

Supplementary Information

P2-seq resolves topology-dependent DNA structural states and their long-range coupling in individual molecules

Nadezda A. Fursova^{1*}, Jan Hoinka^{2*}, Fedor Kouzine^{3*}, Daniel R. Larson¹, Teresa M. Przytycka^{2^}, and David Levens^{3^}

* These authors contributed equally and are listed in alphabetical order

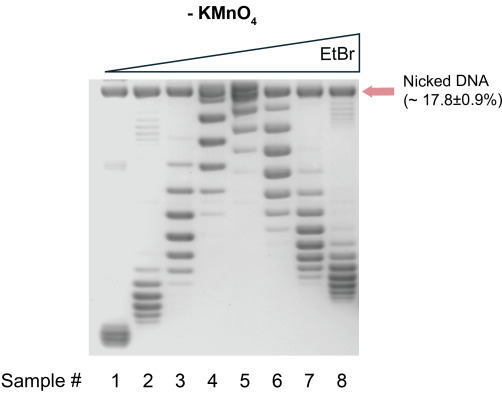
^Co-corresponding authors: przytyck@ncbi.nlm.nih.gov, levensd@mail.nih.gov

¹ Laboratory of Receptor Biology and Gene Expression, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA

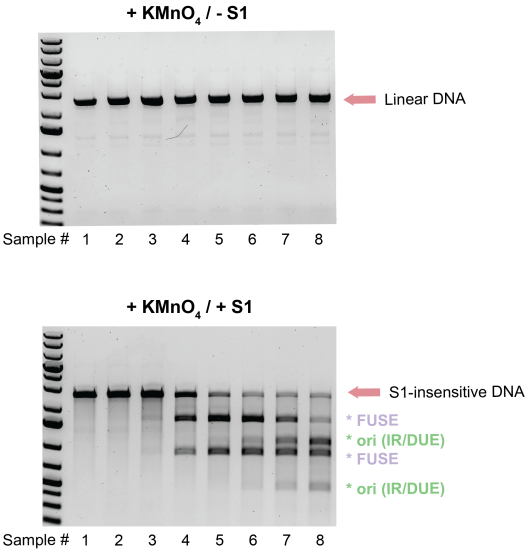
² National Library of Medicine, National Institutes of Health, Bethesda, MD, USA

³ Laboratory of Pathology, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA

A



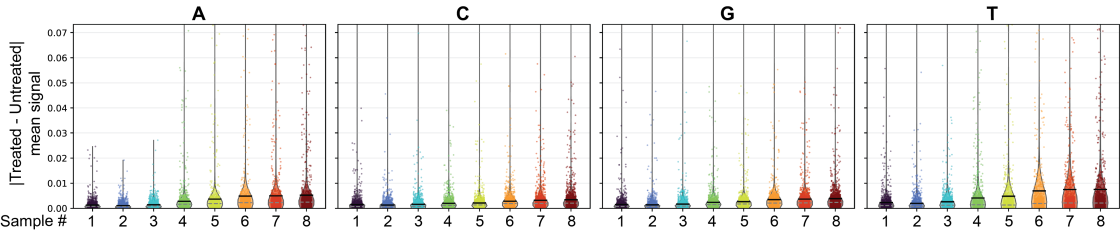
B



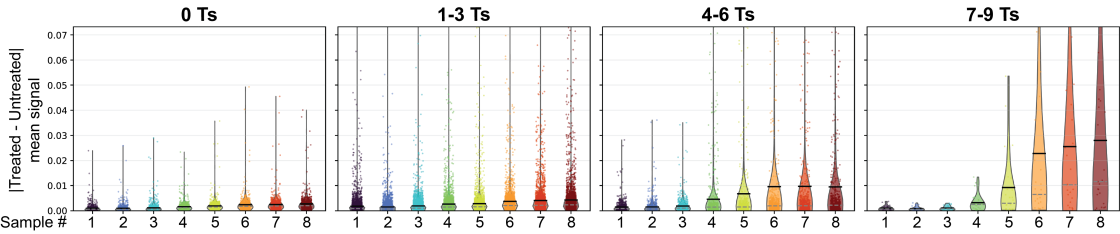
C

Sample #	1	2	3	4	5	6	7	8
ΔLk	0	-5	-9	-14	-18	-23	-27	-31
σ	0	-0.02	-0.04	-0.06	-0.08	-0.1	-0.11	-0.13
% S1-insensitive	86	80	72	46	27	17	16	16

D



E



Supplementary Figure 1. Characterization of supercoiled plasmid samples and supercoiling-dependent KMnO₄ reactivity.

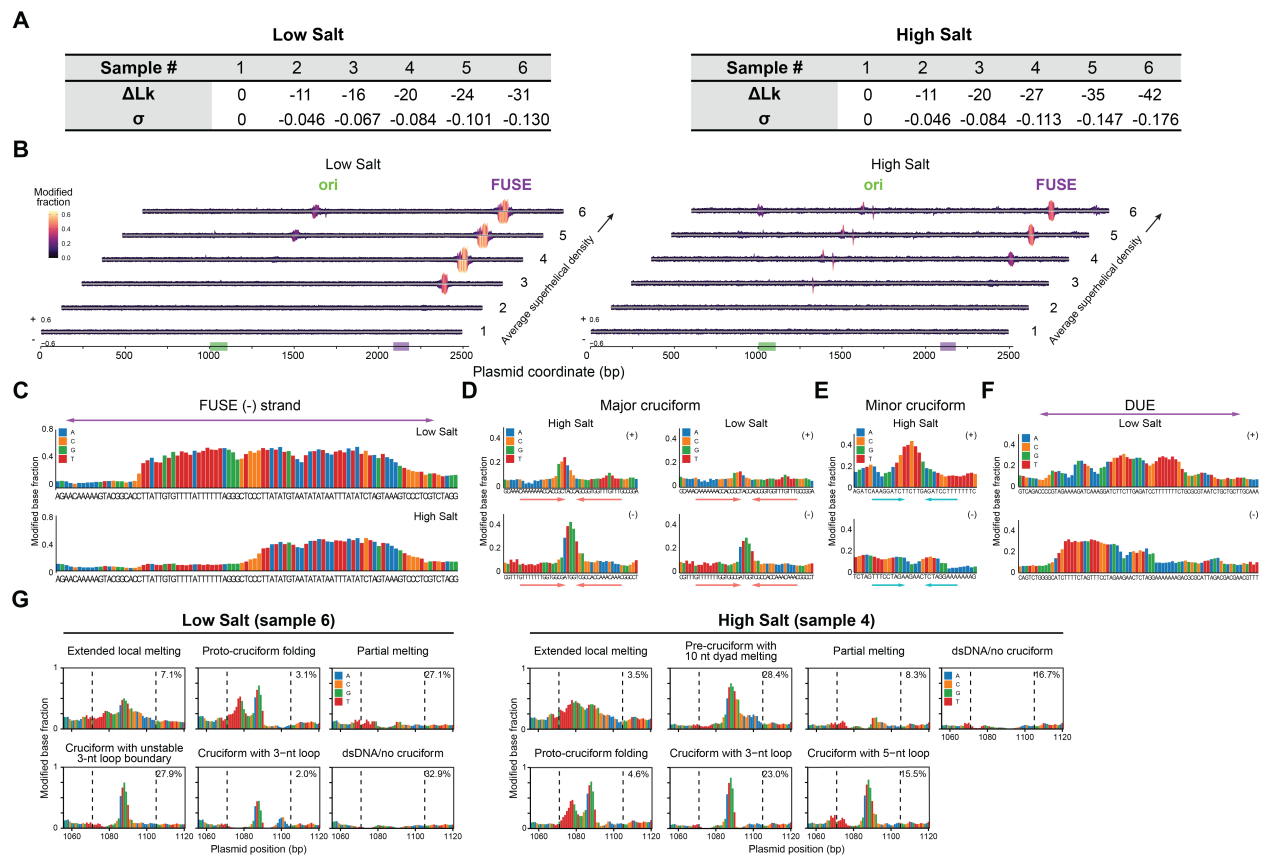
(A) Representative chloroquine agarose gel electrophoresis of plasmid topoisomer populations with defined superhelical densities, generated by topoisomerase relaxation in the presence of increasing concentrations of ethidium bromide (EtBr) under low-salt conditions. The linking-number difference (ΔLk) of the predominant topoisomer in each population was assigned by band counting relative to the relaxed reference topoisomer generated by topoisomerase treatment without EtBr. The arrow indicates nicked DNA, which comprised $17.8 \pm 0.9\%$ of total DNA (mean $\pm$ SE, $n_{\text{samples}} = 8$, $n_{\text{reps}} = 2$).

(B) Permanganate-S1 assay for supercoiling-dependent ssDNA formation. Topoisomer populations from (A) were treated with KMnO₄, linearized, and analyzed by agarose gel electrophoresis before (top) and after (bottom) S1 nuclease digestion. S1 treatment produced increasing amounts of fast-migrating DNA fragments with increasing supercoiling, indicating formation of KMnO₄-reactive single-stranded regions, including FUSE and ori-proximal IR/DUE regions, marked by purple and green asterisks, respectively. The arrows indicate full-length linear DNA in the top gel and the S1-insensitive fraction in the bottom gel. The left lane shows the GeneRuler 1 kb Plus DNA Ladder.

(C) Table summarizing ΔLk , calculated superhelical density σ , and S1-insensitive DNA fraction for each topoisomer population. σ was calculated as $\sigma = 10.5 \times \Delta Lk / N$, where N is the plasmid length (2498 bp). Band intensities were quantified using Image Studio (LI-COR Biosciences). Estimates of the S1-insensitive fraction for low-supercoiling samples 1-3 are approximate due to the low abundance of the S1-sensitive products, making quantification sensitive to background subtraction and integration window selection.

(D) Violin plots showing the absolute difference in mean k-mer-normalized ONT signal between KMnO₄-treated samples and the untreated control. Signals were refined using a k-mer-level table and rescaled before calculating differences at 9-mer resolution across supercoiled samples with increasing superhelical density under low-salt conditions. 9-mers were grouped by the nucleotide at the dominant position. Points represent individual 9-mers, and solid and dashed lines indicate the group mean and median, respectively.

(E) As in (D), for 9-mers grouped by total thymine content.



Supplementary Figure 2. Independent validation of P2-seq using the MinION R9.4.1 sequencing platform.

(A) Independent validation of P2-seq using a separate preparation of the mixed topoisomer populations shown in Fig. 1, sequenced on the MinION R9.4.1 platform. The tables summarize ΔLk and calculated superhelical density σ under low-salt (left) and high-salt (right) conditions, spanning the upper range of superhelical densities examined in Fig. 1. σ was calculated as $\sigma = 10.5 \times \Delta Lk / N$, where N is the plasmid length (2498 bp).

(B) Profiles of $KMnO_4$ -mediated DNA modification levels in mixed topoisomer populations with increasing superhelical density (1, lowest; 6, highest) under low- and high-salt conditions, shown separately for the + (top) and – (bottom) strands. The color scale and y-axis indicate the fraction of reads containing $KMnO_4$ -modified bases, calculated using a 3-nt median moving window. Positions of the FUSE element and plasmid origin of replication (ori) are indicated. Average P2-seq profiles generated on the MinION R9.4.1 flow cell are highly similar to profiles from samples with comparable superhelical densities sequenced on the PromethION R10.4 flow cell in Fig. 1.

(C) P2-seq signal across the FUSE element (single arrow) in the highest-supercoiling sample, shown for low- and high-salt conditions on the minus (-) strand. The y-axis shows the fraction of reads containing $KMnO_4$ -modified bases at each position, colored by nucleotide identity.

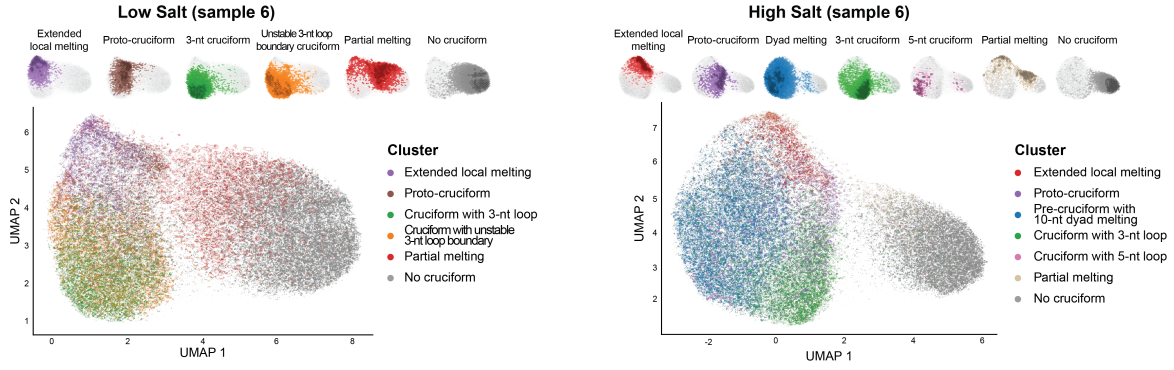
(D) As in (C), for the inverted repeat forming the major cruciform (double arrows), under high- and low-salt conditions, shown separately for the plus (+) and minus (-) strands.

(E) As in (D), for the minor cruciform under high-salt conditions.

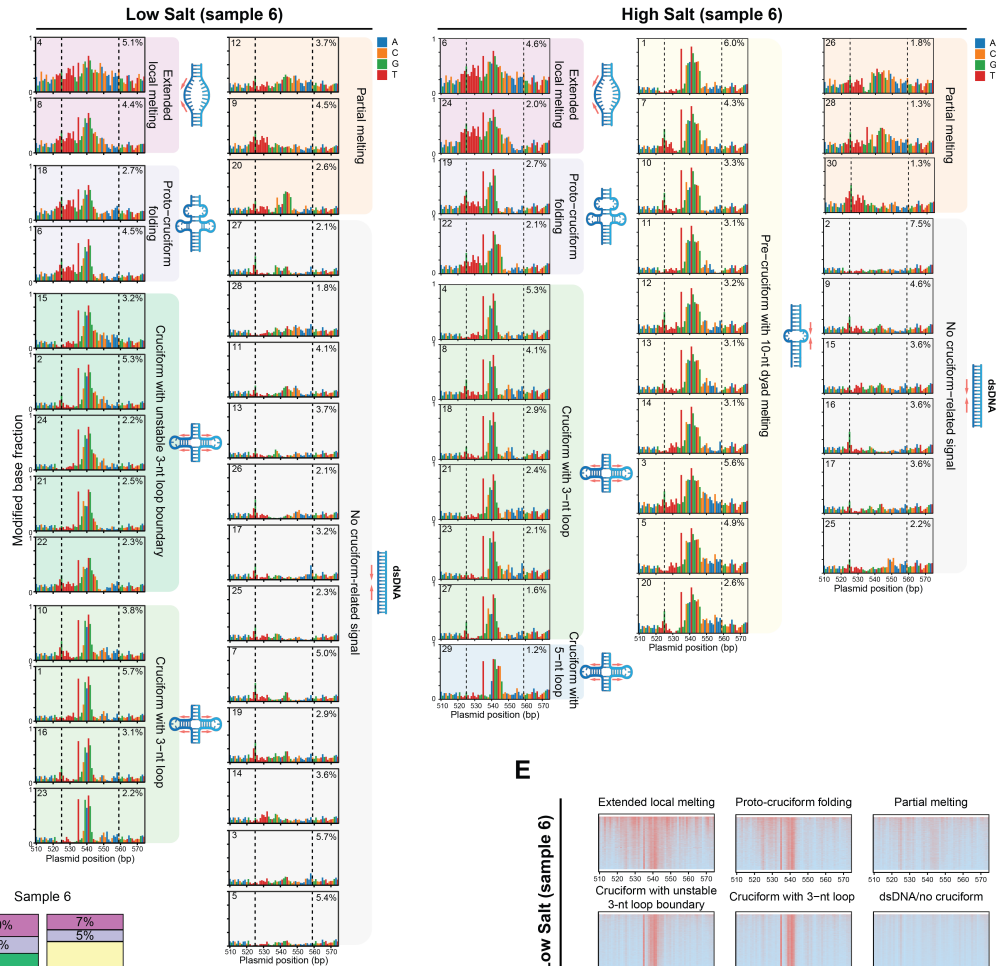
(F) As in (C), for the DNA unwinding element (DUE; single arrow) under low-salt conditions, shown separately for the plus (+) and minus (-) strands.

(G) Bar charts showing average P2-seq profiles for the major cruciform modification state classes and their read fractions, identified using single-read clustering of the inverted repeat region under low-salt (sample 6) and high-salt (sample 4) conditions. The y-axis shows the fraction of minus-strand reads containing $KMnO_4$ -modified bases at each position, with bars colored by nucleotide identity. Vertical dashed lines mark inverted repeat boundaries. This analysis recovers the same major cruciform classes as Fig. 3 (C-type intermediates under low-salt conditions, and both C-type and S-type intermediates for high-salt conditions), with comparable proportions of intermediate, fully extruded, and cruciform-negative states. Minor differences reflect independent plasmid preparations with slightly different superhelical densities and topoisomer distributions, together with differences in R9 chemistry and MinION data processing.

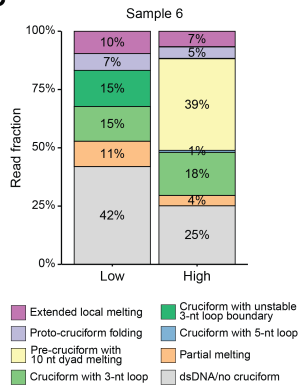
A



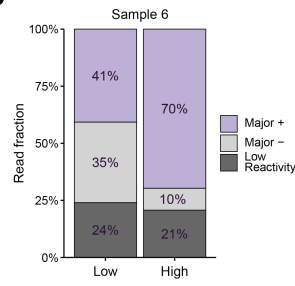
B



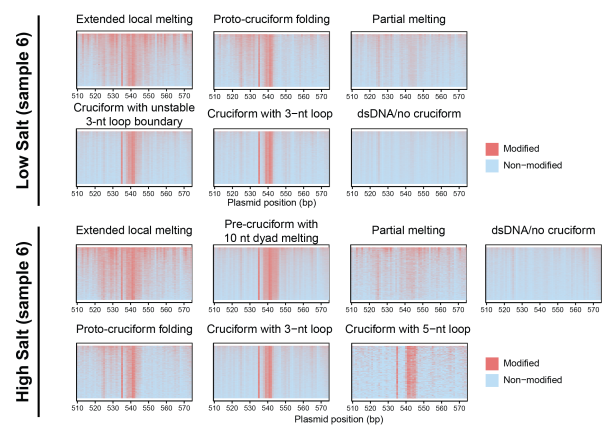
C



D



E



Supplementary Figure 3. Single-read analysis of major cruciform modification states under different ionic conditions.

(A) UMAP visualization of single-read modification scores (Rembo p-values) across the major cruciform-forming region under low- and high-salt conditions (sample 6), showing the separation of identified modification state classes. Main panels: reads are colored by the consolidated modification state class, with point size and opacity increasing with local read density. Top panels: the density distribution for each class is plotted using the same UMAP coordinates, with all other reads shown in gray. Increasing color intensity indicates higher local density, and outlined hexagons mark bins within the top 10% density for that class.

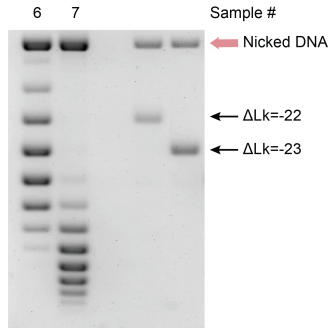
(B) P2-seq profiles of major cruciform KMnO₄ modification classes in the moderately supercoiled topoisomer population (sample 6) under low-salt (left) and high-salt (right) conditions. Each panel shows the fraction of minus-strand reads containing KMnO₄-modified bases at each position, colored by nucleotide identity, for individual clusters. Cluster ID and size (% of total reads) are indicated. Dashed vertical lines mark the inverted repeat boundaries. Clusters were defined by single-read clustering of the major cruciform-forming inverted repeat and assigned to putative structural intermediates based on modification patterns and previously proposed cruciform-folding mechanisms.

(C) Bar charts showing the fraction of reads assigned to the inferred major cruciform structural intermediates defined in (B) under low- and high-salt conditions.

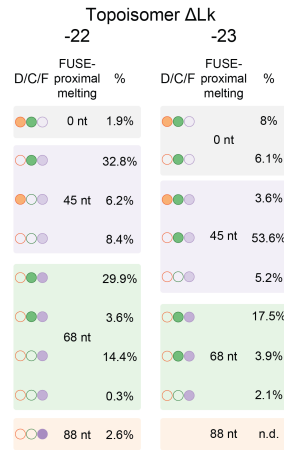
(D) Bar charts showing the fraction of reads assigned to major cruciform-positive (Major+), major cruciform-negative (Major-), and low-reactivity classes in the moderately supercoiled topoisomer population (sample 6) under low- and high-salt conditions. Classes were defined by single-read clustering of the modification patterns for each predominant non-B element, with reads lacking element-specific modification signatures across all elements assigned to the low-reactivity population. Minor deviations from (C) reflect the differences in input samples, target regions, and clustering depth.

(E) Binary heatmaps of single-read KMnO₄ modification profiles for the merged structural classes defined in (B) under low- (top) and high- (bottom) salt conditions. Rows represent individual reads, sorted by total modification score, with each position colored by modification status (red, modified (p-value < 0.02); blue, non-modified). Reads within each class show consistent modification patterns matching the ensemble profile, indicating clean resolution of the modification states used to infer structural intermediates.

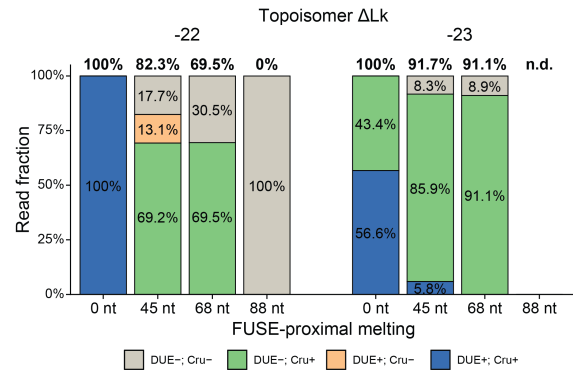
A



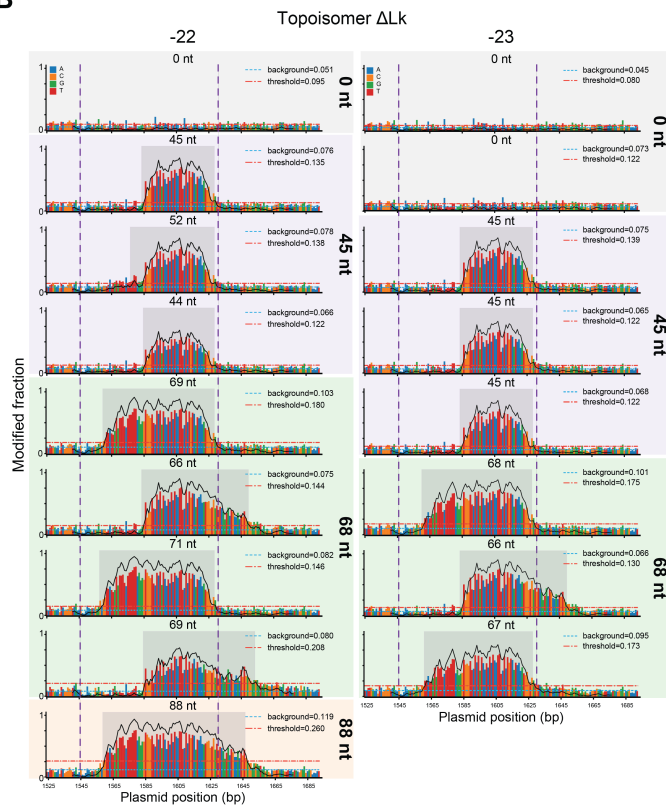
C



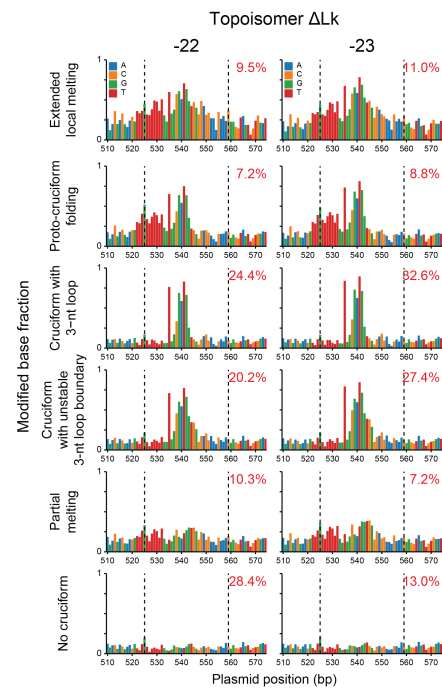
D



B



E



Supplementary Figure 4. Single-molecule analysis of FUSE-proximal melting and major cruciform modification state classes in individual topoisomers.

(A) Individual purified topoisomers $\Delta Lk = -22$ and $\Delta Lk = -23$, shown alongside the topoisomer populations in samples 6 and 7 from Supplementary Fig. 1A.

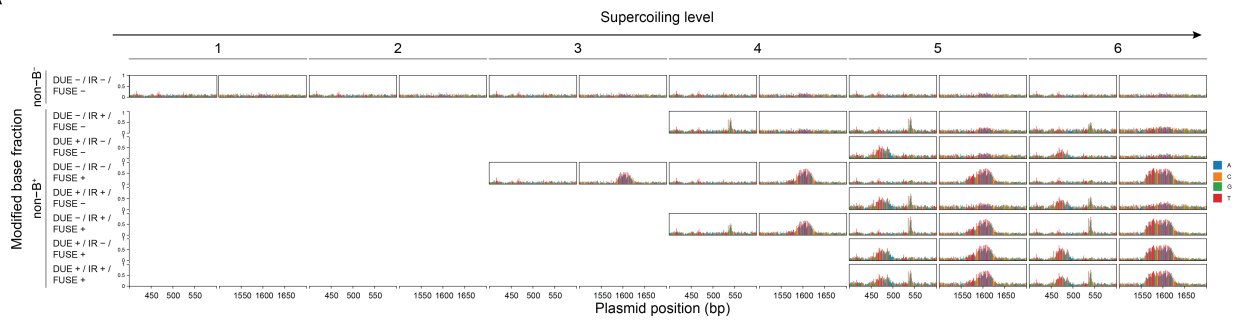
(B) Average P2-seq profiles at the FUSE-proximal region across structural classes defined in Fig. 4A-B for purified topoisomers $\Delta Lk = -22$ (left) and $\Delta Lk = -23$ (right) under low-salt conditions. The value above each panel indicates the estimated length of the FUSE-proximal melted region for that structural class, with classes further grouped by their average melted length across both topoisomers (labels on the right). The fraction of minus-strand reads containing $KMnO_4$ -modified bases at each position, colored by nucleotide identity, is shown in each panel. Dashed vertical lines mark the FUSE element boundaries, shaded gray rectangles highlight the estimated melted region, and horizontal dashed lines indicate the cluster-specific background and detection thresholds.

(C) Fraction of reads in each structural class defined in Fig. 4A-B for purified topoisomers $\Delta Lk = -22$ (left) and $\Delta Lk = -23$ (right), grouped by FUSE-proximal melting state (0, 45, 68, 88 nt). D/C/F filled/open circles denote presence/absence of modification signature at each non-B DNA element (orange = DUE, green = major cruciform, purple = FUSE). For FUSE, fill intensity corresponds to the melting level. n.d. = not detected. Total number of clustered reads is $n = 46,806$ for $\Delta Lk = -22$ and $n = 88,686$ for $\Delta Lk = -23$.

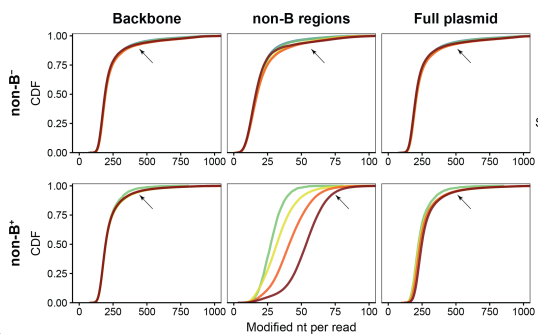
(D) Joint distribution of major cruciform and melted DUE modification signatures across FUSE-proximal melting levels within individual topoisomers. Bars indicate the read fraction in each class based on the joint clustering analysis in Fig. 4. Labels above the bars show the fraction of reads positive for at least one additional non-B signature (DUE melting, major cruciform, or both) at each FUSE melting level. n.d. = not detected.

(E) Average P2-seq profiles of merged major cruciform modification state classes defined in Fig. 4D for purified topoisomers $\Delta Lk = -22$ and $\Delta Lk = -23$ under low-salt conditions. Each panel shows the fraction of minus-strand reads with $KMnO_4$ -modified bases at each position, colored by nucleotide identity, with cluster size (% of total reads) indicated. Dashed vertical lines mark the inverted repeat boundaries. Total number of clustered reads is $n = 51,387$ for $\Delta Lk = -22$ and $n = 85,709$ for $\Delta Lk = -23$.

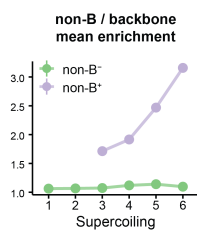
A



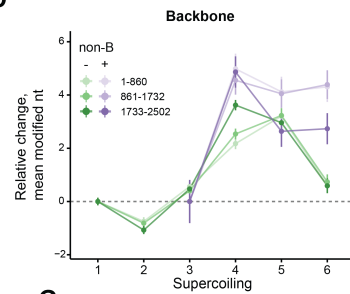
B



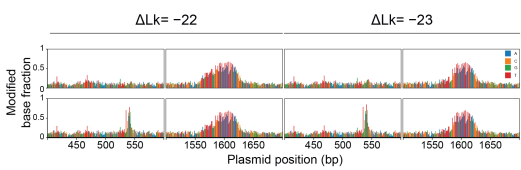
C



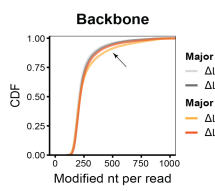
D



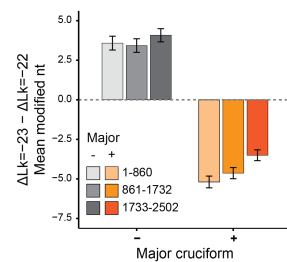
E



F



G



Supplementary Figure 5. Permanganate reactivity of plasmid backbone regions captures supercoiling-induced duplex destabilization.

(A) Average P2-seq profiles across the 400-600 and 1500-1700 plasmid regions for non-B⁻ and non-B⁺ molecules under low-salt conditions. Structural classes were defined by single-read clustering of minus-strand reads based on modification signatures of individual non-B DNA structures. Each panel shows the average fraction of minus-strand reads with KMnO₄-modified bases at each position, colored by nucleotide identity.

(B) Cumulative distributions of the number of minus-strand KMnO₄-modified nucleotides per read for non-B⁻ and non-B⁺ molecules at increasing supercoiling levels, shown across backbone sequences, predominant non-B regions, and the full plasmid. Arrows highlight major differences between the supercoiled samples.

(C) Mean enrichment of minus-strand P2-seq signal at non-B regions relative to backbone sequences in non-B⁻ and non-B⁺ molecules at increasing supercoiling levels. Points show the mean $\pm$ SE. Where not visible, the error bars are smaller than the marker. No point is shown when the corresponding structural class was not detected at a given supercoiling level.

(D) Supercoiling-dependent change in minus-strand KMnO₄ modification signal in non-B⁻ (green) and non-B⁺ (purple) molecules across three consecutive, equally sized plasmid backbone regions, excluding regions containing the predominant non-B structures (1-860, 861-1732, 1733-2502). For each structural class and interval, the mean number of modified nucleotides per read at each supercoiling level was compared to the lowest-supercoiling sample (dashed baseline). Points and error bars show differences in means $\pm$ propagated SE across individual reads. No point is shown when the corresponding structural class was not detected at a given supercoiling level.

(E) As in (A), for purified topoisomers $\Delta Lk = -22$ and $\Delta Lk = -23$, classified by the presence or absence of a major cruciform signature as defined in Fig. 4D.

(F) As in (B), for purified topoisomers $\Delta Lk = -22$ and $\Delta Lk = -23$ and only for backbone sequences, classified by the presence or absence of a major cruciform signature.

(G) As in (D), for purified topoisomers $\Delta Lk = -22$ and $\Delta Lk = -23$, classified by the presence or absence of a major cruciform signature.