Supplementary Information for: Water Mediates Recognition of DNA Sequence *via* Ionic Current Blockade in a Biological Nanopore

Swati Bhattacharya,[†,§] Jejoong Yoo,[†,‡,§] and Aleksei Aksimentiev[*,†,¶]

†*Department of Physics, University of Illinois at Urbana–Champaign, 1110 West Green Street, Urbana, Illinois 61801*

‡*Center for the Physics of Living Cells*

¶*Beckman Institute for Advanced Science and Technology*

§*Contributed equally to this work*

E-mail: aksiment@illinois.edu

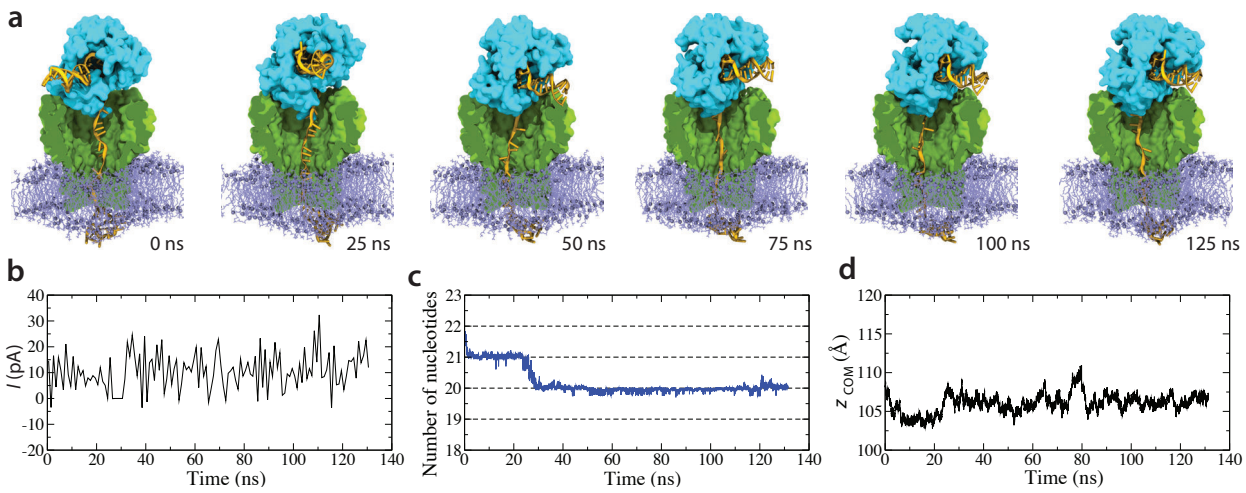
Table of Contents:

Supplementary Methods 1: Setup and simulation of full-length MspA–DNA–phi29 system

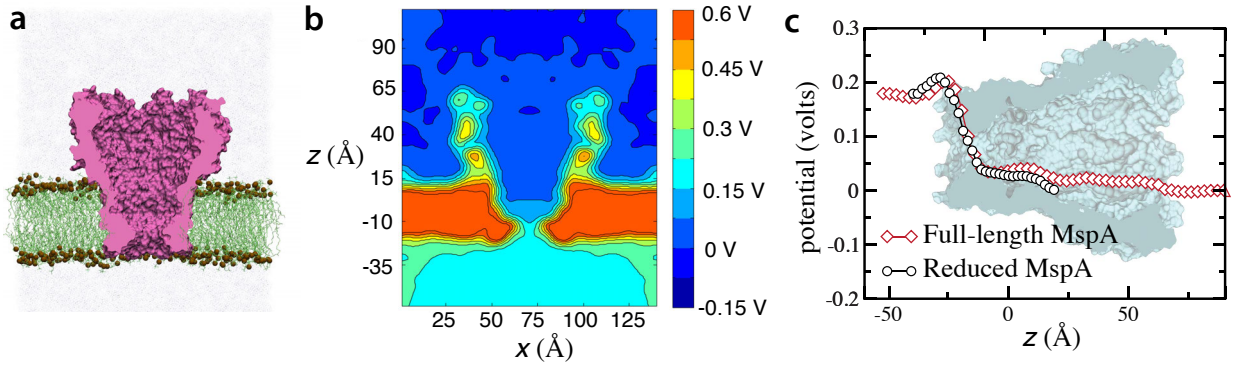
The atomic coordinates of a phi29 DNA polymerase were obtained from the Protein Data Bank (2PYL).[1] The fragments of the primer and template DNA resolved in the structure of the phi29 polymerase were retained. Using the phantom-pore method,[2] a 40-nucleotide fragment of a poly(dA) strand was placed within a full-length M1 MspA nanopore.[3,4] The backbone of the poly(dA) strand was approximately aligned with the axis of the nanopore; the 5′ end of the strand was located at the *trans* side of the membrane (see Figure 1a in the main text). The 3′ end of the poly(dA) fragment was covalently linked to the 5′ end of the template strand of the crystallographically resolved DNA-polymerase complex. The terminal di-deoxyadenosine monophosphate of the primer DNA was removed. The primer-template DNA duplex was extended by the addition of a 12-basepair dsDNA fragment of the following sequence: [5′-TTGCAATTGGGC-3′·5′-GCCCAATTGCAA-3′]. The resulting system was merged with a 12 × 11 nm^2 patch of a 2-oleoyl-1-pamlitoyl-sn-glyecro-3-phosphocholine (POPC) lipid bilayer. The lipid molecules overlapping with the protein were removed; the resulting patch contained 272 lipid molecules. The system was immersed in a volume of water molecules. K^+ and Cl^- ions were added in proportions to produce an electrically neutral system containing 1 M KCl solution. The final system measured 12 × 11 × 22 nm^3 and contained about 226,000 atoms.

All simulations of the polymerase-anchored DNA in the full-length MspA nanopore were performed using the NAMD2 program.[5] The simulations employed periodic boundary conditions, particle mesh Ewald[6] electrostatics and multiple time-stepping.[7] An integration timestep of 2 fs was used with SETTLE[8] and RATTLE[9] algorithms applied to covalent bonds involving hydrogen atoms in water and biomolecules, respectively. The local interaction forces were evaluated every timestep while the full electrostatic forces were evaluated every three timesteps. The van der Waals forces were evaluated using a cutoff of 8 Å and a switching distance of 7 Å.[10] The CHARMM27[11] force field was used for the protein, nucleic

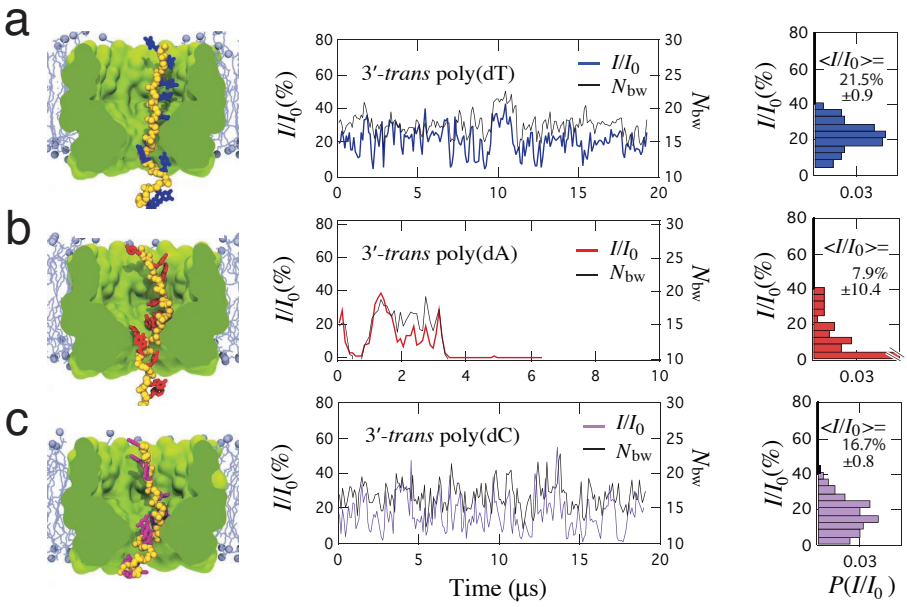
acids, ions and lipids with additional custom NB FIX corrections for the DNA-ion interactions.[12] The Lowe-Anderson thermostat[13] was used to maintain a temperature of 295 K. After initial minimization of 3000 steps, the system was equilibrated for 2 ns in the constant number of particles, pressure and temperature ensemble maintained using the Nosé-Hoover Langevin piston pressure control.[14] Subsequently, the system was simulated for 130 ns in the constant number of particles, volume and temperature ensemble and in the presence of an external electric field applied perpendicular to the lipid membrane, producing a transmembrane bias of 180 mV. In the latter simulation, no restraints were applied to DNA or the phi29 polymerase. In all simulations, the C_α atoms of the MspA nanopore were restrained to their crystallographic coordinates using a harmonic potential with a spring constant of 695 pN/nm.



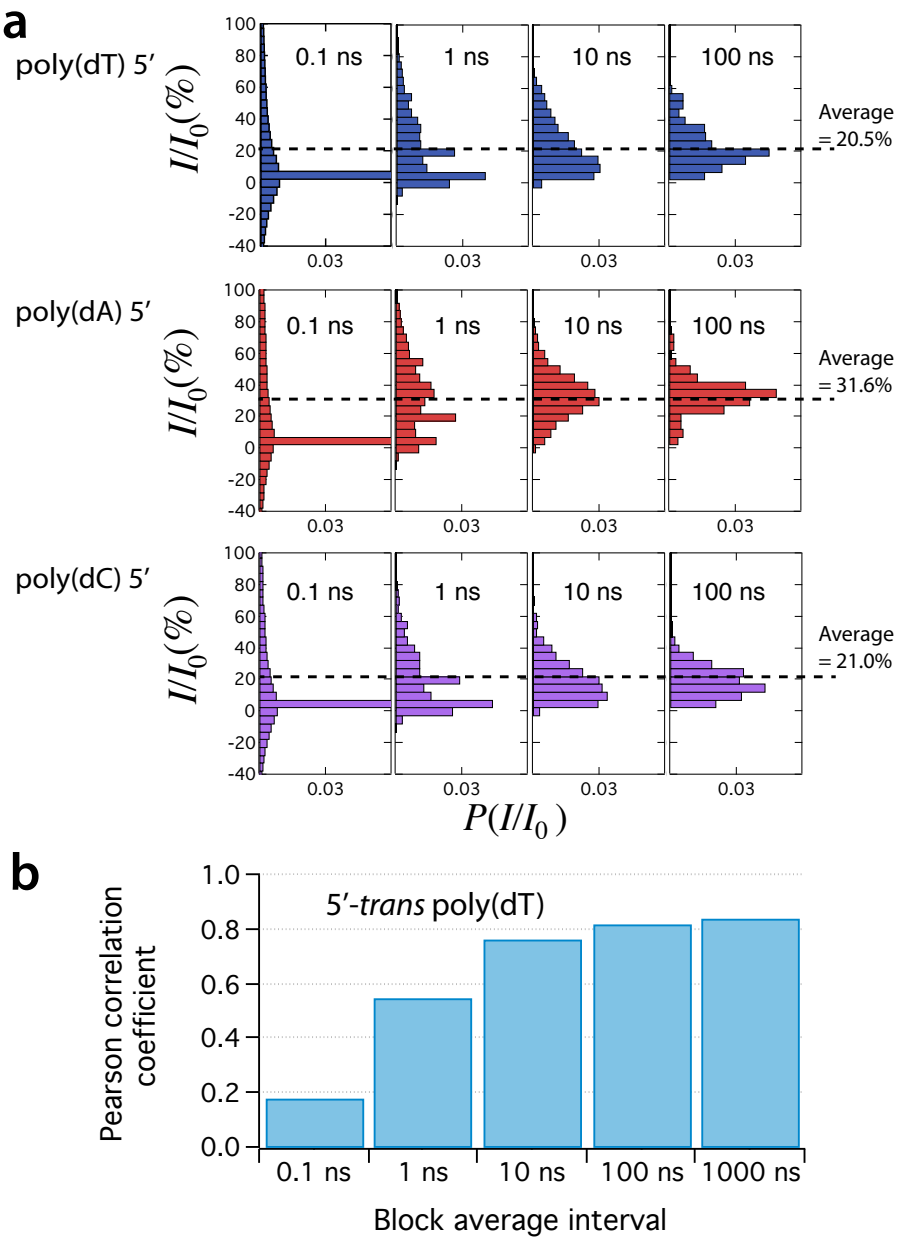
Supplementary Figure 1: MD simulation of full-length MspA–DNA–phi29 system. (a) Snapshots illustrating a 130 ns MD trajectory carried out under a 180 mV transmembrane bias. The MspA nanopore is shown as a green molecular surface cut-away to reveal a template DNA strand (yellow) threaded with its 5′ end through the MspA constriction and its 3′ end through the phi29 polymerase (blue molecular surface). A DNA strand complementary to the template strand forms a DNA duplex between the active site of the DNA polymerase and the 3′ end of the template strand. The MspA nanopore is embedded in a POPC bilayer (purple). Water and ions are not shown. During the course of the simulation, the polymerase and the duplex part of the DNA rotated about the axis of the pore by more than 90°. Such large scale motion of the polymerase did not have a measurable effect on the conformation of the template strand in the MspA constriction. Supplementary Movie 1 illustrates this MD trajectory. **(b)** The nanopore ionic current *versus* simulation time for the MD trajectory of the full-length MspA-DNA-phi29 system. Each data point shows a 1 ns average of the 10 ps sampled trajectory. **(c)** The number of nucleotides between the constriction of MspA (residues 90 and 91) and the junction of single- and double-stranded DNA at the active site of the polymerase. The relatively short (130 ns) duration of the MD trajectory was likely not sufficient to observe a full extension of the template strand under the 180 mV transmembrane bias. **(d)** The z-coordinate of the center of mass of the polymerase. In our coordinate system, the z-axis is parallel to the nanopore axis; the origin is located 11 Å above the center of mass of the MspA constriction (residues 90 and 91), in the MspA vestibule.



Supplementary Figure 2: Electrostatic potential map of MspA. (a) Atomic-scale model of a full-length MspA system embedded in a lipid bilayer membrane. The protein is shown as a cut-away molecular surface, the lipid bilayer is shown as green lines and brown spheres; ions are shown as purple dots, water is not shown. The model contains approximately 167,000 atoms and was built as described elsewhere.[4] **(b)** The electrostatic potential map of the full-length MspA system under a 180 mV transmembrane bias. The map was obtained by averaging the instantaneous distributions of the electrostatic potential over the microscopic conformations observed in a 25 ns MD trajectory and over the eight-fold symmetry of the channel.[4,25] The x and z axes points parallel and normal to the plane of the lipid bilayer. **(c)** The average electrostatic potentials along the symmetry axis of the MspA nanopores observed in MD simulations of the full- and reduced-length MspA models. The electrostatic profiles were computed from all-atom MD simulations of open (no DNA) MspA nanopores at a 180 mV bias. The background image of the nanopore aligns with the profiles of the electrostatic potential.



Supplementary Figure 3: Simulated ionic current blockade traces for the 3′- *trans* MspA systems. **(a)** (Left) A representative conformation of a thymine homopolymer. (Center) Blockade current (color) and the number of bulk-like water molecules (black) in the MspA constriction *versus* simulation time for the minimal systems containing a thymine homopolymer threaded through MspA in the 3′-*trans* orientation of the strand. Each data point shows a 100 ns block-average of the 100 ps-sampled MD trajectory, see Supplementary Note 1. (Right) Normalized histograms of the blockade current traces for thymine homopolymer. The histograms were built using 100 ns block-averages of the 100 ps-sampled current. **(b,c)** Same as in panel a but for adenine and cytosine homopolymers threaded through MspA in the 3′-*trans* orientation of the strand. In the simulation of the 3′-*trans* poly(dA) system, the complete blockade of the ionic current was produced by the 3′-end of the strand blocking the constriction (Supplementary Movie 7).



Supplementary Figure 4: The effect of block averaging interval. **(a)** Histograms of the blockade currents for the three 100-ps sampled MD trajectories block-averaged over the 0.1, 1, 10 and 100 ns intervals. Increasing the duration of block-averaging interval makes the current histograms more compact as thermal fluctuations are averaged out. **(b)** Pearson correlation between the blockade current I/I_0 and the number of bulk-like water molecules N_{bw} for the same MD trajectory of the 5′-*trans* poly(dT) system computed using the specified block-averaging of the raw (100-ps sampled) data. Thermal fluctuations obscure the correlation between I/I_0 and N_{bw} in the case of datasets block-averaged using short ($<$ 10 ns) averaging intervals.

Supplementary Note 1: The effect of block averaging interval on the simulated ionic current distributions.

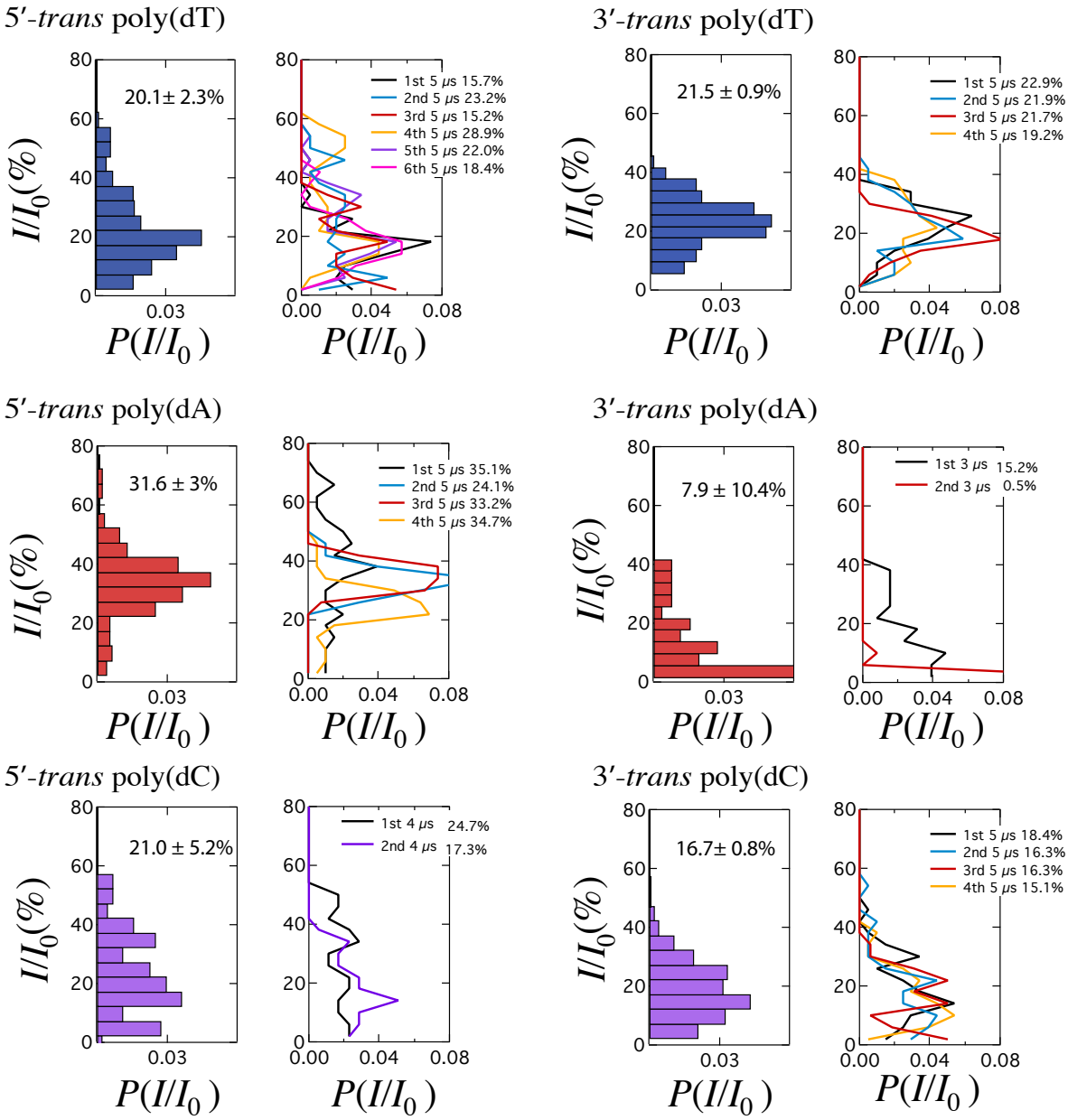
In an MD simulation, instantaneous atomic coordinates of the system are periodically recorded with a certain interval; the practical range being from 1 to 100 ps (the simulation time step is 2 fs). We compute the instantaneous ionic currents by summing up instantaneous displacements of individual ions; the latter are determined by comparing coordinates of the ions between consecutive frames of an MD trajectory.[16] Thus, the instantaneous ionic currents are recorded at the same sampling frequency as the atomic coordinates. For the simulations reported in this work, the sampling frequency was 100 ps.

As we show below, the instantaneous current values at the picosecond time scale are dominated by thermal displacements of ions. To obtain meaningful representation of the ionic current dependence on the simulation time, thermal displacements of ions must be averaged out. To accomplish that, an instantaneous ionic current trace (sampled at 100 ps) is split into fragments of specified length (1000 sampling points for the 100-ns block-average traces reported in Figure 2 and Supplementary Figure 3); each fragment is assigned an average value obtained as the arithmetic mean of the instantaneous values within each averaging interval, the block averaged value. Note that such block averaging procedure does not alter the average value of the entire ionic current trace.

To understand the dependence of the ionic current distributions on the duration of the block-averaging interval, Supplementary Figure 4, let's consider displacements of a single ion in bulk electrolyte solution subject to a uniform electric field along the direction of the field, the z axis. The motion of the ion is a superposition of stochastic displacement (prescribed by the ion's diffusion constant D) and unidirectional drift in the applied field (prescribed by the ion's electrophoretic mobility μ). For a time interval t, the magnitude of average stochastic displacement of the ion is $z_{\text{stoch}} = (2Dt)^{1/2}$; whereas the deterministic component is $z_{\text{det}} = E\mu t$. At high t, the deterministic displacement dominates over stochastic displacement. However, at small enough t, the stochastic displacement can have a larger

amplitude. The crossover time t_{cross} can be found by equating both displacements and solving the equation for t: $z_{\text{det}} = z_{\text{stoch}}$ or $(2Dt)^{1/2} = E\mu t$ leads to $t_{\text{cross}} = 2D/(E\mu)^2$. Using the Einstein relationship, $D = k_{\text{B}}T\mu/q$, where q is the ion's charge, we arrive at $t_{\text{cross}} = 2\,k_{\text{B}}T/(q\mu E^2)$. For $k_{\text{B}} = 1.35{\cdot}10^{-23}$ m^2 kg s^{-2} K^{-1}; $T = 293$ K; $q = 1.6{\cdot}10^{-16}$ Q; $\mu = 10^{-8}$ m^2/(V s); $E = 100$ mV / 4 nm $= 2.5{\cdot}10^7$ V/m, we obtain $t_{\text{cross}} \approx$8 ns. The instantaneous current $I = q\Delta z/\Delta t$ associated with the stochastic and deterministic components are $I_{\text{stoch}} = q(2D/\Delta t)^{1/2}$ and $I_{\text{det}} = qE\mu$. Thus, the amplitude of the instantaneous current produced by stochastic displacement of an ion increases as the sampling time interval decreases whereas the deterministic component remains constant. Because stochastic displacement are equally likely to occur along or against the applied electric field, the instantaneous currents due to stochastic displacement are equally likely to have positive or negative signs.

The dependence of the ionic current histograms on the block-averaging interval, Supplementary Figure 4, is in accord with the above theoretical considerations. At 100 ps sampling frequency, the histogram's tails extend in both negative and positive values. The tails become considerably shorter at 1 ns sampling and disappears at 10 ns sampling, in agreement with our estimate of the crossover time (8 ns). The histograms of 100-ps sampled current exhibit prominent peaks around a zero current value. The appearance of the peak reflects a large fraction of microscopic configurations that had no ions present within the 8-Å slice of the nanopore volume that we used to compute the instantaneous ionic currents. The presence of the peak also signifies that some ions remain relatively immobile along the z axis for a period of time exceeding 100 ps before passing through the pore constriction.



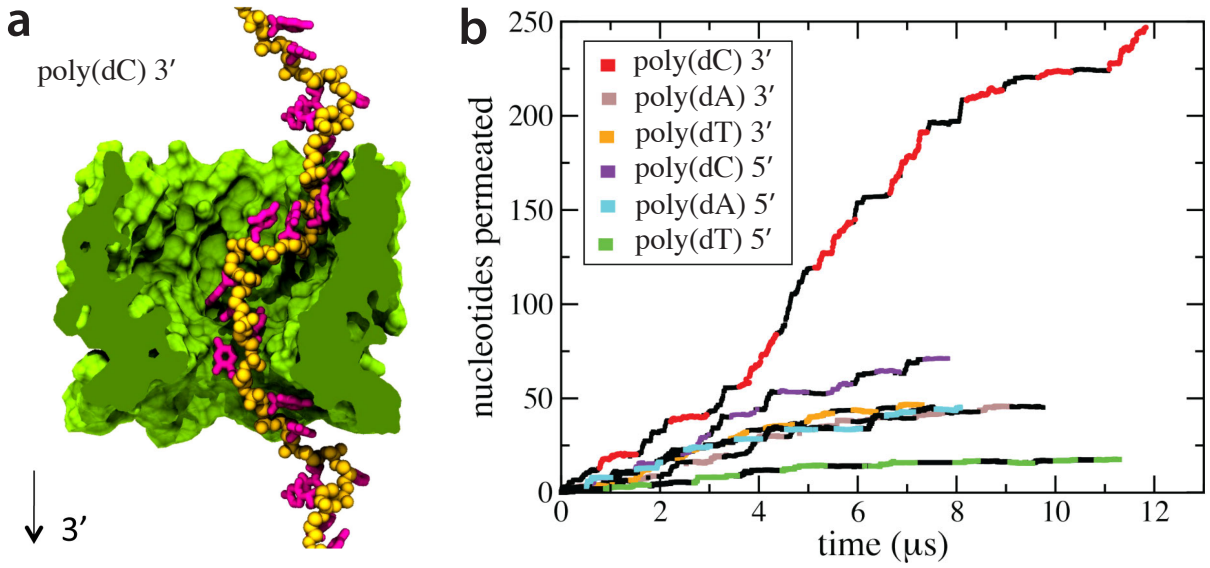
Supplementary Figure 5: Boot-strapping test for the convergence of the MD trajectories. For each simulation system, the left column shows the average value and the distribution of the blockade current computed from the entire MD trajectory, whereas the right column shows the average values and the distributions obtained from the fragments of that trajectory. The standard error was estimated using the average values of the trajectory fragments. Using the Mann-Kendall and von Neumann tests,[?] we determined that ionic current traces block-averaged over 500 ns or longer intervals do not exhibit correlations at a 5% level of statistical significance.

Supplementary Methods 2: Ensemble simulations of ionic current blockades

Ensemble simulations of DNA homopolymers systems were carried out following the protocol described earlier.[4] Briefly, sixteen different DNA configurations were obtained from a 110 ns trajectory of a full-length MspA system at a 1.2 V transmembrane bias.[4] The reduced-length (minimal) MspA system, which contained water, lipid, ions but no DNA, was combined with the DNA strand in each configuration, producing sixteen different systems. DNA segments that extended outside the water box were deleted, the nucleotide composition of the DNA strand was changed to comprise cytosine, adenine or thymine homopolymers. K^+ and Cl^-ions were added at random positions to bring the total charge of each system to zero. Following the usual minimization and equilibration protocols, the ends of each DNA strand were brought together (across the periodic boundary) in a 240 ps steered MD simulation.[15] Subsequently, the ends were covalently joined *via* a custom patch to create an effectively infinite DNA strand joined through periodic boundary conditions. Supplementary Figure 6a illustrates a typical simulation system. The length of the DNA fragment varied among the sixteen systems from 13 to 21 nucleotides.

Supplementary Table 1: Ensemble simulations of ionic current blockades. The ensemble simulations were performed using sixteen replica systems each containing the reduced-length MspA pore and a DNA strand covalently linked to itself across the periodic boundary of the system. Each replica had a unique initial conformation of the DNA strand. Supplementary Figure 6a illustrates one such replica system. The sixteen replicas of each homopolymers system were simulated in parallel at a transmembrane bias of 180 mV. The strands were free to permeate through the nanopore, exploring a range of microscopic conformations. Supplementary Figure 6b specifies the displacements of the DNA homopolymers in the ensemble simulations. To find the average blockade current from the ensemble simulation of a homopolymers system, the average ionic current within each individual trajectory was computed first. The ensemble-average value and the standard error were computed using the sixteen average current values, one from each replica. Individual simulation trajectories ranged in their duration from 300 to 800 ns. The total duration of the sixteen independent trajectories is listed in the last column of the table for each homopolymer system. The open pore ionic current at 180 mV was 910 pA.

System	Average current (pA)	Average current blockade (%)	Total length of 16 trajectories (μs)
5′-*trans* poly(dT)	188 ± 30	20.6 ± 3.2	8.1
5′-*trans* poly(dA)	327 ± 50	35.9 ± 5.5	11.3
5′-*trans* poly(dC)	179 ± 24	19.7 ± 2.6	7.9
3′-*trans* poly(dT)	172 ± 20	18.9 ± 2.2	7.3
3′-*trans* poly(dA)	173 ± 14	19.0 ± 1.5	9.8
3′-*trans* poly(dC)	126 ± 13	13.8 ± 1.4	11.2



Supplementary Figure 6: Ensemble simulations ssDNA transport through MspA. (a) A typical system used in ensemble simulations. The reduced-length MspA pore (same as the one used in the anchored simulations) is shown as a cut-away molecular surface, the backbone of a DNA strand is shown as yellow spheres, the DNA bases are shown as thick magenta lines; lipid, water and ions are not shown. The DNA strand is covalently joined to itself across the periodic boundary, forming an effectively infinite DNA strand. **(b)** Electric field-driven permeation of DNA homopolymers. The number of nucleotides translocated through the constriction of MspA is plotted as a function of time for each of the six ssDNA homopolymers. Each trace is shown in two colors, indicating sixteen individual trajectories (differing by the initial conformation of ssDNA) that were added consecutively to produce the total permeation trace. A transmembrane bias of 180 mV was applied in each simulation. Supplementary Table 1 lists the average ionic current recorded from the ensemble simulations.

Supplementary Note 2: Relationship between simulated and experimental ionic current

Three factors contribute to the mismatch between the absolute ionic currents obtained in our MD simulations of the reduced-length (minimal) MspA systems and experiment: truncation of the MspA channel, bulk conductivity and access resistance.

Truncation of the MspA vestibule increased the conductance of the MspA nanopore. Under identical simulation conditions (180 mV transmembrane bias), the simulated open pore current for the full-length MspA system is approximately 610 pA, which is 300 pA less than the simulated current for the truncated pore model, 910 pA.

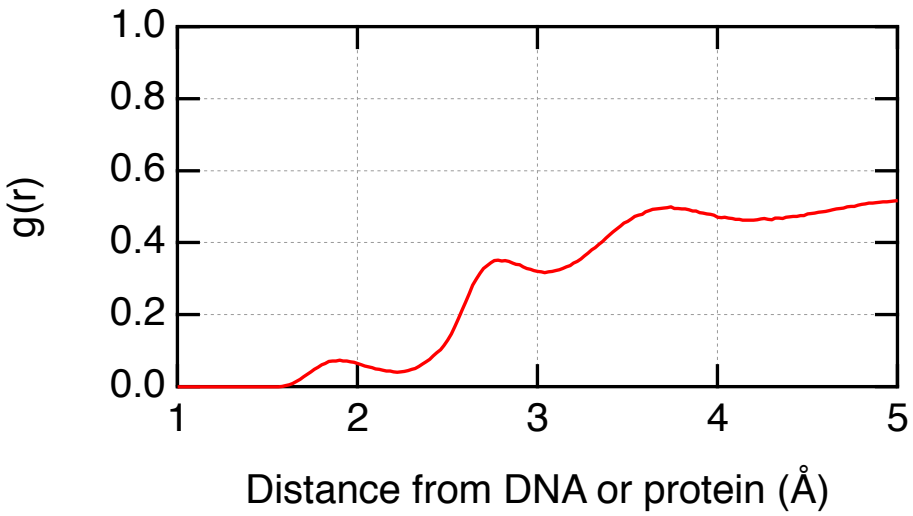
The experimental bulk conductivity of 1 M KCl is 10.5 S/m. The bulk conductivity of 1 M KCl simulated using the MD method is approximately 15 S/m.[10] The absolute conductance of a water-filled channel can be estimated by scaling the simulated conductance with the ratio of bulk conductivities.[16] Because the actual concentration of ions may vary within the channel, the scaling procedure gives an approximate value of the absolute conductance.

In experiment, the measured ionic current reports the overall resistance of the system $R = R_{\text{ch}} + R_{\text{access}}$, where R_{ch} is the channel's resistance and R_{access} is the access resistance.[17] The access resistance of a channel depends on its diameter d and on the bulk conductivity σ as $R_{\text{access}} = 1/(2\sigma d)$ for each side of the channel. In the case of a full-length MspA channel, the total access resistance can be estimated at $0.57 \cdot 10^8$ Ohm using 1 and 5 nanometers as the diameters of the *trans* and *cis* entrances, respectively, and the experimental value of the bulk conductivity ($\sigma = 10.5$ S/m). Experimentally, the ionic current through an open MspA pore at a 180 mV bias is 328 pA.[18] The total resistance of the system is thus $\frac{180 \text{ mV}}{328 \text{ pA}} = 5.49 \cdot 10^8$ Ohm, giving the channel's resistance $R_{\text{ch}} = R - R_{\text{access}} = 4.92 \cdot 10^8$ Ohm. In simulation, the ionic currents are determined under periodic boundary conditions. Hence, the simulations do not properly describe the access resistance of the experimental system.

Taking into account the missing access resistance and the difference between the simulated and experimental bulk conductivity values, the expected value of the simulated ionic current

of a full-length MspA system can be estimated as $\frac{15 \text{ S/m}}{10.5 \text{ S/m}} \cdot \frac{180 \text{ mV}}{4.92 \text{ Ohm}} = 523$ pA, which is within 16% of the value observed in our MD simulation (610 pA). Given the approximate nature of the above estimation, the simulated and measured ionic currents appear to be in very good agreement.

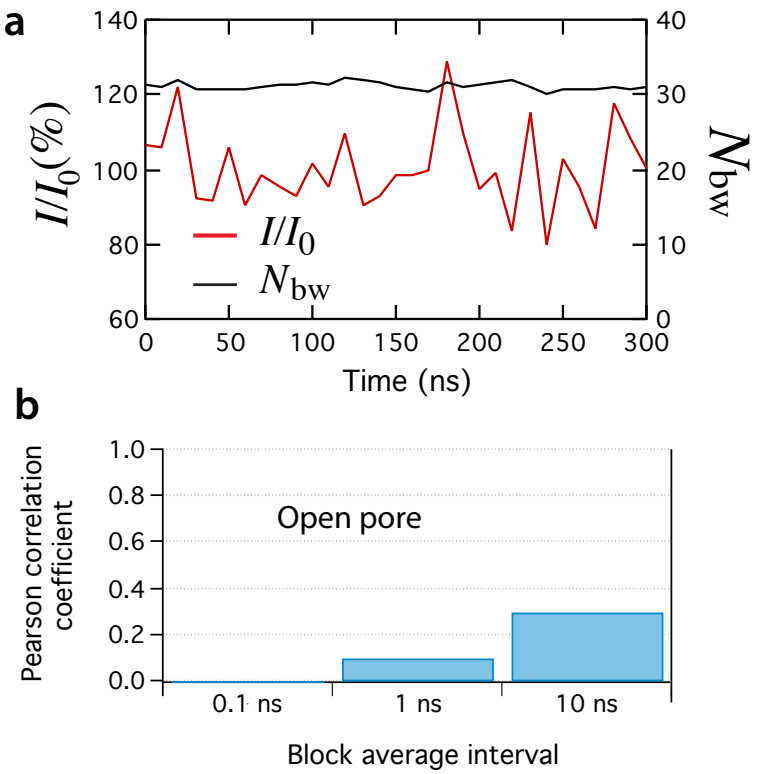
The three factors affecting the absolute value of the simulated ionic current are not expected to influence the ordering of the simulated blockade currents. The missing part of the MspA nanopore is relatively wide ($\sim$5 ns in diameter) and should not affect the DNA sequence specificity of the ionic current blockades (see also Supplementary Figure 1). The sequence specificity of the ionic current should neither depend on the bulk conductivity nor on the access resistance. Taking the above three factors into account would be essential for predicting the absolute value of the blockade ionic current; they are not required to reproduce the ordering of the relative current blockades values.



Supplementary Figure 7: Radial distribution function of water oxygens *versus* distance to DNA or protein atoms. The first peak of the distribution function corresponds to water molecules located within the first solvation shell of the MspA protein or DNA. In this study, we refer to water molecules from the first solvation shell of the protein or DNA (located within 2.5 Å of protein or DNA atoms) as structured water and all other water molecules as bulk-like water.

Supplementary Table 2: Correlation analysis of the ssDNA–MspA MD trajectories. The table lists the Pearson′s correlation coefficients between the nanopore current $I(t)$ and the number of all water molecules in the constriction region of MspA, $N_{\mathrm{w}}(t)$ (colum 2) or the number of bulk-like water molecules in the constriction region of MspA, $N_{\mathrm{bw}}(t)$ (column 3). For each system, the correlation is stronger when only bulk-like water is considered in the analysis. The analysis was performed on $I(t)$, $N_{\mathrm{w}}(t)$ and $N_{\mathrm{bw}}(t)$ timeseries composed from 100-ns block-averages of the 100-ps sampled MD trajectories. Column 4 specifies a power-law fit to the dependence of the nanopore currents I on the number of bulk-like water molecules N_{bw} for each system considered. The dependence was assumed to have the following form: $I = 0.0057 * (N_{\mathrm{bw}} - 9.1)^{\mathrm{X}}$, where the numerical values were derived from a power-law fit to all homopolymer data. The baseline value of 9.1 can be interpreted as a minimum number of bulk-like water molecules required to allow ion passage through the constriction. The duration of individual trajectories is listed in the last column of the table.

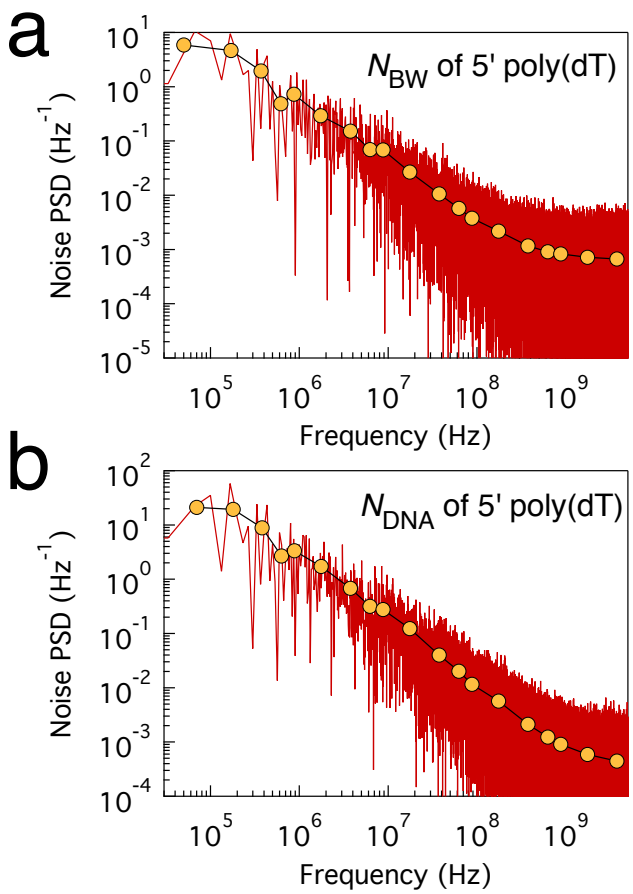
System	Correlation between I and N_{w}	Correlation between I and N_{bw}	X in $I = 0.0057*$ $(N_{\mathrm{bw}} - 9.1)^{\mathrm{X}}$	Total length of the trajectory (μs)
5′-*trans* poly(dT)	0.71	0.82	1.62	30.0
5′-*trans* poly(dA)	0.57	0.70	1.61	18.4
5′-*trans* poly(dC)	0.83	0.90	1.62	8.8
3′-*trans* poly(dT)	0.37	0.56	1.63	21.4
3′-*trans* poly(dA)	0.78	0.88	1.75	6.4
3′-*trans* poly(dC)	0.76	0.87	1.62	21.1
5′-*trans* TTTAAATTTTT	0.68	0.85	1.56	10.0
Average of 7 systems	0.67	0.80	1.63	16.6



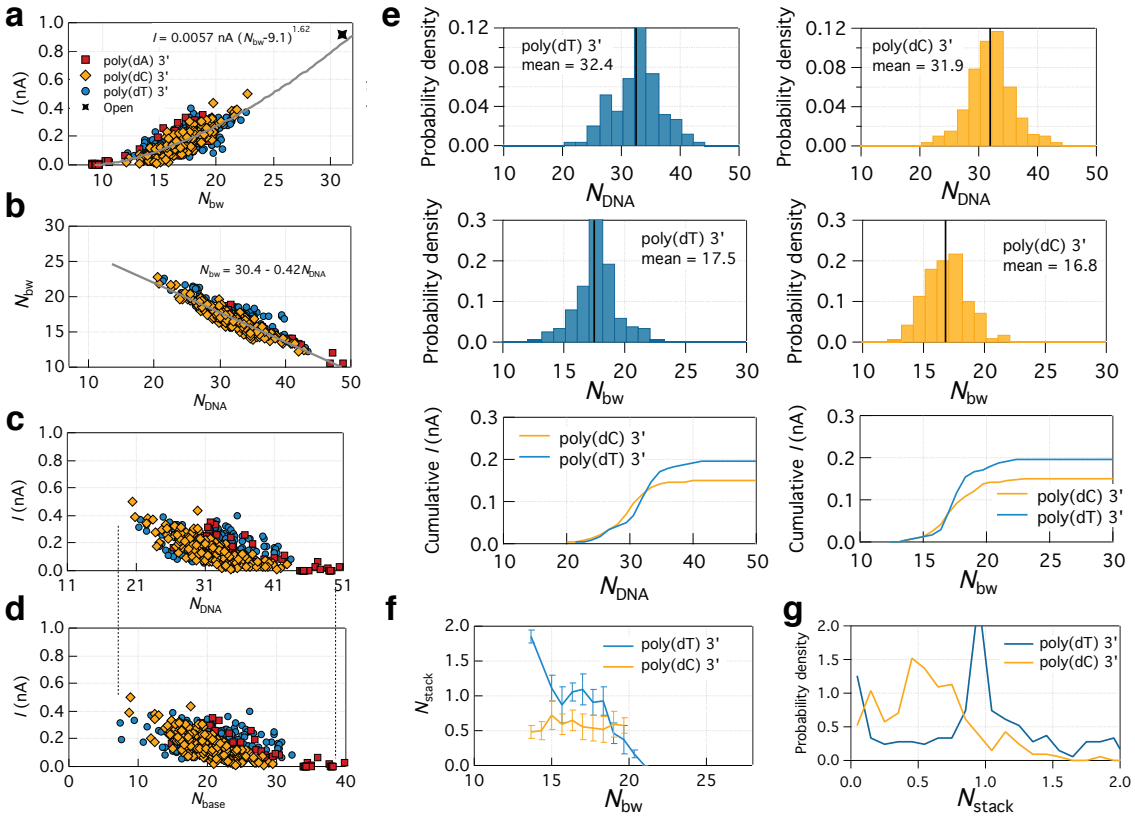
Supplementary Figure 8: Correlation analysis of the open pore MD trajectory. **(a)** The ratio of instantaneous to average open pore current I/I_0 (red) and the number of bulk-like water molecules in the MspA constriction N_{bw} (black) *versus* the simulation time for the minimal MspA systems containing no DNA (the open pore system). Each data point shows a 10-ns block average of the 100-ps sampled MD trajectory. **(b)** Pearson correlation between I/I_0 and N_{bw} within the same MD trajectory of the open pore system computed using the specified block averaging of the raw (100-ps sampled) data. The correlation between the number of bulk-like water molecules and the open pore ionic current is weak.

Supplementary Note 3: Cylindrical model of ionic current blockades

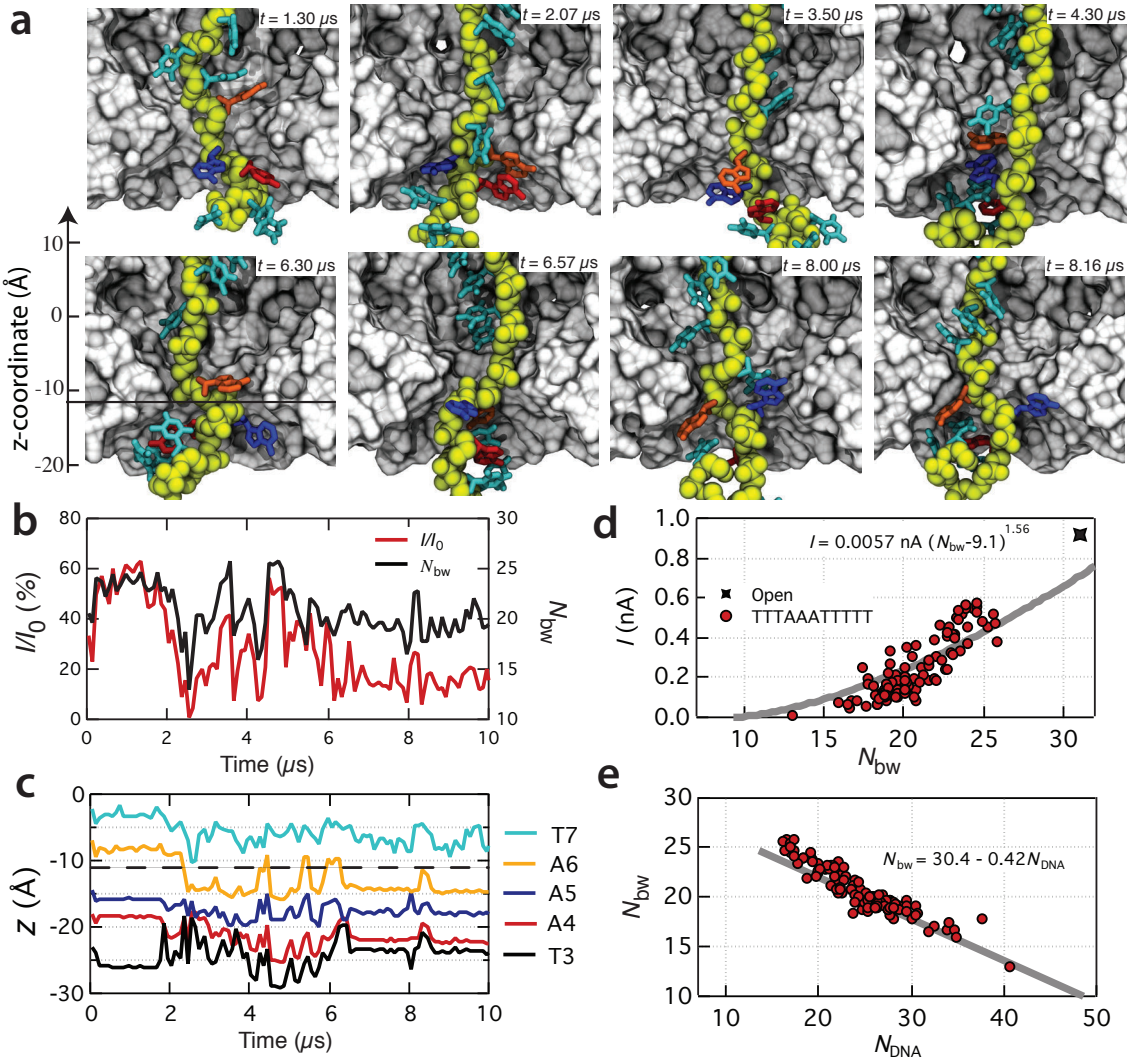
In the cylindrical model of ionic current blockades, the presence of DNA in the MspA constriction is described as a change of the effective cross-sectional area A of a cylinder of a fixed height H, filled with electrolyte of ionic conductivity σ. According to the Ohm's law, the current through a cylindrical resistor $I = \Delta U/R$, where ΔU is the potential difference at the edges of the cylinder and $R = H/(\sigma A)$ is the resistance of the cylinder. The volume of the cylinder V linearly depends on the number of electrolyte molecules inside the cylinder: $V = N\alpha^3$, where α^3 is the volume of one molecule of the electrolyte (water in the case of MspA constriction). Thus, the effective cross-sectional area of the cylinder $A = V/H = N\alpha^3/H$. Substituting this expression for A into Ohm's law gives: $I = \Delta U/R = \sigma \Delta U A/H = \sigma \Delta U N \alpha^3/H^2$, a linear dependence on N.



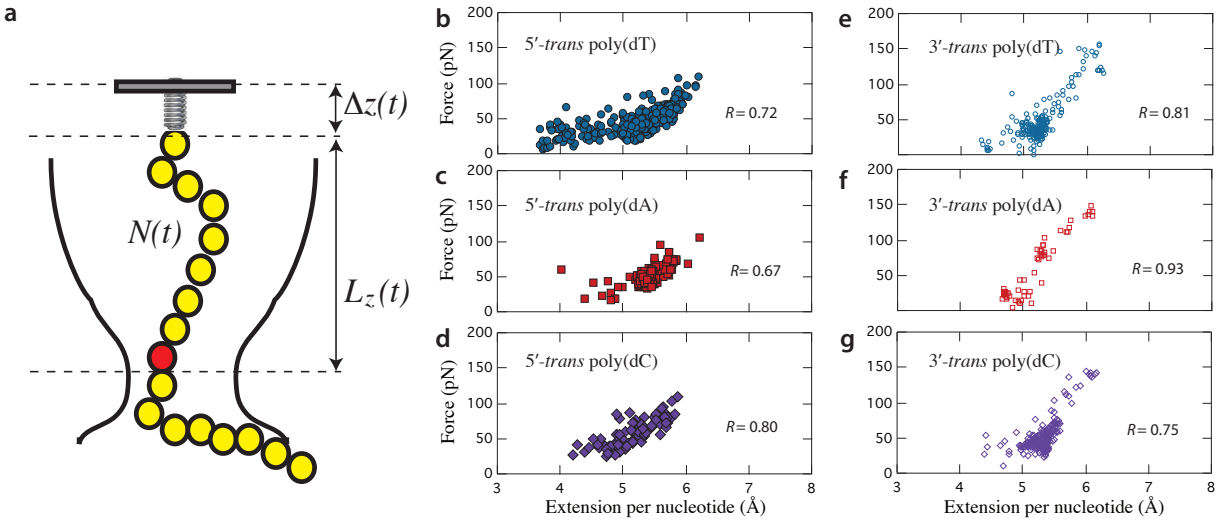
Supplementary Figure 9: The origin of the 1/f noise in the simulated ionic current recordings. (a,b) Power spectral density of the number of bulk-like water molecules in the MspA constriction (panel a) and of the number of non-hydrogen atoms of DNA nucleotides in the constriction (panel b) obtained from the simulations of the minimal MspA system in the presence of DNA. The symbols show the block-average values of the respective PSDs. The 1/f dependence of the PSDs develops in the same frequency range as that of the ionic current PSD, Figure 1c of the main tex.



Supplementary Figure 10: Microscopic mechanism of the ionic current blockades: 3′-*trans* homopolymer data. **(a)** Ionic current *versus* the number of bulk-like water molecules in the constriction of MspA for the 3′-*trans* DNA homopolymer systems. The gray line shows a power-law fit to the data. Supplementary Table 2 lists parameters of the fit for individual trajectories. All data points in panels a–d show 100-ns averages of the 100-ps sampled data from the respective MD trajectories. **(b)** The number of bulk-like water molecules *versus* the number of non-hydrogen DNA atoms in the constriction of MspA. The gray line shows a linear fit to the data. **(c, d)** The ionic current *versus* the number of non-hydrogen atoms of DNA nucleotides (panel c) or DNA bases (panel d) in the constriction of MspA. The dotted lines are guides for the eyes. **(e)** The probability of observing a given number of non-hydrogen atoms of DNA nucleotides (top) and a given number of bulk-like water molecules (bottom) for the 3′-*trans* poly(dT) (right) and 3′-*trans* poly(dC) systems. Vertical black lines indicate the mean values. The bottom panels show the cumulative ionic currents as a function of N_{DNA} (left) and N_{bw} (right). The cumulative ionic currents were computed by integration of the probability density functions weighted by the mean ionic current in each bin. **(f)** The number of base stacks in the constriction *versus* the number of bulk-like water molecules. Bins with fewer than five data points were ignored. Error bars indicate standard deviation of data points in each bin. **(g)** Normalized probability of observing a given number of base-stacked conformations among nucleotides in the constriction of MspA.



Supplementary Figure 11: MD simulation of the 5′-*trans* TTTAAATTTTT system. **(a)** A sequence of snapshots illustrating the simulation of the mixed-sequence ssDNA–MspA system. In each panel, a fragment of the MspA nanopore is shown in grey as a molecular surface. The bases of DNA are shown using the molecular bonds representation; all thymine bases are shown in cyan whereas the 4th, 5th, and 6th adenine base (counting from the anchor point) is shown in red, blue, and orange, respectively. The DNA backbone is shown in yellow using spheres; water and ions are not shown. **(b)** The blockade current (red) and the number of bulk-like water molecules (black) in the MspA constriction *versus* the simulation time for the minimal MspA systems containing the mixed-sequence DNA strand. Each data point shows a 100-ns block average of the 100-ps sampled MD trajectory. The Pearson's correlation coefficient between $I(t)$ and $N_{\mathrm{bw}}(t)$ is 0.85. **(c)** The location of the centers of mass of five DNA nucleotides *versus* simulation time. The color of the lines that denote the locations of adenines corresponds to the color of nucleotides shown in panel a. The z-coordinate is defined in panel a. The black dashed line indicates the location of the middle plane of the MspA constriction ($z = -11$ Å). **(d)** Ionic current *versus* the number of bulk-like water molecules in the constriction of MspA for the mixed-sequence DNA system. The gray line shows a power-law fit to the data. **(e)** The number of bulk-like water molecules *versus* the number of non-hydrogen DNA atoms in the constriction of MspA. The gray line shows a linear fit to the data.



Supplementary Figure 12: Displacement of ssDNA correlates with the force on the restraining anchor. (a) Schematic illustrating analysis of the MD trajectories. The restraining force on the terminal nucleotide $F = k\,\Delta z(t)$, where $\Delta z(t)$ is the displacement of the anchor atom from its initial coordinate (at $t = 0$) along the pore axis and k is the spring constant of the restraining potential. Extension per nucleotide is computed as the ratio of the extension of the DNA fragment located above the constriction along the pore axis $L_z(t)$ to the number of nucleotides located above the constriction $N(t)$. The number of nucleotides was computed by counting the number of backbone atoms above the constriction and dividing that number by the number of backbone atoms in a single nucleotide. **(b-g)** Scatter diagrams showing the correlation between the restraining force and the extension per nucleotide for thymine (b,e), adenine (c,f) and cytosine (d,e) homopolymer systems having either 5′ (b-d) or 3′ (e-g) ends located at the *trans* side of the membrane. The Pearson′s correlation coefficient R between the restraining force and the extension per nucleotide are given for each system in the corresponding plot. All the data points show 100-ns block averages of the 100-ps sampled data of the respective MD trajectories.

Supplementary Note 4: Analysis of the effective force on ssDNA in MspA

In the absence of screening, the force on a single nucleotide can be estimated as $F_{\text{nucl}} = Eq$, where E is the electric field and q is the charge of the nucleotide (equal to the charge of an electron). The electrostatic potential map of MspA in a lipid bilayer membrane reveals that the drop of the transmembrane bias occurs over the ∼1 nm-long constriction of the nanopore, Supplementary Figure 2. The electric field in the constriction can be estimated as $\frac{180 \text{ mV}}{1 \text{ nm}} = 1.8 \cdot 10^8$ V/m. Multiplied by the charge of an electron $e = 1.6 \cdot 10^{-19}$ C, the force on a single nucleotide in the MspA constriction is 28.8 pN. At any given moment, 1 to 3 nucleotides are found in the constriction of the MspA, giving the unscreened force range of ∼30–90 pN.

Several screening mechanisms can lower the force on ssDNA in a nanopore: direct counterion binding, electro-osmotic flow and DNA-pore interactions.[19] Analysis of our MD trajectories has shown no permanent (in excess of 10 ps) binding of ions to DNA in the MspA constriction. In fact, the local electric field in the constriction of MspA is considerably greater than the local electric field produced by a DNA phosphate at the closest distance of approach of a hydrated ion $r = 0.6$ nm:[20] $E_{\text{approach}} = e/(4\pi\epsilon\epsilon_0 r^2) = 0.5 \cdot 10^8$ V/m, where $\epsilon = 80$ is the dielectric constant of water and $\epsilon_0 = 8.85 \cdot 10^{-12}$ F/m is the electric constant.

The electro-osmotic flow is the dominant screening mechanism in the case of nanopores that are six or more nanometers in diameter.[21,22] For a 6 nm-diameter pore, the drag of the electro-osmotic flow is estimated to effectively reduce the charge of a double-stranded DNA by approximately 50%, with the additional 25% reduction coming from ions binding to DNA.[21,23] The drag force of the electro-osmotic flow decreases with the pore diameter,[21,22] increasing the effective force. In the case of MspA, the water flow is not negligible: the average velocity of water molecules in the pore constriction is 0.1 nm/ns, Supplementary Figure 13a,b. There is, however, no substantial correlation between the effective force on the DNA and the water velocity or current, Supplementary Figure 13c,d. Given the very small size of the nanopore in MspA (about 1 nm diameter) and the complex conformation of

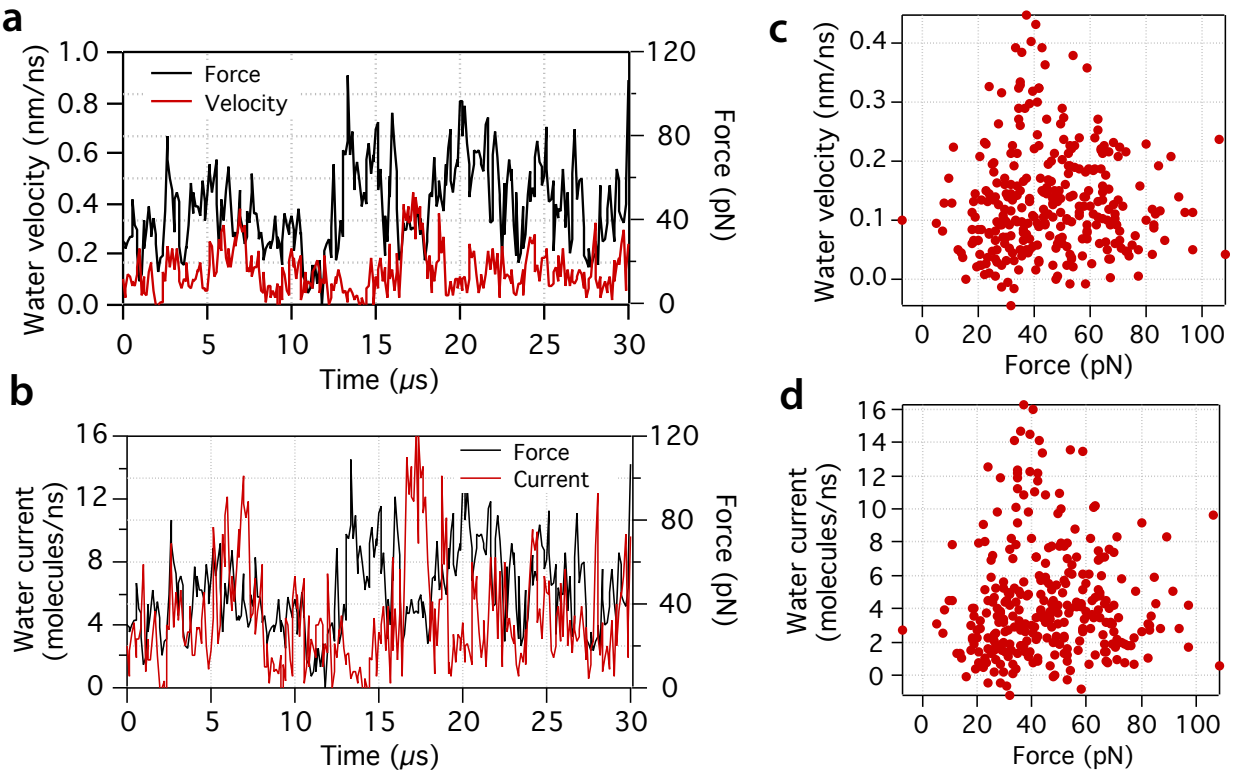
ssDNA in the constriction, it is not possible to put a precise estimate on the effective screening contributed by the electro-osmotic flow. According to the data available for dsDNA in solid-state nanopores, the electro-osmotic screening should not reduce the force of the electric field on DNA by more than 50%. The reduction can be considerably smaller. Newton's relationship for the drag force $F_{\text{drag}} = A\eta\dfrac{dv}{dx}$ predicts a sub-pN force from the electro-osmotic flow, using $A = \pi \cdot 1 \text{ nm} \cdot 1 \text{ nm}$ as the area of the DNA molecule in the constriction of MspA, the dynamic viscosity of water $\eta = 10^{-3}$ N s m^{-2}, and the shear rate of $\dfrac{0.1 \text{ nm ns}^{-1}}{1 \text{ nm}}$.

The effect of direct DNA-nanopore interactions is difficult to ascertain quantitatively, as different chemical groups at the surface of the protein exhibit different affinity to DNA. Binding of the DNA phosphate groups to the amine groups of Asn90 and Asn91 residues was found to have the highest correlation with the force on the anchor point, Supplementary Fig. 14. When tethered by a harmonic potential to the anchor point, the DNA strand undergoes large-amplitude collective motion through the MspA constriction, Figure 5b,c of the main text. As DNA nucleotides pass through the constriction of the pore, their phosphate groups (mostly the phosphates of the 5th and 6th nucleotides, counting from the anchor point) bind to the amine groups of the asparagine side chains Asp90 and Asp91 that define the constriction of MspA (sixteen residues altogether). When unbound from the asparagine residues, the 5th and 6th phosphate groups are located on average above and below the constriction, respectively. Each of the phosphate groups can bind up to three asparagine side chains simultaneously. The simultaneous binding of a phosphate group to three asparagine side chains can produce a long-lasting (up to 2 microseconds) contact, during which the phosphate group can change its binding partners. The high-tension states observed in our simulations correlate with prolonged binding of the 5th phosphate group to the ring of Asn90/Asn91 residues, Supplementary Fig. 14.

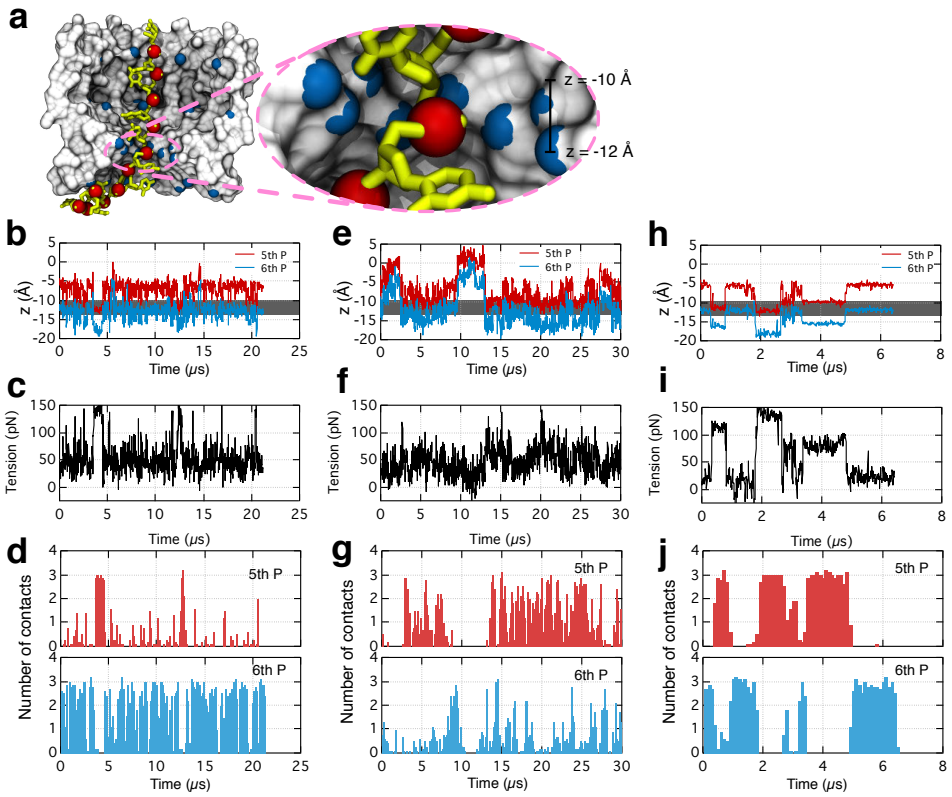
The effective force on ssDNA in MspA was measured in MD simulations performed using a reduced-length model of MspA and periodic boundary conditions. As detailed in Supplementary Note 2, the reduced-length model did not include the resistive elements associated

with the access resistance of the solution. Thus, in comparison to the experimental system, the transmembrane bias drops over a shorter distance in our reduced-length MspA systems, producing a higher electric field in the MspA constriction. As a result, the effective force measured in our MD simulations of reduced-length MspA systems could be overestimated by up to 30%. The values of the force reported in Figure 5d and Supplementary Figures 12, 13 and 14 are not corrected to account for the missing resistive elements and could be up to 30% smaller in the experimental system under the same transmembrane bias.

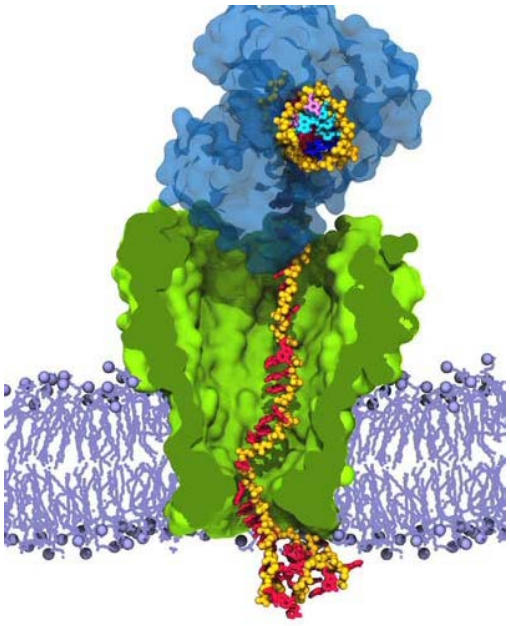
The stall force of the phi29 DNA polymerase—the polymerase used in nanopore experiments with MspA and in our study—is 37 pN,[24] which is well within the range of the forces that we find from our all-atom MD simulations. To achieve the best resolution of the DNA sequence during a nanopore measurement, it may be tempting to operate the DNA polymerase in the synthesis mode at transmembrane biases that nearly stall the polymerase: the high force on DNA reduces the rate the nucleotides are incorporated in the growing DNA strand, increasing accuracy of the ionic current determination. The presence of high-amplitude force fluctuations would, however, raise the probability of DNA polymerase backstepping, increasing the rate of insertion errors.



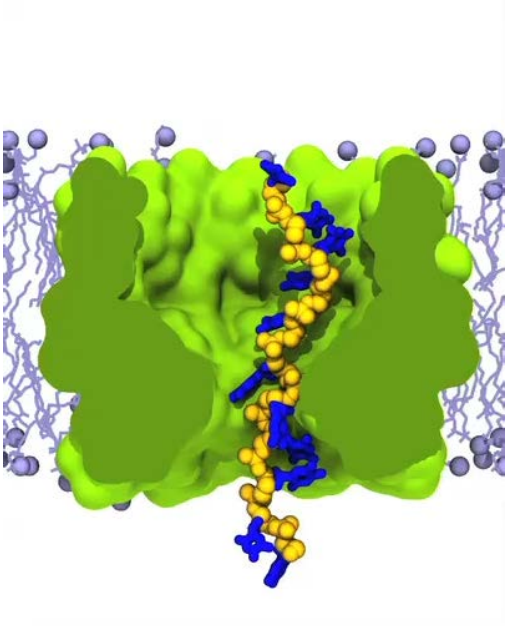
Supplementary Figure 13: Electro-osmotic effect. (a,b) Velocity (a) and current (b) of water molecules in the constriction of MspA during the MD simulation of the 5′-*trans* poly(dT) system. The force on the anchor atom is shown for comparison. To determine the instantaneous velocity and the instantaneous current, the flux (current per unit area) and the density of water oxygen atoms were computed in a volume containing the MspA constriction, $|x| < 5$ Å, $|y| < 5$ Å and $-15 < z < -7$, using a previously described method.[26] The instantaneous water velocity was obtained by dividing the z component of the water flux by the water density; the instantaneous current was obtained by multiplying the z component of the water flux by the $x - y$ cross section of the volume (1 nm^2). Data shown in panels a and b represent 100 ns averages of the instantaneous (100 ps-sampled) velocity and current values. **(c,d)** Scatter plots of the water velocity (c) and water current (d) with respect to the force on the anchor atom. Each data point represents a 100 ns average of 100 ps-sampled data.



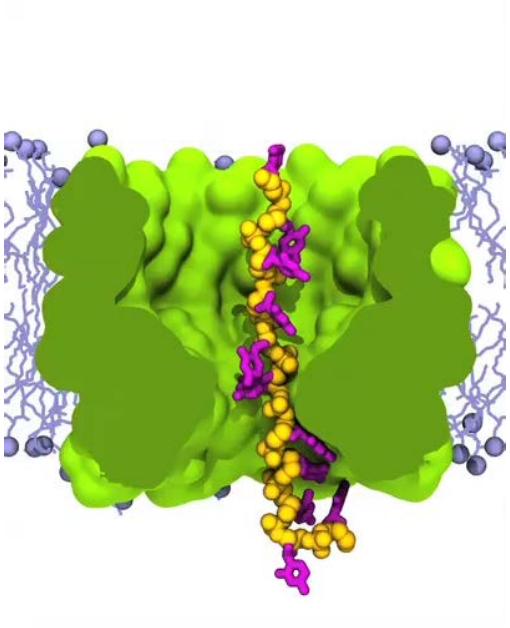
Supplementary Figure 14: Transient binding of DNA phosphates to Asn90/Asn91 residues stabilizes high-tension states. **(a)** A microscopic configuration of the 3′-*trans* poly(dC) system in one of the high-tension states. A fragment of the MspA nanopore is shown in grey as a molecular surface; the DNA strand is shown in yellow using the molecular bonds representation; water and ions are not shown. The blue and red spheres indicate the locations of the nitrogen atoms of the asparagine amine groups and the phosphorus atoms of DNA, P, respectively. In our coordinate system, the amine groups of the Asn90/Asn91 residues are located within the $-12 < z < -10$ Å interval. In the conformation shown, the phosphate group of the 5th nucleotide from the anchor atom makes simultaneous contacts with three amine groups of the Asn90/Asn91 ring. This particular configuration corresponds to a state observed after 4 μs from the beginning of the simulation. **(b)** The location of the phosphorus atom of the 5th and 6th nucleotide (counting from the anchor atom) during the MD simulation of the 3′-*trans* poly(dC) system sampled every 10 ns. The grey box indicates the location of the ring of the Asn90/Asn91 amine groups. **(c)** Measured tension in the DNA strand, same as shown in Figure 3d of the main text. **(d)** The number of simultaneous contacts formed by the phosphate group of the 5th (top) or the 6th (bottom) DNA nucleotide with the amine groups of the Asn90 and Asn91 residues during the MD simulation of the 3′-*trans* poly(dC) system. To count as a contact, the distance between the phosphorus atom of a nucleotide and the nitrogen atom of an amine group must be shorter than 5 Å. The number of contacts formed by the 5th nucleotide anti-correlates with the number of contacts formed by the 6th nucleotide and correlates with the transient increase of the effective force, panel c. **(e–g)** Same as in panels b, c and d, plotted for the 5′-*trans* poly(dT) system. **(h–j)** Same as in panels b, c and d, plotted for the 3′-*trans* poly(dA) system. The average z coordinate of 5th nucleotide's phosphorous atom in the 3′-*trans* poly(dC), 5′-*trans* poly(dT) and 3′-*trans* poly(dA) system is -7.2, -7.0, and -8.1 Å, respectively.



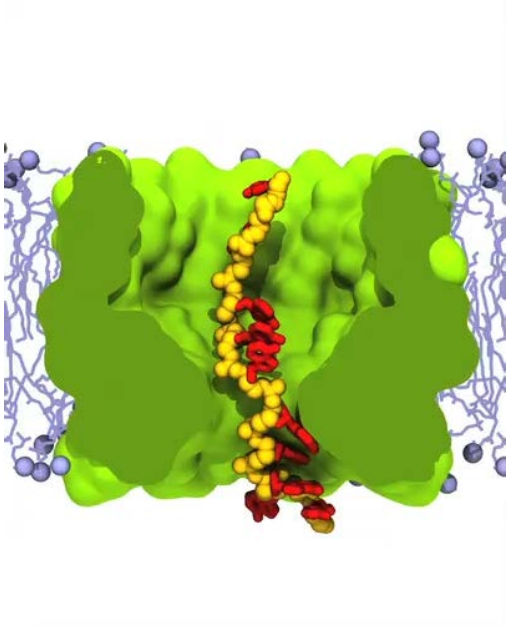
Supplementary Movie 1: Molecular dynamics simulation of the full-length MspA–DNA–phi29 polymerase system under a 180 mV bias. The MspA nanopore is shown as a green molecular surface cut-away to reveal a template DNA strand threaded with its 5′-end through the MspA constriction and its 3′-end through the phi29 polymerase (semitransparent blue molecular surface). A DNA strand complementary to the template strand forms a DNA duplex between the active site of the DNA polymerase and the 3′ end of the template strand. The MspA nanopore is embedded in a POPC bilayer (purple). Water and ions are not shown. During the course of this 130-ns simulation, the polymerase and the duplex part of the DNA rotate about the axis of the pore by more than 90°. Such large scale motion of the polymerase did not have a measurable effect on the conformation of the template strand in the MspA constriction.



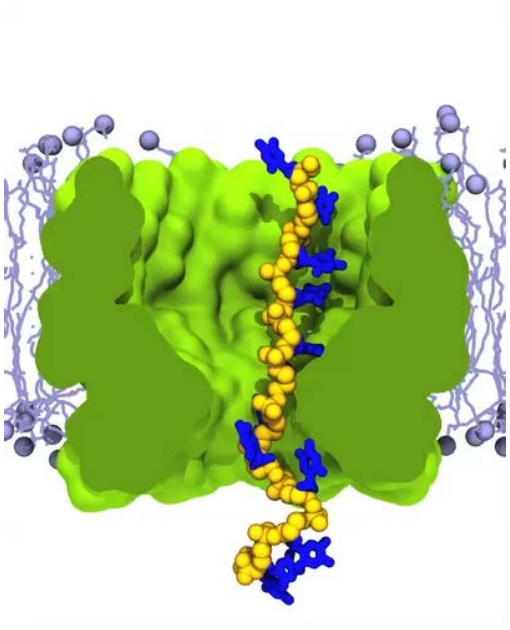
Supplementary Movie 2: Molecular dynamics simulation of the 5′-*trans* poly(dT) system. The reduced-length MspA nanopore is shown as a cut-away green molecular surface, the backbone of the DNA fragment is shown as yellow spheres, and the DNA bases are shown in blue; the lipid bilayer is shown as purple lines and spheres. Water and ions are not shown. The uppermost nucleotide of the DNA strand is harmonically restrained, which mimics the action of a DNA polymerase. This animation illustrates a 9.77 μs fragment of the all-atom trajectory obtained under a 180 mV transmembrane bias.



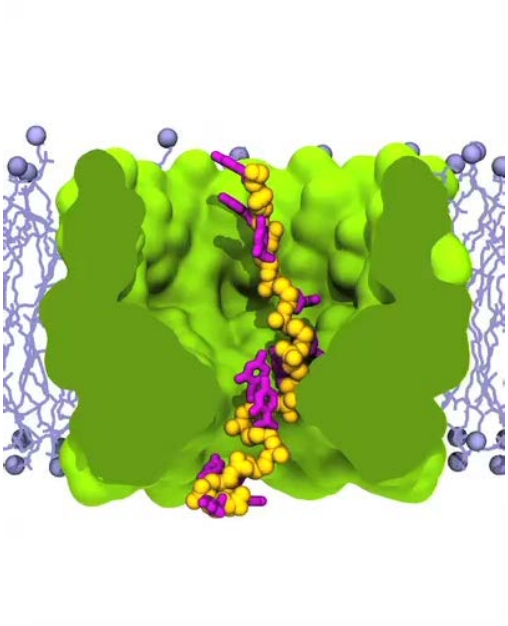
Supplementary Movie 3: Molecular dynamics simulation of the 5′-*trans* poly(dC) system. The reduced-length MspA nanopore is shown as a cut-away green molecular surface, the backbone of the DNA fragment is shown as yellow spheres, and the DNA bases are shown in purple; the lipid bilayer is shown as purple lines and spheres. Water and ions are not shown. The uppermost nucleotide of the DNA strand is harmonically restrained, which mimics the action of a DNA polymerase. This animation illustrates a 8.77 μs fragment of the all-atom trajectory obtained under a 180 mV transmembrane bias.



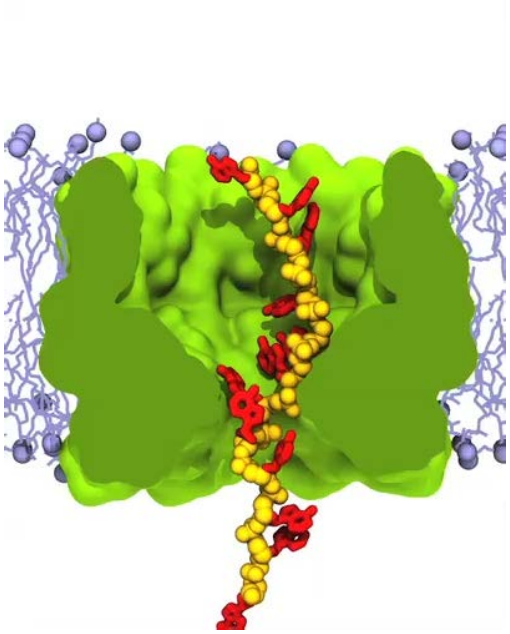
Supplementary Movie 4: Molecular dynamics simulation of the 5′-*trans* poly(dA) system. The reduced-length MspA nanopore is shown as a cut-away green molecular surface, the backbone of the DNA fragment is shown as yellow spheres, and the DNA bases are shown in red; the lipid bilayer is shown as purple lines and spheres. Water and ions are not shown. The uppermost nucleotide of the DNA strand is harmonically restrained, which mimics the action of a DNA polymerase. This animation illustrates a 8.46 μs fragment of the all-atom trajectory obtained under a 180 mV transmembrane bias.



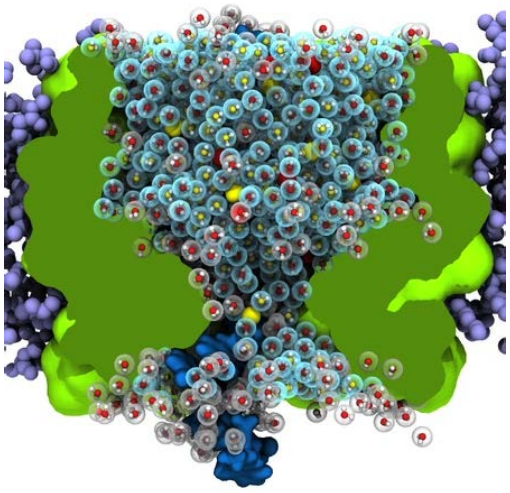
Supplementary Movie 5: Molecular dynamics simulation of the 3′-*trans* poly(dT) system. The reduced-length MspA nanopore is shown as a cut-away green molecular surface, the backbone of the DNA fragment is shown as yellow spheres, and the DNA bases are shown in blue; the lipid bilayer is shown as purple lines and spheres. Water and ions are not shown. The uppermost nucleotide of the DNA strand is harmonically restrained, which mimics the action of a DNA polymerase. This animation illustrates a 6.41 μs fragment of the all-atom trajectory obtained under a 180 mV transmembrane bias.



Supplementary Movie 6: Molecular dynamics simulation of the 3′-*trans* poly(dC) system. The reduced-length MspA nanopore is shown as a cut-away green molecular surface, the backbone of the DNA fragment is shown as yellow spheres, and the DNA bases are shown in purple; the lipid bilayer is shown as purple lines and spheres. Water and ions are not shown. The uppermost nucleotide of the DNA strand is harmonically restrained, which mimics the action of a DNA polymerase. This animation illustrates a 5.76 μs fragment of the all-atom trajectory obtained under a 180 mV transmembrane bias



Supplementary Movie 7: Molecular dynamics simulation of the 3′-*trans* poly(dA) system. The reduced-length MspA nanopore is shown as a cut-away green molecular surface, the backbone of the DNA fragment is shown as yellow spheres, and the DNA bases are shown in red; the lipid bilayer is shown as purple lines and spheres. Water and ions are not shown. The uppermost nucleotide of the DNA strand is harmonically restrained, which mimics the action of a DNA polymerase. This animation illustrates a 6.41 μs all-atom trajectory obtained under a 180 mV transmembrane bias.



Supplementary Movie 8: Microscopic mechanism of ion passage through the MspA constriction. This animation illustrates a short (3 ns) fragment of the 5′-*trans* poly(dT) trajectory. The instantaneous coordinates of the system were recorded every 5 ps for this fragment of the trajectory. The reduced-length MspA channel is shown as a cut-away green molecular surface, the DNA fragment is shown in blue, the lipid bilayer in purple, and potassium and chloride ions as yellow and cyan spheres, respectively. Structural and bulk-like water molecules are shown as semitransparent blue and light grey spheres, respectively. Structural water molecules are defined as those located within 2.5 Å of any protein or DNA atom. Flickering of the water molecules at the periphery of the simulation system is produced by water molecules moving in and out of the volume visualized in this animation.

References

1. Berman, A. J.; Kamtekar, S.; Goodman, J. L.; Lázaro, J. M.; de Viga, M.; Blanco, L.; Salas, M.; Steitz, T. A. Structures of phi29 DNA polymerase complexed with substrate: the mechanism of translocation in B-family polymerases. *EMBO J.* **2007**, *26*, 3494–3505.

2. Mathé, J.; Aksimentiev, A.; Nelson, D. R.; Schulten, K.; Meller, A. Orientation discrimination of single stranded DNA inside the α-Hemolysin membrane channel. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 12377–12382.

3. Butler, T. Z.; Pavlenok, M.; Derrington, I. M.; Niederweis, M.; Gundlach, J. H. Single-Molecule DNA Detection with an Engineered MspA Protein Nanopore. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 20647–20652.

4. Bhattacharya, S.; Derrington, I. M.; Pavlenok, M.; Niederweis, M.; Gundlach, J. H.; Aksimentiev, A. Molecular Dynamics Study of MspA Arginine Mutants Predicts Slow DNA Translocations and Ion Current Blockades Indicative of DNA Sequence. *ACS Nano* **2012**, *6*, 6960–6968.

5. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802.

6. Darden, T. A.; York, D. M.; Pedersen, L. G. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

7. Batcho, P. F.; Case, D. A.; Schlick, T. Optimized Particle-Mesh Ewald/Multiple-Time Step Integration for Molecular Dynamics Simulations. *J. Chem. Phys.* **2001**, *115*, 4003–4018.

8. Miyasaka, T.; Koyama, K.; Itoh, I. Quantum Conversion and Image Detection by a Bacteriorhodopsin-Based Artificial Photoreceptor. *Science* **1992**, *255*, 342–344.

9. Andersen, H. C. RATTLE: A "Velocity" Version of the SHAKE Algorithm for Molecular Dynamics Calculations. *J. Comput. Phys.* **1983**, *52*, 24–34.

10. Wells, D. B.; Bhattacharya, S.; Carr, R.; Maffeo, C.; Ho, A.; Comer, J.; Aksimentiev, A. In *Nanopore-based technology: Single molecule characterization and DNA sequencing*; Gracheva, M. E., Ed.; Humana Press: New York, 2012; Vol. 870; Chapter 10, pp 165–186.

11. MacKerell, Jr., A. D.; Bashford, D.; Bellott, M.; Dunbrack, Jr., R. L.; Evanseck, J.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

12. Yoo, J.; Aksimentiev, A. Improved Parametrization of Li^+, Na^+, K^+, and Mg^{2+} Ions for All-Atom Molecular Dynamics Simulations of Nucleic Acid Systems. *J. Phys. Chem. Lett.* **2012**, *3*, 45–50.

13. Koopman, E. A.; Lowe, C. P. Advantages of a Lowe-Andersen Thermostat in Molecular Dynamics Simulations. *J. Chem. Phys.* **2006**, *124*, 204103.

14. Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant Pressure Molecular Dynamics Algorithms. *J. Chem. Phys.* **1994**, *101*, 4177–4189.

15. Isralewitz, B.; Izrailev, S.; Schulten, K. Binding Pathway of Retinal to Bacterio-opsin: A Prediction by Molecular Dynamics Simulations. *Biophys. J.* **1997**, *73*, 2972–2979.

16. Aksimentiev, A. Deciphering Ionic Current Signatures of DNA Transport through a Nanopore. *Nanoscale* **2010**, *2*, 468–483.

17. Kowalczyk, S. W.; Grosberg, A. Y.; Rabin, Y.; Dekker, C. Modeling the Conductance and DNA Blockade of Solid-State Nanopores. *Nanotech.* **2011**, *22*, 315101.

18. Manrao, E. A.; Derrington, I. M.; Pavlenok, M.; Niederweis, M.; Gundlach, J. H. Nucleotide Discrimination with DNA Immobilized in the MspA Nanopore. *PLoS ONE* **2011**, *6*, e25723.

19. Keyser, U. F.; van Dorp, S.; Lemay, S. G. Tether forces in DNA electrophoresis. *Chem. Soc. Rev.* **2010**, *39*, 939–947.

20. Yoo, J.; Aksimentiev, A. Competitive Binding of Cations to Duplex DNA Revealed through Molecular Dynamics Simulations. *J. Phys. Chem. B* **2012**, *116*, 12946–12954.

21. van Dorp, S.; Keyser, U. F.; Dekker, N. H.; Dekker, C.; Lemay, S. G. Origin of the Electrophoretic Force on DNA in Solid-State Nanopores. *Nat. Phys.* **2009**, *5*, 347–351.

22. Luan, B.; Aksimentiev, A. Electro-Osmotic Screening of the DNA Charge in a Nanopore. *Phys. Rev. E* **2008**, *78*, 021912.

23. Kowalczyk, S. W.; Wells, D. B.; Aksimentiev, A.; Dekker, C. Slowing Down DNA Translocation through a Nanopore in Lithium Chloride. *Nano Lett.* **2012**, *12*, 1038–1044.

24. Ibarr, B.; Chemla, Y. R.; Plyasunov, S.; Smith, S. B.; Lázaro, J. M.; Salas, M.; Bustamante, C. Proofreading dynamics of a processive DNA polymerase. *EMBO J.* **2009**, *28*, 2794–2802.

25. Aksimentiev, A.; Schulten, K. Imaging α-Hemolysin with Molecular Dynamics: Ionic Conductance, Osmotic Permeability and the Electrostatic Potential Map. *Biophys. J.* **2005**, *88*, 3745–3761.

26. Li, C.-Y.; Hemmig, E. A.; Kong, J.; Yoo, J.; Hernández-Ainsa, S.; Keyser, U. F.; Aksimentiev, A. Ionic Conductivity, Structural Deformation and Programmable Anistropy of DNA Origami in Electric Field. *ACS Nano* **2015**, *9*, 1420–1433.