Supplementary Information for

Co-transcriptional R-loop formation by Mfd involves topological partitioning of DNA.

James R. Portman, Gwendolyn M. Brouwer, Jack Bollins, Nigel J. Savery, and Terence R. Strick

Corresponding author: Terence R. Strick
Email: strick@biologie.ens.fr

This PDF file includes:

Supplementary text
Figures S1 to S12
Table S1
SI References

Supplementary Text

DNA Sequences

The transcribed portion of the DNA construct used in the single-molecule experiments consists of a 970 bp transcription cassette containing the T5 phage early N25 promoter, followed by a roughly 900 bp transcript taken from a genic region of *E. coli*, followed by the intrinsic tR2 terminator from phage lambda. The transcription cassette sequence is, from 5' to 3':

ggtaccTTCGAGGGAAATCATAAAAAATTTATTTGCTTTCAGGAAAATTTTTCTGTATAATAGATT
CATAAATTTGAGCGGCGGTGGCAGCGGAATTCGAGGGCAGTTGCGGTCGTGGAACCCACC
GAGTGAAAGTGTGGATGCAGCCCTGTTGCCCAACTTTACCCGTGGCAATGCCCGCGCAGAC
GATCTGGTACGCAATAACGGCTATGCCGCCAACGCCATCCAGCTGCATCAGGATCATATCG
TCGGGTCTTTTTTCCGGCTCAGTCATCGCCCAAGCTGGCGCTATCTGGGCATCGGGGAGGA
AGAAGCCCGTGCCTTTTCCCGCGAGGTTGAAGCGGCATGGAAAGAGTTTGCCGAGGATGAC
TGCTGCTGCATTGACGTTGAGCGAAAACGCACGTTTACCATGATGATTCGGGAAGGTGTGG
CCATGCACGCCTTTAACGGTGAACTGTTCGTTCAGGCCACCTGGGATACCAGTTCGTCGCG
GCTTTTCCGGACACAGTTCCGGATGGTCAGCCCGAAGCGCATCAGCAACCCGAACAATACC
GGCGACAGCCGGAACTGCCGTGCCGGTGTGCAGATTAATGACAGCGGTGCGGCGCTGGGA
TATTACGTCAGCGAGGACGGGTATCCTGGCTGGATGCCGCAGAAATGGACATGGATACCCC
GTGAGTTACCCGGCGGTCGCGCCTCGTTCATTCACGTTTTTGAACCCGTGGAGGACGGGCA
GACTCGCGGTGCAAATGTGTTTTACAGCGTGATGGAGCAGATGAAGATGCTCGACACGCTG
CAGAACACGCAGCTGCAGAGCGCCATTGTGAAGGCGATGTATGCCGCCACCATTGAGAGTG
AGCTGGATACGCAGTCAGCGATGGATTTTATTCTGGGCGCGAACAGTCAGGAGCAGCGGGA
AAGGCTGACCAAGCTTACAAAACGCTCTGGTCGGCCTGGTACTAGTGTAGCACTCTGTGGG
ATATCCTGGAAAGCCCCCGGAAGATGCATCTTCCGGGGGGCTTTTTTTTTGGTTCggtacc

where the -35 and -10 elements of the N25 promoter are highlighted in cyan, the transcription start site is highlighted in red, and the tR2 terminator is highlighted in magenta. In lowercase are shown the KpnI sites used to insert this transcription cassette into the unique KpnI site of a "backbone" 3.6 kbp segment taken from the *T. aquaticus rpoC* gene. The backbone DNA segment was obtained by PCR amplification of the *rpoC* gene using primers containing XbaI and SbfI sites (underlined):

5'GAGAGACCTGCAGGGAAGTCCGCAAGGTCCGCAT 3'
and
5' GAGAGATCTAGAACAGGTCGTAGCCGTAGCAC 3'

and cloned into the unique XbaI and SbfI sites of homemade vector pET21aΔMCS, itself obtained by PCR amplification of pET21a (Novagen) using primers

5' GAGAGACCTGCAGGTATCGCCGACATCACCGATGGGGAAGATCG 3'
and
5' GAGAGATCTAGAGGCTGCTGCCACCGCTGAGCAATAACTAGC 3'

followed by digestion with XbaI and SbfI. The DNA construct bearing an unrelated 90-bp transcript (SI Appendix, Fig. S1e) is described in [1].
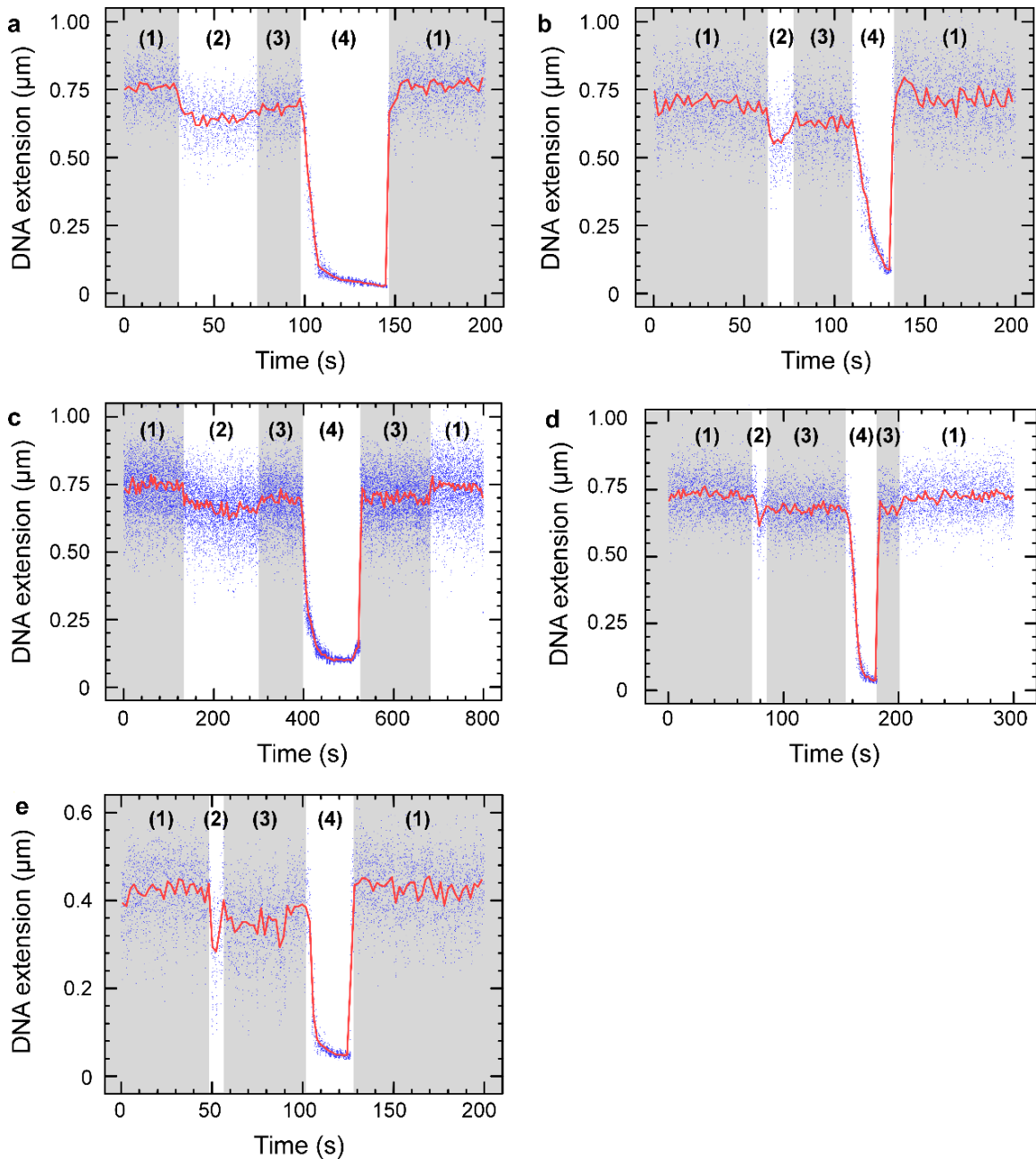
Figure S1: a-d) Additional time-traces showing Mfd interacts with elongating RNAP to form a tripartite supercoiled domain. RNAP and Mfd time-traces begin with (1) positively supercoiled DNA. RNAP then (2) initiates transcription and upon promoter escape converts into (3) an elongation complex, whereupon (4) Mfd 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, pulling the bead down towards the surface. Finally, (a, b) RNAP termination causes dissolution of topological domains, directly returning the DNA to (1) the baseline state. Alternatively, (c, d) loss of Mfd-RNAP interaction or Mfd dissociation from DNA dissolves the topological domains and restores (3) the elongation state which, upon intrinsic transcription termination at tR2, returns the DNA to (1) the baseline state. Blue points – raw data (30 Hz); red line – averaged data (1 s filtering). **e)** Tripartite supercoiled domain formation observed on an unrelated 90 bp transcript sequence with positive supercoiling and the 1-2-3-4 pattern of panels (a,b).
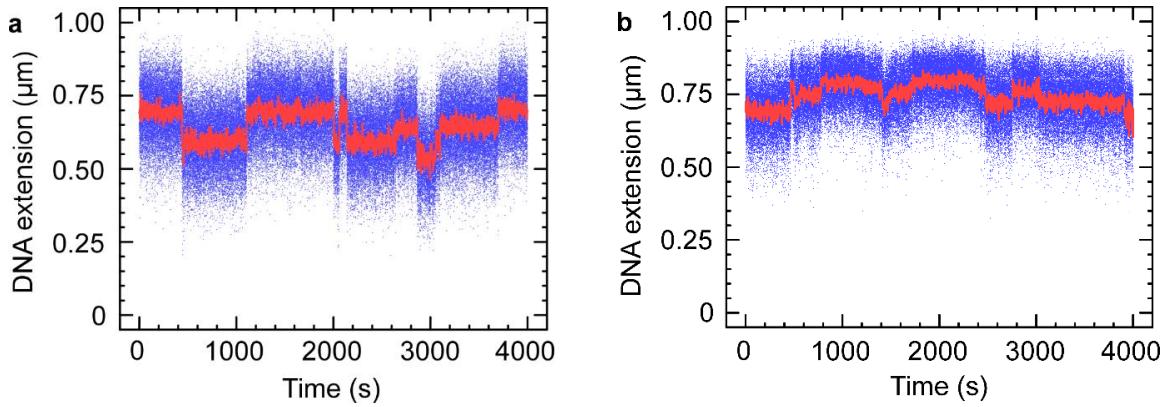
Figure S2: No Mfd controls. **a)** Time-trace showing iterative cycles of transcription by RNAP on positively supercoiled DNA. Across 33 DNA molecules over more than 18000s for each molecule, no topological domain-like events were observed. **b)** Time-trace showing iterative cycles of transcription by RNAP on negatively supercoiled DNA. Across 36 DNA molecules over more than 20000s for each molecule, no topological domain-like events were observed. On negatively supercoiled DNA transcription termination is less efficient and RNAP molecules can build up on the DNA, so occasionally via many ~50nm increases the DNA would reach the maximal extension state. This was readily distinguishable from events observed with Mfd present, and reason for our criteria of only measuring events with amplitudes greater than 200 nm (see Methods).
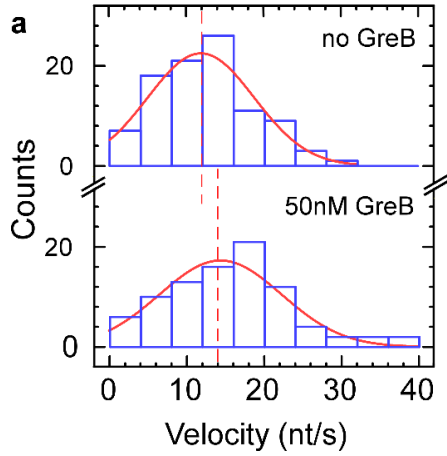
Figure S3: Velocity distributions obtained with wild-type Mfd and RNAP in the absence or presence of GreB. Both experiments were performed at high NTP concentration ([UTP] = [GTP] = [CTP] = [ATP] = 1mM). Histograms of velocities measured (blue) and Gaussian fits (red): (top panel) in the absence of GreB we obtain a mean differential velocity of 12 ± 0.8 nt/s (SEM, n = 96 events); (bottom panel) in the presence of 50nM GreB we obtain a mean differential velocity of 14 ± 1 nt/s (SEM, n = 88 events).
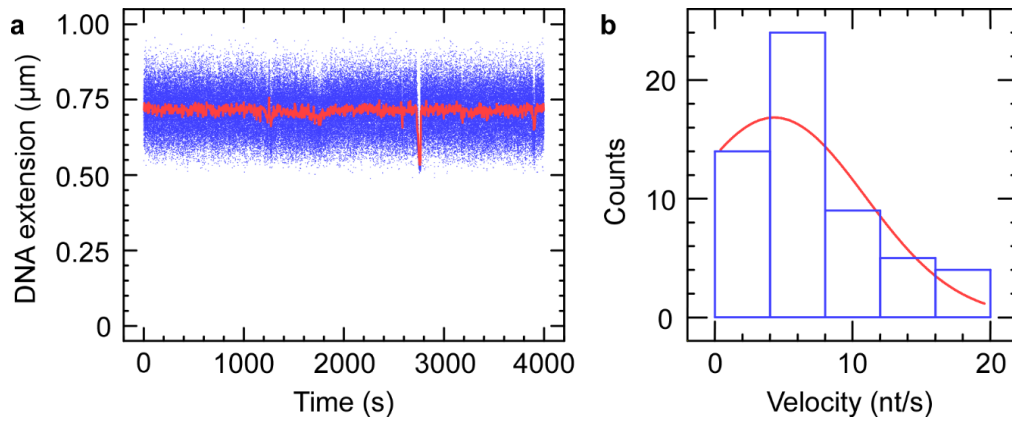
Figure S4: Wild-type Mfd alone control. **a)** Time-trace of 500nM wild-type Mfd and 1mM ATP on positively supercoiled DNA. Apparent unwinding events that lead to an apparent transient gain of positive supercoils are observed (here seen at ~2800s). **b)** Histogram of velocities derived from these unwinding events and Gaussian fit (red, n = 56; average = 4.3 ± 2.6 nt/s (SEM)).
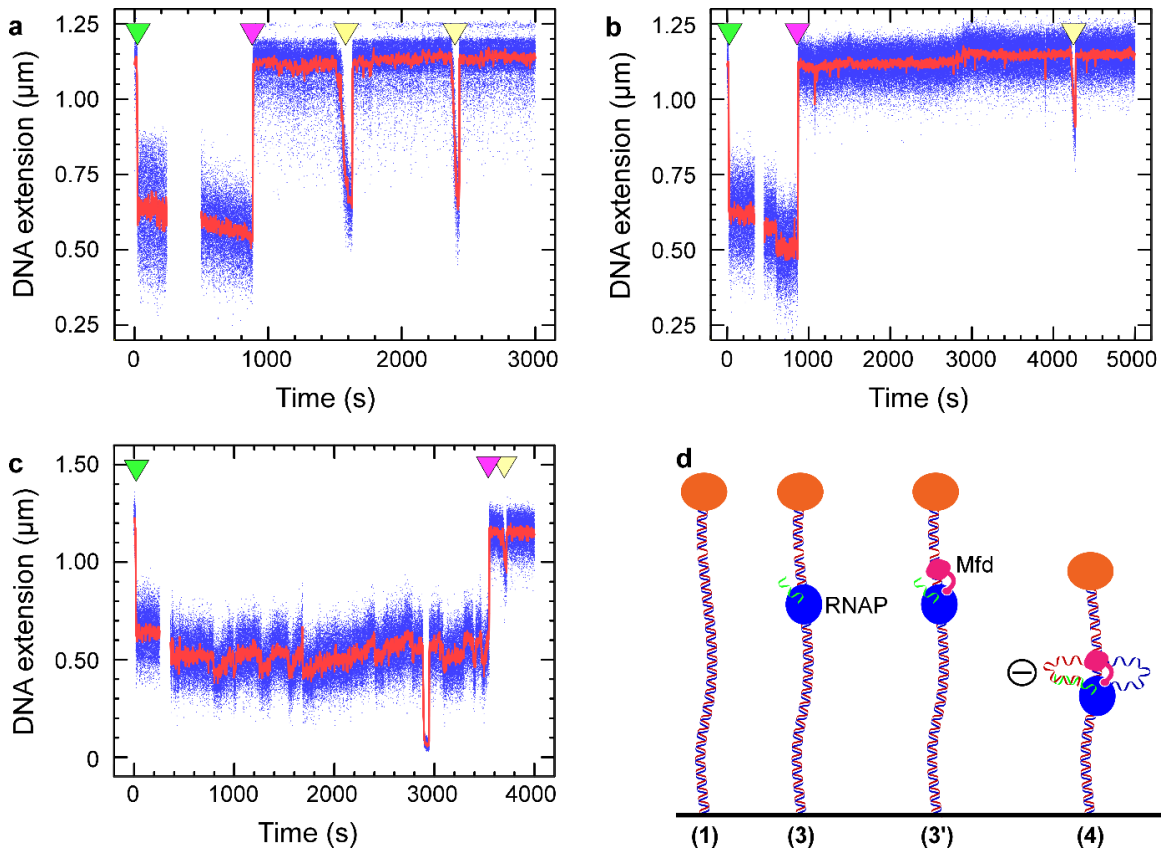
Figure S5: **(a-c)** Time-traces showing DNA compaction in nicked DNA molecules. The green arrow shows the point in time at which the DNA molecule is initially supercoiled from a maximal extension state. The break in data points is when RNAP, Mfd and NTPs were injected into the flow cell. The magenta arrow indicates the point at which the molecule becomes nicked, identified by the rapid loss of supercoils and the return to the maximal extension state. The yellow arrows indicate formation of the tripartite domains wherefore the differential velocity of the two motors reduces the length of DNA in the external domains and increases the length of DNA in the internal domain, causing the DNA extension to decrease (as shown in the model presented in **(d)**). Rates of bead descent for the events shown are a) 9 nm/s and 14 nm/s; b) 6 nm/s; and c) 7 nm/s. At the 0.3 pN extending force used, the DNA end-to-end extension factor is about 0.7x contour length, allowing us to convert bead velocities from nm/s into bp/s, with values of 38 bp/s, 59 bp/s, 25 bp/s and 29 bp/s, respectively.

7
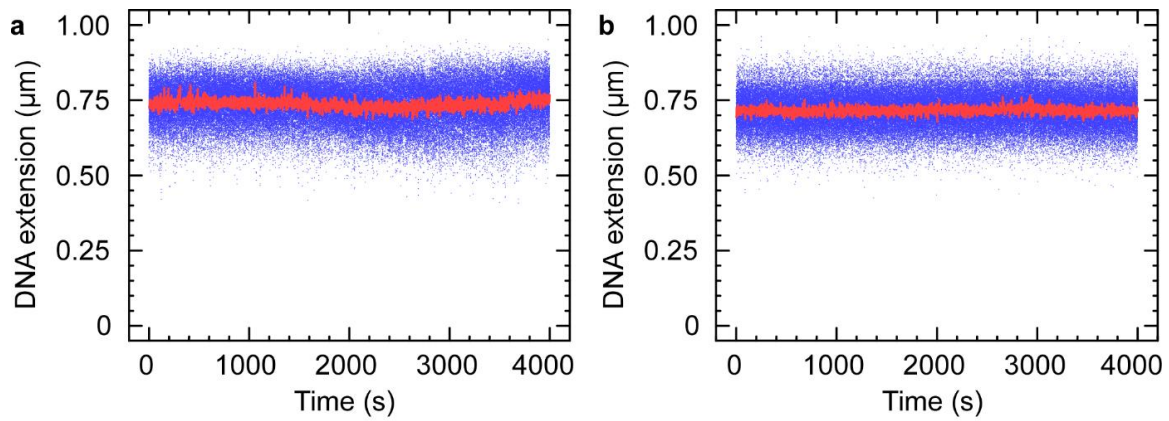
Figure S6: MfdRA953 only controls. **a)** Time-trace of positively supercoiled DNA with 500nM MfdRA953 and 1mM ATP. Across 20 DNA molecules over more than 20000s, no topological domain-like events were observed. **b)** Time-trace of negatively supercoiled DNA with 500nM MfdRA953 and 1mM ATP. Across 20 DNA molecules over more than 28000s, no topological domain-like events were observed.
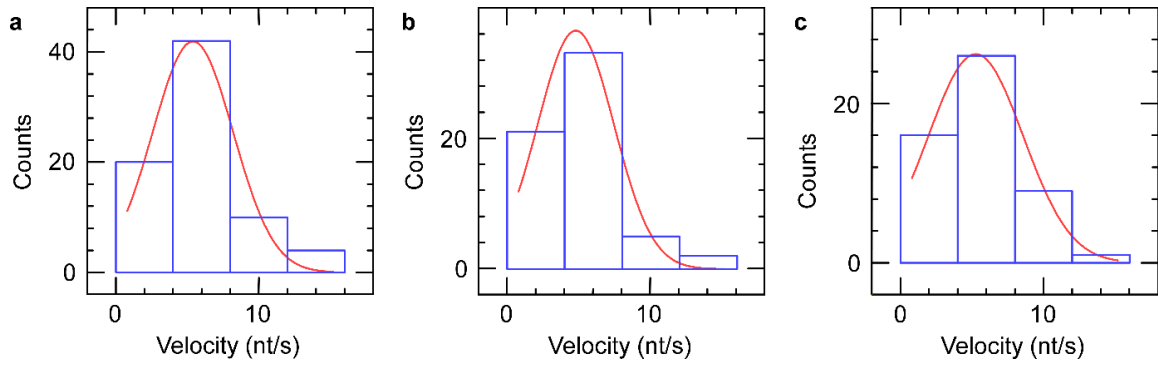
Figure S7: Velocity distributions for bead descent obtained using MfdRA953 and RNAP. **a)** Histogram of velocities measured on positively supercoiled DNA (blue) and Gaussian fit (red, n = 61; average = 4.8 ± 0.4 nt/s (SEM)). **b)** Histogram of velocities measured on negatively supercoiled DNA (blue) and Gaussian fit (red, n = 38; average = 5.4 ± 0.5 nt/s (SEM)). **c)** Histogram of velocities measured on negatively supercoiled DNA in the presence of 50nM GreB (blue) and Gaussian fit (red, n = 52; average = 5.3 ± 0.6 nt/s (SEM)).
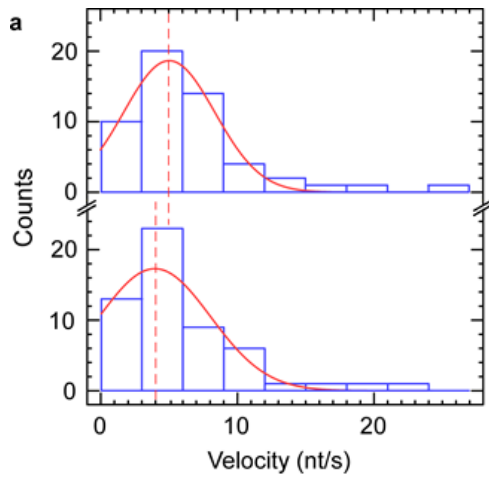
Figure S8: Velocity distributions for bead descent (blue) obtained using MfdLR499 (top panel) without RNAP and (bottom panel) with RNAP. Top panel – MfdLR499 with 1mM ATP on positively supercoiled DNA (n = 53; average = 5.0 ± 0.6 nt/s SEM from Gaussian fit in red). Bottom panel – MfdLR499 with RNAP and 1mM each NTP on positively supercoiled DNA (n = 55; average = 4.0 ± 1.1 nt/s SEM from Gaussian fit in red). There is no significant difference between the two distribution averages, indicating no RNAP-dependent events.
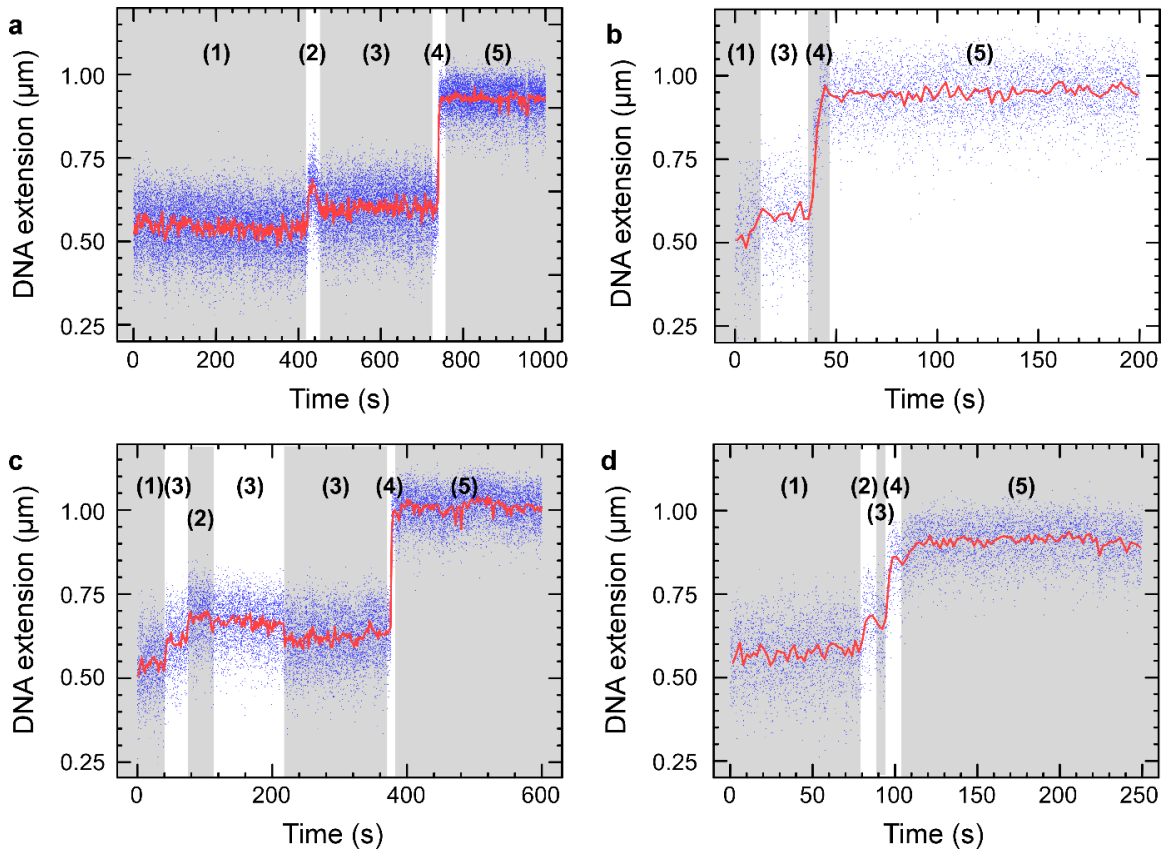
Figure S9: Additional time-traces showing formation of tripartite supercoiled domains leading to R-loops. Time-traces begin showing (1) negatively supercoiled DNA. As RNAP initiates it (2) scrunches and unwinds ~2 turns of DNA, increasing DNA extension by ~100 nm compared to baseline [2]. Successful promoter escape leads to formation of (3) the elongation complex characterized by a ~9 bp transcription bubble, and so DNA extension is increased by only ~50 nm compared to baseline. Then (4) 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, annihilating initial negative supercoiling and allowing the bead extension to increase to the maximal extension state. The long-lived state (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 which can be probed as in Fig. 3c.
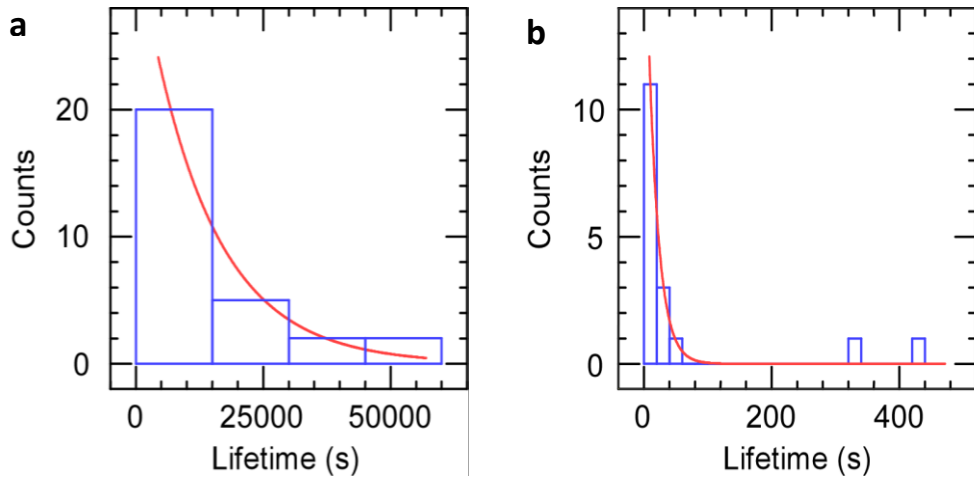
Figure S10: Histograms of R-loop lifetimes. **a)** Histogram of R-loop lifetimes measured in the absence of RNAse HI (blue) and exponential fit (red, n = 29; average = 13000s ± 3000 (SEM)). 50nM GreB was present in all experiments. **b)** Histogram of R-loop lifetimes measured in the presence of 0.025U/µl RNAse HI (blue) and exponential fit (red, n = 17; average = 16 ± 6 s (SEM)).
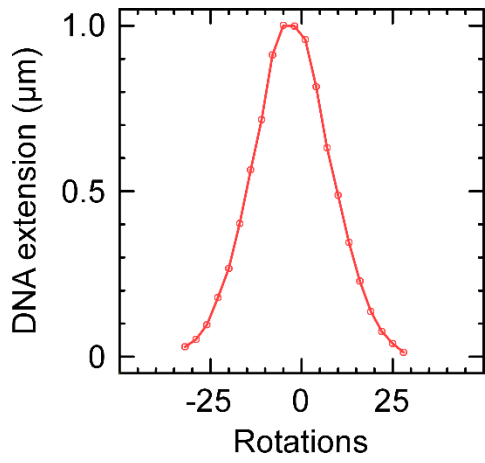
Figure S11: Rotation vs extension curve for the DNA molecule shown in Figure 3c prior to the addition of proteins.
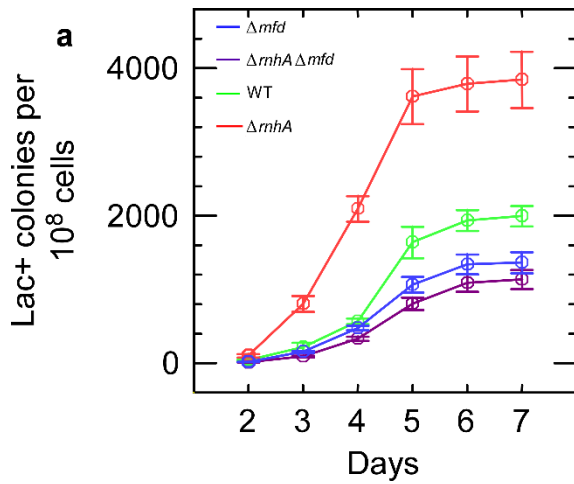
Figure S12: Lac revertant assay control. Cells lacking RNAse HI (PJH683; red) show about double the mutagenesis compared to cells with RNAse HI (SMR4562; green). Cells lacking both Mfd and RNAse HI (PJH946; magenta) show similar levels of mutagenesis as cells lacking Mfd (PJH813 + *pET21a^Lac-^*; blue), indicating R-loops are involved in the Mfd-dependent mutagenesis pathway. Colonies seen before day 2 arise due to replication-associated mutations whilst in liquid media and so are not counted [3]. Data is an average of three independent cultures and error bars represent SEM.

[GreB] (nM)	Total number of events (n)	Number of events passing through maximal extension state	Percentage of events passing through maximal extension state (%)
0	32	7	22 ± 8 (SEM)
50	73	24	33 ± 7 (SEM)

Table S1: Fraction of events with MfdRA953 initiated on negatively supercoiled DNA and that pass through the maximal extension state, resulting in positively supercoiled DNA in the external domain. GreB re-activates backtracked RNAP to enable more events to proceed further and pass the maximal extension state.

SI References

[1]   J. Fan, M. Leroux-Coyau, N. J. Savery, and T. R. Strick. *Reconstruction of bacterial transcription-coupled repair at single-molecule resolution.* Nature, 2016. 536(7615): p. 234–7.

[2]   A. Revyakin, C. Liu, R. H. Ebright, and T. R. Strick. *Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching.* Science, 2006. 314(5802): p. 1139–43,

[3]   J. Cairns and P. L. Foster. *Adaptive reversion of a frameshift mutation in Escherichia coli.* Genetics, 1991. 128(4): p. 695–701.