Science Advances

advances.sciencemag.org/cgi/content/full/2/11/e1601605/DC1

NAAAS

Supplementary Materials for

Holliday junction trap shows how cells use recombination and a junction-guardian role of RecQ helicase

Jun Xia, Li-Tzu Chen, Qian Mei, Chien-Hui Ma, Jennifer A. Halliday, Hsin-Yu Lin, David Magnan, John P. Pribis, Devon M. Fitzgerald, Holly M. Hamilton, Megan Richters, Ralf B. Nehring, Xi Shen, Lei Li, David Bates, P. J. Hastings, Christophe Herman, Makkuni Jayaram, Susan M. Rosenberg

> Published 18 November 2016, *Sci. Adv.* **2**, e1601605 (2016) DOI: 10.1126/sciadv.1601605

The PDF file includes:

- text S1. DNA repair by HR and RuvC specificity for four-way junctions.
- text S2. About half of HJs are detected as RDG foci in living *E. coli* cells, and half of Gam-detectable DSBs result in HJ foci.
- fig. S1. Design of a catalytically inactive RuvC for trapping HJs: RuvCDef.
- fig. S2. Purified RDG binds, does not cleave, and inhibits action of other proteins on synthetic HJs in solution.
- fig. S3. RDG inhibition of resistance to UV light requires induction of transcription of the *ruv*CDef*gfp* gene.
- fig. S4. Titration of RuvC with RDG levels shows minimum RDG/RuvC ratio at which RDG outcompetes RuvC, preventing repair in living cells.
- fig. S5. Spontaneous RDM foci colocalize with RecA-GFP strand exchange protein in *E. coli* cells.
- fig. S6. RDG ChIP-seq localization requires DSBs and specific RuvC antibody.
- fig. S7. Spontaneous RuvCGFP and RDG foci overlap with DNA stain.
- fig. S8. Spontaneous RDG HR/HJ foci per cell correlate with varying chromosome and replication fork numbers.
- fig. S9. Similar growth rates of various mutant strains used.
- fig. S10. Reduced spontaneous RDG/HJ focus levels in *recF*, *recQ*, and *recJ* cells are restored by supplying RecF, RecQ, and RecJ from plasmids.
- fig. S11. RecA overproduction is induced after RDG accumulation in cells.
- fig. S12. Increased *EME1* and *GEN1* HJ resolvase mRNAs in *BLM*overexpressing human cancers of the eight, and six of the eight, most common cancer types, respectively.
- fig. S13. Validation of Mu Gam protein function.
- table S2. *E. coli* K12 strains and plasmids used in this study.
- table S3. Names and locations of new I-sites and alleles.
- table S4. Oligonucleotides used in this study.
- References (*100*–*122*)

Other Supplementary Material for this manuscript includes the following:

(available at advances.sciencemag.org/cgi/content/full/2/11/e1601605/DC1)

 table S1 (Microsoft Excel format). Pairwise mRNA level correlation between human RAD51 with human HJ resolvases and RecQ orthologs in the eight most common cancer types.

text S1. DNA repair by HR and RuvC specificity for four-way junctions

HR is ubiquitous and conserved across the tree of life (*100*). DNA damage that leads to discontinuities such as double-stranded breaks (DSBs) and single-stranded gaps are mended by aligning one strand of discontinuous DNA by basepairing with continuous, complementary ("homologous") DNA elsewhere in the genome (e.g., Fig. 1Aiii). The homologous sequence templates any repair synthesis needed (e.g., Fig. 1Aiv, dashed lines) and, in some HR mechanisms, becomes covalently joined with the previously discontinuous sequence to reconstruct a correct, intact chromosome (Fig. 1Av, solid blue-green junctions). The homologous sequence is commonly a sister chromosome (*6*) but can be any region of sequence identity in the genome. "Strand-exchange" proteins RecA in bacteria, and its orthologs (including RAD51) in eukaryotic and archaeal cells perform the alignment (Fig. 1Aiii). In some HR reactions strand-exchange intermediates progress to a transient four-way DNA junction called a Holliday junction (HJ) (*8*) (Fig. 1Aiv the strands under, above and below each hexagon, Fig. 1B), which must be "resolved", by any of several mechanisms, to re-create functional duplex DNA chromosomes (e.g., endonucleolytic resolution, Fig. 1Aiv-v).

E. coli RuvC is a highly four-way DNA junction-specific binding protein (*30-32*). Fogg et al. compared the dissociation constant (Kd) for RuvC interaction with three-way and four-way DNA junctions. They found that RuvC bound four-way DNA junctions 10-times better than it bound threeway DNA junctions, and 10^3 -10⁴-times better than linear duplex DNA (101). In solution, RuvC alone can bind HJs independently of RuvA (*44*), the stationary anchor of the RuvAB motor responsible for directional branch-migration of four-way junctions.

text S2. About half of HJs are detected as RDG foci in living *E. coli* **cells, and half of Gamdetectable DSBs result in HJ foci**

The ratio of RDG (HJ) foci (Fig. 2D) to GamGFP (DSB) foci (Fig. 2B) is 0.83 and 0.59 *ori*-proximally and –distally, respectively, implying that 83% and 59% of Gam-detectable DSBs become HJs at those sites. Because GamGFP detects about 82% of DSBs (*39*), this implies that roughly 68% of the *ori*proximal and 48% of the *ori-*distal DSBs lead to RDG foci. From these and data with gamma-rayinduced DSBs, we estimate the efficiency of detection of HJs as \geq roughly 50% of HJs detected as RDG foci as follows.

In a dose range in which gamma rays induce 0.014 DSBs/*E. coli* cell/Gy (*102*), we observed 0.005 RDG foci/cell/Gy. The data imply that about 36% (0.005/0.014=36%) of gamma-induced DSBs result in RDG foci (Fig. 2F). This is roughly similar to the estimated 68% and 48% efficiency of HJs detected as RDG foci per *ori-*proximal and -distal I-*Sce*I-induced DSBs, respectively, above and Fig. 2A-E. These are *minimum* estimates of HJ-detection efficiency because only repairing DSBs produce HJs, and many gamma- and I-*Sce*I-induced DSBs may not be repaired to the HJ stage given that both treatments cause some cell death [e.g., ~2% survival at 20 Gy (*102*), used Fig. 2F]. Real HJ-detection efficiency as RDG foci is likely to be higher. These estimates could be two-fold higher if not all DSBs lead to HJs, or lower, if the HJs detected as foci were mostly double rather than single HJs (double HJs illustrated Fig. 1Aiv). However, lower detection is unlikely because single HJs are predicted for most DSB repair in *E. coli* (*103*), and multiple foci per cell are observed (Fig. 2C,D), indicating that all repair HJs do not group into a single focus. The data suggest that RDG detects at least half (\geq 36%-68%) of HJs formed in cells.

fig. S1. Design of a catalytically inactive RuvC for trapping HJs: RuvCDef. Superimposed 3D structures of the active sites of mutant bacteriophage bIL67 RuvC (cyan) and *E. coli* RuvC (green) are shown. Catalytic amino acids D7 and E66 of *E. coli* RuvC are altered to N7 and D66 respectively to create a catalytically inactive protein: RuvCDef, which recapitulates inactivating mutations demonstrated for bIL67 RuvC (*104, 105*). Dashed lines, distance between two atoms (Å).

fig. S2. Purified RDG binds, does not cleave, and inhibits action of other proteins on synthetic HJs in solution. Representative images of agarose gel electrophoresis of the mixtures with DNA visualized with ethidium bromide shown (A-F). For better clarity of the protein-bound DNA bands, inverse-contrast images are shown in E and F. (**A**) RDG and RuvCGFP bind synthetic HJb in solution. HJb contains the 13bp recognition sequence for Flp site-specific recombinase (*36*). Molar ratios of protein to junction were 1.25, 2.5 and 5 (lanes 2-4 and 5-7). (**B**) HJc containing the *Eco*RI recognition sequence is bound completely by RDG or RuvCGFP at a junction to protein molar ratio of 1:10. 10 min incubation on ice, at 23°C or at 37°C saturated binding. (**C**) RuvCGFP, but not RuvCDefGFP (RDG), cleaves synthetic immobile HJ, HJa, in assays per (*104, 105*). (**D**) RDG and RuvCGFP do not inhibit *Eco*RI cleavage of linear duplex DNA containing the *Eco*RI site. The data imply that RDG and RuvCGFP inhibition of *Eco*RI cleavage of HJc (Fig. 1E) reflects binding of the HJ structure, and not a general inhibition of *Eco*RI activity. **Left**, linearized DNA substrate design and representative gel images of *Eco*RI digestion unaffected by RuvCGFP or RDG. **Right**, DNA band intensities normalized to time zero of *Eco*RI treatment, mean (± SEM) of 3 experiments. (**E**) RDG binds and inhibits action of Flp recombinase at FRT sites near a junction center. Two pmoles of the Holliday junction HJb, containing the Flp recognition sequence, were pre-incubated with 20 pmoles of RDG to establish

complete junction binding (lane 2). The bound complex was challenged with 8 and 16 pmoles of Flp, sufficient for ~70% and complete conversion of the junction into the Flp-bound form, respectively, in the absence of RDG (lanes 3 and 4). For competition with 8 pmoles Flp (lane 5), the fraction of HJ-Ruv complex not competed by Flp was $55 \pm 2\%$. With 16 pmoles Flp (lane 6), this fraction was $45 \pm$ 2%. (**F**) For revealing protein-associated DNA bands, a competition assay, performed as in (E) with 16 pmoles of Flp as competitor, was complemented by western blotting (shown G, H). According to quantification of the ethidium bromide-stained DNA bands, ~48% of the RDG-bound junction was refractory to competition by Flp (lane 4). (**G**) DNA-bound proteins from the gel shown in (F) were transferred to a PVDF membrane, which was probed using antibodies (Ab) to GFP. (**H**) The same membrane as in (G) was re-probed 48 h later using Ab to Flp. These data demonstrate the presence of RDG in HJb-RDG complex bands (F,G) and of Flp in the HJb-Flp bands (F, H).

fig. S3. RDG inhibition of resistance to UV light requires induction of transcription of the *ruvC***Def***gfp* **gene.** Negative control for Fig. 1F. Strain designations show native *ruvC* locus, either *ruvC*⁺ or deleted ($\triangle ruvC$) left, and the protein produced from the chromosomal transgene under the control of the PN25*tetO* promoter, right. PN25*tetO*-RDG and PN25*tetO*-RuvCGFP, transgenic chromosomal inducible expression cassettes shown in Fig. 1C and not illustrated, respectively; P_{N25tetO}, chromosomal P_{N25tetO} promoter only. The data show that, as expected, there is no effect of RDG—no dominantnegative UV sensitivity in the *ruvC⁺* strain background—when the *ruvC*Def*gfp* gene is transcriptionally silent (uninduced, no doxycycline, here).

A Titrated IPTG induction of RuvC in doxy-induced RDG-producing cells

fig. S4. Titration of RuvC with RDG levels shows minimum RDG/RuvC ratio at which RDG outcompetes RuvC, preventing repair in living cells*.* (**A**) Production of varying amounts of IPTGinduced RuvC in doxycycline-induced RDG-producing cells and their resulting sensitivity to UV light. 100 ng/ml doxycycline and different doses of IPTG $(0, 10^{-3}, 10^{-1} \text{ mM})$ were added to induce RDG and wild-type RuvC proteins, respectively. Cells were treated with UV doses indicated. Control ∆*ruvC* P*tac* cells are sensitive (all panels); *ruvC*⁺ P*tac* control cells are resistant (all panels), and control ∆*ruvC* P*tac*RuvC cells are also resistant in all panels from leaky expression from the functional P*tac*RuvC construct (left panel, No IPTG, shown in western blot in B) and IPTG-induced production of RuvC (middle and right panels). Error bars, SEM, mean of 3 experiments. (**B**) RDG inhibits RuvC action in living cells when their ratios predict that RDG homodimers are only half as numerous as RuvC homodimers, but not when RuvC homodimers are expected to exceed RDG homodimers by 7.8 ± 0.8 times (mean \pm SEM 3 experiments with 50-58% RuvC versus to 6-8% RDG homodimers). Representative western blot shows levels of RDG and RuvC monomer proteins at the IPTG levels used (above) using anti-RuvC antibody. Because proteins are denatured in western blots, no dimers

(homo or hetero) are seen. The table shows the expected ratios of RuvC to RDG homodimers and heterodimers predicted from the protein levels observed, assuming no bias in dimerization of either protein with itself or the other species. The ratios of RuvC to RDG monomeric units observed were 0.16, 1.53 and 2.5 in the cells with 0, 10^{-3} and 10^{-1} mM IPTG, respectively in the representative experiment shown (one of three performed). (**C**) Western blot showing that with the expression protocol used in most experiments in this work, the levels of RDG induced from the chromosomal transgene (RDG) are about 50-times greater than those of RuvC, produced from the native $ruvC^+$ gene (RuvC non-detectable), such that RDG is expected to protect essentially all HJs from RuvC action (per A, and B).

fig. S5. Spontaneous RDM foci colocalize with RecA-GFP strand exchange protein in *E. coli* **cells**. We moved RDM into a strain carrying the *recAo1403,4136,4155*-*gfp* allele (*46*), which produces a GFP-tagged hypomorphic-mutant version of RecA from the native locus. The mutant RecA has decreased non-DNA foci/filaments and reduced HR activity (*46*). We see significant overlap between spontaneous RDM and RecA-GFP foci (blue arrows; white arrows non-overlapping RDM or RecA-GFP foci). (**A**) Representative images. (**B**) RDM forms foci that overlap significantly with foci of a RecA-GFP fusion protein ($P = 0.002$, two tailed paired *t*-test). 76-83% of RDM spontaneous foci overlap with RecA-GFP foci; the remaining 17-24% are not overlapping (range for 2 experiments). Because foci of proteins bound to specific DNA sites overlap at 10kb and can be distinguished at 13- 55 kb (*39, 47*), the overlap here puts these proteins meaningfully in the same vicinity in the 4.6MB *E. coli* genome. Plots, range of 2 experiments, >600 cells with >50 foci of each kind scored. The absolute frequencies were 0.046-0.056 solo RecA-GFP foci/cell, 0.013-0.022 solo RDM foci/cell, and 0.063- 0.068 RecA-GFP co-localized foci/cell. By contrast, production of the GFP-alone control caused rare green foci $[0.0005$ foci/ cell, similar to a previous report (39)], and RDM (red) foci at 0.038 ± 0.004 per cell (mean \pm range, two experiments), with co-localization of these at only 1.0% \pm 0.2% (mean \pm range). The data imply that RDM foci formed in regions with DNA damage/DNA repair, supporting their representing HJs. The data also show a small increase in RDM foci when RecA-GFP is produced, indicating that this partial-functional RecA fusion (*46*) may prolong repair duration and/or slightly increase DNA damage compared with the wild-type RecA in the GFP-alone control

fig S6. RDG ChIP-seq localization requires DSBs and specific RuvC antibody. (**A**) Negative controls for Fig. 2G. (**i**) RDG enrichment at cleaved I-site L. The negative controls show no I-sitespecific enrichment of RDG—(**ii**) with cutsite present without I-*Sce*I enzyme (DSB-); (**iii**) with nonspecific antibody IgG used in the DSB-producing strain; and (**iv**) with non-specific antibody IgG in the DSB-strain carries I-site L but no I-*Sce*I enzyme. RDG Chip-seq reads are normalized to sequencing depth (median read number) in each sample. (**B**) Negative controls for Fig. 2H. (**i**) RDG accumulates near I-Site J DSB-dependently, in cells that produce I-*Sce*I. (**ii-vi**) RDG ChIP-seq data from strains isogenic to those in Fig. 2H, (ii-vi) DSB-controls with I-*Sce*I enzyme-only (no cutsite) show no RDG accumulation near I-site J in any of the genetic backgrounds examined. RDG ChIP-seq reads were normalized to the median sequencing depth in each sample (relative ChIP-seq reads) and these values were further normalized to the relative genome input at each genomic location (whole-genome sequencing reads normalized to median sequencing depth).

fig. S7. Spontaneous RuvCGFP and RDG foci overlap with DNA stain. Top row: GFP produced alone produces ≤0.03% of cells with a focus (*39*). Middle and bottom rows: RuvCGFP and RDG form foci that overlap with DNA content (DAPI stain, arrows show foci)**.** Membranes are stained with redfluorescent FMTM 4-64FX (F34653, Invitrogen) as described (*106*) and imaged with the microscope far red channel.

fig. S8. Spontaneous RDG HR/HJ foci per cell correlate with varying chromosome and replication fork numbers under varied growth conditions. (**A**) RecA-dependence of most spontaneous RDG and RuvCGFP foci. More foci are observed with RDG than RuvCGFP as predicted by RDG HJ-trap ability in solution and in cells (figs. 1, S2, S4). (**B**) Spontaneous RDG foci occur mostly one per cell, and sometimes two or \geq three per cell. (C) RDG foci were reduced in minimal medium, compared with rich medium, which confer fewer and more chromosomes per cell, respectively, shown with (**D**) flow cytometric chromosome counting. Chromosome copy numbers are measured as DAPI DNA-stain intensities and flow cytometry after replication initiations are blocked, but elongations allowed to continue, producing full chromosomes from each replication bubble (pair of forks) in progress at the time of initiation block (*86*). This technique produces data that reflect numbers of forks in progress. (**E**) Replication-fork numbers derived from chromosome copy numbers in rich and minimal media. In rich medium, 7%, 86%, and 6% of cells have 4, 8, and 16 replication origins on average. In minimal medium, 57%, 41%, and 2% of cells have 2, 4, and 8 origins on average. The analysis shows 9.54 ± 0.09 replication forks per cell on average and 3.06 ± 0.02 replication forks per cell on average in rich and minimal medium, respectively (mean \pm SEM). (**F**) Numbers of RDG foci are constant per replication fork regardless of numbers of forks per cell: 5.0 $x10^{-3} \pm 0.3$ $x10^{-3}$ and 4.2 $x10^{-3} \pm 0.6$ $x10^{-3}$ foci per fork in rich and minimal medium, respectively. (**G**) Relative units of DNA per cell, shown as DAPI fluorescence per cell with flow cytometry, in WT and *dnaA*TS cells at 42°C restrictive temperature, at which *oriC* use is blocked. Counterintuitively, at restrictive temperature, *dnaA*TS cells have more DNA per cell than WT cells because the SOS-DNA damage response is induced, which causes a cell-division block with accumulation of multichromosome cells (*107*). We used the integrated areas under these curves to normalize the numbers of foci per cell in Fig. 3E, to the relative amount of DNA per cell to determine that there are 30.8 ± 0.2 times fewer foci per unit of DNA in *dnaA*TS than WT cells at 42°C.

fig. S9. Similar growth rates of various mutant strains used. Growth curves of the strains tested in Figs. 3A and 4A show similar growth rates. Doxycycline was added in log-phase and foci (Figs. 3A and 4A) were counted in early stationary phase. Three independent experiments were performed and error bars show SEM.

supplying RecF, RecJ, and RecQ from plasmids*.* Strains from left to right: SMR21230, SMR21230, SMR21232, SMR21232, SMR21234, SMR21234. * *P* < 0.05 two tailed unpaired *t*-test.

fig. S11. RecA overproduction is induced after RDG accumulation in cells. The amount of RDG in cells was determined by the intensity of GFP (GFP intensity at 0h is normalized to 1), measured by flow cytometry. IPTG was used to induce the P*tac* promoter that controls the (additional) plasmidborne *recA* copy.

fig. S12. Increased *EME1* **and** *GEN1* **HJ resolvase mRNAs in** *BLM-***overexpressing human cancers of the eight, and six of the eight, most common cancer types, respectively.** Spearman's rank correlation analyses of data from cBioportal (*88, 89*), per Methods. (**A-C**) Each data point represents the mRNA level in one patient sample relative to the reference population (Z score, Methods). Data from 129 - 1100 patient samples were analyzed per cancer type (table S1). (**A**) Increased *EME1* mRNA levels (y axis, Z scores) correlated with increased *BLM* mRNA (x axis, Z scores) in eight out of eight of the most common cancer types (Spearman's rank correlation analysis, $R^2 > 0.25$, $P \le 6.15 \times 10^{-18}$). (**B**) Increased *GEN1* mRNA levels (y axis, Z scores) correlated with increased *BLM* mRNA (x axis, Z scores) in six out of eight of the most common cancer types

(Spearman's rank correlation analysis, $R^2 > 0.25$, $P \le 5.61 \times 10^{-10}$). (**C**) Increased *RECQL4* mRNA levels (y axis, Z scores) correlated with increased *BLM* mRNA (y axis, Z scores) in two out of eight of the most common cancer types. (**D**) Summary: correlation of increased *BLM* mRNA levels with increased levels of *EME1, GEN1* and *RECQL4* tumor RNA-seq data of the eight most common cancers. Numbers in parentheses, number of common cancers of the eight most common types correlated. mRNA Spearman's correlation coefficients calculated between these and other human RecQ orthologs and other HJ resolvases with *BLM* and with each other among the eight most common cancers are summarized in table S1 (\mathbb{R}^2 >0.25 indicates moderate correlation, table S1 for details). (**E**) Summary of Spearman's R^2 values for cancer RNAs of genes both correlated with *BLM* expression ($R^2 > 0.25$ for moderate correlation) and control genes poorly correlated or uncorrelated with *BLM* expression; for example, *ACTB*, which encodes a subunit of actin.

fig. S13. Validation of Mu Gam protein function in phage λ*red gam* plaque-size assay. *gam* is temperature-inducibly controlled by the phage lambda promoter P^R repressed by the λ*c*I*ts857*-encoded temperature-sensitive phage lambda transcriptional repressor. Production of Gam at 37°C showed larger plaques of λred gam (mixed Chi⁺ and Chi⁰) phage, per (39) than at 30°C, at which Mu Gam is not produced.

table S2. *E. coli* **K12 strains and plasmids used in this study.** For all pLC plasmids, genome coordinates (indicated in parentheses) correspond to *E. coli* K12 strain MG1655 genome position (U00096.3).

table S3. **Names and locations of new I-sites and alleles.** Coordinates of the insertion/deletion sites correspond to *E. coli* K12 strain MG1655 genome position (U00096.3), with updated I-site D genome coordinates since its original publication (*39*).

table S4. Oligonucleotides used in this study.

