Single-molecule mass spectrometry in solution using a solitary nanopore

Joseph W. F. Robertson, Claudio G. Rodrigues, Vincent M. Stanford, Kenneth A. Rubinson, Oleg V. Krasilnikov, and John J. Kasianowicz

PNAS published online May 9, 2007;
doi:10.1073/pnas.0611085104

Single-molecule mass spectrometry in solution using a solitary nanopore

Joseph W. F. Robertson*, Claudio G. Rodrigues†, Vincent M. Stanford‡, Kenneth A. Rubinson*§, Oleg V. Krasilnikov†¶, and John J. Kasianowicz*¶

*Electronics and Electrical Engineering Laboratory, Semiconductor Electronics Division, National Institute of Standards and Technology, Gaithersburg, MD 20899-8120; †Laboratory of Membrane Biophysics, Department of Biophysics and Radiobiology, Universidade Federal de Pernambuco, 50670-901, Recife, PE, Brazil; ‡Information Technology Laboratory, Information Access Division, National Institute of Standards and Technology, Gaithersburg, MD 20899-8940; and §Department of Biochemistry and Molecular Biology, Wright State University, Dayton, OH 45435

Edited by Nicholas J. Turro, Columbia University, New York, NY, and approved March 29, 2007 (received for review December 13, 2006)

We introduce a two-dimensional method for mass spectrometry in solution that is based on the interaction between a nanometer-scale pore and analytes. As an example, poly(ethylene glycol) molecules that enter a single α-hemolysin pore cause distinct mass-dependent conductance states with characteristic mean residence times. The conductance-based mass spectrum clearly resolves the repeat unit of ethylene glycol, and the mean residence time increases monotonically with the poly(ethylene glycol) mass. This technique could prove useful for the real-time characterization of molecules in solution.

Gaussian mixture model | poly(ethylene glycol) | α-hemolysin | ion channel | maximum likelihood signal classification

In his landmark patent, Wallace Coulter introduced a method to detect particles suspended in solution and driven through a capillary (1). The number and sizes of the particles are determined by measuring electrical resistance changes in the capillary. With capillaries of diameter $d \sim 100$ μm, the Coulter technique is widely used to count and size red blood cells and platelets. More recently, mesofluidic structures and carbon nanotubes with diameters <1 μm have been used for analysis of macromolecules, colloids, and bioparticles typically >100 nm in diameter (2–4). Further reduction of the portal diameter leads to phenomenological limitations by excluding flow driven with hydrostatic pressure or electroosmosis. As a result, transport through nanostructures must occur by diffusion or electromigration.

The natural useful limit of the capillary diameter is the size of the molecules of interest, and biological ion channels have diameters and lengths commensurate with molecular dimensions (5). The characteristic time for a molecule to diffuse the length of such pores (~10 nm) is 50–500 ns (6), which is inaccessible for meaningful conductance measurements. Thus, the motion of individual molecules that enter the pore must be inhibited by physical or chemical interactions to permit analytical conductance-based measurements of separate events.

Biological channels in lipid bilayers have been used to detect and quantitate a variety of analytes, including H^+ and D^+ in solution (7), single-stranded RNA and DNA (8–13), small organic molecules (14), specific sugar molecules (15), poly(ethylene glycol) (PEG) (16–22), and anthrax toxins (23). More recently, solid-state silicon nitride nanopores were used to detect individual double-stranded DNA molecules (24, 25).

PEG has been used to estimate the size of biological ion channels. Earlier studies demonstrated that PEGs small enough to enter the *Staphylococcus aureus* α-hemolysin (αHL) nanopore decrease the single-channel conductance (16, 18, 19, 21). The mean residence times of PEGs in this channel increase with electrolyte concentration (18, 22). As we show here, the sufficiently long polymer–pore interaction time enables the discrimination of each species within a homologous series of PEG molecules with nearly baseline resolution.

Results and Discussion

Typical single-channel recordings in the absence and presence of PEG are shown in Fig. 1. A large ionic current flows through a single αHL channel ($I = 150$ pA $\pm$ 2.6 pA, $P = 0.001$; for $V = +40$ mV, 4 M KCl, no PEG; Fig. 1 *Top Left* and *Middle Left*). A polydisperse PEG sample (pPEG) [$HO(CH_2CH_2O)_nH$, $25 < n < 50$; *vide infra*] added to the trans compartment causes transient, partial-current blockades (Fig. 1 *Top Right* and *Middle Right*). Each decrease and subsequent increase in the ionic current corresponds to a single PEG molecule entering and exiting the pore, respectively. The duration of a blockade event is the residence time of the polymer in the pore. The PEG-induced blockades are widely distributed in both conductance and residence time (Fig. 1 *Middle Right* and *Bottom*).

A larger sample of blockades (Fig. 2, ~700 events) illustrates the marked differences between current blockades caused by pPEG (Fig. 2 *Upper*) and monodisperse PEG (mPEG) 1294 ($n = 29$, Fig. 2 *Lower*). The all-points histograms of the ionic current (Fig. 2 *Right*) demonstrate that pPEG can be distinguished easily from mPEG. The histogram from a significantly longer time series with $>10^5$ events does not resolve the individual molecular species, even though visual inspection of the blockade events suggests that distinct conductance states exist. Although each conductance state is normally distributed, the individual blockades could not be accurately decoded using a Gaussian mixture model (GMM) fit using expectation maximization (EM) (26, 27).

To resolve and accurately fit the individual components within the mixture, we modified our previous statistical method (28) by adding a signal preconditioning step. Specifically, each blockade event is represented by its mean current value. A histogram made from the mean blockade currents easily resolves ~24 differently sized PEGs (Fig. 3 *Upper*, red). As anticipated, the mean current histogram for mPEG 1294 ($n = 29$) shows a primary peak at $I/I_{open} = 0.250 \pm 0.005$ with a small anisotropy on the higher current side, i.e., lower mass (Fig. 3 *Upper*, blue). The mPEG histogram provides a correlate of PEG molecular mass, thus calibrating the mean current histogram into a mass spectrum. The 1:1 correspondence between this histogram and a MALDI-TOF mass spectrogram for the same pPEG sample (Fig. 3

Author contributions: J.W.F.R., C.G.R., and V.M.S. contributed equally to this work; O.V.K. and J.J.K. designed research; J.W.F.R., C.G.R., and K.A.R. performed research; V.M.S. contributed new analytic tools; J.W.F.R. and V.M.S. analyzed data; and J.W.F.R., K.A.R., V.M.S., and J.J.K. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Abbreviations: αHL, α-hemolysin; GMM, Gaussian mixture model; mPEG, monodisperse PEG; PEG, poly(ethylene glycol); pPEG, polydisperse PEG.

¶To whom correspondence may be addressed. E-mail: kras@ufpe.br or john.kasianowicz@nist.gov.

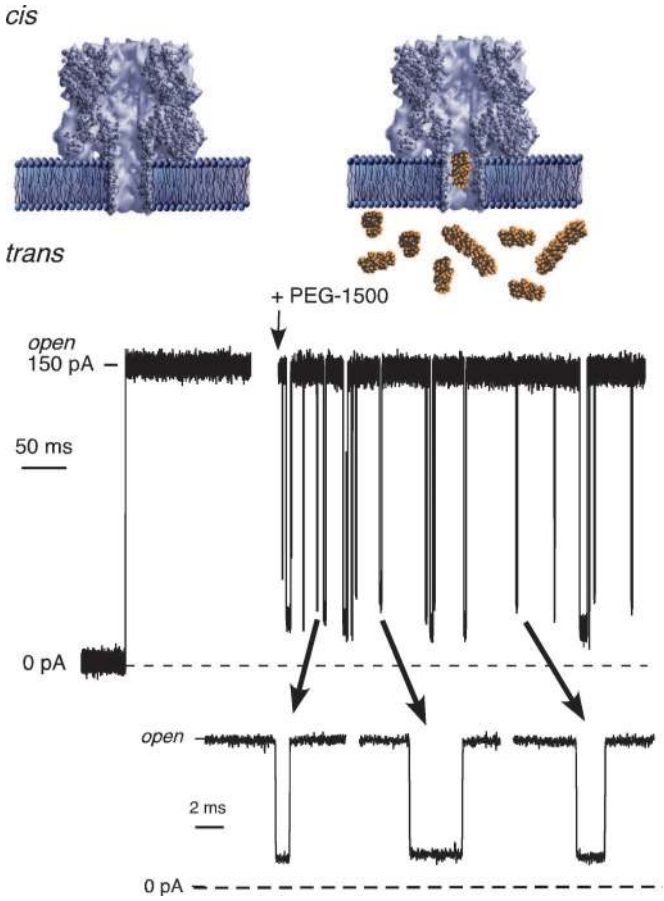

Fig. 1. Neutral polymers cause well defined reductions in the ionic current as they partition into a solitary nanopore in a lipid bilayer membrane. (*Middle Left*) The ionic current, through an αHL channel bathed by a polymer-free solution, is quiescent. Addition of polydisperse PEG (M_r = 1,500 g/mol) (*Top Right*) cause persistent current blockades (*Middle Right* and *Bottom*). The solutions bathing the membrane contained 4 M KCl and 5 mM Tris buffer, pH 7.5. (*Middle* and *Bottom*) The horizontal dashed lines indicate zero current.

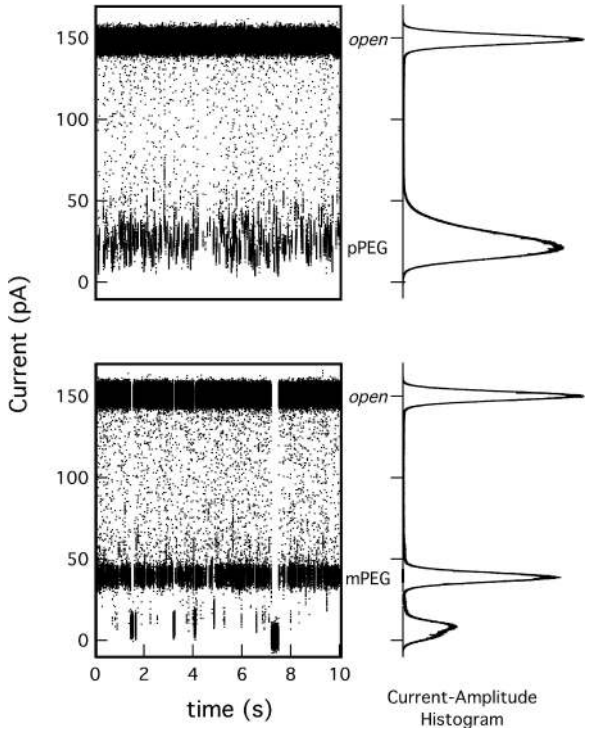

Fig. 2. A single nanopore discriminates between polymers with different molecular masses. The difference between the conductance states caused by polydisperse (M_r = 1,500) (*Lower Left*) and monodisperse (M = 1,294 g/mol, n = 29) (*Upper Left*) PEG is readily apparent. The time series data shown contained ~500 and ~700 events for the poly- and monodisperse PEG samples, respectively. (*Upper* and *Lower Right*) All-points histograms of the ionic current reflect the distinct natures of the two polymer samples. The ionic current histograms for each sample were calculated from $>10^5$ blockade events. The long-lived, small ionic current blockades near zero in the monodisperse PEG time series are most likely caused by impurities in the PEG samples. These events are long-lived but few in number.

Lower) demonstrates the accuracy of a solitary nanopore as a molecular sizing device.

Although the event–mean histogram visibly resolves each component in the pPEG mixture, additional statistical techniques are required to extract key features, such as peak amplitudes, precise peak position, and characteristic residence time of each size of polymer in the pore. To provide an unbiased analysis and ensure an excellent fit, an automated GMM fitting procedure with more components than the number of visually identifiable peaks was used (see *Materials and Methods*). The resultant GMM fits the empirical probability density function well (Fig. 3 *Upper*, solid black line). Of the initial 100 components, the statistical procedure rejected 57 components because of two criteria: low mixture weight (53 components) and excessive width (4 components). The low mixture weight components are insignificant with respect to the total fit, whereas the broad peaks represent an unresolved baseline. The remaining 43 narrow Gaussians are used to assign each individual blockade event to its most likely conductance state by using a maximum likelihood rule. Approximately 25% of the conductance states include a second, minor Gaussian to provide an adequate statistical fit.

Fig. 4 illustrates a typical assignment of the averaged individual current blockades (Fig. 4, red points) to the GMM states (Fig. 4, black lines) by using a simplified hidden Markov model decoding procedure (28–30). This process permits the estimation of the residence times for each PEG-induced current blockade state and thus for each polymer size. The detailed view of the time series (Fig. 4 *Lower*) shows that the larger polymers spend more time in the pore than do the smaller ones. The residence time distribution for each polymer in the homologous series is exponential (Fig. 5, 3 of the 24 residence-time distributions shown). This result suggests first-order binding kinetics between the polymers and the nanopore. As expected, the mean residence times of the pPEG and mPEG for n = 29 agree (e.g., Fig. 5, red and black points, respectively). The ability to identify the residence time for the mass of each analyte provides a second discriminant for multivariate analyses of aqueous molecular species (7, 14).

We have demonstrated a technique for mass discrimination by using a solitary molecular scale pore. Multivariate discriminants could enable analysis of numerous species in solution on the basis of molecular size and chemical functionality of the analytes. This single-molecule analysis technique could prove useful for the real-time characterization of biomarkers (i.e., nucleic acids, proteins, or other biopolymers). With automated, unsupervised analytical and statistical methods, this technique should prove viable as a generalized analytical technique with nanopore arrays containing nanopores both with specific affinities for single biomarkers and with nonspecific transducers such as αHL. *In situ* monitoring of cellular metabolism with such arrays should provide the sensitivity to monitor subtle changes observed through the release of biomarkers.

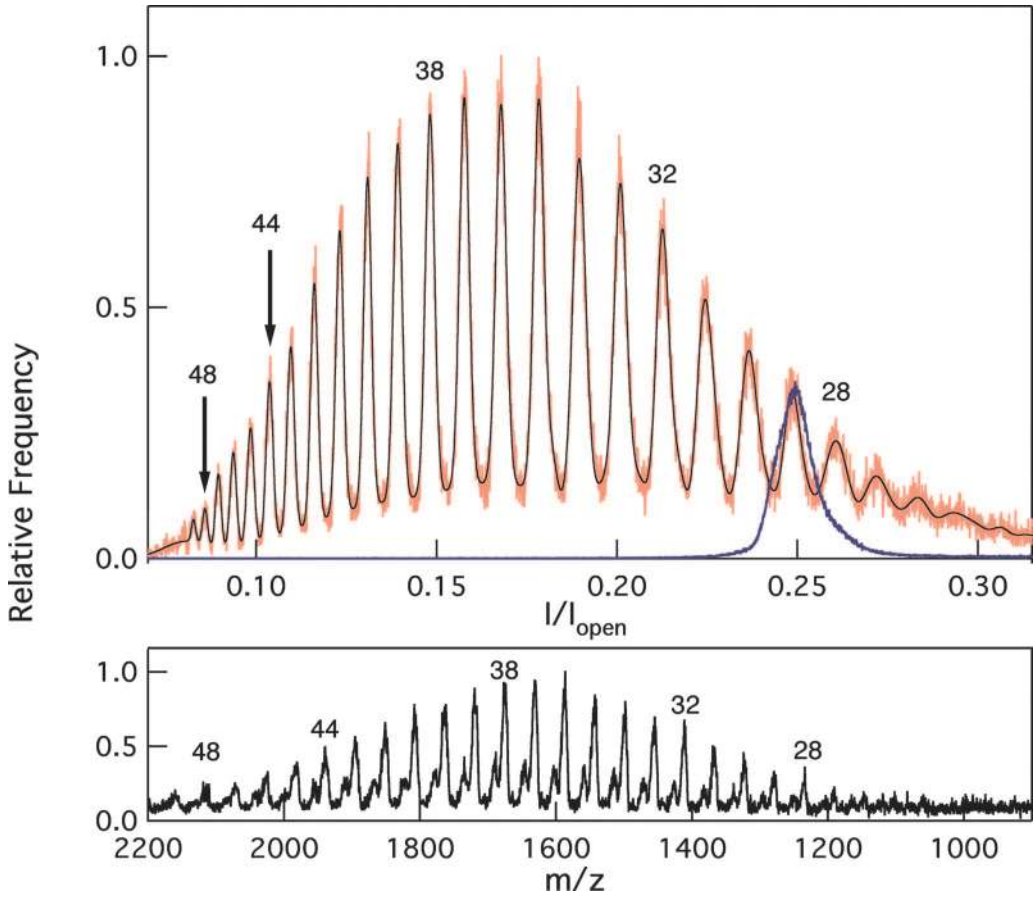

Fig. 3. Mass distributions obtained with a single nanopore (*Upper*) is compared with a conventional MALDI-TOF mass spectrum (*Lower*) for polydisperse PEG (M_r = 1,500 g/mol). The histogram was obtained as described in the text. Greater values of I/I_{open} correspond to lower PEG molecular masses. The histogram of the state-averaged current (red) are overlaid with the GMM fit (black). The model fits the empirical probability density function well with a Kolmogorov–Smirnov goodness of fit statistic, KS = 0.295 (32). The mean conductance-based histogram for monodisperse PEG-1294 (blue) is scaled to the height of the corresponding polydisperse peak. In the MALDI-MS, under the desorption/ionization conditions used, each PEG *n*-mer yields a parent ion peak, MH^+, and a base peak 16–17 units lower in mass, suggesting a loss of −O or −OH.

Materials and Methods

Membrane and Channel Formation. Solvent-free planar lipid bilayer membranes were formed from diphytanoyl phospatidylcholine (1,2-diphytanoyl-*sn*-glycero-3-phosphocholine; Avanti Polar Lipids, Alabaster, AL) in pentane (J. T. Baker, Phillipsburg, NJ) on an ~70-μm diameter hole in a 25-μm thick Teflon partition that separates two identical Teflon chambers. The hole was pretreated with a solution of 1:400 vol/vol hexadecane (Aldrich, St. Louis, MO) in pentane. Both chambers contained 4 M KCl (Mallinckrodt, Paris, KY), 5 mM 2-amino-2-hydroxymethyl-1,3-propanediol (Tris; Schwarz/Mann Biotech, Cleveland, OH), adjusted to pH 7.5 with concentrated citric acid (Fluka, Buchs, Switzerland).

Single channels were formed by adding ~0.25 μg of αHL (List Biological Laboratories, Campbell, CA) to the solution on one side of the partition, here termed cis. After a single channel formed, the cis chamber was rapidly flushed with fresh buffer to prevent further channel incorporation. Unless otherwise stated, the data were obtained with an applied potential of −40 mV with two Ag/AgCl electrodes separated from the bulk electrolyte by Vycor salt bridges (3 M KCl). The current was measured using an Axopatch 200B patch-clamp amplifier (Molecular Devices, Sunnyvale, CA) and filtered at 10 kHz with a four-pole Bessel filter before digitization at 50 kHz.

The αHL toxin can form at least two conformers that have different conductance levels and gating properties (7). Only the higher conductance conformer was used here, which has an approximately ohmic conductance of 3.75 nS between ± 50 mV (data not shown). PEG (pPEG 1500; Fluka; or mPEG 29; Polypure, Oslo, Norway) was added to the trans chamber from stock solutions of 12 mg/ml in electrolyte to a final concentration of 0.045 mg/ml.

MALDI-TOF Analysis of PEG Samples. MALDI-TOF mass spectra of the PEG samples were obtained with a Voyager DE-STR (PerSeptive Biosystems, Framingham, MA) by using the reflectron mode. Desorption/ionization was produced by irradiation with pulsed UV light (337 nm) from a nitrogen laser. The instrument was operated at 25 kV in the positive ion mode by using an extraction delay time set at 600 ns. The final spectra were averaged from 100 shots while moving the laser over the surface of the sample with the laser power set slightly over the threshold for the appearance of each spectrum. The samples were prepared from 1% wt/wt PEG solutions in distilled water. The matrix solution was 1:1 acetonitrile:water saturated with all-*trans*-retinoic acid (Sigma, St. Louis, MO) with 0.1% fluoroacetic acid (Matheson, Joliet, IL) added. The sample and matrix were mixed 1:1 to a total volume of 2 μl before drying.

From quantitative mixtures of PEGs with different nominal molecular weights, the instrumental sensitivity variation was <20% over the range 1,000 units to 10 kilounits. As a result, the instrument sensitivity changes can be ignored over molecular

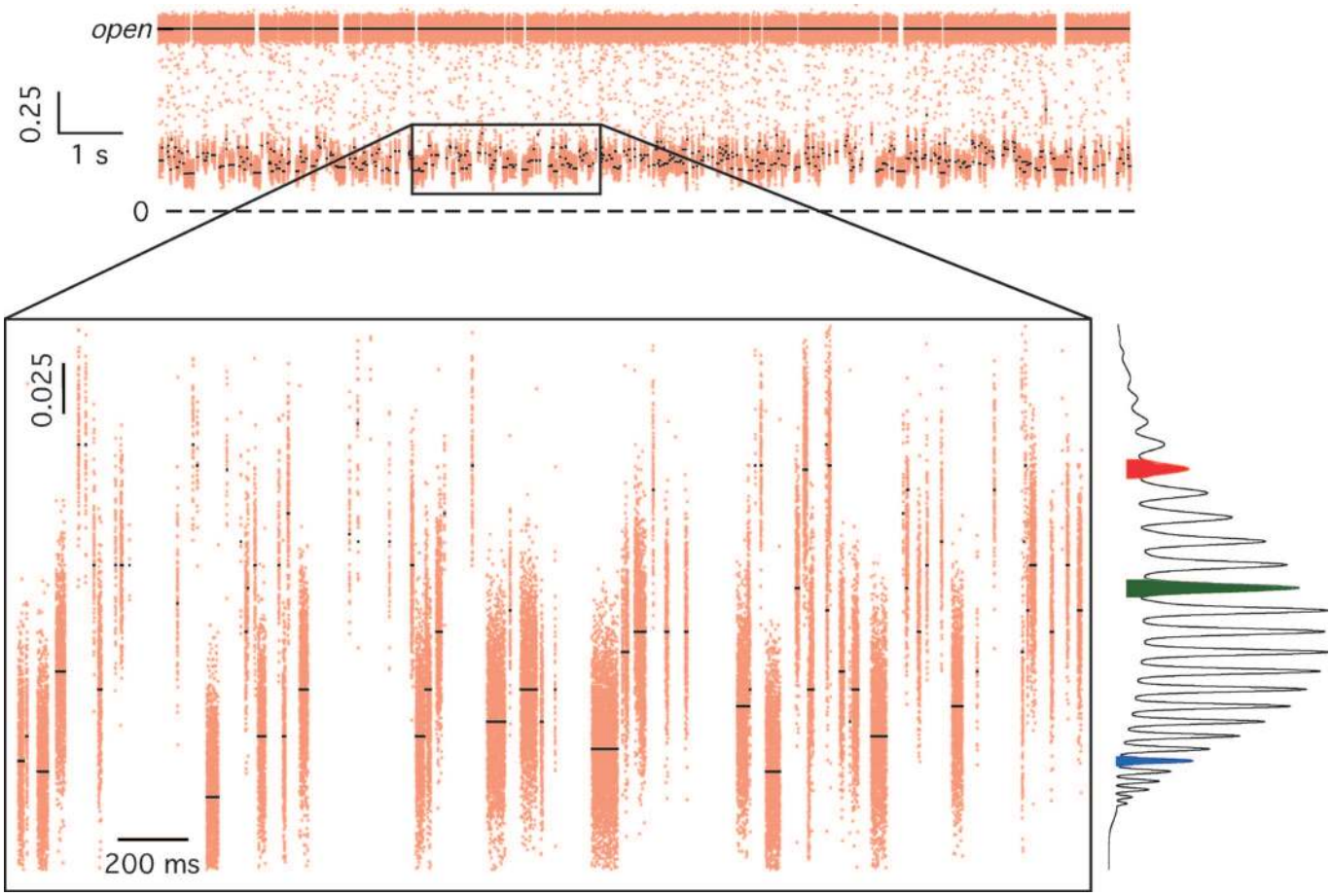

Fig. 4. The current through a solitary nanopore discriminates between individual PEG polymers that have different molecular masses. Ionic current blockades caused by individual molecules are assigned to Gaussian states of the nanopore mass spectrogram (Fig. 3 *Upper*). The GMM permits assignment of individual blockades to the conductance states by maximum likelihood decoding (solid black lines). (*Upper*) A 15-second-long block of data showing the open channel and blockade states. Expansion of the time series data in the highlighted region (*Lower Left*) compared with a histogram made from the GMM fit (*Lower Right*). The colored peaks in the histogram reference the individual polymers in the pPEG that are discussed in Fig. 5.

weight distributions for the nominal molecular weights used here; that is, the relative response to all n-mers in a sample are considered equal within experimental error.

Event-Mean Extraction. We extended the techniques described in ref. 28 to the present case of highly overlapped conductance blockades caused by a pPEG sample. First, we preprocessed the conductance time series, identifying the blockades by a 5 σ deviation from the mean open-channel conductance. Because virtually all (>99.5%) of the PEG-induced blockades could be characterized by a single Gaussian state, we took the mean value of the blockade event as the best estimate of the event amplitude. A histogram of these mean values was made for display (Fig. 3). In contrast to the unresolved unimodal amplitude distribution of the raw conductance signal shown in Fig. 2, we observed a finely resolved structure in the event-mean amplitudes (Fig. 3).

GMM. To extract reliable measurements from the sample of pPEG event means, we fit a multicomponent GMM to the sample. Because the expectation-maximization (EM) procedure can converge to an unsatisfactory local maximum, we started with many more Gaussian components than the number of peaks obvious in the histogram of the event-mean sample. We then discarded those assigned low mixture weights ($<10^{-4}$) by the EM estimation process. This resulted in a GMM fit that cannot

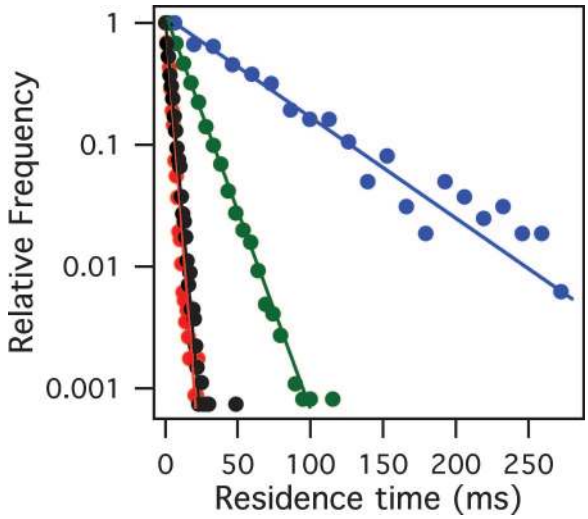

Fig. 5. Residence-time distributions associated with each polymer species vary systematically with the polymer mass. The derived residence time distributions are shown on a semilog plot for three representative states corresponding to the 1,294 (red), 1,558 (green), and 2,042 (blue) g/mol components of pPEG and to mPEG 1294 (black). The mean residence times, estimated from a least-squares fit of a single exponential to each data set are (in milliseconds) as follows: (2.8 ± 0.1), (3.2 ± 0.1), (13.4 ± 0.1), (52 ± 2) for mPEG 1294, pPEG 1294, pPEG 1558, and pPEG 2042, respectively.

be rejected at the 0.05 probability level as measured by the Kolmogorov–Smirnov statistic.

Maximum Likelihood Event State Assignment. We estimate a maximum *a posteriori* (MAP) state sequence for the incomplete data, given the hidden Markov model. In the present case, because we have represented each event by its mean conductance, we use a uniform transition matrix to reflect the fact that no intra-event state transitions are possible for the pPEG event means. This state transition model incorporates an ergodic and equiprobable state transition matrix into our GMM/hidden Markov model architecture to identify the conductance states and their associated mean amplitudes by maximum likelihood. Assignments are then made using the Viterbi decoding algorithm (31). This state assignment procedure partitions the event data into disjoint sets and thus permits estimation of the residence time distributions of PEG-induced conductance states corresponding to each Gaussian mixture component.

We thank Brian Nablo for technical help. This work was supported in part by Conselho Nacional de Desenvolvimento Cientifico e Tecnologico, Brazil (O.V.K.), Rede de Nanotecnologia Molecular e de Interfaces (O.V.K.), Instituto do Milenio de Materiais Complexos (O.V.K.), a National Research Council Research Associateship (to J.W.F.R.), the National Institute of Standards and Technology Single Molecule Manipulation and Measurement Program (J.J.K. and V.M.S.), a National Science Foundation-Nanoscale Interdisciplinary Research Team grant (to J.J.K.), and the National Institute of Standards and Technology Office of Law Enforcement Standards (J.J.K.).

1. Coulter WH (1953) US Patent 2,656,508.
2. Ito T, Sun L, Crooks RM (2003) *Anal Chem* 75:2399–2406.
3. Saleh OA, Sohn LL (2003) *Nano Lett* 3:37–38.
4. Heins EA, Siwy ZS, Baker LA, Martin CR (2005) *Nano Lett* 5:1824–1829.
5. Song LZ, Hobaugh MR, Shustak C, Cheley S, Bayley H, Gouaux JE (1996) *Science* 274:1859–1866.
6. Einstein A (1956) *Investigations on the Theory of the Brownian Movement* (Dover, New York).
7. Kasianowicz JJ, Bezrukov SM (1995) *Biophys J* 69:94–105.
8. Kasianowicz JJ, Brandin E, Branton D, Deamer DW (1996) *Proc Natl Acad Sci USA* 93:13770–13773.
9. Akeson M, Branton D, Kasianowicz JJ, Brandin E, Deamer DW (1999) *Biophys J* 77:3227–3233.
10. Henrickson SE, Misakian M, Robertson B, Kasianowicz JJ (2000) *Phys Rev Lett* 85:3057–3060.
11. Kasianowicz JJ, Henrickson SE, Weetall HH, Robertson B (2001) *Anal Chem* 73:2268–2272.
12. Vercoutere W, Winters-Hilt S, Olsen H, Deamer D, Haussler D, Akeson M (2001) *Nat Biotechnol* 19:248–252.
13. Mathe J, Aksimentiev A, Nelson DR, Schulten K, Meller A (2005) *Proc Natl Acad Sci USA* 102:12377–12382.
14. Gu LQ, Braha O, Conlan S, Cheley S, Bayley H (1999) *Nature* 398:686–690.
15. Kullman L, Winterhalter M, Bezrukov SM (2002) *Biophys J* 82:803–812.
16. Krasilnikov OV, Sabirov RZ, Ternovsky VI, Merzliak PG, Muratkhodjaev JN (1992) *FEMS Microbio Immun* 105:93–100.
17. Bezrukov SM, Vodyanoy I, Parsegian VA (1994) *Nature* 370:279–281.
18. Bezrukov SM, Vodyanoy I, Brutyan RA, Kasianowicz JJ (1996) *Macromolecules* 29:8517–8522.
19. Movileanu L, Bayley H (2001) *Proc Natl Acad Sci USA* 98:10137–10141.
20. Bezrukov SM, Kasianowicz JJ (2002) in *Structure and Dynamics of Confined Polymers*, eds Kasianowicz JJ, Kellermayer MSZ, Deamer DW (Kluwer, Dordrecht, The Netherlands), Vol 3, pp 117–130.
21. Krasilnikov OV (2002) in *Structure and Dynamics of Confined Polymers*, eds. Kasianowicz JJ, Kellermayer MSZ, Deamer DW (Kluwer, Dordrecht, Netherlands), Vol 3, pp 97–116.
22. Krasilnikov OV, Rodrigues CG, Bezrukov SM (2006) *Phys Rev Lett* 97:018301.
23. Halverson KM, Panchal RG, Nguyen TL, Gussio R, Little SF, Misakian M, Bavari S, Kasianowicz JJ (2005) *J Biol Chem* 280:34056–34062.
24. Li J, Stein D, McMullan C, Branton D, Aziz MJ, Golovchenko JA (2001) *Nature* 412:166–169.
25. Storm AJ, Chen JH, Zandbergen HW, Dekker C (2005) *Phys Rev E* 71:051903.
26. Redner RA, Walker HF (1984) *SIAM Review* 26:195–237.
27. Xu L, Jordan MI (1996) *Neural Comput* 8:129–151.
28. Kasianowicz JJ, Henrickson SE, Misakian M, Weetall HH, Robertson B, Stanford V (2002) in *Structure and Dynamics of Confined Polymers*, eds Kasianowicz JJ, Kellermayer MSZ, Deamer DW (Kluwer, Dordrecht, Netherlands), Vol 3, pp 141–164.
29. Baum L (1972) *Inequalities* 3:1–8.
30. Rabiner LR (1989) *Proc IEEE* 77:257–286.
31. Viterbi AJ (1967) *IEEE Trans Inform Theory* IT-13:260–267.
32. Chakaravarti IM, Laha RG, Roy J (1967) *Handbook of Methods of Applied Statistics* (Wiley, Chichester, UK).