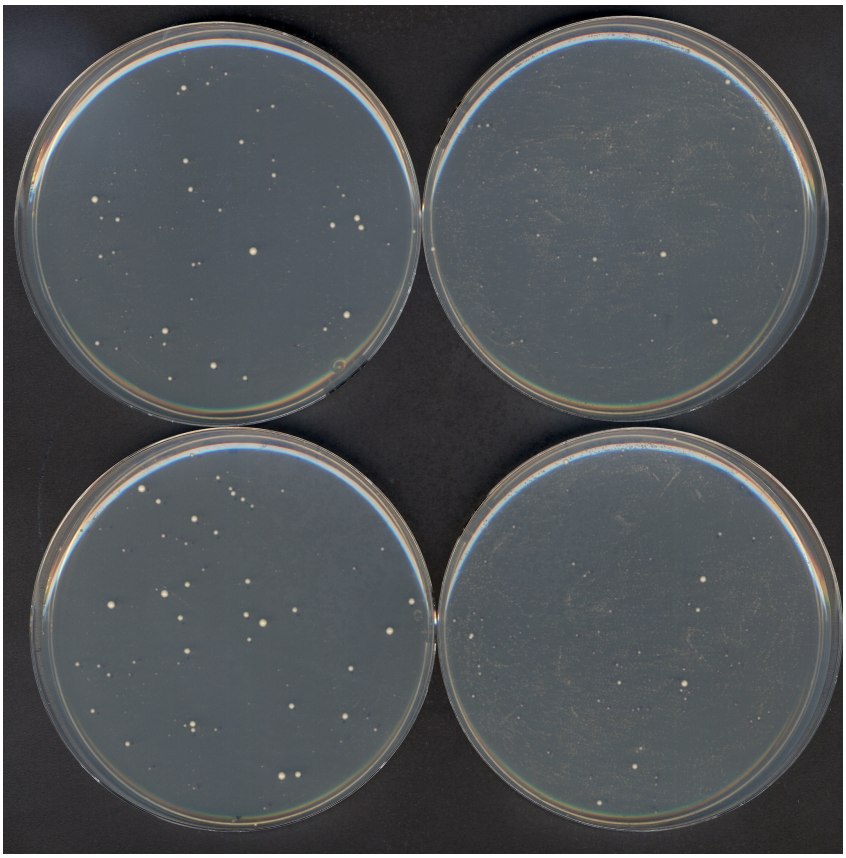

Figure 10. Galactose-Canavanine 60 µg/mL plates on day 4. Large, distinct colonies are seen in the wild type strain. Few distinct colonies but a large number of small background colonies visible in the los1 strain.

Large-Scale CAG Expansion and Holliday Junction Resolvases

A successful knockout was made of *yen1Δ* and confirmed through PCR. Using a *mus81Δ* strain previously created, a *mus81Δyen1Δ* was made and confirmed through PCR. A spot assay was done of *mus81Δ*, *yen1Δ*, and *mus81Δyen1Δ* (Figure 11). The YPD plate shows equal growth between strains, indicating that there are no growth defects and that an approximately equal number of cells were plated for each. The *can1Δ* strain does not contain the *CAN1* cassette, thus showing equal growth on all plates. No difference in growth on galactose-canavanine media was apparent between the knockout strains and the wild type.

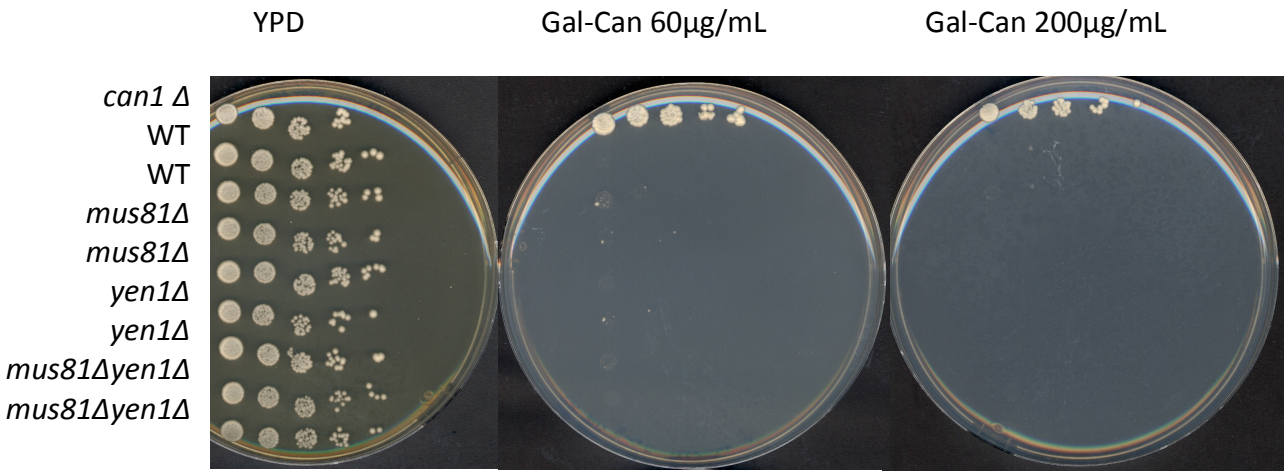

Figure 11. Spot assay of 1:5 dilutions of mus81Δ, yen1Δ, and mus81Δyen1Δ strains. *can1Δ* do not contain the *CAN1* cassette. WT strains contain the *CAN1* cassette with $(CAG)_{140}$ repeats. mus81Δ, yen1Δ, and mus81Δyen1Δ contain the *CAN1* cassette and do not have a functional copy of the respective gene. YPD media contain yeast extract, peptone, and dextrose. Gal-Can media contain yeast extract, peptone, galactose, amino acid dropout mix –arg, and different concentrations of canavanine (60µg/mL or 200µg/mL).

A fluctuation assay was then performed using the gal-can 60µg/mL concentration and the MSS method was used to calculate canavanine resistance rates and expansion rates. There was no significant difference between canavanine resistance rates, as was seen on the spot assay, but there was a significant difference between the expansion rates. The *mus81Δyen1Δ* double mutant showed about a three-fold decrease in expansion rate as compared to the wild type (Figure 12), while the single mutant showed no significant difference.

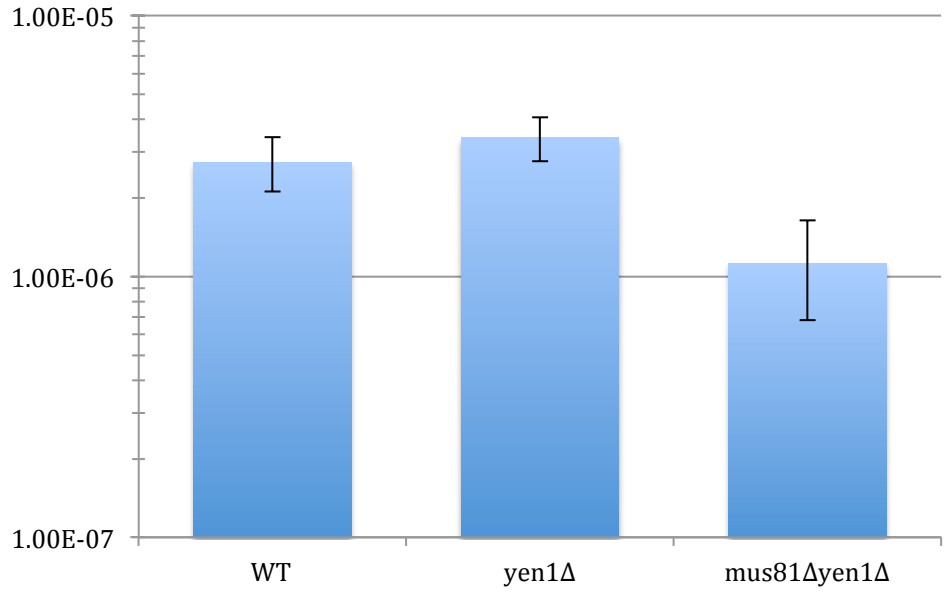

Figure 12. Expansion rates of *yen1Δ*, *mus81Δyen1Δ*, and wild type strains on galactose-canavanine 60µg/mL. Rates were calculated using MSS and show expansion rate and 95% confidence intervals

This fluctuation assay was then performed on a higher concentration of selective media, using the gal-can 200µg/mL plates. Again the same trend was found. There was some fluctuation in the canavanine rates, but all differences found were minor (Figure 13). The single knockout strains were not greatly different from the wild type. But there was a large decrease in the *mus81Δyen1Δ* expansion rate, confirming what was seen in the previous fluctuation assay.

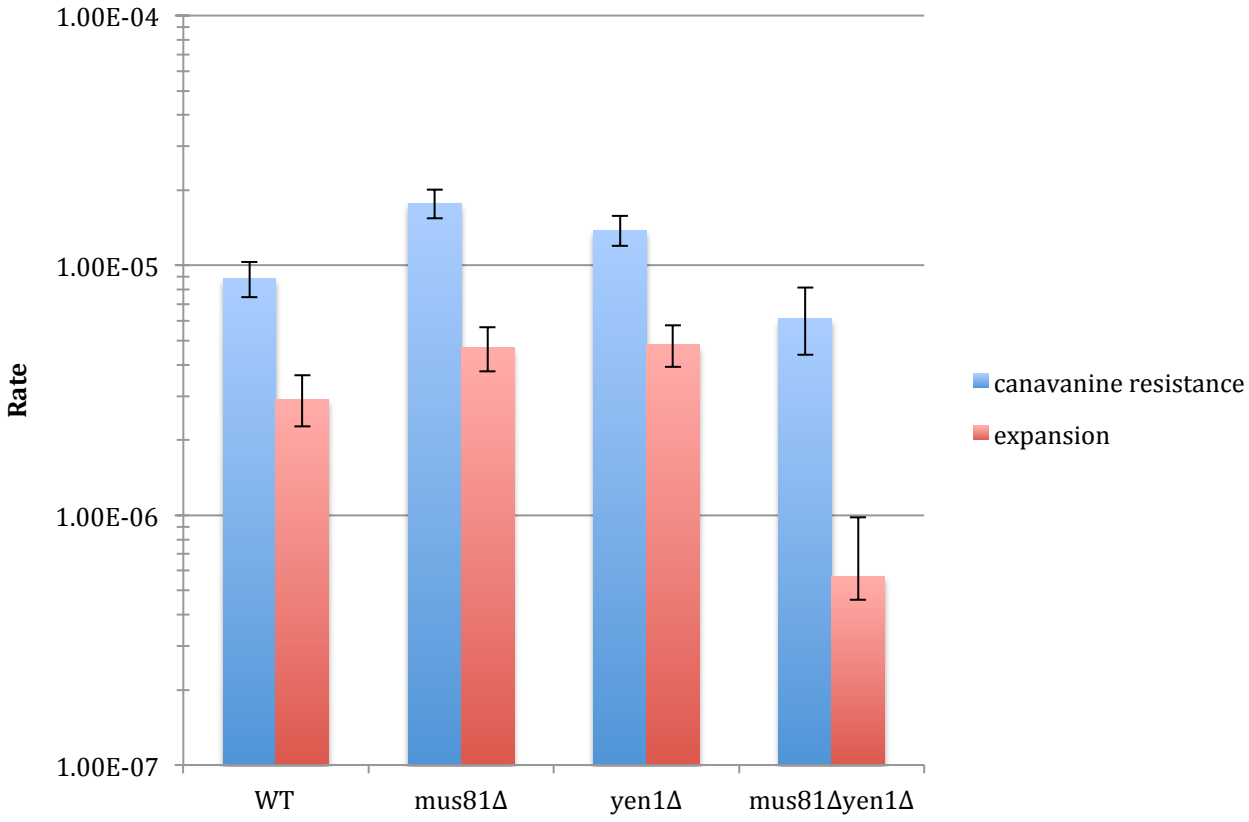

Figure 13. Rates of fluctuation assay performed with *yen1Δ*, *mus81Δyen1Δ*, and wild type strains on galactose-canavanine 200µg/mL. Rates were calculated using MSS and show rate and 95% confidence intervals

Discussion

los1Δ Increases Canavanine Resistance, but Does Not Affect CAG Repeat Expansion

Los1 was identified in an unbiased genetic screen, which used the marker of increased canavanine resistance to identify mutations that affected CAG repeat expansion. As seen in the screen, the *los1Δ* strain showed an increase in canavanine resistance rate in the *CAN1* cassette spot assay (Figure 5). An initial fluctuation assay also showed promising results, suggesting that the increased canavanine resistance corresponded with an increased expansion rate, indicating that Los1 may be involved in preventing repeat expansions (Table 1). Because of this, Los1 was pursued further. The goal was to confirm the increase in expansion rate in the *CAN1* cassette, determine whether *los1Δ* affected non-hairpin forming repeats using the $(GAA)_{100}$ *URA3* cassette, and to determine whether *los1Δ* affected small-scale CAG repeat expansion.

Multiple fluctuation assay trials with the $(CAG)_{140}$ *CAN1* cassette indicated that the initial large increase in canavanine resistance and expansion rate may have been a chance occurrence. The canavanine resistance rates from the later fluctuation assay trials showed a steady decrease with each assay performed (Table 1). The Trial 3 results are most puzzling, as they do not correspond to the increase in canavanine resistance seen on the spot assay. Explanations of this result in this include experimenter error and changes in media, such as a drier media which could result in an increased canvanine concentration. These results add to the overall difficulty in determining Los1's role. Additionally, Trial 1 saw a very large decrease between the canavanine resistance rate and expansion rate. If such a large decrease also occurred in the later trials, it is likely that there would be no significant difference in expansion rate between the *los1Δ* strains and wild type or that

los1Δ would have a decreased expansion rate compared to wild type. To increase the ambiguity of these results, the wild type rates found here are significantly lower than what has been recorded by other experimenters. It is possible that this difference is due to experimenter technique or changes in the media conditions over time.

The data of the small-scale expansion assay also adds to the confusion. There was no significant difference in expansions on galactose (Table 2), which contradicts the results of Trial 1 of the fluctuation assay but corroborates what was suggested by the later fluctuation assay trials: Los1 doesn't affect CAG repeat expansion. However, no effect may have been seen if Los1 is only involved in large-scale expansion events.

The $(GAA)_{100}$ *URA3* fluctuation assay was done to determine whether Los1 affects non-hairpin forming repeats, such as GAA repeats. Los1 is involved in the export of tRNAs from the nucleus, which have a looped secondary structure, so it was hypothesized that Los1 may be able to recognize looped nucleic acid secondary structures, including CAG hairpins. GAA repeats do not form these hairpins, thus this fluctuation assay was meant to determine whether structural recognition is occurring. The fluctuation assay showed a minor decrease in the expansion rate (Figure 8). However, it has been reported that *los1Δ* has decreased resistance to 5-fluorouracil.[18] Most likely, this is the cause for the decrease seen in the 5-FOA resistance rate and expansion rate, suggesting that the differences seen in the fluctuation assay are not truly significant.

The data thus far did not clearly determine the role of Los1 on CAG repeat expansion. Significant differences between the wild type and the knockout strain were found, but it was not clear whether this was truly due to Los1's influence. In the midst of this project, a paper by McCormik et al. identified the Los1 knockout as a mutation that

extends replicative lifespan in yeast. [17] They further characterized the knockout and found increased transcription of amino acid biosynthesis genes, including arginine. This provides a potential complication when working in the *CAN1* cassette. As previously stated, the *CAN1* gene codes for a transporter that allows canavanine into the cell so that it is integrated into proteins in place of arginine, causing proteins to misfold and eventually cell death. If arginine biosynthesis is increased, it is possible that there is competition between arginine and canavanine for integration in proteins. If enough arginine is incorporated into proteins, it is possible that a sufficient number of functional proteins are created for the cell to survive. Thus, it is likely that canavanine resistance is increased and proper selection of cells that have expansions is not occurring in the *los1Δ* strain. With this interference in the counter-selectable marker, no strong conclusion could be drawn from the data that had been collected thus far.

In hopes of determining the cause of the increase in canavanine resistance of the $(CAG)_{140}$ *CAN1* fluctuation assay, a knockout was made of Los1 in a strain containing the $(GAA)_{100}$ *CAN1* cassette. Since the GAA repeats had not seen an increased expansion rate in the *URA3* cassette with *los1Δ*, a similar result would be expected in the *CAN1* cassette if *los1Δ* did not affect canavanine resistance. However, if the *los1Δ* mutation did increase canavanine resistance, then an increased canvanine resistance rate and expansion rate would be expected in the *los1Δ* $(GAA)_{100}$ *CAN1* strain. It was hypothesized that if there were no change in expansion rate in this strain, then the differences in expansion rate observed in the $(CAG)_{140}$ *CAN1* strains must have been caused by an expansion of CAG repeats, rather than by an overall increase in resistance to canavanine.

The *los1Δ* $(GAA)_{100}$ *CAN1* cassette fluctuation assay was performed. A sufficient number of PCRs were completed to determine that the expansion frequencies between the wild type and *los1Δ* strain were the same. Thus, the canvanine resistance rates were used as proxies of the expansion rate. A slight decrease in the *los1Δ* canavanine resistance rate was seen in the *los1Δ* strain with three and four days of growth on selective media (Figure 9). However, with each day of growth, this effect was becoming less intense. On Day 5, there was no longer a difference between *los1Δ* and wild type, clearly showing that the *los1Δ* mutation increased canavanine resistance.

There are two ways this data could be interpreted. One, Los1 is required for GAA repeat expansion, as seen by both the decrease in expansion rate of the *los1Δ* $(GAA)_{100}$ *URA3* cassette and the decrease in canavanine resistance in the $(GAA)_{100}$ *CAN1* cassette. But the effect in the *CAN1* cassette was masked after another day's growth by the increase in canavanine resistance. This interpretation requires that the lower expansion rate in the *los1Δ* $(GAA)_{100}$ *URA3* fluctuation assay was not caused by increased sensitivity to 5-fluourouracil which had been previously reported.[19]

In another interpretation of the data, the $(GAA)_{100}$ *CAN1* fluctuation assay shows that canavanine resistance increases as the *los1Δ* colonies are allowed to grow. Thus, any rates calculated from a strain with the *los1Δ* mutation in a *CAN1* cassette strain, regardless of microsatellite, would be affected and are most likely not representative of the knockout's true effect. This interpretation would explain that neither the *CAN1* cassette nor the *URA3* cassette are suited to studying *los1Δ* due to its interference in sensitivity to the counter selectable marker. Thus, the data collected in the fluctuation assays cannot give a clear indication into the role of Los1 in CAG repeat expansion. Another cassette would have to be

developed in order to fully examine the effect of Los1. Unfortunately, no such cassette has been developed at this time. But because no strong, consistent changes in canavanine resistance rate or expansion rate were seen is the *los1Δ* strains, it is likely that Los1 is not directly involved in large-scale CAG repeat expansion.

These results show the limitations of the genetic screen performed. The screen used to identify Los1 relied on a single marker as an indicator of increased expansion rate. In the case of Los1, the screen gave results of exactly what was screened for: increased canavanine resistance. In order to find genes that are involved in repeat expansion, not just in changing sensitivity to the selectable marker, the genetic screen can be designed to give more accurate results. For example, including two cassettes with different counter selectable markers could return more genes that are involved in repeat expansion. In this scenario, if a mutation occurred that greatly increased the expansion rate, then it is likely that expansions would occur in both cassettes and would be detectable through both counter selectable markers. Whereas if there is a mutation that increases resistance to a selectable marker, the other selectable marker will indicate that this isn't a true expansion-causing mutation. Moving forward, a second cassette with CAG repeat expansions should be developed so that such a screen could be performed.

Large-Scale CAG Repeat Expansion and Holliday Junction Resolvases – Yen1 and Mus81

Before this project, a *mus81Δ* strain was tested for its effect on CAG repeat expansion, and the fluctuation assay showed that it had no effect. Since Yen1 and Mus81 have the same function of resolving Holliday junctions but with temporal separation, it was thought that Yen1 might be repairing Holliday junctions that Mus81 was unable to repair.

Thus, *mus81Δ* would not be showing the actual effect of Holliday junction resolvases on CAG repeat expansion. When the *yen1Δ* strain was created no effect was seen again, indicating that both proteins were unable to affect CAG repeat expansion independently.

In these single knockout strains, double strand breaks are still occurring and are able to be repaired, despite the loss of one of the resolvases. For example, if a double strand break occurs during DNA synthesis, homologous recombination could occur and Mus81 would normally repair the Holliday junction according to Matos and West's theory of temporal regulation.[12] But if this occurred in the *mus81Δ* strain, the Holliday junction could not be resolved here, but could be repaired later during anaphase when Yen1 is reactivated. The same is true in the *yen1Δ* strain. If a break occurred during metaphase, anaphase or G1, homologous recombination could occur and Yen1 would normally repair the Holliday junction. However, in the *yen1Δ* strain this Holliday junction would remain until Mus81 was reactivated during prometaphase and only then could be resolved. This redundancy ensures that the cell will be able to repair all double strand breaks completely, even when one of the proteins responsible for this repair is inactivated. If this were occurring in the single knockout strains, it would explain why no effect on expansion rate was seen despite their role in large-scale CAG expansion.

To determine whether the proteins were compensating for each other, a double mutant strain of *mus81Δyen1Δ* was created. The decrease in the expansion rate of this strain suggests that one protein was able to offset the inactivation of the other in the single mutant strains (Figure 12, Figure 13). It also suggests that Mus81 and Yen1 are required for CAG expansion and that expansions occur when these resolvases repair Holliday junctions.

These results implicate processes in which Holliday junctions occur as causing CAG repeat expansions, providing evidence for a recent model proposed by Dr. Jane Kim (unpublished). This model begins with fork stalling which was described earlier (Figure 1C). When the replication fork stalls, it can reverse to create a chicken foot structure. This structure can isomerize to result in a Holliday junction (Figure 14). This Holliday junction must then be repaired, possibly by resolvases such as Mus81 and Yen1. The resolution of this structure would result in a one-ended break, which would then need to be repaired. It could be repaired by BIR which ends at the end of a chromosome or which ends by running into an oncoming replication fork, which is shown in Figure 14. If BIR reaches another replication fork, another Holliday junction would be created. This again could be resolved by Mus81 or Yen1. Thus, this model reveals two potential places in which Mus81 and Yen1 could act. Though the resolvases are not directly responsible for the expansion (rather, it is the out-of-register invasion that results in expansion), this model explains how these proteins are required for large-scale CAG repeat expansion to occur (Figure 14).

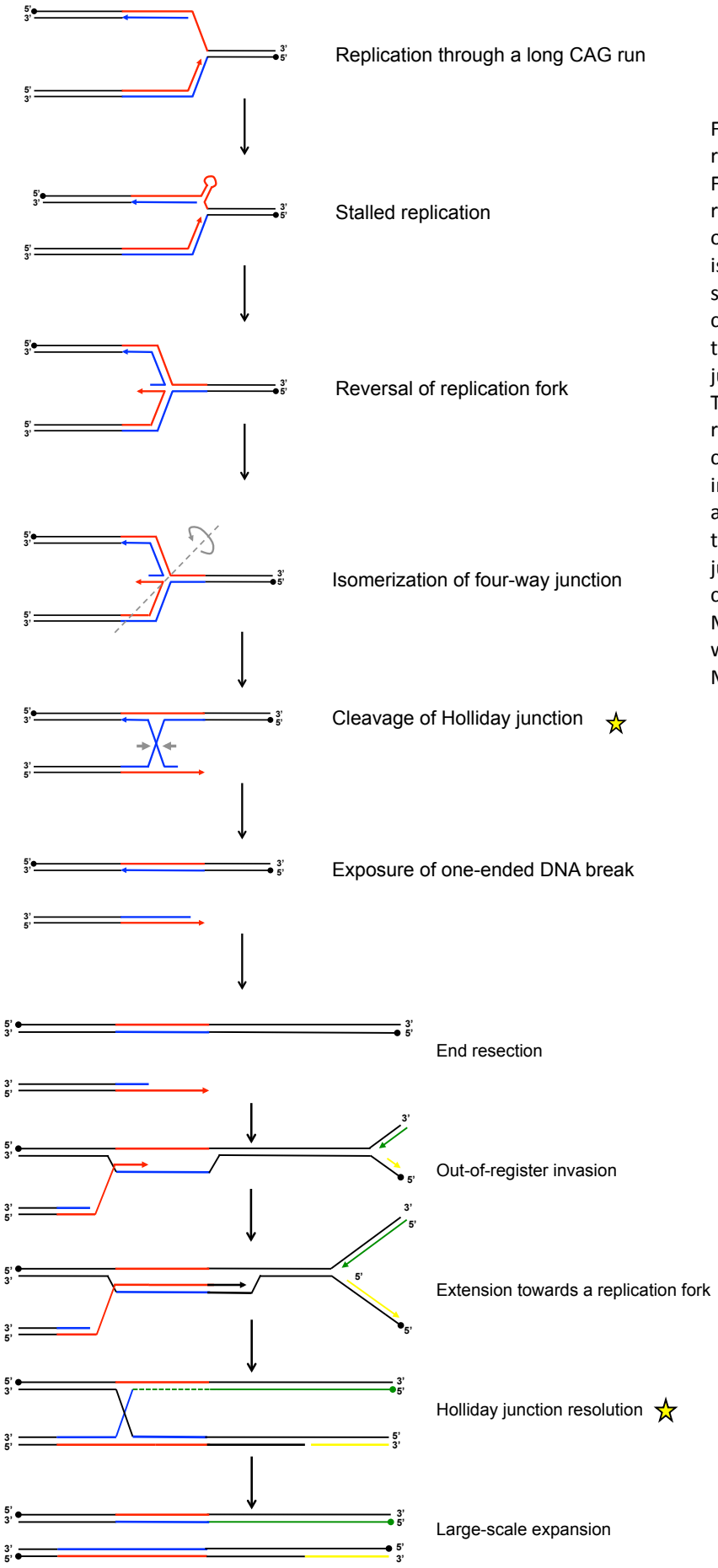

Figure 14. Model of large-scale CAG repeat expansion (Kim, unpublished). First DNA polymerase stalls on a repetitive tract, causing the reversal of the replication fork. Then through isomerization of the chicken-foot structure, a Holliday junction can be created. This Holliday junction would then be repaired, possibly by Holliday junction resolvases Mus81 and Yen1. The resolution of this structure would result in a one-ended break, which could be repaired through break-induced replication that reaches another replication fork, as shown in the diagram. Another Holliday junction will be created when these collide, which could be resolved by Mus81 or Yen1. Stars indicate steps in which Holliday junction resolvases, Mus81 and Yen1, could play a role.

It is also possible that a novel mechanism is occurring to cause these expansions. To give further insight into the exact mechanism, the change in repeat length and other intermediates in the pathway should be analyzed. Another explanation for this decrease in expansion rate is that the strain accumulated stabilizing interruptions within the repetitive tract during the process of creating the knockouts. To eliminate this as a potential cause of the decrease in expansion rate, the *CAN1* cassette in this strain should be sequenced.

Though these results are very promising, there are still some questions remaining. The wild type expansion rate recorded here is significantly lower than what has been described by other experimenters. To address this, a fluctuation assay of the wild type, *mus81Δ*, *yen1Δ*, and *mus81Δyen1Δ* will be performed in parallel with another experimenter. This will confirm the results described here as well as determine whether the difference in wild type data are due to differing experimenter techniques or changes in the wild type strain over time.

Conclusion

The goal of this project was to give insight into possible mechanisms involved in CAG repeat expansion. This was successfully done through the examination of Holliday junction resolvases, Mus81 and Yen1. We found that these proteins were required for repeat expansion to occur, indicating the importance of homologous recombination in large-scale CAG repeat expansion. The other half of the project was focused on Los1, which had been identified through a genetic screen as being involved in CAG repeat expansion. Through many fluctuation assays it became apparent that *los1Δ* was interfering with the counter selectable markers used in both genetic cassettes. Though this clouded the ease of

interpreting the data, there were no large, consistent changes seen in *los1Δ* strains, suggesting that Los1 does not play a direct role in CAG repeat expansion. These results underline the importance of designing genetic screens with robust phenotypic output. Such a screen would give insight into currently unexplored mechanisms of expansion, furthering our understanding of CAG repeat expansion and the diseases associated with it.

Appendix I –

Raw data of fluctuation assays. The number of plates for each fluctuation assay is indicated. The number of colonies were counted after a period of growth, as described in *Methods.* The indicated numbers of PCRs were performed from the colonies growing on selective media. From the PCRs performed, the number of colonies with expansions, contractions, and no change are indicated.

Wild Type $(CAG)_{140}$ *CAN1* YJK146 gal-can 60µg/mL						
Plate	Number of Colonies on Complete Media	Number of Colonies on Selective Media	Expansions	Contractions	No change	Total PCRs
1	172	17	3	1	4	8
2	296	19	1	2	5	8
3	208	33	0	1	7	8
4	184	26	0	1	7	8
5	298	49	1	2	5	8
6	257	35	1	1	6	8
7	224	30	1	2	5	8
8	238	30	0	2	6	8
9	162	31	1	1	6	8
10	147	23	3	1	4	8
11	191	23	2	1	5	8
12	262	31	1	2	4	7
Total			14	17	64	95
Frequency			0.1473684	0.1789473		

Wild Type $(CAG)_{140}$ *CAN1* YJK146 gal-can 200 µg/mL						
Plate	Number of Colonies on Complete Media	Number of Colonies on Selective Media	Expansions	Contractions	No change	Total PCRs
1	521	44	3	0	5	8
2	550	55	0	0	8	8
3	539	4	1	2	1	4
4	499	41	1	1	6	8
5	275	17	0	1	7	8
6	354	56	0	1	7	8
7	465	34	2	0	6	8
8	583	68	1	0	7	8
9	436	33	1	0	7	8
10	214	6	1	1	4	6
11	298	18	2	0	6	8
12	261	21	1	0	7	8
13	415	43	0	1	7	8
Total			13	7	78	98
Frequency			0.1326530	0.07142857		

***los1Δ* (CAG)$_{140}$ *CAN1* Trial 1 YJK209 gal-can 60µg/mL**						
Plate	Number of Colonies on Complete Media	Number of Colonies on Selective Media	Expansions	Contractions	No change	Total PCRs
1	142	38	1	0	7	8
2	137	26	2	0	6	8
3	108	43	2	0	6	8
4	173	32	2	0	6	8
5	126	47	2	0	6	8
6	101	50	1	0	7	8
7	81	25	2	0	6	8
8	137	73	4	1	3	8
9	169	21	2	1	5	8
10	105	18	3	0	5	8
11	159	24	2	0	6	8
12	178	33	1	0	7	8
13	67	44	1	0	7	8
14	97	69	1	0	7	8
Total			26	2	84	112
Frequency			0.232142857	0.0178571		

***los1Δ* (CAG)$_{140}$ *CAN1* Trial 2 YJK209 gal-can 60µg/mL**		
Plate	Number of Colonies on Complete Media	Number of Colonies on Selective Media
1	166	14
2	193	24
3	252	17
4	210	17
5	138	14
6	142	11
7	167	9
8	192	10
9	187	18
10	155	14
11	213	18
12	210	9
13	187	9
14	210	10

***los1Δ* (CAG)$_{140}$ *CAN1* Trial 2 YJK209 gal-can 200 µg/mL**		
Plate	Number of Colonies on Complete Media	Number of Colonies on Selective Media
1	166	37
2	193	59
3	252	44
4	210	42
5	138	48
6	142	52
7	167	48
8	192	44
9	187	70
10	155	44
11	213	51
12	210	26
13	187	30
14	210	38

***los1Δ* (CAG)$_{140}$ *CAN1* Trial 3 YJK209 gal-can 60µg/mL**		
Plate	Number of Colonies on Complete Media	Number of Colonies on Selective Media
1	176	26
2	153	11
3	142	3
4	133	5
5	152	6
6	140	12
7	184	12
8	225	8
9	145	6
10	187	14
11	168	11
12	163	9
13	169	7
14	137	10

***los1Δ* (CAG)$_{140}$ *CAN1* Trial 3 YJK209 gal-can 200 µg/mL**		
Plate	Number of Colonies on Complete Media	Number of Colonies on Selective Media
1	176	18
2	153	13
3	142	8
4	133	6
5	152	11
6	140	12
7	184	12
8	225	10
9	145	7
10	187	3
11	168	3
12	163	7
13	169	5
14	137	4

***los1Δ* (GAA)$_{100}$ *URA3* YJK216**						
Plate	Number of Colonies on Complete Media	Number of Colonies on Selective Media	Expansions	Contractions	No change	Total PCRs
1	78	27	6	0	2	8
2	51	16	8	0	0	8
3	82	24	5	0	3	8
4	55	19	7	0	1	8
5	63	19	7	0	1	8
6	97	46	7	0	1	8
7	116	200	8	0	0	8
8	69	31	8	0	0	8
9	92	34	7	0	1	8
10	109	23	8	0	0	8
11	95	22	6	0	2	8
Total			77	0	11	88
Frequency			0.875	0		

Wild Type $(GAA)_{100}$ *URA3*						
Plate	Number of Colonies on Complete Media	Number of Colonies on Selective Media	Expansions	Contractions	No change	Total PCRs
1	71	44	8	0	0	8
2	60	37	6	0	2	8
3	87	43	6	0	2	8
4	79	57	7	0	1	8
5	68	43	5	0	3	8
6	70	48	6	0	2	8
7	123	80	7	0	1	8
8	83	55	5	0	3	8
9	62	32	6	0	2	8
Total			56	0	16	72
Frequency			0.77777777	0		

Wild Type $(GAA)_{100}$ *CAN1* gal-can 60µg/mL				
		Number of Colonies on Selective Media		
Plate	Number of Colonies on Complete Media	Day 3	Day 4	Day 5
1	92	4	23	49
2	96	5	42	79
3	145	3	26	58
4	169	8	61	94
5	131	4	31	56
6	114	2	36	62
7	128	3	31	65
8	125	3	44	80
9	163	4	33	83
10	99	5	27	70
11	106	1	22	58
12	106	4	21	50

los1Δ* (GAA)$_{100}$ *CAN1 **YJK226** **gal-can 60µg/mL**				
		Number of Colonies on Selective Media		
Plate	Number of Colonies on Complete Media	Day 3	Day 4	Day 5
1	83	0	10	26
2	77	2	14	51
3	107	1	13	40
4	89	0	22	57
5	190	0	6	49
6	122	3	13	63
7	94	1	8	38
8	97	2	11	49
9	103	3	144	216
10	124	0	18	63
11	120	1	15	66
12	134	1	11	53

yen1Δ* (CAG)$_{140}$ *CAN1 **YJK228** **gal-can 60µg/mL**						
Plate	Number of Colonies on Complete Media	Number of Colonies on Selective Media	Expansions	Contractions	No change	Total PCRs
1	273	33	0	0	8	8
2	232	29	1	2	5	8
3	226	34	0	1	7	8
4	205	41	1	3	4	8
5	221	34	0	2	6	8
6	218	20	2	0	6	8
7	284	81	2	1	5	8
8	210	24	1	1	6	8
9	237	56	3	2	3	8
10	173	18	1	5	2	8
11	152	20	2	2	4	8
12	236	21	1	3	4	8
13	224	31	1	1	6	8
14	216	29	1	2	5	8
15	156	25	1	1	6	8
16	225	35	3	0	5	8
Total			20	26	82	128
Frequency			0.15625	0.203125		

mus81Δyen1Δ* (CAG)$_{140}$ *CAN1 **YJK229** **gal-can 60µg/mL**						
Plate	Number of Colonies on Complete Media	Number of Colonies on Selective Media	Expansions	Contractions	No change	Total PCRs
1	90	7	0	0	7	7
2	98	4	0	0	4	4
3	100	18	1	0	7	8
4	99	9	1	0	8	8
5	174	20	1	1	6	8
6	125	21	0	0	8	8
7	102	14	0	1	7	8
8	156	13	1	0	7	8
9	120	12	0	0	8	8
10	102	12	0	1	7	8
11	96	20	0	0	8	8
12	72	118	1	2	5	8
13	86	19	0	1	7	8
14	68	10	0	1	7	8
15	113	15	0	1	7	8
16	82	12	0	0	8	8
Total			4	8	111	123
Frequency			0.0325203	0.0650465		

Bibliography

1. Mirkin, S. M. Expandable DNA repeats and human disease. *Nature* **447,** 932–940 (2007).
2. Orr, H. T. & Zoghbi, H. Y. Trinucleotide Repeat Disorders. *Annu. Rev. Neurosci.* **30,** 575–621 (2007).
3. Andrew, S. E. *et al.* The relationship between trinucleotide (CAG) repeat length and clinical features of Huntington's disease. *Nat. Genet.* **4,** 398–403 (1993).
4. Shishkin, A. A. *et al.* Large-scale expansions of Friedreich's ataxia GAA repeats in yeast. *Mol. Cell* **35,** 82–92 (2009).
5. Kim, J. C. & Mirkin, S. M. The balancing act of DNA repeat expansions. *Curr. Opin. Genet. Dev.* **23,** 280–288 (2013).
6. McMurray, C. T. Mechanisms of trinucleotide repeat instability during human development. *Nat. Rev. Genet.* **11,** 786–799 (2010).
7. Freudenreich, C. H., Kantrow, S. M. & Zakian, V. A. Expansion and Length-Dependent Fragility of CTG Repeats in Yeast. *Science* **279,** 853–856 (1998).
8. Symington, L. S. Role of RAD52 Epistasis Group Genes in Homologous Recombination and Double-Strand Break Repair. *Microbiol. Mol. Biol. Rev.* **66,** 630–670 (2002).

9. RAD52 | SGD. *yeastgenome.org* Available at: http://yeastgenome.org/locus/S000004494/overview#reference. (Accessed: 10th April 2016)

10. Haber, J. E. TOPping Off Meiosis. *Mol. Cell* **57,** 577–581 (2015).

11. Symington, L. S., Rothstein, R. & Lisby, M. Mechanisms and regulation of mitotic recombination in Saccharomyces cerevisiae. *Genetics* **198,** 795–835 (2014).

12. Matos, J. & West, S. C. Holliday junction resolution: Regulation in space and time. *DNA Repair* **19,** 176–181 (2014).

13. Su, X. A., Dion, V., Gasser, S. M. & Freudenreich, C. H. Regulation of recombination at yeast nuclear pores controls repair and triplet repeat stability. *Genes Dev.* **29,** 1006–1017 (2015).

14. Foster, P. L. Methods for Determining Spontaneous Mutation Rates. *Methods Enzymol.* **409,** 195–213 (2006).

15. Zhang, Y. *et al.* Genome-wide Screen Identifies Pathways that Govern GAA/TTC Repeat Fragility and Expansions in Dividing and Nondividing Yeast Cells. *Mol. Cell* **48,** 254–265 (2012).

16. Drake, J. W. A constant rate of spontaneous mutation in DNA-based microbes. *Proc. Natl. Acad. Sci. U. S. A.* **88,** 7160–7164 (1991).

17. McCormick, M. A. *et al.* A Comprehensive Analysis of Replicative Lifespan in 4,698 Single-Gene Deletion Strains Uncovers Conserved Mechanisms of Aging. *Cell Metab.* **22,** 895–906 (2015).

18. Gustavsson, M. & Ronne, H. Evidence that tRNA modifying enzymes are important in vivo targets for 5-fluorouracil in yeast. *RNA N. Y. N* **14,** 666–674 (2008).

19. LOS1 | SGD. Available at: http://yeastgenome.org/locus/S000001688/overview. (Accessed: 12th April 2016)