DNA damage-signaling, homologous recombination and genetic mutation induced by 5-azacytidine and DNA-protein crosslinks in *Escherichia coli*

Julie A. Klaric^a, David J. Glass, Eli L. Perr,^b Arianna D. Reuven^c Mason J. Towne and Susan T. Lovett*

Department of Biology and Rosentiel Basic Medical Sciences Research Center

Brandeis University, Waltham, MA 02454-9110

^aCurrent address: Max Planck Institute for Biology of Ageing, Joseph-Stelzmann-Str. 9b, Cologne, Germany

^bCurrent address: Children's Hospital Boston, 1 Blackfan Circle, Karp 8, Boston, MA 02115

^cCurrent address: Dartmouth College, Remsen Building 5th Floor, 66 College St. Hanover, NH 03755

*Corresponding author: lovett@brandeis.edu, Rosenstiel Center MS029, Brandeis University, 415 South St., Waltham, MA 02454-9110

ABSTRACT

Covalent linkage between DNA and proteins produces highly toxic lesions and can be caused by commonly used chemotherapeutic agents, by internal and external chemicals and by radiation. In this study, using *Escherichia coli*, we investigate the consequences of 5-azacytidine (5-azaC), which traps covalent complexes between itself and the Dcm cytosine methyltransferase protein. DNA protein crosslink-dependent effects can be ascertained by effects that arise in wild-type but not in *dcm* Δ strains. We find that 5-azaC induces the bacterial DNA damage response and stimulates homologous recombination, a component of which is Dcm-dependent. Template-switching at an imperfect inverted repeat (“quasipalindrome”, QP) is strongly enhanced by 5-azaC and this enhancement was entirely Dcm-dependent. The SOS response helps ameliorate the mutagenic effect of 5-azaC but unbalanced

expression of the SOS-induced DNA polymerases, especially PolIV, stimulates QP-associated mutagenesis. In the absence of Lon protease, Dcm-dependent QP-mutagenesis is elevated, suggesting it may play a role in 5-azaC tolerance. Deletions at short tandem repeats, which occur likewise by a replication template-switch, are elevated, but only modestly, by 5-azaC. We see evidence for Dcm-dependent and-independent killing by 5-azaC in sensitive mutants, such as *recA*, *recB*, and *lon*; homologous recombination and deletion mutations are also stimulated in part by a Dcm-independent effect of 5-azaC. Whether this occurs by a different protein/DNA crosslink or by an alternative form of DNA damage is unknown.

Keywords: DNA repair, DNA damage response, genome instability, genetic rearrangements

Highlights:

- 5-azacytidine is broadly mutagenic and recombinogenic
- In *E. coli*, 5-azaC promotes genetic instability through Dcm methyltransferase.
- There are other, unknown lesions induced by 5-azaC besides Dcm/DNA crosslinks
- 5-azaC induces the SOS response, protecting cells from killing and genetic instability

1. INTRODUCTION¹

DNA protein crosslinks (DPCs) are a common spontaneous source of DNA damage and can be formed by both nonenzymatic or enzymatic mechanisms. Nonenzymatic DPCs can be induced by radiation and a wide variety of chemicals (Barker et al. 2005). For example, reactive oxygen species and aldehydes, either encountered from the environment or endogenously formed by metabolic processes, can covalently link

¹ Abbreviations: DPCs, DNA protein crosslinks; CMeT, cytosine methyltransferase; 5-azaC, 5-azacytidine; DSBs, double strand breaks; ssDNA, single-strand DNA; QP, quasipalindrome; QPM, quasipalindrome-associated mutation

proteins to DNA. In addition, a number of conserved cellular enzymes are covalently linked to DNA as a reaction intermediate (Ide et al. 2011). Two examples are DNA topoisomerases and DNA cytosine methyltransferases (“CMeTs”). Certain drugs trap these normally short-lived covalent intermediates. The quinolone class of antibiotics and anticancer drugs etoposide and doxorubicin are topoisomerase type IIA poisons for bacterial and eukaryotic enzymes, respectively, and trap the covalent tyrosyl-topoisomerase/phosphoryl-DNA intermediate (also known as the “cleaved complex”). Camptothecin, another chemotherapeutic drug, is a similar poison for eukaryotic type IB topoisomerases (Drlica and Zhao 1997; Nitiss 2009; Pommier et al. 2010). In addition, cytosine methyl transferases (CMeTs) form a covalent complex with C6 of cytosine in their reaction to methylate the C5 of cytosine in DNA or RNA (Figure 1) (Santi et al. 1984; Friedman 1985). The nucleotide analog 5-azacytidine (Figure 1) inhibits the completion of the reaction, trapping the protein covalently attached to the nucleotide base.

DPCs impede both RNA transcription (Nakano et al. 2017) and replication fork progression (Kuo et al. 2007). In humans and other mammals, the inability to repair DPCs accelerates aging and leads to cancer proneness (Craft et al. 1987; Izzotti et al. 1999; Zahn et al. 1999; Wu et al. 2002; Swenberg et al. 2011; Garaycoechea et al. 2012).

In bacteria, DPCs caused by topoisomerase II poisons or 5-azacytidine can lead to double-strand breaks (DSBs). Accordingly, mutants in double-strand break (DSB) repair factors RecA and RecB are sensitive to both to quinolones and 5-azaC (Bhagwat and Roberts 1987; Lal et al. 1988; Butala et al. 2009; Salem et al. 2009; Ide et al. 2011), although there may be other toxic lesions in addition to DSBs. For topoisomerase II poisons, a cytotoxic lesion is some processed form of the trapped “cleaved complex” since killing can be suppressed by inhibitors of RNA and protein synthesis (reviewed nicely in (Drlica and Zhao 1997)). Even in the absence of frank DSB formation, these covalent complexes can block the

completion of replication, allowing a second replication fork converging onto the site to produce a broken chromosome (Cox et al. 2000).

In bacteria, 5-azaC is a useful tool to study effects of DPCs in vivo whose targets are C5-cytosine methyltransferases (CMeT) in restriction/modification systems. Restriction/modification systems are widespread throughout bacteria, and consist of an endonuclease component and a methyltransferase (Roberts et al. 2015). These systems are considered to be primitive innate immune systems, leading to

the destruction of

invading foreign

DNA. A proportion

of restriction

systems use C5-

cytosine methylation

Figure 1. Cytidine-derived structures

(Fig. 1) to provide immunity to restriction endonuclease cleavage (other systems methylate the N6 of adenine or the N4 of cytosine to confer immunity). Because many of these modification enzymes methylate specific DNA recognition sequences and are nonessential in the laboratory environment, we can target DPCs to particular sites to study their consequence in vivo. In addition, in the absence of the methyltransferase the adduct will not form, providing confidence that any effect is due to the DPC alone. We have recently documented that DPCs in *Escherichia coli* dramatically stimulate a particular class of mutation. This class of mutation arises by a replication template-switch reaction in an imperfect inverted repeat sequence (or “quasipalindrome”, “QP”) (Ripley 1982; Glickman and Ripley 1984; Yoshiyama and Maki 2003; Dutra and Lovett 2006; Lovett 2017). QPs often constitute mutational hotspots, both in bacteria and in eukaryotes (Hampsey et al. 1988; Greenblatt et al. 1996; Bissler 1998; Viswanathan et al.

2000;Yoshiyama et al. 2001;Yoshiyama and Maki 2003). Using QP mutational reporter strains, we discovered that the DPC-producing agents formaldehyde, ciprofloxacin and 5-azaC are potent mutagens for this class of mutations (Laranjo et al. 2018). For 5-azaC, the mutagenic effect requires *E. coli*'s endogenous CMeT, Dcm, indicating that it is the trapped DPC that induces mutation at the QP site (Laranjo et al. 2018).

Surprisingly, although the genes required in DSB repair, *recA* and *recB*, are necessary to avoid to the lethal effects of DPCs, they are not required for their mutagenicity (Laranjo et al. 2018). This suggests that the mutagenic lesions include something other than DSBs. That the mutagenic lesion may be a block to the DNA replication fork is suggested by the fact that impairment in the DnaB fork helicase greatly elevates QP-associated mutation (QPM)(Laranjo et al. 2018). To further explore how DPCs lead to genomic instability, we examine here the consequence of DPCs on genetic recombination and DNA damaging signaling. We further explore the effect of the SOS response, translesion DNA polymerase and the Lon and ClpP proteases on mutagenicity of 5-azaC in *E. coli* at a QP site. We additionally quantify effects on genomic rearrangements at short tandem repeats, which are caused by a different type of template-switching during DNA replication (Lovett et al. 1993; Lovett and Feschenko 1996).

This work finds that DPCs are broadly mutagenic, recombinogenic and inducers of the SOS response. The SOS response is protective in the avoidance of QPM, although unbalanced expression of translesion DNA polymerases, particularly DNA polymerase IV (DinB), is mutagenic. Most of the effects of 5-azaC are due to DPCs formed by the *E. coli* CMeT, Dcm; however, we also find evidence for Dcm-independent DNA lesions that are lethal or mutagenic in certain sensitive genetic backgrounds.

2. MATERIALS AND METHODS

2.1 Strains and growth conditions: All of the strains used are derivatives of *Escherichia coli* K-12 MG1655 (Bachmann 1972; Table 1). Isogenic strains of the indicated genotypes were constructed by P1 vir transduction. DNA transformation by electroporation (Dower et al. 1988) was used to introduce

plasmids and oligonucleotides for recombineering. The strains were grown at 37 °C in Luria broth (LB, Lennox formulation) medium, consisting of 1% Bacto-tryptone, 0.5% yeast extract, 0.5% sodium chloride and, for plates, 1.5% Bacto-agar. Tetracycline (Tc, 15 µg/mL), chloramphenicol (Cm, 30 µg/mL), ampicillin (Ap, 100 µg/mL) and kanamycin (Km, 45-60 µg/mL) were used for genetic selections. Lac⁺ reversion mutants were selected on lactose minimal medium (Lac Min XI) containing 56/2 salts (Willetts et al. 1969), 0.2% lactose (Sigma Aldrich, St. Louis, MO, USA), 0.001% thiamine (Sigma Aldrich), and 2% agar (ThermoFisher, Sparks, MD, USA). X-gal (40 µg/mL) and IPTG (0.1 mM) (both from Gold Bio, St. Louis, MO, USA) were included in lactose selection medium as a visual aid for counting colonies. Glucose minimal (Glu Min XI) control plates were made similarly but contained 0.2% glucose in place of lactose.

2.2 Survival assays: For survival to 5-azacytidine (5-azaC, Sigma Aldrich), MG1655 and isogenic derivative (Table 1) were grown to mid-log phase (OD₆₀₀ of approximately 0.2- 0.4), then split and treated at the indicated concentrations or untreated for 2 hr prior to serial dilution and plating on LB media. Survival experiments were performed for at least 3 replicates.

2.3 Luciferase assays: To measure induction of the SOS response, we used luciferase fusions to the *recA* and *dinB* promoters as previously described (Goldfless et al. 2006) and strain backgrounds MG1655, STL19845, STL21742, and STL22811. A single colony was used to inoculate 2 mL of LB+Ap media and grown at 37° with agitation for 2 hours until OD₆₀₀ ≈ 0.5. The culture was then diluted 1:100 in fresh LB+Ap media and grown again at 37C with agitation for 2 hours. The culture was then diluted 1:100 into fresh LB+Ap media in a Costar 96 Well Assay Plate (Treated Polystyrene, Black Plate, Clear Bottom). The plate was then grown in a BioTek Cytation 1 Plate Reader at 37° shaking for 75 minutes, with OD₆₀₀ and luminescence readings being taken every 15 minutes. After 75 minutes, 5-azacytidine was added to the appropriate wells to a final concentration of 12.5 µg/mL. Cultures were then grown for 3.5 hours shaking at 37° with OD₆₀₀ and luminescence readings taken every 15 minutes. Relative luminescence units (RLU) values were calculated by normalizing the luminescence readings (in counts per second) to the OD₆₀₀ of the cultures.

2.4 Deletion assays: Deletion of 101 bp tandem repeat in *tetA* was performed as described using plasmid pSTL57 (Lovett et al. 1994) and strains MG1655 and STL21742. Deletion of 11 bp tandem repeats in *lacZ* was performed as described (Seier et al. 2011) with strain STL14025. For experiments involving 5-azacytidine (Sigma Aldrich) or formaldehyde (ThermoFisher), mid-log phase cultures were split, and treated with the drug, or not, for 2 hours at the indicated concentrations before serial dilution and plating on appropriate selective media. Data are presented as the fold-change in deletion frequencies of the treated vs. untreated samples.

2.5 Crossover recombination assay: Recombination between 411 bp of homology in plasmids pSTL330 and pSTL36 was selected by Tc-resistance as described (Lovett et al. 2002) and strain backgrounds MG1655, STL21742 and STL19845. Mid-log phase cultures were split, and for two hours were treated with 5-azaC at the indicated concentrations before serial dilution and plating on appropriate selective media. Both wt and *recO* mutants were assayed with 4 independent replicates.

2.6 Gene conversion recombination assay: Gene conversion is selected in this assay by the Lac⁺ progeny of a strain carrying two mutant alleles of *lacZ*, one with an internal deletion at the natural *lacZ* locus (7.83 centisomes) and an internal 500 bp *lacZ* fragment integrated at *attTn7* near *glmS* (84.28 centisomes). The *lacZ* recipient allele was constructed by recombineering using a single-strand DNA oligonucleotide and DNA transformation (Sharan et al. 2009). The 5 nucleotides spanning the active site amino acid codon, from nucleotides 1384-1388 of the *lacZ* ORF, were deleted by transformation of oligonucleotide 5' ATCACCCGAG TGTGATCATC TGGTCGCTGGG GAATAGGCCA CGGCGCTAAT CACGACGCGCTGTATCGC in a MG1655 strain carrying the pSIM6 recombineering plasmid (Sharan et al. 2009), yielding strain STL20478 after curing the pSIM6 plasmid with growth at 42°. To construct the recombination donor locus, we used a Tn7-based system to introduce a DNA fragment carrying 250 bp of homology flanking both sides of the *lacZ* active site deletion into the *attTn7* site, near *glmS* (McKenzie and Craig 2006). The DNA fragment produced by colony PCR with oligonucleotides 5' GGGGccccggg GCTGATGAAG CAGAACAAC TTAACGCCGT 3' and 5' GCttaattaa ACTGTTACCC

ATCGCGTGGG CGTATTCGCA 3' was inserted into temperature-sensitive Tn7 vector pGRG25 (McKenzie and Craig 2006) after cleavage with *PacI* and *XmaI* (New England Biolabs). Transformation of this plasmid into strain STL18091 (MGI655 *bglG::FRT kan*) was selected by Ap-resistance at 30°, after which the plasmid was cured by growth in LB at 42°; integrants of the Tn7 end-flanked fragments were identified by PCR among survivors as described (Laranjo et al. 2017). This allele was transduced into strain STL20478 by cotransduction with *bglG::FRT kan*. The *kan* was then removed using FLP plasmid pCP20 (Datsenko and Wanner 2000). The resulting recombination assay strain carrying both loci (*attTn7::lacZ' lacZ_{asΔ}*) is STL22043. Homologous recombination in this strain generates a *lacZ*⁺ locus in its natural site; the donor locus at *attTn7* is unchanged. Strains carrying only the donor or only the recipient *lacZ* locus never yielded Lac⁺ progeny, nor did *recA* strains carrying both alleles, confirming that *lacZ*⁺ isolates arise by homologous recombination between the loci.

For the recombination assays, cultures were grown in LB to midlog phase, then concentrated 5x by centrifugation and resuspension into LB medium containing 5-azaC or not. Growth of these split cultures was continued for 2 hours after which they were harvested by centrifugation, washed with 1 x 56/2 salts and serially diluted. Dilutions were plated on Lac Min XI to determine the number of Lac⁺ recombinants and on Glucose Min XI and LB for total number of CFU. LB plates were incubated at 37° 1 day and minimal plates for 3-4 days.

2.7 Mutation assays. We used a chromosomal *lacZ* mutational reporters specific for G to C transversion and for quasipalindrome-associated template-switch mutations, QP5 and QP6, previously described (Seier et al. 2011). The strains used are listed in Table 1, with the genotype including *lacZ*(QP5) or *lacZ*(QP6). C to G transversion mutations were measured using STL13814 ((Seier et al. 2011) and a *dcm::FRT kan* transductant of this strain, STL22870. +1 G frameshift mutations were measured using strain STL14095 carrying the indicated plasmids. 5-azaC effects were determined with mid-log phase split cultures and with treatment for 2 hours prior to dilution and plating.

2.8 SOS polymerase plasmids: GATEWAY (Thermo Fisher) recombinational cloning technology was used to clone *dinB*, *polB* and *umuCD'* into the pBAD18 arabinose-controlled expression vector. From MG1655 colonies, the *dinB* gene was amplified with the following PCR primers 5' GGGGACAAGT TTGTACAAAA AAGCAGGCTT CGAAGGAGAT AGAACCatgc gtaaatcattca and 5' GGGGACCACT TTGTACAAGA AAGCTGGGTC tcataatccc agcacc, the *polB* gene with 5' GGGGACAAGT TTGTACAAAA AAGCAGGCTT CGAAGGAGAT AGAACCgtgg cgcaggcagg ttttatc, 5' ggggACCACT TTGTACAAGA AAGCTGGGTC tcaaatagc ccaagttgccc. The *umuD'*C gene was amplified from plasmid MPI (Addgene) with primers 5' GGGGACAAGT TTGTACAAAA AAGCAGGCTT CGAAGGAGAT AGAACCatga agagattgca gctcatg, 5' ggggACCACT TTGTACAAGA AAGCTGGGT Cttatttgac cctcagtaaa tcag and inserted using BP reactions into pDONR201. LR reactions transferred these insert to the pSTL360 plasmid (Dutra and Lovett 2006), a destination plasmid derived from pBAD18 (Guzman et al. 1995).

3. RESULTS

3.1 Recombination pathways and DPC processing: The *Escherichia coli* K-12 genome encodes one cytosine methyl transferase, Dcm ("DNA cytosine methylase"). The function of this gene is not known but it is likely a remnant of a restriction/modification system. Dcm methylates the second C in the sequence "CC[A/T]GG"; this is the same specificity as the EcoRII restriction/modification system, encoded on an *E. coli* antibiotic-resistance conferring-plasmid, derived from a clinical isolate (Bannister and Glover 1968; Yoshimori et al. 1972; Bhagwat et al. 1986; Bhagwat et al. 1990).

E. coli K-12 wild-type cells are not particularly sensitive to killing by 5-azaC and mutants in *dcm* are only modestly more resistant (Fig. 2). The fraction of potential sites methylated by Dcm is unknown but there is evidence that not every potential Dcm site is methylated (Bhagwat et al. 1990; Ringquist and Smith 1992); for example, EcoRII overexpression confers additional sensitivity to 5-azaC in *dcm+* *E. coli* (Bhagwat et al. 1990; Krasich et al. 2015). With only the endogenous Dcm CMeT, *recA* and *recB* mutants, deficient in DSB repair, show enhanced sensitivity (Fig. 2), confirming previous studies (Bhagwat and

Figure 2. Fractional survival of *E. coli* genetic variants at indicated doses of 5-azaC for 2 hours. Data are average of at least 3 experiments; error bars represent standard deviation. A. Recombination mutants B. SOS response mutants C. Protease mutants.

Roberts 1987; Lal et al. 1988; Butala et al. 2009; Salem et al. 2009; Ide et al. 2011). This sensitivity is partially dependent on Dcm. There is, however, residual sensitivity, suggesting Dcm-independent toxicity of 5-azaC. This residual sensitivity of *recA dcm* has also been previously documented (Bhagwat and Roberts 1987).

Figure 3. Induction of SOS promoters, *dinB* and *recA*, fused to luciferase expression reporters. Red treated with 50 μ g/ml 5-azaC; black untreated. 5-azaC was added at time 0 and the culture was followed for 300 minutes post-treatment. RLU, relative luciferase units is the luminescence divided by the OD₆₀₀ of the culture.

RecA is required for homologous recombination in *E. coli*, and it participates in two genetically distinct pathways. One of these, involving the RecBCD nuclease, is specific for recombination arising from DSBs, the other, dependent on RecFOR, is specific for recombinational repair of single-strand gaps (reviewed in (Persky and Lovett 2008)). These pathways are defined by the functions that facilitate the loading of RecA, which RecBCD performs on newly resected DNA and RecFOR on ssDNA gaps, the latter displacing bound single-strand DNA binding protein (SSB). The RecA filament bound to ssDNA not only directly facilitates homology search and DNA strand exchange, but is also the inducing signal for the SOS response, a transcriptional response in *E. coli* to DNA damage (Simmons et al. 2008; Lovett 2010). We show here that RecO mutants exhibit no sensitivity to 5-azaC (Fig. 2A) unlike mutants in RecA or RecBCD.

One difficulty in assessing the role of the RecFOR pathway is that, in its absence, gaps become converted to DSBs and are repaired efficiently by the RecBCD pathway. By survival assays to genotoxins, there may

be no evidence for RecFOR participation. However, genetic measurements of homologous recombination and induction of the SOS response may show that the RecFOR pathway is indeed called into play. For example, the replication chain terminator, azidothymidine (AZT), blocks replication and induces the SOS response through the RecFOR pathway, although *recFOR* mutants show no sensitivity to AZT (Cooper and Lovett 2010).

To ascertain the participation of the RecFOR pathway in processing DPCs, we measured SOS induction in *E. coli* after 5-azaC treatment, using a luciferase transcriptional reporter fused to the *dinB* and *recA* promoters. Both of these genes are regulated as part of the SOS regulon, although *recA* is more highly expressed than *dinB*. In these assays, cells are continuously exposed to sublethal concentrations of 5-azaC, beginning in early log phase and monitored through the growth of the culture into stationary phase. Luminescence of the culture is normalized to OD₆₀₀ of the culture, measured at 15 minute intervals. 5-azaC does indeed induce expression of both *dinB* and *recA* promoters, reaching a maximum about 2 hrs after introduction of the drug. This induction was dependent on *dcm*, confirming that DPCs are responsible for the induction. For both induced and uninduced expression, we saw a growth-dependent increase in expression, peaking at late exponential phase. We do not understand the basis of the effect. This increase was not seen in control cultures with the luciferase operon, lacking the promoter regions, so this signal is dependent on the *recA* or *dinB* promoters. We saw a modest reduction by *recO*, although the effect is obscured by the elevation of expression in *recO* untreated cultures, possibly due to the accumulation of spontaneous DNA lesions in this genetic background.

3.2 Does 5-azaC induce recombination, as well as the SOS response? We measured two types of recombination with genetic assays: one selects for crossovers between 411 bp of homology on low copy plasmids (Lovett et al. 2002) and the other detects nonreciprocal gene conversion between an internally deleted *lacZ* gene and an ectopic internal '*lacZ*' gene segment at a position across the chromosome, near *oriC* (see Materials and Methods).

Figure 4. Crossover recombination induced by 5-azaC, 2 hr treatment at indicated doses, in split cultures of wt or *recO* mutant strains. Bars represent mean values of the replicates.

Crossovers detected in the plasmid assay are primarily dependent on the RecFOR pathway (Lovett et al. 2002); because these plasmids lack Chi sites that protect from degradation by RecBCD (Anderson and Kowalczykowski 1997); there is no detectable contribution of the RecBCD pathway. Using cultures that were split and treated for 2 hr with 5-azaC (Fig. 4), we saw a modest 2- to 3-fold elevation of crossing-over with 5-azaC relative to untreated

cultures. This increase was dependent on *recO*. 5-azaC induced recombination was dependent on *dcm* as well, with *dcm* mean fold-induction values of 0.89 ± 0.18 , 0.60 ± 0.21 , 0.50 ± 0.39 at 25, 50 and 75 µg/ml 5-azaC, respectively (n=4).

Gene conversion between chromosomal *lacZ* alleles was more strongly stimulated by 5-azaC, with an over 30-fold increase at 50 µg/ml. (Figure 5). This was reduced 2-fold by *recO*, although there was considerable variation among replicates. We saw no evidence for Dcm-dependence of this recombination, so there must be a recombinogenic lesion induced by 5-azaC other than Dcm-DPCs. The residual recombination in *recO* mutants is presumably dependent on the alternate RecBCD recombination pathway (Persky and Lovett 2008), but we were unable to test *recB* and *recB recO* mutants in this assay, due to their strong sensitivity to 5-azaC.

3.3 Template-switching at quasipalindromes: effect of recombination functions. We had previously documented a strong stimulatory effect of 5-azaC on mutagenesis detect by template-

Figure 5. Gene conversion recombination induced by 50 μ g/ml 5-azaC, 2 hr treatment at indicated doses, in split cultures of wt or *recO* mutant strains. Bars represent mean values of the replicates and standard deviations.

switching in a quasipalindromic sequence. In this study, we expanded this analysis to include additional mutants and doses of 5-azaC. We examined, in parallel, the effect of this set of homologous recombination factors on QPM using leading and strand strand reporter strains, measuring *lacZ* reversion frequencies with and without treatment with 5-azaC at two doses. For each genetic background, an

isogenic *dcm* derivative was also assayed. There are 8 potential Dcm methylation sites within *lacZ*. As we had seen previously (Laranjo et al. 2018; Klaric et al. 2020), we observed stimulation of QPM with increasing 5-azaC dose, with a stronger effect with the leading strand template-switch reporter QP5. (Fig. 6) Most, but possibly not quite all, of this induction is lost in *dcm* mutants. Mutants in *recA* showed no decrement in mutagenicity; if anything, they showed a slightly higher constitutive and 5-azaC-induced frequencies, at both doses, for both reporters. Mutants in *recA* had a strong effect of 5-azaC on the lagging strand reporter QP6 where it elevated QPM more than an order of magnitude at the higher dose. Again, as for wt, 5-azaC-induced QPM in *recA* was largely absent in *dcm* double mutants. Mutants in *recB* showed a slightly higher constitutive rate of QPM, but 5-azaC induction could not be tested at the higher dose because of poor survival. Mutants in *recO* showed a higher constitutive frequency of QPM in untreated cells but looked virtually indistinguishable from wt after 5-azaC treatments. These results suggest that homologous recombination and/or SOS induction promotes higher survival to 5-azaC, with RecA playing a antimutagenic role for QPM. Elevated constitutive QPM rates for *recB* and

recO mutants would be consistent with the idea that both RecBCD and RecFOR pathways process spontaneous lesions, with a stronger effect of the latter pathway.

Fig. 6. Mutation frequencies with and without 5-azaC treatment in recombination mutants, genotype shown above bracket. Shown are *dcm*⁺ and *dcm* derivatives of wt, *recA*, *recB* and *recO* mutants. QP5 is a leading strand mutational reporter; QP6 a lagging strand reporter. The log scale of the the Y-axis is the same in the two graphs, to allow easier visual comparison. No treatment frequencies in brown, 1 µg/ml 5-azaC in light green, 12.5 µg/ml in dark green. Black bar indicates the median value of the replicates.

3.4 Template-switching

at QPs: effect of SOS

response and

translesion DNA

polymerases: The SOS

response of *E. coli* consists of a number of functions coordinately induced by DNA damage or replication

inhibition (Simmons et al.

2008). The genes for these

functions are transcriptionally

repressed by LexA, but which

are induced by LexA self-

cleavage promoted by the

accumulation of RecA/ssDNA

filaments. By measuring QPM

in *lexA3* strains, in which the SOS response cannot be induced, we had previously shown that the SOS response is anti-mutagenic for QPM, both constitutively and that induced by DPCs ((Laranjo et al. 2018) and Figure 8.) Mutants in *lexA3* are also sensitive to killing by 5-azaC (Figure 2B). Several of the factors induced by the SOS response are the translesion DNA polymerases, Pol II, Pol IV and Pol V (encoded by *polB*, *dinB* and *umuCD* genes, respectively). To determine if these were responsible for the antimutagenic effect of the SOS response, we mutated all three and measured spontaneous QPM rates in a wild-type

Figure 7. QPM frequencies in strains lacking the TLS DNA polymerases (*polB dinB umuC*) with and without the DNA exonucleases Exo I and/or Exo VII. Bars indicate the median value of the replicates.

genetic background and in strains lacking one or both of the 3' to 5' exonucleases Exo I and Exo VII.

These exonucleases abort QPM, with Exo I playing the larger role (Laranjo et al. 2017); hence QPM rates are much higher in their absence and potentially more sensitive to TLS polymerase effects. We saw no effect in any background of TLS polymerases on mutation frequency, for either the leading or lagging strand QPM reporters (Figure 7). The loss of all three TLS polymerases also did not reduce survival to 5-azaC (Figure 2B). Therefore, induction of TLS is not responsible for the antimutagenic aspect of the SOS response.

We examined mutants in which the SOS response was constitutively induced, lacking the LexA repressor.

Because the cell division inhibitor SulA is induced during the SOS response, *lexAΔ* strains must carry an additional knockout of *sulA* to remain viable. With the QP5 reporter, *lexA sulA* null strains showed diminished 5-azaC-induced QPM relative to wt strains (Figure 8). However, the *sulA* mutation by itself

Fig. 8. Mutation frequencies with and without 5-azaC treatment in *lexA* mutants, genotype shown above bracket. Shown are *dcm*⁺ and *dcm* derivatives of *sulA* and *lexAΔ sulA* compared to WT and *lexA3* mutants assayed in parallel. QP5 is a leading strand mutational reporter. No treatment frequencies in brown, 1 $\mu\text{g/ml}$ 5-azaC in light green, 12.5 $\mu\text{g/ml}$ in dark green. Black bar indicates the median value of the replicates.

entirely dependent on *dcm* (Fig. 8), showing that it is promoted by the DPCs formed by Dcm methyltransferase in the presence of 5-azaC.

The most striking effect of *lexAΔ*, however, was a more than 10-fold elevation of constitutive rates of QPM (in the absence of 5-azaC) relative to wt and *sulA* strains (Fig. 8). This suggests that the SOS response also can *promote* mutagenesis in some instances.

We wondered whether the unnatural, full-on induction of TLS DNA polymerases was responsible for this effect. We examined strains overexpressing one of the TLS DNA polymerases from an arabinose-regulated pBAD plasmid, with a copy number approximately 20 per cell, in an otherwise wt strain. In the absence of induction, we saw no effect of the TLS polymerases. After 2 hours of induction in arabinose

Figure 9. Induced expression of TLS DNA polymerase and effects on mutation frequencies with the indicated *lacZ* mutational reporter strains. Bars indicate the median value of the replicates.

medium, there was a strong mutagenic effect of DNA polymerase IV (DinB) for mutations detected by both QP5 and QP6, 34- and 13-fold, respectively. (Figure 9). Pol IV is known to be prone to -1 G

frameshift mutations (Kim et al. 1997) and in a *lac* reporter specific for this mutation, we saw a very large increase in mutation frequency by Pol IV when induced. Induction of Pol II caused about a 3-fold increase in QPM on the leading strand (QP5), and a 2-fold effect on the lagging strand (QP6). Pol IV also was modestly mutagenic. This suggests that the TLS polymerases are mutagenic for QPM when their abundance is unbalanced, particularly Pol IV, a potential explanation for the hypermutable constitutive phenotype of *lexAΔ sulA* strains.

3.5 Template-switching at QPs: effects of proteases: Because DPCs may be removed by proteolysis (Vaz et al. 2017), we examined two major intracellular, energy-dependent proteases of *E. coli* for effects on 5-azaC induced QPM. Mutants in the Lon protease were found to be extremely sensitive

to killing by 5-azaC (Figure 2C) and this sensitivity was abolished completely by a secondary mutation in the cell division inhibitor, *sulA*. During recovery from the SOS response, Lon protease destroys SulA, allowing cells to divide again. In the absence of

Fig. 10. Mutation frequencies with and without 5-azaC treatment in *lon* and *clpP* mutants, genotype shown above bracket. Shown are *dcm⁺* and *dcm* derivatives of *lon* and *clpP* mutants, compared to *wt* assayed in parallel. QP5 is a leading strand mutational reporter. No treatment frequencies in brown, 1 μ g/ml 5-azaC in light green, 12.5 μ g/ml in dark green. Black bar indicates the median value of the replicates.

Lon, SOS induction is lethal because cells never recover the ability to divide (Schoemaker et al. 1984).

Complete suppression of *lon* by *sulA* (Figure 2C) suggests that the failure to turn off the SOS response to resume cell division is predominantly responsible for 5-azaC-induced killing in this background. A secondary mutation in *Dcm* conferred only partial resistance to 5-azaC in *lon* mutants. Therefore, some lethal SOS induction derives from *Dcm* DPCs formed with 5-azaC, but there must be a second SOS-inducing lesion that is *Dcm*-independent. Mutants in *ClpP*, a component of both *ClpAP* and *ClpXP* protease complexes (Gottesman 1996), did not show any sensitivity to 5-azaC (Figure 2C).

With QPM reporter QP5, mutants in *lon* showed hypermutability, 20-fold relative to wt, after treatment with the lower dose of 5-azaC (the higher dose could not be tested because of killing) (Figure 10). This elevated mutagenesis was entirely dependent on *dcm* (Figure 10). Mutants in *clpP* exhibited normal levels of *Dcm*-dependent QPM as assayed with reporter QP5.

3.6 Template-switch mediated deletions: Deletions and expansion of tandem direct repeats occurs by a template-switch mechanism, independent of homologous recombination proteins including *recA* (reviewed in (Lovett 2017)). With 2 different assays for deletion formation (Lovett et al. 1994; Seier et al. 2011), we tested the effects of DPC-agents 5-azaC and formaldehyde. 5-azaC showed a dose-dependent increase in the deletion frequency of 11 bp tandem repeats within the chromosomal *lac* locus, with a 4-fold enhancement of deletion at the highest dose of 100 µg/ml (Figure 11a). Note that with a hundred-fold lower dose of 1 µg/ml 5-azaC, we saw a 5-fold enhancement of QPM *lac* mutagenesis using the QP5 reporter (Fig. 3A), so template-switching that gives rise to short deletions is not as strongly affected by 5-azaC as are QP-templated mutations. The 5-azaC effect was reduced to 2-fold in the *dcm* mutant (Fig. 11b), although the distributions considerably overlap. This indicates that although some of the deletionogenic effect of 5-azaC may occur through *Dcm* DPCs, some 5-azaC-induced deletions are independent of *Dcm*. Formaldehyde had a much stronger effect on deletion in this assay, with an enhancement of over 20-fold in the standard 2-hour treatment (Figure 11a). In a plasmid assay for deletion of 101 bp repeats within the *tetA* gene, conferring tetracycline resistance, we saw that 2 hr of 5-azaC treatment enhanced deletion frequencies 2-fold, which was reduced to 1.2 fold in the *dcm* mutant

Fig. 11. Tandem repeat deletion induced by 5-azaC and FA. Cultures were split and treated for 2 hr with 5-azaC or FA. The mutation frequency in the treated culture relative to that in the untreated is plotted: a value of 1 (indicated by dashed line in A and C) means there is no effect. Bars indicate mean values. A. Deletion of 11 bp repeats in chromosomal *lacZ*. B. Effect of *dcm* on 11 bp *lacZ* deletion. C. Deletion of 101 bp repeats on plasmid-encoded *tetA* and effects of *dcm*.

(Figure 11c). Similar treatment of wt type cells with formaldehyde enhanced deletion in this assay 3-fold and it was not appreciably affected by *dcm*, as expected, since the DPCs induced by FA are nonspecific.

4. DISCUSSION

We used 5-azacytidine to investigate the genetic consequences of DPCs formed by cytosine methyltransferase, Dcm, in *E. coli*. Using a variety of genetic assays, we measured 5-azaC effects on the DNA damage response, recombination and genetic mutations produced by template switching, at inverted or direct DNA repeats.

4.1 The DNA damage response: By transcriptional reporter assays for *dinB* and *recA* promoters, we show that 5-azacytidine induces the SOS response, a transcriptional response triggered by the formation of RecA filaments on single-strand DNA in vivo. This induction was dependent on Dcm, implicating the Dcm protein-DNA crosslink as the inducing lesion. We saw a modest reduction in a RecO mutant, implicating the RecFOR pathway for SOS induction; however, at least for the RecA promoter, there was detectable 5-azaC induction independent of RecO. We suspect that the RecBCD pathway also induces SOS in the absence of RecO, but this dependence could not be tested due to poor survival of *recB* strains to 5-azaC.

4.2 Homologous recombination: A 2-hour treatment with 5-azaC induced homologous recombination about 30-fold, as detected by a gene conversion assay between an internally-deleted *lacZ* gene and a 500 bp homologous *lacZ* fragment integrated across the chromosome, at *attTn7*. This increase in recombination was also seen in a *dcm* mutant, suggesting that lesions other than Dcm-DNA crosslinks can initiate recombination. There was a small, 2-fold, reduction in 5-azaC-induced recombination in a *recO* mutant. A plasmid-based assay for crossover recombination, showed a more modest 2-3 fold stimulation by a 2-hr treatment with 5-azaC, which was reduced by *recO* and *dcm*. Therefore, we found evidence for homologous recombination induced by 5-azaC, some associated with DPCs and some not.

4.3 Mutability at QP sites/ effects of recombination: In the absence of RecA, 5-azaC mutagenicity was enhanced at quasipalindromic (QP) sites, dependent on the Dcm methyltransferase.

The elevation by *recA* was particularly striking for the QP6 reporter for lagging-strand template-switching, where loss of *recA* elevated mutation frequencies, relative to wt, 8- and 68-fold at 1 and 12.5 µg/ml 5-azaC, respectively. This elevation of lagging strand QPM was not due to lack of the SOS response, since *lexA3* mutant strains, likewise unable to mount the SOS response, showed no or 3-fold elevation of QP6 rates at 1 and 12.5 µg/ml 5-azaC, respectively, relative to wt type strains. Therefore, the capacity for homologous recombination, negated by *recA* but not by *lexA3*, helps avoid QPM by 5-azaC but apparently only, or predominantly, on the lagging strand.

4.4 Mutability at QP sites/ effects of the SOS response and TLS DNA polymerases:

The SOS response does have an antimutagenic effect on 5-azaC -induced leading strand QPM, with 4-6-fold enhancement by *lexA3* and 8-fold by *recA*, relative to wt. Our prior work with the natural QP mutational hotspot in the *thyA* gene had indicated that the TLS polymerases, under SOS control, had an antimutagenic effect (Dutra and Lovett 2006), a potential explanation for the antimutagenic nature of the SOS response. However, in this work, using QP sites within the *lacZ* gene, we can find no evidence for such an effect. We saw no effect on mutation rates by introduction of *polB dinB umuC* knockouts, even when ExoI and ExoVII were absent. Rather, these SOS-induced polymerases, when overexpressed, appeared to be mutagenic for QPM. When expressed from the arabinose pBAD promoter on a medium copy number plasmid, induction of any of the three polymerases elevates QPM, with the strongest effect by *dinB*. Because we do not know the extent of the induced expression, this is difficult to interpret, although it does indicate that unbalanced polymerase expression has the potential to cause QPM. This is also consistent with our finding that *lexAΔ* strains show an elevated spontaneous level of QPM. Although Pol IV (DinB) is reported to have a greater mutagenic effect on the lagging strand (Fijalkowska et al. 2012), we see increases in QPM with both leading and lagging strand reporters, although this could be influenced by over-expression.

Surprisingly, a *sulA* mutation, which prevents cell division arrest after DNA damage, reduced the mutagenicity of 5-azaC, as detected on the leading strand QP5 reporter, from 3- to 5-fold. We know of

no precedent for *sulA* effects on mutagenesis, with the exception of a very modest (<2x) effect on stress-induced mutation (Al Mamun et al. 2012). Except for *dcm* mutants, this is the only strain to exhibit lower frequencies of mutation than wt after 5-azaC treatment. We can explain the lower yield of mutants in the *sulA* strain by proposing that potential mutants fail to survive in the absence of cell division arrest, and so are lost in this assay. The loss of *sulA* did not, however, influence survival of the bulk cell population to 5-azaC, so the vulnerable cells in this case must be a subpopulation.

4.5 Mutability at QP sites/ effects of proteases: The Lon protease plays a potential role in the avoidance of 5-azaC induced mutations at quasipalindromes; *lon* mutants were hypermutable by over 50-fold at the lower 5-azaC dose. We documented this elevation using the leading strand QPM reporter, but saw a similar enhancement on with the lagging strand reporter (Klaric, unpublished results). The persistence of a DPC in the absence of Lon may promote template-switching, whereas prompt removal of the DPC by Lon proteolysis leads to a nonmutagenic outcome. Mutants in *lon* were extremely sensitive to killing by 5-azaC, but this appears to be entirely due to the failure to recover after induction of the SOS response; this sensitivity was suppressed completely by an additional mutation in *sulA*. We saw no effect by ClpP.

4.6 5-azaC: not just DPCs? Our data suggest that although Dcm-dependent DPCs largely contribute to genetic effects of 5-aza, there is some Dcm-independent lesion(s) as well. This is particularly apparent in the survival curves of strains sensitive to killing by 5-azaC (*recA*, *recB*, and *lon* mutants), where *dcm* improves survival but does not restore it to wildtype levels. In addition, homologous recombination assayed by gene conversion between *lacZ* alleles is induced strongly by 5-azaC but is largely *dcm*-independent; 5-azaC-induced deletion of 11 bp in *lacZ* is partially *dcm*-independent. Whatever the Dcm-independent lesion is, it must induce the SOS response since *sulA* entirely suppresses 5-azaC killing of *lon* mutants, whereas *dcm* suppresses it only partially. Dcm-independent killing of *recA* and *recB* mutants by 5-azaC suggests that the lesion must also lead to DSBs in DNA. 5-azaC can break down into guanylyurea (Figure 1), which in DNA mispairs with cytosine, leading

to C to G transversion mutations (Jackson-Grusby et al. 1997). In mammalian cells, this C to G transversion correlates with sites of C-methylation, leading to speculation that the CMeT induces the cytosine ring opening that converts it to guanylylurea (Jackson-Grusby et al. 1997). However, more recent work suggest that guanylylurea deoxribonucleotides can be directly incorporated into DNA during replication (Lamparska et al. 2012). Using a specific reporter for C to G transversion mutations (Cupples and Miller 1989; Seier et al. 2011), we confirmed that 5-azaC does indeed have a strong, 500-fold mutagenic effect on C to G transversion (confirming (Cupples and Miller 1989)), and occurs entirely independent of *dcm* (Table S1 and (Doiron et al. 1999)). Therefore, guanylylurea lesions may indeed be a potential contributor to *dcm*-independent effects of 5-azaC, where it may act like a abasic lesion to block DNA synthesis or elicit processing that leads to mutagenesis, template switching or recombination. It is also possible that there are additional unknown CMeTs encoded in the *E. coli* genome that are responsible for the *dcm*-independent effects.

4.7 Differential effects on template-switching: Although a template-switch mechanism has been proposed for both QPM at inverted repeats and deletions/expansion at direct repeats (reviewed in (Lovett 2017)), we saw a much stronger effect on the former than the latter. We assayed two types of deletions that arise between short tandem repeated sequences, 11 bp in the chromosomal *lacZ* gene, and 101 bp in the *tetA* gene on a ColEI plasmid and both were stimulated modestly, 2-4 fold after 2 -hr treatment with 5-azaC. In contrast, comparable treatment yielded stimulation of QPM over 2 orders of magnitude. We do not know the basis for this differential effect but further work may clarify the differences between these template-switching events.

5. Acknowledgments: We thank Hirotada Mori and the National Genetics Institute of Japan for the collection of *E. coli* knockout mutants. Funding: This work was supported by the National Institutes of Health, grants R01 GM51753 and P01 GM105473 to STL, T32 GM007122 to JAK and National Science Foundation REU grant DBI-1359172 to ELP.

TABLE 1: *E. coli* K-12 strains used in this study

Strain number	Genotype	
MGI655	<i>rph-1</i>	Bachmann 1972
STL1464	<i>lexA::Tn5</i>	Lab collection
STL7180	<i>recA::cat</i>	Lab collection
STL7428	<i>sulA::FRT kan</i>	Lab collection
STL12071	<i>lexA3 malF::Tn10 kan</i>	Lab collection
STL13814	<i>lacZ(C1384G) mphC281::Tn10</i>	Seier et al. 2011
STL14025	<i>lacZ(+11) kan mphC281::Tn10</i>	Seier et al. 2011
STL14095	<i>lacZ(-1G1474) mphC-281::Tn10</i>	Seier et al. 2011
STL14776	<i>lacZ(QP5) xonA::FRT xseA::FRT mphC-281::Tn10</i>	Seier et al. 2011
STL15144	<i>lacZ(QP5) lexA3 malF::Tn10 kan mphC281::Tn10</i>	Laranjo et al. 2018
STL15823	<i>lacZ(QP6) xonA::FRT xseA::FRT mphC-281::Tn10</i>	Laranjo et al. 2018
STL16519	<i>lacZ(QP6) lexA3 malF::Tn10 kan mphC281::Tn10</i>	Laranjo et al. 2018
STL17976	<i>lacZ(QP5) xonA::FRT kan mphC-281::Tn10</i>	Laranjo et al. 2018
STL17978	<i>lacZ(QP6) xonA::FRT kan mphC-281::Tn10</i>	Laranjo et al. 2018
STL17980	<i>lacZ(QP5) xseA::FRT kan mphC-281::Tn10</i>	Laranjo et al. 2018
STL17982	<i>lacZ(QP6) xseA::FRT kan mphC-281::Tn10</i>	Laranjo et al. 2018
STL18747	<i>lacZ(QP5) dcm::FRT kan mphC281::Tn10</i>	Laranjo et al. 2018
STL18749	<i>lacZ(QP6) dcm::FRT kan mphC281::Tn10</i>	Laranjo et al. 2018
STL19829	<i>lacZ(QP5) recA::cat mphC281::Tn10</i>	Laranjo et al. 2018
STL19831	<i>lacZ(QP6) recA::cat mphC281::Tn10</i>	Laranjo et al. 2018
STL19843	<i>recB::FRT kan</i>	Laranjo et al. 2018
STL19845	<i>recO::FRT kan</i>	This work
STL20589	<i>lacZ(QP5) mphC-281::Tn10</i>	resolate of STL17685 Laranjo et al. 2018

Strain number	Genotype	
STL20590	<i>lacZ(QP6) mphC-281::Tn10</i>	reisolate of STL15654, Seier et al. 2011
STL20956	<i>lacZ(QP5) dinB::FRT polB::FRT cat umuC::FRT mphC281::Tn10</i>	This work
STL20959	<i>lacZ(QP6) dinB::FRT polB::FRT cat umuC::FRT mphC281::Tn10</i>	This work
STL21382	<i>lacZ(QP5) recO::FRT kan mphC-281::Tn10</i>	This work
STL21384	<i>lacZ(QP6) recO::FRT kan mphC-281::Tn10</i>	This work
STL21742	<i>dcm::FRT kan</i>	This work
STL21845	<i>sulA::FRT</i>	This work
STL21944	<i>dcm::FRT kan recA::cat</i>	This work
STL21945	<i>lacZ(QP5) dcm::FRT kan recA::cat mphC-281::Tn10</i>	This work
STL21947	<i>lacZ(QP6) dcm::FRT kan recA::cat mphC-281::Tn10</i>	This work
STL22123	<i>lon::FRT kan</i>	This work
STL22147	<i>clpP::FRT kan</i>	This work
STL22165	<i>lacZ(QP5) dinB::FRT polB::FRT cat umuC::FRT mphC281::Tn10 xonA::FRT xseA::FRT kan</i>	This work
STL22169	<i>lacZ(QP5) dinB::FRT polB::FRT cat umuC::FRT mphC281::Tn10 xseA::FRT kan</i>	This work
STL22176	<i>lacZ(QP5) clpP::FRT kan mphC-281::Tn10</i>	This work
STL22178	<i>lacZ(QP6) clpP::FRT kan mphC-281::Tn10</i>	This work
STL22196	<i>lacZ(QP5) lon::FRT kan mphC-281::Tn10</i>	This work
STL22198	<i>lacZ(QP6) lon::FRT kan mphC-281::Tn10</i>	This work
STL22299	<i>lacZ(QP5) recB::FRT kan mphC-281::Tn10</i>	This work
STL22300	<i>lacZ(QP6) recB::FRT kan mphC-281::Tn10</i>	This work
STL22644	<i>lacZ(QP6) dcm::FRT recB::FRT kan mphC-281::Tn10</i>	This work
STL22655	<i>lacZ(QP5) dcm::FRT recB::FRT kan mphC-281::Tn10</i>	This work
STL22672	<i>lexA::Tn5 sulA::FRT</i>	This work

Strain number	Genotype	
STL22683	<i>lacZ(QP5) lexA::Tn5 kan sulA::FRT mphC-281::Tn10</i>	This work
STL22684	<i>lacZ(QP6) lexA::Tn5 kan sulA::FRT mphC-281::Tn10</i>	This work
STL22707	<i>lacZ(QP5) sulA::FRT mphC-281::Tn10</i>	This work
STL22709	<i>lacZ(QP6) sulA::FRT mphC-281::Tn10</i>	This work
STL22811	<i>recO::FRT kan dcm::FRT</i>	This work
STL22813	<i>lacZ(QP6) dcm::FRT recO::FRT kan mphC-281::Tn10</i>	This work
STL22824	<i>lacZ(QP5) dcm::FRT recO::FRT kan mphC-281::Tn10</i>	This work
STL22826	<i>lacZ(QP6) dinB::FRT polB::FRT cat umuC::FRT mphC281::Tn10 xseA::FRT kan</i>	This work
STL22864	<i>lacZ(QP6) dinB::FRT polB::FRT cat umuC::FRT mphC281::Tn10 xonA::FRT xseA::FRT kan</i>	This work
STL22869	<i>lacZ(+11) dcm::FRT kan mphC281::Tn10</i>	This work
STL13814	<i>lacZ(C1384G) dcm::FRT mphC281::Tn10</i>	Seier et al. 2011
STL22884	<i>dcm::FRT lon::FRT kan</i>	This work
STL22989	<i>sulA::FRT lon::FRT kan</i>	This work
STL22991	<i>lacZ(QP5) lon::FRT kan sulA::FRT mphC-281::Tn10</i>	This work
STL22993	<i>lacZ(QP6) lon::FRT kan sulA::FRT mphC-281::Tn10</i>	This work
STL23017	<i>lacZ(QP5) dcm::FRT lon::FRT kan mphC-281::Tn10</i>	This work
STL23019	<i>lacZ(QP6) dcm::FRT lon::FRT kan mphC-281::Tn10</i>	This work
STL23067	<i>lacZ(QP5) dcm::FRT kan sulA::FRT mphC-281::Tn10</i>	This work
STL23097	<i>lacZ(QP5) dcm::FRT kan lon::FRT sulA::FRT mphC-281::Tn10</i>	This work
STL23099	<i>lacZ(QP5) clpP::FRT kan dcm::FRT mphC-281::Tn10</i>	This work
STL23124	<i>lacZ(QP5) clpP::FRT kan dcm::FRT mphC-281::Tn10</i>	This work

Table SI. Frequency of C to G transversions with and without 5-azaC, in wild-type and dcm mutants, n= 4

	Mutation frequency x 10⁻⁸	
	wt	dcm
untreated	0.052	<0.05
100 µg/ml 5-azaC	25.3	25.6

Bibliography

- Al Mamun, A. A., M. J. Lombardo, C. Shee, A. M. Lisewski, C. Gonzalez *et al.*, 2012 Identity and function of a large gene network underlying mutagenic repair of DNA breaks. *Science* 338: 1344-1348.
- Anderson, D. G., and S. C. Kowalczykowski, 1997 The recombination hot spot chi is a regulatory element that switches the polarity of DNA degradation by the RecBCD enzyme. *Genes & development* 11: 571-581.
- Bannister, D., and S. W. Glover, 1968 Restriction and modification of bacteriophages by R+ strains of *Escherichia coli* K12. *Biochem Biophys Res Commun* 30: 735-738.
- Barker, S., M. Weinfeld and D. Murray, 2005 DNA-protein crosslinks: their induction, repair, and biological consequences. *Mutat Res* 589: 111-135.
- Bhagwat, A. S., B. Johnson, K. Weule and R. J. Roberts, 1990 Primary sequence of the EcoRII endonuclease and properties of its fusions with beta-galactosidase. *J Biol Chem* 265: 767-773.
- Bhagwat, A. S., and R. J. Roberts, 1987 Genetic analysis of the 5-azacytidine sensitivity of *Escherichia coli* K-12. *Journal of Bacteriology* 169: 1537-1546.
- Bhagwat, A. S., A. Sohail and R. J. Roberts, 1986 Cloning and characterization of the dcm locus of *Escherichia coli* K-12. *J Bacteriol* 166: 751-755.
- Bissler, J. J., 1998 DNA inverted repeats and human disease. *Front Biosci* 3: d408-418.
- Butala, M., S. J. W. Busby and D. J. Lee, 2009 DNA sampling: a method for probing protein binding at specific loci on bacterial chromosomes. *Nucleic Acids Res* 37: e37.
- Cooper, D. L., and S. T. Lovett, 2010 Toxicity and tolerance mechanisms for azidothymidine, a replication gap-promoting agent, in *Escherichia coli*. *DNA Repair (Amst)*.
- Cox, M. M., M. F. Goodman, K. N. Kreuzer, D. J. Sherratt, S. J. Sandler *et al.*, 2000 The importance of repairing stalled replication forks. *Nature* 404: 37-41.
- Craft, T. R., E. Bermudez and T. R. Skopek, 1987 Formaldehyde mutagenesis and formation of DNA-protein crosslinks in human lymphoblasts in vitro. *Mutat Res* 176: 147-155.
- Cupples, C. G., and J. H. Miller, 1989 A set of lacZ mutations in *Escherichia coli* that allow rapid detection of each of the six base substitutions. *Proc Natl Acad Sci U S A* 86: 5345-5349.
- Datsenko, K. A., and B. L. Wanner, 2000 One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A* 97: 6640-6645.
- Doiron, K. M., J. Lavigne-Nicolas and C. G. Cupples, 1999 Effect of interaction between 5-azacytidine and DNA (cytosine-5) methyltransferase on C-to-G and C-to-T mutations in *Escherichia coli*. *Mutat Res* 429: 37-44.
- Dower, W., J. Miller and C. Ragsdale, 1988 High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Research* 16: 6127-6145.
- Drlica, K., and X. Zhao, 1997 DNA gyrase, topoisomerase IV, and the 4-quinolones. *Microbiol Mol Biol Rev* 61: 377-392.
- Dutra, B. E., and S. T. Lovett, 2006 Cis and trans-acting effects on a mutational hotspot involving a replication template switch. *J Mol Biol* 356: 300-311.
- Fijalkowska, I. J., P. Schaaper Rm Fau - Jonczyk and P. Jonczyk, 2012 DNA replication fidelity in *Escherichia coli*: a multi-DNA polymerase affair.
- Friedman, S., 1985 The irreversible binding of azacytosine-containing DNA fragments to bacterial DNA(cytosine-5)methyltransferases. *J Biol Chem* 260: 5698-5705.
- Garaycochea, J. I., G. P. Crossan, F. Langevin, M. Daly, M. J. Arends *et al.*, 2012 Genotoxic consequences of endogenous aldehydes on mouse haematopoietic stem cell function. *Nature* 489: 571-575.
- Glickman, B. W., and L. S. Ripley, 1984 Structural intermediates of deletion mutagenesis: a role for palindromic DNA. *Proc Natl Acad Sci U S A* 81: 512-516.
- Goldfless, S. J., A. S. Morag, K. A. Belisle, V. A. Suter and S. T. Lovett, 2006 DNA repeat rearrangements mediated by DnaK-dependent replication fork repair. *Molecular Cell* 21: 595-604.

- Gottesman, S., 1996 Proteases and their targets in *Escherichia coli*. Annu. Rev. Genet. 30: 465-506.
- Greenblatt, M. S., A. P. Grollman and C. C. Harris, 1996 Deletions and insertions in the p53 tumor suppressor gene in human cancers: confirmation of the DNA polymerase slippage/misalignment model. Cancer Res 56: 2130-2136.
- Guzman, L., D. Belin, M. Carson and J. Beckwith, 1995 Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. J Bacteriol 177: 4121-4130.
- Hampsey, D. M., J. F. Ernst, J. W. Stewart and F. Sherman, 1988 Multiple base-pair mutations in yeast. J Mol Biol 201: 471-486.
- Ide, H., M. I. Shoukamy, T. Nakano, M. Miyamoto-Matsubara and A. M. Salem, 2011 Repair and biochemical effects of DNA-protein crosslinks. Mutat Res 711: 113-122.
- Izzotti, A., C. Cartiglia, M. Taningher, S. De Flora and R. Balansky, 1999 Age-related increases of 8-hydroxy-2'-deoxyguanosine and DNA-protein crosslinks in mouse organs. Mutat Res 446: 215-223.
- Jackson-Grusby, L., P. W. Laird, S. N. Magge, B. J. Moeller and R. Jaenisch, 1997 Mutagenicity of 5-aza-2'-deoxycytidine is mediated by the mammalian DNA methyltransferase. Proc Natl Acad Sci U S A 94: 4681-4685.
- Kim, S. R., G. Maenhaut-Michel, M. Yamada, Y. Yamamoto, K. Matsui *et al.*, 1997 Multiple pathways for SOS-induced mutagenesis in *Escherichia coli*: an overexpression of dinB/dinP results in strongly enhancing mutagenesis in the absence of any exogenous treatment to damage DNA. Proc Natl Acad Sci U S A 94: 13792-13797.
- Klaric, J. A., E. L. Perr and S. T. Lovett, 2020 Identifying Small Molecules That Promote Quasipalindrome-Associated Template-Switch Mutations in *Escherichia coli*. G3 (Bethesda).
- Krasich, R., S. Y. Wu, H. K. Kuo and K. N. Kreuzer, 2015 Functions that protect *Escherichia coli* from DNA-protein crosslinks. DNA Repair (Amst) 28: 48-59.
- Kuo, H. K., J. D. Griffith and K. N. Kreuzer, 2007 5-Azacytidine induced methyltransferase-DNA adducts block DNA replication in vivo. Cancer research 67: 8248-8254.
- Lal, D., S. Som and S. Friedman, 1988 Survival and mutagenic effects of 5-azacytidine in *Escherichia coli*. Mutat Res 193: 229-236.
- Lamparska, K., J. Clark, G. Babilonia, V. Bedell, W. Yip *et al.*, 2012 2'-Deoxyriboguanylurea, the primary breakdown product of 5-aza-2'-deoxyribocytidine, is a mutagen, an epimutagen, an inhibitor of DNA methyltransferases and an inducer of 5-azacytidine-type fragile sites. Nucleic Acids Res 40: 9788-9801.
- Laranjo, L. T., S. J. Gross, D. M. Zeiger and S. T. Lovett, 2017 SSB recruitment of Exonuclease I aborts template-switching in *Escherichia coli*. DNA Repair (Amst) 57: 12-16.
- Laranjo, L. T., J. A. Klaric, L. R. Pearlman and S. T. Lovett, 2018 Stimulation of Replication Template-Switching by DNA-Protein Crosslinks. Genes (Basel) 10.
- Lovett, S. T. (Editor), 2010 *The DNA damage response*. American Society of Microbiology Press, Washington, DC.
- Lovett, S. T., 2017 Template-switching during replication fork repair in bacteria. DNA Repair (Amst) 56: 118-128.
- Lovett, S. T., P. T. Drapkin, V. A. Sutura, Jr. and T. J. Gluckman-Peskind, 1993 A sister-strand exchange mechanism for recA-independent deletion of repeated DNA sequences in *Escherichia coli*. Genetics 135: 631-642.
- Lovett, S. T., and V. V. Feschenko, 1996 Stabilization of diverged tandem repeats by mismatch repair: evidence for deletion formation via a misaligned replication intermediate. Proc Natl Acad Sci U S A 93: 7120-7124.
- Lovett, S. T., T. J. Gluckman, P. J. Simon, V. A. Sutura, Jr. and P. T. Drapkin, 1994 Recombination between repeats in *Escherichia coli* by *arecA*-independent, proximity-sensitive mechanism. Mol Gen Genet 245: 294-300.

- Lovett, S. T., R. L. Hurley, V. A. Suttera, Jr., R. H. Aubuchon and M. A. Lebedeva, 2002 Crossing over between regions of limited homology in *Escherichia coli*. RecA-dependent and RecA-independent pathways. *Genetics* 160: 851-859.
- McKenzie, G. J., and N. L. Craig, 2006 Fast, easy and efficient: site-specific insertion of transgenes into enterobacterial chromosomes using Tn7 without need for selection of the insertion event. *BMC Microbiol* 6: 39.
- Nakano, T., X. Xu, A. M. H. Salem, M. I. Shoukamy and H. Ide, 2017 Radiation-induced DNA-protein cross-links: Mechanisms and biological significance. *Free Radic Biol Med* 107: 136-145.
- Nitiss, J. L., 2009 Targeting DNA topoisomerase II in cancer chemotherapy. *Nat Rev Cancer* 9: 338-350.
- Persky, N. S., and S. T. Lovett, 2008 Mechanisms of recombination: lessons from *E. coli*. *Crit Rev Biochem Mol Biol* 43: 347-370.
- Pommier, Y., E. Leo, H. Zhang and C. Marchand, 2010 DNA topoisomerases and their poisoning by anticancer and antibacterial drugs. *Chem Biol* 17: 421-433.
- Ringquist, S., and C. L. Smith, 1992 The *Escherichia coli* chromosome contains specific, unmethylated dam and dcm sites. *Proc Natl Acad Sci U S A* 89: 4539-4543.
- Ripley, L., 1982 Model for the participation of quasi-palindromic DNA sequences in frameshift mutation. *Proceedings of the National Academy of Sciences of the United States of America* 79: 4128-4132.
- Roberts, R. J., T. Vincze, J. Posfai and D. Macelis, 2015 REBASE--a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res* 43: D298-299.
- Salem, A. M., T. Nakano, M. Takuwa, N. Matoba, T. Tsuboi *et al.*, 2009 Genetic analysis of repair and damage tolerance mechanisms for DNA-protein cross-links in *Escherichia coli*. *J Bacteriol* 191: 5657-5668.
- Santi, D. V., A. Norment and C. E. Garrett, 1984 Covalent bond formation between a DNA-cytosine methyltransferase and DNA containing 5-azacytosine. *Proc Natl Acad Sci U S A* 81: 6993-6997.
- Schoemaker, J. M., R. C. Gayda and A. Markovitz, 1984 Regulation of cell division in *Escherichia coli*: SOS induction and cellular location of the *sulA* protein, a key to lon-associated filamentation and death. *J Bacteriol* 158: 551-561.
- Seier, T., D. R. Padgett, G. Zilberberg, V. A. Suttera, Jr., N. Toha *et al.*, 2011 Insights into mutagenesis using *Escherichia coli* chromosomal *lacZ* strains that enable detection of a wide spectrum of mutational events. *Genetics* 188: 247-262.
- Sharan, S. K., L. C. Thomason, S. G. Kuznetsov and D. L. Court, 2009 Recombineering: a homologous recombination-based method of genetic engineering. *Nat Protoc* 4: 206-223.
- Simmons, L. A., J. J. Foti, S. E. Cohen and G. C. Walker, 2008 The SOS Regulatory Network. *EcoSal Plus* 3.
- Swenberg, J. A., K. Lu, B. C. Moeller, L. Gao, P. B. Upton *et al.*, 2011 Endogenous versus exogenous DNA adducts: their role in carcinogenesis, epidemiology, and risk assessment. *Toxicol Sci* 120 Suppl 1: S130-145.
- Vaz, B., M. Popovic and K. Ramadan, 2017 DNA-Protein Crosslink Proteolysis Repair. *Trends Biochem Sci* 42: 483-495.
- Viswanathan, M., J. J. Lacirignola, R. L. Hurley and S. T. Lovett, 2000 A novel mutational hotspot in a natural quasipalindrome in *Escherichia coli*. *J Mol Biol* 302: 553-564.
- Wu, F. Y., Y. J. Lee, D. R. Chen and H. W. Kuo, 2002 Association of DNA-protein crosslinks and breast cancer. *Mutat Res* 501: 69-78.
- Yoshimori, R., D. Roulland-Dussoix and H. W. Boyer, 1972 R factor-controlled restriction and modification of deoxyribonucleic acid: restriction mutants. *J Bacteriol* 112: 1275-1279.
- Yoshiyama, K., K. Higuchi, H. Matsumura and H. Maki, 2001 Directionality of DNA replication fork movement strongly affects the generation of spontaneous mutations in *Escherichia coli*. *Journal of Molecular Biology* 307: 1195-1206.

Yoshiyama, K., and H. Maki, 2003 Spontaneous hotspot mutations resistant to mismatch correction in *Escherichia coli*: transcription-dependent mutagenesis involving template-switching mechanisms. *J Mol Biol* 327: 7-18.

Zahn, R. K., G. Zahn-Daimler, S. Ax, M. Hosokawa and T. Takeda, 1999 Assessment of DNA-protein crosslinks in the course of aging in two mouse strains by use of a modified alkaline filter elution applied to whole tissue samples. *Mech Ageing Dev* 108: 99-112.