Cotranscriptional R-loop formation by Mfd involves topological partitioning of DNA

James R. Portman[a,b], Gwendolyn M. Brouwer[b,c], Jack Bollins[c], Nigel J. Savery[b,c], and Terence R. Strick[a,b,d,1]

[a]Institut de Biologie de l'Ecole Normale Supérieure (IBENS), Ecole Normale Supérieure, PSL University, INSERM, CNRS, Paris 75005, France; [b]Horizons 2020 Innovative Training Network "DNAREPAIRMAN", Paris 75005, France; [c]DNA-Protein Interactions Unit, University of Bristol, Bristol BS8 1TD, United Kingdom; and [d]Equipe Labellisée de la Ligue Nationale Contre le Cancer, Paris 75103, France

Edited by Aziz Sancar, University of North Carolina at Chapel Hill, Chapel Hill, NC, and approved February 5, 2021 (received for review September 17, 2020)

R-loops are nucleic acid hybrids which form when an RNA invades duplex DNA to pair with its template sequence. Although they are implicated in a growing number of gene regulatory processes, their mechanistic origins remain unclear. We here report real-time observations of cotranscriptional R-loop formation at single-molecule resolution and propose a mechanism for their formation. We show that the bacterial Mfd protein can simultaneously interact with both elongating RNA polymerase and upstream DNA, tethering the two together and partitioning the DNA into distinct supercoiled domains. A highly negatively supercoiled domain forms in between Mfd and RNA polymerase, and compensatory positive supercoiling appears in front of the RNA polymerase and behind Mfd. The nascent RNA invades the negatively supercoiled domain and forms a stable R-loop that can drive mutagenesis. This mechanism theoretically enables any protein that simultaneously binds an actively translocating RNA polymerase and upstream DNA to stimulate R-loop formation.

transcription-coupled repair | R-loops | mutagenesis | single molecule | magnetic trapping

R loops are triple-stranded structures in which one strand of a DNA double helix is annealed to a complementary RNA strand rather than its complementary strand of DNA. The superior stability of the resulting RNA–DNA hybrid over that of the DNA–DNA hybrid (1) partly explains why these R-loops are hotspots for mutagenesis. However, the mechanism whereby an RNA could successfully invade and anneal to regular B-form DNA remains mysterious. Because the process is thermodynamically favored by both the stability of the RNA–DNA hybrid as well as the negative DNA supercoiling of the genome, it is likely that limiting nucleation of such invasion is central to preventing large-scale genome-wide R-loop formation. An example of a pathway involving R-loops that leads to genetic instability is Mfd-dependent mutagenesis.

Mfd is historically well characterized as the transcription-repair coupling factor. Transcription-coupled repair (TCR) deals with bulky lesions in the transcribed strand of actively transcribed DNA (2, 3). Mfd binds to RNA polymerase (RNAP) stalled irreversibly at a DNA lesion and simultaneously binds to DNA immediately upstream of RNAP (4). Mfd uses ATP binding and hydrolysis to bind tightly to and translocate along DNA and against stalled RNAP thus removing RNAP from the DNA; however, remarkably RNAP remains thereafter bound to Mfd as part of a slowly translocating complex (4–6). Upon recruitment of the UvrAB complex, the Mfd–RNAP complex dissociates from the DNA (7, 8). Repair proceeds as UvrB recruits UvrC which nicks the DNA on either side of the lesion, UvrD unwinds the damage-containing oligonucleotide, and DNA polymerase and ligase complete repair (9, 10).

Interestingly, recent work has shown that Mfd can also induce mutagenesis, helping many bacterial organisms evolve to overcome various stresses (11–15). A proposed mechanism for this begins with Mfd-dependent R-loop formation (11, 13, 14). Here, we show that a mechanism by which R-loops are formed involves interactions between Mfd and elongating RNAP and Mfd and upstream DNA. This results in the formation of a closed topological domain between Mfd and RNAP, in which negative supercoiling builds up over time as RNAP continues to translocate. The nascent RNA readily invades this negatively supercoiled domain to form an R-loop. Tellingly, a derivative of Mfd that is incapable of translocation or of displacing RNAP from lesions is competent for R-loop formation confirming that the mechanistic requirements of R-loop formation differ from those of TCR. This represents a potentially universal model of R-loop formation.

Results

Mfd Interacts with Elongating RNAP to Form a Tripartite Supercoiled Domain. To observe the interaction of Mfd with elongating RNAP, we used a 4.6 kbp DNA construct containing a 1 kbp transcription cassette. This DNA construct was manipulated using a magnetic trap, allowing us to observe transcription by individual RNAPs under saturating nucleotide (NTP) conditions (Fig. 1 *A* and *B*) as already described (16). When Mfd was added, we observe transcriptional events interrupted by dramatic bursts of positive DNA supercoiling in which the DNA extension rapidly decreases in a linear fashion and at a rate of ∼100 nm/s (Fig. 1*C* and *SI Appendix*, Fig. S1). The same behavior was also observed using a 2 kbp DNA construct bearing a distinct 100 bp transcription cassette (*SI Appendix*, Fig. S1*E*). No such behavior was observed when Mfd was omitted (*SI Appendix*, Fig. S2).

These events are consistent with topological partitioning of the DNA into three domains in a manner similar to that proposed by

Significance

R-loop structures pose a threat to genomic integrity because their formation can lead to mutagenesis. Mutagenesis drives antibiotic resistance in bacteria and chemotherapy resistance in human cancers; therefore, it is important to understand how R-loops form. Using a combination of in vitro single-molecule experimentation and in vivo mutagenesis assays, we have shown how Mfd, a bacterial protein known for its role in DNA repair, can interact with RNA polymerase to form R-loops. This interaction generates a topological domain in DNA that is highly prone to R-loop formation. The observed mechanism relies on properties of Mfd that are shared by many other proteins, hinting that this is a potentially universal model for R-loop formation.

Author contributions: J.R.P., N.J.S., and T.R.S. designed research; J.R.P., G.M.B., and J.B. performed research; J.R.P., G.M.B., J.B., N.J.S., and T.R.S. analyzed data; and J.R.P., N.J.S., and T.R.S. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: strick@biologie.ens.fr.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2019630118/-/DCSupplemental.

Published April 7, 2021.

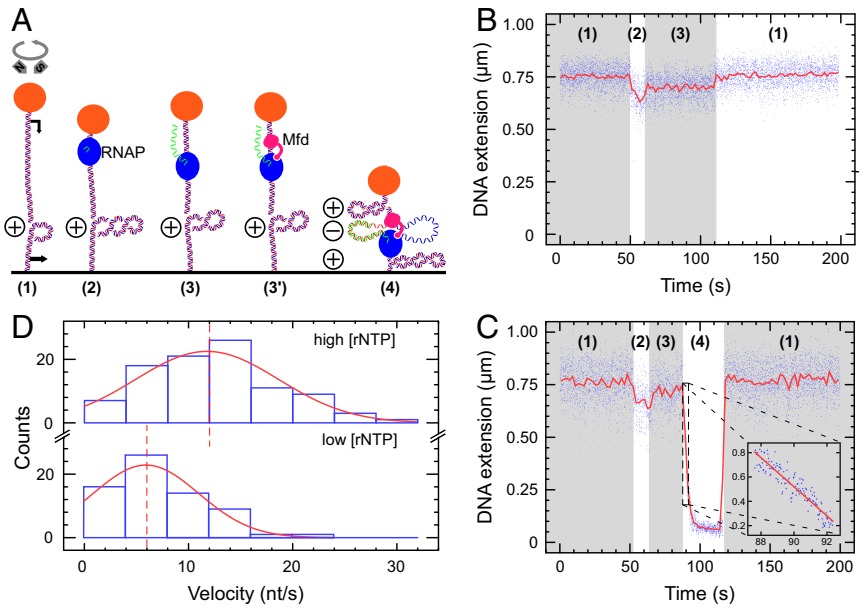

Fig. 1. Mfd interacts with elongating RNAP to form a tripartite supercoiled domain. (*A*) Experimental model. The 4.6 kbp DNA bears a promoter (black arrow, top) and terminator (black arrow, bottom) separated by ∼900 bp of DNA and is tethered between a magnetically trapped bead (orange) and a glass surface. DNA is positively supercoiled prior to addition of components (RNAP: blue, Mfd: pink). During synthesis of RNA (green), Mfd can bind to elongating RNAP and DNA to form a tripartite supercoiled domain. (*B*) RNAP-only time trace showing phase (1) positively supercoiled DNA. In phase (2), as RNAP initiates it scrunches (16) and unwinds ∼2 turns of DNA, reducing DNA extension by ∼100 nm. In phase (3), successful conversion to an elongation complex results in an unwound bubble of ∼9 bp, and so DNA extension is reduced by only ∼50 nm compared with baseline. Upon transcription termination DNA returns to the phase (1) baseline state. (*C*) RNAP and Mfd time-trace showing phase (1) positively supercoiled DNA. Transcription initiation in phase (2) and conversion to the elongation state in phase (3) begins as in *B*, but then in phase (4) Mfd binds to RNAP and DNA to form the tripartite supercoiled domain. Here, ongoing elongation by RNAP causes a gain of positive supercoiling in the external domains, pulling the bead down toward the surface. Finally, loss of Mfd–RNAP interaction or RNAP termination causes the tripartite domain to reunite and dissolve, returning the DNA to the phase (1) baseline state. Blue points are the raw data (30 Hz); and the red line is the averaged data (1 s filtering). Inset focuses on the period of rapid bead descent at the beginning of phase (4) with linear fit (red line) used to measure bead velocity. (*D*) Bead descent rates in nm/s (see inset in prior panel) are converted to differential velocities of motor proteins (in bp/s) using the DNA supercoiling response (20, 25) and represented as histograms (blue) fit to Gaussian distributions (red). (*Top*) At high NTP concentration ([UTP] = [GTP] = [CTP] = [ATP] = 1 mM), we obtain a mean differential velocity of 12 ± 0.8 nt/s (SEM, n = 96 events). (*Bottom*) At lower NTP concentration ([UTP] = [GTP] = [CTP] = [ATP] = 50 μM), we obtain a mean differential velocity of 6 ± 1.0 nt/s (SEM, n = 67 events). Dashed lines indicate mean values.

Liu and Wang (17, 18). Mfd simultaneously binds to RNAP and to upstream DNA in a robust and long-lived manner (6, 8). The former interaction is mediated by the RNAP-interaction domain of Mfd which binds to a conserved site on RNAP's β subunit (19); the latter interaction is mediated by the RecG-like domains of Mfd which allow it to translocate along DNA. This dual interaction generates three distinct topological domains in the DNA: one "internal" domain accounting for the DNA in between Mfd and RNAP and two "external" domains corresponding to the remaining portion of the DNA held between bead and surface in the magnetic trap ("in front" of RNAP and "behind" Mfd). During transcription, DNA is transferred from the external domains, which shrink in length, to the internal domain, which grows in length. Because Mfd acts as an anchor to prevent RNAP from rotating about the DNA helix as it transcribes, the DNA linking number in each of the two domains remains unchanged. The internal domain therefore grows in length but has a fixed linking number. This means it accumulates a negative supercoil for every turn of DNA transcribed, while the external domains conversely accumulate a positive supercoil. Because the external domains are topologically accessible to the magnetic trap, this positive supercoil pulls the bead toward the microscope surface by ∼55 nm, the length of a supercoil under the low extending force used here (see *Materials and Methods*). With transcription occurring at ∼20 bases/s and transcription of 10.5 bases causing accumulation of one supercoil (16), this predicts a rate of descent of ∼100 nm/s, consistent with our observations.

According to this model, the bead's rate of descent should depend on the net velocity difference between RNAP and Mfd. In other words, the difference in velocity between Mfd and RNAP governs the rate at which the internal domain grows and becomes negatively supercoiled. This negative supercoiling is balanced by positive supercoiling in the external domains which pulls the bead to the surface. As a result, the velocities of the proteins control the rate of bead descent. We therefore measured this rate for both high and low NTP concentrations (Fig. 1*D*). In both cases, Mfd is expected to translocate at maximum velocity [∼5 bp/s (7, 8)] due to its strong affinity for ATP [K_M^{ATP}∼10 μM (7)], and so only the velocity of RNAP will change. As we decrease NTP concentration, the distribution of bead descent rates indeed shifts to lower values (Fig. 1*D*). This shows we can modulate the speed of RNAP to alter the rate of supercoil accumulation within the tripartite supercoiled domain.

We note that adding the antibacktracking factor GreB while keeping NTP concentration constant shifted velocity distributions to higher values. This is presumably because GreB reduces the time RNAP spends in backtracked paused states thus increasing the average velocity (*SI Appendix*, Fig. S3). We also note that control experiments carried out with only wild-type Mfd and ATP occasionally result in similar events; however, in such cases, the bead's average rate of descent is only ∼4nt/s (*SI Appendix*, Fig. S4). These Mfd-only events also occur in experiments with RNAP and NTPs but are sufficiently rare and kinetically distinct that they did not interfere with tripartite supercoiled domain measurements.

However, when these control experiments were performed on negatively supercoiled DNA, the frequency of such events increased considerably, interfering with measurements of Mfd-RNAP interactions. For these reasons, experiments on negatively supercoiled DNA were next pursued with the MfdRA953 translocation-deficient mutant (see below).

DNA Compaction without Topological Coupling Is Observed on Nicked DNA Molecules. In nicked DNA molecules, we could still observe a decrease in DNA extension due to the formation of the internal domain (*SI Appendix*, Fig. S5). Indeed, as before, the differential velocity of the two motors reduces the length of DNA in the external domains and increases the length of DNA in the isolated internal domain, causing the DNA extension to decrease. However, in this case, no topological coupling takes place to positively supercoil the external domain. As a result, the rate of bead descent is observed to be roughly 20 times lower than on supercoiled molecules, on the order of 5 nm/s. Given DNA structure and mechanics, this corresponds to a differential velocity of ∼20bp/s, consistent with expectations (20).

Translocase-Deficient MfdRA953 Maintains the Ability to Form Tripartite Supercoiled Domains. To further test the validity of our model, we used a "translocase-deficient" Mfd mutant, MfdRA953. The arginine to alanine substitution at position 953 perturbs the Rec-G–like module's "TRG" motif, rendering Mfd unable to translocate on DNA (21). MfdRA953 and ATP alone on either positively or negatively supercoiled DNA shows no events unlike wild-type Mfd (*SI Appendix*, Fig. S6). This mutant protein can still bind to RNAP and DNA, so it should give rise to tripartite supercoiled domain events (21). Indeed, when transcribing RNAP is exposed to MfdRA953, we observe events similar to those observed with wild-type Mfd (Fig. 2*A*). Although our expectation was that we would observe higher differential velocities with this mutant, we in fact observe a lower velocity differential compared with wild-type (*SI Appendix*, Fig. S7). This could possibly be explained by sliding of MfdRA953 on DNA in the same direction as RNAP.

We also tested MfdRA953–RNAP interactions on DNA that was negatively supercoiled prior to addition of components. In this case, the positive supercoils generated in the external topological domains first annihilate the negative supercoils already present in the DNA, causing an increase in the bead's altitude with respect to the surface. If the tripartite domain lasted long enough, the bead altitude crossed the threshold of maximal DNA extension (i.e., the DNA's torsionally relaxed state) before decreasing as net positive supercoils formed in the external domains (Fig. 2*B*). In addition, the chances of RNAP backtracking increase with the degree of negative supercoiling immediately upstream of RNAP (22). Thus, using GreB to reactivate backtracked complexes increases the fraction of events that pass through the maximal extension state (*SI Appendix*, Table S1) and go on to form net positive supercoils (23) in the external topological domains.

MfdLR499, Deficient in Binding to RNAP, Lacks the Ability to Form Tripartite Supercoiled Domains. To verify that Mfd must bind RNAP for this process to take place, we tested an Mfd mutant, MfdLR499, deficient in its RNAP-binding ability (24). MfdLR499 in the presence of ATP alone could be observed to interact with positively supercoiled DNA in a manner similar to that seen with wild-type Mfd (*SI Appendix*, Fig. S8, *Top*). However, no topological partitioning was observed using MfdLR499 with RNAP and NTPs on positively supercoiled DNA (*SI Appendix*, Fig. S8, *Bottom*). We therefore confirm that RNAP binding is absolutely required for the tripartite supercoiled domain phenomena.

Backtracked RNAP Is Not a Major Source of Tripartite Supercoiled Domain Formation. Lastly, we inquired whether elongating or transiently paused RNAP is the initial substrate for Mfd, rather than RNAP durably arrested in a backtracked state. We therefore defined a metric, $N_{complete}$, corresponding to the number of complete transcription cycles which occur on a given DNA before the next transcription cycle is interrupted by formation of a tripartite supercoiled domain. If backtracked RNAP is the main substrate for Mfd, then the addition of GreB should increase $N_{complete}$ as GreB lowers the time RNAP spends in the backtracked state. Experiments carried out in the presence or absence of 50 nM GreB (Fig. 2*C*) show that $N_{complete} = 4$ in both cases. We conclude that Mfd initially interacts with elongating or transiently paused RNAP rather than backtracked RNAP.

Formation of Tripartite Supercoiled Domains Leads to R-Loops. We next consider whether the negatively supercoiled internal domain

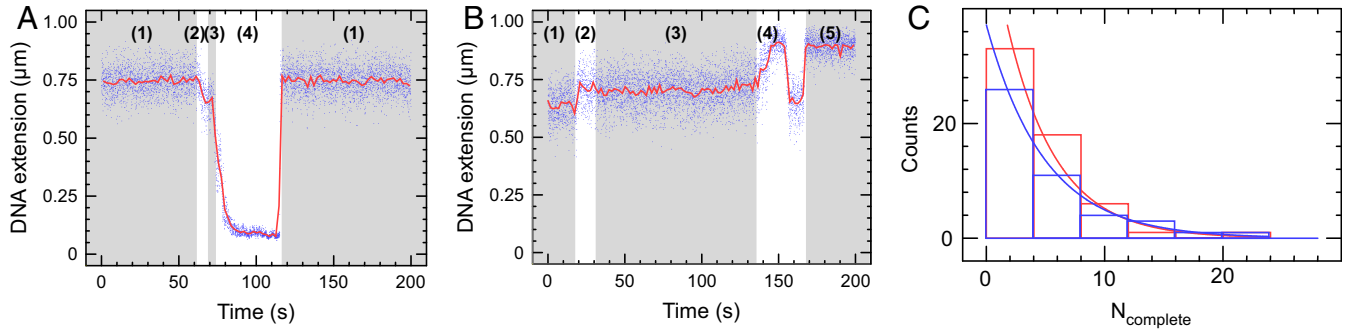

Fig. 2. Translocase-deficient MfdRA953 maintains the ability to form tripartite supercoiled domain. (*A*) Time-trace showing phase (1) positively supercoiled DNA, phase (2) RNAP initiation and phase (3) elongation. In phase (4), MfdRA953 binds to elongating RNAP and DNA to form the tripartite supercoiled domain. As RNAP terminates, the domains dissolve and annihilate, and the DNA returns to the phase (1) baseline extension. (*B*) Time-trace showing phase (1) negatively supercoiled DNA, phase (2) RNAP initiation, and phase (3) elongation. These signals are in the opposing direction to *A* because introduction of positive supercoiling into negatively supercoiled DNA leads to an increase in DNA extension. In phase (4), MfdRA953 binds to elongating RNAP and DNA to form the tripartite supercoiled domain. The gain of positive supercoils first annihilates the negative supercoils (DNA extension increases) before reaching the maximal extension state, after which point there is a net gain of positive supercoils which causes the bead to descend toward the surface. When RNAP terminates or Mfd dissociates from either RNAP or DNA, in phase (5) a new high-extension intermediate state is observed for negatively supercoiled DNA substrate only. (*C*) The frequency distribution of tripartite supercoiled domain events is fitted to an exponential distribution. $N_{complete}$ is defined for each DNA molecule as the number of uninterrupted RNAP transcription cycles completed on that given DNA molecule before a tripartite supercoiled domain event interrupts a transcription cycle; a low $N_{complete}$ indicates an elevated frequency of transcription events displaying tripartite supercoiled domain formation. With 50 nM GreB (blue), the average of $N_{complete}$ is 4.2 ± 0.6 (SEM, $n = 59$) and without GreB (red) the average of $N_{complete}$ is 5.0 ± 0.9 (SEM, $n = 46$).

contained between Mfd and RNAP could be a substrate for R-loop formation (25). In order to observe R-loops, experiments were carried out using MfdRA953 on negatively supercoiled DNA. This is because an R-loop formed in a positively supercoiled background will be ejected when the domain dissolves (e.g., upon transcription termination or dissociation of Mfd), yet will be maintained if formed in a negatively supercoiled background (25). Indeed, events observed on negatively supercoiled DNA with MfdRA953 essentially always display a new intermediate state (Fig. 3A, labeled 5), which appears upon dissolution of the tripartite supercoiled domain and which is followed by a return to the baseline state (Fig. 3*A* and *SI Appendix*, Fig. S9). This intermediate state is extremely long lived, with a mean lifetime on the order of ∼13,000 s (Fig. 3*B* and *SI Appendix*, Fig. S10*A*). It is also extremely sensitive to RNase HI, which specifically degrades the RNA in RNA–DNA hybrids (26), displaying a mean lifetime of only ∼15 s in the presence of 0.025U/μL RNase HI (Fig. 3*B* and *SI Appendix*, Fig. S10*B*).

Furthermore, we successfully interrogated the topology of the intermediate state. We awaited formation of the intermediate state and measured the supercoiling response of the DNA while in this state (Fig. 3*C*). This supercoiling response measurement is termed a "rotation scan" and involves iteratively rotating the magnets and recording the extension of the molecule (*SI Appendix*, Fig. S11). In Fig. 3*C*, the DNA rotation scan taken while the DNA is in the intermediate state is shown in red and begins by exploring the response of DNA in the intermediate state to negative supercoiling. Indeed, once this scan proceeds into the positively supercoiled regime, the R-loop is ejected (25). Overlaid in blue is a second scan, taken immediately after the first scan. The difference in the negative portion of the scan is indicative of the presence (and subsequent absence) of an R-loop, as shown in other single-molecule hybridization experiments (25). According to these single-molecule hybridization experiments, the size of the shift between the two states corresponds to the size of the R-loop formed. For the molecule in Fig. 3*C*, the shift of ∼7 turns indicates the formation of a ∼70 bp R-loop.

Both Wild-Type and Translocation-Deficient Mfd Support R-Loop–Dependent Mutagenesis In Vivo. To probe our model in vivo, we used an assay that quantifies stress-induce mutagenensis by measuring the emergence of bacteria that have mutated a defective lacZ gene and are thus able to utilize lactose as the sole carbon source (Lac+ revertant assay) (13, 27). This assay quantifies starvation stress-induced mutagenesis by measuring the emergence of bacteria that have mutated a defective *lacZ* gene and are thus able to utilize lactose as the sole carbon source. The chromosomal *lac* operon of these bacteria is deleted, and they contain an F′ plasmid bearing a *lacI–lacZ* fusion which produces nonfunctional β-galactosidase due

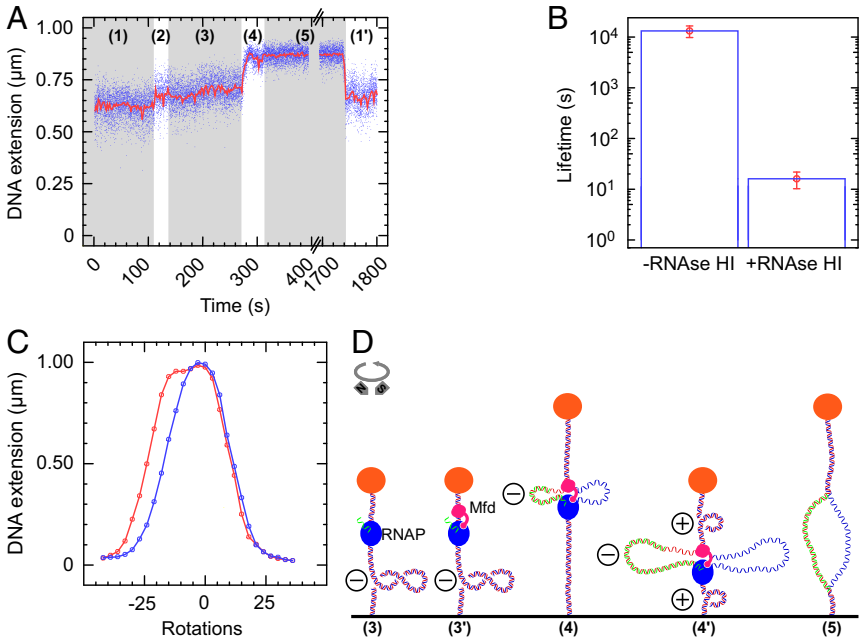

Fig. 3. Formation of tripartite supercoiled domains leads to R-loops. (*A*) Time-trace showing phase (1) negatively supercoiled DNA. In phase (2), as RNAP initiates, it scrunches and unwinds ∼2 turns of DNA, increasing DNA extension by ∼100 nm. After promoter escape, phase (3) elongation proceeds with an unwound bubble of ∼9 bp, and so DNA extension is increased by only ∼50 nm compared with the baseline. Then MfdRA953 binds to RNAP and DNA to form the tripartite supercoiled domain. Ongoing elongation by RNAP causes a gain of positive supercoiling in the external domains, in phase (4), pulling the bead to the surface after passage through the maximal extension state. There follows phase (5), a new state in the DNA extension after the tripartite supercoiled domain event ends. This state, phase (5), is the maximal-extension (i.e., torsionally relaxed) state of the DNA for the extending force used and is consistent with the presence of an R-loop (25). While the DNA is in the maximal-extension state, it is possible for more RNAPs to initiate transcription, and so once the intermediate state ends, the DNA does not always return exactly to a baseline extension but can instead return to a near-baseline extension, phase (1′), indicative of a new ongoing transcription cycle, as we see here. (*B*) The lifetime of the maximal extension "R-loop" state without RNase HI ($n = 29$; average = 13,000 ± 3,000 s [SEM]) and with 0.025U/μL RNase HI ($n = 17$; average = 16 ± 6 s [SEM]); see *SI Appendix*, Fig. S10 for distributions and fits. (*C*) Rotation scan of a DNA molecule captured while in the intermediate state, phase (5). A first scan (in red) is taken immediately following formation of the intermediate state [i.e., while the DNA is in the R-loop state, phase (5)], from −42 turns through to +33 turns. Once the DNA becomes positively supercoiled, the R-loop is ejected, and a second scan (blue) is taken from +33 turns to −42 turns. The difference between the two curves in the negative branch of the curve is indicative of an R-loop (25). A rotation scan taken before the experiment began is shown in *SI Appendix*, Fig. S11. (*D*) Experimental model. In phase (3), RNAP transcribing on negatively supercoiled DNA is [phase (3′)] bound by Mfd to form the tripartite supercoiled domain. In phase (4), positive supercoiling builds up in the external domain which annihilates the negative supercoiling, increasing end-to-end extension until a maximal extension state is reached. In phase (4′), as net positive supercoils are generated the end-to-end extension again decreases until phase (5) when one of the motors dissociates revealing the R-loop maintained in the DNA.

to a +1 frameshift mutation in the *lacZ* gene. This strain cannot use lactose as a carbon source unless a point mutation or amplification occurs (28).

First, as seen by others (13), we confirmed that deletion of the gene encoding RNase H promoted mutation in the wild-type background. Knockout of Mfd caused a decrease in mutation relative to wild-type, and in the knockout background, further deletion of the gene encoding RNase H had no further effect on mutation rates (*SI Appendix*, Fig. S12). As described previously, these results indicate that R-loops are generated by Mfd and are intermediates in the Mfd-dependent mutagenesis pathway (13).

We next asked whether the effect of *mfd* deletion on R-loop–dependent mutagenesis could be complemented by expressing wild-type or mutant Mfd proteins from a plasmid ($pET21a^{Lac-}$) under control of the wild-type *mfd* promoter. Fig. 4 shows that Δ*mfd* bacteria transformed with a vector expressing wild-type Mfd showed a higher mutation frequency than those transformed with an empty vector. Expression of MfdRA953, which lacks DNA translocation activity and is unable to displace stalled RNAP from DNA, also complemented the *mfd* deletion and increased mutation frequency to a level similar to that seen with wild-type Mfd. In contrast, expression of MfdLR499, which is unable to bind stably to RNAP (24), had no effect on the observed mutation frequency in this assay. These results indicate that, in accordance with the single-molecule findings, the RNAP-binding function of Mfd is essential for the phenomenon while its translocase function is not required, and suggest that the mechanism for R-loop formation identified in our single-molecule experiments in vitro underpins mutagenic R-loop formation in vivo.

Discussion

Our work demonstrates robust cotranscriptional formation of an R-loop within the negatively supercoiled portion of a tripartite supercoiled domain. We reveal the key mechanistic features enabling Mfd, a factor first identified for its role in transcription-coupled DNA repair (29), to ultimately be involved in R-loop mediated mutagenesis and adaptation to stress. The mechanism is based on already well-defined aspects of Mfd biochemistry, namely, binding to RNAP and binding to DNA. This ability of Mfd to simultaneously interact with both elongating RNAP and DNA results in the formation of a tripartite supercoiled domain including massive negative supercoiling in the relatively small domain bookended by RNAP and Mfd.

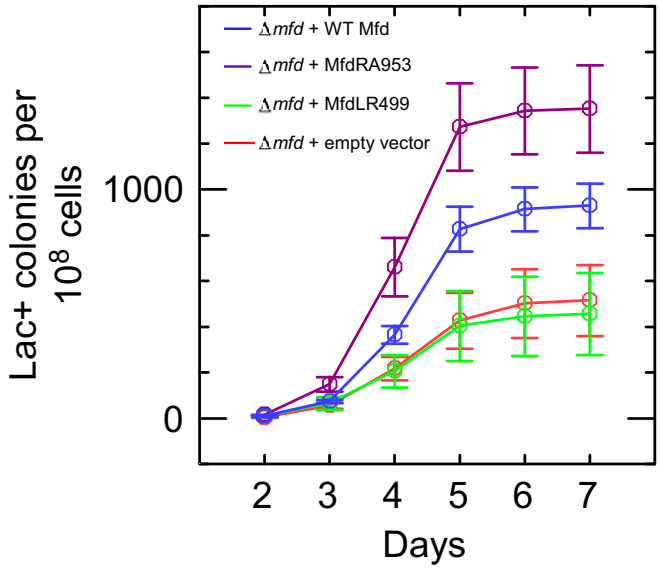

Fig. 4. Wild-type Mfd and MfdRA953 induce mutations via R-loops in cells under starvation stress. Number of mutants (number of colonies counted on minimal lactose media) observed over time in strains lacking Mfd (PJH813 + $pET21a^{Lac-}$; red) or expressing wild-type Mfd (PJH813 + $pET21a^{Lac-}$*Mfd*; blue), MfdRA953 (PJH813 + $pET21a^{Lac-}$*MfdRA953*; magenta) or MfdLR499 (PJH813 + $pET21a^{Lac-}$*MfdLR499*; green). Colonies seen before day 2 arise due to replication-associated mutations while in liquid media and are not counted (27). Data are an average of three independent cultures, and error bars represent SEM.

Given that Mfd binds to DNA immediately upstream of RNAP (4), the linking number of this domain when it is formed is small, on the order of one or two. As a result, transcription of (e.g., only 10) base pairs by RNAP is expected to result in a massive degree of negative supercoiling in this domain (normalized supercoiling $\sigma \sim -0.5$), conditions under which DNA is not expected to remain double stranded (20). This negatively supercoiled DNA is a powerful substrate for the formation of R-loops which are expected to be extremely stable in the context of negative chromosomal supercoiling. At the same time, the exact process of R-loop annealing remains poorly understood. The genetics work of Wimberly et al. shows that reduced accessibility of the RNA 5′ end resulting from ribosome loading indeed reduces the frequency of this phenomenon but does not eliminate it (13). R-loop formation is presumably a stochastic process which occurs in competition with occupancy of the RNA by ribosomes.

Mfd's role in R-loop formation is separate and distinct from its role in TCR. In TCR, Mfd translocates against irreversibly stalled RNAP to dislodge it from the lesion site. Prior magnetic trap experiments did not observe R-loop formation because they were probing displacement of irreversibly stalled RNAP. For those assays, RNAP was immobilized at +20 from the transcription start site (i.e., immediately after promoter escape), leaving no opportunity for formation of topological domains of any significant size. We also note that to observe this process in as many as 25% of transcription events, we used an Mfd concentration compatible with the upper range of Mfd concentrations reported in vivo [from 100 to 500 nM (30, 31)]. Thus, although infrequent, this phenomenon is robust, as underscored by its importance in vivo.

Interestingly, recent optical trapping experiments have also probed the movement of Mfd along the DNA contour, both alone and in the presence of RNAP (32). However, the optical trapping experiments implemented did not probe DNA topology and did not look for R-loop formation nor ascertain whether there was a physical link between Mfd and RNAP that persisted during translocation of the two motors. Thus at least three specific classes of translocating Mfd complexes have been identified to date: Mfd alone (*SI Appendix*, Fig. S4 and ref. 32), Mfd associated with a transcription elongation complex (ref. 4, ref. 33, and this work), and Mfd associated with RNA polymerase that has been dissociated from a stall site and no longer retains RNA (6–8).

Taken together, the collection of single-molecule data obtained to date converge on the fact that the velocity of Mfd, on the order of ~5 bp/s, is essentially the same in all of these contexts [Fig. 1*D*, *SI Appendix*, Fig. S4, and (7, 8, 32)]. We note, however, that the processivity of Mfd translocating alone on DNA tends to be short, in the range of a few hundred base pairs (*SI Appendix*, Fig. S4 and ref. 32). The processivity of Mfd translocating on DNA in the context of TCR can reach into the thousands of base pairs if downstream repair factors UvrA and UvrB are not present to remove stable Mfd–RNAP complexes from DNA before carrying out downstream steps of DNA repair (7, 8). Mfd processivity could not be measured in the experiments presented here because of the limited size of the transcription unit. The activity of Mfd in each of these three classes likely differs as a function of the competence for transcription of the targeted RNAP and the availability of the downstream repair components.

Ultimately, the topological mechanism we identify stands in stark contrast to the classical twin-supercoiled domain phenomenon first discussed by Liu and Wang (17) in which RNAP,

prevented from rotating about the helix due to ribosomal torsional drag, for example, induces positive supercoiling downstream and negative supercoiling upstream of RNAP. Although transcription has since been broadly recognized as a potential driver of negative supercoiling behind of RNAP, the large size of chromosomal topological domains (or TADs) and the presence of topoisomerases are likely to dilute this effect. Thus, most importantly, in the twin-domain model, upstream negative supercoiling is dissipated on the kbp scale, whereas in the context of the tripartite supercoiled domain, negative supercoiling is highly focused onto a very small region of DNA. Secondly, our model also has important consequences for overall chromosome topology. Whereas positive supercoils form only downstream of RNAP in the twin-domain model, they form both upstream and downstream of RNAP in the triple-domain model, with the potential of greatly perturbing the overall topology of TADs (34).

Our work highlights the docking site of Mfd on RNAP β subunit as a potential target for inhibiting evolution of antimicrobial resistance. RNAP-associated DNA translocases are found across all kingdoms of life (35–37), and the differential-translocation mechanism identified in this work may drive R-loop formation in species beyond *Escherichia coli*. In humans, for instance, the CSB homolog of Mfd is found to be overexpressed in cancers and drives tumor survival in the face of chemotherapeutic agents (38).

Finally, RNAP-binding proteins have not yet been mechanistically linked to R-loop formation. Concomitant binding to RNAP and DNA will initially generate a small topological domain and then focus negative supercoiling in this small domain, resulting in its near-complete unwinding. The relatively small size of this domain likely makes it harder for topoisomerases to locate it via diffusion. These features are likely to enhance successful R-loop formation in vivo, possibly explaining how Mfd acquired this promutagenic function. Yet as we have shown, DNA translocation by the secondary protein is not a requirement for this phenomenon. Therefore, we propose that numerous transcription factors, which simultaneously bind RNAP and upstream DNA, could be involved in promoter proximal R-loop formation via this mechanism, with ramifications for genetic regulation and stability.

Materials and Methods

DNA Constructs. The DNA construct used in single-molecule experiments is 4.6 kbp long, bearing a "transcription cassette" (T5 N25 promoter, roughly 900 bp of transcript, followed by a tr2 terminator; see *SI Appendix* for details and full sequence) and was cloned into homemade vector pET21aΔMCS via sites for the restriction enzymes XbaI and SbfI. The resulting plasmid was extracted from bacterial cultures using a Nucleobond Xtra Midiprep kit (Macherey-Nagel), digested with the restriction enzymes XbaI, SbfI-HF, and EcoRI (New England Biolabs), and the resulting 4.6 kbp band purified via agarose gel electrophoresis followed by extraction from the gel (NucleoSpin Gel and PCR Clean-up, Macherey-Nagel). This purified DNA was ligated at the XbaI end to a 1 kbp DNA handle bearing multiple biotin groups on both strands and ligated at the SbfI end to a 1 kbp DNA handle bearing multiple digoxigenin groups on both strands [see ref. 39 for details]. These handles are homemade, using the *Thermus aquaticus rpoC* gene as template for a PCR using the following primers: 5′ GAGAGACCTGCAGGACATCAAGGACGAGGTGTGGG 3′ and 5′ GAGAGATCTAGATCCTCAAAGTTCTTGAAGACCGCCTGG 3′.

The PCR is done twice using the Expand High Fidelity PCR system (Roche), once with dUTP-biotin (Roche) in the dNTP mix and the other with alkali-stable dUTP-digoxigenin (Roche) in the dNTP mix. The biotin-labeled DNA was digested with XbaI while the digoxigenin-labeled DNA was digested with SbfI-HF. As in ref. 39, this DNA is tethered first to streptavidin-coated 1 μm diameter superparamagnetic beads (Dynabeads MyOne C1, Thermofisher) and then deposited onto a polystyrene-coated borosilicate glass coverslip (Menzel Gläser) functionalized with anti-digoxigenin (Roche). The DNA-bead mixture incubates on the surface for 10 min, after which free magnetic beads and DNA are washed out with buffer.

Proteins. RNAP core, σ^{70}, GreB, and wild-type Mfd were purified as previously described (6, 8). Core RNAP was saturated with a fivefold excess of σ^{70} to ensure maintenance as holoenzyme. MfdRA953 and MfdLR499 were made by site-directed mutagenesis (QuikChange II, Agilent Technologies) on the pAD6 wild-type Mfd expression plasmid using the following primers:

MfdRA953.

5′ CACGATCTGGAGATTGCGGGCGCGGGTGAACTG 3′

5′ CAGTTCACCCGCGCCCGCAATCTCCAGATCGTG 3′

MfdLR499.

5′ GGAATGACCACGAGGGAAGCGGGTGGCATTACTGG 3′

5′ CCAGTAATGCCACCCGCTTCCCTCGTGGTCATTCC 3′

Successful mutants were verified by sequencing and expressed and purified exactly as wild-type Mfd. Protein concentrations were determined using the Folin–Lowry assay and spectrophotometry.

Flow Cell Preparation. Flow cells derivatized with anti-digoxigenin for single-molecule assays were prepared as previously described (39). Different surfaces were used for experiments with wild-type Mfd and MfdRA953 to avoid cross-contamination.

Reaction Conditions. Single-molecule assays were performed at 34 °C in reaction buffer containing 40 mM K·Hepes pH 8.0, 100 mM KCl, 8 mM $MgCl_2$, 0.5 mg/mL bovine serum albumin, 0.1% wt/vol Tween 20, and 1 mM 1,4-Dithiothreitol (7). Unless stated otherwise, concentrations of components in reactions were as follows: 300 pM RNAP holoenzyme (including a total of 1.5 nM of σ^{70}), 500 nM Mfd/MfdRA953/MfdLR499, 50 nM GreB, 1 mM ATP, 100 μM CTP, 100 μM UTP, 100 μM GTP. Collecting the specified number of events, *n*, typically requires more than three experimental runs involving normally 20 to 50 DNA molecules simultaneously.

Tethered-DNA Assays. Only single molecules of intact, unnicked double-stranded DNA are initially selected for experimentation, as verified by their ability to supercoil and the form of the rotation-extension curve (40). DNA was under a constant extending force of ∼0.3 pN and with +10 turns or −12 turns (for experiments done at positive or negative supercoiling, respectively). For addition of components, DNA was positively supercoiled, and the force was increased to ∼2 pN to prevent DNA molecules becoming stuck to the surface under the flow of components. Once the flow equilibrated, the force was reduced, and the DNA supercoiled to the relevant value. Tripartite supercoiled domain events were chosen based on the following criteria: RNAP initiation must be observed prior to the event, and the event amplitude must be greater than 200 nm (to avoid recording the theoretically possible rapid successive initiation of two RNAPs).

Lac$^+$ Revertant Assay. The Δ*mfd E.coli* strain (PJH813) (13) was transformed with an Mfd-expression plasmid (pETMfd2^{Lac-}, pETMfd2LR499^{Lac-}, pETMfd2RA953^{Lac-}) or pET21a^{Lac-}. The pETMfd2^{Lac-} and pET21a^{Lac-} plasmids are derivatives of pETMfd2 (24) (and variants expressing Mfd with the indicated substitutions) and pET21a (Novagen). They were created by excising the *lacI^Q* gene from the parent plasmids using restriction enzymes FspAI and EcoNI and religating the plasmid backbone following blunt-ending with Klenow DNA polymerase. The *mfd* gene in these plasmids is expressed from the native *mfd* promoter. All cultures and plates were supplemented with 100 μg/mL ampicillin throughout the experiment to maintain the plasmid. A single colony was grown until prolonged stationary phase (∼36 to 48 h) at 37 °C in M9 media supplemented with 0.1% glycerol and 20 μg/mL thiamine. A total of 100 μL of culture was plated onto M9 agar supplemented with 0.1% lactose and 20 μg/mL thiamine and incubated at 37 °C for 7 d. New Lac$^+$ colony-forming units (CFUs) were counted on each day. Previous research shows that colonies arising before day 2 mutated during liquid growth and thus were not counted (13). The number of CFUs plated was determined by serial dilutions on Luria-Bertani agar. Control experiments using wild-type (SMR4562), Δ*mfd* (PJH813), Δ*rnhA* (PJH683), and Δ*rnhA*Δ*mfd* (PJH946) *E.coli* strains (13) were carried out as above; however, as there was no plasmid incorporation, ampicillin was omitted throughout.

Data Acquisition and Analysis. Data were collected on a homebuilt magnetic trap using the PicoJai software suite (PicoTwist SARL). Real-time bead tracking was carried out at 30 Hz; in time traces, raw bead position is shown in blue and ∼1 s average is shown in red. Histograms were fit to a single exponential or Gaussian fit as stated. Number of events (n) and the values of the fits with errors are stated in the figure legends. Data analysis was done using the Xvin software suite (PicoTwist SARL). Velocities were measured manually by locating the linear portion of an event (linear section of the averaged data, with the same averaging values used in all cases) and fitting

the slope of this line. This value was converted from nm/s to nt/s as 55 nm = 1 turn = 10.5 base pairs (40).

Data Availability. Some study data available upon request.

ACKNOWLEDGMENTS. We thank Peter Hastings for supplying the various reporter strains used in the mutagenesis experiments, T.R.S. laboratory members and members of the Bristol DNA-protein interactions unit for stimulating discussions, and Xi Yang for assistance with figure design software. J.R.P. and G.M.B. are supported by graduate scholarships of the Horizons 2020 Innovative Training Network "DNAREPAIRMAN." J.B. was supported by a studentship from the Biotechnology and Biological Sciences Research Council South West Biosciences Doctoral Training Partnership (BBSRC SWBio DTP). J.R.P. acknowledges support from the Frontières du Vivant–Programme Bettencourt doctoral program. Research on this topic in the T.R.S. laboratory was supported by grants from the French Agence Nationale de la Rechrche (ANR-17-CE11-0042) and the Ligue Nationale Contre le Cancer program for Core Research Teams ("Equipes Labellisées"), the NanoRep grant funded by Paris Sciences et Lettres University, as well as core funding from the CNRS, the Ecole Normale Supérieure, and INSERM.

1. M. Thomas, R. L. White, R. W. Davis, Hybridization of RNA to double-stranded DNA: Formation of R-loops. *Proc. Natl. Acad. Sci. U.S.A.* **73**, 2294–2298 (1976).
2. I. Mellon, P. C. Hanawalt, Induction of the Escherichia coli lactose operon selectively increases repair of its transcribed DNA strand. *Nature* **342**, 95–98 (1989).
3. C. P. Selby, A. Sancar, Molecular mechanism of transcription-repair coupling. *Science* **260**, 53–58 (1993).
4. J. S. Park, M. T. Marr, J. W. Roberts, E. coli transcription repair coupling factor (Mfd protein) rescues arrested complexes by promoting forward translocation. *Cell* **109**, 757–767 (2002).
5. C. P. Selby, A. Sancar, Structure and function of transcription-repair coupling factor. I. Structural domains and binding properties. *J. Biol. Chem.* **270**, 4882–4889 (1995).
6. K. Howan *et al*., Initiation of transcription-coupled repair characterized at single-molecule resolution. *Nature* **490**, 431–434 (2012).
7. 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).
8. E. T. Graves *et al*., A dynamic DNA-repair complex observed by correlative single-molecule nanomanipulation and fluorescence. *Nat. Struct. Mol. Biol.* **22**, 452–457 (2015).
9. 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).
10. I. Husain, B. Van Houten, D. C. Thomas, M. Abdel-Monem, A. Sancar, Effect of DNA polymerase I and DNA helicase II on the turnover rate of UvrABC excision nuclease. *Proc. Natl. Acad. Sci. U.S.A.* **82**, 6774–6778 (1985).
11. M. Gómez-Marroquín *et al*., Stationary-phase mutagenesis in stressed Bacillus subtilis cells operates by Mfd-dependent mutagenic pathways. *Genes (Basel)* **7**, 33–45 (2016).
12. J. Han, O. Sahin, Y. W. Barton, Q. Zhang, Key role of Mfd in the development of fluoroquinolone resistance in Campylobacter jejuni. *PLoS Pathog.* **4**, e1000083 (2008).
13. H. Wimberly *et al*., R-loops and nicks initiate DNA breakage and genome instability in non-growing Escherichia coli. *Nat. Commun.* **4**, 2115 (2013).
14. M. N. Ragheb *et al*., Inhibiting the evolution of antibiotic resistance. *Mol. Cell* **73**, 157–165.e5 (2019).
15. S. Million-Weaver *et al*., An underlying mechanism for the increased mutagenesis of lagging-strand genes in Bacillus subtilis. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E1096–E1105 (2015).
16. A. Revyakin, C. Liu, R. H. Ebright, T. R. Strick, Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science* **314**, 1139–1143 (2006).
17. L. F. Liu, J. C. Wang, Supercoiling of the DNA template during transcription. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 7024–7027 (1987).
18. H. Y. Wu, S. H. Shyy, J. C. Wang, L. F. Liu, Transcription generates positively and negatively supercoiled domains in the template. *Cell* **53**, 433–440 (1988).
19. L. F. Westblade *et al*., Structural basis for the bacterial transcription-repair coupling factor/RNA polymerase interaction. *Nucleic Acids Res.* **38**, 8357–8369 (2010).
20. T. R. Strick, J. F. Allemand, D. Bensimon, A. Bensimon, V. Croquette, The elasticity of a single supercoiled DNA molecule. *Science* **271**, 1835–1837 (1996).
21. A. L. Chambers, A. J. Smith, N. J. Savery, A DNA translocation motif in the bacterial transcription–Repair coupling factor, Mfd. *Nucleic Acids Res.* **31**, 6409–6418 (2003).
22. J. Ma, L. Bai, M. D. Wang, Transcription under torsion. *Science* **340**, 1580–1583 (2013).
23. J. Ma *et al*., Transcription factor regulation of RNA polymerase's torque generation capacity. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 2583–2588 (2019).
24. A. M. Deaconescu *et al*., Structural basis for bacterial transcription-coupled DNA repair. *Cell* **124**, 507–520 (2006).
25. T. R. Strick, V. Croquette, D. Bensimon, Homologous pairing in stretched supercoiled DNA. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 10579–10583 (1998).
26. P. Hausen, H. Stein, Ribonuclease H. An enzyme degrading the RNA moiety of DNA-RNA hybrids. *Eur. J. Biochem.* **14**, 278–283 (1970).
27. J. Cairns, P. L. Foster, Adaptive reversion of a frameshift mutation in Escherichia coli. *Genetics* **128**, 695–701 (1991).
28. P. J. Hastings, H. J. Bull, J. R. Klump, S. M. Rosenberg, Adaptive amplification: An inducible chromosomal instability mechanism. *Cell* **103**, 723–731 (2000).
29. 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).
30. C. P. Selby, A. Sancar, Transcription-repair coupling and mutation frequency decline. *J. Bacteriol.* **175**, 7509–7514 (1993).
31. A. Schmidt *et al*., The quantitative and condition-dependent Escherichia coli proteome. *Nat. Biotechnol.* **34**, 104–110 (2016).
32. T. T. Le *et al*., Mfd dynamically regulates transcription via a release and catch-up mechanism. *Cell* **172**, 344–357.e15 (2018).
33. H. N. Ho, A. M. van Oijen, H. Ghodke, Single-molecule imaging reveals molecular coupling between transcription and DNA repair machinery in live cells. *Nat. Commun.* **11**, 1478 (2020).
34. J. Nuebler, G. Fudenberg, M. Imakaev, N. Abdennur, L. A. Mirny, Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E6697–E6706 (2018).
35. J. E. Walker, O. Luyties, T. J. Santangelo, Factor-dependent archaeal transcription termination. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E6767–E6773 (2017).
36. M. Tijsterman, J. Brouwer, Rad26, the yeast homolog of the cockayne syndrome B gene product, counteracts inhibition of DNA repair due to RNA polymerase II transcription. *J. Biol. Chem.* **274**, 1199–1202 (1999).
37. C. Troelstra *et al*., ERCC6, a member of a subfamily of putative helicases, is involved in Cockayne's syndrome and preferential repair of active genes. *Cell* **71**, 939–953 (1992).
38. M. Caputo *et al*., The CSB repair factor is overexpressed in cancer cells, increases apoptotic resistance, and promotes tumor growth. *DNA Repair (Amst.)* **12**, 293–299 (2013).
39. C. Duboc, J. Fan, E. T. Graves, T. R. Strick, Preparation of DNA substrates and functionalized glass surfaces for correlative nanomanipulation and colocalization (NanoCOSM) of single molecules. *Methods Enzymol.* **582**, 275–296 (2017).
40. T. R. Strick, J. F. Allemand, D. Bensimon, V. Croquette, Behavior of supercoiled DNA. *Biophys. J.* **74**, 2016–2028 (1998).