Supporting Material

High Bandwidth Protein Analysis Using Solid-State Nanopores

Joseph Larkin,[†] Robert Henley,[†] Murugappan Muthukumar,[‡] Jacob K. Rosenstein,[¶] and Meni Wanunu[†*]

E-mail: wanunu@neu.edu

Table of Contents

SM1. ***Data analysis details*** Current traces were analyzed using OpenNanopore, an open source translocation software package developed by the Radenovic Group at EPFL.(1) The code was modified to allow data importing from Chimera Instruments format. OpenNanopore fits each translocation event with a rectangular pulse, as shown in Figure S1, and extracts three parameters for each event. The dwell time, t_d, is the residence time of a molecule in the pore. The current blockage, ΔI, is the amount of current the molecule excludes from the pore. The inter-event time, δt, is the length of time between the beginning of one event and the beginning of the next. Since multi-level blockade pulses were extremely rare, the code was modified to bypass multi-level events.

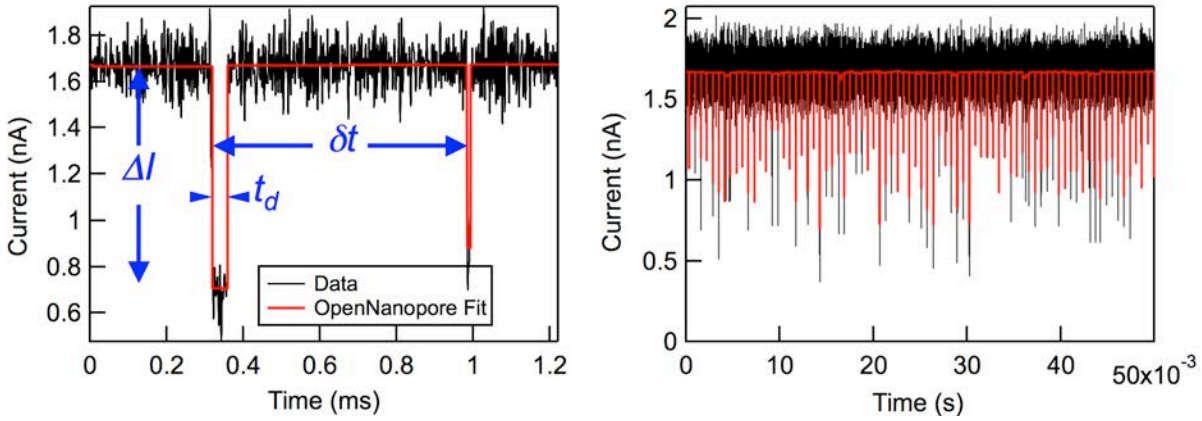

Figure S1. Data Analysis. (left) Close-up trace of two successive events for protK at V = -100 mV. OpenNanopore, the analysis software used, fits each translocation even to a rectangular pulse, and extracts the ΔI, t_d, and δt parameters, displayed in blue. Raw data is in black and the fit from OpenNanopore appears in red. (right) Series of analyzed events for ProtK. All experiments were carried out in the dilute regime, i.e., $\delta t >> t_d$.

Current signals from nanopores are usually considered in the time domain, while random noise is best described in the frequency domain. Nanopore traces are invariably low-pass filtered to reduce noise, but this can filter out some of the signal as well. In thinking about the detection efficiency of brief protein translocations, it can be useful to consider a signal representation which includes both time- and frequency-domain information. Fig. S1b shows a spectrogram produced from a recording of protK translocations through an HfO_2 nanopore. (A spectrogram plots the Fourier transform of a short segment of the data as a function of time; Fig. S1b was produced using a 64-microsecond window computed every 16 microseconds.) The spectral content of longer-duration events is concentrated at lower frequencies, and thus the features of these events are impacted less by the low-pass filter. Faster events' energy is spread more widely across the spectrum, and thus low-pass filtering can remove significant fractions of the signal. It is clear, for example, that 10 kHz bandwidth would be inadequate to capture the full energy of the signals in Fig. S1b. In this work we used a digital low-pass filter with a 250 kHz cutoff frequency.

A range of different algorithms can be used to identify and characterize nanopore events. The simplest is a low-pass filter followed by a threshold-crossing level detector. This approach is not ideal for datasets with appreciable noise, and more advanced techniques can be utilized which incorporate statistical

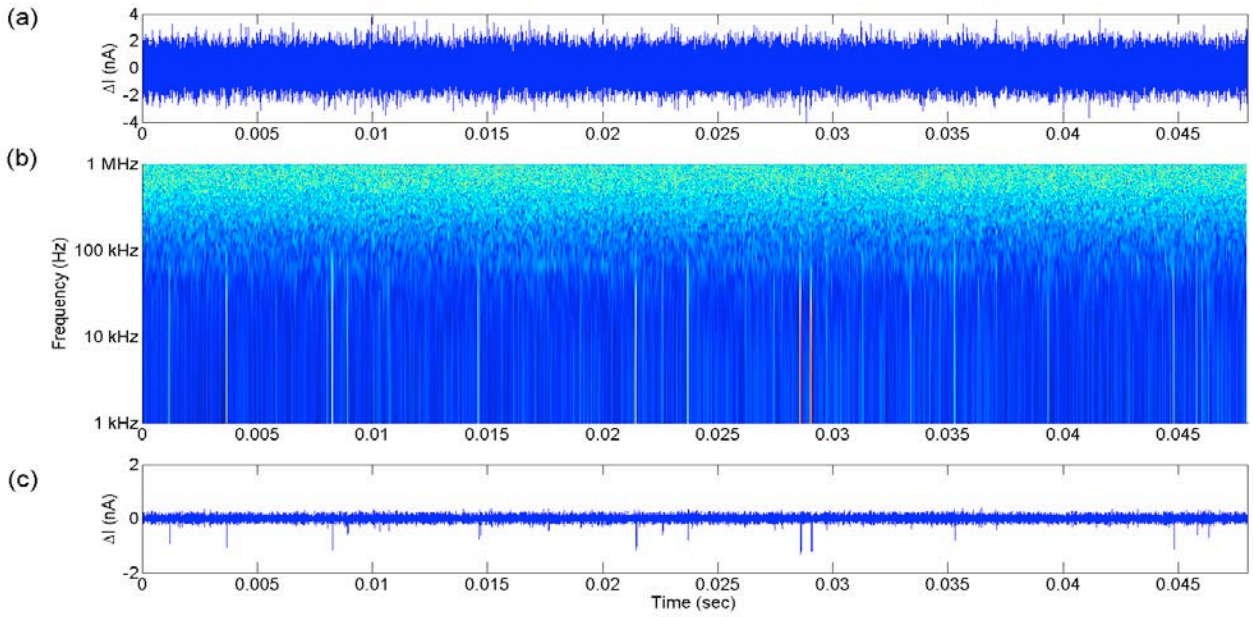

Figure S1b. Frequency content of nanopore current signals. (a) A segment of raw data recorded with a bandwidth of 1 MHz. The signals are obscured by noise. (b) A spectrogram of the trace, showing the frequency content as a function of time (rainbow color scheme, intensity increases from blue to red). (c) The same signal after a digital low-pass filter with a cutoff of 200 kHz reveals the events.

models for both the signal and noise. The OpenNanopore software package used in this work implements a level detector that is optimized for fitting abrupt stepwise signals in the presence of Gaussian noise.

SM2. Determination of pore dimensions We evaluate the pore dimensions (diameter and height) by first translocating a short, 100-bp fragment of double-stranded DNA (dsDNA, see Figure S6 insets). Once translocation data has been obtained with dsDNA and the average ΔI has been measured, the following set of equations, which are derived from a geometric model of pore conductance, were used to evaluate the pore dimensions.(2, 3)

$$I_o = V\sigma\left(\frac{4h_{eff}}{\pi d^2}+\frac{1}{d}\right)^{-1} \quad ; \quad \Delta I = V\sigma\left(\frac{4h_{eff}}{\pi(2.2nm)^2}+\frac{1}{2.2nm}\right)^{-1}$$

where I_O is the measured open pore current, V is the applied trans-membrane voltage, σ is the buffer solution conductivity (measured to be 0.0968 S/cm @ 25°C), and ΔI is the mean current blockage amplitude of the nucleic acid. The unknowns, h_{eff} and d, are the effective height of the pore and the pore diameter respectively. In this model, the current blockage due to the DNA is attributed entirely to the volume of current it excludes from the nanopore. The dsDNA signal, ΔI, may then be used to solve the second equation for h_{eff}, which may then be substituted into the first equation to obtain d. Using this method, the HfO_2 pore in our study was found to have h_{eff} = 7 nm and d = 5.2 nm, and the SiN pore had h_{eff} = 6.2 nm and d = 4.8 nm.

SM3. Size, charge of proteins and effect of electro-osmosis ProtK and RNAse volumes were independently determined using two methods. ZetaPALS analysis was performed at Brookhaven Instruments (Holtsville, NY), which measured effective diameters of 4.76±0.05 nm for ProtK and 3.68±0.03 nm for RNAse A (see results tabulated in **Figure S2**). Similarly, the effective diameter of the RNase:ProtK complex was measured at 1M KCl buffer to be 8.00±0.42 nm. Additionally, the volume of each protein was computed with the Vorlume algorithm.(4) In this computation, a solvent-accessible model was used, where each residue's radius was expanded by a 1.4 Å van der Waals shell. The algorithm returned a solvent-accessible volume of 45.9 nm^3 for ProtK and 32.7 nm^3 for RNAse A.

Both Proteinase K and RNAse A are positively charged in the 1 M KCl, 10mM Tris, 1mM EDTA, pH ~7.7 experimental buffer. Support for this comes from early measurements of the pI for each protein, which obtained pI = 9.6 for RNAse A(5) and pI = 8.9 for Proteinase K.(6) In our experimental pH of 8.1, we would expect these molecules to be positively charged.

Brookhaven Instruments

Summary Statistics Report

Type	Start Date/Time	Sample ID	Eff. Diam. (nm)	Polydispersity	Baseline Index
DLS	7/23/2013 2:14:52 PM	Proeinase K Solution NorthEastern diluted - 3	4.73	0.044	9.5
DLS	7/23/2013 2:09:50 PM	Proeinase K Solution NorthEastern diluted - 2	4.73	0.071	9.4
DLS	7/23/2013 2:04:49 PM	Proeinase K Solution NorthEastern diluted - 1	4.82	0.097	8.5
		Mean:	4.76	0.071	9.1
		Std Err:	0.03	0.015	0.3
		Std Dev:	0.05	0.027	0.6

Brookhaven Instruments

Summary Statistics Report

Type	Start Date/Time	Sample ID	Eff. Diam. (nm)	Polydispersity	Baseline Index
DLS	7/23/2013 11:36:27 AM	RNAse A Solution NorthEastern (f0.02um) - 3	3.65	0.111	9.3
DLS	7/23/2013 11:31:25 AM	RNAse A Solution NorthEastern (f0.02um) - 2	3.68	0.089	8.2
DLS	7/23/2013 11:26:24 AM	RNAse A Solution NorthEastern (f0.02um) - 1	3.71	0.074	7.8
		Mean:	3.68	0.091	8.4
		Std Err:	0.02	0.011	0.4
		Std Dev:	0.03	0.019	0.8

Brookhaven Instruments

Summary Statistics Report

Type	Start Date/Time	Sample ID	Eff. Rad. (nm)	Polydispersity	Baseline Index
DLS	12/2/2013 2:56:50 PM	Protease K + RNAse A, ~ 5 mg/mL, 90deg - 5	3.78	0.227	0.0
DLS	12/2/2013 2:51:48 PM	Protease K + RNAse A, ~ 5 mg/mL, 90deg - 4	4.03	0.245	0.0
DLS	12/2/2013 2:46:46 PM	Protease K + RNAse A, ~ 5 mg/mL, 90deg - 3	4.27	0.242	0.0
DLS	12/2/2013 2:36:43 PM	Protease K + RNAse A, ~ 5 mg/mL, 90deg - 1	3.91	0.224	0.0
		Mean:	4.00	0.234	0.0
		Std Err:	0.11	0.005	0.0
		Std Dev:	0.21	0.011	0.0

Figure S2. Dynamic light scattering measurement of the hydrodynamic diameters of proteinase K, RNAse A, and a 1:1 mixture of the two proteins at 1M KCl. Note that for the complex the report inadvertently displayed the effective radius, while for the isolated proteins measurements displayed are the effective diameters.

SM4. Dwell time distributions and estimation of percentage of events detected, F_{obs}: The t_d distribution used in the paper comes from a first passage calculation using a drift-diffusion model outlined in previous studies.(7, 8) Figure S3a displays the fits to this distribution for ProtK and RNase at several voltages in the 5.2 nm HfO_2 pore. Once we have fit a given set of t_d values to this distribution, we may use the distribution equation to estimate the fraction of detected events. Our time resolution limits detection to events of ~2.5 µs or more. Shorter events are either severely distorted, which makes them erroneously appear in the first 1-2 bins, or not detected at all. However, we can fit our measured dwell times to the distribution equation from the drift-diffusion model. Once we have this total distribution function of dwell times, we may estimate the fraction of translocation events that we detect. We do this by dividing the area under the distribution curve between t = 2.5 µs and t = ∞ (the temporal region we can detect) by the total area under the curve. This process is illustrated in **Figures S3a**. Based on this, using bulk D values we miss 76-96% of the events for RNAse and 62-89% of the events for ProtK, as shown in **Figure S3b**.

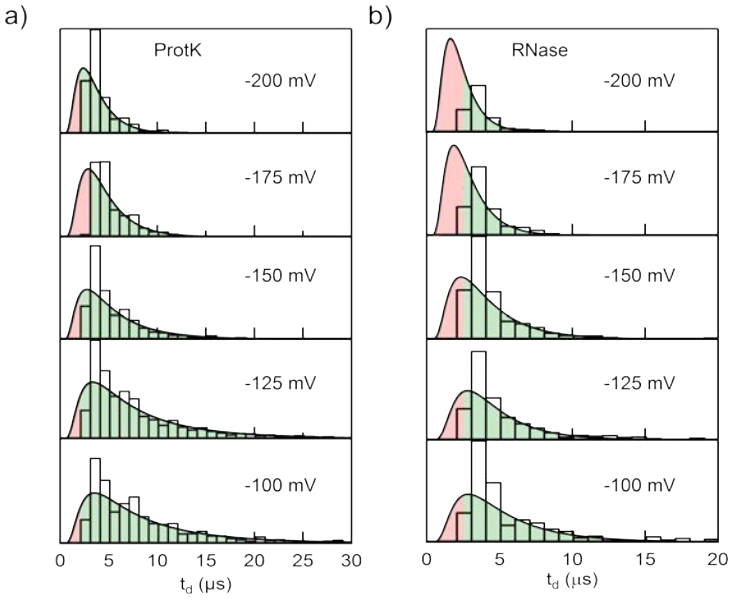

Figure S3a. Fitted t_d distributions for RNAse A (left) and Proteinase K (right) using freely varying *D* values. Fitted distributions are shown by the black curve. These curves are used to estimate the fraction of detected events (F_{Obs}) at each voltage. This fraction is the area under the curve from *t* = 2.5 µs to t=∞, represented by the green region, divided by the total area under the curve.

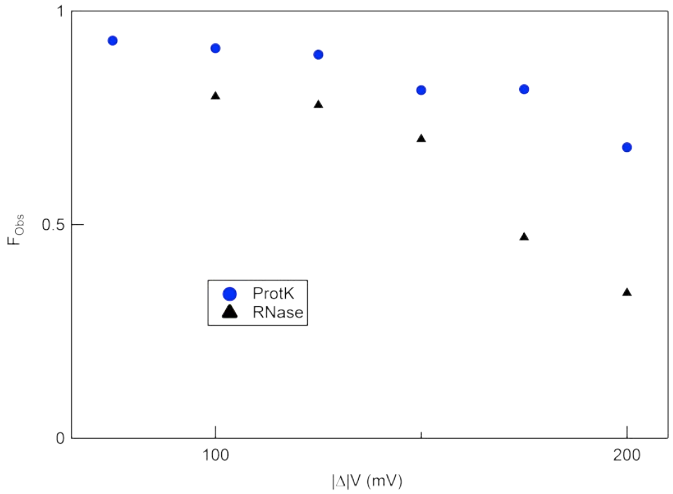

Figure S3b. Estimated fraction of observed events, F_{Obs}, for ProtK and RNase in the 5.2 nm HfO_2 pore discussed in the main text.

SM5. Continuous electrical traces of protein translocations

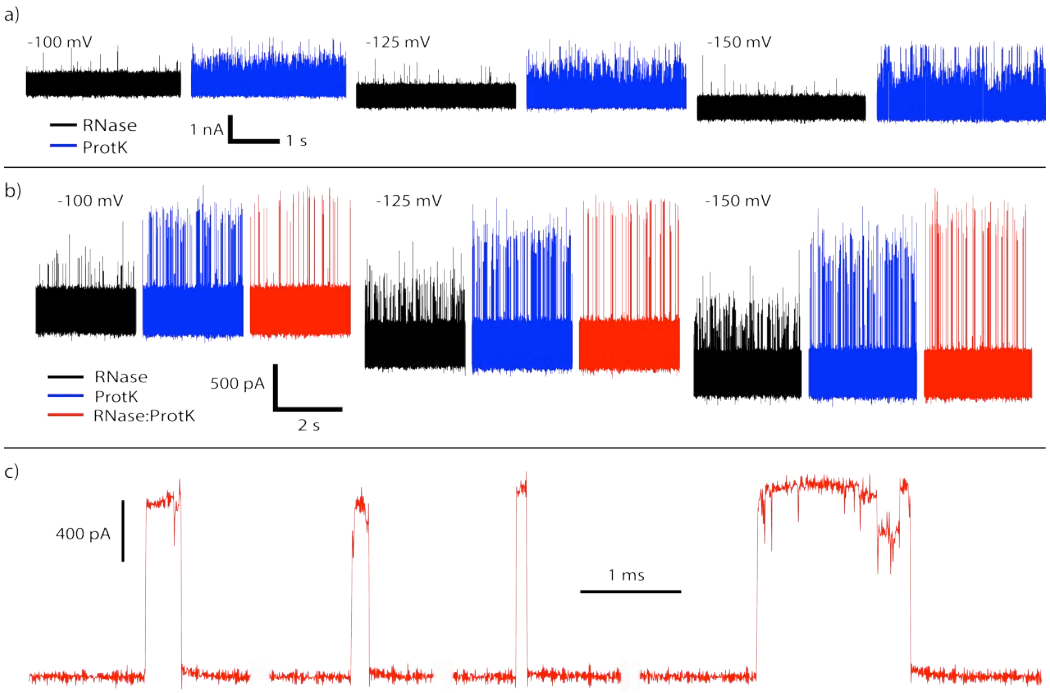

Figure S4. a) Continuous, 3-second traces current traces for RNAse (1 nM) and ProtK (69 nM) at ΔV = -100 mV, -125 mV, and -150 mV in a 5.2 nm HfO_2 pore. B) Traces of RNase (~30 nM), ProtK (~40 nM), and a 1:1 mixture of the two through a 4.8 nm SiN nanopore. C) Sample events from the RNase:ProtK complex (all traces low-pass filtered to 250 kHz).

SM6. Protein analysis using silicon nitride (SiN) pore Here we present supplementary data for the 4.8 nm diameter SiN pore discussed in the main article. Figure S5a shows scatterplots of translocation data for RNase and ProtK through this pore.

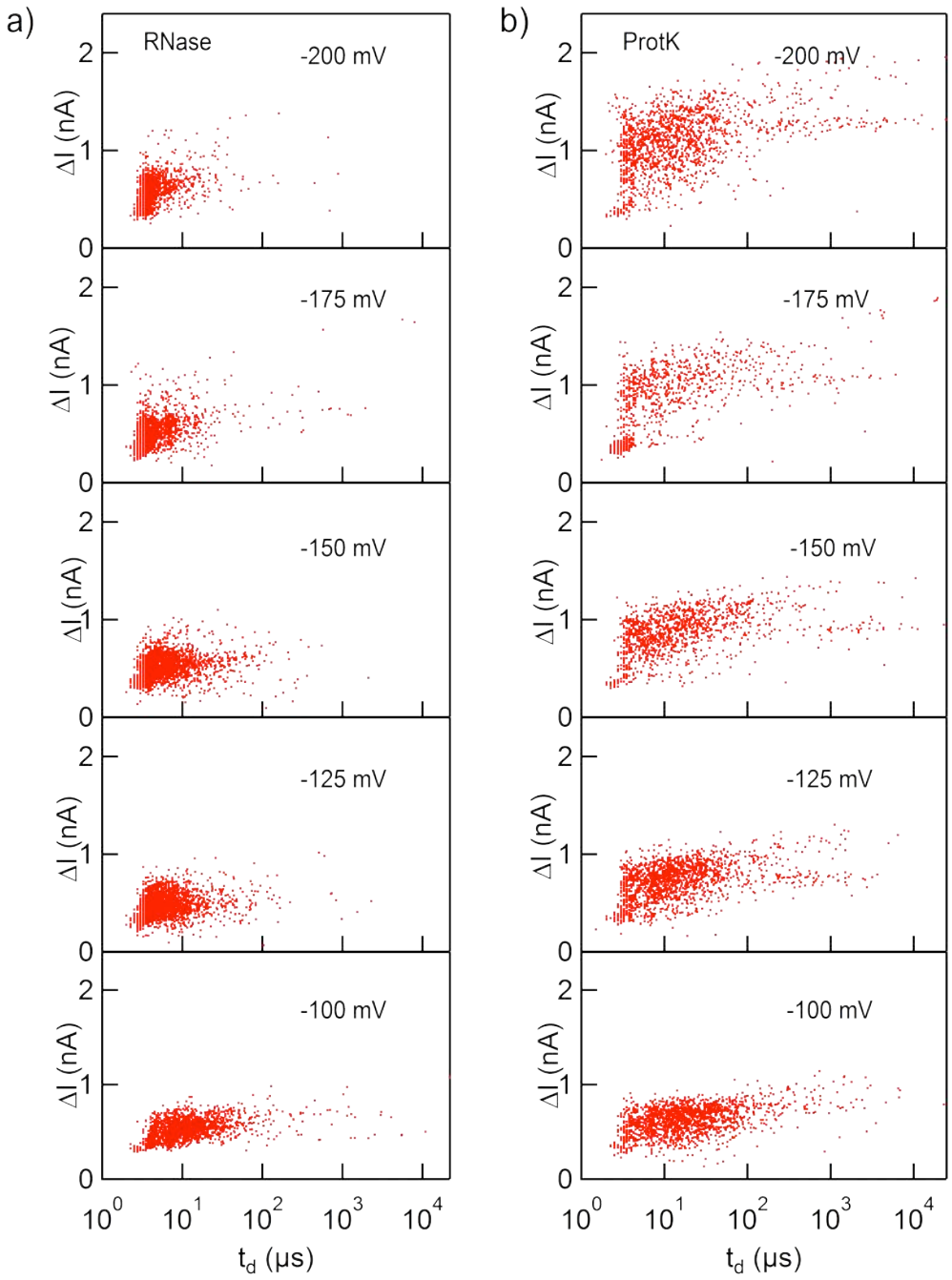

Figure S5a. Scatterplots of a) RNase and b) ProtK translocations from -100 mV to -200 mV in a 4.8 nm SiN pore.

In Figure S5b, we present the dwell time distributions for the SiN data, along with the corresponding fits to the 1-D first passage model (Eq. 1 in main text of the paper).

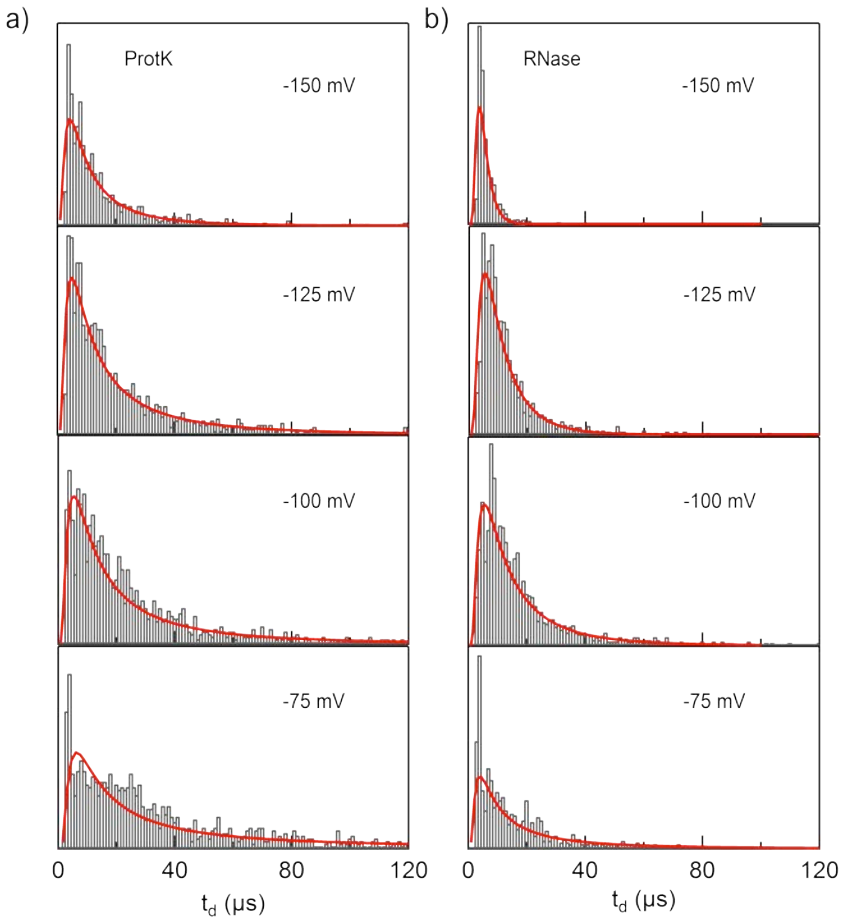

Figure S5b. Dwell time distributions for the 4.8 nm SiN_x nanopore at ΔV = -75 mV, -100 mV, -125 mV, and -150 mV. D and v values obtained from these fits are displayed in Figure 5 of the main manuscript text.

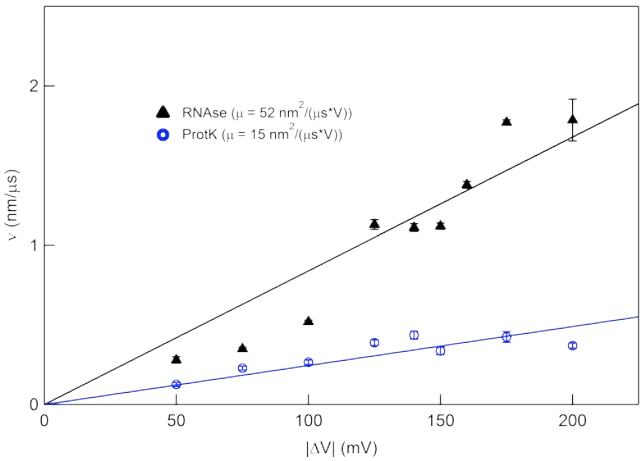

Figure S5c. Drift velocity (*v*) vs. applied voltage (ΔV) for the SiN nanopore with *d* = 4.8 nm, h_{eff} = 6.2 nm. Results from these data are plotted in the scatter plot in Figure 5 of the main text.

Mobilities of proteins in SiN pore: As with the 5.2 nm HfO_2 pore, from the distributions fitted in Figure S5b, we extract drift velocities. These velocities increase linearly with magnitude of applied voltage, as shown below in Figure S5c. Although the RNase data somewhat deviates for this pore at low voltage, the general trend is expected for electrophoretic transport through the pore. Mobility values for the SiN pore are slightly smaller than for the HfO2 pore, although the mobility difference relationship between the two proteins is maintained.

Dwell-time distribution of 1:1 RNase:ProtK complex: The measured volume of the complex resulting from RNase/ProtK mixture is nearly the sum of the two proteins' volumes (see above). However, we see in Figure S5d that the mixture displays dwell times that are orders of magnitude longer than either protein individually. Because the pore in this study has *d* = 4.8 nm, we believe the complex formed from the RNase/ProtK mixture is too large to fit through the pore unhindered. These long dwell times may be due to dissociation or unfolding of the complex.

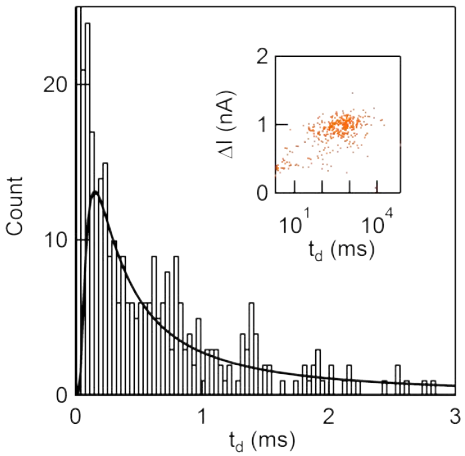

Figure S5d. Dwell time distribution and scatterplot (inset) of translocation data for a 1:1 mixture of RNase and ProtK at *ΔV* = -125 mV (shown in Figure 6 in main text) in a SiN pore illustrates the orders of magnitude longer dwell times for the 1:1 complex than for either protein by itself.

SM7. DNA Dwell Time Distributions Figure S6 displays the measured dwell time distributions and resulting fits for 100 base pair DNA translocating through a 5 nm HfO_2 pore. The drift velocities from these fits are displayed in Figure 4 in the main text.

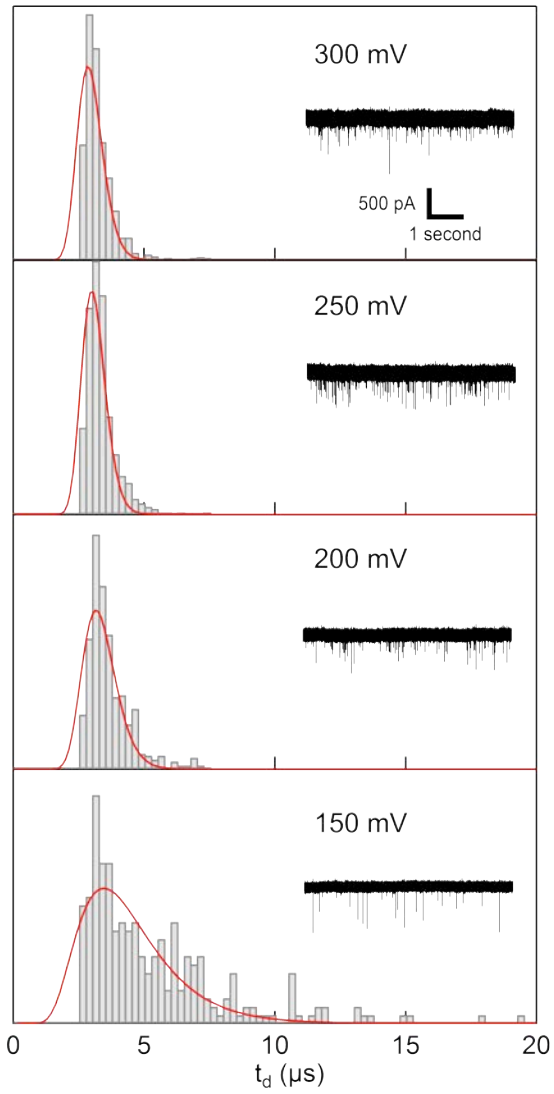

Figure S6. Distributions of dwell times for 100 base pair DNA in a 5nm HfO_2 pore. Insets provide continuous, 3-second example traces for 100 base pair at each voltage. As with the proteins, apparent event rate drops at the highest voltage as events become too fast for the displayed bandwidth.

References:

1. Raillon, C., P. Granjon, M. Graf, L.J. Steinbock, and A. Radenovic, Fast and automatic processing of multi-level events in nanopore translocation experiments. *Nanoscale*, 2012. 4(16): 4916-4924.
2. Wanunu, M., T. Dadosh, V. Ray, J.M. Jin, L. McReynolds, and M. Drndic, Rapid electronic detection of probe-specific microRNAs using thin nanopore sensors. *Nat Nanotechnol*, 2010. 5(11): 807-814.
3. Kowalczyk, S.W., A.Y. Grosberg, Y. Rabin, and C. Dekker, Modeling the conductance and DNA blockade of solid-state nanopores. *Nanotechnology*, 2011. 22(31): 315101.
4. Cazals, F., H. Kanhere, and S. Loriot, Computing the Volume of a Union of Balls: A Certified Algorithm. *Acm T Math Software*, 2011. 38(1).
5. Tanford, C. and J.D. Hauenstein, Hydrogen Ion Equilibria of Ribonuclease1. *J Am Chem Soc*, 1956. 78(20): 5287-5291.
6. Ebeling, W., N. Hennrich, M. Klockow, H. Metz, H.D. Orth, and H. Lang, Proteinase K From Tritirachium-Album. *Eur J Biochem*, 1974. 47(1): 91-97.
7. Muthukumar, M., Polymer Translocation. 2011: CRC Press.
8. Ling, D.Y. and X.S. Ling, On the distribution of DNA translocation times in solid-state nanopores: an analysis using Schrodinger's first-passage-time theory. *J Phys: Condens Matter*, 2013. 25: 375102.