



# Genome-wide strand-specific UV mutagenesis in *Escherichia coli* is directed by the Mfd translocase

Ogün Adebali<sup>a,b,c,1</sup> , Piotr A. Mieczkowski<sup>d</sup> , Aziz Sancar<sup>c,1</sup> , and Christopher P. Selby<sup>c,1</sup>

Affiliations are included on p. 7.

Contributed by Aziz Sancar; received August 23, 2025; accepted October 20, 2025; reviewed by Thomas J. Santangelo and Roel M. Schaaper

**Transcription-coupled repair in *Escherichia coli* which is mediated by the Mfd translocase is responsible for higher repair rate in *lacZ* and *lacI* genes upon induction of transcription. Here, we analyze the entire *E. coli* genome for the effect of Mfd on UV-induced mutagenesis. We find genome-wide preferential repair of the transcribed strand (TS) over the nontranscribed strand (NTS), and consequently, fewer mutations are caused by cyclobutane pyrimidine dimers in the TS than the NTS, in a manner proportional to transcription rate. In *mfd*<sup>−</sup> cells, most mutations are in the TS, caused by RNA polymerase stalled at template strand damage inhibiting repair. These findings are pertinent to *mfd*<sup>−</sup> phenotypes involving gene expression, recombination, stationary phase mutagenesis, and drug resistance.**

mutagenesis | transcription-coupled repair | DNA excision repair | Mutation frequency decline | UV

Ultraviolet (UV) light induces two major photoproducts in DNA, the cyclobutane pyrimidine dimer (CPD), and the pyrimidine (6-4) pyrimidone photoproduct [(6-4)PP]. In *Escherichia coli*, these damages are repaired by nucleotide excision repair, which requires the concerted action of proteins encoded by three genes that confer UV resistance (*uvrA*, *uvrB* and *uvrC*). The UvrA, UvrB, and UvrC subunits of the excision nuclease incise the damaged strand on both sides of the damage (1). Following dual incision, DNA helicase II (UvrD) facilitates removal of the 12- to 13-nucleotide-long damage-containing oligomer, as well as the repair subunits. The process is completed with repair synthesis by DNA Pol I and ligation by DNA ligase.

Early studies of UV damage response revealed that if irradiated cells were held in minimal media before plating, the frequency of certain suppressor mutations declined. This phenomenon, termed mutation frequency decline (Mfd), was found to be dependent upon excision repair, and a mutant was isolated in which no decline was observed. The gene responsible was mapped (2), and the encoded Mfd protein was isolated and found to be a translocase that interacts with both RNA polymerase stalled at damage in the transcribed strand (TS), and with UvrAB, and mediates rapid damage recognition and repair of transcription-blocking lesions in the TS (3–6).

This process of preferential excision repair of DNA damage in the template strand of an actively transcribed gene is known as transcription-coupled repair (TCR) (7). It occurs in eukaryotes and prokaryotes, and Mfd has been found to be necessary and sufficient for TCR in *E. coli* (6, 8). In *E. coli*, TCR occurs genome-wide at the same time as basal, transcription-independent repair but at a faster rate, on the order of 3- to 10-fold faster depending upon transcription rate. This enhanced rate is seen in the repair of CPDs, which are relatively slowly repaired by global repair. However, (6-4)PPs, which cause more pronounced helical distortions than CPDs, are recognized and repaired very rapidly by the basal repair pathway, apparently faster than RNA polymerase becomes blocked at (6-4)PPs, and thus TCR has no significant role in (6-4)PP repair (6, 9–11).

The reduction in suppressor mutations that is the hallmark of mutation frequency decline presumably results from rapid TCR of transcription-blocking CPDs in the anticodon-encoding region of certain tRNAs (12). In studies of drug resistance induced by DNA damage, forward mutations were found to be increased in *mfd*<sup>−</sup> cells, also consistent with a role of Mfd in repair (2). Interestingly, studies mapping individual forward mutations in the *lacI* gene revealed a strandedness in mutagenesis (13, 14). In wild-type cells, more mutations were observed in the nontranscribed strand (NTS) of *lacI*, presumably due to rapid TCR of the TS, and in *mfd*<sup>−</sup> cells, more mutations were observed in the TS, where RNA polymerase stably blocked at damage is known to inhibit repair of the damage (13). In this investigation, we have examined UV-mutagenesis genome-wide, in order to determine the generality of the strandedness of UV mutagenesis and the role of Mfd. Our findings extend to the entire genome the observations made in *lacI*, and

## Significance

Cytotoxic and mutagenic DNA-damaging agents are ubiquitous. Bulky DNA damage is removed genome-wide in bacteria by excision repair, and from transcribed strands (TS) by transcription-coupled repair (TCR), in which the Mfd protein targets transcription-blocking damage for accelerated excision repair. In analyzing UV-induced mutations in *Escherichia coli*, genome-wide, we find fewer TS mutations in active genes consistent with removal of premutagenic damage by TCR. *mfd*<sup>−</sup> cells exhibit more mutations in the TS, reflecting inhibition of repair by RNA polymerase that remains stalled at template strand lesions. These effects are dependent upon transcription level and are meaningful in characterizing the mutagenic phenotype associated with Mfd, which has additional, incompletely characterized roles in recombination, gene regulation, and drug resistance.

Author contributions: O.A., P.A.M., A.S., and C.P.S. designed research; P.A.M. and C.P.S. performed research; O.A. and P.A.M. analyzed data; and O.A., P.A.M., A.S., and C.P.S. wrote the paper.

Reviewers: T.J.S., Colorado State University; and R.M.S., National Institute of Environmental Health Sciences.

The authors declare no competing interest.

Copyright © 2025 the Author(s). Published by PNAS. This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

<sup>1</sup>To whom correspondence may be addressed. Email: oadebali@sabanciuniv.edu, aziz\_sancar@med.unc.edu, or christopher\_selby@med.unc.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2523368122/-/DCSupplemental>.

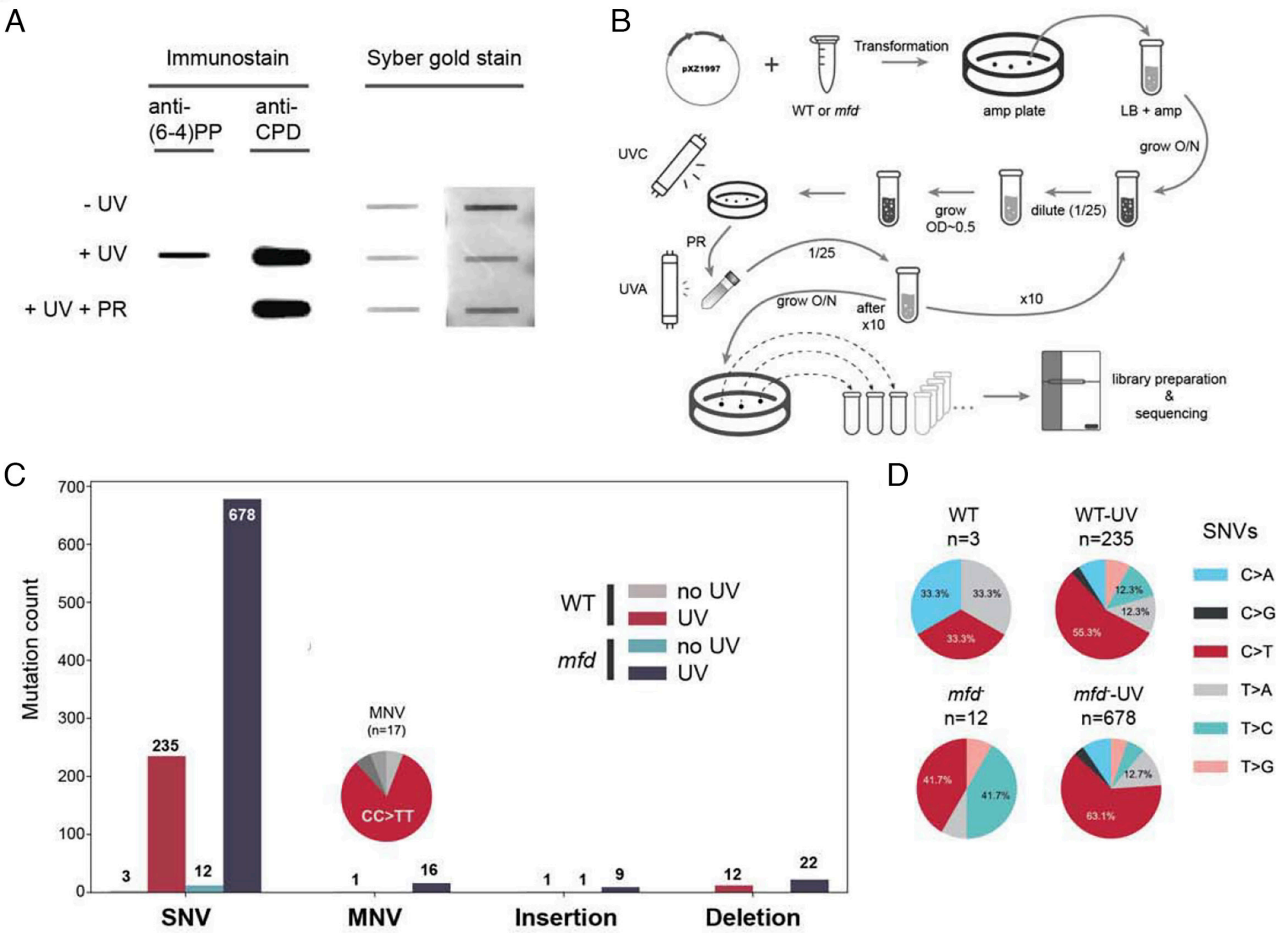
Published November 13, 2025.

furthermore indicate a positive association between the extent of this strandedness and transcription level, and also show that the strandedness is indeed due to changes in mutagenesis of the TS.

Results

**Generation of Mutants.** With the goal of obtaining substantial numbers of UV-induced mutations suitable for genome-wide strand-bias analysis, we employed a procedure of repeated rounds of UV irradiation of mid-log phase cultures, with cell dilution and regrowth after each round, to allow replication and fixation of mutations, as has been done previously with *E. coli* (15). In addition, of the two major UV photoproducts, the pyrimidine (6-4) pyrimidone (6-4) photoproducts [(6-4)PPs] and CPDs, only the latter is subject to TCR (9). Therefore, in this investigation, we removed (6-4)PPs by selective photoreactivation. To do this, we used wild-type and *mfd*<sup>-</sup> cells with the genomic CPD photolyase gene deleted (16), and in addition, cells were transformed with a plasmid that expresses (6-4)PP photolyase. Results of testing a wild-type clone for selective (6-4)PP photoreactivation are shown in Fig. 1A.

The overall procedure, illustrated in Fig. 1B, involved UV irradiation of mid-log phase cells on ice. We used a dose of 75 J/m<sup>2</sup>, comparable to the 72 J/m<sup>2</sup> dose employed in a previous relevant mutagenesis study (13). Following UV irradiation, cells were photoreactivated on ice. Cells were then diluted and grown overnight in selective media to allow repair, replication, and fixation of mutations. Cells were then diluted and grown to mid-log phase in a low level of isopropyl-β-D-thiogalactopyranoside (IPTG) (30 μM) to induce (6-4)PP photolyase expression and then irradiated again. Ten rounds of irradiation were completed, and unirradiated controls were processed in parallel. Following 10 rounds, overnight cultures were plated, and individual colonies were grown for DNA isolation and sequencing. For wild-type cells, 22 clones (-UV) and 24 clones (+UV) were processed, and for *mfd*<sup>-</sup> cells, 48 clones (-UV) and 48 clones (+UV) were processed. Samples of DNA from the original transformed cells were also isolated for sequencing and represent “unmutated” or “round zero” samples. Note that although the cells were *phr*<sup>-</sup> and bore a plasmid-encoded (6-4)PP photolyase gene, for simplicity the two strains used in this report are referred to as wild-type and *mfd*<sup>-</sup> since the intention is to examine the role of



**Fig. 1.** Removal of (6-4) photoproducts, UV-mutagenesis procedure, and initial mutation characterization. (A) Selective photoreactivation of (6-4)PPs. Samples were loaded in duplicate, and the membrane was cut in half to immunostain separately with the anti-(6-4)PP and anti-CPD antibodies. Damage is absent in unirradiated cells, and photoreactivation of (6-4)PPs was successful in these cells that lack CPD photolyase. (B) Procedure for mutation induction. Plasmid pXZ1997 expressing (6-4)PP photolyase was transformed into wild-type (WT) and *mfd*<sup>-</sup> cells. A (6-4)PP photolyase proficient clone of each was grown overnight in selective LB. The culture was then subject to 10 rounds of growth to mid-log phase, UV irradiation, and photoreactivation, then dilution, and overnight growth. Following 10 rounds, cultures were plated and single colonies were grown for isolation and sequencing of the genome. (C) Mutation counts by strain and UV irradiation status. Bar plots show the number of single-nucleotide variants (SNVs), multinucleotide variants (MNVs), insertions, and deletions identified in wild-type and *mfd*<sup>-</sup> strains with and without UV treatment. A pie chart shows the MNV composition among UV-treated samples (1 WT and 16 *mfd*<sup>-</sup>). (D) Distribution of SNV types by condition. Pie charts represent the relative proportions of six SNV categories (in context of pyrimidine to purine) for each sample group: WT (no UV), WT+UV, *mfd*<sup>-</sup> (no UV), and *mfd*<sup>-</sup>+UV. The designation of SNVs as pyrimidine to purine identifies the base originally damaged because CPDs [and (6-4)PPs] form at dipyrimidines. Thus, both C > T and G > A variants in the raw dataset are designated as C > T since the C (in the case of G > A, the C opposite G) was the damaged base that led to the variant sequence following replication.

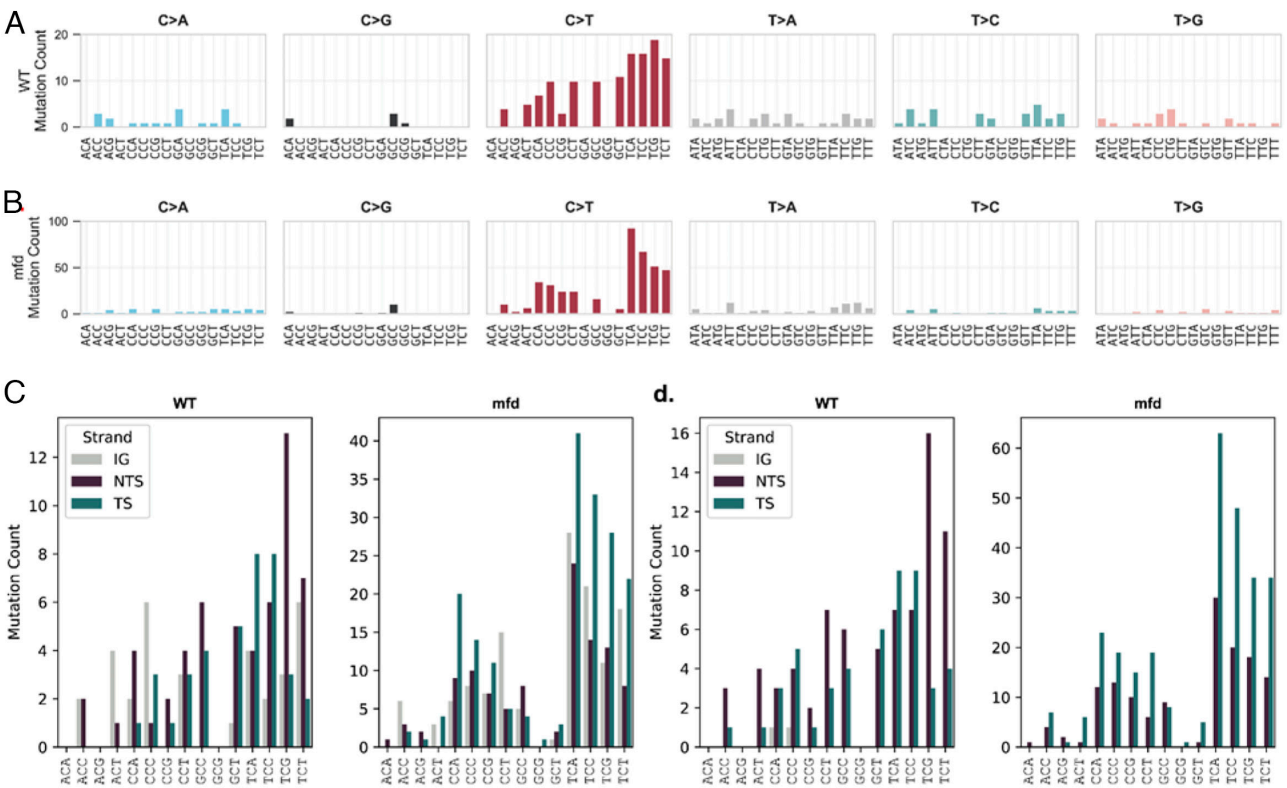
Mfd in mutagenesis. In preliminary experiments, under the conditions described above, UV survival in wild-type cells was 39% and 5% following 50 and 100 J/m<sup>2</sup>, respectively. Following a dose of 75 J/m<sup>2</sup>, survival was 15% (wild type) and 1% (*mfd*<sup>-</sup>), comparable to previously reported values (13). Following photoreactivation of (6-4)PPs, survival was 37% (wild type) and 7% (*mfd*<sup>-</sup>).

**Numbers, Characteristics, and Location of Mutations Obtained.**

Sequence variants were identified in genomic DNA, but not in the plasmid bearing the (6-4)PP photolyase gene. Dataset S1 lists the variants identified along with associated characteristics. Fig. 1C shows the genomic DNA variants sorted by mutation type. The number of UV-independent mutations comprised a modest background compared to UV-induced mutations. UV-induced single-base substitutions were by far the predominant mutation type and were the subject of further analysis. Adjusting for the number of clones sequenced (*mfd*<sup>-</sup>, 48; wild type, 24), there were 1.44× more mutations in *mfd*<sup>-</sup> cells as compared to wild type, a trend previously seen in studies that employed different methods and endpoints (2, 13). The pie charts in Fig. 1D show that C to T transitions were the most abundant UV-induced substitution in both strains. Fig. 2 shows the frequency of the different single base substitutions as a function of sequence context for wild-type (Fig. 2A) and *mfd*<sup>-</sup> (Fig. 2B) cells. The pattern of C to T substitutions was generally similar between wild-type and *mfd*<sup>-</sup> cells and are almost entirely located in sequences with adjacent dipyrimidine target sites, which are necessary for photoproduct formation. These C to T mutations in Fig. 2A and B were analyzed in the following panels of Fig. 2 and in all subsequent figures. In

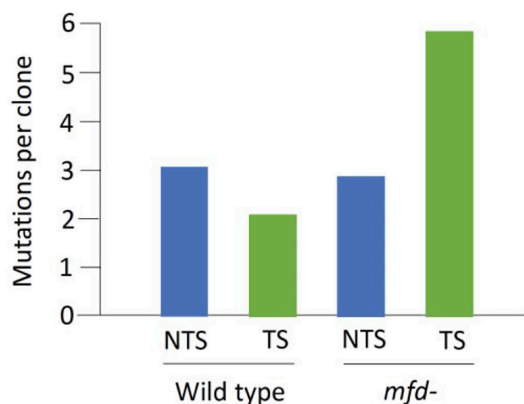
Fig. 2C and D, the C to T substitutions in each sequence context are plotted, with respect to the strand in which the premutagenic damage was located. In Fig. 2C, mutation sites are plotted as TS, NTS, or intergenic according to their genomic annotation. The plot shows that numerous sites are located in intergenic regions. The abundance of mutations in intergenic regions may be partly due to their selection during the mutagenesis procedure as sites that are not deleterious to the cell. Since practically the entire genome is transcribed to some degree, it was possible to designate most of the mutations in intergenic regions using RNA-seq data available for the wild-type and *mfd*<sup>-</sup> cells (9). This was done in Fig. 2D by designating each mutation as TS if it was in a strand transcribed more than the opposing strand, and NTS if transcribed less. Importantly, Fig. 2C and D both illustrate a trend toward relatively more TS mutations in *mfd*<sup>-</sup> cells, and possibly fewer TS mutations in wild-type cells. These trends in strandedness of mutagenesis are more clearly illustrated in Fig. 3 by plotting the number of mutants per clone sequenced by strand and strain, and the trends seen here genome-wide confirm and extend results previously seen in *lacI*.

**Strandedness as a Function of Transcription Level.** Fig. 3 implies a role of transcription on mutagenesis, so in Fig. 4A we plotted each mutation as a dot according to its associated transcription level (x-axis), strand, and strain (y-axis). Importantly, mutations in the TS show substantial differences in their associated transcription level distributions in both wild-type and *mfd*<sup>-</sup> cells. Going left to right, in wild-type cells, with a higher transcription level, a scarcity of mutations is evident. In contrast,



**Fig. 2.** Sequence context and strand distribution of UV-induced mutations in wild-type and *mfd*<sup>-</sup> strains. (A and B) Mutation counts for each SNV type in wild-type (WT) (A) and *mfd*<sup>-</sup> (B) cells are displayed by trinucleotide sequence context. The mutated base is centered, and flanking bases are shown on either side. Separate panels correspond to the six possible base substitution types. The colors shown in Fig. 1D are retained in Fig. 2A and B. (C) Mutation counts in selected C > T contexts categorized by their genomic annotation: intergenic (IG, gray), NTS (purple), and TS (teal) in wild-type and *mfd*<sup>-</sup> strains. (D) Mutation counts for the same sequence contexts as in (C), but using transcriptional strand designation derived from RNA-seq expression data derived from WT and *mfd*<sup>-</sup> strains (9). Sites were assigned as TS or NTS based on the strand with higher expression, with remaining sites classified as intergenic (IG). Mutation counts are plotted separately for wild-type and *mfd*<sup>-</sup> samples.





**Fig. 3.** UV-induced C to T mutations sorted by strand and strain. In wild type cells there were 75 (NTS) and 50 (TS) C to T mutations, and in *mfd*<sup>-</sup> cells there were 138 (NTS) and 283 (TS) C to T mutations. 24 wild-type clones and 48 *mfd*<sup>-</sup> clones were sequenced, and the number of mutations per clone is plotted. Strands were designated as in Fig. 2D using the terms TS and NTS to describe the mutations based upon the strand with the higher transcription level, although some of the mutations were in fact in regions annotated as intergenic.

in the TS of *mfd*<sup>-</sup> cells, there is a less perceptible but significant overrepresentation of high transcription-associated mutations. These trends are apparent and significant when comparing the wild-type and *mfd*<sup>-</sup> TS data, and also when comparing the TS with NTS data for each strain. Median values for transcription level of mutations (vertical purple bars) also demonstrate these trends. In contrast, the NTS data points for the two strains appear similarly distributed, are not significantly different, and their median values are comparable.

To quantify these data, mutations were binned into quartiles based upon the distribution of NTS mutations, which appear to be unbiased, and the numbers of NTS and TS mutations in each of these quartiles are plotted in Fig. 4B and C. The corresponding TS/NTS ratios were calculated per quartile and are plotted in Fig. 4D and E. The lowest transcription levels are in quartile 1 and the highest are in quartile 4. The plots illustrate the strong association between transcription level and abundance of TS mutations.

The transcriptional bias in TS mutations was then modeled by first plotting in Fig. 4F and G the cumulative fraction of mutations in the TS (y-axis) as a function of decreasing transcription level left to right on the x-axis. The first four dots on the left side of panel F are plotted at zero indicating that these four mutations, associated with the four highest levels of transcription, were all in the NTS (0% cumulative TS). The fifth dot plotted at 0.2 = 1/5 indicates that the mutation associated with the fifth highest transcription level was located in the TS (20% cumulative TS). Thus plotted, the data overall show the association between high transcription level (to the left) and low TS mutation in wild-type cells (panel F), and high TS mutation in *mfd*<sup>-</sup> cells (panel G). Curves were fitted to the data points, and the data were modeled by plotting in Fig. 4H the ratio of the two curves in panels F and G as TS mutation in *mfd*<sup>-</sup>/wild-type cells. This plotted ratio thus models the overall strand switch in mutation between the two cell strains as a function of transcription. From Fig. 4I, the frequency distribution plot showing the number of mutations observed at each transcription level, it can be seen from the data to the left of the arbitrarily drawn vertical dotted line that 16% of mutation sites (panel I) are modeled to be responsible for 20-fold and higher (y-axis, panel H) differences in mutation incidence. This illustrates the genomic region in which the largest variations in mutation incidence occur between the two cell strains.

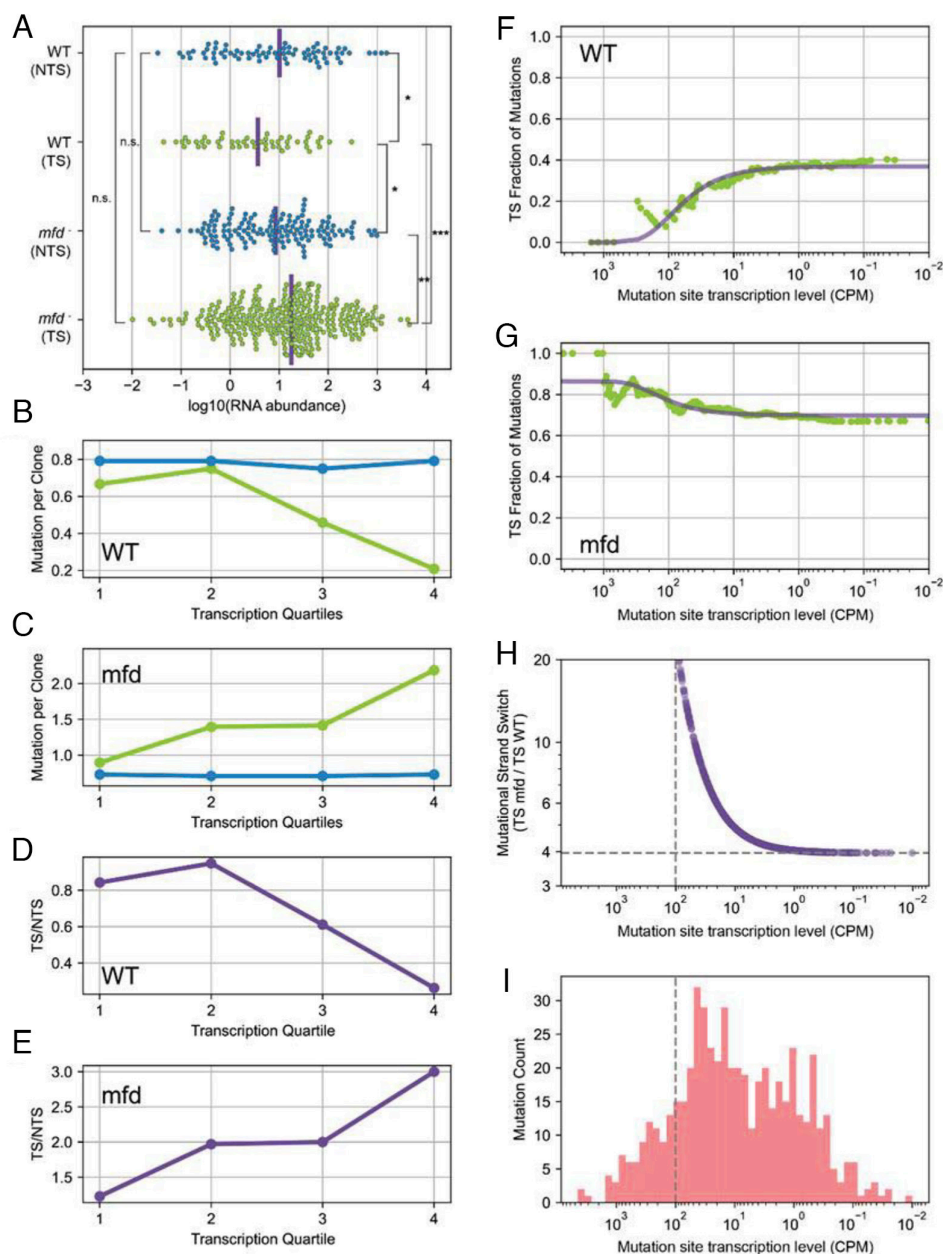
## Discussion

This investigation substantially advances our knowledge about strandedness of mutagenesis in *E. coli*. The iterative steps of irradiation, replication, and mutation fixation applied here successfully produced sufficient numbers of mutations to detect significant differences in the distribution of mutations as a function of transcription level in both wild-type and *mfd*<sup>-</sup> cells. Both strains demonstrated a unique phenotype attributable to transcription and repair dynamics. As depicted in Fig. 5, in highly transcribed regions, when RNAP is blocked by TS damage, in wild-type cells TCR follows, leading to fewer TS mutations, and in *mfd*<sup>-</sup> cells RNAP remains stably bound and inhibits repair, leading to more TS mutations. Damage in the NTS does not impede RNAP (17), and the data suggest that NTS repair and mutagenesis are unaffected by transcription.

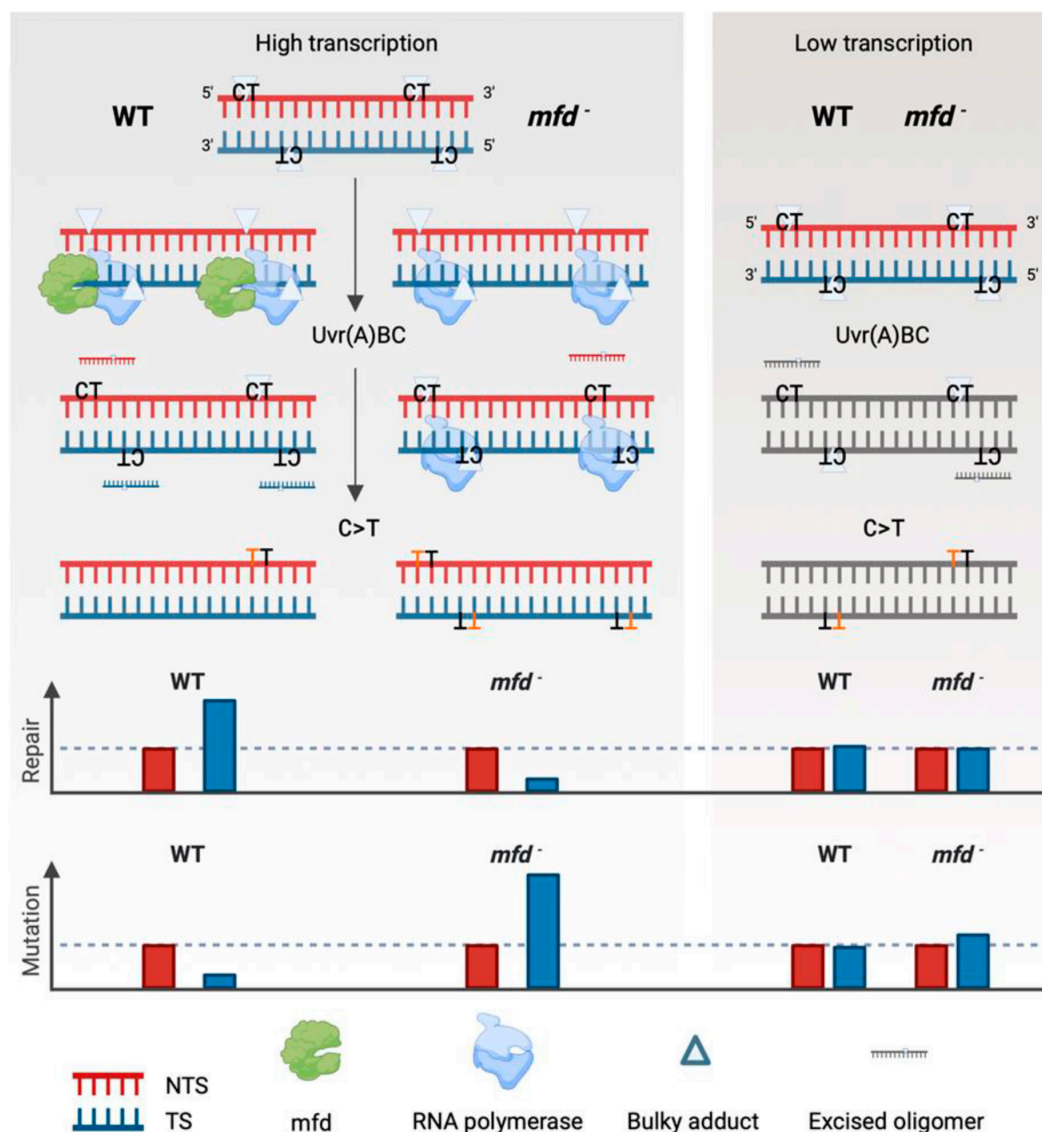
At the outset of this investigation, it was unclear whether transcription- and Mfd-dependent mutagenic phenotypes, while anticipated, would actually be observed. To be consistent with prior studies, cells were grown in rich broth in which case many biosynthetic genes are likely quiescent. Also, it was not possible to estimate with certainty the number of mutations to expect, so multiple rounds of irradiation and regrowth were employed to boost variant numbers. The results in Fig. 3 do in fact show the anticipated trends in strandedness, namely, few mutations in the TS of wild-type cells and more mutations in the TS of *mfd*<sup>-</sup> cells. These trends are consistent with prior mutagenesis studies in *lacI* (13, 14) and thus compelling in the aggregate. Importantly, after sorting the results by transcription level, the number of mutations in the TSs and NTSs varied by as much as approximately twofold to fourfold, and the trends in mutation as a function of transcription level were statistically significant in both wild-type and *mfd*<sup>-</sup> cells. For comparison, the prior study of mutation in episome-located *lacI* controlled by the strong *lacI*<sup>Q</sup> promoter found a 3.2-fold bias in favor of mutation in the TS in wild-type cells, and a 4.5-fold bias in favor of mutations in the NTS. It is notable that while the *lacI* study and prior repair studies have demonstrated strand biases in mutation and repair in terms of ratios, this study shows that Mfd- and transcription level-dependent changes in mutations occur specifically in the template strand, as illustrated in Fig. 5.

The mutagenesis and repair data taken together indicate that TCR of the TS occurs as a function of transcription level simultaneously with global repair of essentially all (6-4)PPs, and many CPDs in low or nontranscribed DNA. As previously discussed, this distribution of repair with most repair apparently by the global pathway is consistent with the limited UV sensitivity of *mfd*<sup>-</sup> cells. In *mfd*<sup>-</sup> cells, the disposition of RNA polymerase stalled at damage is an important consideration since it could become an impediment to replication and contribute to lethality. A possible mitigating factor is the Rho hexamer, which is involved in transcription and transcription termination (18, 19), and has been reported to remove RNAP stalled at DNA damage, and may thus contribute to survival of irradiated *mfd*<sup>-</sup> cells (20). Interestingly, a comparison of transcription termination by Rho and Mfd, which has been shown to remove RNAP stalled by several mechanisms including nucleotide starvation, DNA-bound protein, and DNA damage, showed that Mfd did not terminate transcription at Rho-dependent or -independent sites (21), consistent with the lethality of *rho* deletion in many bacteria.

Mfd has been reported to have multiple functions in the cell (6). The ability of Mfd to remove RNAP blocked by various impediments other than DNA damage has been well documented, and regulation of gene expression in certain instances may follow from the removal of elongating RNAP blocked by DNA-binding



**Fig. 4.** Distribution of UV mutations as a function of strand and transcription level in wild-type and *mfd* cells. (A) Swarm plot showing the transcription level (x-axis) associated with each C > T mutation (in dipyrimidine context), categorized by strand (transcribed strand [TS], green; nontranscribed strand [NTS], blue) and genotype (wild-type and *mfd*), y-axis). Each point represents an individual mutation. Vertical purple bars indicate the median transcription level of mutations in each group. Statistical comparisons between groups were performed using the two-sided Mann-Whitney *U* test. Significance levels are indicated as follows:  $P = 0.00904$  (\*\*): *mfd* TS vs. *mfd* NTS;  $P = 0.142$  (n.s.): *mfd* TS vs. WT (wild type) NTS;  $P = 7.43 \times 10^{-5}$  (\*\*\*): *mfd* TS vs. WT TS;  $P = 0.735$  (n.s.): *mfd* NTS vs. WT NTS;  $P = 0.0225$  (\*): *mfd* NTS vs. WT TS;  $P = 0.0297$  (\*): WT NTS vs. WT TS. (B and C) C > T mutation counts per clone of NTS (blue) and TS (green) mutations per quartile of transcription level in wild-type (B) and *mfd* (C) strains. Quartiles were defined based on the distribution of transcription levels of NTS mutations. The lowest transcription levels are in quartile 1 and the highest are in quartile 4. (D and E) Ratios of TS to NTS mutations per transcription quartile (based on the NTS mutations) in wild-type (D) and *mfd* (E) samples. (F and G) Cumulative fraction of mutations (in green) located on the TS plotted as a function of decreasing transcription level. For each plot, each fraction is plotted as a point in order from highest to lowest transcription level (left to right on the x-axis), and the cumulative fraction of TS mutations is calculated incrementally (y-axis). Each data point is plotted on the y-axis as the average fraction of mutations occurring in the TS by averaging the fraction of TS mutations for each data point and all data points to the left of it. Thus, the first data point on the Left in panel F indicates that the one mutation at the most highly transcribed site was in the NTS, and the last data point on the Right indicates that about 40% of all mutations were in the NTS. High variability is seen toward the left where each additional data point value can substantially affect the average. Fitted lines (in purple) modeled through logistic regression represent a model of the cumulative trend.  $R^2$  values for WT and *mfd* are 0.905 and 0.7852, respectively. (H) Ratio of cumulative TS mutation logistic regression curves from panels (F and G), representing the relative increase in TS mutations in *mfd* compared to wild-type with increasing transcription level (increasing right to left). The vertical dashed line indicates a transcription level threshold above which the TS to NTS mutation ratio exceeds 20, an arbitrarily chosen cutoff to highlight regions with extreme strand bias between the two strains. The y-axis was truncated at this threshold to accommodate the asymptotic nature of the ratio curve, which diverges toward infinity as TS mutations dominate in highly transcribed regions. The horizontal dashed line represents the genome-wide average TS/NTS mutation ratio, serving as a baseline independent of transcription level. (I) Histogram showing the frequency distribution of transcription levels at mutation sites in both wild-type and *mfd* strains. The vertical dashed line marks the transcription level threshold above which a 20-fold difference in TS mutation incidence was observed, corresponding to the top 16% of C > T mutation sites by expression level.



**Fig. 5.** Role of transcription level, blocked RNAP, and TCR on mutational strand bias following induction of CPD damage in DNA. The formation of TT mutations at CT sites of UV damage are illustrated schematically in the top portion of the illustration and plotted quantitatively below along with quantitative representations of repair. Icons are defined at the bottom. With low transcription (*Right*), there is little to no blockage of RNAP or TCR, UvrABC excision nuclease repairs both strands to a comparable level, and residual damage leads to comparable levels of mutations in each strand. With high transcription, RNAP is blocked by damage in the TS but not the NTS. In wild-type (WT) cells, damage that blocks RNAP is preferentially removed by TCR resulting in fewer mutations in the TS, while in *mfd*<sup>−</sup> cells, stably blocked RNAP delays repair of TS damage, leading to more mutations in the TS.

proteins, DNA structures, or other factors (22). A role in recombination in *B. subtilis* has also been described (23). Interestingly, damage-independent mutagenesis has been associated with Mfd, which has been referred to as an evolvability factor that may contribute toward drug resistance (24). The role of Mfd in some of these phenotypes is not well characterized and perhaps the findings reported here can aid research in these important areas.

It is possible that the strandedness in UV mutagenesis observed in *E. coli* will also be seen in eukaryotes, though eukaryotic transcription-repair coupling factors (CSA and CSB) and excision repair proteins do not share homology with the corresponding bacterial proteins and the reactions mechanisms differ (6, 7). Excision repair in archaea has not been well characterized, though the archaeal RNA polymerase, as is the case with bacterial and mammalian RNA polymerases, is stalled by TS but not NTS damage, and a putative archaeal transcription-repair coupling factor, eta, has been found that is capable of removing stalled RNA polymerase, and aids in UV-resistance (25, 26). Thus,

eta-dependent strandedness in mutagenesis may be found to occur in archaea.

## Methods

**Cells.** For wild-type cells we used CPD photolyase-deficient MGP (MG1655 *phr::kan*). *mfd*<sup>−</sup> cells were obtained by deleting the *kan* gene from MGP to make MPd and then deleting the *mfd* gene from MPd to make MPdM (MPd *mfd::kan*) (16). Plasmid pXZ1997, which expresses *Drosophila* (6-4)PP photolyase fused to maltose binding protein (27), was linearized with EcoRI, religated, and transformed into wild-type and *mfd*<sup>−</sup> cells with selection on ampicillin plates. Isolated transformants were grown, frozen stocks were made of each, and photoreactivation capacity of clones was confirmed as described below.

**Generation of Mutants.** Streaks were made from the above stocks, and a wild-type and an *mfd*<sup>−</sup> clone were grown overnight. From wild-type and *mfd*<sup>−</sup> cultures, samples representing "Round 0 -UV cells" were frozen and stored at −80 °C for isolation of DNA at a later date. Samples of wild-type and *mfd*<sup>−</sup> overnight cultures were also used to start 10 rounds of UV/control treatments to generate mutants.



For round 1, cultures were inoculated at 1/25 dilution into LB +amp (200 µg/mL) +IPTG (30 µM) and grown to an OD of approximately 0.5. Cultures were placed on ice, pelleted, and washed and resuspended in ice-cold PBS. 6.7 mL volumes of cells in R-100 tissue culture dishes were irradiated on ice with 75 J/m<sup>2</sup> of UVC. Cells were then transferred to ice-cold 15 mL centrifuge tubes and irradiated for 90 min with UVA light filtered through a glass plate and the tubes, with the tubes floating in ice water mixed occasionally. Irradiated and control unirradiated cultures were then diluted into fresh LB +amp (200 µg/mL) +kan (10 µg/mL) and grown overnight. The second round was then initiated by inoculating LB +amp +IPTG as in round 1. After 10 rounds, overnight cultures were spread on rifampicin (200 µg/mL) plates. Individual colonies were grown in approximately 1 mL +amp +kan and DNA was isolated from each using the QIAamp DNA Mini Kit (31506). DNA was isolated from two round 0 samples of wild-type cells and *mfd*<sup>−</sup> cells. The number of DNA samples isolated for sequencing after 10 rounds were, for wild-type cells, 22 (−UV) and 24 (+UV), and for *mfd*<sup>−</sup> cells, 48 (−UV) and 48 (+UV). DNA concentrations were measured using the Qubit broad range system. In planning the experiment, not knowing whether a sufficient number of genomic mutations would be obtained, we decided to include rifampicin selection in the final round to insure that some mutations would be obtained, and in addition as an endpoint for troubleshooting if needed. Practically each clone in fact did have a mutation in the *rpoB* gene, as expected, however, since the selection process may influence the *rpoB* mutations obtained, examination of these was not followed up as sufficient mutations were obtained genome-wide for our analysis.

**Assessment of Photoreactivation by Slot Blot.** Five or more mL volumes of cells prepared and irradiated (or control unirradiated) as described above were used for slot-blot analysis of CPD and (6-4)PP levels in DNA. After irradiation, cells were cooled on ice, pelleted, and transferred to an Eppendorf tube on ice. 330 µL of cold TE was added, and cells were resuspended. 40 µL of sodium dodecyl sulfate (SDS) (10%) was added and mixed gently with the cells, which were then incubated at room temperature for 20 to 25 min. 100 µL of NaCl (5 M) was added, and after mixing, samples were incubated at 4 °C overnight. Samples were centrifuged for 1 h at 4 °C. DNA was purified from pellets using the QIAamp DNA Mini Kit. During the incubation with kit buffer ATL plus proteinase K, clumps were resuspended by pipetting. After resuspending the purified DNA in kit elution buffer, 4 µL RNaseA was added and samples were incubated for 1 h at 37 °C. Samples were then purified using the QIAquick PCR Purification Kit (28106). DNA concentration was measured by Nanodrop. DNA was then mixed with 12 µL of DNA Mini Kit elution buffer and water for 250 ng total in 250 µL. DNA was then melted at 94 °C for 10 min, cooled in ice water, and mixed with cold 250 µL ammonium acetate (2 M). A slot blot apparatus was then used to apply the DNA to a nitrocellulose membrane (GE Healthcare Hybond ECL RPN3030D). The DNA was fixed to the membrane by heating to 80 °C in a vacuum oven for 90 min. The membrane was blocked with 5% milk, washed and probed overnight with anti-(6-4)PP antibody (CosmoBio NM-DND-002) at 1/4,000 dilution, or anti-CPD antibody (CosmoBio NM-DND-001) at 1/8,000 dilution. After washing, blots were incubated in secondary horseradish peroxidase linked sheep anti-mouse antibody from Cytiva at 1/5,000 dilution for at least an hour, washed, and developed using the BioRad Clarity ECL substrates. Following immunostaining, membranes were stained for DNA content with syber gold (Invitrogen S11494) at 1/20,000 dilution. Imaging was with a Bio-Rad ChemiDoc Imaging System.

**Identification of Mutations.** Whole-genome DNA libraries were prepared from bacterial samples using 80 ng of input DNA and the Watchmaker DNA Library Preparation Kit with Fragmentation (Watchmaker Genomics), incorporating stubby adapters and 10-nt Unique Dual Index primers from Integrated DNA Technologies. The resulting libraries had a median insert size of 224 bp (±66 bp SD). Prepared libraries were pooled and converted for sequencing using the Complete Genomics platform and sequenced on a DNBSEQ-T7 flowcell (paired-end 2× 150 bp;

Complete Genomics). Demultiplexed FASTQ files were imported into CLC Genomics Workbench v25.02 (Qiagen) for analysis. After quality control, reads were trimmed to remove adapter sequences. Trimmed reads were then aligned to the *E. coli* reference genome (ASM1942v1) using standard alignment settings, with nonspecific matches ignored. PCR duplicates were removed from the aligned reads. All samples achieved >100× genome coverage. Variant calling was performed using the Fixed Ploidy Variant Detection 2.6 tool with standard parameters for haploid genomes and a minimum variant frequency threshold of 50%. To exclude background mutations, variants present in control samples (round 0 for wild-type and *mfd*<sup>−</sup> strains) were removed, retaining only those with control read counts below 10 (CLC v1.6). Variants were then grouped into four categories: wild-type and *mfd*, with or without UV treatment. Identified variants were annotated for gene orientation, amino acid changes, and flanking sequences using CLC's annotation tools.

**Strand-of-Origin Inference.** Although a fixed SNV is present on both DNA strands after replication, the mutation type still reveals the strand of origin of the lesion. A C→T transition indicates that the cytosine on that strand was the damaged base, whereas a G→A substitution reflects damage to the cytosine on the opposite strand. By combining this base-specific information with gene orientation and RNA-seq-defined transcription direction, we can assign each mutation to the transcribed or NTS, despite the eventual symmetry of the mutation in double-stranded DNA postreplication.

**RNA-seq Data Processing and Quantification.** Raw RNA-seq data for *E. coli* wild-type and *mfd*<sup>−</sup> strains were retrieved from the Sequence Read Archive (SRA) under accession numbers SRR30505829 and SRR30505832 (wild-type), and SRR30505866 and SRR30505841 (*mfd*<sup>−</sup>). Paired-end reads were downloaded using fasterq-dump and quality-trimmed with Trimmomatic (28), applying adapter removal and quality filters with the following parameters: ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36.

Trimmed reads were aligned to the *E. coli* reference genome (ASM1942v1) using STAR (v2.7) (29). The genome index was built prior to alignment using a splice junction database overhang of 99. Output BAM files were sorted by coordinate and filtered using samtools to retain only properly paired alignments.

Strand-specific genome-wide coverage profiles were generated using deepTools (30) bamCoverage, with CPM (counts per million) normalization and a bin size of 1 bp. Forward and reverse strand coverages were exported separately as bedGraph files for downstream analyses and visualization. All steps were executed using a reproducible Snakemake (31) workflow, with isolated software environments managed through Conda.

To quantify strand-specific transcription levels at each mutation site, we parsed the strand-specific bedGraph coverage files for each replicate. Signal values, in Dataset S1, were averaged across replicates for wild-type and *mfd*<sup>−</sup> samples separately. For each mutation site, the higher-expressed strand was designated as the TS, and the lower-expressed strand as the NTS. The strand difference was quantified as the absolute difference in CPM between TS and NTS at each site.

**Data, Materials, and Software Availability.** The dataset is deposited to the SRA database (PRJNA1302349) (32).

**ACKNOWLEDGMENTS.** This work was funded by the NIH (GM118102 and ES0033414 to A.S.).

Author affiliations: <sup>a</sup>Molecular Biology, Genetics and Bioengineering Program, Faculty of Engineering and Natural Sciences, Sabanci University, Istanbul 34956, Türkiye; <sup>b</sup>Department of Computational Science-Biological Sciences, Scientific and Technological Research Council of Türkiye Research Institute for Fundamental Sciences, Gebze 41470, Türkiye; <sup>c</sup>Department of Biochemistry and Biophysics, University of North Carolina School of Medicine, Chapel Hill, NC 27599-7260; and <sup>d</sup>Department of Genetics, University of North Carolina School of Medicine, Chapel Hill, NC 27599-7264

1. A. Sancar, W. D. Rupp, A novel repair enzyme: UVRABC excision nuclease of *Escherichia coli* cuts a DNA strand on both sides of the damaged region. *Cell* **33**, 249–260 (1983).
2. E. M. Witkin, Mutation frequency decline revisited. *Bioessays* **16**, 437–444 (1994).
3. C. P. Selby, A. Sancar, Gene- and strand-specific repair in vitro: Partial purification of a transcription-repair coupling factor. *Proc. Natl. Acad. Sci. U.S.A.* **88**, 8232–8236 (1991).
4. C. P. Selby, E. M. Witkin, A. Sancar, *Escherichia coli* *mfd* mutant deficient in "mutation frequency decline" lacks strand-specific repair: In vitro complementation with purified coupling factor. *Proc. Natl. Acad. Sci. U.S.A.* **88**, 11574–11578 (1991).

5. C. P. Selby, A. Sancar, Molecular mechanism of transcription-repair coupling. *Science* **260**, 53–58 (1993).
6. C. P. Selby, L. A. Lindsey-Boltz, W. Li, A. Sancar, Molecular mechanisms of transcription-coupled repair. *Annu. Rev. Biochem.* **92**, 115–144 (2023).
7. P. C. Hanawalt, G. Spivak, Transcription-coupled DNA repair: Two decades of progress and surprises. *Nat. Rev. Mol. Cell Biol.* **9**, 958–970 (2008).
8. E. Kaja et al., Comparing Mfd- and UvrD-dependent models of transcription coupled DNA repair in live *Escherichia coli* using single-molecule tracking. *DNA Repair* **137**, 103665 (2024).

9. O. Adebali, A. Sancar, C. P. Selby, Dynamics of transcription-coupled repair of cyclobutane pyrimidine dimers and (6–4) photoproducts in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2416877121 (2024).
10. J. Fan, M. Leroux-Coyau, N. J. Savery, T. R. Strick, Reconstruction of bacterial transcription-coupled repair at single-molecule resolution. *Nature* **536**, 234–237 (2016).
11. D. J. Crowley, P. C. Hanawalt, Induction of the SOS response increases the efficiency of global nucleotide excision repair of cyclobutane pyrimidine dimers, but not 6–4 photoproducts, in UV-irradiated *Escherichia coli*. *J. Bacteriol.* **180**, 3345–3352 (1998).
12. R. C. Bockrath, J. E. Palmer, Differential repair of premutational UV-lesions at tRNA genes in *E. coli*. *Mol. Gen. Genet.* **156**, 133–140 (1977).
13. A. R. Oller, I. J. Fijalkowska, R. L. Dunn, R. M. Schaaper, Transcription-repair coupling determines the strandedness of ultraviolet mutagenesis in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 11036–11040 (1992).
14. S. Kunal, D. E. Brash, Excision repair at individual bases of the *Escherichia coli* lacI gene: Relation to mutation hot spots and transcription coupling activity. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 11031–11035 (1992).
15. A. Shibai *et al.*, Mutation accumulation under UV radiation in *Escherichia coli*. *Sci. Rep.* **7**, 14531 (2017).
16. O. Adebali, A. Sancar, C. P. Selby, Mfd translocase is necessary and sufficient for transcription-coupled repair in *Escherichia coli*. *J. Biol. Chem.* **292**, 18386–18391 (2017).
17. C. P. Selby, A. Sancar, Transcription preferentially inhibits nucleotide excision repair of the template DNA strand in vitro. *J. Biol. Chem.* **265**, 21330–21336 (1990).
18. V. Molodtsov, C. Wang, E. Firlar, J. T. Kaelber, R. H. Ebright, Structural basis of Rho-dependent transcription termination. *Nature* **614**, 367–374 (2023).
19. A. Ray-Soni, M. J. Bellecourt, R. Landick, Mechanisms of bacterial transcription termination: All good things must end. *Annu. Rev. Biochem.* **85**, 319–347 (2016).
20. S. Jain, R. Gupta, R. Sen, Rho-dependent transcription termination in bacteria recycles RNA polymerases stalled at DNA lesions. *Nat. Commun.* **10**, 1207 (2019).
21. C. P. Selby, A. Sancar, Structure and function of transcription-repair coupling factor. II Catalytic properties. *J. Biol. Chem.* **270**, 4890–4895 (1995).
22. M. N. Ragheb, C. Merrikh, K. Browning, H. Merrikh, Mfd regulates RNA polymerase association with hard-to-transcribe regions in vivo, especially those with structured RNAs. *Proc. Natl. Acad. Sci. U.S.A.* **118**, 1–10 (2021).
23. S. Ayora, F. Rojo, N. Ogasawara, S. Nakai, J. C. Alonso, The Mfd protein of *Bacillus subtilis* 168 is involved in both transcription-coupled DNA repair and DNA recombination. *J. Mol. Biol.* **256**, 301–318 (1996).
24. M. N. Ragheb *et al.*, Inhibiting the evolution of antibiotic resistance. *Mol. Cell* **73**, 157–165.e5 (2019).
25. J. E. Walker, O. Luyties, T. J. Santangelo, Factor-dependent archaeal transcription termination. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E6767–E6773 (2017).
26. A. M. Gehring, T. J. Santangelo, Archaeal RNA polymerase arrests transcription at DNA lesions. *Transcription* **8**, 288–296 (2017).
27. X. Zhao *et al.*, Reaction mechanism of (6–4) photolyase. *J. Biol. Chem.* **272**, 32580–32590 (1997).
28. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
29. A. Dobin *et al.*, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
30. F. Ramirez *et al.*, deepTools2: A next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
31. J. Koster, S. Rahmann, Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* **34**, 3600 (2018).
32. O. Adebali, P. A. Mieczkowski, A. Sancar, C. P. Selby, Genome-wide strand-specific mutagenesis in *E. coli* is directed by the Mfd translocase, Sequence Read Archive of the National Center for Biotechnology Information (NCBI). <https://ncbi.nlm.nih.gov/bioproject/PRJNA1302349>. Deposited 7 July 2025.