HHS Public Access

Author manuscript

Cell Rep. Author manuscript; available in PMC 2018 June 21.

Published in final edited form as:

Cell Rep. 2018 January 30; 22(5): 1200–1210. doi:10.1016/j.celrep.2018.01.014.

Distinct Mechanisms of Nuclease-Directed DNA-Structure-Induced Genetic Instability in Cancer Genomes

Junhua Zhao[1,4], **Guliang Wang**[1,4], **Imee M. del Mundo**[1], **Jennifer A. McKinney**[1], **Xiuli Lu**[1], **Albino Bacolla**[1], **Stephen B. Boulware**[1], **Changsheng Zhang**[2], **Haihua Zhang**[3], **Pengyu Ren**[2], **Catherine H. Freudenreich**[3], and **Karen M. Vasquez**[1,5,*]

[1]Division of Pharmacology and Toxicology, College of Pharmacy, The University of Texas at Austin, Dell Pediatric Research Institute, 1400 Barbara Jordan Boulevard, Austin, TX 78723, USA

[2]Department of Biomedical Engineering, The University of Texas at Austin, 107 W Dean Keeton Street, Austin, TX 78712, USA

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

SUMMARY

Sequences with the capacity to adopt alternative DNA structures have been implicated in cancer etiology; however, the mechanisms are unclear. For example, H-DNA-forming sequences within oncogenes have been shown to stimulate genetic instability in mammals. Here, we report that H-DNA-forming sequences are enriched at translocation breakpoints in human cancer genomes, further implicating them in cancer etiology. H-DNA-induced mutations were suppressed in human cells deficient in the nucleotide excision repair nucleases, ERCC1-XPF and XPG, but were stimulated in cells deficient in FEN1, a replication-related endonuclease. Further, we found that these nucleases cleaved H-DNA conformations, and the interactions of modeled H-DNA with ERCC1-XPF, XPG, and FEN1 proteins were explored at the sub-molecular level. The results suggest mechanisms of genetic instability triggered by H-DNA through distinct structure-specific, cleavage-based replication-independent and replication-dependent pathways, providing critical evidence for a role of the DNA structure itself in the etiology of cancer and other human diseases.

In Brief

*Correspondence: karen.vasquez@austin.utexas.edu.

[4]These authors contributed equally

[5]Lead Contact

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, and one table and can be found with this article online at https://doi.org/10.1016/j.celrep.2018.01.014.

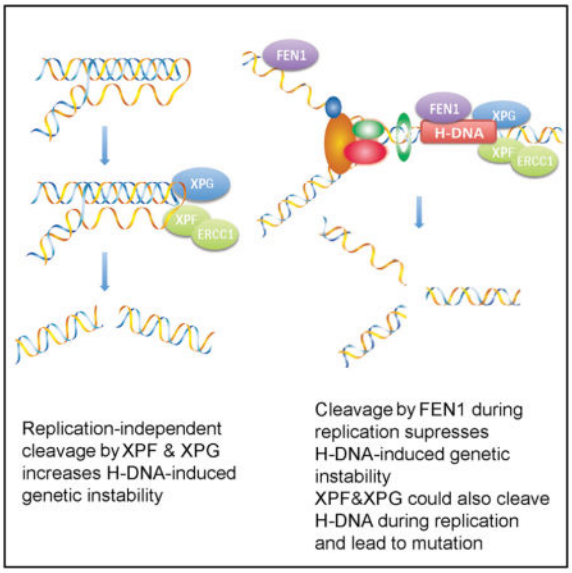
AUTHOR CONTRIBUTIONS

J.Z., G.W., and K.V. designed the study; J.Z., G.W., I.M.d.M., J.A.M., A.B., and S.B.B. performed experiments; H.Z. and C.H.F. provided materials and methods on yeast artificial chromosomes; X.L., C.Z., G.W., K.M.V., and P.R. designed and performed the structural simulation studies; and J.Z., G.W., A.B., X.L., and K.M.V. analyzed data and wrote the manuscript. All authors discussed the results and commented on the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

DNA sequences that can adopt alternative structures, such as H-DNA, have been implicated in cancer etiology. Zhao et al. found that such sequences are enriched at translocation breakpoints in human cancer genomes and that repair-and replication-related nucleases cleave H-DNA in both error-free “replication-related” and mutagenic “replication-independent” mechanisms.



INTRODUCTION

Genetic instability often occurs at particular endogenous “hotspot” regions in the genome and is associated with the occurrence of cancer and other human diseases (Popescu, 2003); however, the mechanisms involved in generating these hotspots are largely unknown. Whereas malfunction of trans-factors, such as DNA repair proteins, is likely to contribute to global genome-wide instability events, *cis* elements must play a role in locally accumulated DNA damage and mutation at mutation hotspots. Interestingly, repetitive DNA elements often co-localize with endogenous mutation and breakpoint hotspots in cancer and neurological disorders, and many of them can adopt structures that differ from canonical B-form DNA (non-B DNA) (Choi and Majima, 2011; Pestov et al., 1991; Wells et al., 2005). In addition to important biological roles, such as regulating DNA replication, transcription, and recombination, non-B-DNA-forming sequences have been shown to stimulate mutagenesis in the absence of exogenous DNA damage (Kinniburgh, 1989; Vasquez and Wang, 2013; Wang and Vasquez, 2006, 2014; Zhao et al., 2010). Thus, we speculated that non-B DNA may be involved in generating endogenous mutation hotspots.

An example includes a 23-bp polypurine/polypyrimidine H-DNA-forming mirror repeat in the human *c-MYC* gene found at a translocation hotspot in Burkitt lymphoma (Mirkin et al., 1987). H-DNA forms at polypurine-polypyrimidine regions with mirror repeat symmetry, where half of the tract is separated into single-stranded DNA during DNA metabolic processes (e.g., replication, transcription, and repair) and one of the single strands winds back and pairs with the purine strand in the remaining duplex via Hoogsteen hydrogen bonding, forming a three-stranded helix (Frank-Kamenetskii and Mirkin, 1995; Lyamichev

et al., 1986). We found that this H-DNA-forming sequence stimulated the formation of DNA double-strand breaks (DSBs) and induced genetic instability in mammals (Wang et al., 2008; Wang and Vasquez, 2004), providing strong evidence that H-DNA is a causative factor in genetic instability.

Despite its role in genetic instability, the biological significance of H-DNA (and other non-B DNA structures) in cancer development, and the mechanisms involved are still not clear. We have established a computer program to search for non-B-DNA-forming sequences, including potential H-DNA-forming sequences in any given sequence (Wang et al., 2013). In this study, we searched a human cancer genome database and found that H-DNA-forming sequences are enriched at chromosomal translocation breakpoints, supporting the roles of H-DNA in cancer etiology. We further demonstrated that the DNA-repair-related nucleases ERCC1-XPF and XPG cleaved H-DNA and were involved in a mechanism of genetic instability independent of DNA replication, whereas the DNA-replication-related flap endonuclease 1 (FEN1) protein was able to cleave H-DNA and was involved in suppressing H-DNA-induced mutagenesis in a replication-dependent fashion.

RESULTS

H-DNA-Forming Sequences Are Enriched Near Chromosomal Translocation Hotspots in Human Cancers

To determine the impact of H-DNA in human cancer etiology, we mapped potential H-DNA (triplex)-forming repeats (TFRs) within ±100 bp (thereafter referred as to bins) of 19,956 translocation breakpoints from sequenced cancer genomes (COSMIC at http://cancer.sanger.ac.uk/cosmic). We determined the distributions of each TFR midpoint relative to the breakpoint positions (taken to be 0). Both the percent of bins harboring TFRs ≥6 bp on each mirror repeat arm and the number of TFRs (normalized to 20,000 bins) were greater for translocations (35.3 ± 6.6 and 6,676) than for those in 20,000 random-pick control sequences extracted from the reference human genome (24.9 ± 4.9 and 4,715; t test; p value 4.2×10^{-50}; Figure 1A). In addition, the number of TFRs in each bin surrounding the translocation breakpoints was consistently higher than in controls (Figure 1B). Thus, H-DNA-forming sequences (TFRs) are enriched surrounding human cancer translocation breakpoints, implicating H-DNA in endogenous chromosomal instability and cancer. Notably, the 23-bp TFR in the *c-MYC* gene near a translocation breakage hotspot in lymphoma, as discussed above, was found in three translocation events in the human cancer translocation dataset (data not shown). Although this G-rich repeat also has the potential to form a G-quadruplex structure, we recently demonstrated that this particular sequence preferentially forms a stable H-DNA structure (Del Mundo et al., 2017). Moreover, we previously found that a variant of this sequence that forms H-DNA only, but not the sequence that supports G-quadruplex formation only, stalls transcription similar to the native sequence (Belotserkovskii et al., 2007). Thus, we used this sequence as a model H-DNA-forming sequence in the following genetic and biochemical mechanistic studies.

Nucleotide Excision Repair Proteins Are Involved in H-DNA-Induced DSBs and Genetic Instability in Yeast and Human Cells

To determine the mechanisms involved in H-DNA-induced muta-genesis *in vivo*, the 23-bp TFR from the human *c-MYC* promoter or a B-DNA control sequence was inserted into a yeast artificial chromosome (YAC) (Callahan et al., 2003) adjacent to the selectable *URA3* gene (Figure 2A). H-DNA-induced DSBs on the YACs result in loss of *URA3*, rendering the cells resistant to treatment with 5-fluoroorotic acid (5-FOA^R). We used this system to screen a yeast deletion library (GSA-5 from ATCC, Manassas, VA) to identify genes involved in H-DNA-induced mutagenesis *in vivo*, and the results were confirmed in human cells.

In wild-type (WT) BY4742 yeast cells, the H-DNA-forming sequence stimulated chromosome breakage ~9.4-fold above the control sequence (Figure 2B; 11.60×10^{-5} versus 1.24×10^{-5}). PCR and Southern blotting confirmed that the YAC underwent DSBs and lost the distal end of the chromosome in the majority of the *URA3*-deficient cells (data not shown), consistent with our prior findings in mammalian cells (Wang et al., 2008; Wang and Vasquez, 2004). Strikingly, in *rad1Δ* (*XPF*) and *rad10Δ* (*ERCC1*) yeast strains that are deficient in the nucleotide excision repair (NER) endonuclease complex Rad1-Rad10, H-DNA-induced chromosomal breakage was reduced from 9.4-fold above background in the WT to 3.3-fold and 2.0-fold above control, respectively (Figure 2B). These results clearly indicate that the NER endonuclease complex Rad1-Rad10 (ERCC1-XPF) is necessary for H-DNA-induced chromosomal breaks and subsequent genetic instability *in vivo*.

To characterize the effects of NER on the mutagenic events induced by H-DNA in human cells, we measured H-DNA-induced mutagenesis in a *supF*-reporter system (Wang and Vasquez, 2004; Figure 2C) in cells derived from human xeroderma pigmentosum (XP) patients (Evans et al., 1997; Wu et al., 2007). Consistent with our results in yeast, the H-DNA-induced mutations were substantially reduced from 11.1-fold above control in WT to 2.0-fold in XPF-deficient human cells (Figure 2D; t test of raw data; p value < 0.05). H-DNA-induced mutagenesis was also substantially reduced (from 11.1-fold to 2.9-fold) in XPA-deficient human cells compared to isogenic controls (Figure 2D; t test of raw data; p value < 0.05), suggesting a role for functional NER in the mutagenic processing of H-DNA.

Moreover, analysis of the H-DNA-induced mutants produced in the human cells revealed substantial differences between XPF-deficient and XPF-proficient cells; the vast majority (77%) of H-DNA-induced mutations in human XPF-proficient cells were deletions, which were substantially reduced to 41% in the XPF-deficient cells (Table 1; χ^2 test; p value < 0.01). These results suggest that XPF is involved in H-DNA processing, leading to DNA breakage and deletions near the H-DNA structure-forming sequences. In addition, we found that XPG, another structure-specific endonuclease required for NER, was also involved in H-DNA-induced mutagenesis in human cells; although its absence did not significantly reduce H-DNA-induced mutation frequencies in human cells (Figure S1), XPG deficiency resulted in fewer deletions (33%) compared to XPG-proficient cells (71%: Table 1; χ^2 test; p value < 0.05).

H-DNA-induced large deletions that contained microhomologies at the junctions in human cells (or arm loss in yeast) suggested that H-DNA stimulated the formation of DSBs *in vivo*.

In fact, DSB hotspots were detected near the H-DNA-forming sequences on substrates recovered from XPF-proficient cells by ligation-mediated PCR (LM-PCR). Importantly, the absence of XPF altered the distribution of H-DNA-induced DSB hotspots in human cells. In XPF-proficient cells, a strong PCR signal of ~230 bp was detected (Figure 2E). From the location of the upstream specific primer, the ~230-bp PCR product mapped DSBs at or near the center of the H-DNA-forming sequence (in the linker sequence between the two mirror repeats). This band, along with an ~180-bp band that mapped to a breakpoint 40 bp upstream of the H-DNA-forming sequence, was substantially reduced in XPF-deficient cells, whereas a PCR product of ~210 bp was enhanced in XPF-deficient cells (Figure 2E, compare lanes 3 and 4). The PCR products shown in Figure 2E were purified and cloned into the pGEM-T vector, and ~30 colonies from each sample (XPF+ and XPF−) were sequenced. The breakpoints were mapped, as shown in Figure 2F, and the distributions of the breakpoints approximately matched the lengths of the LM-PCR products in Figure 2E. These results suggest a role for ERCC1-XPF cleavage in processing H-DNA structures to generate DSB intermediates, resulting in deletions and subsequent genetic instability.

ERCC1-XPF and XPG Nucleases Cleave H-DNA

The biological and genetic studies described above suggest that ERCC1-XPF and XPG are involved in DNA breakage and mutation caused by H-DNA, but whether H-DNA also represents a substrate for these nucleases has not been investigated. The preferred substrates for ERCC1-XPF and XPG include NER pre-incision bubbles containing double-stranded DNA (dsDNA)/single-stranded DNA (ssDNA) junctions, stem loops, and 3′ and 5′ flaps (Evans et al., 1997). To determine whether H-DNA is a substrate for ERCC-XPF, we incubated an H-DNA substrate formed by a 5′-radiolabeled oligonucleotide, MCR2–5′ (Figures 2G and S2), in XPF-proficient and -deficient human cell extracts. The NER cleavage activities in the human cell extracts were examined by their activities on a Y-splayed substrate, which is a preferred structure for XPF cleavage. As expected, XPF-proficient cell extracts cleaved the Y-splayed structure on the duplex region, −1 and −2 to the junction (Figure S3). In XPF-proficient human cell extracts, we found that the intra-molecular H-DNA was cleaved at the loop between the two Hoogsteen hydrogen-bonded strands, resulting in a 37-nt cleavage product (Figure 2H). This cleavage differs from the incision sites of XPF on Y-type junctions, which would either be on the duplex near the Watson-Crick hydrogen-bonded loop region (“TCCC” sequence upstream to the “TTTTT” loop on the right side in Figure 2G) or the “GGGG” region in Watson-Crick hydrogen-bonded duplex adjacent to the “CGCTTA” loop (loop on the left side in Figure 2G). Therefore, this cleavage appears to be specific to H-DNA with incision at the loop adjacent to the Hoogsteen hydrogen-bonded triplex region. This cleavage was dramatically reduced in human XPF-deficient cell extracts (Figure 2H), revealing a direct activity of XPF on H-DNA. Purified human recombinant XPG also cleaved the H-DNA substrate, removing the 5′-ssDNA flap at the junction adjacent to the Watson-Crick base-paired duplex (Figure 2I; marked in Figure 2G). Thus, our results demonstrate that ERCC1-XPF and XPG cleave H-DNA, implicating these NER nucleases in the mutagenic processing of H-DNA.

FEN1 Cleaves H-DNA and Suppresses H-DNA-Induced Mutation In Vivo

To determine whether other structure-specific endonucleases process H-DNA structures, we performed mutagenesis assays in a *RAD27*-deleted yeast strain, a homolog of human *FEN1*, which shares a conserved active site with XPG (Hohl et al., 2007). Surprisingly, in contrast to the effects of XPG and XPF on H-DNA-induced genetic instability, RAD27 deficiency dramatically *increased* H-DNA-induced mutagenesis by ~794% in the *rad27Δ* strain relative to WT cells (Figure 3A; t test; p value < 0.01). Likewise, depletion of FEN1 (<10% of WT levels; Figure S4) in human HeLa cells increased H-DNA-induced genetic instability to ~485% over WT cells (Figure 3B; t test; p value < 0.05). Analysis of the H-DNA-induced mutants generated in FEN1-depleted HeLa cells revealed a significant increase in large deletions (100%) compared to WT cells (Table 1; χ^2 test; p value < 0.01), suggesting that FEN1 may be involved in resolving H-DNA structures. Indeed, we found that human recombinant FEN1 (Shen et al., 1997) cleaved the MCR2–5′-derived H-DNA substrate at the base of the 5′ flap, yielding a 10-nt cleavage product (Figure 3C) similar to that detected by XPG (marked in Figure 2G). Thus, we have identified H-DNA as an unexpected substrate for FEN1. These results indicate that FEN1 prevents genomic instability at H-DNA regions, perhaps by resolving mutagenic H-DNA structures.

Prediction and Simulation of the Interactions of NER Factors and the FEN1 Protein with an H-DNA Structure

To obtain detailed insight into the interactions of these proteins with H-DNA, we proposed structural models of human hFEN1, hXPG, and hERCC1-XPF bound to H-DNA. The X-ray crystal structure of hFEN1 was downloaded from the PDB. The human XPG catalytic core (hXPG 1–155 and 762–984) and the human ERCC1-XPF complex were constructed from homology modeling. Similarly, the H-DNA structure for the sequence used in this work (Figure 4) was assembled by using available experimental structures as templates and further refined by using molecular dynamics simulations. The modeling details are described in the Experimental Procedures. The program HADDOCK (Dominguez et al., 2003; van Zundert et al., 2016) was then used to predict the complex structure and interactions between each protein with H-DNA.

The results of *in silico* docking and molecular dynamic simulations revealed that the $(HhH)_2$ domain of ERCC1 in the ERCC1-XPF complex interacts with the Hoogsteen-Loop (H-loop in Figure 4A; the loop formed when the homopurine strand winds back to serve as a third strand in the complex) of the H-DNA. The metal ion site on hXPF interacts with the H-loop at the cleavage site regions (as identified in the cleavage assay shown in Figures 2H and 2G). The active site residues (K716 and D720) of the nuclease domain of hXPF are located in close proximity to these cleavage sites. The α2 (S717-N727) helix (Figure 4A, middle, yellow helix) of the hXPF nuclease domain inserts into the minor groove between the homopyrimidine strand and the homopurine strand of the Watson-Crick base-paired duplex within the H-DNA structure. This model suggests a potential mechanism for the protein-DNA interaction and cleavage.

The nuclease core domain of human XPG (hXPG) was modeled, and we illustrated how the H-DNA traverses the path of basic residues on the electrostatic surface of hXPG and cations

(Figure 4B). In this model, the β hairpin between β6 and β7 of hXPG (Figure 4B, middle, cyan sticks) interacts with the upstream region of the homopyrimidine strand and the H-loop. The helix-2turn-helix (H2TH) motif (Figure 4B, middle, yellow residues) binds to the downstream region of the homopurine third strand. The one-metal ion active site (Figure 4B, middle, yellow ball) between the thumb and forefinger binds to the cleavage sites on the homopyrimidine strand upstream to the triple-stranded structure as shown in Figure 2G. The α1 helix (Figure 4B, middle, red helix) and the highly conserved active residues, α4 K84 and R91, are also located in close proximity to the cleavage sites on H-DNA (Figure 4B, right).

Human FEN1 (hFEN1) has a mixed α/β structure containing a seven-stranded twisted β sheet core and 15 α helices. The active site contains two Mg^{2+} ions, and the overall shape resembles a left-handed boxing glove (Figure 4C; Tsutakawa et al., 2011). In the simulated model of the hFEN1-H-DNA complex, H-DNA contacts the “palm and fingers” along a positively charged patch spanning the width of the boxing glove. The direction of H-DNA in the hFEN1-H-DNA complex is opposite to that seen in the hXPG-H-DNA complex. The electrostatic surface (−65 mV to +65 mV) of hFEN1 and cations shows that the H-DNA interfaces with the basic residues (Figure 4C, left). The homopurine third strand of H-DNA is nicely embraced by hFEN1 via its positively charged patch. In this model, the β hairpin between β6 and β7 (Figure 4C, middle, cyan sticks) interacts with the 3′ end of the third strand and the homopurine strand in the duplex. The H2TH motif at the glove base (Figure 4C, middle, yellow sticks) binds to the 5′ region of the third strand and the 5′ region of the homopyrimidine strand within the duplex. The α1 helix of hFEN1 (Figure 4C, middle, red helix) interacts with the homopurine strand in the duplex and the homopurine third strand. The two-metal ion active site (Figure 4C, middle, green) between the thumb and forefinger binds the cleavage sites on the homopyrimidine strand in the duplex, upstream to the triplex region (see Figure 2G for cleavage sites). Figure 4C (right panel) shows the atomic details of how the two Mg2+ ions and the highly conserved active residues, α4 K93 and R100, are coordinated near the cleaved sites on H-DNA to facilitate electrostatic rate acceleration of the phosphodiester hydrolysis (Sengerová et al., 2010). A recent study suggested that the electrostatic interaction between the phosphate backbone and the positive FEN1 residues, particularly K93, inverts the phosphodiester backbone of ssDNA, a key mechanism in promoting 5′ flap specificity and incision (Tsutakawa et al., 2017). This is consistent with our observation that the K93 residue is close to the H-DNA phosphate group, although the distance is a bit further and the interaction is not as strong as those observed in the FEN1-ssDNA complex shown in Tsutakawa et al. (2017). This difference is not surprising because ssDNA has greater flexibility to undergo significant conformational changes than densely packed H-DNA.

NER Proteins and FEN1 Interact with H-DNA In Vivo

The interactions of NER and FEN1 proteins with H-DNA were confirmed by chromatin immunoprecipitation (ChIP) assays in yeast. Using primers flanking the human H-DNA-forming sequence on the YAC (Table S1), we detected a 12-fold enrichment of Rad27 (FEN1) at the H-DNA region over the control B-DNA region (Figure 5A). Enrichment of Rad1 (XPF) and Rad2 (XPG) was also detected at the H-DNA region, 12-fold and 14-fold

above the control region, respectively (Figure 5A). Notably, the absence of Rad14 (XPA) reduced Rad1 (XPF) binding to the H-DNA region by ~4-fold (from 32-fold to 9-fold above control; Figure 5B), consistent with a requirement for functional NER, whereby XPA recruits ERCC1-XPF to H-DNA. In contrast, enrichment of Rad27 (FEN1) remained constant in the presence or absence of Rad14 (XPA; Figure 5B). Taken together, these results support a role for NER in the recognition and processing of H-DNA *in vivo*, a mechanism distinct from that of FEN1.

FEN1 and NER Proteins Are Involved in DNA Replication-Dependent and Replication-Independent Pathways of H-DNA-Induced Mutagenesis In Vivo

A primary cellular function of FEN1 is to process Okazaki fragments during DNA replication, and as such, it is under tight cell-cycle regulation (Henneke et al., 2003). Because DNA replication plays an important role in DNA-structure-induced mutagenesis (Iyer and Wells, 1999; Kang et al., 1995; Miret et al., 1998; Pelletier et al., 2003; Trinh and Sinden, 1991; Wang and Vasquez, 2009), we assessed whether the effects of FEN1 and ERCC1-XPF on H-DNA-induced mutagenesis were dependent on DNA replication and whether the contrasting effects of FEN1 and ERCC1-XPF might be explained by their associations with replication. Of note, FEN1 was active in all assays; only the templates were replication incompetent in the replication-deficient systems used in this study. In both replication-proficient (Figure 5C) and replication-deficient (Figure 5D) systems, the H-DNA-forming sequence stimulated mutations over background levels. However, the effect of XPF deficiency in diminishing H-DNA-induced mutagenesis was more substantial in the replication-deficient systems than in the replication-proficient assays (Figures 5C and 5D). In contrast, the effect of FEN1 on H-DNA-induced mutagenesis was dependent on DNA replication, because FEN1 depletion in human cells increased mutations only on replicating templates, irrespective of the XPF status (Figures 5C and 5D). These results suggest that the role of FEN1 in H-DNA-induced mutagenesis is largely contingent upon active DNA replication.

DISCUSSION

The discovery that alternative non-B DNA conformations can stimulate genetic instability in the absence of exogenous DNA damage has changed our understanding of the mechanisms involved in genetic instability, evolution, and disease etiology. For example, long GAA repeats have the potential to adopt H-DNA and can stimulate repeat expansions and deletions in Friedreich’s ataxia patients, perhaps via slippage events during replication (Gacy et al., 1998). However, the H-DNA-forming sequence tested in this study is not a simple repeat, and we found that it stimulates the formation of DSBs and deletions in mammalian cells and in mice (Wang et al., 2008; Wang and Vasquez, 2004). Thus, it is important to explore the biological significance of H-DNA-forming sequences in human disease and the mechanisms of H-DNA-induced mutation.

H-DNA-Forming Sequences Are Enriched at Chromosomal Translocation Breakpoints in Human Cancer Genomes

We found that sequences with the potential to adopt H-DNA were significantly enriched at human cancer translocation breakpoints. Not only were more H-DNA-forming sequences identified within ±100 bp of the translocation breakpoints, but the potential H-DNA-forming sequences near the cancer breakpoints were longer than those from randomly selected control regions, suggesting that such sequences may form stable H-DNA structures surrounding the translocation breakpoints, implicating H-DNA as a cause of chromosome breakage in human cancers. Taken together, these data provide evidence of the biological significance of H-DNA-induced mutagenesis in human cancer etiology.

Replication-Related and Replication-Independent Mutations Induced by H-DNA

An appropriate H-DNA-forming sequence only warrants the potential to form an H-DNA conformation; to adopt a structure with a higher energy status than B-DNA, the formation of H-DNA requires negative supercoiling and unwinding of the DNA duplex. DNA replication unwinds DNA from histones or other binding proteins, opens the DNA helix, and leaves long single-stranded DNA regions on the lagging strand, which could promote the B- to H-DNA transition. In fact, DNA replication plays an important role in mutagenesis induced by repetitive elements, such as hairpin-forming triplet repeats (Kroutil and Kunkel, 1999; Lujan et al., 2012; Pelletier et al., 2003; Pollard et al., 2004). In this study, we detected H-DNA-induced genetic instability in both replication-proficient and replication-deficient systems, providing evidence for at least two distinct mechanisms of H-DNA-induced mutation in mammalian cells: a replication-dependent pathway and a replication-independent, structure-specific cleavage mechanism. We further demonstrated that the NER proteins, XPA, ERCC1-XPF, and XPG, and the DNA replication-related nuclease FEN1 were involved in H-DNA-induced mutagenesis. The NER proteins stimulated H-DNA-induced mutagenesis, whereas FEN1 was implicated in a mechanism that suppressed H-DNA-induced mutagenesis *in vivo*.

NER Proteins ERCC1-XPF and XPG Are Involved in Replication-Independent Mutagenesis Induced by H-DNA

The triple-stranded region in H-DNA shares some structural similarities to intermolecular triplexes formed by triplex-forming oligonucleotides, which are known to be mutagenic (Vasquez and Wilson, 1998), and the NER protein XPA is required for this mutagenic processing in mammalian cells (Vasquez et al., 2002). Thus, we hypothesized that H-DNA-induced distortions to the DNA helix may be recognized as “damaged DNA” by the NER machinery (Vasquez and Wang, 2013). In support of this notion, the NER damage/distortion recognition complexes XPA-RPA and XPC-RAD23B have been shown to bind intermolecular DNA triplexes with high affinity (Jain et al., 2008; Thoma et al., 2005; Vasquez et al., 2002). Once DNA damage is recognized and verified by the NER complexes, the ERCC1-XPF complex incises the dsDNA 5′ to the lesion, whereas XPG incises the dsDNA 3′ to the damage. Interestingly, in our study, XPF cleaved H-DNA at the loop between the two Hoogsteen hydrogen-bonded strands, an area that is specific to a triple-stranded structure, rather than the loop at the Watson-Crick hydrogen-bonded duplex (Figure

2H), suggesting that the cleavage is H-DNA structure specific. In our protein-H-DNA complex structural models, the ERCC1-XPF complex interacts with the H-loop, and the active residues and the metal ion site on XPF are located near the cleavage site, providing a possible mechanism for the protein-DNA interaction and cleavage (Figure 4A). Such H-DNA-specific endonuclease activity of ERCC1-XPF has not yet been reported. Consistent with our biochemical assays, deficiency in either XPF or XPA resulted in a substantial reduction in H-DNA-induced mutagenesis in human and yeast cells, particularly the deletion events (Figures 2B and 2D). XPF deficiency also dramatically altered the distribution of H-DNA-induced breakpoints in human cells (Figure 2E). Moreover, the effects of XPF deficiency in suppressing H-DNA-induced mutagenesis were more pronounced in the absence of DNA replication (Figure 5D), suggesting that XPF is recruited to and cleaves H-DNA structures independent of replication. Taken together, these results clearly reveal a role of the XPF protein in cleaving H-DNA, further supporting an NER-dependent "DNA damage" repair model (Jain et al., 2008; Vasquez and Wang, 2013; Wang and Vasquez, 2004). We posit that, in non-replicating DNA, H-DNA cleavage by ERCC1-XPF results in DNA breaks that are processed by error-prone microhomology-mediated end-joining mechanisms (Kha et al., 2010; Mao et al., 2008; Vasquez and Wang, 2013).

XPG incises the DNA 3′ to the damage during NER processing. In this study, we did not detect incisions near the H-DNA-specific loop between the two Hoogsteen hydrogen-bonded strands, as we observed with XPF. Rather, XPG cleaved the H-DNA substrate at a site similar to that of FEN1 (Figure 2I). In our structural models, hXPG interacts with H-DNA via basic residues on its electrostatic surface, and the metal ion active site between the thumb and forefinger and the highly conserved active residues α4 K84 and R91 bind to the homopyrimidine strand upstream of the triple-stranded region (Figure 4B) near the cleavage sites determined by the *in vitro* assay (Figure 2I). XPG cleavage may not be the "rate-limiting step" in H-DNA-induced mutation *in vivo*, because the absence of XPG did not have a significant effect on H-DNA-induced mutagenesis; however, the absence of XPG altered the mutation spectrum dramatically, implicating XPG in the formation of H-DNA-induced deletions (Table 1). It is plausible that XPG processes the repair intermediates generated by other enzymes (e.g., XPF).

FEN1 Is Involved in Replication-Dependent Genetic Instability at H-DNA-Forming Sequences

Based on the modeled structures, the interactions between hFEN1 and H-DNA suggest that the positively charged "palm and fingers" on the protein interact with the homopurine third strand of H-DNA, and the active site between the thumb and forefinger binds the cleavage sites on the homopyrimidine strand upstream to the triple-stranded region to release the flapped 5′ single-stranded region (Figure 4C), which is exactly at the incision site demonstrated in the cleavage assay (Figure 2G). Although the incision sites on the H-DNA template were similar with FEN1 and XPG, FEN1 deficiency resulted in a substantial increase rather than a decrease in H-DNA-induced mutagenesis in both yeast and human cells (Figure 3). Further, the impact of FEN1 on H-DNA-induced mutagenesis was dependent on DNA replication (Figures 5C and 5D). This is consistent with the well-established function of FEN1 in replication (removing the 5′ flap of RNA/DNA primers in

Okazaki fragments during DNA replication). H-DNA can stall replication forks *in vivo* (our unpublished data; Boyer et al., 2013), which can lead to the exposure of long single-stranded DNA templates and immature Okazaki fragments. The absence of FEN1 could leave these unstable intermediates unprocessed, resulting in genetic instability. Alternatively, an H-DNA structure similar to the substrate used in this study (Figure 2G) could form on a single-stranded DNA region during lagging-strand synthesis, and FEN1 deficiency could inhibit Okazaki fragment maturation and ligation. Nevertheless, these data support a model where FEN1 processing of H-DNA structures may assist in continuous DNA replication and thereby reduce replication-dependent genetic instability at these sequences. A recent report also suggested that a defect of the *Rad27* domain near the two Mg^{2+} ions and the active residues identified in this study led to expansion of GAA repeats, which can potentially form small loops or H-DNA structures (Tsutakawa et al., 2017).

In summary, we have identified (1) a DNA replication-independent model of DNA-structure-induced mutagenesis, where H-DNA is recognized as “damage” and cleaved by NER nucleases, resulting in strand breaks and subsequent deletions, and (2) a DNA replication-dependent model, where FEN1 suppresses H-DNA-induced mutagenesis as depicted in Figure 6. More gene products are likely involved in H-DNA-induced DNA breakage, repair, and mutagenesis in human cells, perhaps similar to large gene networks involved in mutagenic repair of DNA breaks described in prokaryotes (Al Mamun et al., 2012). Further work to identify additional candidate genes/pathways involved in H-DNA-induced mutagenesis is warranted, as it will provide critical information for understanding the mechanisms of DNA-structure-induced genetic instability in human diseases.

EXPERIMENTAL PROCEDURES

For detailed methods, please see Supplemental Experimental Procedures.

Bioinformatics Analyses

19,957 unique breakpoints of translocations in human cancer genomes were obtained from COSMIC at http://cancer.sanger.ac.uk/cosmic. TFRs were defined as two adjacent mirror symmetric purine or pyrimidine runs of ≥6 bases each separated by a loop size of 0–7 bases (Mirkin and Frank-Kamenetskii, 1994). TFRs identified within ±100 bp from the breakpoints (bins) were recorded.

YAC Fragility Assays

H-DNA or control sequences were inserted between the telomere seed G4T4 and *URA3* in the YAC in the donor yeast strain K213 (Callahan et al., 2003). To measure H-DNA-induced chromosome fragility, YACs were transferred to canavanine-resistant recipient cells from a yeast deletion library via Kar-cross (Callahan et al., 2003), and the *URA3* mutants were selected on plates with 5-fluoro-orotic acid (5-FOA). The mutation frequencies were calculated as the number of FOA-resistant (FOA^R) colonies divided by the number of total colonies.

H-DNA-Induced Mutagenesis Assays in Human Cells

The H-DNA-containing (pMexY) or B-DNA control (pCex) plasmids were transfected into HeLa cells, human XPF-proficient or -deficient cells (Wu et al., 2007), XPG-proficient or -deficient cells (Evans et al., 1997), or XPA-proficient and -deficient cells (Jones et al., 1992). To deplete FEN-1 in human cells, human FEN1 or non-targeting small interfering RNAs (siRNAs) were transfected 48 hr before mutation reporter transfection and also co-transfected with shuttle reporters. Shuttle vectors were recovered 48 hr after transfection and transformed into MBM7070 cells, and mutants were screened by blue/ white screening as described (Wang et al., 2009). To address the effect of replication on H-DNA-induced mutagenesis, plasmids with or without the SV40 replication origin (pMexY and pCex; pMexY-SV40 and pCex-SV40, respectively) were transfected into human cells and mutants were screened as described above.

LM-PCR Analysis of H-DNA-Induced DSBs

Transfected plasmids were isolated from human cells, treated with Pol I Klenow fragment, and ligated to the linkers. PCR-amplified products between the specific primer and the breakpoint (linker) were separated on 2% agarose gels, and fragments of interest were purified from the gels and sequenced to map the DSB sites as described (Wang and Vasquez, 2004; Wang et al., 2009).

ChIP Assays

The Simple Chip Enzymatic Chromatin IP Kit (Cell Signaling Technology, Santa Cruz, CA) was used for ChIP assays according to the manufacturer's recommendations and as described (Zhao et al., 2009). Recovered ChIP and input DNA were used for PCR amplification using T3 and T7 primers surrounding the H-DNA-forming or control sequences (Table S1).

Cleavage Assays

Formation of the various DNA structures was confirmed by native acryl-amide gel electrophoresis and/or circular dichroism (Del Mundo et al., 2017). Cell extracts were prepared from XPF-proficient or -deficient cells (Evans et al., 1997) using a NucBuster kit (EMD Millipore, Temecula, CA). For cleavage assays, 6×10^{-8} M radiolabeled H-DNA, duplex DNA, and flap substrates were incubated with 20 ng of purified human recombinant FEN1 protein (kindly provided by Dr. Binghui Shen, City of Hope), 48 ng of purified human recombinant XPG protein (kindly provided by Dr. Richard D. Wood, The University of Texas M.D. Anderson Cancer Center), or cell extract at 30°C for the times indicated in their specific reaction buffers (see Supplemental Experimental Procedures for details). Cleavage products were separated on 20% denaturing polyacrylamide gels and visualized using a Typhoon PhosphorImager.

Modeling of Protein-DNA Complexes

X-ray crystal structures of human proteins were obtained from the PDB (http://www.rcsb.org; https://doi.org/10.2210/pdb3q8l/pdb). The structures of the catalytic core of the human XPG protein, the ERCC1 central domain, and the nuclease domain of hXPF were

used for modeling. The H-DNA structure (based on the sequence used in this study) was assembled, and the final structure was refined by using 5-ns molecular dynamics simulations with AMBER force field and software (http://ambermd.org; Case et al., 2015). The H-DNA structure was solvated in a water box with physiological concentrations of counter ions in the molecular dynamics simulations.

Statistical Methods

The FOA^R frequencies in the YAC fragility assay and the frequencies of mutation in the *supF* reporter gene in mammalian cells were calculated by dividing the numbers of mutants (FOA^R or *supF*-deficient white colonies) by the numbers of total colonies. Student's t test or χ^2 test was used to determine significant differences between groups. p values of less than 0.05 were considered statistically significant. For each experiment, at least 30,000 total colonies were counted and were independently repeated at least 3 times.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Drs. Richard Wood and Binghui Shen for providing purified proteins and Dr. Rick A. Finch for comments. This work was supported by an NIH/NCI grant to K.M.V. (CA093729) and by a Robert A. Welch Foundation to C.Z. and P.R. (F-1691).

References

Al Mamun AA, Lombardo MJ, Shee C, Lisewski AM, Gonzalez C, Lin D, Nehring RB, Saint-Ruf C, Gibson JL, Frisch RL, et al. Identity and function of a large gene network underlying mutagenic repair of DNA breaks. Science. 2012; 338:1344–1348. [PubMed: 23224554]

Belotserkovskii BP, De Silva E, Tornaletti S, Wang G, Vasquez KM, Hanawalt PC. A triplex-forming sequence from the human c-MYC promoter interferes with DNA transcription. J Biol Chem. 2007; 282:32433–32441. [PubMed: 17785457]

Boyer AS, Grgurevic S, Cazaux C, Hoffmann JS. The human specialized DNA polymerases and non-B DNA: vital relationships to preserve genome integrity. J Mol Biol. 2013; 425:4767–4781. [PubMed: 24095858]

Callahan JL, Andrews KJ, Zakian VA, Freudenreich CH. Mutations in yeast replication proteins that increase CAG/CTG expansions also increase repeat fragility. Mol Cell Biol. 2003; 23:7849–7860. [PubMed: 14560028]

Case, DA., Berryman, JT., Betz, RM., Cerutti, DS., Cheatham, TE., III, Darden, TA., Duke, RE., Giese, TJ., Gohlke, H., Goetz, AW., et al. AMBER 2015. San Francisco: University of California; 2015.

Choi J, Majima T. Conformational changes of non-B DNA. Chem Soc Rev. 2011; 40:5893–5909. [PubMed: 21901191]

Del Mundo IMA, Zewail-Foote M, Kerwin SM, Vasquez KM. Alternative DNA structure formation in the mutagenic human c-MYC promoter. Nucleic Acids Res. 2017; 45:4929–4943. [PubMed: 28334873]

Dominguez C, Boelens R, Bonvin AM. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. J Am Chem Soc. 2003; 125:1731–1737. [PubMed: 12580598]

Evans E, Fellows J, Coffer A, Wood RD. Open complex formation around a lesion during nucleotide excision repair provides a structure for cleavage by human XPG protein. EMBO J. 1997; 16:625–638. [PubMed: 9034344]

Frank-Kamenetskii MD, Mirkin SM. Triplex DNA structures. Annu Rev Biochem. 1995; 64:65–95. [PubMed: 7574496]

Gacy AM, Goellner GM, Spiro C, Chen X, Gupta G, Bradbury EM, Dyer RB, Mikesell MJ, Yao JZ, Johnson AJ, et al. GAA instability in Friedreich's Ataxia shares a common, DNA-directed and intraallelic mechanism with other trinucleotide diseases. Mol Cell. 1998; 1:583–593. [PubMed: 9660942]

Henneke G, Koundrioukoff S, Hübscher U. Phosphorylation of human Fen1 by cyclin-dependent kinase modulates its role in replication fork regulation. Oncogene. 2003; 22:4301–4313. [PubMed: 12853968]

Hohl M, Dunand-Sauthier I, Staresincic L, Jaquier-Gubler P, Thorel F, Modesti M, Clarkson SG, Schärer OD. Domain swapping between FEN-1 and XPG defines regions in XPG that mediate nucleotide excision repair activity and substrate specificity. Nucleic Acids Res. 2007; 35:3053–3063. [PubMed: 17452369]

Iyer RR, Wells RD. Expansion and deletion of triplet repeat sequences in Escherichia coli occur on the leading strand of DNA replication. J Biol Chem. 1999; 274:3865–3877. [PubMed: 9920942]

Jain A, Wang G, Vasquez KM. DNA triple helices: biological consequences and therapeutic potential. Biochimie. 2008; 90:1117–1130. [PubMed: 18331847]

Jones CJ, Cleaver JE, Wood RD. Repair of damaged DNA by extracts from a xeroderma pigmentosum complementation group A revertant and expression of a protein absent in its parental cell line. Nucleic Acids Res. 1992; 20:991–995. [PubMed: 1549511]

Kang S, Jaworski A, Ohshima K, Wells RD. Expansion and deletion of CTG repeats from human disease genes are determined by the direction of replication in E. coli. Nat Genet. 1995; 10:213–218. [PubMed: 7663518]

Kha DT, Wang G, Natrajan N, Harrison L, Vasquez KM. Pathways for double-strand break repair in genetically unstable Z-DNA-forming sequences. J Mol Biol. 2010; 398:471–480. [PubMed: 20347845]

Kinniburgh AJ. A cis-acting transcription element of the c-myc gene can assume an H-DNA conformation. Nucleic Acids Res. 1989; 17:7771–7778. [PubMed: 2678005]

Kroutil LC, Kunkel TA. Deletion errors generated during replication of CAG repeats. Nucleic Acids Res. 1999; 27:3481–3486. [PubMed: 10446236]

Lujan SA, Williams JS, Pursell ZF, Abdulovic-Cui AA, Clark AB, Nick McElhinny SA, Kunkel TA. Mismatch repair balances leading and lagging strand DNA replication fidelity. PLoS Genet. 2012; 8:e1003016. [PubMed: 23071460]

Lyamichev VI, Mirkin SM, Frank-Kamenetskii MD. Structures of homopurine-homopyrimidine tract in superhelical DNA. J Biomol Struct Dyn. 1986; 3:667–669. [PubMed: 3271043]

Mao Z, Bozzella M, Seluanov A, Gorbunova V. DNA repair by nonhomologous end joining and homologous recombination during cell cycle in human cells. Cell Cycle. 2008; 7:2902–2906. [PubMed: 18769152]

Miret JJ, Pessoa-Brandão L, Lahue RS. Orientation-dependent and sequence-specific expansions of CTG/CAG trinucleotide repeats in Saccharomyces cerevisiae. Proc Natl Acad Sci USA. 1998; 95:12438–12443. [PubMed: 9770504]

Mirkin SM, Frank-Kamenetskii MD. H-DNA and related structures. Annu Rev Biophys Biomol Struct. 1994; 23:541–576. [PubMed: 7919793]

Mirkin SM, Lyamichev VI, Drushlyak KN, Dobrynin VN, Filippov SA, Frank-Kamenetskii MD. DNA H form requires a homopurine-homopyrimidine mirror repeat. Nature. 1987; 330:495–497. [PubMed: 2825028]

Pelletier R, Krasilnikova MM, Samadashwily GM, Lahue R, Mirkin SM. Replication and expansion of trinucleotide repeats in yeast. Mol Cell Biol. 2003; 23:1349–1357. [PubMed: 12556494]

Pestov DG, Dayn A, Yu Siyanova E, George DL, Mirkin SM. H-DNA and Z-DNA in the mouse c-Ki-ras promoter. Nucleic Acids Res. 1991; 19:6527–6532. [PubMed: 1754390]

Pollard LM, Sharma R, Gómez M, Shah S, Delatycki MB, Pianese L, Monticelli A, Keats BJ, Bidichandani SI. Replication-mediated instability of the GAA triplet repeat mutation in Friedreich ataxia. Nucleic Acids Res. 2004; 32:5962–5971. [PubMed: 15534367]

Popescu NC. Genetic alterations in cancer as a result of breakage at fragile sites. Cancer Lett. 2003; 192:1–17. [PubMed: 12637148]

Sengerová B, Tomlinson C, Atack JM, Williams R, Sayers JR, Williams NH, Grasby JA. Brønsted analysis and rate-limiting steps for the T5 flap endonuclease catalyzed hydrolysis of exonucleolytic substrates. Biochemistry. 2010; 49:8085–8093. [PubMed: 20698567]

Shen B, Nolan JP, Sklar LA, Park MS. Functional analysis of point mutations in human flap endonuclease-1 active site. Nucleic Acids Res. 1997; 25:3332–3338. [PubMed: 9241249]

Thoma BS, Wakasugi M, Christensen J, Reddy MC, Vasquez KM. Human XPC-hHR23B interacts with XPA-RPA in the recognition of triplex-directed psoralen DNA interstrand crosslinks. Nucleic Acids Res. 2005; 33:2993–3001. [PubMed: 15914671]

Trinh TQ, Sinden RR. Preferential DNA secondary structure mutagenesis in the lagging strand of replication in E. coli. Nature. 1991; 352:544–547. [PubMed: 1865910]

Tsutakawa SE, Classen S, Chapados BR, Arvai AS, Finger LD, Guenther G, Tomlinson CG, Thompson P, Sarker AH, Shen B, et al. Human flap endonuclease structures, DNA double-base flipping, and a unified understanding of the FEN1 superfamily. Cell. 2011; 145:198–211. [PubMed: 21496641]

Tsutakawa SE, Thompson MJ, Arvai AS, Neil AJ, Shaw SJ, Algasaier SI, Kim JC, Finger LD, Jardine E, Gotham VJB, et al. Phosphate steering by Flap Endonuclease 1 promotes 5′-flap specificity and incision to prevent genome instability. Nat Commun. 2017; 8:15855. [PubMed: 28653660]

van Zundert GCP, Rodrigues JPGLM, Trellet M, Schmitz C, Kastritis PL, Karaca E, Melquiond ASJ, van Dijk M, de Vries SJ, Bonvin AMJJ. The HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes. J Mol Biol. 2016; 428:720–725. [PubMed: 26410586]

Vasquez KM, Wang G. The yin and yang of repair mechanisms in DNA structure-induced genetic instability. Mutat Res. 2013; 743–744:118–131.

Vasquez KM, Wilson JH. Triplex-directed modification of genes and gene activity. Trends Biochem Sci. 1998; 23:4–9. [PubMed: 9478127]

Vasquez KM, Christensen J, Li L, Finch RA, Glazer PM. Human XPA and RPA DNA repair proteins participate in specific recognition of triplex-induced helical distortions. Proc Natl Acad Sci USA. 2002; 99:5848–5853. [PubMed: 11972036]

Wang G, Vasquez KM. Naturally occurring H-DNA-forming sequences are mutagenic in mammalian cells. Proc Natl Acad Sci USA. 2004; 101:13448–13453. [PubMed: 15342911]

Wang G, Vasquez KM. Non-B DNA structure-induced genetic instability. Mutat Res. 2006; 598:103–119. [PubMed: 16516932]

Wang G, Vasquez KM. Models for chromosomal replication-independent non-B DNA structure-induced genetic instability. Mol Carcinog. 2009; 48:286–298. [PubMed: 19123200]

Wang G, Vasquez KM. Impact of alternative DNA structures on DNA damage, DNA repair, and genetic instability. DNA Repair (Amst). 2014; 19:143–151. [PubMed: 24767258]

Wang G, Carbajal S, Vijg J, DiGiovanni J, Vasquez KM. DNA structure-induced genomic instability in vivo. J Natl Cancer Inst. 2008; 100:1815–1817. [PubMed: 19066276]

Wang G, Zhao J, Vasquez KM. Methods to determine DNA structural alterations and genetic instability. Methods. 2009; 48:54–62. [PubMed: 19245837]

Wang G, Gaddis S, Vasquez KM. Methods to detect replication-dependent and replication-independent DNA structure-induced genetic instability. Methods. 2013; 64:67–72. [PubMed: 23954565]

Wells RD, Dere R, Hebert ML, Napierala M, Son LS. Advances in mechanisms of genetic instability related to hereditary neurological diseases. Nucleic Acids Res. 2005; 33:3785–3798. [PubMed: 16006624]

Wu Y, Zacal NJ, Rainbow AJ, Zhu XD. XPF with mutations in its conserved nuclease domain is defective in DNA repair but functions in TRF2-mediated telomere shortening. DNA Repair (Amst). 2007; 6:157–166. [PubMed: 17055345]

Zhao J, Jain A, Iyer RR, Modrich PL, Vasquez KM. Mismatch repair and nucleotide excision repair proteins cooperate in the recognition of DNA interstrand crosslinks. Nucleic Acids Res. 2009; 37:4420–4429. [PubMed: 19468048]

Zhao J, Bacolla A, Wang G, Vasquez KM. Non-B DNA structure-induced genetic instability and evolution. Cell Mol Life Sci. 2010; 67:43–62. [PubMed: 19727556]

Highlights

- H-DNA-forming sequences are enriched at translocation breakpoints in human cancer
- FEN1 suppresses H-DNA-induced genetic instability during replication
- ERCC1-XPF and XPG are involved in replication-independent H-DNA-induced mutagenesis
- H-DNA is a substrate for ERCC1-XPF, XPG, and FEN1

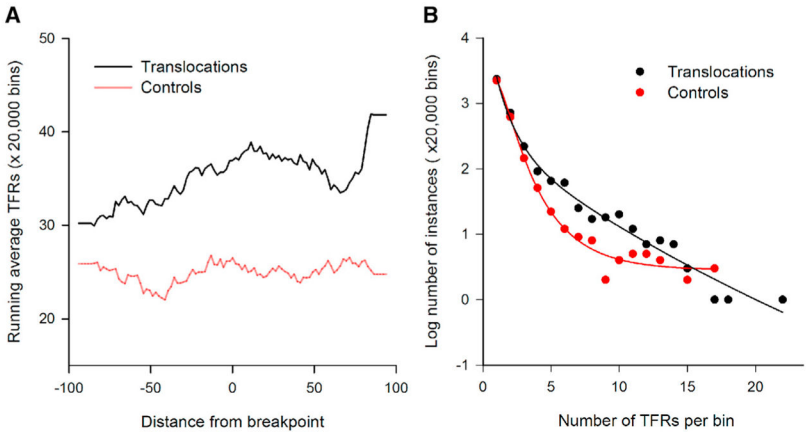


Figure 1. H-DNA TFRs Are Enriched at Translocation Breakpoints in Human Cancer Genomes

(A) Running average of triplex-forming repeats (TFRs) in the cancer translocation breakpoint (black) and control (red) datasets, whose center loop positions are located at each base along the ±100 bp flanking the cancer translocation breakpoints or control sites.

(B) Distribution of number of bins/20,000 bins each containing various numbers of TFRs in cancer translocation breakpoint (black) and control (red) datasets.

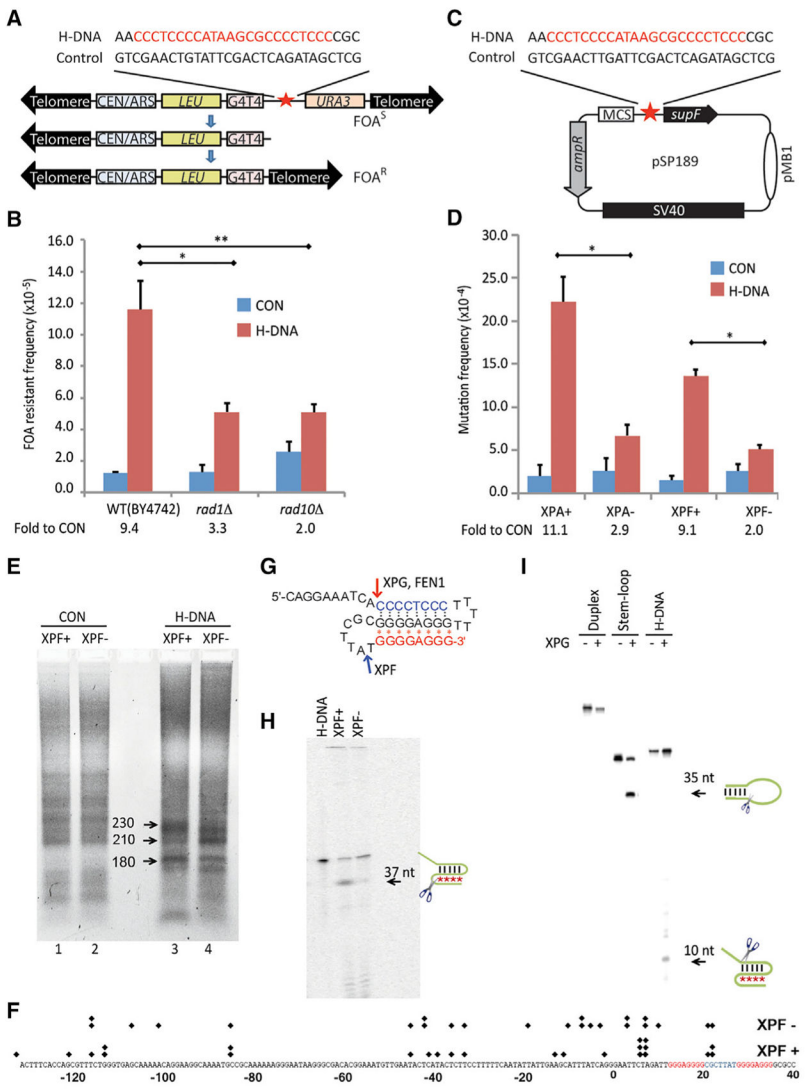


Figure 2. NER-Associated H-DNA-Induced Genetic Instability in Yeast and Human Cells

(A) Schematic structure of the YACs used for measuring H-DNA-induced genetic instability in yeast. Human H-DNA-forming or control B-DNA sequences are shown.

(B) Frequencies of H-DNA-induced fragility in YACs, as assessed by FOA resistance, in WT (BY4742), *rad1Δ*, or *rad10Δ* strains from the yeast single-gene deletion library GSA-5. The numbers below the graph show the fold increases in H-DNA-induced genetic instability above the control in each cell line (fold to CON). At least 30,000 colonies were screened in each assay. Error bars represent the SD from three repeats; two-sided t test; p values < 0.05 (*); <0.01 (**).

(C) Schematic of the *supF* mutation reporter containing human B-DNA- or H-DNA-forming sequences.

(D) H-DNA-induced mutation frequencies in NER-proficient (WT) and NER-deficient human cells. Fold to CON was calculated as described in (B). At least 30,000 colonies were screened in each assay. Error bars represent the SD from three repeats; two-sided t test; p values < 0.05 (*).

(E) LM-PCR detection of DSBs on plasmids recovered from XPF-proficient or -deficient human cells 24 hr after transfection. PCR products were separated on a 1.5% agarose gel. The 230-bp PCR fragment maps to a DSB hotspot within the H-DNA-forming sequence.
(F) PCR products were purified and sequenced to map the breakpoints. Each "◆" represents a breakpoint identified from the sequencing data.
(G) Schematic of the H-DNA structure of oligonucleotide MCR2–5′. Black dots represent Watson-Crick hydrogen bonds, and red "*" represent Hoogsteen hydrogen bonds. Cleavage sites of XPF, XPG, and FEN1 from (H), (I), and Figure 3C are indicated on the structure in the schematic.
(H) Cleavage of H-DNA in XPF-proficient and -deficient human cell extracts. The 37-nt cleavage product is marked by an arrow.
(I) H-DNA cleavage by purified human recombinant XPG protein. A stem-loop substrate serves as a positive control. The cleavage products of 35 nt from the stem-loop and 10 nt from the H-DNA substrate are marked by arrows.

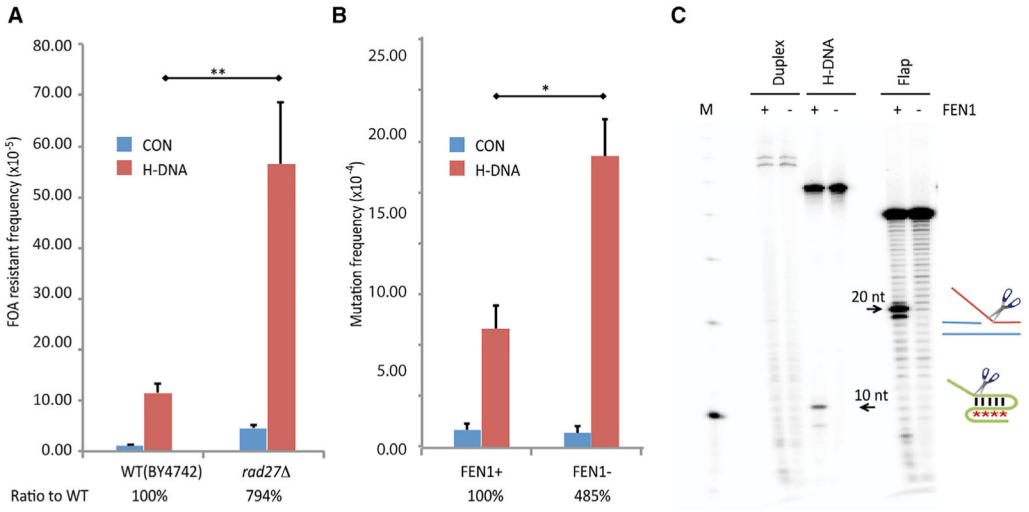


Figure 3. FEN1 Cleavage and Inhibition of H-DNA Structure-Induced Mutagenesis

(A) H-DNA-induced FOA resistance on YACs in WT (BY4247) or *rad27Δ* yeast strains. Ratio to WT: the relative H-DNA-induced genetic instability in repair-deficient cells (after subtracting the frequencies from the control B-DNA in the same cell line) above that in WT cells is shown. Error bars represent the SD from three repeats; two-sided t test; p values < 0.01 (**).

(B) H-DNA-induced mutation frequencies on *supF*-mutation reporters in WT or FEN1-depleted HeLa cells. Ratio to WT: the relative H-DNA-induced genetic instability in repair-deficient cells above that in WT cells, as in (A), is shown. Error bars represent SD from three repeats; two-sided t test; p values < 0.05 (*).

(C) Cleavage of H-DNA formed by oligonucleotide MCR2–5′ and a flap structure (Table S1) by purified human recombinant FEN1. The 20-nt cleavage product from the preferred substrate (a 5′ flap) and the 10-nt product from H-DNA are marked by arrows. M, 10-bp marker.

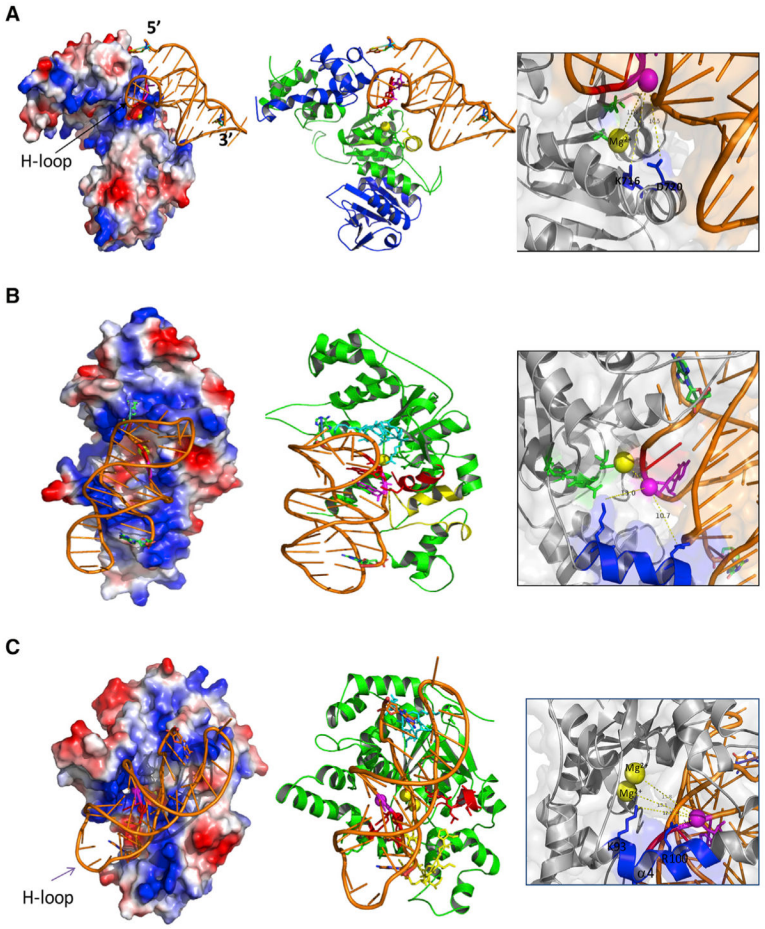


Figure 4. Structural Models of DNA Repair and Replication Proteins Interacting with H-DNA

(A–C) Modeled structures of an (A) hERCC1-XPF-H-DNA complex, (B) an hXPG-H-DNA complex, and (C) an hFEN1-H-DNA complex were generated by HADDOCK. Left: electrostatic surface presentation of the protein-H-DNA complex is shown. The positive and negative charged surfaces are colored in blue and red, respectively. Exposed hydrophobic residues are in white. Middle: cartoon presentation of the protein-H-DNA complex is shown. The protein is colored in green (ERCC1 in the hERCC1-XPF complex is in blue). The cleavage sites are colored in magenta. Right: detail of residues involved in cleavage sites of H-DNA is shown. The oxygen of the involved phosphodiester in H-DNA is depicted as a magenta sphere. The magnesium ion is colored in yellow. The active residues of proteins are shown as blue and green sticks. The distances between the cleavage sites and active sites are measured and shown as dotted yellow lines.

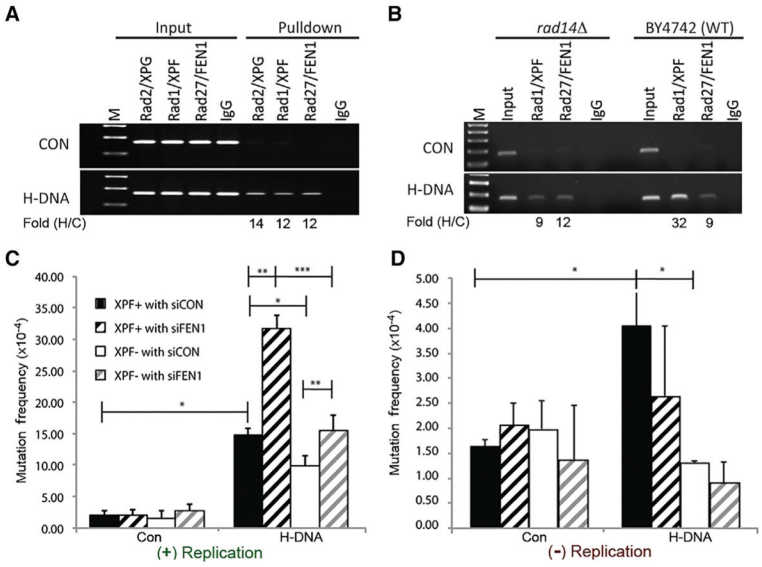


Figure 5. Association of Endonucleases with H-DNA and Effects of Replication on H-DNA-Induced Mutagenesis

(A) Yeast ChIP results demonstrate enrichment of Rad1 (XPF), Rad2 (XPG), and Rad27 (FEN1) at H-DNA relative to control B-DNA. The ChIP signal from the PCR assay of the H-DNA-forming sequence was normalized to the control sequence for each antibody used for the pull-down; Fold (H/C). M, 100-bp marker.

(B) ChIP results in *rad14Δ* (*XPA*) yeast demonstrate a requirement for Rad14 (XPA) in the enrichment of Rad1 (XPF), but not Rad27 (FEN1) on H-DNA.

(C and D) Effects of XPF and FEN1 on H-DNA-induced mutagenesis in replicating (C) and non-replicating systems (D). Control and H-DNA-containing mutation-reporter plasmids were transfected into human XPF-proficient (XPF+) or -deficient cells (XPF−) that were FEN1 depleted (siFEN1) or FEN1 proficient (siCON). Mutations were analyzed 48 hr after transfection. Error bars represent SD from three repeats; two-sided t test; p values < 0.05 (*); <0.01 (**).

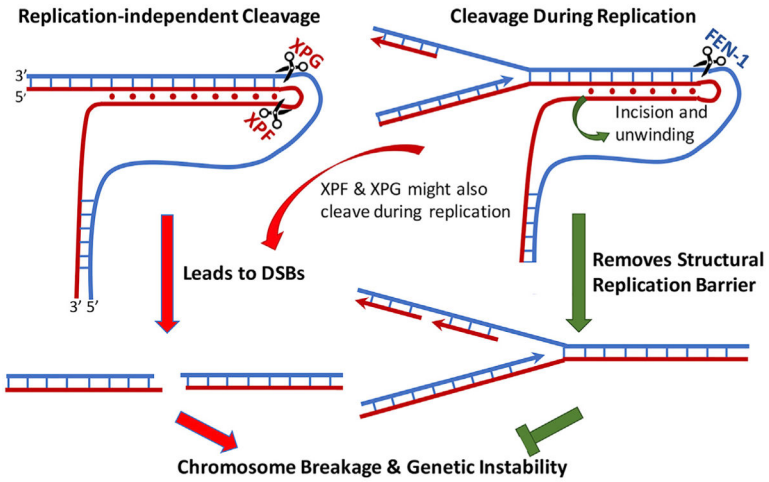


Figure 6. Model of Replication-Independent and Replication-Dependent H-DNA-Induced Genetic Instability in Eukaryotes

Left: schematic of a replication-independent mechanism of H-DNA-induced mutagenesis, where the alternative DNA conformation recruits DNA repair cleavage enzymes (e.g., ERCC1-XPF and XPG). The incision sites shown correspond to the cleavage sites of XPF and XPG in Figure 2G. Right: replication-related genetic instability where the H-DNA structure impedes the DNA replication machinery, which could lead to DSBs, is shown. XPF and XPG could also cleave H-DNA during replication, leading to DSBs and genetic instability. FEN1 can cleave H-DNA to resolve the mutagenic structure, thereby suppressing H-DNA-induced mutagenesis.

Table 1

Analysis of H-DNA-Induced and Spontaneous (B-DNA) Mutations in Human Cells

		H-DNA		B-DNA Control	
Cells	**Replication Status**	**Deletions**	**Point Mutations**	**Deletions**	**Point Mutations**
XPF+	+	77.8% (28/36)	22.2% (8/36)	38.9% (7/18)	61.1% (11/18)
XPF−	+	41.4%[a] (12/29)	58.6% (17/29)	53.8% (14/26)	46.2% (12/26)
XPG+	+	71.4% (15/21)	28.6% (6/21)	41.2% (7/17)	58.8% (10/17)
XPG−	+	33.3%[b] (7/21)	66.7% (14/21)	12.5% (2/16)	87.5% (14/16)
XPA+	+	71.4% (10/14)	28.6% (4/14)	11.1% (2/18)	88.9% (16/18)
XPA−	+	56.5% (13/23)	43.5% (10/23)	26.7% (4/15)	73.3% (11/15)
HeLa, siCON	+	77.8% (14/18)	22.2% (4/18)	66.7% (6/9)	33.3% (3/9)
HeLa, siFEN	+	100.0%[a] (43/43)	0.0% (0/43)	92.9% (13/14)	7.1% (1/14)
XPF+, siCON	−	15.0% (3/20)	85.0% (17/20)	0.0% (0/21)	100.0% (21/21)
XPF+, siCON	+	47.4% (9/19)	52.6% (10/19)	4.4% (1/23)	95.6% (22/23)
XPF−, siCON	−	12.5% (1/8)	87.5% (7/8)	19.0% (4/21)	81.0% (17/21)
XPF−, siCON	+	34.8% (8/23)	65.2% (15/23)	0.0% (0/19)	100.0% (19/19)
XPF+, siFEN	−	30.8% (4/13)	69.2% (9/13)	25.0% (2/18)	75.0% (16/18)
XPF+, siFEN	+	28.0% (7/25)	72.0% (18/25)	6.2% (1/16)	93.8% (17/16)
XPF−, siFEN	−	14.3% (1/7)	85.7% (6/7)	83.3% (10/12)	16.7% (2/12)
XPF−, siFEN	+	15.8% (3/19)	84.2% (16/19)	5.3% (1/19)	94.7% (18/19)

[a] p value < 0.01 in χ^2 test

[b] p value < 0.05