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 silicone 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.

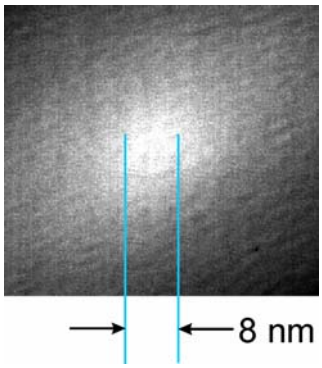


Figure S1. TEM image of the nanopore used for figure 2.

Molecules used and sample preparation.

Bovine β-lactoglobulin variant a (βLGa) was obtained from Sigma-Aldrich and used without further purification. Final protein concentrations were verified by UV-vis absorption. The concentration of βLGa in the cis chamber used in this study was 35±15 nM. HPr (Histidine-containing Phosphocarrier Protein, 85 aa, 9120 dalton, -2*e* at pH 7) was obtained from Jeremy S. Lee [3]. 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.

βLGa is monomeric in the cis *chamber*

βLGa is dimeric under physiological conditions (pH 7, [βLGa]>50 μM)[4] with a dissociation constant of ~20 μM. The dimer has semi-axes of 1.8 and 3.5 nm (dm=3.6 and 7 nm)[5]. Our measurements were performed at a very low concentration of βLGa protein, in the *cis* chamber it was ~30 nM, three orders of magnitude lower than the physiological condition. Dimer association has Keq= 5.36 104 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 μs and t_d >6000 μs were not selected in this work. Blockage events with long rise times or large slops 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 2001[7] 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 Perkins1986[8] 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 σ=0.115/(Ω·cm)= 1.1 x10^{-8}/(Ω·nm), ψ =0.12 V, and H_{eff} = 20 nm as determined below.

1. Using the Radical Planes method from Table 3 and 4 in Nadassy 2001 [7] the atomic volume of AT and GC pairs are Λ (G+C)=0.6066 nm^3 and Λ (A+T)=0.6183 nm^3. The quotient of the average volume of a base pair Λ_{bp}=0.61245 nm^3 and the rise in the helix per base pair, 0.34 nm, give the excluded volume per unit length, A_{DNA} =0.61245 nm^3 /(0.34 nm,) = 1.8 nm^2. Using the mode of the corrected dsDNA current blockage histogram gives a calibration of H_{eff} =20 nm.

2. Using the numbers from the paper of Zwolak[9] below for the volume per Nucleotide, in cubic nanometers Λn=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: Λ_{bp}=0.6855 nm^3 = A_{DNA} x 0.34 nm. Using these numbers, the A_{DNA}=2.02 nm^2. Using the mode of the corrected dsDNA current blockage histogram gives a calibration of H_{eff} =22 nm.

The modest difference in DNA volume estimates suggests that a ~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.

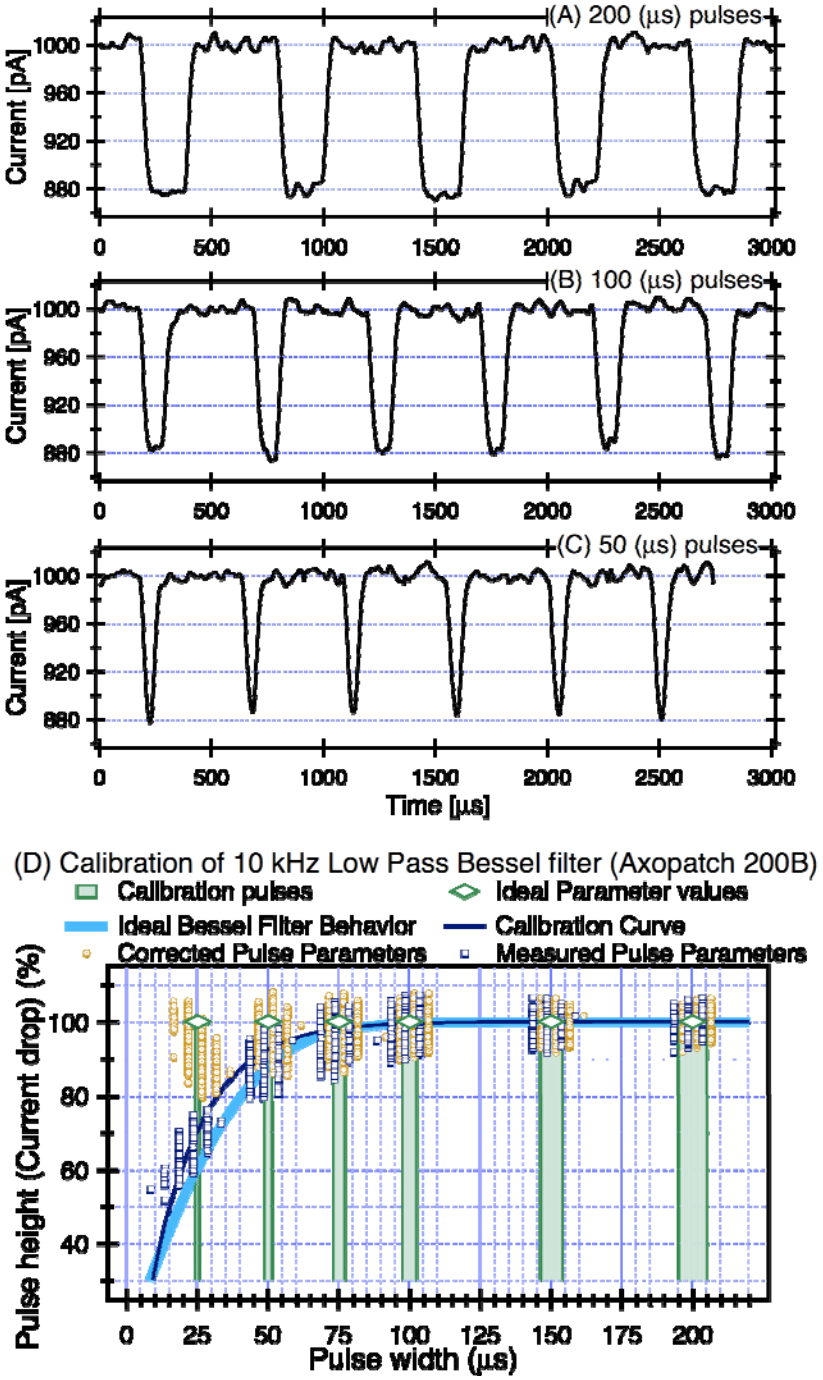


Figure S2:

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_a was calculated) verses 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:

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

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 Tanford [10]. Translocation volumes were similarly calculated using the consensus volume estimate for each amino acid published in Perkins1986 [8]. The translocation potential was calculated by integrating the charge over the electric field present assuming that the electostatic 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 ßLGa translocates as a folded protein, we can predict the observed translocation signals by treating ßLGa as a charged ellipsoid. At pH 7.0 ßLGa has an electric monopole of approximately q=9e and an electric dipole of approximately 700 Debye.

Dipolar orientations.

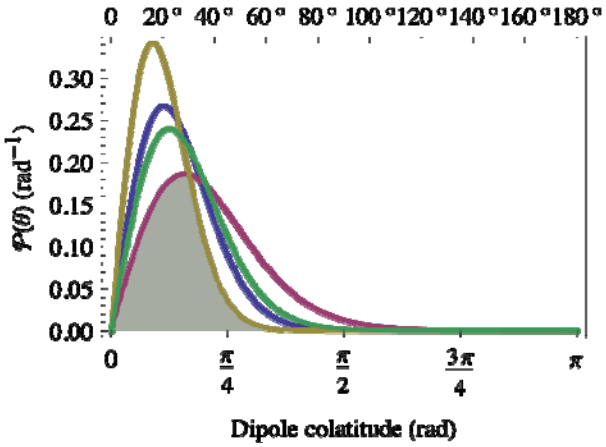
The dipolar enegy is $\Delta U_d = \vec{\mu}.\vec{\mathcal{E}} = \mu\mathcal{E}\cos(\theta)$ where is the angle between the dipole and the nanopore axis. This quantity must be statistically averaged over the spherical angles. The probability density of a particular orientation is:

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

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. ΔS = -((8.2 J)/(K mol)). Using the distribution above to evaluate the expectation value for the energy gives: ΔU = -((14.437 kJ)/mol). Giving a free energy contribution from the dipole of ΔG = -((12.0 kJ)/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 βLGa 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_{eff}=10 nm. (See the blue curve in Figure S3)

Boltzmann Distribution of dipoles



Blue : Monmer H_{eff} = 10 nm, Red : Monmer H_{eff} = 20 nm, Yellow : Dimer H_{eff} = 10 nm, Green : Dimer H_{eff} = 20 nm

Figure S3. Boltzmann Distribution of dipoles.Yellow (dimer dipole Heff= 10 nm), Blue (monomer dipole Heff= 10 nm), Green (dimer dipole Heff= 20 nm), Red (monomer dipole Heff= 20 nm).

Exit barrier due to dipolar orientation

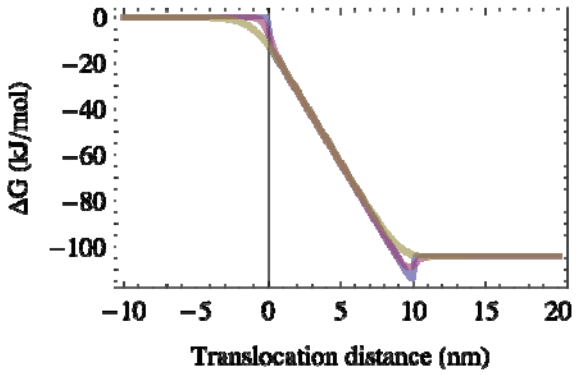


Figure S4. The free energy vs translocation distance for hard spheres of different effective radii. The barrier is quite apparent when the protein is treated as a point charge and dipole (blue curve), If the charges are spread over a diameter of 1 nm, the barrier softens considerably (red curve). Once 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 βLGa 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 βLGa 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 βLGa 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 $\varepsilon = \Delta\phi/H_{\text{eff}}$. 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.

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

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, $P(d_{trans},t)= 0$—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 x0:

$$\mathcal{P}(x,t) = \frac{e^{-\frac{(x-u_e\mathcal{E}t)^2}{4\mathcal{D}t}}}{\sqrt{4\pi\mathcal{D}t}} - \frac{\mathcal{A}e^{-\frac{(x-u_e\mathcal{E}t-x0)^2}{4\mathcal{D}t}}}{\sqrt{4\pi\mathcal{D}t}}.$$

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

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

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

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

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

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

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

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.