Electronic Supplementary Material (ESI) for Nanoscale.

Supporting Information

Discrimination of RNA Fiber Structures Using Solid-State Nanopores

Prabhat Tripathi[1], Morgan Chandler[3], Christopher Michael Maffeo[5], Ali Fallahi[2], Amr Makhamreh[2], Justin Halman[3], Aleksei Aksimentiev[4,5*], Kirill A. Afonin[3*], and Meni Wanunu[1,2*]

[1]Department of Physics, [2]Department of Bioengineering, Northeastern University, Boston, MA, 02115, [2] Department of Chemistry, University of North Carolina at Charlotte, Charlotte, NC 28223, USA [3]Beckman Institute for Advanced Science and Technology and [4]Department of Physics, University of Illinois at Urbana- Champaign, Urbana, Illinois, 61801, USA

List of Sequences

Fiber strands assemble via ~180° HIV kissing loop motifs in an AB system as previously described[1-2]. A-NF and B-NF denote the two non-functionalized fiber strands which are the components of NF fibers. A-F and B-F denote functionalized fiber strands which are comprised of the NF fiber core sequences with the 3'-end addition of an RNA antisense sequence designed against green fluorescent protein (GFP). To this, the complementary (sense RNA) strand is added to complete the functional moiety. All RNA sequences used in this work were prepared *via* in vitro transcription except for complement strand, which was purchased directly from Integrated DNA Technologies, Inc. The underlined region in functionalized strands denotes the complementary region to complementary strand.

A-NF:
5'GGGAAUCCAAGGAGGCAGGAUUCCCGUCACAGAAGGAGGCACUGUGAC

B-NF:
5'GGGAACGUAAGCCUCCAACGUUCCCGGAUGCUAAGCCUCCAAGCAUCC

A-F:
5'GGGAAUCCAAGGAGGCAGGAUUCCCGUCACAGAAGGAGGCACUGUGACUUUGGUGGUGCAGAUGAACUUCAGGGUCA

B-F:
5'GGGAACGUAAGCCUCCAACGUUCCCGGAUGCUAAGCCUCCAAGCAUCCUUUGGUGGUGCAGAUGAACUUCAGGGUCA

Complement for A-F (B-F):
5'/5Phos/ACCCUGAAGUUCAUCUGCACCACCG

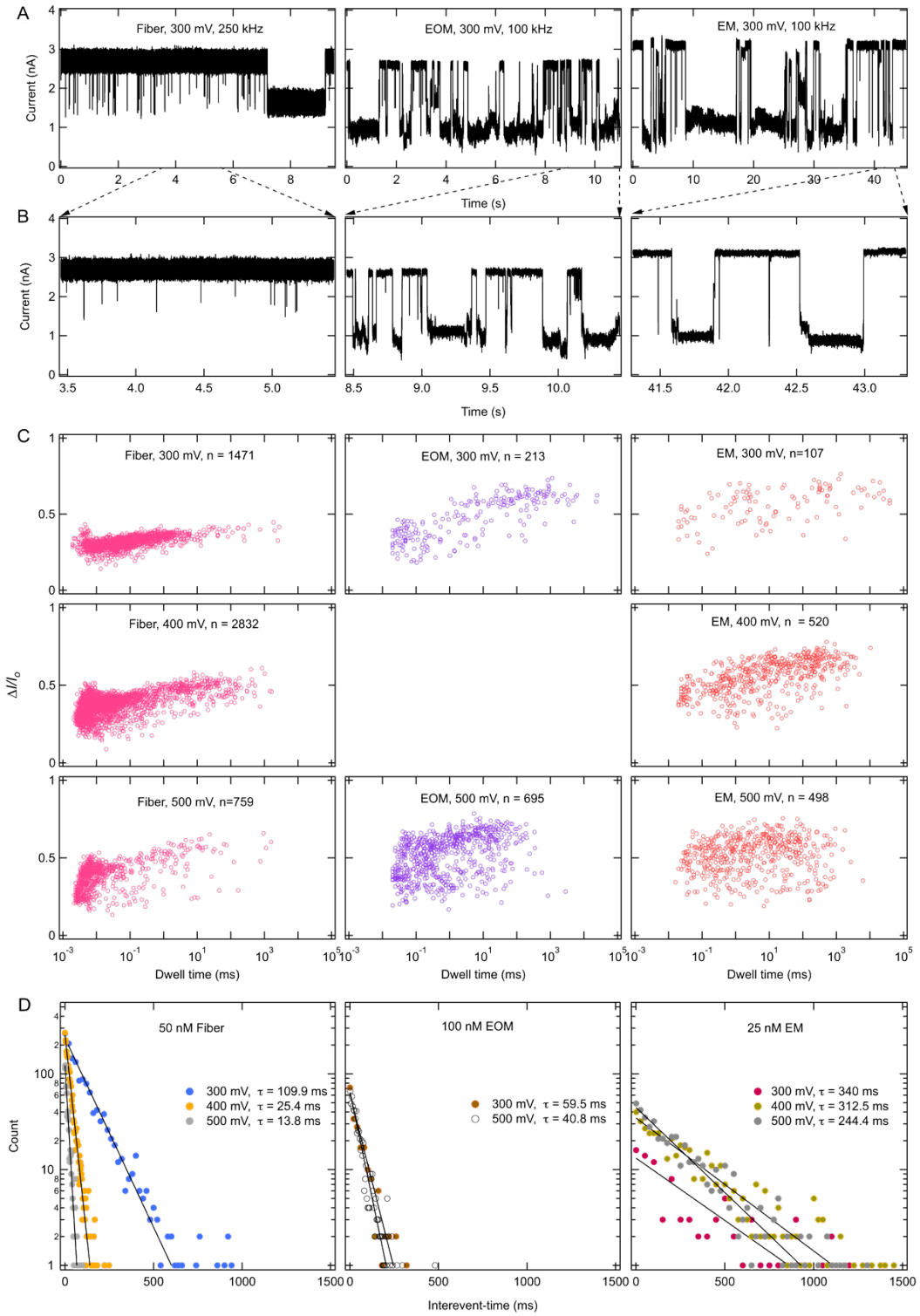

Figure S1i (**A**) Ionic current traces for each fiber at 300 mV, measured with a 4.5 nm pore in 1 M KCl, 10 mM HEPES, 2 mM $MgCl_2$, 7.5 pH. **(B)** Expanded view of ionic current traces for each fiber type in (A) with same time scale. (**C**) Scatter plots for each fiber at different voltages. Number of events collected for conditions are indicated in each panel. (**D**) Inter-event time distributions for each fiber at different voltages.

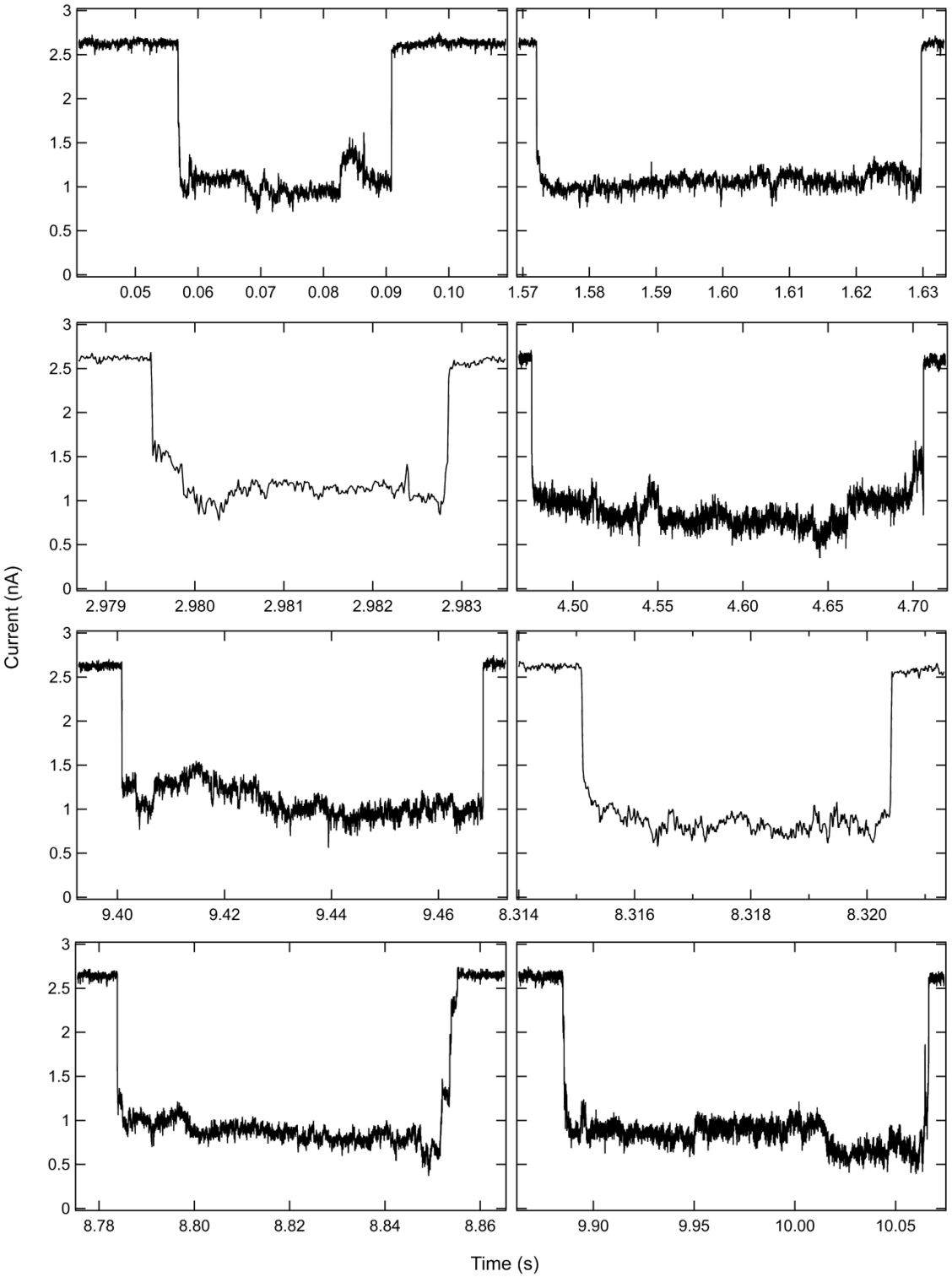

Figure S2ii. Closer look of events for the current traces shown for EOM fiber in Figure S1i at -300 mV. These events suggest that the ultra-fast oscillatory current behavior occurring at the ~ 1 μs time scale in the simulation is difficult to observe in the experiment due to limited bandwidth (100 kHz) of the measurement.

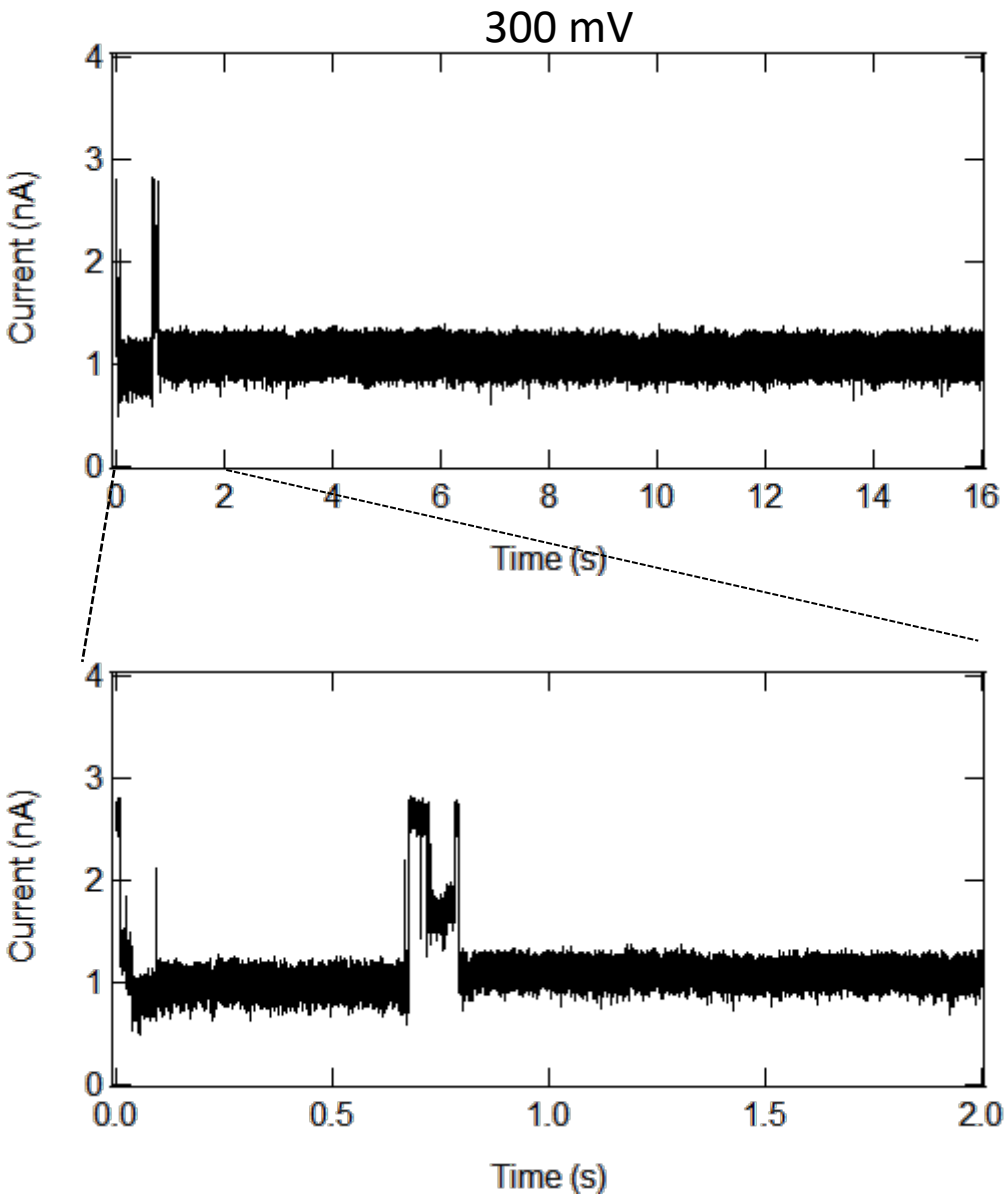

Figure S2 (**A**) Example ionic current traces for a 4.5 nm pore where long and clogging-like events were observed for EOM fibers (and were also observed for EM fibers but data not shown) in 1 M KCl, 10 mM HEPES, 2 mM $MgCl_2$, 7.5 pH measured at 300 mV.

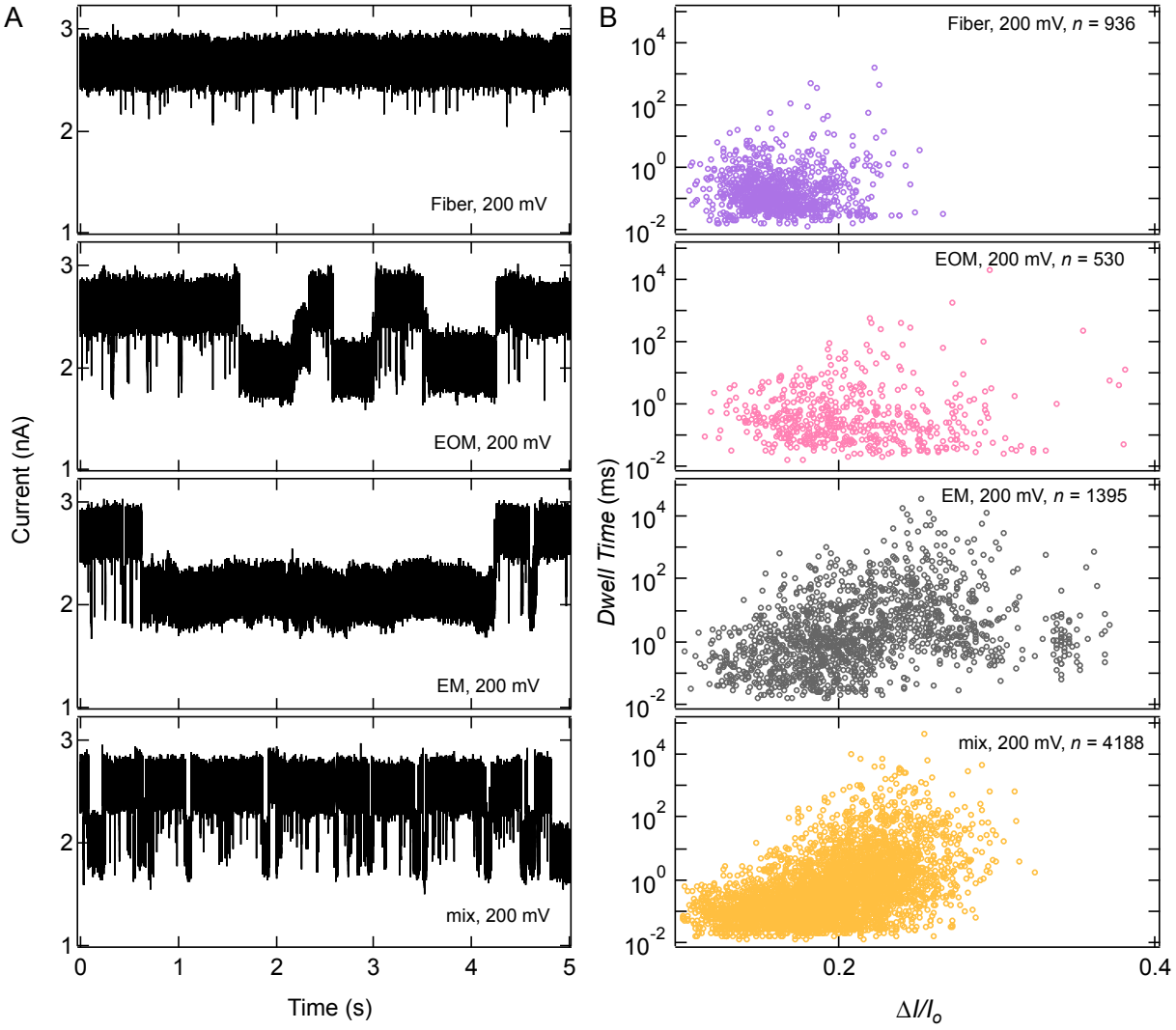

Figure S3 (**A**) Ionic current traces for 50 nM of NF, EOM, and EM fibers, and 1:1:1 mixture of each at 200 mV, measured with a 6 nm pore in 0.4 M KCl, 10 mM HEPES, 2 mM $MgCl_2$, 7.5 pH. (**B**) Scatter plots for Fiber, EOM, EM, and 1:1:1 mixture of each at 200 mV. Number of events collected for conditions are indicated in each panel.

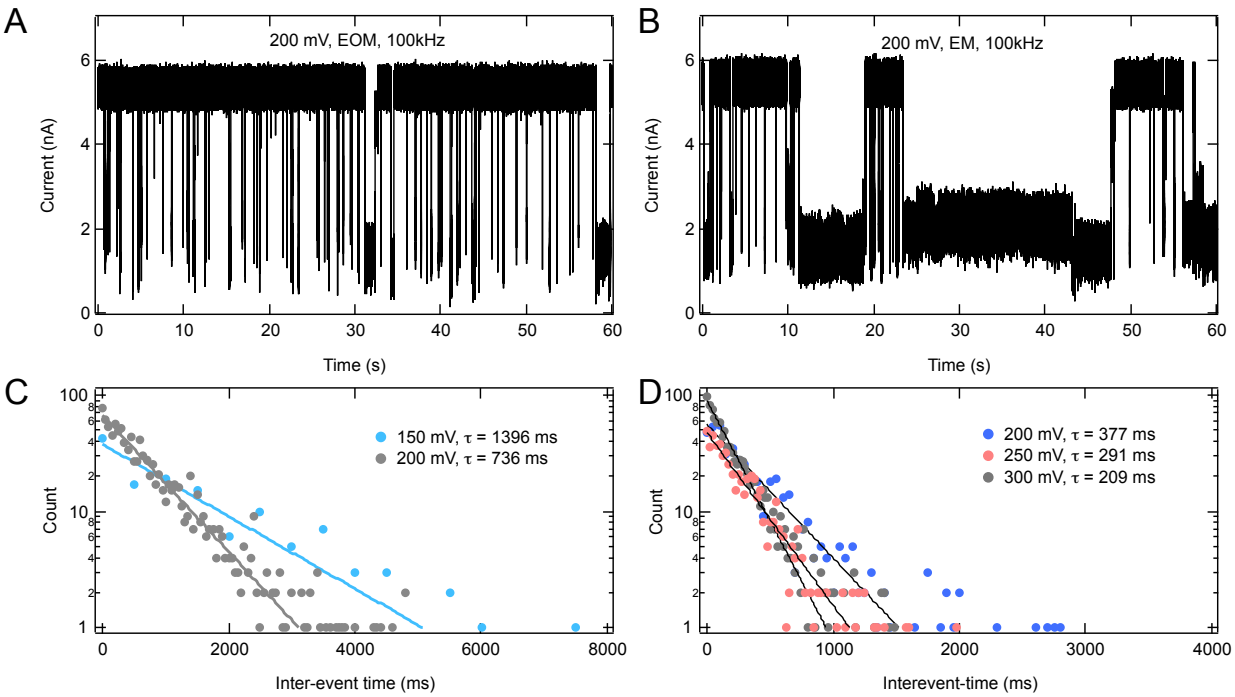

Figure S4 (**A**) Ionic current traces for 10 nM EOM fibers at 200 mV, measured with a 5 nm pore in 2 M KCl, 10 mM HEPES, 2 mM $MgCl_2$, 7.5 pH. Traces were recorded at a sampling rate of 250 kHz and filtered at 100 kHz. (**B**) Ionic current traces for 20 nM EM fibers at 200 mV, measured with a 5 nm pore in 2 M KCl, 10 mM HEPES, 2 mM $MgCl_2$, 7.5 pH. (**C**) Inter-event time distribution for EOM fibers at 150-200 mV. (**D**) Inter-event time distribution for EM fibers at 200-300 mV.

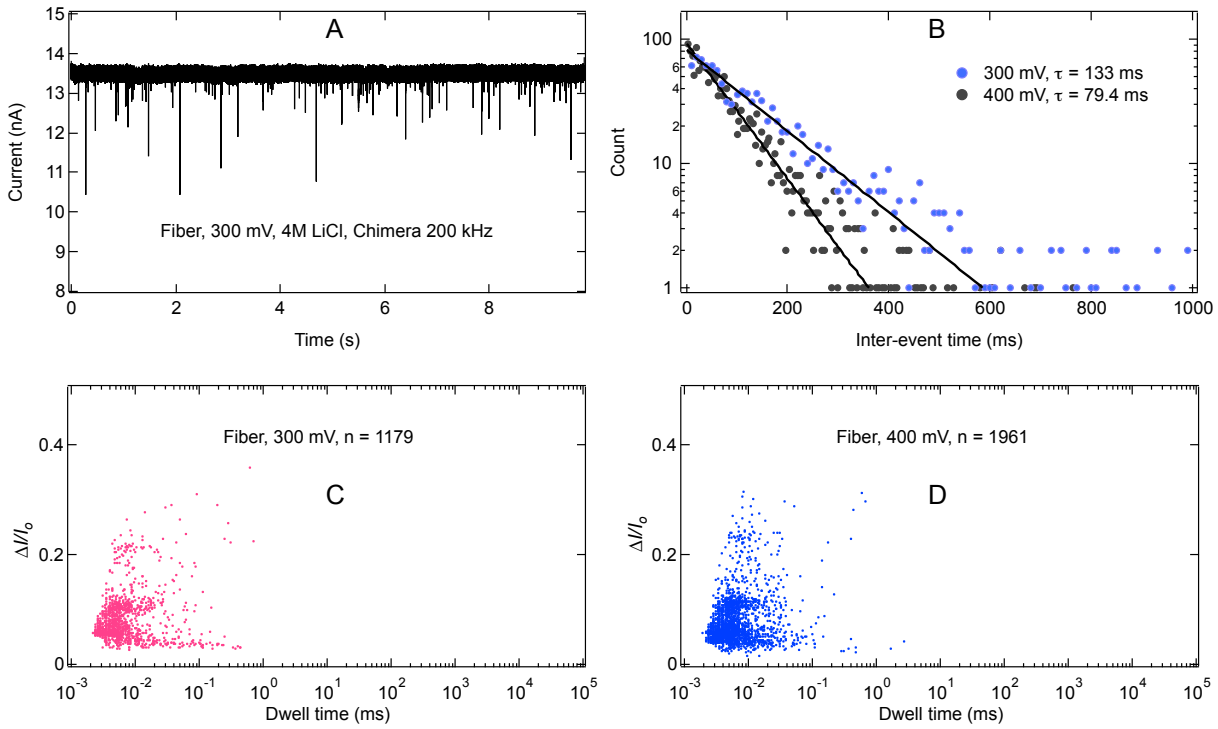

Figure S5 (**A**) Ionic current traces for 10 nM NF fibers at 200 mV, measured with a 6.5 nm pore in 4 M LiCl, 10 mM HEPES, 2 mM $MgCl_2$, 7.5 pH. Traces were recorded using a custom instrument (Chimera VC) at a sampling rate of 4167 kHz and filtered at 200 kHz. (**B**) Inter-event time distribution at 300 mV and 400 mV respectively. Solid curves represent fit with single exponential distribution and time constants are indicated in annotation. (**C**) and **(D)** Scatter plots for NF fibers at 300 mV and 400 mV respectively. Number of events collected (n) in each fiber are indicated.

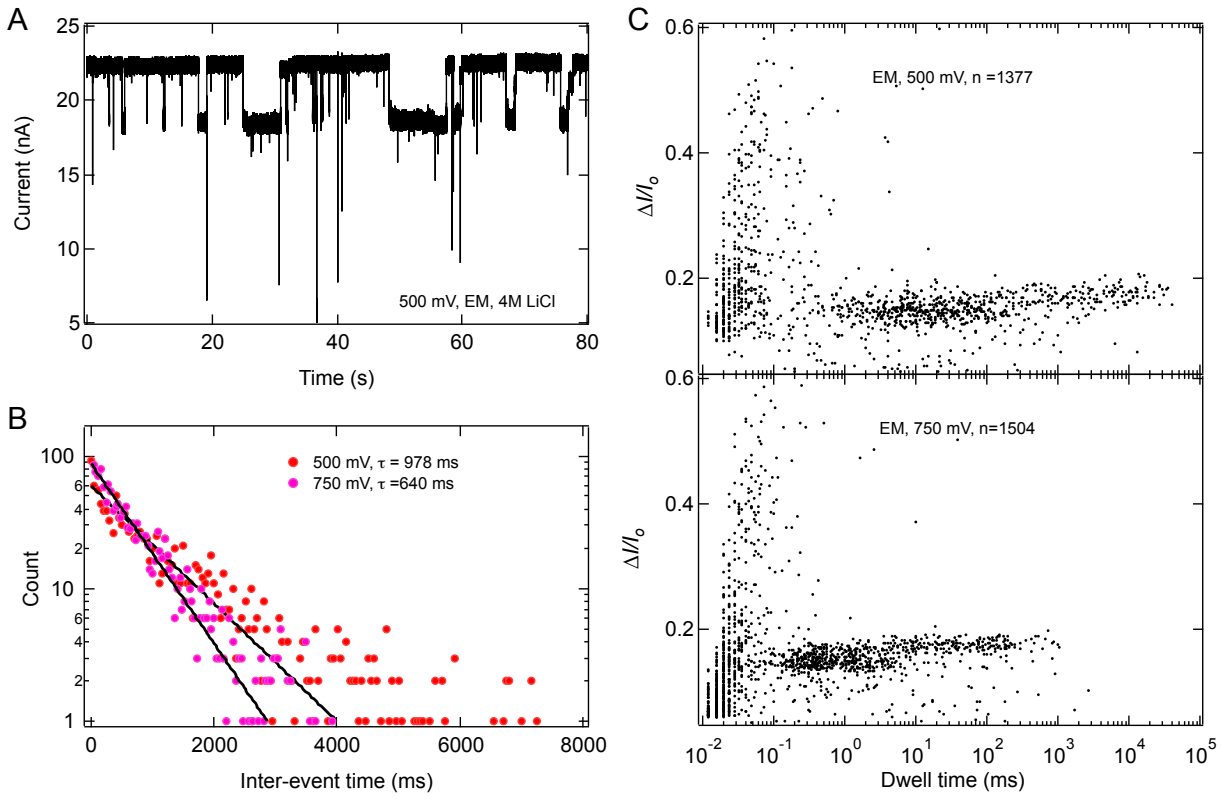

Figure S6 (**A**) Ionic current traces for 10 nM EM fibers at 500 mV, measured with a 6 nm pore in 2 M KCl, 10 mM HEPES, 2 mM $MgCl_2$, 7.5 pH. Traces were recorded at a sampling rate of 250 kHz and filtered at 100 kHz. (**B**) Inter-event time distribution for EM fibers at 500 and 750 mV. (**C**) Scatter plots of EM fibers at 500-750 mV, where number of events are indicated in each panel.

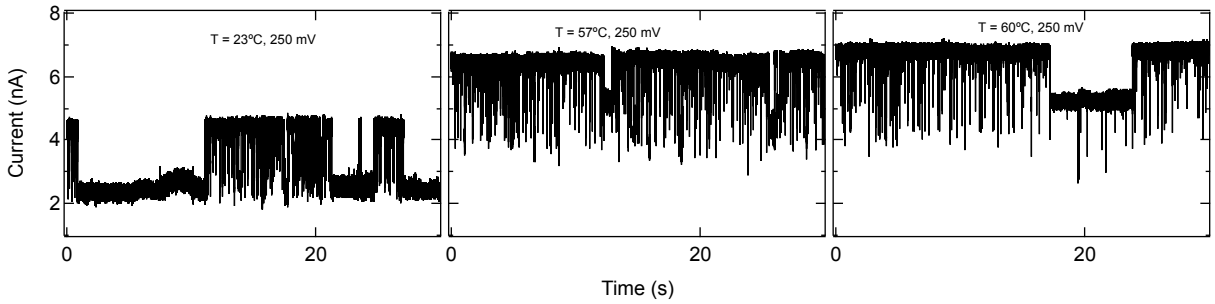

Figure S7 (**A**) Ionic current traces for 100 nM EM fibers at 250 mV at different temperatures, measured with a 5 nm pore in 1 M KCl, 10 mM HEPES, 2 mM $MgCl_2$, 7.5 pH. Traces were recorded at a sampling rate of 250 kHz and filtered at 100 kHz.

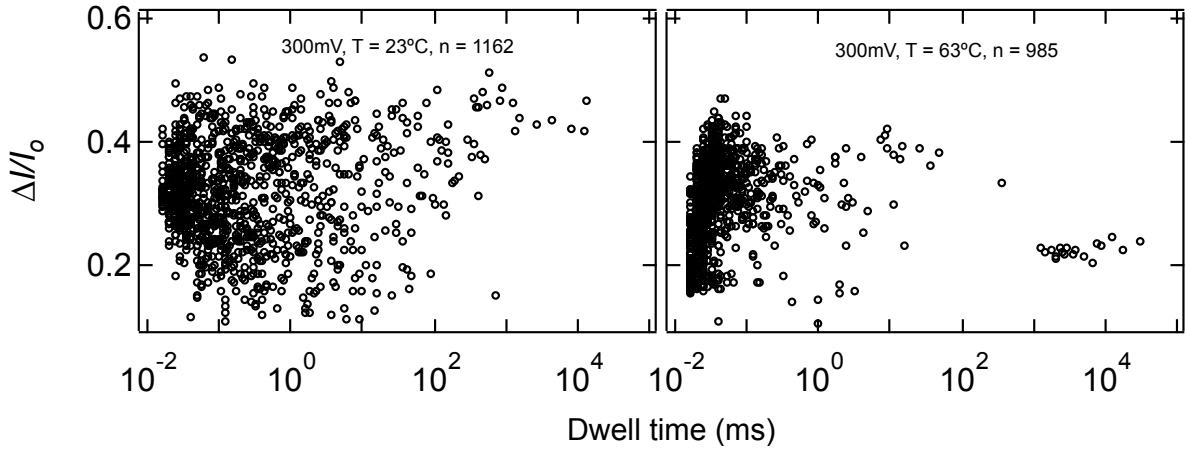

Figure S8 Scatter plots of current blockage ratio and dwell time for 100 nM EM fibers at 300 mV at 23 ℃ and 63℃ respectively, measured with a 5 nm pore in 1 M KCl, 10 mM HEPES, 2 mM $MgCl_2$, 7.5 pH.

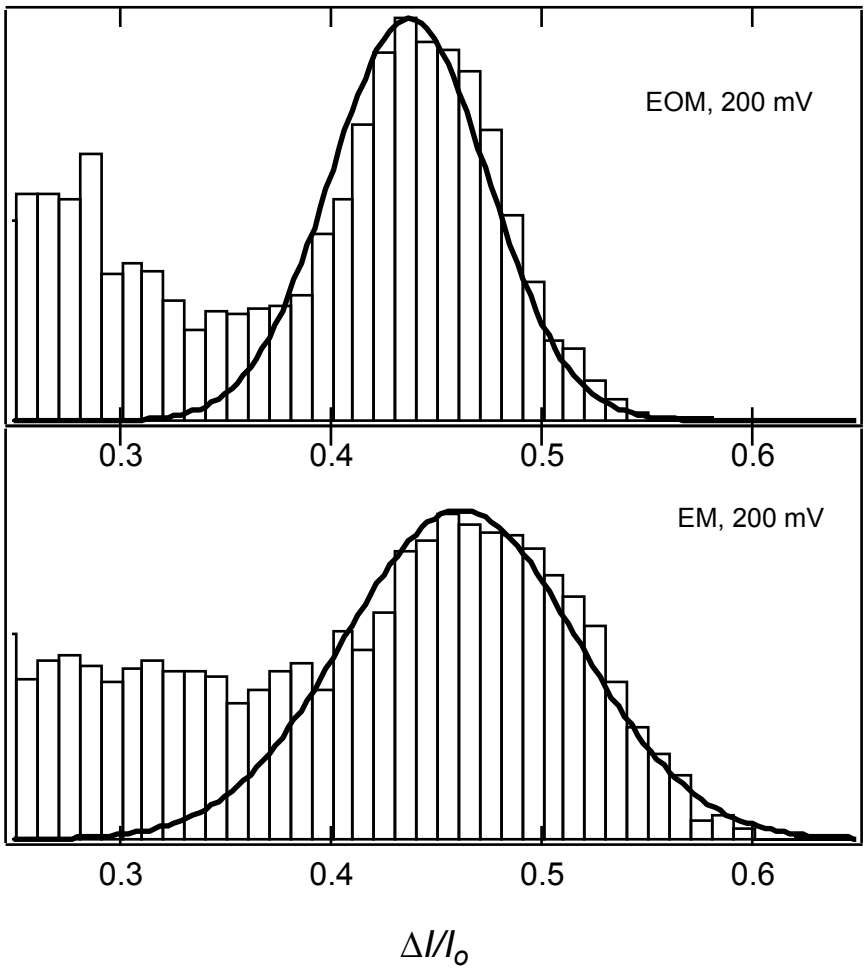

Figure S9. Histograms of current blockade ratio measured with a 6 nm pore at 200 mV for EOM and EM. Solid-curves represents fit with gaussian distribution function with parameters shown below.

Table S1: Fitting parameters in Figure S9 above

Parameters	EOM	EM
Amplitude (A)	354.260	155.100
Mean (x)	0.440	0.462
Width (w)	0.053	0.080

To determine the confidence interval for the differences in mean value of blockade ratio

$$\Delta\frac{\Delta I}{I_O}^{EM-EOM} = x_{EM} - x_{EOM} \pm z \times \sqrt{\frac{\left((A_{EM}-1)\times {w_{EM}}^2\right)+\left((A_{EOM}-1)\times {w_{EOM}}^2\right)}{(A_{EM}+A_{EOM}-2)}} \times \sqrt{\frac{1}{(A_{EM}+A_{EOM}-2)}}$$

Where, z is score of a confidence interval, for 99% confidence its value is 2.58. Putting z-score and using parameters in tableS1. The differences in mean blockade ratio between EM and EOM at 99% confidence would be

$$\Delta\frac{\Delta I}{I_O}^{EM-EOM} = 0.022 \pm 0.007$$

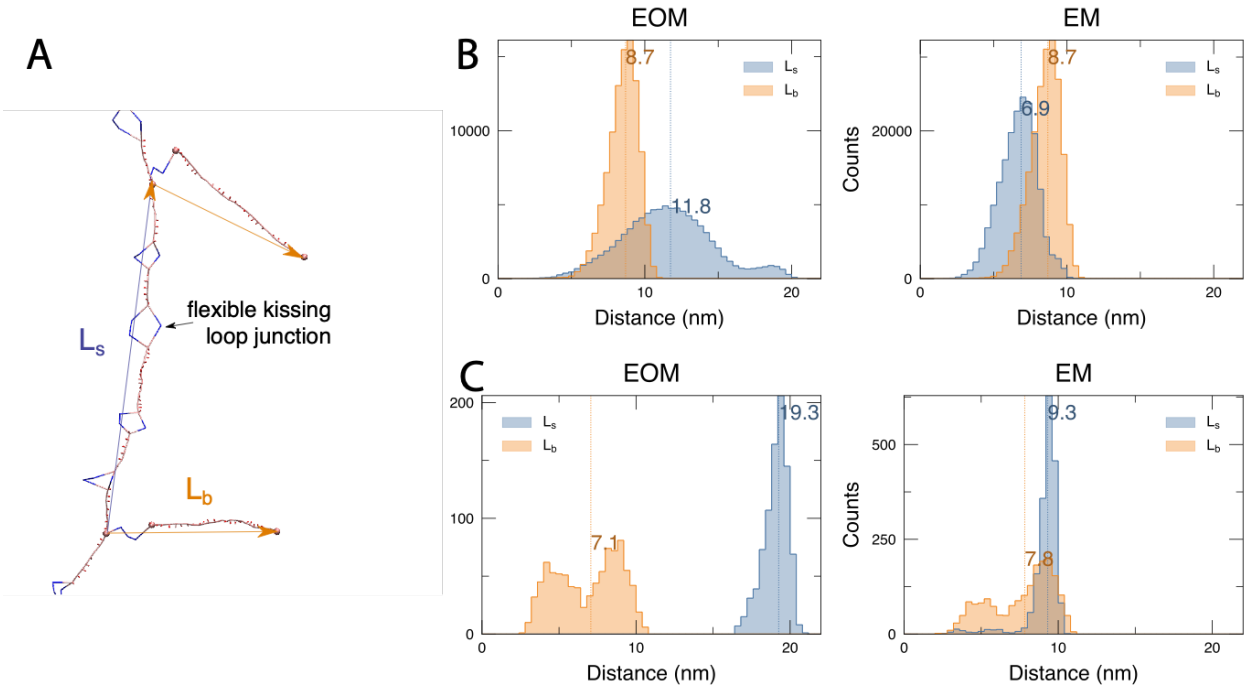

Figure S10. Analysis of fiber branch spacing and length in 3-D model of the computer simulation. (A) Definition of branch spacing (L_s) and length (L_b). (B) Distributions of L_s and L_b measured in bulk solution. (C) Distributions measured in the pore during the translocation. Distribution suggest that the oscillations observed in the simulation is caused by substantially increases of L_s in the pore because the helical axis across kissing loop junctions (these include flexible unstructured ssRNA region, see blue beads in A) adopts a 180 degree angle when the RNA is in or a bit above the pore.

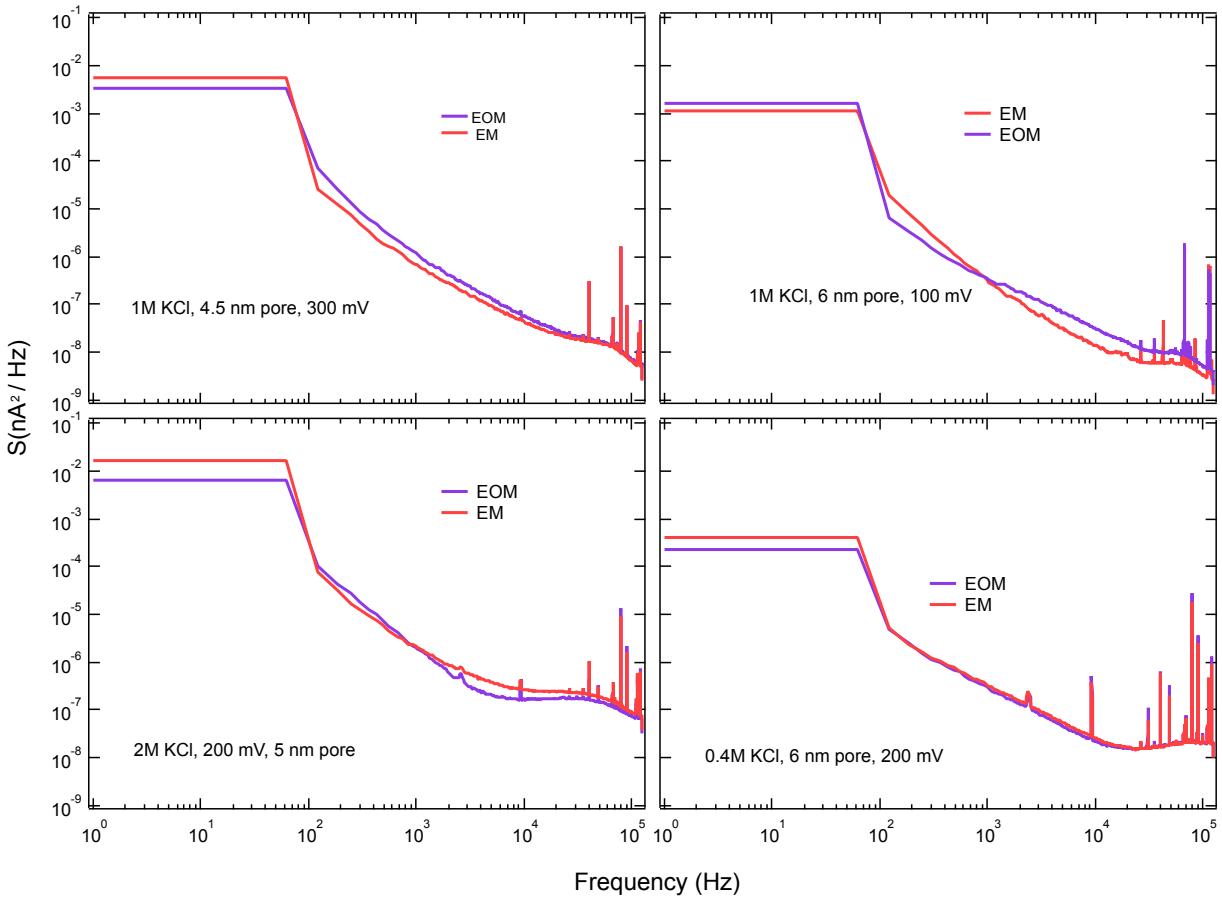

Figure S11. Power spectrum of ionic current traces in different salt environment and for different pore diameters.

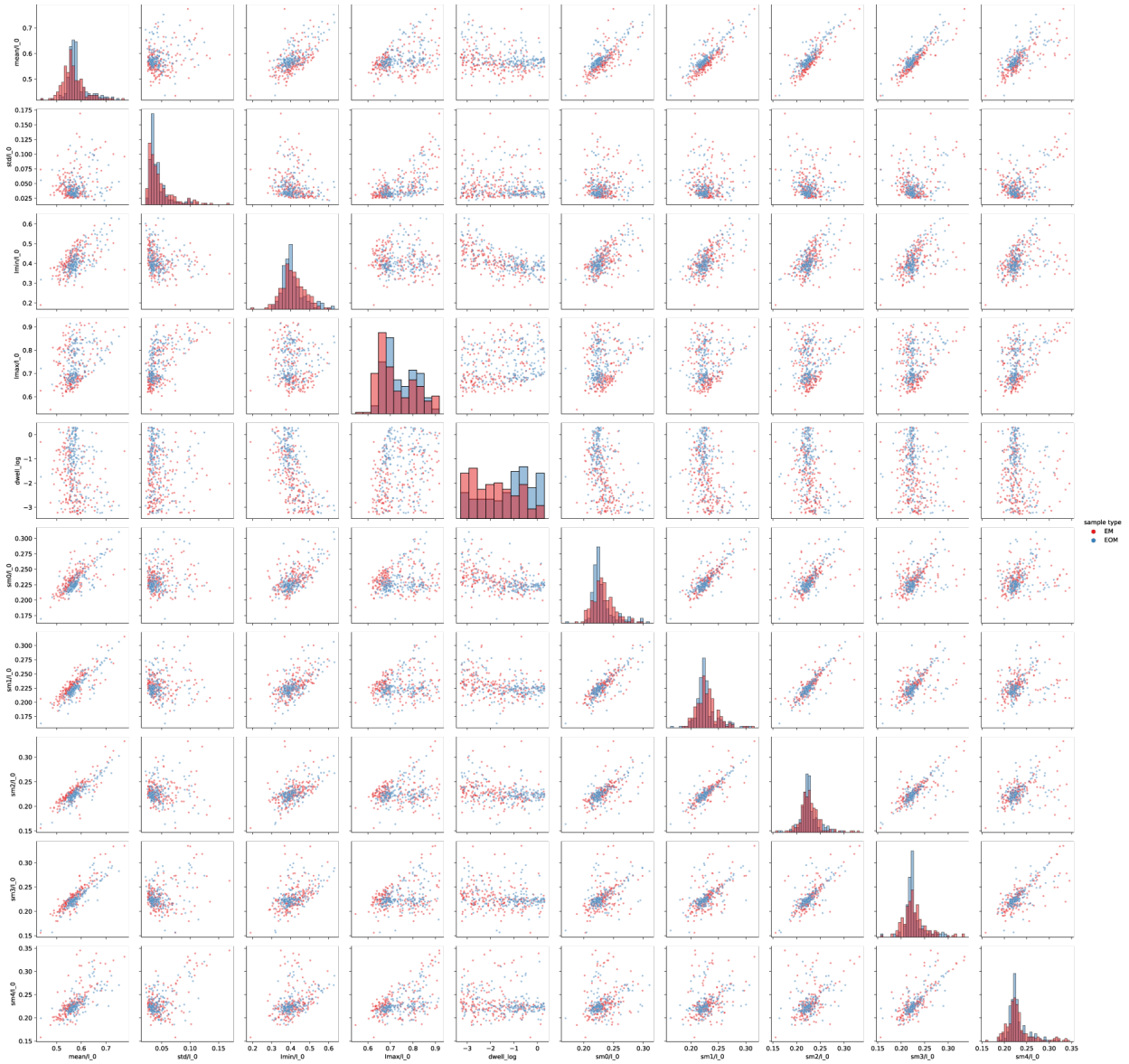

Figure S12: Comparison of 10 features extracted for SVM analysis. The diagonals are histogram distributions of each feature, while the off-diagonal scatterplots are pairwise comparisons (see labeled axes). The dataset in this figure is obtained from 100 mV recordings of EM and EOM samples using a 6nm pore, and the criteria for selection of translocation events was (1) 500 μs<event duration<2s, and (2) normalized minimum current $I_{min}/I_0 < 0.63$. 187 events from each data set were selected (limited by EOM event count) and all were plotted in this graph. Note that the separation of any individual feature is insignificant. However, pairwise comparisons begin to show higher degrees of separation. This graph demonstrates the presence of fundamental signal variability among the two samples, while emphasizing the need for multi-variate analysis such as the SVM classification conducted in this work.

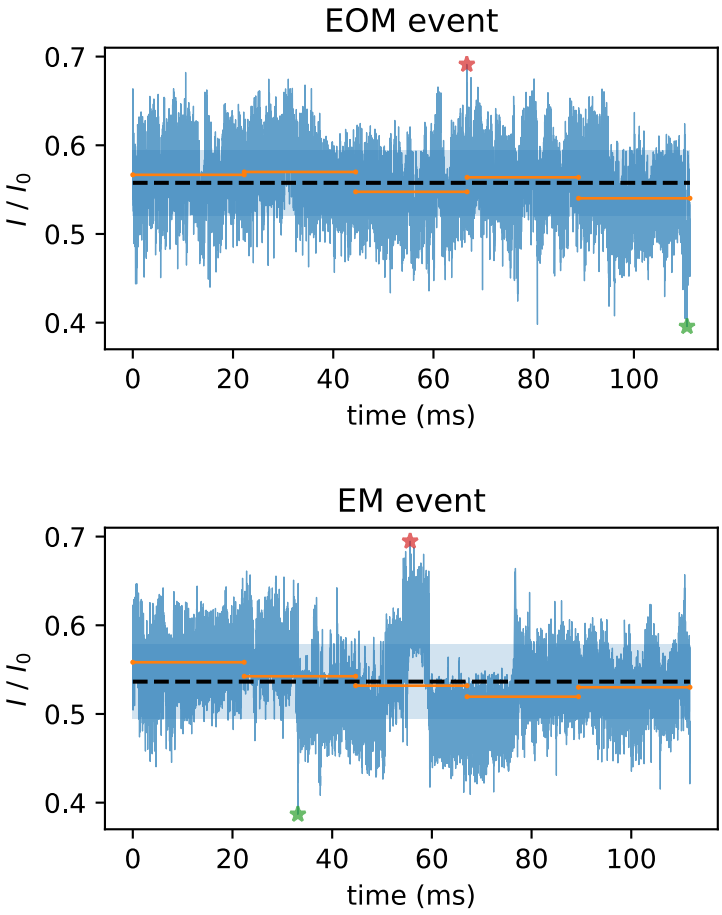

Figure S13: Example normalized trace of events for EM and EOM measured with a 6 nm pore at 100 mV. Each event was divided into five equal-duration segments.

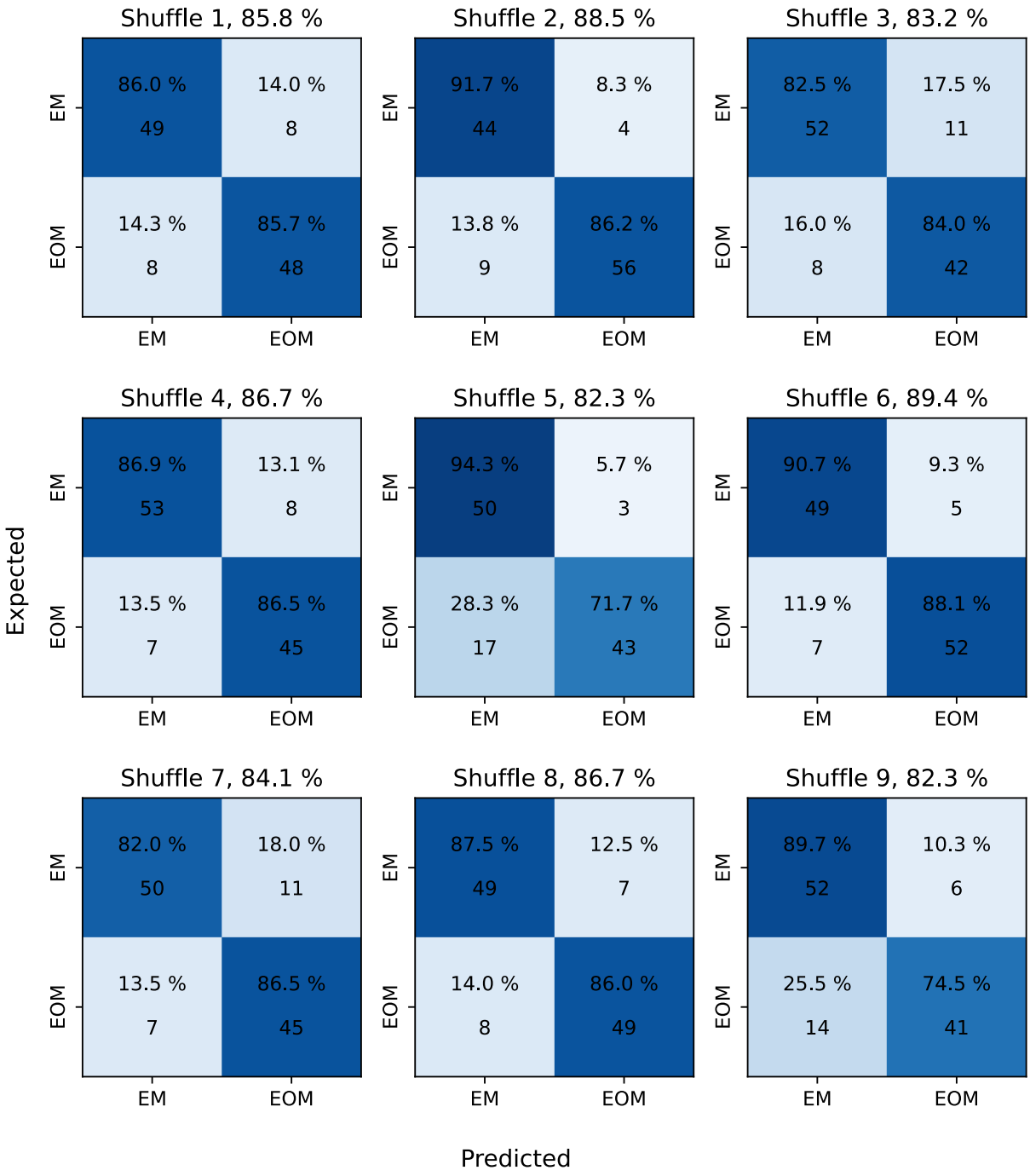

Figure S14: Support vector machine (SVM) testing results with 9 unique random samples of the selected events into train/test groups. The train/test ratio was 70/30%. This repeat was performed to ensure that accuracy of one particular shuffle is not biased due to a given lucky sampling. The percentages (and color scheme) show confusion matrix values normalized with respect to the total samples in the row (True labels). The number below the percentage represent snumber of identified samples in each class.

References:

1. Rackley, L.; Stewart, J. M.; Salotti, J.; Krokhotin, A.; Shah, A.; Halman, J. R.; Juneja, R.; Smollett, J.; Lee, L.; Roark, K.; Viard, M.; Tarannum, M.; Vivero-Escoto, J.; Johnson, P. F.; Dobrovolskaia, M. A.; Dokholyan, N. V.; Franco, E.; Afonin, K. A., RNA Fibers as Optimized Nanoscaffolds for siRNA Coordination and Reduced Immunological Recognition. *Advanced Functional Materials* **2018,** *28* (48), 1805959.
2. Grabow, W. W.; Zakrevsky, P.; Afonin, K. A.; Chworos, A.; Shapiro, B. A.; Jaeger, L., Self-Assembling RNA Nanorings Based on RNAI/II Inverse Kissing Complexes. *Nano Letters* **2011,** *11* (2), 878-887.