MOSAIC: A Modular Single-Molecule Analysis Interface for Decoding Multistate Nanopore Data

Jacob H. Forstater, § Kyle Briggs, § Joseph W. F. Robertson, † Jessica Ettedgui, † Olivier Marie-Rose, † Canute Vaz, John J. Kasianowicz, ‡ Vincent Tabard-Cossa, § and Arvind Balijepalli * †

†Physical Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, United States
‡Department of Chemical Engineering, Columbia University, New York, New York 10027, United States
§Department of Physics, University of Ottawa, Ottawa, Ontario K1N 6N5, Canada
||Information Technology Laboratory, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, United States

ABSTRACT: Biological and solid-state nanometer-scale pores are the basis for numerous emerging analytical technologies for use in precision medicine. We developed Modular Single-Molecule Analysis Interface (MOSAIC), an open source analysis software that improves the accuracy and throughput of nanopore-based measurements. Two key algorithms are implemented: ADEPT, which uses a physical model of the nanopore system to characterize short-lived events that do not reach their steady-state current, and CUSUM+, a version of the cumulative sum statistical method optimized for longer events that do. We show that ADEPT detects previously unreported conductance states that occur as double-stranded DNA translocates through a 2.4 nm solid-state nanopore and reveals new interactions between short single-stranded DNA and the vestibule of a biological pore. These findings demonstrate the utility of MOSAIC and the ADEPT algorithm, and offer a new tool that can improve the analysis of nanopore-based measurements.

Protein and solid-state nanopores (Figure 1A) are the basis for single-molecule measurements of a variety of analytes including ions, 1−10 single-stranded RNA and DNA, 11−13 double-stranded DNA, 14−19 proteins, 20−23 synthetic polymers, 24−26 and metallic nanoparticles. 24−26 The method is conceptually simple. An electric potential applied across a nanopore that spans two electrically isolated chambers (filled with electrolyte solutions) results in an ionic current with a mean value (⟨i⟩) (Figure 1B). Single molecules that reversibly partition into the pore cause a series of pulses or current blockades (Figure 1B). The change in pore conductance is caused by the volume exclusion of mobile ions from the pore, and interactions between the ions and the analyte. 5,21,26 The change in conductance, 5,20,21,26,27 and the residence times of analytes in the pore, 8,20,21 are used to estimate the analyte size, 20,21 effective charge, 21, and dipole moment. 28

Analyte-induced events appear as single or multiple conductance state levels, arising from changes in the analyte conformation or interactions in the pore. 9,11,20,21,27,29,30 (Figure 1C). Multiple conductance level events differ from the gating of ion channels, where the channel fluctuates between two states, open and closed. 31 These fluctuations are well characterized using hidden Markov models 32−34 and kinetic simulations. 34,35 On the other hand, several analysis techniques have been applied to analyze nanopore-based single-molecule data including threshold detection, 36,37 slope- or area-based techniques, 38 the cumulative sum (CUSUM) algorithm, 39 charge conservation, 40 and probabilistic machine-learning techniques. 36 While these approaches are effective when the residence times of analytes in the nanopore are long (compared to the characteristic time constant of the system), they are not useful for characterizing short-lived events. To more accurately characterize short events, we developed a technique that models the ionic current response with an equivalent electrical circuit. 36,41 This algorithm, when applied to the interaction of a polydisperse mixture of a synthetic polymer with the Staphylococcus aureus α-hemolysin (αHL) nanopore, recovered 18-fold more events per unit time at high measurement bandwidth (B = 100 kHz), reduced the constraints on data acquisition by permitting polymers to be separated at lower bandwidth (B = 10 kHz), and improved the resolving power in the low mass regime (to polymers with molecular weight ≈ 370 g/mol).

Received: September 20, 2016
Accepted: October 31, 2016
Published: October 31, 2016
Here, we describe Modular Single-Molecule Analysis Interface (MOSAIC), an improved data analysis tool for analyzing nanopore data. We implemented two algorithms in MOSAIC: ADEPT, the equivalent electrical circuit model described above,\(^{36,41}\) and an improved version of CUSUM that is suitable for analyzing events with relatively long residence times in the pore. The software is extensible, and allows many commonly used data formats, signal conditioning, and data processing algorithms to be seamlessly integrated. Below, we demonstrate the features and utility of MOSAIC when applied to data measured with both biological and solid-state nanopores.

## MATERIALS AND METHODS

### Solid-State Nanopore Measurements.
Solid-state nanopore data were reanalyzed from Briggs et al.\(^{42}\) Briefly, nanopores were fabricated in 50 \(\mu\)m \(\times\) 50 \(\mu\)m, 10 nm thick low-stress silicon nitride (Si\(_3\)N\(_4\)) TEM windows (Norcada, Canada) via the controlled breakdown (CBD) method\(^{47}\) in a 1 M NaCl buffer (pH 10, 10 mM NaHCO\(_3\)). Nanopore measurements of double-stranded DNA (dsDNA) were performed in 3.6 M LiCl, 10 mM HEPES (pH 8) using highly purified 50 base pair (bp) dsDNA fragments (NoLimits no. SM1421, Life Technologies). Data were low-pass filtered at 100 kHz with a hardware 4-pole Bessel filter (Axopatch 200B) and digitized by a National Instruments PCI-6120 DAQ card with a sampling rate \(\approx 500\) kHz.

### Biological Nanopore Measurements.
Nanopore measurements were performed using quartz capillaries with a \(\approx 1 \mu\)m diameter aperture on one end\(^{26,43,44}\) within a custom polycarbonate test cell with \(\approx 200 \mu\)L volume (Electronic Biosciences, San Diego, CA). Analytes were dissolved in the working buffer and added directly into the capillary or to the external test cell.

For single-stranded DNA measurements, the quartz capillary was filled with a 10–20 \(\mu\)M solution of different length homopolymeric adenosine d\(_A\)\(_{20}\), d\(_A\)\(_{30}\), or d\(_A\)\(_{100}\) (Integrated DNA Technologies, Coralville, IA) dissolved in 1 M NaCl, 1X TE buffer (10 mM Tris, 1 mM EDTA in DNase-free water, titrated to pH 7.2 with 3 M HCl).

For poly(ethylene glycol) measurements (PEG), data from two previous studies that span a wide range of polymer sizes were combined.\(^{22,26}\) In both cases, the capillary was filled with a solution containing a combination of polydisperse PEG (Fluka, Switzerland) and a highly purified calibration standard (Polypure, Oslo, Norway), dissolved in 4 M KCl (Sigma-Aldrich, St. Louis, MO), buffered with 10 mM Tris (Schwarz/Mann Biotech, Cleveland, OH) and titrated to pH 7.2 with 3 M citric acid. The two different solutions were as follows: (a) 20 \(\mu\)M PEG-600 (MW\(_{avg}\) = 600 g/mol), 40 \(\mu\)M PEG-400 (MW\(_{avg}\) = 400 g/mol) and 2 \(\mu\)M purified PEG-502 (MW\(_w\) = 502 g/mol) or (b) 30 \(\mu\)M PEG-1000 (MW\(_{avg}\) = 1000 g/mol), 30 \(\mu\)M PEG-1500 (MW\(_{avg}\) = 1500 g/mol), and 1 \(\mu\)M purified PEG-1251 (MW\(_w\) = 1251 g/mol).

Planar lipid bilayers were formed across the quartz capillary aperture using a 10 mg/mL solution of 1,2 diphytanoyl-sn-glycero-3-phosphatidylcholine (DPhyPC; Avanti Polar Lipids, Alabaster, AL) in n-decane (Sigma-Aldrich).\(^{36}\) Subsequently, wild-type S. aureus \(\alpha\)-Hemolysin (\(\alpha\)HL) was introduced to the test cell by adding a solution containing either \approx 250 ng of monomeric \(\alpha\)HL (List Biological Laboratories, Campbell, CA) or \approx 2.5 ng of purified preformed heptamers. To facilitate channel incorporation, the bilayer was thinned and enlarged by applying a transmembrane potential of \approx 300 mV and a static back pressure within the capillary. Following the insertion of a single channel, the static pressure was reduced and the voltage decreased to the value used for the measurement to prevent further channel incorporation.

The potential was applied across the membrane by a pair of Ag/AgCl electrodes. Immediately prior to use, the electrode placed within in the capillary was prepared by abrading an Ag wire (Alfa Aesar) with 600 grit sandpaper and soaking it in bleach for \approx 10 min. The external electrode in the test cell bath was a 2 mm Ag/AgCl disk electrode (E202, In Vivo Metric). Data were acquired with a custom high-impedance amplifier system (Electronic BioSciences, San Diego, CA) and conditioned with a low-pass antialiasing filter. The analog signal was digitized by a National Instruments PCI-6120 DAQ card with a sampling rate \(F_s\) of 1 MHz, further conditioned using a software-based 8-pole low pass Bessel filter with a cutoff frequency of 100 kHz and resampled at 500 kHz.

### Data Processing and Analysis.
Nanopore data were processed using a Python based program (MOSAIC) developed in-house. The software implements the ADEPT and CUSUM\(^+\) algorithms, which are described below. The compiled program and source code are freely available at https://pages.nist.gov/mosaic/. MOSAIC consists of a modular data processing pipeline which allows users to analyze ionic current data from single-molecule nanopore experiments. The software is designed using object-oriented principles, which ensures that modules remain interoperable. This also makes it straightforward to implement new features into the software, such as alternative analysis algorithms or custom data formats.

---

**Figure 1.** (A) Schematic illustration of DNA translocation through a solid-state nanopore. An electric potential applied across the pore produces an ionic current. (B) The partitioning of DNA into the pore causes well-defined current reductions with different mean current blockade amplitudes (e.g., \(i_o\) and \(i_b\)). (C) A single level event is characterized by the ratio of the mean currents for the occupied and fully open pore \((i_o/i_b)\) and the level residence time \((\Delta t)\). (D) A histogram illustrating the relative current blockade depth for two species obtained by analyzing the events. (E) Residence time distributions for the two blockade depth populations. The mean residence times are estimated from fits to the distributions.
In many cases, users can interact with MOSAIC using a front end graphical user interface.

**Algorithms.** MOSAIC consists of a pipeline with five modules: (i) load data, (ii) optional signal conditioning and filtering, (iii) event detection, (iv) event analysis, and (v) results storage. A detailed description of the software architecture is presented in the Supporting Information. In this section, we discuss two algorithms implemented in MOSAIC: (i) CUSUM+, an improved version of the CUSUM algorithm that provides robust statistical analysis of events which converge to a steady-state, and (ii) ADEPT, an implementation of a previously developed theory that uses a physical model of the nanopore system to accurately characterize very short events that do not approach a steady-state ionic current.

**Cumulative Sum Analysis (CUSUM+).** Cumulative Sum (CUSUM) is a commonly used method to detect step-like changes in time-series data but was only recently implemented in nanopore analysis. It assumes that the interaction of an analyte with the pore causes a series of instantaneous changes in the ionic current from its baseline value (defined as states and well-approximated by step functions), and the ionic current noise follows a known distribution (e.g., Gaussian). A statistical test identifies when the current level changes. The instantaneous log-likelihood ratios of sequential data points for both positive and negative step changes are calculated. The positive values of these ratios are independently summated and a negative log-likelihood resets the sum to zero. These form a two-sided decision function, which detects level changes that correspond to either an increase or decrease in the current level. A new state is identified when one of the decision functions exceeds a threshold determined automatically by the software. The locations of state changes are determined from the minima of related functions, and the mean ionic current between sequential states is calculated and used to determine the local blockade depth, defined as ratio of the ionic current when the pore is occupied to that of the open pore \( (i(i)/i_0) \) (Figure 1C).

We implemented an improved version of the CUSUM algorithm in MOSAIC (CUSUM+), which is less sensitive to artifacts that can be falsely identified as a state change. This is achieved by specifying a minimum time between successive triggers (to exclude transients from the state change detection and blockade depth calculations) and by requiring that identified state levels differ by a minimum value (corresponding to a physically significant change). The efficiency is improved by eliminating or reducing redundant computations (e.g., maintaining running calculations of the mean and variance).

**Adaptive Time-Series Analysis (ADEPT).** For very long events (>5τ, Figure 2 left), the blockade depth is easily estimated (e.g., with CUSUM+). However, that process fails for short-lived events (<5τ, Figure 2 right) that do not reach a steady-state mean value. In this case, the blockade depths are estimated by fitting the data to an electrical circuit model of the nanopore (implemented as ADEPT in MOSAIC). The algorithm assumes a molecule partitioning into the nanopore instantaneously increases the nanopore resistance \( R_{\text{n}} \) by \( \Delta R \). However, the system capacitance causes the ionic current to change over a finite time. For a constant applied voltage, \( V_0 \), the predicted ionic current is \( i(t) = i_0 - \beta (1 - e^{-\frac{t}{\tau}}) \), where \( i_0 = \frac{V_0}{R_{\text{n}} + R_{\text{p}}} \), \( \beta = \frac{V_0 \Delta R}{(R_{\text{n}} + R_{\text{p}} + \Delta R)(R_{\text{n}} + R_{\text{p}})} \), and \( \tau = \frac{C R_{\text{n}} (R_{\text{n}} + \Delta R)}{R_{\text{n}} + R_{\text{p}} + \Delta R} \). The difference between the time constants \( (\tau) \) leading up to and following an event are much shorter than typical sampling rates (see the Supporting Information for the calculation). We therefore use a single fit parameter for \( \tau \), which reduces the degrees of freedom. There is an option to override this constraint.

**RESULTS AND DISCUSSION**

**Analysis of Short dsDNA Fragments Measured with SiN\textsubscript{x} Nanopores.** We compared the results of CUSUM+ and ADEPT on measurements of 50 base pair (bp) double-stranded DNA (dsDNA) translocating through a \( \approx 2.4 \) nm diameter SiN\textsubscript{x} nanopore (\( \approx 2800 \) events). At an applied potential of 400 mV, the mean residence time of dsDNA in the pore is \( \approx 440 \) μs, more than an order of magnitude longer than the characteristic time constant of the system \( (\tau = 10 \) μs, \( B = 100 \) kHz). Both algorithms produce two distinct peaks in the blockade depth histogram \( (i(i)/i_0) \approx 0.070 \pm 0.001 \) and 0.488 ± 0.004). Peak positions were obtained using an error-weighted Gaussian fit and are reported with an expanded uncertainty, \( k = 2 \) (see Supporting Information for a full listing of the analysis and fit parameters). The leftmost peak corresponds to DNA translocation, whereas the rightmost peak is likely due to the helical structure of dsDNA unwinding to transition from the B-form to the S-form dsDNA, where the chain elongates by 1.7 fold because of the strong electric field gradient across the pore.

At 800 mV, the blockade depth histogram produced by ADEPT has two overlapping peaks (Figure 3B) consisting of a narrow peak \( (i(i)/i_0) = 0.080 \pm 0.002 \), the expected location for B-form dsDNA, and a broader peak \( (i(i)/i_0) = 0.138 \pm 0.002 \). The latter is comprised of short, single-level events, which likely result from transient interactions between the dsDNA and the access region outside the pore. This was not accurately identified in our previous analysis (see below). A third, low amplitude, broad peak is visible at \( (i(i)/i_0) = 0.226 \pm 0.002 \), Figure 2. ADEPT and CUSUM+ analysis applied to a simulated nanopore measurement (gray). Two events with the same current blockade (red) but different residence times \( (t_{res}) \), with respect to the system characteristic relaxation time \( (\tau) \) are shown. (Left) For a long event \( (t_{res} \geq 5\tau) \), the ionic current converges close to its steady-state value, and the current levels estimated by ADEPT and CUSUM+ are equivalent. (Right) For short events \( (t_{res} \approx 2\tau) \), the current does not reach the steady-state value of the idealized pulse (red). In this case, CUSUM+ and other algorithms used in nanopore analysis systematically underestimate the steady-state current (blue) by an amount \( \Delta i \) (gray; dashed). In contrast, the physical model underlying ADEPT allows the algorithm to accurately estimate an event’s steady-state current.
0.018. It is probable that this peak is associated with these access-region interactions. Alternatively, it is possible that these events could represent partial unwinding of the dsDNA secondary structure at forces below the $B^{-}S$ stretching transition threshold.

As seen in Figure 3B, CUSUM+ also detects the first peak, $⟨i⟩/⟨i_0⟩ = 0.090 \pm 0.002$ (albeit slightly shifted compared to the ADEPT value). However, it only characterizes $\approx 20\%$ of the events in the second peak ($⟨i⟩/⟨i_0⟩ = 0.134 \pm 0.010$) where the mean residence time of the events ($⟨t_{\text{res}}⟩ = 47 \pm 4 \mu s$) is less than $5\tau$. In addition, CUSUM+ misses most of the events from the third peak detected with ADEPT. While CUSUM+ could be allowed to characterize events with lifetimes less than $5\tau$, it will underestimate the blockade depth ratios (see the Supporting Information, Figure S3). Both algorithms identify the fourth peak ($⟨i⟩/⟨i_0⟩ \approx 0.56$), which arises from the stretching transition of dsDNA noted above ($S$-form of dsDNA).

CUSUM+ recovers more events here than ADEPT, which utilizes a fitting routine that may not converge for some very long events ($t_{\text{res}} > 25$ ms; 50,000 points; $F_s = 500$ kHz). Clearly, the results would be improved if ADEPT is used for relatively short-lived events and CUSUM+ is used on events with residence times $>5\tau$.

We use ADEPT to determine the blockade depth ratio histograms for three different length single-stranded DNA (ssDNA) homopolymers ($dA_{100}$, $dA_{40}$, and $dA_{20}$) entering an $\alpha$HL nanopore from the cis side. The applied potential is $V = 140$ mV and the polynucleotides are added to the cis side of the pore.

Figure 3. Blockade depth histograms for a 50 base pair double-stranded DNA measured with a 2.4 nm diameter SiN$_x$ nanopore. (A) At 400 mV, both CUSUM+ (gray), and ADEPT (red) are in excellent agreement. (B) At 800 mV, the mean residence time decreases to $\approx 36 \mu s$ (estimated from ADEPT analysis). The analyses by ADEPT and CUSUM+ are markedly different.

Figure 4. Normalized blockade depth histograms for (A) $dA_{100}$, (B) $dA_{40}$, and (C) $dA_{20}$ single-stranded DNA interacting with the $\alpha$HL nanopore estimated using the ADEPT algorithm. The applied potential is $V = 140$ mV and the polynucleotides are added to the cis side of the pore.

(Supporting Information, Figure S4). This functionality will be implemented in a future version of MOSAIC.

While both CUSUM+ and ADEPT produce comparable results for events with residence times $>5\tau$, CUSUM+'s statistical approach is on average $\approx 10\times$ faster than the Levenberg–Marquardt least-squares fitting used in ADEPT.

Therefore, CUSUM+ is preferred for events with mean residence times considerably longer than the recovery time of the system ($\gg 5\tau$). Furthermore, the processing time per event for each algorithm scales linearly with the residence time, and therefore the number of data points in an event (see Supporting Information, Figure S5).

Analysis of Single-Stranded DNA Oligonucleotides with ADEPT. We use ADEPT to determine the blockade depth ratio histograms for three different length single-stranded DNA (ssDNA) homopolymers ($dA_{100}$, $dA_{40}$, and $dA_{20}$) entering an $\alpha$HL nanopore from the cis side.

We consider events with up to 6 discrete states, with each state containing at least 5 data points ($t_{\text{res}} > 10 \mu s, F_s = 500$ kHz). Events are partitioned from the time series data with a thresholding algorithm that identifies the current when the current deviates by more...
than 5 standard deviations from the mean open channel current. A complete listing of the analysis parameters is shown in Supporting Information, Table S3.

The current blockades observed with poly(dA) appear as either one level (a shallow or deep blockade) or two levels (a shallow blockade followed by a deeper one). Approximately 60% of the events have two levels. As shown in the blockade depth ratio histograms for dA100, dA40, and dA20 (Figure 4), these two observed levels are comprised of several different states.

The location of these two peaks agrees with previous measurements of dA100 where two highly overlapping peaks were observed at these locations. In contrast to earlier measurements we resolve the 3′ and 5′ events with a separation better than 3σ. The differences between the 3′ and 5′ blockade depth peaks (Figure 4A, two leftmost peaks) are progressively more difficult to discern for the shorter polynucleotides (Figure 4B, C). The amplitude of the 5′ peak decreases substantially for dA40 (Figure 4B) and is not resolved for dA20 (Figure 4C). These results are likely due to the lower probability of the 5′ end entering the pore and the decreasing residence time of shorter ssDNA molecules.

Interestingly, the shallow blockade level (⟨i⟩/⟨i₀⟩ ≈ 0.5) is characterized by a single peak for dA100 and dA40 and two peaks for dA20. Previous studies have either not reported this peak or only noted it for molecules as short as dA50. Furthermore, the sharp decrease in residence time associated with shorter polymers complicates their analysis and has thus far limited the analysis of polynucleotides as short as dA20.

The algorithms within MOSAIC improve the characterization of short polynucleotides. Figure 5 shows the voltage-dependent behavior of the dA20 shallow blockade peaks and their residence time distributions. The shallow blockade depth distributions are qualitatively different than those measured for longer polymers (Figure 4), as noted previously. Furthermore, we observe a change in the morphology of the peaks with increasing voltage as seen in Figure 5A. In particular, increasing the magnitude of the applied potential: (i) shifts the peaks to smaller ⟨i⟩/⟨i₀⟩ values, i.e., the polynucleotide blocks more
current (Figure 5A), (ii) increases the residence times (Figure 5B) (in contrast to the mean residence time of the deep blockades in Figure 4 that are associated with translocation), \(4, 5\) and (iii) increases the number of events observed per unit time (capture rate) (see the Supporting Information, Figure S7).

Moreover, with increasing voltage, the shallow blockades were more likely to exhibit two states with the shallow blockade preceding a deep blockade. The above results strongly suggest that the shallow blockade peaks correspond to \(\text{dA}_{20}\) interacting with the vestibule but not translocating through the pore. Furthermore, the observed increase in the residence time of the shallow blockade (Figure 5B) with voltage suggests that the change in the leftmost peak amplitude in Figure 5A is likely due to interactions between the analyte and different regions of the vestibule, rather than a loss of signal. Interestingly, of the three measured analytes (\(\text{dA}_{100}, \text{dA}_{40}, \text{dA}_{20}\)), this voltage-dependent change in peak structure was only observed with \(\text{dA}_{20}\) indicating that the phenomenon may be length-dependent.

**Example of Using MOSAIC: Single Molecule Mass Spectrometry with a Biological Nanopore.** We show a typical analysis using MOSAIC’s graphical user interface. Specifically, we use both CUSUM+ and ADEPT to separate monomers in polydisperse PEG samples with an \(\alpha\)HL nanopore. \(4, 20, 22, 26, 55, 56\) Previous studies showed that the blockade depth ratio \(\langle i \rangle / \langle i_0 \rangle\) and mean residence time of these events scale monotonically with polymer size.

Analysis of the PEG data is set up using the GUI shown in Figure 6 and is configured using drop down menus. After selecting a data source (MOSAIC accepts most common electrophysiology data formats: Axon ABF, QUB QDF, as well as raw binary and comma separated value, CSV, data), a segment of the time series is displayed, which assists in determining the mean open channel current \(\langle i_0 \rangle\), noise \(\sigma_{i_0}\) and threshold values for preliminary event identification.

A key feature of MOSAIC is the ability to integrate custom algorithms into the processing pipeline. Within the GUI, the user can select the analysis algorithm. The PEG data were analyzed independently using both the ADEPT and CUSUM+ algorithms. The blockade depth histogram of the events and the processing statistics are presented in real time. Fits of the physical model (ADEPT) or detected states (CUSUM+) of individual analyzed events are also displayed to monitor the progress and quality of the analysis. The results are stored in a SQLite database (or can be exported as a CSV file from within the GUI) for further analysis.

This example further illustrates the differences between the CUSUM+ and ADEPT algorithms. Only events that deviate from the open channel current baseline by at least \(2.7\sigma\) were analyzed. Events shorter than \(5\tau\) were excluded from the CUSUM+ analysis (the default value when running CUSUM+...
from the GUI). On the other hand, when using ADEPT, we excluded events shorter than 2.5r (25 μs) to minimize fitting errors (set with the Advanced Settings dialogue in the GUI).

Both algorithms produce well-resolved peaks for PEGs larger than 17-mers (Figure 7 and Supporting Information, Figure S8). For smaller PEGs that have shorter mean residence times, CUSUM+ shows a single broad peak whereas ADEPT easily excludes events shorter than 2.5r (Figure 7, gray). As expected, for both algorithms, where the peaks are resolved, the peak positions are in good agreement. For PEGs (n < 17), the signal-to-noise ratio of the CUSUM+ peaks is lower than those recovered by ADEPT (Supporting Information, Figure S8), consistent with the lower number of events recovered by CUSUM+ in this region.

CONCLUSIONS
We developed a new open source platform for the analysis of single-molecule data (MOSAIC) and implemented two robust optimized algorithms (ADEPT and CUSUM+) for biological or solid-state nanopore measurements. When applied to dsDNA measurements with a 2.4 nm SiN₃ nanopore, MOSAIC found previously undetected states most likely arising from the transient interactions between dsDNA and the access region of the solid-state pore. Additionally, when measuring short oligonucleotides poly(dA)ₙ, MOSAIC accurately analyzed events with residence times <5r, thereby characterizing previously unreported interactions of dA₂₀ with the αHL nanopore. Such analysis can be used to provide greater insight into the underlying physics of analyte-nanopore interactions.

ASSOCIATED CONTENT
Supporting Information
The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.6b03725.

Additional analysis of ssDNA and dsDNA blockade depth histograms and extended discussion of MOSAIC software contents and structure (PDF)

AUTHOR INFORMATION
Corresponding Author
*E-mail: arvin.d.balijepalli@nist.gov. Phone: (301) 975-3526.

Author Contributions
The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes
Certain commercial entities, equipment, or materials may be identified in this document in order to describe an experimental procedure or concept adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the entities, materials, or equipment are necessarily the best available for the purpose. The authors declare no competing financial interest.

ACKNOWLEDGMENTS
This work was supported in part by Grant R01HG007415 from the National Human Genome Research Initiative (J.J.K) and by the National Sciences and Engineering Research Council of Canada (NSERC). K.B. acknowledges the financial support provided by NSERC and the Vanier CGS program for postgraduate fellowships. We gratefully acknowledge the Ju Laboratory (Columbia University) for providing the preheptamerized alpha-hemolysin used in some of these studies. We thank the numerous research groups who provided feedback on the software during its development.

REFERENCES


(31) Bezanilla, F. Physiol Rev. 2000, 80 (2), 555−592.


(42) Briggs, K.; Kwok, H.; Tabard-Cossa, V. Small 2014, 10 (10), 2077−2086.


(45) Page, E. S. Biometrika 1954, 41 (1/2), 100.


(50) Butler, T. Z.; Gundlach, J. H.; Troll, M. Biophys. J. 2007, 93 (9), 3229−3240.


