Distinguishing single- and double-stranded nucleic acid molecules using solid-state nanopores

Gary M. Skinner*, Michiel van den Hout*, Onno Broekmans, Cees Dekker, and Nynke H. Dekker

Supplementary Information

Determination of current blockade

The histograms presented in Figures 2-4 were composed from current traces sampled at a frequency of 150 kHz or 200 kHz. Generally, the traces were digitally low-pass filtered at a frequency determined by the criteria discussed below (see **Filtering**), prior to event detection. For high voltage data (V_{bias} > 500 mV), low-pass filtering was frequently not necessary. Events were then selected according to the following algorithm:

1. The filtered current trace is parsed into segments of 5000 data points I_x, and for each segment the average <I> and standard deviation σ are determined (Figure S1). Within each segment, events are identified whenever an excursion of amplitude 4.5σ from the average <I> is detected.
2. The duration of each event is determined by finding the positions x_{start} and x_{end}, at which the current values are within 0.5σ from the average <I>. The duration of the event is defined as (x_{end} - x_{start})*Δt (where Δt is the minimum time step, Δt = 1/f_{sample}). Events shorter than a minimum duration t_{min} and longer than t_{max} are discarded (typical values: t_{min} = 40 µs, t_{max} = 10 ms). For each remaining event, the data corresponding to the indices [x_{start} - n, x_{end} + n] is selected (Figure S1, bottom panel) and added to the final output trace. A check is performed that overlapping points for neighboring events are removed.

3. The output traces for each segment are then merged into a final trace, from which a histogram of the current values is deduced (see Figures 2, 4, and S4). Gaussian functions are then fitted to each peak and the current blockade is determined from the difference in the positions of the peaks.

Filtering

The frequency at which the original current trace was low-pass filtered prior to event detection varied slightly for different datasets. Its value was selected depending on the average noise level of the current in the nanopore, as well as the signal-to-noise ratio of the individual events. For lower voltages, the filter frequency was chosen such that peaks could clearly be distinguished from the average noise level, all the while verifying that they were not significantly distorted by the filter.

To prevent possible filtering distortion, we analyzed the current blockade deduced for datasets low-pass filtered at different frequencies. An example of the resulting current blockade as a function of filter frequency is shown in Figure S2a for poly(A) at 600 mV. One can observe that the values level off at a filter frequencies of 30 kHz and higher. At lower frequencies, the computed current blockades are distorted and therefore cannot be trusted. In Figures S2b,c, a similar analysis is performed for poly(C) and DNA at 600 mV, respectively. For this poly(C) dataset, the minimum low-pass filter frequency that can be applied equals 55 kHz. This higher value is likely a consequence of the shorter mean lengths obtained in poly(C) synthesis (see Methods), which result in a concomitant reduction of the translocation times.

In a few cases the minimum filter frequency could not easily be determined via this method, because at higher filter frequencies no clear peaks could be determined from the histograms. In these cases, a different method for avoiding event distortion as a result of filtering was employed. Specifically, the criterion for the minimum duration t_{min} for each event was adapted, so that only events with

translocation times longer than $2*(1/f_{filter})$ were selected. Events satisfying this condition are also not significantly distorted by the influence of the filter.

Event rate

To ensure that all peaks in the current histograms correspond to contributions from single molecules only, all measurements were performed with low molecule concentrations (1-10 ng/µl). In Figure S3, we plot the typical event rates as a function of voltage for poly(A), poly(C), and dsDNA: the maximum rate is approximately 10 events/s. As the average translocation time is on the order of 100 µs, there is typically a 0.1% chance of finding a molecule in the pore. The likelihood of two molecules simultaneously traversing the pore is therefore negligible, which implies that the observed peaks in the histograms reflect the contributions from single molecules.

Translocation Time

The translocation times are plotted as a function of the applied bias voltage for all molecule types sampled (Figure S5). From this representative dataset, it can be seen that the translocation time roughly scales as 1/V, as expected. The relatively long translocation times of the poly(U) and poly(C) homopolymers were rather unexpected, and may reflect interactions with the nanopore. However, for other datasets we have observed significantly shorter translocation times for poly(C), indicating that for interactions with the nanopore surface may be more or less pronounced. Our datasets do not, however, demonstrate any correlation between the translocation time and the conductance blockade per events.

Modeling

In the main text, we have proposed a model based on entropic stretching to account for the observed decrease in ΔG as a function of voltage for the homopolymers. In this model, we assume that

the conductance blockade ΔG is inversely proportional to the amount of polymer material inside the nanopore, and thus inversely proportional to the relative extension x/L_0. We therefore express the conductance blockade as:

$$\Delta G = \frac{b}{\frac{x}{L_0} + c}, \text{ or conversely: } \frac{x}{L_0} = \frac{b}{\Delta G} - c$$

where *b* and *c* are fit parameters. Second, we assume that the stretching force is proportional to the applied bias voltage: $F = \sigma V$, where σ is a conversion factor in pN/mV. For any flexible polymer, the force–extension relationship is well described by the worm-like chain model[40]:

$$F(\frac{x}{L_0}) = \frac{k_B T}{L_P}\left\{\frac{1}{4(1-\frac{x}{L_0})^2} - \frac{1}{4} + \frac{x}{L_0}\right\}$$

Using the expressions above, this can be rewritten as:

$$V(\Delta G) = a\left\{\frac{1}{4(1-\frac{b}{\Delta G}+c)^2} - \frac{1}{4} + \frac{b}{\Delta G} - c\right\}, \text{ with } a = \frac{k_B T}{L_P \sigma} \text{ a free parameter}$$

This expression can be fitted to the homopolymeric data in Figure 3a. From the fit, we extract values for the fit parameters $\Delta G_{\min} = \frac{b}{c+1}$, the conductance blockade at full extension; $\Delta G_{\max} = \frac{b}{c}$, the conductance blockade at zero extension (or V = 0); and $\sigma = \frac{k_B T}{L_p a}$, the conversion factor in pN/mV. These fit parameters have the following values:

	ΔG_{min}	ΔG_{max}	σ
Poly(A)	0.65 ± 0.17 nS	1.86 ± 0.48 nS	0.0076 ± 0.0029 mV/pN
Poly(C)	0.39 ± 0.29 nS	5.53 ± 4.25 nS	0.0062 ± 0.0014 mV/pN
Poly(U)	0.29 ± 0.20 nS	4.22 ± 1.77 nS	0.0089 ± 0.0038 mV/pN

These values are discussed in the main text.

Supplementary Figures

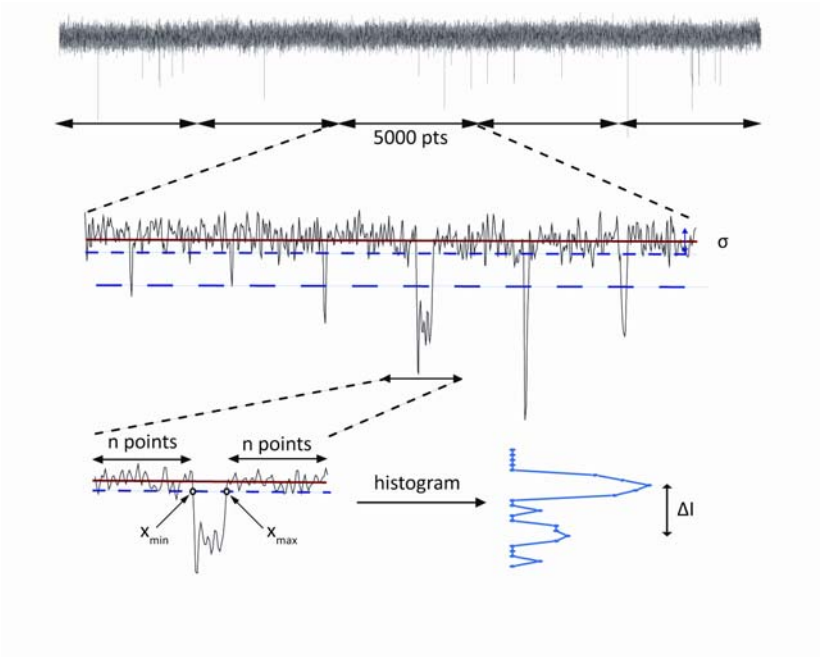

Figure S1 Explanation of the event detection algorithm; see text above for a full description.

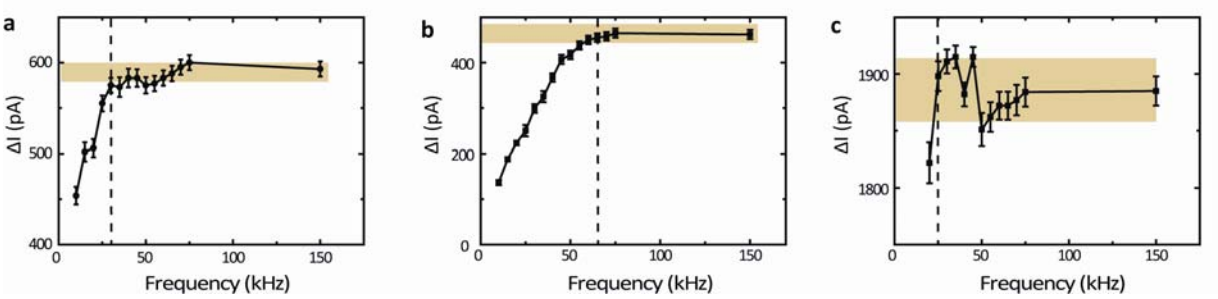

Figure S2 Influence of filtering the raw data before event detection: ΔI as determined from the histograms at different filter frequencies for poly(A) (**a**), poly(C) (**b**) and DNA (**c**) at 600 mV. In this manner, the minimum filter frequency at which the value ΔI is not distorted by the filter can be determined for each dataset.

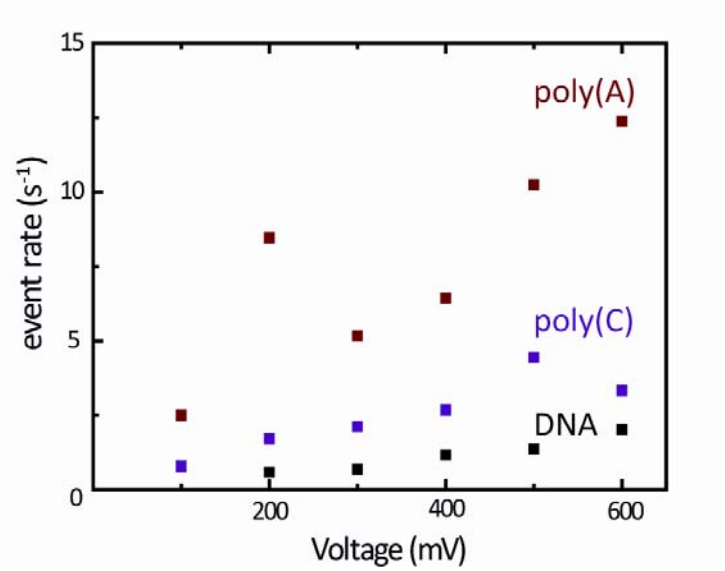

Figure S3 Graph showing the typical average event rate for representative molecules. In our experiments, the molecule concentrations were chosen such that the event rate was always low enough that it is very unlikely that two molecules simultaneously traverse the pore.

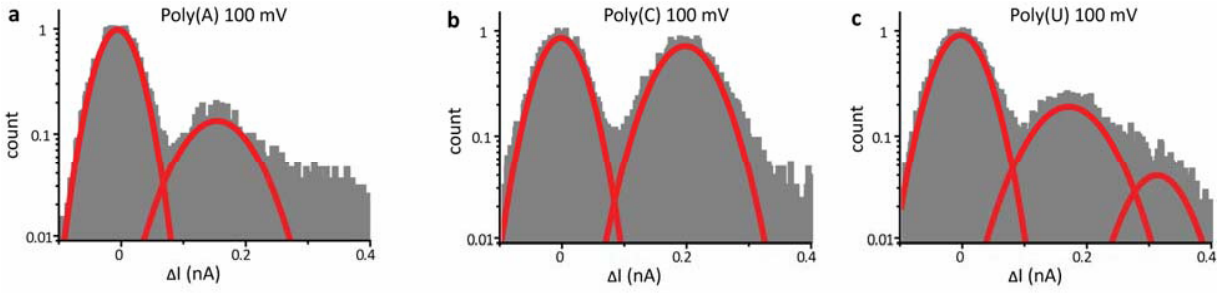

Figure S4 Typical histograms of the observed current blockade for the homopolymers poly(A) **(a)**, poly(C) **(b)** and poly(U) **(c)** at 100 mV bias voltage. The peak at ΔI=0 nA corresponds to the baseline current, while the other peaks in the current blockade are due the presence of molecules in the nanopore. In subfigures **(a)** and **(c),** a peak corresponding to two strands in the pore can be observed.

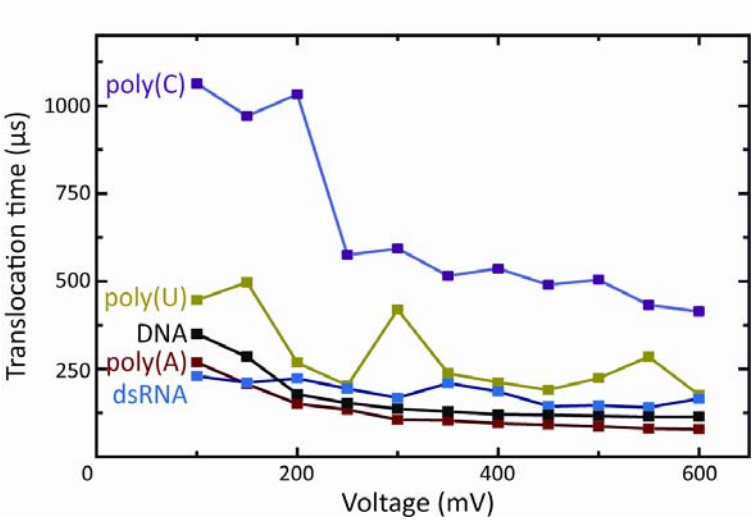

Figure S5 Examples of the average translocation time for each of the molecules versus the applied bias voltage. In general, the translocation time decreases at higher voltages. The long translocation time for the poly(C) data is in this particular dataset probably due to interactions with the nanopore surface.

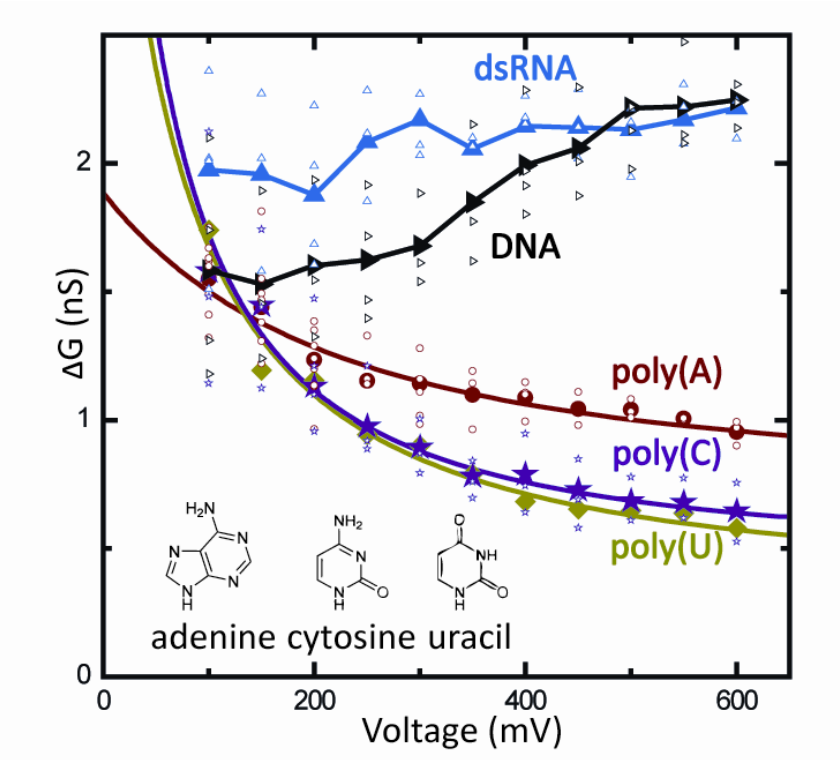

Figure S6 This plot contains the same information as Figure 3 in the main text, but explicitly shows all measured data points to indicate their full spread: open symbols represent individual measurements, whereas the closed symbols correspond to the average values. The double-stranded molecules showed more variation than the single-stranded homopolymers. The mixing experiment from Figure 4 is not included in this graph.