Supporting Information

Ionic Current-Based Mapping of Short Sequence Motifs in Single DNA Molecules using Solid-State Nanopores

Kaikai Chen,[†,‡] Matyas Juhasz,[§] Felix Gularek,[§] Elmar Weinhold,[§] Yu Tian,[‡] Ulrich F. Keyser[,†] and Nicholas A. W. Bell[*,†,∥]*

[†]Cavendish Laboratory, University of Cambridge, JJ Thomson Avenue, Cambridge, CB3 0HE, United Kingdom

[‡]State Key Laboratory of Tribology, Tsinghua University, Beijing 100084, China

[§]Institute of Organic Chemistry, RWTH Aachen University, Landoltweg 1, D-52056 Aachen, Germany

[∥]Present address: Department of Chemistry, University of Oxford, Oxford, OX1 3TA, United Kingdom

*E-mail: ufk20@cam.ac.uk, nbell@cantab.net

Fax: +44 (0)1223 337000

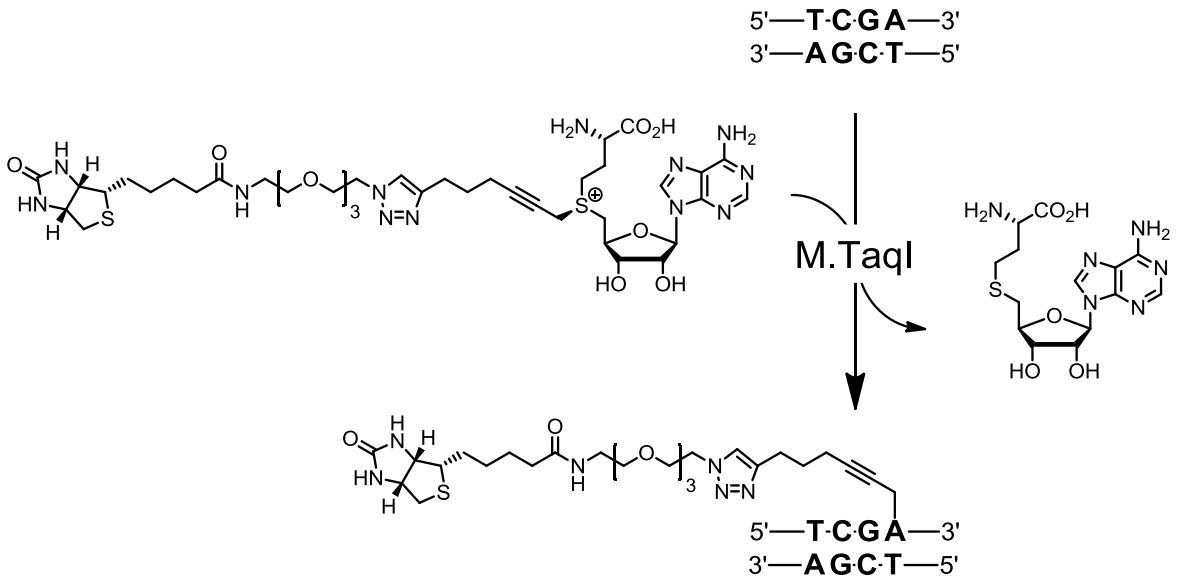
S1 Materials and methods

Chemicals.

LiCl powders were purchased from Sigma-Aldrich. The 4 M LiCl solution was buffered with 1×TE (10 mM Tris, 1 mM EDTA, pH 8.0, Sigma-Aldrich). Monovalent streptavidin was generously provided by the Howarth Lab.[1] Recombinant DNA MTase M.TaqI was produced as previously described.[2]

Sequence-specific biotinylation of plasmid DNA.

Biotin-labelled DNA was prepared by incubating double-stranded plasmids (500 ng/µL pBR322, 50 ng/µL ΦX174 or 50 ng/µL M13mp18, New England BioLabs (NEB), Ipswich, MA, USA), biotinylated AdoMet analogue (80 µM) and M.TaqI (pBR322: 1.24 µM, ΦX174: 0.13 µM, M13mp18: 0.12 µM, 1 eq. M.TaqI with respect to 5'-TCGA-3' recognition sequences on the plasmids) in NEB buffer 4 (pBR322: 50 µL, ΦX174 or M13mp18: 100 µL, 20 mM Tris-HCl, 50 mM KOAc, 10 mM $Mg(OAc)_2$, 1 mM DTT, pH 7.9) at 65 °C for 1 h (Scheme S1). Plasmids were purified using the QIAquick PCR purification kit (QIAGEN, Hilden, Germany) according to the instructions of the manufacturer. Complete labelling was verified by protection of the modified plasmids against cleavage by the cognate restriction endonuclease R.TaqI. DNA samples of each labelling reaction were supplemented with R.TaqI (10 units/µg DNA, New England BioLabs, Ipswich, MA, USA), incubated at 65 °C for 1 h and analyzed by agarose gel (1%) electrophoresis (0.5×TBE buffer, 1 h, 90 V, 0.01% GelRed) (Figure S1).



Scheme S1. M.TaqI-catalyzed transfer of an extended side chain with a biotin residue from an *S*-adenosyl-L-methionine cofactor analogue to adenine within the double-stranded 5'-TCGA-3' DNA sequence.

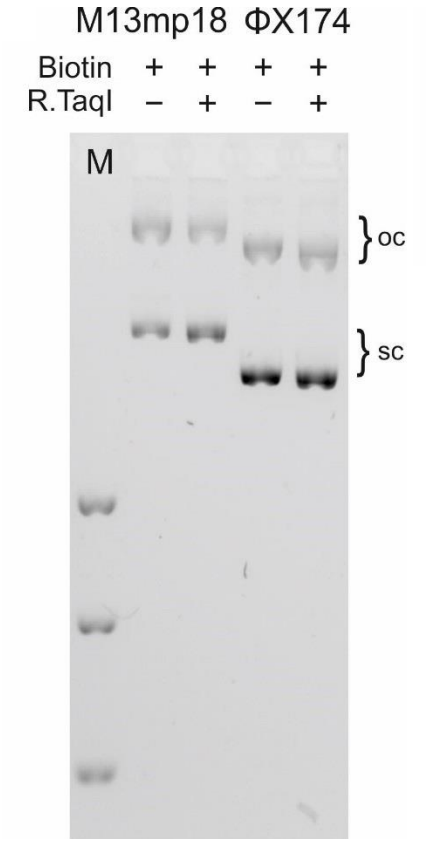


Figure S1: Analysis of plasmid modification by agarose gel electrophoresis for ΦX174 DNA and M13mp18 DNA. Biotinylated M13mp18 (Biotin +) and ΦX174 (Biotin +) are protected against cleavage by R.TaqI (R.TaqI +). Only bands for supercoiled DNA (sc) and open-circular DNA (oc, nicked DNA) are observed as with biotinylated M13mp18 (Biotin +) and ΦX174 (Biotin +) in the absence of R.TaqI (R.TaqI -).

Glass nanopore fabrication.

The glass nanopores were pulled from quartz capillaries (with outer diameter 0.5 mm and inner diameter 0.2 mm) using a pipette puller (P2000, Sutter Instrument Co., Novato, CA, USA). The inner pore diameters 14 ± 3 nm (mean ± s. d.) at the tip were estimated based on a previous characterization of our fabrication method.[3]

Setup and nanopore measurements.

Biotinylated plasmids were linearized with restriction endonucleases having unique recognition sites (pBR322: R.AhdI, ΦX174: R.BaeI, M13mp18: R.MscI) and diluted to match the measurement buffer (4 M LiCl, 1×TE, pH 8.0). The samples were added to the reservoir outside the nanopore at a concentration of 0.2-2 nM. For DNA detection a positive voltage (600 mV) was applied in all experiments. The ionic current was recorded with an amplifier (Axopatch 200B, Molecular Devices, Sunnyvale, CA, USA), filtered with an external 50 kHz Bessel filter (Frequency Devices) and digitized at a 250 kHz sampling rate with a data card (PCI 6251; National Instruments, Austin, TX, USA).

S2 Data analysis for protein-labelled pBR322 DNA translocations

1) Plots of all the translocation events caused by the protein-labelled pBR322 DNA.

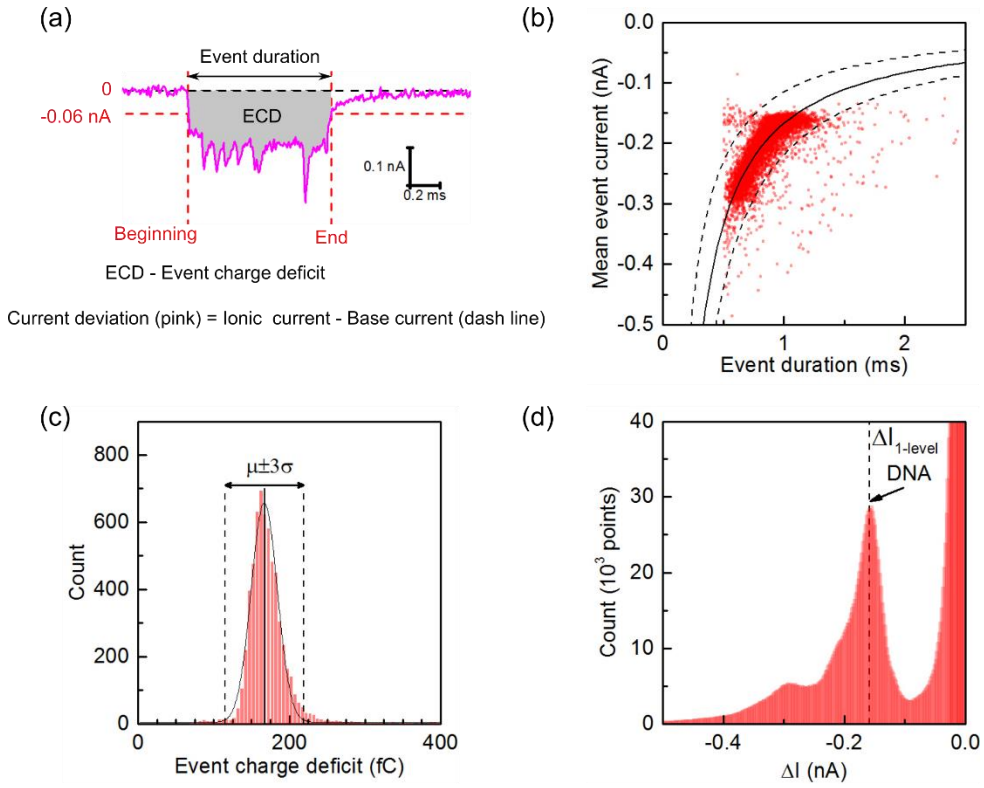


Figure S2. Translocation events caused by the protein-labelled pBR322 DNA through a ~14 nm-diameter glass nanopore. (a) An example event. The event beginning and end are set at the points where the current deviation crosses -0.06 nA. The area of the shadow represents the event charge deficit (ECD). (b) Scatter plot of 6169 events. The mean event current is inversely proportional to the event duration (translocation time). (c) Histogram of event charge deficit of the events shown in (b). The main peak at approximately 167 fC is due to the translocation of full length labelled pBR322 DNA. Only events within μ±3σ (5794 out of 6169) are retained for further analysis. (d) Histogram of the current deviation for all the event points (250 points/ms). We define the bump caused by the DNA as the 1_level current $\Delta I_{1\text{-level}}$ (-0.158 nA here).

2) Selection of unfolded events

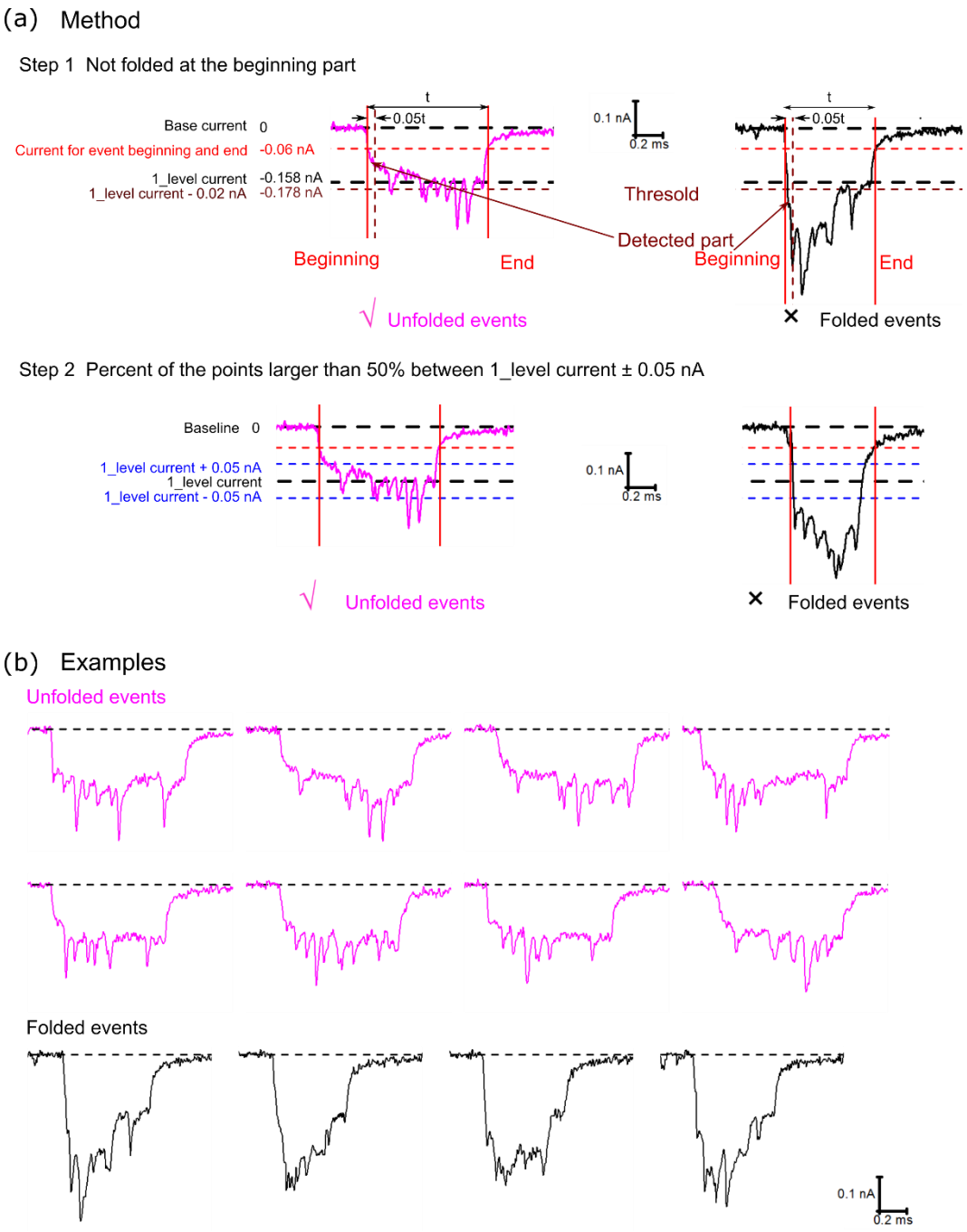


Figure S3. Selection of unfolded events. (a) Method for selection of unfolded events. In the first step, the minimum current during the first 5% of the translocation is compared to the threshold (-0.02 nA smaller than the 1_level current -0.158 nA) to distinguish the unfolded event (left, pink)

and the folded event (right, black). In the second step, we only use events where 50% or more current points lie between 1_level current ± 0.05 nA (blue dashed lines). (b) Examples of unfolded events and folded events. We had 1772 unfolded events out of the 5794 events from Figure S2.

3) Event peak analysis

An algorithm was written to determine the positions of peaks as shown in Figure S4.

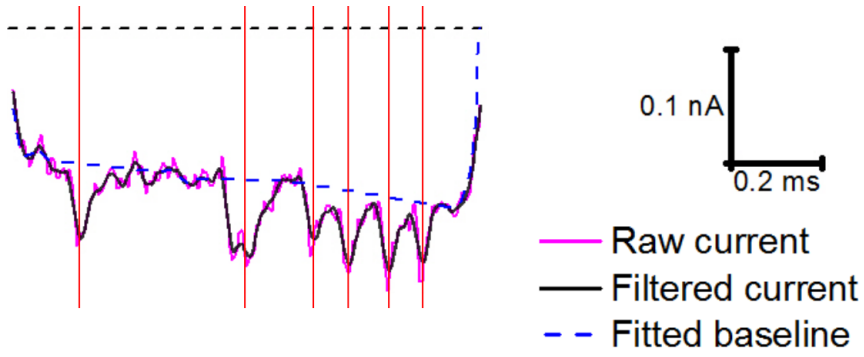


Figure S4. Method for peak detection with pBR322 DNA as an example. The raw current (pink) is first filtered by a Savitzky-Golay filter with six side points and polynomial order 3 (the black line shows the filtered current). The baseline (blue dashed line) is then determined by an automated algorithm which replaces sections of the trace which have high signal derivative with a straight line. Peaks are then detected if the current threshold is 0.2 times of the 1_level current less than the fitted baseline using Labview's in-built Peak Detector routine with the width for peak detection set at 12 points. The red lines indicate positions of detected peaks.

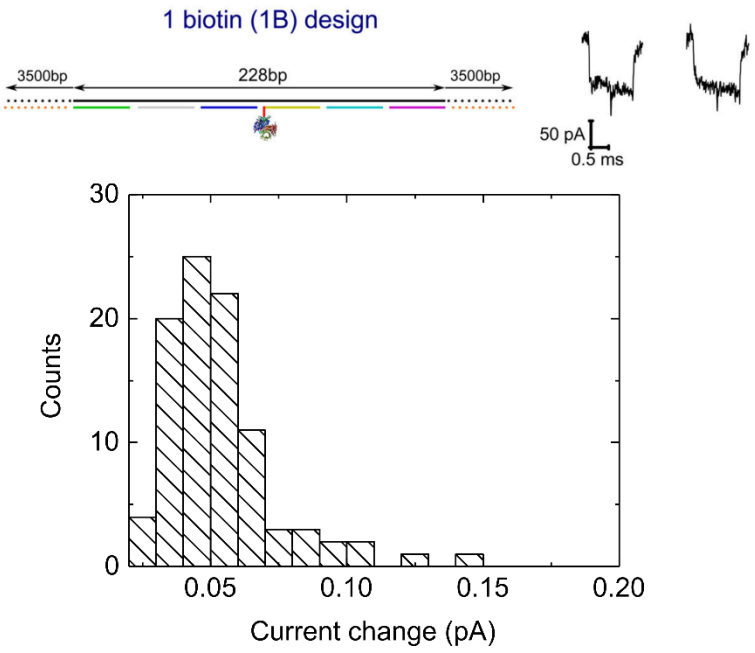


Figure S5. Analysis of the signal from a single streptavidin (data taken from Bell *et al*[4]). A 7.2 kbp DNA double-strand was engineered with a single biotin tag on one oligonucleotide at its centre and an excess of streptavidin added. Example translocations are shown together with a histogram of the current change from the streptavidin signal. Experiments were performed in 4M LiCl electrolyte. A unimodal distribution was observed.

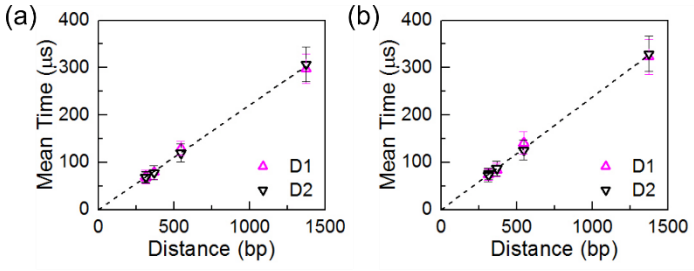


Figure S6. Analysis of drift during the course of the nanopore experiment presented in Figure 2 in the main text. The data from Figure 2f in the main text was split into two – the graph on the left (a) shows the kinetics of the first half of the events measured and the graph on the right (b)

shows the kinetics of the second half of the events measured. The close match indicates no considerable drift during the time course of the experiment.

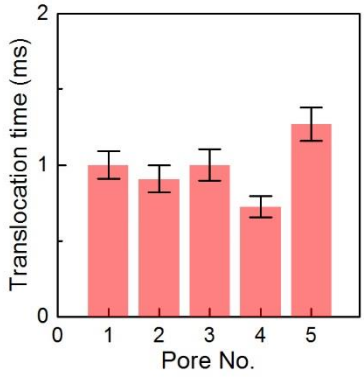


Figure S7. Translocation times of the protein-labelled pBR322 DNA measured with five separate nanopores. The times are the mean values of the Gaussian fits to the histograms of translocation time for unfolded events and the error bars are the standard deviations.

S3 Data analysis for the mixture of unlabelled pBR322 DNA and protein-labelled pBR322 DNA

1) Plots of all the translocation events caused by the mixture

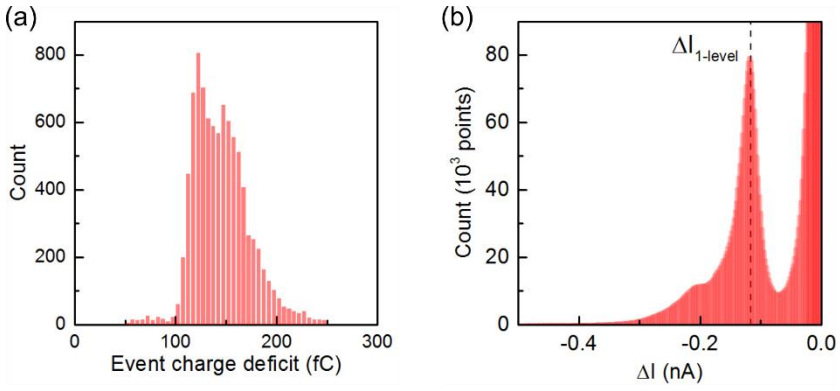


Figure S8. Histograms of the event charge deficit (a) and current deviation ΔI (b) for 9238 events caused by a mixture of unlabelled and protein-labelled pBR322 DNA.

2) Selection of unfolded events

The method for the selection of unfolded events is the same as shown in Figure S3. We had 4053 unfolded events out of the 9238 events.

3) Event peak analysis

The method for detecting the peaks is the same as shown in Figure S4.

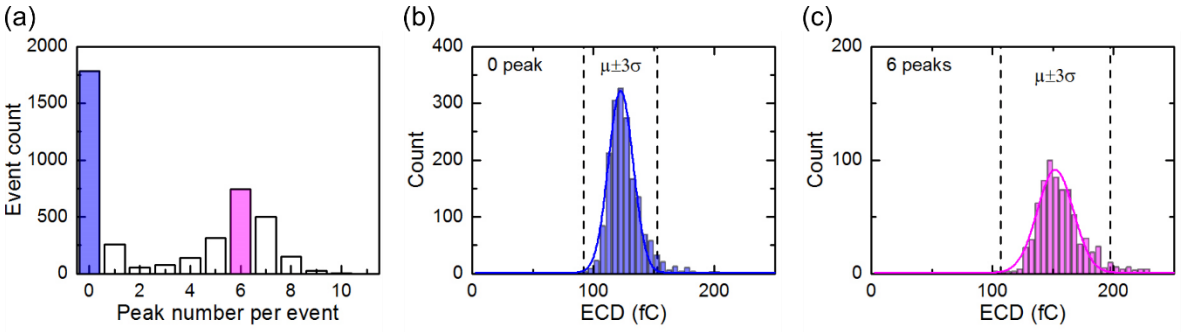


Figure S9. Selection of the events caused by a mixture of unlabelled and labelled pBR322 DNA. (a) Histogram of peak number detected per event. (b) and (c) show ECD histograms of the events detected with 0 and 6 peaks respectively. Only events within μ±3σ are retained for the translocation time comparison.

S4 Data analysis for the protein-labelled ΦX174 DNA translocations

1) Plots of all the translocation events caused by the protein-labelled ΦX174 DNA.

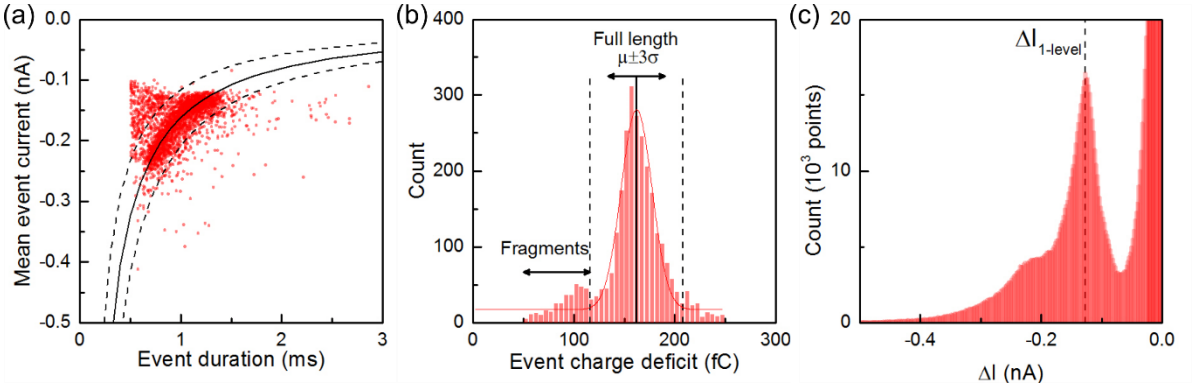


Figure S10. 2968 translocation events through a ~14 nm-diameter glass nanopore caused by the labelled ΦX174 DNA. (a) Scatter plot of the events. (b) Histogram of event charge deficit for the events. The main peak at approximately 162 fC is due to the translocation of ΦX174 DNA with full length attached with streptavidin on all binding sites. The tail in the histogram is attributed to DNA fragments. Only events within μ±3σ are retained for further analysis (2381 out of 2968 here). (c) Histogram of the current deviation for all the event points (250 points/ms).

2) Selection of unfolded events

The method for the selection of unfolded events is the same as shown in Figure S3. We had 807 unfolded events out of the 2381 events.

3) Event peak analysis

The method for analyzing the peak positions is the same as shown in Figure S4.

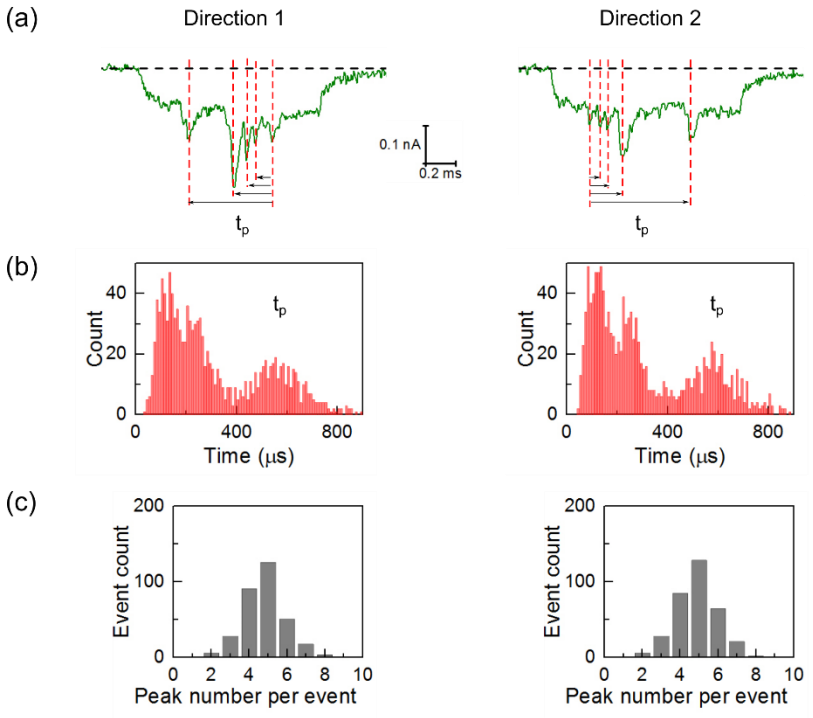


Figure S11. Peak analysis for events caused by the protein-labelled ΦX174 DNA. (a) Schematic of times of peaks measured with respect to the last (left column) and the first (right column) peaks for the two possible directions. (b) Histograms of the times of all peaks measured, (c) Number of peaks detected per event. The left column show the data in Direction 1 and the right Direction 2.

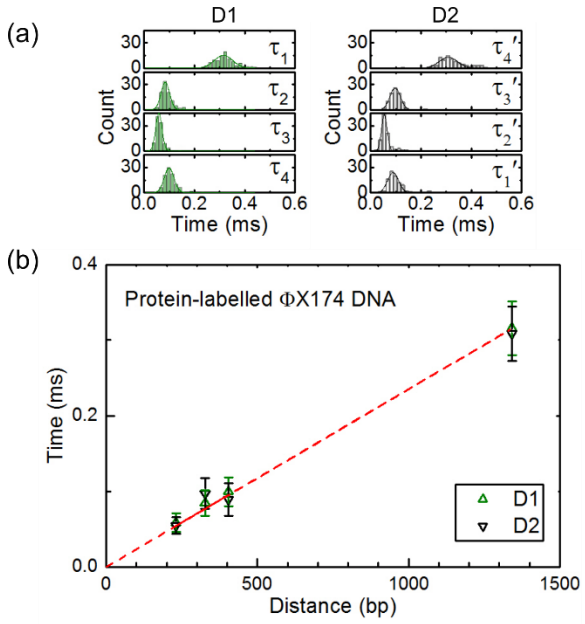


Figure S12. Translocation times between adjacent peaks for events with 5 peaks caused by labelled ΦX174 DNA. (a) Histograms of the translocation times between adjacent recognition sites in two orientations (D1 and D2) obtained from 100 and 90 events respectively. (b) A linear fit to the scatter plots of translocation time as a function of distance. The times and error bars are obtained from the mean values and standard deviations of the Gaussian fits to the histograms in (a). Sites 1 and 2 and Sites 3-7 are regarded as single sites respectively and the center positions are used as the site positions.

S5 Data analysis for the protein-labelled M13mp18 DNA translocations

1) Plots of all the translocation events caused by the protein-labelled M13mp18 DNA

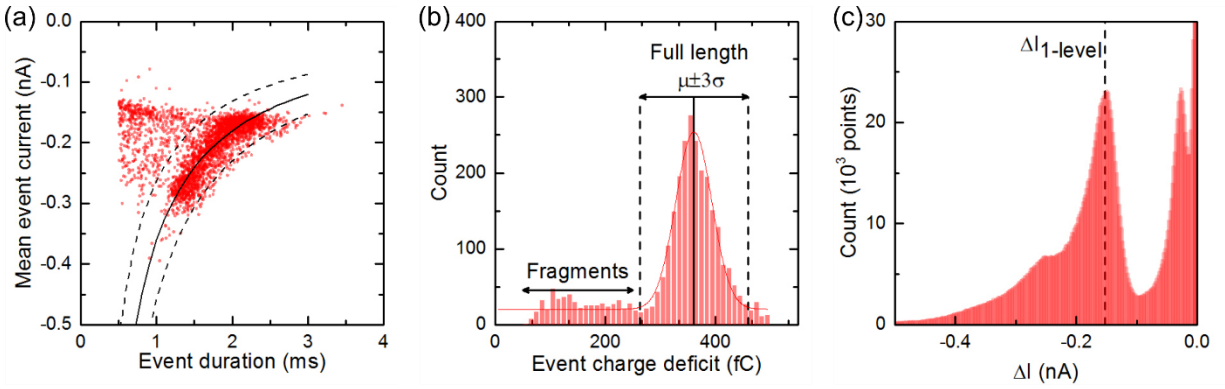


Figure S13. 2944 translocations through a ~14 nm-diameter glass nanopore caused by labelled M13mp18 DNA. (a) Scatter plot of the events. (b) Histogram of event charge deficit for the events shown in (a). The main peak at approximately 361 fC is due to the translocation of M13mp18 DNA with full length attached with streptavidin on all binding sites. The tail of the histograms is caused by DNA fragments. Only events within μ±3σ are retained for further analysis (2317 out of 2944 here). (c) Histogram of the deviation of the current for all the event points (250 points/ms).

2) Selection of unfolded events

The method for the selection of unfolded events is the same as shown in Figure S3 except that the first 2% part of the event was used as the detected portion in the first step since the first site occurs closer to the end of the DNA. We had 803 unfolded events out of the 2317 events.

3) Event peak analysis

The method for analyzing the peak positions is the same as shown in Figure S4 with the peak detection width 12 points.

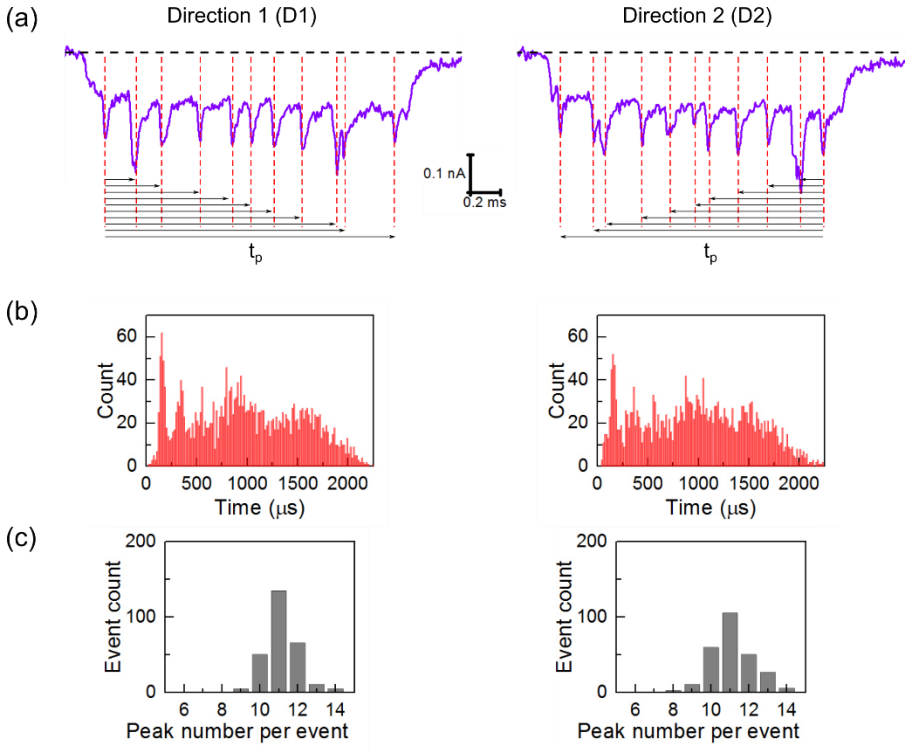


Figure S14. Peak analysis for events caused by the protein-labelled M13mp18 DNA. (a) Schematics of the time. (b) Histograms of peak timepoints measured with respect to the first peak (left) and last peak (right) recorded. (c) Number of peaks detected per event. The left column show the data in Direction 1 and the right Direction 2.

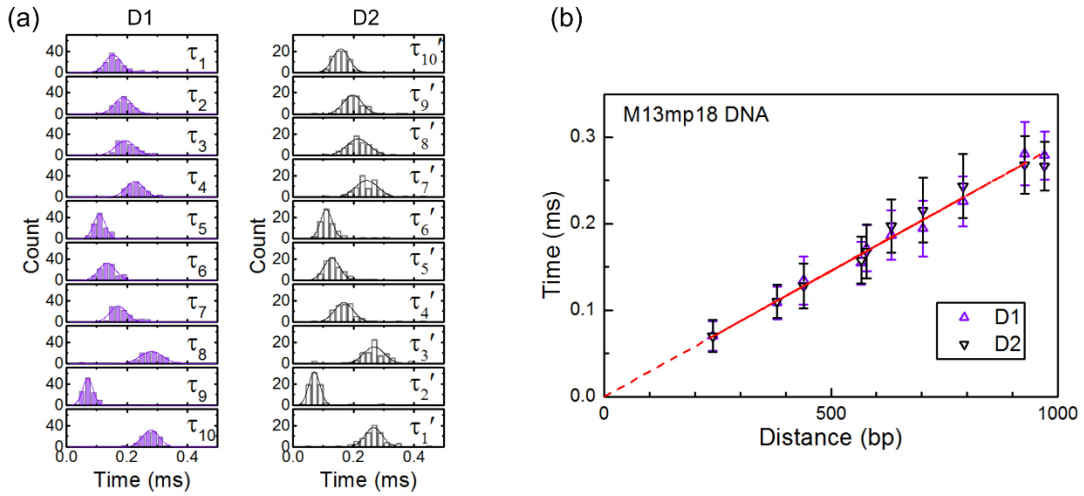


Figure S15. Translocation times between adjacent peaks for events with 11 peaks caused by labelled M13mp18 DNA. (a) Histograms of the translocation times between adjacent recognition sites in two directions obtained from 119 and 73 events respectively. (b) A linear fit to the scatter plots of translocation time as a function of actual distance. The times and error bars are obtained from the mean values and standard deviations of the Gaussian fits to the histograms in (a). Sites 2 and 3 are regarded as one and the centre is used as the site.

S6 Electric field in the glass nanopore

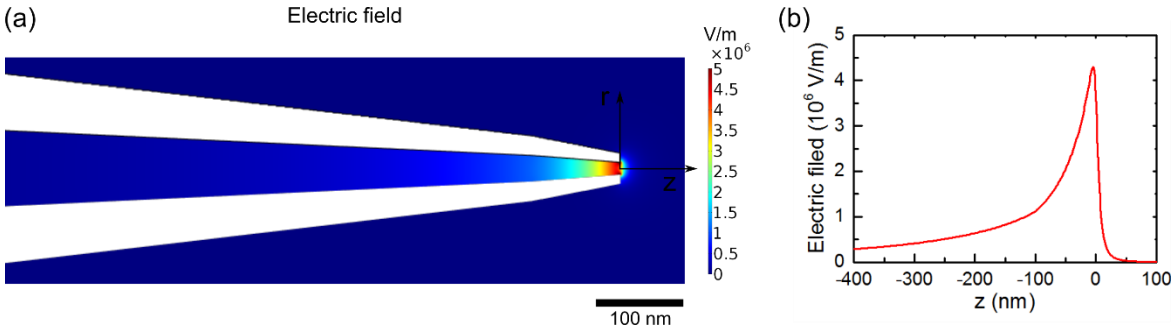


Figure S16. Electric field in the glass nanopore at a voltage 600 mV. (a) Distribution of the electric filed around the tip. (b) Electric field along the center axis (z shown in (a) with $z = 0$ at the very tip) of the pore. The results are obtained by numerical simulations with COMSOL Multiphysics 4.4 using an average geometry of the nanopore as characterized in ref 3.

S7 Data analysis for the mixture of unlabelled pBR322 DNA, protein-labelled pBR322 DNA and protein-labelled M13mp18 DNA

1) Histogram of the event charge deficit caused by the mixture for all the events

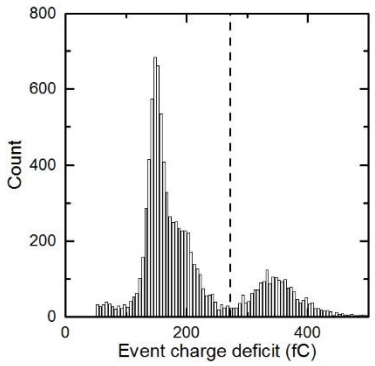


Figure S17. ECD Histogram of 9244 events caused by the mixture of unlabelled pBR322 DNA, protein-labelled pBR322 DNA and protein-labelled M13mp18 DNA. The right group is caused by the labelled M13mp18 DNA and the left group is caused by the unlabelled and labelled pBR322 DNA. Events with ECD within $\mu \pm 2\sigma$ are retained in the right group and $\mu - 2\sigma$ is set as the left threshold for the left group. The ECD values shown by the dashed lines are 120 fC, 272 fC and 414 fC resprectively.

2) Selection of unfolded events

The method for the selection of unfolded events is the same as shown in Figure S3. We had 747 unfolded events out of 1803 events in the right group (272 fC<ECD<414 fC) and 2835 events unfolded events out of 5883 events in the left group (120 fC<ECD<272 fC) in Figure S17.

3) Event peak analysis

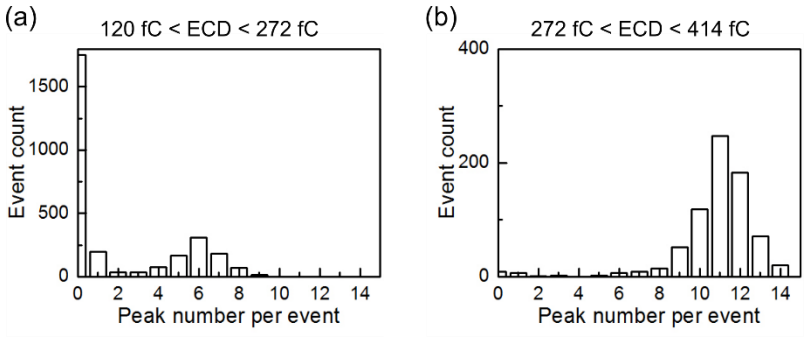


Figure S18. Histogram of peak number detected per event for the unfolded events with 120 fC < ECD < 272 fC (a) and events with 272 fC < ECD < 414 fC (b).

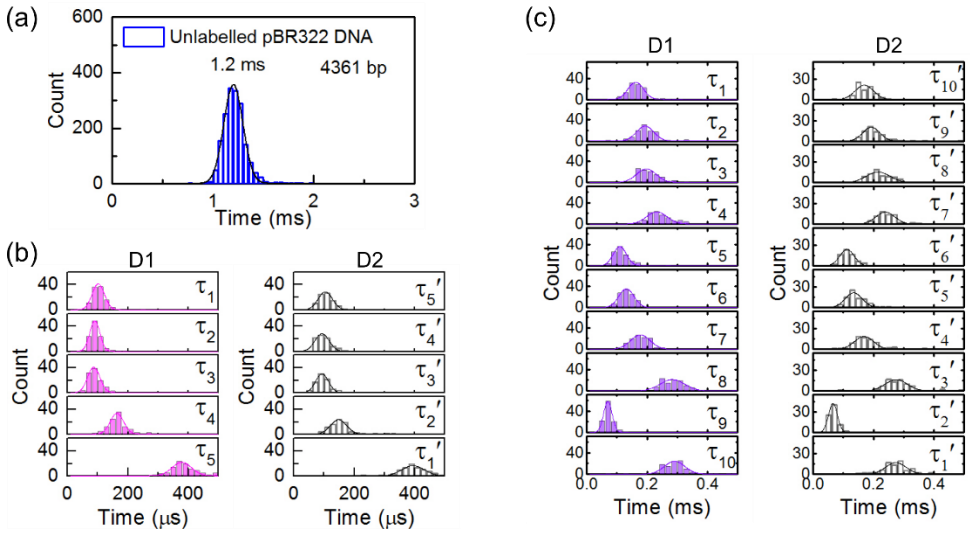


Figure S19. Histograms of the translocation times. (a) Histogram of the translocation time of the unlabelled pBR322 DNA. (b) Histograms of the translocation times between adjacent peaks in two directions for the events with 6 peaks caused by labelled pBR322 DNA. (c) Histograms of the translocation times between adjacent peaks in two directions for the events with 11 peaks caused by labelled M13mp18 DNA.

4) Estimation of the distances between adjacent sites

Direction 1						
Theoretical adjacent distances (bp)		368	315	313	546	1376
Estimated	Mean of the Gussian fit (bp)	370	328	315	594	1375
	S. d. of the Gussian fit (bp)	74	61	74	80	117
Mean error (%)		0.52	4.07	0.70	8.92	-0.02
Direction 2						
Theoretical adjacent distances (bp)		368	315	313	546	1376
Estimated	Mean of the Gussian fit (bp)	378	338	338	538	1434
	S. d. of the Gussian fit (bp)	86	86	78	99	137
Mean error (%)		2.65	7.26	7.89	-1.45	4.24

Table S1. Estimation of the distances between adjacent sites for the event detected with 6 peaks and comparisons with theoretical values for pBR322 DNA. The mean error is calculated as the normalized difference between the estimated mean value and theoretical value ((Estimated mean value – Theoretical value) / Theoretical value).

Direction 1											
Theoretical adjacent distances (bp)		567	633	703	791	381	440	579	927	239	971
Estimated	Mean of the Gussian fit (bp)	589	702	717	842	397	476	645	1030	253	1058
	S. d. of the Gussian fit (bp)	92	105	118	121	78	91	117	145	51	121
Mean error (%)		3.8	10.9	2.0	6.5	4.3	8.3	11.4	11.1	5.9	9.0
Direction 2											
Theoretical adjacent distances (bp)		567	633	703	791	381	440	579	927	239	971
Estimated	Mean of the Gussian fit (bp)	615	699	781	868	406	489	612	999	238	995
	S. d. of the Gussian fit (bp)	111	102	141	113	85	102	114	132	54	123
Mean error (%)		8.6	10.4	11.1	9.8	6.6	11.1	5.6	7.8	-0.4	2.5

Table S2. Estimation of the distances between adjacent sites for the event detected with 11 peaks and comparisons with theoretical values for M13mp18 DNA. The mean error is calculated as the normalized difference between the estimated mean value and theoretical value ((Estimated mean value – Theoretical value) / Theoretical value).

REFERENCES

(1) Howarth, M.; Chinnapen, D. J.; Gerrow, K.; Dorrestein, P. C.; Grandy, M. R.; Kelleher, N. L.; El-Husseini, A.; Ting, A. Y. *Nat. Methods* **2006,** *3*, 267-273.
(2) Goedecke, K.; Pignot, M.; Goody, R. S.; Scheidig, A. J.; Weinhold, E. *Nat. Struct. Mol. Biol.* **2001,** *8*, 121-125.
(3) Bell, N. A.; Keyser, U. F. *Nat. Nanotechnol.* **2016,** *11*, 645-651.
(4) Bell, N. A.; Keyser, U. F. *JACS* **2015,** *137*, 2035-2041.