Short inverted repeats are hot spots for genetic instability in mammalian cells

Steve Lu

In collaboration with Dr. Guliang Wang and Scott Spitzer

In partial fulfillment of the requirements for graduation with the Dean's Scholars Honors Degree in Biochemistry

_____ _____

Dr. Karen M. Vasquez Date

Supervising Professor

_____ _____

Dr. David Laude Date

Co-Supervising Professor

_____ _____

Dr. John F. Stanton Date

Honors Advisor in Biochemistry

Abstract

Analyses of chromosomal aberrations in genetics disorders such as leukemia and lymphoma have provided compelling evidence that segments of the human genome containing repetitive sequences can be prone to breakage and chromosomal rearrangements. While long inverted repeat (IR) sequences have been shown to be mutagenic in mammalian cells, the mutagenic potential of short IR sequences (≤30 bp), which are very abundant in eukaryotic genomes and often co-localize near chromosomal breakpoints, has not been investigated. Herein, we demonstrate that cruciform structures formed at short perfect IRs are mutagenic in mammalian cells. We found that DNA double-strand breaks (DSBs) occurred *in vivo* at or near the IR sequences, and the majority of the mutations induced by short IRs were large-scale deletions spanning the palindromic sequence. Analyses of the mutant junctions revealed that ~80% contained microhomologies, which are characteristic of an error-prone microhomology-mediated end-joining repair mechanism. We found that cruciform-induced mutations and DNA strand breaks may occur through 1) a replication-dependent mechanism involving DNA replication fork stalling or 2) replication-independent, structure-specific enzymes that facilitate cleavage at positions adjacent to the IR sequences. Taken together, our results demonstrate that short inverted repeat sequences play a role in genetic instability, and provide a plausible mechanistic explanation for the co-localization of IRs with chromosome breakage points in human disease. These findings also support the hypothesis that alternatively structured DNA is a feature of genome plasticity and may be a contributor to human disease and a driving force in the evolution of the human genome by providing a means for diversity within the population.

Background

Genetic analyses of cancer-related translocation events in human cells have detected certain regions of the genome that are more susceptible to breakage, which may lead to the deregulation of proto-oncogenes or inactivation of tumor suppressor genes (1, 2). Interestingly, many of these hot-spot regions contain sequences that have the potential to form non-B DNA structures such as left-handed Z-DNA, triple helical H-DNA, or cruciform DNA (3, 4). Accumulating evidence support the existence of non-B DNA structures *in vivo*. For example, anti-cruciform monoclonal antibodies that protect stem-loop structures on plasmid DNA from cleavage have been identified (5). Although the mechanisms are still unclear, studies have shown that these unusual DNA secondary conformations are a source of non-random genetic instability in bacteria, yeast, and mammals (4, 6-9). We previously demonstrated that triplex-forming sequences from the human *c-MYC* gene and Z-DNA-forming sequences resulted in large deletions and rearrangements in mammalian cells and mice (9, 10). These studies provide evidence that endogenous human sequences are mutagenic in mammals and suggest a role for DNA structure in translocation-mediated disease.

Cruciform DNA structures can form at inverted repeat (IR) or palindromic sequences, i.e., sequences that are followed directly by an exact mirror image on the complementary strand (11). This sequence motif can either base pair to the complementary strand forming a B-DNA duplex or undergo intrastrand base pairing and fold on itself. When only one exposed strand folds over on itself, it is designated as a hairpin structure; however when both strands form hairpins, the four-arm junction is referred to as a cruciform DNA structure (3). Energetically the formation of a cruciform is not favored over a perfectly base-paired Watson-Crick helix, in

part due to less base pairing at the loop region. However negative supercoiling, a common status in mammalian genomic sequences, potentially provides the energy for the formation and stabilization of the cruciform structure (11-13).

IR sequences are very common in both prokaryotic and eukaryotic genomes, with IRs ≥21 bp estimated to occur once every 5,600 bp in humans (14). The pervasiveness of these sequences in genomes indicates that they may have important biological functions such as structural roles in tRNAs and ribozymes (11, 15). In addition to the detection of IRs at or near replication origins in bacterial and mammalian cells, the eukaryotic 14-3-3 cruciform-binding proteins have also been identified as conserved regulators of DNA replication (15, 16). Typically, IRs are excluded from the coding regions in both *E. coli* as well as eukaryote genomes (17), suggesting possible regulatory roles or mechanistic functions in genome evolution.

Studies focused on role of IRs in mutagenesis have primarily been on neurological diseases such as Hungtinton's in which secondary hairpin loops are proposed to support the expansion of quasipalindromic CNG repeats (18). While the mutagenic potential of hairpin structures containing mismatches in the stem has been well documented, many questions remain about the presence of cruciform structures and their role in pathogenic rearrangements *in vivo* (3). Few studies have focused on cancer-associated chromosomal aberrations due to perfect palindromes (no mismatches in the stem), but there is compelling evidence that perfect palindromic sequences may serve as structural platforms that result in gene amplification and contribute to subsequent tumor progression (19). The centers of large unique AT-rich palindromes (>400 bp) on chromosomes 11 and 22 that form cruciform structures have also

been studied extensively and were found to generate DSBs, simulating recurrent constitutional translocations in sperm leading to chromosomally unbalanced offspring (20).

Although previous research has shown that long palindromic sequences can lead to mutations, short IR sequences are more abundant in the human genome, and their mutagenic potential has not been well characterized. In this study, we demonstrate that cruciform structures that form at short perfect IRs (≤30 bp) confer genetic instability in mammalian cells. Our data provides evidence that cruciforms formed at short IRs can cause DSBs *in vivo* perhaps by stalling DNA replication or activating enzymes that can recognize and cleave the secondary structures. The DSBs appeared to be processed via a microhomology-mediated end-joing (MMEJ) mechanism, resulting in large-scale deletions spanning the IR sequence. These findings provide a plausible mechanistic explanation for the colocalization of IRs with chromosome breakage points in human disease, and support the hypothesis that the formation of non-B DNA structures may be involved in genetic instability and disease etiology.

Materials & Methods

Construction of Cruciform-forming Plasmids. AT-rich and CG-rich inverted repeats, 30 bp and 28 bp, respectively, and a control sequence unable to adopt non-B DNA structures were subcloned into a lacZ' mutation-reporter shuttle vector pUCNIM at an EcoRI-SalI cassette (according to standard protocols). The sequences of the inserts (Fig. 1A) were designed such that cruciforms were the most likely non-B DNA structures formed. After the transformation of ligation products into DH5α bacterial cells, restriction analysis and direct DNA sequencing

confirmed the recombinants. The plasmid constructs were named pUCON, pUCG+, and pUAT+, corresponding to their inserts.

Probing the Cruciform Structure in Modified Plasmids. Because IRs can form hairpins and cruciforms, we used four-way junction specific T7 endonuclease I, following the methods described by *Kurahashi et al*. (8), to determine if the modified plasmids could specifically form cruciform *in vitro*. ScaI restriction enzyme cleaved the plasmid twice (1.2 kb upstream and 1 kb kb downstream from the inserts). T7-induced DSBs presumably resulting from cleavage of the cruciform structures were indicated by a 1.2 kb ScaI-T7 fragment and a 1 kb T7-ScaI fragment (Fig. 1B). A 2.2 kb ScaI-Sca1 fragment served as the loading control on a 1% agarose gel.

In Vivo **Mutagenesis Assay in Mammalian COS-7 Cells.** The background mutation frequencies of the cruciform-forming plasmids in DH5α bacterial cells did not differ significantly from the control. Plasmid DNA was transfected into mammalian COS-7 cells using AMAXA™ Nucleofector™ Kit V by Lonza (Houston, TX) according to the recommended protocol. After 48 hours, the plasmids were recovered using Qiagen QIAprep Spin Miniprep (Valencia, CA) and treated with DpnI to remove unreplicated plasmids. Mutants were identified in a DH5α blue-white screening. Actual cruciform-induced mutation frequencies in mammalian cells were adjusted by subtracting the background bacterial mutation frequencies from measured frequencies. A Students t-test determined the statistical significance.

Detection of DSBs in Mammalian Cells and HeLa Cell-free Extracts by Ligation-mediated PCR (LM-PCR). LM-PCR was carried out as described by Rodriguez et al. (21) with minor alterations. The region between the upstream primer and IR insert was ≈200 bp. Amplified PCR products were separated on a 1.8% agarose gel, purified using Promega Wizard® SV Gel and PCR Clean-Up System (Madison, WI), cloned into the Promega pGEM-T® (Madison, WI) vector system, and finally transformed into DH5α cells for sequencing to map the breakpoints.

***In Vitro* Mutagenesis Assay in HeLa Cell-free Extracts.** The pUCNIM reporter plasmid contains an SV40 origin of replication and can only be replicated in the presence of the SV40 large T antigen. SV40 large T antigen-deficient HeLa cell-free extracts from a CHIMERx DNA replication assay kit (Milwaukee, WI) were employed to determine the relationship between replication and cruciform-induced genetic instability. Purified large T antigen (CHIMERx) was added to the manufacturer's recommended reaction mixture to allow for replication. 50 ng of plasmid DNA was incubated for 6 hours at 37°C and then digested with DpnI (if from replicating extracts). Cruciform-induced mutants were determined by blue-white screening and analyzed by restriction digestion and sequencing.

Two-dimensional Gel Electrophoresis of Replication Intermediates in Mammalian Cells. Identical to the mutagenesis assay, the modified plasmids were transfected into mammalian COS-7 cells. However the plasmids used the *supF* system rather than the *lacZ* system. Because the control and IR sequences were identical to those in the *lacZ* plasmids, the

plasmids were named pCON, pCG+, and pAT+ according to their inserts. After 24 hours, replication intermediates were isolated and subjected to 2D gel electrophoresis according to Krasilnikova and Mirkin (22).

Mutation Analysis. *LacZ'* mutants were randomly selected and characterized by restriction digestion analysis and sequencing. The recovered plasmids from mammalian cells were amplified and subsequently digested with Eag1 and BssS1 into a total of 7 fragments separated on a 1.5% agarose gel. A 877 bp fragment contained the cruciform-forming (or control) sequence, and a 2,035 bp fragment contained the SV40 origin of replication which also contains IRs (see Fig. 2B). The plasmid DNA was also sequenced to further analyze the mutations.

Results

Probing the Cruciform Structure in Modified Plasmids. A control plasmid and two cruciform-forming plasmids containing AT- or CG-rich IRs (29, 30 bp with a 1, 2 bp spacer, respectively) were constructed to determine whether short IR sequences were mutagenic in mammalian cells. In order to confirm that the sequences were capable of cruciform extrusion, the plasmids were first treated with T7 endonuclease I to cleave at the four-way junction between B-DNA and the cruciform stems. Digestion with ScaI released ScaI-ScaI fragments containing the IR inserts. The shorter Sca1-T7 and T7-Sca1 fragments resulting from T7-induced DSBs at the cruciform structure colocalized with the IR insert (see Fig. 1B). While both

IRs formed cruciform structure, the AT-rich IR appeared to form cruciform structure on the plasmid more readily *in vitro* than the CG-rich IR. Thus, our data suggest that short CG-rich IRs appear to have the capacity to form cruciform structures *in vitro*, even though the formation of cruciforms at such sequences has been reported to be kinetically unfavorable due to the formidable CG bond barrier (12).

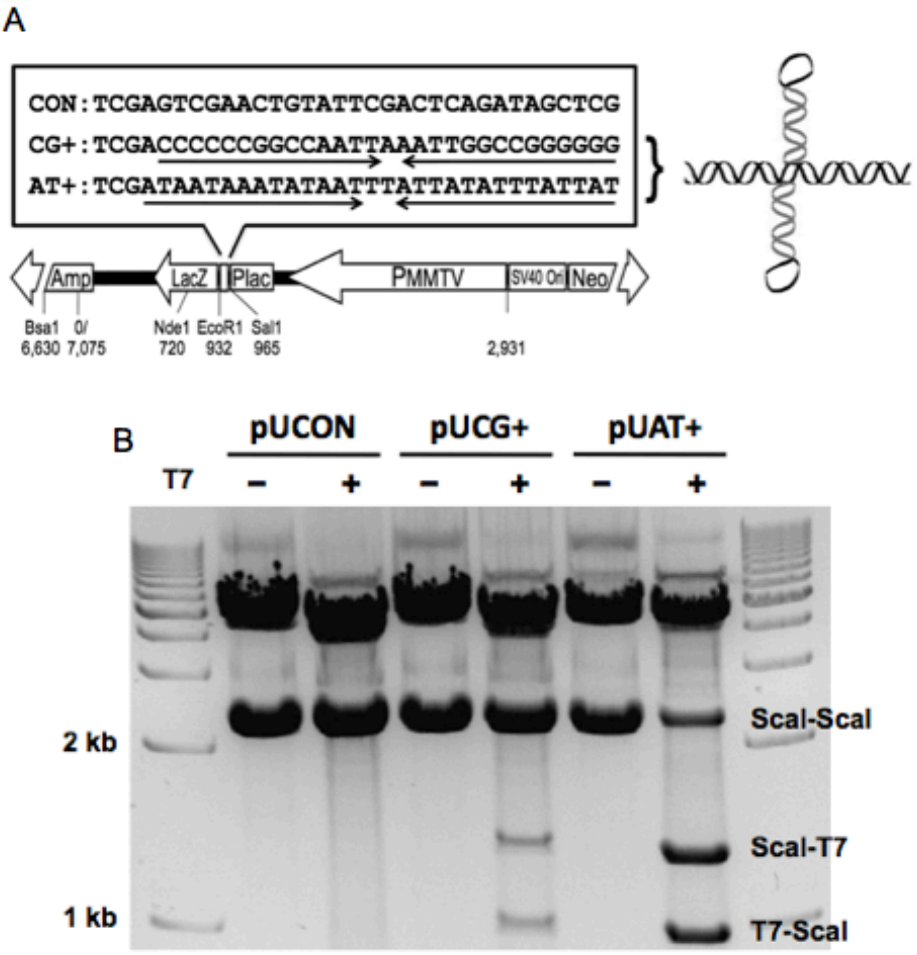

Figure 1. Inverted repeats form cruciform structures on reporter plasmids *in vitro*. (A) Schematic diagram of the reporter plasmid containing IR sequences; the cloning site is indicated, as are the restriction sites used throughout the study. (B) Detection of secondary DNA structures as assessed by T7 endonuclease sensitivity. Incubation with four-way junction-specific T7 endonuclease was followed by ScaI restriction enzyme digestion. The 1.2 kb ScaI-T7 and 1 kb T7-ScaI fragments indicated cruciform extrusion at the IR sequence. The ScaI-ScaI fragment served as a loading control.

Cruciform-induced Mutagenesis in Mammalian COS-7 Cells. To determine the mutagenic potential of cruciform structures formed at short IR sequences in mammalian cells, cruciform-forming and control plasmids were transfected into mammalian COS-7 cells and were allowed to replicate/repair for 48 hours. After plasmid recovery from mammalian cells, the frequencies of *lacZ* mutations were determined by blue-white screening (Fig. 2A). Both pUCG+ (9.2×10^{-3})

and pUAT+ (10.1×10^{-3}) had at least a three-fold higher mutation frequency above the control, pUCON (2.9×10^{-3}) in the lacZ' system. The low background frequency of mutations of all three plasmids generated in DH5α bacterial cells ($<1 \times 10^{-4}$) indicated that cruciform-forming sequences were either relatively stable, repaired without impacting the *lacZ'* gene, or that the mutations were not sensitive through the *LacZ'* system in *E. coli*. These findings indicate that both CG-rich ($p < 0.003$; Student's t-test) and AT-rich ($p < 0.001$; Student's t-test) IRs are mutagenic in mammalian COS-7 cells.

Cruciforms Induce Large-scale Deletions in Mammalian Cells. Twenty randomly picked mutants were subjected to restriction digestion analysis and/or sequencing to reveal the mutation spectrum. Greater than 90% of the cruciform-induced mutants generated in COS-7 cells contained large-scale deletions (Fig. 2C and Table 1). In the majority of the mutants, the deleted regions were greater than 200 bp and included the cruciform-forming sequences. In addition to spontaneous point mutations (20%), the control B-DNA insert usually remained intact in spite of large-scale deletions (74%). This supports the idea that the cruciform-forming sequences were responsible for the mutants generated from the modified plasmids (pUCG+ and pUAT+) in mammalian cells. Interestingly, some mutants generated in the cruciform-forming plasmids contained either a duplication of the single-stranded loop region or a deletion of only one arm of the stem (Fig. 2B). These events are strong indicators of cruciform structure-induced mutations. Furthermore, many junction breakpoints contained 1-6 bp microhomologies, suggesting that they were generated from DSBs that were then processed by

error-prone MMEJ repair (Fig. 3), the predominant form of DSB repair found in mammalian cells.

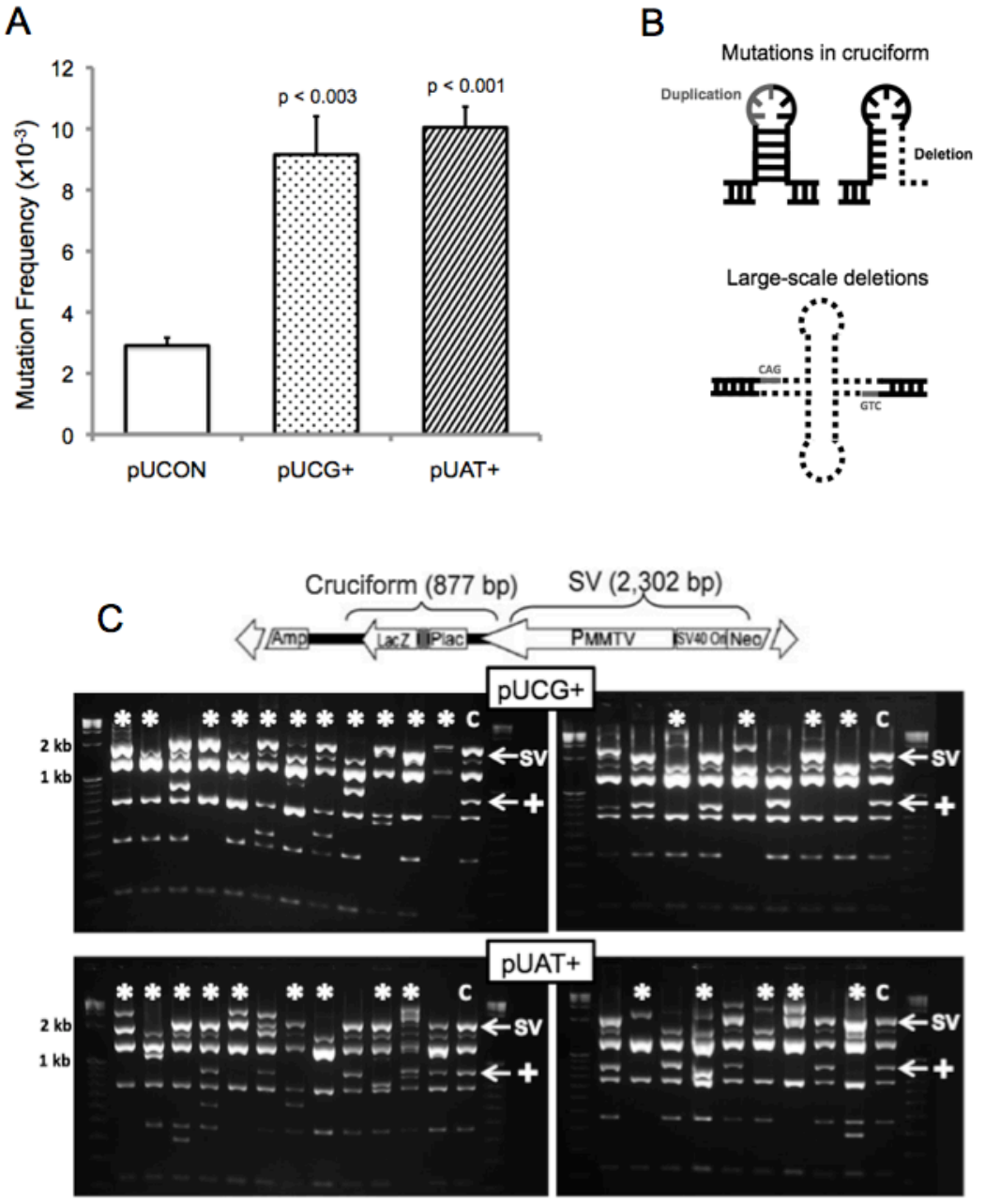

Figure 2. Cruciform-induced mutations in reporter plasmids in mammalian cells. (A) Mutation frequencies in the LacZ' system after 48 hrs of incubation in mammalian COS-7 cells. Error bars indicate SEM of 4 to 5 replicates. (B) Restriction digestion analysis of cruciform-induced *lacZ'* mutants in COS-7 cells. Plasmids were isolated from COS-7 cells 48 hours post-transfection and digested into 7 fragments after incubation with Eag1 and BssS1 restriction enzymes. An 877-bp fragment contains the cruciform-forming sequence (labeled with a '+'), and a 2,302-bp fragment contains the SV40 origin (labeled 'SV'). The lane marked "C" is DNA isolated from a wild-type colony used as a control.

Mutation	Cruciform-forming plasmids		Total	pUCON
	pUCG+	pUAT+		
Point Mutations	**0 (0%)**	**1 (3%)**	**1 (1%)**	**5 (26%)**
Mutations in the Cruciform*	**2 (5%)**	**1 (3%)**	**3 (4%)**	**0 (0%)**
Large-scale Deletions**	**39 (95%)**	**31 (94%)**	**70 (95%)**	**14 (74%)**
Total	**41 (100%)**	**33 (100%)**	**74 (100%)**	**19 (100%)**

Table 1. Mutation spectra on plasmids replicated in mammalian COS-7 cells. Plasmids pUCG+ and pUAT+ contain cruciform-forming sequences; plasmid pUCON contains canonical B-DNA-forming sequences and serves as a control. *Mutations in the cruciform sequence included a duplication of the stem region and "one-sided deletions" where only half of the stem was deleted are shown. **>50bp.

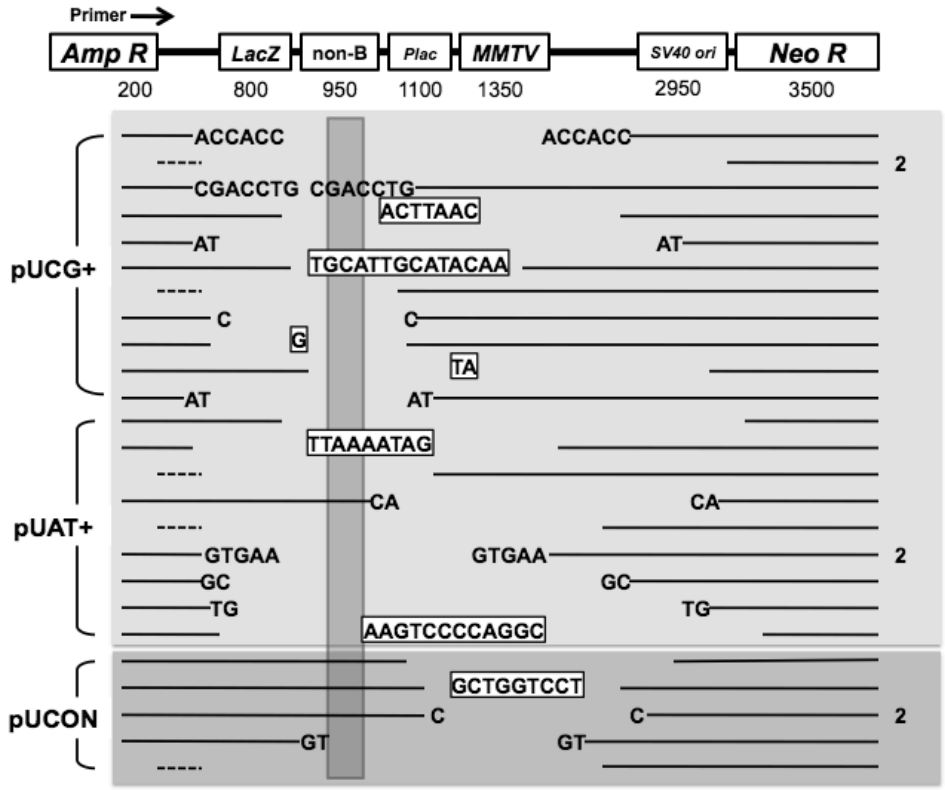

Figure 3. Schematic diagram of deletion mutants. The majority of cruciform-induced mutants characterized contained deletions that included the IR sequences. Letters that are boxed indicate insertions within the deleted regions. Numbers on the right designate the number of identical mutants. Dotted lines represent deletions that spanned near the primer sequence. Unboxed letters represent microhomologies.

Cruciform-forming Sequences Induce DSBs in Mammalian Cells. To provide direct evidence that cruciform-forming sequences led to DSBs in mammalian cells, ligation-mediated PCR was performed as previously described (21) using an upstream primer (≈200 bp from the IR) and a second primer within the short DNA duplex linker. Unique bands ≈200 bp from both cruciform-containing sequences indicated that these sequences were hot spots for DSBs (Fig. 4). Additional products ≈140 bp from the IR were either intermediates of DSB processing or the result of another hot-spot region upstream of the cruciform-forming sequences. No identifiable distinct bands for the DNA fragments were recovered from the control plasmid, suggesting that the DSBs were the result of processing of the cruciform-forming sequences. Sequencing of pGEM-T clones revealed that the majority of break sites were mapped 1 bp upstream of the AT-rich IR, while the breakpoints in the pUCG+ plasmid were 1-10 bp upstream of the sequence. This result suggested the presence of structure-specific nucleases that cleave at the base of the cruciform stem. Other possibilities include strand breaks as a result of polymerase stalling or increased susceptibility to DNA damage at single-stranded regions in the structure.

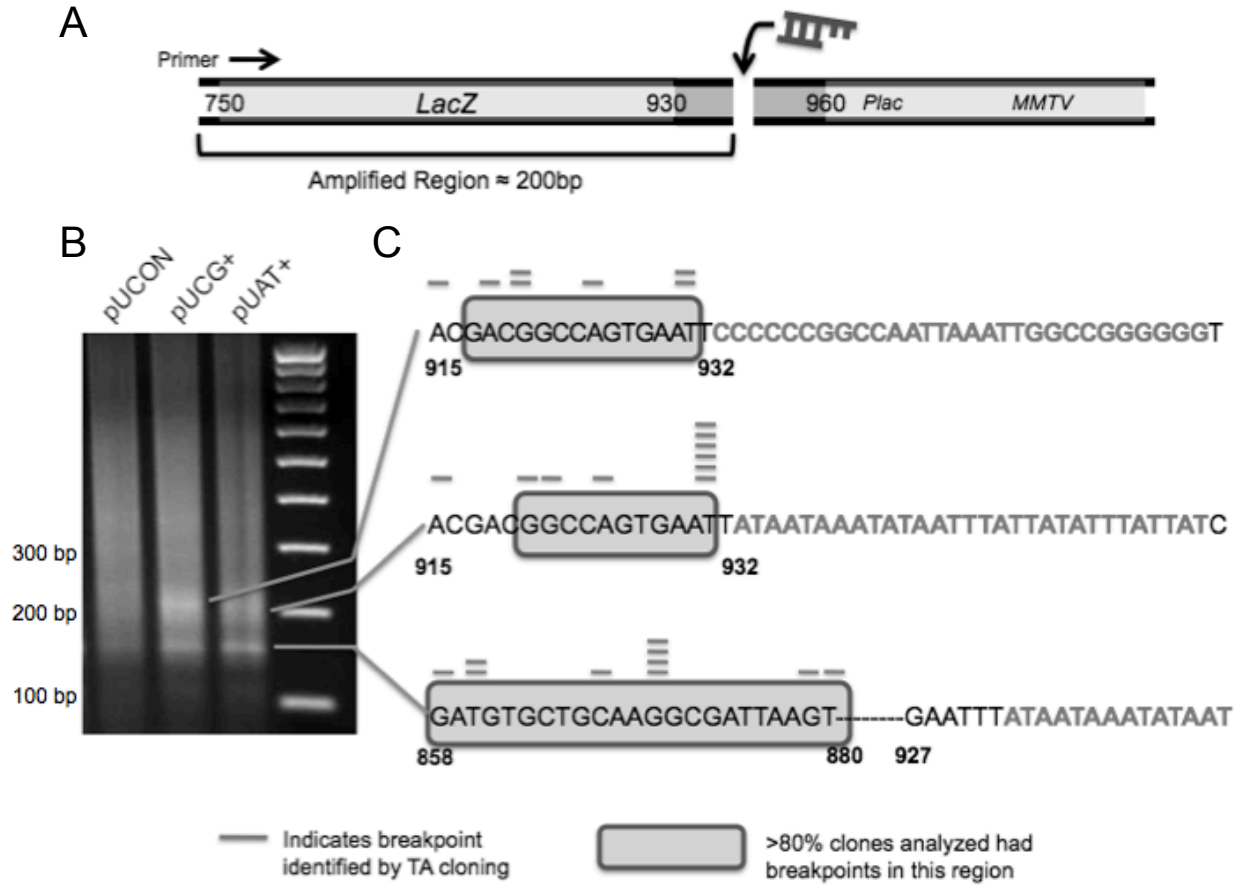

Figure 4. Mapping of cruciform-induced DNA double-strand breaks. (A) Schematic map of the LM-PCR amplified region between the upstream primer and the linker ligated at the breakpoint. (B) Electrophoresis of LM-PCR products using plasmids pUCON, pUCG+, and pUAT+ recovered from COS-7 cells 48 hours after transfection. (C) Sequencing of cloned PCR products purified from agarose gels, as indicated. Gray lines represent the incidence of breakpoints mapped at this position.

Short Inverted Repeats Can Stall DNA Replication. Two-dimensional gel electrophoresis of replication intermediates recovered from mammalian COS-7 cells determined if short cruciform-forming sequences can cause DNA replication fork stalling. The replication intermediates of pCON, pCG+, and pAT+ were purified using DpnI to removed unreplicated DNA and cleaved with NdeI and BsaI to isolate the cruciform-containing region on the plasmid.

First separated by mass, the intermediates were then resolved by mass and shape in the second dimension agarose gel to form a smooth replication arc. The arc generated from pCG+ sample had a bulge, indicating the accumulation of a replication fork intermediate at the short IR sequence (Fig. 5). Interestingly, the short AT-rich IR sequence did not appear to stall the DNA polymerase significantly compared to the control sequence. Therefore *in vivo*, short CG-rich IRs can cause detectable DNA replication blockage.

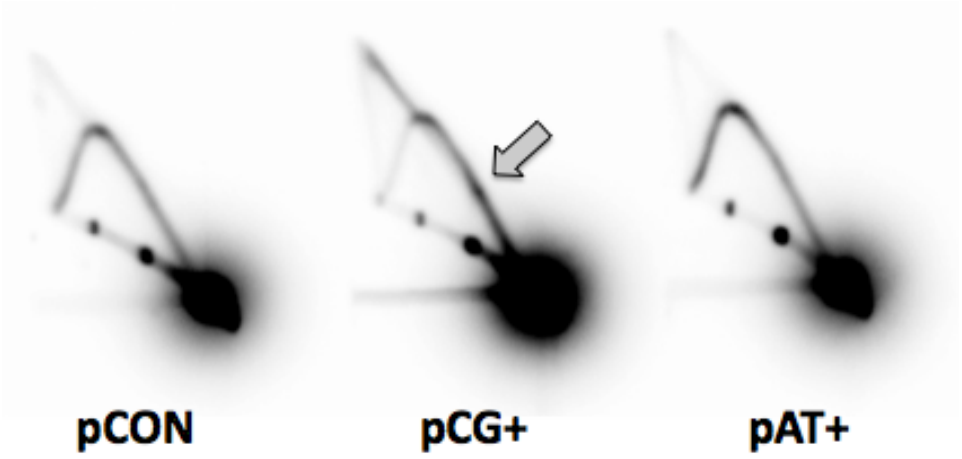

Figure 5. Short inverted repeats stall DNA replication. 2D gel electrophoresis of replication intermediates indicated that short CG-rich IRs cause significant replication fork stalling. The AT-rich IR did not cause substantial replication blockage compared to the control sequence. The arrow designates the accumulating of stalled replication fork intermediate.

Cruciform-induced Mutagenesis is Not Dependent on DNA Replication. Because previous studies have implicated the secondary structure formed at long IRs as the cause of replication fork stalling *in vivo* (23), SV40 large T antigen-deficient HeLa cell-free extracts were used to determine the role of replication on cruciform-induced genetic instability in mammalian cells. Following the incubation of modified plasmids in the extracts, blue-white screening and sequencing revealed that the CG-rich (4.9×10^{-4} vs. 1.5×10^{-4}: $p < 0.02$ SV40+; 4.3×10^{-4} vs. 0.9×10^{4}: $p < 0.01$ SV40-) and AT-rich (11.5×10^{-4}: $p < 0.01$ SV40+; 5.4×10^{-4}: $p <$

0.02 SV40-) IRs were more mutagenic than the control plasmid *in vitro* (Fig. 6A and 6B) and yielded similar mutation spectra regardless of the presence of large T antigen. Approximately 60% of the mutations detected were large-scale deletions (>500 bp) with microhomologies at the breakpoint junctions. Yet, LM-PCR indicated that the position of the majority of DSBs occurring in cell-free extracts differed with or without replication (Fig. 6C). The position of DSBs shifted upstream by approximately 50 bp from the IR when replication did not occur. With replication, the DSBs mapped directly adjacent to the IRs.

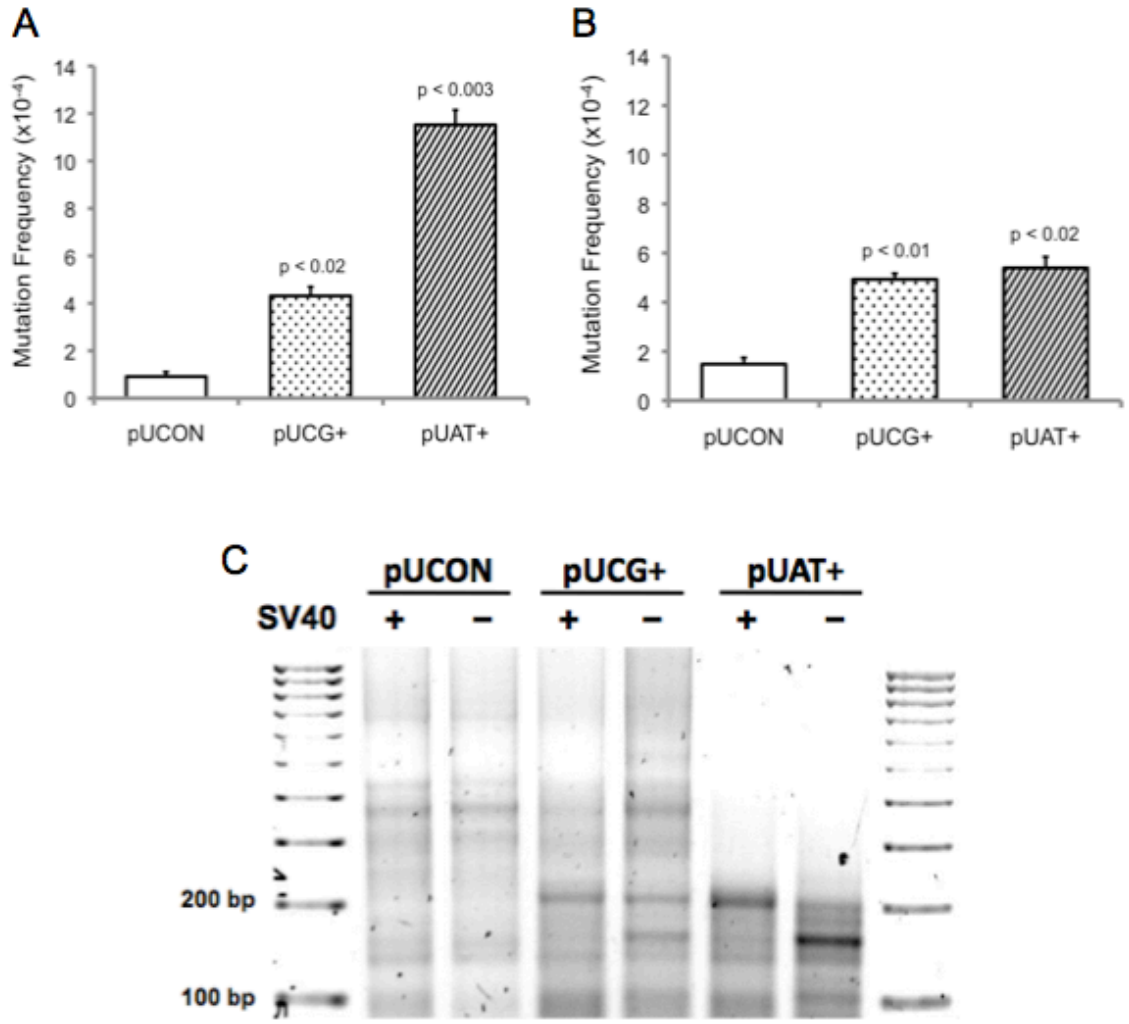

Figure 6. Cruciform-induced mutagenesis in replication-proficient or replication-deficient cell extracts. Plasmids were incubated in (A) replication-proficient (SV40+) or (B) replication-deficient (SV40-) HeLa cell-free extracts for 6 hours. Replication was allowed by addition of SV40 large T antigen. Error bars represent the SEM of 3 experiments. (C) LM-PCR mapped the difference in DSB positions depending the absence or presence of replication. The ≈200bp band in replicating extracts mapped directly adjacent to the IR insert. An additional unique band corresponding to ≈150 bp was also identified.

Discussion

The potential for hairpin/cruciform structures to form in the human genome is quite high, with IRs ≥21 bp found approximately once every 5,600 bp in the human genome (14). In addition to binding targets for homodimeric proteins or tumor suppressor p53, identification of IRs in origins of replication in both prokaryotic and eukaryotic genomes has provided support for a potential role of cruciform/hairpin secondary structures as binding sites for regulatory proteins (11, 24). Although it is clear that IRs have key biological functions *in vivo*, the extrusion of hairpin/cruciform structures, which can serve as physical barriers to normal cellular functions, has been associated with mutagenesis and human disease.

Palindromic IR sequences have been linked to recurrent translocations and various pathologies in mammalian cells. In humans, large palindromes (>400 bp) have been found in or near sites that undergo gross chromosomal rearrangements and the resolution of cruciform extrusions is being considered as a source of genetic instability (8, 19). Bioinformatics studies mapping breakpoints in cancer cell line genomes have implicated cruciforms as active elements in the genome that provide platforms for translocations or deletions (19). While triplet repeats are associated with several human neurological disorders, the information gathered from these simple repeats may not be directly applicable to perfect IR sequences since mismatches in the hairpin stems may influence the recognition and processing of these structures in the cell (25, 26). In this study, we examined the ability of short cruciform-forming sequences to induce genetic instability in both bacterial and mammalian cells. This present study fills in one of the gaps in understanding palindrome-mediated mutagenesis in mammalian cells; we demonstrate

that aside from large palindromes, the much more common short IRs can also form structures that mediate DSBs and deletions in mammalian cells.

Sequencing and restriction analyses revealed that >90% of the mutations generated were large-scale deletions (>50 bp) containing microhomologies, characteristic of DSB processing by MMEJ. IRs found in eukaryotic genomes are >60% AT-rich while most IRs in prokaryotes are <60% AT-rich (14); however, in this study neither IR (CG-rich nor AT-rich) had a significantly higher mutation frequency than that of the other. The low number of cruciform-induced mutants detected in bacteria suggests that these sequences are either relatively stable in prokaryotes or readily repaired using error-free homologous recombination. In contrast, MMEJ in mammalian cells has been implicated as a pathway responsible for chromosomal translocations and rearrangements associated with cancer (27). The two distinct DSB repair pathways operating in cells may explain the difference in mutation frequencies between prokaryotes and eukaryotes, since most prokaryotes generally employ an error-free homologous recombination mechanism to process DSBs, whereas mammalian cells can also repair DSBs efficiently through a second pathway, MMEJ (27). We recently reported that Z-DNA-forming sequences induce different types of mutations depending on the availability of HR or MMEJ repair systems in the cells (28).

Currently, there are two predominant mechanisms being proposed by which cruciform extrusion can induce genetic rearrangements in mammalian cells: 1) stalling of proteins involved in DNA metabolic processes such as replication or transcription when encountering the structure; and/or 2) structure-specific nucleases that cleave at the base of the cruciform. The differences in the DSB positions, mapped by LM-PCR, in the presence or absence of

replication in HeLa cell-free extracts suggests that both replication-dependent and replication-independent mechanisms exist. During replication or transcription, significant negative supercoiling is generated at regions trailing behind the polymerases due to unwinding of the DNA; it is this alleviation of negative superhelical tension that is the driving force for cruciform extrusion (13). Consequently, formation of these structures may be as transient as the dynamic levels of supercoiling within the cell, but their presence is sufficient to elicit processing that results in mutagenesis. Protein binding including that of repair proteins could also stimulate or stabilize formation of these alternative DNA structures in cells (24).

If not resolved by helicases or other components of the replication and transcriptional machinery, hairpin/cruciform structures can cause stalling of DNA polymerases and result in strand breaks at the replication fork (15, 23). These hairpin structures may also allow the polymerase to completely bypass the sequence, introducing deletions or insertions depending on whether the structures form on the template or nascent strand. Cruciform extrusion is possible under this situation particularly since the entire length of the short IR may be located within one Okazaki fragment and can be exposed as single-stranded DNA on the lagging-strand template. Evidence to support this idea has been demonstrated by the slow progression of replication forks at stem-loop structures when Alu-IRs are introduced into bacteria, yeast, and mammals (23). Even though small IRs can stall DNA replication, the lack of point mutations in our study indicates that the replication slippage model was not the dominant mechanism of cruciform-induced instability in our system. However, the majority of DSBs at the IRs in replication-proficient extracts suggests that the DNA breaks were likely the result of DNA polymerase stalling at the cruciform structure. The significance of the replication-dependent

model in cruciform-induced mutagenesis remains unclear, since the mutation frequencies and mutations identified in our system were similar in the presence or absence of replication. It is also possible that RNA polymerases can be stalled at non-B DNA structures, and consequentially, recruit repair proteins that cause DNA breaks as intermediates in the processing of the structures (29).

Alternatively, IRs may mediate deletions via direct processing of the unusual conformation by structure-specific endonucleases (9). Several proteins that can recognize and potentially cleave hairpin/cruciform structures have already been identified; the cruciform-binding protein, CBP (16), and topoisomerase II (30) represent two such examples. In addition to releasing torsional stress in DNA, topoisomerase II demonstrates structure-specific nuclease activity at the base of the hairpin that was determined significantly by both the size of the stem and loop regions (30). The cruciform structure and intraduplex angles resemble Holliday junctions formed during homologous recombination in mitosis, and therefore could present a substrate for Holliday junction resolvases (31, 32). Other likely candidates include the Mre11/Rad50/Nbs1 complex (33), the MSH2-MSH3 complex, and nucleotide excision repair factors. Recent work on SbcCD in *E.coli* (bacterial analog to eukaryotic Mre11 and Rad50) has demonstrated increased chromosomal rearrangements in the absence of the SbcCD complex in bacteria (34). It has also been suggested that cruciform formation may induce nucleosome phasing, which would make the structure itself and the surrounding DNA sequences accessible to replication factors and/or DNA damaging agents (15). With the mapping of DSBs ≈50 bp preceding the IRs (as shown in Fig. 6C), it can be suggested that cruciform resolution may be caused with the binding of repair proteins followed by cleavage upstream of the protein. The

observation that the cruciform-forming sequences used in this study were mutagenic in the absence of replication implicates a replication-independent pathway in palindrome-mediated mutagenesis, and is consistent with structure-specific proteins initiating DSBs at these sites.

The identification of mutants containing duplications of the spacer region between the IRs or deletions of the IR (see Table 1) provided further support for a structure-specific cleavage model of the secondary structures. For example, these mutations may have been initiated by a "center-break mechanism" where the cruciform tips are endonucleolytic targets (32). These deletion events together with results from LM-PCR (see Fig. 4 and 6C) suggest that either Holliday junction resolvases or other structure-specific/repair nucleases contributed to cruciform-induced mutagenesis in mammalian cells.

Collectively, the results from this study suggest that short palindromic cruciform-forming sequences are intrinsically mutagenic likely via two mechanisms: 1) DNA replication fork stalling and 2) structure-specific cleavage events that are not dependent on replication in mammalian cells. Resulting DSBs are then repaired in an error-generating fashion by a MMEJ mechanism. Our findings support the idea that palindrome-mediated rearrangements are a universal feature of genome plasticity and may be a contributor to human disease and a driving force in the evolution of the human genome by providing a means for diversity within the population. The contribution of short IRs to genome instability may be underappreciated relative to our knowledge of the stimulation of persistent rearrangements by rarer, longer IRs. A better understanding of the factors that contribute to genomic instability provides the potential for the development of novel therapeutic strategies to advance current methodologies used to treat or prevent diseases of genetic instability.

Acknowledgements

Dr. Guliang Wang helped with the training and guidance through this study. Scott Spitzer helped considerably with the methods in the two-dimensional gel electrophoresis and performed the Southern blot. This work was supported by an NCI grant to K.M.V. (CA093729), and an NIEHS Center grant ES07784 for Facility Core Services.

References

1. Popescu NC (2003) Genetic alterations in cancer as a result of breakage at fragile sites. *Cancer Lett* 192(1):1-17.
2. Burrow AA, Williams LE, Pierce LC, & Wang YH (2009) Over half of breakpoints in gene pairs involved in cancer-specific recurrent translocations are mapped to human chromosomal fragile sites. *BMC Genomics* 10:59.
3. Bacolla A & Wells RD (2004) Non-B DNA conformations, genomic rearrangements, and human disease. *J Biol Chem* 279(46):47411-47414.
4. Raghavan SC, Tsai A, Hsieh CL, & Lieber MR (2006) Analysis of non-B DNA structure at chromosomal sites in the mammalian genome. *Methods Enzymol* 409:301-316.
5. Frappier L, Price GB, Martin RG, & Zannis-Hadjopoulos M (1989) Characterization of the binding specificity of two anticruciform DNA monoclonal antibodies. *J Biol Chem* 264(1):334-341.
6. Collins J (1981) Instability of palindromic DNA in Escherichia coli. *Cold Spring Harb Symp Quant Biol* 45 Pt 1:409-416.
7. Nasar F, Jankowski C, & Nag DK (2000) Long palindromic sequences induce double-strand breaks during meiosis in yeast. *Mol Cell Biol* 20(10):3449-3458.
8. Kurahashi H, *et al.* (2004) Cruciform DNA structure underlies the etiology for palindrome-mediated human chromosomal translocations. *J Biol Chem* 279(34):35377-35383.
9. Wang G & Vasquez KM (2006) Non-B DNA structure-induced genetic instability. *Mutat Res* 598(1-2):103-119.
10. Wang G, Carbajal S, Vijg J, DiGiovanni J, & Vasquez KM (2008) DNA structure-induced genomic instability in vivo. *J Natl Cancer Inst* 100(24):1815-1817.
11. Bissler JJ (1998) DNA inverted repeats and human disease. *Front Biosci* 3:d408-418.
12. Courey AJ & Wang JC (1983) Cruciform formation in a negatively supercoiled DNA may be kinetically forbidden under physiological conditions. *Cell* 33(3):817-829.
13. Lilley DM (1980) The inverted repeat as a recognizable structural feature in supercoiled DNA molecules. *Proc Natl Acad Sci U S A* 77(11):6468-6472.
14. Schroth GP & Ho PS (1995) Occurrence of potential cruciform and H-DNA forming sequences in genomic DNA. *Nucleic Acids Res* 23(11):1977-1983.
15. Pearson CE, Zorbas H, Price GB, & Zannis-Hadjopoulos M (1996) Inverted repeats, stem-loops, and cruciforms: significance for initiation of DNA replication. *J Cell Biochem* 63(1):1-22.
16. Zannis-Hadjopoulos M, Yahyaoui W, & Callejo M (2008) 14-3-3 cruciform-binding proteins as regulators of eukaryotic DNA replication. *Trends Biochem Sci* 33(1):44-50.
17. Cox R & Mirkin SM (1997) Characteristic enrichment of DNA repeats in different genomes. *Proc Natl Acad Sci U S A* 94(10):5237-5242.
18. Sinden RR, *et al.* (2002) Triplet repeat DNA structures and human genetic disease: dynamic mutations from dynamic DNA. *J Biosci* 27(1 Suppl 1):53-65.
19. Tanaka H, Bergstrom DA, Yao MC, & Tapscott SJ (2006) Large DNA palindromes as a common form of structural chromosome aberrations in human cancers. *Hum Cell* 19(1):17-23.

20. Kurahashi H & Emanuel BS (2001) Unexpectedly high rate of de novo constitutional t(11;22) translocations in sperm from normal males. *Nat Genet* 29(2):139-140.
21. Rodriguez H, *et al.* (2000) Mapping oxidative DNA damage using ligation-mediated polymerase chain reaction technology. *Methods* 22(2):148-156.
22. Krasilnikova MM & Mirkin SM (2004) Analysis of triplet repeat replication by two-dimensional gel electrophoresis. *Methods Mol Biol* 277:19-28.
23. Voineagu I, Narayanan V, Lobachev KS, & Mirkin SM (2008) Replication stalling at unstable inverted repeats: interplay between DNA hairpins and fork stabilizing proteins. *Proc Natl Acad Sci U S A* 105(29):9936-9941.
24. Jagelska EB, Pivonkova H, Fojta M, & Brazda V (2010) The potential of the cruciform structure formation as an important factor influencing p53 sequence-specific binding to natural DNA targets. *Biochem Biophys Res Commun* 391(3):1409-1414.
25. Owen BA, *et al.* (2005) (CAG)(n)-hairpin DNA binds to Msh2-Msh3 and changes properties of mismatch recognition. *Nat Struct Mol Biol* 12(8):663-670.
26. Mirkin SM (2007) Expandable DNA repeats and human disease. *Nature* 447(7147):932-940.
27. McVey M & Lee SE (2008) MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends Genet* 24(11):529-538.
28. Kha DT, Wang G, Natrajan N, Harrison L, & Vasquez KM (2010) Pathways for double-strand break repair in genetically unstable Z-DNA-forming sequences. *J Mol Biol* 398(4):471-480.
29. Ditlevson JV, *et al.* (2008) Inhibitory effect of a short Z-DNA forming sequence on transcription elongation by T7 RNA polymerase. *Nucleic Acids Res* 36(10):3163-3170.
30. Froelich-Ammon SJ, Gale KC, & Osheroff N (1994) Site-specific cleavage of a DNA hairpin by topoisomerase II. DNA secondary structure as a determinant of enzyme recognition/cleavage. *J Biol Chem* 269(10):7719-7725.
31. Lewis SM & Cote AG (2006) Palindromes and genomic stress fractures: bracing and repairing the damage. *DNA Repair (Amst)* 5(9-10):1146-1160.
32. Lobachev KS, Rattray A, & Narayanan V (2007) Hairpin- and cruciform-mediated chromosome breakage: causes and consequences in eukaryotic cells. *Front Biosci* 12:4208-4220.
33. Trujillo KM & Sung P (2001) DNA structure-specific nuclease activities in the Saccharomyces cerevisiae Rad50*Mre11 complex. *J Biol Chem* 276(38):35458-35464.
34. Darmon E, *et al.* (2010) E. coli SbcCD and RecA control chromosomal rearrangement induced by an interrupted palindrome. *Mol Cell* 39(1):59-70.