Investigating *cis* and *trans* factors involved in common fragile site breakage and healing using Flex1, a subregion of FRA16D

A thesis

Submitted by

Simran Kaushal

In partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Biology

TUFTS UNIVERSITY

August 2018

Advisor: Dr. Catherine Freudenreich

Abstract

Common Fragile Sites (CFSs) are regions of DNA that display gaps and breaks in metaphase chromosomes under replication stress. While CFSs are not normally expressed, or broken, in individuals, conditions of replication stress result in their breakage. Evidence of such sites has been found in organisms ranging from yeast to humans. Further, CFSs are often sites of rearrangement in cancer cell lines, therefore understanding their expression is very relevant to human disease. CFSs are late replicating, and conditions of replication stress result in their under-replication at the point of chromatin condensation in the cell cycle. CFSs are enriched in DNA sequences that can form abnormal, or secondary, structures, which could also play a role in their breakage. There are several theories for the cause of CFS breakage, each with varying levels of support, and this is largely due to the difficulty of studying complicated DNA sequences in mammalian cells. Here, we studied Flex1, a roughly 300 bp subregion of CFS FRA16D in *S. cerevisiae* that has a perfect AT repeat that is highly polymorphic in humans and predicted to form stable secondary structures *in vivo*. Flex1 also stalls replication in an AT repeat length-dependent manner. Working in yeast allows precise genetic control of DNA sequences, making it an excellent model system to investigate current theories for CFS breakage. We have found that breakage at Flex1 is dependent on structure-specific endonuclease (SSE) complexes Mus81-Mms4, Slx1-Slx4, and Rad1-Rad10, similar to what has been found for human SSEs at FRA16D. Thus, Flex1 serves as a model system for studying breakage and healing at a CFS sequence. The cleavage of Flex1 by Mus81 is dependent on

the formation of a secondary structure by the AT repeats. By comparing the effects of different Flex1 sequences on breakage and healing, we have evidence to propose a new theory for CFS expression: CFSs are not only prone to breakage but also impaired in their ability to heal following fragility. We propose that breakage in FRA16D is initiated by SSE cleavage at Flex1, followed by difficulty healing after fragility due to the propensity of adjacent sequences to form hairpins. We also discovered a role for the fork stabilization function of the Mrc1 (hClaspin protein) in protecting against fragility at Flex1, supporting an important link between fork stalling and fragility.

Acknowledgements

Thank you to my committee chair and advisor, Dr. Catherine Freudenreich. I greatly appreciate your faith in the exciting things this project could teach us, even when it was very technically challenging. Thank you for your mentorship, support, and encouragement for the past 7 years. I have learned so much from being in your lab.

Thank you to my committee members for your guidance throughout my graduate career:

Dr. Sergei Mirkin for helpful scientific conversations and thoughtful and entertaining conversations about everything else in life.

Dr. Mitch McVey for your support and scientific conversations. Thank you for being an excellent model of a university professor who is very engaged in mentorship both in the lab and in the classroom. I hope to follow many of your examples someday.

Thank you also to my outside examiner, Dr. Sharon Cantor, for taking the time to read my thesis and contribute to the defense process.

I'd like to thank Dr. Nealia House, Dr. Xiaofeng Allen Su, Samantha Regan, Alice Haouzi, Charles Wollmuth, Julia Haft, Ruby Ye, Sasha Khristich, Michael Sigouros, and Alison Soo Mi Lee for scientific contributions to this project.

To Tufts Biology department members: Stephen Fuchs, Juliet Fuhrman, Susan Ernst, Eric Tytell, Francie Chew, Michael Levin, and George Ellmore, thank you for making my time at Tufts interesting and fruitful both academically and personally. I thank you all for your guidance and leaving me with very fond memories of Tufts and its faculty members. Thanks to Michael Grossi, Deirdre McCann, Anisha Chandra Mohan, Michael Doire, Eileen Magnant, Elizabeth Palmer, and Sanjukta Ghosh for administrative and day-to-day support.

Thank you to my inspirations before Tufts: Rima Garg –you're next! Sanjay Vashee, for being the most patient and inspiring mentor I could have hoped for in starting off my molecular biology research. Remember when designing a primer was the hardest thing imaginable for me? And last, but not least, to my dearest Rahul Mamu, Dr. Rahul Jasuja: my earliest memories were of your graduation at Tufts and I am so proud to be following in your footsteps nearly 20 years later. I finally understand that you were studying genes and not jeans all that time.

Thank you to my dear friends Ellie Camlin, Nealia C.M. House, Angeli Sivaraman, Biney Singh, Jessica Arbuthnot, Kara Constantine, my baby friends Maya, Iselin, Emmett, Willow, and Orson (and the Base Pair Robins), and to department and labmates for support.

Thank you to the McGintys for helping me find a home in Massachussetts. And finally, shukriya to Atul, Superna, and Sohum Kaushal, Nani and Nanu, and Daadi for being such a loving, honest, and funny family and for teaching me the value of hard work and education.

List of Abbreviations

ATR Ataxia Telangiectasia and Rad3 related

BACH1=FANCJ BRCA1-Associated C-terminal Helicase 1 = Fanconi Anemia complementation group J

BIR Break-Induced Replication

bp Base-Pair

BTR BLM helicase-Topo IIIα-RMI1-RMI2 complex in humans

BrdU BromoDeoxyUridine

CAA ChlorAcetAldehyde

CFSs Common Fragile Sites

CldU ChlorodeoxyUridine

ChIP Chromatin ImmunoPrecipitation

Ctrl Control strains used in various assays

CtIP C-terminal binding protein Interacting Protein

DDRA Direct Duplication Recombination Assay

DNA DeoxyRibonucleic Acid

dsDNA Double-Stranded DNA

G_1 phase Gap 1 Phase

G_2 phase Gap 2 Phase

hg Human Genome build number

HU Hydroxyurea

IdU IododeoxyUridine

IR Inverted Repeat

kb KiloBases of DNA

M phase Mitotic Phase

Mb MegaBases of DNA

MCM2-7 mini-chromosome maintenance complex components 2-7

MCM3 mini-chromosome maintenance complex component 3

MiDAS Mitotic DNA Synthesis

NHEJ Non-Homologous End Joining

nt NucleoTide

ORC Origin Recognition Complex

PATRRs Palindromic AT-Rich Repeats

PCR Polymerase Chain Reaction

qPCR Quantitative Polymerase Chain Reaction

RACE Rapid Amplification of cDNA Ends

RNA RiboNucleic Acid

Rrm3 rDNA Recombinational Mutation 3 protein

S phase Synthesis phase

SCE Sister Chromatid Exchange

SDSA Synthesis-Dependent Strand-Annealing

SEM Scanning Electron Microscopy

Sgs1 Slow Growth Suppressor 1 gene

SMARD Single Molecule Analysis of Replicated DNA

SMX SLX1-SLX4, MUS81-EME1, and XPF-ERCC1 super complex in mammals

SSA Single-Strand Annealing

ssDNA Single-Stranded DNA

STR Sgs1-Top3-Rmi1 complex in *S. cerevisiae*

TLS TransLesion Synthesis

UFB Ultra Fine Bridge

WT WildType

WWOX/FOR WW-domain containing OXidoreductase/Fragile site FRA16D OxidoReductase

YAC Yeast Artificial Chromosome

Table of Contents

Chapter 1

Introduction: Common Fragile Site Breakage: Causes and Consequences

Abstract

Common Fragile Sites (CFSs) are regions of DNA that display gaps and breaks in metaphase chromosomes under replication stress. CFSs tend to be AT-rich and have been hypothesized to form secondary structures that play a role in their fragility. This chapter will review the background and significance of studying CFSs, the role that DNA sequence and structure plays in replication fork stalling, and the consequences of CFS breakage and having secondary structures in the genome.

Fragile sites

In 1970, the term fragile site was used to refer to recurrent breaks found near the haptoglobin locus of human chromosome 16 (Magenis et al., 1970). Since then, the term fragile site has been expanded to define loci that exhibit chromosome fragility as gaps and breaks on metaphase chromosomes. There are two main classes of fragile sites: rare and common fragile sites (CFSs), which are defined by their frequency of expression (or breakage) in the population. Rare fragile sites are present in less than 5% of individuals, are usually caused by expanded repetitive DNA elements, and are inherited in a Mendelian fashion (Durkin and Glover, 2007, Schwartz et al., 2006). Common fragile sites are present in all

individuals and are only expressed after partial inhibition of DNA synthesis. CFS expression is often induced by drugs that cause replication stress, such as aphidicolin, which inhibits DNA polymerase α and DNA polymerase δ elongation, or folate deficiency plus hydroxyurea (HU), which depletes dNTP pools (Yan et al., 1987, Glover et al., 1984), and caffeine, which inhibits the kinase Ataxia telangiectasia and Rad3 related (ATR, which detects stalled forks and activates the DNA damage checkpoint) (Durkin and Glover, 2007). Over 75 CFSs have been identified to date (NCBI).

Common Fragile Sites

Common fragile sites were first described in the literature in 1984 by Glover et al. as sites prone to gaps and breaks at certain locations of the same chromosomes under conditions of partial DNA synthesis inhibition (Glover et al., 1984). CFSs are very large regions of DNA, usually hundreds of kilobases in length, hence their visibility on metaphase chromosome spreads. The two most highly broken, or expressed, human CFSs are FRA3B and FRA16D. Interestingly, there is evidence for CFSs in many species, including primates (Arlt et al., 2003, Ruiz-Herrera et al., 2004) and mice (Shiraishi et al., 2001, Krummel et al., 2002, Rozier et al., 2004). Further, some CFSs are conserved across mice, DT40 chicken cells, and mammals (Le Tallec et al., 2013). Even *S. cerevisiae* have replication slow zones (RSZs) which are akin to fragile sites (Cha and Kleckner, 2002, Cimprich, 2003, Lemoine et al., 2005). Considering their conservation across species, CFSs may serve an important adaptive purpose in evolution.

In humans, CFSs are located in ~50% of reoccurring cancer-associated deletions (Le Tallec et al., 2013) and cancer-specific chromosomal translocations (Dillon et al., 2010), and therefore they are very relevant to human health. While CFSs have been studied for almost 35 years, the molecular basis for their expression is still poorly understood. There are currently several different theories for CFS expression that are not mutually exclusive (Figure 1-1).

Common fragile sites and replication impairment

The best-supported theory for CFS fragility is an inability to complete replication in S phase. CFS expression can be induced by drugs that impair replication, such aphidicolin, folate deficiency plus HU, or by drugs that inhibit DNA repair such as caffeine, 1-β-D-arabinofuranosyl-cytosine, and 5-floroxodeoxyuridine (Li et al., 1986a, Li and Zhou, 1985, Li et al., 1986b, Zhou et al., 1984, Glover et al., 1984). Thymidine analog bromodeoxyuridine (BrdU) incorporation colocalization with CFS FRA3B FISH probes revealed the region is normally replicated late in S phase and can remain unreplicated into G_2 phase; replication is further delayed by exposure to aphidicolin (Le Beau et al., 1998). Quantitative PCR (qPCR) of BrDU-labeled DNA indicated that FRA16D is also replicated in late S and G_2 phases (Palakodeti et al., 2004). Numerous other CFSs have also been shown to be late replicating (Glover et al., 2017, Durkin and Glover, 2007). Molecular combing experiments indicated that replication proceeds through CFS FRA16C at a slower rate (1.67 kb/min) than bulk genome replication (2.2 kb/min) (Ozeri-Galai et al., 2011).

Inability to fire backup origins at CFSs

The Kerem lab conducted molecular combing concurrent with sequential pulse labeling with thymidine analogs iododeoxyuridine (IdU) followed by chlorodeoxyuridine (CldU) to map the average origin density genome-wide and at CFS FRA16C under normal and aphidicolin-induced replication stress conditions (Ozeri-Galai et al., 2011). Genome-wide, the distance between two sister forks is significantly reduced from 109 ± 7 kb to 90 ± 5 kb under conditions of replication stress. This is expected, as under normal conditions only some licensed replication origins are activated (Gilbert, 2007). After replication stress, dormant origins are activated to lessen the replication load through the region (Ge et al., 2007, Ibarra et al., 2008), resulting in a shorter distance between origins. Aphidicolin-induced mild replication stress resulted in a significant reduction in fork distance genome-wide, while the fork distance at FRA16C was the same under normal growth (81 ± 10 kb) and replication stress conditions (81 ± 16 kb). These results indicated that at FRA16C, all dormant origins are already activated under normal conditions and there are no additional origins to activate under perturbed replication conditions (Ozeri-Galai et al., 2011). siRNA knockdown of MCM3, one of the proteins needed to license replication origins, resulted in an increase in expression of FRA3B, FRA16C, and FRA16D, indicating that a high density of activated origins is needed to prevent fragility at CFSs (Ozeri-Galai et al., 2011). Altogether, these data show that some CFSs have perturbed replication even in normal cellular conditions and maintain stability by firing additional origins. Under conditions of replication stress, few to no additional origins are

able to be fired at CFSs, leading to their under-replication and the expression of gaps and breaks on metaphase chromosomes. A similar study also found a reduction in fork speed and inter-origin distance at CFS FRA6E (Palumbo et al., 2010). A 50 kb subregion of FRA3B showed four origins that were active but not as efficient as those in nonfragile regions by analyzing nascent DNA abundance by microarrays (Palakodeti et al., 2010).

Several translesion synthesis (TLS) polymerases are recruited to CFSs in order to prevent their expression (Bhat et al., 2013, Rey et al., 2009, Mansilla et al., 2016). The Eckert lab investigated human polymerase δ holoenzyme transit throughout several predicted structure-forming subregions from CFSs. They investigated polymerase progression *in vitro* through a non-AT rich inverted repeat (IR) next to a A19 run from FRA16D, an (AT)25 repeat near an A22 run from FRA3B, and an interrupted (AT)24 repeat followed by a A28 run (Flex5) from FRA16D. They found that human translesion synthesis (TLS) polymerases η and κ can switch with the replicative polymerase δ holoenzyme when it is paused specifically at inverted repeats (IRs) and mononucleotide runs. (Barnes et al., 2017).

Paucity of origins at CFSs

The origin recognition complex (ORC) binds open DNA at replication origins in order to assemble the pre-replication complex. There is some evidence to support that CFSs are origin-poor, for example the association of 73% of CFSs with ORC2-poor regions (Miotto et al., 2016). In 2011, the Debatisse lab investigated replication dynamics through FRA3B and their results supported an alternative

theory regarding replication origins at CFSs. This study utilized molecular combing across a 1.5 Mb region of FRA3B DNA in lymphoblastoid and fibroblast cells as FRA3B is fragile in the former cell type but not the latter. They found that FRA3B was origin-poor specifically in the fragile cell type and found no change in replication fork speed across the site, but rather a paucity of replication of origins compared to neighboring nonfragile DNA. This would require replication forks to cover very long distances in order to replicate CFSs, which could lead to their under-replication come the end of the cell cycle (Letessier et al., 2011). In support of this hypothesis, the Schildkraut lab found that FRA16D and FRA6E have a paucity of replication origins in lymphocytes and that FANCD2 assists in the firing of dormant origins to complete replication through both sites, possibly due to a role in chromatin looping or histone chaperone activity (Madireddy et al., 2016). Replication origins are altered throughout differentiation (Dazy et al., 2006, Hiratani et al., 2008), are conserved across cell types, and are thought to be controlled by differences in epigenetic marks, chromatin domains (Hansen et al., 2010, Ryba et al., 2010), and transcriptional usage (Gregoire et al., 2006, Hiratani et al., 2008) in different cell types. Replication origin usage could be one explanation for differences in CFS breakage in different cell types (Le Tallec et al., 2011, Hosseini et al., 2013, Le Tallec et al., 2013).

FRA3B was mapped to a region of about 4.5 Mb, so neither replication dynamics study (Palakodeti et al., 2010, Letessier et al., 2011) evaluated replication across the entire CFS sequence (Becker et al., 2002). Thus, it is possible that both

theories can serve as explanation breakage at CFSs. See Figure 1-1 for illustrations of the various theories for CFS expression. The top left part of Figure 1-1 illustrates the consequences of either paucity of origins or inability to fire backup origins. It has been established that replication speed determines chromatin looping and origin usage in mammalian cells, further connecting the proposed explanations of replication dynamics at CFSs (Courbet et al., 2008).

In 2012, the Eckert and Makova labs performed a genome-wide analysis of CFSs by converting aphidicolin-induced CFS cytogenetic locations (Mrasek et al., 2010) to genomic coordinates. In agreement with previous work, they found that CFSs were very large, ranging from 0.7-25 Mb in length. Several of their findings were different from what was previously known about CFSs. Genome-wide, they did not find a correlation between late replication or low origin density and CFSs. They found that aphidocolin-induced CFSs are located distant from centromeres, not in G-bands, have high DNA flexibility, and are enriched in *Alu*-repeats (Fungtammasan et al., 2012).

However, there are several caveats of the Eckert and Makova study. Notably, this study did not evaluate FRA16D and 3 other CFSs because the mapping they did was not validated by comparison to more detailed studies using FISH mapping to create a higher resolution map. Thus, it is possible that the method of converting aphidicolin-induced CFS cytogenetic locations to genomic coordinates may not be as reliable for defining CFS regions. Replication timing changes by cell type and developmental stage, and the three replication timing studies the Fungtammasan et al. referenced all had their own caveats. Notably, simple repeats

were excluded in a 2010 replication timing study they referenced (Hansen et al., 2010). While Fungtamassan et al. did not find any correlation between replication origin density and aphidicolin-induced CFSs, the computational and experimental origin-mapping data they referenced did not agree with each other in terms of origin location. This brings into question the quality of the computational and experimental origin mapping data they were comparing the CFSs to; if the data from the two methods don't agree with one another, then each method is likely only capturing a certain subset of origins. These findings also don't take into account replication origin usage and efficiency. It will be interesting to know how higher quality sequences of repetitive DNA and CFSs in the human genome will affect the results of genome-wide CFS studies.

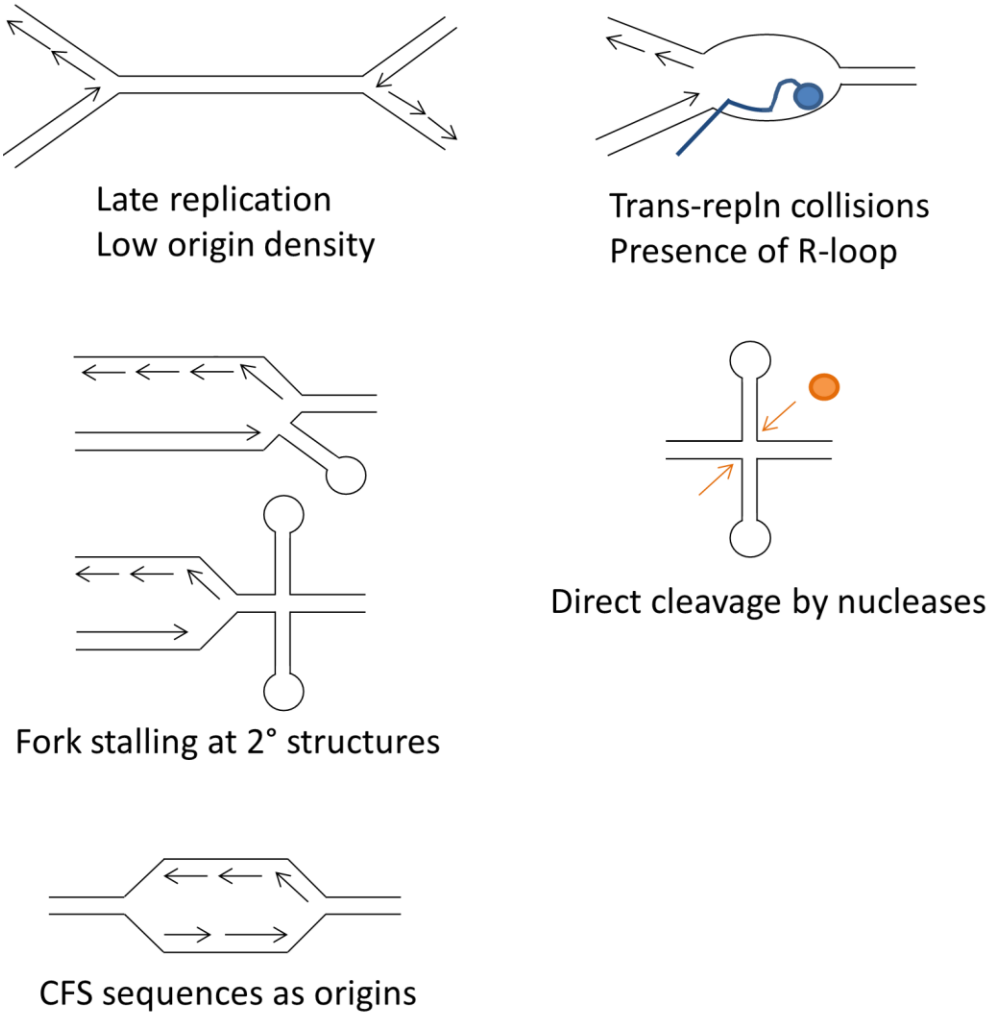


Figure 1-1. Proposed theories for CFS fragility. CFS fragility may be due to issues with replication fork stalling at secondary structures in CFS sequences, a low density of replication origins, transcription-replication conflicts, and direct cleavage by nucleases. Note that all theories may not be mutually exclusive in accounting for CFS expression.

CFS expression due to transcription: a controversial theory

Transcription and transcription-replication collisions have also been linked to CFS expression. CFSs tend to be in present in large genes between 300 kb and 2 Mb in length (Durkin and Glover, 2007, Smith et al., 2006, Gao and Smith, 2014, Le Tallec et al., 2013, McAvoy et al., 2007). CFS expression also varied by cell type when tested in epithelial, erythroid, and fibroblast cell lines (Hosseini et al., 2013, Le Tallec et al., 2011, Le Tallec et al., 2013) as it is known that transcriptional profiles vary across cell types and thus transcription through CFSs was proposed to play a role in their fragility.

There is some evidence to support a role for R-loops in CFS fragility. R-loops are RNA:DNA hybrids, which can directly cause fragility by impairing replication fork and transcription bubble procession through the regions (Aguilera and Garcia-Muse, 2012). Helmrich et al. found evidence for *in vivo* R-loops at FHIT/FRA3B which were reduced upon the addition of RNaseH (Helmrich et al., 2011). Further, siRNA knockdown of RNaseH1 resulted in an increase in the number of breaks detected at CFSs FRA3B, FRA16D, and FRA7K, indicating that RNaseH1 functions to stabilize CFS regions by removing RNA:DNA hybrids (Helmrich et al., 2011).

Some very recent publications have connected transcription to replication origin usage. The Debatisse lab found that high transcription of large genes resulted in a shift in their replication pattern from late to mid-S phase, likely giving the cells more time to complete synthesis of the regions before M phase (Blin et al., 2018

BioRxiv.org https://doi.org/10.1101/286807). Duncan Smith's lab has also recently found that replication stress results in redistribution of replication termination relative to transcription (Chen et al., 2018 BioRxiv.org https://doi.org/10.1101/324079). The Schildkraut lab found evidence that FANCD2 facilitates replication through CFSs FRA16D and FRA6E by resolving RNA:DNA hybrids (Madireddy et al., 2016). Thus, there seems to be a feedback loop between transcription and replication, which could explain conflicting evidence for each molecular process's role in CFS expression.

The role of transcription in CFS expression is still not well understood. It is possible that transcription does contribute to fragility in certain cellular contexts: perhaps at certain fragile sites or subregions of fragile sites, fragility is governed by transcription.

DNA secondary structures and their association with CFSs

Originally, CFS boundaries were cytogenetically defined using colocalization of FISH probes, even before the entirety of the human genome sequence was available (Wilke et al., 1996). CFSs and repetitive DNA are notoriously difficult to subclone and sequence. AT-rich sequences are difficult to subclone and maintain in plasmids, which is often an intermediate step in generating libraries for sequencing. Repetitive sequences are also difficult to PCR amplify and sequence. The fragment size of most popular NextGen sequencing technologies is roughly 150 bp – 300 bp, while repetitive DNA can be larger than those fragment lengths – this makes it impossible to align and quantify the number of repeats in

repetitive DNA. These technical challenges have resulted in poor coverage and sequence quality of repetitive regions of genomes. It is our hope that the advent of sequencing technologies with a longer fragment length such as Oxford Nanopore will aid in the proper sequencing coverage needed to understand the exact sequence composition at fragile regions of the genome.

CFSs tend to be AT-rich (Zlotorynski et al., 2003, Mishmar et al., 1998). Studies of DNA replication dynamics indicated that AT-rich sequences with the potential to form secondary structures block replication forks in FRA6E, FRA16C, and FRA16D (Palumbo et al., 2010, Ozeri-Galai et al., 2011, Madireddy et al., 2016). Thus, it was postulated that CFSs may be composed of multiple subregions subject to fragility fork stalling.

The Kerem lab developed the FlexStab program, which uses flexibility between dinucleotides in 100 bp windows to predict regions of high flexibility by DNA sequence (Mishmar et al., 1998). They studied DNA sequences enriched for flexibility between bases as this may affect protein-DNA interactions and chromatin condensation. It is our opinion that flexibility between nucleotides is not the determining factor in CFS fragility. Rather, the FlexStab program often has peaks at AT-rich regions, especially regions containing perfect AT dinucleotides because the A to T base-pair step has the highest predicted flexibility of all base-pair combinations. While it is true that such sequences are enriched for flexibility between bases, they are also more likely to participate in intrastrand base pairing to form secondary structures. Research from the Freudenreich lab and many others supports the importance of secondary structure

forming capability of sequences in replication fork stalling and fragility, therefore we feel the secondary structure hypothesis is a better explanation for the contributing factor of sequence to CFS expression.

An early study by the Kerem lab closely monitored CFS FRA7H. By FISH, they show integration of an SV40 site into FRA7H – this supports the claim that CFSs are hotspots for viral integration (see Consequences of CFS fragility section below). They defined FRA7H as a 161 kb region and sequenced it. They saw some evidence of unusual chromatin organization at FRA7H because of FISH signals that deviated in orientation from the expected orientation, but these could also be due to genomic rearrangements so this was not definitive. They also analyzed published FRA3B sequence and found clusters of high flexibility and low stability, meaning higher likelihood of forming secondary structure. (Mishmar et al., 1998).

A few years later, the Kerem lab used FISH colocalization near gaps and constrictions to map FRA7E (Zlotorynski et al., 2003). They saw evidence of 2 fragility "hot spots" that were too small to be spotted by cytogenetics, which is evidence for the idea of multiple regions of secondary structure and fragility in CFSs. They find that G-band fragile sites are enriched in clusters of high flexibility peaks compared to nonfragile G-band DNA (the same had previously been shown for R-band fragile sites). The flexibility peaks were AT-rich and enriched in AT and TA dinucleotides, similar to known rare FSs, further supporting that these flexibility prediction programs are enriching for AT-rich DNA which can nucleate to form secondary structures. The Richards lab used the

TwistFlex program (a newer version of the old Kerem lab program FlexStab), which identified the Flex1 and Flex5 subregions of FRA16D (Ried et al., 2000).

Flex1 is the highest flexibility peak predicted in FRA16D by the FlexStab program, and many cancer cell breakpoints correspond with some of the predicted FlexStab fragility peaks (Ried et al., 2000, Finnis et al., 2005, Mangelsdorf et al., 2000). Flex1 is a roughly 300 bp subregion of FRA16D that contains a perfect AT dinucleotide repeat that is highly polymorphic in humans (Finnis et al., 2005). Work from the Freudenreich lab showed that as AT repeat length increased at Flex1, replication fork stalling increased as measured by 2D gels (Zhang and Freudenreich, 2007) (see Figure 1-2). While these data collectively show an indirect correlation between structure-forming capability and fragility, there was still no direct evidence for the role of those regions in CFS expression, unlike the copious connections that have been found in rare fragile sites (see next section).

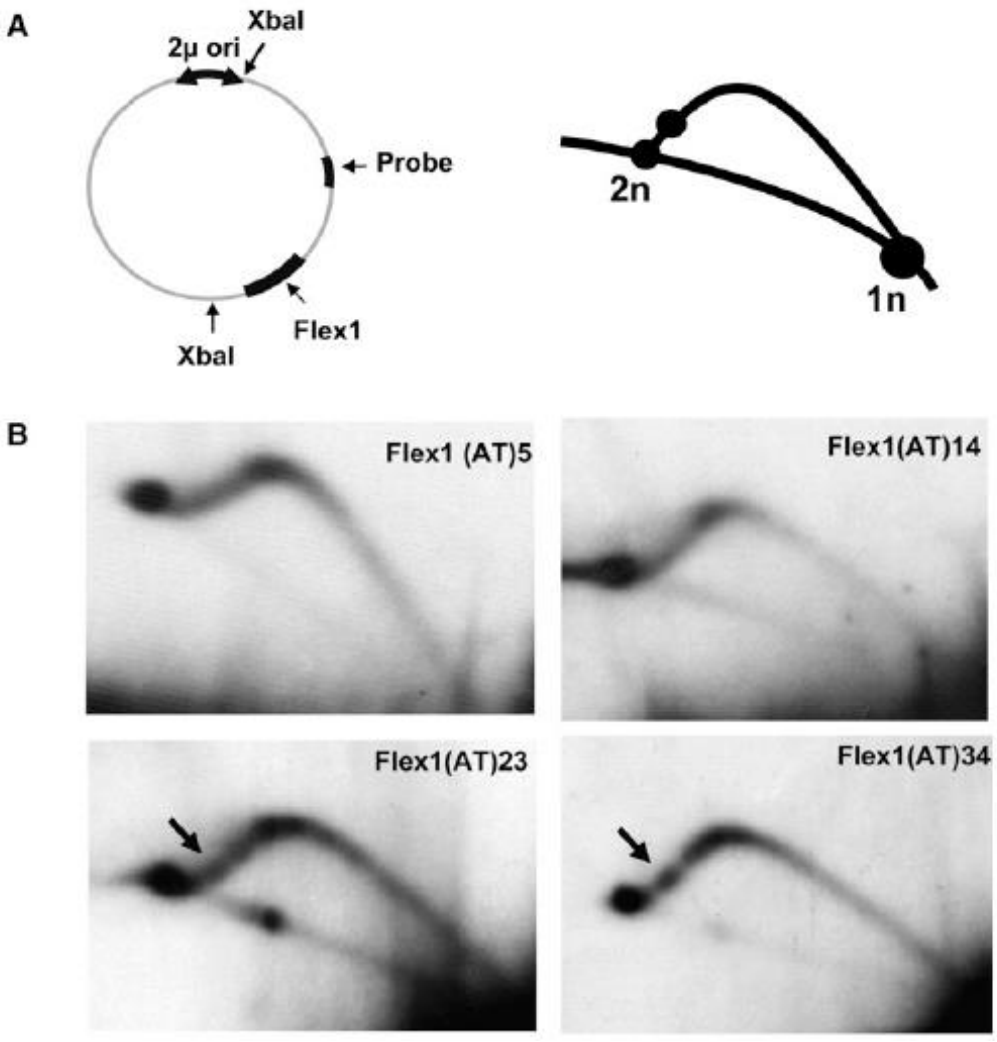


Figure 1-2. Flex1 AT repeats stall replication forks in a length dependent manner on *S. cerevisiae* plasmids. A) Diagrams of the plasmid used for 2D gels. B) 2D gels showing replication intermediates at Flex1 (AT)n. Figure adapted from (Zhang and Freudenreich, 2007).

Yuh-Hwa Wang's lab used Mfold (Zuker, 2003) to predict secondary structure forming potential of 300 nt segments (this is the size of an Okazaki initiation zone in mammalian cells and thus could reasonably be single-stranded during DNA replication) with a 150 nt shift window. Then they grouped 50 of these windows together to look for clusters of sequences forming stable secondary structures with a more negative ΔG and found CFSs FRA10G, FRA10D, and FRA10F had an enrichment in low free-energy segments per section compared to non-fragile DNA. They used Mfold to query the region of FRA3B and FRA16D and each had eight and three regions, respectively, that are predicted to form stable secondary structures; these regions also correlate with known breakpoint and LOH regions in various types of cancers. They tested the ability of sequences to form a secondary structure by an *in vitro* assay using DNA fragments that are predicted to form secondary structures, denaturing them and allowing them to re-duplex in various concentrations of NaCl to allow re-annealing of single strands after denaturation and then running out on a gel with and without the re-duplexing treatment – these data validate the use of Mfold to predict secondary structure formation by DNA sequence. They found a correlation between predicted fragile regions by their threshold with known deletion/insertion/point mutation/copy number alteration sites in many different types of cancers. They did not see an overrepresentation of LINEs, SINEs, LTRs, DNA transposons, or simple repeats in aphidicolin-induced CFSs on chr10 compared to the human genome as a whole, suggesting that there are no obvious repeat sequences responsible for their

fragility. The variation of DNA sequence features was also found when comparing the 6 most commonly expressed CFSs (Ried et al., 2000).

Secondary structures and fragility at non-CFS sequences

When trinucleotide repeats expand beyond a certain threshold they can play a causative role in many diseases. For example, expansions of CGG repeats cause expression of a rare fragile site. Two well-studied CGG expansions cause fragile X syndrome and FRAXE mental retardation. Expanded CGG repeats can adopt abnormal secondary structures including hairpin and G-quadruplex structures, and CGG repeats cause replication fork stalling in a plasmid-based system in yeast and primate cells (Voineagu et al., 2009). Expanded CGGs also cause chromosome fragility in the yeast DDRA system used in Kaushal Thesis Chapter 2 (Balakumaran et al., 2000).

CAG repeat expansion is a cause of heritable degenerative diseases such as Huntington's disease and myotonic dystrophy type 1 (Usdin et al., 2015). Contractions and expansions of the repeats occur due to their propensity to form stable hairpins when the DNA becomes singe stranded during processes like replication and transcription (Usdin et al., 2015, Polleys et al., 2017). CAG/CTG repeats cause fork stalling in mammalian cells (Cleary et al., 2002) and cause fork stalling and reversed forks when evaluated in *S. cerevisiae* (Nguyen et al., 2017). Expanded CAG/CTG repeats also cause fragility when inserted on a yeast chromosome (Freudenreich et al., 1998) and cells from myotonic dystrophy

patients containing expanded CTG repeats have increased formation of micronuclei, a consequence of chromosome breakage (Usdin et al., 2015).

Expanded GAA repeats within the frataxin gene cause Friedrich's Ataxia. Expanded GAA repeats can form triplex DNA and cause fork stalling in *S. cerevisiae* (Krasilnikova and Mirkin, 2004). GAA repeats are also fragile sites on yeast chromosomes (Kim et al., 2008) and in human cells (Kumari et al., 2015).

Inverted repeats (IRs) are two sequences with complementary DNA on the same strand that are facing towards one another. IRs can form hairpin-like secondary structures, even when both repeats are separated by an intervening DNA sequence. IRs can be formed by many different types of sequences, including (AT)n repeats and *Alu*-repeats. In mammalian cells, IRs cause replication fork stalling as analyzed by 2D gels and they also underwent mutagenesis (Lu et al., 2015). In 90% of the cases, mutagenesis was caused by a deletion of the entire IR sequence, however 10% of mutants had duplications in the stem and small deletions or IR arms, evidence of the ability of IRs to form a cruciform structure. Computational analysis of almost 20,000 breakpoints from cancer genomes revealed that small IRs of 7-30 bp in length were enriched within 200 bp of translocation breakpoints. It was postulated that collision of incoming replication forks with stalled forks at IRs could be responsible for DSBs and rearrangements (Lu et al., 2015). Finally, IRs are fragile sites both on yeast chromosomes (Lobachev et al., 1998) and in human cells (Lu et al., 2015).

Palindromic AT-rich repeats (PATRRs) can form large secondary structures and are found at the breakpoints of many recurrent translocations, such as the t(11;22) translocation between PATRR11 and PATRR22. Mild replication stress leads to deletions at PATRR11, similar to what has been found at CFSs (Kurahashi et al., 2009). Interestingly, PATRR11 resides within the CFS FRA11G sequence (Fechter et al., 2007).

Secondary structures can also play important roles in replication initiation and gene expression. The 2D3 antibody binds the base of a cruciform (Steinmetzer et al., 1995). The antibody enhanced the initiation of DNA replication at cruciforms in monkey CV-1 cells (Zannis-Hadjopoulos et al., 1988) and specifically labelled nuclei from the G_1/S boundary and throughout S phase in mammalian cells (Ward et al., 1990).

It has been shown that CFSs are breakage-prone regions of DNA likely to be enriched in secondary structure forming capability. The Kerem lab's computational analysis revealed that CFSs have high 78% A/T content and are enriched in interrupted AT/TA dinucleotide repeats, similar to the expanded AT-rich repeats associated with BrDU and distamycin A-induced rare fragile sites FRA10B (Hewett et al., 1998) and FRA16B (Yu et al., 1997). A multiple sequence alignment revealed that rare and common fragile sites were similar in sequence and organization. Distamycin A or BrDU were used to induce expression of rare fragile sites FRA16B and FRA10B, and FISH revealed that they colocalize to the same regions as the common fragile sites FRA16C and FRA10E, respectively (Zlotorynski et al., 2003, Yu et al., 1997). However, unlike

rare fragile sites, a direct link between DNA structure and fragility has yet to be demonstrated at CFS sequences, an aim of the studies in Chapter 2 (Kaushal et al. 2018, submitted). The FRA3B sequence that the Debatisse lab evaluated contained the 50 kb region with inefficient origins studied by (Palakodeti et al., 2010) but did not include two long AT-rich sequences that would be predicted to cause fork stalling and impair fork speed due to an ability to form abnormal secondary DNA structures. Therefore, the two theories of replication dynamics (inability to fire backup origins versus paucity of origins) across CFSs are not mutually exclusive.

AT-rich sequences can function as origins

Spinocerebellar ataxia type 10 is a progressive ataxia associated with unsteady gait and upper limb ataxia, in addition to scanning dysarthria and dysphagia. Its phenotype is associated with an expanded ATTCT repeat at the *ATXN10* locus. ATTCT repeats do not form secondary structures, but rather function as DNA unwinding elements; the repeats form an unpaired duplex that is detectable by atomic force microscopy (Potaman et al., 2003). The repeats can function as a replication origin in lymphoblastoid cells (Liu et al., 2007). ATTCT repeats are also fragile, expand in length, and result in premature transcription termination when studied in *S. cerevisiae* (Cherng et al., 2011).

In her thesis, Haihua Zhang from the Freudenreich found that a plasmid with CFS FRA16D subregion Flex1 (AT)34 (but not plasmids with shorter repeat lengths) are able to self-replicate in yeast (Haihua Zhang Thesis Figure 3.5). Therefore, it

was hypothesized that the AT repeats could function as a replication origin. 2D gel analysis of plasmid replication intermediates showed strong double Y structures (Haihua Zhang Thesis Figure 3.6), implying that Flex1 (AT)34 plasmid are maintained by recombination activity, however AT repeats functioning as an origin was not completely ruled out as a possibility.

The Debatisse lab found that mammalian DNA replication origins are enriched in flexibility peaks, as has been found for CFSs (Toledo et al., 2000). The same study also found that two mammalian DNA replication origins had an increased sensitivity to aphidicolin treatment, strengthening the connection between replication origin sequences and CFS fragility (Toledo et al., 2000).

Structure-specific endonucleases and Mitotic DNA Synthesis at CFSs

Work from the Debatisse, Rosselli, and Hickson labs showed that human complexes MUS81-EME1 (*S. cerevisiae* Mus81-Mms4) and XPF-ERCC1 are required for sister chromatid separation and CFS expression (Naim et al., 2013, Ying et al., 2013), pointing to a role for structure-specific endonucleases (SSEs) at CFSs. The Hickson lab proposed that SSEs induce local fragility at CFSs to avoid global genomic fragility if the sister chromatids are mechanically separated while still unreplicated and attached to one another, see Figure 1-3 (Ying et al., 2013). If CFSs do not have SSE-induced cleavage and sister chromatid separation, DNA bridges can persist during nuclear division, resulting in chromosome missegregation and 53BP1 body formation in G_1 phase (Ying et al., 2013). Mammalian cells depleted for MUS81 by short hairpin RNA had increased

FRA16D-containing micronuclei (Figure 1-3) (Ying et al., 2013). Since these studies used whole chromosomes, they could not point to a specific DNA sequence or sequences that were recruiting SSEs.

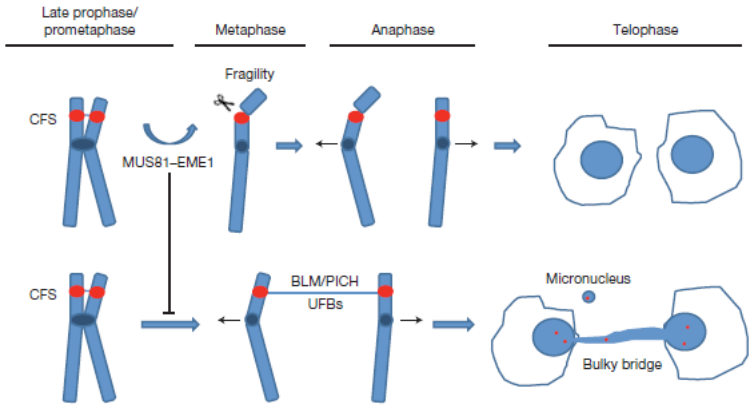


Figure 1-3. hMUS81 activity at sister chromatids. MUS81 acts to create local fragility in order to avoid worse global consequences. Figure is from (Ying et al., 2013).

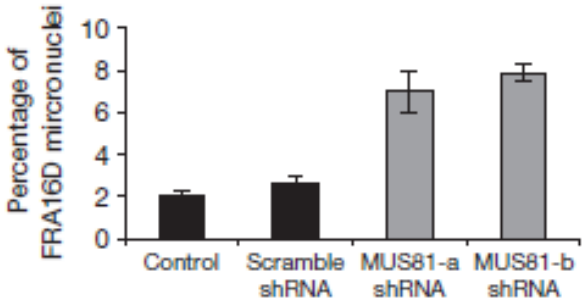


Figure 1-4. Reduction of hMUS81 expression results in decreased FRA16D fragility. Figure is from (Ying et al., 2013).

CFSs have difficulty replicating, and the Hickson lab found that under replication stress, CFSs incorporate the DNA synthesis label 5-ethynyl-2'-deoxyuridine (EdU) during mitosis. This mechanism, Mitotic DNA Synthesis (MiDAS), is dependent on MUS81 activity. MiDAS is a mechanism for cells to avoid the negative consequences of CFS expression, including chromosomal ultra-fine DNA bridges connecting sister chromatids and 53BP1 nuclear bodies that form around unprocessed CFSs. They also found that the SSE scaffold SLX4 helps recruit MUS81 to CFSs. (Minocherhomji et al., 2015). The Hickson lab hypothesizes that chromatin condensation at the end of G_2 triggers SSE cleavage and resumption of DNA synthesis via POLD3-dependent MiDAS. However, it was not understood whether a specific sequence or structure was being recognized by and cleaved by SSEs.

Molecular consequences of CFS expression

CFSs are hotspots for sister chromatid exchange (SCE) and viral integration, which are unsurprising consequences of their fragility (Glover and Stein, 1987, Wilke et al., 1996). In mammalian cells cultured under replication stress conditions, they are hot spots translocations and deletions (Wang et al., 1997, Glover and Stein, 1987, Glover and Stein, 1988). They also initiate breakage events involved in gene amplification via breakage-fusion-bridge cycles, which is a common occurrence in cancer progression (Coquelle et al., 1997).

CFSs are preferentially sensitive to oncogene-induced DNA damage in pre-neoplastic lesions (reviewed in (Negrini et al., 2010)). Inhibition of ATR results in CFS expression (Casper et al., 2002).

FRA3B is located within the *FHIT* gene and is considered the most highly expressed CFS (Huebner et al., 1998). FHIT is a putative tumor suppressor, and large intragenic deletions have been found in FRA3B in various tumor cells (Huebner et al., 1998, Glover, 1998).

FRA16D is considered the second most highly expressed CFSs in humans and is located within the putative tumor suppressor gene WWOX (Glover et al., 1984). It is fragile in multiple cell types, indicating that its fragility is governed by something inherently fragile in its sequence (Le Tallec et al., 2013). WWOX-associated deletions within FRA16D have been identified in breast, colon, esophageal, lung, ovarian, and prostate carcinomas (Bednarek et al., 2000, Mangelsdorf et al., 2000, Paige et al., 2000, Ried et al., 2000). Adenocarcinomas

of the colon, breast, lung, stomach and ovary have been observed to possess deletions at FRA16D (Bignell et al., 2010).

One analysis of 3,132 specimens from 26 cancer types identified three of the ten most frequent deletion locations were present within the genes FHIT, PARK2 and WWOX, each containing CFSs FRA3B, FRA6E, and FRA16D, respectively (Beroukhim et al., 2010). FRA16D is also located near the oncogene c-*MAF*, which has been associated with translocations and gene mis-expression in multiple myelomas (Chesi et al., 1998).

Precancerous and cancer cells have altered replication timing and progression compared to normal cells (Hills and Diffley, 2014). Cancer cells are also well-known for their propensity to undergo rearrangements and deletions. Since delayed replication is associated with CFSs, it is unsurprising that deletions at CFSs are found in cancer cell lines. Deletions of CFS sequences could give tumor cells a replicative advantage.

Ultra fine bridges (UFBs) form when sister chromatids remain connected due to incomplete replication during nuclear division, which can lead to breakage, chromosome rearrangements, or nondisjunction (Glover et al., 2017). CFS fragility can result in chromosomal deletions, translocations, and loss of heterozygosity that can be a stepping stone in the progression of cancer. This claim is also supported by the fact that CFSs are often located near cancer breakpoints. Breakage at CFSs can lead to gene misexpression, which is relevant

since many CFSs are located within tumor supressors. Finally, CFS fragility could also result in the ultimate cellular consequence: death.

Proteins and genetic factors associated with CFS expression

There are several well-studied genetic factors that are involved in CFS expression. The ataxia telangiectasia and Rad3-related (ATR) kinase is required to prevent CFS expression during normal, unperturbed replication. In lymphocytes and fibroblasts, moderate replication stress triggers chromatin loading of sensors/mediators of ATR pathway but doesn't activate Chk1 or p53 to avoid fork disassembly. ATR depleted cells have increased ssDNA formation upon Mre11-dependent resection of collapsed forks (Koundrioukoff et al., 2013). Downstream ATR effectors CHK1, HUS1, Claspin (*S. cerevisiae* Mrc1), and SMC1 also prevent CFS fragility (Durkin and Glover, 2007, Glover et al., 2017). These data imply that some sort of fork stalling is occurring at CFSs even without the addition of drugs such as aphidicolin.

Replication fork damage sensing proteins are also important in prevent CFS expression. RAD51, RAD52, BRCA1 are all associated with homologous recombination repair and CFS stability. Fork sensing proteins FANCD2 and FANCJ are also associated with CFS fragility (Durkin and Glover, 2007); in fact, FANCD2 twin foci on connected sister chromatids are a hallmark of detecting CFS fragility (Minocherhomji et al., 2015). It is very possible that secondary structures within CFSs stall forks. Since proteins known to respond to stalled

forks are associated with CFS expression, this is evidence that stalled and not just slowed forks are present at CFSs.

Bloom's Syndrome and Werner's Syndrome are two genetic diseases that result in predisposition to cancer, among many other phenotypes. Bloom's syndrome is named for mutations in the associated Bloom syndrome helicase BLM and Werner's is named for mutations in the WRN helicase. The RecQ1 helicase is also important for CFS expression. Since helicases unwind all DNA, and some helicases have specialized roles in unwinding secondary structures to facilitate replication, the importance of helicases at CFSs can be easily understood.

The CFS FRA16D lies within tumor suppressor gene WWOX

My thesis focuses on studying the breakage and healing of Flex1, a subregion of CFS FRA16D. CFS FRA16D is located between exons 8 and 9 of the WW-domain containing oxidoreductase gene (WWOX, also known as FOR); this is an evolutionary conserved, pleiotropic gene expressed in all human tissues (Gilad et al., 2006, Nunez et al., 2006). The WWOX gene is highly conserved in many organisms, including mice and *Drosophila melanogaster* (O'Keefe et al., 2011). FRA16D was originally mapped by FISH to be roughly 200 kb long (Mangelsdorf et al., 2000). By arranging subclone tiles in a directed (rather than random) manner, the FRA16D region was further extended to roughly 270 kb (Ried et al., 2000); however, even in this study there is only 4-fold sequence coverage, which is low coverage compared to the rest of the human genome.

The YES proto-oncogene and p53 binding protein-2 are both known to bind WW-domain containing proteins (Pirozzi et al., 1997). WWOX depletion has been associated with central nervous system pathology (Abu-Remaileh et al., 2015).

WWOX affects the DNA damage response, cellular proliferation, cellular metabolism, extra-cellular matrix composition, angiogenesis, and apoptosis (Aqeilan et al., 2014, Aqeilan et al., 2004a, Aqeilan et al., 2004b, Gourley et al., 2009, Dayan et al., 2013, Abu-Odeh et al., 2014, Chang et al., 2001, Bouteille et al., 2009, Wen et al., 2017).

There is a lot of data to support that WWOX functions as a tumor suppressor that may play a role in both tumor initiation and progression (Abu-Remaileh et al., 2015). WWOX depletion has been found in 9 of 9 pancreatic cancer cell lines (Kuroki et al., 2004). Heterozygous WWOX depletion in mice results in an increased propensity for tumor formation and progression (Abu-Remaileh et al., 2015, Abdeen et al., 2011). Reviewed in (Aldaz et al., 2014). Homogyzous WWOX depletion in humans can result in ataxia, epilepsy, and mental retardation (Mallaret et al., 2014).

An analysis of 746 cancer cell lines determined that WWOX (FRA16D) was the third most common site of hemi and homozygous deletions in the entire human genome (Bignell et al., 2010). Rapid Amplification of cDNA Ends (RACE) and Northern blots indicated that WWOX has multiple alternatively spliced transcripts in different tissues. RT-PCR and RACE indicated that these various WWOX

transcripts were differentially expressed in normal versus tumor cells. (Ried et al., 2000).

In 2005, the Richards lab did a careful evaluation of CFS FRA16D. They found that the Flex1 subregion of FRA16D has a highly polymorphic AT dinucleotide repeat ranging from 11-88 perfect ATs. Many cancer cell lines show deletions and breakpoints near the WWOX/FOR gene (Mangelsdorf et al., 2000, Ried et al., 2000, Finnis et al., 2005, Bignell et al., 2010). FRA16D deletions were present in a primary carcinoma and a secondary metastasis, showing instability at FRA16D early in cancer. Further, AT repeats and AT-rich sequences were near the deletion endpoints. Most cell lines with FRA16D homozygous deletions also have FRA3B deletions, therefore the conditions leading to their fragility and perhaps inhibition of healing may be similar for both of the most highly expressed CFSs. In one study, the Richards lab found no correlation between AT repeat copy number at Flex1 and fragile site expression, however metaphase chromosome spreads are notoriously inconsistent and therefore this may not have been the most accurate measure of fragility (Finnis et al., 2005). One gap my thesis aims to fill is the relationship between AT repeat copy number and Flex1 fragility.

CFSs are highly relevant regions for studying DNA replication and fragility in order to prevent the progression of cancer. My Thesis extensively studies the Flex1 subregion of CFS FRA16D. I did find a correlation between AT length, fork stalling, and fragility at Flex1. Thus, I aimed to investigate the role of several CFS theories in *S. cerevisiae* by using Flex1 as a model for CFS breakage and healing.

I present evidence linking secondary structures, fork stalling, and SSE cleavage to fragility at Flex1. I also evaluated the requirement of many trans factors for Flex1 fragility and healing to evaluate the relevance of CFS expression theories at Flex1. One goal of my unpublished results chapter (Kaushal Thesis Chapter 3) is to investigate whether transcription and R-loops play a role in Flex1 fragility.

References

ABDEEN, S. K., SALAH, Z., MALY, B., SMITH, Y., TUFAIL, R., ABU-ODEH, M., ZANESI, N., CROCE, C. M., NAWAZ, Z. & AQEILAN, R. I. 2011. Wwox inactivation enhances mammary tumorigenesis. *Oncogene,* 30**,** 3900-6.

ABU-ODEH, M., SALAH, Z., HERBEL, C., HOFMANN, T. G. & AQEILAN, R. I. 2014. WWOX, the common fragile site FRA16D gene product, regulates ATM activation and the DNA damage response. *Proc Natl Acad Sci U S A,* 111**,** E4716-25.

ABU-REMAILEH, M., JOY-DODSON, E., SCHUELER-FURMAN, O. & AQEILAN, R. I. 2015. Pleiotropic Functions of Tumor Suppressor WWOX in Normal and Cancer Cells. *J Biol Chem,* 290**,** 30728-35.

AGUILERA, A. & GARCIA-MUSE, T. 2012. R loops: from transcription byproducts to threats to genome stability. *Mol Cell,* 46**,** 115-24.

ALDAZ, C. M., FERGUSON, B. W. & ABBA, M. C. 2014. WWOX at the crossroads of cancer, metabolic syndrome related traits and CNS pathologies. *Biochim Biophys Acta,* 1846**,** 188-200.

AQEILAN, R. I., ABU-REMAILEH, M. & ABU-ODEH, M. 2014. The common fragile site FRA16D gene product WWOX: roles in tumor suppression and genomic stability. *Cell Mol Life Sci,* 71**,** 4589-99.

AQEILAN, R. I., KUROKI, T., PEKARSKY, Y., ALBAGHA, O., TRAPASSO, F., BAFFA, R., HUEBNER, K., EDMONDS, P. & CROCE, C. M. 2004a. Loss of WWOX expression in gastric carcinoma. *Clin Cancer Res,* 10**,** 3053-8.

AQEILAN, R. I., PEKARSKY, Y., HERRERO, J. J., PALAMARCHUK, A., LETOFSKY, J., DRUCK, T., TRAPASSO, F., HAN, S. Y., MELINO, G., HUEBNER, K. & CROCE, C. M. 2004b. Functional association between Wwox tumor suppressor protein and p73, a p53 homolog. *Proc Natl Acad Sci U S A,* 101**,** 4401-6.

ARLT, M. F., CASPER, A. M. & GLOVER, T. W. 2003. Common fragile sites. *Cytogenet Genome Res,* 100**,** 92-100.

BALAKUMARAN, B. S., FREUDENREICH, C. H. & ZAKIAN, V. A. 2000. CGG/CCG repeats exhibit orientation-dependent instability and orientation-independent fragility in Saccharomyces cerevisiae. *Hum Mol Genet,* 9**,** 93-100.

BARNES, R. P., HILE, S. E., LEE, M. Y. & ECKERT, K. A. 2017. DNA polymerases eta and kappa exchange with the polymerase delta holoenzyme to complete common fragile site synthesis. *DNA Repair (Amst),* 57**,** 1-11.

BECKER, N. A., THORLAND, E. C., DENISON, S. R., PHILLIPS, L. A. & SMITH, D. I. 2002. Evidence that instability within the FRA3B region extends four megabases. *Oncogene,* 21**,** 8713-22.

BEDNAREK, A. K., LAFLIN, K. J., DANIEL, R. L., LIAO, Q., HAWKINS, K. A. & ALDAZ, C. M. 2000. WWOX, a novel WW domain-containing protein mapping to human chromosome 16q23.3-24.1, a region frequently affected in breast cancer. *Cancer Res,* 60**,** 2140-5.

BEROUKHIM, R., MERMEL, C. H., PORTER, D., WEI, G., RAYCHAUDHURI, S., DONOVAN, J., BARRETINA, J., BOEHM, J. S., DOBSON, J., URASHIMA, M., MC HENRY, K. T., PINCHBACK, R. M., LIGON, A. H., CHO, Y. J., HAERY, L., GREULICH, H., REICH, M., WINCKLER, W., LAWRENCE, M. S., WEIR, B. A., TANAKA, K. E., CHIANG, D. Y., BASS, A. J., LOO, A., HOFFMAN, C., PRENSNER, J., LIEFELD, T., GAO, Q., YECIES, D., SIGNORETTI, S., MAHER, E., KAYE, F. J., SASAKI, H., TEPPER, J. E., FLETCHER, J.

A., TABERNERO, J., BASELGA, J., TSAO, M. S., DEMICHELIS, F., RUBIN, M. A., JANNE, P. A., DALY, M. J., NUCERA, C., LEVINE, R. L., EBERT, B. L., GABRIEL, S., RUSTGI, A. K., ANTONESCU, C. R., LADANYI, M., LETAI, A., GARRAWAY, L. A., LODA, M., BEER, D. G., TRUE, L. D., OKAMOTO, A., POMEROY, S. L., SINGER, S., GOLUB, T. R., LANDER, E. S., GETZ, G., SELLERS, W. R. & MEYERSON, M. 2010. The landscape of somatic copy-number alteration across human cancers. *Nature,* 463**,** 899-905.

BHAT, A., ANDERSEN, P. L., QIN, Z. & XIAO, W. 2013. Rev3, the catalytic subunit of Polzeta, is required for maintaining fragile site stability in human cells. *Nucleic Acids Res,* 41**,** 2328-39.

BIGNELL, G. R., GREENMAN, C. D., DAVIES, H., BUTLER, A. P., EDKINS, S., ANDREWS, J. M., BUCK, G., CHEN, L., BEARE, D., LATIMER, C., WIDAA, S., HINTON, J., FAHEY, C., FU, B., SWAMY, S., DALGLIESH, G. L., TEH, B. T., DELOUKAS, P., YANG, F., CAMPBELL, P. J., FUTREAL, P. A. & STRATTON, M. R. 2010. Signatures of mutation and selection in the cancer genome. *Nature,* 463**,** 893-8.

BOUTEILLE, N., DRIOUCH, K., HAGE, P. E., SIN, S., FORMSTECHER, E., CAMONIS, J., LIDEREAU, R. & LALLEMAND, F. 2009. Inhibition of the Wnt/beta-catenin pathway by the WWOX tumor suppressor protein. *Oncogene,* 28**,** 2569-80.

CASPER, A. M., NGHIEM, P., ARLT, M. F. & GLOVER, T. W. 2002. ATR regulates fragile site stability. *Cell,* 111**,** 779-89.

CHA, R. S. & KLECKNER, N. 2002. ATR homolog Mec1 promotes fork progression, thus averting breaks in replication slow zones. *Science,* 297**,** 602-6.

CHANG, N. S., PRATT, N., HEATH, J., SCHULTZ, L., SLEVE, D., CAREY, G. B. & ZEVOTEK, N. 2001. Hyaluronidase induction of a WW domain-containing oxidoreductase that enhances tumor necrosis factor cytotoxicity. *J Biol Chem,* 276**,** 3361-70.

CHERNG, N., SHISHKIN, A. A., SCHLAGER, L. I., TUCK, R. H., SLOAN, L., MATERA, R., SARKAR, P. S., ASHIZAWA, T., FREUDENREICH, C. H. & MIRKIN, S. M. 2011. Expansions, contractions, and fragility of the spinocerebellar ataxia type 10 pentanucleotide repeat in yeast. *Proc Natl Acad Sci U S A,* 108**,** 2843-8.

CHESI, M., BERGSAGEL, P. L., SHONUKAN, O. O., MARTELLI, M. L., BRENTS, L. A., CHEN, T., SCHROCK, E., RIED, T. & KUEHL, W. M. 1998. Frequent dysregulation of the c-maf proto-oncogene at 16q23 by translocation to an Ig locus in multiple myeloma. *Blood,* 91**,** 4457-63.

CIMPRICH, K. A. 2003. Fragile sites: breaking up over a slowdown. *Curr Biol,* 13**,** R231-3.

CLEARY, J. D., NICHOL, K., WANG, Y. H. & PEARSON, C. E. 2002. Evidence of cis-acting factors in replication-mediated trinucleotide repeat instability in primate cells. *Nat Genet,* 31**,** 37-46.

COQUELLE, A., PIPIRAS, E., TOLEDO, F., BUTTIN, G. & DEBATISSE, M. 1997. Expression of fragile sites triggers intrachromosomal mammalian gene amplification and sets boundaries to early amplicons. *Cell,* 89**,** 215-25.

COURBET, S., GAY, S., ARNOULT, N., WRONKA, G., ANGLANA, M., BRISON, O. & DEBATISSE, M. 2008. Replication fork movement sets chromatin loop size and origin choice in mammalian cells. *Nature,* 455**,** 557-60.

DAYAN, S., O'KEEFE, L. V., CHOO, A. & RICHARDS, R. I. 2013. Common chromosomal fragile site FRA16D tumor suppressor WWOX gene expression and metabolic reprograming in cells. *Genes Chromosomes Cancer,* 52**,** 823-31.

DAZY, S., GANDRILLON, O., HYRIEN, O. & PRIOLEAU, M. N. 2006. Broadening of DNA replication origin usage during metazoan cell differentiation. *EMBO Rep,* 7**,** 806-11.

DILLON, L. W., BURROW, A. A. & WANG, Y. H. 2010. DNA instability at chromosomal fragile sites in cancer. *Curr Genomics,* 11**,** 326-37.

DURKIN, S. G. & GLOVER, T. W. 2007. Chromosome fragile sites. *Annu Rev Genet,* 41**,** 169-92.

FECHTER, A., BUETTEL, I., KUEHNEL, E., SAVELYEVA, L. & SCHWAB, M. 2007. Common fragile site FRA11G and rare fragile site FRA11B at 11q23.3 encompass distinct genomic regions. *Genes Chromosomes Cancer,* 46**,** 98-106.

FINNIS, M., DAYAN, S., HOBSON, L., CHENEVIX-TRENCH, G., FRIEND, K., RIED, K., VENTER, D., WOOLLATT, E., BAKER, E. & RICHARDS, R. I. 2005. Common chromosomal fragile site FRA16D mutation in cancer cells. *Hum Mol Genet,* 14**,** 1341-9.

FREUDENREICH, C. H., KANTROW, S. M. & ZAKIAN, V. A. 1998. Expansion and length-dependent fragility of CTG repeats in yeast. *Science,* 279**,** 853-6.

FUNGTAMMASAN, A., WALSH, E., CHIAROMONTE, F., ECKERT, K. A. & MAKOVA, K. D. 2012. A genome-wide analysis of common fragile sites: what features determine chromosomal instability in the human genome? *Genome Res,* 22**,** 993-1005.

GAO, G. & SMITH, D. I. 2014. Very large common fragile site genes and their potential role in cancer development. *Cell Mol Life Sci,* 71**,** 4601-15.

GE, X. Q., JACKSON, D. A. & BLOW, J. J. 2007. Dormant origins licensed by excess Mcm2-7 are required for human cells to survive replicative stress. *Genes Dev,* 21**,** 3331-41.

GILAD, Y., OSHLACK, A., SMYTH, G. K., SPEED, T. P. & WHITE, K. P. 2006. Expression profiling in primates reveals a rapid evolution of human transcription factors. *Nature,* 440**,** 242-5.

GILBERT, D. M. 2007. Replication origin plasticity, Taylor-made: inhibition vs recruitment of origins under conditions of replication stress. *Chromosoma,* 116**,** 341-7.

GLOVER, T. W. 1998. Instability at chromosomal fragile sites. *Recent Results Cancer Res,* 154**,** 185-99.

GLOVER, T. W., BERGER, C., COYLE, J. & ECHO, B. 1984. DNA polymerase alpha inhibition by aphidicolin induces gaps and breaks at common fragile sites in human chromosomes. *Hum Genet,* 67**,** 136-42.

GLOVER, T. W. & STEIN, C. K. 1987. Induction of sister chromatid exchanges at common fragile sites. *Am J Hum Genet,* 41**,** 882-90.

GLOVER, T. W. & STEIN, C. K. 1988. Chromosome breakage and recombination at fragile sites. *Am J Hum Genet,* 43**,** 265-73.

GLOVER, T. W., WILSON, T. E. & ARLT, M. F. 2017. Fragile sites in cancer: more than meets the eye. *Nat Rev Cancer,* 17**,** 489-501.

GOURLEY, C., PAIGE, A. J., TAYLOR, K. J., WARD, C., KUSKE, B., ZHANG, J., SUN, M., JANCZAR, S., HARRISON, D. J., MUIR, M., SMYTH, J. F. & GABRA, H. 2009. WWOX gene expression abolishes ovarian cancer tumorigenicity in vivo and decreases attachment to fibronectin via integrin alpha3. *Cancer Res,* 69**,** 4835-42.

GREGOIRE, D., BRODOLIN, K. & MECHALI, M. 2006. HoxB domain induction silences DNA replication origins in the locus and specifies a single origin at its boundary. *EMBO Rep,* 7**,** 812-6.

HANSEN, R. S., THOMAS, S., SANDSTROM, R., CANFIELD, T. K., THURMAN, R. E., WEAVER, M., DORSCHNER, M. O., GARTLER, S. M. & STAMATOYANNOPOULOS,

J. A. 2010. Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proc Natl Acad Sci U S A,* 107**,** 139-44.

HELMRICH, A., BALLARINO, M. & TORA, L. 2011. Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. *Mol Cell,* 44**,** 966-77.

HEWETT, D. R., HANDT, O., HOBSON, L., MANGELSDORF, M., EYRE, H. J., BAKER, E., SUTHERLAND, G. R., SCHUFFENHAUER, S., MAO, J. I. & RICHARDS, R. I. 1998. FRA10B structure reveals common elements in repeat expansion and chromosomal fragile site genesis. *Mol Cell,* 1**,** 773-81.

HILLS, S. A. & DIFFLEY, J. F. 2014. DNA replication and oncogene-induced replicative stress. *Curr Biol,* 24**,** R435-44.

HIRATANI, I., RYBA, T., ITOH, M., YOKOCHI, T., SCHWAIGER, M., CHANG, C. W., LYOU, Y., TOWNES, T. M., SCHUBELER, D. & GILBERT, D. M. 2008. Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS Biol,* 6**,** e245.

HOSSEINI, S. A., HORTON, S., SALDIVAR, J. C., MIUMA, S., STAMPFER, M. R., HEEREMA, N. A. & HUEBNER, K. 2013. Common chromosome fragile sites in human and murine epithelial cells and FHIT/FRA3B loss-induced global genome instability. *Genes Chromosomes Cancer,* 52**,** 1017-29.

HUEBNER, K., GARRISON, P. N., BARNES, L. D. & CROCE, C. M. 1998. The role of the FHIT/FRA3B locus in cancer. *Annu Rev Genet,* 32**,** 7-31.

IBARRA, A., SCHWOB, E. & MENDEZ, J. 2008. Excess MCM proteins protect human cells from replicative stress by licensing backup origins of replication. *Proc Natl Acad Sci U S A,* 105**,** 8956-61.

KIM, H. M., NARAYANAN, V., MIECZKOWSKI, P. A., PETES, T. D., KRASILNIKOVA, M. M., MIRKIN, S. M. & LOBACHEV, K. S. 2008. Chromosome fragility at GAA tracts in yeast depends on repeat orientation and requires mismatch repair. *EMBO J,* 27**,** 2896-906.

KOUNDRIOUKOFF, S., CARIGNON, S., TECHER, H., LETESSIER, A., BRISON, O. & DEBATISSE, M. 2013. Stepwise activation of the ATR signaling pathway upon increasing replication stress impacts fragile site integrity. *PLoS Genet,* 9**,** e1003643.

KRASILNIKOVA, M. M. & MIRKIN, S. M. 2004. Replication stalling at Friedreich's ataxia (GAA)n repeats in vivo. *Mol Cell Biol,* 24**,** 2286-95.

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 Chromosomes Cancer,* 34**,** 154-67.

KUMARI, D., HAYWARD, B., NAKAMURA, A. J., BONNER, W. M. & USDIN, K. 2015. Evidence for chromosome fragility at the frataxin locus in Friedreich ataxia. *Mutat Res,* 781**,** 14-21.

KURAHASHI, H., INAGAKI, H., KATO, T., HOSOBA, E., KOGO, H., OHYE, T., TSUTSUMI, M., BOLOR, H., TONG, M. & EMANUEL, B. S. 2009. Impaired DNA replication prompts deletions within palindromic sequences, but does not induce translocations in human cells. *Hum Mol Genet,* 18**,** 3397-406.

KUROKI, T., YENDAMURI, S., TRAPASSO, F., MATSUYAMA, A., AQEILAN, R. I., ALDER, H., RATTAN, S., CESARI, R., NOLLI, M. L., WILLIAMS, N. N., MORI, M., KANEMATSU, T. & CROCE, C. M. 2004. The tumor suppressor gene WWOX at FRA16D is involved in pancreatic carcinogenesis. *Clin Cancer Res,* 10**,** 2459-65.

LE BEAU, M. M., RASSOOL, F. V., NEILLY, M. E., ESPINOSA, R., 3RD, GLOVER, T. W., SMITH, D. I. & MCKEITHAN, T. W. 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**,** 755-61.

LE TALLEC, B., DUTRILLAUX, B., LACHAGES, A. M., MILLOT, G. A., BRISON, O. & DEBATISSE, M. 2011. Molecular profiling of common fragile sites in human fibroblasts. *Nat Struct Mol Biol,* 18**,** 1421-3.

LE TALLEC, B., MILLOT, G. A., BLIN, M. E., BRISON, O., DUTRILLAUX, B. & DEBATISSE, M. 2013. Common fragile site profiling in epithelial and erythroid cells reveals that most recurrent cancer deletions lie in fragile sites hosting large genes. *Cell Rep,* 4**,** 420-8.

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**,** 587-98.

LETESSIER, A., MILLOT, G. A., KOUNDRIOUKOFF, S., LACHAGES, A. M., VOGT, N., HANSEN, R. S., MALFOY, B., BRISON, O. & DEBATISSE, M. 2011. Cell-type-specific replication initiation programs set fragility of the FRA3B fragile site. *Nature,* 470**,** 120-3.

LI, N., WU, Y. & ZHOU, X. T. 1986a. Human chromosome hot points. V. The effect of four nucleosides on chromosomes in folate-free medium. *Hum Genet,* 74**,** 101-3.

LI, N. & ZHOU, X. T. 1985. Human chromosome hot points. IV. Uridine-induced hot-point breaks at 3p14 and 16q23-24 and increased expression of fragile site Xq27 in folate-free medium. *Hum Genet,* 71**,** 363-5.

LI, X. Z., YAN, Z. A. & ZHOU, X. T. 1986b. The effect of 1-beta-D-arabinofuranosyl-cytosine on the expression of the common fragile site at 3p14. *Hum Genet,* 74**,** 444-6.

LIU, G., BISSLER, J. J., SINDEN, R. R. & LEFFAK, M. 2007. Unstable spinocerebellar ataxia type 10 (ATTCT*(AGAAT) repeats are associated with aberrant replication at the ATX10 locus and replication origin-dependent expansion at an ectopic site in human cells. *Mol Cell Biol,* 27**,** 7828-38.

LOBACHEV, K. S., SHOR, B. M., TRAN, H. T., TAYLOR, W., KEEN, J. D., RESNICK, M. A. & GORDENIN, D. A. 1998. Factors affecting inverted repeat stimulation of recombination and deletion in Saccharomyces cerevisiae. *Genetics,* 148**,** 1507-24.

LU, S., WANG, G., BACOLLA, A., ZHAO, J., SPITSER, S. & VASQUEZ, K. M. 2015. Short Inverted Repeats Are Hotspots for Genetic Instability: Relevance to Cancer Genomes. *Cell Rep*.

MADIREDDY, A., KOSIYATRAKUL, S. T., BOISVERT, R. A., HERRERA-MOYANO, E., GARCIA-RUBIO, M. L., GERHARDT, J., VUONO, E. A., OWEN, N., YAN, Z., OLSON, S., AGUILERA, A., HOWLETT, N. G. & SCHILDKRAUT, C. L. 2016. FANCD2 Facilitates Replication through Common Fragile Sites. *Mol Cell,* 64**,** 388-404.

MAGENIS, R. E., HECHT, F. & LOVRIEN, E. W. 1970. Heritable fragile site on chromosome 16: probable localization of haptoglobin locus in man. *Science,* 170**,** 85-7.

MALLARET, M., SYNOFZIK, M., LEE, J., SAGUM, C. A., MAHAJNAH, M., SHARKIA, R., DROUOT, N., RENAUD, M., KLEIN, F. A., ANHEIM, M., TRANCHANT, C., MIGNOT, C., MANDEL, J. L., BEDFORD, M., BAUER, P., SALIH, M. A., SCHULE, R., SCHOLS, L., ALDAZ, C. M. & KOENIG, M. 2014. The tumour suppressor gene WWOX is

mutated in autosomal recessive cerebellar ataxia with epilepsy and mental retardation. *Brain,* 137**,** 411-9.

MANGELSDORF, M., RIED, K., WOOLLATT, E., DAYAN, S., EYRE, H., FINNIS, M., HOBSON, L., NANCARROW, J., VENTER, D., BAKER, E. & RICHARDS, R. I. 2000. Chromosomal fragile site FRA16D and DNA instability in cancer. *Cancer Res,* 60**,** 1683-9.

MANSILLA, S. F., BERTOLIN, A. P., BERGOGLIO, V., PILLAIRE, M. J., GONZALEZ BESTEIRO, M. A., LUZZANI, C., MIRIUKA, S. G., CAZAUX, C., HOFFMANN, J. S. & GOTTIFREDI, V. 2016. Cyclin Kinase-independent role of p21(CDKN1A) in the promotion of nascent DNA elongation in unstressed cells. *Elife,* 5.

MCAVOY, S., GANAPATHIRAJU, S. C., DUCHARME-SMITH, A. L., PRITCHETT, J. R., KOSARI, F., PEREZ, D. S., ZHU, Y., JAMES, C. D. & SMITH, D. I. 2007. Non-random inactivation of large common fragile site genes in different cancers. *Cytogenet Genome Res,* 118**,** 260-9.

MINOCHERHOMJI, S., YING, S., BJERREGAARD, V. A., BURSOMANNO, S., ALELIUNAITE, A., WU, W., MANKOURI, H. W., SHEN, H., LIU, Y. & HICKSON, I. D. 2015. Replication stress activates DNA repair synthesis in mitosis. *Nature,* 528**,** 286-90.

MIOTTO, B., JI, Z. & STRUHL, K. 2016. Selectivity of ORC binding sites and the relation to replication timing, fragile sites, and deletions in cancers. *Proc Natl Acad Sci U S A,* 113**,** E4810-9.

MISHMAR, D., RAHAT, A., SCHERER, S. W., NYAKATURA, G., HINZMANN, B., KOHWI, Y., MANDEL-GUTFROIND, Y., LEE, J. R., DRESCHER, B., SAS, D. E., MARGALIT, H., PLATZER, M., WEISS, A., TSUI, L. C., ROSENTHAL, A. & KEREM, B. 1998. Molecular characterization of a common fragile site (FRA7H) on human chromosome 7 by the cloning of a simian virus 40 integration site. *Proc Natl Acad Sci U S A,* 95**,** 8141-6.

MRASEK, K., SCHODER, C., TEICHMANN, A. C., BEHR, K., FRANZE, B., WILHELM, K., BLAUROCK, N., CLAUSSEN, U., LIEHR, T. & WEISE, A. 2010. Global screening and extended nomenclature for 230 aphidicolin-inducible fragile sites, including 61 yet unreported ones. *Int J Oncol,* 36**,** 929-40.

NAIM, V., WILHELM, T., DEBATISSE, M. & ROSSELLI, F. 2013. ERCC1 and MUS81-EME1 promote sister chromatid separation by processing late replication intermediates at common fragile sites during mitosis. *Nat Cell Biol,* 15**,** 1008-15.

NEGRINI, S., GORGOULIS, V. G. & HALAZONETIS, T. D. 2010. Genomic instability--an evolving hallmark of cancer. *Nat Rev Mol Cell Biol,* 11**,** 220-8.

NGUYEN, J. H. G., VITERBO, D., ANAND, R. P., VERRA, L., SLOAN, L., RICHARD, G. F. & FREUDENREICH, C. H. 2017. Differential requirement of Srs2 helicase and Rad51 displacement activities in replication of hairpin-forming CAG/CTG repeats. *Nucleic Acids Res,* 45**,** 4519-4531.

NUNEZ, M. I., LUDES-MEYERS, J. & ALDAZ, C. M. 2006. WWOX protein expression in normal human tissues. *J Mol Histol,* 37**,** 115-25.

O'KEEFE, L. V., COLELLA, A., DAYAN, S., CHEN, Q., CHOO, A., JACOB, R., PRICE, G., VENTER, D. & RICHARDS, R. I. 2011. Drosophila orthologue of WWOX, the chromosomal fragile site FRA16D tumour suppressor gene, functions in aerobic metabolism and regulates reactive oxygen species. *Hum Mol Genet,* 20**,** 497-509.

OZERI-GALAI, E., LEBOFSKY, R., RAHAT, A., BESTER, A. C., BENSIMON, A. & KEREM, B. 2011. Failure of origin activation in response to fork stalling leads to chromosomal instability at fragile sites. *Mol Cell,* 43**,** 122-31.

PAIGE, A. J., TAYLOR, K. J., STEWART, A., SGOUROS, J. G., GABRA, H., SELLAR, G. C., SMYTH, J. F., PORTEOUS, D. J. & WATSON, J. E. 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 Res,* 60**,** 1690-7.

PALAKODETI, A., HAN, Y., JIANG, Y. & LE BEAU, M. M. 2004. The role of late/slow replication of the FRA16D in common fragile site induction. *Genes Chromosomes Cancer,* 39**,** 71-6.

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. *Hum Mol Genet,* 19**,** 99-110.

PALUMBO, E., MATRICARDI, L., TOSONI, E., BENSIMON, A. & RUSSO, A. 2010. Replication dynamics at common fragile site FRA6E. *Chromosoma,* 119**,** 575-87.

PIROZZI, G., MCCONNELL, S. J., UVEGES, A. J., CARTER, J. M., SPARKS, A. B., KAY, B. K. & FOWLKES, D. M. 1997. Identification of novel human WW domain-containing proteins by cloning of ligand targets. *J Biol Chem,* 272**,** 14611-6.

POLLEYS, E. J., HOUSE, N. C. M. & FREUDENREICH, C. H. 2017. Role of recombination and replication fork restart in repeat instability. *DNA Repair (Amst),* 56**,** 156-165.

POTAMAN, V. N., BISSLER, J. J., HASHEM, V. I., OUSSATCHEVA, E. A., LU, L., SHLYAKHTENKO, L. S., LYUBCHENKO, Y. L., MATSUURA, T., ASHIZAWA, T., LEFFAK, M., BENHAM, C. J. & SINDEN, R. R. 2003. Unpaired structures in SCA10 (ATTCT)n.(AGAAT)n repeats. *J Mol Biol,* 326**,** 1095-111.

REY, L., SIDOROVA, J. M., PUGET, N., BOUDSOCQ, F., BIARD, D. S., MONNAT, R. J., JR., CAZAUX, C. & HOFFMANN, J. S. 2009. Human DNA polymerase eta is required for common fragile site stability during unperturbed DNA replication. *Mol Cell Biol,* 29**,** 3344-54.

RIED, K., FINNIS, M., HOBSON, L., MANGELSDORF, M., DAYAN, S., NANCARROW, J. K., WOOLLATT, E., KREMMIDIOTIS, G., GARDNER, A., VENTER, D., BAKER, E. & RICHARDS, R. I. 2000. Common chromosomal fragile site FRA16D sequence: identification of the FOR gene spanning FRA16D and homozygous deletions and translocation breakpoints in cancer cells. *Hum Mol Genet,* 9**,** 1651-63.

ROZIER, L., EL-ACHKAR, E., APIOU, F. & DEBATISSE, M. 2004. Characterization of a conserved aphidicolin-sensitive common fragile site at human 4q22 and mouse 6C1: possible association with an inherited disease and cancer. *Oncogene,* 23**,** 6872-80.

RUIZ-HERRERA, A., GARCIA, F., FRONICKE, L., PONSA, M., EGOZCUE, J., CALDES, M. G. & STANYON, R. 2004. Conservation of aphidicolin-induced fragile sites in Papionini (Primates) species and humans. *Chromosome Res,* 12**,** 683-90.

RYBA, T., HIRATANI, I., LU, J., ITOH, M., KULIK, M., ZHANG, J., SCHULZ, T. C., ROBINS, A. J., DALTON, S. & GILBERT, D. M. 2010. Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. *Genome Res,* 20**,** 761-70.

SCHWARTZ, M., ZLOTORYNSKI, E. & KEREM, B. 2006. The molecular basis of common and rare fragile sites. *Cancer Lett,* 232**,** 13-26.

SHIRAISHI, T., DRUCK, T., MIMORI, K., FLOMENBERG, J., BERK, L., ALDER, H., MILLER, W., HUEBNER, K. & CROCE, C. M. 2001. Sequence conservation at human and mouse

orthologous common fragile regions, FRA3B/FHIT and Fra14A2/Fhit. *Proc Natl Acad Sci U S A,* 98**,** 5722-7.

SMITH, D. I., ZHU, Y., MCAVOY, S. & KUHN, R. 2006. Common fragile sites, extremely large genes, neural development and cancer. *Cancer Lett,* 232**,** 48-57.

STEINMETZER, K., ZANNIS-HADJOPOULOS, M. & PRICE, G. B. 1995. Anti-cruciform monoclonal antibody and cruciform DNA interaction. *J Mol Biol,* 254**,** 29-37.

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 Res,* 28**,** 4805-13.

USDIN, K., HOUSE, N. C. & FREUDENREICH, C. H. 2015. Repeat instability during DNA repair: Insights from model systems. *Crit Rev Biochem Mol Biol,* 50**,** 142-67.

VOINEAGU, I., SURKA, C. F., SHISHKIN, A. A., KRASILNIKOVA, M. M. & MIRKIN, S. M. 2009. Replisome stalling and stabilization at CGG repeats, which are responsible for chromosomal fragility. *Nat Struct Mol Biol,* 16**,** 226-8.

WANG, L., PARADEE, W., MULLINS, C., SHRIDHAR, R., ROSATI, R., WILKE, C. M., GLOVER, T. W. & SMITH, D. I. 1997. Aphidicolin-induced FRA3B breakpoints cluster in two distinct regions. *Genomics,* 41**,** 485-8.

WARD, G. K., MCKENZIE, R., ZANNIS-HADJOPOULOS, M. & PRICE, G. B. 1990. The dynamic distribution and quantification of DNA cruciforms in eukaryotic nuclei. *Exp Cell Res,* 188**,** 235-46.

WEN, J., XU, Z., LI, J., ZHANG, Y., FAN, W., WANG, Y., LU, M. & LI, J. 2017. Decreased WWOX expression promotes angiogenesis in osteosarcoma. *Oncotarget,* 8**,** 60917-60932.

WILKE, C. M., HALL, B. K., HOGE, A., PARADEE, W., SMITH, D. I. & GLOVER, T. W. 1996. FRA3B extends over a broad region and contains a spontaneous HPV16 integration site: direct evidence for the coincidence of viral integration sites and fragile sites. *Hum Mol Genet,* 5**,** 187-95.

YAN, Z. A., LI, X. Z. & ZHOU, X. T. 1987. The effect of hydroxyurea on the expression of the common fragile site at 3p14. *J Med Genet,* 24**,** 593-6.

YING, S., MINOCHERHOMJI, S., CHAN, K. L., PALMAI-PALLAG, T., CHU, W. K., WASS, T., MANKOURI, H. W., LIU, Y. & HICKSON, I. D. 2013. MUS81 promotes common fragile site expression. *Nat Cell Biol,* 15**,** 1001-7.

YU, S., MANGELSDORF, M., HEWETT, D., HOBSON, L., BAKER, E., EYRE, H. J., LAPSYS, N., LE PASLIER, D., DOGGETT, N. A., SUTHERLAND, G. R. & RICHARDS, R. I. 1997. Human chromosomal fragile site FRA16B is an amplified AT-rich minisatellite repeat. *Cell,* 88**,** 367-74.

ZANNIS-HADJOPOULOS, M., FRAPPIER, L., KHOURY, M. & PRICE, G. B. 1988. Effect of anti-cruciform DNA monoclonal antibodies on DNA replication. *EMBO J,* 7**,** 1837-44.

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. *Mol Cell,* 27**,** 367-79.

ZHOU, X. T., XU, B. H., CHU, C. L., XIA, G. F., LI, N. & SHA, R. 1984. Human chromosome hot points. 1. Hot point at 3p14 in three populations. *Hum Genet,* 67**,** 249-51.

ZLOTORYNSKI, E., RAHAT, A., SKAUG, J., BEN-PORAT, N., OZERI, E., HERSHBERG, R., LEVI, A., SCHERER, S. W., MARGALIT, H. & KEREM, B. 2003. Molecular basis for expression of common and rare fragile sites. *Mol Cell Biol,* 23**,** 7143-51.

ZUKER, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res,* 31**,** 3406-15.

Sequence and nuclease requirements for breakage and healing of a structure-forming (AT)n sequence within fragile site FRA16D

Simran Kaushal[1], Charles E. Wollmuth[1], Samantha B. Regan[1], Alice Haouzi[1,3], Ryan P. Barnes[4], Soo Mi Lee[1,5,], Nealia C. M. House[1,6], Kristin A. Eckert[4], Catherine H. Freudenreich[1,2,*]

[1] Department of Biology, Tufts University, Suite 4700, 200 Boston Ave, Medford, MA 02155, USA

[2] Program in Genetics, Sackler School of Graduate Biomedical Sciences, Tufts University, Boston, MA 02111, USA

[3] Present address: Emory University School of Medicine, 100 Woodruff Circle, Atlanta, GA 30322, USA

[4] Department of Pathology, The Jake Gittlen Laboratories for Cancer Research, Penn State University College of Medicine, Hershey, PA, USA

[5] Present address: Program in Virology, Graduate School of Arts and Sciences, Harvard University, Cambridge, Massachusetts.

[6] Present address: Department of Radiation Oncology, Division of Genomic Stability, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA 02215

*Corresponding author
Catherine Freudenreich Tel +1 617 627 4037; Fax +1 617 627 0309; E-mail catherine.freudenreich@tufts.edu

This manuscript has been sent out for review.

Abstract

Common fragile sites (CFSs) are genomic regions that display gaps and breaks in human metaphase chromosomes under replication stress and are often deleted in cancer cells. We studied a ~300 basepair subregion (Flex1) of human CFS FRA16D in yeast, and found it recapitulated characteristics of CFS fragility in human cells. Flex1 fragility was dependent on the ability of a variable-length AT repeat, predicted to form a cruciform structure, to stall replication. Fragility at Flex1 is initiated by structure-specific endonuclease Mus81, likely acting within an Slx1-4- Rad1-10 complex, while Yen1 protects Flex1 against breakage. Sae2 is required for healing of Flex1 after breakage. Our study shows that breakage within a CFS can be initiated by nuclease cleavage of forks stalled at DNA structures. Furthermore, our results suggest that CFSs are not just prone to breakage but also impaired in their ability to heal, and this deleterious combination accounts for their fragility.

Introduction

Common fragile sites (CFSs) are highly unstable human chromosomal regions that are prone to breakage and the formation of cancer associated deletions. The gaps can be visualized cytogenetically, and generally span hundreds of kilobases of DNA. Over 75 CFSs have been discovered to date. CFS breakage, or expression, can be induced in most individuals and therefore they are considered a normal part of chromosome structure. The molecular basis for their fragility is still not well understood, though an inability to complete replication during S

phase is an important component. CFS expression can be induced by drugs that inhibit polymerase progression, such as aphidicolin and hydroxyurea (Glover et al., 2017). The two most commonly expressed CFSs in human cells, FRA3B and FRA16D, replicate late in S and into G2 phase (Le Beau et al., 1998, Palakodeti et al., 2004, Letessier et al., 2011). CFSs can undergo Mitotic DNA synthesis (MiDAS) in order to finish replicating these regions before nuclear division occurs (Minocherhomji et al., 2015).

There is evidence that CFS expression varies by cell type (Le Tallec et al., 2013) and there is conflicting evidence that that gene expression levels may correlate with fragility levels (Helmrich et al., 2011, Le Tallec et al., 2013). FRA16D, located within a large intron of the WWOX tumor suppressor gene, is one of the most breakage-prone CFSs, as it was expressed in all four cell types tested (Le Tallec et al., 2013). This suggests that FRA16D is inherently fragile, even under varied levels of transcription. It has recently been shown that the protein FANCD2 facilitates replication through FRA16D by inducing dormant origin firing and suppressing DNA:RNA hybrid formation (Madireddy et al., 2016).

Although CFSs are a normal part of chromosome structure, they are vulnerable parts of the genome as they are frequently the locations of homogyzous and hemizygous deletions in many cancer cell lines (Finnis et al., 2005, Bignell et al., 2010). DNA breaks can initiate deletions or breakage-fusion-bridge cycles, both of which can cause mis-expression of the affected genes (Coquelle et al., 1997, Finnis et al., 2005). CFSs are also hotspots of *de novo* copy number variations (CNVs) in many tumor types, likely occurring due to replication stress followed

by aberrant repair (Durkin et al., 2008, Zack et al., 2013, Glover et al., 2017). Breakage at CFSs is an early event in tumor progression (Halazonetis et al., 2008). Additionally, oncogene overexpression leads to replication stress (oncogene-induced replication stress) that can then result in CFS breakage, deletions, and rearrangements (Macheret and Halazonetis, 2015, Miron et al., 2015, Taylor and Lindsay, 2016, Glover et al., 2017).

CFSs tend to be AT-rich, making their DNA easier to unwind to form unusual or non-B DNA secondary structures, which could play a role in their fragility. Computational analysis of CFSs by multiple groups using different approaches have identified a higher density of sequences with potential to form stable secondary structures compared to controls (Mishmar et al., 1998, Zlotorynski et al., 2003, Fungtammasan et al., 2012, Dillon et al., 2013) and secondary structures at both rare and common fragile sites have connections to human disease and cancer (Thys et al., 2015).

Flex1 is a ~300 bp AT-rich subregion of human common fragile site FRA16D. Flex1 contains a polymorphic perfect AT repeat that ranges from 11-88 copies in humans tested and is frequently deleted in FRA16D associated tumor cell lines (Finnis et al., 2005). Due to their weak base-pairing, AT repeats can nucleate an unwound region of double-stranded supercoiled DNA to form cruciforms *in vivo* once they exceed a length of around 22 repeat units (McClellan et al., 1990, Dayn et al., 1991, Bowater et al., 1991). Cruciform cleavage and resolution has been implicated in multiple common chromosomal translocations (Inagaki et al., 2013, Kato et al., 2014). AT repeats and short inverted repeats are prevalent in the

human genome and have been implicated in driving genomic rearrangements in evolution. Also, they are enriched near cancer translocation and deletion breakpoints (Babcock et al., 2007, Lu et al., 2015, Bacolla et al., 2016).

Our lab has previously shown that the Flex1 sequence caused chromosome fragility when inserted into an artificial chromosome in *S. cerevisiae*, and that a Flex1 sequence containing (AT)34 caused replication fork stalling (Zhang and Freudenreich, 2007). It was hypothesized that a secondary structure at Flex1 is causing replication fork stalling and contributing to FRA16D breakage.

In this study, we show that Flex1 is a significant contributor to overall FRA16D breakage, and that Flex1 fragility increases with AT repeat length in a nonlinear fashion. This supports the hypothesis that cruciform structures at longer AT lengths are the cause of fragility. Additionally, the Flex1 (AT)34 repeat pauses human polymerase δ *in vitro*. Structure-specific endonuclease (SSE) Mus81 is required for Flex1 fragility, in agreement with the known requirement of human MUS81 for FRA16D expression in human cells (Naim et al., 2013, Ying et al., 2013). Importantly, we find that Mus81 induces Flex1 fragility only at AT lengths long enough to form a cruciform, implying that Mus81 is specifically acting at secondary structures in FRA16D, cleaving either the cruciform or a resulting structure such as a stalled replication fork. Slx1-Slx4 and Rad1-Rad10 nucleases also play a role in causing Flex1 fragility, and our data provide evidence for the existence of an Slx1-Slx4-Mus81-Mms4-Rad1-Rad10 (SMR) super complex in *S. cerevisiae*. In contrast, Yen1, which acts in late mitosis, has a role in preventing (rather than causing) fragility at Flex1. Our data suggest that coordinated cleavage

by SSEs of forks stalled by DNA structures at CFSs may account for their characteristic expression of gaps and breaks in mitosis. Furthermore, Pol32-mediated synthesis is required to prevent Flex1 fragility, analogous to the requirement for POLD3-dependent MiDAS at CFSs in human cells (Minocherhomji et al., 2015). Finally, we identify that it is not only the AT repeat length but also the flanking Flex1 sequences that play a role in the expression of breaks at Flex1, as they inhibit efficient healing of the broken DNA. Therefore, we propose a new theory that the DNA sequences at CFSs have both an increased tendency to break and a reduced ability to heal following breakage, contributing to their persistence into M phase and their propensity to instigate large deletions and translocations.

Results

Flex1 is a crucial element causing FRA16D sequence breakage *in vivo*

To determine whether Flex1 was responsible for a substantial amount of FRA16D fragility, we deleted the Flex1 sequence from YAC 801B6, which contains 1.4 Mb of human chromosome 16 sequence including FRA16D (Figures 2-1A and 2-1C). Despite deleting only ~300 bp of the 1.4 Mb human sequence (0.02%), we observed a significant decrease in frequency of YAC end loss (measured by FOA^R) (Figure 2-1B). These results indicate that Flex1 accounts for a significant and measurable fraction of breakage events within 801B6, highlighting the importance of the Flex1 sequence in contributing to overall FRA16D fragility. Nonetheless, based on the level of fragility of the adjacent sequence (972D3)

(Figures 2-1B and 2-1C), we speculate that other fragile elements may combine with Flex1 to account for the full fragility of the entire region.

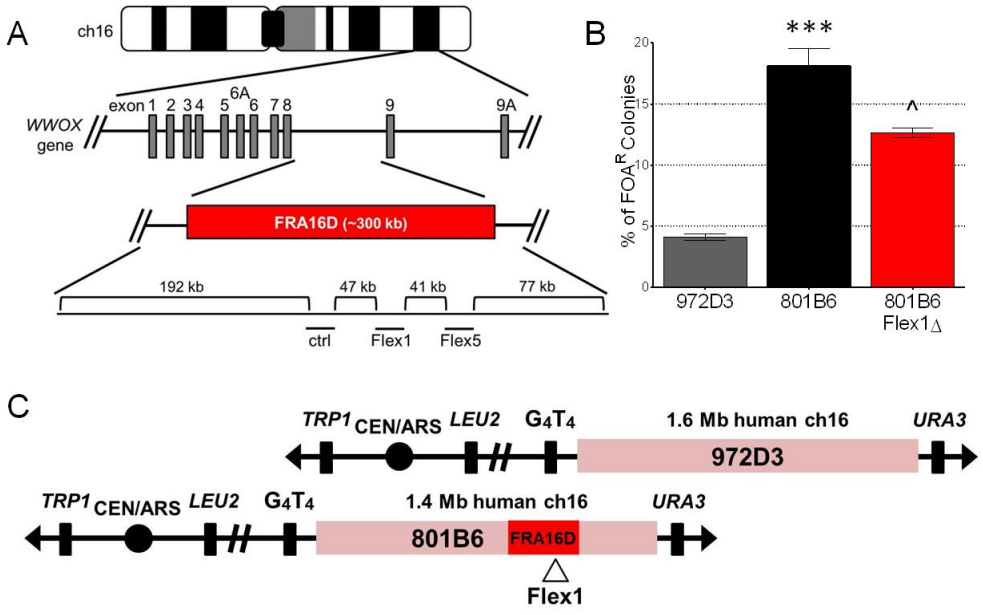


Figure 2-1: Flex1 is important for FRA16D fragility. (A) Schematic of the chromosomal locations of FRA16D and subregions Flex1, Flex5, and the control sequence, which has 372 bp of AT-rich FRA16D sequence not predicted to form a secondary structure. Two YACs from the CEPH YAC library containing 1.4 Mb of human chromosome 16 including FRA16D (801B6) or 1.6 Mb of human chromosome 16 adjacent to FRA16D (972D3) were altered to contain a $(G_4T_4)_{13}$ telomere seed (Zhang and Freudenreich, 2007). An alternate version of the FRA16D YAC was created where Flex1 was replaced by a selectable marker; YAC integrity was verified by pulsed-field gel electrophoresis and PCR of subregions (Figure 2-S1 and Table 2-S1). (B) Fragility of various FRA16D YACs was measured by percent of FOA resistance (% FOA^R), indicating loss of the

URA3 marker. *** = p< 0.001 compared to 972D3; ^ = p<0.05 compared to 801B6 with Flex1, by unpaired t-test; see also Table 2-S2. *URA3* loss was confirmed by PCR in a subset of FOA^R colonies. (C) Diagram of YACs containing human chromosome 16 sequences. Chromosome 16 boxes are lined up according to their genomic coordinates.

Flex1 is fragile in an (AT)n repeat length-dependent manner

Because the Flex1 sequence contains a polymorphic (AT)n repeat in humans, it was important to address the role that AT repeat length plays in Flex1 fragility. Our group previously demonstrated an increase in fork stalling with increasing AT length, and there was a trend of increasing fragility as Flex1 AT length increased from 5 to 14 to 23 (Zhang and Freudenreich, 2007). Thus, it was hypothesized that as AT repeat length increases, a secondary structure at the repetitive sequence becomes more stable, resulting in replication fork stalling and fragility. Due to the severity of the replication fork stalling at Flex1(AT)34, it was hypothesized that the AT repeats of this length formed a cruciform, although the sequence could also form hairpins on either strand. To test the hypothesis that the AT repeats are forming a secondary structure *in vivo* to cause fragility, two approaches were taken. First, flanking Flex1 sequences on either side of the AT repeats were standardized, and second, fragility was tested using two genetic assays that utilize different mechanisms of recovering the broken DNA. To evaluate the role of AT tract length in Flex1 fragility, we inserted three different sequences of varying AT lengths but standardized short 5' and short 3' (S5' and S3', respectively) Flex1 flanking sequences into a direct duplication recombination assay fragility system on yeast chromosome II (DDRA fragility assay) (Figures 2-2A and 2-S2A). Breakage at Flex1 can stimulate recombination between flanking homologous *ADE2* sequences via single strand annealing (SSA), which results in loss of the intervening *URA3* gene and 5-FOA resistant, Ade^{+} cells (Freudenreich et al., 1998, Paeschke et al., 2011, Polleys and

Freudenreich, 2018). Recombination rates were also measured for a control sequence, which is a roughly 380 bp sequence from FRA16D that is not predicted to form a stable secondary structure (Figure 2-1A) (Zhang and Freudenreich, 2007). This assay mimics the types of deletion events that have been shown to occur naturally in cancer cells and other cells under replication stress, with the benefit of the deletions being selectable. In these constructs, the recombination rate increased significantly with increasing AT length, consistent with a repeat length-dependent increase in fragility (Figure 2-2B). The significant increase in recombination rate of (AT)23 and (AT)34 coincides with the known propensity of AT repeats to form a cruciform structure with much higher frequency when their length exceeds roughly 22 repeats (McClellan et al., 1990, Dayn et al., 1991, Bowater et al., 1991). The dramatic increase in fragility upon adding only 11 additional (AT)n repeats together with the severity of the Flex1 (AT)34 replication fork stalling *in vivo* (Zhang and Freudenreich, 2007) strongly supports that a fork-blocking cruciform is frequently forming at this length.

To test another possible fragile element from FRA16D, we measured fragility of Flex5 using the DDRA fragility assay (Figures 2-1A and 2-S3). Flex5 is a subregion of FRA16D that has an interrupted (AT)24 repeat ((AT)24i) and a 28 bp polyA sequence. Flex5 has been shown to cause polymerase δ pausing *in vitro* and replication through Flex5 can be improved by using polymerases η and κ (Shah et al., 2010, Walsh et al., 2013, Barnes et al., 2017). We found that Flex5 sequences did not stimulate recombination; the rate is no different from the control, which is not predicted to form secondary structures (Figure 2-S3, -HU

conditions). The Flex5 (AT)24i is very similar in sequence to Flex1 (AT)23, yet shows less fragility, similar to Flex1 (AT)14's recombination rate - this could indicate that the interruptions in Flex5's AT repeats reduce their propensity to form stable secondary structures under *in vivo* conditions.

To determine the effect of replication stress on fragility of CFS FRA16D subregions, the rate of FOA^R of Flex1 and Flex5 was also measured after treatment with hydroxyurea (HU) (Figure 2-S3). HU-mediated replication fork stalling is expected to further exacerbate fragility, which was observed in all sequences tested. However, the effect of HU was much stronger for sequences predicted to form no DNA structure or a weak DNA structure, such as the control (7.2-fold over no HU), Flex5 (6.8-fold (o1) and 9.4-fold (o2) over no HU), and Flex1 (AT)14 (5.8-fold over no HU), compared to sequences predicted to form a stable hairpin or cruciform structure (~2-fold over no HU for Flex1 (AT)23 and Flex1 (AT)34). These data suggest that additional replication stress is not required for fragility of sequences that can form a stable-enough structure to stall replication in their normal cellular context, though it further increases the likelihood of stalling and chromosome breakage.

An (AT)34 repeat causes human polymerase delta pausing in vitro

Previously, Zhang and Freudenreich (2007) found that the Flex1 (AT)34 sequence causes replication fork stalling during plasmid replication in *S. cerevisiae*, and it was hypothesized that a secondary structure formed by the (AT)34 sequence would impede human polymerase δ transit through the region. Human 4-subunit polymerase δ holoenzyme DNA synthesis through Flex1 with (AT)34 was

measured using an *in vitro* primer extension assay as described in (Shah et al., 2010, Walsh et al., 2013, Barnes et al., 2017). Polymerase synthesis was measured on ssDNA templates containing the Flex1 (AT)34 sequence in the presence of RFC-loaded PCNA, and pausing was identified as sites of accumulated primer extension reaction products. Note that in this assay only template hairpins would be able to form, and not cruciforms that require double-stranded DNA. As highlighted, human polymerase δ struggles to replicate the AT repeats and significant stalling is detected throughout the AT tract (Figure 2-2C, red box). The same result was observed during polymerase δ synthesis using the opposite Flex1 strand as a template (Figure 2-S4). We conclude that the Flex1 (AT)34-dependent replication fork stalling previously observed *in vivo* is due to stalling specifically by the AT repeat, and can explain the AT-length dependent fragility results.

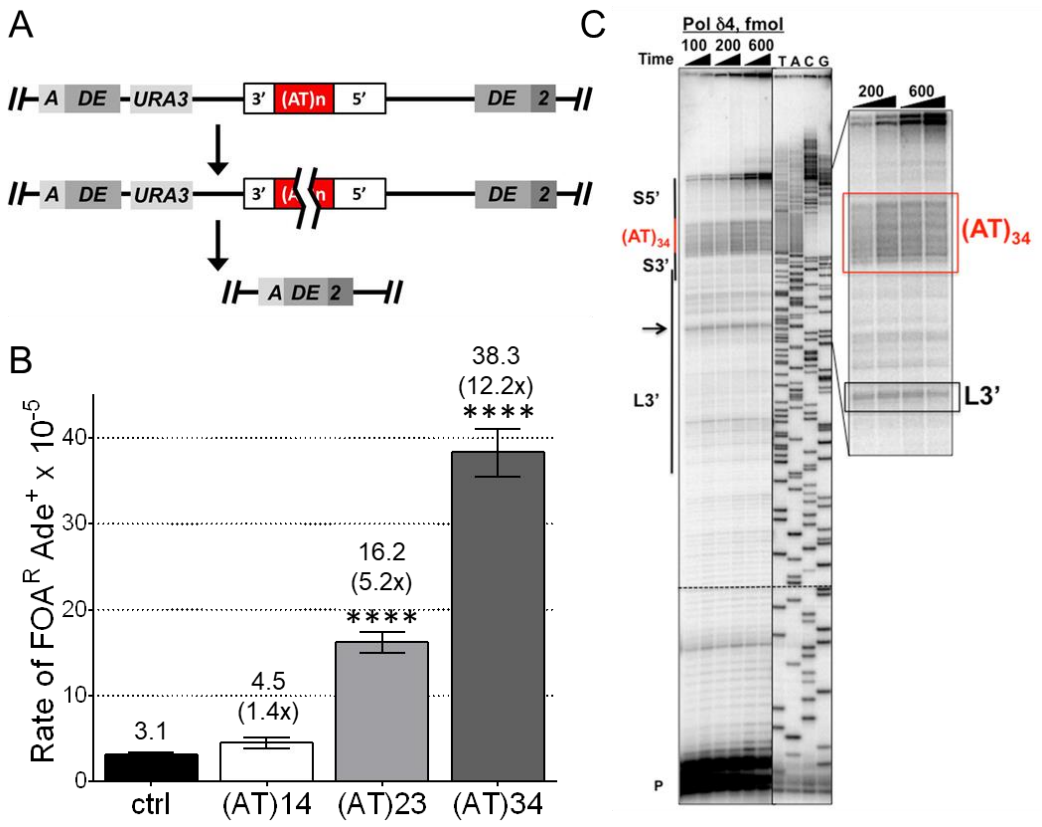


Figure 2-2: Flex1 is fragile in an AT-length dependent manner. (A) Schematic of the DDRA fragility assay. See Figure S2A for details. (B) Recombination rate increases as Flex1 AT repeat length increases; strains were tested for significant deviation from the control using an unpaired t-test; **** = $p<0.0001$ compared to the control (ctrl). Orientation 1 data is shown. Rates are reported above the appropriate bar with fold over the control in parentheses; see also Table 2-S3. (C) *In vitro* DNA synthesis of Flex1 with (AT)34 and a L3' flanking sequence by the 4-subunit human polymerase δ holoenzyme (Pol δ4), showing pause sites at the (AT)34 repeat (red box) and the base of the L3' hairpin (arrow and black box). Sequence outside of the marked area is composed of the plasmid backbone. TACG, dideoxy sequencing ladder of the DNA template.

Several structure-specific endonucleases cause fragility at Flex1

In human cells, the structure-specific endonucleases (SSEs) MUS81-EME1 and XPF-ERCC1 were shown to promote FRA16D expression in mitotic cells that had experienced replication stress, presumably by causing cleavage of a persistent replication intermediate (Ying et al., 2013, Naim et al., 2013). Human MUS81 and EME1 are also needed to promote PolD3-mediated mitotic DNA synthesis (MiDAS) at CFSs (Minocherhomji et al., 2015). However, these results were obtained using whole chromosomes, thus it was unclear where the cleavage was occurring. SSEs can act on substrates at stalled or reversed forks that form in S phase (and possibly persist into G2/M), and in G2 and M phases SSEs can act on homologous recombination (HR) intermediates or unreplicated DNA to allow chromosome separation (Rass, 2013, Symington et al., 2014, Wyatt and West, 2014, Dehe and Gaillard, 2017). Since our data indicate that longer AT repeats within the Flex1 sequence form cruciform structures that could resemble SSE substrates and also stall replication forks, it was of interest to determine if SSEs act at Flex1. SSEs could either cause the observed fragility by directed cleavage, or alternatively protect against fragility by allowing proper resolution of stalled forks or recombination intermediates.

Upon deletion of the *MUS81* gene, Flex1 (AT)34 fragility significantly decreases in both the DDRA fragility assay and the previously used YAC end loss assay (Figures 2-3A, 2-4, and 2-S2) (Zhang and Freudenreich, 2007). Healing in the YAC assay usually occurs by resection to the G_4T_4 telomere seed sequence and subsequent telomere addition, which results in loss of selectable markers (Polleys

and Freudenreich, 2018). Using the DDRA fragility assay, we evaluated the effect of *mus81*Δ on Flex1 containing various AT tract lengths and the control. The recombination rate is significantly decreased in *mus81*Δ only when Flex1's AT stem length exceeds 22 bp, the threshold for forming cruciforms; *mus81*Δ had no effect on the recombination rate of the control or Flex1 (AT)14 strains (Figure 2-3A). We conclude that Mus81 is specifically acting to cleave either directly at a DNA structure formed by the AT repeat or at a stalled fork caused by the structure. These data suggest that the requirement for MUS81 for FRA16D expression in human cells is due to structure-mediated events at Flex1, or perhaps at Flex1 in combination with other fork stalling regions.

MUS81 nuclease activity is required for the initiation of POLD3-mediated MiDAS in human cells (Minocherhomji et al., 2015), hence we sought to test the effect of deleting the yeast homolog of POLD3 (*POL32*, a subunit of Polδ) on Flex1 fragility. Lack of Pol32 resulted in a large increase in recombination rate (Figure 2-3B), demonstrating the importance of Pol32-mediated synthesis in preventing fragility and deletion of Flex1.

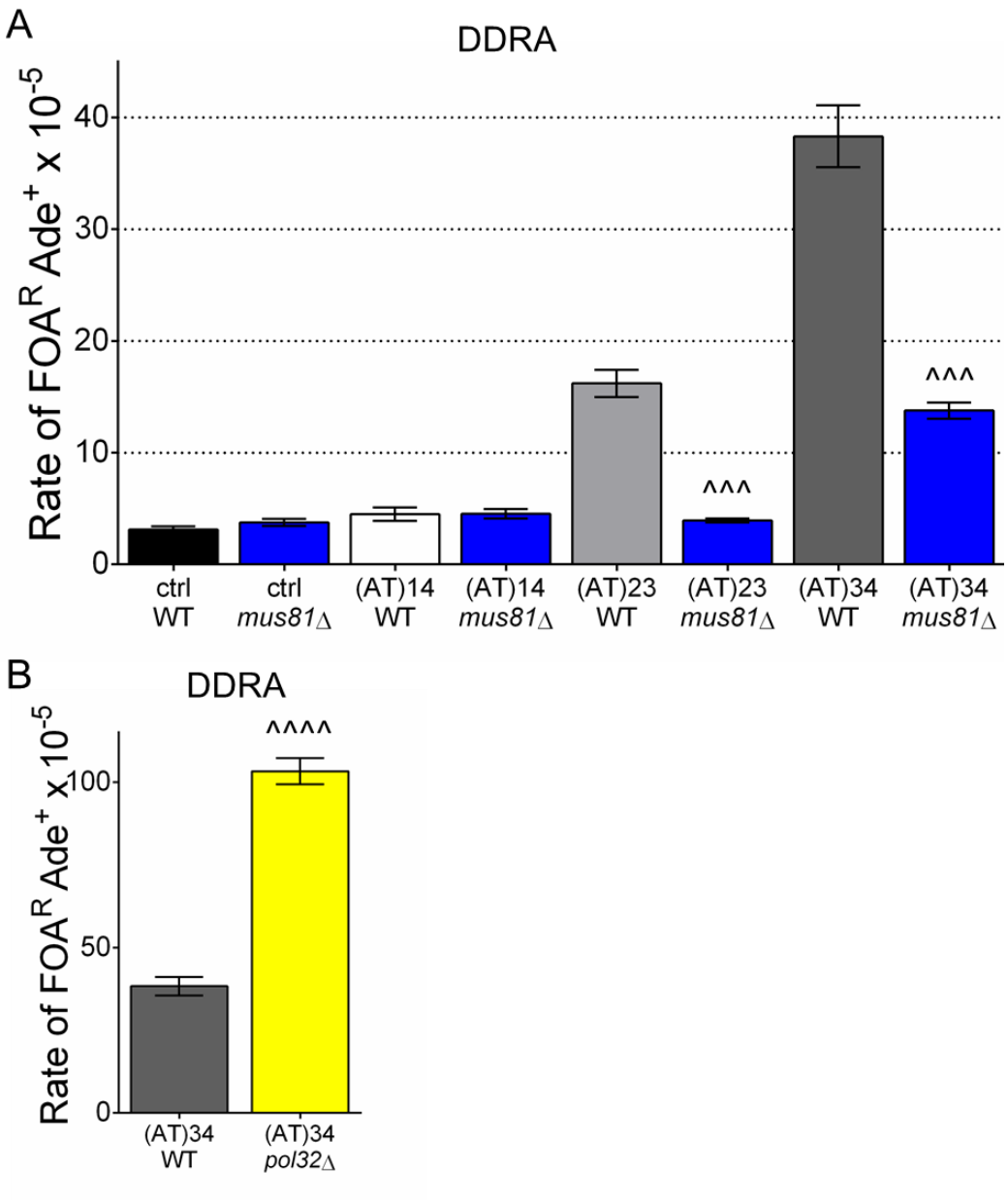


Figure 2-3: Mus81 causes fragility at Flex1 sequences that form secondary structures and stall replication. (A) The DDRA fragility assay was used to measure recombination rates of Flex1 orientation 1 with various AT lengths in *mus81*Δ strains and (B) *pol32*Δ strains. ^^^ p<0.001; ^^^^p<0.0001 compared to WT same Flex1 tract using an unpaired t-test. See also Table 2-S3.

In humans, SLX4 recruits MUS81-EME1 to CFSs in order to allow MiDAS to occur to avoid more deleterious anaphase bridges and 53BP1 body formation (Guervilly et al., 2015, Minocherhomji et al., 2015, Ouyang et al., 2015). SLX4 is a scaffolding protein that recruits multiple enzymes, including MUS81-EME1, SLX1, and XPF-ERCC1, to enhance their activity and coordinate SSE action timing and pathway choice (Sarbajna et al., 2014, Dehe and Gaillard, 2017). There is evidence for a super complex of SLX1-SLX4, MUS81-EME1, and XPF-ERCC1 in mammals, called the SMX DNA repair trinuclease (Wyatt et al., 2017). In *S. cerevisiae,* Slx4 interacts directly with Slx1 to form an endonuclease and also serves as a scaffold for the Mus81-Mms4 and Rad1-Rad10 complexes (human MUS81-EME1/2 and XPF-ERCC1, respectively) (Cussiol et al., 2017). We reasoned that secondary structures, stalled forks, or recombination structures induced by Flex1 may be substrates for other SSEs acting with Mus81 and tested this hypothesis by creating mutants in both genetic systems.

Strains lacking either Slx4 or Rad1 showed a significant decrease in fragility for both the control and Flex1 (AT)34 in the DDRA fragility assay (Figures 2-4A and 2-4B). The decrease in recombination in the control strain in the *slx4*Δ or *rad1*Δ backgrounds indicates that both Slx4 and Rad1 are required for SSA, as shown previously (Freudenreich et al., 1998, Mimitou and Symington, 2009, Dehe and Gaillard, 2017) . However, the deletion of both proteins had a much more dramatic fold decrease compared to wild-type for the Flex1 (AT)34 sequence compared to the control sequence (12-fold vs. 3.2-fold for *slx4*Δ; 6.1 vs. 1.8-fold for *rad1*Δ). These data indicate that the Slx4 complex and the Rad1-Rad10

nuclease may have an additional role in induction of fragility at Flex1, aside from their role in SSA. We confirmed this conclusion by deleting *SLX4* and *RAD1* in the Flex1 (AT)34 YAC end loss assay strain, as healing in this assay should not require Slx4 or Rad1. Indeed, fragility was significantly decreased in both backgrounds, verifying the importance of these proteins in preventing Flex1 fragility (Figure 2-4C). A *mus81*Δ *rad1*Δ strain had about the same level of fragility as each single mutant (Figure 2-4C), suggesting that they are working in the same pathway to cause Flex1 (AT)34 fragility.

In both yeast and human cells, the Slx1 (hSLX1) nuclease associates with Slx4 (hSLX4) and targets branched DNA structures (Fricke and Brill, 2003, Svendsen et al., 2009). Therefore Slx1 may also be required to process structures formed by or because of Flex1. Indeed, *slx1*Δ mutants had a decrease in fragility to a level similar to *mus81*Δ in the YAC end loss assay (Figure 2-4C), though the decrease was less dramatic at the internal chromosome location in the DDRA assay (Figure 2-4A). This decrease was specific to Flex1, as the control rate was unchanged in the *slx1*Δ mutant (Figure 2-4B). Thus Slx1 also contributes to Flex1 fragility. The *mus81*Δ *slx1*Δ Flex1 (AT)34 fragility rate is similar to that of a *mus81*Δ single mutant in the DDRA fragility assay (Figure 2-4A). Also, the *mus81*Δ *rad1*Δ and *slx1*Δ *rad1*Δ double mutants showed similar Flex1 (AT)34 fragility levels as each single mutant in the YAC end loss assay (Figure 2-4C). Overall, these results indicate that Mus81, Slx4, Slx1, and Rad1 are all working in the same pathway to cause fragility at Flex1 (AT)34, suggesting that they are functioning together to cause cleavage of a structure induced by this sequence. The consistent reduction

in fragility upon single and double knockouts of members suggest that all members of the super complex must be present for efficient cleavage and fragility occur. However, some fragility still remains which could be accounted for by the remaining nucleases or a separate source of fragility.

Yen1 protects Flex1 against fragility

S. cerevisiae Yen1 (human GEN1) is an SSE that only gains access to the DNA in mitosis, and prefers perfect 4-way junctions such as Holliday junctions (Minocherhomji and Hickson, 2014). Since Yen1 and Mus81 have overlapping substrates, Yen1 could act as a backup for Mus81 to cleave the Flex1 (AT)34 sequence. Surprisingly, removal of *YEN1* results in a significant increase in Flex1 (AT)34 fragility in the DDRA fragility assay system (Figure 2-4A). This result prompted us to investigate the order of action of Mus81 and Yen1 by creating a double mutant. In the *mus81*Δ *yen1*Δ double mutant, the recombination rate was reduced to *mus81*Δ levels, indicating that Mus81 acts upstream of Yen1 (Figure 2-4A). These results suggest that Mus81 acts before anaphase to cleave the Flex1 (AT)34 sequence whereas Yen1 has an entirely different role, for example to resolve problems persisting into anaphase. Interestingly, a *yen1*Δ had no effect on Flex1(AT)34 fragility in the YAC assay, where the repeat is near the end of a chromosome with no converging replication fork. This result suggests that Yen1 resolves a structure created from two ends, such as two replication forks that have not merged or a two-ended recombination structure, which is consistent with its ability to cleave Holliday junctions.

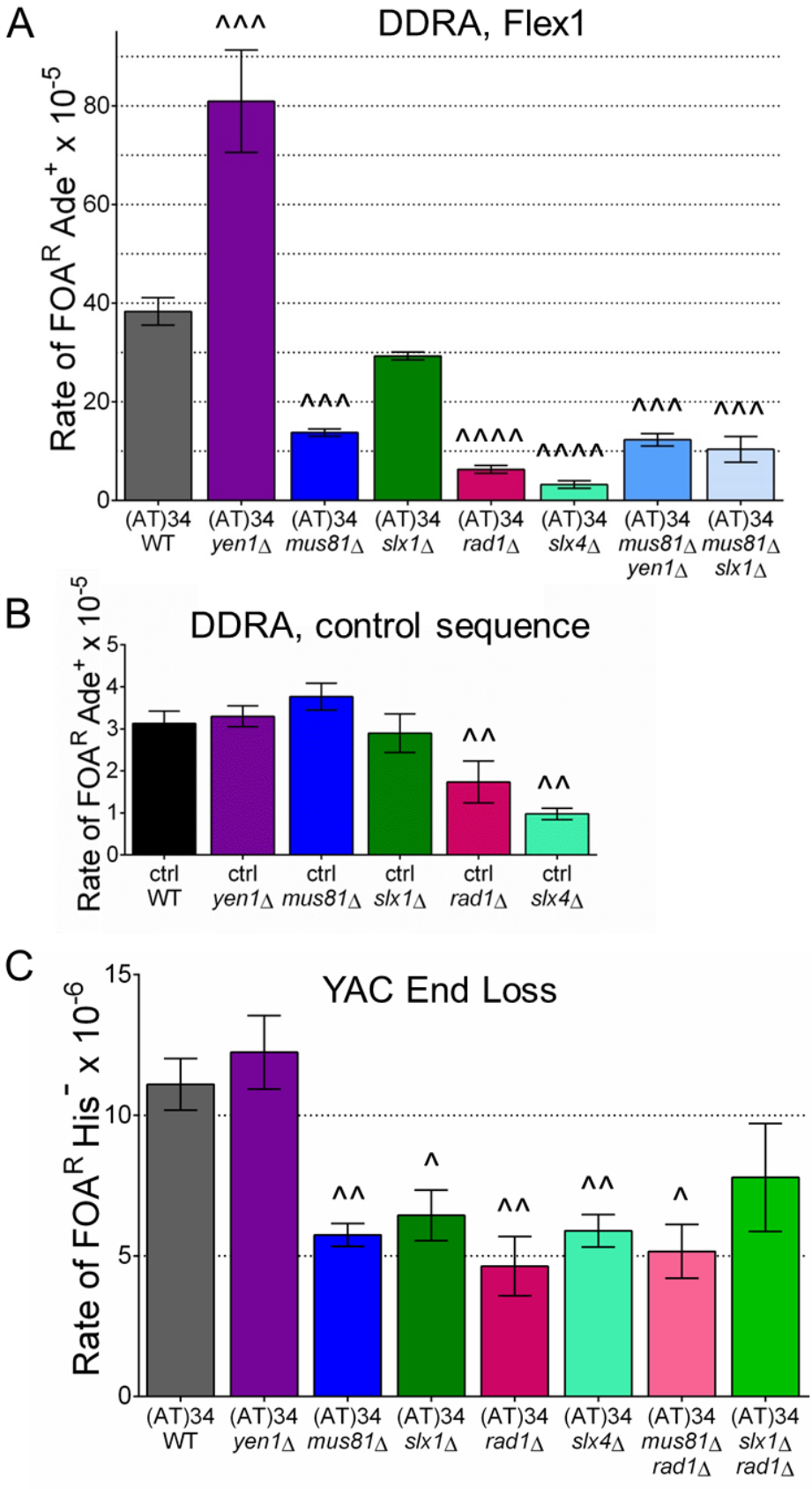

A
DDRA, Flex1
Rate of FOA^R Ade^+ x 10^-5
0
20
40
60
80
^^^
^^^
^^^^
^^^^
^^^
^^^
(AT)34 WT
(AT)34 yen1Δ
(AT)34 mus81Δ
(AT)34 slx1Δ
(AT)34 rad1Δ
(AT)34 slx4Δ
(AT)34 mus81Δ yen1Δ
(AT)34 mus81Δ slx1Δ
B
DDRA, control sequence
Rate of FOA^R Ade^+ x 10^-5
0
1
2
3
4
5
^^
^^
ctrl WT
ctrl yen1Δ
ctrl mus81Δ
ctrl slx1Δ
ctrl rad1Δ
ctrl slx4Δ
C
YAC End Loss
Rate of FOA^R His^- x 10^-6
0
5
10
15
^^
^
^^
^^
^
(AT)34 WT
(AT)34 yen1Δ
(AT)34 mus81Δ
(AT)34 slx1Δ
(AT)34 rad1Δ
(AT)34 slx4Δ
(AT)34 mus81Δ rad1Δ
(AT)34 slx1Δ rad1Δ

Figure 2-4: Fragility of AT repeat-dependent structures is dependent on Mus81, Slx4, Slx1, and Rad1 nucleases, but not Yen1. (A) Effect of deleting structure-specific endonucleases in Flex1 (AT)34 orientation 1 DDRA fragility assay strains and (B), control DDRA fragility assay strains. See also Table 2-S3. (C) Effect of deleting nucleases in indicated YAC end loss assay strains, see also Table 2-S4. Rates were tested for significant deviation from the WT of the same Flex1 construct using an unpaired t-test; ^ $p<0.05$, ^^ $p<0.01$, ^^^ $p<0.001$, and ^^^^ $p<0.0001$.

Sae2 is required for healing of Flex1

We hypothesize that the AT-rich nature of CFSs makes them more likely to form secondary structures, which can both cause fragility and inhibit healing after breakage has occurred. If this is true, proteins involved in end resection should be important for healing breaks at CFSs and minimizing CFS expression. Sae2 is required to stimulate the MRX nuclease to cleave hairpin-capped DNA ends to facilitate resection and prevent palindromic gene amplification and other deleterious rearrangements (Mimitou and Symington, 2009, Cejka, 2015), and in TALEN-induced DSB repair of CTG repeats, which form hairpins (Mosbach et al., 2018). The human homolog, CtIP, is also needed at hairpin-capped ends (Makharashvili et al., 2014). In the DDRA fragility assay a *sae2*Δ mutant had decreased healing specifically for Flex1-containing constructs, but the recombination rate of the control was unchanged (Figure 2-5A). These data indicate that Sae2 is not required for resection of non-structured DNA, but is crucial for repair of Flex1 (AT)34-induced breaks and damage recovery. MRX-Sae2 activity could be required to respond to a number of hairpin-capped structures that could form at Flex1 (see Figure 2-7 and Discussion).

Deletion of *MUS81* and *SAE2* both result in a decrease in Flex1 (AT)34 fragility, yet our interpretation of the cause is different (Figures 2-3A, 2-4, and 2-5A). Based on their known protein functions, it is likely that Mus81 induces breaks at structures formed by Flex1 while Sae2 is required to heal after breakage at Flex1. To test this hypothesis, we measured growth of the *mus81*Δ and *sae2*Δ strains in a microcolony assay. *sae2*Δ strains had a decreased microcolony area and an 8-fold

increase in delayed growth compared to the wildtype (AT)34 strain (Figures 2-5B and 2-5C), which is consistent with a defect in healing resulting in checkpoint arrest and delayed cell division. However, *mus81*Δ cells had both the same microcolony area and percentage of cells with delayed growth as wildtype, indicating that Mus81 is not needed for healing of breaks that occur at Flex1 (Figures 2-5B and 2-5C).

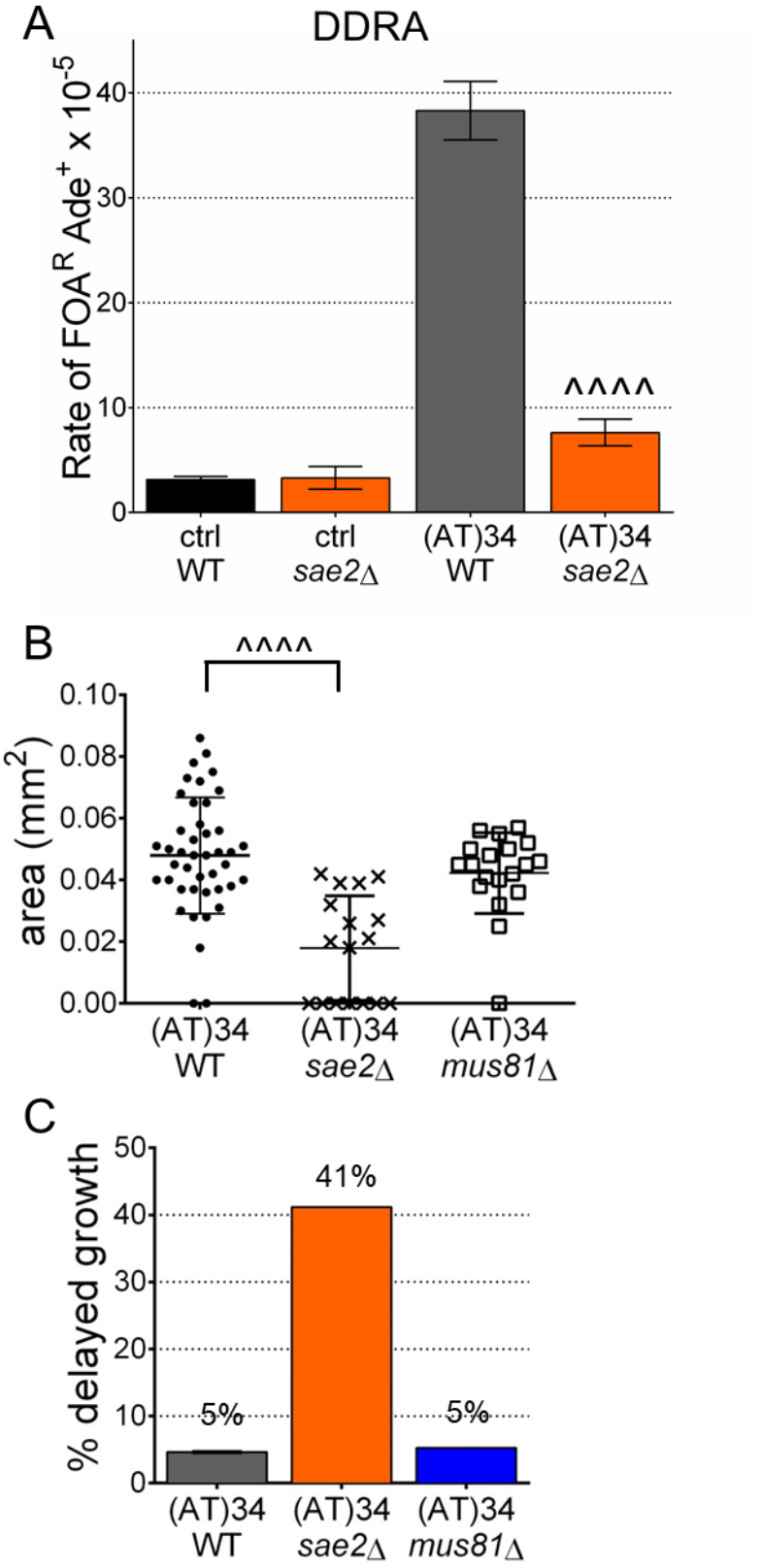

A
DDRA
Rate of FOA^R Ade^+ x 10^-5
40
30
20
10
0
^^^^
ctrl WT
ctrl sae2Δ
(AT)34 WT
(AT)34 sae2Δ
B
^^^^
area (mm^2)
0.10
0.08
0.06
0.04
0.02
0.00
(AT)34 WT
(AT)34 sae2Δ
(AT)34 mus81Δ
C
% delayed growth
50
40
30
20
10
0
41%
5%
5%
(AT)34 WT
(AT)34 sae2Δ
(AT)34 mus81Δ

Figure 2-5: Sae2 is important for healing of breaks at Flex1. (A) DDRA fragility assay data for Flex1 (AT)34 orientation 1 and control strains with *sae2*Δ; strains were tested for significant deviation from the WT with the same Flex1 AT tract length using an unpaired t-test; ^^^^ = $p<0.0001$. See also Table 2-S3. (B) Microcolony area at 30 hrs of growth and (C) percent delayed growth for WT, *mus81*Δ or *sae2*Δ strains with the indicated Flex1 (AT)34 sequences at the YAC location. n= 43 for WT (AT)34-S3', n= 23 for WT (AT)34-L3', n= 19 for *mus81*Δ (AT)34-S3', and n= 17 for *sae2*Δ (AT)34-S3'. Delayed growth cells were those with an area <0.015 mm^2 after 30 hours.

Flex1 flanking sequences affect healing in two different fragility assay systems

In the Zhang and Freudenreich study (2007), Flex1 with (AT)34 gave a significantly lower level of FOA^R than the control, which was unexpected. The data suggested that a lower efficiency of healing could be the cause of the decreased recovery of YAC end loss events, and it was hypothesized that the longer 3' (L3') flanking sequence in the Flex1 (AT)34 construct compared to the other constructs could play a role (Zhang and Freudenreich, 2007).The L3' flanking sequence is 102 bp longer than the short 3' flanking sequence (S3') and the Mfold program (Zuker, 2003) predicts that this extra 102 bp can form a stable hairpin with a ΔG of -6.7 (Figure 2-S5). Human polymerase δ was paused at the L3' sequence, providing evidence that a secondary structure is forming at that sequence (Figure 2-2C, black arrow and black box).

We evaluated the rate of FOA^R of Flex1 (AT)34 strains with L3' and S3' flanking sequences in the DDRA fragility assay, with the prediction that strains with a L3' flanking sequence will have lower rates due to the hairpin inhibiting proper resection and healing. As predicted, the strain with the L3' flank has a significantly decreased level of FOA^R compared to the S3' strain, supporting the hypothesis that the L3' flanking sequence inhibits healing (Figure 2-6B). In the YAC assay, if the additional sequence in the L3' flanking sequence forms a hairpin that inhibits this leftward resection to the telomere seed sequence, the rate of FOA^R should decrease upon the presence of the L3' only in orientation 1, since leftward resection proceeds through the 3' flanking sequence after breakage at or near the AT repeat only in this orientation (Figures 2-6A and 2-S2B). Indeed, the presence of the L3' sequence inhibits healing in orientation 1, as FOA^R His^- rates are significantly decreased compared to the orientation 1 S3' strain (Figure 2-6C). The identity of the 3' sequence did not affect recovery in the YAC assay when it was to the right of the AT repeat in orientation 2, which is consistent with breakage occurring at the AT repeat, followed by leftward resection.

While there is compelling genetic evidence to support that the L3' flanking sequence forms a hairpin that inhibits healing after breakage in our assays, it was also possible that the presence of the L3' flanking sequence actually reduces fragility. To distinguish these possibilities, the Flex1 AT repeat was replaced by an I-SceI recognition sequence in the DDRA fragility assay system so that DSBs could be induced adjacent to the Flex1 flanking sequences. Three strains were created: (1) one with only the I-SceI recognition sequence (breakage without any

expected healing impairments), (2) one with the I-SceI recognition sequence flanked by the S5' and the L3' Flex1 flanking sequences (breakage with healing impairment by L3' hairpin(s)), and (3) one with the I-SceI recognition sequence flanked by the Flex1 S5' and S3' sequences (breakage without much healing impairment by flanks) (Figure 2-6D). The S5'-I-SceI-L3' strain had a reduced recombination rate compared to the I-SceI or S5'-I-SceI-S3' strains (Figure 2-6E). These results further support the conclusion that the hairpin structure(s) present in the long 3' flanking sequence reduces healing by inhibiting resection.

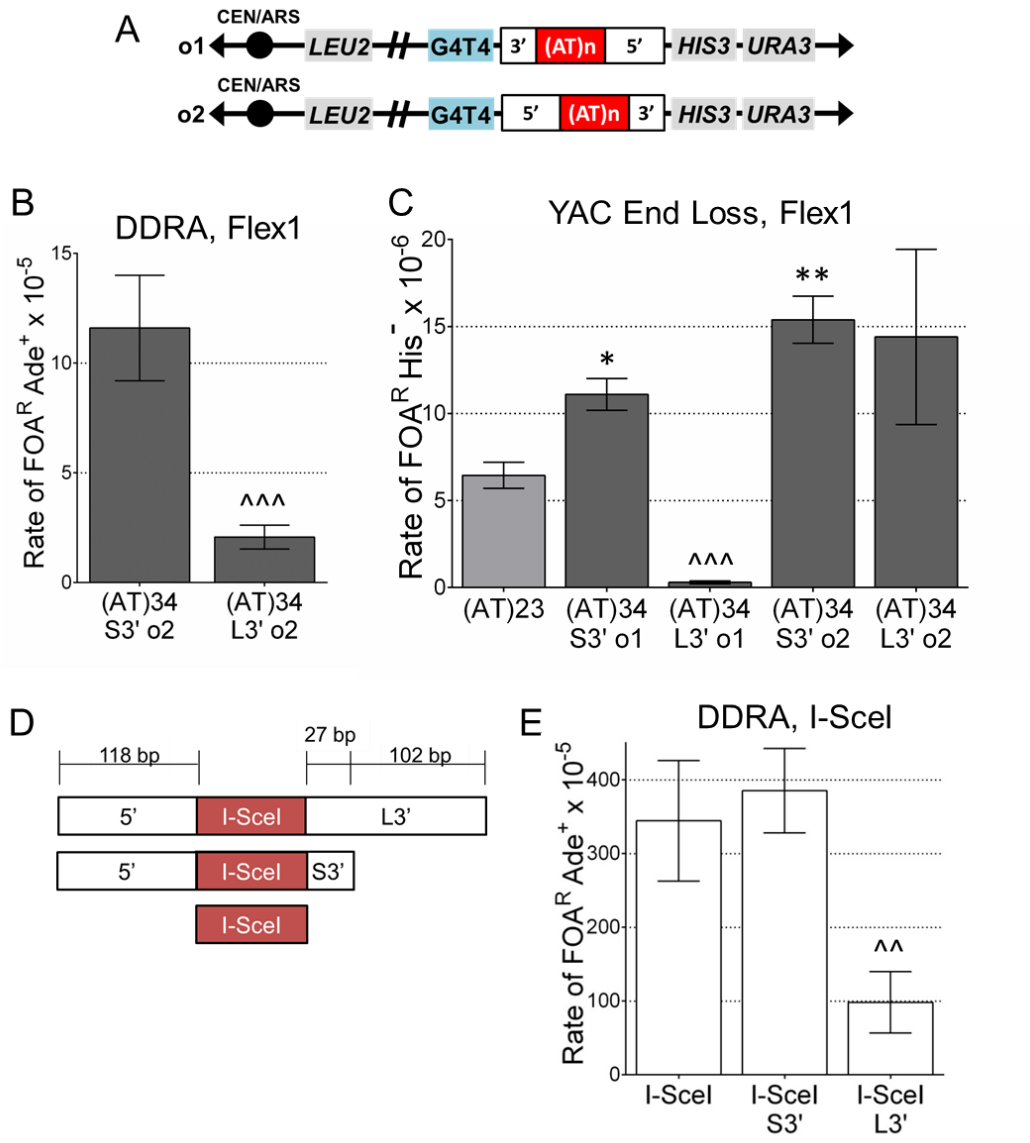


Figure 2-6: Flex1 has flanking hairpin sequences which inhibit healing of DSBs. (A) Schematic of YAC end loss assay showing Flex1 with its 5' and 3' flanking sequences in orientations 1 and 2; long and short 3' designated as L3' and S3', respectively. (B) DDRA fragility assay data for Flex1 (AT)34 in orientation 2 and (C) YAC end loss assay data showing lower rates of recombination and YAC end loss when resection must proceed through the L3' flanking sequence. L3' strains were tested for significant deviation from the same

orientation S3’ strain using an unpaired t-test; ^^^ $p<0.001$ compared to (AT)34-S3’ same orientation. *$p<0.05$, **$p<0.01$ compared to (AT)23. See also Tables 2-S3 and 2-S4. (D) Schematic of three I-SceI strains created. Either an I-SceI recognition sequence only or an I-SceI recognition sequence with Flex1 flanking sequences was inserted into the DDRA fragility assay locus in orientation 1. (E) DDRA fragility assay data for all three I-SceI strains under ~50% galactose induction of I-SceI breaks, showing lower rates of recovery of FOA^R Ade+ recombinants when the Flex1 L3’ flanking sequence is present; ^^ $p<0.01$ compared to I-SceI with S3’ flanking sequence using an unpaired t-test. See also Table 2-S3.

Discussion

Flex1 is an important component of FRA16D fragility

We have demonstrated that Flex1, a roughly 300 bp subregion, is an important determinant of FRA16D breakage *in vivo*, as large FRA16D-containing YACs with Flex1 replaced go from 18.0% to 12.6% chromosome end loss. This indicates that Flex1 could account for roughly 30% of the breaks happening at FRA16D even though it accounts for only 0.02% of the sequence on the large FRA16D-containing YAC. Flex1 fragility is dependent on AT length and the Mus81, Slx1-Slx4, and Rad1-Rad10 nucleases, and is increased in the absence of Pol32 or Yen1. Notably, Flex1 is acting very similarly to FRA16D in the context of a whole human chromosome, where human MUS81, SLX4, and XPF-ERCC1 are required for full CFS expression and MiDAS (Naim et al., 2013, Ying et al., 2013, Minocherhomji et al., 2015). Our data indicate that Flex1 could be one of the specific fork-stalling regions causing MUS81-EME1 nuclease activity at FRA16D in humans. Comparing recovery of breaks at Flex1 with different flanking sequences provided support for a novel hypothesis for CFS expression: CFSs are not only enriched in sequences that cause fragility but deficient in their ability to heal after break induction.

Long uninterrupted AT repeats cause chromosome fragility and polymerase stalling

We find that Flex1 is fragile in an AT repeat length-dependent manner when the flanking sequences are standardized in our DDRA fragility assay. Our findings are consistent with previous studies showing that AT repeats form cruciforms on

plasmids *in vivo* when the AT stem exceeds 22 bp (McClellan et al., 1990, Dayn et al., 1991, Bowater et al., 1991, Cote and Lewis, 2008). Because cruciforms form in dsDNA, they need to overcome the energy of base-pairing to form and they exhibit non-linear properties. Therefore the dramatic AT-length dependence of fragility is consistent with cruciform formation. In contrast, hairpin formation typically occurs in ssDNA and is therefore governed more by the pairing strength (ΔG) of the base-pairs in the stem. Human polymerase δ holoenzyme exhibited significant stalling at the Flex1 (AT)34 repeat tract *in vitro*. Since these assays were performed on ssDNA, they indicate that AT hairpins can also be a significant replication barrier. In contrast, the Flex5 sequence, which is an *in vitro* replication barrier (Shah et al., 2010, Walsh et al., 2013, Barnes et al., 2017), lacked an *in vivo* fragility phenotype in our DDRA fragility assay, despite Flex5 (AT)24i differing from Flex1 (AT)23 by just a few bases. These data support AT cruciform formation as the most relevant *in vivo* structure causing fragility at Flex1, as the interruptions present in the Flex5 AT tract are predicted to significantly reduce the likelihood of cruciform but not hairpin formation. The differing flanking sequences may play an additional role in explaining the fragility exhibited by Flex1 but not Flex5 (see below). The model of a fork encountering a pre-formed cruciform is strengthened by the result that HU treatment is not required for fragility or fork stalling at Flex1 (AT)34 sequences (Figure 2-S3) (Zhang and Freudenreich, 2007). For example, a cruciform could arise during transcription due to increased negative supercoiling caused by passage of RNA polymerase, and then block replication without the need for an

additional stressor. The strong AT length dependence observed suggests that individuals with longer AT repeats at Flex1 will be at a significantly greater risk of chromosome fragility and associated deletions or rearrangements at FRA16D.

The protein Pol32 is needed for many replication and repair pathways, and thus the increase in Flex1 DDRA upon its deletion has many interpretations. Pol3 and Pol31 make up the *S. cerevisae* polymerase δ complex, and Pol32 serves as a processivity factor to prevent frequent polymerase pausing (Burgers and Gerik, 1998). If Pol32's processivity factor role is the cause of the increase in fragility at Flex1 upon its deletion, we should see a similar 2.7-fold increase in fragility in *pol32*Δ ctrl strains compared to WT ctrl strains. Pol32 may function at Flex1 in its capacity as a break-induced replication (BIR) factor (Lydeard et al., 2007, Anand et al., 2013) in order to restart broken replication forks when one end of a DSB shares homology with a donor sequence. To determine whether the pol32 effect is due to a replication versus repair issue, we can make *rad51*Δ and *pol32*Δ *rad51*Δ knockouts, as Rad51 is needed strand invasion and D-loop formation, which is needed during BIR. If *rad51*Δ single mutants show a similar increase in fragility to *pol32*Δ single mutants, this would support that the *pol32* phenotype is from a lack of repair. The Hickson lab showed that MUS81 cleavage is needed in order for POLD3-mediated MiDAS to occur (Minocherhomji et al., 2015). Thus, a *mus81*Δ *pol32*Δ double deletion can determine the order of action of each protein at Flex1 in *S. cerevisiae*. If the *mus81*Δ *pol32*Δ double deletion rate is as low as the rate of the *mus81*Δ single, then Mus81 cleavage is needed for Pol32-mediated synthesis at Flex1, and this could provide the first evidence that MiDAS may also

be an active process at structure-forming DNA in *S. cerevisiae* as has been found in humans. If the *mus81*Δ *pol32*Δ double deletion rate is high, similar to the *pol32*Δ mutant, then a different mechanism is protecting against fragility at Flex1 in yeast versus humans. There is also evidence that a Rev3-Rev7-Pol31-Pol32 complex is responsible for translesion synthesis (TLS) in yeast (Makarova et al., 2012). Therefore, it is possible that TLS synthesis is needed for faithful replication through secondary structure forming regions of Flex1. If this is the case, there should not be an increase in point mutations via reporter mutagenesis assay. Further, a similar DDRA rate increase should be seen upon a *rev3*Δ deletion and a *rev3*Δ *pol32*Δ double deletion in the Flex1 background, as Rev3 is the catalytic subunit of DNA polymerase ζ. The absence of Pol32 can have negative impacts on so many different pathways that can be protecting against fragility at Flex1. Thus, without Pol32 there is likely increased fragility at Flex1, and then a shift towards the SSA pathway of healing, resulting in an increased DDRA rate in our system. A shift to end-joining pathways is also likely in *pol32*Δ mutants, although these events likely wouldn't allow for the specific ADE^{+} FOA^{R} phenotype selected for in the DDRA.

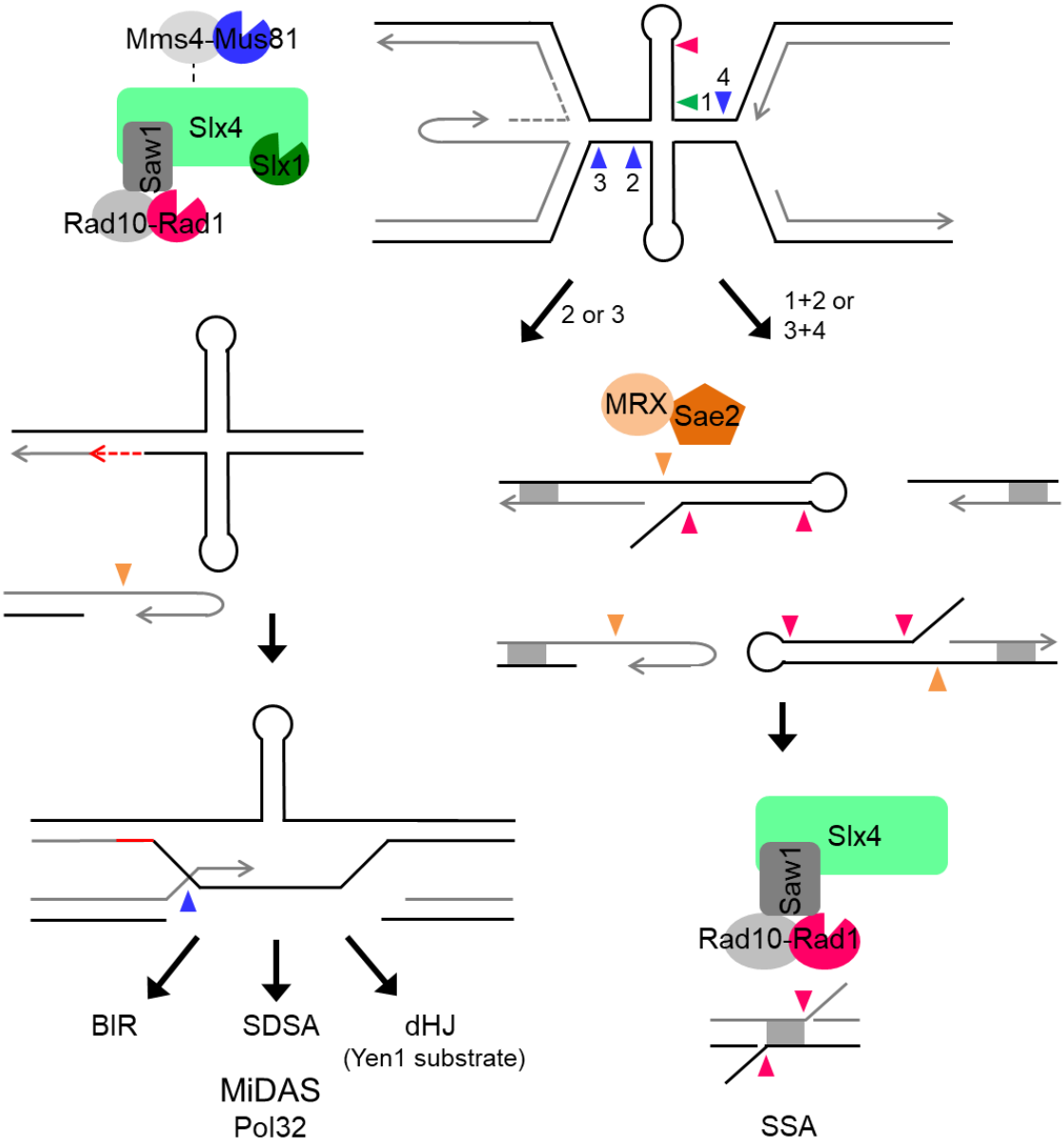


Figure 2-7: A model for Slx4 super complex cleavage of stalled fork substrates formed by secondary structures at Flex1. Secondary structure forming sequences at Flex1 cause replication fork stalling, which can potentially result in a reversed fork and convergence of the fork approaching from the right in the chromosome II situation. We propose that either a DNA structure and/or the stalled fork is cleaved by an Slx1-Slx4-Mus81-Mms4-Rad1-Rad10 complex,

acting together or sequentially; possible cleavage positions of various nucleases in the complex are designated with appropriately colored arrows. Only two cleavage outcomes are depicted, though others are possible. Mus81 cleavage at a stalled fork approaching from the left (arrow 3) will produce a one-ended break (left pathway). The broken end, which may be processed by MRX-Sae2, can invade the intact sister (repaired by gap filling) to initiate repair by homologous recombination. Repair outcomes include MiDAS-like break-induced replication (BIR), synthesis dependent strand annealing (SDSA), or second-end capture and double Holliday junction resolution by Yen1. Alternatively, if cleavage at stalled forks on either side of the cruciform occurs (arrows 3 and 4) or at the cruciform 4-way junction (by coordinated Slx1-Mus81 cleavage, arrows 1 and 2), 4 ends will be produced. The hairpin-capped ends can be processed by MRX-Sae2. Rad1 could also process hairpin loops and/or non-homologous flaps. Recombinants are recovered by SSA at homologous sequences (e.g. “DE” region of homology denoted by grey box), resulting in deletion of intervening sequences. 3’ non-homologous flaps created during SSA require Rad1-Rad10 and Slx4 nucleases for processing.

Mus81, Slx4, Slx1, and Rad1 structure-specific endonucleases cause AT-repeat length-dependent cleavage

The human SSEs MUS81-EME1 and XPF-ERCC1 and the SLX4 protein have roles in CFS expression and are known to cleave stalled forks and recombination intermediates (Ying et al., 2013, Naim et al., 2013, Symington et al., 2014, Dehe and Gaillard, 2017), hence we sought to determine the roles of their *S. cerevisiae* homologs at Flex1. Our data indicate that either the AT repeat itself or a resulting stalled fork or recombination structure is cleaved by Mus81, potentially acting in an Slx1-Slx4-Mus81-Mms4-Rad1-Rad10 (SMR) DNA repair super complex, however it is also possible that the nucleases could be acting separately and sequentially.

A fork stalled by a cruciform structure presents several potential SSE substrates (Figure 2-7). First, the cruciform base or loops could be targeted. A cruciform formed by a perfect AT repeat on a plasmid is cleaved by Mus81 in *S. cerevisiae* (Cote and Lewis, 2008). Slx4, via its interaction with Slx1 and (indirectly) with Mus81, could relax the substrate specificity of each component nuclease and allow for an Slx1-mediated nick (Figure 2-7, green arrow, number 1) followed by a Mus81-mediated counter nick (Figure 2-7, blue arrow, number 2) to cleave at the base of the cruciform, as found for their human counterparts by the West lab (Wyatt et al., 2013, Wyatt et al., 2017). Alternatively, the 3-way junction of a stalled fork or 4-way junction of a reversed fork could be targeted for cleavage (arrows 3 and 4). The nearest replication fork in the DDRA fragility assay system on chromosome II approaches Flex1 from the left (Figure 2-S2A) (Freudenreich

et al., 1997), but if the fork is unable to restart, a converging fork could approach from the right (Figure 2-7, top). Either coordinated cleavage at the cruciform base or cleavage of both forks would result in two hairpin-capped ends that would require processing by the Sae2-Mre11 nuclease (Figure 2-7, right pathway, orange arrows); formation of hairpins by the flanking sequence could provide an additional resection barrier. This pathway would favor SSA and recovery in our assay. Rad1-10 could target the hairpin loop, as was recently shown for human XPF-ERCC1 at an inverted repeat structure (Lu et al., 2015), and would also be required for cleavage of non-homologous flaps at hairpin-capped ends or during SSA (Figure 2-7, pink arrows).

If Mus81 cleaves only one side of a stalled or reversed fork, this would result in a one-ended break (Figure 2-7, left pathway). The uncleaved sister chromatid can be repaired by DNA synthesis across the gap, possibly via TLS polymerase activity. The cleaved sister chromatid's free 3' end could invade the uncleaved sister chromatid to finish repair replication via a MiDAS-like mechanism requiring the Pol32 subunit of Pol δ (mammalian ortholog POLD3) – the synthesis could continue to the end of the chromosome or until it meets a converging fork. The requirement for Pol32 to prevent Flex1 (AT)34 fragility is consistent with MiDAS being an important pathway for recovery of cleaved forks. Mus81 could also act to cleave and resolve a repair synthesis intermediate (Figure 2-7, bottom left), however this model does not fit with the requirement for Mus81 for POLD3 recruitment and MiDAS in human cells, and recruitment of SLX4 and MUS81-EME1 to CFSs in early mitosis before POLD3 (Naim et al., 2013,

Minocherhomji et al., 2015, Bhowmick et al., 2016). It would also require an alternative mechanism of generating the broken fork.

If a second end becomes available (e.g. from the converging fork), the free 3' end could also participate in synthesis dependent strand annealing (SDSA), or second-end capture and double Holliday junction (dHJ) resolution. The dHJ could be dissolved by the Sgs1-Top3-Rmi1 complex (human BTR) or branch migrated and resolved by Yen1 (human GEN1) in late mitosis, consistent with the requirement for Yen1 to protect against fragility in a pathway dependent on Mus81.

Potential action of an SMR trinuclease complex at DNA structures

The nuclease action at Flex1 could be through separate Mus81-Mms4, Slx1-Slx4, and Rad1-Rad10 complexes or through a larger complex of all proteins recruited by Slx4. The *mus81*Δ *slx1*Δ and *mus81*Δ *rad1*Δ double mutant fragility assay rates resembled that of a *mus81*Δ single knockout, supporting their activity in a super complex similar to the SMX DNA repair trinuclease complex that has been observed in human cells *in vitro* (Wyatt et al., 2017). In *S. cerevisiae*, there is some evidence that Slx4 binds either Slx1 or Rad1-Rad10, but not both, as reviewed in (Dehe and Gaillard, 2017). However, more recent studies indicate that sumoylation of Saw1 may coordinate Slx1-Slx4 and Rad1-Rad10 cleavage in response to UV (Sarangi et al., 2014) and therefore this super complex may exist in yeast under conditions of DNA damage. The reduction in fragility of *slx4*Δ mutants in both genetic assays supports the importance of the Slx4 scaffolding protein for coordinating cleavage at DNA structures caused by Flex1. *slx1*Δ mutants showed a significant decrease in the YAC end loss assay and a small,

non-significant, decrease in fragility in the DDRA fragility assay, implying that Slx1 might be less important than Mus81 in a situation where a converging fork is present. Mus81-dependent cleavage of a resected fork is expected to have a lesser dependence on Slx1 compared to cleavage of an intact 4-way junction, which lends support to a stalled fork as the relevant substrate.

Yen1 protects Flex1 against fragility

Yen1 (human GEN1) is sometimes considered a backup nuclease to Mus81. Like *mus81*Δ, the *yen1*Δ effects are specific to Flex1 (Figure 2-4). However, unlike the other SSEs tested, *yen1*Δ mutants had an increase in fragility, indicating that Yen1 protects against Flex1-induced fragility. Interestingly, the effect of deleting Yen1 was much more evident when the Flex1 (AT)34 sequence was in the middle of chromosome II (DDRA assay) compared to the end of a chromosome (YAC assay). Yen1 could function to resolve a double Holliday junction, a situation more likely to arise when there is a second end capture, which is not available when there is no incoming fork as on the end of the YAC. On chromosome II, Flex1 (AT)34 *yen1*Δ *mus81*Δ mutant fragility is equivalent to *mus81*Δ fragility levels, indicating that Mus81 acts upstream of Yen1 in the same pathway. This is consistent with the known timing of Mus81/MUS81 action earlier in the cell cycle than Yen1/GEN1 (Blanco and Matos, 2015, Dehe and Gaillard, 2017). For example, one scenario that fits with our data is that SMR cleaves a secondary structure and/or stalled fork at Flex1, which is then resolved by either SSA (Figure 2-7, right pathway) or recombination (Figure 2-7, left pathway). If recombination occurs, a dHJ intermediate may result, requiring cleavage by Yen1.

Alternatively, incomplete MiDAS by Pol δ might leave connected sister chromatids that would require Yen1 resolution, and in its absence mechanical chromosome breakage could occur. SSA is a pathway that could rescue these breaks, resulting in deletions and recovery in our assay.

Structures that flank a fragile site impair resection and alter repair outcomes

Our data show that a hairpin predicted to form in the flanking sequence of Flex1 inhibits healing in both of our genetic assays or when placed adjacent to an induced DSB, and causes polymerase δ pausing in a primer extension assay. These data bring about a new hypothesis for common fragile site fragility: CFS expression could be a combination of cleavage and processing of stalled forks, and inefficient healing due to the presence of multiple contiguous sequences that form secondary structures.

Sae2 is required to process breaks that occur at Flex1, and the absence of Sae2 severely reduces recovery of broken chromosomes and negatively impacts cell growth and division. These results are consistent with the known activity of Sae2 in stimulating Mre11 nuclease processing of hairpin-capped ends (Mimitou and Symington, 2009, Cejka, 2015). These ends could result from SMR cleavage near the base of the cruciform to produce AT hairpin-capped ends, or from fold-back of flanking hairpins (for example on a reversed fork end) (Figure 2-7). In a similar assay in mammalian cells, CtIP was found to be essential for recovering breaks at Flex1 by SSA but was not required at clean I-SceI DSBs (Wang et al., 2014), and CtIP also functions as a co-factor of MRN nuclease in mammalian cells (Anand et al., 2016). Therefore, this appears to be a conserved pathway, and is likely

operating at naturally occurring breaks at FRA16D in human cells. Indeed, deletion of the Rad50 component of yeast MRX caused increased death of cells containing FRA16D on the large 801B6 YAC that was exacerbated by replication stress (Zhang and Freudenreich, 2007). Since Sae2 prevents translocations in yeast (Deng et al., 2015), it is likely that MRN-CtIP prevents genomic rearrangements at Flex1 in FRA16D.

Our data explain the propensity for FRA16D, the Flex1 region in particular, to be deleted in cancer cell lines. Since resection is an important feature of almost all cellular DSB repair mechanisms, our results predict that breaks that occur within structure-forming DNA in human cells will have a reduced efficiency of healing, which may favor alternative and less conservative repair pathways that generate translocations or large deletions.

Implications for genome stability and cancer initiation

Our data show that nuclease cleavage is only relevant for Flex1 sequences with 23 or more AT repeats, which corresponds to a size that can form a cruciform and stall replication forks *in vivo*. This predicts that individuals with longer AT alleles at Flex1 will be more reliant on the SMX nuclease to process stalled forks and prevent deleterious translocations and deletions in the FRA16D region. Thus, the ability to respond to replication stress caused by DNA structures by the regulated action of nucleases may be an important cancer protective mechanism (Fragkos and Naim, 2017). Our study points to Flex1 as a valid therapeutic target to reduce genome instability in individuals with greater than 23 ATs at Flex1. Cleavage of other naturally occurring palindromes in human cells has been shown to occur *in vivo*, leading to translocations (Kato et al., 2014, Inagaki et al., 2013), which have also been found at FRA16D in multiple myeloma patients (Ried et al., 2000). Thus, the mechanisms described here could be generally applicable to many cruciform-forming structures in the human genome.

Acknowledgements

We thank Brian Lenzmeier for sharing reagents to construct the ADE2 version of the DDRA assay, Francesca Storici for the inducible I-SceI system, Ryan McGinty, Julia Haft, and Michael Sigouros for help with yeast strain construction, verification, and some assays, Suzanne Hile for help with analysis of the *in vitro* extension data, Allen Su, Keerthana Gnanapradeepan, and Adam Snider for help with cloning. Funding was provided by NSF MCB1330743, NIH GM105473, NIH GM122880 and Tufts Dean's Fund awards to CHF, Arnold and Mabel

Beckman Foundation award to CHF and CEW, and Tufts Graduate Student Research Award to SK.

Author Contributions

Conceptualization, S.K., and C.H.F.; Methodology, S.K., K.A.E. and C.H.F.; Validation, S.K., K.A.E. and C.H.F; Formal Analysis, S.K.; Investigation, S.K., C.E.W., S.B.R., A.H., R.P.B., S.M.L., and N.C.M.H.; Resources, C.H.F., K.A.E., B.L, F.S., X.S, K.G., A.S.; Writing – Original Draft, S.K. and C.H.F.; Writing – Review & Editing, S.K., C.H.F., S.M.L., N.C.M.H, C.E.W., S.B.R., R.P.B., K.A.E; Visualization, S.K.; Supervision, S.K., K.A.E. and C.H.F.; Project Administration, K.A.E., and C.H.F.; Funding Acquisition, S.K., C.E.W., K.A.E. and C.H.F.

Declaration of Interests

The authors declare no competing interests.

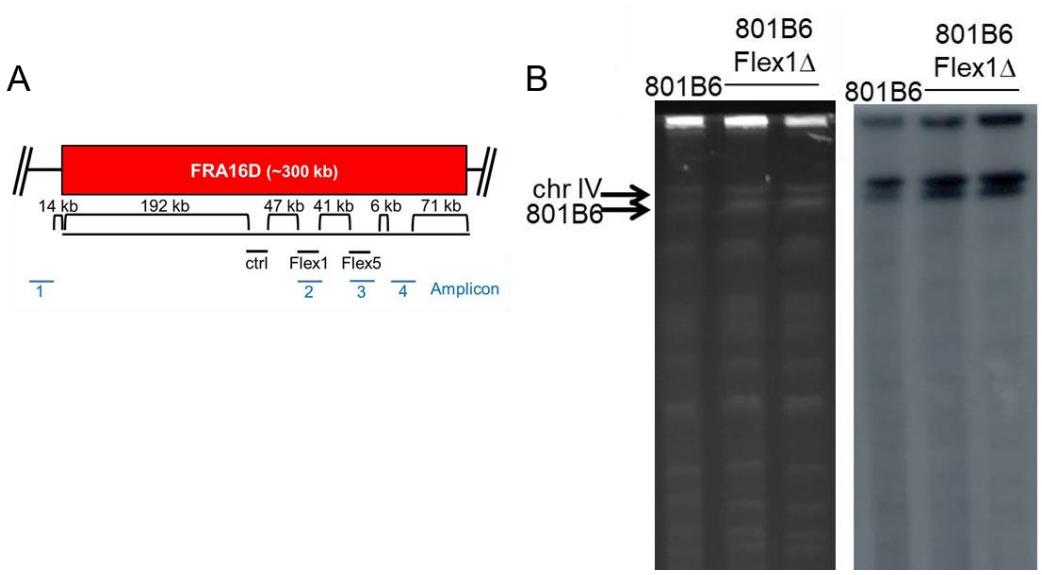


Figure 2-S1 (related to Figure 2-1): Confirmation of FRA16D YAC integrity. (A) Large FRA16D YAC structure was verified by PCR amplifying the indicated amplicons. See PCR results in Table 2-S1. (B) Overall size of the 801B6 YAC (~1400 kb) was verified by pulsed field gel electrophoresis (left side of figure) of intact chromosomes followed by a Southern blot using a probe to *TRP1* (right side of figure). The probe binds to the *TRP1* marker on the YAC (~1500 kb) as well as the *trp1-289* allele on chromosome IV. The 801B6 YAC contains Flex1 (AT)34 by PCR and sequencing (data not shown).

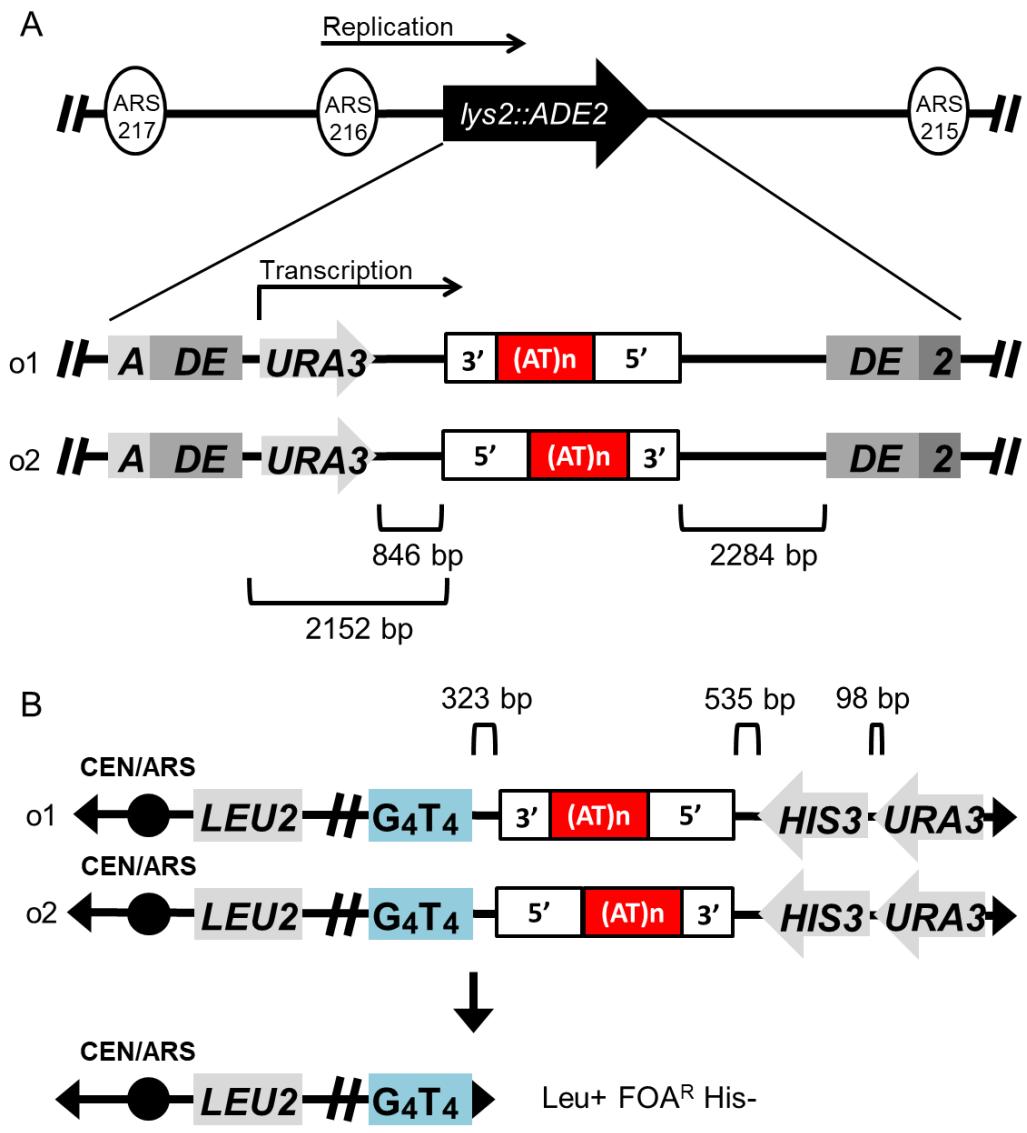


Figure 2-S2 (related to Figures 2-2A, 2-2C, 2-3, 2-4, 2-5A, 2-6B, 2-6C, and 2-S3): Assay constructs. (A) A detailed depiction of the DDRA fragility assay cassette at the *LYS2* locus on chromosome II is shown. (B) A detailed depiction of the YAC end loss assay and the yeast artificial chromosome showing Flex1 in orientations 1 (o1) and 2 (o2).

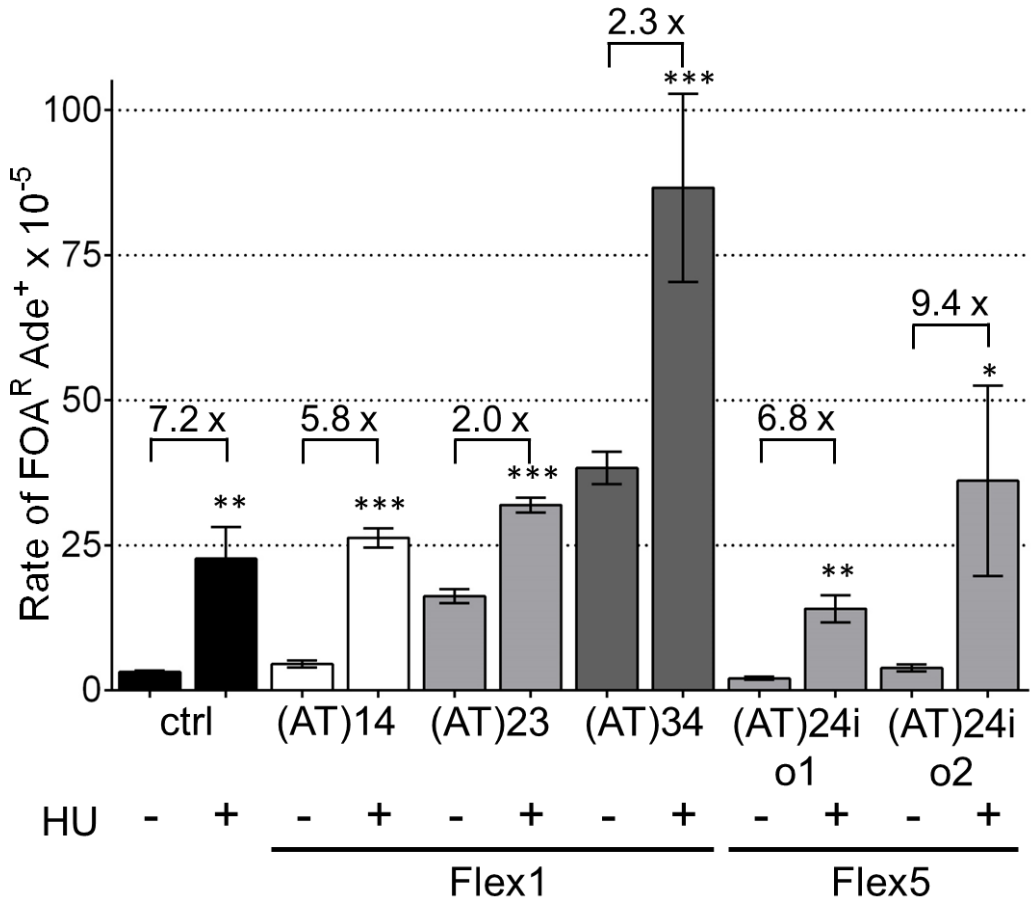


Figure 2-S3 (related to Figures 2-1A, 2-2A and 2-2B): HU Increases fragility at all FRA16D subregion sequences tested. The DDRA assay was used to evaluate fragility of Flex1 orientation 1 and Flex5 when cells were grown in the presence of and absence of 100 mM HU. The sequence of Flex5 (AT)24i is: $(TA)_{20}AA(TA)_3T$. HU-treated strain recombination rates were tested for significant deviation from the same strain grown in non-HU conditions using an unpaired t-test; * $p<0.05$, ** $p<0.01$, *** $p<0.001$, and **** $p<0.0001$. The fold increase in recombination rate upon HU treatment is reported above each appropriate pair of rates.

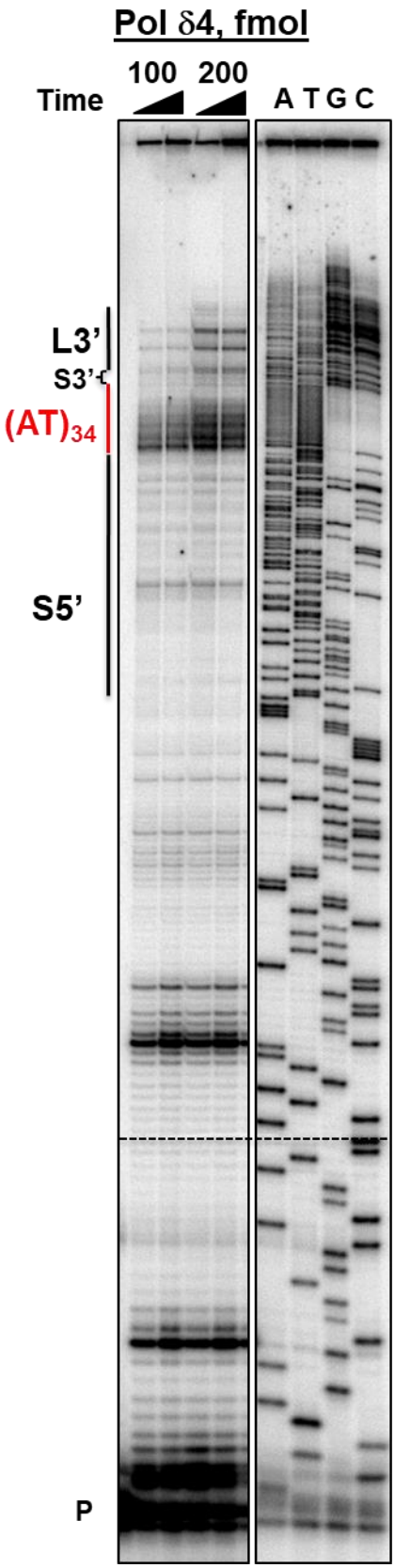

Pol δ4, fmol
100
200
Time
A T G C
L3'
S3'
(AT)34
S5'
P

Figure 2-S4 (related to Figure 2-2C): Flex1 (AT)34 stalls human polymerase delta on both DNA strands. *In vitro* DNA synthesis of Flex1 with (AT)34 and a L3' flanking sequence by the 4-subunit human polymerase δ holoenzyme (Pol δ4), showing pause sites at the (AT)34 repeat. This reaction utilizes the opposite ssDNA template for synthesis compared to Figure 2-2C. Sequence outside of the marked area is composed of the plasmid backbone. TACG, dideoxy sequencing ladder of the DNA template.

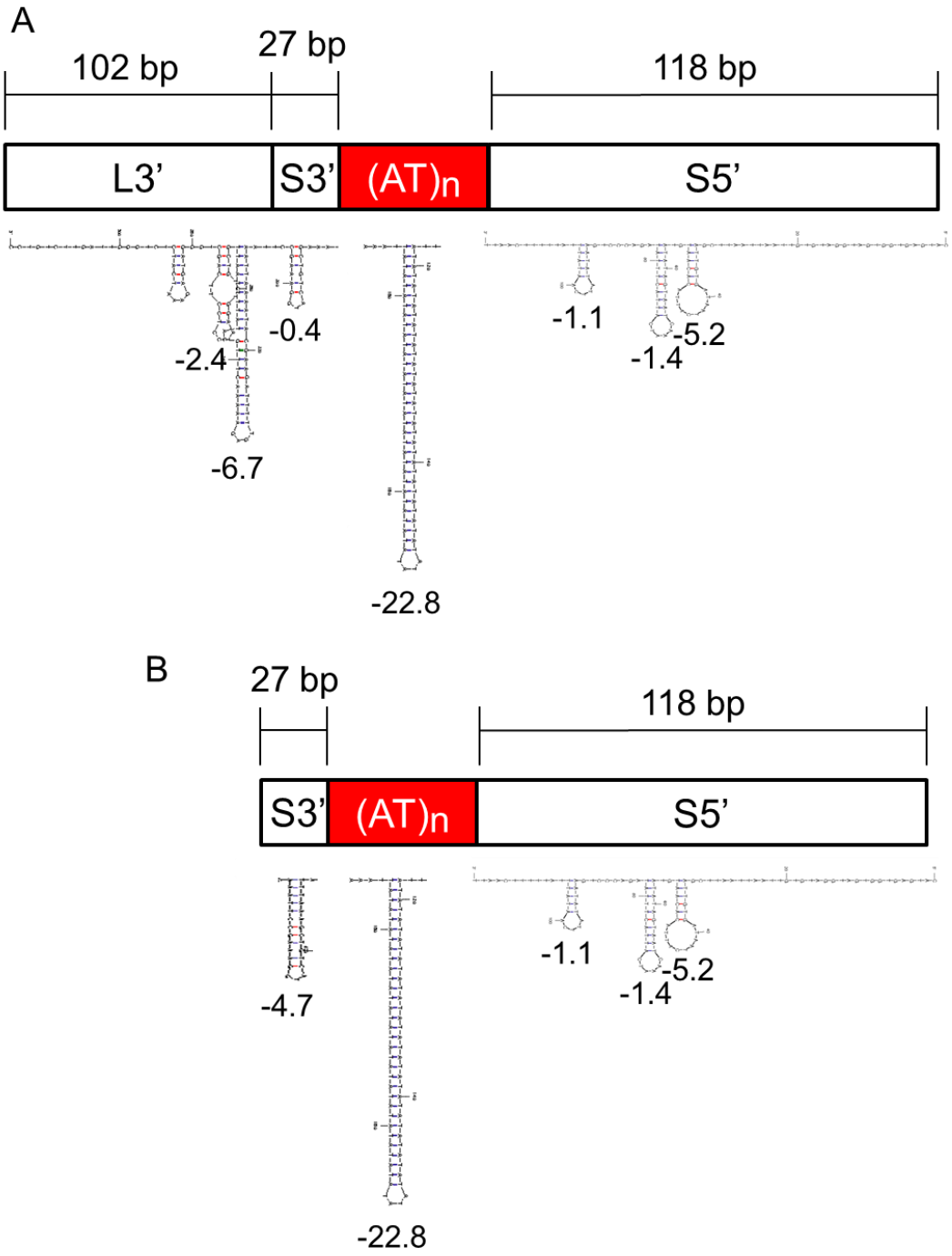


Figure 2-S5 (related to Figure 2-6): Secondary structure predictions for Flex1 with various flanking sequences. Secondary structure predictions for sequences contained within Flex1 with a L3' (A) and S3' (B) flanking sequence.

ΔG values of each predicted hairpin are reported below the structure. Note that the sequence between hairpins is non-contiguous for illustration purposes.

Table 2-S1. Related to Figures 2-1 and 2-S1. FRA16D YAC Subregion PCR Results.

Strain #	**Amplicon**	**PCR Product Expected?**	**PCR Product?**
#1051 801B6	1	Yes	Yes
	2	Yes	Yes
	3	Yes	Yes
	4	Yes	Yes
#3076 801B6 Flex1Δ	1	Yes	Yes
	2	No	No
	3	Yes	Yes
	4	Yes	Yes
#3077 801B6 Flex1Δ	1	Yes	Yes
	2	No	No
	3	Yes	Yes
	4	Yes	Yes

Table 2-S2. Related to Figure 2-1B. % FOA^R colonies in large FRA16D YACs.

YAC strain	**# of Experiments**	**Average % FOA^R**	**SEM**	**p value**	**p compared to**
972D3	3	4.1	0.2646		
801B6	6	18.1	1.4241	0.0003	972D3
801B6 Flex1Δ	4	12.6	0.4033	0.0166	801B6

Table 2-S3. Related to Figures 2-2B, 2-3, 2-4A, 2-4B, 2-5A, 2-6C, 2-6B and 2-S3. DDRA fragility assay data. All Flex1 constructs contain the S3' flanking sequence and are in orientation 1 unless otherwise noted.

FRA16D sequence	**Deleted gene(s) or treatment**	**# of Experiments**	**Average FOA^R x 10^{-5}**	**SEM**	**p value**	**p compared to**
ctrl		6	3.1	0.2883		
ctrl	+HU	4	22.7	5.4501	0.0020	ctrl
Flex1 (AT)14		3	4.5	0.6028	0.0499	ctrl
Flex1 (AT)14	+HU	3	26.3	1.6586	0.0002	Flex1 (AT)14
Flex1 (AT)23		5	16.2	1.2178	<0.0001	ctrl
Flex1 (AT)23	+HU	3	31.9	1.2785	0.0002	Flex1 (AT)23
Flex1 (AT)34		7	38.3	2.7815	<0.0001	ctrl
Flex1 (AT)34	+HU	3	122.4	37.0016	0.0058	Flex1 (AT)34
Flex5 o1		4	2.1	0.2780	0.0365	ctrl
Flex5 o1	+HU	3	14.0	2.3483	0.0002	Flex5 o1
Flex5 o2		5	3.9	0.5953	0.2626	ctrl

Flex5 o2	+HU	3	34.8	9.5620	0.0046	Flex5 o2
ctrl	*mus81*Δ	3	3.8	0.3180	0.2203	ctrl
Flex1 (AT)14	*mus81*Δ	3	4.2	0.3606	0.6913	Flex1 (AT)14
Flex1 (AT)23	*mus81*Δ	3	3.9	0.1667	0.0003	Flex1 (AT)23
Flex1 (AT)34	*mus81*Δ	3	13.8	0.7219	0.0005	Flex1 (AT)34
Flex1 (AT)34	*pol32*Δ	3	103.3	3.9633		
Flex1 (AT)34	*yen1*Δ	4	80.9	10.3907	0.0007	Flex1 (AT)34
Flex1 (AT)34	*slx1*Δ	3	29.3	0.7881	0.0750	Flex1 (AT)34
Flex1 (AT)34	*rad1*Δ	3	6.3	0.7937	<0.0001	Flex1 (AT)34
Flex1 (AT)34	*slx4*Δ	5	3.2	0.7736	<0.0001	Flex1 (AT)34
Flex1 (AT)34	*mus81*Δ *yen1*Δ	3	12.3	1.2583	0.0004	Flex1 (AT)34
Flex1 (AT)34	*mus81*Δ *slx1*Δ	3	10.4	2.5989	0.0003	Flex1 (AT)34
ctrl	*yen1*Δ	3	3.3	0.2517	0.7238	ctrl
ctrl	*slx1*Δ	3	2.9	0.4583	0.6659	ctrl
ctrl	*rad1*Δ	3	1.7	0.4978	0.0341	ctrl
ctrl	*slx4*Δ	3	1.0	0.1362	0.0016	ctrl
ctrl	*sae2*Δ	3	3.3	1.0817	0.8455	ctrl
Flex1 (AT)34	*sae2*Δ	4	7.6	1.2743	<0.0001	Flex1 (AT)34
Flex1 (AT)34 S3’ o2		4	11.6	2.4052	<0.0001	Flex1 (AT)34
Flex1 (AT)34 L3’ o2		9	2.1	0.5431	0.0002	Flex1 (AT)34 S3' o2
I-SceI only		6	344.5	81.6916		
I-SceI S3’		7	385.3	57.1993	0.6832	I-SceI only

I-SceI L3’		7	98.3	41.6164	0.0016	I-SceI S3'

Table 2-S4. Related to Figures 2-4C and 2-6C. YAC fragility assay data. All Flex1 constructs contain the S3' flanking sequence and are in orientation 1 unless otherwise noted.

FRA16D sequence	**Deleted gene(s)**	**# of Experiments**	**Average $FOA^{R}His^{+}$ x 10^{-6}**	**SEM**	**p value**	**p compared to**
Flex1 (AT)34 S3' o1		3	11.1	0.9207	0.0167	(AT)23-S3' o1
Flex1 (AT)34	*yen1*Δ	3	12.2	1.3043	0.5287	(AT)34-S3' o1
Flex1 (AT)34	*mus81*Δ	3	5.8	0.4177	0.0061	(AT)34-S3' o1
Flex1 (AT)34	*slx1*Δ	3	6.5	0.9244	0.0232	(AT)34-S3' o1
Flex1 (AT)34	*rad1*Δ	3	4.6	1.0366	0.0094	(AT)34-S3' o1
Flex1 (AT)34	*slx4*Δ	3	5.9	0.5859	0.0087	(AT)34-S3' o1
Flex1 (AT)34	*mus81*Δ *rad1*Δ	5	5.2	0.9528	0.0108	(AT)34-S3' o1
Flex1 (AT)34	*slx1*Δ *rad1*Δ	3	7.8	1.9150	0.2511	(AT)34-S3' o1
Flex1 (AT)23 S3'		3	6.4	0.7513		
Flex1 (AT)34 L3'		3	0.3	0.0876	0.0003	(AT)34-S3' o1
Flex1 (AT)34 S3' o2		3	15.4	1.3528	0.0044	(AT)23-S3' o1
Flex1 (AT)34		3	14.4	5.0560	0.8578	(AT)34-S3'

L3' o2						o2

Table 2-S5. Yeast strains.

Strain Name	**Background**	**CFY #**	**Genotype**	**Source**
972D3 YAC	AB1380	1087	*MATa, ura3-52, his5, trp1-289, lys2-1, can1-100, ade2-1* YAC*: LEU2 C_4A_4 URA3 TRP1*	(Albertsen et al., 1990; Zhang and Freudenreich, 2007)
801B6 YAC	AB1380	1086	YAC*: LEU2 C_4A_4 URA3 TRP1*	(Albertsen et al., 1990; Zhang and Freudenreich, 2007)
801B6 YAC Flex1Δ	CFY# 1086	3076, 3077	YAC*: LEU2 C_4A_4 URA3 TRP1* Flex1*::KANMX6*	this study
lys2::ADE2	YPH499	2268	*MATa, leu2-Δ1, ura3-52, his3-Δ200, trp1-Δ63, ade2Δ::hisG (salmonella), lys2::ADE2*	this study
ctrl	CFY# 2268	2863, 2864	*lys2::ADE2*::*URA3*-no repeat control	this study
Flex1 (AT)14	CFY# 2268	3917, 3921	*lys2::ADE2*::*URA3*-Flex1(AT)14	this study
Flex1 (AT)23	CFY# 2268	3445, 3473	*lys2::ADE2*::*URA3*-Flex1(AT)23	this study
Flex1 (AT)34	CFY# 2268	2525, 2712	*lys2::ADE2*::*URA3*-Flex1(AT)34	this study
Flex5 o1	CFY# 2268	3526, 3527	*lys2::ADE2*::*URA3*-Flex5 o1	this study
Flex5 o2	CFY# 2268	3528, 3529	*lys2::ADE2*::*URA3*-Flex5 o2	this study
ctrl *mus81*Δ	CFY# 2863	3375	*mus81::KANMX4*	this study

Flex1 (AT)14 *mus81*Δ	CFY# 3917	4326, 4327	*mus81::KANMX4*	this study
Flex1 (AT)23 *mus81*Δ	CFY# 3445	3799, 3800	*mus81::KANMX4*	this study
Flex1 (AT)34 *mus81*Δ	CFY# 2525	3377, 3378	*mus81::KANMX4*	this study
Flex1 (AT)34 *pol32*Δ	CFY# 2525	4349, 4350	*pol32::KANMX6*	this study
Flex1 (AT)34 *yen1*Δ	CFY# 2525	3987, 3988, 4063	*yen1::TRP1*	this study
Flex1 (AT)34 *slx1*Δ	CFY# 2525	4138, 4139	*slx1::KANMX4*	this study
Flex1 (AT)34 *rad1*Δ	CFY# 2525	4584,4585	*rad1::TRP1*	this study
Flex1 (AT)34 *slx4*Δ	CFY# 2525	4022, 4023	*slx4::KANMX6*	this study
Flex1 (AT)34 *mus81*Δ *yen1*Δ	CFY# 4063	4203, 4204	*mus81::KANMX4*, *yen1::TRP1*	this study
Flex1 (AT)34 *mus81*Δ *slx1*Δ	CFY# 4139	4238, 4239	*slx1::KANMX4*, *mus81::TRP1*	this study
ctrl *yen1*Δ	CFY# 2863	4125, 4126	*yen1::TRP1*	this study
ctrl *slx1*Δ	CFY# 2864	4340, 4341	*slx1::KANMX4*	this study
ctrl *rad1*Δ	CFY# 2863	4582, 4583	*rad1::TRP1*	this study
ctrl *slx4*Δ	CFY# 2863	4328, 4329	*slx4::KANMX6*	this study

ctrl *sae2*Δ	CFY# 2863	3607, 3608	*sae2::KANMX6*	this study
Flex1 (AT)34 *sae2*Δ	CFY# 3106	3520, 3521	*sae2::KANMX6*	this study
Flex1 (AT)34 S3' o2	CFY# 2268	3106, 3202-3204	*lys2::ADE2*::*URA3*-Flex1 (AT)34 S3' o2	this study
Flex1 (AT)34 L3' o2	CFY# 2268	2372- 2375	*lys2::ADE2*::*URA3*-Flex1 (AT)34 L3' o2	this study
no-I-SceI cut site	CFY# 2268	3518	*ILV1::pGAL*-I-SceI nuclease	this study
I-SceI only	CFY# 3518	4439, 4440, 4342	*lys2::ADE2*::*URA3*-I-SceI only	this study
S5' I-SceI S3'	CFY# 3518	3989, 4323	*lys2::ADE2*::*URA3*-S5'-I-SceI-S3'	this study
S5' I-SceI L3'	CFY# 3519	3991, 3992	*lys2::ADE2*::*URA3*-S5'-I-SceI-L3'	this study
WT strain with YAC CF1 (no Flex1)	BY4705	765	*MAT α, leu2Δ0, ura3Δ0, his3Δ200, trp1Δ63, ade2Δ::hisG, lys2Δ0, met15Δ0,* YAC CF1: *ade3-2p* ARS1 CEN4 *LEU2* $(G_4T_4)_{13}$ *URA3*	(Callahan et al., 2003)
Flex1 (AT)34 S3' o1 on YAC	CFY #765	3457, 3458	YAC*: LEU2* Flex1(AT)34 *HIS3 URA3* (this and all YACs in this study are modified from YAC CF1; only relevant markers and added sequence are listed)	this study
Flex1 (AT)34 on	CFY# 3457	4315, 4316	*yen1::TRP1*	this study

YAC *yen1*Δ				
Flex1 (AT)34 on YAC *mus81*Δ	CFY# 3458	4284, 4285	*mus81::KANMX4*	this study
Flex1 (AT)34 on YAC *slx1*Δ	CFY# 3458	4313, 4314	*slx1::KANMX4*	this study
Flex1 (AT)34 on YAC *rad1*Δ	CFY# 3458	4351, 4352	*rad1::KANXMX6*	this study
Flex1 (AT)34 on YAC *slx4*Δ	CFY#3457	4550, 4551	*slx4::KANMX4*	this study
Flex1 (AT)34 on YAC *mus81*Δ *rad1*Δ	CFY# 4284	4408, 4409	*rad1::TRP1* *mus81::KANMX4*	this study
Flex1 (AT)34 on YAC *slx1*Δ *rad1*Δ	CFY# 4313	4425, 4426	*rad1::TRP1* *slx1:: KANMX4*	this study
Flex1 (AT)23 S3' o1 on YAC	CFY #765	1239, 1240	YAC*: LEU2* Flex1(AT)34 *HIS3 URA3*	(Zhang and Freudenreich, 2007)
Flex1 (AT)34 L3' on YAC	CFY #765	1241, 1242	YAC*: LEU2* Flex1(AT)34 *HIS3 URA3*	(Zhang and Freudenreich, 2007)
Flex1 (AT)34 S3' o2 on YAC	CFY #765	3884, 3885	YAC*: LEU2* Flex1(AT)34 *HIS3 URA3*	this study
Flex1 (AT)34 L3' o2 on YAC	CFY #765	3455, 3456	YAC*: LEU2* Flex1(AT)34 *HIS3 URA3*	this study

Table 2-S6. Oligonucleotides.

Oligo Name	CF Oligo Stock #	Purpose	Sequence
TRP1_222b_int_for	1711	Southern TRP1 probe for	GGCGTGTTTC GTAATCAACC
TRP1_127bp_int_rev	1712	Southern TRP1 probe rev	GGCGTCAGTC CACCAGCTAA
P1_for_252 bp_chk	1807	FRA16D amplicon 1 for	GCATATGAGA ATACTCATACT CAG TGCTGC
P1_110bp_chk	1704	FRA16D amplicon 1 rev	CCATGCACTCT GGTGTACCA
P3_for_642 bp_chk	1840	FRA16D amplicon 2 for	GTGTGAATAC CAGGTGGTAG GGATTATGTG
P3_rev_120 bp_chk	1841	FRA16D amplicon 2 rev	ACAGAACTAA CCCAGAGATG GTTTCTCATC
F5His_For	1545	FRA16D amplicon 3 for	GGGAGTCCTA GATCAAGGTG
P4_rev_752 bp_chk	1809	FRA16D amplicon 3 rev	GAACTCAGAT AAAGATAAGG CCTATGGTTC
P5P5B_for_672bp_chk	1810	FRA16D amplicon 4 for	AAAACTTTGC TGGAGAACAT CACCAATCAC
P5P5B_rev_428bp_chk	1811	FRA16D amplicon 4 rev	TTCTGAGAAA CTGTCACAGC CAAGAAGATG
F1_420down	1267	Checking Flex1::KANMX6 in FRA16D YAC	GCTGAAGTCA CAAGATCTTA GGATGGGGTG

pBL007for	679	Screening for pBL007 transformants with insert	AAGCATATTT GAGAAGATGC GGCCAGC
pBL007rev	680	Screening for pBL007 transformants with insert	GGAATAAGGG CGACACGGAA ATGTTGA
Flex1_pBL007_seq_For	1032	PCR and sequencing of insert in pBL007 and chrII locus	ACTCACTATA GGGCGAATTG
Flex1_pBL007_seq_Rev	1033	PCR and sequencing of insert in pBL007 and chrII locus	CCAACTGATC TTCAGCATCT
5'LYS2_pBL007_integr_For	1028	PCR of 5' cassette in chrII locus	AAGTAACAAG CAGCCAATAG
5'LYS2_pBL007_integr_Rev	1029	PCR of 5' cassette in chrII locus	CATGTGTCAG AGGTTTTCAC
3'LYS2_pBL007_integr_For	1030	PCR of 3' cassette in chrII locus	CTCGGAATTA ACCCTCACTA
3'Lys2junctionrev	1047	PCR of 3' cassette in chrII locus	GCAAAGTGGT GATAGAGTTC
T7	2	PCR and sequencing of insert in pHZ-HIS3MX6 and YAC	TAATACGACT CACTATAGGG
M13R	1343	PCR and sequencing of insert in pHZ-HIS3MX6 and YAC	CAGGAAACAG CTATGACC
His3Revsk	375	PCR from *HIS3MX6* to *URA3* to confirm modified YAC	TTAGATAAAT CGACTACGGC AC

URA3 for	832	PCR from *HIS3MX6* to *URA3* to confirm modified YAC	CAGTACTCTG CGGGTGTATA CAG
ILV1_for	1465	PCR of 5' junction of pGAL-I-SceI nuclease cassette	CTCTGCGCTAT ATCTTTGGG
GAL1,10_chk	1466	PCR of 5' junction of pGAL-I-SceI nuclease cassette	CGCTTCGCTG ATTAATTACCC CAG
I-SceI_for2	1511	Creation of I-SceI insert for cloning (3' end anneals to 1512)	gatctaGAATTCg gtactgcgggatatcgt ccattccgacagTAG GGATAACAGG GTAAT
I-SceI_rev2	1512	Creation of I-SceI insert for cloning (3' end anneals to 1511)	tatcgaGAATTCa gcgcgacgtcgcttgc ggtattcggATTAC CCTGTTATCCC TActgt
I-SceI_for2_short	1513	Creation of I-SceI insert for cloning	gatctaGAATTCg gtactgc
I-SceI_rev2_short	1514	Creation of I-SceI insert for cloning	tatcgaGAATTCa gcgcgac

Table 2-S7. Plasmids.

Plasmid	**CF Plasmid stock#**	**Description**	**Source**
pFA6a-KANMX6	136	Template for one-step gene replacement by PCR	(Wach et al., 1994)
pBL007	223	*ADE2* nt 512-1480 *URA3*	this study

pBL007+ctrl	387/388	*ADE2* nt 512-1480 *URA3*-EcoRI-ctrl-BamHI	this study
pBL007+S5'-(AT)14-S3' o1	565/566	*ADE2* nt 512-1480 *URA3*-EcoRI-Flex1(AT)14-EcoRI	this study
pBL007+S5'-(AT)23-S3' o1	516/517	*ADE2* nt 512-1480 *URA3*-EcoRI-Flex1(AT)23-EcoRI	this study
pBL007+S5'-(AT)34-S3' o1	351	*ADE2* nt 512-1480 *URA3*-EcoRI-Flex1(AT)34-EcoRI	this study
pHZ-*HIS3MX6*	466	G_4T_4 *HIS3MX6* *URA3*	this study
pHZ-*HIS3MX6*+S5'-(AT)34-S3' o1	513	G_4T_4 *HIS3MX6* *URA3* EcoRI-Flex1(AT)34-S3' o1-EcoRI	this study
pHZ-*HIS3MX6*+S5'-(AT)34-S3' o2	559, 560	G_4T_4 *HIS3MX6* *URA3* EcorI-Flex1(AT)34-S3' o2-EcoRI	this study
pHZ-*HIS3MX6*+S5'-(AT)34-L3' o2	512	G_4T_4 *HIS3MX6* *URA3* EcoRI-Flex1(AT)34-L3' o1-EcoRI	this study
pBL007+I-SceI	519	*ADE2* nt 512-1480 *URA3*-EcoRI-I-SceI-EcoRI	this study
pBL007+S5'-I-SceI-S3'	571	*ADE2* nt 512-1480 *URA3*-EcoRI-Flex1 S5'-I-SceI-S3'-	this study

		EcoRI	
pBL007+S5'-I-SceI-L3'	581	*ADE2* nt 512-1480 *URA3*-EcoRI-Flex1 S5'-I-SceI-L3'-EcoRI	this study
pGSHU	524	pFA6a-pGAL1-I-SceI-HYG-klURA3	(Storici et al., 2003)

Table 2-S8. I-SceI cloning gBlocks.

gBlock Name	**Sequence**
EcoRI-S5-I-SceI-S3-EcoRI	AGCGTAGAATTCTGTTACCATGAGTGGTGATGGATGTG TTAATTAATTCGATTGTGATAATCATTACACAATGTAT ATAGTAATCAAATCATTACTTTATAGACCCTGAATATA TTCAATATTTATTTTTCAATTTAGGGATAACAGGGTAA TTTAAAGCTGTCATGGAAAGCCTTAAAGCAGTATGAAT TCTCTGAC
EcoRI-S5-IsceI-L3-EcoRI	AGCGTAGAATTCTGTTACCATGAGTGGTGATGGATGTG TTAATTAATTCGATTGTGATAATCATTACACAATGTAT ATAGTAATCAAATCATTACTTTATAGACCCTGAATATA TTCAATATTTATTTTTCAATTTAGGGATAACAGGGTAA TTTAAAGCTGTCATGGAAAGCCTTAAAGTTAAAATACG AAGATTTTTGAGAAAAACTTTGCATATTTTAATTGCTG TCTGGAATCCTCCTTCAGCTGGGATGAGAAATCATCTC TGGGTTAGTTCTGTCCCAGTATGAATTCTCTGAC

Table 2-S9. Chemicals, peptides, and recombinant proteins needed for polymerase pausing assay.

Chemical, peptide, or recombinant protein	**SOURCE**	**IDENTIFIER**
Polyethylene Glycol (Avg Mol Wt of 8000)	Sigma	Cat# P5413; CAS 25322-68-3
Equilibrated Phenol, pH 8.0, Ultrapure	Affymetrix/Thermo-Fisher	Cat# AAJ75829AN; CAS 108-95-2
γ^{32}P ATP (6000Ci/mmol)	Perkin-Elmer	Cat# BLU002Z001MC
T4 Polynucleotide Kinase	Thermo-Fisher	Cat# 18004010
Recombinant human PCNA	Laboratory of Marietta Lee	Biochemistry 2001 40: 4512-4520
Recombinant yeast RFC	Laboratory of Linda Bloom	*The Journal of Biological Chemistry 2012* 287: 2203-9
Sequenase 2.0	Affymetrix/Thermo-Fisher	*Cat# 70775Y*
Exo- Klenow Polymerase	Affymetrix/Thermo-Fisher	*Cat# 70057Z*

Table 2-S10. Key Resources Table.

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Bacterial and Virus Strains		
R408 Helper Phage	Promega	Cat# P2291
SURE 2 Supercompetent Cells	Agilent Technologies	Cat# 200152

Biological Samples		
Chemicals, Peptides, and Recombinant Proteins		
EcoRI-HF	NEB	Cat# R3101S
BamHI-HF	NEB	Cat# R3136S
AhdI	NEB	Cat# R0584S
XbaI	NEB	Cat# R0145S
Chemicals, etc., for polymerase pausing assay, see Table S9.		
Critical Commercial Assays		
Deposited Data		
Experimental Models: Cell Lines		
Experimental Models: Organisms/Strains		
S. cerevisiae derivatives, see Table S2.	This study	N/A
Oligonucleotides		
Oligonucleotides, see Table S3.	This study	N/A
I-SceI cloning gBlocks, see Table S5.	Integrated DNA Technologies	N/A
G40-16mer, PAGE-Purified 5'-GCA TGC CTG CAG GTC G -3'	Integrated DNA Technologies	N/A

Recombinant DNA		
Plasmids, see Table S4.	This study	N/A
pGEM-3Zf(-) Vector	Promega	Cat# P2261
Software and Algorithms		
FALCOR	Hall et al., 2009	http://www.keshavsingh.org/protocols/FALCOR.html
ImageQuant version 5.2	GE Healthcare	N/A
Other		
Illustra Microspin G-50 column	GE Healthcare	Cat# 27-5330-1

Contact for reagent and resource sharing

Further information and requests for resources and reagents should be directed to corresponding author, Catherine H. Freudenreich (Catherine.freudenreich@tufts.edu).

Methods

Experimental model and subject details

Yeast strains, oligonucleotides, and plasmids used in this study are listed in Tables 2-S5, 2-S6, and 2-S7, respectively. All yeast strains were grown at 30^{o} C and all bacterial strains were grown at 37^{o} C.

Large FRA16D YAC strains were as used as previously described (Zhang and Freudenreich, 2007). Overall YAC length was confirmed using pulsed field gel

electrophoresis followed by a Southern blot using a TRP1 probe (Figure 2-S1B). Flex1 was replaced in FRA16D with the *KANMX* marker, which was confirmed by PCR. Intact YAC structure was also verified using PCR of subregions across FRA16D (Figure 2-S1A and Table 2-S1).

Chromosome II Flex1 strains were created by modifying the pBL007 plasmid, which has a *URA3* marker and nucleotides 512-1480 of *ADE2* (designated DE in diagrams). The FRA16D subregions of interest were inserted into the EcoRI only or BamHI and EcoRI sites in the MCS of pBL007. Orientation was confirmed by PCR and sequencing. Plasmids were digested with XbaI to linearize them for transformation into *lys2::ADE2* yeast strains, replacing *ADE2* with the ADE-*URA3*-Flex1-DE2 cassette. All chromosome II yeast strains were checked by PCR of the pBL007 cassette junctions and sequencing to confirm correct sequence and orientation.

The Flex1 subregion YACs (AT)23-S3' and (AT)34-L3' in orientation 1 were created previously (Zhang and Freudenreich, 2007). Flex1 (AT)34-S3' in o1 and o2 and Flex1 (AT)34-L3' o2 YAC strains were made by modifying the pHZ-*HIS3MX6* plasmid. The Flex1 subregion of interest was inserted by EcoRI-based subcloning into the MCS of pHZ-*HIS3MX6*. Correct Flex1 sequence insertion in the right orientation was confirmed by PCR and sequencing. Plasmids were digested with AhdI to linearize them for transformation into CFY #765 BY4705 yeast strains containing *URA3* marked YAC CF1 (Callahan et al., 2003) and selecting for His+ transformants. Correct structure of the Flex1 YACs was

confirmed by PCR of the pHZ-*HIS3MX6* cassette junctions (primers 375 and 832 in Table 2-S6) and sequencing to confirm Flex1 sequence and orientation.

Chromosome II I-SceI strains were created by modifying the pBL007 plasmid. The I-SceI only insert was created by PCR with primers 1511 and 1512, whose 3' ends anneal to one another at the I-SceI recognition sequence; that PCR product was then used as a template for PCR with primers 1513 and 1514 to complete generation of the insert. S5'-I-SceI-S3' and S5'-I-SceI-L3' inserts were synthesized as gBlocks (Table 2-S8) (Integrated DNA Technologies, Coralville, Iowa) flanked by EcoRI restriction sites and contained S5'and S3' from Flex1 flanking sequences and an I-SceI restriction site, or S5' and L3' from Flex1 flanking sequences and an I-SceI restriction site. The inserts were cloned into the EcoRI site of pBL007. Correct I-SceI recognition sequence insertion into the plasmid was confirmed by PCR and sequencing. A yeast strain with a galactose-inducible I-SceI nuclease was created by transformation of a PCR product from the pGSHU plasmid (Storici et al., 2003) into the *ILV1* locus in a *lys2::ADE2* strain (CF stock #2268). Insertion of the galactose-inducible I-SceI nuclease was confirmed by hygromycin resistance and PCR of the 5' junction of the cassette. XbaI-linearized pBL007+I-SceI DNA was transformed into the *lys2::ADE2* strain with the galactose-inducible I-SceI nuclease. Yeast strains were confirmed by PCR and sequencing as stated above.

All gene deletion mutants were created using one-step gene replacement. Primers with homology to regions directly upstream and downstream of ORF for gene replacement were used to amplify gene replacement fragments from either the

pFA plasmid series or yeast genomic DNA of a previously made gene replacement strain. Proper gene replacement was confirmed by PCR using primer sets: (1) that hybridize to the marker gene and a genomic region outside of the gene to be replaced and (2) are located within the open reading frame (ORF) to be replaced to confirm ORF absence. Sequences of primers used are available upon request.

Method details

Large FRA16D YAC Breakage Assay

Large FRA16D YAC strains with confirmed YAC structure were patched onto YC-Ura-Leu-Trp plates and then plated for single colonies on YC-Ura-Leu-Trp and grown for 2 days at 30^{o} C. A portion of 10 single colonies was used to inoculate ten 1 mL YC-Leu cultures at 0.02-0.04 OD which were grown at 30^{o} C for 6-7 divisions (~16 hours). 100 uL of a 10^{-4} dilution of each culture was plated on FOA-Leu to query for cells that had lost *URA3* gene function, potentially by breakage within FRA16D and YAC end loss. 100 uL from each culture were combined, diluted to 10^{-4}, and plated on YC-Leu media to obtain a total cell count. Plates were grown for 3 days at 30^{o} C. Breakage frequency was calculated.

DDRA Fragility Assay

DDRA fragility assay strains were patched onto YC-Ura to maintain selection for the *ADE2* recombination assay cassette in the starting strains. Cells from a YC-Ura patch were plated for single colonies on YEPD non-selective media for 3 days at 30^{o} C to allow breakage to occur. Individual colonies were resuspended in

400 uL diH_2O, diluted as appropriate (varies by strain and mutant), and plated on FOA-Ade media to select for cells that have undergone breakage and recombination of the chromosome II cassette. 100 uL from each colony suspension were combined, diluted to either 10^{-4} or 10^{-5}, and plated on YEPD media to obtain a total cell count. A rate of FOA^R Ade^+ was calculated using the method of the median (Lea and Coulson, 1949) using the FALCOR online calculator (Hall et al., 2009).

I-SceI DDRA fragility assays were performed in the same manner, except all media was supplemented with 10x isoleucine, 10x leucine, and 10x valine to compensate for disruption of the *ILV* locus. YEP plates were made with 1.5% galactose and 0.5% glucose to induce ~50% cutting of I-SceI. For hydroxyurea DDRA fragility assays, the YEPD plates were supplemented with 100 mM HU.

***In vitro* polymerase δ pausing assay**

Templates for polymerase reactions were created by cloning the 315 bp Flex1 sequence (S5' AT34 L3') into the MCS/BamH1 site of the pGEM3Zf(-) vector (Promega, P2261). Inserts in two orientations were isolated in order to purify ssDNA templates of both strands. For each construct, single-stranded DNA was isolated after R408 helper phage (Promega, P2291) infection of plasmid-bearing SURE cells (e14-(McrA-), Δ(mcrCB-hsdSMR-mrr)171, endA1, gyrA96, thi-1, supE44, relA1, lac, recB, recJ, sbcC, umuC::Tn5 (Kanr) uvrC [F' proAB lacIqZΔM15 Tn10 (Tetr) Amy Camr]; Agilent Technologies, 200152). Log phase plasmid-bearing SURE cells in 2XYT media were infected with 1/50th volume of R408 (titer of phage stock was >1 x 10^{11} plaque forming units (pfu)/mL) and

incubated in a 37°C shaker for 7 hours. After pelleting the bacterial cells, virus particles in the supernatant were precipitated on ice for 30 min with a polyethylene glycol (Sigma, P5413)/ammonium acetate solution at final concentrations of 4 % and 0.75 M, respectively. Virus was pelleted and resuspended in an appropriate volume of Phenol Extraction Buffer (PEB; 100 mM Tris, pH 8.0, 300 mM NaCl, 1 mM EDTA, pH 8.0). DNA was extracted one time with two volumes of phenol (Affymetrix/Thermo-Fisher, AAJ75829AN) saturated with PEB, one time with one volume of phenol, and one time with half volume 24:1 chloroform: isoamyl alcohol. After extraction, DNA was precipitated with ammonium acetate at 2.0 M final concentration and 2 volumes of ethanol and resuspended in 10 mM Tris and 1 mM EDTA, pH 8.0. Small ssDNA preparations from independent clones were sequenced (dideoxy sequencing) to verify integrity of the insert prior to large scale purification of ssDNA templates.

DNA synthesis templates were created by 32P end-labeling (γ32P ATP (6000Ci/mmol); Perkin-Elmer, BLU002Z001MC) a PAGE-purified 16mer oligonucleotide (G40-16mer, Integrated DNA Technologies) using T4 Polynucleotide Kinase (Thermo-Fisher, 18004010) according to the manufacturer's instructions and hybridizing to ssDNA at a 1:1 molar ratio in 1X SSC buffer (150 mM NaCl and 15 mM sodium citrate). The G40 oligonucleotide initiates synthesis 14 nucleotides downstream of the Flex1 insert. To remove unincorporated radionucleotide, the hybridized primer-templates were purified over illustra Microspin G-50 columns (GE Healthcare, 27-5330-01). Primer extension reactions contained 100 fmol of primed ssDNA substrate, 400 fmol

human recombinant PCNA (Xu et al., 2001), 1700 fmol yeast RFC (Thompson et al., 2012), 20 mM Tris HCl, pH 7.5, 8 mM MgCl2, 5 mM DTT, 40 µg/ml BSA, 150 mM KCl, 5% glycerol, 0.5 mM ATP, and 250 uM dNTPS, and were preincubated at 37°C for 3 min. Synthesis was initiated upon addition of the indicated fmol purified 4-subunit recombinant human Pol δ4 (Zhou et al., 2012). Negative controls were performed as described, with the omission of RF-C. Aliquots were removed at 3, 7, and 15 minutes, quenched in 1 volume STOP dye (Formamide, 5 mM EDTA pH 8.0, 0.1% xylene cylanol, 0.1% bromophenol blue) and reaction products were separated on an 8% denaturing polyacrylamide gel and quantitated using a Molecular Dynamics STORM 860 Phosphoimager. A control for the percent of primers productively hybridized to each primer-template substrate (% Hyb) was performed using excess Exo- Klenow polymerase (Affymetrix/Thermo-Fisher, 70057Z), and a background control for primer impurities (no Pol) was performed by incubating unextended primer-template substrate in reaction buffer without addition of polymerase. Dideoxy sequencing reactions were carried out simultaneously with the Pol δHE reactions, using the same primer-template substrates and Sequenase 2.0 (Affymetrix/Thermo-Fisher, 70775Y).

Flex1 subregion YAC end loss fragility assay

Fragility assays were performed on the YACs as previously described (Zhang and Freudenreich, 2007). Cells were plated onto YC-Leu-Ura plates in order to select for both arms of the YAC. Ten 1 mL YC-Leu liquid cultures of 0.02-0.04 starting OD_{600} were inoculated from YC-Leu-Ura patches and grown overnight at 30^{o} C

for 6-7 divisions (~16 hours for wildtype strains; longer for some mutants). A portion of each culture (100 uL for WT strains; less for strains with high fragility rates) was plated on FOA-Leu to query for cells that had lost *URA3* gene function, potentially by breakage within Flex1 and YAC end loss. Plates were grown for 5 days at 30^{o} C. Total cell counts were obtained by combining 100 uL from each YC-Leu overnight culture and plating 10^{-4} and 10^{-5} dilutions on YC-Leu. FOA-Leu plates were replica plated onto YC-His; any colonies growing on YC-His did not lose the right arm of the YAC and were removed from colony counts. A rate of FOA^{R} His^{-} was calculated using the method of the median using the Fluctuation Analysis Calculator (FALCOR).

Pulsed Field Gel Electrophoresis

Large FRA16D YAC length was verified using CHEF gels (Bio-Rad) and Southern blot hybridization. Cells were grown to early log phase in YC-Leu-Ura-Trp media and whole chromosomal DNA was isolated in 0.8% agarose plugs (Bio-Rad Clean Cut agarose). Plugs were run on a 1.2% gel, 5V/cm, 60-120 switch, for 48 hours. The Southern blot was performed using a *TRP1* probe to the YAC (see Figure 1C for relative TRP1 location on the YAC).

Quantification and Statistical Analysis

DDRA fragility and YAC end loss assays were all a minimum of 3 assays, usually from 2 independently created strains. Strains were tested for significant deviation from the appropriate control using a t-test. Average rates are graphed with error bars indicating the standard error of the mean (see Tables 2-S2, 2-S3 and 2-S4).

References

ANAND, R., RANJHA, L., CANNAVO, E. & CEJKA, P. 2016. Phosphorylated CtIP Functions as a Co-factor of the MRE11-RAD50-NBS1 Endonuclease in DNA End Resection. *Mol Cell,* 64**,** 940-950.

ANAND, R. P., LOVETT, S. T. & HABER, J. E. 2013. Break-induced DNA replication. *Cold Spring Harb Perspect Biol,* 5**,** a010397.

BABCOCK, M., YATSENKO, S., STANKIEWICZ, P., LUPSKI, J. R. & MORROW, B. E. 2007. AT-rich repeats associated with chromosome 22q11.2 rearrangement disorders shape human genome architecture on Yq12. *Genome Res,* 17**,** 451-60.

BACOLLA, A., TAINER, J. A., VASQUEZ, K. M. & COOPER, D. N. 2016. Translocation and deletion breakpoints in cancer genomes are associated with potential non-B DNA-forming sequences. *Nucleic Acids Res,* 44**,** 5673-88.

BARNES, R. P., HILE, S. E., LEE, M. Y. & ECKERT, K. A. 2017. DNA polymerases eta and kappa exchange with the polymerase delta holoenzyme to complete common fragile site synthesis. *DNA Repair (Amst),* 57**,** 1-11.

BHOWMICK, R., MINOCHERHOMJI, S. & HICKSON, I. D. 2016. RAD52 Facilitates Mitotic DNA Synthesis Following Replication Stress. *Mol Cell,* 64**,** 1117-1126.

BIGNELL, G. R., GREENMAN, C. D., DAVIES, H., BUTLER, A. P., EDKINS, S., ANDREWS, J. M., BUCK, G., CHEN, L., BEARE, D., LATIMER, C., WIDAA, S., HINTON, J., FAHEY, C., FU, B., SWAMY, S., DALGLIESH, G. L., TEH, B. T., DELOUKAS, P., YANG, F., CAMPBELL, P. J., FUTREAL, P. A. & STRATTON, M. R. 2010. Signatures of mutation and selection in the cancer genome. *Nature,* 463**,** 893-8.

BLANCO, M. G. & MATOS, J. 2015. Hold your horSSEs: controlling structure-selective endonucleases MUS81 and Yen1/GEN1. *Front Genet,* 6**,** 253.

BOWATER, R., ABOUL-ELA, F. & LILLEY, D. M. 1991. Large-scale stable opening of supercoiled DNA in response to temperature and supercoiling in (A + T)-rich regions that promote low-salt cruciform extrusion. *Biochemistry,* 30**,** 11495-506.

BURGERS, P. M. & GERIK, K. J. 1998. Structure and processivity of two forms of Saccharomyces cerevisiae DNA polymerase delta. *J Biol Chem,* 273**,** 19756-62.

CALLAHAN, J. L., ANDREWS, K. J., ZAKIAN, V. A. & FREUDENREICH, C. H. 2003. Mutations in yeast replication proteins that increase CAG/CTG expansions also increase repeat fragility. *Mol Cell Biol,* 23**,** 7849-60.

CEJKA, P. 2015. DNA End Resection: Nucleases Team Up with the Right Partners to Initiate Homologous Recombination. *J Biol Chem,* 290**,** 22931-8.

COQUELLE, A., PIPIRAS, E., TOLEDO, F., BUTTIN, G. & DEBATISSE, M. 1997. Expression of fragile sites triggers intrachromosomal mammalian gene amplification and sets boundaries to early amplicons. *Cell,* 89**,** 215-25.

COTE, A. G. & LEWIS, S. M. 2008. Mus81-dependent double-strand DNA breaks at in vivo-generated cruciform structures in S. cerevisiae. *Mol Cell,* 31**,** 800-12.

CUSSIOL, J. R., DIBITETTO, D., PELLICIOLI, A. & SMOLKA, M. B. 2017. Slx4 scaffolding in homologous recombination and checkpoint control: lessons from yeast. *Chromosoma,* 126**,** 45-58.

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. *J Bacteriol,* 173**,** 2658-64.

DEHE, P. M. & GAILLARD, P. H. 2017. Control of structure-specific endonucleases to maintain genome stability. *Nat Rev Mol Cell Biol,* 18**,** 315-330.
DENG, S. K., YIN, Y., PETES, T. D. & SYMINGTON, L. S. 2015. Mre11-Sae2 and RPA Collaborate to Prevent Palindromic Gene Amplification. *Mol Cell,* 60**,** 500-8.
DILLON, L. W., PIERCE, L. C., NG, M. C. & WANG, Y. H. 2013. Role of DNA secondary structures in fragile site breakage along human chromosome 10. *Hum Mol Genet,* 22**,** 1443-56.
DURKIN, S. G., RAGLAND, R. L., ARLT, M. F., MULLE, J. G., WARREN, S. T. & GLOVER, T. W. 2008. Replication stress induces tumor-like microdeletions in FHIT/FRA3B. *Proc Natl Acad Sci U S A,* 105**,** 246-51.
FINNIS, M., DAYAN, S., HOBSON, L., CHENEVIX-TRENCH, G., FRIEND, K., RIED, K., VENTER, D., WOOLLATT, E., BAKER, E. & RICHARDS, R. I. 2005. Common chromosomal fragile site FRA16D mutation in cancer cells. *Hum Mol Genet,* 14**,** 1341-9.
FRAGKOS, M. & NAIM, V. 2017. Rescue from replication stress during mitosis. *Cell Cycle,* 16**,** 613-633.
FREUDENREICH, C. H., KANTROW, S. M. & ZAKIAN, V. A. 1998. Expansion and length-dependent fragility of CTG repeats in yeast. *Science,* 279**,** 853-6.
FREUDENREICH, C. H., STAVENHAGEN, J. B. & ZAKIAN, V. A. 1997. Stability of a CTG/CAG trinucleotide repeat in yeast is dependent on its orientation in the genome. *Mol Cell Biol,* 17**,** 2090-8.
FRICKE, W. M. & BRILL, S. J. 2003. Slx1-Slx4 is a second structure-specific endonuclease functionally redundant with Sgs1-Top3. *Genes Dev,* 17**,** 1768-78.
FUNGTAMMASAN, A., WALSH, E., CHIAROMONTE, F., ECKERT, K. A. & MAKOVA, K. D. 2012. A genome-wide analysis of common fragile sites: what features determine chromosomal instability in the human genome? *Genome Res,* 22**,** 993-1005.
GLOVER, T. W., WILSON, T. E. & ARLT, M. F. 2017. Fragile sites in cancer: more than meets the eye. *Nat Rev Cancer,* 17**,** 489-501.
GUERVILLY, J. H., TAKEDACHI, A., NAIM, V., SCAGLIONE, S., CHAWHAN, C., LOVERA, Y., DESPRAS, E., KURAOKA, I., KANNOUCHE, P., ROSSELLI, F. & GAILLARD, P. H. 2015. The SLX4 complex is a SUMO E3 ligase that impacts on replication stress outcome and genome stability. *Mol Cell,* 57**,** 123-37.
HALAZONETIS, T. D., GORGOULIS, V. G. & BARTEK, J. 2008. An oncogene-induced DNA damage model for cancer development. *Science,* 319**,** 1352-5.
HALL, B. M., MA, C. X., LIANG, P. & SINGH, K. K. 2009. Fluctuation analysis CalculatOR: a web tool for the determination of mutation rate using Luria-Delbruck fluctuation analysis. *Bioinformatics,* 25**,** 1564-5.
HELMRICH, A., BALLARINO, M. & TORA, L. 2011. Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. *Mol Cell,* 44**,** 966-77.
INAGAKI, H., OHYE, T., KOGO, H., TSUTSUMI, M., KATO, T., TONG, M., EMANUEL, B. S. & KURAHASHI, H. 2013. Two sequential cleavage reactions on cruciform DNA structures cause palindrome-mediated chromosomal translocations. *Nat Commun,* 4**,** 1592.
KATO, T., FRANCONI, C. P., SHERIDAN, M. B., HACKER, A. M., INAGAKAI, H., GLOVER, T. W., ARLT, M. F., DRABKIN, H. A., GEMMILL, R. M., KURAHASHI, H. & EMANUEL, B. S. 2014. Analysis of the t(3;8) of hereditary renal cell carcinoma: a palindrome-mediated translocation. *Cancer Genet,* 207**,** 133-40.

LE BEAU, M. M., RASSOOL, F. V., NEILLY, M. E., ESPINOSA, R., 3RD, GLOVER, T. W., SMITH, D. I. & MCKEITHAN, T. W. 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**,** 755-61.

LE TALLEC, B., MILLOT, G. A., BLIN, M. E., BRISON, O., DUTRILLAUX, B. & DEBATISSE, M. 2013. Common fragile site profiling in epithelial and erythroid cells reveals that most recurrent cancer deletions lie in fragile sites hosting large genes. *Cell Rep,* 4**,** 420-8.

LEA, D. E. & COULSON, C. A. 1949. The distribution of the numbers of mutants in bacterial populations. *J Genet,* 49**,** 264-85.

LETESSIER, A., MILLOT, G. A., KOUNDRIOUKOFF, S., LACHAGES, A. M., VOGT, N., HANSEN, R. S., MALFOY, B., BRISON, O. & DEBATISSE, M. 2011. Cell-type-specific replication initiation programs set fragility of the FRA3B fragile site. *Nature,* 470**,** 120-3.

LU, S., WANG, G., BACOLLA, A., ZHAO, J., SPITSER, S. & VASQUEZ, K. M. 2015. Short Inverted Repeats Are Hotspots for Genetic Instability: Relevance to Cancer Genomes. *Cell Rep*.

LYDEARD, J. R., JAIN, S., YAMAGUCHI, M. & HABER, J. E. 2007. Break-induced replication and telomerase-independent telomere maintenance require Pol32. *Nature,* 448**,** 820-3.

MACHERET, M. & HALAZONETIS, T. D. 2015. DNA replication stress as a hallmark of cancer. *Annu Rev Pathol,* 10**,** 425-48.

MADIREDDY, A., KOSIYATRAKUL, S. T., BOISVERT, R. A., HERRERA-MOYANO, E., GARCIA-RUBIO, M. L., GERHARDT, J., VUONO, E. A., OWEN, N., YAN, Z., OLSON, S., AGUILERA, A., HOWLETT, N. G. & SCHILDKRAUT, C. L. 2016. FANCD2 Facilitates Replication through Common Fragile Sites. *Mol Cell,* 64**,** 388-404.

MAKAROVA, A. V., STODOLA, J. L. & BURGERS, P. M. 2012. A four-subunit DNA polymerase zeta complex containing Pol delta accessory subunits is essential for PCNA-mediated mutagenesis. *Nucleic Acids Res,* 40**,** 11618-26.

MAKHARASHVILI, N., TUBBS, A. T., YANG, S. H., WANG, H., BARTON, O., ZHOU, Y., DESHPANDE, R. A., LEE, J. H., LOBRICH, M., SLECKMAN, B. P., WU, X. & PAULL, T. T. 2014. Catalytic and noncatalytic roles of the CtIP endonuclease in double-strand break end resection. *Mol Cell,* 54**,** 1022-33.

MCCLELLAN, J. A., BOUBLIKOVA, P., PALECEK, E. & LILLEY, D. M. 1990. Superhelical torsion in cellular DNA responds directly to environmental and genetic factors. *Proc Natl Acad Sci U S A,* 87**,** 8373-7.

MIMITOU, E. P. & SYMINGTON, L. S. 2009. DNA end resection: many nucleases make light work. *DNA Repair (Amst),* 8**,** 983-95.

MINOCHERHOMJI, S. & HICKSON, I. D. 2014. Structure-specific endonucleases: guardians of fragile site stability. *Trends Cell Biol,* 24**,** 321-7.

MINOCHERHOMJI, S., YING, S., BJERREGAARD, V. A., BURSOMANNO, S., ALELIUNAITE, A., WU, W., MANKOURI, H. W., SHEN, H., LIU, Y. & HICKSON, I. D. 2015. Replication stress activates DNA repair synthesis in mitosis. *Nature,* 528**,** 286-90.

MIRON, K., GOLAN-LEV, T., DVIR, R., BEN-DAVID, E. & KEREM, B. 2015. Oncogenes create a unique landscape of fragile sites. *Nat Commun,* 6**,** 7094.

MISHMAR, D., RAHAT, A., SCHERER, S. W., NYAKATURA, G., HINZMANN, B., KOHWI, Y., MANDEL-GUTFROIND, Y., LEE, J. R., DRESCHER, B., SAS, D. E., MARGALIT, H.,

PLATZER, M., WEISS, A., TSUI, L. C., ROSENTHAL, A. & KEREM, B. 1998. Molecular characterization of a common fragile site (FRA7H) on human chromosome 7 by the cloning of a simian virus 40 integration site. *Proc Natl Acad Sci U S A,* 95**,** 8141-6.

MOSBACH, V., POGGI, L., VITERBO, D., CHARPENTIER, M. & RICHARD, G. F. 2018. TALEN-Induced Double-Strand Break Repair of CTG Trinucleotide Repeats. *Cell Rep,* 22**,** 2146-2159.

NAIM, V., WILHELM, T., DEBATISSE, M. & ROSSELLI, F. 2013. ERCC1 and MUS81-EME1 promote sister chromatid separation by processing late replication intermediates at common fragile sites during mitosis. *Nat Cell Biol,* 15**,** 1008-15.

OUYANG, J., GARNER, E., HALLET, A., NGUYEN, H. D., RICKMAN, K. A., GILL, G., SMOGORZEWSKA, A. & ZOU, L. 2015. Noncovalent interactions with SUMO and ubiquitin orchestrate distinct functions of the SLX4 complex in genome maintenance. *Mol Cell,* 57**,** 108-22.

PAESCHKE, K., CAPRA, J. A. & ZAKIAN, V. A. 2011. DNA replication through G-quadruplex motifs is promoted by the Saccharomyces cerevisiae Pif1 DNA helicase. *Cell,* 145**,** 678-91.

PALAKODETI, A., HAN, Y., JIANG, Y. & LE BEAU, M. M. 2004. The role of late/slow replication of the FRA16D in common fragile site induction. *Genes Chromosomes Cancer,* 39**,** 71-6.

POLLEYS, E. J. & FREUDENREICH, C. H. 2018. Methods to Study Repeat Fragility and Instability in Saccharomyces cerevisiae. *Methods Mol Biol,* 1672**,** 403-419.

RASS, U. 2013. Resolving branched DNA intermediates with structure-specific nucleases during replication in eukaryotes. *Chromosoma,* 122**,** 499-515.

RIED, K., FINNIS, M., HOBSON, L., MANGELSDORF, M., DAYAN, S., NANCARROW, J. K., WOOLLATT, E., KREMMIDIOTIS, G., GARDNER, A., VENTER, D., BAKER, E. & RICHARDS, R. I. 2000. Common chromosomal fragile site FRA16D sequence: identification of the FOR gene spanning FRA16D and homozygous deletions and translocation breakpoints in cancer cells. *Hum Mol Genet,* 9**,** 1651-63.

SARANGI, P., ALTMANNOVA, V., HOLLAND, C., BARTOSOVA, Z., HAO, F., ANRATHER, D., AMMERER, G., LEE, S. E., KREJCI, L. & ZHAO, X. 2014. A versatile scaffold contributes to damage survival via sumoylation and nuclease interactions. *Cell Rep,* 9**,** 143-152.

SARBAJNA, S., DAVIES, D. & WEST, S. C. 2014. Roles of SLX1-SLX4, MUS81-EME1, and GEN1 in avoiding genome instability and mitotic catastrophe. *Genes Dev,* 28**,** 1124-36.

SHAH, S. N., OPRESKO, P. L., MENG, X., LEE, M. Y. & ECKERT, K. A. 2010. DNA structure and the Werner protein modulate human DNA polymerase delta-dependent replication dynamics within the common fragile site FRA16D. *Nucleic Acids Res,* 38**,** 1149-62.

STORICI, F., DURHAM, C. L., GORDENIN, D. A. & RESNICK, M. A. 2003. Chromosomal site-specific double-strand breaks are efficiently targeted for repair by oligonucleotides in yeast. *Proc Natl Acad Sci U S A,* 100**,** 14994-9.

SVENDSEN, J. M., SMOGORZEWSKA, A., SOWA, M. E., O'CONNELL, B. C., GYGI, S. P., ELLEDGE, S. J. & HARPER, J. W. 2009. Mammalian BTBD12/SLX4 assembles a Holliday junction resolvase and is required for DNA repair. *Cell,* 138**,** 63-77.

SYMINGTON, L. S., ROTHSTEIN, R. & LISBY, M. 2014. Mechanisms and regulation of mitotic recombination in Saccharomyces cerevisiae. *Genetics,* 198**,** 795-835.

TAYLOR, E. M. & LINDSAY, H. D. 2016. DNA replication stress and cancer: cause or cure? *Future Oncol,* 12**,** 221-37.

THOMPSON, J. A., MARZAHN, M. R., O'DONNELL, M. & BLOOM, L. B. 2012. Replication factor C is a more effective proliferating cell nuclear antigen (PCNA) opener than the checkpoint clamp loader, Rad24-RFC. *J Biol Chem,* 287**,** 2203-9.

THYS, R. G., LEHMAN, C. E., PIERCE, L. C. & WANG, Y. H. 2015. DNA secondary structure at chromosomal fragile sites in human disease. *Curr Genomics,* 16**,** 60-70.

WALSH, E., WANG, X., LEE, M. Y. & ECKERT, K. A. 2013. Mechanism of replicative DNA polymerase delta pausing and a potential role for DNA polymerase kappa in common fragile site replication. *J Mol Biol,* 425**,** 232-43.

WANG, H., LI, Y., TRUONG, L. N., SHI, L. Z., HWANG, P. Y., HE, J., DO, J., CHO, M. J., LI, H., NEGRETE, A., SHILOACH, J., BERNS, M. W., SHEN, B., CHEN, L. & WU, X. 2014. CtIP maintains stability at common fragile sites and inverted repeats by end resection-independent endonuclease activity. *Mol Cell,* 54**,** 1012-21.

WYATT, H. D., LAISTER, R. C., MARTIN, S. R., ARROWSMITH, C. H. & WEST, S. C. 2017. The SMX DNA Repair Tri-nuclease. *Mol Cell,* 65**,** 848-860 e11.

WYATT, H. D., SARBAJNA, S., MATOS, J. & WEST, S. C. 2013. Coordinated actions of SLX1-SLX4 and MUS81-EME1 for Holliday junction resolution in human cells. *Mol Cell,* 52**,** 234-47.

WYATT, H. D. & WEST, S. C. 2014. Holliday junction resolvases. *Cold Spring Harb Perspect Biol,* 6**,** a023192.

XU, H., ZHANG, P., LIU, L. & LEE, M. Y. 2001. A novel PCNA-binding motif identified by the panning of a random peptide display library. *Biochemistry,* 40**,** 4512-20.

YING, S., MINOCHERHOMJI, S., CHAN, K. L., PALMAI-PALLAG, T., CHU, W. K., WASS, T., MANKOURI, H. W., LIU, Y. & HICKSON, I. D. 2013. MUS81 promotes common fragile site expression. *Nat Cell Biol,* 15**,** 1001-7.

ZACK, T. I., SCHUMACHER, S. E., CARTER, S. L., CHERNIACK, A. D., SAKSENA, G., TABAK, B., LAWRENCE, M. S., ZHSNG, C. Z., WALA, J., MERMEL, C. H., SOUGNEZ, C., GABRIEL, S. B., HERNANDEZ, B., SHEN, H., LAIRD, P. W., GETZ, G., MEYERSON, M. & BEROUKHIM, R. 2013. Pan-cancer patterns of somatic copy number alteration. *Nat Genet,* 45**,** 1134-40.

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. *Mol Cell,* 27**,** 367-79.

ZHOU, Y., MENG, X., ZHANG, S., LEE, E. Y. & LEE, M. Y. 2012. Characterization of human DNA polymerase delta and its subassemblies reconstituted by expression in the MultiBac system. *PLoS One,* 7**,** e39156.

ZLOTORYNSKI, E., RAHAT, A., SKAUG, J., BEN-PORAT, N., OZERI, E., HERSHBERG, R., LEVI, A., SCHERER, S. W., MARGALIT, H. & KEREM, B. 2003. Molecular basis for expression of common and rare fragile sites. *Mol Cell Biol,* 23**,** 7143-51.

ZUKER, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res,* 31**,** 3406-15.

Chapter 3: Investigating the Causes and Consequences of FRA16D and Flex1 Fragility, Unpublished Work

Abstract

Flex1 is a subregion of CFS FRA16D that forms secondary structures and plays a role in CFS FRA16D expression. In these unpublished works, I have used the Flex1 system to follow up on some of the SSE requirements for Flex1 fragility. I also used the Flex1 system to carefully investigate some of the theories proposed for CFS expression in humans, especially the roles of replication fork stabilizers and of transcription and R-loops at Flex1. I have mentored several undergraduates and graduate students in the Freudenreich lab who have contributed to the works below.

The CAG data from this chapter is from the following paper currently in submission, referred to as Gellon et al., submitted.

Mrc1 and Tof1 prevent fragility and instability at long CAG repeats by their fork stabilizing function

Authors: Lionel Gellon[1], Simran Kaushal[1], Jorge Cebrián[1], Mayurika Lahiri[1,2], Sergei Mirkin[1], and Catherine H. Freudenreich[1,*]

[1] Department of Biology, Tufts University, Suite 4700, 200 Boston Ave, Medford, MA, 02155, USA

[2] Current address: Indian Institute of Science Education and Research, Pune 411 021, Maharashtra, India

*Corresponding author

Catherine Freudenreich Tel +1 617 627 4037; Fax +1 617 627 0309; E-mail catherine.freudenreich@tufts.edu

Secondary Structures and CFS Expression: Introduction

It has long been hypothesized that CFSs are enriched in secondary structure forming sequences (reviewed in Kaushal Thesis Chapter 1), and the FlexStab program has been useful in identifying subregions of interest within various CFSs. While this program has identified some interesting CFS subregions, we feel its intent to identify highly flexible subregions is not the most relevant hypothesis for approaching CFS sequence analysis.

In 2009, Dr. Catherine Freudenreich, Dr. Sergei Mirkin, Dr. Lenore Cowen, and Dr. Anoop Kumar designed a palindrome prediction program to identify subregions within CFSs that are likely to form secondary structures and therefore play a role in their fragility. This program had several parameters: the length of the stem must be at least 20 bp, 3 or less mismatches are allowed in the stem, 2 or less inserts or gaps are allowed in the stem, and the length of the loop must be less than 12 bases (the length can be up to 30 bp in AT-rich regions, defined as a region with >80% A or T bases).

This palindrome prediction program identified several subregions of FRA16D that may form stable secondary structures, including the Flex1 and Flex5 subregions previously identified using FlexStab (Figure 3-1 and 3-2). Palindrome 1 (P1) is a 127 bp palindrome predicted to form just outside of what is considered the FRA16D fragility core; it is possible that secondary structures at P1 could initiate fragility, thereby extending our definition of the boundaries of the FRA16D core, or it could prevent healing after fragility within the FRA16D core.

Flex1, as previously mentioned, is AT-rich and has a polymorphic AT repeat that is 34 repeats ((AT)34) in human genome version 18 (hg18). Flex4 has a short 8 AT repeat sequence and a 17 bp polyA repeat and was previously shown to not increase fragility in the YAC system (Zhang and Freudenreich, 2007). Flex5 has an interrupted (AT)24i repeat and a 27 bp polyA tail; it was a polymerase δ pausing sequence *in vitro* (Bergoglio et al., 2013). P5 and P5b are 90 bp and 70 bp hairpins very close together (170 bp apart) and are not AT-rich.

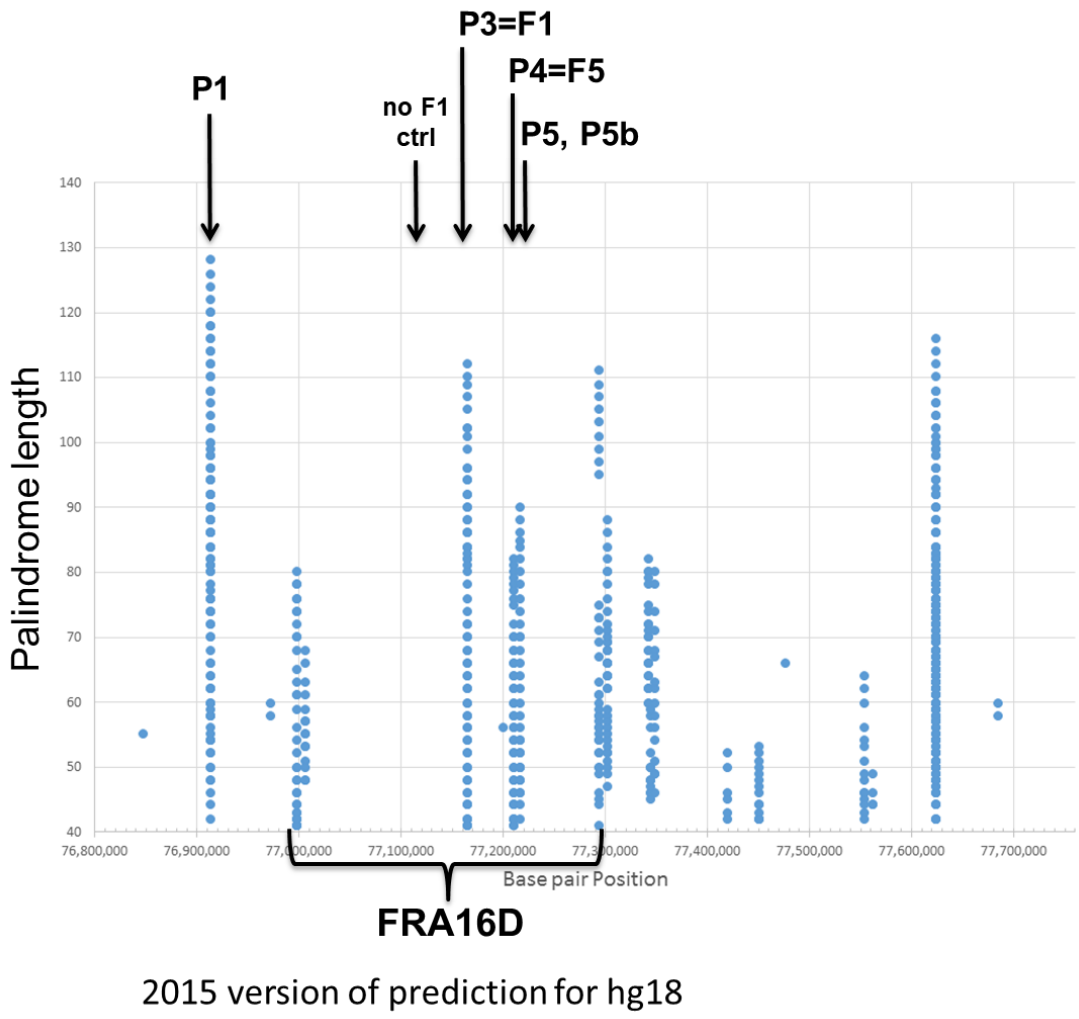


Figure 3-1. Subregions of human chromosome 16 (build hg18) predicted to form stable secondary structures. The computer program designed and executed by Freudenreich, Mirkin, Cowen, Kumar identified subregions of FRA16D likely to form long palindromes. Palindrome peaks of interest are labelled. The longer the palindrome length, the more stable the predicted secondary structure.

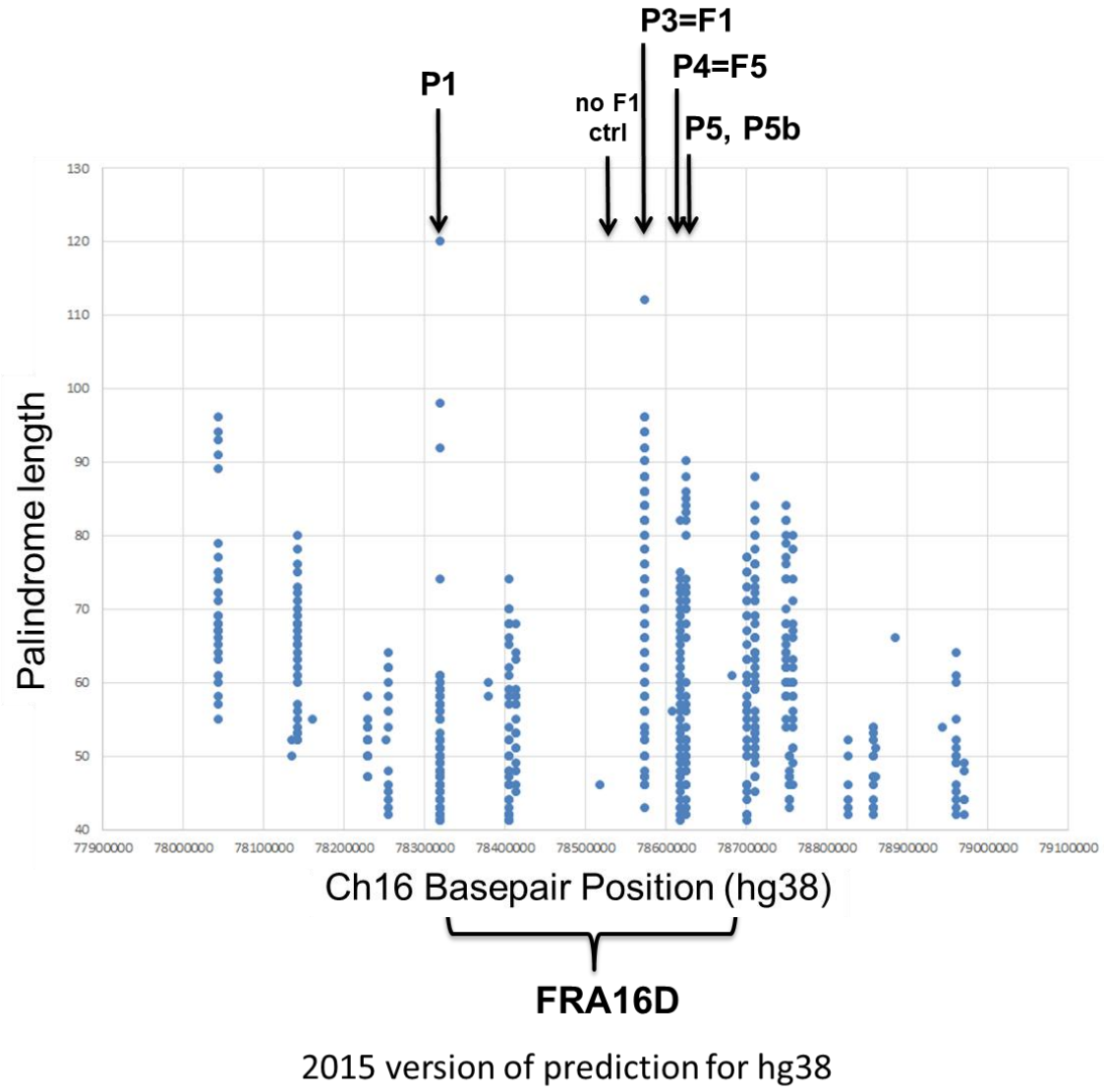


Figure 3-2. Subregions of human chromosome 16 (build hg38) predicted to form stable secondary structures. The computer program designed and executed by Freudenreich, Mirkin, Cowen, Kumar identified subregions of FRA16D likely to form long palindromes. Palindrome peaks of interest are indicated.

Secondary Structures and CFS Expression: Results

Author contributions: Strains were created by Alice Haouzi, these PCR checks and PFGE were performed by Simran Kaushal.

Our goal has been to individually replace each of the potential structure-forming sequences within the large FRA16D YAC and measure breakage. Replacement of Flex1 with a selectable marker in the context of the FRA16D 801B6 YAC (hereafter referred to as a Flex1Δ FRA16D YAC strain) resulted in a measurable and significant decrease in breakage frequency (Chapter 2, Figure 2-1). Future plans are to make single and double replacements of combinations of P1, Flex1, and Flex5 to understand if multiple secondary structure forming sequences play an additive or synergistic role in FRA16D fragility.

Table 3-1. FRA16D Subregion PCR Primers and Amplification Information

Region amplified	**CHF Lab Primer #s**	**Tm**	**Expected Size**	**PCR Amplification Conditions**
P1	1807	58.6 C	635 bp	53 C 0:30 annealing, 68 C 1:00 ext
	1704	57.1 C		
P3 (F1)	1840	60.0 C	875 bp	61.1 C 0:30 annealing, 68 C 1:30 ext
	1841	59.5 C		
P4 (F5)	2362	63 C	1475bp WT F5	61 C 1:00 annealing, 72 C 2:00 ext
	2363	63 C	1522bp F5::his	
P5	1810	59.7 C	1241 bp	61.1 C 0:30 annealing, 68 C 1:30 ext
	1811	60.1 C		

Table 3-2. FRA16D YAC PCR Strain Confirmation Results Summarized. See Addendum Table 3-1 for large FRA16D YAC % FOA^R of all strains. The integrity of the strains in red is in question.

Strain Name	**CFY#**	**PFGS**	**P1 PCR**	**P3 (F1) PCR**	**P4 (F5) PCR**	**P5 PCR**
FRA16D	1051	intact	yes	yes	Yes	yes
FRA16D F1::kan	3076	intact	yes	no	Yes	yes
FRA16D F1::kan	3077	intact	yes	no	Yes	yes
FRA16D P5P5b::hyg	3517	intact	yes	yes	no	no
FRA16D F5::his	3513	intact	yes	yes	Yes?	yes
FRA16D F5::his	3514	intact	yes	yes	Yes?	yes
FRA16D F5::his	3587	shortened	yes	yes	No	no
FRA16D F5::his	3588	intact	yes	yes	Yes?	yes

Secondary Structures and CFS Expression: Discussion

As shown in Chapter 2, removal of Flex1 from a YAC containing FRA16D results in a dramatic decrease in fragility, illustrating Flex1's importance in FRA16D's fragility. Additional replacement strains were shown to have unreliable YAC structure by either PFGE or PCR. The FRA16D YAC structure is questionable for strains in red in Table 3-2 and thus should be remade for proper evaluation of the effect on FRA16D breakage.

It would be interesting to replace the gene *MUS81* in the large FRA16D YAC with and without Flex1. If a further reduction in fragility is found in the FRA16D Flex1Δ *mus81*Δ large FRA16D YAC strain, this could indicate that Mus81 is acting at multiple secondary structure forming sequences at FRA16D to initiate CFS expression.

Replication through Flex1 and FRA16D: Introduction

Author contributions: Simran Kaushal made Flex1 *mrc1*Δ strains and Flex1 *tof1*Δ strain and did initial DDRAs. Charles E. Wollmuth created and checked the *tof1*Δ ctrl DDRA strain and Charles E. Wollmuth and Samantha Regan completed the assays. CAG strain creation and assays were completed by Simran Kaushal, Lionel Gellon, and Mayurika Lahiri. Simran Kaushal and Julia Haft performed the *mrc1AQ* strain transformations together. Julia Haft created *RAD30* and *CTF4* mutants and performed the DDRAs. Simran Kaushal created strains for γH2AX ChIP and performed assays.

The replication fork travels with many accessory proteins to fulfill different functions under different replication impairment circumstances. Since AT-rich sequences are known to cause fork stalling at CFSs (summarized in Kaushal Thesis Chapter 1), and human polymerase δ pausing at the AT repeat and long 3' Flex1 sequences was detected (Kaushal Thesis Chapter 2 Figure 2-6), we wanted to understand the role of replication fork stabilizers, translesion synthesis polymerases, and checkpoint proteins on Flex1 (AT)34 fragility.

Replication fork stabilizers

The S phase cell cycle checkpoint monitors DNA replication progression and can detect replication issues such as depleted dNTP pools or stalled forks. In some conditions, the S phase checkpoint can then be induced and block late origin firing and activate various repair pathways to complete replication before M phase begins, in order to avoid large-scale DNA rearrangements. Mec1 recruitment to RPA-coated ssDNA activates the S phase checkpoint, which in turn activates mediator proteins Mrc1, Rad9, Tof1, and Csm3. These proteins in turn activate Rad53, which results in replication fork stabilization and prevention of any additional origin firing. After replication is completed, the S phase checkpoint must be deactivated in order for the cell cycle to resume. Since human Claspin (Mrc1) is associated with CFS expression (Glover et al., 2017) and it was recently found that Mrc1 interacts with Mcm7 to facilitate origin firing (Masai et al., 2017), we wanted to investigate whether it has a role in Flex1 fragility.

S. cerevisiae Mrc1 (hClaspin) has 2 independent functions: it plays a fork stabilization role during replication (Katou et al., 2003) by interacting with polymerase ε on the leading strand (Lou et al., 2008) and also signals the S phase checkpoint in response to DNA damage (Osborn and Elledge, 2003). In response to DNA damage, Mrc1 is phosphorylated at several serine-glutamine/threonine-glutamine (SQ/TQ) motifs, which results in Rad53 recruitment, further inducing the S phase checkpoint (Osborn and Elledge, 2003, Pike et al., 2004). We tested the Flex1 DDRA rate in both *mrc1*Δ and *mrc1AQ* checkpoint deficient mutants (Osborn and Elledge, 2003) to investigate the importance of both Mrc1 functions in maintaining Flex1 stability. Ctf4 is thought to stabilize the replisome on the lagging strand via its binding to both the Mcm2-7 replicative helicase and polymerase α (Gambus et al., 2009), therefore its role at Flex1 was also investigated.

Tof1 (hTimeless) also travels with the replication fork; in *S. cerevisiae*, Tof1-Csm3 recruits Mrc1 (Bjergbaek et al., 2005, Uzunova et al., 2014). Mrc1 and Tof1-Csm3 are thought to restrain replication fork progression under certain DNA damage conditions (Pardo et al., 2017). When Mrc1, Tof1, and Csm3 are absent, the DNA synthesis rate decreases and the leading strand polymerase ε becomes uncoupled from the Mcm2-7 helicase, which leads to excessive DNA unwinding and ssDNA formation at stalled forks (Katou et al., 2003, Lou et al., 2008, Bjergbaek et al., 2005).

Since the conditions after deletion of *MRC1* and *TOF1* are similar to CFS induction conditions, we investigated the role of Mrc1 and Tof1 in Flex1 fragility.

I also investigated the role of Mrc1 and Tof1 on fragility of CAG repeats of short, medium, and long lengths. In the CAG YAC strains, *mrc1*Δ results in a non-specific increase in fragility. This Gellon et al. paper is currently in submission.

Non replicative polymerases at Flex1

The *RAD30* gene encodes DNA polymerase η, a translesion synthesis (TLS) polymerase involved in lesion bypass (Haracska et al., 2001). Human polymerase η is recruited to CFSs, including subregions of FRA16D, to enable replication in mammalian cells by exchanging with replicative polymerase δ (Bergoglio et al., 2013, Barnes et al., 2017).

In Thesis Chapter 2, we found that *pol32*Δ mutants had an increase in Flex1 fragility in the DDRA, in line with a role for its human homolog POLD3 in MiDAS at CFSs. However, this knockout must be made in the control strain to determine if this effect is Flex1-dependent and therefore secondary structure forming and fork stalling sequence dependent.

Replication through Flex1 and FRA16D: Results

Author contributions: Alice Haouzi created, checked, and did assays on the large FRA16D YAC strains. Simran Kaushal performed subregion PCR and PFGE verification of the FRA16D YAC strains as in (Kaushal Thesis Chapter 2).

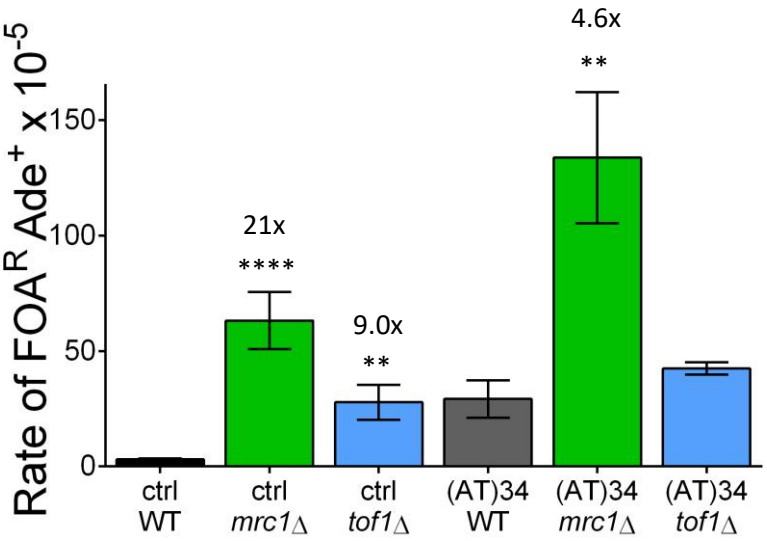
MRC1 mutants had a significant increase in DDRA rates of both Flex1 and ctrl strains (Figure 3-3 A). Ctrl *tof1*Δ mutants had a significant increase in fragility, yet no change in Flex1 (AT)34 fragility was detected. Since only one ctrl *tof1*Δ

strain had been made, 2 new strains should be made and assayed in the ctrl background.

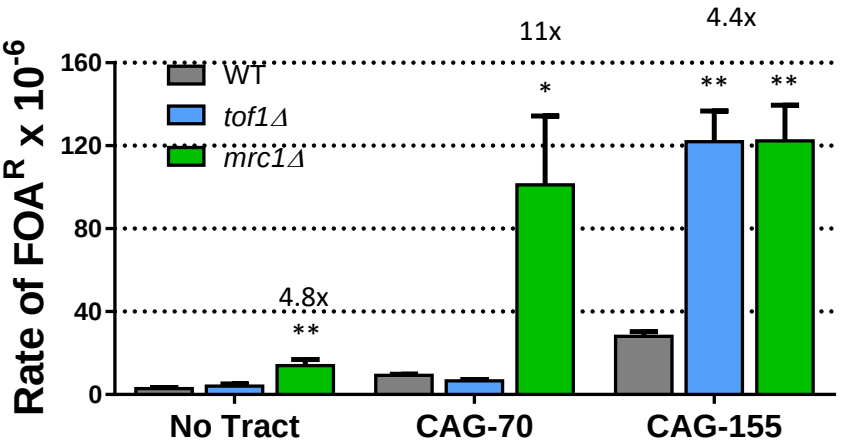
I contributed to testing the effect of *mrc1*Δ and *tof1*Δ mutants in CAG YAC fragility (Gellon et al., in submission.; see excerpted Figure 3-3B). Loss of Mrc1 also resulted in an increase in fragility, regardless of the sequence tested. However, the change in fragility was much less dramatic in the ctrl sequence (CAG-0) than in the Flex1 ctrl sequence (21-fold increase in Flex1 strain background in the DDRA assay versus 4.8-fold increase in the CAG-0 YAC strains). From these data, Mrc1 has a general role stabilizing all forks (as indicated by the increase in fragility upon *mrc1*Δ in control strains in both systems). The role of Mrc1 is also evident in strains containing structure-forming CAG repeats.

The role of Tof1 in preventing fork breakage is specific to medium CAG tracts greater than 85 repeats (Gellon et al., in submission; see excerpted Figure 3-3C) and implies that there may be different DNA substrates formed by CAG repeats as they pass a threshold of over 85 repeats. Tof1 does not seem important for preventing Flex1 fragility. It is unclear why the *tof1*Δ ctrl strain has a 9.0-fold increase in fragility compared to the WT ctrl, as this was not seen in the CAG-0 *tof1*Δ strains. The *tof1*Δ DDRA data are only from one transformant, thus it may be an issue in the background of the tranformant causing an unexpectedly high level of fragility. Currently, new *tof1*Δ control strains are being made and assayed for DDRA rates. Tof1 is known to promote fork pausing at protein-mediated replication barriers by counteracting the Rrm3 helicase (Mohanty et al., 2006), so

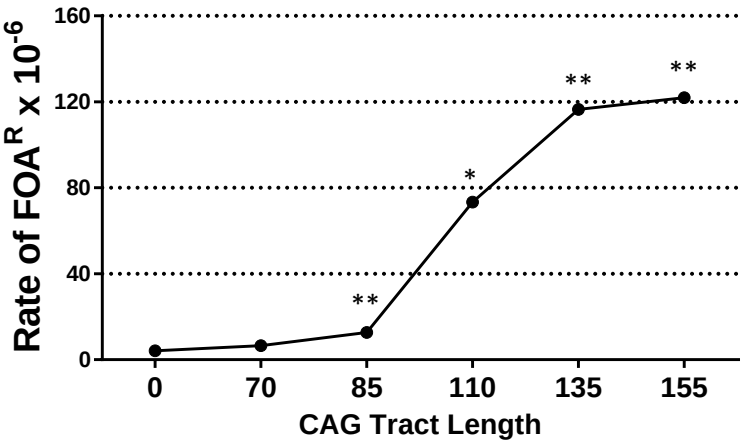
these data may indicate that proteins are not forming any barriers to pause replication and cause fragility at Flex1(AT)34.


Rate of FOA^R Ade$^+$ x 10^{-5}
150
100
50
0
21x
****
9.0x
**
4.6x
**
ctrl WT
ctrl mrc1Δ
ctrl tof1Δ
(AT)34 WT
(AT)34 mrc1Δ
(AT)34 tof1Δ

A.


Rate of FOA^R x 10^{-6}
160
120
80
40
0
WT
tof1Δ
mrc1Δ
4.8x
**
11x
*
4.4x
**
**
No Tract
CAG-70
CAG-155

B.


Rate of FOA^R x 10^{-6}
160
120
80
40
0
**
*
**
**
0
70
85
110
135
155
CAG Tract Length

C.

Figure 3-3. Mrc1 and Tof1 mutants affect Flex1 and CAG repeats differently. A) Flex1 DDRA data for ctrl and mutant strains. B) CAG YAC fragility assay data for *mrc1*Δ and *tof1*Δ mutants. YACs with no CAGs, medium length CAGs, and long CAGs were tested. *MRC1* mutants show an increase in fragility at all sequences tested. Tof1 had no effect on Flex1 fragility and a C) length-dependent effect at only long CAG repeat tracts. * = P<0.05, **= P < 0.01 and ****= P < 0.0001 when compared to same background. See Tables 3-2, 3-3, and 3-5 for raw data.

Mrc1, through either its fork stabilization or checkpoint signaling response, is needed to facilitate faithful replication through Flex1 and CAG repeats. We tested the importance of the replication fork stabilization function by testing *mrc1AQ* mutants in both ctrl and Flex1 strains. Briefly, *mrc1AQ* mutant plasmids from the Elledge lab were transformed into *mrc1*Δ strains and DDRAs were performed under normal conditions. The presence of *mrc1AQ* did not change Flex1 (AT)34 fragility from WT strains, indicating that the it is the fork stabilization function of Mrc1 that causes breakage at Flex1. Surprisingly, the *mrc1AQ* mutant gave a significant increase in the ctrl DDRA rate (Figure 3-4). However, one major caveat exists when interpreting the *mrc1AQ* data. The *mrc1AQ* plasmid, marked with LEU2, was transformed into mrc1 strains in order to create *mrc1AQ* strains. DDRAs were performed without maintaining selection for the *mrc1AQ* plasmid, therefore it is possible that cells in the assays lost the *mrc1AQ* mutant plasmid and thus were expressing an *mrc1*Δ phenotype and potentially incurring breaks. In the future, DDRAs should be repeated while maintaining *LEU2* selection throughout each step of the assay to get a trustworthy snapshot of the role of Mrc1's checkpoint activation function on fragility of the control sequence. Nonetheless, we can conclude that the replication fork stabilization function of Mrc1 is likely responsible for its protective role at Flex1 (Figure 3-4). The fork stabilization function of Mrc1 also seems to be the protective function of Mrc1 to prevent CAG fragility (Figure 3-5).

CTF4 mutants also showed a modest but significant increase in DDRA rate for Flex1 strains, therefore both leading and lagging strand fork stabilization is

important for preventing fragility and deletions at Flex1. It is possible that the coupling function of Mrc1 on the leading strand is more critical in avoiding Flex1than the coupling function of Ctf4 on the lagging strand. See Figure 3-6.

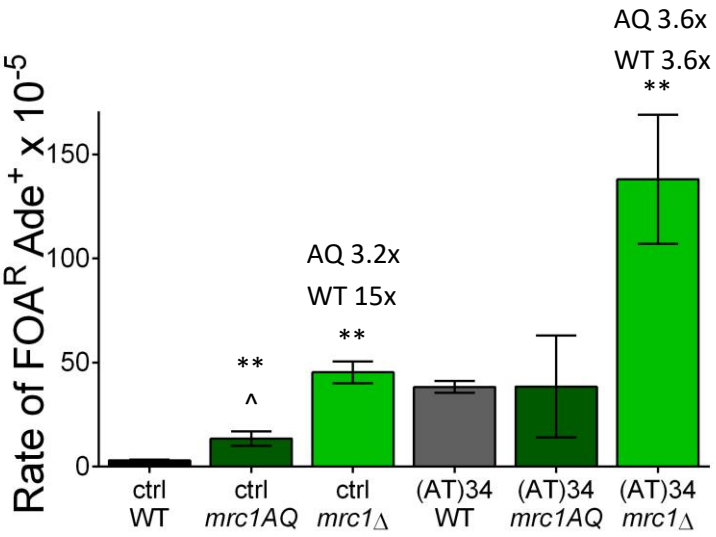


Figure 3-4. Mrc1 and mrc1AQ DDRA rates for ctrl and Flex1 sequences. The replication fork stabilization function of Mrc1 is likely responsible for its protective role at Flex1. **= P < 0.01 when compared to same background, ^ = P < 0.05 when compared to mrc1AQ with same region of interest. See Kaushal Chapter 3 Addendum Table 3-2 for raw values.

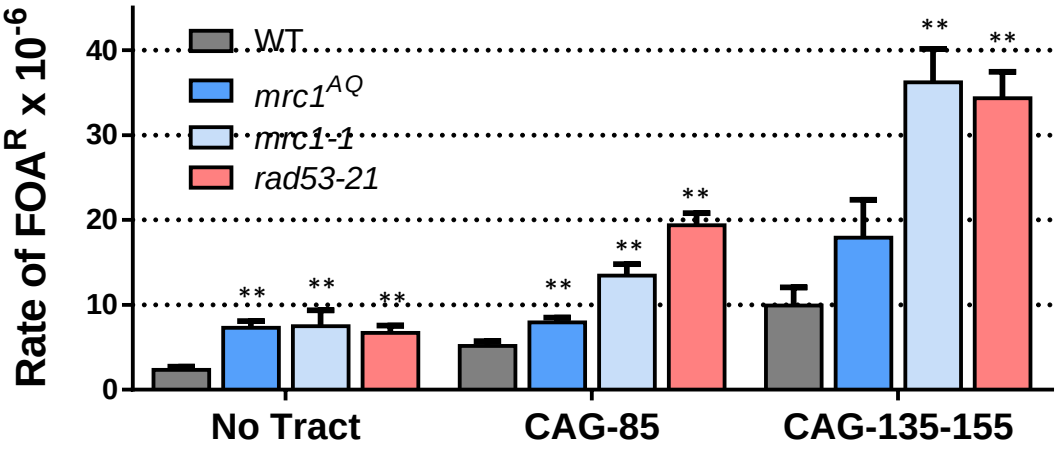


Figure 3-5. The checkpoint function of Mrc1 is dispensable for preventing fragility at CAG repeats. Our data indicate that loss of the Mrc1's fork stabilization role is the primary cause of fragility in *mrc1*Δ mutants. Figure adapted from Gellon et al. paper in submission. See Kaushal Thesis Chapter 3 Addendum Table 3-4 for raw data.

Replication through Flex1 is not dependent on polymerase η, as indicated by no change in the DDRA rate (Figure 3-6).

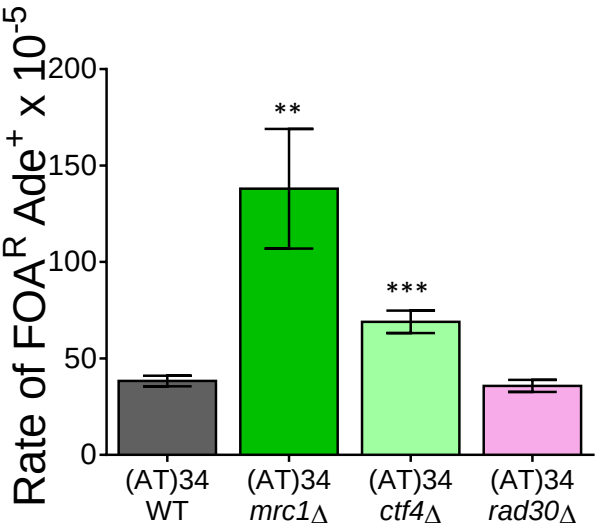


Figure 3-6. Flex1 DDRA rates with various fork stabilizers and one TLS polymerase mutant. **= P < 0.01 and ***= P < 0.001 when compared to same background.

We sought to test for the histone modification γH2AX, which is a marker of DSBs and stalled replication forks. It was hypothesized that the histone modification may be enriched at Flex1 compared to a control strain and also may indicate at which phase of the cell cycle DNA damage is occurring at Flex1. We used chromatin immunoprecipitation (ChIP) to test for γH2AX enrichment at Flex1 using primers binding on either side of the repeat.

Our results show no such enrichment (Figure 3-7), however the ChIP may not have worked, since there is no enrichment of IP sample over input. The DNA shearing size was not optimal (larger than the 300-500 bp fragments desired for ChIP), and a positive control was not included, therefore this experiment should be repeated, preferably with a Flex1 (AT)34 S3' strain. See raw and analyzed data in Figure 3-7.

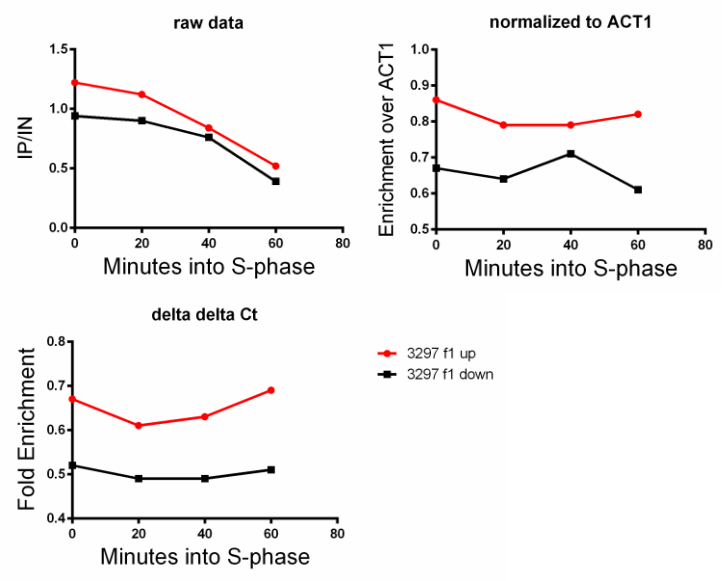


Figure 3-7. Raw and analyzed γH2AX ChIP data. ChIP was performed with the phospho129 antibody on Flex1 (AT)34 L3' strains with qPCR primers hybridizing to either side of the repeat construct. For raw data numbers, see Kaushal Thesis Chapter 3 Addendum Table 3-7.

Replication through Flex1 and FRA16D: Discussion

An *mrc1*Δ mutant gave an increased fragility rate in controls in both the Flex1 DDRA and the CAG YAC systems, supporting that the protein plays a general fork stabilization role needed for faithful replication through all DNA, regardless of its structure-forming potential. However, the dramatic 21x increase in fragility in the *mrc1*Δ DDRA control strain has a higher fold-increase in fragility than the CAG-0 control on the YAC and any of the strains containing structure-forming repeats, so this strain should be freshly made and retested. If we see the same result again, it is possible that Mrc1 is more important in facilitating replication through an internal chromosome location (DDRA) versus replicating DNA towards the end of a chromosome (as in the YAC) system. *TOF1* mutants did not have an increase in fragility in the CAG-0 control but did for the Flex1 control sequence – however, since these data are from one strain I do not consider them trustworthy and thus plan to make two fresh *tof1*Δ control DDRA strains to test its true effect on breakage. Assuming that freshly made *tof1*Δ control DDRA strains still show an increase in the DDRA rate, this would support the conclusion that the Mrc1-Tof1-Csm3 complex is more important for replicating an internal location.

The *mrc1*Δ fold increase in fragility in the AT versus CAG system is similar and thus may implicate a role for Mrc1 at DNA structures. Tof1 activity is only implicated in preventing fragility at very specific DNA structures, as it is not important at Flex1 (AT)34 or medium-length CAGs (CAG-70 to 85) but is important in preventing fragility at structures formed by long CAGs (CAG-110 to

155). These data may mean that Flex1 (AT)34 is not as strong of a replication barrier as long CAG repeats. Further, since Tof1 has a role in removing protein impediments to DNA replication, these data may imply that long CAG repeats tightly bind proteins that must be removed with help from Tof1 in order to prevent fragility.

Data from *mrc1AQ* mutants in both the DDRA and YAC systems indicate that the checkpoint function is more necessary for preventing DNA fragility at control sequences than structure-forming sequences. Comparing *mrc1*Δ versus *mrc1AQ* rates for various tract lengths in the CAG and Flex1 AT systems shows an increase in fragility at structure-forming sequences when the fork stabilization function of Mrc1 is compromised, but not in the *mrc1AQ* mutant. It is likely that due to the decoupling of the helicase and polymerase in the absence of Mrc1, long stretches of ssDNA can arise and form secondary structures, resulting in fragility.

The increase in Flex1 DDRA rate upon deletion of both leading strand and lagging strand specific fork stabilization genes (*MRC1* and *CTF4*, respectively) supports that secondary structures are forming on both strands of Flex1, however it remains to be determined if these structures are hairpins, slipped-strand structures, or a cruciform engaged with both strands. The difference in Flex1 DDRA rates between *mrc1*Δ and *ctf4*Δ mutants may imply that more secondary structures are forming at Flex1 on the leading strand, and that these leading strand structures are more deleterious to the cell and are more likely to result in fragility. It is also possible that deleting *MRC1* has a greater effect on fork destabilization than the deletion of *CTF4*.

Strikingly, it was recently found that the *S. cerevisiae* protein Dia2 degrades Mrc1 to restart stalled forks after checkpoint induction in a pathway involving Sgs1 and Mph1 (Chaudhury and Koepp, 2017). Dia2 may be a key regulator of the DNA damage response and replication resumption at Flex1 and therefore is an attractive and easy genetic target to replace in either Flex1 fragility system.

In the Eckert lab 2017 study, polymerase δ and polymerase η exchange occurred at a region of polyA runs and interrupted AT repeats. In Chapter 2 of this thesis, I present data showing modest increases in fragility at interrupted Flex5 (AT)24i repeats versus perfect Flex1 (AT)23 repeats. Therefore, the two types of repeats seem to have very different characteristics. Rad30 was not needed for replication through Flex1, which contradicts results from the Eckert lab showing a role for the polymerase in replicating through the Flex5 interrupted AT repeat in FRA16D (Barnes et al., 2017, Bergoglio et al., 2013). (See Figure 3-6). Since the Flex5 sequence has a polyA run, it would be interesting to see if the Flex5 sequence causes an increased DDRA rate in a *rad30*Δ mutant, an experiment currently in progress.

In the future, it would also be of interest to make mutants in other yeast TLS polymerases pol ζ (catalytic subunit in yeast is Rev3) and Rev1 to determine if any other TLS polymerases are important for replicating through Flex1. Even if we see no change in DDRA rate upon individual TLS polymerase gene replacements, we could measure DDRA rates in a triple mutant to see if multiple TLS polymerases have functional redundancy to facilitate replication through Flex1.

I saw no enrichment or change in γH2AX recruitment at Flex1, however there were several caveats with my data. Therefore, this is not necessarily a negative result and it is yet to be determined if the modification will be detectable at Flex1.

Investigating when Mus81 is cleaving structures at Flex1:

Introduction

In *S. cerevisiae*, Mus81's binding partner Mms4 is phosphorylated by Cdc5/Cdc28 and Cdc7/Dbf4 in order to hyperactivate its nucleolytic activity in G_2/M phase (Blanco and Matos, 2015, Wild and Matos, 2016, Princz et al., 2017). To determine whether hyperactivation of Mus81 activity is required for Flex1 cleavage, we used a nonphosphorylatable *mms4* mutant (*mms4-np*) lacking phosphorylation sites, as it significantly reduces M phase Mus81 nuclease activity (Gallo-Fernandez et al., 2012, Saugar et al., 2013). The nuclease activity reduction upon the introduction of the nonphosphorylatable mutations is shown in Figure 3-8B, adapted from (Gallo-Fernandez et al., 2012).

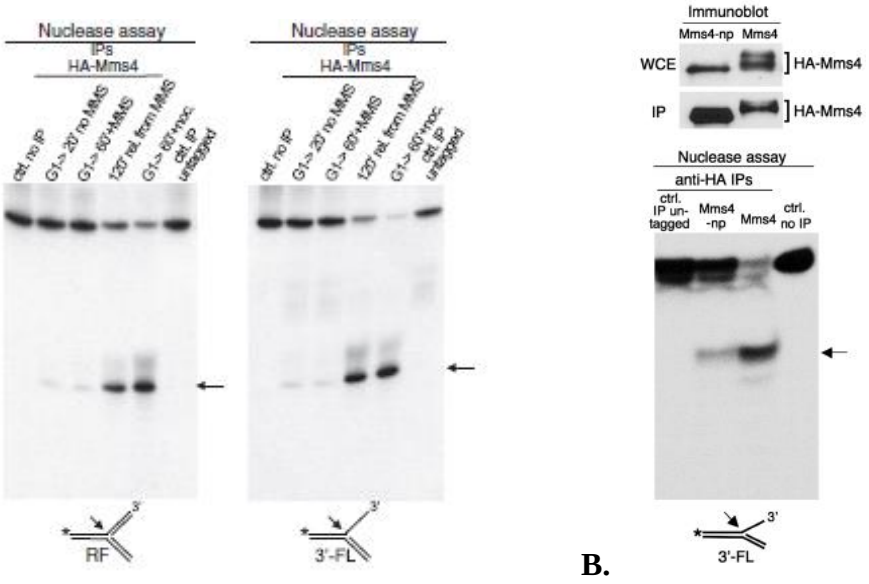


Figure 3-8. Mms4 is phosphorylated after completion of S-phase and *mms4-np* mutant has reduced nuclease activity. (A) This image is modified from (Saugar et al., 2013). The panel indicates that Mus81 has some minimal cleavage activity even without Mms4 phosphorylation. (B) Mus81 nuclease activity is reduced but not fully ablated in *mms4-np* mutants. From (Gallo-Fernandez et al., 2012).

Investigating when Mus81 is cleaving structures at Flex1: Results

Author contributions: Undergraduate senior thesis student Charles Wollmuth created the *mms4-np*,strains and performed the DDRAs.

If Mus81 cleavage activity in M phase were responsible for Flex1 fragility, we would expect to see a decrease in DDRA rate to *mus81*Δ levels in an *mms4-np* mutant. However, the recombination rate at Flex1 (AT)34 was unchanged in the *mms-np* mutant (Figure 3-9). Therefore, these data may cautiously imply that Mus81 cleavage activity in S phase is sufficient to induce the Flex1 fragility rates we see. However, there are many caveats of these unexpected results, see Discussion.

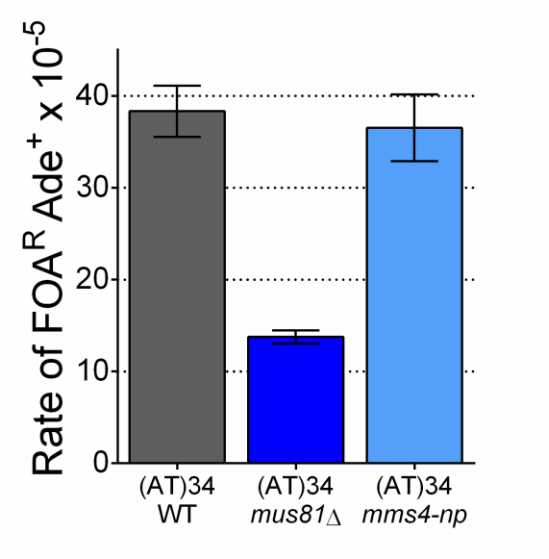


Figure 3-9. Flex1 *mms4-np* data. Flex1 strains with a nonphosphorylatable Mms4 show no change in DDRA rate.

Investigating when Mus81 is cleaving structures at Flex1: Discussion

The Flex1 *mms4-np*data may indicate that the S phase activity level of Mus81 is sufficient to cause Flex1 (AT)34 fragility, therefore it could be cleaving a structure generated in S phase. However, since it is well-supported that the Mus81-Mms4-Slx4 complex does not form until M phase (Princz et al., 2017), it is also possible that minimal Mus81 activity in M phase is sufficient to induce fragility at Flex1 (see Figure 3-8). Mus81 is thought to interact with the Slx4 scaffold through Dpb11 and Rtt107 proteins. Therefore it would be useful to make single, double, and triple knockouts of *rtt107*Δ and *dpb11*Δ along with *mus81*Δ to understand if it is Mus81's recruitment to Slx4 or as-yet uncharacterized Mus81 S phase activity working at Flex1.

The role of healing in CFS expression: Introduction

I-SceI strains with and without Flex1 flanking sequences were created so that DSBs could be induced within Flex1 and healing by SSA could be observed. We found that the presence of the L3' Flex1 flanking sequence resulted in a significantly lower DDRA rate, likely because it forms a secondary structure that inhibits efficient resection and healing after fragility (Kaushal Thesis Chapter 2 Figures 2-6B and 2-6E). I also measured DDRA rates of each strain in a *sae2*Δ background to investigate healing without a secondary structure-induced break.

The role of healing in CFS expression: Results

Author contributions: all strain construction and assays done by Simran Kaushal.

I hypothesized that the DDRA rate may be reduced in I-SceI-L3' *sae2*Δ strains compared to I-SceI *sae2*Δ and I-SceI-S3' *sae2*Δ as was seen for Flex1 (AT)34 (Kaushal Thesis Chapter 2 Figure 2-6E), as the activity may be needed to unwind secondary structures formed in the L3 flanking sequence for efficient healing after a DSB.

I measured the DDRA rates of all 3 constructs with and without Sae2 in the presence of glucose ("gluc", no DSB induced) and galactose ("gal", 100% DSB induction at I-SceI recognition sequence). I saw an increase in DDRA rate for all I-SceI *sae2*Δ strains relative to WT (Figure 3-10). Sae2 could be needed for healing in a manner that does not lead to deletions in a mechanism independent of structure, explaining the increase in DDRA rate of all strains tested. There is still

a decrease in I-SceI-L3' *sae2*Δ DDRA rates compared to the I-SceI only and I-SceI-S3' strains, which supports that the L3' is causing a defect in healing when located near a DNA break.

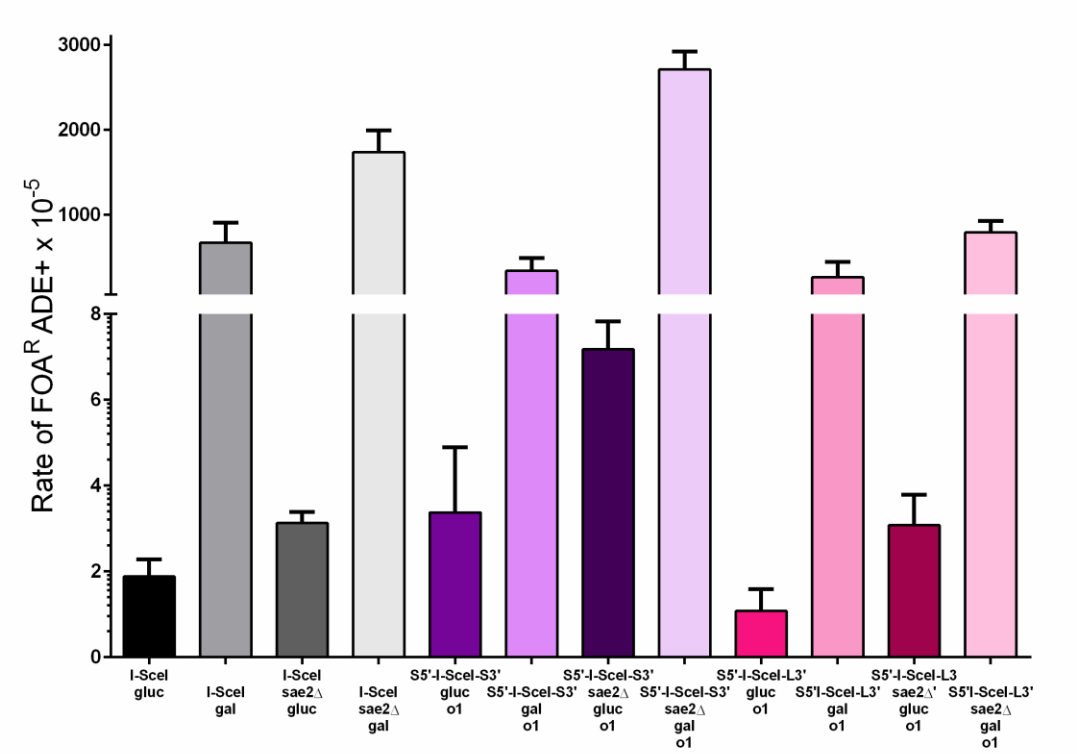


Figure 3-10. Preliminary DDRA rates of I-SceI WT and *sae2* mutants under DSB induction and no induction conditions. Breaks are induced in galactose conditions. See Raw data in Table 3-3. See caveats of data below.

Table 3-3. I-SceI WT and *sae2*Δ preliminary data under 2% glucose (no break induction) and 2% galactose (100% break induction).

I-SceI x 10^{-5}			
Strain #	Gluc	Gal	Fold change
3570 (S-plate)	0.3	263.1^	877
3570	2.3	337.1^	147
3570	2.1^	646.7^	308
3585	2.4^	540.5^	225
3585	2.3	1562.5^	679

All data
Ave gluc: 1.88 x 10^{-5}
Ave gal: 670 x 10^{-5}
Ave fold change: 447

Best data only
Ave gluc: 1.6
Ave gal: -
Ave fold change: -

I-SceI sae2Δ x 10^{-5}			
Strain #	Gluc	Gal	Fold Change
4100	3.6^	2196.7^	610
4101	2.9^	2133^	735
4100	3.5	1486.6	424
4101	2.5	1132.3	452

All data
Ave gluc: 3.1 x 10^{-5}
Ave gal: 1737 x 10^{-5}
Ave fold change: 555

Best data only
Ave gluc: 3.0
Ave gal: 1309 x 10^{-5}
Ave fold change: 438

When resequenced, I-SceI parent strain and parent plasmid had 2 I-SceI sites

S5'-I-SceI-S3' o1 x 10^{-5}			
Strain #	Gluc	Gal	Fold Change
3989	5.9	224.7^	38
3989	6.1^	750.8^	123
3990	0.9^	31.5	35
3990	0.58	341.9^	589

All data
Ave gluc: 3.37 x 10^{-5}
Ave gal: 337 x 10^{-5}
Ave fold change: 196

Best data only
Ave gluc: 3.24
Ave gal: 31.5
Ave fold change: -

S5'-I-SceI-S3' o1 sae2Δ x 10^{-5}			
Strain #	Gluc	Gal	Fold Change
4102	8.2	3007.2^	366
4103	6.8	2086.2^	306
4102	8.2	2914.1	355
4103	5.5	2841.5	516

All data
Ave gluc: 7.2 x 10^{-5}
Ave gal: 2712.3 x 10^{-5}
Ave fold change: 386

Best data only
Ave gluc: 6.9 x 10^{-5}
Ave gal: 2877.8 x 10^{-5}
Ave fold change: 435.5

S5'-I-SceI-L3' o1 x 10^{-5}			
Strain #	Gluc	Gal	Fold Change
3992	0.18^^	2.3	13
3991	1.5^	203.8^	136
3992	0.4	9.3	23
3991	2.9	958.9^	330
3992	0.4	146.4^	366

All data
Ave gluc: 1.1 x 10^{-5}
Ave gal: 264.1 x 10^{-5}
Ave fold change: 173

Best data only
Ave gluc: 1.2
Ave gal: 5.8
Ave fold change: 4.8

S5'-I-SceI-L3' o1 sae2Δ x 10^{-5}			
Strain #	Gluc	Gal	Fold Change
4107	5.2	595.7^	114
4108	2.6	1088.7^	418
4107	2.3	542.3	235
4108	2.2	938.4	426

All data
Ave gluc: 3.1 x 10^{-5}
Ave gal: 791 x 10^{-5}
Ave fold change: 298

Best data only
Ave gluc: 2.3
Ave gal: 740
Ave fold change: 330

^ = data inaccurate due to too many colonies on FOA plates

The role of healing in CFS expression: Discussion

Overall, my data show that Sae2 is primarily required when the break is initiated by Flex1 (AT)34, as the reduction in healing in the absence of Sae2 was no longer observed when the AT repeat was not present. This indicates that the main role for Sae2 at Flex1 is to process hairpin-capped ends formed by the AT repeat. In contrast, when no AT repeat was present, the absence of Sae2 produced the opposite effect: an increase in FOAR. The small but consistent increase in SSA healing in the absence of Sae2 when an I-Sce1 break is induced suggests that Sae2 is required for healing a break without structures in a way that does not bias towards deletions. Additionally, the consistent reduction in healing in the presence or absence of Sae2 when the L3' flanking sequence is present suggests that Sae2 cannot deal with the L3' structure well.

However, there are several caveats to consider in regards to the preliminary I-SceI data in this chapter. In the experiment presented in Chapter 2, I-SceI DDRA data was obtained under 1.5% galactose, 0.5% glucose conditions, in which roughly 50% break induction is expected according to (Escalante-Chong et al., 2015); therefore these data are not directly comparable. The I-SceI *sae2*Δ DDRA data was obtained under 2% galactose conditions, where 100% break induction is expected. Under 2% galactose conditions, the levels of breaks may be so high that differences in healing may be harder to detect. The I-SceI strains 3570 and 3585 contained two I-SceI recognition sequences, which could alter breakage and recombination frequencies and therefore these data may not be trustworthy. Further, the parent strains for the I-SceI *sae2*Δ mutants also may have had these

two recognition sequences. Finally, many of the assays were not high quality data as most plates had too many colonies to count with accuracy, as noted by the red hats in Table 3-3. The I-SceI *sae2*Δ strains should be remade and assays should be conducted in tandem on the same batch of plates and 5-FOA media to reduce the large variation we tend to find in this assay.

Investigating the role of transcription in Flex1 fragility:

Introduction

There are many connections between transcription, secondary structure formation, and fragility. During transcription, RNA polymerase II travels along the DNA template and generates excess negative supercoiling ahead of the fork, which can allow for DNA secondary structure formation that can stall replication and induce fragility. Since CFSs are late replicating and present in large genes, it is more probable that transcription and replication are not spatially and temporally separated over CFSs in human cells. Further, transcription can cause replication fork stalling (Garcia-Muse and Aguilera, 2016).

Long transcripts and slow replication could result in increased transcription-replication collisions, resulting in CFS fragility. One study used qRT-PCR of FACS-separated cell fractions to determine that transcription through CFSs FRA3B, FRA16D, and FRA7K (present in the FHIT, WWOX, and IMMP2L, respectively) takes more than one cell cycle. Transcripts were identified starting in the G_2/M phase and continued throughout one complete cell cycle and finished transcription in the next G_1/early S phase. They found that CFS fragility correlated with gene expression by comparing expression of all three sites in cell lines with high and low transcript levels. The CFS-containing genes tested were all replicated in late S phase, showing that transcription and replication are occurring at the same time over the same template in the CFSs tested. They found

breaks increased genome-wide when RNaseH1 was depleted with siRNA, prolonging the livelihood of R-loops (Helmrich et al., 2011).

In contrast, a study from the Debatisse lab saw no correlation between gene expression observed by RNA transcript levels using RT-qPCR and CFS expression by the frequency of breaks visible on metaphase chromosomes. They also note that breakage frequencies vary throughout a gene, across different cell lines, and even across multiple isolates of the same cell type; all of these data indicate that at least for some fragile sites, there is no correlation between a transcriptional unit and breakage. (Le Tallec et al., 2013).

Madireddy et al. measured replication dynamics through FRA16D and FRA6E using single molecule analysis of replicated DNA (SMARD) in lymphoblasts, which have high fragile site expression, and fibroblasts, which have low fragile site expression. They found that FANCD2 allows for normal bidirectional replication through the fragile sites from both sides by aiding the firing of dormant replication origins to assist replication through the fragility core regions of both sites only in lymphoblasts, thus providing more evidence supporting cell-type specificity of CFSs and showing that FANCD2 function is needed specifically in the cell type with a paucity of initiation events. FANCD2 may function to remove R-loop replication impediments as overexpression of RNaseH1 allowed for replication to proceed bidirectionally in *FANCD2*$^{-/-}$ cells. This FANCD2 function seems to be in conjunction with the downstream Fanconi Anemia pathway proteins BRCA2 and FANCD1, notably which are involved in replication fork restart (Raghunandan et al., 2015). (Madireddy et al., 2016).

Investigating the role of transcription in Flex1 fragility: Results

Author contributions: all RT-qPCR was performed by Simran Kaushal.

The DDRA gives 10-fold higher rates than the YAC assay for strains with the same Flex1 tract. This difference could be due to differences in healing efficiency in the two different genetic assays. We also hypothesized that the internal chromosome II locus in the DDRA may be more highly transcribed than the YAC, resulting in more cruciform extrusion and Flex1 fragility. We supported this hypothesis by performing RT-qPCR of cDNA generated from Flex1 (AT)34 strains in both genetic assay backgrounds. Fitting with our hypothesis, the internal chromosome is more highly transcribed at Flex1 than the YAC (Figure 3-11).

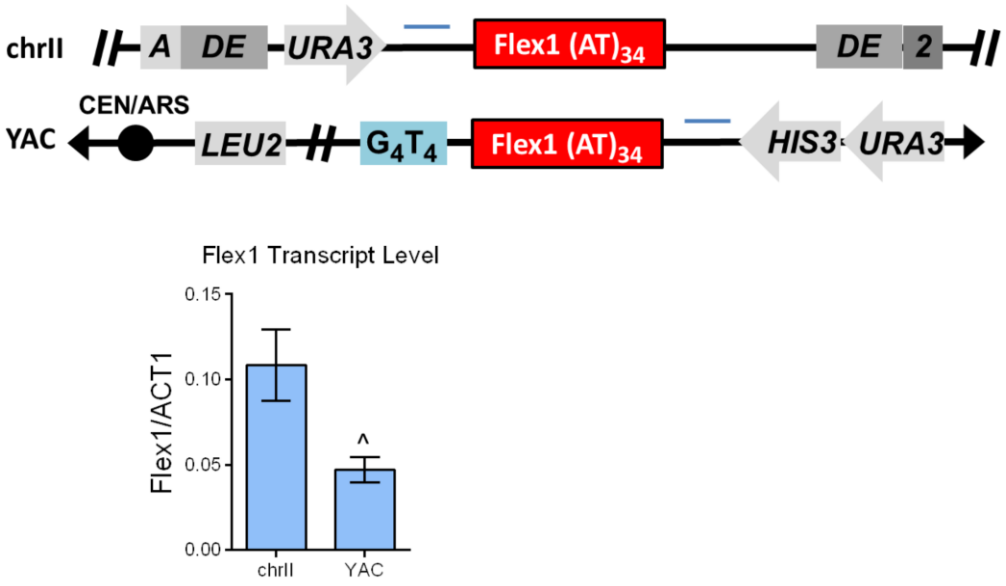


Figure 3-11: qRT-PCR of both Flex1 (AT)34 genetic constructs to determine the level of RNA transcripts entering Flex1. There are more Flex1 transcripts present on chrII than on the YAC when compared to ACT1. Chromosome II and YAC data is combined from 5 experiments from 3 different RNA and subsequent cDNA preps. P= 0.0264. For raw data see Kaushal Addendum Table 3-8.

In *S. cerevisiae*, Top1 is needed to relieve torsional stress during transcription (Wang, 2002). In Chapter 2, we hypothesize that a cruciform is forming at the Flex1 AT repeats *in vivo*. Since AT repeats can extrude to form cruciforms during transcription (Dayn et al., 1992), it is possible that Flex1 AT repeat cruciform extrusion could occur during transcription of FRA16D. As an initial test, I created a *top1*Δ Flex1 (AT)34 DDRA strain and tested fragility (see Figure 3-12). Flex1 (AT)34 fragility was not affected in a *top1* mutant. This result indicates that either transcription is not important in causing Flex1 fragility or that another topoisomerase is able to compensate for the loss of Top1 in the mutant.

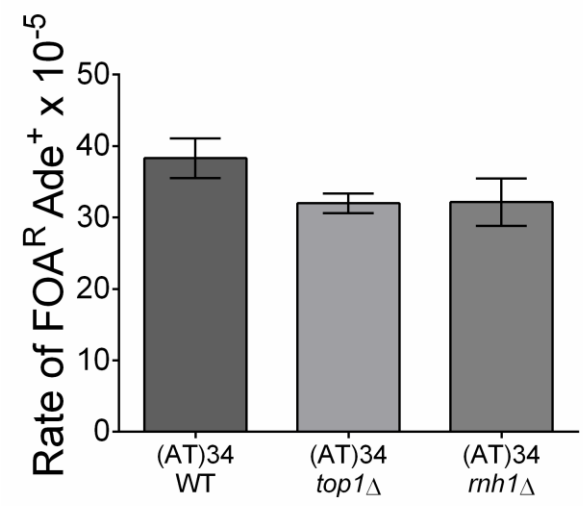


Figure 3-12. Mutants investigating the role of transcription and R-loops in Flex1 fragility. Preliminary data may indicate that transcription and R-loops do not play a role in fragility at Flex1. See Kaushal Thesis Chapter 3 Addendum Table 3-6 for raw data.

Investigating the role of RNaseH and R-loops in Flex1 fragility

Author contributions: *RNH1* and *RNH201* mutants were created by rotation student Alexandra (Sasha) Khristich and preliminary data was obtained by Simran Kaushal. *rnh1*Δ DDRA data was obtained by rotation student Ruby Ye.

The role of R-loops in Flex1 fragility was tested by deleting both R-loop cleaving enzymes in *S. cerevisiae*, *RNH1* and *RNH201*, which encode RNaseH1 and RNaseH201, respectively (Zimmer and Koshland, 2016). With the deletion of these genes, R-loops should persist at Flex1. I was never able to get a measurable DDRA rate for Flex1 as I never got any FOA^R colonies, no matter how I altered the plating conditions for the DDRA. These results may imply that removal of R-loops is the sole cause of fragility at Flex1, and that R-loops are actually protective at Flex1 (perhaps by preventing cruciform extrusion, as proposed for Rrm3). There also may be something wrong with the strain backgrounds, although both myself and the rotation student Ruby Ye verified the integrity of several markers of interest in the double mutant strain. We decided to measure the DDRA rate of *rnh1*Δ mutants (see Figure 3-12), as this is the primary enzyme involve in removing RNA:DNA hybrids, whereas RNaseH201 is primarily needed for ribonucleotide removal from DNA. Since Flex1 fragility did not change in strains lacking Rnh1, it seems unlikely that R-loop removal causes fragility at Flex1.

Investigating the role of transcription in Flex1 fragility: Discussion

Our RT-qPCR may indicate that transcription and cruciform formation can explain the higher fragility rates seen in the DDRA system versus the YAC. However, data from *TOP1* and *RNH1* mutants may indicate that transcription and R-loops do not play a role in fragility at Flex1, and it is possible these findings could be extended to CFSs in general. In the future, inducing transcription through Flex1 should give us a definitive answer as to the role of transcription in Flex1 fragility.

Recent publications have connected transcription to replication origin usage. The Debatisse lab found that large genes that are highly transcribed switch their replication pattern from late to mid-S phase, likely to give the cells more time to complete synthesis of the DNA before M phase (Blin et al., 2018 BioRxiv.org https://doi.org/10.1101/286807). Replication stress also results in a redistribution of replication termination relative to transcription (Chen et al., 2018 BioRxiv.org https://doi.org/10.1101/324079). Further, cells exposed to oncogene-induced replication stress have the inappropriate activation of intergenic origins, leading to replication fork and collapse (Macheret and Halazonetis, 2018). As previously mentioned, the Schildkraut lab found evidence of FANCD2 resolution of RNA:DNA hybrids to allow the activation of dormant origins (Madireddy et al., 2016). Perhaps our data better support the transcription-replication origin model

for CFS fragility. Our system is also readily adaptable to study the role of transcription in fragility of other sequences of interest as well, such as Flex5.

Methods

Genetic mutants

Genetic mutants were created and confirmed as described in the Methods of Chapter 2. *mms4-np* strain was created by PCR amplification of the mms4-np-HIS3 fragment from the pMG2 plasmid, followed by integration into the genome. Nonphosphorylation mutations were confirmed by PCR and sequencing.

Direct Duplication Recombination Assays

Assays were performed as described in the Methods of Chapter 2, with the following changes:

- Colony suspensions from *mrc1AQ* DDRAs were plated on YC-Leu to determine of the *mrc1AQ* plasmid was maintained in the general population during the DDRA conditions.
- I-SceI *sae2*Δ DDRAs were performed as I-SceI DDRAs were performed, except that YEPD media was supplemented with either 2% glucose (no break induction conditions) or 2% galactose (100% break induction conditions).

Chromatin Immunoprecipitation

Cells from strain #3297 were grown to saturation overnight in YC-Ura-Leu media. Cells were diluted to 0.2OD in 200 mL YC-Ura-Leu media and grown at 30C until reaching OD 0.4. Cells were spun down and resuspended in YEPD media with 1µM alpha factor and incubated until cells were synchronized in G_1 as visualized by microscopy (1.5 hours maximum). Alpha factor was washed off twice and cells were released into YEPD. Time points were taken starting at 0 minutes and then every 20 minutes up until 60 minutes into S phase. At each time point, 45 mL of culture was cross-linked in 1% final concentration of formaldehyde for 20 minutes. Cross-linking was quenched by adding 2.5 M glycine. Time point cross-linked culture aliquots were kept on ice until the 60 minute time point was completed. Cells were washed twice with ice-cold 1x TBS and once with ice-cold FA lysis buffer. Cells were resuspended in 1mL ice-cold FA lysis buffer/2mM PMSF (added fresh). Cells were lysed for 3 minutes with a mini bead beater at maximum speed. Samples were transferred to microcentrifuge tubes and spun down at 4C at maximum speed for 15 minutes. The supernatant was discarded at and resuspended in ice-cold lysis buffer. Samples were sonicated for 10 s at 4C using a continuous pulse alternating with 10 s incubation on ice. Phospho 129S antibody was added to the chromatin sample and incubated overnight at 4C with gentle agitation. Protein A or G sepharose beads were added to the chromatin-antibody sample and incubated for 2 hours at 4C with gentle agitation. Beads were washed twice with FA lysis buffer plus 0.5M NaCl, twice with wash buffer, and once with 1x TE. ChIP elution buffer was added to the

beads, followed by a 65C incubation for 25 minutes. Crosslinking was reversed by adding 20 uL 20 mg/mL Pronase in TBS. DNA was isolated by phenol chloroform extraction and qPCR was performed. qPCR was performed (Roche SYBR Green Master Mix with Rox kit).

RT-qPCR

DNA was isolated from log phase cultures grown in Yeast Complete medium for chromosome II strains and YC-Leu-Ura media for YAC strains, using random hexamers for RT-PCR. The locations of the qPCR primer pairs used are indicated by the blue bars in Figure 3-11 (POWER SYBR Green Master Mix, Thermo Scientific). Primers 1254 (5'aactgttgggaagggcgatc 3') and 1255 (5'tgagtcgtattacaattcactggc 3') were used for qPCR.

References

BARNES, R. P., HILE, S. E., LEE, M. Y. & ECKERT, K. A. 2017. DNA polymerases eta and kappa exchange with the polymerase delta holoenzyme to complete common fragile site synthesis. *DNA Repair (Amst),* 57**,** 1-11.

BERGOGLIO, V., BOYER, A. S., WALSH, E., NAIM, V., LEGUBE, G., LEE, M. Y., REY, L., ROSSELLI, F., CAZAUX, C., ECKERT, K. A. & HOFFMANN, J. S. 2013. DNA synthesis by Pol eta promotes fragile site stability by preventing under-replicated DNA in mitosis. *J Cell Biol,* 201**,** 395-408.

BJERGBAEK, L., COBB, J. A., TSAI-PFLUGFELDER, M. & GASSER, S. M. 2005. Mechanistically distinct roles for Sgs1p in checkpoint activation and replication fork maintenance. *EMBO J,* 24**,** 405-17.

BLANCO, M. G. & MATOS, J. 2015. Hold your horSSEs: controlling structure-selective endonucleases MUS81 and Yen1/GEN1. *Front Genet,* 6**,** 253.

CHAUDHURY, I. & KOEPP, D. M. 2017. Degradation of Mrc1 promotes recombination-mediated restart of stalled replication forks. *Nucleic Acids Res,* 45**,** 2558-2570.

DAYN, A., MALKHOSYAN, S. & MIRKIN, S. M. 1992. Transcriptionally driven cruciform formation in vivo. *Nucleic Acids Res,* 20**,** 5991-7.

ESCALANTE-CHONG, R., SAVIR, Y., CARROLL, S. M., INGRAHAM, J. B., WANG, J., MARX, C. J. & SPRINGER, M. 2015. Galactose metabolic genes in yeast respond to a ratio of galactose and glucose. *Proc Natl Acad Sci U S A,* 112**,** 1636-41.

GALLO-FERNANDEZ, M., SAUGAR, I., ORTIZ-BAZAN, M. A., VAZQUEZ, M. V. & TERCERO, J. A. 2012. Cell cycle-dependent regulation of the nuclease activity of Mus81-Eme1/Mms4. *Nucleic Acids Res,* 40**,** 8325-35.

GAMBUS, A., VAN DEURSEN, F., POLYCHRONOPOULOS, D., FOLTMAN, M., JONES, R. C., EDMONDSON, R. D., CALZADA, A. & LABIB, K. 2009. A key role for Ctf4 in coupling the MCM2-7 helicase to DNA polymerase alpha within the eukaryotic replisome. *EMBO J,* 28**,** 2992-3004.

GARCIA-MUSE, T. & AGUILERA, A. 2016. Transcription-replication conflicts: how they occur and how they are resolved. *Nat Rev Mol Cell Biol,* 17**,** 553-63.

GLOVER, T. W., WILSON, T. E. & ARLT, M. F. 2017. Fragile sites in cancer: more than meets the eye. *Nat Rev Cancer,* 17**,** 489-501.

HARACSKA, L., JOHNSON, R. E., UNK, I., PHILLIPS, B., HURWITZ, J., PRAKASH, L. & PRAKASH, S. 2001. Physical and functional interactions of human DNA polymerase eta with PCNA. *Mol Cell Biol,* 21**,** 7199-206.

HELMRICH, A., BALLARINO, M. & TORA, L. 2011. Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. *Mol Cell,* 44**,** 966-77.

KATOU, Y., KANOH, Y., BANDO, M., NOGUCHI, H., TANAKA, H., ASHIKARI, T., SUGIMOTO, K. & SHIRAHIGE, K. 2003. S-phase checkpoint proteins Tof1 and Mrc1 form a stable replication-pausing complex. *Nature,* 424**,** 1078-83.

LE TALLEC, B., MILLOT, G. A., BLIN, M. E., BRISON, O., DUTRILLAUX, B. & DEBATISSE, M. 2013. Common fragile site profiling in epithelial and erythroid cells reveals that most recurrent cancer deletions lie in fragile sites hosting large genes. *Cell Rep,* 4**,** 420-8.

LOU, H., KOMATA, M., KATOU, Y., GUAN, Z., REIS, C. C., BUDD, M., SHIRAHIGE, K. & CAMPBELL, J. L. 2008. Mrc1 and DNA polymerase epsilon function together in linking DNA replication and the S phase checkpoint. *Mol Cell,* 32**,** 106-17.

MACHERET, M. & HALAZONETIS, T. D. 2018. Intragenic origins due to short G1 phases underlie oncogene-induced DNA replication stress. *Nature,* 555**,** 112-116.
MADIREDDY, A., KOSIYATRAKUL, S. T., BOISVERT, R. A., HERRERA-MOYANO, E., GARCIA-RUBIO, M. L., GERHARDT, J., VUONO, E. A., OWEN, N., YAN, Z., OLSON, S., AGUILERA, A., HOWLETT, N. G. & SCHILDKRAUT, C. L. 2016. FANCD2 Facilitates Replication through Common Fragile Sites. *Mol Cell,* 64**,** 388-404.
MASAI, H., YANG, C. C. & MATSUMOTO, S. 2017. Mrc1/Claspin: a new role for regulation of origin firing. *Curr Genet,* 63**,** 813-818.
MOHANTY, B. K., BAIRWA, N. K. & BASTIA, D. 2006. The Tof1p-Csm3p protein complex counteracts the Rrm3p helicase to control replication termination of Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A,* 103**,** 897-902.
OSBORN, A. J. & ELLEDGE, S. J. 2003. Mrc1 is a replication fork component whose phosphorylation in response to DNA replication stress activates Rad53. *Genes Dev,* 17**,** 1755-67.
PARDO, B., CRABBE, L. & PASERO, P. 2017. Signaling pathways of replication stress in yeast. *FEMS Yeast Res,* 17.
PIKE, B. L., TENIS, N. & HEIERHORST, J. 2004. Rad53 kinase activation-independent replication checkpoint function of the N-terminal forkhead-associated (FHA1) domain. *J Biol Chem,* 279**,** 39636-44.
PRINCZ, L. N., WILD, P., BITTMANN, J., AGUADO, F. J., BLANCO, M. G., MATOS, J. & PFANDER, B. 2017. Dbf4-dependent kinase and the Rtt107 scaffold promote Mus81-Mms4 resolvase activation during mitosis. *EMBO J,* 36**,** 664-678.
RAGHUNANDAN, M., CHAUDHURY, I., KELICH, S. L., HANENBERG, H. & SOBECK, A. 2015. FANCD2, FANCJ and BRCA2 cooperate to promote replication fork recovery independently of the Fanconi Anemia core complex. *Cell Cycle,* 14**,** 342-53.
SAUGAR, I., VAZQUEZ, M. V., GALLO-FERNANDEZ, M., ORTIZ-BAZAN, M. A., SEGURADO, M., CALZADA, A. & TERCERO, J. A. 2013. Temporal regulation of the Mus81-Mms4 endonuclease ensures cell survival under conditions of DNA damage. *Nucleic Acids Res,* 41**,** 8943-58.
UZUNOVA, S. D., ZARKOV, A. S., IVANOVA, A. M., STOYNOV, S. S. & NEDELCHEVA-VELEVA, M. N. 2014. The subunits of the S-phase checkpoint complex Mrc1/Tof1/Csm3: dynamics and interdependence. *Cell Div,* 9**,** 4.
WANG, J. C. 2002. Cellular roles of DNA topoisomerases: a molecular perspective. *Nat Rev Mol Cell Biol,* 3**,** 430-40.
WILD, P. & MATOS, J. 2016. Cell cycle control of DNA joint molecule resolution. *Curr Opin Cell Biol,* 40**,** 74-80.
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. *Mol Cell,* 27**,** 367-79.
ZIMMER, A. D. & KOSHLAND, D. 2016. Differential roles of the RNases H in preventing chromosome instability. *Proc Natl Acad Sci U S A,* 113**,** 12220-12225.

Kaushal Thesis Chapter 3 Addendum

Data in colors was generated by other contributors as noted in text.

Table 3-1. Large FRA16D YAC % FOA^R of all strains. * = p<0.05 compared to WT same tract ** = p<0.01; *** = p<0.001; **** = p<0.0001

YAC strain	Individual Experiments						Average	SEM	p value	p compared to	p	p value	p compared to	p
972D3	3.7	4.6	4				4.1	0.265						
801B6	13.3	24.1	18.8	17.56	17.6	17.2	18.1	1.424	0.0003	972D3	***			
801B6 Flex1Δ	12.55	11.9	13.78	12.33			12.6	0.403	<0.0001	972D3	****	0.01663	801B6	^
801B6 Flex5Δ	6.9	1.2	1.8	14	10.9		7.0	2.499	0.4244	972D3	**	0.0029	801B6	^^
801B6 P5ΔP5bΔ	16.7	16					16.4	0.350	<0.0001	972D3	****	0.5275	801B6	

Table 3-2. Replication fork stabilizer Flex1 and ctrl DDRA data. Rate of FOA^R x 10^{-6} is shown.

Strain	Individual assays							Average	SEM	p value	p compared to	p	p value	p compared to	p
ctrl	2.9	3.8	1.9	3.7	3	3.5		3.1	0.288						
ctrl	37	44	55					45.3	5.239	<0.0001	ctrl	****			

mrc1Δ															
ctrl tof1Δ	18.1	42.7	22.7					27.8	7.558	0.0016	ctrl	**			
ctrl mrc1AQ	17	10						13.5	3.500	0.0010	ctrl	***	0.0219	ctrl mrc1Δ	^
Flex1 (AT)34	45	39.7	29.5	45.5	45.5	30	33	38.3	2.782	<0.0001					
Flex1 (AT)34 mrc1Δ	100	97	229	126				138.0	31.024	0.0018	Flex1 (AT)34	**			
Flex1 (AT)34 tof1Δ	47	43	38					42.7	2.603	0.3781	Flex1 (AT)34				
Flex1 (AT)34 mrc1AQ	63	14						38.5	24.500	0.9879	Flex1 (AT)34				
Flex1 (AT)34 ctf4Δ	71	78	58					69.0	5.860	0.0006	Flex1 (AT)34	***			
Flex1 (AT)34 rad30Δ	29	35	37	47	31			35.8	3.137	0.5654	Flex1 (AT)34				

Table 3-3. Replication fork stabilizer full knockout (CAG)n YAC data. Rate of FOA^R x 10^{-6} is shown.

* = p<0.05 compared to WT same tract ** = p<0.01; *** = p<0.001; **** = p<0.0001.

Strain	Individual assays							Average	SEM	p value	p compared to WT same tract length
CAG-0											
WT	2.2	2.6	4.4	4.7	2.3	0.97		2.86	0.582		
tof1	4.09	5.96	2.3					4.12	1.057	0.2902	
mrc1	9.71	12.37	19.72					13.93	2.993	0.0013	**
CAG-70											
WT	8.7	8.7	10.4					9.27	0.567		
tof1	7.64	7.57	5.14	6.2				6.64	0.599	0.02739	
mrc1	66.26	167.48	69.53					101.09	33.208	0.0506	*
CAG-155											
WT	32	24	28					28.00	2.309		
tof1	183	94	141	103.3	87.6	84.1		121.99	14.886	0.0040	**
mrc1	156	111	100					122.35	17.134	0.005480071	**

Table 3-4. (CAG)n Replication fork stabilizer mutant data. Rate of FOA^R x 10^{-6} is shown. ** = $p<0.01$

Strain	Individual assays							Average	SEM	p value compared to WT same tract length	p compared to WT same tract length	p compared to no tract same genotype	p compared to no tract same genotype
CAG-0													
WT	1.8	3.9	1.4	3.5	1.1	2.0	2.6	2.3257	0.394				

mrc1-1	4.8	11.2	6.4					7.4633	1.902	0.00414571	**		
mrc1AQ	5.9	6.7	9.6	6.9				7.2825	0.803	0.000143438	**		
rad53-21	7.4	5.7	4.0	7.3	9.2			6.7060	0.868	0.000467229	**		
CAG-85													
WT	4	7.5	5.4	3.3	5.65	4.06	6.28	5.15143	0.554			0.0013	**
mrc1-1	12.5	12.2	11.7	17.0				13.4475	1.339	8.18024E-05	**	0.0446	*
mrc1AQ	*7*	9.2	7	8.6				7.945	0.560	0.0097	**	0.5240	
rad53-21	22.7	21.2	15.3	21.2	16.5			19.352	1.458	1.21024E-06	**	7.2433E-05	****
CAG-155													
WT	**6.1**	**7.5**	**5.8**	14.2	16			9.92	2.153			0.0021	***
mrc1-1	34.3	30.5	43.76					36.2067	3.931	0.0006	**	0.0028	**
mrc1AQ (actually CAG-145)	9.8	16.4	7.7	23.9	31.7			17.8940	4.465	0.1464		0.0766	
rad53-21 (actually CAG-135)	24.5	35.7	36.9	25	45.5	43	29.6	34.3171	3.137	0.000160021	**	2.90975E-05	****

Table 3-5. All CAG tract length *tof1*Δ YAC fragility data. Rate of FOA^R x 10^{-6} is shown.

***tof1*Δ**	**Bkgd**	**Individual assays**	**Average**	**SEM**

strain tract										
CAG-0	BY4705	4.09	5.96	2.3					4.1	1.057
CAG-70	BY4705	7.64	7.57	5.14	6.2				6.6	0.599
CAG-85	W303	13.5	9.9	14.7					12.7	1.442
CAG-110	BY4705	63.8	48.8	107.3					73.3	17.543
CAG-135	W303	196.3	80.1	76.9	112.5				116.5	27.796
CAG-155	BY4705	182.9	93.86	141.04	160.39	103.3	87.6	84.1	121.9	14.886

Table 3-6. Flex1 Mutant DDRA data. Rate of FOA^R Ade+ x 10^{-5} is shown.

***** = p<0.05 compared to WT same tract ** = p<0.01; *** = p<0.001; **** = p<0.0001.

Strain	Individual assays							Average	SEM	p value	p compared to	p
Flex1 (AT)34 *mms4-np*	32.2	31.3	29.7	40.3	49.1			36.5	3.640	0.6983	Flex1 (AT)34	
Flex1 (AT)34 top1Δ	31.4	35.1	28.6	32.9				32.0	1.364	0.1393	Flex1 (AT)34	
Flex1 (AT)34 rnh1Δ	25.1	24.2	37.1	44.6	26.6	35.3		32.2	3.332	0.1798	Flex1 (AT)34	

Table 3-7. Flex1 Raw γH2AX ChIP data from 6-2-14.

	Sample name	Ct	Mean Ct						
F1 upstream primers	3297 T0 IN 1:500	21.4113	21.4298		Sample	ΔCt	%Input	ΔΔCt	Fold enrichment
	3297 T0 IN 1:500	21.4483			S129P (AT)34 0min	-0.46497	0.724485	0.57605	1.490762
	3297 T20 IN 1:500	21.6754	21.64175		S129P (AT)34 20min	-0.33612	0.792168	0.70375	1.628733
	3297 T20 IN 1:500	21.6081			S129P (AT)34 40min	0.102878	1.073914	0.67685	1.598645
	3297 T40 IN 1:500	21.2535	21.26075		S129P (AT)34 60min	0.821528	1.767277	0.545	1.45902
	3297 T40 IN 1:500	21.268							
	3297 T60 IN 1:500	21.163	21.17125						
	3297 T60 IN 1:500	21.1795							
	3297 T0 IP 1:50	17.6031	17.6429						
	3297 T0 IP 1:50	17.6827							
	3297 T20 IP 1:50	17.9682	17.9837						
	3297 T20 IP 1:50	17.9992							
	3297 T40 IP	17.9382	18.0417						

	1:50								
	3297 T40 IP 1:50	18.1452							
	3297 T60 IP 1:50	18.639	18.67085						
	3297 T60 IP 1:50	18.7027							
F1 downstream	3297 T0 IN 1:500	20.8406	20.9237		**Sample**	**ΔCt**	**%Input**	**ΔΔCt**	**Fold enrichment**
	3297 T0 IN 1:500	21.0068			S129P (AT)34 0min	-0.08817	0.940714	-0.95285	0.516611
	3297 T20 IN 1:500	21.1507	21.13445		S129P (AT)34 20min	-0.01172	0.991908	-1.02815	0.490339
	3297 T20 IN 1:500	21.1182			S129P (AT)34 40min	0.466028	1.381301	-1.04	0.486327
	3297 T40 IN 1:500	20.6059	20.61455		S129P (AT)34 60min	1.258628	2.392681	-0.9821	0.506242
	3297 T40 IN 1:500	20.6232							
	3297 T60 IN 1:500	20.3609	20.3948						
	3297 T60 IN 1:500	20.4287							
	3297 T0 IP 1:50	17.4944	17.5136						
	3297 T0 IP 1:50	17.5328							

	3297 T20 IP 1:50	17.7911	17.8008						
	3297 T20 IP 1:50	17.8105							
	3297 T40 IP 1:50	17.5503	17.5389						
	3297 T40 IP 1:50	17.5275							
	3297 T60 IP 1:50	18.3101	18.3315						
	3297 T60 IP 1:50	18.3529							
ACT1 primers	3297 T0 IN 1:500	21.6523	21.7253		**Sample**	**ΔCt**	**%Input**		
	3297 T0 IN 1:500	21.7983			S129P (AT)34 0min	-1.04102	0.485983		
	3297 T20 IN 1:500	22.0022	22.0244		S129P (AT)34 20min	-1.03987	0.486371		
	3297 T20 IN 1:500	22.0466			S129P (AT)34 40min	-0.57397	0.671765		
	3297 T40 IN 1:500	21.4565	21.41925		S129P (AT)34 60min	0.276528	1.211276		
	3297 T40 IN 1:500	21.382							
	3297 T60 IN 1:500	21.3454	21.40535						
	3297 T60 IN	21.4653							

	1:500								
	3297 T0 IP 1:50	17.3657	17.36235						
	3297 T0 IP 1:50	17.359							
	3297 T20 IP 1:50	17.6294	17.6626						
	3297 T20 IP 1:50	17.6958							
	3297 T40 IP 1:50	17.5622	17.52335						
	3297 T40 IP 1:50	17.4845							
	3297 T60 IP 1:50	18.3721	18.35995						
	3297 T60 IP 1:50	18.3478							

Table 3-8. RT-qPCR data.

	Mean	**SEM**	**N**
chrII	0.108	0.021	5
YAC	0.047	0.007	5

Table 3-9. Plasmid Strains from Kaushal Thesis Chapter 3.

Plasmid	CF Plasmid stock#	Description	Source
pAO139 *mrc1AQ LEU2*	614	A^{pr} *LEU2 mrc1*^{AQ}	(Osborn and Elledge, 2003)
pMG2		Template for *mms4-np::HISMX*	(Gallo-Fernandez et al., 2012)

Table 3-10. Yeast Strains from Kaushal Thesis Chapter 3.

Strain Name	Background	CFY #	Genotype	Source
801B6 YAC Flex5Δ	AB1380, CFY#1086	3513, 3514, 3587, 3588	*MATa, ura3-52, his5, trp1-289, lys2-1, can1-100, ade2-1* Flex5*::HISMX* YAC*: LEU2 C_4A_4 URA3 TRP1*	this study
801B6 YAC P5P5bΔ	AB1380, CFY#1086	3517	*MATa, ura3-52, his5, trp1-289, lys2-1, can1-100, ade2-1* P5P5b*::hyg* YAC*: LEU2 C_4A_4 URA3 TRP1*	this study
ctrl (DDRA)	YPH499, CFY#2268	2863/2864	*MATa, leu2-Δ1, ura3-52, his3-Δ200, trp1-Δ63, ade2Δ::hisG (salmonella), lys2::ADE2*::*URA3*-no repeat control	this study
ctrl mrc1Δ (DDRA)	CFY #2864	3721/3722	*mrc1::HIS5(S. pombe)*	this study
ctrl tof1Δ (DDRA)	CFY#2864	4078	*tof1::KANMX6*	this study
ctrl rrm3Δ (DDRA)	CFY#2863	4590/4591	*rrm3::KANMX*	
Flex1 (AT)34 (DDRA)	CFY#2268	2525/2712	*MATa, leu2-Δ1, ura3-52, his3-Δ200, trp1-Δ63, ade2Δ::hisG (salmonella), lys2::ADE2*::*URA3*-Flex1(AT)34	this study
Flex1 (AT)34	CFY#2525	3212/3629	*mrc1::HIS5(S. pombe)*	this study

mrc1Δ (DDRA)				
Flex1 (AT)34 tof1Δ (DDRA)	CFY#2525	3213/3214	*tof1::KANMX6*	this study
Flex1 (AT)34 ctf4Δ (DDRA)	CFY#2525	4157/4158	*ctf4::KANMX*	this study
Flex1 (AT)34 rad30Δ (DDRA)	CFY#2525	4159/4160	*rad30::KANMX*	this study
Flex1 (AT)34 srs2Δ (DDRA)	CFY#2525	3950/3951	*srs2::KANMX*	this study
Flex1 (AT)34 srs2Δ (YAC)	CFY#3457	4558/4559	*srs2::TRP1*	this study
Flex1 (AT)34 sgs1Δ (YAC)	CFY#4521	4521, 4537, 4592	*sgs1::KANMX*	this study
Flex1 (AT)34 mph1Δ (YAC)	CFY#3457	4538, 4539	*mph1::KANMX*	this study
Flex1 (AT)34 *mms4-np* (DDRA)	CFY# 2525	4586, 4587	*mms4Δ::mms4-np* (S54A; S55A; S183A; S200A; S220A; S221A; S300A; T301A; S402A)-*HIS3*	this study
Flex1 (AT)34 rad52Δ (YAC)	CFY#3457	4599/4600	*rad52::NATMX*	this study
Flex1 (AT)34 mus81Δ rad52Δ (YAC)	CFY#4284	4601	*mus81::KANMX*	this study
Flex1 (AT)34 top1Δ (DDRA)	CFY#2525	4330/4331	*top1::TRP1*	this study
Flex1 (AT)34 rnh1Δ (DDRA)	CFY#2525	4244/4245	*rnh1::HIS3*	this study
S5'(AT)34L3' o1	CFY#3219	3297	BY4705 + YAC: *LEU2, Flex1, HIS3, URA3*	this study

bar1::KANMX				
I-SceI (2 I-SceI sites)	CFY#3518	3570/3585	*ILV1::pGAL*-I-SceI nuclease, *lys2::ADE2*::*URA3*-I-SceI only	this study
I-SceI sae2Δ	CFY#3750	4100/4101	*sae2::KANMX6*	this study
I-SceI-S3'	CFY#3518	3989/3990	*lys2::ADE2*::*URA3*-S5'-I-SceI-S3'	this study
I-SceI-S3' sae2Δ	CFY#3989	4102/4103	*sae2::KANMX6*	this study
I-SceI-L3'	CFY#3519	3991/3992	*lys2::ADE2*::*URA3*-S5'-I-SceI-L3'	this study
I-SceI-L3' sae2Δ	CFY#3991	4107/4108	*sae2::KANMX6*	this study
WT CAG-0	BY4705	765	*Mat α, ade2Δ::hisG, his3Δ200, leu2Δ0, lys2Δ0, met15Δ0, trp1Δ63, ura3Δ0, can^R;* YAC: *LEU2, URA3*, CAG-0	(Sundararajan et al. 2009)
CAG-0 tof1Δ	CFY#765	2610	*tof1::HIS3*	Gellon et al., in prep
CAG-0 mrc1Δ	CFY#765	2287/2288	*mrc1::HIS3*	Gellon et al., in prep
CAG-0 mrc1AQ	W303	1104/1105	*Mat a, ade2-1, his3-11,15, leu2-3,112, trp1-1, ura3-1, can1-100 HIS::mrc1AQMYC13;* YAC: *LEU2, URA3*, CAG-0	(Osborn and Elledge, 2003), Gellon et al., in prep
CAG-0 mrc1-1	W303, Y300	947-949	*Mat a, ade2-1, his3-11,15, leu2-3,112, trp1-1, ura3-1, can1-100, mrc1-1;* YAC: *LEU2, URA3*, CAG-0	(Freudenreich and Lahiri, 2004)
CAG-0 rad53Δ	W303, Y300	815/816	*rad53-21*	Gellon et al., in prep
WT CAG-70	BY4705	766	YAC: *LEU2, URA3*, CAG-70	(Sundararajan et al. 2009)
CAG-70 tof1Δ	CFY#766	2615	*tof1::HIS3*	Gellon et al., in prep
CAG-70 mrc1Δ	CFY#766		*mrc1::HIS3*	(Osborn and Elledge, 2003), Gellon et al., in prep
CAG-85 mrc1AQ	W303	1195/1196	*HIS::mrc1AQMYC13*	Gellon et al., in prep
CAG-85 mrc1-1	W303, Y300	950-952	*mrc1-1*	(Freudenreich and Lahiri, 2004)

CAG-85 rad53Δ	W303, Y300		*rad53-21*	Gellon et al., in prep
CAG-110 tof1Δ	CFY#765	2614	*tof1::HIS3*	Gellon et al., in prep
CAG-135 tof1Δ	W303	1214	*tof1::HIS3*	Gellon et al., in prep
WT CAG-155	BY4705	767	YAC: *LEU2*, *URA3*, CAG-155	(Sundararajan et al. 2009)
CAG-155 tof1Δ	CFY#767	2590/2617	*tof1::HIS3*	Gellon et al., in prep
CAG-155 mrc1Δ	CFY#767	2591	*mrc1::HIS3*	Gellon e al., in prep
CAG-145 mrc1AQ	W303	1107	*HIS::mrc1AQMYC13*	(Osborn and Elledge, 2003), Gellon et al., in prep
CAG-155 mrc1-1	W303, Y300	953-955	*mrc1-1*	(Freudenreich and Lahiri, 2004)
CAG-155 rad53Δ	W303, Y300		*rad53-21*	Gellon et al., in prep

Chapter 4

Causes and consequences of common fragile site expression at FRA16D: perspectives and future directions

Abstract

Flex1 has proven to be an excellent model system for breakage and healing of structure-forming CFS sequences in *S. cerevisiae*. We have begun investigating the relevance of most proposed CFS hypotheses on this sequence's fragility and healing. Most importantly, we have proposed a new theory in the CFS field: that CFSs may have difficulty healing after fragility, resulting in their expression phenotype. Our genetic systems are readily adaptable to study other CFS subregions of interest from FRA16D and other CFSs.

Secondary structures at Flex1 and CFSs and replication impairment

Before the Kaushal et al. 2018, submitted, study (Kaushal Thesis Chapter 2), there were connections between secondary structure forming sequences and fragility at rare fragile sites but not CFSs. Our results support previous evidence of *in vivo* structure formation at the Flex1 subregion of CFS FRA16D, and we have found that fragility at this structure is likely initiated by an Slx1-Slx4-Mus81-Mms4-Rad1-Rad10 (SMR) DNA repair super complex. The AT length polymorphism at Flex1 seems highly relevant to structure formation and CFS fragility. Our data indicate that Flex1 is likely forming a cruciform structure, although it could also be forming slipped-strand hairpins on both strands. The SMR complex can either be targeting secondary structures for direct cleavage or replication forks stalled at secondary structures. Since Mus81 becomes more active throughout the cell cycle

and peaks at G_2/M phase, it is likely that the SMR complex is targeting structures or stalled forks at Flex1 during those phases. However, our *mms4-np* data may imply that Mus81 (and thus possibly SMR complex) activity in S phase is sufficient to cause Flex1 fragility. Our discovery that Yen1 protects against Flex1 fragility is consistent with its known peak activity in M phase, as Mus81 activity drops.

The knowledge gained from my research has two important translational consequences. The first is that it points to a region of FRA16D that is specifically susceptible to Mus81 cleavage in *S. cerevisiae* (and likely hMUS81 cleavage as well), which adds a direct target to the Hickson lab's model for FRA16D fragility. Our findings also indicate that individuals with more than 22 ATs at Flex1 may be more susceptible to fragility at FRA16D and therefore the subregion could serve as a possible therapeutic target.

Comparing the DDRA rates of Flex1 (AT)23 to Flex5 (AT)24i may indicate that interruptions in perfect dinucleotide AT repeats drastically reduce their secondary structure forming capability and consequential fragility. Interestingly, interruptions in rare fragile site secondary structures can lessen the disease symptoms (Thys et al., 2015, Pearson et al., 1998, Weisman-Shomer et al., 2000, Jarem et al., 2010). Thus, our works may also point to perfect repeat subregions of CFSs as especially difficult for the cell to faithfully replicate, and therefore the most important therapeutic and structure-forming regions responsible for CFS fragility.

It has been proposed that upon decoupling of the replicative polymerases and helicase can allow for stretches of ssDNA to accumulate and fold into secondary structures, resulting in fragility and characteristic CFS expression. We found that Mrc1 plays an important role in preventing fragility at control, Flex1, and CAG repeat sequences. Further, by testing the effect of an *mrc1AQ* mutant we determined that it is likely the fork stabilization and not checkpoint function of Mrc1 that makes it so crucial in preventing fragility, which supports previous theories of CFS structure formation. Fork stabilization on both the leading and lagging strands seems important for avoiding fragility at Flex1, as both *mrc1*Δ and *ctf4*Δ mutants showed an increase in Flex1 recombination rates in yeast. This implies that secondary structures are forming and impeding replication on both Flex1 strands, which supports that a cruciform may be forming at that location (though it does not rule out the possibility of slipped-strand DNA structures at the locus as well). The Tof1 fork stabilizer is important at long CAG tract lengths and thus seems important after the formation of a certain type or length of secondary structure. When tested in the Flex1 DDRA, Tof1 protected against fragility at a control sequence but not Flex1, however these results must be recapitulated in freshly made strains. Preliminarily, it is possible that Tof1 is protective specifically at structures formed by CAG repeats versus AT repeats, which may provide additional evidence for the formation of a cruciform rather than a hairpin structure forming at the Flex1 (AT)34 repeat. *In vitro* data investigating replication through FRA16D indicated the importance of TLS polymerases in synthesizing through structure-forming repeats (Barnes et al., 2017), but we have

not seen an importance at Flex1 of the one TLS polymerase investigated thus far, Rad30. Thus, replication through CFSs is a dynamic and complex process that may be variable across CFSs due to the structure-forming potential of different subregions.

Fragility at Flex1

Flex1 is predicted to form a stable cruciform when AT lengths exceed 22 bp and the Flex1 (AT)-flanking DNA also is AT-rich and is predicted to form structures. These secondary structures may form during transcription or possibly during periods of single-strandedness during replication. The secondary structures can then block polymerase progression, resulting in a stalled fork and incomplete DNA replication. Our data provide evidence for the existence of a Slx1-Slx4-Mus81-Mms4-Rad1-Rad10 complex that responds to secondary structures and/or fork stalling at Flex1. This is consistent with *in vitro* evidence supporting the formation of an SLX1-SLX4, MUS81-EME1, and XPF-ERCC1 super complex (SMX) DNA repair trinuclease in mammals (Wyatt et al., 2017) and the requirement of MUS81, ERCC1 (*S. cerevisiae* Rad1) and SLX4 for CFS fragility in human cells (Naim et al., 2013, Ying et al., 2013, Minocherhomji et al., 2015). Thus, Flex1 could serve as a very important region of fork stalling and nuclease cleavage at FRA16D and could be a causative factor in the breaks and gaps seen at FRA16D on human metaphase chromosomes.

Healing at Flex1 and CFSs

We find evidence for a new theory for the expression of CFSs: that they are prone to fragility and deficient in healing due to the presence of multiple secondary structures. SSE-induced fragility has been proposed to initiate breakage at CFSs to allow for sister chromatids to faithfully separate. However, broken DNA at CFSs leaves them vulnerable to deletions and rearrangements. Thus, CFSs must heal quickly and efficiently in order to prevent deleterious genomic consequences. My work has brought an important new hypothesis to the CFS field that CFSs are deficient in their ability to heal. After the initial fragility, interrupted repeats with hairpin-forming capability could impede healing, titrating Sae2/hCtIP under replication stress conditions, resulting in continued expression of fragility. Previous data from our lab showed that FRA16D-containing YACs have slower growth under normal conditions, and that cell death was greatly exacerbated in the presence of the *rad50*Δ mutation and exposure to HU (Zhang and Freudenreich, 2007). This supports all of our findings that FRA16D is enriched in secondary structure-forming sequences, which especially result in fragility under replication stress and must be healed in order to prevent cell death. Our data suggest that CFSs may initiate breakage at structure-forming AT-rich or perfect AT repeats, and subsequent resection and healing can be affected by the ability of adjacent sequences to form secondary structures. Sae2 is important for healing after fragility at Flex1, consistent with the importance of human CtIP and yeast Sae2 in responding to hairpin-capped DNA ends. In the future, it would be interesting to repeat the I-SceI break *sae2*Δ recombination assays under 50%

induction conditions and from the same batch of media to confirm that Sae2 is not important at clean I-SceI ends in yeast as was found in human cells. Therapeutically, this healing requirement can be exploited in therapy of precancerous cells by inactivating DNA repair proteins such as CtIP, which would result in cell death rather than potentially deleterious- and cancer-progressing-rearrangements at CFSs.

Transcription and Flex1 fragility

RT-qPCR showed that Flex1 is transcribed in both the internal chromosome (DDRA) and YAC assay locations where it was studied; therefore we hypothesized that a secondary structure, specifically a cruciform, was forming during transcription. This hypothesis was further supported by comparing levels of Flex1 transcription in the two systems, which shows that Flex1 is more highly transcribed in the system in which it is most fragile (the DDRA). However, different fragility levels could also be explained by genomic location within the middle of a chromosome (DDRA) versus the end of a chromosome (YAC), or could be explained by different genetic backgrounds of the yeast strains in the two assay systems. I demonstrated that gene replacement of *top1*Δ did not decrease Flex1 DDRA rates, either meaning that transcription is not the causative event in Flex1 fragility or that other topoisomerases are compensating for the loss of Top1. In the future, it would be interesting to induce transcription through Flex1 to definitively determine the effect of transcription on Flex1 fragility. Our data may also support that the transcription effect seen at CFSs is not due to R-loops by rather due to changes in origin usage.

The role of chromatin modifications at CFS fragility

Since CFS expression varies across cell types, epigenetic factors are an attractive candidate for playing a role in their expression. It is well known that chromatin compaction varies across cell type, and this process can be controlled by various histone modifications. Chromatin compaction could impede replication through CFSs or prevent access of repair machinery to CFS DNA, and it could also affect secondary structure formation. Flex1 and the DDRA system could serve as an excellent model system to study how histone modifications and nucleosome dynamics play a role in CFS fragility.

Some studies indicate that TA dinucleotides, due to their high DNA flexibility, kink the minor groove of the DNA to allow for tight wrapping around the nucleosome (Wu et al., 2010). After distamycin exposure, the 33 bp repeated AT-rich sequence from rare fragile site FRA16B excludes nucleosomes. However, it is difficult to determine whether this was due to DNA sequence identity or fragility and degradation of the DNA due to induction of fragility using distamycin (Hsu and Wang, 2002). . In our DDRA, the extra 102 bp present in the L3' but not S3' flanking sequence of Flex1 could play a role in chromatin positioning and dynamics which could in turn affect healing dynamics in the SSA pathway required for healing in the DDRA. It is possible that a secondary structure formed by AT repeats excludes nucleosomes from the region, which could be a factor in their fragility or difficulty repairing after a DSB.

Fork restart after HU-induced checkpoint depends on the chromatin remodeling proteins Ino80 and Isw2 (Shimada et al., 2008, Travesa et al., 2008, House et al.,

2014). In fact, both proteins were shown to increase chromatin access in replicating regions (Lee et al., 2015), making it likely that they could be acting at CFS locations where MiDAS is occurring. Ino80 and Isw2 may be enriched at Flex1 due to its secondary structure and fork stalling capabilities. Interestingly, the Freudenreich lab recently found that the chromatin remodeler Isw1 prevents CAG repeat expansion by correctly depositing and spacing nucleosomes on the DNA after transcription (Koch et al., 2018). We could also test to see if Isw1 is needed for nucleosome positioning around the Flex1 ATs and flanking sequences. However, since AT repeats are not known as strong nucleosome positioning elements the way CAG repeats are, we may not see any fragility change in a Flex1 *isw1*Δ mutant. We can directly test if Flex1 is a nucleosome position or nucleosome refractory sequence by performing MNase assays as in (Koch et al., 2018).

Perspectives

The sequence of Flex1 is remarkably well-conserved between humans and gorillas, including the presence of an (AT)n repeat. If CFSs are vulnerable parts of our genome, why have they persisted throughout evolution? It is possible that breakage at CFSs is not a driving force in tumorigenesis but rather a symptom of replication instability in cancer cells. They may have remained in our genomes because individuals normally do not experience CFS fragility and deleterious consequences until they get cancer later in life, after reproducing. However, it is likely that CFS fragility and rearrangements are a stepping stone in the progression of cancer, and that the deletion of late-replicating fragile sites gives

cancer cells a replicative advantage. CFSs may serve as triggers for genome rearrangements under dire conditions of replication stress, where it is beneficial for the organism to undergo rapid genome rearrangement in hopes of adapting to the stress. CFSs are also present in long genes and tumor suppressors, and their fragility may promote the formation of alternative transcripts under replication stress, which could provide a cellular advantage. Thus, as the picture of secondary structures and fragility at CFSs emerge, it will be interesting to investigate their evolutionary purpose, if there is one.

Overall, the Flex1 serves as an excellent model for CFS fragility and healing requirements. It has corroborated many of the known genetic requirements for CFS fragility. Further, availability of many different assay systems makes the study of Flex1 in *S. cerevisiae* an excellent method to carefully investigate various hypotheses for CFS breakage without the added complications of studying very large entire CFS regions in mammalian cells.

References

BARNES, R. P., HILE, S. E., LEE, M. Y. & ECKERT, K. A. 2017. DNA polymerases eta and kappa exchange with the polymerase delta holoenzyme to complete common fragile site synthesis. *DNA Repair (Amst),* 57**,** 1-11.

HOUSE, N. C., KOCH, M. R. & FREUDENREICH, C. H. 2014. Chromatin modifications and DNA repair: beyond double-strand breaks. *Front Genet,* 5**,** 296.

HSU, Y. Y. & WANG, Y. H. 2002. Human fragile site FRA16B DNA excludes nucleosomes in the presence of distamycin. *J Biol Chem,* 277**,** 17315-9.

JAREM, D. A., HUCKABY, L. V. & DELANEY, S. 2010. AGG interruptions in (CGG)(n) DNA repeat tracts modulate the structure and thermodynamics of non-B conformations in vitro. *Biochemistry,* 49**,** 6826-37.

KOCH, M. R., HOUSE, N. C. M., COSETTA, C. M., JONG, R. M., SALOMON, C. G., JOYCE, C. E., PHILIPS, E. A., SU, X. A. & FREUDENREICH, C. H. 2018. The Chromatin Remodeler Isw1 Prevents CAG Repeat Expansions During Transcription in Saccharomyces cerevisiae. *Genetics,* 208**,** 963-976.

LEE, L., RODRIGUEZ, J. & TSUKIYAMA, T. 2015. Chromatin remodeling factors Isw2 and Ino80 regulate checkpoint activity and chromatin structure in S phase. *Genetics,* 199**,** 1077-91.

MINOCHERHOMJI, S., YING, S., BJERREGAARD, V. A., BURSOMANNO, S., ALELIUNAITE, A., WU, W., MANKOURI, H. W., SHEN, H., LIU, Y. & HICKSON, I. D. 2015. Replication stress activates DNA repair synthesis in mitosis. *Nature,* 528**,** 286-90.

NAIM, V., WILHELM, T., DEBATISSE, M. & ROSSELLI, F. 2013. ERCC1 and MUS81-EME1 promote sister chromatid separation by processing late replication intermediates at common fragile sites during mitosis. *Nat Cell Biol,* 15**,** 1008-15.

PEARSON, C. E., EICHLER, E. E., LORENZETTI, D., KRAMER, S. F., ZOGHBI, H. Y., NELSON, D. L. & SINDEN, R. R. 1998. Interruptions in the triplet repeats of SCA1 and FRAXA reduce the propensity and complexity of slipped strand DNA (S-DNA) formation. *Biochemistry,* 37**,** 2701-8.

SHIMADA, K., OMA, Y., SCHLEKER, T., KUGOU, K., OHTA, K., HARATA, M. & GASSER, S. M. 2008. Ino80 chromatin remodeling complex promotes recovery of stalled replication forks. *Curr Biol,* 18**,** 566-75.

THYS, R. G., LEHMAN, C. E., PIERCE, L. C. & WANG, Y. H. 2015. DNA secondary structure at chromosomal fragile sites in human disease. *Curr Genomics,* 16**,** 60-70.

TRAVESA, A., DUCH, A. & QUINTANA, D. G. 2008. Distinct phosphatases mediate the deactivation of the DNA damage checkpoint kinase Rad53. *J Biol Chem,* 283**,** 17123-30.

WEISMAN-SHOMER, P., COHEN, E. & FRY, M. 2000. Interruption of the fragile X syndrome expanded sequence d(CGG)(n) by interspersed d(AGG) trinucleotides diminishes the formation and stability of d(CGG)(n) tetrahelical structures. *Nucleic Acids Res,* 28**,** 1535-41.

WU, B., MOHIDEEN, K., VASUDEVAN, D. & DAVEY, C. A. 2010. Structural insight into the sequence dependence of nucleosome positioning. *Structure,* 18**,** 528-36.

WYATT, H. D., LAISTER, R. C., MARTIN, S. R., ARROWSMITH, C. H. & WEST, S. C. 2017. The SMX DNA Repair Tri-nuclease. *Mol Cell,* 65**,** 848-860 e11.

YING, S., MINOCHERHOMJI, S., CHAN, K. L., PALMAI-PALLAG, T., CHU, W. K., WASS, T., MANKOURI, H. W., LIU, Y. & HICKSON, I. D. 2013. MUS81 promotes common fragile site expression. *Nat Cell Biol,* 15**,** 1001-7.

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. *Mol Cell,* 27**,** 367-79.