The Effect of the Polymorphic (AT)n Dinucleotide Repeat Length on the Fragility of the Flex1 Region within the Human Common Fragile Site FRA16D

Soo-Mi Alison Lee
Senior Honors Thesis for the Department of Biology
Tufts University, 2011

Advisor: Dr. Catherine Freudenreich
Committee: Dr. Sergei Mirkin

ACKNOWLEDGEMENTS

First, I would like to thank my advisor Dr. Catherine Freudenreich and my thesis committee member, Dr. Sergei Mirkin, for their guidance and advice throughout this project.

I would also like to thank my research mentor, Nealia House and to the members of the Freudenreich laboratory: Melissa Koch, Lionel Gellon, Ranjith Anand, Adam Snider, Lucas Schlager, Laura Sloan, Jennifer Nguyen, Scott Shuldiner, Allen Su, Robyn Jong, Jordana Laks, Lauren Verra, and Kelly Beagan, for their helpful suggestions. Special thanks to Nealia House for help in constructing yeast strains and critical reading of the manuscript; to Melissa Koch for many helpful conversations related to this project and helpful suggestions for the manuscript; to Adam Snider for help in Flex1 cloning.

Finally, I would like to thank my family and friends, for their unwavering support and encouragement throughout my thesis project.

TABLE OF CONTENTS

ABSTRACT

Common fragile sites are chromosomal regions that are susceptible to breakage. The second most highly expressed common fragile site, FRA16D, is located within the WWOX tumor suppressor gene and has been associated with chromosomal instability in various cancers. Flex1 is a ~500 bp AT-rich region located within FRA16D and contains a polymorphic (AT)n dinucleotide repeat. The Flex1 region has been shown to increase chromosome fragility in a yeast model system. We used a direct repeat recombination assay in *S. cerevisiae* to examine the effect of the length of the perfect AT repeat embedded within the Flex1 region on chromosome breakage. We observed an increase in the rate of chromosome breakage that is dependent on the length of the perfect AT repeat. However, for the Flex1 sequences that lack 102 bp of the 3'-end sequence that is predicted to easily extrude to form a 17 bp hairpin, the rate of breakage was significantly lower. These results indicate that the 17 bp hairpin-prone sequence strongly inhibits recovery of broken chromosomes in the assay. Though the exact mechanism of inhibition is unknown, a potential mechanism is that the 17 bp hairpin interferes with resection at a double stranded break needed to expose complementary single stranded DNA for homologous recombination. The main finding of this study is that there is a correlation between the rate of chromosome breakage and the secondary structure-forming propensities among the perfect AT repeats within the AT-rich Flex1 sequence. Therefore, we propose that a hairpin (from a single-stranded DNA) and/or cruciform (from dsDNA) structure formed by the AT repeat in the Flex1 region contributes to chromosome breakage.

INTRODUCTION

The maintenance of genomic stability is essential for normal cellular function and cell viability. Loss of genomic stability is causally associated with a number of human diseases including cancers. A major threat to genome integrity is DNA double-strand breaks (DSBs) because, if repaired improperly, DSBs can result in mutations or other DNA instability events such as rearrangements, translocations, and cell death (Aguilera & Gómez-González, 2008; Negrini et al., 2010). Common fragile sites are chromosomal regions that are especially susceptible to breakage under conditions that partially inhibit DNA replication. These regions are observed as non-random gaps or breaks within metaphase chromosomes under conditions of replication stress, such as in the presence of low doses of aphidicolin, a specific inhibitor of eukaryotic DNA polymerase α (Figure 1A). Common fragile sites are highly unstable regions of the genome, and have been found to be hotspots for increased sister chromatid exchanges (Figure 1B) (Glover & Stein, 1987) and preferential sites of viral integration (Ferber et al., 2003; Popescu et al., 1990). Despite their instability, common fragile sites are found on the chromosomes of all individuals,

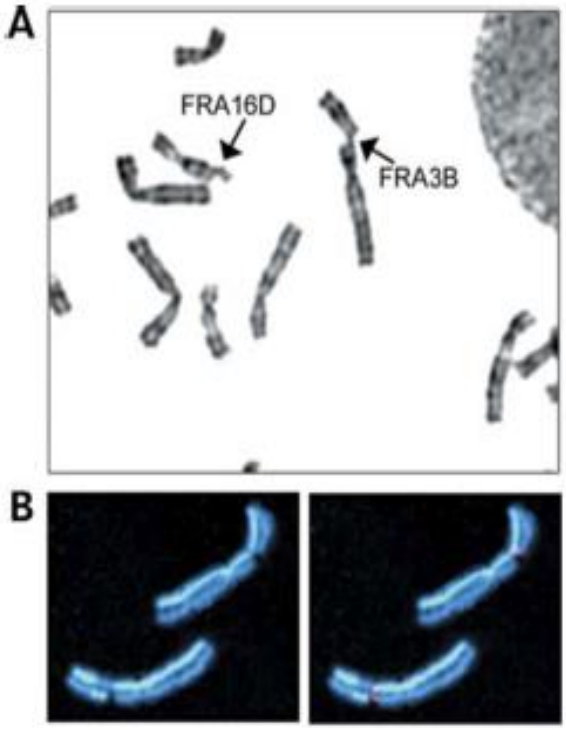


Figure 1. Examples of Common Fragile Sites
(A) Breaks at fragile sites FRA3B and FRA16D in metaphase chromosomes. (B) Homologs of chromosome 3 with breaks at FRA3B stained to show sister chromatid exchanges (Glover et al., 2005).

and are thus considered to be a normal component of chromosome structure (Glover et al., 2005). Furthermore, they are conserved across a wide range of species, which implies that these sequences are important for a still unknown function (Glover et al., 1998). It has been proposed that conserved AT-rich high-flexibility regions are characteristic of common fragile sites (Matsuyama et al., 2003). To date, over 80 common fragile sites have been identified in the human genome, and the 20 most readily expressed common fragile sites account for more than 80% of DNA breakage following treatment with low doses of aphidicolin (Glover et al., 1984). Common fragile site FRA3B located at chromosome region 3p14.2 is the most highly expressed and other highly expressed common fragile sites include: FRA16D at 16q23, FRAXB at Xp22.3, FRA6E at 6q26, and FRA7H at 7q32.3 (Glover et al., 2005). It is known that specific environmental and dietary factors (e.g. caffeine, ethanol, and tobacco), and other stress factors (e.g. hypoxia) can induce common fragile sites (Richards et al., 2008; Gandhi et al., 2010).

Common Fragile Sites are Associated with Genomic Instability in Tumorigenesis

Recurrent chromosomal deletions, translocations, and loss of heterozygosity that are hallmarks of particular kinds of cancer have been shown to be located near or within common fragile site loci (Yunis & Soreng, 1984; Huebner & Croce, 2001; Richards, 2001; Dillon et al., 2010). Over half of all known cancer-specific recurrent translocation breakpoints have been mapped to common fragile sites (Burrow et al., 2009). For example, homozygous deletions at the second most readily induced common fragile site, FRA16D, have been detected in adenocarcinomas of the lung, breast, ovary, and colon (Paige et al., 2000). In addition, translocation breakpoints have been mapped within

FRA16D in ~25% of cases of multiple myeloma, and breakpoints for four out of five of these cases are located within the FOR tumor suppressor gene spanning FRA16D (Chesi et al., 1998; Mangelsdorf et al., 2000; Krummel et al., 2002). Other fragile sites, FRA3B and FRA7G, have also been found to be associated with chromosomal instability in various cancers including epithelial ovarian cancer and lung cancer (Huebner et al., 1998; Huang et al., 1999). For example, homozygous deletions at the most highly expressed common fragile site, FRA3B, is extremely common in carcinomas of the lung, stomach, kidney, and cervical carcinomas, especially in those exposed to highest levels of environmental mutagens, and such deletions often inactivate the FHIT gene spanning FRA3B (Huebner et al., 1998). These results suggest that common fragile sites are indeed highly susceptible to DNA instability in cancer. Moreover, a recent study provides direct evidence linking breakage at common fragile sites to the formation of a cancer-specific rearrangement in human cells. Gandhi et al. (2010) showed that when human thyroid cells are exposed to fragile site-inducing chemicals, DNA breaks occur within fragile sites FRA10C and FRA10G, leading to the generation of RET/PTC1 rearrangement, which is known to contribute to papillary thyroid carcinoma development. These results are supportive of a causative role for common fragile site-associated genomic instability in cancer, rather than a consequence of widespread genomic instability after tumor progression. In addition, common fragile sites are often located near or within tumor suppressor genes, so it has been proposed that disruption of tumor suppressor genes at common fragile sites could directly contribute to cancer progression (Hubner & Croce, 2001; O'Keefe & Richards, 2006). For example, the WW domain-containing oxidoreductase (WWOX) gene encodes a tumor suppressor that has a role in apoptosis and is often inactivated by deletions at FRA16D, resulting in loss of

WWOX/FOR function, and thus leading to tumorigenesis (Finnis et al., 2005). Studies have suggested that common fragile sites predispose to specific chromosomal breakage associated with DNA instability (e.g. deletions, rearrangements, or translocations) in certain forms of cancer (reviewed in Mangelsdorf et al., 2000).

AT-Rich Regions are Characteristic of Common Fragile Sites: Implications for the Mechanism of Common Fragile Site Induction

Common fragile sites can extend over hundreds to thousands of kilobases of DNA with gaps or breaks occurring throughout, and the specific sequence elements that are responsible for fragility of common fragile sites have not yet been completely elucidated (reviewed in Shah et al., 2009). However, common fragile sites contain regions of high flexibility, termed 'flexibility peaks,' which are highly AT-rich as determined by a computer program that predicts the flexibility of the DNA helix based on the twist angle between consecutive base pairs (Richards et al., 2008; Finnis et al., 2005). Studies have shown that the flexible sequences are composed of interrupted runs of AT-dinucleotides (Zlotorynski et al., 2003; Toledo et al., 2000; Dillon et al., 2010). Flexible DNA regions within common fragile sites often coincide with the location of mapped deletion breakpoints in cancer cells, suggesting their role in common fragile site expression. However, many studies have demonstrated that the flexible regions are not necessary for expression of common fragile sites (reviewed in O'Keefe & Richards, 2006), and another study disputes that common fragile sites are enriched for highly flexibility peaks (Helmrich et al., 2007). Instead, the flexible regions have been proposed to be hotspots of chromosomal fragility because AT-repeat sequences are predicted to form secondary DNA structures that could impede DNA replication, thus causing

fragility (Zhang & Freudenreich, 2007; Burrow et al., 2010; Palakodeti et al., 2010; reviewed in Richard et al., 2008). It is widely believed that DNA sequences that can adopt unusual secondary structures such as the AT repeats at FRA16D are associated with hotspots of genomic instability because such structures *in vivo* can cause replication fork stalling and double stranded DNA breaks (reviewed in Richard et al., 2008; reviewed in Voineagu et al., 2009). For example, DNA hairpins stall replication fork progression in bacteria, yeast, and mammalian cells, causing chromosomal fragility (Voineague et al., 2008; Krasilnikova & Mirkin, 2004). Furthermore, DNA sequences that can form stable structures have been shown to induce various forms of genomic instability, such as gross chromosomal rearrangements, and chromosomal breaks both in humans and in model organisms including yeast and bacteria (reviewed in Voineague et al., 2009). These results suggest that genomic instability at common fragile sites is driven by the ability to form unusual DNA structures.

Common Fragile Sites are Late-Replicating Regions of the Genome

A feature of common fragile sites is that they are late-replicating regions of the genome (reviewed in Freudenreich, 2007). For example, sequences at FRA3B replicate very late and the addition of aphidicolin further delays replication with ~16.5% of FRA3B sites remaining unreplicated in G2 phase (Le Beau et al., 1998). The delayed replication at rare fragile sites have been explained by the observation that structure-prone expandable repeats such as (GAA)n repeats (Krasilniokva & Mirkin, 2004), and (CGG)n•(CCG)n and (CAG)n•(CTG)n (Samadashwily et al., 1997) found at these sites form secondary structures that slow progression of replication forks in bacterial, yeast, and mammalian cells. These results suggest that other fragile site sequences that can

adopt secondary structures may be late replicating regions because of the formation of secondary structures that slow or block progression of replication forks. Studies have shown that the common fragile sites FRA16D and FRA7H also replicate late because they experience difficulty in replication fork progression (Palakodeti et al., 2004; Hellman et al., 2000). These studies support a model in which common fragile site regions initiate replication normally, but are slow to complete replication, thus leading to unreplicated, single stranded DNA regions that give rise to the breaks and gaps observed in metaphase chromosomes. Further supporting this model is the finding that cells deficient for the replication checkpoint kinase, Ataxia-Telangiectasia and Rad3-Related (ATR), whose function is to stabilize stalled replication forks, show a significant increase in gaps and breaks at fragile sites (Casper et al., 2002). In contrast to the fork barrier hypothesis, it has been proposed that the fragility of common fragile sites does not depend on fork slowing or stalling but on a paucity of initiation events in the core of common fragile sites. This study has shown that in lymphoblastoid cells, the FRA3B core region is deficit in initiation events, and supports a model in which common fragile sites are the latest initiation-poor regions to complete replication in a given cell type, which leads to incompletely replicated common fragile sites that result in DNA breaks (Letessier et al., 2011).

DNA Damage Checkpoint Responses at Common Fragile Sites & Model for the Mechanisms of Instability at Fragile Sites

Studies have been conducted to understand how DNA damage checkpoint responses function at structure-prone regions that are associated with genomic instability (reviewed in Voineagu et al., 2009). The DNA damage response is a signal-transduction

cascade that is composed of two major branches in *S. cerevisiae*: one pathway is dependent on Tel1 (mammalian ATM kinase) and responds mainly to double-stranded breaks such as those induced by ionizing radiation (reviewed in Lavin & Kozlov, 2007), and another pathway is dependent on Mec1 (mammalian ATR kinase) and is activated primarily to replication fork stalling such as stalls caused by aphidicolin, hydroxyurea, and hypoxia (reviewed Cimprich & Cortez, 2008; reviewed in Glover et al., 2005). The study by Casper et al. (2002) demonstrated that ATR-deficient cells show increased instability at common fragile sites with and without the addition of aphidicolin, suggesting that ATR is necessary for the maintenance of stability at common fragile sites both during partially inhibited and normal DNA replication (Casper et al., 2002; reviewed in Glover et al., 2005). Thus, this finding has confirmed that the ATR checkpoint pathway is associated with control of common fragile site expression, and has linked cell-cycle checkpoint function with fragile site stability.

The model for instability at common fragile sites predicts that unreplicated, single stranded regions that arise from delayed or stalled replication at these sites become coated with RPA, which activates the ATR pathway (Figure 2) (Glover et al., 2005). Then ATR is recruited by its cofactor ATRIP, which recognizes RPA-coated single stranded DNA. It requires further activation by the replication processivity clamp-like complex, Rad17, Mec3, Ddc1 (mammalian Rad9, Rad1, and Hus1), which is loaded by the clamp loader-like complex Rad24-RFC (mammalian Rad17-RFC). Then, ATR kinase phosphorylates the effector kinase CHK1, which in turn phosphorylates downstream target proteins to trigger checkpoint responses: cell cycle arrest, replisome stabilization, and inhibition of replication origin firing (reviewed in Aguiléra et al., 2008; reviewed in Voineagu et al., 2009).

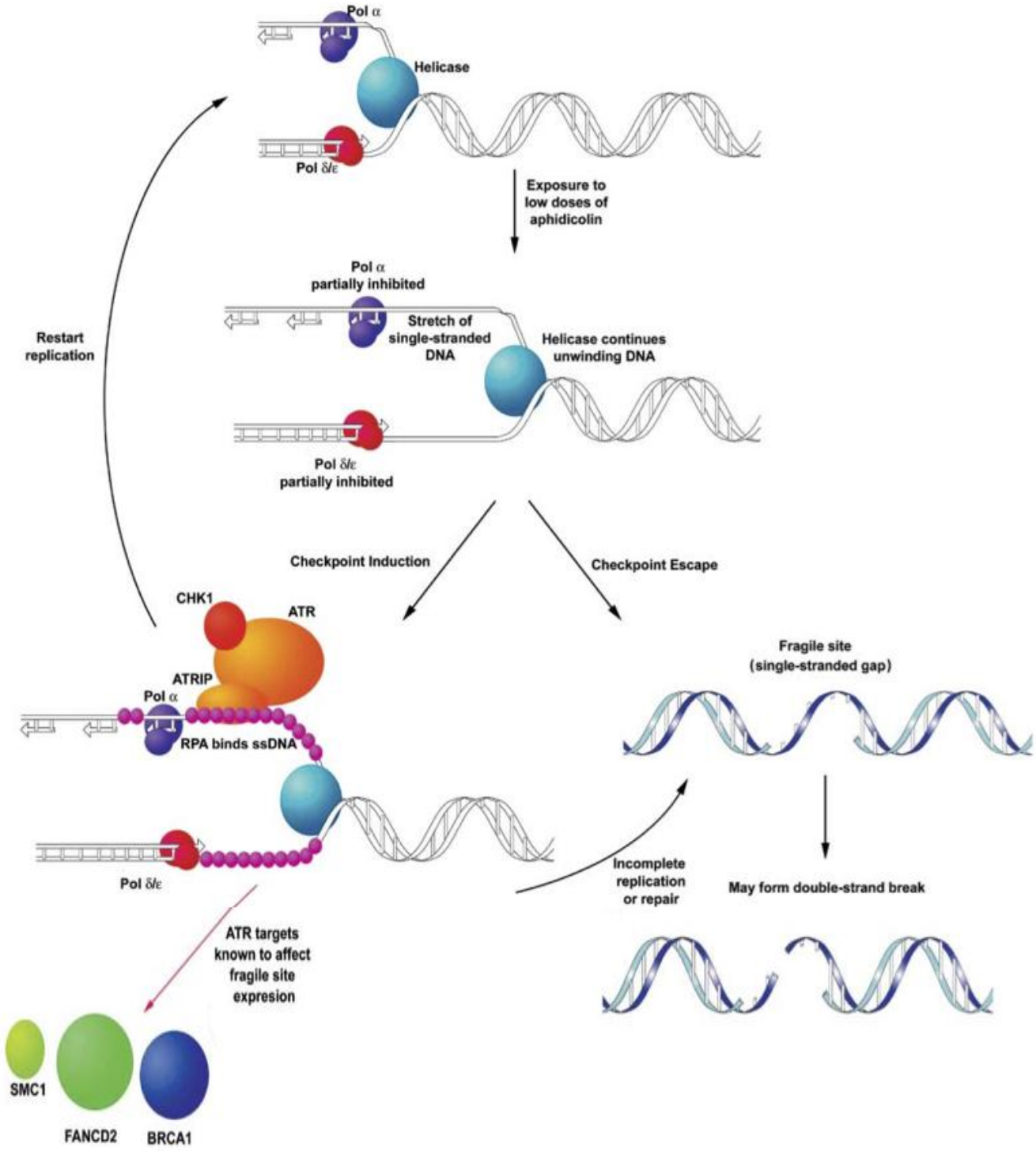


Figure 2. Model for Instability at Common Fragile Sites
Shown are the two pathways, either checkpoint activation or checkpoint escape, which can result from unreplicated, single stranded DNA that arise from delayed or stalled replication at common fragile sites. Checkpoint escape will result in increased fragile site instability and checkpoint activation will lead to maintenance of fragile site stability (Glover et al., 2005).

Subsequent investigations have identified a number of targets of the ATR pathway that influence fragile site fragility, including BRCA1, SMC1, and the Fanconi anemia (FA) pathway proteins (reviewed in Glover et al., 2005). BRCA1 has a number of important functions in the DNA damage response including roles in checkpoint and repair processes; SMC1 is a member of a protein family that has crucial roles in chromosome condensation, DNA repair, and sister chromatid cohesion; The FA pathway protein functions are unknown, but they are thought to be involved in the repair of unusual DNA structures (Sobeck et al., 2007). Some types of DNA damage that occurs at common fragile site sequences may escape checkpoint control, resulting in increased fragile site instability (e.g. chromosomal breakage), whereas other types of DNA damage may activate checkpoint control more efficiently, resulting in maintenance of fragile site stability (reviewed in Voineagu et al., 2009).

A ~500 bp AT-Rich Region (Flex1) within the Common Fragile Site FRA16D Increases Chromosome Fragility

The second most highly expressed common fragile site, FRA16D, is contained within a 270 kb region of chromosome 16q23.2, which is located within intron 8 of the WWOX gene (Shah et al., 2010). Flex1 is a ~500 bp AT-rich region within FRA16D and contains a perfect (AT)n dinucleotide repeat that is predicted to easily extrude to form secondary structures. Using a yeast artificial chromosome (YAC) breakage assay in yeast, Zhang and Freudenreich (2007) demonstrated that the Flex1 sequence significantly increase chromosome fragility, which was further enhanced by either addition of a replication inhibitor or absence of the DSB repair protein, Rad52. The YAC breakage assay allows a quantitative measure of the amount of breakage of Flex1

sequences by utilizing a YAC that contains the Flex1 sequence, the *URA3* and *LEU2* genes as selectable markers at each end, and a telomeric seed sequence (C_4A_4) to rescue broken YACs (Figure 3).

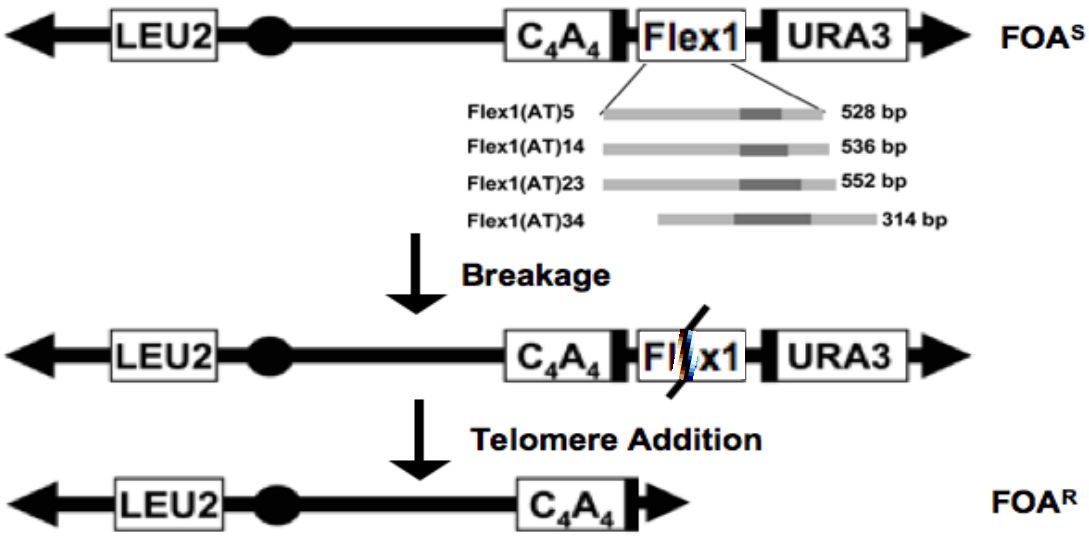


Figure 3. Schematic Diagram of the YAC Breakage Assay
Cells containing a YAC with the Flex1 sequences with different numbers of AT repeats (AT-5, -14, -23, or -34) and the *URA3* gene are FOA^S. If breakage occurs inside the Flex1 region, the broken YAC can be rescued by *de novo* telomere addition to the C_4A_4 telomere seed sequence. In this process, the cells become FOA^R due to the elimination of the *URA3* gene (Zhang & Freudenreich, 2007).

When breaks occur at or near the Flex1 sequence and fail to be healed by a normal DSB repair pathway, exonucleases will degrade broken ends of breakage intermediates to expose the telomeric seed sequence, and telomere addition by telomerase will result in rescue of the broken YAC. In this process, the *URA3* gene is lost and cells are converted from 5-floroorotic acid (FOA) sensitive (FOA^S) to FOA resistant (FOA^R). Thus, a rate of FOA^R correlates with a rate of Flex1 breakage.

The Flex1 region within the human common fragile site FRA16D is highly AT rich (65-75%) and contains sequences flanking a frequently deleted region in the

HC1116 tumor cell line. In addition, Flex1 contains a (AT)n dinucleotide repeat that shows a high degree of polymorphism of repeat number in the human population. Thus, Flex1 with different number of AT repeats (AT-5, -14, -23, -34) were tested for their ability to induce chromosome fragility. Flex1(AT)5, Flex1(AT)14, and Flex1(AT)23 showed increased chromosome breakage in a manner dependent on the length of the (AT)n dinucleotide repeat compared to the 386 bp no repeat control sequence. Interestingly, Flex1(AT)34 showed a significantly decreased chromosome breakage compared to the control and the other Flex1 sequences with shorter AT repeats. It was predicted that repeats of (AT)21-34 in an AT-rich region such as Flex1 can form a stable secondary structure *in vivo*, whereas lower numbers of AT repeats do so much less efficiently. Thus, Flex1 sequences with longer AT repeats, Flex1(AT)23 and Flex1(AT)34), were predicted to form more stable secondary structures than those with shorter AT repeats, Flex1(AT)5 and Flex1(AT)14). The Flex1(AT)34 can form a hairpin (Tm=67°C), and a second 17 bp hairpin (Tm=52.7°C) 23 bp away, as predicted by the Mfold program (Zuker, 2003). Therefore, the authors hypothesized that the Flex1(AT)34 sequence breaks at a higher rate but forms a stable secondary structure that interferes with the exonuclease activity to expose the telomeric seed sequence for telomere addition, resulting in poor recovery of broken YACs. If this is the case, it was predicted that the breakage rate of the Flex1(AT)23 is also underestimated.

Goals of this Project

The ultimate goal of this project was to determine the effect of the length of the AT repeat on the FRA16D Flex1 fragility by using a different assay for chromosome fragility. The first major objective was to determine whether a direct repeat

recombination assay overcomes the problem of the more structure-prone Flex1 sequences interfering with the exonuclease activity to expose the telomeric seed sequence in the assay that the authors used (Zhang & Freudenreich, 2007). The second objective was to examine the effect of 5' and 3' sequences flanking the AT repeat on chromosome fragility. The Flex1 sequences with different numbers of AT repeats have different 5'- and 3'-flanking sequences (Figure 11). Therefore, it was important to account for the influence of this difference on chromosome fragility in order to accurately determine the effect on fragility solely from the AT repeats. To test this, yeast strains containing Flex1 sequences with various combinations of 5'- and 3'-flanking sequences were constructed (Table 2 & Figure 14). The third goal was to determine whether increasing the AT-repeat copy numbers has an additive effect on fragility. Since FRA16D has been shown to contain several AT-rich sequences and perfect AT repeats very close to the Flex1 AT-rich sequence, it was important to examine the effect of multiple copies of AT-rich sequences on chromosome fragility. To test this, yeast strains containing Flex1 sequences with multiple copies of AT repeats were constructed.

MATERIALS AND METHODS

Yeast Strains

All strains are in CHF yPH499 (stock #2268), which is a derivative of yBL3100 background (*MAT* a, *ura3-52, lys2-801amber, trp1-Δ63, his3-Δ200, and leu2-Δ1*). The pBL007 vector containing Flex1(AT)n insert was linearized within the *ADE2* sequence with XbaI and was then integrated into the *ADE2* locus on chromosome II in the yPH499 strain.

Preparation of Flex1(AT)n Inserts for Cloning

The Flex1(AT)n fragments were previously cloned into the EcoRI site of the pYES2 vector (Figure 4). Bacteria containing the plasmids were streaked onto LB Amp plates (100 µg/ml) and grown overnight at 37°C. Each colony from the plate was inoculated into 2 ml of liquid LB Amp media (100 µg/ml) and grown overnight at 37°C with shaking. A Zippy™ plasmid miniprep kit (Zymo Research) was used to isolate the plasmids from the bacteria according to the manufacturer's instructions. The isolated plasmids were digested with EcoRI in a 20 µl volume, containing 150 ng of plasmid DNA, 10 units of EcoRI, and 1X EcoRI buffer to excise Flex1(AT)n fragments from the plasmids. This mixture was incubated at 37°C overnight and then incubated at 65°C for 20 minutes to inactivate EcoRI. Next, the digested DNA were separated on 1.5% agarose gels at 80V

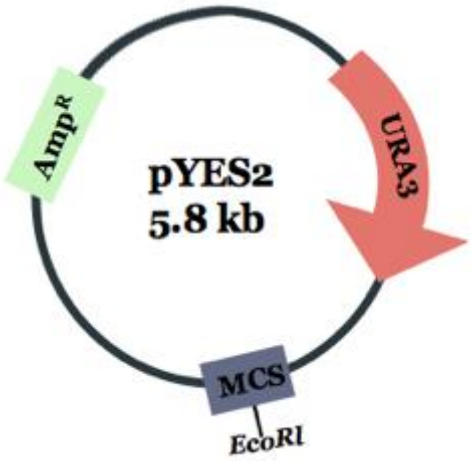


Figure 4. Schematic for pYES2 Vector
The vector contains a *URA3* gene and an ampicillin resistance gene. The Flex1(AT)n fragments that were previously cloned into the EcoRI site in the multiple cloning site (MCS) of the pYES2 vector were excised by treatment with EcoRI.

for 1 hour to isolate the Flex1(AT)n inserts from undigested vector. The Flex1(AT)n fragments were purified from the gel using the Axygen Gel Clean-Up Kit according to the manufacturer's instructions.

Preparation of the pBL007 Vector for Cloning

The pBL007 vector (Figure 5) was digested with EcoRI in a 20 µl volume, containing 150 ng of plasmid DNA, 10 units of EcoRI, and 1X EcoRI buffer. The mixture was incubated at 37°C overnight and then incubated at 65°C for 20 minutes to inactivate EcoRI . Next, to isolate digested pBL007 vector from undigested vector, the digested DNA were fractionated on 1.0% agarose gels at 80V for 1 hour. Then, the EcoRI-linearized pBL007 vector was purified from the gel using Axygen Gel Clean-Up Kit according to the manufacturer's instructions. To reduce the ability of the EcoRI-digested vector to re-circularize without incorporating the Flex1(AT)n inserts, the digested vector was dephosphorylated in a 15 µl volume, containing 200 ng of EcoRI-digested pBL007 vector, 1 unit of shrimp alkaline phosphatase (SAP), and 1X SAP buffer. The mixture was incubated at 37°C for 1 hour, and then an additional 1 µl of SAP was added, followed by incubation at 37°C for an additional 30 minutes. The mixture was then incubated at 65°C for 20 minutes to inactivate SAP.

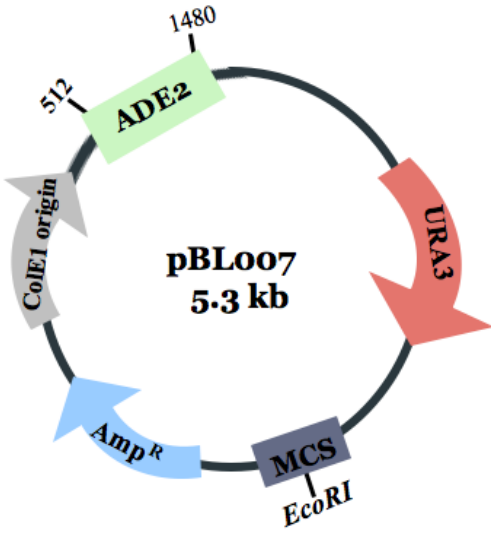


Figure 5. Schematic for pBL007 Vector The vector contains an ampicillin resistance gene, a *URA3* gene, and a fragment of the *ADE2* gene (nts 512-1480). The Flex1(AT)n fragments were cloned into the EcoRI site in the multiple cloning site (MCS) of the pBL007 vector.

Cloning of Flex1(AT)n sequences

Ligations were performed in a 20 µl volume, containing 1 Weiss unit of DNA ligase, 1X ligation buffer, and a 3:1 molar ratio of insert to vector. A control ligation was performed using de-phosphorylated pBL007 vector, DNA ligase, and ligation buffer, but without the Flex1(AT)n insert. The mixtures were incubated at 4°C overnight.

The ligated DNA was transformed into DH5α cells by heat shock. Between 5:1 and 7:1 volumetric ratios of competent cells to ligated DNA were used. First, 50-70 µl of competent cells were incubated on ice with 10 µl of ligated DNA for 30 minutes, then heat shocked at 42°C for 30 seconds, followed by incubation on ice for 5 minutes. The cells were allowed to recover in 900 µl of nonselective LB liquid media at 37°C for 1 hour with shaking. After recovery, 150-200 µl of the transformed cells were spread onto LB-agar plates containing 100µg/ml ampicillin and cultured at 37°C for 1 day to select for cells containing the pBL007 plasmid.

Verification of Bacterial Transformation

Bacterial transformants were subjected to PCR screening and sequencing to find correct transformants. First, colonies that grew on LB Amp plates were cultured in 2 ml LB broth at 37°C overnight, and the plasmids were purified using a Zippy^{TM} plasmid miniprep kit (Zymo Research). Then, the isolated plasmids were subjected to Taq Colony PCR amplification with a forward primer (primer #679) ~500 bp upstream of the multiple cloning site (MCS) of pBL007 vector and a reverse primer (primer #680) ~500 bp downstream of the MCS. The Taq Colony PCR reaction protocol contained 4.25 µl diH_2O, 2.5 µl 5X buffer (w/o $MgCl_2$), 0.75 µl 25mM $MgCl_2$, 1.25 µl 10 pmol/µl of the

forward primer (primer #679), 1.25 µl 10 pmol/µl of the reverse primer (primer #680), 0.25 µl 10mM dNTPs, 0.25 µl Phoenix Taq polymerase, and 50 ng of plasmid DNA. The PCR cycles were as follows: heat at 94°C for 1 min, 35 cycles at 94°C for 20 sec, at 54°C for 1 min, and at 68°C for 3 min, followed by a final incubation at 68°C for 3 min. The PCR products were subjected to electrophoresis in 1.5% agarose gels at 80V for 1 hour. In addition, bacterial transformants that generated a correct PCR band size were sequenced using a reverse primer that anneals 20 bp between SacI site on the 5'-end and EcoRI site on the 3'-end in the multiple cloning site of pBL007 vector (primer #876). All the correct bacterial transformants were stored in glycerol stock at -70°C (Appendix Table 1&2).

Preparation of Flex1 Fragments for Integrative Yeast Transformation

Plasmids were isolated from the verified bacterial transformants using a Zippy^{TM} plasmid miniprep kit (Zymo Research), and then digested with XbaI in a 20 µl volume, containing 150 ng of plasmid DNA, 10 units of XbaI, 1X BSA, and 1X NEBuffer 4. The mixture was incubated at 37°C overnight and then incubated at 65°C for 20 minutes to inactivate XbaI. Next, to isolate digested vector from undigested vector, the digests were separated on 1.0% agarose gels at 80V for 1 hour. Then, the XbaI-linearized plasmids were purified from the gel by Axygen Gel Clean-Up Kit according to the manufacturer's instructions.

Integrative Yeast Transformation of the Flex1 Sequences

The XbaI-linearized pBL007 plasmids were transformed into yPH499 strains following a high-efficiency version of the lithium acetate/single-stranded carrier

DNA/PEG method of transformation of *S. cerevisiae* (Gietz et al., 2007) with the following modifications: The overnight culture of *ADE2*$^{+}$ yPH499 yeast strain grown in YEPD liquid media was reinoculated to an OD_{600} of ~0.35 and incubated in the shaking incubator at 30°C at 225 rpm until the OD_{600} reached ~1.5. The transformation mix contained 240 µl PEG 3350 (50% (w/v)), 36 µl LiAc (1.0M), 50 µl single-stranded carrier DNA (2.0 mg ml^{-1}), and 1.2 µg plasmid DNA. Yeast strains were heat shocked at 42°C for 20 minutes and they were allowed to recover in 1mL YEPD liquid media at 30°C for 3 hours. Subsequently, 50, 100, 150, and 200 µl of the cell suspension were plated onto YC-Ura plates, in order to select for yeast that integrated *URA3* gene.

Verification of Integration of Flex1(AT)n at *ADE2* Locus in Yeast

Integration of Flex1(AT)n fragments was verified by PCR using a forward primer (primer #878) that anneals 22-42 bp upstream of *LYS2* start codon and a reverse primer (primer #879) that anneals 16-36 bp downstream from *URA3* start codon. The Taq Colony PCR reaction contains 8.25 µl distilled H_2O, 1.25 µl 10X buffer (w/ $MgCl_2$), 1.25 µl 10 pmol/µl of the forward primer, 1.25 µl 10pmol/µl of the reverse primer, 0.25 µl 10mM dNTPs, 0.25 µl Phoenix Taq polymerase, and a small section of the yeast colony. The expected size of a PCR product of a correct transformant was ~3.0 kb.

Direct Repeat Recombination Assay

The direct repeat recombination assay was performed on each yeast strain in order to measure the amount of breakage of the Flex1 sequences (Figure 6). Yeast strains with Flex1(AT)n inserted at the LYS2 locus of chromosome II were grown to single colonies on non-selective YEPD plates at 30°C for 72 hours. Then, ten colonies

were randomly chosen from each strain and the entire colony was resuspended in 400 µl dH_2O and each suspension was split into two fractions. The first fraction was used to determine the number of cells in a colony: 10 µl from each 400 µl of colony suspension was diluted to 10^{-5} and 100 µl of the 10^{-5} dilution was plated on a YEPD plate and the number of cells on that plate after 3-4 days at 30°C was multiplied by 10^5 to determine the number of cells in a quarter of a colony. The second fraction of each colony was used to determine the number of cells in that specific colony that underwent recombination between the *ADE2* direct repeats. Since only cells that lost *URA3* and regained *ADE2* could grow on FOA (1g/L) plates lacking adenine, 10 µl from the remaining 400 µl of colony suspension was diluted in 190 µl dH_2O (1:20 dilution) and the entire volume of 200 µl was plated onto FOA-Ade plates such that between 50 and 500 colonies of approximately 2-3 mm in diameter would grow after 4-5 days at 30°C. The number of cells that grew on selective plates was multiplied by 10 to determine the number of mutants in a quarter of a colony. Next, the method of maximum likelihood was used to determine the rate of FOA^R as a measure of recombination rate using fluctuation analysis calculator (FALCOR) (Hall et al., 2009).

RESULTS

Direct Repeat Recombination Assay to Study Fragility of the FRA16D Flex1 Region in Yeast

The first objective was to test the hypothesis generated by Zhang and Freudenreich (2007) that the Flex1(AT)23 and Flex1(AT)34 breakage rates were underestimated because the (AT)23 and (AT)34 repeats form DNA structures that inhibit *de novo* telomere addition and recovery of broken YACs. To test this, we modified a direct repeat recombination assay that our lab previously used to analyze breakage of an expanded CTG repeat sequence (Freudenreich et al., 1998). The first goal was to construct yeast strains containing the same Flex1 sequences that had been used in the YAC fragility assay into the recombination assay. This assay utilizes the fact that a double-strand break between two direct repeats stimulates the rate of recombination between the repeats, with elimination of the intervening segments. Yeast strains were constructed with the Flex1 sequences and a *URA3* gene between two direct repeats of the *ADE2* gene (Figure 6). If breaks occur at or near the Flex1 region, homologous recombination between the duplicated segments of *ADE2* can occur, thereby restoring an intact *ADE2* gene and eliminating the intervening segment containing the *URA3* gene. The resulting cells will be *Ade*$^+$ and FOA^R, so recombination events were selected on FOA plates lacking adenine, since only cells that lost *URA3* and regained *ADE2* could grow on them. The direct repeat recombination assay does not require *de novo* telomere addition to rescue breakage intermediates, so it was predicted that the formation of DNA structures by Flex1(AT)23 and Flex1(AT)34 would not interfere with recovery of broken chromosomes. Therefore, the Flex1(AT)23 and Flex1(AT)34 breakage rates would be higher using this assay if Zhang and Freudenreich's hypothesis is correct.

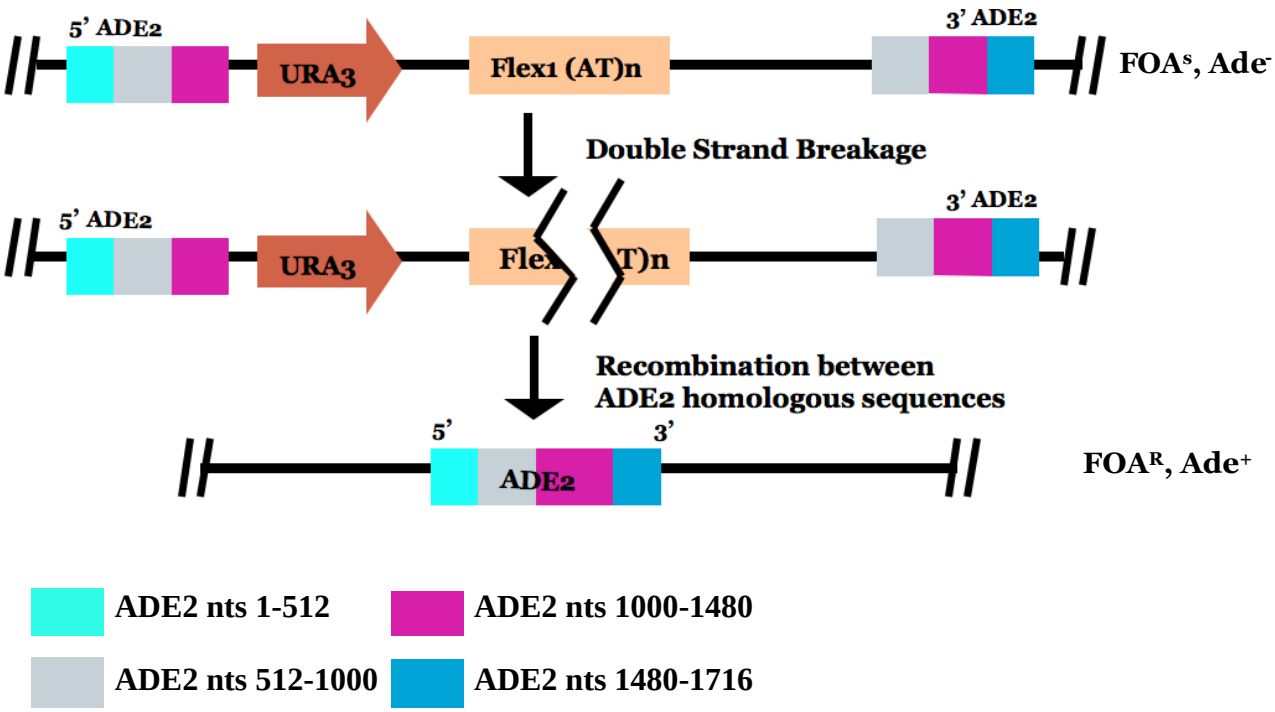


Figure 6. Schematic Diagram of the Direct-Repeat Recombination Assay
Structure of the modified *ADE2* locus is drawn approximately to scale. The *URA3* gene (1.17 kb) and the Flex1 sequence (~500 bp) were placed between a 1480-bp 5' fragment of the *ADE2* gene and a 1204-bp 3' fragment of the *ADE2* gene to create a 968-bp duplicated segments of *ADE2*. If breakage occurs in the 4.8-kb region between the duplicated segments of *ADE2*, recombination can occur between them, thereby restoring an intact *ADE2* gene and eliminating the intervening sequences. The resulting cells will be *Ade*$^+$ and FOA^R. Thus, the rate of FOA^R correlates with a rate of chromosome breakage.

Cloning the FRA16D Flex1 sequences into the pBL007 vector

In order to test the hypothesis generated by Zhang and Freudenreich, the initial goal was to clone the same Flex1 sequences that had been used in the YAC breakage assay into the direct repeat recombination assay. The first step was to clone these Flex1 sequences into the pBL007 vector, transform the ligated DNA into competent bacteria, and then verify correct transformants. The second step was to integrate the XbaI-linearized Flex1 sequences at the *ADE2* locus in yPH499 yeast strain (Figure 7).

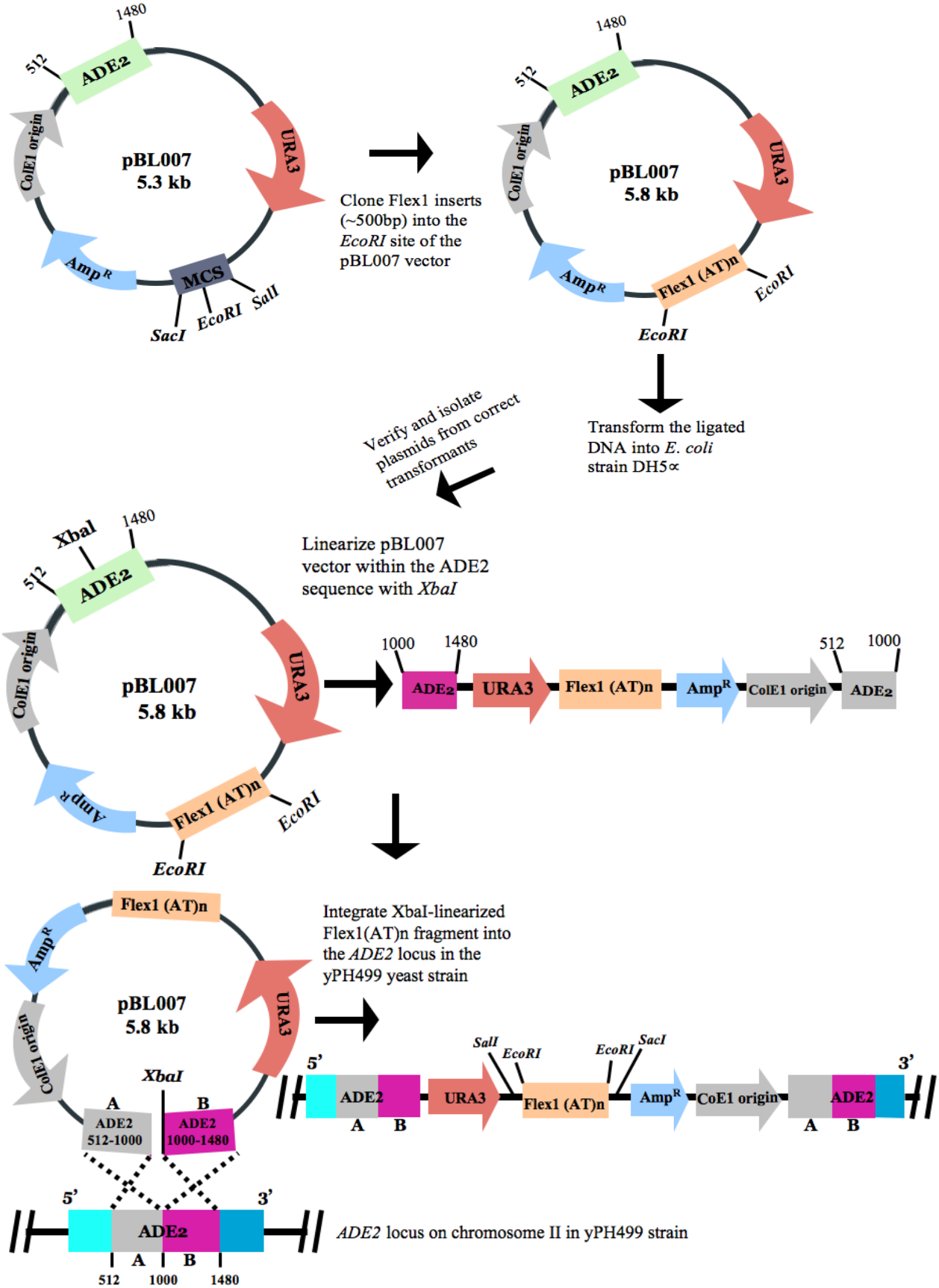


Figure 7. Schematic for Constructing Yeasts for the Recombination Assay Flex1 is cloned into the pBL007 vector and the ligated DNA is transformed into bacteria. The isolated plasmids from bacterial transformants are linearized with XbaI and then integrated into the *ADE2* locus on chromosome II in yPH499 yeast strain.

Flex1 sequences that were previously cloned into pBL007 vector were sequenced using a reverse primer that anneals 20 bp between SacI site on the 5'-end and EcoRI site on the 3'-end in the multiple cloning site of pBL007 vector (primer #876). The sequencing results confirmed three correct transformants (Figure 8). Interestingly, a transformant that was supposed to be Flex1(AT)34 had undergone a shortening of the AT-repeat length during transformation and became Flex1(AT)28. Since the Flex1(AT)23 clone was not obtained, Flex1(AT)28 was used instead. Also, it was not possible to obtain the 386 bp no-repeat control transformant because the 386 bp no-repeat control fragment in the pYES2 vector was not excised with EcoRI as predicted.

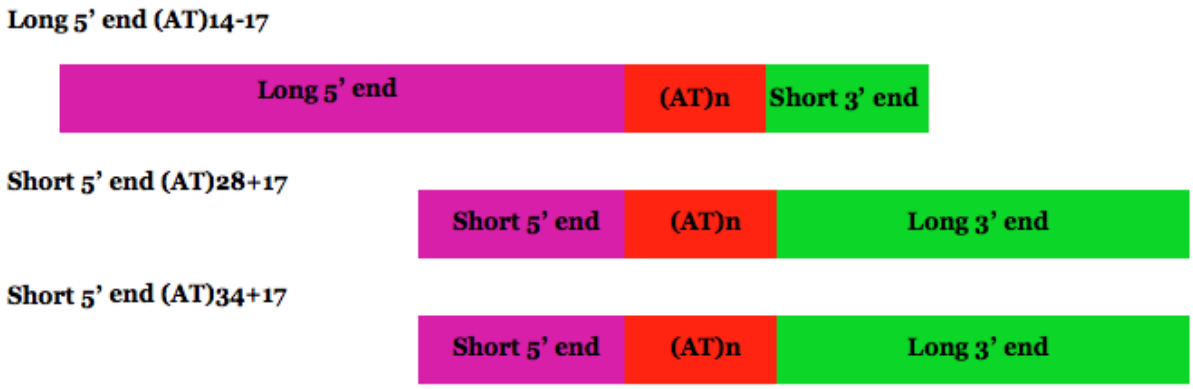


Figure 8. Sequencing Confirmed Three Correct Bacterial Transformants
The verified transformants were Flex1(AT)14, Flex1(AT)28, and Flex1(AT)34. Flex1(AT)28 and Flex1(AT)34 are exactly the same except for the AT repeat number: Flex1(AT)14 contains 362 bp of extra 5' end sequence and lacks 102 bp of the 3' end sequence. Thus, Flex1(AT)14 will be referred to as long 5' end (AT)14-17, whereas Flex1(AT)28 and Flex1(AT)34 will be referred to as short 5' end (AT)28+17 and short 5' end (AT)34+17, respectively.

Integrative Yeast Transformation of the Flex1 Sequences

Once correct bacterial transformants were identified, the next step was to linearize the isolated pBL007 plasmids within the *ADE2* fragment (nts 512-1480) with XbaI to target integration into the *ADE2* locus in yPH499 yeast. The transformed yeast

colonies were picked and screened by PCR using a forward primer (primer #878) that anneals 22-42 bp upstream of *LYS2* start codon and a reverse primer (primer #879) that anneals 16-36 bp downstream from *URA3* start codon (Figure 9A). The expected size of PCR products was ~3.0 kb for all the transformants in lanes 3-15. Therefore, the ~3.0 kb PCR products in lanes 3-15 indicated the presence of correct transformants (Figure 9B). Lane 2 contains a negative control, the *ADE2+* cells, the starting strain before transformation, so it did not give any bands as expected since it lacks the binding site for the reverse primer (primer #879).

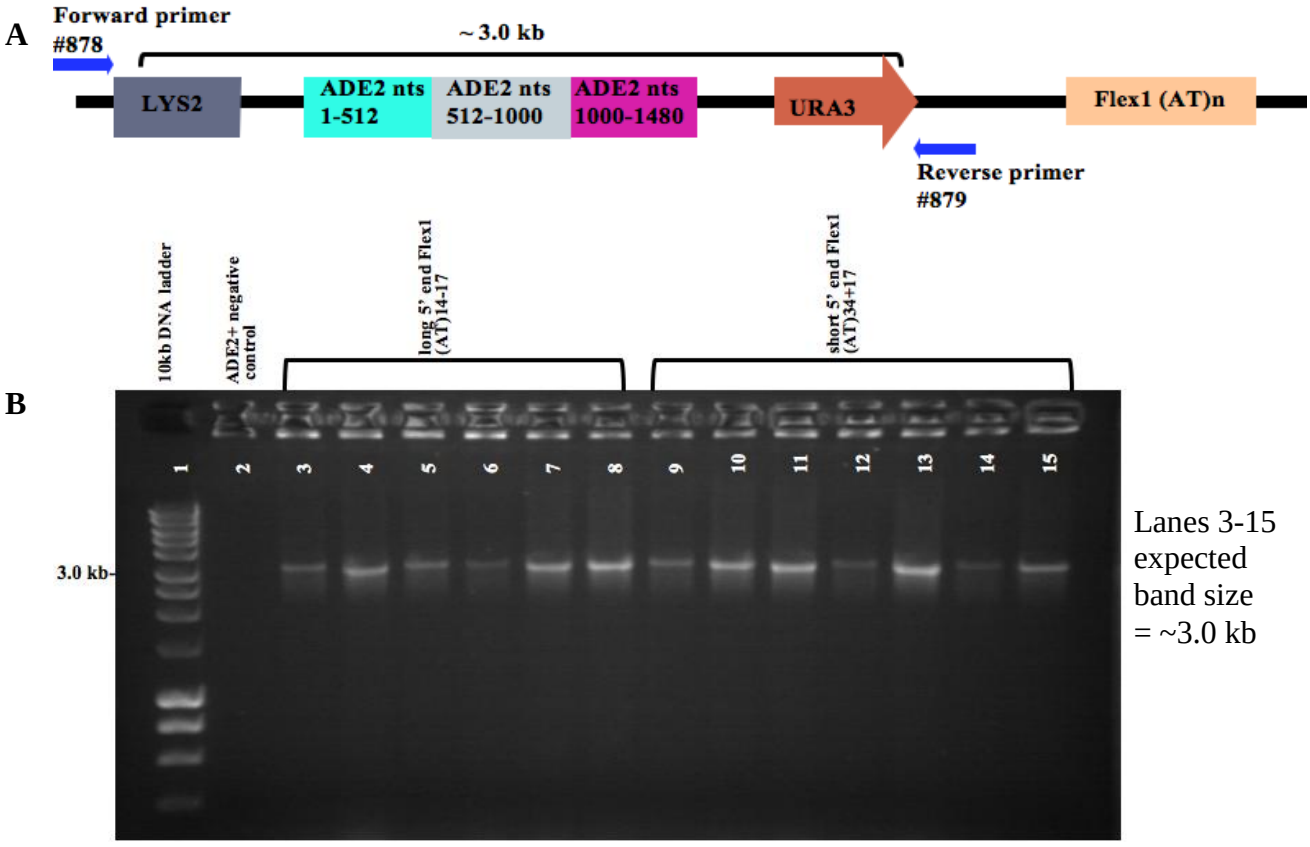


Figure 9. PCR Verification of Yeast Transformation
(A) Schematic for forward and reverse primers used to detect correct yeast transformants; the forward primer (primer #878) anneals 22-42 bp upstream of *LYS2* start codon and the reverse primer (primer #879) anneals 16-36 bp downstream from *URA3* start codon. Thus, the expected PCR product is ~3.0 kb.
(B) Lane 1: 10 kb DNA ladder
Lane 2: ADE2+ negative control (no PCR product expected)
Lanes 3-8: Long 5' end Flex1(AT)14-17 transformant
Lanes 9-15: Short 5' end Flex1(AT)34+17 transformant

Direct Repeat Recombination Assay

The next goal was to determine whether chromosome fragility increases with increasing numbers of Flex1 AT repeats. To do this, we used fluctuation analysis to determine the rate of recombination between the duplicated *ADE2* sequences in yeasts with three different Flex1 sequences with different numbers of AT repeats (AT-14, -28, -34). Two trials of the recombination assays were performed on the three transformants obtained: Flex1(AT)14, Flex1(AT)28, and Flex1(AT)34 (Figure 10). All three transformants showed an increase in the rate of FOA^R (the rate of chromosome breakage) compared to the no repeat control sequence, indicating that the Flex1 region is inherently fragile. Compared to the no repeat control, the rate was ~5-fold greater for Flex1(AT)14, ~2-fold greater for both Flex1(AT)28, and Flex1(AT)34. Interestingly, the rate of FOA^R of Flex1(AT)14 (~5-fold over control) was consistently greater than Flex1(AT)28 and Flex1(AT)34. This was similar to the trend seen in the study by Zhang and Freudenreich (2007). The fact that the rate of (AT)28 construct was similar to that of the (AT)34 construct and that both were lower than (AT)14 construct mean either (i) fragility was not influenced by the AT repeat length after all or (ii) the 5' and/or 3' flanking sequences were influencing fragility and thus the rate of recombination. Therefore, this result prompted us to test whether the 5' and/or 3' sequences flanking the AT repeat have any effect on chromosome fragility.

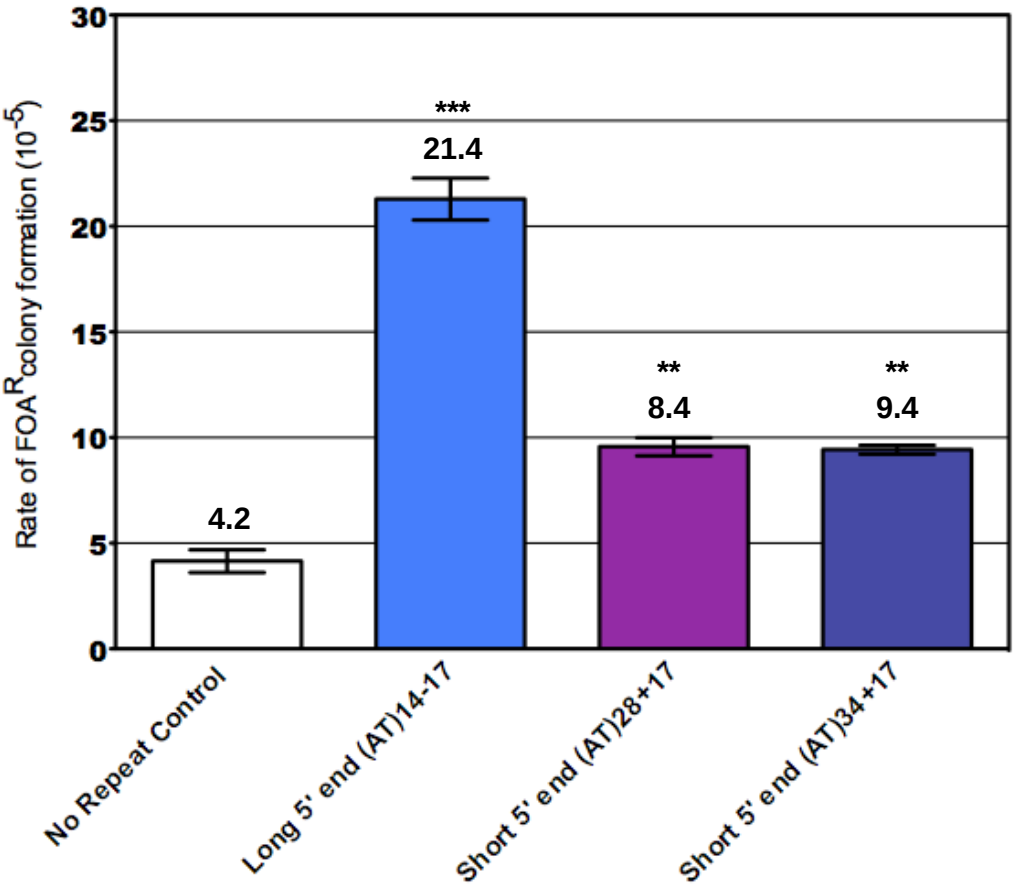


Figure 10. Results of the Recombination Assay for the No Repeat Control, Flex1(AT)14, Flex1(AT)28, and Flex1(AT)34

The rate of generation of FOA^R colonies is shown for each strain. Values are the average of two experiments for each strain except for the no repeat control. The data for the no repeat control can be found in Appendix Table 3. The no repeat control experiments were conducted using a recombination assay that contained a 700-bp duplication. Error bars represent the standard error of the mean (SEM). The labeled values are the mean values for each strain. The ** symbol indicates $p<0.01$ in comparison to the no repeat control, and *** indicates $p<0.001$ (pooled variance t-test).

The Effect of Sequences Flanking the AT repeats on the FRA16D Flex1 Fragility

The next goal was to test the hypothesis that the 5'- and 3'-flanking sequences affect chromosome fragility. The Flex1 sequences with different numbers of AT repeats have different 5' and 3' sequences flanking the AT repeat (Figure 11). Therefore, it was important to determine if this difference has any effect on chromosome fragility. This

was accomplished by cloning various Flex1 sequences with different combinations of 3' and 5' sequences into the recombination assay (Table 2).

Another goal was to determine whether increasing the AT-repeat copy numbers has an additive effect on fragility. Since FRA16D has been shown to contain several AT-rich sequences and perfect AT repeats very close to the Flex1 AT-rich sequence, it was important to examine the effect of multiple copies of AT-rich sequences on fragility. To test this, a yeast strain containing multiple copies of AT repeats was constructed.

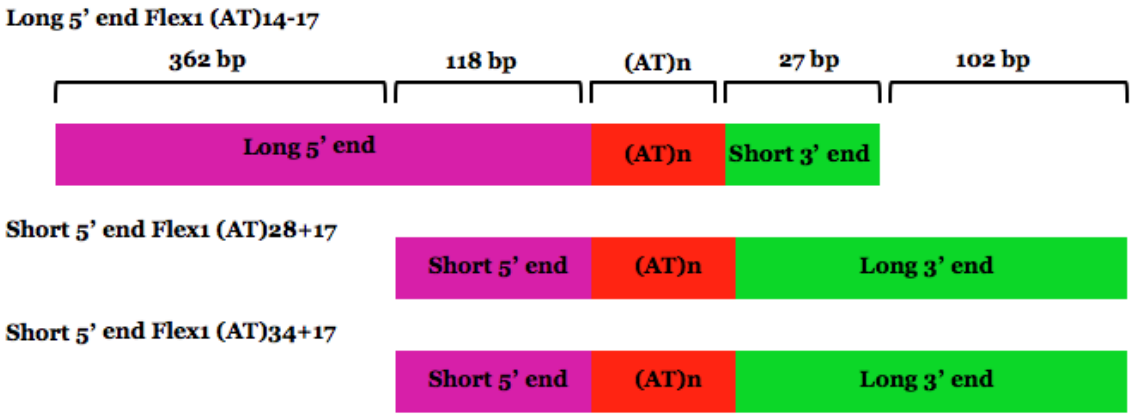


Figure 11. Schematic Diagram of Different Parts of the Flex1 Sequences
Flex1(AT)28 and Flex1(AT)34 are exactly the same except for the AT repeat number: Flex1(AT)14 contains 362 bp of extra 5' end sequence and lacks 102 bp of the 3' end sequence that is predicted to form a 17 bp hairpin by Mfold (Zuker, 2003). Thus, for the sake of consistency, the following nomenclature will be used throughout: Flex1(AT)14 is labeled as "long 5' end (AT)14-17, whereas Flex1(AT)28 and Flex1(AT)34 are labeled as "short 5' end (AT)28+17 and "short 5' end (AT)34+17, respectively. It is unknown whether the 17 bp hairpin structure actually forms *in vivo*, but -17 or +17 is used for the purpose of simple nomenclature.

The first step in cloning Flex1 sequences with various combinations of 3' and 5' flanking sequences and multiple copies of AT repeats, was to digest the pBL007 vector with EcoRI, de-phosphorylate, and then gel-purify. The expected size of a linearized-

pBL007 vector is ~5.8 kb, so the bands at ~5.8 kb in lanes 2-4 indicated the presence of the EcoRI-linearized pBL007 vector (Figure 12).

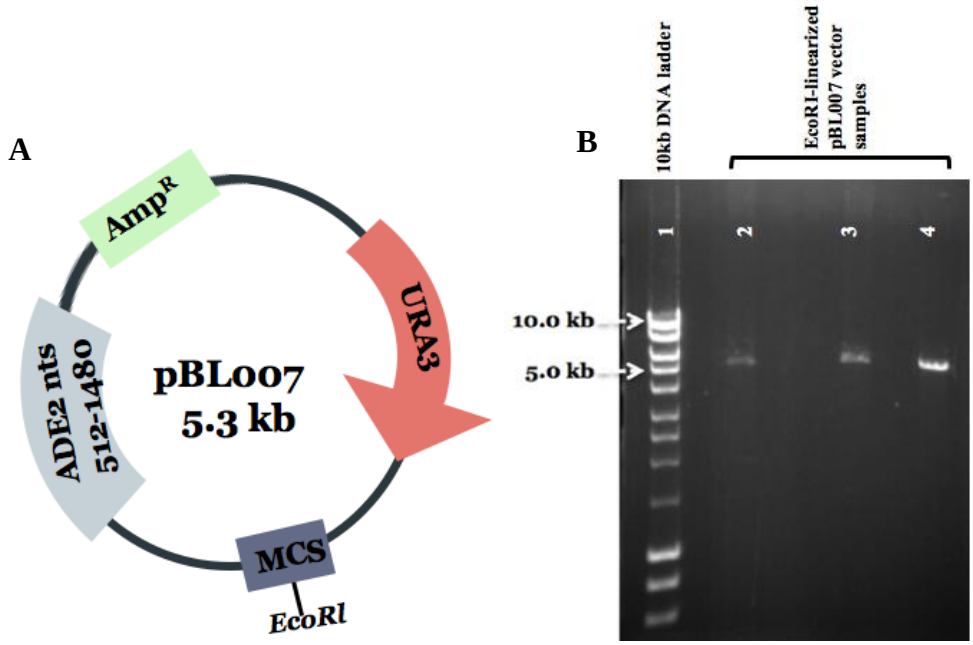


Figure 12. Confirmation of EcoRI-linearized pBL007 vector
(A) Schematic for pBL007 vector. EcoRI digests pBL007 vector within the multiple cloning site (MCS). (B) The EcoRI-digested pBL007 vector was separated from undigested vector by gel electrophoresis. Lanes 2-4 gave bands at ~5.3 kb, which is the expected size for the EcoRI-linearized pBL007 vector.
Lane 1: 10 kb DNA ladder
Lanes 2-4: EcoRI-linearized pBL007 vector

Ligations were performed to insert the Flex1 sequences into the linearized pBL007 vector. The Flex1 inserts with different combinations of 5' and 3' end sequences were variations of the Flex1 sequence that had previously been cloned into the pYES2 vector (K. Shah and C. H. Freudenreich, unpublished data). For transformations, the EcoRI-digested, de-phosphorylated pBL007 vector was used as a negative control and the undigested pBL007 vector was used as a positive control. For the target transformations, an average of ~55 colonies were obtained for each transformation, for all the negative controls, fewer than 10 colonies were obtained, and

for all the positive controls, a lawn of colonies was observed (Table 1). A low number of colonies for the negative control indicated high de-phosphorylation efficiency and a high number of colonies for the positive control indicated high transformation efficiency.

Table 1. Summary of Bacterial Transformation Results

Flex1 Transformation (Flex1 Insert from pYES2 Vector)	No. of Colonies (Target Transformation)	No. of Colonies (Positive Control)	No. of Colonies (Negative Control)	Transformant Insert in pBL007 Vector Verified by Sequencing
(AT)14+17	59	Lawn	6	Long 5' end (AT)21-17
(AT)23+17	51	Lawn	5	Short 5' end (AT)23+17 (in both orientations)
				Short 5' end (AT)24+17
(AT)23-17	62	Lawn	2	Long 5' end (AT)23-17
(AT)34-17	47	Lawn	4	Short 5' end (AT)34-17
(AT)34+17+(AT)23+17	55	Lawn	9	[Short 5' end (AT)23+17] X 3

The table shows a list of Flex1 inserts from the pYES2 vector that were excised and cloned into pBL007 vector. Initially, the 5' sequence flanking the AT-repeat was unknown in these Flex1 inserts in the pYES2 vector. After subcloning the Flex1 inserts from the pYES2 vector into the pBL007 vector, transformant Flex1 inserts were sequenced and the 5' end sequences were determined.

Transformed bacterial colonies were screened by PCR with a forward primer (primer #679) ~500 bp upstream of the multiple cloning site of pBL007 vector and a reverse primer (primer #680) ~500 bp downstream of the multiple cloning site (Figure 13A). Given that the Flex1(AT)n insert is ~500 bp, a correct transformant would produce a ~1.5 kb band, whereas a transformant without an insert would produce a 1.0 kb band. Therefore, the 1.5 kb PCR products in lanes 3-6, 8, 9, 11, 12, 14, and 15 indicated correct transformants (Figure 13B). Subsequently, the transformants were verified by sequencing and transformants in lanes 4, 6, 8-9, 12, 14-15 were confirmed as correct transformants, and were saved as laboratory stocks and used for the

recombination assay. The sequencing results confirmed seven transformants (Figure 14). Interestingly, a transformant that was supposed to be (AT)34+17+(AT)23+17 had undergone an increase in the AT repeat copy number and turned out to be (AT)23+17+(AT)23+17+(AT)23+17 (a.k.a. [(AT)23+17]X 3). Also, a transformant that was supposed to be (AT)14+17 had undergone an increase in the AT repeat length and lost 102 bp of the 3' sequence and became (AT)21-17. A complete list of all the new Flex1 clones is shown in Table 2.

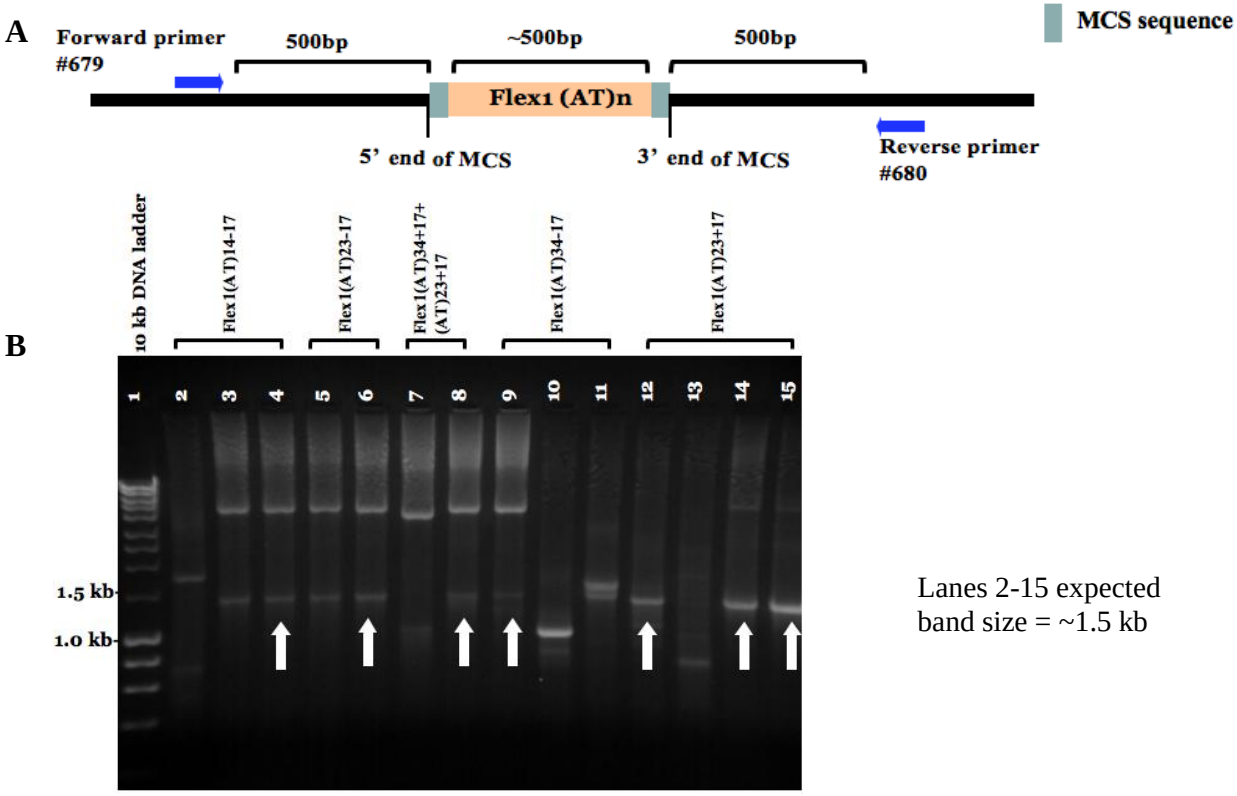


Figure 13. PCR Verification of Bacterial Transformation

(A) Schematic for forward and reverse primers used to detect correct transformants. The forward primer (primer #679) anneals ~500 bp upstream of the multiple cloning site (MCS) of pBL007 and a reverse primer (primer #680) anneals ~500 bp downstream of the MCS. Since Flex1(AT)n insert is ~500 bp, the expected PCR product of a correct transformant is ~1.5 kb. The white arrows point to transformants verified by sequencing, and were saved as laboratory stocks and used for the recombination assay.

(B) Lane 1: 10 kb DNA ladder

Lanes 2-4: (AT)14+17 transformant

Lanes 5, 6: (AT)23-17 transformant

Lanes 7, 8: (AT)34+17+(AT)23+17 transformant

Lanes 9-11: (AT)34-17 transformant

Lanes 12-15: (AT)23+17 transformant

Table 2. New Flex1 Bacterial Clones Verified by Sequencing

Flex1 Sequence in pBL007 Vector Verified by Sequencing	Plasmid Stock Number	Insert Source
Long 5' end (AT)21-17	#352	pYES2 (AT)14+17
Short 5' end (AT)23+17 (in both orientations)	#354 #355	pYES2 (AT)23+17
Short 5' end (AT)24+17	#350	
Long 5' end (AT)23-17	#349	pYES2 (AT)23-17
Short 5' end (AT)34-17	#351	pYES2 (AT)34-17
[Short 5' end (AT)23+17]X 3	#353	pRS426 (AT)34+17+(AT)23+17

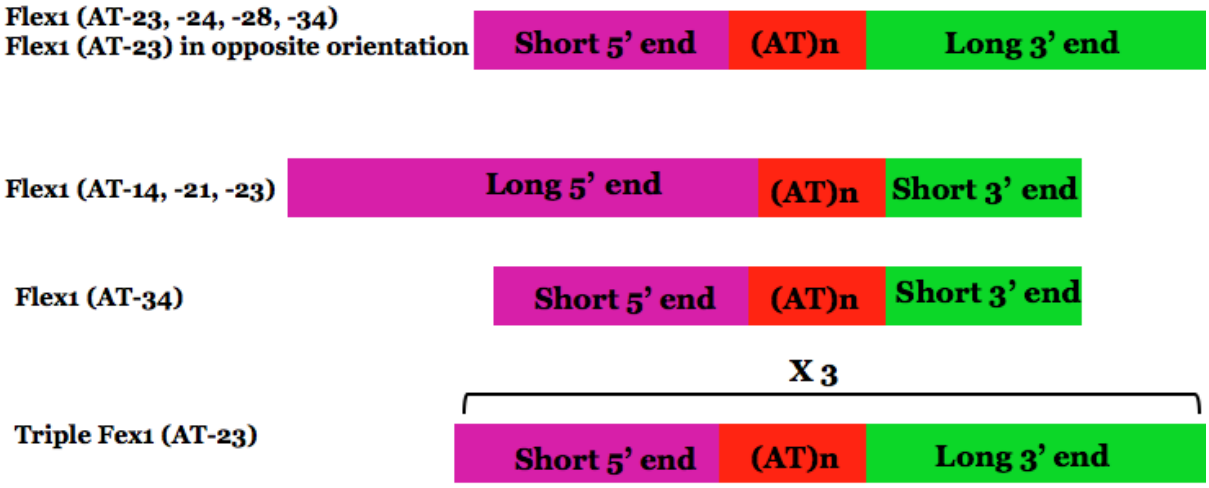


Figure 14. Summary of all the Bacterial Transformants Obtained
Sequencing results verified seven new bacterial transformants shown above.

Subsequently, pBL007 plasmids isolated from the transformants were linearized with XbaI and were transformed into the yPH499 yeast strain. Then, the yeast transformants were selected on YC-Ura plates. I was unable to screen yeast transformant colonies by PCR due to an inability of the PCR procedure to give clear bands for PCR products. I tried several PCR conditions including different extension temperatures, but these did not increase PCR efficiency. I noticed that bands get fainter as PCR product sizes increase beyond 2 kb. Thus, one explanation for the trouble obtaining the desired

PCR product (expected size ~3.0 kb) is that PCR becomes inefficient in amplifying large PCR products beyond > ~2.0 kb. Thus an alternative screening method was used. In this method, a small amount of transformant colonies was patched onto YC-Ade plates to identify correct transformants. This was done because the correct transformants on YC-Ade plates will either not grow or barely grow, displaying a red sectoring, since the *ADE2* gene is disrupted during transformation. Using this method, three yeast transformants were identified: short 5' end (AT)34-17, long 5' end (AT)23-17, and [short 5' end (AT)23+17]X 3.

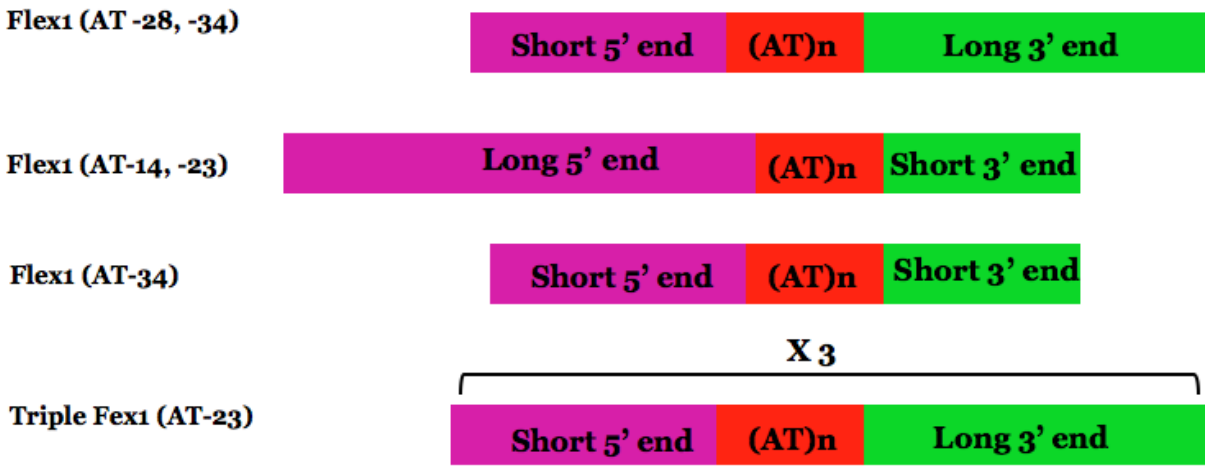
Direct Repeat Recombination Assay

The direct repeat recombination assays were performed on the new Flex1 constructs (Figure 15A) and on the previously obtained constructs (Figure 11). All of the strains showed an increase in the rate of FOA^R compared to the control, and the previous constructs (long 5' end (AT)14-17, short 5' end (AT)28+17, and short 5' end (AT)34+17) showed the same trend in the rate of breakage as before (Figure 15B). However, the surprising finding was that short 5' end (AT)34-17 showed a significant increase in the rate of breakage (~55-fold over control) compared to the other Flex1 sequences tested. Compared to the control, the rate was ~5-fold greater for long 5' end (AT)14-17, ~2-fold greater for both short 5' end (AT)28+17 and short 5' end (AT)34+17, and ~ 55-fold greater for short 5' end (AT)34-17. The only difference between short 5' end (AT)34+17 strain and short 5' end (AT)34-17 strain is that the latter one lacks 102 bp of the 3' sequence that is predicted to form a 17 bp hairpin. From this difference alone, short 5' end (AT)34-17 showed a rate of breakage ~24-fold greater

than short 5’ end (AT)34+17 indicating that the 17 bp hairpin-prone sequence strongly inhibits recovery of broken chromosomes in the recombination assay.

For all the Flex1 sequences lacking the17 bp hairpin-prone sequence, we observed an increase in the rate of chromosome breakage that is dependent on the length of the AT repeat (Figure 15B). Compared to the control, the rate was ~5-fold greater for long 5’ end (AT)14-17, ~7 fold greater for long 5’ end (AT)23-17, and ~55-fold greater for short 5’ end (AT)34-17. Though the result for long 5’ end (AT)23-17 is from a single experiment, the Flex1 sequences with a longer AT repeat showed a higher rate of breakage. Interestingly, [short 5’ end (AT)23+17]X 3, which has three copies of (AT)23 repeats showed a rate of breakage that is ~1.4 fold lower than the breakage rate for long 5’ end (AT)23-17, which has only one (AT)23 repeat, suggesting that the multiple copies of 17 bp hairpin-prone sequence in the [short 5’ end (AT)23+17]X 3 strain strongly inhibit recovery of broken chromosomes in the assay. Furthermore, the breakage rate for [short 5’ end (AT)23+17]X 3 was ~2.6 fold greater than the short 5’ end (AT)28+17 rate, indicating some additive effect of multiple AT repeats on fragility.

A



Schematic diagram of all the Flex1 constructs used in the recombination assay

B

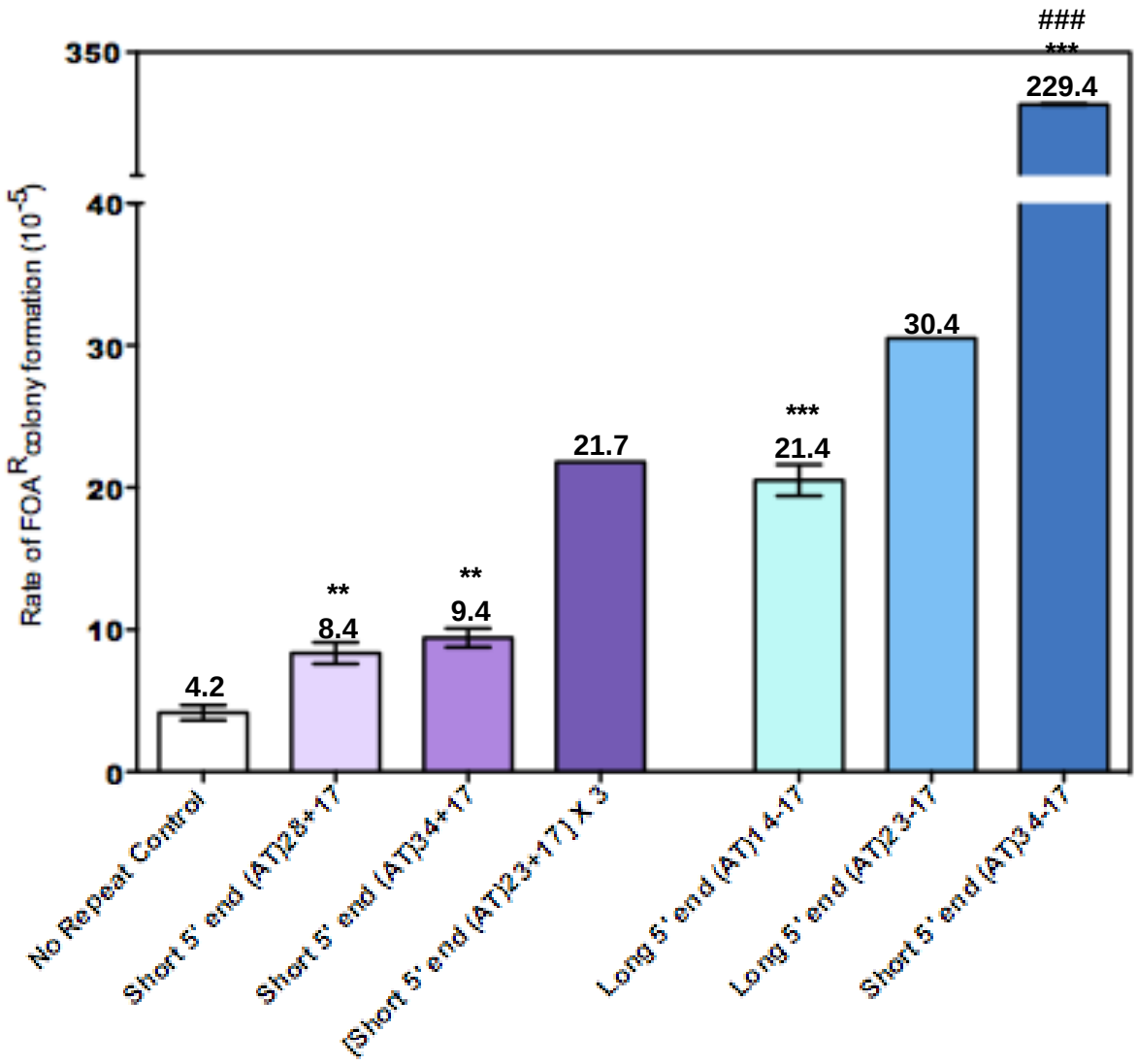


Figure 15. Results of the Recombination Assay for the No Repeat Control, Short 5' end (AT)28+17, Short 5' end (AT)34+17, [Short 5' end (AT)23+17]X3, Long 5' end (AT)14-17, Long 5' end (AT)23-17, Short 5' end (AT)34-17

(A) Schematic diagram of all the Flex1 constructs used in the recombination assay.

(B) The rate of generation of FOA^R colonies is shown for each strain. Values for long 5' end (AT)14-17, short 5' end (AT)28+17, and short 5' end (AT)34+17 are the average of four experiments. Values for short 5' end (AT)34-17 are the average of two experiments. Values for long 5' end (AT)23-17, and [short 5' end (AT)23+17]X 3 are from a single experiment. The no repeat control values are the average of five experiments. The no repeat control data can be found in Appendix 2. The error bars represent the standard error of the mean (SEM). For short 5' end (AT)34-17, the range is 228 to 230 X 10^{-5}. The breaks indicate a discontinuity in the y axis and the labeled values are the mean values for each strain. The ** symbol indicates $p<0.01$ in comparison to the no repeat control, and *** indicates $p<0.001$ (pooled variance t-test). The ### symbol indicates $p<0.001$ in comparison to short 5' end (AT)34+17 (pooled variance t-test).

DISCUSSION

The initial goal of my project was to determine if the direct repeat recombination assay can overcome the problem of the structure-prone Flex1 region interfering with *de novo* telomere addition in the YAC breakage assay used by Zhang and Freudenreich (2007). The direct repeat recombination assay was performed on Flex1(AT)14, Flex1(AT)28, and Flex1(AT)34. Compared to the no repeat control, the rate was ~5-fold greater for Flex1(AT)14, ~2-fold greater for both Flex1(AT)28, and Flex1(AT)34 (Figure 10). Interestingly, the rate of FOA^R of Flex1(AT)14 (~5-fold over control) was consistently greater than Flex1(AT)28, and Flex1(AT)34. This result is consistent with the study by Zhang and Freudenreich (Zhang & Freudenreich, 2007). The fact that the Flex1(AT)28 construct rate was similar to the Flex1(AT)34 construct and that both were lower than Flex1(AT)14 indicated that either (i) fragility was not influenced by the length of the AT repeat after all, or (ii) the 5'- and/or 3'-flanking sequences were influencing the rate of recombination. Thus, this result prompted us to test the hypothesis that sequences flanking the AT repeat within Flex1 affect chromosome fragility.

Cloning Flex1(AT)n Sequences

To examine the effect of the 5' and 3' flanking sequences on chromosome fragility, Flex1 sequences with various combinations of 5' and 3' sequences were cloned into the pBL007 vector (Figure 14). The Flex1 sequences were difficult to clone, possibly because these sequences are highly AT rich (65-75%) and contain perfect AT repeats embedded within the AT-rich sequence. The difficulty of cloning AT-rich genomic DNA is well documented (Gardner et al., 2002). Nearly every step of the

cloning process had to be modified to obtain successful results. The main problem was high background levels of empty vector transformants and relatively few recombinants. I tried several different cloning conditions, including different concentrations of enzymes, different reaction times for digestion and ligation, and different molar ratios of insert to vector. After increasing the concentration of restriction enzymes and the duration of digestion and ligation reactions, the background levels dropped slightly. The background levels finally dropped substantially when the EcoRI-digested vector was separated from the undigested vector and purified by gel electrophoresis. The sequencing results of bacterial transformants indicated that some transformants did not preserve the Flex1 sequences stably in plasmids and had either duplication and/or deletion of the AT repeats (Table 1).

The 17 bp Hairpin-Forming Sequence Strongly Inhibits Recovery of Broken Chromosomes

The Flex1 sequences were integrated into the *ADE2* locus on chromosome II in yPH499 yeast strain and chromosome fragility was quantified using the direct repeat recombination assay. Strikingly, cells that contained short 5' end (AT)34-17 showed ~24-fold greater rate of breakage compared to cells that had short 5' end (AT)34+17. The only difference between these two strains is that the strain that gave ~24-fold higher rate of breakage lacks 102 bp of the 3' end sequence, which is predicted to easily extrude to form a 17 bp hairpin structure (Figure 18C & D). These data clearly suggest that the 17 bp hairpin-forming sequence strongly inhibits recovery of broken chromosomes in the assay. The inhibitory effect is again demonstrated by the difference in the breakage rates between short 5' end (AT)28+17 and long 5' end (AT)23-17. The

long 5’ end (AT)23-17 strain that lacks the 17 bp hairpin-prone sequence gave ~3-fold greater breakage rate than the short 5’ end (AT)28+17 strain that contains the 17 bp hairpin-prone sequence (Figure 15). This is further supported by the finding that despite the correlation between the length of the AT repeat and the breakage rate observed in the Flex1 sequences without the 17 bp hairpin-forming sequence, the [short 5’ end (AT)23+17]X3 strain, which has three copies of the (AT)23 repeat gave a lower rate of breakage than long 5’ end (AT)23-17 that has only one copy of the (AT)23 repeat (Figure 15). A possible explanation for the result is that because the [short 5’ end (AT)23+17]X3 strain has three copies of the 17 bp hairpin-prone sequence, it experiences an additive effect of inhibition by the 17 bp hairpin-prone sequences, thereby further inhibiting recovery of broken chromosomes in the assay.

The Effect of Multiple AT Repeats on Fragility

The common fragile site FRA16D has been shown to contain several AT-rich sequences and perfect AT repeats very close to the Flex1 AT-rich sequence. Therefore, it was important to determine the effect of multiple copies of AT-rich sequences on fragility. The two constructs that both have the 17 bp hairpin-forming sequence and the same 5’ end sequence, but differ only in the copy number of the AT repeat are [short 5’ end (AT)23+17]X 3 and short 5’ end (AT)28+17 strains. The [short 5’ end (AT)23+17]X 3 strain that has three copies of the (AT)23 repeat showed a rate of breakage ~2.6 fold greater than short 5’ end (AT)28+17 that has one copy of the (AT)28 repeat, indicating some additive effect of multiple AT repeats on chromosome fragility. This result supports the idea that the formation of multiple hairpin and/or cruciform structures by multiple AT repeats contributes to chromosome fragility.

The 17 bp Hairpin Structure May Interfere with Double-Strand Break Resection by Sae2 and the MRX Complex

One explanation for the inhibitory effect of the 17 bp hairpin-forming sequence is that it interferes with resection of the DSB ends needed to expose the homologous *ADE2* fragments, thereby preventing homologous recombination (HR) and recovery of broken chromosomes in our assay. DSB repair can occur through non-homologous end joining (NHEJ) or HR. When a DSB occurs between direct repeats, its repair is primarily achieved by a particular kind of HR known as single strand annealing (SSA) (Pâques & Haber, 1999; Clikeman et al., 2001; Lehman & Carroll, 1991; Petrini et al., 1997). Because of the presence of homologous *ADE2* direct repeats on either side of the intervening region that contains the Flex1 sequence and the *URA3* gene, a DSB in the intervening region would stimulate recombination by SSA and would be mainly repaired by a Rad1- and Rad52-dependent SSA. In fact, Freudenreich et al. (1998) showed that DSB repair occurs primarily by SSA in a direct repeat recombination assay. SSA requires 5' to 3' degradation of the DSB ends to expose complementary single-stranded DNA followed by Rad52 annealing of the repeats. Subsequently, nonhomologous long 3'-ended single-stranded residual tails are removed by the Rad1 nuclease and the resulting gaps are filled in by DNA repair synthesis and ligation, resulting in deletion of the intervening sequence (Figure 16).

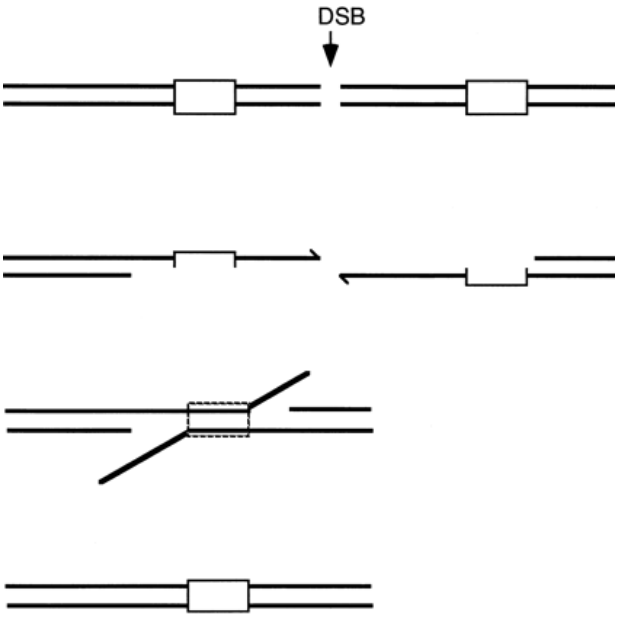


Figure 16. Single Strand Annealing
A DSB between two direct repeats stimulates recombination by SSA. Resection of the ends of DSB generates complementary single stranded DNA that are annealed by Rad52. Subsequently, the nonhomologous 3'-ended single stranded tails are removed followed by gap filling (Pâques & Haber, 1999).

Studies have suggested a two-step mechanism for DSB resection during homologous recombination in budding yeast. First, Sae2 and the MRX complex act cooperatively to remove a small oligonucleotide(s) from the DSB ends to produce a partly resected intermediate. Second, this intermediate is further processed by Exo1 or Dna2 and the Sgs1 helicase by a combination of exonucleolytic and unwinding activities to produce 5' resected ends (Figure 17). It has been suggested that the initial processing step by Sae2 and MRX may be important to prevent binding by the Ku complex, leading to repair by error-prone NHEJ. Furthermore, studies have suggested that Sae2 interacts with MRX to process hairpin-containing DNA intermediates during DSB resection, in which MRX facilitates Sae2 endonuclease activity to cleave single-stranded DNA

adjacent to the hairpins (Mimitou & Symington, 2008; Lengsfeld et al., 2007; Clerici et al., 2005; Fiorentini et al., 1997).

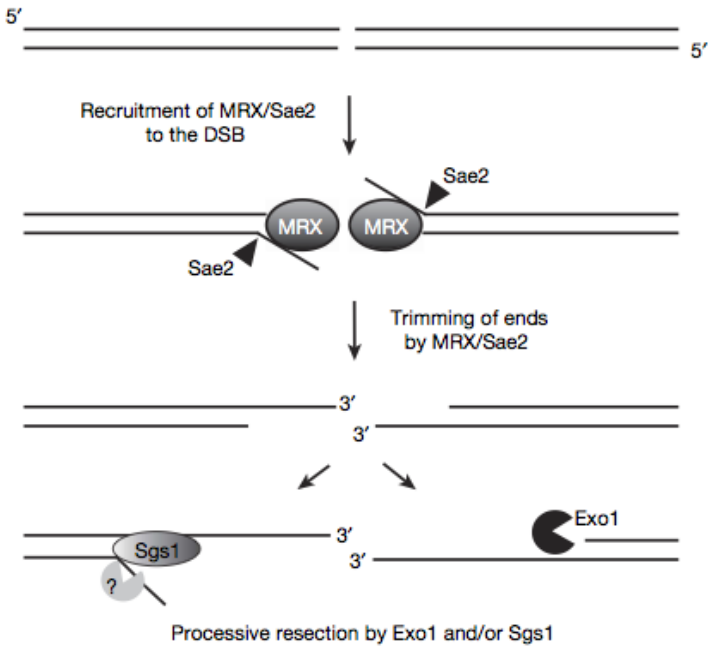


Figure 17. Two-Step Mechanisms for DSB Processing in Homologous Recombination
After a DSB is formed, Sae2 and the MRX complex remove a small oligonucleotide(s) from the ends to generate a partly resected intermediate, which is further processed by the 5' to 3' exonucleolytic activity of Exo1 and the Sgs1 helicase (Mimitou & Symington, 2008).

The inhibitory effect of the 17 bp hairpin-forming sequence may be due to the 17 bp hairpin interfering with resection of the DSB ends needed to initiate HR. When a DSB occurs in Flex1 sequences that lack the 17 bp hairpin-forming sequence, resection of the ends of DSB would occur normally and subsequent HR would rescue broken chromosomes in the assay. In contrast, a DSB in the Flex1 sequences that form the 17 bp hairpin ~23 bp away from a secondary structure formed by the AT repeat may give rise to unusual structures such as DSBs covalently closed by the hairpin structure (e.g. a

hairpin fold-back structure or hairpin-capped ends) that can inhibit or reduce the efficiency of DSB resection. This will result in a failure to initiate HR in Flex1 sequences that form the 17 bp hairpin and thus recovery of broken chromosomes in the assay. Subsequently, this will underestimate the rate of breakage at Flex1 sequences that generate the 17 bp hairpin. It is not known what the structures of the ends are immediately after the DSB because the breaks are spontaneous and the exact location of the breaks are unknown. However, SSA is inherently sensitive to defects in DSB resection, so any DNA structure that can interfere with DSB resection will inhibit SSA. Moreover, failure to process the ends of DSB covalently closed by the hairpin structure could result in binding of the Ku complex and subsequent repair by NHEJ without loss of the *URA3* marker.

The Fragility of the FRA16D Flex1 Region is Dependent on the Length of the AT Repeat

When only the Flex1 regions lacking the 17 bp hairpin-prone sequence were tested for chromosome fragility, we observed an AT-repeat length dependent increase in the rate of chromosome breakage (Figure 15); the Flex1 sequences with longer AT repeats showed a higher rate of breakage. This result strongly suggests that a variable AT-repeat length, which correlates with the ability to form a secondary structure, is a crucial factor in the propensity to break. This finding is supported by the study by Zhang and Freudenreich (2007), which demonstrated that the Flex1 regions cause replication fork stalling in an AT-repeat length dependent manner; it was observed that longer Flex1 AT alleles caused a stronger replication stalling. Therefore, the authors concluded that a secondary structure formed by the AT repeat acts as a replication fork barrier causing

fork stalling, which is partly responsible for the increased chromosome breakage in the FRA16D Flex1 region (Zhang & Freudenreich, 2007).

Potential Mechanisms for Chromosome Breaks at the FRA16D Flex1 Region

Common fragile sites are extremely susceptible to breakage and are hotspots for translocation and deletion breakpoints in various cancer cells. The deletion or translocations within common fragile sites often result in the disruption of proto-oncogenes, leading to tumorigenesis. Often, the breakpoint sequences are characterized by extremely AT-rich sequences. One of the recurrent breakpoints in human common fragile site FRA16D is mapped within the ~500 bp AT-rich (65-75%) region (Flex1) that contains a perfect AT repeat (Zhang & Freudenreich, 2007). Studies have shown that inverted repeats located in AT-rich regions form a cruciform structure *in vivo* in many experimental organisms including *Escherichia coli*, yeasts, and mice. These studies also demonstrated that highly AT-rich sequences, particularly alternating (AT)n sequence, do not have kinetic barrier to cruciform extrusion *in vitro* (Dayn et al., 1991; reviewed in Kogo et al., 2007). Thus, the characteristics of the Flex1 sequences, including a high AT content (65-75%) and perfect (AT)n dinucleotide repeats, should help in forming a stable cruciform structure *in vivo* by the force of negative supercoiling in the chromosomal context. Specifically, cruciform formation can occur *in vivo* by processes that increase DNA supercoiling, including active transcription and replication (Dayn et al., 1991). The putative cruciform structure involving both strands of Flex1(AT)34 is shown (Figure 18A). These studies also showed that a longer inverted repeat within an AT-rich region is more likely to form a cruciform structure than a shorter one in the same context (Dayn et al., 1991; Kogo et al., 2007; Kurahashi et al., 2010). Therefore,

the longer AT-rich inverted repeats of Flex1 are more likely to adopt a cruciform structure.

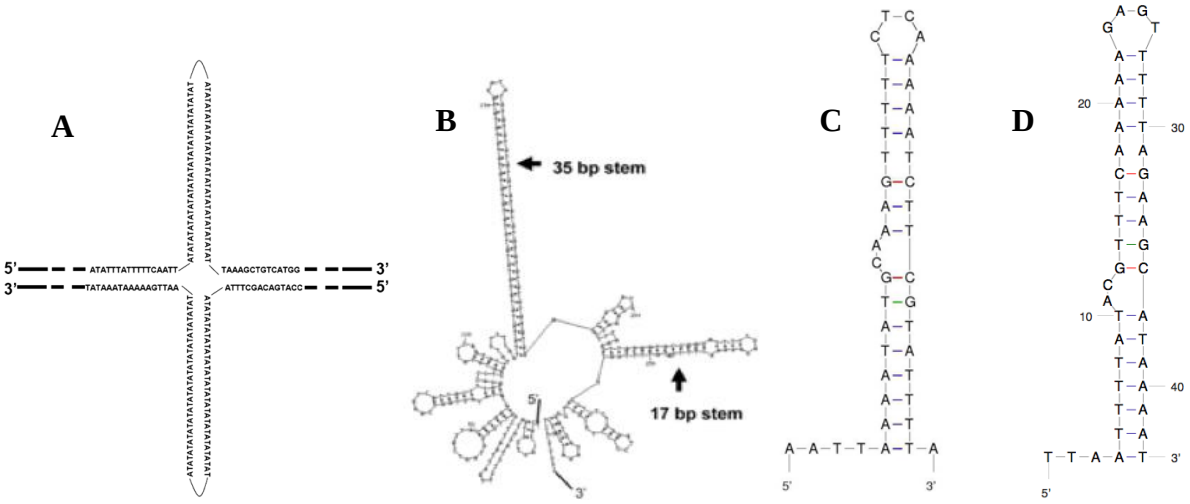


Figure 18. Potential Secondary Structures Formed by Flex1 Sequences
(A) Flex1(AT)34 has a potential to form the following cruciform structure through intrastrand-base pairing within single-stranded DNA. Other Flex1(AT)n sequences with lower numbers of AT repeats are predicted to form cruciform structures, but with less efficiency (Zhang & Freudenreich, 2007). **(B)** The potential secondary structure with a highest melting point (Tm 60°C) formed by single-stranded Flex1(AT)34 sequence predicted by Mfold (Zuker, 2003; Zhang & Freudenreich, 2007). **(C)** The 17 bp hairpin (Tm 48.4°C) that is predicted to form 23 bp away from the hairpin formed by the AT-repeat of the Flex1 region containing 101 bp of extra 3' end sequence (Zuker, 2003). **(D)** The 17 bp hairpin (Tm 52.2°C) that is predicted to form by the complementary strand.

Correlation between the rate of chromosome breakage and the secondary structure-forming propensity among these Flex1 sequences strongly supports that the formation of the secondary structures is the aetiology for the chromosome fragility, leading to genomic instability. This concept is further supported by the observation that longer Flex1 AT alleles caused a stronger replication stalling, which can lead to chromosome breakage (Zhang & Freudenreich, 2007). Thus, as the authors have hypothesized, fragility of the Flex1 region is likely to be mediated by the stalling of DNA replication at the perfect AT repeat that forms a stable secondary structure. This is further supported by the observation that breakage was exacerbated by treatment with

hydroxyurea, which slows replication. When replication is slowed, there are large single-stranded regions, which increase the probability of forming a secondary structure that can stall the replication fork. In fact, studies have shown that sequences prone to forming secondary structures cause replication fork stalling, leading to DNA breaks (reviewed in Voineagu et al., 2009).

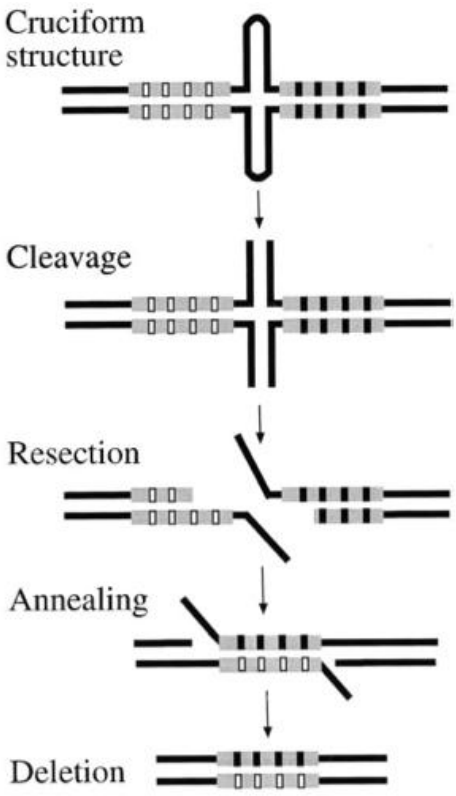


Figure 19. Processing of the Cruciform Followed by Single Strand Annealing
Cleavage at the cruciform structure by a structure-directed nuclease such as SbcCD would result in two DNA molecules. Subsequently, DNA end resection followed by annealing and ligation restores two continuous DNA strands. SbcCD's exonuclease activity may also resect the 3' strand until it encounters regions of homology (Bzymek & Lovett, 2001).

Another potential pathway for chromosome breakage at the Flex1 region is that the AT-rich inverted repeat becomes extruded as a cruciform, which is then cleaved by structure-directed nucleases, such as a DNA resolvase. A structure-specific nuclease

cleavage may initiate the DNA breakages followed by resection of the DNA ends and annealing through SSA pathway (Figure 19). This is a plausible mechanism because cruciform extrusion and subsequent processing of the cruciform structure has been observed in a yeast model (reviewed in Lemoine et al., 2005).

In the Flex1 region, length polymorphism was observed at the perfect AT repeat in the human population with alleles varying from 11 to 88 AT repeats. Since chromosome fragility was shown to be dependent on the length of the AT repeat, individuals with longer Flex1 AT alleles may be more susceptible to genomic instability in FRA16D observed in various cancer cells.

Overall, my data show that 102 bp of extra 3'-end sequence prone to forming the 17 bp hairpin strongly inhibits recovery of broken chromosomes in the direct repeat recombination assay. Though the exact mechanism of inhibition is unknown, it is likely that the 17 bp hairpin interferes with DSB resection by Sae2 and the MRX complex. Once that sequence was removed, we observed a correlation between the rate of chromosome breakage and the secondary structure-forming propensities among the Flex1 sequences. Potential pathways by which Flex1 induces chromosome fragility may include chromosome breakage from replication fork stalling at structures formed by Flex1 and cruciform-mediated DSB that is independent of replication.

An important next step for this project is to determine if the 5'-end sequence flanking the perfect AT repeat has any effect on chromosome fragility. Some Flex1 constructs have 362 bp of extra 5'-end sequence and some do not, so it is important to determine whether that extra 362 bp sequence at the 5' end has any influence on fragility. To determine this, we could compare the rate of breakage between Flex1 clones that have the same sequence except for the 5'-end sequence. Specifically, we

could compare the fragility between two Flex1 clones that have the same number of AT repeat and the same 3'-end sequence, but differ only in their 5'-end sequences (either a long 5' end or short 5' end). For example, we could make new constructs and compare the rate of breakage between short 5' end (AT)14-17 and long 5' end (AT)14-17 clones, short 5' end (AT)23-17 and long 5' end (AT)23-17 clones, and short 5' end (AT)34-17 and long 5' end (AT)34-17. If there is no difference in the rate of breakage between Flex1 clones that differ only in the 5'-end sequence, it would indicate that the 5' sequence flanking the AT repeat has no effect on fragility.

Another important experiment for this project is to determine preferred sites of breakage in chromosomes with Flex1 sequences by pulsed-field gel electrophoresis. This method will enable direct detection of breakage intermediates and allow us to verify whether the perfect AT repeat within Flex1 is indeed the preferred site of breakage.

Finally we would like to test whether the fragility of Flex1 sequences depends on the orientation the 17 bp hairpin-forming sequence relative to the replication origin. It has been suggested that a structure-prone sequence in the lagging strand template is more likely to form a secondary structure than when a structure-prone sequence is in the leading strand. This is because a portion of the lagging strand template called the Okazaki Initiation Zone remains transiently single stranded, which could facilitate the formation of secondary structures by structure-prone sequences. Once the structure forms, it will stall replication fork and a stalled fork could collapse to generate chromosome breakage (Mirkin, 2006). It would be interesting to compare the effects of the 17 bp hairpin-prone sequence in the lagging strand versus when it is in the leading strand.

BIBLIOGRAPHY

Aguiléra, A. & Gómez-González, B. (2008). Genome instability: a mechanistic view of its causes and consequences. *Nature Reviews,* 9, 204-217.

Bowater, R. P., Aboul-ela, F., & Lilley, D. M. (1994). Large-scale opening of A+T rich Regions within supercoiled DNA molecules is suppressed by salt. *Nucleic Acids Research,* 22(11), 2042-2050.

Branzei, D., & Foiani, M. (2010). Maintaining genome stability at replication fork. *Nature Reviews*, 11, 208-219.

Burrow, A. A., Marullo, A., Holder, L. R., & Wang, Y. (2010). Secondary structure formation and DNA instability at fragile site FRA16B. *Nucleic Acids Research,* 38(9), 2865-2877.

Burrow, A. A., Williams, L. E., Pierce, L. C. T., & Wang, Y. (2009). Over half of breakpoints in gene pairs involved in cancer-specific recurrent translocations are mapped to human chromosomal fragile sites. *BMC Genomics*, 10(59), 1-11.

Bzymek, M., & Lovett, S. T. (2001). Evidence for two mechanisms of palindrome-stimulated deletion in *Escherichia coli*: single-strand annealing and replication slipped mispairing, *Genetics*, 158, 527-540.

Casper, A. M., Nghiem, P., Arlt, M. F., & Glover, T. W. (2002). ATR regulates fragile site stability. *Cell*, 111, 779-789.

Chesi, M., Bergsagel, P. L., Shonukan, O. O., Martelli, M. L., Brents, L. A., Chen, T. et al. (1998). Frequent dysregulation of the c-MAF proto-oncogene at 16q23 by translocation to an Ig locus in multiple myeloma. Blood, 91, 4457-4463.

Cimprich, K. A., & Cortez, D. (2008). ATR: An essential regulator of genome integrity. *Nat. Rev. Mol. Cell.* Biol., 9, 616-627.

Clerici, M., Mantiero, D., Lucchini, G., & Longhese, M. P. (2005). The *Saccharomyces cerevisiae* Sae2 protein promotes resection and bridging of double strand break ends. *The Journal of Biological Chemistry*, 280(46), 38631-38638.

Clikeman, J. A., Khalsa, G. J., Barton, S. L., & Nickoloff, J. A. (2001). Homologous recombination repair of double-strand breaks in yeast is enhanced by MAT heterozygosity through yKU-dependent and –independent mechanisms. *Genetics*, 157, 579-589.

Dayn, A., Malkhosyan, S., Duzhy, D., Lyamichev, V., Panchenko, Y., & Mirkin, S. (1991). Formation of (dA-dT)n cruciforms in *Escherichia coli* cells under different environmental conditions. *Journal of Bacteriology*, 173(8), 2658-2664.

Dillon, L. W., Burrow, A. A., & Wang, Y. (2010). DNA instability at chromosomal fragile sites in cancer. *Current Genomics*, 11, 326-337.

Ferber, M. J., Thorland, E. C., Brink, A. A., Rapp, A. K., Phillips, L. A., & McGovern, R. et al. (2003). Preferential integration of human papillomavirus type 18 near the c-myc locus in cervical carcinoma. Oncogene, 22, 7233-7242.

Finnis, M., Dayan, S., Hobson, L., Chenevix-Trench, G., Friend, K., & Ried, K. et al. (2005). Common chromosomal fragile site FRA16D mutation in cancer cells. *Human Molecular Genetics*, 14(10), 1341-1349.

Fiorentini, P., Huang, K. N., Tishkoff, D. X., Kolodner, R. D., & Symington, L. S. (1997). Exonuclease I of *Saccharomyces cerevisiae* functions in mitotic recombination in vivo and in vitro. *Molecular and Cellular Biology*, 17(5), 2764-2773.

Focarelli, M. L., Soza, S., Mannini, L., Paulis, M., Montecucco, A., & Musio, A. (2009). Claspin inhibition leads to fragile site expression. *Genes, Chromosomes & Cancer,* 48, 1083-1090.

Freudenreich, C. H (2005). Molecular Mechanisms of Chromosome Fragility. *Chemtraicks-Biochemistry and Molecular Biology*, 18(3), 141-152.

Freudenreich, C. H. (2007). Chromosome fragility: molecular mechanisms and cellular consequences. *Frontiers in Bioscience*, 12, 4911-4924.

Freudenreich, C. H., Kantrow, S. M., & Zakian, V. A. (1998). Expansion and length-dependent fragility of CTG repeats in yeast. *Science,* 279, 853-856.

Gandhi, M., Dillon, L. W., Pramanik, S., Nikiforov, Y. E., & Wang Y. (2010). DNA breaks at fragile sites generate oncogenic RET/PTC rearrangements in human thyroid cells. *Oncogene*, 29, 2272-2280.

Gietz, R. D., & Schiestl, R. H. (2007). High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nature Protocols*, 2(1), 31-35.

Glover, T. W., Arlt, M. F., Casper, A. M., & Durkin, S. G. (2005). Mechanisms of common fragile site instability. *Human Molecular Genetics*, 14(2), R196-R205.

Glover, T. W., Berger, C., Coyle, J., & Echo, B. (1984). DNA polymerase α inhibition by aphidicolin induces gaps and breaks at common fragile sites in human chromosomes. *Human Genetics*, 67, 136-142.

Glover, T. W., & Stein, C. K. (1987). Induction of sister chromatic exchanges at common fragile sites. *Am. J. Hum. Genet*., 41, 882-890.

Hall, B. M., Ma, C., Liang, P., & Singh, K. K. (2009). Fluctuation analysis calculator (FALCOR): a web tool for the determination of mutation rate using Luria-Delbruck fluctuation analysis. *Bioinformatics*, 25(12), 1564-1565.

Hellman, A., Rahat, A., Scherer, S. W., Darvasi, A., Tsui, L. –C., & Kerem, S. (2000). Replication delay along FRA7H, a common fragile site on human chromsome 7, leads to chromosomal instability. *Mol. Cell. Biol*., 20, 4420-4427.

Helmrich, A., Stout-Weider, K., Matthaei, A., Hermann, K., Heiden, T., & Schrock, E. (2007). Identification of the human/house syntenic common fragile site FRA7K/Fra12C1—relation of FRA7K and other human common fragile sites on chromosome 7 to evolutionary breakpoints. *International Journal of Cancer*, 120(1), 48-54.

Huang, H., Reed, C. P., Lomberk, G., Wang, L., Shridhar, V., Hartmann, L. et al. (1999). Frequent deletions within FRA7G at 7q31.2 in invasive epithelial ovarian cancer. *Genes Chromosomes Cancer*, 24(1), 48-55.

Huebner, K., & Croce, C. M. (2001). FRA3B and other common fragile sites: the weakest links. *Nat. Rev. Cancer*, 1, 214-221.

Huebner, K., Garrison, P. N., Barnes, L. D., & Croce, C. M. (1998). The role of the FHIT/FRA3B locus in cancer. *Annual Rev. Genet*., 32, 7-31.

Inagaki, H., Ohye, T., Kogo, H., Yamada, K., Kowa, H., Shaikh, T. H., & Emanuel, B. S. et al. (2005). Palindromic AT-rich repeat in the NF1 gene is hypervariable in humans and evolutionarily conserved in Primates. *Human Mutation,* 26(4), 332-342.

Jiang, Y., Lucas, I., Young, D. J., Davis, E. M., Karrison, T., & Rest, J. S. et al. (2009). Common fragile sites are characterized by histone hypoacetylation. *Human Molecular Genetics,* 18(23), 4501-4512.

Kogo, H., Inagaki, H., Ohye, T., Kato, T., Emanuel, B., & Kurahashi, H. (2007). Cruciform extrusion propensity of human translocation-mediating palindromic AT-rich repeats. *Nucleic Acids Research*, 35(4), 1198-1208.

Krasilnikova, M. M., & Mirkin, S. M. (2004). Replication Stalling at Friedreich's Ataxia (GAA)n Repeats in Vivo. *Molecular and Cellular Biology*, 24(6), 2286-2295.

Krummel, K. A., Denison, S. R., Calhoun, E., Phillips, L. A., & Smith, D. I. (2002). The common fragile site FRA16D and its associated gene WWOX are highly conserved in the mouse at Fra8E1. Genes Chrom. Cancer, 34, 154-167.

Kurahashi, H., Inagaki, H., Kato, T., Hosoba, E., Kogo, H., & Ohye, T. et al. (2009). Impaired DNA replication prompts deletions within palindromic sequences, but does not induce translocations in human cells. *Human Molecular Genetics,* 18(18), 3397-3406.

Kurahashi, H., Inagaki, H., Ohye, T., Kogo, H., Tsutsumi, M., & Kato, T. et al. (2010). The constitutional t(11;22): implications for a novel mechanism responsible for gross chromosomal rearrangements. *Clinical Genetics*, 78, 299-309.

Lavin, M. F., & Kozlov, S. (2007). ATM activation and DNA damage response. *Cell Cycle*, 6, 931-942.

Lange, S. S., Takata, K., & Wood, R. D. (2011). DNA polymerase and cancer. *Nature reviews*, 11, 96-110.

Le Beau, M. M., Rassool, F. V., Neilly, M. E., Espinosa, R., III, Glover, T. W., & Smith, D. I. et al. (1998). Replication of a common fragile site, FRA3B, occurs late in S phase and is delayed further upon induction: implications for the mechanism of fragile site induction. *Hum. Mol. Genet.,* 7, 755-761.

Lehman, C. W., & Carroll, D. (1991). Homologous recombination catalyzed by a nuclear extract from *Xenopus* oocytes. *Proc. Natl. Acad. Sci. USA,* 88, 10840-10844.

Lemoine, F. J., Degtyareva, N. P., Lobachev, K., & Petes, T. D. (2005). Chromosomal translocations in yeast induced by low levels of DNA polymerase: a model for chromosome fragile sites. *Cell*, 120, 587-598.

Lengsfeld, B. M., Rattray, A. J., Bhaskara, V., Ghirlando, R., & Paull, T. T. (2007). Sae2 Is an endonuclease that processes hairpin DNA cooperatively with the Mre11/Rad50/Xrs2 complex. *Molecular Cell*, 28, 638-651.

Letessier, A., Millot, G. A., Koundrioukoff, S., Lachagès, A., Vogt, N., & Hansen, R. S. et al. (2011). Cell-type-sepcific replication initiation programs set fragility of the FRA3B fragile site. *Nature Letter*, 470, 120-124.

Mangelsdorf, M., Ried, K., Woollatt, E., Dayan, S., Eyre, H., & Finnis, M. et al. (2000). Chromosomal fragile site FRA16D and DNA instability in cancer. *Cancer Research,* 60, 1683-1689.

Mani, R., & Chinnaiyan, A. M. (2010). Triggers for genomic rearrangements: insights into genomic, cellular and environmental influences. *Nature Reviews*, 11, 819-830.

Matsuyama, A., Shiraishi, T., Trapasso, F., Kuroki, T., Alder, H, & Mori, M. et al. (2003). Fragile site orthologs FHIT/FRA3B and Fhit/Fra14A2: Evolutionarily conserved but highly recombinogenic. *Proc. Natl. Acad. Sci. USA*, 100(25), 14988-14993.

Mimitou, E. P., & Symington, L. S. (2008). Sae2, Exo1 and Sgs1 collaborate in DNA double-strand break processing. *Nature*, 455, 770-775.

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

Negrini, S., Gorgoulis, V. G., & Halazonetis, T. D. (2010). Genomic instability – an evolving hallmark of cancer. *Nature Reviews*, 11, 220-228.

O'Keefe, L. V., & Richards, R. I. (2006). Common chromosomal fragile sites and cancer: Focus on FRA16D. *Cancer Letters*, 232, 37-47.

Paige, A. J. W., Taylor, K. J., Stewart, A., Sgouros, J., Gabra, H., & Sellar, G. C. et al. (2000). A 700-kb physical map of a region of 16q23.2 homozygously deleted in multiple cancers and spanning the common fragile site FRA16D. *Cancer Research*, 60, 1690-1697.

Palakodeti, A., Han, Y., Jian, Y., & Le Beau, M. M. (2004). The role of late/slow replication of the FRA16D in common fragile site induction. *Genes Chromosomes Cancer*, 39, 71-76.

Palakodeti, A., Lucas, I., Jiang, Y., Young, D. J., Fernald, A. A., Karrison, T., & Le Beau, M. M. (2010). Impaired replication dynamics at the FRA3B common fragile site. Human Molecular Genetics, 19(1), 99-110.

Pâquest, F., & Haber, J. E. (1999). Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiology and Molecular Biology Reviews*, 63(2), 349-404.

Petrini, J. H. J., Bressan, D. A., & Yao, M. S. (1997). The RAD52 epistasis group in mammalian double strand break repair. *Seminars in Immunology*, 9, 181-188.

Popescu, N. C., Zimonjic, D., & DiPaolo, J. A. (1990). Viral integration, fragile sites, and proto-oncogenes in human neoplasia. *Human Genetics*, 84(5), 383-386.

Ragland, R. L., Glynn, M. W., Arlt, M. F., & Glover, T. W. (2008). Stably transfected common fragile site sequences exhibit instability at ectopic Sites. *Genes, Chromosomes & Cancer*, 47, 860-872.

Ricchetti, M., Fairhead, C., & Dujon, B. (1999). Mitochondrial DNA repairs double-strand breaks in yeast chromosomes. Nature, 402, 96-100.

Richards, R. I. (2001). Fragile and unstable chromosomes in cancer: causes and consequences. *TRENDS in Genetics*, 17(6), 339-345.

Richard, G., Kerrest, A., & Dujon, B. (2008). Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiology and Molecular Biology Reviews*, 72(4), 686-727.

Samadashwily, G. M., Raca, G., Mirkin, S. M. (1997). Trinucleotide repeats affect DNA replication in vivo. *Nat. Genet*, 17, 298-304.

Shah, S. N., Opresko, P. L., Meng, X., Lee, M. Y. W. T., & Eckert, K. A. (2009). DNA structure and the Werner protein modulate human DNA polymerase delta-dependent replication dynamics within the common fragile site FRA16D. *Nucleic Acids Research*, 38(4), 1149-1162.

Sidorova, J. M. (2008). Roles of the Werner syndrome RecQ helicase in DNA replication. *DNA Repair*, 7, 1776-1786.

Sobeck, A, Stone, S., & Hoatlin, M. E. (2007). DNA structure-induced recruitment and activation of the fanconi anemia pathway protein FANCD2. *Molecular and Cellular Biology*, 27(1), 4283-4292.

Toledo, F., Coquelle, A., Svetlova, E., & Debatisse, M. (2000). Enhanced flexibility and aphidicolin-induced DNA breaks near mammalian replication origins: implications for replicon mapping and chromosome fragility. *Nucleic Acids Research*, 28, 4805-4813.

Tuduri, S., Crabbe, L., Conti, C., Tourriere, H., Holtgreve-Grez, H., & Jauch, A. et al. (2009). Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. *Nature Cell Biology*, 11(11), 1315-1323.

Voineagu, I., Freudenreich, C. H., & Mirkin, S. M. (2009). Checkpoint responses to unusual structures formed by DNA repeats. *Molecular Carcinogenesis*, 48, 309-318.

Voineagu, I., Narayanan, V., Lobachev, K. S., & Mirkin, S. M. (2008). Replication

stalling at unstable inverted repeats: Interplay between DNA hairpins and fork stabilizing proteins. *PNAS*, 105(29), 9936-9941.

Wan, C., Kulkarni, A., & Wang, Y. (2010). ATR preferentially interacts with common fragile site FRA3B and the binding requires its kinase activity in response to aphidicolin treatment. *Mutation Research*, 686, 39-46.

Yunis, J. J., & Soreng, A. L. (1984). Constitutive fragile sites and cancer. *Science*, 226, 1199-1204.

Zhang, H., & Freudenreich, C. H. (2007). An AT-rich sequence in human common fragile site FRA16D causes fork stalling and chromosome breakage in *S. cerevisiae*. *Molecular Cell*, 27, 367-379.

Zhao, J., Bacolla, A., Wang, G., & Vasquez, K. M. (2010). Non-B DNA structure-induced genetic instability and evolution. *Cellular and Molecular Life Sciences*, 67, 43-62.

Zlotorynski, E., Rahat, A., Skaug, J., Ben-Proat, N., Ozeri, E., & Hershberg, R. et al. (2003). Molecular basis for expression of common and rare fragile sites. *Mol. Cell. Biol*., 23, 7143-7151.

Zucker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, 31, 3406-3415.

APPENDIX

Table 1. Previous Flex1 Bacterial Clones

Flex1 Sequences	Stock # (Date made)	Insert Source (Stock #)	Vector Source (Stock #)	Orientation
Long 5' end (AT)14-17	#338 (7/30/09)	pYES2 (AT)14-17 (#297)	pBL007 (#223)	1
Short 5' end (AT)28+17	#339 (7/30/09)	pYES2 (AT)34+17 (#295)	pBL007 (#223)	1
Short 5' end (AT)34+17	#340 (7/30/09)		pBL007 (#223)	1

Sequencing previously cloned Flex1 sequences verified three correct bacterial transformants: Long 5' end (AT)14-17, short 5' end (AT)28+17, and short 5' end (AT)34+17. Orientation refers to the orientation of the cloned Flex1 sequences with respect to the direction of plasmid replication. Orientation 1 indicates the orientation of the cloned Flex1 in which replication proceeds through the Flex1 sequence from its 5'-end. Orientation 2 indicates the orientation of the cloned Flex1 in which replication proceeds through the Flex1 sequence from its 3'-end, which is the same orientation as the direction of plasmid replication.

Table 2. New Flex1 Bacterial Clones

Flex1 Sequences	Plasmid Stock # (Date made)	Insert Source (Plasmid Stock #)	Vector Source (Plasmid Stock #)	Orientation
Long 5' end (AT)21-17	#352 (3/19/11)	pYES2 (AT)14+17 (#292)	pBL007 (#223)	2
Short 5' end (AT)23+17 (in both orientations)	#354, #355 (3/19/11)	pYES2 (AT)23+17 (#293)	pBL007 (#223)	#354 (1) #355 (2)
Short 5' end (AT)24+17	#350 (3/19/11)		pBL007 (#223)	1
Long 5' end (AT)23-17	#349 (3/19/11)	pYES2 (AT)23-17 (#298)	pBL007 (#223)	2
Short 5' end (AT)34-17	#351 (3/19/11)	pYES2 (AT)34-17 (#294)	pBL007 (#223)	2
[Short 5' end (AT)23+17]X3	#353 (3/19/11)	pRS426 (AT)34+17+(AT)23+17 (#248)	pBL007 (#223)	1

Showing Flex1 sequences with different combinations of 5' and 3' end sequences flanking the perfect (AT)n dinucleotide repeat in the Flex1 region. Sequencing confirmed eight correct bacterial transformants shown in the table. Orientation refers to the orientation of the cloned Flex1 sequences with respect to the direction of plasmid replication. Orientation 1 indicates the orientation of the cloned Flex1 in which replication proceeds through the Flex1 sequence from its 5'-end. Orientation 2 indicates the orientation of the cloned Flex1 in which replication proceeds through the Flex1 sequence from its 3'-end, which is the same orientation as the direction of plasmid replication.

Table 3. No Repeat Control Fragility Summary

	Rate of FOA^R (per 10^{-5}) for No Repeat Control	**Data Source**
1	3.84	Sara Kantrow (1996-1997)
2	5.50	Sara Kantrow (1996-1997)
3	5.28	Mayurika Lahiri (2003)
4	2.82	Mayurika Lahiri (2003)
5	3.32	Mayurika Lahiri (2003)
Average	4.15	

Showing values and sources for the no repeat control fragility data from five experiments, two of which are conducted Sara Kantrow, and the rest conducted by Mayurika Lahiri.

Table 4. Flex1 Fragility Summary

Flex1 Sequence	Rate of FOA^R (per 10^{-5})	Date
Long 5' end (AT)14-17	17.50	11/14/2010
	25.65	3/1/2011
	20.30	3/19/2011
	22.28	3/24/2011
Average	21.43	
Short 5' end (AT)28+17	7.72	11/14/2010
	6.56	3/1/2011
	9.14	3/19/2011
	9.99	3/24/2011
Average	8.35	
Short 5' end (AT)34+17	7.69	11/14/2010
	10.85	3/1/2011
	9.22	3/19/2011
	9.64	3/24/2011
Average	9.35	
Short 5' end (AT)34-17	230.04	3/19/2011
	228.67	3/24/2011
Average	229.36	
Long 5' end (AT)23-17	30.43	4/19/2011
[Short 5' end (AT)23+17]X3	21.74	4/19/2011

The rate of generation of FOA^R colonies is show for each strain.