

Single-Molecule Protein Unfolding in Solid State Nanopores

David S. Talaga^{*,†} and Jiali Li[‡]

Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854, and Department of Physics, University of Arkansas, Fayetteville, Arkansas 72701

Received February 18, 2009; E-mail: talaga@rutgers.edu

Abstract: We use single silicon nitride nanopores to study folded, partially folded, and unfolded single proteins by measuring their excluded volumes. The DNA-calibrated translocation signals of β -lactoglobulin and histidine-containing phosphocarrier protein match quantitatively with that predicted by a simple sum of the partial volumes of the amino acids in the polypeptide segment inside the pore when translocation stalls due to the primary charge sequence. Our analysis suggests that the majority of the protein molecules were linear or looped during translocation and that the electrical forces present under physiologically relevant potentials can unfold proteins. Our results show that the nanopore translocation signals are sensitive enough to distinguish the folding state of a protein and distinguish between proteins based on the excluded volume of a local segment of the polypeptide chain that transiently stalls in the nanopore due to the primary sequence of charges.

1. Introduction

Protein translocation across nanometer-scale pores is of fundamental importance both in basic science and in biotechnology. Nanopore translocation is a common event in trafficking of proteins between eukaryotic organelles. Emerging technologies have enabled progress in understanding protein translocation, and our understanding is still at an early stage.¹ Particles suspended in solution can be characterized^{2–4} by the drop in ionic current they cause during translocation through a small pore. This current drop is caused by the decrease in conductivity resulting from the displacement of a specific volume of electrolyte from the pore. The current drop persists for as long as the particle remains in the pore; this sojourn time depends on particle transport properties and charge as well as nanopore shape and electrical bias. For a protein particle, the current drop signal contains information on the protein's shape or folding state. If the protein molecule is unfolded, the signal depends on the amino acid sequence through the volume and charge of the polypeptide segment present inside the nanopore. Ion channels or protein pores reconstituted in lipid membranes have been utilized to characterize single polymers, DNA/RNA molecules,^{5–7} and proteins.^{8–11} More recently, single nanometer-scale solid-state pores have been fabricated in insulating silicon nitride (SiN_x) and silicon dioxide membranes^{12–16} and

used to study chain-like DNA molecules^{17–22} and native-state protein molecules.^{23,24}

Recent studies of protein translocation by Han et al.^{24,25} through larger-diameter synthetic nanopores, 55 nm²⁴ and 24–28 nm,²⁵ were performed in salt solutions without denaturing agents and without presenting event distributions. Fologea

[†] Rutgers, The State University of New Jersey.

[‡] University of Arkansas.

- (1) Wickner, W.; Schekman, R. *Science* **2005**, *310*, 1452–1456.
- (2) Gregg, E. C.; Steidley, K. D. *Biophys. J.* **1965**, *5*, 393–405.
- (3) DeBlois, R. W.; Bean, C. P. *Rev. Sci. Instrum.* **1970**, *41*, 909–916.
- (4) Bezrukov, S. M. *J. Membr. Biol.* **2000**, *174*, 1–13.
- (5) Bezrukov, S. M.; Vodyanoy, I.; Parasegian, V. A. *Nature* **1994**, *370*, 279–281.
- (6) Kasianowicz, J. J.; Brandin, E.; Branton, D.; Deamer, D. W. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 13770–13773.
- (7) Meller, A.; Nivon, L.; Brandin, E.; Golovchenko, J.; Branton, D. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 1079–1084.

- (8) Stefureac, R.; Waldner, L.; Howard, P.; Lee, J. S. *Small* **2008**, *4*, 59–63.
- (9) Movileanu, L.; Schmittschmitt, J. P.; Scholtz, J. M.; Bayley, H. *Biophys. J.* **2005**, *89*, 1030–1045.
- (10) Oukhaled, G.; Mathe, J.; Biance, A.-L.; Bacri, L.; Betton, J.-M.; Lairez, D.; Pelta, J.; Auvray, L. *Phys. Rev. Lett.* **2007**, *98*, 158101.
- (11) Robertson, J. W. F.; Rodrigues, C. G.; Stanford, V. M.; Rubinson, L. A.; Krasilinikov, O. V.; Kasianowicz, J. J. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 8207–8211.
- (12) Li, J.; Stein, D.; McMullan, C.; Branton, D.; Aziz, M. J.; Golovchenko, J. A. *Nature* **2001**, *412*, 166–169.
- (13) Storm, A. J.; Chen, J. H.; Ling, X. S.; Zandbergen, H. W.; Dekker, C. *Nat. Mater.* **2003**, *2*, 537–540.
- (14) Heng, J. B.; Ho, C.; Kim, T.; Timp, R.; Aksimentiev, A.; Grinkova, Y. V.; Sliagar, S.; Schulten, K.; Timp, G. *Biophys. J.* **2004**, *87*, 2905–11.
- (15) Cai, Q.; Ledden, B.; Krueger, E.; Golovchenko, J. A.; Li, J. *J. Appl. Phys.* **2006**, *100*, 024914.
- (16) Dekker, C. *Nat. Nanotechnol.* **2007**, *2*, 209–215.
- (17) Fologea, D.; Gershow, M.; Ledden, B.; McNabb, D. S.; Golovchenko, J. A.; Li, J. *Nano Lett.* **2005**, *5*, 1905–1909.
- (18) Fologea, D.; Uplinger, J.; Thomas, B.; McNabb, D. S.; Li, J. *Nano Lett.* **2005**, *5*, 1734–1737.
- (19) Li, J.; Gershow, M.; Stein, D.; Brandin, E.; Golovchenko, J. A. *Nat. Mater.* **2003**, *2*, 611–615.
- (20) Storm, A. J.; Chen, J. H.; Zandbergen, H. W.; Dekker, C. *Phys. Rev. E* **2005**, *71*, 051903.
- (21) Storm, A. J.; Storm, C.; Chen, J.; Zandbergen, H.; Joanny, J.-F.; Dekker, C. *Nano Lett.* **2005**, *5*, 1193–1197.
- (22) Gershow, M.; Golovchenko, J. A. *Nat. Nanotechnol.* **2007**, *2*, 775–779.
- (23) Fologea, D.; Ledden, B.; McNabb, D. S.; Li, J. *J. Appl. Phys. Lett.* **2007**, *91*, 053901.
- (24) Han, A.; Schurmann, G.; Mondin, G.; Bitterli, R. A.; Hegelbach, N. G.; de Rooij, N. F.; Staufer, U. *Appl. Phys. Lett.* **2006**, *88*, 093901–093903.
- (25) Han, A.; Creus, M.; Schurmann, G.; Linder, V.; Ward, T. R.; Rooij, N. F. d.; Staufer, U. *Anal. Chem.* **2008**, *80*, 4651–4658.

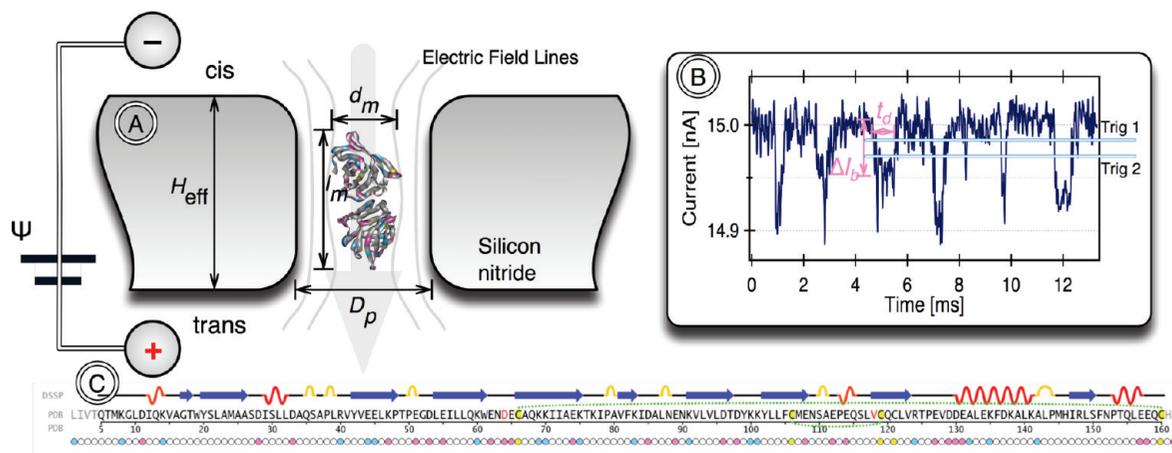


Figure 1. A) Schematic diagram of a nanopore translocation experiment. The three-dimensional structure of β LGa dimer derived from PDB file 2AKQ is shown approximately to scale with typical SiN_x nanopores. Positive and negative residues at pH 7 are colored blue and red respectively. B) Several recorded β LGa current blockage events (event driven mode). C) The primary sequence of β LGa.

et al.²³ used 16–18 nm diameter pores to study larger proteins BSA and Fibrinogen observing that both gave two clusters of events, though only the larger ΔI_b cluster was analyzed.

Protein nanopores have been used to study proteins and peptides.^{8–10,26} Most of these results showed multiple peaks in current drop amplitudes. Oukhaled et al.¹⁰ observed short and long duration current blockages. They concluded that short blockages are due to the passage of completely unfolded proteins, as their frequency increases as the concentration of the denaturing agent increases. The duration of the long passages was attributed to the wait time for protein unfolding.

In this article, we seek to measure single protein molecules at different concentrations of denaturant (urea) as they translocate through voltage-biased silicon nitride nanopores. We aimed to distinguish the native, partially folded, and unfolded states of bovine β -lactoglobulin variant a (β LGa) and to evaluate the ability of a nanopore to determine and/or influence the conformational state of a protein.

β LGa is a lipocalin present in whey, the stability and folding of which has been extensively studied.²⁷ Its principle biological role is apparently to provide a source of protein in milk. It may also enhance the solubility of fat and fat-soluble nutrients. Interest in β LGa unfolded structures and aggregation has been driven by basic science and by the importance of β LGa to the dairy and food processing industries. Translocation of lipocalin–ligand adducts through nanochannels provides a potential mechanism for remediation of hydrophobic and amphiphilic targets.

The unfolding of β LGa by urea has been studied by fluorescence^{28,29} dynamic light scattering (DLS)²⁹ and urea gradient gel electrophoresis.³⁰ Urea can be used to change the conformational state of β LGa from folded at 0 M to a partially folded intermediate at 5 M and unfolded at 8 M.^{28,29} Based on DLS diffusion measurements we can infer a mean hydrodynamic

volume of 24.4 ± 4.1 , 51.0 ± 16.6 , and 165 ± 44 nm³ at 0, 5, and 8 M urea, respectively, with the errors reflecting dispersion in the measurement.²⁹ Such changes are easily within the resolution of nanopore measurements. Since urea can lower the energy barrier to unfolding we expect this could influence the conformational integrity of the protein during translocation.

The main component of a nanopore sensing system was a single nanopore fabricated in a silicon nitride membrane that separated two salt-solution-filled chambers whose only electrical connection was *via* the electrolyte solution inside the nanopore (Figure 1A). A pair of Ag/AgCl electrodes was embedded in each chamber solution. When charged protein molecules were added to the *cis* chamber and the correct polarity voltage was applied to the electrodes, the protein molecules were captured by the electric field near the nanopore and driven through the nanopore to the *trans* chamber. The protein molecule interacting with, or translocating through, the nanopore caused a transient current drop (blockage). The current blockage events were recorded with an Axopatch 200B (Molecular Devices) integrated system (10 kHz low pass 4-pole Bessel filter, event-driven mode). At this setting, the nanopore measuring system was tested and calibrated with synthetic current blockages: ideal square pulses of pulse height 100 pA and width from 25 to 300 μ s, generated from a function generator (Agilent 33250A). (See Supporting Information for calibration details.) The recorded data were analyzed with the same MatLab routines as those used for DNA and protein translocation. When the pulse width was between 25 and 100 μ s, the pulse height was attenuated, but the time durations remained correct under our data analysis procedure. The preservation of the square pulse shape is the design property of multipole low-pass Bessel filters. The current blockage amplitudes (ΔI_b as shown in Figure 1B) presented in this work were corrected with this calibration. Nanopores were fabricated by ion beam sculpting.^{12,15} The *cis* and *trans* chambers were cast in PDMS (polydimethylsiloxane).¹⁷

Previous studies with different shaped particles translocated in pores of varying sizes^{2–4,31} have shown that the instantaneous amplitude of current blockage $\Delta I_b(t)$ is approximately proportional to the instantaneous excluded atomic volume $\Lambda(t)$ of a translocation particle inside the pore and can be written as

- (26) Sutherland, T. C.; Long, Y.-T.; Stefureac, R.-I.; Bediako-Amoa, I.; Kraatz, H.-B.; Lee, J. S. *Nano Lett.* **2004**, *4*, 1273–1277.
 (27) Sawyer, L.; Kontopidis, G. *Biochim. Biophys. Acta* **2000**, *1482*, 136–148.
 (28) Yagi, M.; Sakurai, K.; Kalidas, C.; Batt, C. A.; Goto, Y. *J. Biol. Chem.* **2003**, *278*, 47009–47015.
 (29) Giurleo, J. T.; He, X.; Talaga, D. S. *J. Mol. Biol.* **2008**, *381*, 1332–1348.
 (30) Beringhelli, T.; Eberini, I.; Galliano, M.; Pedoto, A.; Perduca, M.; Sportiello, A.; Fontana, E.; Monaco, H. L.; Gianazza, E. *Biochemistry* **2002**, *41*, 15415–15422.

- (31) Henriquez, R. R.; Ito, T.; Sun, L.; Crooks, R. M. *Analyst* **2004**, *129*, 478–482.

$$\Delta I_b(t) = -\frac{\sigma\psi}{H_{\text{eff}}^2} \Lambda(t) \left[1 + f\left(\frac{d_m}{D_p}, \frac{l_m}{H_{\text{eff}}}\right) \right] \quad (1)$$

Here σ is the solution conductivity, and ψ is the applied voltage to the electrodes. Over the range 0.5–3.0 M KCl the conductivity of a single nanopore is nearly linearly proportional to the KCl concentration.^{18,32} Consequently, we used the measured bulk conductivity of the buffered salt solutions for the nanopore conductivities in this work. These measured values were $\sigma = 112, 169, 144,$ and 110 mS/cm for 1 M KCl, 2 M KCl, 2 M KCl + 5 M urea, and 2 M KCl + 8 M urea, respectively. Urea reduces the magnitude of current blockage values in proportion to its effect on solution conductivity in eq 1. As illustrated in Figure 1A, d_m is the diameter and l_m is the length of a protein molecule, and D_p is the average diameter of a cylindrical nanopore. $f(d_m/D_p, l_m/H_{\text{eff}})$ is a correction factor that depends primarily on the relative geometry of the particle and the pore but also includes other possible parameters we have ignored. The physical thickness of the nanopores fabricated by ion beam sculpting was estimated to be 10 to 15 nm.^{15,33} The calibrated effective thickness, H_{eff} , accounts for any extension of the electric field lines beyond the physical limits of the nanopore. Using eq 1, the instantaneous excluded volume of a molecule can be estimated from $\Delta I_b(t)$ with $\Lambda(t) \approx (\Delta I_b(t) * H_{\text{eff}}^2) / (\sigma\psi)$. For a long ($l_m \gg H_{\text{eff}}$) approximately cylindrical dsDNA molecule, eq 1 can be simplified as $\Delta I_b \approx \sigma\psi A_{\text{DNA}} / H_{\text{eff}}$, where A_{DNA} is the mean atomic volume per unit length of a dsDNA molecule and its value can be calculated from the four different nucleic acids.^{34,35} A value of $A_{\text{DNA}} = 1.8$ nm² per nm is used in this work (see SI for details). We calibrated H_{eff} using ΔI_b measured from a linear 2706 base pair dsDNA (pNEB206, NEB) to allow the excluded atomic volume of a protein in a nanopore to be estimated.

The current blockage duration, t_d , is taken to be the sojourn time of a protein in a nanopore. The distribution of t_d reflects the fundamental physics of translocation. If the translocation is dominated by viscous diffusion under a constant bias, the distribution of charged particle sojourn times can be solved analytically. We modeled the sojourn time as the first passage time (fpt) for one-dimensional diffusion of a charged particle in a constant electric field (Ψ/H_{eff}) from the entrance to the exit of the nanopore. The electrophoretic drift velocity (v) and the diffusion constant (D) allow derivation (see Supporting Information) of the sojourn time distribution:

$$P_{\text{fpt}}(t_d) = \frac{\exp\left\{\frac{(d - vt_d)^2}{4Dt_d}\right\}(vt_d + d)}{t_d\sqrt{4Dt_d\pi}} \quad \text{with} \quad d = \begin{cases} H_{\text{eff}} & l_m < H_{\text{eff}} \\ H_{\text{eff}} + l_m & l_m \geq H_{\text{eff}} \end{cases} \quad (2)$$

Equation 2 is appropriate for translocating objects that are approximately uniformly charged. Equations 1 and 2 have ignored many complex issues like electro-osmotic flow in and

near the pore, interactions between a charged molecule and the pore, and variable charge along the length of the translocating molecule.

2. Results

2.1. β LGa Current Blockage Events and Event Polarity.

2.1.1. Translocation Event Distribution under Different Conditions. The structure and primary sequence of β LGa appear in Figure 1A and C. The reported pI of β LGa ranges from 4.8–5.5, and it has a charge of $\sim 8e$ at pH 7.²⁷ β LGa was measured in 2 M KCl 10 mM buffer above (pH 7.0) and below (pH 4.6) the pI. After adding 35 ± 15 nM β LGa to the *cis* chamber at pH 7, we initially observed current blockage events only when the *trans* chamber electrode was positively biased, consistent with negative β LGa. Transient blockage events were typically detected at an average rate of ~ 10 s⁻¹. The distribution of event parameters did not change over the range of protein concentrations used in the *cis* chamber. Figure 1B shows typical current blockage events recorded in event driven mode (i.e., gaps between events are not recorded); the depth and duration of the events varied significantly. If event recording had previously been conducted for several hours, translocation events with the *trans* chamber negatively biased could be observed at a greatly reduced rate, indicating β LGa had translocated to the *trans* chamber.

Each recorded blockage event was characterized by its time duration t_d and its mean current drop amplitude ΔI_b as defined by the levels in Figure 1B. The joint distributions of recorded events for β LGa in 2 M KCl are shown in Figure 2A for $\Psi = 120$ mV and Figure 2D for $\Psi = 60$ mV. Each dot in the figures represents the t_d and ΔI_b of one blockage event. The joint distribution of ΔI_b and t_d for both biases showed two major clusters of events labeled as cluster 1 for smaller ΔI_b and cluster 2 for larger ΔI_b . The values of ΔI_b were smaller for $\Psi = 60$ mV compared to $\Psi = 120$ mV. The population of cluster 2 was $\sim 50\%$ of the total events for $\Psi = 60$ mV, and it was $\sim 20\%$ for $\Psi = 120$ mV in this measurement. The sojourn time histograms at the bottom axis show that the dwell time for β LGa at $\Psi = 60$ mV was somewhat shorter than those at 120 mV. The distribution of 60 mV events may be limited by the open pore noise.

2.1.2. Polarity Changes when pH < pI. β LGa at pH 4.6 in a clean apparatus produced current blockage events only when the *trans* chamber electrode was negatively biased, consistent with positively charged β LGa. The results of β LGa at pH 4.6 ($\Psi = -120$ mV) is shown in Figure 2E. The lowest values of the excluded volume were consistent with the volumes observed for cluster 1a and 1b of the pH 7.0 experiments. The volumes of clusters 2 and 3 are also consistent with the volume observed at pH 7.0; however the translocation times are substantially longer suggesting that the dynamics and mechanism of translocation may be very different at pH 4.6.

2.1.3. Nanopore Diameter. Similar measurements performed for β LGa proteins with different diameter (D_p) nanopores showed that the peak values of ΔI_b varied with D_p , suggesting changes in H_{eff} or in the correction factor f in eq 1 (data not shown). In this work, only results from the same nanopore are directly compared.

2.2. Event Analysis and β LGa Protein in 0, 5 and 8M Urea. Following the measurement shown in Figure 2A and 2D, the *cis* chamber solution was flushed out with 2 M KCl + 5 M urea at pH 7; β LGa was added, and the event distribution from

(32) Smeets, R. M.; Keyser, U. F.; Krapf, D.; Wu, M.-Y.; Nynke, H. D.; Dekker, C. *Nano Lett.* **2006**, *6*, 89–95.

(33) King, G. M.; Golovchenko, J. A. *Phys. Rev. Lett.* **2005**, *95*, 216103.

(34) Zwolak, M.; Ventra, M. D. *Rev. Mod. Phys.* **2008**, *80*, 141.

(35) Nadassy, K.; Tomás-Oliveira, I.; Alberts, I.; Janin, J.; Wodak, S. J. *Nucleic Acids Res.* **2001**, *29*, 3362–3376.

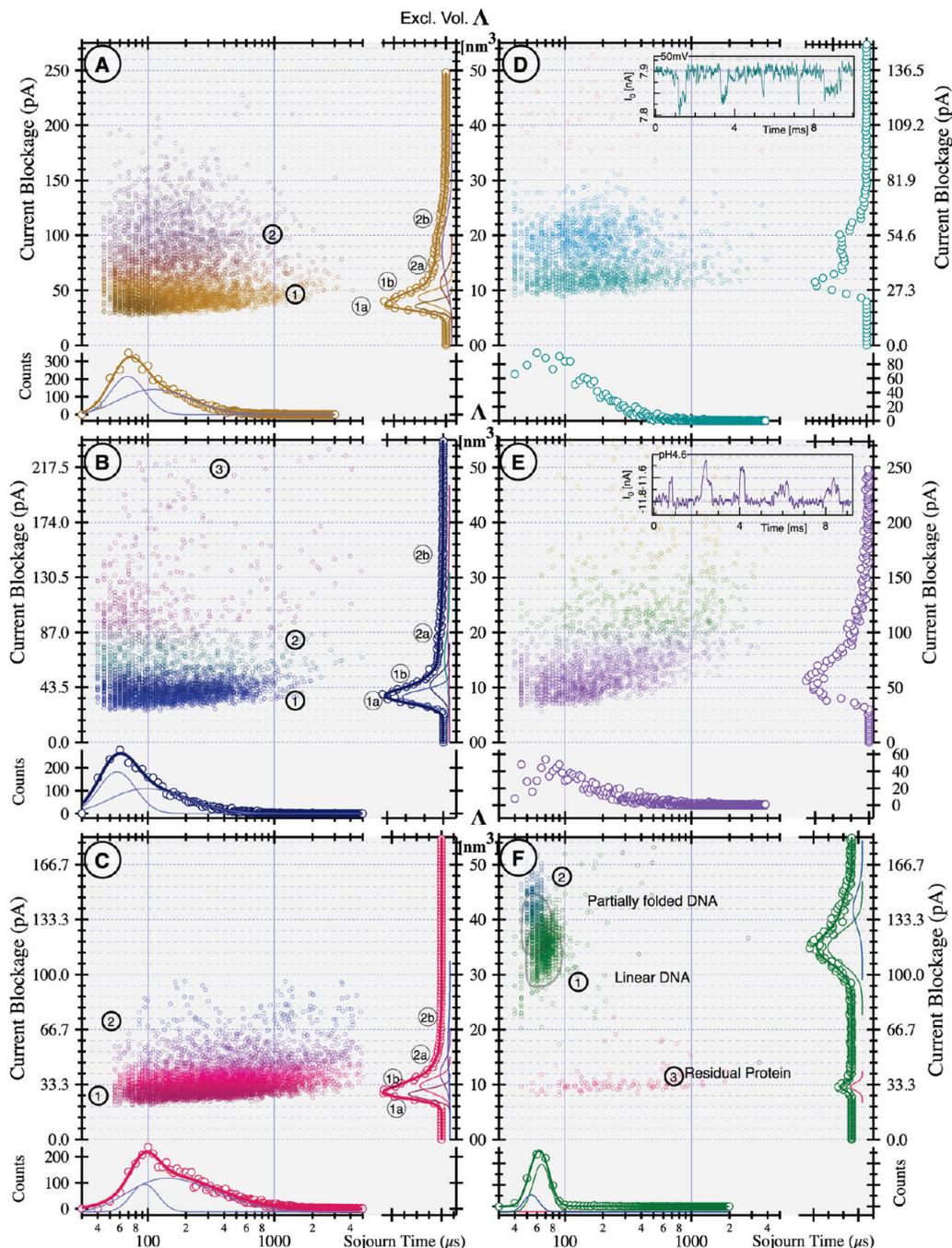


Figure 2. Event distributions. β LGa in: (A) 2 M KCl with no urea at pH 7 120 mV, (B) 2 M KCl 5 M urea at pH 7 120 mV, (C) 2 M KCl 8 M urea at pH 7 120 mV, (D) 2 M KCl with no urea at pH 7 60 mV inset: Sample events, (E) 2 M KCl with no urea at pH 4.6 –120 mV inset: Sample events. (F) 2.7 kbp dsDNA in 1 M KCl 120 mV. The solid curves on the right axes are fits the marginal distributions (circles) of ΔI_b with individual Gaussian contributions shown. The solid curves on the bottom axes are fits to the marginal distributions (circles) of t_d using the biased diffusion model. Circled numbers annotate clusters in the joint and marginal distributions as discussed in the text and Table 1. The scatter plots of the event joint distributions are colored by statistical contribution to the different clusters fit in the marginal current blockage distributions. All plots share the same excluded volume axis in the center. The diameter of nanopore used for the series measurements was $D_p = 8 \pm 2$ nm as imaged by TEM. (See SI) The open pore current I_0 was 15.5, 7.8, 6.5, 7.9, 12, and 7.5 nA from panels A to F respectively.

this measurement is shown in Figure 2B. The same procedure was repeated for 2 M KCl + 8 M urea, and the event distribution is shown in Figure 2C. In summary, these urea concentration measurements ($\Psi = 120$ mV) showed the following:

Without urea (2A), the distribution of ΔI_b comprised two major clusters of events. Cluster 1 had a most probable value of $\Delta I_b \approx 40$ pA with t_d across multiple scales from ~ 70 to 200 μ s. Cluster 2 had a most probable $\Delta I_b \approx 100$ pA with a very

broad distribution in ΔI_b . The population of cluster 2 was $\sim 20\%$ of the total events in this measurement.

In 5 M urea (2B), the most probable ΔI_b of cluster 1 was at ~ 36 pA, and the population of cluster 2 events had decreased to $\sim 10\%$. A third low-probability cluster was observed for only the 5 M urea data with ΔI_b 2–3 times that of cluster 2.

In 8 M urea (2C), the most probable value of ΔI_b of cluster 1 was at ~ 28 pA. Very few cluster 2 events were observed.

Table 1. Excluded Volumes and Sojourn Time Distribution Fit Parameters^a

Fit	Figure 2A	Figure 2B	Figure 2C	Figure 2F	Figure 3A	Figure 3B	units
	β LGa 0 M urea	β LGa 5 M urea	β LGa 8 M urea	DNA 0 M urea	β LGa 8 M urea	HPr 8 M urea	
Excluded Volume	$\Lambda_{1a} = 7.1 \pm 0.15$	$\Lambda_{1a} = 7.3 \pm 0.1$	$\Lambda_{1a} = 8.24 \pm 0.13$	$\Lambda_1 = 34.9 \pm 0.2$	$\Lambda_{1a} = 7.5 \pm 0.2$	$\Lambda_{1a} = 6.8 \pm 0.1$	nm^3
	$\Lambda_{1b} = 8.6 \pm 0.7$	$\Lambda_{1b} = 8.7 \pm 0.2$	$\Lambda_{1b} = 9.8 \pm 0.9$	—	$\Lambda_{1b} = 9.5 \pm 0.8$	$\Lambda_{1b} = 8.3 \pm 0.6$	nm^3
	$\Lambda_{2a} = 11.1 \pm 1.7$	$\Lambda_{2a} = 10.8 \pm 2.7$	$\Lambda_{2a} = 12. \pm 5.$	$\Lambda_2 = 41.9 \pm 0.8$	$\Lambda_{2a} = 12.3 \pm 2.1$	—	nm^3
	$\Lambda_{2b} = 17.6 \pm 1.4$	$\Lambda_{2b} = 17.4 \pm 2.1$	$\Lambda_{2b} = 16. \pm 5.$	$\Lambda_3 = 9.6 \pm 0.1$	—	—	nm^3
	$d = 20$	$d = 20$	$d = 20$	$d = 950$	$d = 20$	$d = 20$	nm
Biased Diffusion	$a_1 = 7.1 \pm 0.15$ $D_1 = 0.25 \pm 0.02$ $v_1 = 0.26 \pm 0.01$	$a_1 = 0.27 \pm 0.05$ $D_1 = 0.35 \pm 0.03$ $v_1 = 0.31 \pm 0.01$	$a_1 = 0.10 \pm 0.01$ $D_1 = 0.12 \pm 0.01$ $v_1 = 0.20 \pm 0.01$	$a_1 = 0.21 \pm 0.01$ $D_1 = 150 \pm 5$ $v_1 = 17.2 \pm 0.03$	$a_1 = 0.09 \pm 0.02$ $D_1 = 0.20 \pm 0.03$ $v_1 = 0.28 \pm 0.01$	$a_1 = 0.31 \pm 0.06$ $D_1 = 0.22 \pm 0.02$ $v_1 = 0.26 \pm 0.01$	— $\text{nm}^2/\mu\text{s}$ $\text{nm}/\mu\text{s}$
Biased Diffusion	$a_2 = 7.1 \pm 0.15$ $D_2 = 0.49 \pm 0.06$ $v_2 = 0.11 \pm 0.01$	$a_2 = 0.73 \pm 0.05$ $D_2 = 0.6 \pm 0.1$ $v_2 = 0.12 \pm 0.01$	$a_2 = 0.90 \pm 0.01$ $D_2 = 0.46 \pm 0.02$ $v_2 = 0.06 \pm 0.01$	$a_2 = 0.79 \pm 0.01$ $D_2 = 162 \pm 3$ $v_2 = 14.2 \pm 0.02$	$a_2 = 0.91 \pm 0.02$ $D_2 = 0.50 \pm 0.04$ $v_2 = 0.09 \pm 0.01$	$a_2 = 0.69 \pm 0.06$ $D_2 = 0.38 \pm 0.09$ $v_2 = 0.11 \pm 0.01$	— $\text{nm}^2/\mu\text{s}$ $\text{nm}/\mu\text{s}$
Activated Barrier/Exponential	$b_1 = 190 \pm 16$ $t_1 = 53 \pm 4$ $b_2 = 158 \pm 17$ $t_2 = 177 \pm 11$	$b_1 = 180 \pm 7$ $t_1 = 56 \pm 2$ $b_2 = 89 \pm 7$ $t_2 = 210 \pm 10$	$b_1 = 66 \pm 4$ $t_1 = 33 \pm 4$ $b_2 = 160 \pm 3$ $t_2 = 322 \pm 7$	— — — —	$b_1 = 195 \pm 2$ $t_1 = 230 \pm 3$ — —	$b_1 = 137 \pm 1$ $t_1 = 113 \pm 2$ — —	— μs — μs

^a Excluded volume distributions were fit with enough (2–4) Gaussian components to give flat residuals. Protein sojourn time distributions were fit to both the biased diffusion model (a , D , v) and to the activated barrier crossing model (b , t). Subscripts on the parameters correlate to the cluster labels in Figure 2. The errors are fitting errors.

The decrease in ΔI_b of cluster 1 was consistent with the decrease in solution conductivity σ with increasing urea concentration. Using the same data analysis parameters, no events were detected in a recorded trace of open pore current before β LGa was added, consistent with good discrimination of open pore noise from *bona fide* current blockages. However, we observed that adding β LGa to the *cis* chamber caused an increase in the open pore current fluctuations and generated short-lived spike events as shown in Figure 1B.

2.3. Calibration of Nanopore H_{eff} Using dsDNA. After the β LGa measurements, we calibrated the same nanopore using 2706 bp dsDNA in 1 M KCl at pH 7. Figure 2F shows the most probable ΔI_b for the dsDNA was $\Delta I_b \approx 120$ pA with $t_d \approx 70$ μs . This was consistent with our previous results on dsDNAs measured in nanopores fabricated the same way. Using $\sigma = 0.112/(\Omega \cdot \text{cm})$, $\Psi = 120$ mV, and $A_{\text{DNA}} = 1.8 \text{ nm}^2$,³⁵ a value of $H_{\text{eff}} = 20 \pm 2$ nm was calculated ($\Lambda^{\text{dsDNA}} = 36 \text{ nm}^3$) for this nanopore. Assuming $\Lambda \approx (\Delta I_b * H_{\text{eff}}^2)/(\sigma\Psi)$ holds for both DNA and β LGa protein translocations, we converted the ΔI_b to an estimated excluded volume Λ for β LGa as labeled on the center shared vertical axes of Figure 2. The differences in current drop scales arise from the different solution conductivities (σ) used in eq 1.

2.4. Distribution of the ΔI_b or Λ of β LGa Molecules. The marginal distribution of ΔI_b or Λ showed multiple peaks. We fit the histograms of ΔI_b with multiple Gaussian peaks, labeled in Figures 2 and 3 as 1a, 1b, 2a, 2b, and 3. The fitted peak positions of Λ values are listed in Table 1. The multiple peak fitting worked well to characterize the DNA translocation in two geometries. One peak located at $\Delta I_b = 120$ pA and $t_d = 70$ μs represents linear dsDNA translocation. The other peak located at $\Delta I_b = 133$ pA and $t_d = 60$ μs represents partially folded dsDNA translocation.¹⁹

For β LGa in 2 M KCl, the most probable excluded volume for the cluster 2 events was $\Lambda^{\beta\text{LGa}} \approx 20 \text{ nm}^3$. The most probable excluded volume for cluster 1 events in 0, 5, and 8 M urea was $\Lambda^{\beta\text{LGa}} \approx 8 \text{ nm}^3$. The excluded volume of the cluster 1 events was $\sim 40\%$ of the cluster 2 events.

2.5. Comparison of Unfolded β LGa and HPr. Figure 3 shows results from two different proteins, β LGa and HPr (85 amino

acids, $-2e$ at pH 7),³⁶ measured in the same nanopore at 2 M KCl and 8 M urea. HPr has no cysteine and is completely denatured in 8 M urea. The open pore current was the same ($I_0 = 3.2$ nA) during these measurements. Only cluster 1 events were observed for both of these proteins in 8 M urea. The most probable current drop values were $\Delta I_b(\text{HPr}) = 22 \pm 3$ pA and $\Delta I_b(\beta\text{LGa}) = 26 \pm 6$ pA. The position and width of the peaks in the β LGa current drop distribution found for the 8 M urea matched well between data sets (Figure 2C and 3B) taken on different days with different nanopores, suggesting that these two different pores have similar H_{eff} . The same value of $H_{\text{eff}} = 20$ nm was used to convert ΔI_b to Λ (right axes).

The marginal distribution of HPr ΔI_b events measured at 8 M urea showed an asymmetric peak that could be decomposed into two Gaussian contributions. The marginal distribution of β LGa ΔI_b events measured at 8 M urea was also asymmetric and required three Gaussian contributions for a satisfactory fit. Peak 1 of HPr and β LGa gave similar excluded volumes even though β LGa is nearly twice the size of HPr. Fit parameters for Figures 2 and 3 are summarized in Table 1.

2.6. Multiple Time Scales for Dwell Time Distributions. Fits of the marginal distribution of sojourn times to eq 2 appear as the solid lines through the experimental distributions at the bottom of the panels in Figures 2 and 3. Fit parameters are summarized in Table 1. For peak 1 of DNA, the parameters for linear translocation through the nanopore were $D_1 \approx 162 \text{ nm}^2/\mu\text{s}$ with $v_1 \approx 14 \text{ nm}/\mu\text{s}$ ($t_d \approx 70$ μs). For peak 2 of folded dsDNA,¹⁹ $D_2 \approx 150 \text{ nm}^2/\mu\text{s}$ with $v_2 \approx 17 \text{ nm}/\mu\text{s}$.

The distributions of nanopore sojourn times for proteins were substantially more heterogeneous than we observed for the DNA control. Fits of the observed distributions of sojourn times for the proteins required extremely low values for the diffusion constant and electrophoretic velocities. For β LGa, the fitting parameters are on the order of $D \approx 10^{-1} \text{ nm}^2/\mu\text{s}$ and $v \approx 10^{-1} \text{ nm}/\mu\text{s}$ ($t \approx 200$ μs). Both parameters are 3 orders of magnitude smaller than the bulk values²⁹ for diffusion constants ($\sim 10^2 \text{ nm}^2/\mu\text{s}$) and for drift velocity ($\sim 10^2 \text{ nm}/\mu\text{s}$), indicating that they are either significantly altered in the nanopore or the translo-

(36) Jia, Z.; Quail, J. W.; Waygood, E. B.; Delbaeres, L. T. J. *J. Biol. Chem.* **1993**, *268*, 22490–22501.

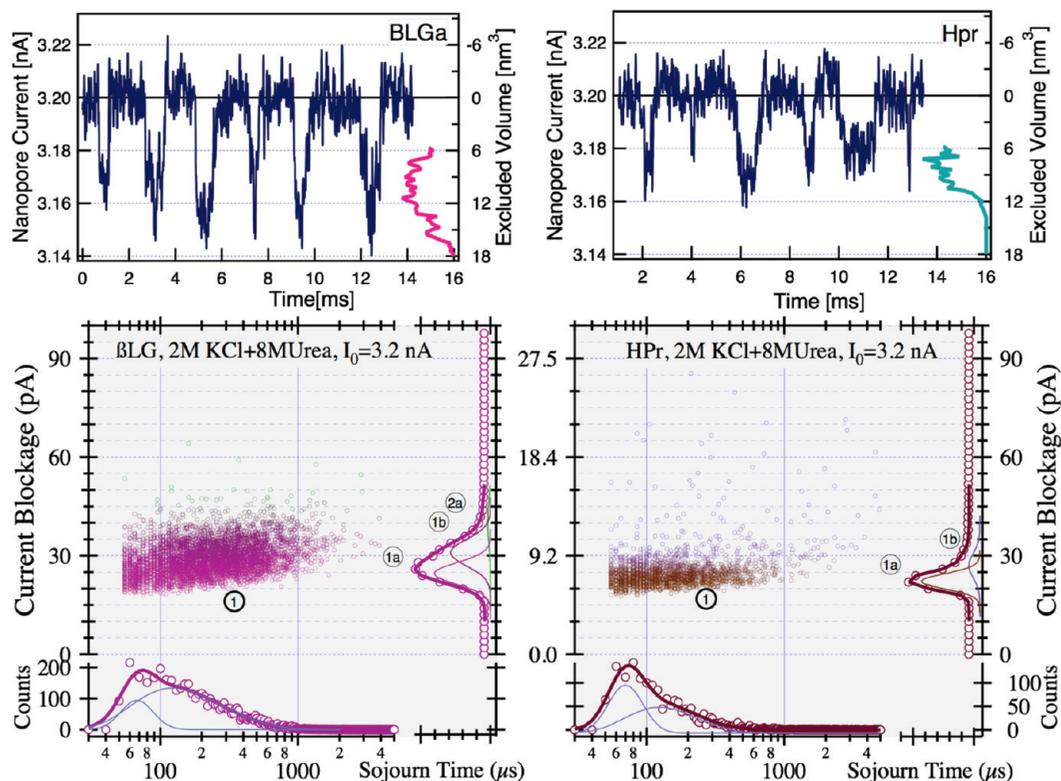


Figure 3. HPr (right panels) and β LGa (left panels) measured in the same nanopore. The nanopore had an average diameter of $D_p = 4 \pm 1$ nm and the open pore current was $I_0 = 3.2$ nA in 8 M urea and 2 M KCl.

cation process is not consistent with the biased diffusion mechanism. Using β LGa bulk values³⁷ for D and v in eq 2 predicts average sojourn times on the order of <1 μ s.

The anomalously long sojourn times suggested that barriers to exiting the nanopore exist and that thermal activation may be the limiting step for translocation. Activated sequential barrier crossing predicts sequential first-order kinetic steps resulting in a multiexponential first passage time distribution. Therefore we also fit the sojourn time distributions to such functions. Average exponential decay time parameters (t_1 and t_2) associated with the long time side of the distribution appear in Table 1. The short time scale rise ($t_d < 30$ μ s) in the exponential first passage time distribution was not accurately resolved due to the 10 kHz Bessel filter which reduced the amplitude of the resistive pulses to below the threshold for detection.

Both cluster 1 and cluster 2 in the β LGa event histograms required two Gaussian contributions to fit the excluded volumes (or ΔI_b) and either two biased diffusion contributions or two exponentials to fit sojourn times (t_d). The parameters describing the sojourn times for cluster 1 and cluster 2 were similar except for the relative weights in a given urea concentration. The presence of multiple sojourn time scales and the observation of multiple current blockage levels during translocation lead us to conclude that multiple states of β LGa occur during nanopore translocation.

3. Analysis and Discussion

In this section we discuss the nature of folded and unfolded β LGa and how this gives rise to the heterogeneous translocation event distributions shown in Figures 2 and 3. We relate these

observations to the sojourn time of the translocating molecule in the nanopore and the instantaneous volume displaced by the molecule during the translocation.

3.1. Excluded volume is not consistent with globular protein translocation. Adding the volume of all 161 β LGa amino acids gives an excluded volume of $\Lambda = 22.8$ nm³ per β LGa monomer.^{38,39} This volume is expected to generate at least $\Delta I_b \approx 110$ pA in 2 M KCl solution (Figure 2A). Folded β LGa has a large void calyx that should increase the true excluded volume from this value. Partial or complete unfolding increases a protein's hydrodynamic radius. One might naively expect that the nanopore translocation volume would similarly increase, but the opposite trend was observed.

The cluster 1 events for different urea concentrations have similar excluded volumes of $\Lambda^{\beta\text{LGa}} = 6.7\text{--}8.8$ nm³, approximately 33% that of the entire protein. Early studies² in large pores found that the current blockage was larger for spheres than for cylinders of the same volume. By neglecting this correction (f in eq 1) our DNA calibration procedure should *overestimate* the volume of globular β LGa. Under all conditions β LGa translocation was dominated by events with excluded volumes *smaller* than that of the monomeric folded protein.

The partial volumes of the 85 amino acids in HPr add to 11.2 nm³, or about half that of β LGa. The difference in blockage depth between HPr and β LGa at 8 M urea was only 10%. The similarity of blockage depth of β LGa and HPr is more like the linear translocation of DNA/RNA. For linear translocation (Figure 4D), the depth of current blockage is expected to be largely independent of contour length.

Figure 4B compares the excluded volume calculated from a 20 nm contour length segment of HPr and β LGa as a function

(37) Guzey, D.; McClements, D. J. *Food Hydrocolloids* **2006**, *20*, 124–131.

(38) Perkins, S. J. *Eur. J. Biochem.* **1986**, *157*, 169–180.

(39) Zamyatin, A. A. *Prog. Biophys. Mol. Biol.* **1972**, *24*, 109–123.

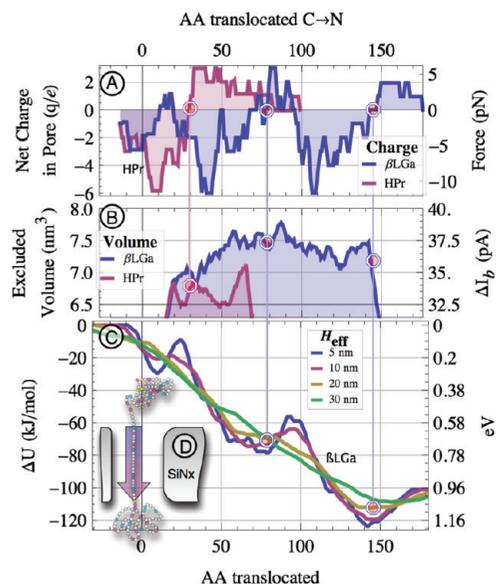


Figure 4. (A) The net charge at pH 7.0 of β GLa and HPr as predicted by treating the ionization of the individual amino acids as independent as a function of the number of residues translocated. (B) The calculated Λ assuming a contour length equal to $H_{\text{eff}} = 20$ nm for β GLa and HPr as a function of number of amino acids translocated. The octagons mark the location of stall points in the translocation. (C) The electrostatic contribution to the potential energy of β GLa as a function of the number of amino acids translocated through the nanopore for various values of H_{eff} as labeled in the figure. This potential is for the translocation of the C terminus first. (D) A schematic of the linear translocation geometry for β GLa with positive and negative residues colored blue and red, respectively.

of the number of amino acids translocated (0.38 nm³/residue). The values predicted by this analysis ($\Lambda^{\text{HPr}} = 6.75$ nm³, $\Lambda^{\beta\text{GLa}} = 7.48$ nm³) match quantitatively with the values of $\Lambda^{\text{HPr}} = 6.8 \pm 0.1$ nm³ and $\Lambda^{\beta\text{GLa}} = 7.5 \pm 0.2$ nm³ observed in Figure 3 for the smallest-volume contribution to the HPr and β GLa event distributions. The observed differences in excluded volume arise because β GLa has bulkier amino acids and therefore a larger excluded volume for a segment inside the nanopore. The variability of the β GLa translocation volume is also bigger, consistent with observations. This suggests that the cluster 1 translocation events were dominated by protein that was essentially linear inside the pore (Figure 4D), and the folded or globular protein translocation could account for no more than 20% of the events in 2 M KCl even when no urea is present (Figure 2A).

We have found that at 5 M urea β GLa forms disulfide-linked dimers.²⁹ Cluster 3 in Figure 2B is consistent with partially unfolded disulfide-linked dimers. The broad distribution of sojourn times suggests that the translocating dimer is not in the folded state. Cross-linking at 5 M urea is consistent with the observation that disulfide cross-linking does not occur when β GLa is folded.²⁹

3.2. Protein Charge Sequence and Stalling. In the β GLa nanopore translocation measurement, we have observed low displaced volume values and very slow translocation times. One explanation for the observations is that the protein is being unfolded and pulled through as a linear polymer. For linear translocation, the primary sequence of charged amino acids will be more important than that the net charge of the protein.

Figure 4A shows the calculated charge and force of β GLa inside the pore as a function of the number of residues translocated for $\Psi = 120$ mV and $H_{\text{eff}} = 20$ nm. This analysis predicted several locations where the net force is zero, sug-

gesting stalling points or locations of metastability. Two positive peaks in the curve suggest that *at two points during the linear translocation the electrophoretic force opposes translocation.* This can be more readily visualized in the translocation potential plot of Figure 4C.

The presence of multiple stall points provides an explanation for the anomalous t_d results when fitting to eq 2; eq 2 is only valid in the absence of energetic barriers. If the barriers are large enough, Kramers reaction rate theory⁴⁰ predicts that the distribution of sojourn times should be multiexponential according to the number of barriers present. One or two exponential contributions were adequate in all cases, consistent with the presence of no more than two activated barrier crossings during translocation.

Since the barrier at $\Psi = 60$ mV is smaller compared to $\Psi = 120$ mV, a shorter t_d is predicted by the thermally activated barrier model. The peak values of t_d for β GLa appear to decrease as the voltage was decreased from $\Psi = 120$ mV (Figure 2A) to $\Psi = 60$ mV (Figure 2D) instead of increase, as predicted by voltage driven translocation. This also supports the thermally activated barrier model of the sojourn time distribution.

The presence of positive charges at the N terminus and negative charges at the C terminus suggests that the insertion of β GLa into the nanopore will favor the C-terminus. The dipole of β GLa has been experimentally measured to be ~ 700 D.⁴¹ Using PDB entry 2akq, a computational estimate for the monomeric dipole is 795 D⁴² and is approximately oriented from the C-terminus to the N-terminus of the folded protein. This also suggests that dipolar orientation of the intact monomer would favor C-terminal-first insertion. An expansion of this analysis of the electric field orientation of β GLa appears in the Supporting Information.

Figure 4C shows that the barriers during translocation are maximized for thin pores and become smoother for thicker pores. The pattern of translocation barriers is highly dependent on the sequence of charged residues. These observations are in sharp contrast to DNA translocation. When calculating the driving force for DNA translocation, the thickness of the pore cancels out because of the uniform charge density. There is essentially no DNA sequence effect on the translocation driving force. By contrast, our analysis suggests that the translocation of proteins is highly sequence dependent. The distribution of translocation times depends strongly on protein sequence, applied voltage, and H_{eff} .

We have neglected two potentially important contributions to the translocation potential (Figure 4C): entropy and folding. β GLa is unfolded in 8 M urea. Random coil polymers are expected to resist confinement. The protein must give up conformational entropy to enter the confined region of the nanopore. If the unfolded state is behaving as a random coil, then entropic considerations would suggest a barrier centered at \sim residue 81 where the chain entropy would be most-reduced by the confinement of a region in the nanopore.

For a folded protein, both partial unfolding at the *cis* chamber side and refolding at the *trans* chamber side could create additional energetic contributions to translocation.

3.3. The Electrical Force Exerted on a β GLa may Partially Unfold it. Our observations are consistent with bulk measurements that showed that β GLa protein is completely

(40) Kramers, H. A. *Physica (Utrecht)* **1940**, *7*, 284–304.

(41) Ferry, J. D.; Oncley, J. L. *J. Am. Chem. Soc.* **1941**, *63*, 272–278.

(42) Felder, C. E.; Prilusky, J.; Silman, I.; Sussman, J. L. *Nucleic Acids Res.* **2007**, *35*, W512–W521.

denatured in 8 M urea and partially denatured in 5 M urea.^{28–30} Looped translocation geometries for the unfolded protein (8 M urea) suggest some persistence of structure. In 5 M urea, oxidative dimerization of β LGa (peak 3) starts as soon as the sample solution was made. The presence of unfolded translocation at 0 M urea suggests electric-field induced unfolding of β LGa.

Research on force unfolding of proteins has shown some proteins rupture at elongation forces larger than ~ 5 pN.^{43,44} The electrical field strength in the nanopores is $E \approx \Psi/H_{\text{eff}} = 6 \times 10^4$ V/cm; this electric field is at least 2 orders of magnitude larger than it is in a regular electrophoresis experiment. The force per charge estimated is $eE = 1.0$ pN/e. β LGa has 27 negative and 18 positive residues at pH 7.0. When a β LGa molecule is entering the nanopore, opposite charges in the protein will be driven in opposite directions by the electric field (see Figure 4A.) This electric force is much larger than the strength of the hydrogen bonds that hold a protein in a folded shape; thus some of the hydrogen bonds in a β LGa could be broken at the entrance of a nanopore, at least partially denaturing it. The increased population of cluster 2 events (folded or looped β LGa) at lower voltage, $\sim 50\%$ at $\Psi = 60$ mV ($\sim 20\%$ at $\Psi = 120$ mV), supports our force unfolding hypothesis above since the weaker electrical field strength provides less driving force to unfold a β LGa protein.

Figure 4C predicts that a thicker nanopore would manifest smaller electrostatic energy barriers to linear translocation and smaller t_d as a result. A thicker nanopore would have more amino acids present at the stall point. For a nanopore of a given thickness, this can also predict the effect of the protein being less than fully extended near the stall points. That is, the effective number of amino acids in the pore would be larger. Figure 4C shows that, at a stall point, the energy is higher if more amino acids are in the pore. Therefore increasing the extension would be energetically favorable near the stall points. This suggests that proteins at a stall point will tend to elongate to their most extended form.

This argument neglects entropy. Elongation in the pore would decrease the entropy of the segment inside the pore but increase the entropy of segments outside the pore. The net entropic effect will depend, to first approximation, on the translocation position through its effect on the number of residues in the different regions of the nanopore apparatus. Thus different stall points may have different entropic contributions to the free energy.

3.4. pH Effects on Translocation. Several changes in the 2D event distribution were observed at pH 4.6 as compared to pH 7.0. The principal effect of reducing the pH is the partial protonation of carboxylic acid amino acid side chains. At pH 4.6 the acidic residues are approximately 50% protonated. This substantially changes the distribution of stall points and the dipole moment of the protein. The polarity of translocation events is observed to be opposite of those at pH 7.0 in accordance with expectations. Because there are fewer total charges on the protein, the driving force for electrostatic unfolding is expected to be reduced. Overall the distribution of cluster 1 translocation events is peaked earlier in t_d , consistent with smaller electrostatic barriers to translocation. The distribution of events appears to favor larger excluded volumes,

suggesting a greater contribution from the looped translocation structure. This is consistent with a reduced driving force for electrostatic extension of the protein. The increase in translocation time for cluster 2 and 3 events may be due to the specific arrangement of charges in the loops of translocating protein or due to other effects. For example, since the protonation of the carboxylic acid is near the midpoint of the equilibrium, any electric-field effect on the pK_a of the side groups could substantially alter the electrostatics of the translocating protein. Resolution of these effects awaits future investigation.

3.5. Evaluation of Bumping/Collision Hypothesis. The shallow blockage shown by the proteins is consistent with $\sim 30\%$ of the protein volume contributing to the blocked current. One hypothesis is that a globular β LGa molecule only partially entered the pore and then went back to the *cis* chamber. The cluster at the lowest current blockage values, therefore, might be expected to be due to collisions. In DNA experiments, we observed short spikes with small ΔI_b and t_d ; they usually appear at the left bottom corner in an event distribution plot. We attributed this type of event to long DNA molecules ($\sim \mu\text{m}$) that were captured in the middle of the long chain rather than at the end. The DNA must then be bent at the entrance by the electric field. If the bending failed, the DNA would be bounced back to the *cis* chamber.

In this work, β LGa and HPr proteins were smaller than the size of nanopores used, and short events did not appear in the histogram. For the cluster 1 events to be due to globular protein that partially entered a pore, persisted for the long sojourn times, and then return to the *cis* chamber, there must be some trap for the protein in the vicinity of the nanopore. Otherwise, diffusion would normally carry the protein away far more rapidly than the time scales we observed in the sojourn time distributions. Given the net charge of a β LGa protein ($-9e$), and the large electrical field strength in the nanopore, a trap deep enough to localize globular β LGa protein for the times observed is unlikely.

3.6. Protein Adsorption. An alternative interpretation of the presence of the multiple time scales for translocation is that proteins adsorb to the nanopore surface causing a change in current. The unusually low translocation volume would then arise because, for the vast majority of the translocation time, the protein is only measured when in the surface layer of the nanopore. This hypothesis cannot easily account for the similarity of the β GLa translocation volumes and the HPr translocation volumes. Though the calculation of the net effect on the current of a protein adsorbed to the nanopore is not well understood, one would expect the effect to be approximately in proportion to either the volume displaced or the surface area contacted on the nanopore. In either of these cases we expect the difference between HPr and β GLa to be larger than we observed. The translocation time distribution would be dominated by the desorption reaction. To explain the observed translocation time distributions multiple adsorption modes would be required with different desorption rates. The protein is not expected to adsorb to the SiN_x surface in the presence of 8 M urea. Protein adsorption to the surface of the nanopore/membrane may be involved in the long-time changes (\sim hours) of the nanopore electrical properties; however it does not appear to be a likely explanation for the observed translocation signals.

3.7. Translocation Event Heterogeneity and Protein Sequence. Linear translocation of dsDNA was much less heterogeneous than that of proteins. In this section we discuss

(43) Lellermayer, M. S. Z.; Smith, S. B.; Granzer, H. L.; Bustamante, C. *Science* **1997**, *276*, 1112.

(44) Bechtluft, P.; Leeuwen, R. G. H. v.; Tyeman, M.; Tomkiewicz, D.; Nouwen, N.; Tepper, H. L.; Driessen, A. J. M.; Tans, S. J. *Science* **2007**, *318*, 1458–1461.

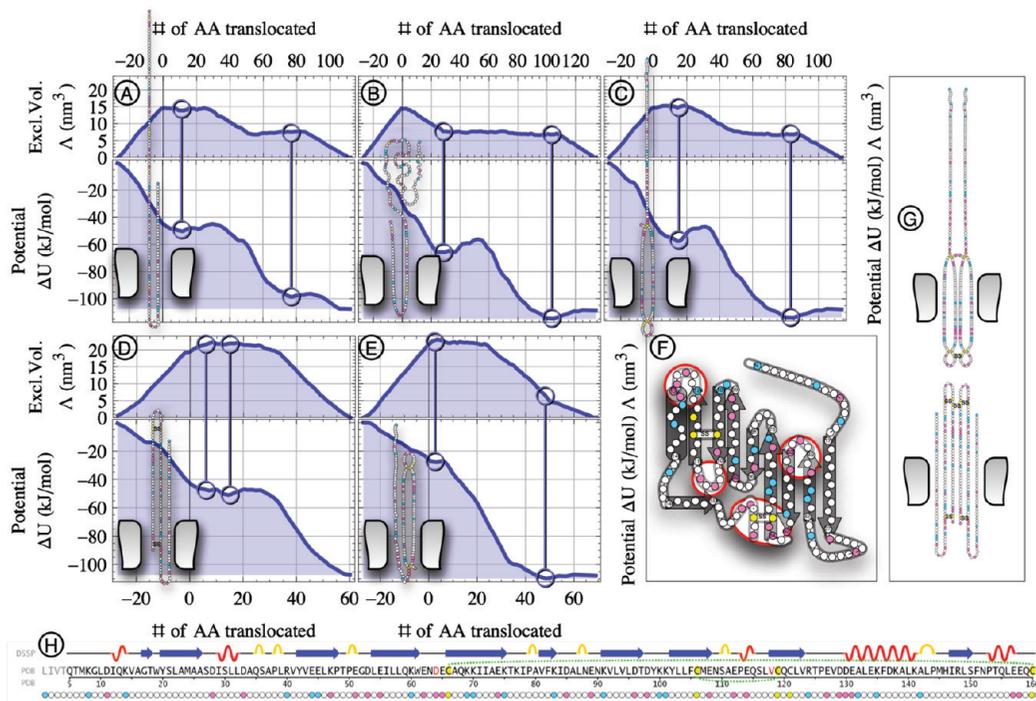


Figure 5. (A, B) Possible looped translocation geometries for the unfolded protein with no disulfide bonds intact. These loops are performed in the native structure as shown in Figure 1. (A) The E51 loop insertion has two stall locations with different excluded volumes. (B) Two stall points for the L133 loop occur at similar excluded volumes. (C, D, E, G) Possible translocation geometries for the partially unfolded protein with native disulfide bonds intact. (C) Two stall points for the C106–C119 loop occur at different excluded volumes. (D) The C66–C161 insertion has only a shallow stall point near the exit and is expected to translocate more rapidly. (E) If the negatively charged loop at E51 inserts first the result is a broad stall with translocation volume close to that of the full protein. (F) Diagram of β LgA showing the location and relationship of secondary structural elements. The small circles show the charge of residues as red for negative and blue for positive. The large red circles highlight clusters of negatively charged (at pH 7.0) amino acids present on turns at the surface of the protein that may be “hooks” for unfolding translocation with a positively biased trans chamber. Native disulfide bonds between C106–C119 and C66–C161 are shown in yellow. (G) Disulfide linked dimers will have more complicated translocation patterns with peak volumes ranging from 4–6 strands. These translocation diagrams were calculated for a $H_{\text{eff}} = 20$ nm pore. (H) The primary sequence of β LgA.

how the increase in translocation event heterogeneity can be related to the details of the protein sequence.

3.7.1. Primary Sequence Effects. The individual amino acids vary substantially in their partial molecular volumes. This difference was quantitatively observed in comparing unfolded β LgA and HPr. This suggests that nanopores could, in principle, distinguish proteins based on the primary sequence variability of the excluded volume and the location of stall points.

3.7.2. Loops. Highly charged loops present on the surface of β LgA were identified based on the folded structure and the topological constraints imposed by disulfide bonds. These putative structures are a possible explanation for the range of excluded volumes observed. Two mechanisms could make β LgA be looped during translocation. (1) Unreduced native disulfide bridges will enforce the presence of a loop. (2) Nanopores could preferentially capture negatively charged turns/clusters on the surface of β LgA as shown by the red circles in Figure 5F. As with translocation of linearized protein, the sequence details of the loop inserted into the nanopore will dictate the dynamic nature of the signal measured. The translocation potentials and excluded volume profiles for several likely loops appear in Figure 5. Figure 5A shows the capture of the loop at residue 133. Stall points are predicted with $\Lambda^{\beta\text{LgA}} = 14$ and 8 nm^3 . The measured current (and therefore the excluded volume) of such an event would depend on the time-weighted mean of the volumes Λ_1 and Λ_2 at the two stall points:

$$\langle \Lambda \rangle = \int \int P(t_1) P(t_2) \frac{t_1 \Lambda_1 + t_2 \Lambda_2}{t_1 + t_2} dt_1 dt_2 \quad (3)$$

$P(t_i)$ is the distribution of stall times at each metastable position in the translocation. A close examination of individual events shows that many of them have multiple levels (Figure 1B, 2D, 2E, 3). Since the present analysis did not attempt to distinguish substeps from noise, only the average excluded volume was reported. Figure 5B shows the capture of a loop at residue 51. Both stall points give similar translocation volumes. Events like these would be expected to give similar volumes to the completely linear translocation. Figure 5C shows the capture of a cluster at residue 113. The native disulfide bridges are compatible with this translocation, and two stall points are predicted with $\Lambda^{\beta\text{LgA}} = 15$ and 8 nm^3 similar to Figure 5A. These events are consistent with the larger excluded volume contributions to cluster 1 in the β LgA event distributions. The upper edge of the cluster 1 $\Lambda^{\beta\text{LgA}} \approx 15 \text{ nm}^3$, and the lower edge is $\Lambda^{\beta\text{LgA}} \approx 7.5 \text{ nm}^3$. The time at each stall point will produce a range of observed average volumes between these two limits according to eq 3, consistent with our observed distribution. Figure 5D shows the capture of a charged cluster of amino acids at residue 51 and the resulting loops that are present due to intact native disulfide bonds. Figure 5E shows the capture of a loop at residue 63 that is favored by native disulfide bonds. These geometries are examples of how partially folded translocation events can give rise to the displaced volumes of 17–20 nm^3 observed in cluster 2 of the 0 and 5 M urea experiments. Figure 5F shows disulfide linked dimers will have more complicated translocation patterns with peak volume contributions from 4–6 strands, consistent with the events in cluster 3 in Figure 2D. Dimers formed at 5 M urea are likely to be one

of the early structures formed when β LGa is incubated under amyloidogenic conditions, though formation of higher order oxidative aggregates is relatively unimportant.²⁹

3.8. Correlation between ΔI_b and t_d . The various translocation events illustrated in Figure 5 not only provide an explanation for the heterogeneity but also provide a mechanism for correlation between ΔI_b and the charge sequence of a protein which determines t_d . The presence of multiple translocation shapes provides multiple clusters of events providing an overall correlation between ΔI_b and t_d . However, the presence of multiple stall points with different excluded volumes suggests a mechanism of correlation between ΔI_b and t_d within a single cluster through eq 3. As the time spent in the deeper well increases, the mean will become increasingly weighted by the volume at that position.

When a nonflexible particle translocates, the current drop amplitude and sojourn time are independent random variables. There should be no correlation between ΔI_b and t_d . Correlation between ΔI_b and t_d requires that there be multiple species with different electrophoretic mobilities and excluded volumes. For example, a shorter t_d represents partially folded DNA with larger ΔI_b (Figure 2F). In the present case correlation can only arise due to conformation changes occurring in the protein either prior to or during the translocation process. These changes must affect both the electrophoretic mobility and the excluded volume of the protein.

We discuss two possible explanations for the presence of multiple current blockage levels in a single transient. One possibility is the simultaneous translocation of multiple molecules that reside in the pore for different lengths of time. In this case there could be up to three levels observed corresponding to the excluded volume of each molecule alone and the molecules together. The other possibility is that a single molecule could show multiple current blockage values. We find the second explanation more compelling since the presence of steps in the current blockage profile depended on the character of the molecule being translocated and not on the concentration of protein. Furthermore, the translocation of a protein as an unfolded chain leads to multiple translocation current levels in a single event.

3.9. Comparison with Polynucleotide Translocation. The average current blockage value, ΔI_b , is a time-averaged quantity. This average depends on the amount of time the translocating molecule spends at each different value of excluded volume during the translocation. A significant difference between polynucleotide translocation and polypeptide translocation is the degree of variability inherent in the polymer. The charge variability of proteins can result in a variable translocation driving force as a function of displacement, an effect not important to polynucleotide translocation. The excluded volume of amino acids varies by a factor of 3.8, whereas that of dsDNA (or ssDNA) varies by less than 2% (or 10%). This variability in volume maps directly to variability in ΔI_b . Furthermore, proteins are structurally more complex than DNA. The persistence length of DNA is much longer than that of proteins suggesting that proteins are more likely to loop. Induced and persistent tertiary and secondary structures in proteins will increase the heterogeneity of protein translocation signals in much the same way as is observed for polynucleotides. However, for proteins, the importance and variety of such contributions is expected and is observed to be greatly increased.

3.10. Single-Molecule Protein Unfolding. The electrical detection of single protein molecules in various states of folding

invites comparison to other single molecule approaches to folding such as fluorescence^{45,46} and force measurements.^{47,48} The different physical principles that underlie the different methods make them not directly comparable. The use of fluorescence energy transfer, for example, allows probing one or more folding coordinates.^{49–52} AFM studies of beta-sheet proteins have suggested that the direction of force applied can influence the rupture threshold of the protein.⁴⁸ The tendency for proteins with (large) net dipoles to orient in an electric field (see Supporting Information) would provide an optimal geometry for electric-field induced unfolding. The strong electric field in the nanopore and distribution of positive and negative charges along the sequence of the protein provides the driving force for unfolding.

One of the challenges of studying protein folding on the single molecule level remains being able to observe both the folded and unfolded states under a given set of experimental conditions. A significant benefit for protein folding studies using nanopores is the electrical signal that is synchronized with the translocation and thus unfolding process. Translocation of unfolded proteins by nanopores provides a Maxwell's Daemon-like synchronization signal that could be very useful for incorporation into fluorescence-electrical hybrid experiments.

The present study suggests that nanopore approaches to protein folding may have some advantages that are complementary to existing methods. The distribution of the current drop appears to contain information about the stable loops that are present in partially folded states. Sampling of these loops is at the electrostatic stall points. The nanopore experiment places structurally disrupting forces onto the protein according to its distributions of charges rather than at the location of a tether. This is either a benefit or a limitation depending on one's perspective

During translocation, the nanopore appears to perturb the distribution of protein conformation states in favor of extended states. The similarity of this to biological nanopore protein translocation may be important. However, the natural electrochemical potential across biological membranes often can produce electric fields comparable to those expected in synthetic nanopores. The commonplace translocation of unfolded proteins through natural nanopores suggests that folding following nanopore translocation may be a particularly physiologically relevant approach to single molecule protein unfolding/refolding studies. All single molecule approaches to single molecule protein folding require artificial introduction of an additional driving force to manipulate the distribution of folded and unfolded states. This can be any combination of solution conditions including chemical denaturants and/or the application of an external mechanical force. It is not well understood if these driving forces leave intact the free energy landscape upon

(45) Michalet, X.; Weiss, S.; Jger, M. *Chem. Rev.* **2006**, *106*, 1785–813.

(46) Schuler, B. *ChemPhysChem.* **2005**, *6*, 1206–1220.

(47) Borgia, A.; Williams, P.; Clarke, J. *Annu. Rev. Biochem.* **2008**, *77*, 101–25.

(48) Brockwell, D. J.; Paci, E.; Zinober, R. C.; Beddard, G. S.; Olmsted, P. D.; Smith, D. A.; Perham, R. N.; Radford, S. E. *Nat. Struct. Biol.* **2003**, *10*, 731–737.

(49) Deniz, A.; Laurence, T.; Dahan, M.; Chemla, D.; Schultz, P.; Weiss, S. *Annu. Rev. Phys. Chem.* **2001**, *52*, 233–53.

(50) Jia, Y.; Talaga, D. S.; Lau, W.; Lu, H.; DeGrado, W.; Hochstrasser, R. M. *Chem. Phys.* **1999**, *247*, 69–83.

(51) Schuler, B.; Eaton, W. A. *Curr. Opin. Struct. Biol.* **2008**, *18*, 16–26.

(52) Talaga, D. S.; Lau, W.; Roder, H.; Tang, J.; Jia, Y.; DeGrado, W. F.; Hochstrasser, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 13021–13026.

which the folding reaction occurs. The nanopore accomplishes this with an electrostatic potential that is comparable to those present across biological membranes.

Summary

We used SiN_x nanopores to examine the differences between folded, partially unfolded, and unfolded single protein molecules. β LGa translocations were heterogeneous showing multiple time scales and multiple translocation current blockages. The calibrated excluded volume was smaller in most cases than that expected for globular β LGa translocations. The sojourn times for β LGa translocations in all cases were 2–3 orders of magnitude longer than expected in folded globular states. This suggests that the general understanding of nanopore translocation that has been formulated based on polynucleotides needs to be modified for proteins.

Our analysis suggests that the events we measured are consistent with linear translocation and looped translocation of proteins even under folded conditions. Evaluating the potential as a function of linear translocation predicts that the protein will stall at different sequence positions during a translocation. This arises because of the heterogeneity of charge along the amino acid sequence, an effect not present in polynucleotides. Stalling at different locations can explain the long sojourn times, the heterogeneity in the current blockage histogram, and correlation between sojourn times and excluded volumes.

The results in this work have demonstrated that a solid-state nanopore sensor can be used to evaluate the folding state or shape of a protein by measuring its excluded volume. The excluded volume of the amino acids present at the stall points can vary enough to distinguish β LGa and HPr. Our experiments and analysis open the way to study the shape and sequence variability of single protein molecules in solid-state nanopores. Since translocation appears to entirely disrupt the calyx, further investigation is required to allow intact nanopore transport of β LGa-ligand complexes for nanobioremediation applications.

Acknowledgment. We thank Bradley Ledden and Ryan Rollings for nanopore preparation, Professor Jene A. Golovchenko for assisting with FIB hole fabrication, and Professor Jeremy S. Lee for the HPr protein. This work is supported by NIH R21HG003290 and NIH R01GM071684.

Supporting Information Available: The contents of Supporting Information include the following: Details of the protein and nanopore materials used; methods used to extract the event parameters, calibrate the size of the nanopore, correct for the effects of Bessel filtering, and calculate translocation profiles; detailed analysis and discussion of dipolar orientation effects in nanopores; and a derivation of the first passage time distribution for biased diffusion to a sink. This information is available free of charge via the Internet at <http://pubs.acs.org/>.

JA901088B

Supporting Information for :

“Single-molecule Protein Unfolding in Solid State Nanopores”

by David S. Talaga, Jiali Li

SI.1 MATERIALS

Nanopores used

Nanopores were fabricated in a free standing 280 nm thick low stress silicon nitride membrane supported by a 380 μm thick silicon substrate using a combination of focussed ion beam milling and feedback controlled ion beam sculpting^{1,2}. We found 4-10 nm diameter pores to be most effective for detecting small proteins like β -lactoglobulin. Depending on the samples and solution conditions used, these pores can be operated for many hours and even days in some cases before becoming very noisy or irreversibly blocked. TEM images show the projected view of the nanopore's perimeter. Since the diameter of the pore is not the same throughout its length, its precise contours cannot be known. Hence, the absolute value of the current blockages caused by identical DNA or protein molecules can vary from one nanopore to the next by as much as 20% even for similar diameters. Because of pore-to-pore variations in thickness, diameter, shape, or surface charge, a single nanopore was used for the full set of experiments reported in figure 2 of this work. Figure S1 shows the TEM image of the nanopore used for figure 2. A smaller nanopore was used for the experiments shown in figure 3. Pore-to-pore variability resulted in replicate experiments giving similar but not identical results.

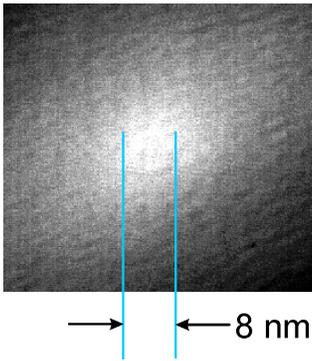


Figure S1. TEM image of the nanopore used for figure 2.

Molecules used and sample preparation.

Bovine β -lactoglobulin variant a (β LGa) was obtained from Sigma-Aldrich and used without further purification. Final protein concentrations were verified by UV-vis absorption. The concentration of β LGa in the cis chamber used in this study was 35 ± 15 nM. HPr (Histidine-containing Phosphocarrier Protein, 85 aa, 9120 dalton, $-2e$ at pH 7) was obtained from Jeremy S. Lee³. Linear dsDNA (2,706 base pairs, pNEB206A, New England Biolabs) was used as a standard to calibrate H_{eff} . All measurements were performed in 1M or 2M KCl containing 10 mM Tris or phosphate and 1 mM EDTA.

β LGa is monomeric in the cis chamber

β LGa is dimeric under physiological conditions (pH 7, $[\beta\text{LGa}] > 50 \mu\text{M}$)⁴ with a dissociation constant of $\sim 20 \mu\text{M}$. The dimer has semi-axes of 1.8 and 3.5 nm ($d_m = 3.6$ and 7 nm)⁵. Our measurements were performed at a very low concentration of β LGa protein, in the *cis* chamber it was ~ 30 nM, three orders of magnitude lower than the physiological condition. Dimer association has $K_{\text{eq}} = 5.36 \cdot 10^4 \text{ M}^{-1}$ giving $< 0.2\%$ dimer at the low concentrations of our experiments. Translocation is therefore expected to be dominated by monomeric events. The free energy of unfolding in the absence of urea is 21.1 kcal/mol (0.92 eV or 88 kJ/mol)⁶.

SI.2 METHODS

Analysis of current blockage events.

Current blockage events are characterized by their average current drop amplitude ΔI_b and their time duration t_d as illustrated in Fig 1. B. The current blockage events were recorded in event driven mode and then were analyzed using custom Matlab® routines. These routines include baseline correction, events classification, and calculation of ΔI_b and t_d . For events detection and classification, the start of a current blockage event was defined as one that caused the nanopore current to drop monotonically below two thresholds; the end of the event was signaled by the current trace climbing monotonically back to the open channel current past both of these thresholds. The event duration, t_d , was defined by the time between the current drop across the second threshold. The second threshold was set to be at 50% of the most probable peak value of the current blockages. The arithmetic mean of the current blockage value (ΔI_b) was calculated within the range between the crossings of the second threshold. Current blockage events with $t_d < 40 \mu s$ and $t_d > 6000 \mu s$ were not selected in this work. Blockage events with long rise times or large slopes between trig 1 and 2 were also filtered out.

Calibration of the effective thickness H_{eff} using DNA

The calibration procedure was based on using known excluded atomic volumes of dsDNA molecule to calibrate the parameter H_{eff} . The value for the DNA cross-sectional area we use is not based on the crystallographic edge-to-edge distance of 2.1 nm. We used the data in Nadassy 2001⁷ to determine a mean excluded volume per unit length of dsDNA. This has the same dimensions as the cross-sectional area, but is more applicable to our excluded volume calibration procedure. One significant difference is that the volume of the grooves is not included. Similar data from Perkins1986⁸ is used to provide volume estimates for the proteins to compare to the experimental nanopore-derived excluded volumes. In this way we are comparing similar quantities between the dsDNA and the proteins.

The formula used for the excluded volume was $\Lambda(t) \approx (\Delta I_b(t) * H_{eff}^2) / (\sigma \psi)$. For the calibration of figure 2 the parameters used were $\sigma = 0.115 / (\Omega \cdot cm) = 1.1 \times 10^{-8} / (\Omega \cdot nm)$, $\psi = 0.12 V$, and $H_{eff} = 20 nm$ as determined below.

1. Using the Radical Planes method from Table 3 and 4 in Nadassy 2001⁷ the atomic volume of AT and GC pairs are $\Lambda (G+C)=0.6066 \text{ nm}^3$ and $\Lambda (A+T)=0.6183 \text{ nm}^3$. The quotient of the average volume of a base pair $\Lambda_{\text{bp}}=0.61245 \text{ nm}^3$ and the rise in the helix per base pair, 0.34 nm, give the excluded volume per unit length, $A_{\text{DNA}}=0.61245 \text{ nm}^3 / (0.34 \text{ nm}) = 1.8 \text{ nm}^2$. Using the mode of the corrected dsDNA current blockage histogram gives a calibration of $H_{\text{eff}}=20 \text{ nm}$.

2. Using the numbers from the paper of Zwolak⁹ below for the volume per Nucleotide, in cubic nanometers $\Lambda_n=0.349, 0.359, 0.324,$ and 0.339 for A, G, C, and T.⁹ The average volume for a base pair of DNA: $\Lambda_{\text{bp}}=0.6855 \text{ nm}^3 = A_{\text{DNA}} \times 0.34 \text{ nm}$. Using these numbers, the $A_{\text{DNA}}=2.02 \text{ nm}^2$. Using the mode of the corrected dsDNA current blockage histogram gives a calibration of $H_{\text{eff}}=22 \text{ nm}$.

The modest difference in DNA volume estimates suggests that a $\sim 11\%$ systematic error may be present in the calibration of the excluded volume. On a practical level this difference is negligible since the difference in calibrated H_{eff} only changes the number of amino acids that must be included in calculating the volume in the pore. Since the systematic error is the same for all translocations on a given nanopore, the data for a set of experiments is directly comparable.

Calibration of 10kHz Low-Pass Bessel-Filter Response

Ionic current signal through solid state nanopores was measured and recorded using an integrated Axopatch 200B patch-clamp amplifier system (Molecular devices) in resistive feedback mode. The 10 kHz low pass Bessel filter in the Axopatch 200B was selected for some measurements in our work. The filter is implemented as an analog circuit, therefore to determine its influence on our measurements we performed a series of control experiments. The whole measurement system was tested and calibrated with artificial current drops generated by ideal square pulses from a function generator (Agilent 33250A). The calibration signal and analysis appears in Figure S2 below.

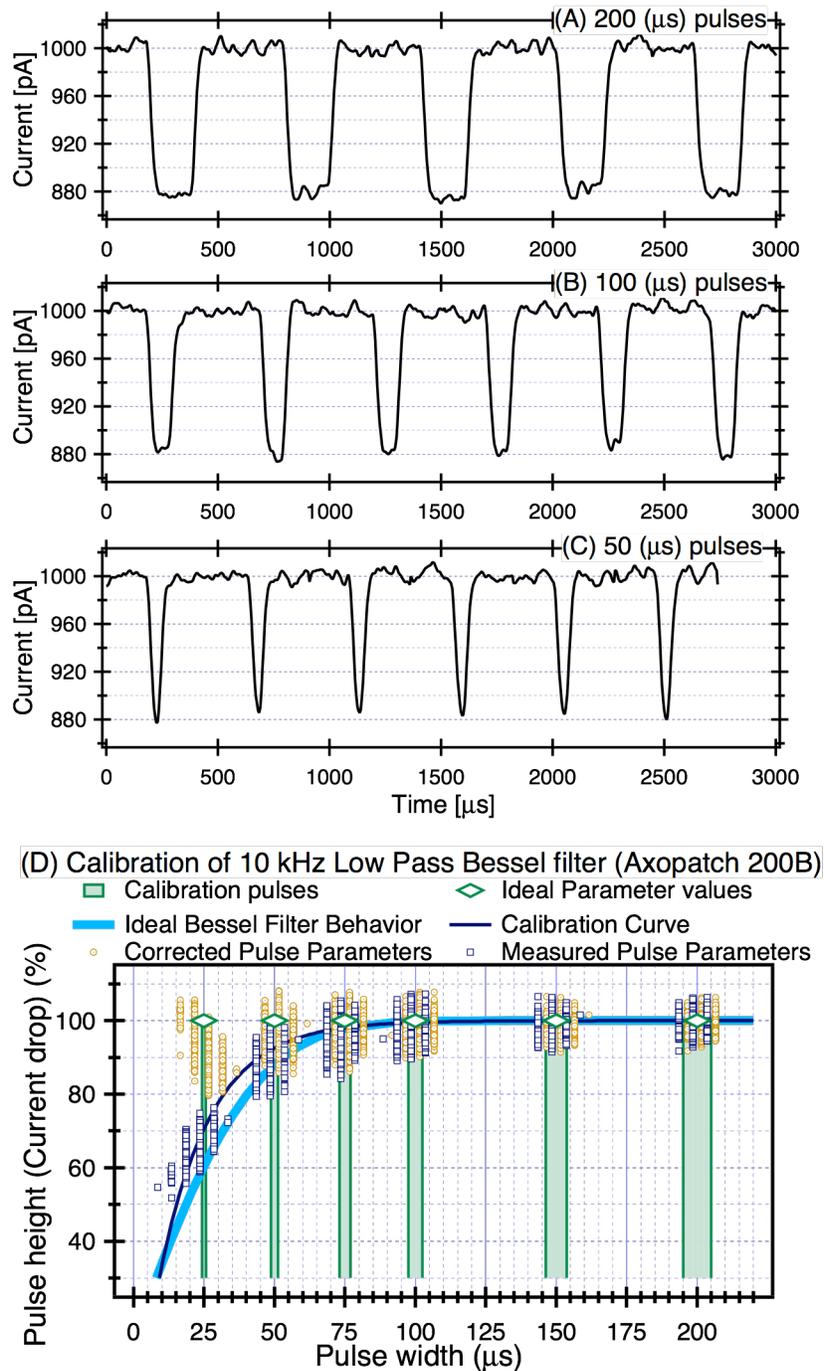


Figure S2:

The ideal pulse widths generated were 25, 50, 75, 100, 150, 200, and 300 μsec. The pulse height was about 120 pA with a rise time of 5 ps. Examples of the 10 kHz filter responses to these pulses recorded are shown in (S2A) for 200 μs, (S2B) for 100 μs, and (S2C) for 50 μs pulses. The recorded artificial current drop events was analyzed with the same MatLab routines used for analyzing real DNA and

protein current drop signals. The measured mean pulse heights ΔI_b (averaged between the two points that the t_d was calculated) versus the pulse widths t_d are plotted in the figure (S2D, blue squares).

This calibration shows that when the pulse width is less than 100 μsec , the calculated mean pulse height will be attenuated, but the time durations (the width of half height) remain correct up to 25 μsec pulses. When the time duration measured in our work was less than 100 μs , the current blockage amplitude can be corrected (S2D, tan circles) with this calibration (thin dark blue line) as shown.

The filtered pulse shape for an n -pole low pass Bessel filter at frequency ω for an ideal square pulse of width τ is:

$$\mathcal{F}_s^{-1} \left[\frac{((-1)^n 2^{2n} \Gamma(n + \frac{1}{2})) e^{\frac{is}{\omega}} (a \tau \text{sinc}(\frac{s\tau}{2}))}{\sqrt{\pi} \Gamma(n + 1) L_n^{-2n-1}(\frac{2is}{\omega})} \right] (t)$$

Where F^{-1} is the inverse Fourier transform. Using $t=0$ in this equation gives the peak intensity value of a square pulse after being filtered which is shown as the light blue line in Figure S2 above.

Calculation of translocation profiles

The charge on the segment of protein being translocated was calculated with a contour length of 0.38 nm per residue and adding the charge contribution for each amino acid based on its average ionization state determined by the acid-based equilibrium using the pKa data of Tanford¹⁰. Translocation volumes were similarly calculated using the consensus volume estimate for each amino acid published in Perkins1986⁸. The translocation potential was calculated by integrating the charge over the electric field present assuming that the electrostatic potential only changes within the confines of the pore, i.e. the bulk solution behaves as a conductor. Programs were implemented in Mathematica 6.

SI.3 SUPPORTING ANALYSIS

Dipole Barriers to translocation

If β LGa translocates as a folded protein, we can predict the observed translocation signals by treating β LGa as a charged ellipsoid. At pH 7.0 β LGa has an electric monopole of approximately $q=9e$ and an electric dipole of approximately 700 Debye.

Dipolar orientations.

The dipolar energy is $\Delta U_d = \vec{\mu} \cdot \vec{\mathcal{E}} = \mu \mathcal{E} \cos(\theta)$ where θ is the angle between the dipole and the nanopore axis. This quantity must be statistically averaged over the spherical angles. The probability density of a particular orientation is:

$$\mathcal{P}(\phi, \theta) = \frac{\mathcal{E}\mu}{4\pi kT} \operatorname{csch}\left(\frac{\mathcal{E}\mu}{kT}\right) e^{\frac{\mathcal{E}\mu \cos(\theta)}{kT}} \sin(\theta)$$

ΔS due to dipolar orientation.

Comparing this distribution to the expected unbiased distribution in the *cis* chamber allows calculation of the entropic barrier to entering the nanopore due to orientation of the protein. $\Delta S = -((8.2 \text{ J})/(\text{K mol}))$. Using the distribution above to evaluate the expectation value for the energy gives: $\Delta U = -((14.437 \text{ kJ})/\text{mol})$. Giving a free energy contribution from the dipole of $\Delta G = -((12.0 \text{ kJ})/\text{mol})$. This is quite a bit above thermal energy of 2.4 kJ/mol. Also it is above the 3kT of energy partitioned into translations and rotations suggesting that β LGa should be oriented and directed through the nanopore opening by this effect. The expected average monomer orientation is ± 22 degrees about the nanopore axis for $H_{\text{eff}}=10$ nm. (See the blue curve in Figure S3)

Boltzmann Distribution of dipoles

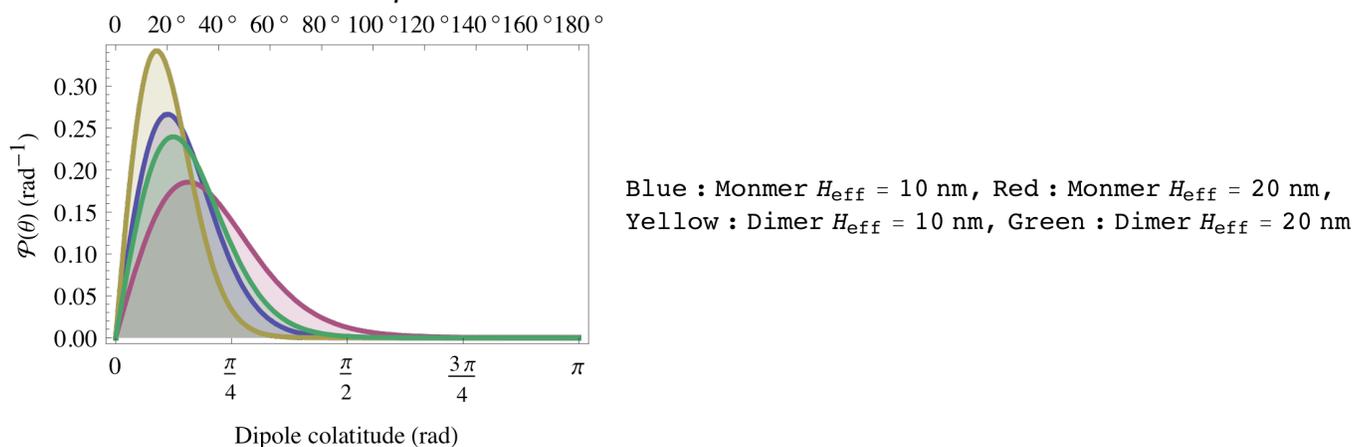


Figure S3. Boltzmann Distribution of dipoles. Yellow (dimer dipole $H_{\text{eff}} = 10$ nm), Blue (monomer dipole $H_{\text{eff}} = 10$ nm), Green (dimer dipole $H_{\text{eff}} = 20$ nm), Red (monomer dipole $H_{\text{eff}} = 20$ nm).

Exit barrier due to dipolar orientation

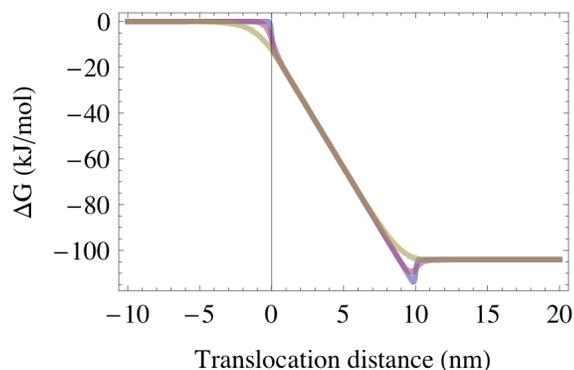


Figure S4. The free energy vs translocation distance for hard spheres of different effective radii. The barrier is quite apparent when the protein is treated as a point charge and dipole (blue curve), If the charges are spread over a diameter of 1 nm, the barrier softens considerably (red curve). Once the true size of the protein is included, the barrier disappears (yellow curve).

Translational entropy can be neglected for a single particle once it is in the nanopore. The presence of a large dipole on β LGa in the folded state provides an additional source of energy in the presence of the electric field. The dipole energy is lost upon exiting the field present in the nanopore. In principle this could create a barrier to exiting the nanopore and could account for the anomalously long translocation

times. Figure S4 shows a plot of the free energy vs translocation distance for hard spheres of different effective radii. The dipole is carried by charges that are spread over the protein. This substantially softens the barrier as illustrated in Figure S4. Given the relative sizes of β LGa and the nanopore, we do not expect dipole orientation to cause a significant barrier to translocation.

Native state dipolar contribution to unfolding force

If the dipole comes from net charges on the surface it is consistent with about 4 charges (i.e. 2 positive, 2 negative) separated by 3.6 nm. This dipole moment would stabilize the folding structure of a β LGa molecule without the presence of an electric field. However, in a 120 mV biasing potential across a 20 nm pore, the net force to pull the protein apart would be 8 pN due to the dipole alone. This dipole will increase if the protein deforms as a result of the applied electric field, further increasing the pulling force.

SI.4 Derivation of Biased Diffusion First Passage Time Distribution

Here we derive the first passage time distribution for a charged particle that has electrophoretic mobility u and a diffusion constant D that is located initially at position 0 to travel to a sink (trans chamber) located a distance H_{eff} away when driven by an electric field $\mathcal{E} = \psi/H_{\text{eff}}$. The drifting speed of the particle is $V_d = u_e \mathcal{E}$.

We start with the Fokker-Planck equation for the evolution of the position-time probability distribution function. We solve the 1D Smoluchowski type diffusion equation below.

$$\frac{\partial \mathcal{P}(x, t)}{\partial t} = u_e \mathcal{E} \frac{\partial \mathcal{P}(x, t)}{\partial x} + D \frac{\partial^2 \mathcal{P}(x, t)}{\partial x^2}$$

This biased diffusion model explicitly includes linear diffusion along the direction of translocation. The transmission rate for diffusing out of the pore against the potential bias is very small for experimentally relevant bias potentials. This informs the initial and boundary conditions for the problem.

A general normalized solution with initial condition $P(x,0)=\delta(x)$ is

$$\mathcal{P}(x,t) = \frac{e^{-\frac{(x-u_e \mathcal{E}t)^2}{4\mathcal{D}t}}}{\sqrt{4\pi\mathcal{D}t}}$$

To account for the boundary condition at the exit of the nanopore, $P(d_{\text{trans}},t)=0$ —we consider the exit to be a sink beyond which the particle cannot return (absorbing boundary)—we introduce an image sink of amplitude A at position x_0 :

$$\mathcal{P}(x,t) = \frac{e^{-\frac{(x-u_e \mathcal{E}t)^2}{4\mathcal{D}t}}}{\sqrt{4\pi\mathcal{D}t}} - \frac{Ae^{-\frac{(x-u_e \mathcal{E}t-x_0)^2}{4\mathcal{D}t}}}{\sqrt{4\pi\mathcal{D}t}}.$$

The solution that satisfies this boundary condition is

$$\mathcal{P}(x,t) = \sqrt{\frac{1}{4\pi t\mathcal{D}}} \left(e^{-\frac{(x-t\mathcal{E}u_e)^2}{4t\mathcal{D}}} - e^{-\frac{(-2d+x+t\mathcal{E}u_e)^2}{4t\mathcal{D}}} \right)$$

The probability that a particle has *not* translocated (reached the boundary d , survived) is

$$\mathcal{CDF}(t) = \int_{-\infty}^d \sqrt{\frac{1}{4\pi t\mathcal{D}}} \left(e^{-\frac{(x-t\mathcal{E}u_e)^2}{4t\mathcal{D}}} - e^{-\frac{(-2d+x+t\mathcal{E}u_e)^2}{4t\mathcal{D}}} \right) dx$$

$$\mathcal{CDF}(t) = \text{erf} \left(\frac{d-t\mathcal{E}u_e}{2\sqrt{t\mathcal{D}}} \right).$$

The probability that a particle has translocated (reached the boundary and left the system) is

$$1 - \mathcal{CDF}(t) = \text{erfc} \left(\frac{d-t\mathcal{E}u_e}{2\sqrt{t\mathcal{D}}} \right)$$

Substituting the definition of the drift velocity $v = \mathcal{E} u_e$, the probability density function of the particle reaching the boundary in a given time t is

$$\mathcal{P}_{\text{fpt}}(t) = \frac{\partial}{\partial t} \text{erfc} \left(\frac{d-tv}{2\sqrt{t\mathcal{D}}} \right) = \frac{e^{-\frac{(d-tv)^2}{4t\mathcal{D}}} (d+tv)}{t\sqrt{4\pi t\mathcal{D}}}.$$

This is the sojourn time distribution. d is the distance to be translocated. For a particle that is small with respect to H_{eff} d is just H_{eff} . For a long polymer like DNA d is the contour length of the polymer plus the length of the pore.

To account for any experimentally observable "backing up" we would need to invoke additional physics in the models. One mechanism that could provide a driving force for "backing up" is a persistent change in ionization state during translocation. However, we do not currently observe any experimental data that compels more complicated models.

References

- (1) Li, J.; Stein, D.; McMullan, C.; Branton, D.; Aziz, M. J.; Golovchenko, J. A. *Nature* **2001**, *412*, 166-169.
- (2) Stein, D. M.; McMullan, C. J.; Jiali Li; Golovchenko, J. A. *Review of Scientific Instruments* **2004**, *75*, 900-905.
- (3) Jia, Z.; Quail, J. W.; Waygood, E. B.; Delbaeresn, L. T. J. *The Journal of Biological Chemistry* **1993**, *268*, 22490-22501.
- (4) Gottschalk, M.; Nilsson, H.; Roos, H.; Halle, B. *Protein Science* **2003**, *12*, 2404-2411.
- (5) Piazza, R.; Iacopini, S.; Galliano, M. *Europhysics Letter* **2002**, *59*, 149-154.
- (6) Yagi, M.; Sakurai, k.; Kalidas, C.; Batt, C. A.; Goto, Y. *The Journal of Biological Chemistry* **2003**, *278*, 47009-47015.
- (7) Nadassy, K.; Tomás-Oliveira, I.; Alberts, I.; Janin, J.; Wodak, S. J. *Nucleic Acids Research* **2001**, *29*, 3362-3376.
- (8) Perkins, S. J. *European Journal of Biochemistry* **1986**, *157*, 169-180.
- (9) Zwolak, M.; Ventra, M. D. *Rev. Mod. Phys.* **2008**, *80*.
- (10) Tanford, C.; Swanson, S. A.; Shore, W. S. *Journal of American Chemical Society* **1955**, *77*, 6414-6421.