Electronic Supplementary Material (ESI) for Nanoscale.

SUPPLEMENTARY FIGURES & TABLES

Fingerprinting Branches on Supercoiled Plasmid DNA Using Quartz Nanocapillaries

M. Sumanth Kumar, Koushik S. & G. V. Soni*
Raman Research Institute, Bangalore, INDIA

This supplementary file contains:

Figures: S1-S9

Tables: S1-S5

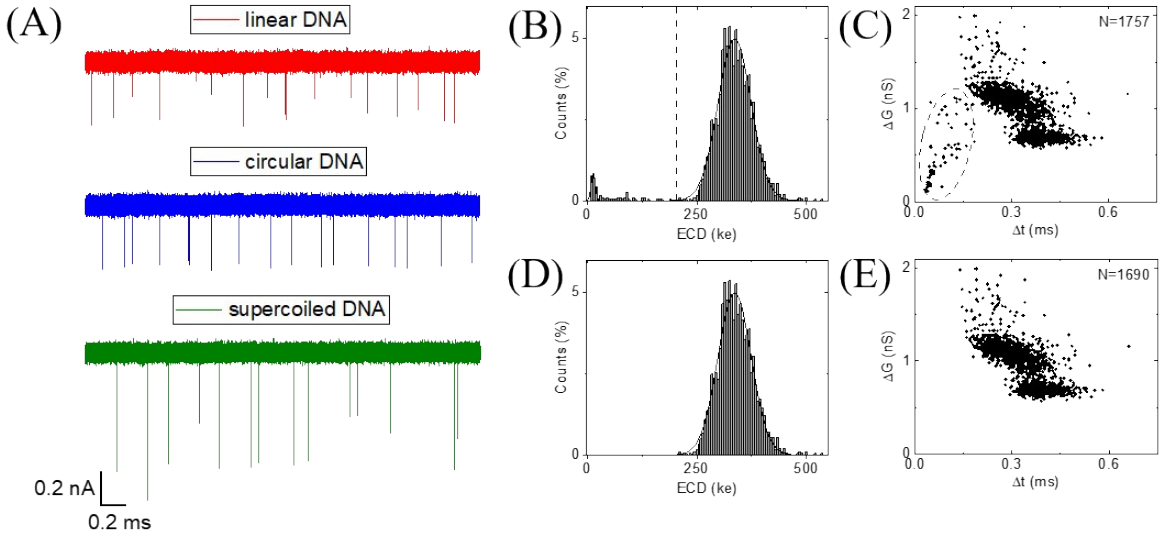

Figure S1: Current trace, identification and elimination of collision peaks from nanopore data. Data presented here is for nanopore P31 of 19 nm, measuring translocation events of linear plasmid DNA at 300 mV. (A) Current trace for different conformations of plasmid DNA from P31. (B) & (C) show the ECD histogram and ΔG-Δt scatter plot respectively for lin-DNA without any ECD filter. We can see the presence of collision events. (D) & (E) show the ECD histogram and scatter plot respectively for the same data after filtering events only above the ECD threshold.

It must be noted that collision peaks are seen in all nanopore experiments and can interfere with the interpretation of results. It is important to identify and eliminate them. We find that ECD shows two peaks at 14 ± 5 ke- and 336 ± 38 ke- ⁻. The events corresponding to the higher value are due to translocation and the events with lower value correspond to collision peaks. We filter out the collision events from the dataset by choosing only the events with $ECD > \mu - 3.5 * \sigma$. Here, μ and σ are the mean and standard deviation of the ECD peak for our sample.

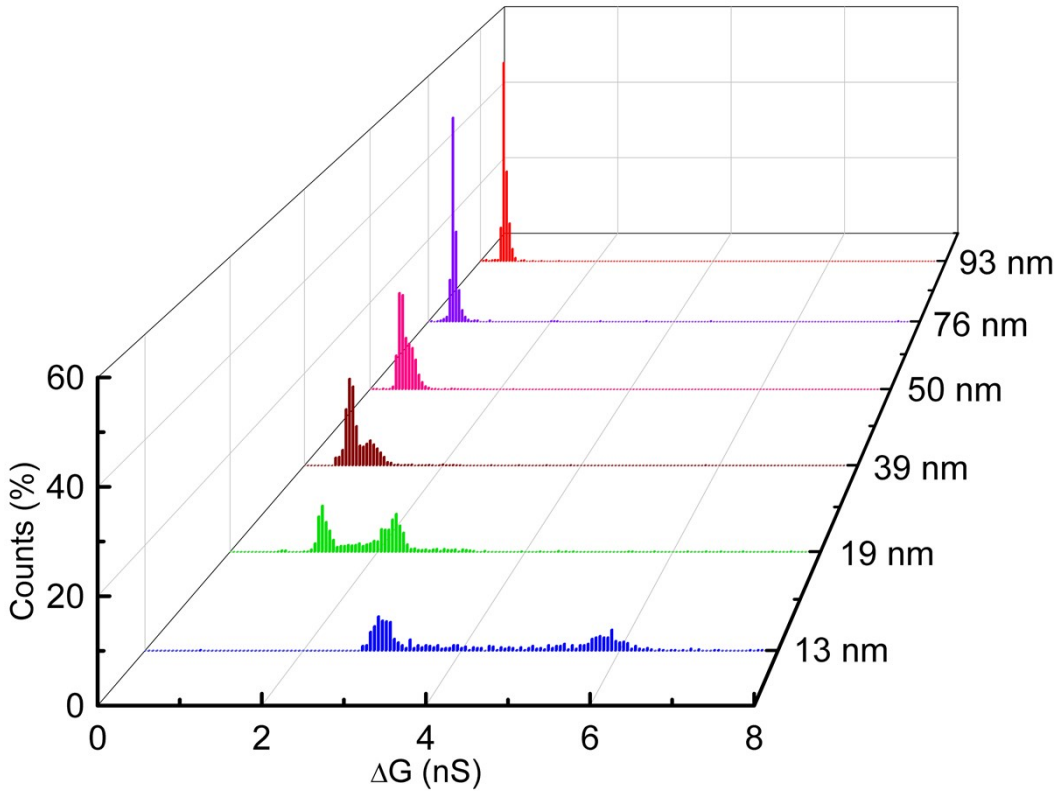

Figure S2: Role of nanopore size in resolving branched populations of supercoiled DNA. Histograms of conductance drop are compared for various pore diameters – Pores used here are: P24 (13 nm), P31 (19 nm), P30 (39 nm), P08 (50 nm), P02 (76 nm), and P03 (93 nm) (see also Fig 3 in main text).

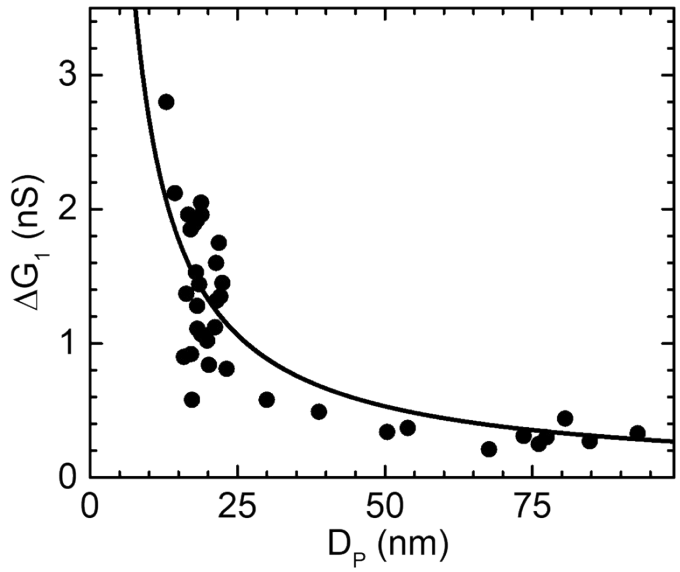

Figure S3: Conductance drop (1st peaks) (300 mV) during translocation of a supercoiled plasmid is plotted as a function of pore diameter when measured for N = 37 different nanopores. The solid line is the fit to eqn (S1) (Fig 3A (inset) is reproduced here for detailed explanation).

We performed translocation of sc-plasmid DNA in a total of 37 nanopores with a spectrum of sizes ranging from 13 nm to 93 nm and the conductance drop due to unbranched form of supercoiled DNA is fitted to the following equation,

$$\Delta G_1(D_p) = \sigma\left(\left(\frac{L_{eff}}{A_P}+\frac{1}{2D_P}\right)^{-1}-\left(\frac{L_{eff}}{(A_P-A_{sc-DNA})}+\frac{1}{2\sqrt{\frac{A_P-A_{sc-DNA}}{\pi}}}\right)^{-1}\right)$$

$$= \sigma\left(\left(\frac{4K}{\pi D_P}+\frac{1}{2D_P}\right)^{-1}-\left(\frac{4K}{\pi D_P\left(1-\frac{2D_{DNA}^2}{D_P^2}\right)}+\frac{1}{2\sqrt{(D_P^2-2D_{DNA}^2)}}\right)^{-1}\right) \quad \text{(S1)}$$

Here, $A_P = \pi D_P^2/4$, $A_{sc-DNA} = 2 \times \pi D_{DNA}^2/4$ and $K = L_{eff}/D_P$ where L_{eff} is the effective length of the pore. Substituting the diameter of lin-DNA (D_{DNA}) which is 2.2 nm, the fitting gives us $L_{eff} = (3.94 \pm 0.26) \times D_P$. As we expect, the ΔG of the first peak (ΔG_1) decreases with an increase in pore size. The variation in ΔG_1 values for similar pore sizes is due to the variation in internal profiles of the capillaries which are not accessible to us. A cluster of data points is visible around 20 nm because our experimental investigations are primarily focused on pores with diameters in this size range.

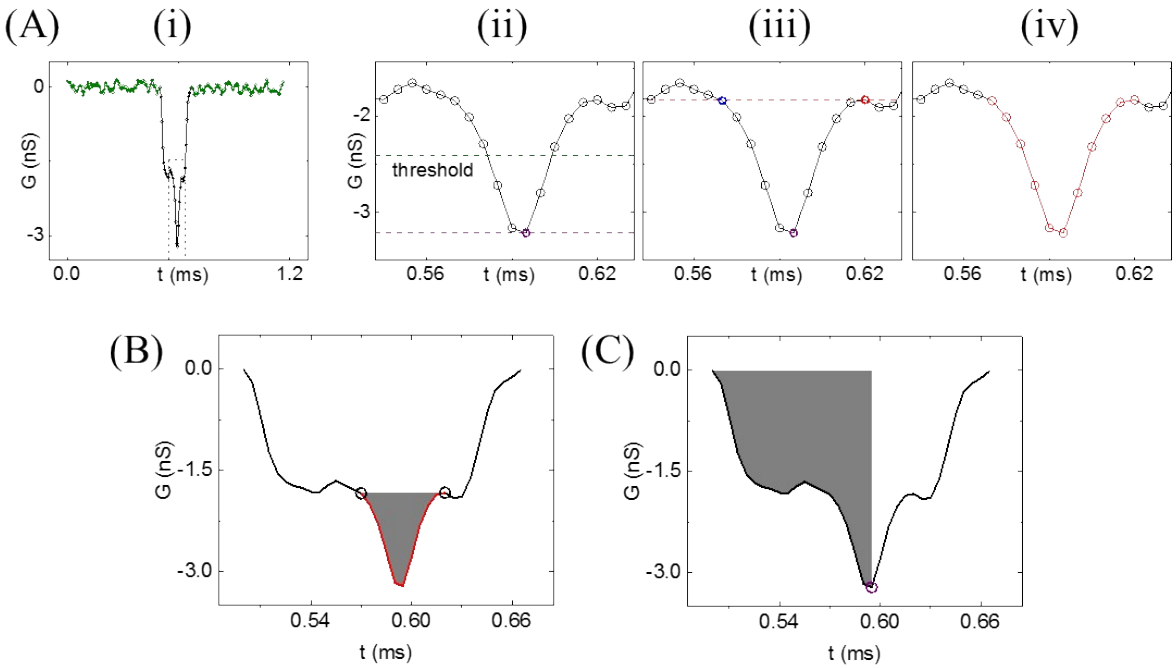

Figure S4: Branch identification and characterization from spikes in translocation event: (A) shows representative event (i) of sc-DNA translocating through a 17 nm nanopore (P25). The green trace is the baseline and the black trace is the translocation event with the current spike due to the branched region marked as the dotted region. Zoom of this current spike (dotted region) is shown on the right in steps (ii) – (iv), describing the three-step event processing. (B) & (C) shows the full event (black trace in (A)(i)) with the ECD of the branched region (red) and the ECD up to the location of the current spike (purple circle) shaded along the raw event.

We exploit the ECD for characterizing features on DNA for both linear and supercoiled DNA samples. We collect all events with spikes by considering events with $\Delta G > \Delta G_1 + 2\sigma_1$ and $\Delta G < \Delta G_2 + 2\sigma_2$ as branched (sc-DNA) events. The events with $\Delta G > \Delta G_2 + 2\sigma_2$ have faster dwell times and do not contain resolvable spikes as they possibly represent highly compact structures. Here σ_1 and σ_2 are the standard deviations in the gaussian fits to first and second ΔG peaks respectively. We employ a simple algorithm to detect the spike within each event for quantitative analysis.

(ii): Find the data point index and the value of the global minimum (purple circle). Starting from a threshold (green dotted line, $T_h = 0.25 \times \Delta G$) above this global minimum, find on either side, the first data point where the slope of the curve changes sign.

(iii): Horizontal line from this data point defines the begin (blue circle) and the end (red circle) of the spike.

(iv): Trace between the start and end points defines the spike region for ECD calculation.

(B) Branch length (shaded region) is calculated as follows

$$L_b = \frac{ECD_b}{ECD_{event}} \times L_{sc}$$

where ECD_{event} and ECD_b are respectively the ECDs of the entire event and the branched region only, L_{sc} is the linear length of sc-DNA which is 514.25 nm for 3025 bp pGEM3z plasmid.

(C) Relative position Z_b (shaded region) of the branch is calculated as follows:

$$Z_b = \frac{ECD_Z}{ECD_{event}}$$

where, ECD_Z is the ECD of the event up to the global minimum.

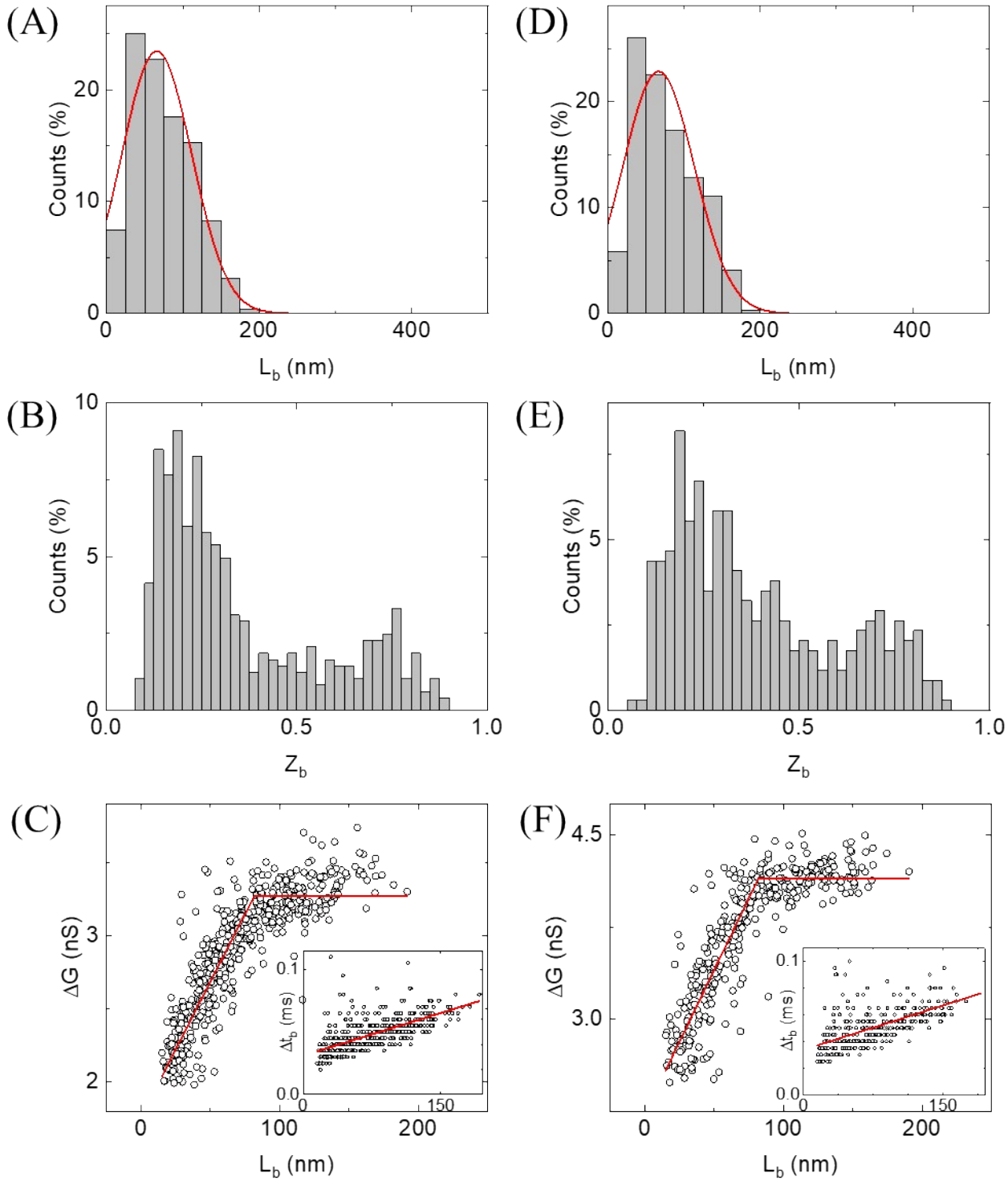

Figure S5: Analysis of branches in sc-DNA as measured in nanopores P12 (n = 1193), P19 (n = 745)

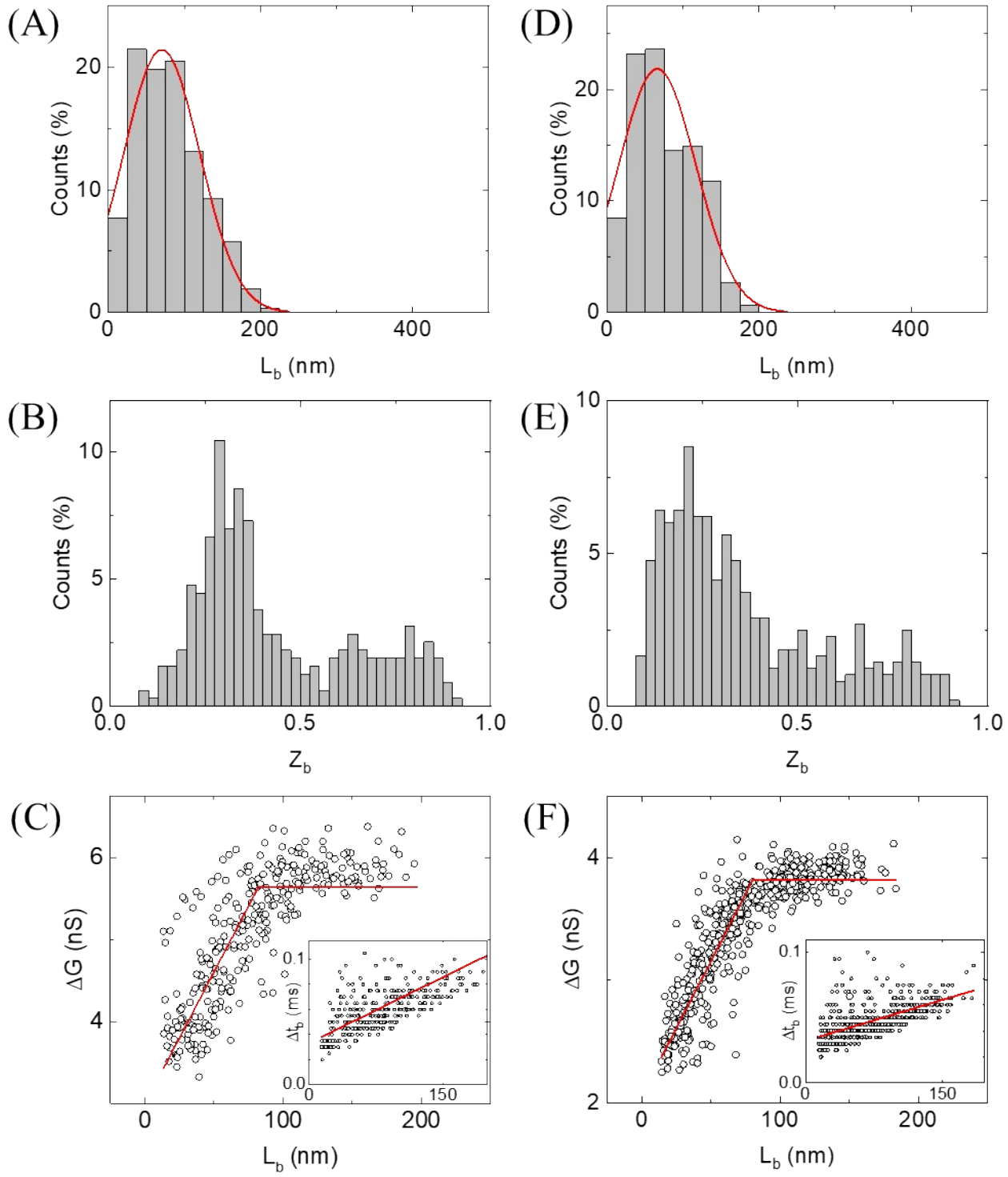

Figure S6: Analysis of branches in sc-DNA as measured in nanopores P24 (n = 755), P26 (n = 1009)

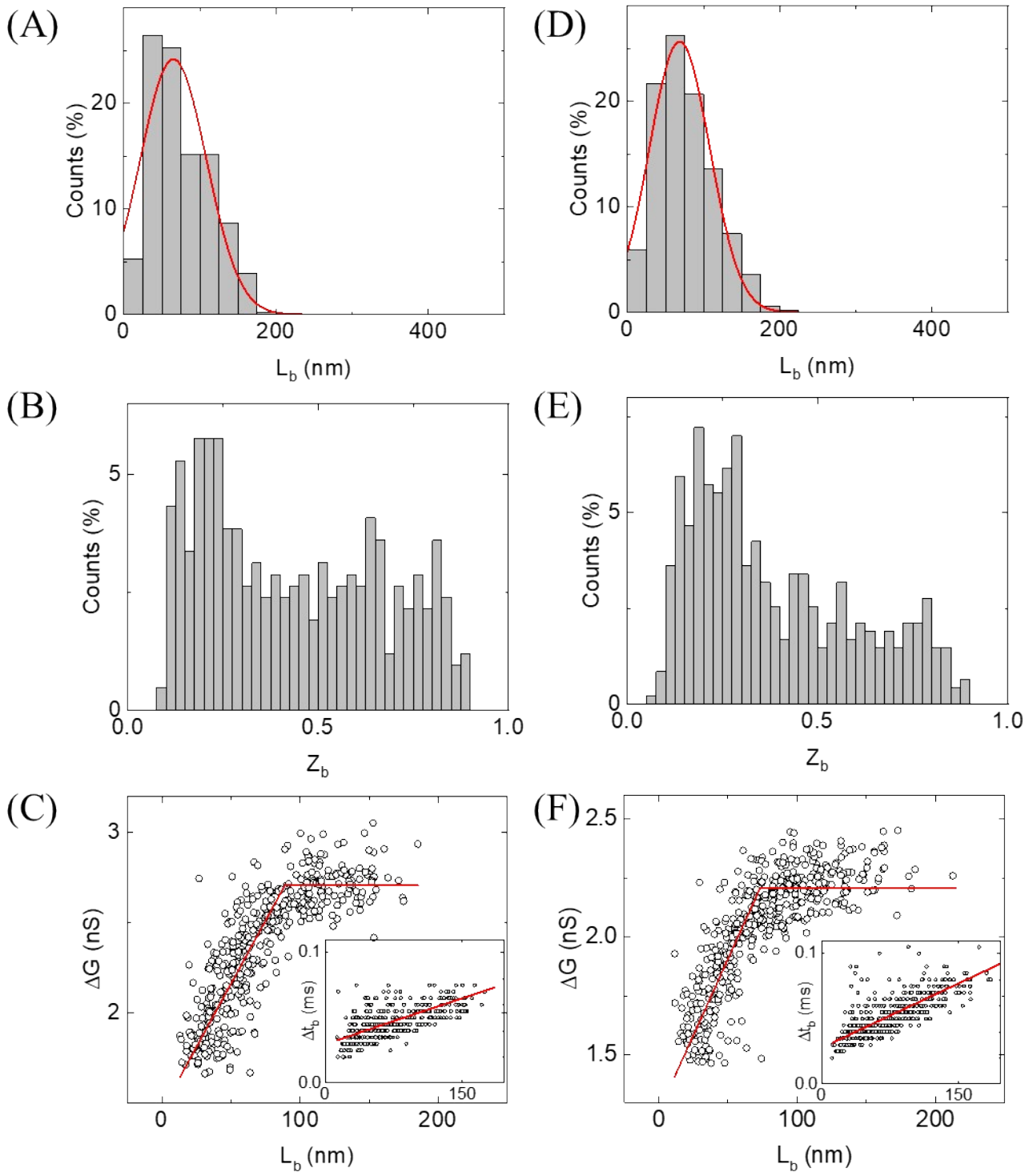

Figure S7: Analysis of branches in sc-DNA as measured in nanopores P29 (n = 1249), P31 (n = 960)

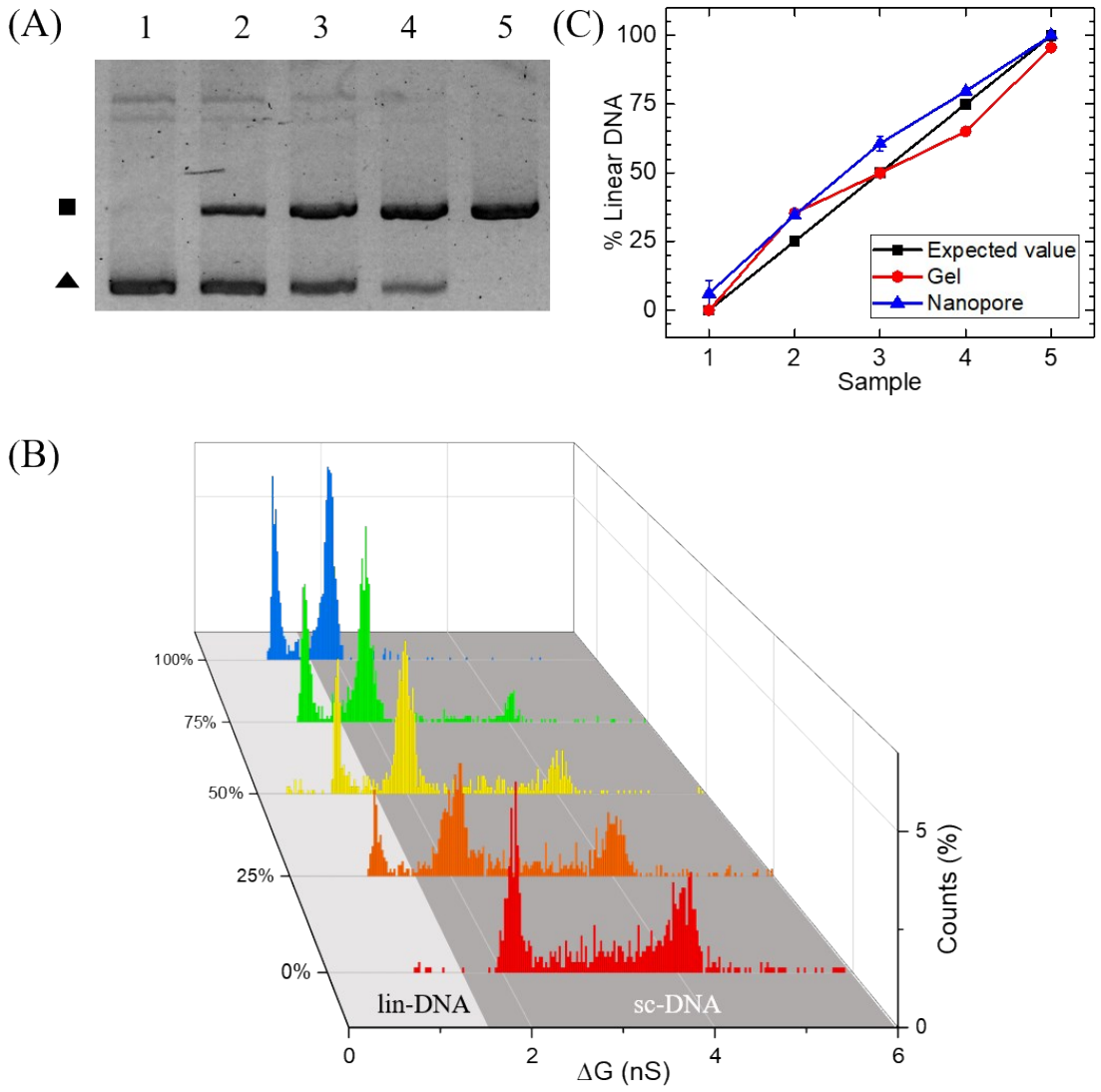

Figure S8: Demonstration of nanopore based quantification of linear plasmid using a known mixture of lin- and sc-DNA plasmid forms. (A) Gel image with lanes 1-5 showing bands corresponding to supercoiled (triangle) and linear (square) plasmid in the mixed samples with 0% (completely supercoiled), 25%, 50%, 75%, & 100% (completely linearized) of linear plasmid respectively. (B) Nanopore data from P37 (19 nm) showing ΔG histograms for the same mixtures as (A). Here n = 925 (0%), 919 (25%), 872 (50%), 895 (75%), 668 (100%). (C) Comparison of the amount of lin-DNA calculated from gel band (red circles) analysis and nanopore (N=2, from P36 and P37) data (blue triangles) analysis. Black squares show the expected values.

Here, we demonstrate the quantification of lin-plasmid DNA in a sample with a binary mixture of lin- and sc-DNA samples. We prepare samples containing a known mixture of linear and supercoiled forms with 0%, 25%, 50%, 75%, 100% of linear DNA. All the samples are analysed using gel electrophoresis (Fig S8A) and nanopore (Fig S8B). 100% linear plasmid shows two peaks in ΔG histograms with a first peak corresponding to the unfolded linear DNA and the second peak corresponding to folded linear DNA which also coincides with the first ΔG peak for supercoiled plasmid. However, since a known percentage of events (Table S1) are in the unfolded ΔG peak of a 100% linear DNA sample, we can use this correction factor to estimate the total amount of linear DNA in any sample. We use the following expression to accurately quantify the percentage of linear DNA in a mixture of plasmid conformations.

$$\%\ Linear\ DNA\ in\ a\ mixture = \%\ Unfolded\ events\ in\ a\ mixture \times \frac{100}{\%\ Unfolded\ events\ for\ 100\%\ linear\ Di} \quad \text{(S2)}$$

We find that this nanopore peak analysis method produces quantitatively accurate results when compared to the gel electrophoresis (Fig S8C).

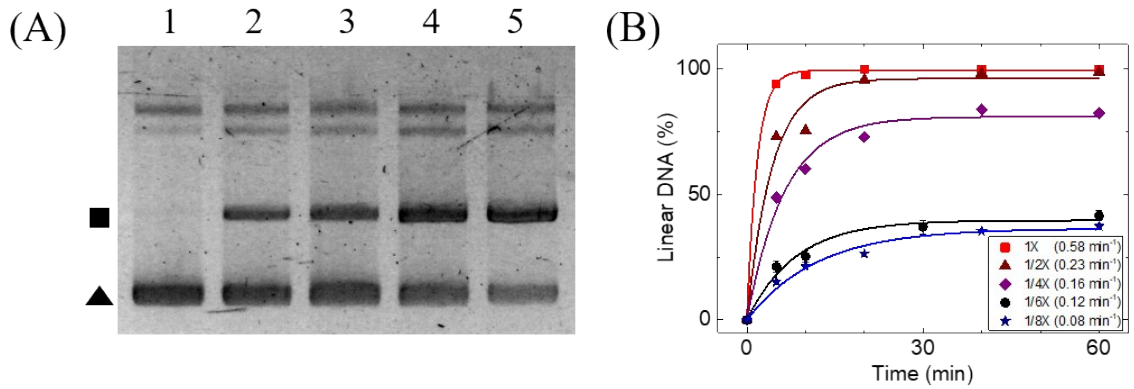

Figure S9: Gel quantification of rate constant for DNA digestion at different NdeI concentrations: (A) Gel image shows time-dependent linearization of sc-DNA plasmid by NdeI restriction enzyme at 1/6X concentration. Lanes 1-5 shows the result of restriction digestion assay when the reaction was stopped at 0, 5, 10, 30- and 60-min time points. The marker triangle shows sc-DNA and the square shows linearized DNA. (Fig 5A is reproduced here for clarity). (B) Plot of the amount of lin-DNA as a function of reaction time for NdeI concentrations of 1X, 1/2X, 1/4X, 1/6X, 1/8X. Solid lines are the fits to kinetics model (eqn (4)) and the first-order rate constants calculated for each reaction condition are mentioned in their respective parenthesis in the legend.

Pore ID	Sample	ΔG (nS)	Δt (ms)	ECD (ke^-)	% events in peak1	% events in peak2
P28	lin-DNA	0.7 ± 0.1 1.3 ± 0.1	0.25 ± 0.04 0.36 ± 0.05	425 ± 46	34.1	65.9
P29	lin-DNA	0.6 ± 0.1 1.2 ± 0.1	0.23 ± 0.03 0.32 ± 0.06	373 ± 37	37.7	62.3
P31	lin-DNA	0.5 ± 0.1 1.0 ± 0.1	0.26 ± 0.03 0.37 ± 0.05	336 ± 38	33.4	66.6
P31	cir-DNA	1.0 ± 0.1	0.22 ± 0.03	312 ± 26	90.3	9.7
P31	sc-DNA	1.1 ± 0.1 2.1 ± 0.1	0.21 ± 0.04	348 ± 35	34.9	65.1
P32	lin-DNA	0.31 ± 0.04 0.6 ± 0.1	0.27 ± 0.05 0.36 ± 0.04	178 ± 18	39.0	61.0
P35	lin-DNA	0.4 ± 0.1 0.8 ± 0.1	0.24 ± 0.04 0.33 ± 0.05	232 ± 21	41.9	58.1
P36	lin-DNA	0.65 ± 0.1 1.3 ± 0.1	0.23 ± 0.03 0.36 ± 0.08	437 ± 61	35.2	64.8
P37	lin-DNA	0.9 ± 0.1 1.8 ± 0.1	0.19 ± 0.01 0.27 ± 0.05	457 ± 50	34.2	65.8

Table S1: Summary of population analysis for various pores used. The data in this table correspond to Fig 2, Fig 5 and Fig S8.

Pore ID	D_P (nm)	G (nS)	ΔG_1 (nS)	ΔG_2 (nS)	Δt (ms)
P01	68	147.71	0.2		0.34
P02	76	207.04	0.25		0.56
P03	93	235.85	0.3		0.56
P04	77	242.13	0.3		0.30
P05	85	208.77	0.3		0.32
P06	74	255.75	0.3		0.41
P07	81	138.31	0.4	0.8	0.28
P08	50	174.83	0.3	0.4	0.32
P09	54	183.82	0.4	0.6	0.40
P10	18	81.43	1.9	3.8	0.15
P11	21	101.42	1.1	2.2	0.31
P12	21	84.75	1.6	3.2	0.20
P13	22	83.40	1.75	3.4	0.27
P14	17	87.34	2.0	4.0	0.20
P15	19	78.31	1.4	2.7	0.16
P16	21	87.64	1.3	2.5	0.17
P17	16	50.35	0.9	1.8	0.21
P18	20	97.94	1.0	2.0	0.18
P19	19	84.53	2.05	4.0	0.17
P20	22	100.60	1.4	2.6	0.14
P21	18	81.04	1.5	3.0	0.20
P22	30	110.25	0.6	1.1	0.26
P23	23	100.50	0.8	1.5	0.22
P24	13	58.69	2.8	5.6	0.20
P25	17	66.14	1.85	3.7	0.21
P26	18	77.46	1.9	3.8	0.15
P27	17	60.75	0.9	1.7	0.16
P28	16	64.85	1.4	2.8	0.17
P29	18	77.82	1.3	2.6	0.19
P30	39	159.24	0.5	0.9	0.22
P31	19	68.45	1.1	2.1	0.21
P32	17	61.80	0.6	1.1	0.19
P33	14	54.59	2.1	4.3	0.13
P34	18	48.10	1.1	2.1	0.20
P35	20	54.95	0.9	1.7	0.18
P36	22	88.97	1.5	2.9	0.15
P37	19	62.07	2.0	3.9	0.13

Table S2: Summary of sc-DNA translocation data obtained from a total of 37 nanopores as shown in Fig 3A (inset).

				ΔG **vs** L_b**for sc-DNA**					
Pore ID	Pore diameter (nm)	ΔG_1 (nS)	ΔG_2 (nS)	L_b (nm)	Expected L_{eff} (nm) (3.94 x D_P)	L_{eff} (nm)	Slope1 (nS/nm)	Slope2 (nS/nm)	ΔG_{sat} (nS)
P12	21	1.6	3.2	66 ± 46	83	81	0.019	0	3.3
P19	19	2.05	4.0	66 ± 47	75	82	0.023	0	4.1
P24	13	2.8	5.6	70 ± 50	51	82	0.032	0	5.65
P25	17	1.85	3.7	68 ± 50	67	71	0.025	0	3.75
P26	18	1.9	3.8	66 ± 51	71	80	0.022	0	3.8
P29	18	1.3	2.6	65 ± 43	71	89	0.014	0	2.7
P31	19	1.1	2.1	68 ± 40	75	73	0.013	0	2.2

Table S3: Summary of ΔG vs L_b fitting from branch analysis for supercoiled DNA from a total of 7 nanopores.

		Δt_b **vs** L_b **for sc-DNA**			
Pore ID	Pore diameter (nm)	Slope $\times 10^4$ (ms/nm)	Intercept (ms)	Velocity (μm/s)	Velocity (bp/μs)
P12	21	2.28	0.031	4396	13
P19	19	2.22	0.034	4507	13
P24	13	3.50	0.033	2856	8
P25	17	2.58	0.033	3871	11
P26	18	2.10	0.032	4763	14
P29	18	2.35	0.030	4248	12
P31	19	3.31	0.028	3025	9

Table S4: Summary of Δt_b vs L_b fitting from branch analysis for supercoiled DNA from a total of 7 nanopores.

Reaction time	% events in peak1 after correction
0 min	3.8 ± 1.4
5 min	22.4 ± 2.9
30 min	41.5 ± 3.8
60 min	55.2 ± 7.0

Table S5: This is the average of results obtained from 3 datasets in 2 nanopores for quantification of enzyme activity (as shown in Fig 5C). Two datasets were obtained on P32 and one dataset on P35.