Genome instability induced by triplex forming mirror repeats in *S.cerevisiae*

A Dissertation

Presented to

The Academic Faculty

By

Hyun-Min Kim

In partial fulfillment

of the requirements for the degree

Doctor of Philosophy in the School of Biology

Georgia Institute of Technology

May 2009

Genome instability induced by triplex forming mirror repeats in *S.cerevisiae*

Approved by:

Dr. Kirill Lobachev

Academic advisor

School of Biology

Georgia Institute of Technology

Dr. Nicholas Hud

School of Chemistry and Biochemistry

Georgia Institute of Technology

Dr. Francesca Storici

School of Biology

Georgia Institute of Technology

Dr. Yury Chernoff

School of Biology

Georgia Institute of Technology

Dr. SooJin Yi

School of Biology

Georgia Institute of Technology

Date approved; March 20 2009

Acknowledgements

I would like to extend my earnest appreciation to my academic advisor Dr. Kirill Lobachev for his scientific guidance and ambitious motivation during my PH.D program.

I would like to thank Dr. Natasha Degtyareva for her help with the scientific guide and critical review.

I would like to thank my thesis committee members Dr. Nicholas Hud, Dr. Soojin Yi, Dr. Yuri Chernoff, and Dr. Francesca Storici.

I would like to thank Katy Bruce for the proofreading and Gary Newnam for his help.

I would like to thank all the undergraduate students, especially Shadeah Suleiman and George Laskar.

I would like to thank my parents Soo-Woong Kim and Cha-Soon Hwang and my brother Hyun-Dong Kim for their unconditional supports.

Table of Contents

List of Figures

CHAPTER I

List of Tables

Chapter II

Chapter I. Background and significance

1.1. Repetitive sequences that can adopt secondary structures induce genome instability

Many cancers and hereditary diseases are characterized by chromosomal anomalies such as deletions, inversions, translocations and gene amplifications (Duker, 2002; Lengauer et al, 1998). Chromosomal aberrations can be a consequence of cell exposure to exogenous factors that cause DNA damage (e.g. ionizing radiation), or can be a result of malfunctioning of the endogenous systems, for example those involved in DNA metabolism.

Recent studies reveal that certain DNA motifs can also pose a threat to genome stability (Gordenin & Resnick, 1998; Lobachev et al, 2000).Unstable repetitive sequences such as triplet repeats, inverted repeats, and AT- and GC-rich micro- and minisatellites known as "at-risk motifs" (ARMs) promote genome rearrangements (Gordenin & Resnick, 1998). Their ability to induce genome instability strongly depends on their potential to adopt non-canonical secondary structures (Callahan et al, 2003; Lobachev et al, 2000), Figure 1). Hairpin and cruciform structures can be formed by intra strand base pairing in one and in both DNA strands respectively. Because of internal symmetry of the inverted, AT, CG di-nucleotides and CTG/CAG, CGG/CCG triplet repeats these sequences could form hairpins and cruciforms (Sinden, 1994). In addition, some triplet repeats adopt intramolecular triplex DNA (GAA/TTC and CGG/CCG) or quadruplex DNA (CGG/CCG) structures via formation of Hoogsteen base-pairs between single or double

strands with the double helix (Sinden, 1994). The propensity to adopt stable secondary structures directly correlates with the size of repeats or repeat tracks (Mitas, 1997; Pearson et al, 1998; Sutherland et al, 1998). Among the above mentioned repetitive sequences, triplet repeats are under heavy scrutiny owing to their known ability to expand and consequently affect human health.

1.2. Triplet repeats expansion occurrences and consequences

1.2.1. Triplet repeats occur naturally in the human genome and expanded tracks lead to neurodegenerative diseases

Trinucleotide repeats can be found in both coding and non coding regions of the human genome. Expansions of CTG/CAG, CCG/CGG or GAA/TTC repeats are associated with nearly 25 human neuromuscular diseases and the list continues to grow (Table. 1) (Mirkin, 2007), (Pearson et al, 1998). The normal (non-disease) repeat size ranges from 5 to ~45 for CTG/CAG, 7 to ~60 for CCG/CGG and 7 to ~33 for GAA/TTC. Individuals with critical threshold length defined as "premutation size" of triplet repeats are more prone for expansions. For example, in individuals suffering from *Friedreich's* ataxia, a common inherited ataxia, expanded GAA/TTC repeats of up to 1700 triplets can be found in intron 1 of the FRDA gene (Figure 1).

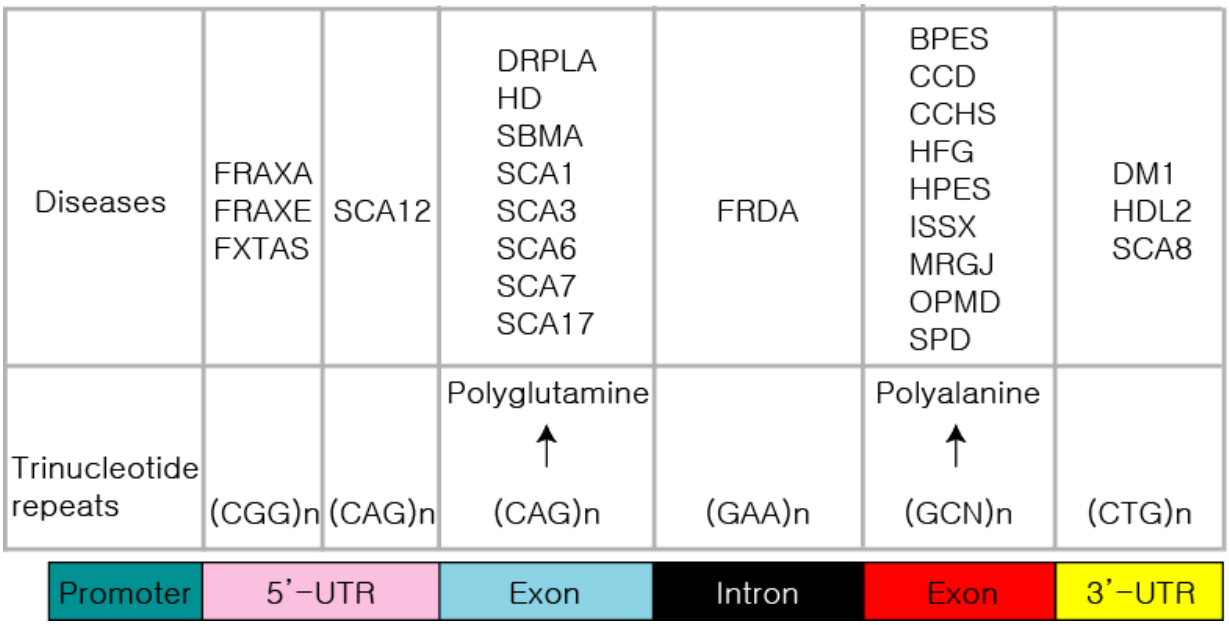


Figure 1. Locations of expandable trinucleotide repeats responsible for human diseases.

BPES, blepharophimosis and epicanthus inversus; CCD, cleidocranial dysplasia; CCHS, congenital central hypoventilation syndrome; DM, myotonic dystrophy; DRPLA, dentatorubral–pallidoluysian atrophy; FRAXA, fragile X syndrome; FRAXE, fragile X mental retardation associated with *FRAXE* site; **FRDA, Friedreich's ataxia**; FXTAS, fragile X tremor and ataxia syndrome; HD, Huntington's disease;HDL2, Huntington's-disease-like 2; HFG, hand–foot–genital syndrome;HPE5, holoprosencephaly 5; ISSX, X-linked infantile spasm syndrome; MRGH, mental retardation with isolated growth hormone deficiency; OPMD, oculopharyngeal muscular dystrophy; SBMA, spinaland bulbar muscular atrophy; SCA, spinocerebellar ataxia;SPD, synpolydactyly.

1.2.2. Expanded triplet repeats affect gene expression and/or protein function.

Expansion of triplet repeats found in non-coding regions of genes can result in loss of protein function. For instance, expansion of CGG repeats (more than 200 repeats) in the 5' UTR of the fragile X mental retardation 1 gene (FMR1) results in transcriptional silencing of the gene and loss of protein, causing alterations in dendritic functions in the diseased individuals. Polyglutamine diseases represent altered protein function disorders where expansion of triplet repeats in the coding regions of the gene leads to aggregation of the protein products. An example of such disorder is Huntington's disease, which is caused by the expansion of CAG repeats in exon1 of the corresponding gene. Expanded triplet repeats can also lead to altered RNA function, as evidenced in diseases such as myotonic dystrophy, which is caused by expansion of CAG repeats in the 3' of untranslated region.

1.3. Long tracks of triplet repeats (CAG/CTG and CCG/CGG) induce double stranded breaks and chromosome rearrangements.

Triplet repeat expansions can pose a threat to human health via adversely affecting gene/protein function or by affecting chromosome stability. Expanded triplet repeats act as fragile sites *in vitro* as revealed by constrictions in metaphase chromosomes of cells exposed to replication inhibitors (Nelson, 1995; Sutherland et al, 1998). Evidence for fragile site-mediated chromosome breakage in humans was first demonstrated in studies addressing the molecular mechanisms leading to Jacobsen syndrome. This disease is

caused by expansion of CCG repeats. The chromosome deletions in cells from the patients with the syndrome coincide with the location of the expanded repeats. Also, expanded tracks of CCG/CGG and CTG/CAG repeats accentuate chromosome breakage not only at the site of the repeats expansion but also in the vicinity of the location of the repeats in yeast chromosomes (Balakumaran et al, 2000; Sutherland et al, 1998).

Number of observations supports an idea that the abnormal secondary structure of the triplet repeats act as inducer for chromosome instability. Both CAG/CTG and CCG/CGG trinucleotide repeats are prone to adopt non-canonical secondary structures such as hairpin DNA, cruciform structure, slipped strand structure and/or quadruplex (Figure 2). These expanded trinucleotide repeat tracts can cause replication arrest *in vivo* in yeast (Pelletier et al, 2003).

1.4. Indirect evidence suggests GAA/TTC expansion - induced chromosome fragility.

GAA/TTC trinucleotide repeats (TNR) are prone to adopt non-canonical secondary structures which are considered to be inducers of genome instability. For instance, the expanded GAA/TTC tracts can adopt abnormal secondary structures such as triplex DNA, sticky DNA, and/or hairpin DNA which could interfere with DNA replication, transcription, and/or recombination. Although GAA/TTC repeats have strong potential for inducing genome instability, their ability to induce chromosome fragility has never

been proven so far. The overall goal of this study is to investigate the mechanisms by which long tracks of GAA/TTC repeats can induce chromosome fragility.

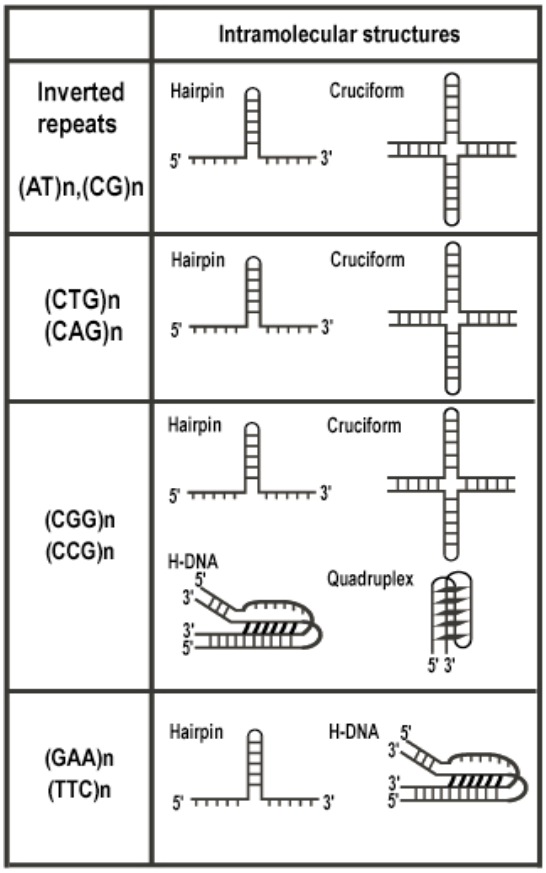


Figure 2. Examples of non-canonical DNA secondary structures adopted by repetitive sequences

Previous studies showed that GAA/TTC repeat blocks DNA replication on a plasmid. Often replication arrest can be converted to chromosome break which triggers chromosome rearrangements. As genome rearrangements are hallmark for cancer and human diseases, it is very important to understand the mechanism of chromosome fragility induced by GAA/TTC repeats. Moreover, it has been demonstrated that GAA/TTC repeats are highly abundant in the human genome (Clark et al, 2006; Clark et al, 2004; Clark et al, 2007), therefore they have high potential for induction of chromosome instability in multiple genomic locations which amplifies the significance of studying chromosome fragility potential caused by GAA/TTC repeats.

1.4.1. Expanded GAA/TTC repeats can adopt non-canonical secondary structure 'triplex DNA': *Pu* Pu Py and *Py* Pu Py

Human genetics and model organism studies leave no doubts that non-canonical secondary structures, such as hairpin or cruciform DNA, triplex DNA, and quadruplex are strong inducers for chromosome instability (Wells, 2008). Since expanded tracts of GAA/TTC repeats are prone adaptation of non-canonical secondary structure such as triplex DNA, sticky DNA and hairpin DNA we hypothesize that these abnormal secondary structures can act as inducers of genome instability. Dependence of the level of transcriptional repression of frataxin and the extent of genomic instability in patients'

cells on the size of GAA/TAA repeats in non-coding region of *FXN* gene strongly suggest causative role of abnormal DNA secondary structures in pathology of the disease (Campuzano et al, 1996; Clark et al, 2004; Clark et al, 2007). (The property of GAA/TTC repeats to inhibit frataxin transcription and the predisposition for genetic instability are dependent on the size of the repeats, which propose abnormal DNA secondary structures mediate genetic instability and transcription inhibition of frataxin.)

GAA/TTC trinucleotide track is a polypurine·polypyrimidine mirror repeat sequence. The polypurine·polypyrimidine mirror repeats can adopt two non-B-DNA structures: triplex (or H-DNA, intramolecular triplex DNA) and sticky DNA (Wells, 2008). Triplet repeat-containing DNA can adopt sticky DNA conformation under very strict conditions. It requires two long tracts of $(GAA/TTC)_n$ (n = 59–270) repeat in a single DNA and two GAA/TTC tracts must be located in the direct repeat orientation (Son et al, 2006), (Vetcher et al, 2002). It has been suggested that GAA/TTC tracts could form hairpin structures, however it has only been demonstrated in one *in vitro* study so far (Heidenfelder et al, 2003). GAA/TTC tracts could fold into triplex DNA structures under conditions less stringent that those necessary for "sticky DNA". Triplex DNA has been detected both *in vitro* and *in vivo* in independent studies utilizing various methods such as two-dimensional gel electrophoresis and antibody recognition of triplex DNA conformation (Wells, 2008).

Intramolecular triplexes in DNA can exist in either *Purine(R)* Purine(R) Pyrimidine(Y) or *Pyrimidine(Y)* Purine(R) Pyrimidine(Y) configurations (Figure 3, A). *Y*·R·Y structure forms only in acidic condition (~pH 4) and does not easily form at neutral pH (~pH 7). Under physiological conditions, the *R*·R·Y triplex is formed more readily and is more stable than the *Y*·R·Y conformation in GAA/TTC tracks (Sakamoto et al, 1999). The formation of such structures depends not only on pH of the reaction, but also on such factors as presence of divalent metal ions, length and homogeneity of the repeat tracts. Its formation is strongly preferential under conditions of negative superhelicity, which *in vivo* can be provided by processes that require separation of the two strands of the DNA duplex such as replication, transcription, and repair (Frank-Kamenetskii & Mirkin, 1995).

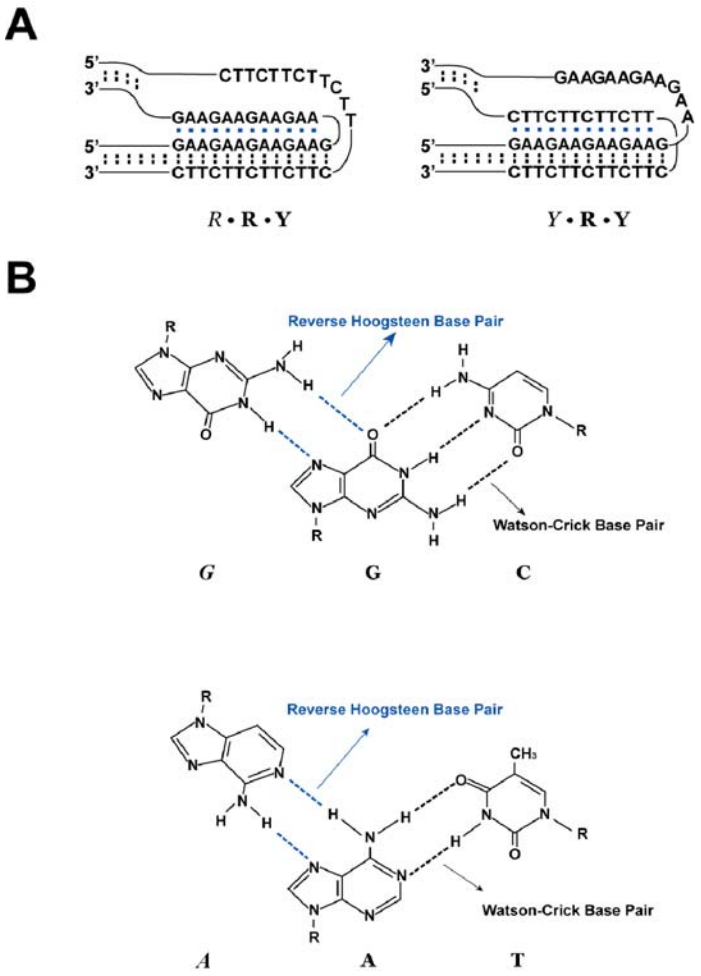


Figure 3. Triplex DNA structures formed by GAA/TTC repeats.

A. Two probable H-DNA configurations adopted by GAA repeats. The *R*.R.Y triplex is more stable and more versatile than *Y*.R.Y. B. Base pairing within G.G.G and A.A.T triplexes. The Hoogsteen hydrogen bonds are shown as blue dots and the Watson-Crick base parings are represented as black dots.

1.4.2. GAA/TTC repeats induce intra- and intermolecular recombination in *E. coli.*

Previous studies in *E.coli* and yeast have shown that GAA/TTC repeats have strong potential for genome instability induction. Intramolecular and intermolecular recombination studies showed that the frequency of recombination between the GAA/TTC tracts was 15 times higher than the non-repeating control sequences in *E.coli* (Napierala et al, 2004).

1.4.3. GAA/TTC repeats lead to replication stalling on a plasmid *in vitro and in vivo* in *S. cerevisiae.*

Several studies on the molecular pathology of GAA/TTC repeats destabilization in Friedreich's ataxia revealed that premutation- and disease-size repeat in yeast stalled the replication fork progression *in vitro* and *in vivo*, while normal-size repeats did not affect replication (Gacy et al, 1998; Krasilnikova & Mirkin, 2004). It is known that replication blockage is often apt to double stranded breaks formation (Bierne & Michel, 1994),(Michel et al, 1997). These findings argue that expanded GAA/TTC repeats have a potential to cause chromosome breakage and, consequently, fragility at the loci of repeats expansion..

1.5. Goals of the research

Overall, these observations suggest that GAA/TTC repeats, as well as CAG/CTG and CCG/CGG trinucleotide repeat, could create a hot spot of recombination at the site of expansion. Based on these observations, we hypothesize that expanded GAA/TTC repeats have strong potential to induce double stranded breaks and result in gross chromosome rearrangements (GCR) in eukaryotic genome.

The main goal of this research is to understand molecular mechanisms of GAA/TTC-associated genetic instability in a model eukaryotic organism, *S. cerevisiae.* We have focused our research on identifying the key players for chromosome fragility induced by GAA/TTC repeats. We demonstrate that expanded GAA/TTC repeats represent a threat to eukaryotic genome integrity by triggering double-strand breaks and gross chromosomal rearrangements. The fragility potential strongly depends on the length of the tracts and orientation of the repeats relative to the replication origin, i.e. with their propensity to adopt triplex structure and to block replication fork progression. In a course of our studies we discovered a novel function of mismatch repair machinery in modulation of triplex-induced chromosome fragility and demonstrated that MutSβ complex and endonuclease activity of MutLα play an important role in facilitation of fragility. We suggest that the mechanism of GAA/TTC-induced chromosomal aberrations defined in yeast can also operate in human carriers with expanded tracts (see Chapter II).

In addition to GAA/TTC triplex forming repeats, non-GAA polypurine polypyrimidine mirror repeats that are prone to the formation of similar structures were found to be hotspots for rearrangements in humans and other model organisms. These include H-DNA forming sequences located in the major breakpoint cluster region at *BCL2* (Raghavan et al, 2005), intron 21 of *PKD1* (Blaszak et al, 1999), (Patel et al, 2004) and promoter region of *C-MYC* (Wang & Vasquez, 2004) (Michelotti et al, 1996). We have examined the fragility potential of triplex (H-DNA) forming mirror motifs and compared to that of GAA/TTC trinucleotide repeats (for more information, see Chapter III).

Lastly, we have investigated the effect of the triplex-binding small molecules, azacyanines, on GAA-mediated fragility using the chromosomal arm loss assay. We have found that *in vivo*, azacyanines stimulate (GAA/TTC)-mediated arm loss in a dose dependent manner in actively dividing cells. Azacyanines treatment enhances the GAA-induced replication arrest. We discovered that also, azacyanines at concentrations that induce fragility also inhibit cell growth. Over 60% of yeast cells are arrested at G2/M stage of the cell cycle. This implies an activation of DNA-damage checkpoint response.(for more information, see Chapter IV).

1.6. References

Balakumaran BS, Freudenreich CH, Zakian VA (2000) CGG/CCG repeats exhibit orientation-dependent instability and orientation-independent fragility in Saccharomyces cerevisiae. *Hum Mol Genet* **9**(1)**:** 93-100

Bierne H, Michel B (1994) When replication forks stop. *Mol Microbiol* **13**(1)**:** 17-23

Blaszak RT, Potaman V, Sinden RR, Bissler JJ (1999) DNA structural transitions within the PKD1 gene. *Nucleic Acids Res* **27**(13)**:** 2610-2617

Callahan JL, Andrews KJ, Zakian VA, Freudenreich CH (2003) Mutations in yeast replication proteins that increase CAG/CTG expansions also increase repeat fragility. *Mol Cell Biol* **23**(21)**:** 7849-7860

Campuzano V, Montermini L, Molto MD, Pianese L, Cossee M, Cavalcanti F, Monros E, Rodius F, Duclos F, Monticelli A, Zara F, Canizares J, Koutnikova H, Bidichandani SI, Gellera C, Brice A, Trouillas P, De Michele G, Filla A, De Frutos R, Palau F, Patel PI, Di Donato S, Mandel JL, Cocozza S, Koenig M, Pandolfo M (1996) Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science* **271**(5254)**:** 1423-1427

Clark RM, Bhaskar SS, Miyahara M, Dalgliesh GL, Bidichandani SI (2006) Expansion of GAA trinucleotide repeats in mammals. *Genomics* **87**(1)**:** 57-67

Clark RM, Dalgliesh GL, Endres D, Gomez M, Taylor J, Bidichandani SI (2004) Expansion of GAA triplet repeats in the human genome: unique origin of the FRDA mutation at the center of an Alu. *Genomics* **83**(3)**:** 373-383

Clark RM, De Biase I, Malykhina AP, Al-Mahdawi S, Pook M, Bidichandani SI (2007) The GAA triplet-repeat is unstable in the context of the human FXN locus and displays age-dependent expansions in cerebellum and DRG in a transgenic mouse model. *Hum Genet* **120**(5)**:** 633-640

Duker NJ (2002) Chromosome breakage syndromes and cancer. *Am J Med Genet* **115**(3)**:** 125-129

Frank-Kamenetskii MD, Mirkin SM (1995) Triplex DNA structures. *Annu Rev Biochem* **64:** 65-95

Gacy AM, Goellner GM, Spiro C, Chen X, Gupta G, Bradbury EM, Dyer RB, Mikesell MJ, Yao JZ, Johnson AJ, Richter A, Melancon SB, McMurray CT (1998) GAA instability in Friedreich's Ataxia shares a common, DNA-directed and intraallelic mechanism with other trinucleotide diseases. *Mol Cell* **1**(4)**:** 583-593

Gordenin DA, Resnick MA (1998) Yeast ARMs (DNA at-risk motifs) can reveal sources of genome instability. *Mutat Res* **400**(1-2)**:** 45-58

Heidenfelder BL, Makhov AM, Topal MD (2003) Hairpin formation in Friedreich's ataxia triplet repeat expansion. *J Biol Chem* **278**(4)**:** 2425-2431

Krasilnikova MM, Mirkin SM (2004) Replication stalling at Friedreich's ataxia (GAA)n repeats in vivo. *Mol Cell Biol* **24**(6)**:** 2286-2295

Lengauer C, Kinzler KW, Vogelstein B (1998) Genetic instabilities in human cancers. *Nature* **396**(6712)**:** 643-649

Lobachev KS, Stenger JE, Kozyreva OG, Jurka J, Gordenin DA, Resnick MA (2000) Inverted Alu repeats unstable in yeast are excluded from the human genome. *Embo J* **19**(14)**:** 3822-3830

Michel B, Ehrlich SD, Uzest M (1997) DNA double-strand breaks caused by replication arrest. *EMBO J* **16**(2)**:** 430-438

Michelotti GA, Michelotti EF, Pullner A, Duncan RC, Eick D, Levens D (1996) Multiple single-stranded cis elements are associated with activated chromatin of the human c-myc gene in vivo. *Mol Cell Biol* **16**(6)**:** 2656-2669

Mirkin SM (2007) Expandable DNA repeats and human disease. *Nature* **447**(7147)**:** 932-940

Mitas M (1997) Trinucleotide repeats associated with human disease. *Nucleic Acids Res* **25**(12)**:** 2245-2254

Napierala M, Dere R, Vetcher A, Wells RD (2004) Structure-dependent recombination hot spot activity of GAA.TTC sequences from intron 1 of the Friedreich's ataxia gene. *J Biol Chem* **279**(8)**:** 6444-6454

Nelson DL (1995) The fragile X syndromes. *Semin Cell Biol* **6**(1)**:** 5-11

Patel HP, Lu L, Blaszak RT, Bissler JJ (2004) PKD1 intron 21: triplex DNA formation and effect on replication. *Nucleic Acids Res* **32**(4)**:** 1460-1468

Pearson CE, Eichler EE, Lorenzetti D, Kramer SF, Zoghbi HY, Nelson DL, Sinden RR (1998) Interruptions in the triplet repeats of SCA1 and FRAXA reduce the propensity and complexity of slipped strand DNA (S-DNA) formation. *Biochemistry* **37**(8)**:** 2701-2708

Pelletier R, Krasilnikova MM, Samadashwily GM, Lahue R, Mirkin SM (2003) Replication and expansion of trinucleotide repeats in yeast. *Mol Cell Biol* **23**(4)**:** 1349-1357

Raghavan SC, Chastain P, Lee JS, Hegde BG, Houston S, Langen R, Hsieh CL, Haworth IS, Lieber MR (2005) Evidence for a triplex DNA conformation at the bcl-2 major breakpoint region of the t(14;18) translocation. *J Biol Chem* **280**(24)**:** 22749-22760

Sakamoto N, Chastain PD, Parniewski P, Ohshima K, Pandolfo M, Griffith JD, Wells RD (1999) Sticky DNA: self-association properties of long GAA.TTC repeats in R.R.Y triplex structures from Friedreich's ataxia. *Mol Cell* **3**(4)**:** 465-475

Sinden (1994) *DNA structure and Function*, San Diego: Academic Press.

Son LS, Bacolla A, Wells RD (2006) Sticky DNA: in vivo formation in E. coli and in vitro association of long GAA*TTC tracts to generate two independent supercoiled domains. *J Mol Biol* **360**(2)**:** 267-284

Sutherland GR, Baker E, Richards RI (1998) Fragile sites still breaking. *Trends Genet* **14**(12)**:** 501-506

Vetcher AA, Napierala M, Iyer RR, Chastain PD, Griffith JD, Wells RD (2002) Sticky DNA, a long GAA.GAA.TTC triplex that is formed intramolecularly, in the sequence of intron 1 of the frataxin gene. *J Biol Chem* **277**(42)**:** 39217-39227

Wang G, Vasquez KM (2004) Naturally occurring H-DNA-forming sequences are mutagenic in mammalian cells. *Proc Natl Acad Sci U S A* **101**(37)**:** 13448-13453

Wells RD (2008) DNA triplexes and Friedreich ataxia. *FASEB J* **22**(6)**:** 1625-1634

Chapter II

Chromosome fragility at GAA/TTC tracts in yeast depends on repeat orientation and requires mismatch repair system

- The study results in Chapter II were published. *The EMBO Journal* (2008) 27, 2896–2906, 'Chromosome fragility at GAA tracts in yeast depends on repeat orientation and requires mismatch repair', Hyun-Min Kim[1], Vidhya Narayanan[1], Piotr A Mieczkowski[2,5], Thomas D Petes[2], Maria M Krasilnikova[3], Sergei M Mirkin[4] and Kirill S Lobachev[1*].

1 School of Biology and Institute for Bioengineering and Bioscience, Georgia Institute of Technology, Atlanta, GA, USA
2 Department of Molecular Genetics and Microbiology, Duke University, Durham, NC, USA
3 Department of Biochemistry and Molecular Biology, Penn State University, University Park, PA, USA
4 Department of Biology, Tufts University, Medford, MA, USA
* Corresponding author

- All experiments were done by me with an assist of V. Narayanan and Dr. K. Lobachev.

- All comparative genome hybridization (CGH) were done by Dr. Mieczkowski at Petes' lab at Duke University.

2.1. Introduction

Expansion of GAA/TTC trinucleotide repeats was recognized as a detrimental polymorphism in the human genome with the discovery of the molecular mechanisms underlying Friedreich's ataxia (FRDA) (Campuzano et al, 1996). FRDA is an autosomal recessive disease caused by the inheritance of two mutant alleles of the frataxin (*FXN*) gene from heterozygous parents (reviewed in (De Biase et al, 2006)). In most cases (98%), inactivation of the *FXN* function in both alleles results from inhibition of gene expression by abnormal GAA repeat expansion occurring within the first intron. While chromosomes from unaffected individuals have less than 65 triplets, disease-causing FRDA alleles contain 66-1,700 GAA repeats. Premutation (34-65 triplets) and mutant (>66 triplets) alleles exhibit high levels of instability (expansions and contractions) in somatically dividing and non-dividing cells in a tissue- and an age-dependent manner (Al-Mahdawi et al, 2004). Premutation and disease alleles are also highly unstable during intergenerational transmission often undergoing both contractions and expansions, with hyperexpansions reaching up to a 10-fold increase in one generation (reviewed in (De Biase et al, 2006)).

Systematic analysis of the human genome revealed that the *FXN* locus is not the only location where GAA tracts can expand (Clark et al, 2004; Clark et al, 2007). Almost 1,000 loci containing more than eight GAA repeats, including 29 loci with premutation size tracts, have been identified. Nine out of 29 premutation alleles are highly

polymorphic and prone to large expansions which can reach up to 140 copies. To date, these expansions have not been shown to be associated with diseases.

The property of GAA repeats to inhibit *FXN* transcription and its predisposition for genetic instability are dependent on the size of the expanded tracts, which in fact reflects the ability of the repeats to adopt non-canonical DNA secondary structures (Wells, 2008).The GAA triplet repeat is a polypurine·polypyrimidine (R·Y) sequence exhibiting mirror symmetry (reviewed in (Frank-Kamenetskii & Mirkin, 1995)). Such R·Y tracts can predominantly adopt two non-B-DNA structures: triplex (or H-DNA) and sticky DNA. Triplex is formed as a result of overlaying a third strand into the major grove of the DNA double helix. The third strand pairs with the double helix via Hoogsteen or reverse Hoogsteen hydrogen bonds, thus leaving the complementary strand (either R or Y) unpaired (For more information, see Chapter I).

Triplexes were detected *in vitro* and *in vivo* studies in model systems (reviewed in (Bissler, 2007)). The formation of such structures is dependent on the homogeneity of the GAA tract and is strongly favored under conditions of negative superhelicity (For more information, see Chapter I). Studies *in vitro* and in model organisms show that stable secondary structures in turn can hinder transcription (Bidichandani et al, 1998) which can account for the GAA length-dependent inactivation of *FXN* gene function in FRDA patients (Campuzano et al, 1996). It was also found that triplexes formed by GAA repeats stall the progression of replication fork both *in vitro* and *in vivo* (Gacy et al, 1998;

Krasilnikova & Mirkin, 2004), providing possible explanations for the GAA/TTC-associated genetic instability and the origin of expanded alleles.

In this study, we demonstrate that, expanded GAA/TTC repeats are strong inducers of DSBs and gross chromosomal rearrangements (GCRs), in yeast. The fragility potential depends on the length of the tract and the orientation of the repeats relative to the replication origin which correlates with their propensity to adopt triplex and to block replication fork movement. Mutants defective in the function of MutSβ and the endonuclease activity of MutLα exhibit reduced levels of GCRs and DSB formation, indicating that mismatch repair machinery (MMR) might trigger the fragility by processing the triplex structure. GCRs resulting from the GAA-mediated breaks have a specific pattern: terminal deletions coupled with non-reciprocal translocations involving expanded GAA/TTC tracts and GAA/TTC–rich regions located on non-homologous chromosomes. We propose that the mechanism of genome destabilization caused by GAA/TTC repeats defined in yeast might operate in carriers with expanded tracts at the *FXN* and other loci in the human genome.

2.2. Results

2.2.1. Experimental system

To evaluate the potential of the expanded GAA/TTC repeats to induce chromosomal fragility we have employed two experimental assays that monitor the induction of gross chromosome rearrangements (GCR)s and mitotic ectopic recombination. The GCR assay

is based on the loss of *CAN1* and *ADE2* genes located on chromosome V (Figure 1). This experimental assay was previously used to characterize the specific pattern of GCRs resulting from hairpin-capped breaks induced by inverted repeats (Narayanan et al., 2006). Haploid yeast strains were constructed where the left arm of chromosome V in the region of *CAN1* gene was modified. *LYS2* cassettes with GAA/TTC repeats were placed centromere-proximal to *CAN1*. The region between *LYS2* and the telomere does not contain essential genes and can be deleted. The *ADE2* gene was moved telomere-distal to *CAN1*. The *LYS2* cassettes contain homogeneous GAA/TTC repeats of length 20 (corresponding to normal allele size in humans), 60 (pre-mutation size), 120, 230 or 340 (mutant sizes); all GAA/TTC insertions result in loss of *LYS2* function. Repeats were inserted into the chromosome in two different orientations with respect to the direction of replication. Replication is initiated at the *ARS507* origin and proceeds from right to left in this region (Raghuraman et al., 2001; Yabuki et al., 2002, see also replication origin database at http://www.oridb.org/). The lagging strand template contains GAA repeats in "GAA-orientation" and TTC repeats in "TTC-orientation". A DSB in the *LYS2* region can cause deletion of the chromosome V region including *CAN1* and *ADE2,* resulting in canavanine-resistant red colonies ($Can^{R}Ade^{-}$). Such GCR isolates can be distinguished from canavanine-resistant white colonies that are produced due to point mutations or small deletions in *CAN1*.

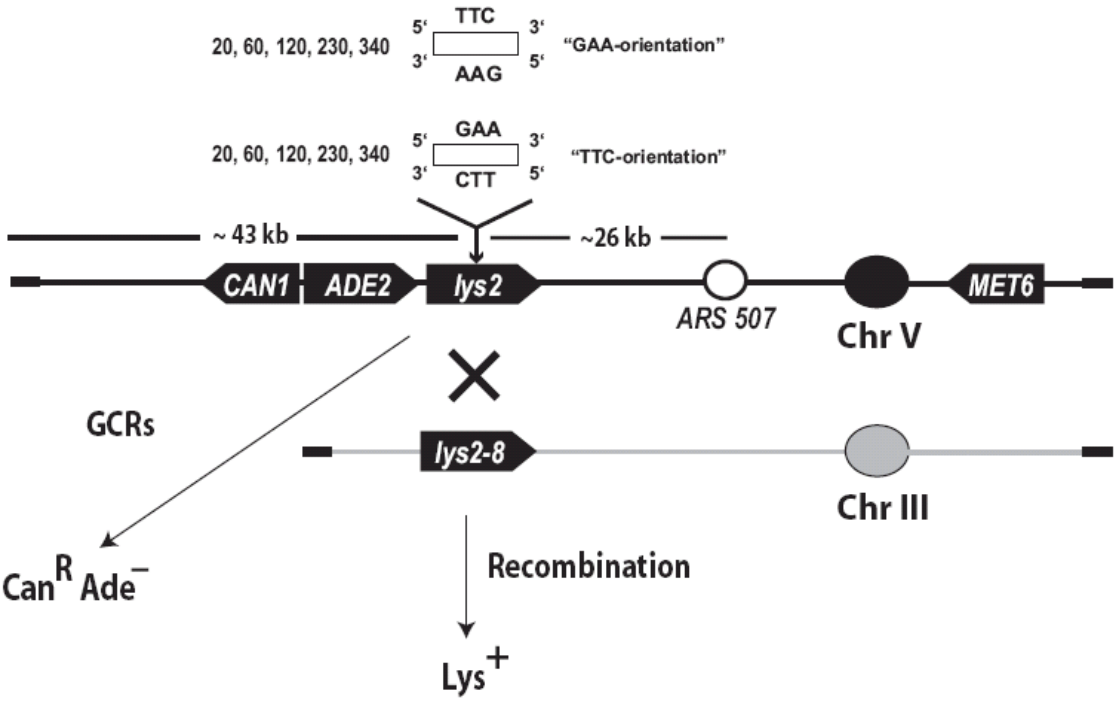


Figure 1. Experimental system to study chromosomal fragility induced by expanded tracts of GAA/TTC repeats.

The breakage at the location of GAA/TTC tracts can lead to 43 kb telomere-proximal deletion resulting in $Can^{R}Ade^{-}$ clones. In a separate set of strains, the *lys*2-8 allele was integrated into chromosome III, allowing us to measure the level of homologous recombination induced by GAA/TTC repeats. The 'X' denotes a recombination event generating a wild-type *LYS2* allele.

To determine if the GAA/TTC tracts can stimulate mitotic ectopic recombination, we integrated a *lys2-8* allele (Lobachev et al., 1998) at the *LEU2* locus of chromosome III. Recombination between *lys2*::*GAA/TTC* and the *lys2-8* generates Lys^+ prototrophs, primarily through gene conversion of the insert-containing allele.

2.2.2. GAA repeats induce GCRs in a size- and an orientation-dependent manner

Expanded tracts of GAA/TTC repeats strongly increased the rate of chromosome V arm loss (Table 1). The degree of stimulation depended on both the size and the orientation of the repetitive tracts. The strains with $(GAA)_{20}$ and $(TTC)_{20}$ tracts exhibited low rates of arm loss events, similar to that of strains containing direct *Alu* repeats which cannot adopt secondary structures (Narayanan et al., 2006). There was a mild increase (~3 fold) in levels of *CAN1* region loss for pre-mutation size alleles (60 repeats) over normal size alleles for both orientations. However, alteration in the size of the repeat tracts from 60 to 120 led to a tremendous change in their ability to trigger chromosomal arm loss events. There were a 215-fold and a 646-fold increases in TTC- and GAA-orientation, respectively. Interestingly, further increments in the repeat tract lengths had different effects on arm loss rates for different orientations. TTC repeats of 120, 230 and 340 tract lengths had similar rates, whereas the 120, 230 and 340 GAA tracts stimulated GCRs 1,937-, 7,338- and 82,629 times more frequently than $(GAA)_{20}$, respectively.

2.2.3. Structural organization of rearranged chromosomes in Can^{R}Ade$^-$ isolates

To directly determine what structural changes were acquired by chromosome V as a result of GCR, we analyzed the molecular karyotypes of 12 independent Can^{R}Ade$^-$ clones isolated from strains containing $(TTC)_{230}$ and $(GAA)_{230}$ repeats. Chromosomes from these isolates were separated using CHEF gel electrophoresis and chromosome V was examined by hybridization with a right arm-specific probe (Figure 2A). Based on the mobility of altered chromosomes, several different recurrent classes of rearrangements were detected for both repeat orientations. In the majority of cases, the novel chromosomes were larger than wild type chromosome V, suggesting that the arm loss events were accompanied by the gain of genetic material. This conclusion was confirmed when genomic DNA from Can^{R}Ade$^-$ isolates was analyzed using comparative genomic hybridization (CGH) on microarrays (Figure 2B).

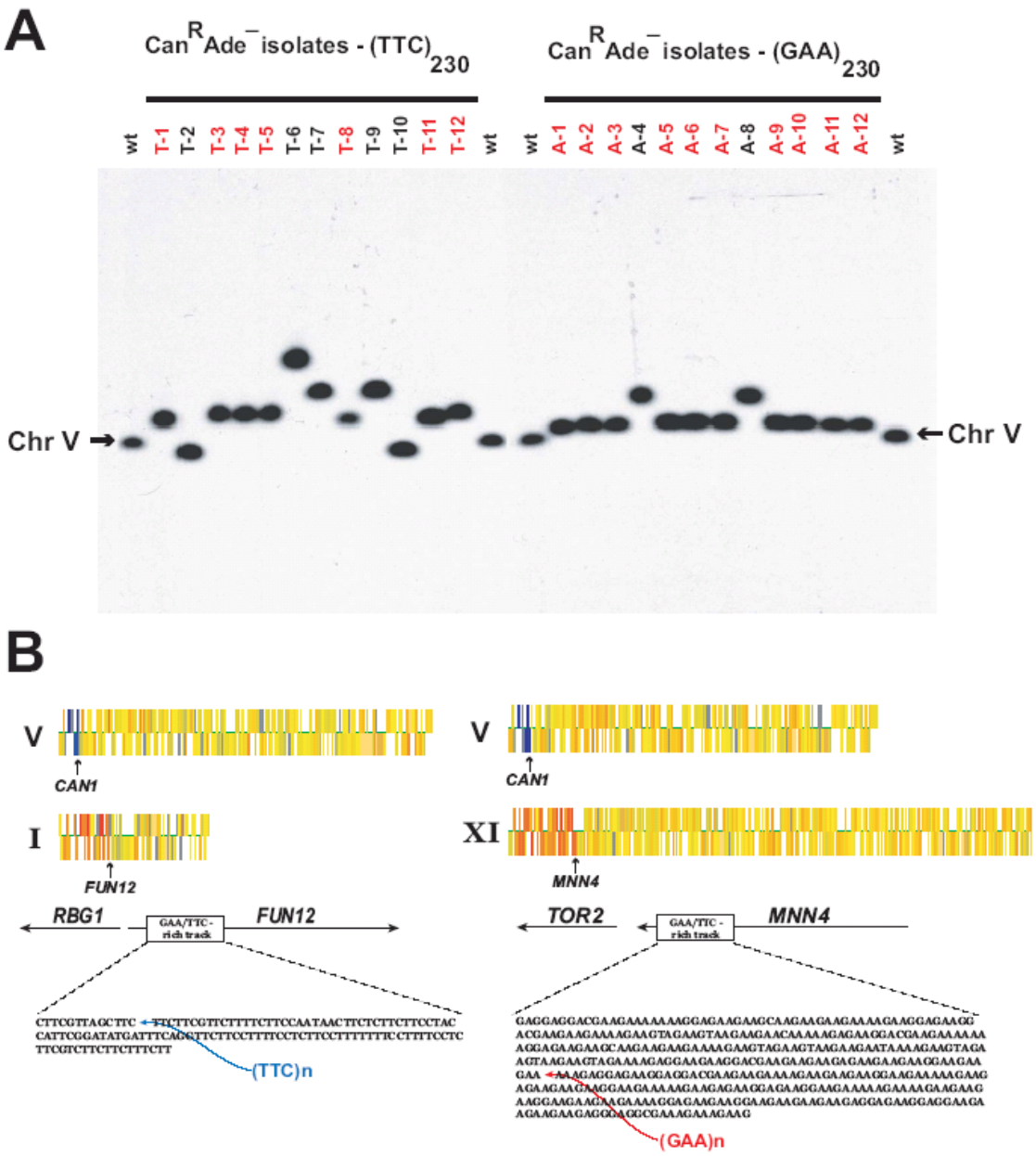


Figure 2. Structural analysis of chromosomal arm loss events stimulated by $(GAA/TTC)_{230}$ tracts.

(A) Analysis of rearranged chromosome Vs in $Can^R Ade^-$ isolates by CHEF gels and Southern blotting. The right arm of chromosome V was highlighted using a *MET6*-specific probe in Southern analysis. Lanes labelled with 'wt' are strains containing wild-type chromosome V with $(TTC)_{230}$ and $(GAA)_{230}$ repeats. Lanes T-1 to T-12 are $Can^R Ade^-$ isolates from $(TTC)_{230}$ strains.

Lanes A-1 and A-12 are $Can^{R}Ade^{-}$ isolates from strains with $(GAA)_{230}$ repeats. The primary GCR classes are labelled in red. (B) CGH and breakpoint analysis of the most frequent rearrangements resulting from $(TTC/GAA)_{230}$ -mediated breaks. Upper panels are the microarray analysis of arm loss events. DNAs from experimental strain and control strain were labelled with different fluorescent nucleotides and hybridized in competition to DNA microarrays with yeast genes and intergenic regions. Each vertical bar corresponds to one ORF in Watson (upper bars) and in Crick (bottom bars) orientations. Color coding is as follows: grey, repeated genomic elements; yellow, sequences present in the same dosage in the wild-type and control strains; red, sequences that were duplicated in the experimental strain relative to the control; blue, sequences that were deleted in the experimental strain relative to the control. Only those chromosomes that had a deletion or duplication are shown in this figure. Complete data for these experiments is online at GEO database (accession number GSE11425). Bottom panels depict the structure of the translocation breakpoints on chromosomes I and XI. The donor sites for BIR are shown. Blue and red arrows indicate the breakpoint junctions between GAA/TTC tracts from chromosome V and GAA/TTC-rich regions on donor chromosomes (examples are shown). The left panel is the analysis of a major class of GCRs in $(TTC)_{230}$ strains (isolates T-1, T-3, T-4, T-5, T-8, T-11 and T-12). The right panel is the analysis of a major class of GCRs in $(GAA)_{230}$ strains (isolates A-1, A-2, A-3, A-5, A-6, A-7, A-9, A-10, A-11 and A-12). The complete analysis of the breakpoints for all isolates is presented in the **Table 2.**

Among the 12 analyzed $Can^{R}Ade^{-}$ isolates from $(TTC)_{230}$ strains, only two (T-2 and T-10) had a terminal deletion of V with a breakpoint near *CAN1* locus. This pattern likely reflects *de novo* telomere addition to the broken molecule following DSB induction at the repetitive tracts. The remaining ten isolates had a deletion of the centromere-distal *CAN1* region coupled with a duplication of telomere-proximal regions of non-homologous chromosomes. One likely mechanism for generating such rearrangements is the induction of the break at the location of GAA/TTC repeats on chromosome V, followed by healing of the broken end via break-induced replication (BIR) (Malkova et al., 1996) involving homology or microhomology (Figure 6). Non-reciprocal translocations were confirmed by PCR analysis with primers annealing to the regions on non-homologous chromosomes that flank the breakpoints. We sequenced the PCR fragments containing the breakpoint junctions for one or several representatives of each GCR class. All such junctions were chimeric with a GAA/TTC pure repeat region from chromosome V fused with non-homogeneous GAA/TTC-rich stretches in chromosome I (T-1, T-3, T-4, T-5, T-8, T-11 and T-12), chromosome XIII (T-7 and T-9) or chromosome II (T-6). The non-homogeneous GAA/TTC-rich tracts were all configured such that a BIR event initiated by a break in the pure GAA/TTC tract on chromosome V in the TTC orientation would produce a monocentric chromosome.

Similarly, in isolates derived from $(GAA)_{230}$ strains the broken end was stabilized via BIR involving chromosome XI (A-1, A-2, A-3, A-5, A-6, A-7, A-9, A-10, A-11 and A-12) or chromosome XIII (A-4 and A-8). The translocation breakpoints were mapped to 420 bp GAA/TTC-rich tract in *MNN4* gene on chromosome XI and to a 120 bp GAA/TTC-containing tract in *FPR3* gene on chromosome XIII (Figure 2B and Table 2).

In summary, these results indicate that repair of breaks triggered by GAA/TTC repeats generate specific patterns of rearrangements, wherein translocations are the primary outcome of the GAA/TTC-mediated fragility.

2.2.4. Induction of homologous recombination between *lys*2 alleles depends on the orientation of repeat tracts

The disparity in the GCR potential of GAA and TTC tracts can be explained by the different propensities of the repeats to adopt secondary structures that are processed to DSBs (discussed below). Alternatively, the bias can be attributed to the efficiency of subsequent steps in the recombination process (processing of the broken ends, invasion of the broken end into a homologous template, etc.). In addition, the number of genomic templates that are "invadable" by GAA/TTC tracts, their lengths, degree of sequence divergence, and orientation with respect to telomere could be contributing factors to account for the differences in GCR rates. To determine whether the disparity in the GCR

potential reflects DSB formation or a subsequent step in the repair, we analyzed recombination between a *lys2-8* allele integrated at the *LEU2* locus of chromosome III and the *lys2* alleles containing 20, 60, 120 and 230 repeats in both orientations. The *lys2-8* allele serves as a uniform template for DSB repair and allows to measure the true repeat fragility potential in an unbiased manner. Consistent with GCR data, we found that the GAA repeats stimulate recombination more strongly than the TTC repeats. For example, the $(GAA)_{230}$ tract induces recombination between *lys2* alleles about 200 times more efficiently than the $(TTC)_{230}$ tract (Table 1). These results indicate that the observed orientation dependence is likely to reflect differences in the propensity of the GAA and TTC repeats for breakage.

2.2.5. Orientation-dependent blockage of replication by expanded GAA tracts

To get better insights into the molecular mechanisms of GAA/TTC-associated instability, we analyzed the progression of the replication fork through the chromosomal region containing GAA/TTC tracts using two-dimensional gel electrophoresis (2D) (Figure 3). Replication progression across $(GAA)_{20\text{-}230}$, $(TTC)_{230}$ and $(TTC)_{340}$ tracts was monitored in wild type strains. Owing to the inherent instability associated with $(GAA)_{340}$ repeats (see below), wild type strains rapidly accumulate a mix of truncated tracts upon propagation, making the 2D results unclear (data not shown). Therefore, the

analysis of replication forks in this strain was carried out in *Δmsh2* background that prevents large changes in the tract size (see below).

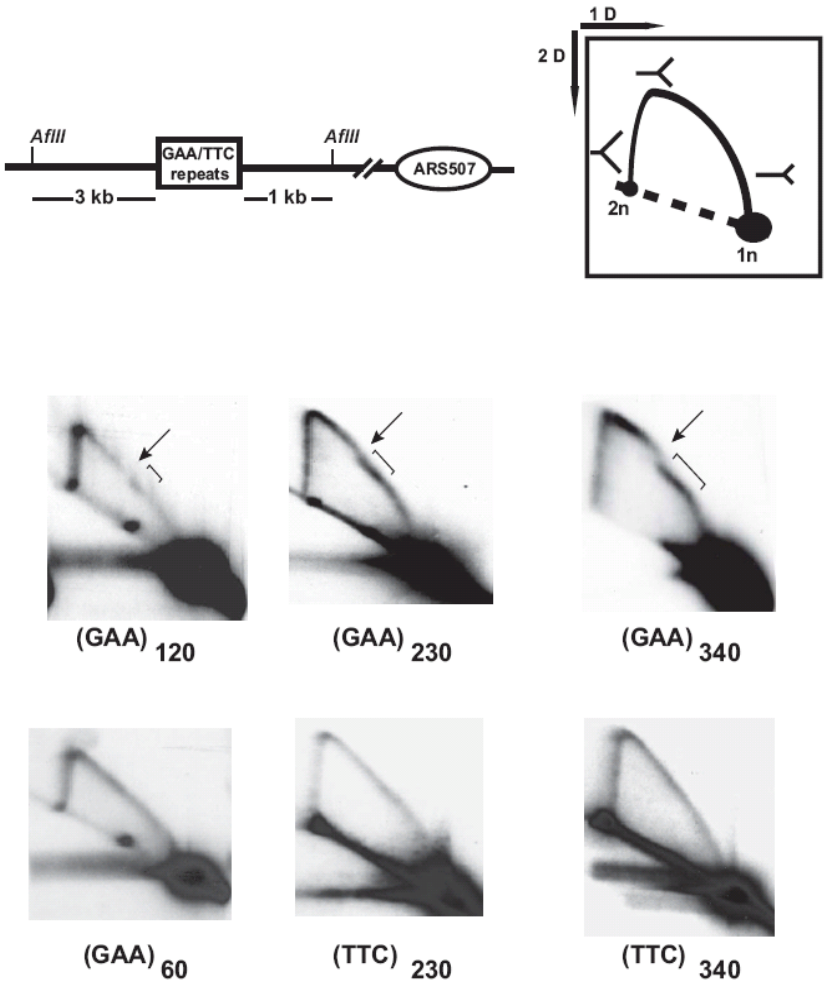


Figure 3. 2D analysis of replication intermediates in strains containing GAA/TTC repeats. Neutral/neutral 2D electrophoresis was used to resolve unreplicated molecules and Y-like structures (Brewer and Fangman, 1987). Replication initiated at ARS507 proceeds from right to left through the region containing the repeat tracts. Cleavage with *Afl*II positions the GAA/TTC repeats on the long shoulder of the Y-arc. The 4 kb *Afl*II-digested *LYS2* fragment was used as a probe in Southern blot hybridization. Accumulation of the replication intermediates leads to the appearance of bulges on the replication arc. Replication pausing zones are indicated by brackets. Arrows point to the Y-arc interruptions coinciding with the centre of the GAA tracts.

We have found that replication stalling occurs at $(GAA)_{120}$, $(GAA)_{230}$ and $(GAA)_{340}$ but not at $(GAA)_{60}$ or TTC tracts. The inhibition zone coincides with the location of repeat tracts. These results are consistent with the previous report wherein GAA repeats arrested replication of two-micron plasmids in a length- and an orientation-dependent manner (Krasilnikova and Mirkin, 2004). It should be noted that, in this study, the blockage zone of the Y arc is shifted causing the arc interruption. This discontinuity co-localizes with the center of the GAA tracts. This particular migration pattern of Y intermediates might be explained by the presence of secondary structures, such as H-DNA, at the arrested forks.

Based on these results, we suggest that the expanded GAA repeats, when present on lagging strand template, lead to the formation of triplex DNA structure that blocks the progression of the replication fork. It has been shown that the homopurine tracts are poor substrates for replication protein A (RPA) binding and for primer synthesis by the Polα-primase complex (Frick and Richardson, 2001; Wold, 1997). Hence, it is likely that GAA repeats on the template of lagging strand hinder the synthesis of Okazaki fragments, and therefore, generate long regions of single-stranded DNA, providing optimal conditions for secondary structure formation. In addition, the *R*·R·Y triplex that would be formed by folding the GAA-rich strand is expected to be more stable than the *Y*·R·Y conformation at the physiological pH and ionic conditions (Frank-Kamenetskii and Mirkin, 1995).

Overall these results demonstrate that repeats in the orientation most prone for fragility also block the replication fork progression, suggesting that DSB formation and replication arrest are related events.

2.2.6. Effect of inverting the *LYS2* cassette on fragility and the replication block potential of $(GAA/TTC)_{230}$ repeats

It is formally possible that the orientation-dependent bias in the fragility of the GAA repeats might be attributed to the direction of transcription in *LYS2* gene. Besides replication, transcription is another polar cellular process wherein the duplex DNA is unwound creating regions of negative superhelicity. In the GAA-orientation, GAA repeats are located on the transcribed strand of *LYS2*. It has been demonstrated that the *FXN* gene expression is blocked (likely by triplex DNA) when expanded GAA tracts are on the sense strand (Bidichandani et al., 1998; Campuzano et al., 1997). In addition, it has been shown that halted transcription can attenuate replication fork progression (Krasilnikova et al., 1998). Hence, it is possible that replication arrest and subsequent breakage could result from defect in transcription elongation rather than from impaired lagging strand synthesis. It should be noted, in plasmid-based studies, the GAA-induced replication block was not dependent on transcription through the repeats (Krasilnikova and Mirkin, 2004). Consistent with this study, we found that disruption of *LYS2* promoter with the *KanMX* cassette in strains containing $(GAA)_{230}$ and $(TTC)_{230}$ repeats did not

affect either the GCRs rates or the replication fork progression across repeat tracts (data not shown).

To directly assess if the presence of GAA repeats on lagging strand template is responsible for the observed repeat orientation-dependent fragility, we constructed strains in which the orientation of the *LYS2* gene with $(GAA)_{230}$ was changed with respect to direction of replication from *ARS507* (Figure 4A). The flipped cassette now places the GAA repeats on the leading strand template. The arm loss events were ~9 fold lower than in strains containing original orientation of *LYS2* cassette wherein GAA repeats are situated on the lagging strand template. It should be noted that this fold difference in arm loss events is comparable to that detected in strains with $(GAA)_{230}$ and $(TTC)_{230}$ with the *LYS2* cassette is in the original orientation. As expected, when the *LYS2* with the $(TTC)_{230}$ tract was inverted, the GCR tendency of the tract was also reversed. Moreover, the flipped GAA repeats did not compromise replication fork progression while replication arrest was readily detected at the flipped TTC tracts (Figure 4B).

This observed change in GCR potential of the repeat tract upon the inversion of the *LYS2* cassette, demonstrates that the fragility is independent of sequences that flank the repetitive tracts, undermines the contribution of transcription and strongly implicates the role of replication.

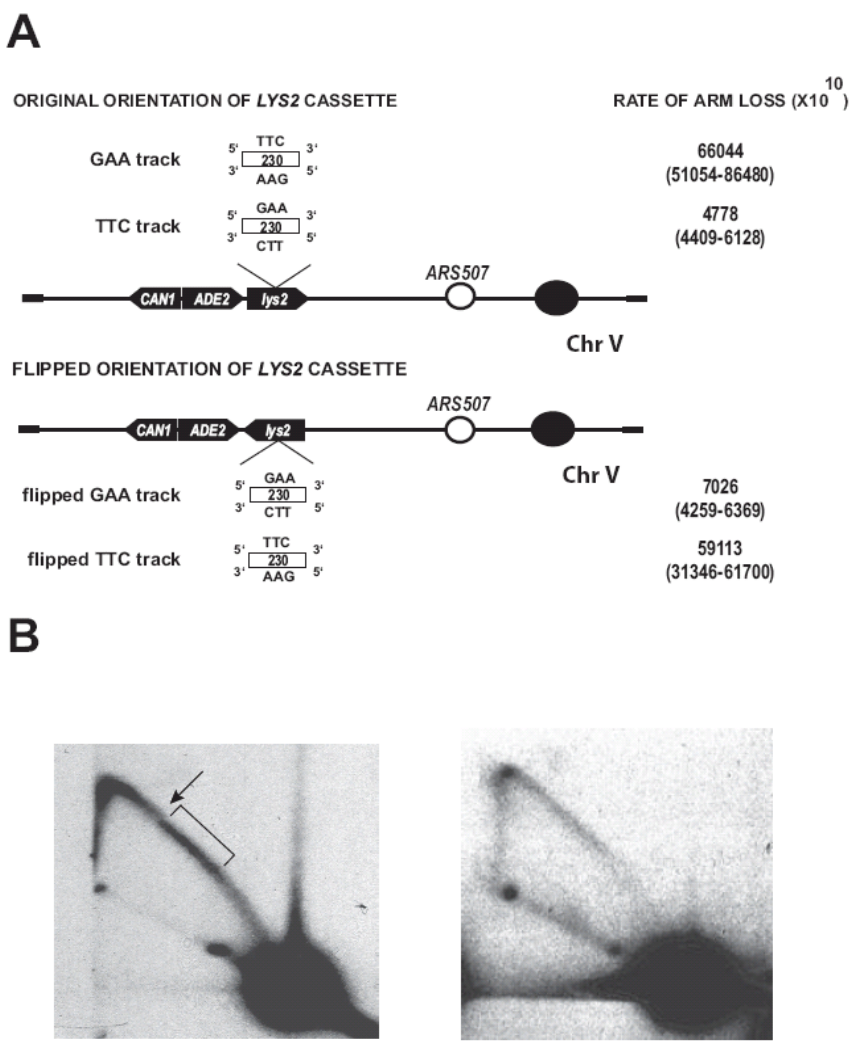


Figure 4. Induction of GCR and the ability to block fork progression by GAA/TTC repeats are affected by their orientation relative to the origin of replication.

(A) GCR rates of original and flipped constructs. The schematic diagram of the original and the flipped *LYS2* cassette containing GAA/TTC tracts is shown on the left. The corresponding GCR rates are shown on the right. (B) Replication fork progression across flipped GAA/TTC tracts. 2D analysis was performed as described in **Figure 3**. The replication pause zone across the flipped TTC tracts is indicated by brackets.

2.2.7. GAA repeat size variations and fragility are dependent on the mismatch repair system

Increases in the size of repeat tracts led to elevated levels of tract length variations with deletions observed more frequently than expansions (Table 3). The most unstable are the $(GAA)_{340}$ tracts that exhibit 80% of large contractions upon propagation. We found that disruptions of *MSH2*, *MSH3*, *MLH1* and *PMS1,* but not *MSH6,* genes resulted in decreased levels of large deletions in $(GAA)_{340}$ tracts, however, there was a marked increase in the levels of small deletions (Figure 7 and data not shown). This data are consistent with previous studies in yeast where MMR deficiency was shown to cause elevated levels of small deletions and additions in tracts of repetitive DNA (for example, Sia et al., 1997).

Defects in MMR also strongly reduced $(GAA)_{230}$, and $(TTC)_{230}$ repeat-induced chromosomal arm loss (Figure 5A). Strains of the *Δmsh2* or *Δmsh2Δmsh6* genotypes had about 15-fold reductions in the ability of $(GAA)_{230}$ tracts to trigger GCRs. Disruption of *MSH3* led to a 6-fold reduction in GCR rates, whereas*Δmsh6* strains had a modest but statistically significant decrease (1.6-fold) in the level of arm loss events. GCR rates in *mlh1* and *pms1* strains were comparable to those observed in *Δmsh2* mutants. We also examined the effect of *msh2*-G693A which impairs the ATPase activity of Msh2p, but not the ability of Msh2p to form complexes with other MutS proteins (Drotschmann et

al., 1999). In addition, we assessed the effects of the *pms1*-E707K mutation, which disrupts the newly-discovered endonuclease function of MutLα but not other activities of the complex (Kadyrov et al., 2006; Kadyrov et al., 2007). Both of these point mutations reduced the frequencies of the GAA-induced GCRs to about the same extent observed in the *Δmsh2* strain. Disruption of *MSH2* also affected GCRs in TTC strains, leading to a 5-fold decrease (data not shown), indicating that *Y* R·Y triplexes are also targeted by MMR.

These results indicate that MutSβand MutLα heterodimeric complexes are required for both GAA-mediated chromosomal fragility and tract length variations, while the contribution of MutSα is minor. Importantly, the ATPase function of Msh2p and the endonuclease activity of Pms1p are both necessary for the induction of GCRs.

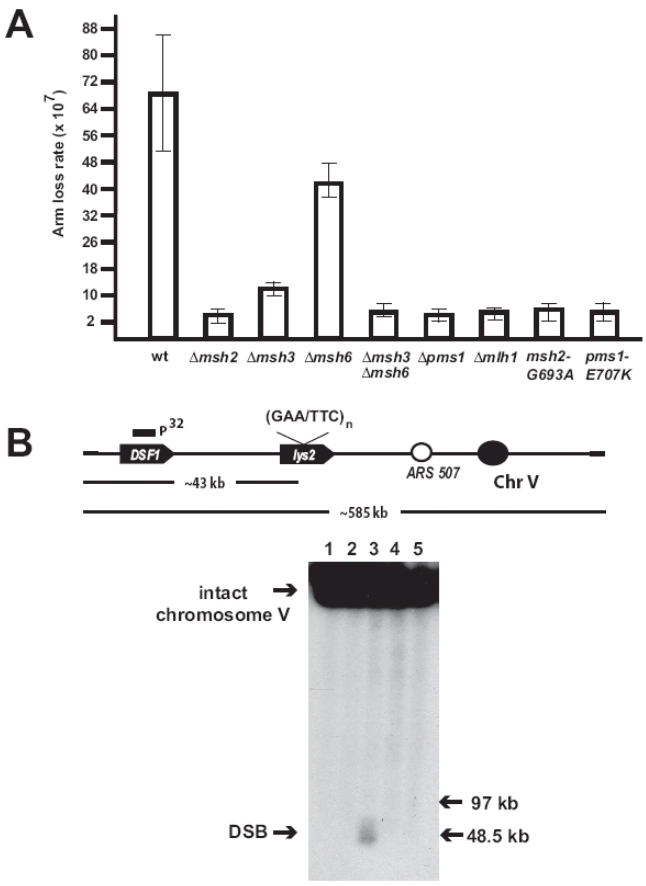


Figure 5. Chromosomal fragility at GAA/TTC tracts requires MMR.
(A) MMR mutants strongly affect GCRs induced by $(GAA)_{230}$ tracts. Values are median rates determined in fluctuation tests using at least 14 cultures. Error bars indicate 95% confidence intervals. (B) Breakage of chromosome V in strains containing GAA/TTC repeats. The position of the GAA/TTC tracts on chromosome V is shown. Chromosomes were separated on the CHEF gel, transferred to a nylon membrane and hybridized with *DSF1*-specific probe to highlight the intact chromosome (~ 585 kb) and broken fragment (~ 43 kb). l ladder was used as a molecular size standard shown on the right. The positions of the marker bands were determined on the ethidium bromide-stained gel prior to Southern blot hybridization. The lanes are: 1, wild-type strain with $(GAA)_{20}$; 2, wild-type strain with $(TTC)_{230}$; 3, wild-type strain with $(GAA)_{230}$; 4, Δ*msh2* strain with $(GAA)_{230}$; 5, *pms1-E707K* strain with $(GAA)_{230}$.

2.2.8. MMR triggers chromosomal breakage at $(GAA)_{230}$ tracts

Induction of GCRs in strains containing GAA/TTC tracts is likely a consequence of the DSB formation at the location of repeats. The rate of GCRs is significantly reduced in MMR-deficient strains, indicating that MMR might be responsible for the breakage. Alternatively, the defect in MMR may negatively affect the repair of the broken molecules by hampering the resection of the DSB intermediates or by reducing the formation and/or extension of heteroduplex intermediates during BIR. To address this issue directly, we analyzed the chromosomal DSB formation in the repeat-containing strains (Figure 5B). No DSBs were detected in strains with $(TTC)_{230}$ but were visible in strains carrying $(GAA)_{230}$ repeats (lane 3, Figure 5B) consistent with their different potential to trigger GCRs and homologous recombination (see above). The breakage in $(GAA)_{230}$ strains was compromised in *Δmsh2* and *pms1*-E707K mutants, suggesting that MMR machinery is required for efficient DSB formation and is not involved in the processing or healing of the broken ends.

2.3. Discussion

Polypurine-polypyrimidine sequences that have potential to adopt triplex secondary structure are highly polymorphic and abundant in eukaryotic genomes, ranging from yeast to humans (Cox and Mirkin, 1997). We have found that in yeast, expanded

GAA/TTC tracts that belong to this class of sequence motifs strongly stimulate chromosomal fragility in a size- and orientation-dependent manner, often culminating in translocations. The mismatch repair machinery is a key player in the repeat-mediated breakage. This study unravels a novel role of MMR and also shows that the triplex-forming repeats can be a potent source of chromosomal aberrations similar to those observed in tumors.

2.3.1. Mechanism of chromosomal fragility induced by the expanded GAA/TTC tracts

In GCR and homologous recombination assays, repeats in both orientations exhibit strong breakage potential, although the fragility is more pronounced when the expanded GAA repeats are present on the lagging strand template during DNA replication (Table 1 and Figure 4A). Consistently, in strains with repeats in the GAA orientation, we detect a prominent replication fork arrest and accumulation of DSBs (Figure 3, 4B and5B). One possible explanation for this orientation bias is that the purine-rich DNA template is not an ideal substrate for the proteins involved into lagging strand DNA synthesis such as RPA and Polα-primase (Frick and Richardson, 2001; Wold, 1997). Hampered Okazaki fragment synthesis would generate long single-stranded regions that could loop out and forming triplexes with the double-stranded region ahead of the fork (Figure 6).

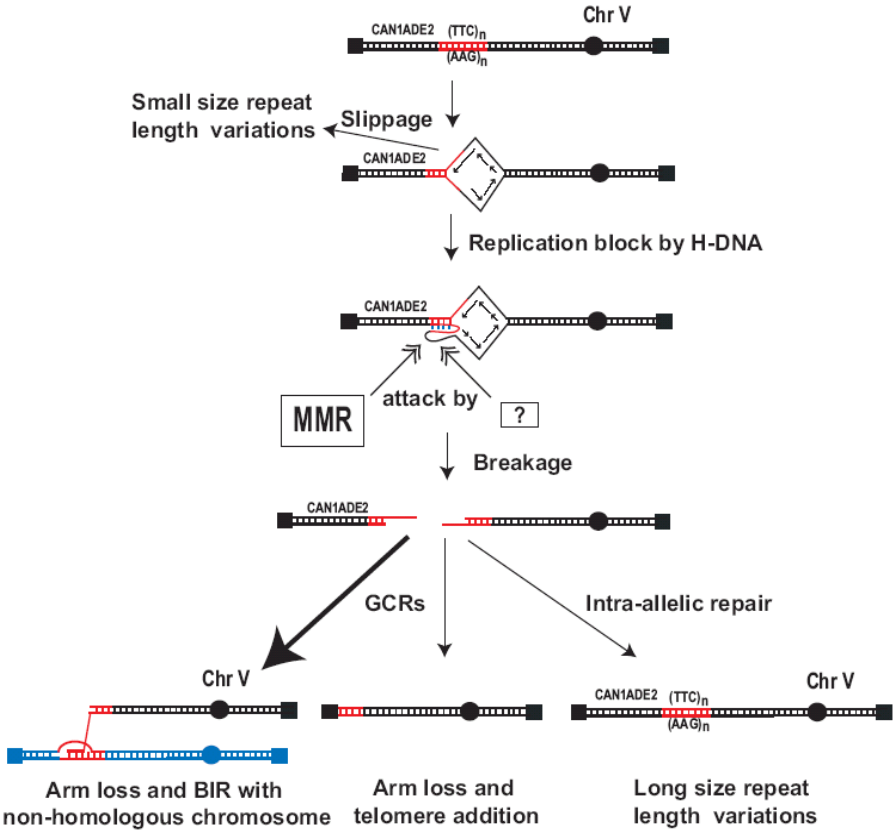


Figure 6. Model for chromosomal fragility and rearrangements mediated by triplex-forming GAA/TTC repeats.

The GAA/TTC tracts are shown (not to scale) in red. Telomeres (filled rectangles) and centromeres (solid circles) are also shown. Diamonds with arrows are the bidirectional replication forks. A non-homologous chromosome is depicted in blue. GAA/TTC tracts are microsatellites that are prone to slippage during DNA synthesis. In MMR-deficient strains, this instability is manifested as small size repeat variations. We hypothesize that a triplex structure will be adopted preferentially when the GAA repeats are located on the lagging strand template. Triplex can arrest replication progression. We suggest that MMR system recognizes and processes the H-DNA leading to DSBs. DSBs can also be introduced by an alternative, MMR- independent minor pathways indicated by the boxed '?'. Following DSB induction, the broken end can be healed through intra-allelic repair (such as NHEJ or single-strand annealing) or homologous recombination with a repetitive tract on the sister chromatid, leading to large size variations.

Alternatively, centromere-containing broken fragment can be repaired by BIR with GAA/TTC-rich regions on non-homologous chromosomes, resulting in arm loss and non-reciprocal translocation. Rarely, the broken end can also be capped by de novo telomere addition.

Alternatively, the difference in the breakage potential could be accounted by the greater stability of the *R*·R·Y secondary structure adopted by repeats in GAA orientation versus the *Y*·R·Y triplex formed due to TTC strand folding. Hence *R*·R·Y H-DNA can be a stronger barrier for replication fork. In both of the cases, arrested fork intermediates are expected to contain the secondary structure that can explain the observed migration pattern of Y molecules in 2D gels (Figure 3).

The triplex can be recognized and targeted by MMR resulting in tract length variations and DSB formation. It should be noted that although MMR is the primary player in the fragility, *Δmsh2* does not completely eliminate the GAA/TTC-induced GCRs (Figure 5A), indicating that the triplex and/or arrested fork can lead to breakage via an alternative MMR-independent pathway. A DSB occurring within the GAA tract is expected to split chromosome V into acentric and centromere-containing fragments. There are several pathways to repair such a break. If the broken ends are repaired by recombination with the allelic unbroken GAA tract on the sister chromatid or if the broken ends are re-joined by NHEJ or single-strand annealing, one would expect to get larger or smaller tracts without an associated translocation. Alternatively, it is possible that the acentric fragment would be lost and the centromere-containing fragment would invade GAA/TTC-rich genomic sequences located on non-homologous chromosomes. This pathway of repair would result in non-reciprocal translocations.

2.3.2. The role of MMR in triplex-mediated instability

Contribution of MMR to trinucleotide repeat instability has been extensively studied, in both prokaryotes and eukaryotes, for CNG tracts. In *Escherichia coli*, defects in MMR lead to a decreased level of large deletions but an elevated rate of small-size alterations in the hairpin-forming CAG/CTG repeats (Jaworski et al., 1995; Parniewski et al., 2000; Schmidt et al., 2000; Schumacher et al., 1998; Wells et al., 1998). In yeast, loss of MMR results in elevated rates of small tract alterations of CAG/CTG and CCG/CGG repeats, but has little effect on the rates of large deletions or insertions (reviewed by Lenzmeier and Freudenreich, 2003). In mice, MMR proteins are involved in regulating somatic and germline instability of CAG/CTG repeats (promoting both expansions and contractions), MutSβ being the major player (Foiry et al., 2006 and references therein). *In vitro* studies show that, although Msh2p or Msh2p-Msh6p and Msh2p-Msh3p complexes efficiently bind DNA hairpins, repair does not occur (Bowers et al., 2000; Pearson et al., 1997; Owen *et al.*, 2005). Owen et al*.,* proposed that inactive MutSβ bound to the secondary structure might prevent its processing thereby promoting tract-length changes.

We have found that similar to observations made with CAG/CTG repeats, disruptions of *MSH2*, *MSH3*, *MLH1* and *PMS1* but not *MSH6* in yeast alter the stability of $(GAA)_{340}$ repeats: decreasing the rate of large deletions and increasing the rate of small deletions (Figure 7 and data not shown). In addition, MMR deficiency reduces the

GAA/TTC-associated fragility. Based on these results, we suggest that besides hairpins, the triplex secondary structure might be another substrate for MMR recognition. However, unlike hairpins, H-DNA is actively processed by MutSβ and MutLα resulting in repeat size variations and DSB formation. This is strongly supported by our data that *msh2*-G693A mutants defective in the ATPase activity of Msh2-complexes or *pms1*-E707K mutants that lack the endonuclease activity of the MutLα exhibit compromised fragility and long size repeat variations.

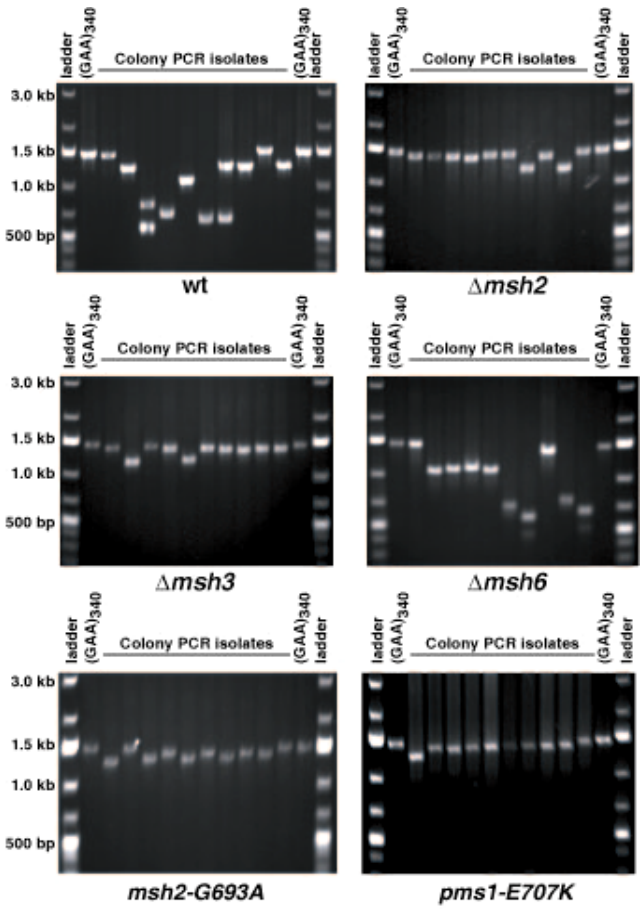


Figure 7. Effect of mutations in MMR on $(GAA)_{340}$ tract stability.

To determine the frequency of expansions and contractions of repeat tracts during mitotic divisions, we re-streaked yeast colonies that have been verified for the presence of $(GAA)_{340}$ full size repeats on complete media. Ten colonies were then selected for PCR amplification to look for changes in the length of the repetitive tracts. The lanes labeled with red asterisks denote unchanged tracts length. The band on the lane labeled $(GAA)_{340}$ is 1389 bp and corresponds to the full-size repeats plus flanking sequences. The primer-pair that reliably worked in colony PCR reaction was determined experimentally. Analysis of tract length variations in *Δmlh1*, *Δpms1, msh2-G693A* and *pms1-E707K* strains yielded results similar to those observed in *Δmsh2* and *Δmsh3* strains.

MMR proteins process multiple DNA distortions that arise during replication, DNA repair, and recombination (reviewed in Jiricny, 2006). Which feature of the triplex secondary structure is recognized by MMR? It is possible that the Hoogsteen base pairs formed between the duplex and the folded strand are a good target for MutSβ binding. Alternatively, the looped-out junction at the border between the duplex and triplex can be a substrate. Biochemical characterization of the binding and the cleavage of the defined triplex substrates by purified MMR proteins, along with solving the structure of MutS complexes (especially MutSβ) bound to H-DNA templates, might help to differentiate between these scenarios.

Along with an important role in the maintaining the integrity of prokaryotic and eukaryotic genomes, MMR is also implicated in the DSB generation as a consequence of 'futile cycles of repair' in cells treated with alkylating agents or antimetabolites (reviewed in Bignami et al., 2003). We suggest that aberrant attempt to repair the triplex structure by MMR during replication can also culminate in DSB formation. The nature of the substrate might dictate the outcome of the MMR attack. Either extensive removal of the third strand involved in the Hoogsteen interaction or nicking of the loop region of the triplex might cause DSBs since the targeted strand lacks the complementary chain.

Eukaryotic genomes contain, besides GAA/TTC tracts, other triplex-forming homopurine·homopyrimidine mirror repeats (Cox and Mirkin, 1997). In humans, several regions that contain non-GAA H-DNA adopting sequences are hotspots for

rearrangements. These include the major breakpoint cluster region at the *BCL2* (Raghavan et al., 2005a), intron 21 of *PKD1* (Blaszak et al., 1999; Patel et al., 2004) and promoter region of *C-MYC* (Michelotti et al., 1996; Wang and Vasquez, 2004). The susceptibility of these regions for aberrations was attributed to the ability of the secondary structures to impede replication progression (Raghavan et al., 2005b) and cause DSBs (Patel et al., 2004). Although the Rag1/Rag2 endonuclease was implicated in promoting DSBs at the bcl2-Mbr locus (Raghavan et al., 2005b), our data strongly suggests that MMR might be an additional player in the breakage formation at the location of triplex structures. It is conceivable that the mechanisms governing GAA-instability might be the same for other H-DNA adopting sequences. Hence, it would be important to assess whether MMR besides GAA/TTC triplexes, can also target non-GAA H-DNA substrates.

2.3.3. Implications for the stability of the human genome

We find that expanded GAA/TTC repeats in yeasts are potent inducers of DSBs and chromosomal aberrations; orientation of the tracts relative to the replication origin is an important factor governing the instability. These data suggest that the human carriers of the expanded tracts such as Friedreich's ataxia patients might be at risk for the formation of chromosome aberrations. It is also conceivable that triplex forming GAA/TTC tracts can act as canonical fragile sites in cytogenetic analyses either spontaneously or upon

induction with chemicals that stabilize triplex structures such as polycyclic compounds (Chan and Glazer, 1997).

Understanding of the molecular mechanisms that govern the stability of the eukaryotic genomes is important for studying the etiology of cancers and hereditary diseases. Based on this study, we propose that chromosomal regions in human carriers containing long triplex-forming repeats are predisposed for breakage and GCRs. We suggest that the length of the repetitive tracts, their location in the genome, and the genetic background may be important factors that determine the susceptibility of the individuals to tumorigenic aberrations.

2.4. Experimental Procedures

2.4.1. Strains and Plasmids

All strains in this study were isogenic to KT119 strain (*MATa*, *his7-2*, *leu2-3,112, trp1-Δ*, *ura3-Δ*, *lys2-Δ, ade2-Δ, bar1-Δ, sfa1-Δ, cup1-1-Δ*, *yhr054c-Δ cup1-2-Δ*, *lys2*::*kanMXURA3*, *ADE2*, *CUP1*, *SFA1* derivative of TP strains described in Narayanan et al., 2006. GAA/TTC repeats of length 20, 60, 120 and 230 were integrated into *LYS2* in two orientations using the *dellito perfetto* technique (Storici et al., 2001). GAA repeats located on the plasmids (Krasilnikova and Mirkin, 2004) were used as the source for PCR amplification. Strains with 340 repeats in both orientations were a result of natural expansions from 230 repeats. The repeats were verified for the absence of interruptions

by sequencing from both ends. MMR genes were disrupted with the *kanMX* cassette (Wach et al., 1994). *msh2-G693A* and *pms1-E707K* alleles were introduced using *dellito perfetto* technique.

2.4.2. Genetic Techniques

The rates and 95% confidence intervals of the arm loss and recombination between *lys2* alleles were estimated in fluctuation tests using at least 14 independent cultures. Single colonies taken into the test were prescreened for the presence of full-sized tracts using colony PCR. The canavanine-containing media was made with a low concentration of adenine (5mg/L) to allow color detection.

2.4.3. Structural Analysis of the Genome Rearrangements

Chromosome aberrations were characterized using CHEF (contour-clamped homogeneous electric field) gels and Southern Blot Hybridization with *MET6*-specific probe. Chromosomal DNA was embedded into agarose plugs using the CHEF Genomic DNA plug Kit from Bio-Rad. Gels were run in 0.5X TBE at 14°C using the Bio-Rad CHEF Mapper XA for 40 hours with switch times of 36.63s-2m6.67s for the analysis of arm loss events.

For DSB detection, chromosomal DNA was embedded into agarose plugs using the CHEF Genomic DNA plug Kit from Bio-Rad (~6 x 10^8 cells/1ml of plug). Gels were

run in 0.5X TBE at 14°C using the Bio-Rad CHEF Mapper XA for 19.44 hours with switch times of 2.08s-42.91s. Southern blot hybridization was performed using 350 bp probe homologous to *DSF1* gene.

Genomic DNA preparation and subsequent CGH microarray analysis of the $Can^{R}Ade^{-}$ isolates were performed according to procedures described (Lemoine et al., 2005). Arrays were analyzed using GenePix pro 4.1 (Axon Instruments) and Gene Spring ®5.1 (Silicon Genetics). A complete analysis of the microarrays can be found on line at GEO database (accession number GSE11425). All microarray analyses were done with the collaboration of Dr. Tom Petes' lab at Duke University. The translocation breakpoints were PCR amplified and sequenced with primers designed based on microarray data.

2.4.4. 2D Analysis of Replication Fork Intermediates

Cells were arrested in G1 with α-factor and released synchronously into S-phase. At 40 min after release, chromosomal DNA was extracted using ultra centrifugation (Friedman and Brewer, 1995). Neutral/neutral 2-D analysis was carried out in according to (Brewer and Fangman, 1987). The *Afl*II-digested DNA was separated in the first dimension on a 0.4% gel without ethidium bromide in 1X TBE buffer at 1V/cm for 38 hours in first dimension. The second dimension gel was run at 6 V/cm in 1X TBE buffer containing 0.3 μg/ml ethidium bromide for 13 hours. Southern blot hybridization using 4 kb *Afl*II-digested *LYS2* fragment was performed to highlight replication intermediates.

2.5. References

Al-Mahdawi S, Pinto RM, Ruddle P, Carroll C, Webster Z, Pook M (2004) GAA repeat instability in Friedreich ataxia YAC transgenic mice. Genomics 84: 301–310

Bidichandani SI, Ashizawa T, Patel PI (1998) The GAA triplet-repeat expansion in Friedreich ataxia interferes with transcription and may be associated with an unusual DNA structure. Am J Hum Genet 62: 111–121

Bignami M, Casorelli I, Karran P (2003) Mismatch repair and response to DNA-damaging antitumour therapies. Eur J Cancer 39: 2142–2149

Bissler JJ (2007) Triplex DNA and human disease. Front Biosci 12: 4536–4546

Blaszak RT, Potaman V, Sinden RR, Bissler JJ (1999) DNA structural transitions within the PKD1 gene. Nucleic Acids Res 27: 2610–2617

Bowers J, Tran PT, Liskay RM, Alani E (2000) Analysis of yeast *MSH2–MSH6* suggests that the initiation of mismatch repair can be separated into discrete steps. J Mol Biol 302: 327–338

Campuzano V, Montermini L, Lutz Y, Cova L, Hindelang C, Jiralerspong S, Trottier Y, Kish SJ, Faucheux B, Trouillas P (1997) Frataxin is reduced in Friedreich ataxia patients and is associated with mitochondrial membranes. Hum Mol Genet 6: 1771–1780

Campuzano V, Montermini L, Molto MD, Pianese L, Cossee M, Cavalcanti F, Monros E, Rodius F, Duclos F, Monticelli A (1996) Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. Science 271: 1423–1427

Chan PP, Glazer PM (1997) Triplex DNA: fundamentals, advances, and potential applications for gene therapy. J Mol Med 75:267–282

Clark RM, Bhaskar SS, Miyahara M, Dalgliesh GL, Bidichandani SI (2006) Expansion of GAA trinucleotide repeats in mammals. Genomics 87: 57–67

Clark RM, Dalgliesh GL, Endres D, Gomez M, Taylor J, Bidichandani SI (2004) Expansion of GAA triplet repeats in the human genome: unique origin of the FRDA mutation at the center of an Alu. Genomics 83: 373–383

Clark RM, De Biase I, Malykhina AP, Al-Mahdawi S, Pook M, Bidichandani SI (2007) The GAA triplet-repeat is unstable in the context of the human FXN locus and displays age-dependent expansions in cerebellum and DRG in a transgenic mouse model. Hum Genet 120: 633–640

Cossee M, Schmitt M, Campuzano V, Reutenauer L, Moutou C, Mandel JL, Koenig M (1997) Evolution of the Friedreich's ataxia trinucleotide repeat expansion: founder effect and premutations. Proc Natl Acad Sci USA 94: 7452–7457

Cox R, Mirkin SM (1997) Characteristic enrichment of DNA repeats in different genomes. Proc Natl Acad Sci USA 94: 5237–5242

De Biase I, Rasmussen A, Bidichandani SI (2006) Evolution and instability of the GAA triplet repeat sequence in Friedrich's ataxia. In Genetic Instabilities and Neurological Diseases, Robert RD, Ashizawa T (eds), pp 305–320. New York: Academic Press

Drotschmann K, Clark AB, Tran HT, Resnick MA, Gordenin DA, Kunkel TA (1999) Mutator phenotypes of yeast strains heterozygous for mutations in the MSH2 gene. Proc Natl Acad Sci USA 96: 2970–2975

Foiry L, Dong L, Savouret C, Hubert L, te Riele H, Junien C, Gourdon G (2006) Msh3 is a limiting factor in the formation of intergenerational CTG expansions in DM1 transgenic mice. Hum Genet 119: 520–526

Frank-Kamenetskii MD, Mirkin SM (1995) Triplex DNA structures. Annu Rev Biochem 64: 65–95

Frick DN, Richardson CC (2001) DNA primases. Annu Rev Biochem 70: 39–80

Gacy AM, Goellner GM, Spiro C, Chen X, Gupta G, Bradbury EM, Dyer RB, Mikesell MJ, Yao JZ, Johnson AJ (1998) GAA instability in Friedreich's ataxia shares a common, DNA-directed and intraallelic mechanism with other trinucleotide diseases. Mol Cell 1: 583–593

Grabczyk E, Usdin K (2000a) Alleviating transcript insufficiency caused by Friedreich's ataxia triplet repeats. Nucleic Acids Res 28: 4930–4937

Grabczyk E, Usdin K (2000b) The GAA*TTC triplet repeat expanded in Friedreich's ataxia impedes transcription elongation by T7 RNA polymerase in a length and supercoil dependent manner. Nucleic Acids Res 28: 2815–2822

Jaworski A, Rosche WA, Gellibolian R, Kang S, Shimizu M, Bowater RP, Sinden RR, Wells RD (1995) Mismatch repair in Escherichia coli enhances instability of CTGn

triplet repeats from human hereditary diseases. Proc Natl Acad Sci USA 92: 11019–11023

Jiricny J (2006) The multifaceted mismatch-repair system. Nat Rev Mol Cell Biol 7: 335–346

Kadyrov FA, Dzantiev L, Constantin N, Modrich P (2006) Endonucleolytic function of MutLalpha in human mismatch repair. Cell 126: 297–308

Kadyrov FA, Holmes SF, Arana ME, Lukianova OA, O'Donnell M, Kunkel TA, Modrich P (2007) Saccharomyces cerevisiae MutLalpha is a mismatch repair endonuclease. J Biol Chem 282: 37181–37190

Krasilnikova MM, Mirkin SM (2004) Replication stalling at Friedreich's ataxia GAAn repeats in vivo. Mol Cell Biol 24: 2286–2295

Krasilnikova MM, Samadashwily GM, Krasilnikov AS, Mirkin SM (1998) Transcription through a simple DNA repeat blocks replication elongation. EMBO J 17: 5095–5102

Lenzmeier BA, Freudenreich CH (2003) Trinucleotide repeat instability: a hairpin curve at the crossroads of replication, recombination, and repair. Cytogenet Genome Res 100: 7–24

Lobachev KS, Shor BM, Tran HT, Taylor W, Keen JD, Resnick MA, Gordenin DA (1998) Factors affecting inverted repeat stimulation of recombination and deletion in Saccharomyces cerevisiae. Genetics 148: 1507–1524

Malkova A, Ivanov EL, Haber JE (1996) Double-strand break repair in the absence of RAD51 in yeast: a possible role for breakinduced DNA replication. Proc Natl Acad Sci USA 93: 7131–7136

Michelotti GA, Michelotti EF, Pullner A, Duncan RC, Eick D, Levens D (1996) Multiple single-stranded cis elements are associated with activated chromatin of the human c-myc gene in vivo. Mol Cell Biol 16: 2656–2669

Mirkin SM (2007) Expandable DNA repeats and human disease. Nature 447: 932–940

Napierala M, Dere R, Vetcher A, Wells RD (2004) Structure-dependent recombination hot spot activity of GAA.TTC sequences from intron 1 of the Friedreich's ataxia gene. J Biol Chem 279: 6444–6454

Narayanan V, Mieczkowski PA, Kim HM, Petes TD, Lobachev KS (2006) The pattern of gene amplification is determined by the chromosomal location of hairpin-capped breaks. Cell 125: 1283–1296

Ohshima K, Kang S, Larson JE, Wells RD (1996) Cloning, characterization, and properties of seven triplet repeat DNA sequences. J Biol Chem 271: 16773–16783

Ohshima K, Montermini L, Wells RD, Pandolfo M (1998) Inhibitory effects of expanded GAA.TTC triplet repeats from intron I of the Friedreich ataxia gene on transcription and replication in vivo. J Biol Chem 273: 14588–14595

Owen BA, Yang Z, Lai M, Gajek M, Badger II JD, Hayes JJ, Edelmann W, Kucherlapati R, Wilson TM, McMurray CT (2005) CAGn-hairpin DNA binds to Msh2–Msh3 and changes properties of mismatch recognition. Nat Struct Mol Biol 12: 663–670

Pandolfo M (2006) Friedreich's ataxia. In Genetic Instabilities and Neurological Diseases, Robert RD, Ashizawa T (eds), pp 277–296. New York: Academic Press

Parniewski P, Jaworski A, Wells RD, Bowater RP (2000) Length of CTG.CAG repeats determines the influence of mismatch repair on genetic instability. J Mol Biol 299: 865–874

Patel HP, Lu L, Blaszak RT, Bissler JJ (2004) PKD1 intron 21: triplex DNA formation and effect on replication. Nucleic Acids Res 32: 1460–1468

Pearson CE, Ewel A, Acharya S, Fishel RA, Sinden RR (1997) Human MSH2 binds to trinucleotide repeat DNA structures associated with neurodegenerative diseases. Hum Mol Genet 6:1117–1123

Pollard LM, Sharma R, Gomez M, Shah S, Delatycki MB, Pianese L, Monticelli A, Keats BJ, Bidichandani SI (2004) Replicationmediated instability of the GAA triplet repeat mutation in Friedreich ataxia. Nucleic Acids Res 32: 5962–5971

Raghavan SC, Chastain P, Lee JS, Hegde BG, Houston S, Langen R, Hsieh CL, Haworth IS, Lieber MR (2005a) Evidence for a triplex DNA conformation at the bcl-2 major breakpoint region of the t14;18 translocation. J Biol Chem 280: 22749–22760

Raghavan SC, Swanson PC, Ma Y, Lieber MR (2005b) Double-strand break formation by the RAG complex at the bcl-2 major breakpoint region and at other non-B DNA structures in vitro. Mol Cell Biol 25: 5904–5919

Raghuraman MK, Winzeler EA, Collingwood D, Hunt S, Wodicka L, Conway A, Lockhart DJ, Davis RW, Brewer BJ, Fangman WL (2001) Replication dynamics of the yeast genome. Science 294: 115–121

Rothstein R, Michel B, Gangloff S (2000) Replication fork pausing and recombination or 'gimme a break'. Genes Dev 14: 1–10

Sakamoto N, Chastain PD, Parniewski P, Ohshima K, Pandolfo M, Griffith JD, Wells RD (1999) Sticky DNA: self-association properties of long GAA.TTC repeats in R.R.Y triplex structures from Friedreich's ataxia. Mol Cell 3: 465–475

Schmidt KH, Abbott CM, Leach DR (2000) Two opposing effects of mismatch repair on CTG repeat instability in Escherichia coli. Mol Microbiol 35: 463–471

Schumacher S, Fuchs RP, Bichara M (1998) Expansion of CTG repeats from human disease genes is dependent upon replication mechanisms in Escherichia coli: the effect of long patch mismatch repair revisited. J Mol Biol 279: 1101–1110

Sia EA, Kokoska RJ, Dominska M, Greenwell P, Petes TD (1997) Microsatellite instability in yeast: dependence on repeat unit size and DNA mismatch repair genes. Mol Cell Biol 17: 2851–2858

Smith CE, Llorente B, Symington LS (2007) Template switching during break-induced replication. Nature 447: 102–105

Sutherland GR (2003) Rare fragile sites. Cytogenet Genome Res 100: 77–84

Wang G, Vasquez KM (2004) Naturally occurring H-DNA-forming sequences are mutagenic in mammalian cells. Proc Natl Acad Sci USA 101: 13448–13453

Wells RD (2008) DNA triplexes and Friedreich ataxia. FASEB J 22: 1625–1634

Wells RD, Parniewski P, Pluciennik A, Bacolla A, Gellibolian R, Jaworski A (1998) Small slipped register genetic instabilities in Escherichia coli in triplet repeat sequences associated with hereditary neurological diseases. J Biol Chem 273: 19532–19541

Wold MS (1997) Replication protein A: a heterotrimeric, singlestranded DNA-binding protein required for eukaryotic DNA metabolism. Annu Rev Biochem 66: 61–92

Yabuki N, Terashima H, Kitada K (2002) Mapping of early firing origins on a replication profile of budding yeast. Genes Cells 7: 781–789

Table 1. Length and orientation-dependent induction of GCRs and homologous recombination by GAA/TTC repeats.

[a] The loss of *CAN1* and *ADE2*–containing region was measured in strains that do not have the *lys2-8* allele

[b] Numbers in parentheses correspond to the 95% confidence interval

[c] ND - not determined

Insertion in *LYS2*		Arm loss rate (x 10^{10})[a]	Recombination rate (x 10^{7})
Orientation	Tract length		
TTC	20	8 (4-17)[b]	5 (4-9)
	60	21 (19-35)	8 (6-11)
	120	4515 (3226-6149)	19 (14-23)
	230	4778 (4409-6128)	20 (17-36)
	340	4851 (4133-6659)	ND[c]
GAA	20	9 (5-15)	7 (5-9)
	60	27 (20-43)	17 (12-27)
	120	17436 (9120-24770)	287 (170-373)
	230	66044 (51054-86480)	3808 (2410-4384)
	340	743659 (468938-1085186)	ND

Table 2. Can^{R}Ade^{-} isolates of strains with $(TTC)_{230}$ and $(GAA)_{230}$ repeats

[a] The percentage of isolates that belong to the GCR class is indicated in parentheses.

[b] Blue and red arrows indicates the point of junction with the invaded TTC or GAA strands from chromosome V, respectively. Multiple arrows indicate invasion point identified for independent isolates

Isolate	Rearrangement	Breakpoint ORF	GAA/TTC-rich sequence used for invasion and BIR[b]
T-1, T-3, T-4, T-5, T-8, T-11, T-12	deletion+ non reciprocal translocation (Left arm, Chr.I), 7/12 (58%)[a]	*FUN12*	CTTCGTTAGCTTC←TTCTTCGTTCTT←TTCTTCCAATAACTTCTCTTCTTCCTACCATTCGGATATGATTTCAGGTTCTTCCTTTTCCTCTTCCTTTTTTTCCTTTTCCTCTTCGTCTTCTTCTTTCTT
T-7, T-9	deletion+non reciprocal translocation (Left arm, Chr.XIII), 2/12 (17%)	*ORC1*	CTCCTTCTTCTTCTGCTTCT←TCTTCTGCTTCTGCTTCTTCTCTTTCTT
T-6	deletion+non reciprocal translocation (Left arm, Chr.II), 1/12(8%)	*SCT1*	TCCCTTCTCCTTCTCCTTCTTCTTCTCCTTCTTC←TTCTTCTCCTTCTTCTTCTTCTTCCCTTTCTTCCTC
T-2, T-10	deletion+telomere addition, 2/12 (17%)	-	-
A-1,A-2, A-3, A-5, A-6, A-7, A-9, A-10, A-11, A-12	deletion+non reciprocal translocation (Left arm, Chr.XI), 10/12 (83%)	*MNN4*	GAGGAGGACGAAGAAAAAAAGG←AGAAGAAGCAAGAAGAAGAAAAGAAGGAGAAGGACGAAGAAGAAAAGAAGTAGAAGTAAGAAGAACAAAAAGAGAAGGACGAAGAAAAAAAGGAGAAGAAGCAAGAAGAAGAAAAGAAGTAGAAGTAAGAAGAATAAAAGAAGTAGAAGTAAGAAGTAGAAAAGAGGAAGAAGGACGAAGAAGAAGA←GAAGAAGAAGGAAGAAGAA←AAAGAGGAGAAGGAGGACGAAGAAGAAAAGAAGAAGAAGGAAGAAAAAGAAGAGAAGAAGAAGGAAGAAAAAGAAAAGAAGAAGAAGGAAGAAGAAAAGGAGAAGAAGGAAGAAGAAGAAGAGGAGAAGGAGGAAGAAGAAGAAGAGGGAGGCGAAAGAAAGAAG

A-4, A-8	deletion+non reciprocal translocation (Left arm, Chr.XIII), 2/12 (17%)	*FPR3*	GAAGAAATGAAAAAATCGAAAGAAAA GAAGAAGAA←AGAGAA←GCACGAAA GAGAAAAAAAGGAAAAACGAAAAGAA TCCAAGACCGAATTGAAGAAGAAAAA CAAGAAGAAGAAGAAGAAG

Table 3. Length- and orientation-dependent size variations in expanded GAA and TTC repeat tracts.

* The single colonies verified for the presence of full-size repeat tracts were re-streaked on YPD media and size changes in the progeny colonies were determined by PCR analysis on genomic DNA (20-50 ng was used as template). Colonies derived from three independent isolates for each tract length were examined. Expansions and contractions of the repeats were scored if the PCR analysis revealed the presence of a novel band different from the original repeat length. In most of the cases, the PCR products contained only one discrete band.

Insertion in *LYS2*	Number of colonies examined	Contraction* (%)	Expansion* (%)	Unchanged (%)
$(GAA)_{60}$	120	<1	<1	100
$(GAA)_{230}$	120	3.3	1.7	95
$(GAA)_{340}$	120	88.3	5.8	5.8
$(TTC)_{60}$	60	<1.7	<1.7	100
$(TTC)_{230}$	120	2.5	1.7	95.8
$(TTC)_{340}$	120	4.2	3.3	92.5

Chapter III

Triplex forming non-GAA/TTC mirror repeats induce chromosome fragility in *S.cerevisiae*.

- The study results in chapter III are not published yet.

- All experiments were done by me.

3.1. Introduction

3.1.1. Non-canonical secondary structures are the sources of genome instabilities

Prokaryotic and eukaryotic genomes are enriched with repetitive sequences (Cox & Mirkin, 1997). One class of these repeat sequences can adopt non-canonical secondary DNA structures (non-B DNA structures). Studies from model organisms and human cells implicate the non-canonical secondary structures forming sequences to be hot spots for deletions, amplification and translocations (Wells, 2007).

Several studies have provided a link between secondary structures and genome rearrangements. Palindrome (Hairpin- and cruciform-forming repetitive sequences) mediates rearrangements by triggering in specific patterns (terminal deletions coupled with adjacent duplications, (Narayanan, 2006)) and facilitate an increase in copy number (Figure 1), (Tanaka et al, 2007). These types of secondary structure include inverted repeats (IRs), AT- and GC-rich micro and mini satellites. In human, palindromic AT-rich repeats (PATRRs) are hotspots translocations. These translocations can result in reproductive problems in carriers or in an inherited syndrome, neurofibromatosis type I. Three types of PATRRs (11, 17, and 22) that induce human genome instability have been identified so far (reviewed in (Kurahashi et al, 2006)). Also, amplified ~33bp and ~42bp AT rich micro and mini satellites are co localized at two rare fragile sites (FRA10B and FRA16B) in human cells. In addition to palindromic AT-rich repeats, CAG/CTG or CCG/CGG trinucleotide repeats can adopt hairpin or cruciform structures and often found in rare fragile sites in human (Sutherland, 2003). CCG-rich micro satellite (up to 2000 copies repeats) are co localized at six rare fragile sites (FRAXA, FRAXE, FRAXF, FRA10A, FRA11B and FRA16A) out of eight total. G-quartet is another type of non-canonical secondary

structure suggested to be formed at CCG/CGG trinucleotide repeat and GC rich mini satellites (Figure 1).

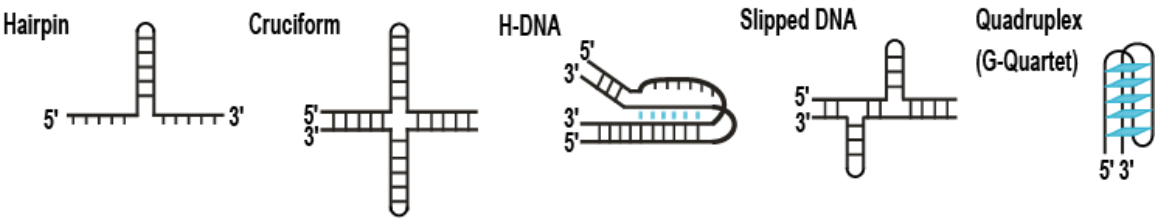


Figure 1. Examples of non-canonical DNA secondary structures.

Black line depicts Watson-Crick hydrogen base pairing. Blue line represents Hoogsteen hydrogen base pairing.

3.1.2. Triplex forming non-GAA/TTC mirror repeats have been reported to be hotspots for translocations or mutations in the human genome.

Triplex DNA-forming GAA/TTC tracts are another type of non canonical secondary structures which can induce genome fragility. For examples, long tracts of GAA/TTC repeats block DNA replication and are strong hot-spots for double-stranded break (DSB) formation and gross chromosome rearrangements (Kim et al, 2008; Krasilnikova & Mirkin, 2004; Wells, 2008).

In addition to GAA/TTC repeats, non-GAA/TTC sequences with triplex-forming potential have been identified as potent inducers of genome aberration in the human genome. These include major breakpoint cluster region (MBR) at the *BCL-2* gene (Raghavan et al, 2005a), intron 21 of *PKD1* (Blaszak et al, 1999; Patel et al, 2004) and promoter region of *C-MYC* (Michelotti et al, 1996; Wang & Vasquez, 2004).

Major breakpoint region of *BCL2*

The *BCL-2* major breakpoint region (MBR, 150bp), involved in the t(14;18) translocation of nearly all follicular lymphomas, is the most common of all lymphoid cell translocation sites ((Raghavan & Lieber, 2006), Figure2). After break happens at the immunoglobulin heavy chain locus, normally the D and J are joined together. In the t(14;18), the D and J coding ends fail to join mutually. In addition to this error, a break occurs at the 3' UTR of *BCL-2* gene. Around 75% of breaks are localized in a 150 bp major breakpoint region.

The *BCL-2* MBR can adopt a non-B form structure *in vivo* (Raghavan et al, 2004). It

contains three breakage hotspots and their sequences considered as interacting with the VDJ recombinase complex even though no homology with the traditional heptamer/space/nonamer sequences exist. It was suggested that the creation of DSBs is dependent on triplex DNA secondary structure and the RAG complex (RAG1, RAG2 and HMG1). RAG complex generates DSBs by creating independent nicks close to one another on the two strands at the MBR in the human cell (Raghavan et al, 2005b). Although RAG complex was implicated in promoting DSBs at the MBR locus, our data suggested that mismatch repair (MMR) might be an additional player in the break formation at the location of GAA/TTC triplex structures. It would be important to assess whether MMR, besides GAA/TTC triplexes, can also target non-GAA/TTC triplex DNA (Kim et al, 2008).

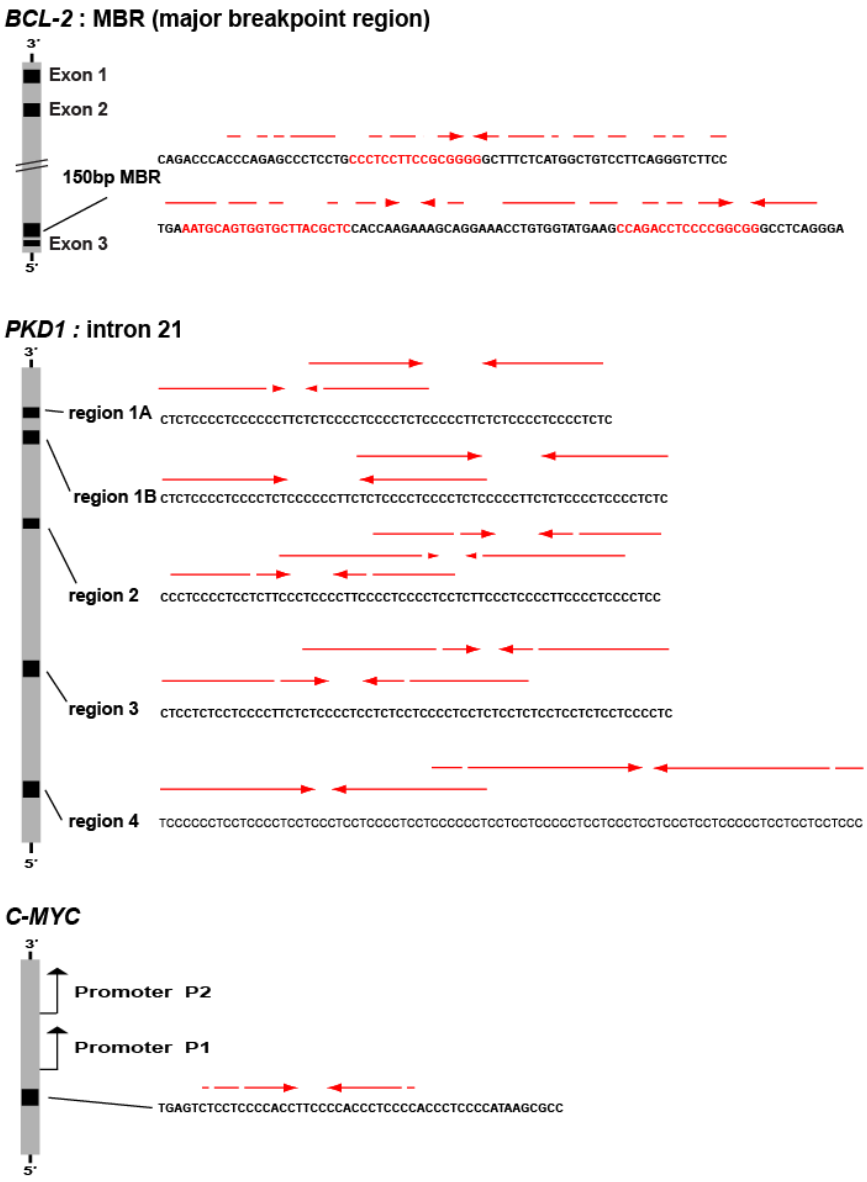


Figure 2. Triplex forming mirror repeats in *BCL-2*, *PKD1* and *C-MYC* regions

Vertical gray ideograms on the left depict the location of mirror repeats (black bars) within the three regions. The red arrows indicate mirror repeats. The red color in nucleotide sequence represents peak regions in MBR.

PKD1

Polycystic kidney disease (PKD) is an autosomal dominant disease and affects over 500,000 Americans. 85 % of PKD patients have mutations in the *PKD1* gene. The nature of cyst formation has been attributed to instability in the *PKD1* gene.

Intron 21 of *PKD1* gene contains the largest polypurine - polypyrimidine tract in the human genome (Figure 2). Several studies represented that this 2.5 kbp long poly purine –pyrimidine mirror repeat is capable of forming multiple non-B-DNA structures (Blaszak et al, 1999) and causes genome instability. Polymerase arrests within the Pu-Py tracts only in one direction of replication by primer extension study. Also Pu-Py tracts induce a replication blockage and double-stranded breaks in a SV40 *in vitro* replication assay with HeLa cell extracts (Patel et al, 2004). Bacolla *et al*., suggested that nucleotide excision repair (NER) recognize the triplex DNA structures formed by the *PKD1* poly(R.Y) tracts as 'lesions' and cleaved them on a plasmid in *E.coli* (Bacolla et al, 2001). It would be interesting to determine whether mismatch repair (MMR) recognize triplex DNA structures formed at intron21 region of *PKD1*.

Promoter region of *C-MYC*

Myc (c-Myc) gene encodes for a transcription factor that regulates expression of many genes (~15%) through binding on enhancer box sequences and recruiting histone acetyltransferases (HATs) (Ruf et al, 2001). A mutated Myc doesn't bind to DNA of other genes correctly and permanently expressed. This leads to the unregulated

expression of many genes. Some of which are involved in cell proliferation and this results in the formation of tumor (Aulmann et al, 2002; Pestov et al, 1991). Translocation sites at Myc are found at t(8:14) Burkitt's lymphoma (Saglio et al, 1993) and t(12;15) BALB/c plasmacytoma (Kovalchuk et al, 1997). Translocation and their breakpoints are clustered at the ~27 base pair promoter regions of *C-MYC* gene in cancer cells. This short promoter regions have been suggested as the triplex forming sequences ((Wang & Vasquez, 2004), Figure 2). On the contrary, Simosson et al., suggested that the promoter of *C-MYC* forms G-quartet structure *in vitro* rather than triplex DNA (Simonsson et al, 1998) (Rangan et al, 2001). Possible secondary structures formed by promoter region of *C-MYC* remains to be verified.

3.1.3. *Friedreich's ataxia* GAA/TTC tracts are strong hot-spots for DSB formation and gross chromosome rearrangements.

We have demonstrated that triplex DNA-forming *Friedreich's ataxia* GAA/TTC tracts are strong hot-spots for DSB formation and gross chromosome rearrangements (Kim et al, 2008). The chromosome fragility strongly depends on the length of the tracts and orientation of the repeats relative to the replication origin. This correlates with their propensity to adopt triplex structure and to block replication progression. We showed that fragility is mediated by mismatch repair machinery and requires the MutSβ and endonuclease activity of MutLα. We have suggested that the mechanism of GAA/TTC-induced chromosomal aberrations in yeast model system.

In this study, we have assessed chromosome fragility induced at non-GAA/TTC mirror repeat tracts and have compared their potential to that of GAA/TTC tracts. Although DSB formation has been studied at non-GAA/TTC triplex forming sequences, the underlying mechanism how breaks occurring and what are the key players required for fragility are yet to be illustrated comprehensively. Study results from this research will help to elucidate the mechanism underlying the fragility of non-GAA/TTC mirror repeats tracts by comparing with the mechanism of GAA/TTC.

3.2. Results

3.2.1. Experimental system

To monitor the breakage at the site of the secondary structure forming repeat tracts we have employed a sensitive assay system based on *S.cerevisiae*. This experimental assay was identical as described in chapter II 2.2.1 (Figure 3.).

In brief, we have built non-GAA/TTC mirror repeats on chromosome V in yeast. The *LYS2* cassettes contain three different non-GAA/TTC mirror repeat motifs. 1) 50 bps promoter region of *C-MYC*, 2) 150 bps MBR of *BCL-2* , 3) 88 bps Peak 4 of *PKD1* (Peak 4 is the longest mirror repeat among 4 peak regions at intron 21 of *PKD1* gene, Figure 2). All cloned mirror repeats were sequenced and verified. All mirror repeat insertions result in loss of *LYS2* function. Repeats were inserted into the chromosome in two different orientations with respect to the direction of replication except peak4 of *PKD1*. Replication is initiated at the *ARS507* origin and proceeds from right to left in this region ((Raghuraman et al, 2001; Yabuki et al, 2002). The lagging strand template contains

>50% purines in “Purine-orientation” and > 50% pyrimidines in “Pyrimidine-orientation”. In MBRL-Pu and MBRL-Py strains, we have replaced the potential hotspots in the peak regions {Patel, 2004 #52} with normal duplex forming sequences to verify the chromosome instability induced from triplex DNA secondary structure.

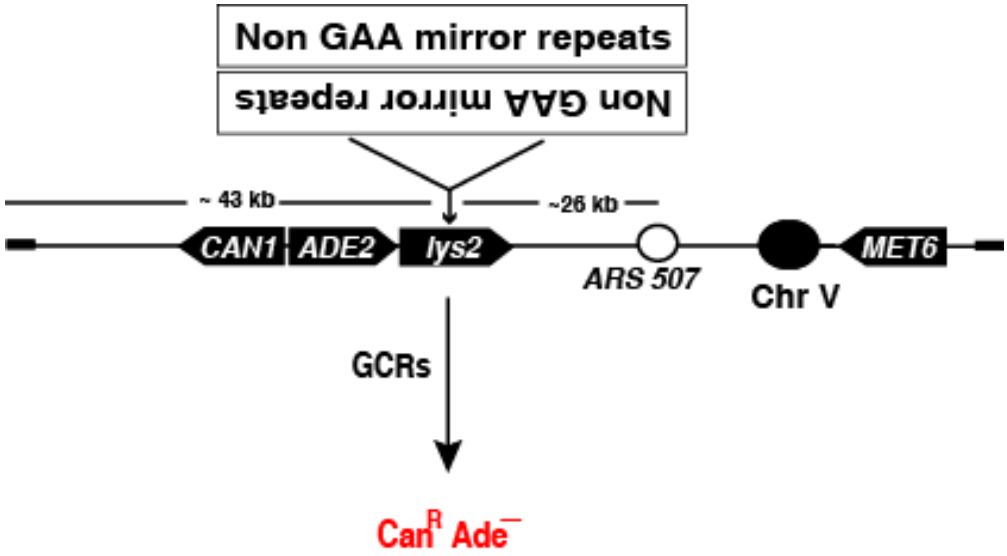


Figure 3. Experimental system to study chromosomal fragility induced by non-GAA/TTC Pu•Py mirror repeats.

Experimental system to study chromosomal fragility induced by expanded tracts of triplex forming mirror repeats. The breakage at the location of mirror repeats can lead to 43 kb telomere-proximal deletion resulting in $Can^{R}Ade^{-}$ clones. The *LYS2* cassettes contain three different non-GAA/TTC mirror repeat motifs were built on *LYS2* region. 1) 50 bps promoter region of *C-MYC*, 2) 150 bps MBR of *BCL-2* , 3) 88 bps Peak 4 of *PKD1* (Peak 4 is the longest mirror repeat among 4 peak regions at intron21 of *PKD1* gene).

3.2.2. Triplex forming mirror repeats induce GCRs in an orientation-independent manner

There was an increase (~26 fold) in levels of *CAN1* region loss for MBR-Pu and MBR-Py triplex forming motifs over strains containing $(GAA)_{20}$ {Kim, 2008 #41}. However, no significant orientation dependency was observed (Table 1).

MBRL-Pu and MBRL-Py strains which contain normal Watson-Crick duplex sequences replacing triplex forming hotspots inside the peak regions show ~ 2 fold reduction compared to MBR-Pu and MBR-Py strains. However the arm loss rate for MBR-Pu and MBRL-Py strains were ~ x14 higher than $(GAA)_{20}$. The degree of stimulation was not notably depended on the orientation of the repetitive tracts in these cases.

Similar to MBR-Pu and MBR-Py strains, the peak4 region of *PKD1*-Pu induced fragility by ~27 fold over the control strains. Also, *C-MYC*-Py and -Pu motif sequences induced rate of arm loss by 5 – 6 folds over control strains. However, no significant difference in the rate of arm loss was observed in purine and pyrimidine orientation.

3.3. Discussion

MBR-Pu, MBR-Py and the peak4 region of *PKD1* mirror repeat tracts increase arm loss by 26~27 fold over the strains containing $(GAA)_{20}$. *C-MYC*-Py and -Pu sequences motifs induced fragility by 5~6 fold over control strains. In all cases, however, no significant orientation dependency was observed in Pu and Py orientation (Table 1).

3.3.1. The fragilities of non-GAA/TTC mirror repeats were not dependent on orientation significantly.

It is interesting to note that the fragilities of non-GAA/TTC mirror repeats were not dependent on orientation significantly. This might imply an idea that there might be two distinctive mechanisms of fragility at the triplex forming mirror repeats. One simple idea is that GAA/TTC tracts contain repetitive three nucleotides and they might form mismatch containing- triplex secondary structures. Unlike GAA/TTC repeats, non-GAA/TTC mirror repeats tracts do not have any repeat tracts and form mismatch free triplex DNA. Consequently, their mechanism of genome fragility can be different from that of GAA/TTC repeats.

Alternatively, the fragility of non-GAA/TTC mirror repeat tracts might not strong enough to trigger orientation dependency (Table 1). In case of GAA/TTC trinucleotide repeats, orientation dependency was started to be observed from $(GAA/TTC)_{120}$ tracts (Table 1), (Kim et al, 2008). As the fragilities of non-GAA/TTC mirror repeat tracts are lower than $(GAA)_{120}$ and/or $(TTC)_{120}$ by 100 ~ 300 fold, they might not strong enough to reveal their nature of orientation dependency. It is likely that

all non-GAA/TTC mirror repeat tracts are not long enough to form stable triplex DNA structures compared to $(GAA)_{120}$ or $(TTC)_{120}$ repeats (50 bp of *C-MYC,* 150 bp of MBR, 88 bp peak4 region of *PKD1,* and 360 bp of $(GAA)_{120}$ or $(TTC)_{120}$). Hence, the triplex DNA secondary structure of non-GAA/TTC mirror repeat tracts might not long and strong enough to trigger orientation dependency (Table 1). In this case, this barrier can be overcome by mutant backgrounds which can increase fragility. Also, small molecules which might stabilize the triplex DNA specifically and induce fragility might be used as useful tools to study mechanism of fragility governed by non-GAA/TTC mirror repeats (In chapter IV, we have investigated the effect of triplex-specific small molecules on GAA-mediated fragility).

In MBRL-Pu and MBRL-Py strains, Watson-Crick duplex sequences replacing triplex forming peak regions and they exhibited x14 higher fragility over the control strains and 2 fold reductions over the parental MBR-Pu and MBR-Py strains. This result suggests that it is not only hotspots in peak regions but also flanking mirror repeat sequences can still contribute to fragility induced at MBR (Figure 2).

We are unable to compare orientation dependency as we have successfully build *PKD1*-Py orientation in our system yet to. This might imply that *PKD1*-Py orientation is more prone to adopt the secondary structure than the other orientation, however, its precise explanation need to be investigated.

Another interesting question is to indentify replicaion blockage or DSBs formation at the location of non-GAA/TTC mirror repeat tracts. Although the fragility potential of non-GAA/TTC mirror repeats are not strong enough to be detected on a gel based on our previous experiments with GAA/TTC repeats. Ligation mediated PCR can be helpful techniques to detect DSBs {Xu, 1996 #131}.

The effect of mismatch repair defective mutation on the fragility of non-GAA/TTC mirror repeats genome instability will be interesting in the future. This result will help us to determine if fragility potential of non-GAA/TTC tracts is similar to that of the GAA triplet repeats.

3.4. References

Aulmann S, Bentz M, Sinn HP (2002) C-myc oncogene amplification in ductal carcinoma in situ of the breast. *Breast Cancer Res Treat* **74**(1)**:** 25-31

Bacolla A, Jaworski A, Connors TD, Wells RD (2001) Pkd1 unusual DNA conformations are recognized by nucleotide excision repair. *J Biol Chem* **276**(21)**:** 18597-18604

Blaszak RT, Potaman V, Sinden RR, Bissler JJ (1999) DNA structural transitions within the PKD1 gene. *Nucleic Acids Res* **27**(13)**:** 2610-2617

Cox R, Mirkin SM (1997) Characteristic enrichment of DNA repeats in different genomes. *Proc Natl Acad Sci U S A* **94**(10)**:** 5237-5242

Goni JR, de la Cruz X, Orozco M (2004) Triplex-forming oligonucleotide target sequences in the human genome. *Nucleic Acids Res* **32**(1)**:** 354-360

Hackett JA, Feldser DM, Greider CW (2001) Telomere dysfunction increases mutation rate and genomic instability. *Cell* **106**(3)**:** 275-286

Kim HM, Narayanan V, Mieczkowski PA, Petes TD, Krasilnikova MM, Mirkin SM, Lobachev KS (2008) Chromosome fragility at GAA tracts in yeast depends on repeat orientation and requires mismatch repair. *EMBO J* **27**(21)**:** 2896-2906

Kovalchuk AL, Muller JR, Janz S (1997) Deletional remodeling of c-myc-deregulating chromosomal translocations. *Oncogene* **15**(19)**:** 2369-2377

Krasilnikova MM, Mirkin SM (2004) Replication stalling at Friedreich's ataxia (GAA)n repeats in vivo. *Mol Cell Biol* **24**(6)**:** 2286-2295

Kurahashi H, Inagaki H, Ohye T, Kogo H, Kato T, Emanuel BS (2006) Chromosomal translocations mediated by palindromic DNA. *Cell Cycle* **5**(12)**:** 1297-1303

Michelotti GA, Michelotti EF, Pullner A, Duncan RC, Eick D, Levens D (1996) Multiple single-stranded cis elements are associated with activated chromatin of the human c-myc gene in vivo. *Mol Cell Biol* **16**(6)**:** 2656-2669

Narayanan BA (2006) Chemopreventive agents alters global gene expression pattern: predicting their mode of action and targets. *Curr Cancer Drug Targets* **6**(8)**:** 711-727

Patel HP, Lu L, Blaszak RT, Bissler JJ (2004) PKD1 intron 21: triplex DNA formation and effect on replication. *Nucleic Acids Res* **32**(4)**:** 1460-1468

Pestov DG, Dayn A, Siyanova E, George DL, Mirkin SM (1991) H-DNA and Z-DNA in the mouse c-Ki-ras promoter. *Nucleic Acids Res* **19**(23)**:** 6527-6532

Raghavan SC, Chastain P, Lee JS, Hegde BG, Houston S, Langen R, Hsieh CL, Haworth IS, Lieber MR (2005a) Evidence for a triplex DNA conformation at the bcl-2 major breakpoint region of the t(14;18) translocation. *J Biol Chem* **280**(24)**:** 22749-22760

Raghavan SC, Lieber MR (2006) DNA structures at chromosomal translocation sites. *Bioessays* **28**(5)**:** 480-494

Raghavan SC, Swanson PC, Ma Y, Lieber MR (2005b) Double-strand break formation by the RAG complex at the bcl-2 major breakpoint region and at other non-B DNA structures in vitro. *Mol Cell Biol* **25**(14)**:** 5904-5919

Raghavan SC, Swanson PC, Wu X, Hsieh CL, Lieber MR (2004) A non-B-DNA structure at the Bcl-2 major breakpoint region is cleaved by the RAG complex. *Nature* **428**(6978)**:** 88-93

Raghuraman MK, Winzeler EA, Collingwood D, Hunt S, Wodicka L, Conway A, Lockhart DJ, Davis RW, Brewer BJ, Fangman WL (2001) Replication dynamics of the yeast genome. *Science* **294**(5540)**:** 115-121

Rangan A, Fedoroff OY, Hurley LH (2001) Induction of duplex to G-quadruplex transition in the c-myc promoter region by a small molecule. *J Biol Chem* **276**(7)**:** 4640-4646

Ruf IK, Rhyne PW, Yang H, Borza CM, Hutt-Fletcher LM, Cleveland JL, Sample JT (2001) EBV regulates c-MYC, apoptosis, and tumorigenicity in Burkitt's lymphoma. *Curr Top Microbiol Immunol* **258:** 153-160

Saglio G, Grazia Borrello M, Guerrasio A, Sozzi G, Serra A, di Celle PF, Foa R, Ferrarini M, Roncella S, Borgna Pignatti C, et al. (1993) Preferential clustering of chromosomal breakpoints in Burkitt's lymphomas and L3 type acute lymphoblastic leukemias with a t(8;14) translocation. *Genes Chromosomes Cancer* **8**(1)**:** 1-7

Schroth GP, Ho PS (1995) Occurrence of potential cruciform and H-DNA forming sequences in genomic DNA. *Nucleic Acids Res* **23**(11)**:** 1977-1983

Simonsson T, Pecinka P, Kubista M (1998) DNA tetraplex formation in the control region of c-myc. *Nucleic Acids Res* **26**(5)**:** 1167-1172

Sutherland GR (2003) Rare fragile sites. *Cytogenet Genome Res* **100**(1-4)**:** 77-84

Tanaka H, Cao Y, Bergstrom DA, Kooperberg C, Tapscott SJ, Yao MC (2007) Intrastrand annealing leads to the formation of a large DNA palindrome and determines the boundaries of genomic amplification in human cancer. *Mol Cell Biol* **27**(6)**:** 1993-2002

Wang G, Vasquez KM (2004) Naturally occurring H-DNA-forming sequences are mutagenic in mammalian cells. *Proc Natl Acad Sci U S A* **101**(37)**:** 13448-13453

Wells RD (2007) Non-B DNA conformations, mutagenesis and disease. *Trends Biochem Sci* **32**(6)**:** 271-278

Wells RD (2008) Discovery of the role of non-B DNA structures in mutagenesis and human genomic disorders. *J Biol Chem*

Xu Y, Sugiyama H (2006) Photochemical approach to probing different DNA structures. *Angew Chem Int Ed Engl* **45**(9)**:** 1354-1362

Yabuki N, Terashima H, Kitada K (2002) Mapping of early firing origins on a replication profile of budding yeast. *Genes Cells* **7**(8)**:** 781-789

3.5. Materials and methods

3.5.1. Strains

KT119 strain (*MATa, his7-2, leu2-3,112, trp1-D, ura3-D, lys2-D, ade2-D, bar1-D, sfa1-D, cup1-1-D, yhr054c-D, cup1-2-D, lys2::kanMXURA3,ADE2, CUP1* and *SFA1*) is a derivative of TP strains described in Kim *et al* (2008). Nucleotide sequences of the primers used for integrations and disruptions are available upon request.

3.5.2. Genetic techniques

The rates and 95% confidence intervals of the arm loss in *lys2* alleles were estimated in fluctuation tests using at least 12 independent cultures. The canavanine containing media was made with a low concentration of adenine (5mg/L) to allow color detection; strains with an *ade2* mutations form red colonies in medium with low levels of adenine.

3.7. Tables

Table 1. Triplex forming non-GAA/TTC mirror repeats induce chromosome arm loss.

Triplex forming motifs	Rate of arm loss
$(GAA)_{20}$	**$9.0x10^{-10}$**
95% confidence intervals	$5.0x10^{-10}$~ $1.5x10^{-9}$
$(TTC)_{60}$	**$2.1x10^{-9}$**
	$1.9x10^{-9}$~ $3.5x10^{-9}$
$(GAA)_{60}$	**$2.7x10^{-9}$**
	$2.0x10^{-9}$~ $4.3x10^{-9}$
$(TTC)_{120}$	**$4.5x10^{-7}$**
	$3.2x10^{-7}$~ $6.1x10^{-7}$
$(GAA)_{120}$	**$17.4x10^{-7}$**
	$9.1x10^{-7}$~ $24.7x10^{-7}$
c-*MYC*-Py	**$4.8x10^{-9}$**
	$2.6x10^{-9}$~ $7.9x10^{-9}$
c-*MYC*-Pu	**$5.7x10^{-9}$**
	$3.8x10^{-9}$~ $8.9x10^{-9}$
MBRL-Py	**$1.4x10^{-8}$**
	$9.3x10^{-9}$~ $1.7x10^{-8}$
MBRL-Pu	**$1.4x10^{-8}$**
	$8.3x10^{-9}$~ $1.6x10^{-8}$
MBR-Py	**$2.6x10^{-8}$**
	$2.0x10^{-8}$~ $3.6x10^{-8}$
MBR-Pu	**$2.6x10^{-8}$**
	$2.0x10^{-8}$~ $4.0x10^{-8}$
***PKD1*-4**	**$2.7x10^{-8}$**
	$2.2x10^{-8}$~ $4.0x10^{-8}$

Chapter IV

Triplex stabilizing small molecules augment GAA/TTC repeat-induced chromosome fragility *in vivo.*

- The results presented in chapter IV are not published yet.

- All small molecules (Azacyanines and coralyne chloride) were synthesized and purified by Dr. Ozgul Persil in Dr. Nikolas Hud Lab.

- All experiments described in this chapter were done by me.

4.1. Introduction

4.1.1. Small molecules have been studied for drug design by the virtue of binding specificity.

Numbers of small molecules that bind to DNA are clinically proven as therapeutic agents such as daunomycin, doxorubicin, echinomycin, imidazoacridone, distamycin, quarfloxin, and mitoxantrone (Lockhart et al, 2004; Mazerska et al, 2001; Yamori et al, 1999; Kong et al, 2005; Tauchi et al, 2003). In general, these small molecules bind to DNA and inhibit the replication, transcription, or the activity of topoisomerase (Fornari et al, 1994; Leng & Leno, 1997).

Rationally, small molecules need to bind to the secondary structure tightly but also selectively in order to reduce non-specific side effects. As of this requirements, the binding of small molecules to non-canonical nucleic acid structures, such as triplex and quadruplex DNA, has been a major focus of drug design owing to these molecules activity as therapeutic agents (Chaires, 2005) (Waring, 2003). Conformations of nucleic acids which can adopt non canonical secondary structures such as triplex DNA and/or G-quartet can provide us opportunities for the design of selective and tight-binding small molecules (Chaires, 2008; Ren et al, 2001).

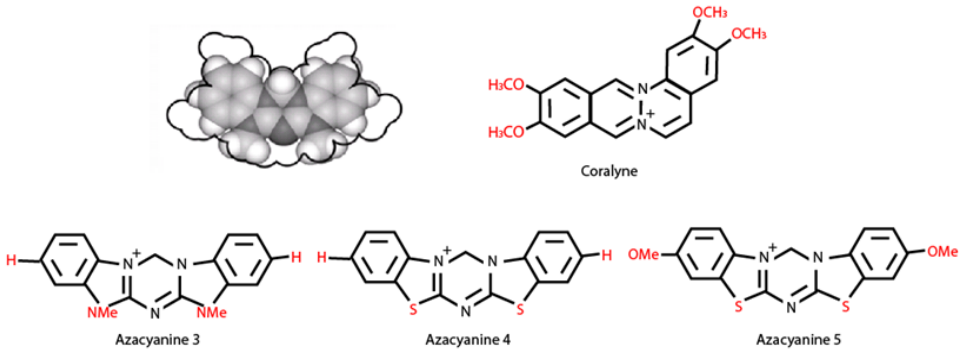
4.1.2. Small molecules can promote triplex DNA *in vitro*.

For several years Hud's laboratory has been investigating the nature of coralyne binding to a variety of DNA and RNA structures (Jain et al, 2003; Polak & Hud, 2002).

The small molecule coralyne (Figure 1A) was recently found to bind poly(dA) with one coralyne molecule per four adenine bases. It has been demonstrated that coralyne promotes the formation of an antiparallel homo-adenine duplex in poly(dA) (Persil, 2008; Persil et al, 2004). Similar to coralyne, another type of small molecules, azacyanine 3,4, and 5 also bind to poly(dA), however they have very low binding specificity to duplex DNA (Persil, 2008), Figure 1. Overall, coralyne and azacyanines have crescent shape, they are positively charged and too bulky to intercalate between Watson-Crick hydrogen bonds. These features make these small molecules potent and very specific in promoting and stabilizing the secondary structures DNA which can possibly employing Hoogsteen base pairing *in vitro*. This implies an idea that these small molecules might stabilize secondary structures such as triplex DNA conformations.

Inspired by these *in vitro* results, we have investigated the effect of small molecules on chromosomal fragility mediated by triplex forming GAA/TTC repeats using arm loss assay.

A



B

	Association constant, poly (dA)
Coralyne	1×10^{-7} M
Azacyanine 3	3.8×10^{-5} M
Azacyanine 4	4.8×10^{-4} M
Azacyanine 5	2.5×10^{-5} M

Figure 1. Chemical structure of Azacyanine molecules.

A, The top right figure represents a space-filling model of azacyanine overlaid with an outline of a space filling model of coralyne. The others represent chemical structures of small molecules. The red color represents different substituent from molecules. B, Association binding constants for poly(dA) (Persil, 2008).

4.2. Results

4.2.1. Experimental system to study chromosomal fragility resulting from GAA/TTC repeats with small molecule treatments.

We have employed previously developed experimental system based on the loss of *CAN1* and *ADE2* genes located on chromosome V to monitor the chromosomal break formation induced by (GAA/TTC) repeats (for full information, see Figure 1 in chapter II, and Kim et al, 2008).

4.2.2. Azacyanines stimulate $(GAA/TTC)_{230}$-mediated arm loss in a concentration-dependent manner.

We have found that *in vivo*, azacyanines 3, 4, and 5 but not coralyne stimulate $(TTC)_{230}$ and $(GAA)_{230}$-mediated arm loss. Compared to azacyanine 3 and 4, azacyanine 5 has the strongest synergistic effect on fragility with the expanded $(GAA)_{230}$ and $(TTC)_{230}$ repeat tracts at 0.2 mM concentration. Rates of arm loss were elevated by ~ 11 fold when the 0.2mM azacyanine 5 is supplemented to the media. Treatment with azacyanines 3, 4 and 5 induces GAA/TTC-mediated arm loss in a dose-dependent manner (Table 2A, and Figure2). Azacyanines induce chromosomal fragility, not only in the case of expanded repeats tracts but also when strains containing short $(GAA)_{20}$ and $(TTC)_{20}$ repeat tracts

are treated. Fragility was increased by ~3.5 fold and 2.3 fold with the strain containing $(GAA)_{20}$ and $(TTC)_{20}$ tracts, respectively (Figure 2, Table 1)

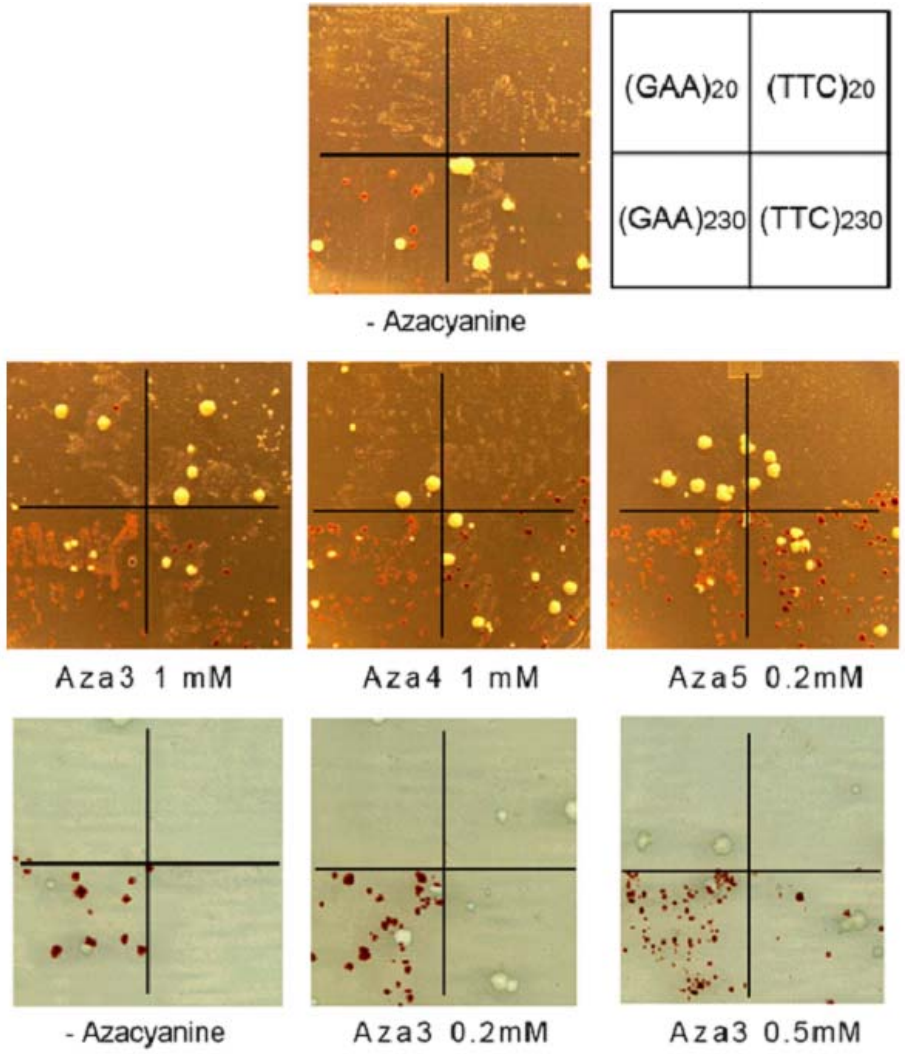


Figure 2. Azacyanines stimulate $(GAA/TTC)_{230}$ mediated arm loss in a concentration-dependent manner.

Yeast cells were grown on media supplemented with azacyanines and was then replica plated to canavanine media.

In order to investigate whether the azacyanine effect is specific to triplex forming GAA/TTC repeat tracts, we have constructed a control strain that does not contain any repeats in the *LYS2*. No significant increase in fragility was detected upon azacyanine treatment (Table 1).

To verify that the effect small molecules on genome fragility are connected to DNA replication, yeast cells were synchronized and released in the presence of azacyanine 5. Cells were collected from the first and the second cellular divisions and were assessed for the fragility level. Upon treatment, 3 to 4 fold higher fragility was observed from zero to the first cell division (~2.5h) with the strain containing $(GAA)_{230}$ or $(TTC)_{230}$ repeat tracts. Similarly, fragility measured from first to second cell division was increased by approximately 3 to 4 folds (Figure 3).

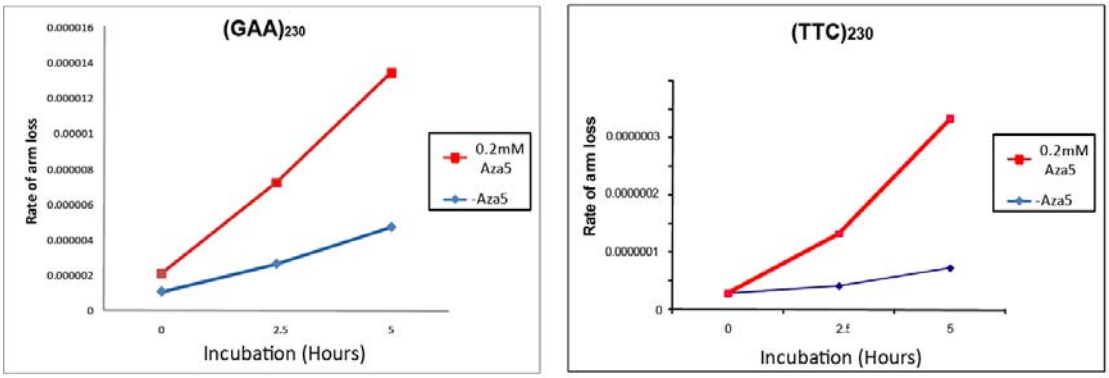


Figure 3. Azacyanines stimulate arm loss in actively dividing cells.

Yeast cell cycles were synchronized and released with Azacyanine 5 treatment. Cells were prepared from the first and the second cell division and were measured for fragility.

4.2.3. Azacyanines augment replication blockage by expanded GAA/TTC repeats

Previously, we demonstrated that the fragility potential of GAA/TTC strongly correlates with their propensity to adopt secondary structure and to block replication progression (Kim et al, 2008). We did not observed replication arrest zone at $(GAA)_{20}$ and $(GAA)_{60}$ repeat tracts. $(GAA)_{120}$ repeats were a mild replication barrier while $(GAA)_{230}$ and $(GAA)_{340}$ tracts blocked replication progression profoundly exhibiting characteristic Y-arc interruptions. In this study, we investigated whether azacyanine also can enhance replication blockage and fragility at the location of GAA/TTC repeats. We employed strains with $(GAA)_{120}$ tracts to investigate azacyanine effects on replication blockage.

Yeast cells were synchronized with alpha factor and released with 0.2 mM azacyanine 5 for 50-55 minutes and used for 2D gel replication analysis. We have found that azacyanine 5 clearly promotes replication arrest zone at $(GAA)_{120}$ tracts compared to the same strain with no treatment (Figure 4).

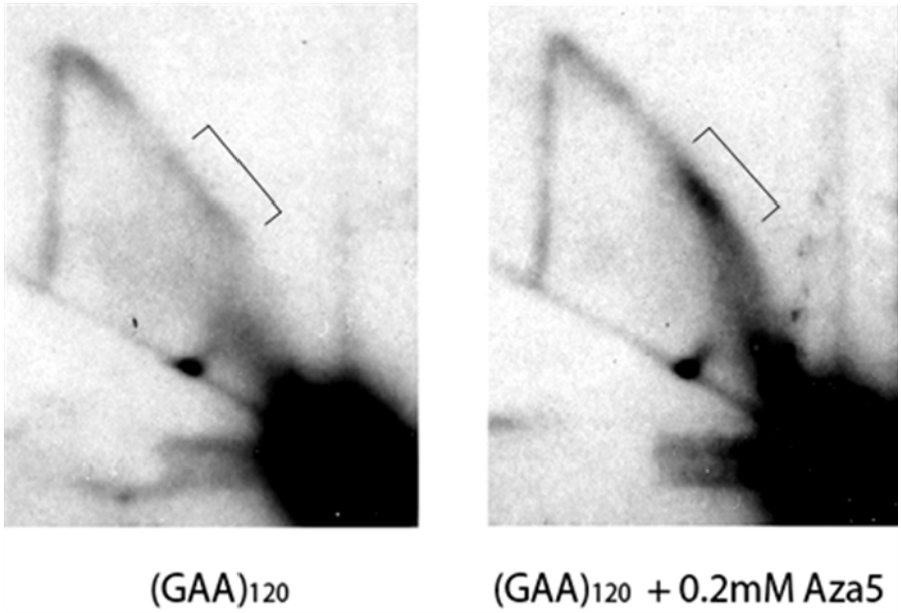


Figure 4. Azacyanines augment replication blockage by expanded GAA/TTC repeats.

Yeast cells were synchronized with alpha factor and released with 0.2 mM azacynine5 for 50-55 minutes and used for 2-D gel replication analysis.

4.2.4. Azacyanines compromise cellular viability and cause checkpoint response

Azacyanines at concentrations that induce fragility also inhibit cell growth. With azacyanine 5 treatment, yeast cells grow poorly on SD complete media at 0.2 mM concentration. At 0.5 mM concentration, growth is greatly repressed. Similarly, azacyanine 3 and 4 inhibit cell growth at 0.5mM concentration (Figure 5). Azacyanine 5 inhibits cell growth at the lowest concentration (0.2mM) compared to Azacyanine 3 and 4 (0.5mM) (Table 2-B).

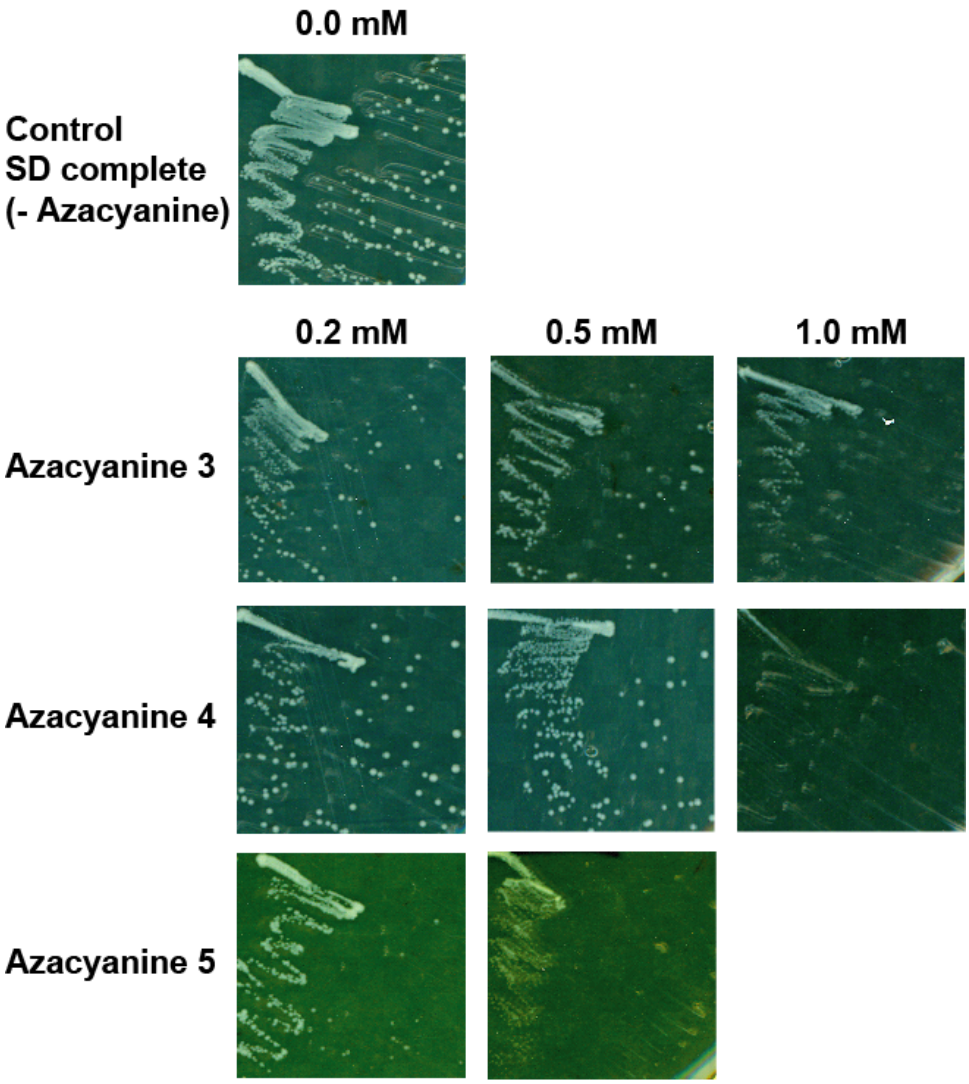
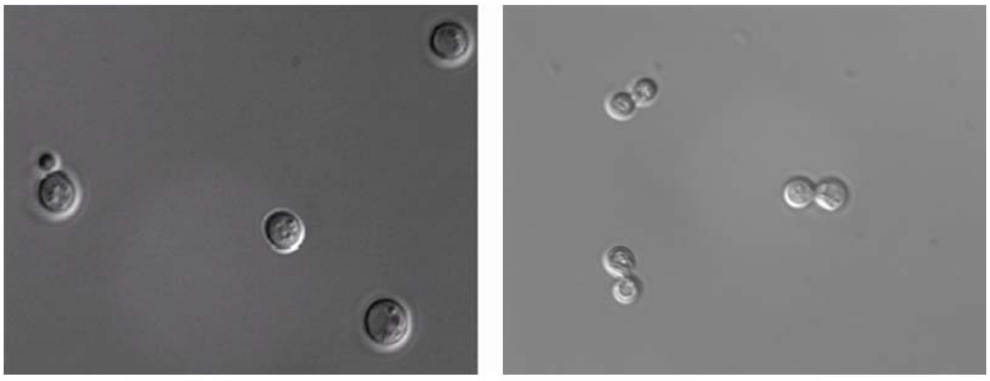


Figure 5. Azacyanines inhibit cell growth in a dose- dependent manner.

Yeast cells are streaked out on SD complete media supplemented with different concentrations (0, 0.2, 0.5, and 1.0mM) of azacyanines.

Over 60% of the yeast cells at concentrations of azacyanines that inhibit cellular divisions were arrested at G2/M stage of cell cycle which is indicative of DNA-damage activated checkpoint response (Figure 6).



- Azacyanine 0.2mM Aza5

	G1	S	G2	Total
Control	229	84	87	400 cells
% of Cells	57 %	21 %	22 %	100 %
Aza5	106	27	267	400 cells
% of Cells	27 %	7 %	66 %	100 %

Figure 6. Treatment of yeast cells with azacyanines lead to G2/M arrest indicative of checkpoint response

Wild type cells were arrested in G2/M phase upon treatment with 0.2mM azacyanine 5. Cells were imaged following propagation in liquid synthetic complete medium with and without the

drug. Over 60% of the cells were arrested with large buds following the incubation with the azacyanine 5. Treatment with azacyanine 3 and 5 yielded similar results (data not shown).

Mre11 is a subunit of a complex with Rad50p and Xrs2p (M/R/X complex), and one of its functions is to repair DNA double-stranded breaks. Rad52 stimulates strand exchange by facilitating Rad51p binding to single-stranded DNA and helps repairing double-stranded breaks (reviewed in Symington, 2002). Double-stranded break repair deficient mutants (Δ*rad52* or Δ*mre11)* are highly sensitive to azacyanine 3, 4 or 5 (Figure 7). These observations indicate that azacyanines might exacerbate cell growth by triggering multiple breaks in the genome (Figure 8).

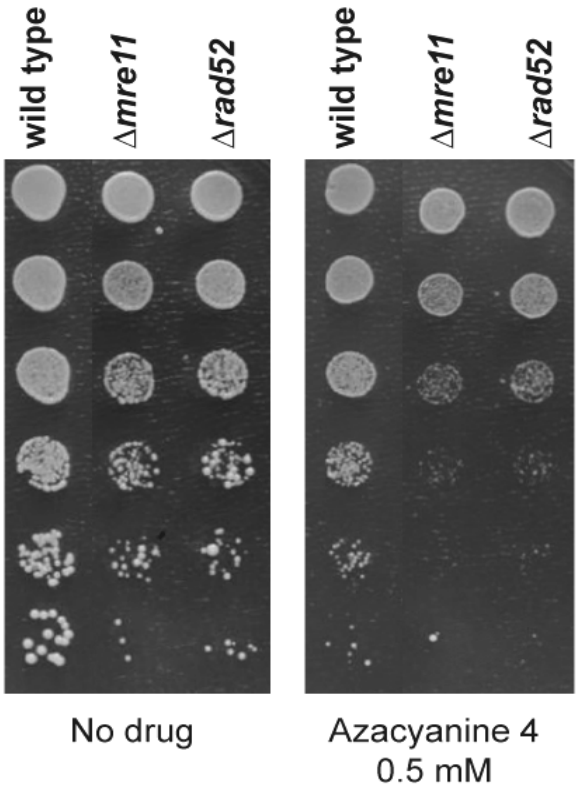


Figure 7. DSB repair mutants are sensitive to treatment with azacyanines

Sensitivity of wild type, *Δmre11* and *Δrad52* strains to azacyanine 4. Equal number of cells for each strain was spotted on synthetic complete medium containing either no drug (left panel) or 0.5mM azacyanine 4 (right panel). 6 fold serial dilutions from top to the bottom were spotted

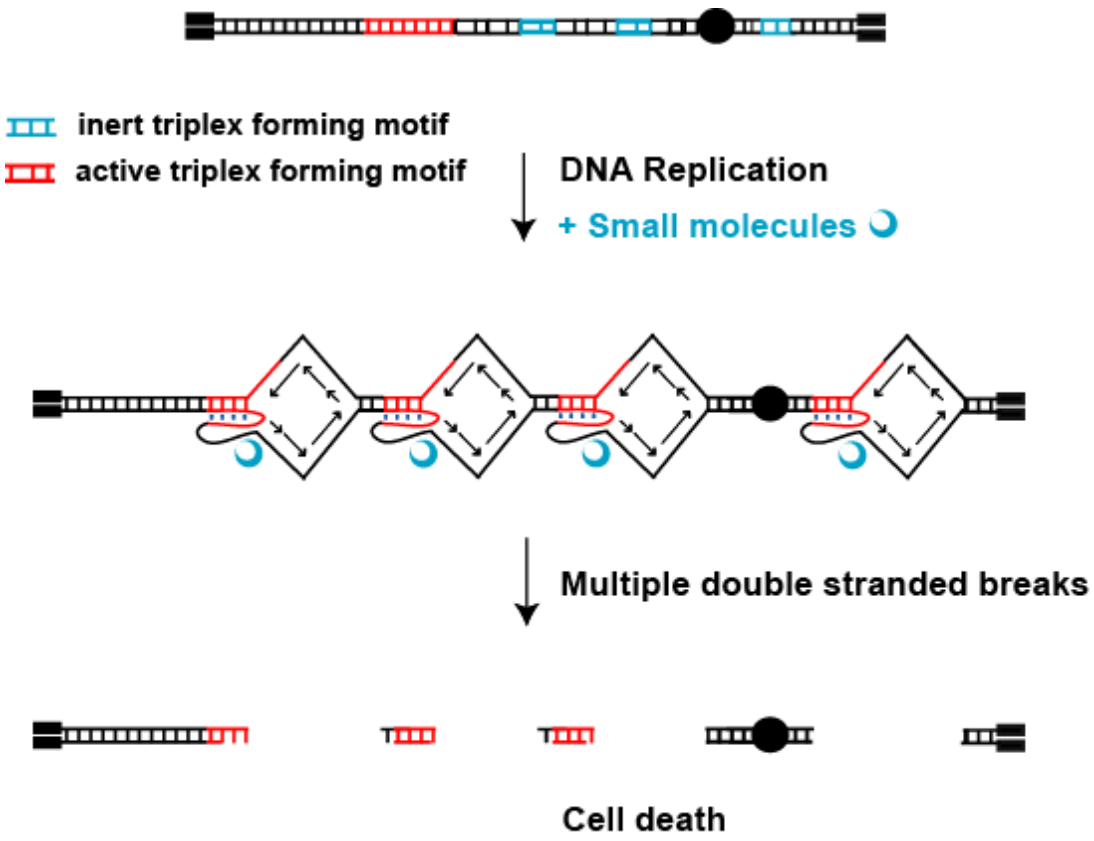


Figure 8. Model to show how azacyanine augmented triplex-mediated fragility leads to cell death.

4.3. Discussion

Triplex forming mirror repeat motifs are highly polymorphic and abundant in eukaryotic genomes, ranging from yeast to human (Cox & Mirkin, 1997). We have found that azacyanines can stabilize triplex DNA secondary conformation at GAA/TTC repeat tracts and induce fragility and senescence *in vivo.*

Azacyanine treatment induces GAA/TTC mediated arm loss in a dose- dependent manner. Fragility was increased as the concentration of azacyanine was elevated. Although coralyne chloride showed triplex DNA binding specificity and had strongest binding affinity to poly d(A)s *in vitro* (Persil, 2008), it did not affect chromosome fragility in our *in vivo* assay (data not shown). Azacyanine 5 has the strongest effect on fragility mediated by GAA/TTC repeats compared to azacyanine 3 and 4 *in vivo*. One explanation for the different effect of these compounds is that azacyanines unlike coralyne do not have exocyclic groups on the rings and compared to coralyne possess smaller surface area (Persil, 2008). The bulky shape of azacyanine might fit more tightly to triplex DNA conformation than coralyne and therefore stabilize the secondary structure which promotes genome fragility. Alternatively, the secondary structures adopted at poly d(A)s might not be the triplex DNA at GAA/TTC repeats. The binding potential of azacyanines for triplex forming DNA remains to be investigated.

Another interesting observation is that azacyanine 5 induces fragility even at the small repeat tracts such as $(GAA)_{20}$ and $(TTC)_{20}$. This result indicates, with an aid of small molecules, triplex DNA can be efficiently formed and represent a problem for replication progression even by short GAA/TTC tracts that normally behave as neutral sequences.

We suggested that fragility induced by GAA/TTC repeats can cause size variation of repeat tracts (Kim et al, 2008). Consistently with our previous hypothesis, azacyanine not only induce fragility but also trigger size variations of GAA/TTC repeats. Both contractions and expansions were increased by 5-foldfold and 7.5-fold respectively upon treatment (Table 3, Figure 9).

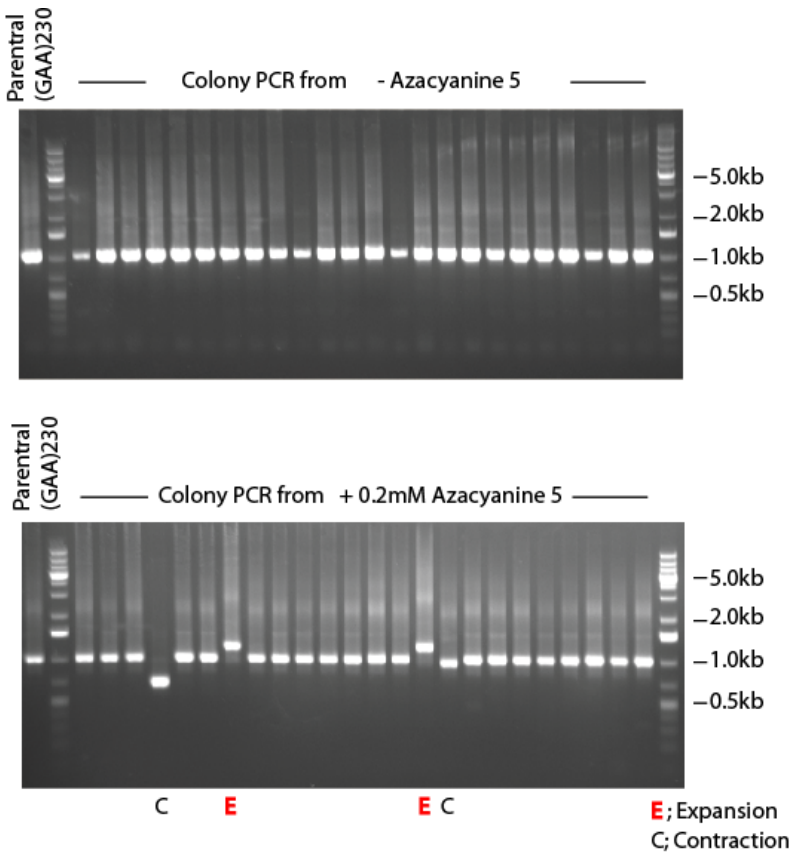


Figure 9. Azacyanine5 induces size variation of GAA/TTC repeats.

Yeast strain containing $(GAA)_{230}$ was streaked out on YPD supplemented with 0.2mM azacyanine5 then streaked out again on YPD containing no azacyanine. Twenty four colonies were picked randomly and used for PCR.

We have demonstrated that azacyanine 5 augments the replication arrest zone at $(GAA)_{120}$ tracts. This result strongly supports an idea that azacyanine induces genome fragility by stabilizing triplex DNA *in vivo* during DNA replication.

It would be very interesting to see whether azacyanines induce fragility at non-GAA/TTC triplex-forming motifs such as MBR region of *BCL-2* or intron21 of *PKD1*.

Our data suggest that azacyanines stabilize triplex DNA *in vivo,* and this might trigger multiple DSBs during the S-phase which are sensed by the checkpoint surveillance system (Figure 6). First, they boost replication blockage along with chromosome fragility (Figure 4). Second, more than sixty percent of cells are arrested at G2/M stage which is an indication of a DNA-damage checkpoint response (Figure 6). Third, double-stranded break repair deficient mutants such as Δ*rad52* or Δ*mre11* are highly sensitive to azacyanine 3, 4 or 5 treatments (Figure7). Fourth, Azacyanines inhibit cell growth in a concentration-dependent manner. These observations support an idea that azacyanines exacerbate cell growth by triggering multiple breaks in the genome in actively dividing cells. We propose that these small molecules can be the basis for the development of novel antitumor drugs that act via the inhibition of cellular proliferation (Figure8 and Figure 5). Moreover, azacyanines do not induce mutation which is consistent with their *in vitro* features being as poor DNA intercalators (Figure 10).

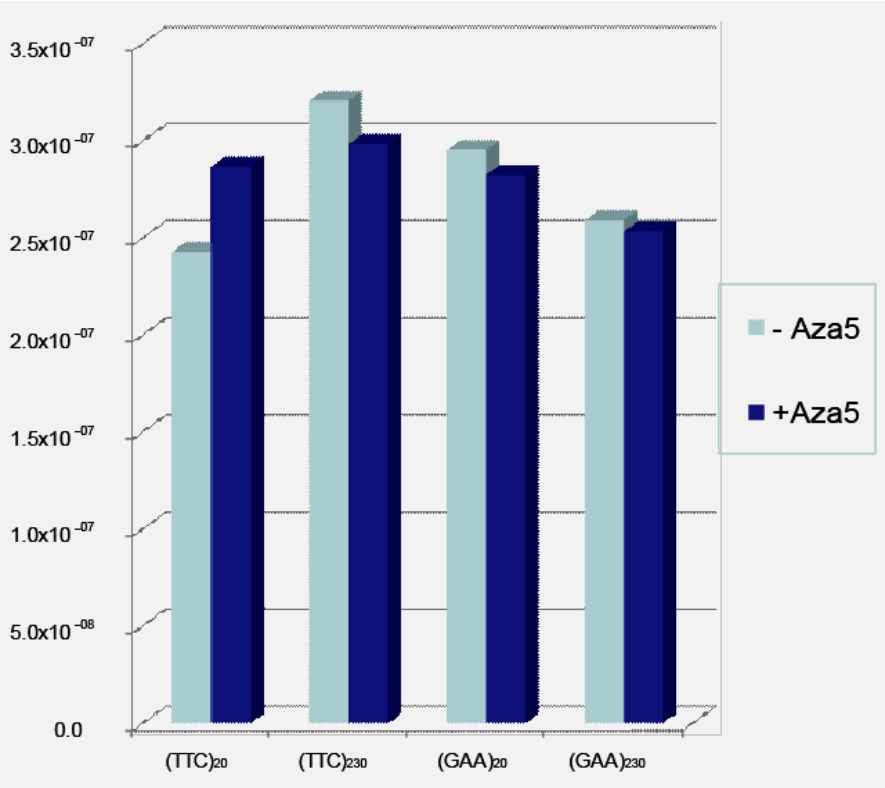


Figure 10. Azacyanines are not mutagens

Yeast strain containing $(TTC)_{20,}$ $(TTC)_{230,}$ $(GAA)_{20,}$ and $(GAA)_{230,}$was streaked out on synthetic complete medium supplemented with 0.2mM azacyanine5 then used for fluctuation test to count simple mutations in *CAN1* gene on canavanine media.

Azacyanines provide a promising new template for the design of new small molecules due to their 1) ease and one-pot synthesis; 2) low binding affinity to Watson-Crick duplex DNA *in vitro* and *in vivo*; 3) absence of mutagenic effect. Modifications (substitution or addition of new functional groups to the ring system) of azacyanine might enhance specificity to target triplex DNA and result in fewer side effects (non mutagenic), to cells. Potentially, they can be the basis for the development of novel antitumor drugs (Figure 8). Azacyanines can be also a useful tool for highlighting and identifying triplex-containing regions in human cells.

4.4. References

Arimondo PB, Thomas CJ, Oussedik K, Baldeyrou B, Mahieu C, Halby L, Guianvarc'h D, Lansiaux A, Hecht SM, Bailly C, Giovannangeli C (2006) Exploring the cellular activity of camptothecin-triple-helix-forming oligonucleotide conjugates. *Mol Cell Biol* **26**(1)**:** 324-333

Brewer BJ, Fangman WL (1987) The localization of replication origins on ARS plasmids in S. cerevisiae. *Cell* **51**(3)**:** 463-471

Chaires JB (2005) *Structural selectivity of drug-nucleic acid interactions probed by competition dialysis, in DNA Binders and Related Subjects*

Springer-Verlag Berlin.

Chaires JB (2008) Allostery: DNA does it, too. *ACS Chem Biol* **3**(4)**:** 207-209

Cox R, Mirkin SM (1997) Characteristic enrichment of DNA repeats in different genomes. *Proc Natl Acad Sci U S A* **94**(10)**:** 5237-5242

Fornari FA, Randolph JK, Yalowich JC, Ritke MK, Gewirtz DA (1994) Interference by doxorubicin with DNA unwinding in MCF-7 breast tumor cells. *Mol Pharmacol* **45**(4)**:** 649-656

Frank-Kamenetskii MD, Mirkin SM (1995) Triplex DNA structures. *Annu Rev Biochem* **64:** 65-95

Glover TW, 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**(2)**:** 136-142

Goni JR, de la Cruz X, Orozco M (2004) Triplex-forming oligonucleotide target sequences in the human genome. *Nucleic Acids Res* **32**(1)**:** 354-360

Gottesfeld JM (2007) Small molecules affecting transcription in Friedreich ataxia. *Pharmacol Ther* **116**(2)**:** 236-248

Jain SS, Polak M, Hud NV (2003) Controlling nucleic acid secondary structure by intercalation: effects of DNA strand length on coralyne-driven duplex disproportionation. *Nucleic Acids Research* **31**(15)**:** 4608-4615

Kim HM, Narayanan V, Mieczkowski PA, Petes TD, Krasilnikova MM, Mirkin SM, Lobachev KS (2008) Chromosome fragility at GAA tracts in yeast depends on repeat orientation and requires mismatch repair. *EMBO J* **27**(21)**:** 2896-2906

Kong D, Park EJ, Stephen AG, Calvani M, Cardellina JH, Monks A, Fisher RJ, Shoemaker RH, Melillo G (2005) Echinomycin, a small-molecule inhibitor of hypoxia-inducible factor-1 DNA-binding activity. *Cancer Res* **65**(19)**:** 9047-9055

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA,

Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ (2001) Initial sequencing and analysis of the human genome. *Nature* **409**(6822)**:** 860-921

Leng F, Leno GH (1997) Daunomycin disrupts nuclear assembly and the coordinate initiation of DNA replication in Xenopus egg extracts. *J Cell Biochem* **64**(3)**:** 476-491

Lockhart AC, Howard M, Hande KR, Roth BJ, Berlin JD, Vreeland F, Campbell A, Fontana E, Fiorentini F, Fowst C, Paty VA, Lankford O, Rothenberg ML (2004) A phase I dose-escalation and pharmacokinetic study of brostallicin (PNU-166196A), a novel DNA minor groove binder, in adult patients with advanced solid tumors. *Clin Cancer Res* **10**(2)**:** 468-475

Mazerska Z, Dziegielewski J, Konopa J (2001) Enzymatic activation of a new antitumour drug, 5-diethylaminoethylamino-8-hydroxyimidazoacridinone, C-1311, observed after its intercalation into DNA. *Biochem Pharmacol* **61**(6)**:** 685-694

Persil O (2008) SMALL MOLECULE RECOGNITION OF HOMOPURINE

NUCLEIC ACID STRUCTURES. *Persil, O PhD Thesis, DrHud lab, Georgia Institute of Technology*

Persil O, Santai CT, Jain SS, Hud NV (2004) Assembly of an antiparallel homo-adenine DNA duplex by small-molecule binding. *J Am Chem Soc* **126**(28)**:** 8644-8645

Polak M, Hud NV (2002) Complete disproportionation of duplex poly(dT)center dot poly(dA) into triplex poly(dT)center dot poly(dA)center dot poly(dT) and poly(dA) by coralyne. *Nucleic Acids Research* **30**(4)**:** 983-992

Ren J, Qu X, Dattagupta N, Chaires JB (2001) Molecular recognition of a RNA:DNA hybrid structure. *J Am Chem Soc* **123**(27)**:** 6742-6743

Seenisamy J, Bashyam S, Gokhale V, Vankayalapati H, Sun D, Siddiqui-Jain A, Streiner N, Shin-Ya K, White E, Wilson WD, Hurley LH (2005) Design and synthesis of an expanded porphyrin that has selectivity for the c-MYC G-quadruplex structure. *J Am Chem Soc* **127**(9)**:** 2944-2959

Sutherland GR (1988) The role of nucleotides in human fragile site expression. *Mutat Res* **200**(1-2)**:** 207-213

Symington LS (2002) Role of RAD52 epistasis group genes in homologous recombination and double-strand break repair. *Microbiol Mol Biol Rev* **66**(4)**:** 630-670, table of contents

Tauchi T, Shin-Ya K, Sashida G, Sumi M, Nakajima A, Shimamoto T, Ohyashiki JH, Ohyashiki K (2003) Activity of a novel G-quadruplex-interactive telomerase inhibitor, telomestatin (SOT-095), against human leukemia cells: involvement of ATM-dependent DNA damage response pathways. *Oncogene* **22**(34)**:** 5338-5347

Waring MJaLPGW (2003) *Forty Years On, in Small Molecule DNA and RNA binders*, Berlin: Wiley-VCH Verlag GmbH & Co.

Yamori T, Matsunaga A, Sato S, Yamazaki K, Komi A, Ishizu K, Mita I, Edatsugi H, Matsuba Y, Takezawa K, Nakanishi O, Kohno H, Nakajima Y, Komatsu H, Andoh T, Tsuruo T (1999) Potent antitumor activity of MS-247, a novel DNA minor groove binder, evaluated by an in vitro and in vivo human cancer cell line panel. *Cancer Res* **59**(16)**:** 4042-4049

4.5. Materials and methods

4.5.1. Strains

All strains in this study were isogenic to KT119 KT119 strain (*MATa, his7-2, leu2-3,112, trp1-D, ura3-D, lys2-D, ade2-D, bar1-D, sfa1-D, cup1-1-D, yhr054c-D, cup1-2-D, lys2::kanMXURA3,ADE2, CUP1* and *SFA1*) is a derivative of TP strains described in Kim *et al.,* (2008).

4.5.2. Genetic techniques

The rates and 95% confidence intervals of the arm loss in *lys2* alleles were estimated in fluctuation tests using at least 12 independent cultures. The media containing canavanine was made with a low concentration of adenine (5mg/L) to allow color detection; strains with an *ade2* mutations form red colonies in medium with low levels of adenine. *MRE11* and *RAD52* genes were disrupted with the *kanMX* cassette with homologous tails (Kim et al, 2008)

4.5.3. 2D Analysis of replication fork intermediates

Cells were arrested in G1 with α-factor and released synchronously into S-phase. At 40 minutes after release, chromosomal DNA was extracted, and neutral/neutral 2-D analysis was carried out

in according to Brewer et al., (Brewer & Fangman, 1987). The *Afl*II-digested DNA was separated in the first dimension on a 0.45% agarose gel without ethidium bromide in 1X TBE buffer at 1V/cm for 36 hours in first dimension. The second dimension gel was run at 6 V/cm in 1X TBE buffer containing 0.3 µg/ml ethidium bromide for 12 hours. Southern blot hybridization using 4 kb *Afl*II-digested *LYS2* fragment was performed to highlight replication intermediates.

4.8. Tables

Table 1. Azacyanine augment chromosomal arm loss induced by (GAA) and (TTC) repeat orientations.

Yeast strain containing $(GAA)_{0,}$ $(TTC)_{20,}$ $(GAA)_{20,}$ $(TTC)_{230,}$ and $(GAA)_{230}$ were streaked out on SD complete media supplemented with 0.2mM azacyanine5 then used for fragility assays in canavanine media.

Insertion in *LYS2*	Arm loss rate (/ 10^{10})	
	- Azacyanine	0.2mM Azacyanine 5
$(GAA)_0$	9 (4.8 - 15)	9 (4.2 -17)
$(TTC)_{20}$	8 (4-17)	19 (15-38)
$(GAA)_{20}$	9 (5-15)	33 (21-50)
$(TTC)_{230}$	4780 (4410-6130)	53954 (36910-113130)
$(GAA)_{230}$	66999 (51100 - 86480)	754760 (459260 - 1103700)

Table 2. Azacyanines inhibit cell growth and augment genome fragility in a dose-dependent manner.

A

Chromosome arm loss with azacyanine treatment (canavanine media)

	0.2mM	0.5mM	1.0mM
Azacyanine 3	no change	+ +	+ + +
Azacyanine 4	no change	+ +	+ + +
Azacyanine 5	+ +	+ + +	no growth

\+ : intensity of arm loss changes compared to no azacyanine

B

Growth inhibition of azacyanine (SD complete media)

	0.2mM	0.5mM	1.0mM
Azacyanine 3	no change	-	- - -
Azacyanine 4	no change	-	- - -
Azacyanine 5	- -	- - -	no growth

\- : growth inhibition compared to no azacyanine

Table 3. Azacyanine 5 treatment triggers size variation of $(GAA)_{230}$ repeat tracks.

A yeast strain containing $(GAA)_{230}$ was streaked out on YPD supplemented with 0.2mM azacyanine5 and was re-streaked on YPD containing no azacyanine. Colonies are chosen for PCR randomly.

Insertion in *LYS2*	**Number of colonies examined**	**Contraction (%)**	**Expansion (%)**	**Unchanged (%)**
$(GAA)_{230}$	120	2.0	1.0	83
$(GAA)_{230}$ + 0.2mM Aza5	120	10.0	7.5	83

Chapter V

5.1. Overall conclusions

During my Ph.D program, I have investigated chromosome fragility induced by the triplex forming mirror repeats including GAA/TTC trinucleotide repeat and triplex forming non-GAA/TTC mirror repeats. The following are the conclusions summarized from the graduate work:

I. Expanded GAA/TTC repeats are strong inducers of double-stranded breaks and gross chromosomal rearrangements in yeast. The fragility potential depends on the length of the GAA/TTC tracts and the orientation of the repeats relative to the replication origin. This correlates with their propensity to adopt triplex and to block replication fork movement. Breaks triggered by GAA/TTC repeats generate specific patterns of rearrangements. Non reciprocal translocations are the primary outcome of the GAA/TTC-mediated fragility. Mismatch repair machinery might trigger the fragility by processing the triplex structure.

II. Triplex forming non-GAA/TTC mirror repeats such as major breakpoint cluster region (MBR) at the *BCL-2* gene, intron 21 of *PKD1* and promoter region of *C-MYC* induce chromosome fragility in *S.cerevisiae*. There is an increase (~26 fold) in levels of *CAN1* region loss for MBR-Pu and MBR-Py triplex forming motifs over strains containing

$(GAA)_{20}$. No significant orientation dependency is observed.

III. Azacyanines 3, 4, and 5 but not coralyne stimulate (GAA/TTC) - mediated arm loss in a dose dependent manner *in vivo*. Azacyanines at concentrations that induced fragility also inhibit cell growth. Over sixty percents of the yeast cells are arrested at G2/M stage of cell cycle indicative of DNA-damage activated checkpoint response. Mutants defective in double-stranded break repair show hyper sensitivity to the azacyanines.

5.2. List of publications

Published or in press

Hyun-Min Kim, Vidhya Narayanan, Piotr A. Mieczkowski, Thomas D. Petes, Maria M. Krasilnikova, Sergei M. Mirkin and Kirill S. Lobachev, 'Chromosome fragility at the expanded GAA/TTC tracts in yeast depends on repeat orientation and requires the mismatch repair system', ***EMBO journal,*** EMBO J. 2008 Nov 5;27(21):2896-906

Vidhya Narayanan, Piotr Mieczkowski, **Hyun-Min Kim**, Thomas D. Petes, and Kirill S. Lobachev, 'The Pattern of Gene Amplification is determined by the Chromosomal Location of Hairpin-capped Breaks' ***Cell***, 2006 June 30, Vol. 125, 1283-1296

In preparation

Victoria Liston, **Hyun-Min Kim**, Kirill Lobachev, Igor. B. Rogozin and Youri I.Pavlov, 'Estimation of Okazaki fragment size in yeast *in vivo*'.