Cell Reports, Volume *22*

Supplemental Information

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

Junhua Zhao, Guliang Wang, Imee M. del Mundo, Jennifer A. McKinney, Xiuli Lu, Albino Bacolla, Stephen B. Boulware, Changsheng Zhang, Haihua Zhang, Pengyu Ren, Catherine H. Freudenreich, and Karen M. Vasquez

Supplemental Experimental Procedures:

Bioinformatics Analyses. The dataset containing the breakpoint locations of translocations and deletions in cancer genomes mapped to the human genome assembly GRCh37/hg19 was obtained from COSMIC at http://cancer.sanger.ac.uk/. The total number of unique breakpoints for translocations and deletions was 19,957 and 46,372, respectively. Triplex-forming repeats (TFRs) located within $\pm$100 bp from the breakpoints (bins), comprising two adjacent purines or pyrimidine runs of length $\geq$6 bases each, displaying mirror symmetry, and separated by a loop size of 0-7 bases, were identified using custom scripts. This constraint was set based on the observation of S1 nuclease sensitivity of such native sequences, suggesting their potential to adopt H-DNA conformations (Mirkin and Frank-Kamenetskii, 1994). For overlapping TFRs only the longest match was output. The bedtools utility was used to generate a set of 20,000 non-gap-matching sequences (bins), each 200 bases long.

Yeast Strains and YACs. Yeast strain 213 (MATa *Kar1-1, his7, leu2-3, 112, ura3-52*) was used for YAC construction and transfer into other yeast strains. Yeast strain BY4742 (*MATα, his3Δ1, leu2Δ0, lys2Δ0, ura3Δ0*) and their derivatives (Yeast deletion library GSA-5, ATCC, Manassas, VA) were used in mutation variation screening. H-DNA or control sequences were cloned between the telomere seed G4T4 and the *URA3* gene on a replication-defective plasmid derived from pRS306. Then the plasmids were linearized and used to construct YACs by homologous recombination with YAC VS5 containing a point mutation in the *URA3* gene (Callahan et al., 2003). H-DNA or control sequences and a functional *URA3* gene were selected on yeast minimal media SD base (Clontech, Mountain View, CA) with CSM-Ura (MP Biomedicals, Santa Ana, CA) plates

and confirmed by PCR with primers T7 and T3 (oligonucleotide sequences are listed in Table S1), followed by direct DNA sequencing.

YAC Transfer (kar-cross). Donor cells K213 containing YACs with human H-DNA or control sequences were grown from a single colony in SD base with CSM-Ura-Leu medium (MP Biomedicals) overnight at 30°C. Canavanine resistant recipient cells from the yeast deletion library were grown in yeast complete media YPD at the permissive temperature overnight. YACs were transferred from donor cells to recipient cells via kar-crosses as previously described in Callahan et al. 2003 (Callahan et al., 2003).

YAC Fragility Assays. Single yeast cells harboring YACs containing human H-DNA or control sequences were used to inoculate cultures and were grown for 20 h at the permissive temperature in SD base with CSM-Leu medium. 50 µl of each culture was plated on the SD base with CSM-Leu plates with 5-fluoroorotic acid (5-FOA, Zymo Research, Irvine, CA) to select for breakage events. 10 µl of each sample was diluted and plated on the SD base with CSM-Leu plates for a total cell number count. 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. HeLa cells, human XPF-proficient or XPF-deficient cells (Wu et al., 2007), and human XPG-proficient or XPG-deficient cells (Evans et al., 1997b) were maintained in Dulbecco's Modified Eagle Medium (DMEM, Life Technologies, Carlsbad, CA) with 10% fetal bovine serum (FBS) and antibiotics. Human XPA-proficient and XPA-deficient cells (Jones et al., 1992) were maintained in RPMI medium with 10% FBS and antibiotics. The H-DNA-containing (pMexY) or control (pCex) plasmids were transfected into human cells using

GenePORTER (Genlantis Inc., San Diego, CA) according to the manufacturer's recommendations. *E. coli* MBM7070 cells (F-*lacZ (am)CA7020, lacY1, hsdR−, hsdM+, araD139 Δ(araABC-leu)7679, galU, galK, rpsL, thi*) were used in the *supF*-based mutation frequency assays. Shuttle vectors carrying human H-DNA or control DNA sequences were extracted from human cells and transformed into MBM7070 cells mutation frequencies were assayed by blue/white screening as previously described (Wang et al., 2009). DNA from individual mutant colonies was sequenced to characterize the mutations/deletions near the H-DNA or control B-DNA regions.

siRNA Depletion in Cultured Human Cells. Human FEN1 or non-targeting siRNAs (Dharmacon, GE Healthcare Lafayette, CO) were transfected into cultured XPF-proficient or XPF-deficient human cells (Wu et al., 2007) with RNAimax (Invitrogen, Life Technologies Grand Island, NY) using the manufacturer's recommended protocol. A second siRNA transfection together with the reporter plasmids was carried out 48 h after the first transfection using GenePORTER (Genlantis Inc., San Diego, CA). Cells were harvested 48 h later for Western blotting to examine the FEN1 protein levels, and plasmids were extracted for mutation analysis.

Replication Effects on H-DNA-induced Mutation Frequency. The SV40 origin of replication in the plasmids carrying the human H-DNA-forming or control sequences was deleted by restriction digestion (using NcoI) and ligation. Plasmids with or without the SV40 replication origin (pMexY and pCex; pMexY-SV40 and pCex-SV40, respectively) were transfected into human cells using a Nucleofector (Lonza, Walkersville, MD). Cells were collected 24 or 48 hours after transfection for blue/white screening in *E. coli* MBM7070 cells as described above.

Ligation-mediated PCR (LM-PCR) Analysis of H-DNA-induced DSBs. Plasmids were recovered from human cells using a modified Hirt extraction method described previously (Wang and Vasquez, 2004; Wang et al., 2009). Briefly, extracted plasmids were treated with Pol I Klenow fragment to blunt the DNA ends, and ligated to the PCR Linkers (annealed from oligonucleotides LMPCR1 and LMPCR2). The ligation products were used as templates in the PCR reaction with a specific primer (LMsupF183) to the *supF*-reporter gene and a linker primer (LMPCR2) (Table S1). Amplified products were separated on a 2% agarose gel, and fragments of interest were purified from the gel and sequenced to map the DSB sites.

Chromatin Immunoprecipitation (ChIP) Assays. ChIP assays were performed using a Simple Chip Enzymatic Chromatin IP Kit (Cell Signaling, Inc., Santa Cruz, CA) according to the manufacturer's recommended protocol and as described previously (Zhao et al., 2009). Some modifications in protocols were specific to yeast cells. Yeast cells containing YACs with human H-DNA-forming or control sequences were grown overnight at a permissive temperature and 5 mL of culture was fixed with 1% formaldehyde for 8 min at room temperature to crosslink DNA and proteins. After quenching and washing, the YACs from the yeast cells were extracted using lysis buffer (0.1 M Tris-Cl pH 8.0, 50 mM ethylenediaminetetraacetic acid, 1% SDS) and Zymolyase (Zymo Research, Irvine, CA), followed by sonication to obtain an average DNA fragment length <500 bp. Approximately 1% of the total chromatin was stored as "input DNA" and the remaining chromatin was incubated with 5 μg of specific antibody (α-FEN1, α-XPG, α-XPF, Abcam, Cambridge, MA) or control α-IgG antibody and samples were incubated overnight at 4°C on a rotator. The rest of steps were performed according to the ChIP Kit protocols. Fractions of purified ChIP DNA and input DNA were used for PCR analysis.

The primers for PCR amplification were T3 and T7 surrounding the H-DNA-forming or control sequences (Table S1). Amplified products were separated on 1.5% agarose gels containing ethidium bromide and visualized on a ChemiDoc Imager (Bio-Rad, Hercules, CA).

Substrate Preparation. The H-DNA substrate was formed using a single-stranded oligonucleotide MCR2-5' (Table S1); the duplex DNA substrate was formed by annealing oligonucleotides 71 and 72 (Zhao et al., 2009); the FEN1 cleavage control substrates were formed by annealing oligonucleotides FENS1, FENS2, and FENS3 (Storici et al., 2002); the stem-loop substrate was formed by a single-stranded oligonucleotide SL46 (Evans et al., 1997b). The Y-splayed structure was formed by annealing Splayed-1 and Splayed-2 oligoes (Table S1). The H-DNA fold-back oligonucleotide MCR2-5' was incubated in triplex-forming buffer (20 mM NaCacodylate, 0.1 mM EDTA, 10 mM $MgCl_2$, 100 mM NaCl), and the other substrates were formed by annealing oligonucleotides in buffer (10 mM Tris, pH 8.0, 50 mM NaCl, 1 mM EDTA) boiled for 10 min, and slowly cooled to room temperature. Formation of the various structures was confirmed by native acrylamide gel electrophoresis and/or circular dichroism.

Cleavage Assays. $6x10^{-8}$ M radiolabeled H-DNA-forming fold-back oligonucleotides, 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) in reaction buffer (40 mM Tris-HCL pH 7.5, 10 mM $MgCl_2$, 5 mM DTT, 200 µg/mL BSA) (Storici et al., 2002) at 30°C for 15 min. Cleaved products were separated on a 20% denaturing polyacrylamide gel, and visualized using a Typhoon PhosphorImager (GE Healthcare Life Sciences, Pittsburgh, PA). Forty-eight ng of purified human recombinant XPG protein (kindly provided by Dr. Richard D. Wood, The University of Texas M.D.

Anderson Cancer Center) or BSA was incubated with $4x10^{8}$ M DNA substrates at 30°C overnight in a buffer containing 25 mM HEPES (pH 6.8), 4 mM $MnCl_2$, 1 mM DTT, 250 µg/mL BSA, and 10% glycerol. Formamide/dye was added to stop the reactions and the cleavage products were resolved on a 12% denaturing polyacrylamide gel. Cell extracts were prepared from XPF-proficient or XPF-deficient cells (Evans et al., 1997a) using a NucBuster kit (EMD Millipore, Temecula, CA). The H-DNA structure formed by oligonucleotide MCR2-5' (Table S1) was incubated with cell extract in reaction buffer (40 mM Tris-HCL pH 7.5, 10 mM $MgCl_2$, 5 mM DTT, 200 µg/mL BSA) at 30°C for 1 hour. Formamide/dye was added to stop the reactions and the cleavage products were resolved on a 12% denaturing polyacrylamide gel.

Modeling of Protein-DNA Complexes. The X-ray crystal structure of human FEN1 (hFEN1) was obtained from the Protein Data Bank (PDB, http://www.rcsb.org, DOI: 10.2210/pdb3q8l/pdb). The structure of the catalytic core of the human XPG protein (hXPG 1-155 and 762-984) was modeled through SwissModel (http://swissmodel.expasy.org/) using the catalytic core of its homologous Rad2p as a template (http://www.rcsb.org, DOI: 10.2210/pdb4q0z/pdb). The model of the complex of ERCC1-XPF was made using the program Pymol (http://www.pymol.org). The model was constructed from the known structures of the human XPF-ERCC1 HhH domains (http://www.rcsb.org, DOI: 10.2210/pdb1z00/pdb), the ERCC1 central domain (http://www.rcsb.org, DOI: 10.2210/pdb2jpd/pdb) and the nuclease domain of hXPF modeled through SwissModel using the archaeal XPF homodimer bound to dsDNA as the template (http://www.rcsb.org, DOI: 10.2210/pdb2bgw/pdb).

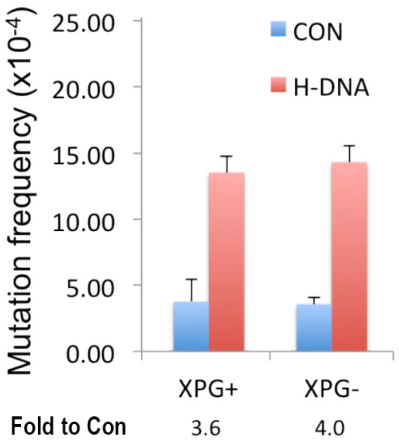
The H-DNA structure for the sequence used in the *in vitro* nuclease activity assay was assembled from 4 parts: the triple-stranded region, the H-loop, the turn between the two Watson-Crick paired duplex, and the 5’ terminal single strand DNA upstream to the

triple-stranded region. The H-DNA substrate was directly constructed from the solution structure of purine.purine.pyrimidine DNA triplet-repeat containing G.GC and T.AT triples (http://www.rcsb.org, DOI: 10.2210/pdb135d/pdb). For the TA-A triple structure, the TA was assigned as a canonical Watson Crick pair, and the A-A was assigned as a trans-Hoogsteen pair where both of the two As used the Hoogsteen edge, and the amino group of one A hydrogen bonds to the N7 of the other A (Leontis et al., 2002). The H-loop was modeled similarly as that in the available model (http://www.rcsb.org, DOI: 10.2210/pdb135d/pdb, and DOI: 10.2210/pdb1b4y/pdb). The turn between the Watson-Crick paired duplex regions was modeled as an anticondon U-turn (Gutell et al., 2000). The final structure was then refined by using 5-nanosecond molecular dynamics simulations with AMBER force field and software (http://ambermd.org) (D.A. Case, 2015). In the molecular dynamics simulations, the H-DNA structure was solvated in a water box with physiological concentrations of counter ions.

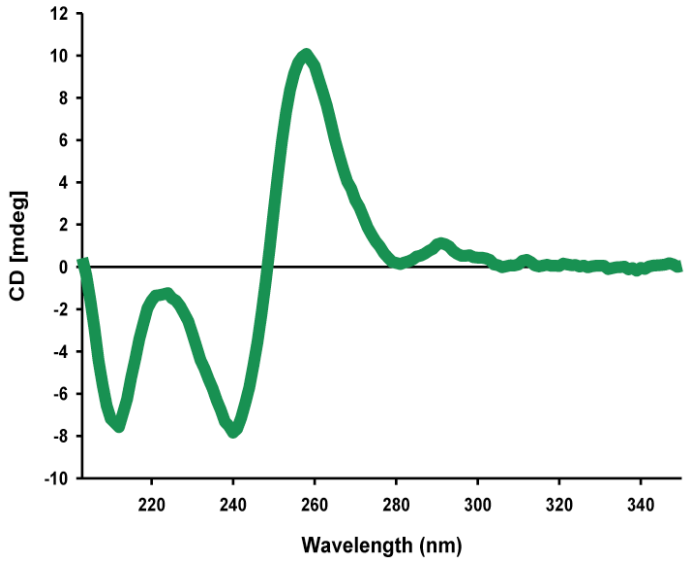
Supplementary Table 1. List of oligonucleotides used in this study, Related to Experimental Procedures

Name	Sequence
MCR2-5’	5’-cag gaa atc acc cct ccc ttt ttg gga ggg gcg ctt atg ggg agg g
71	5’-ggt acc gaa ttt cgg ccg agg ggg agg ggg tgg tgg ggg ggg aag gat tcg aac ctt
72	5’-aag gtt cga atc ctt ccc ccc cca cca ccc cct ccc cct cgg ccg aaa ttc ggt acc
FENS1	5’-gtc atg ata gat ctg atc gct cga att cct gca gcc cgg
FENS2	5’-ccg ggc tgc agg aat tcg ata tca agc tta tcg ata ccg tcg acc tcg a
FENS3	5’-tcg agg tcg acg gta tcg ata agc ttg ata
T7	5’-taa tac gac tca cta tag gg
T3	5’-att aac cct cac taa agg ga
LMPCR1	5’-cgt aca ttc aca acg ata gcg act ga
LMPCR2	5’-gct atc gtt gtg aat gta cg
LMsupF183	5’-aga tcc agt tcg atg taa cc
Splayed-1	5'-cgc tca agt tta tat taa aaa gca gag tt-3'
Splayed-2	3’-gcg agt tca aat ata atg ata tcg acc gt-5’

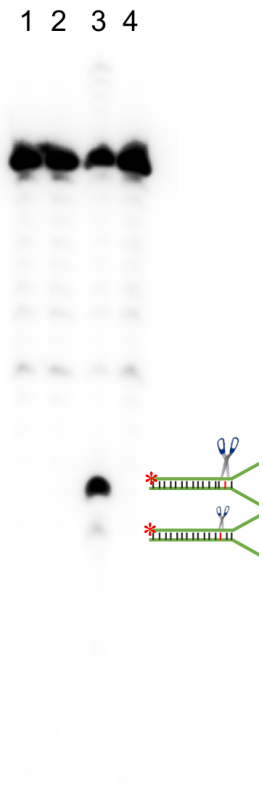
Supplementary Data:



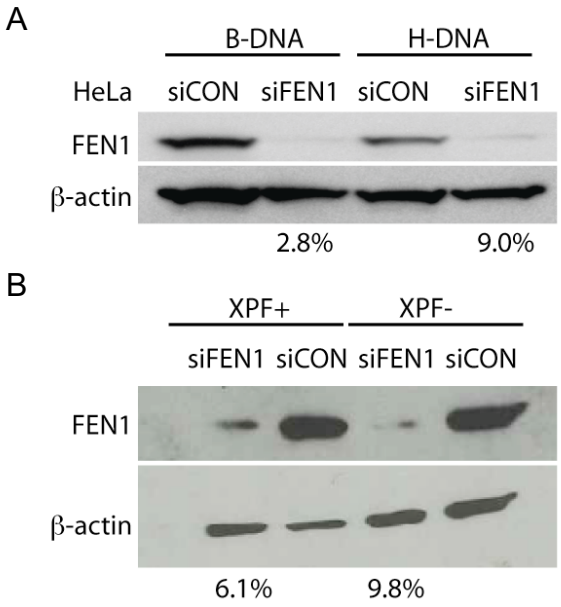
Supplementary Figure 1. The NER protein XPG does not affect H-DNA-induced mutation frequency in human Cells. Related to Figure 2 and Table 1. H-DNA-induced mutation frequencies in XPG-proficient (XPG+) and XPG-deficient (XPG-) human cells. The numbers below the graph show the fold increases in H-DNA-induced genetic instability above the B-DNA control in each cell line (Fold to CON). At least 10,000 colonies were screened in each of three assays.



Supplementary Figure 2. Circular dichroism (CD) spectrum of the H-DNA fold-back oligonucleotide MCR2-5'. Related to Figure 2. Blank and sample-containing quartz cuvettes were matched prior to use. The CD spectrum of 10 mM MCR2-5' was measured at pH 7.0 in 20 mM HEPES, 100 mM NaCl, 0.1 mM EDTA and 10 mM $MgCl_2$ at 25°C. Samples were evaluated on a Jasco J-815 CD Spectrometer (JASCO, Inc., Easton, MD) equipped with a Peltier temperature controller and scanned from 200 – 350 nm. Data were processed using the JASCO software and exported to Sigma Plot 10. The negative peak at 210-220 nm represents the formation of an H-DNA structure.



Supplementary Figure 3. Cleavage of Y-splayed structure by XPF in human cell-free extracts. Related to Figure 2. 5'-radiolabeled ssDNA (**lanes 1&2**) or Y-shape splayed substrate radiolabeled on the 5'-end of the top strand (**lanes 3&4**), were subjected to cleavage assays in human cell free extracts. **Lanes 1&3**: XPF-proficient cell extract; **lanes 2&4**: XPF-deficient cell extract. The 16-nt and 15-nt cleavage products are marked, suggesting cleavage sites -1 and -2 nt to the junction (shown in red lines on the right).



Supplementary Figure 4. siRNA-mediated depletion of FEN1 in human cells. Related to Figure 3. **(**A) Detection of FEN1 protein in HeLa cells treated with FEN1 siRNA (siFEN) or non-targeting control siRNA (siCON) by immunoblotting 48 hours post-treatment. (B) Detection of FEN1 protein in XPF-proficient or deficient human cells with FEN1 siRNA (siFEN) or non-targeting control siRNA (siCON) by immunoblotting 48 hours post-treatment. The amount of FEN1 expression was quantified using ImageJ. The percentages shown below the gel were compared to the expression levels in control knockdown samples.

Supplementary References:

Callahan, J.L., Andrews, K.J., Zakian, V.A., and Freudenreich, C.H. (2003). Mutations in yeast replication proteins that increase CAG/CTG expansions also increase repeat fragility. Mol Cell Biol *23*, 7849-7860.

D.A. Case, J.T.B., R.M. Betz, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, T. Luchko, R. Luo, B. Madej, K.M. Merz, G. Monard, P. Needham, H. Nguyen, H.T. Nguyen, I. Omelyan, A. Onufriev, D.R. Roe, A. Roitberg, R. Salomon-Ferrer, C.L. Simmerling, W. Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, D.M. York and P.A. Kollman (2015). AMBER 2015. University of California, San Francisco.

Evans, E., Fellows, J., Coffer, A., and Wood, R.D. (1997a). Open complex formation around a lesion during nucleotide excision repair provides a structure for cleavage by human XPG protein. The EMBO journal *16*, 625-638.

Evans, E., Moggs, J.G., Hwang, J.R., Egly, J.M., and Wood, R.D. (1997b). Mechanism of open complex and dual incision formation by human nucleotide excision repair factors. The EMBO journal *16*, 6559-6573.

Gutell, R.R., Cannone, J.J., Konings, D., and Gautheret, D. (2000). Predicting U-turns in ribosomal RNA with comparative sequence analysis. J Mol Biol *300*, 791-803.

Jones, C.J., Cleaver, J.E., and Wood, R.D. (1992). 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 research *20*, 991-995.

Leontis, N.B., Stombaugh, J., and Westhof, E. (2002). The non-Watson-Crick base pairs and their associated isostericity matrices. Nucleic Acids Res *30*, 3497-3531.

Mirkin, S.M., and Frank-Kamenetskii, M.D. (1994). H-DNA and related structures. Annu Rev Biophys Biomol Struct *23*, 541-576.

Storici, F., Henneke, G., Ferrari, E., Gordenin, D.A., Hubscher, U., and Resnick, M.A. (2002). The flexible loop of human FEN1 endonuclease is required for flap cleavage during DNA replication and repair. The EMBO journal *21*, 5930-5942.

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

Wang, G., Zhao, J., and Vasquez, K.M. (2009). Methods to determine DNA structural alterations and genetic instability. Methods *48*, 54-62.

Wu, Y., Zacal, N.J., Rainbow, A.J., and Zhu, X.D. (2007). XPF with mutations in its conserved nuclease domain is defective in DNA repair but functions in TRF2-mediated telomere shortening. DNA Repair (Amst) *6*, 157-166.

Zhao, J., Jain, A., Iyer, R.R., Modrich, P.L., and Vasquez, K.M. (2009). Mismatch repair and nucleotide excision repair proteins cooperate in the recognition of DNA interstrand crosslinks. Nucleic acids research *37*, 4420-4429.