ARTICLES
PUBLISHED ONLINE: 25 JULY 2016 | DOI: 10.1038/NNANO.2016.120
nature nanotechnology

Reading the primary structure of a protein with 0.07 nm^3 resolution using a subnanometre-diameter pore

Eamonn Kennedy[1†], Zhuxin Dong[1†], Clare Tennant[2] and Gregory Timp[3*]

The primary structure of a protein consists of a sequence of amino acids and is a key factor in determining how a protein folds and functions. However, conventional methods for sequencing proteins, such as mass spectrometry and Edman degradation, suffer from short reads and lack sensitivity, so alternative approaches are sought. Here, we show that a subnanometre-diameter pore, sputtered through a thin silicon nitride membrane, can be used to detect the primary structure of a denatured protein molecule. When a denatured protein immersed in electrolyte is driven through the pore by an electric field, measurements of a blockade in the current reveal nearly regular fluctuations, the number of which coincides with the number of residues in the protein. Furthermore, the amplitudes of the fluctuations are highly correlated with the volumes that are occluded by quadromers (four residues) in the primary structure. Each fluctuation, therefore, represents a read of a quadromer. Scrutiny of the fluctuations reveals that the subnanometre pore is sensitive enough to read the occluded volume that is related to post-translational modifications of a single residue, measuring volume differences of ~0.07 nm^3, but it is not sensitive enough to discriminate between the volumes of all twenty amino acids.

Proteins are the machinery that makes biology work. The 3D structure of a protein—how a protein binds to itself—determines its function. One of the key factors that dictates the protein's 3D structure is the primary structure, which consists of a linear sequence of amino acids (AAs) linked by peptide bonds. Thus, sequencing a protein is essential to proteomics, the next step beyond genomics, in the analysis of biology[1,2]. However, the two methods commonly used for sequencing proteins, mass spectrometry and Edman degradation, are subject to limitations. Edman degradation does not work if the N-terminal AA is chemically modified or buried in the folded protein and it only sequences peptides that are about 30–50 residues long. On the other hand, mass spectrometry can sequence a protein of any size, but it does not provide information on the complete sequence. As mass spectrometry relies on enzymatic digestion, it becomes computationally demanding to reassemble the sequence from its constituents as the size increases. Moreover, mass spectrometry lacks sensitivity[3].

Sequencing a single protein molecule with a nanopore is among the alternatives that have been proffered[4–17]. Single molecule sensitivity develops when a protein that is immersed in an electrolyte is impelled through the nanopore by an electric field, producing a blockade in the current that depends on the occluded volume and informs on the chemical constituency in the pore. So far nanopores have been used to detect and analyse proteins, but not to sequence them. This is because the higher-order (secondary, tertiary, quaternary) structure of the protein confounds the interpretation of the blockade current. Moreover, the charge distribution along the native protein is not uniform, which frustrates the systematic control of the translocation kinetics by the electric field in the pore[17]. Recent developments have focused mainly on the use of biological pores for sequencing DNA[18–21]. Nanopore sequencing of DNA is distinguished from other methodologies by kilobase-long reads of single molecules[20]. Although the read fidelity is low[20,21]—for example, the Oxford MinION v7 chips show only about a 68% correct per-read average—with multiple reads per site, that is, high coverage (30×), it is practicable to sequence this way. This methodology for sequencing DNA cannot be extended to proteins, however, because the pores are too large—lacking chemical specificity—and the chemical agents needed for denaturation would adversely affect a biological nanopore.

To sequence a protein with a pore, several technical hurdles have to be overcome. First the protein has to be denatured to eliminate the higher-order structure, leaving only the primary structure, to facilitate the interpretation of the blockade current associated with the AAs[22]. Second, the deficient chemical sensitivity of a pore (which is related to the fractional volume occluded by the molecule[23]) has to be improved. Third, if the electric force field in the pore is to be used to systematically impel the molecule through the pore, the charge distribution along the protein has to be uniform. To overcome these hurdles, we sputtered subnanometre-pores (subnanopores) through thin inorganic silicon nitride membranes and used them to analyse single protein molecules that were denatured by heat, sodium dodecyl sulfate (SDS) and β-mercaptoethanol (BME).

Subnanopore fabrication and characterization

Electron beam-induced sputtering[24] in a scanning transmission electron microscope (STEM) was used to produce pores smaller than the size of an α-helix (which has a diameter of <0.5 nm and a rise per residue of 0.15 nm)—a common secondary structure found in proteins that is comparable in size to a hydrated ion[25]. The small size was doubtless the key to improved chemical specificity. The topographies of the subnanopores were inferred from transmission electron microscopy (Fig. 1a–c, (i) and Supplementary Fig. 1). To accurately assess the topography, each micrograph was

[1]Electrical Engineering, University of Notre Dame, Notre Dame, Indiana 46556, USA. [2]Chemical Engineering, University of Notre Dame, Notre Dame, Indiana 46556, USA. [3]Electrical Engineering and Biological Science, University of Notre Dame, Notre Dame, Indiana 46556, USA.
[†]These authors contributed equally to this work. *e-mail: gtimp@nd.edu

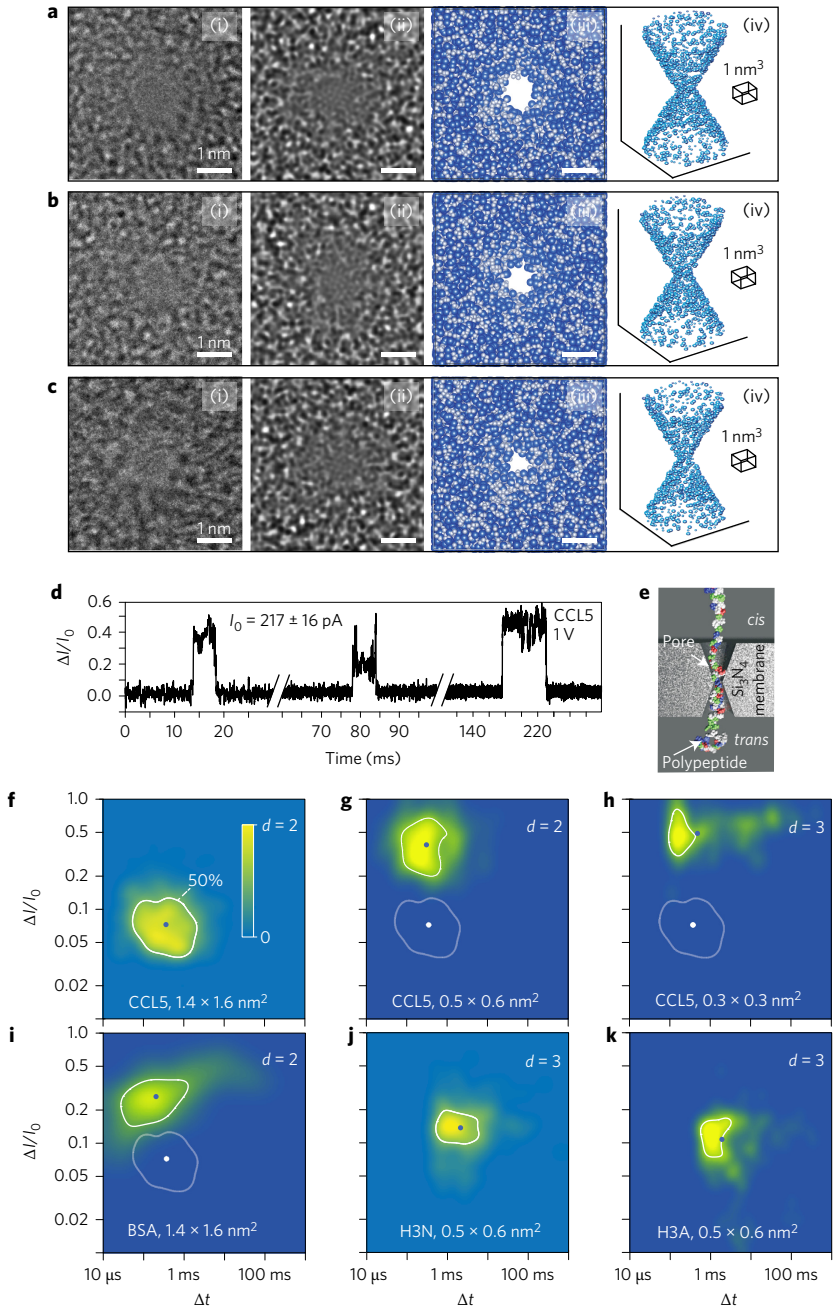

Figure 1 | Detecting single proteins using a subnanopore. a–c, TEM images (i), multislice simulations (ii), 2D projections (iii) and 3D representations (iv) of subnanopores with nominal diameters of 0.7 nm (**a**), 0.5 nm (**b**) and 0.3 nm (**c**), respectively. The subnanopores were sputtered through silicon nitride membranes that were nominally 10 nm thick. The shot noise is associated with electron transmission through the pore. The multislice simulations of the TEM images are consistent with the experimental conditions. The simulations all assume a bi-conical pore with a 20° cone angle and a defocus (see Methods) of −40 nm. The projections correspond to pores with 0.7 × 0.8 nm^2, 0.5 × 0.6 nm^2 and 0.3 × 0.4 nm^2 cross-sections at the waist. The close correspondence between the simulations and the actual TEM images indicates that the models accurately reflected the actual pore topography. The 2D projections (top view) through the model indicate the atomic distribution near the pore waist. The atoms are represented by a space-filling model in which Si is a sphere with a diameter of 0.235 nm and *N* is a sphere with a diameter of 0.13 nm. The 3D space-filled representations of the pore models show only the atoms on the pore surface for clarity. **d**, Consecutive current traces are shown that illustrate the distribution of the duration and fractional blockade currents associated with translocations of single molecules of CCL5 through a 0.5 × 0.6 nm^2 pore at 1 V. Higher values correspond to larger blockade currents. **e**, Schematic of the translocation of a protein through a subnanopore. Denaturing agents impart a uniform negative charge to the protein, resulting in a rod-like structure. **f–h**, Heat maps that characterize the distribution of fractional blockades relative to the open pore current ($\Delta I/I_0$) versus the blockade duration (Δt) are shown, associated with denatured CCL5 translocating through pores with a 1.4 × 1.6 nm^2 (**f**), 0.5 × 0.6 nm^2 (**g**) and 0.3 × 0.3 nm^2 (**h**) cross-section, respectively, at 1 V. The white contour indicates a region that contains 50% of the blockades. The same CCL5 contour in **f** is represented in **g–i**, in grey to contrast the median fractional blockades. **i**, As for **f**, but a heat map showing the blockade current distribution that is associated with denatured BSA translocating through the same 1.4 × 1.6 nm^2 pore at 1 V. This distribution was easily distinguished from CCL5 because the energy distance was large, that is, $\Delta = 1.6 \times 10^{-4}$. **j,k**, Heat maps that characterize the distribution of the blockades that are associated with denatured H3N and H3A, respectively, translocating through a pore with a 0.5 × 0.6 nm^2 cross-section at 0.7 V. These distributions could not be easily distinguished because the energy distance was only $\Delta = 3.8 \times 10^{-5}$.

imitated by multislice simulations (Fig. 1a–c,(ii)). The simulations reproduced the actual imaging conditions while accounting for dynamic scattering of the electron beam by the membrane. From the close correspondence between the images and simulations, it was inferred that the models (Fig. 1a–c,(iii,iv)) were realistic representations of the actual pores. The simulations indicated that the pores were bi-conical, with cone-angles θ that ranged around $15 \pm 5°$, and irregular, with subnanometre cross-sections at the waist. Electrolytic conductance measurements, along with finite element simulations (FES) of the measurements, confirmed the subnanometre size and bi-conical topography (Supplementary Note 1 and Supplementary Fig. 2a–d), after accounting for pore charge (Supplementary Note 2 and Supplementary Fig. 3). Furthermore, FES of the subnanopores revealed that the bi-conical topography crowded the current and focused the electric field at the waist into a region about 1.5 nm in extent (Supplementary Fig. 2c), which is approximately four AA residues long (as there is about 0.38 nm per AA on the peptide chain.)

Another salient aspect of the device involved the use of silicon nitride membranes (nominally 10 nm thick) that were resistant to chemical agents such as SDS and BME and to the high temperature used for denaturation. SDS is an anionic detergent that works, in combination with heat (45–100 °C) and reducing agents like BME, to impart a nearly uniform negative charge to the protein that stabilized denaturation. Although the exact structure of the aggregate formed by SDS and the proteins remains unsolved, from the several models that have been proposed[26–32], we adopted a 'rod-like' model in which the SDS molecules form a shell along the length of the protein backbone[26]. The resulting uniform charge on the protein was supposed to facilitate the electrical control of the translocation kinetics. Owing to its size, however, it is unlikely that the SDS remained bound to the protein; it was probably cleaved from the protein by the steric constraints imposed by the pore topography above the waist.

Protein discrimination from blockade current measurements

Nine types of protein were analysed by measuring the blockade currents through subnanopores: two recombinant chemokines RANTES (CCL5) and CXCL1; bovine serum albumin (BSA); three biotinylated, subtly different variants of the tail peptides of the H3 histone (residues 1–20); two homopolymers that were used as negative controls, poly-L-glutamic acid sodium (20 AAs long; E20) and poly-L-arginine hydrochloride (30 AAs long; R30); and a custom-made, poly-AA block co-polymer, block-$p(R)_{10}$-*co*-$(G_3S)_3$-G (denoted by R-G; 23 AAs long). Of special interest are the three tail peptides: one was native (denoted by H3N) and the other two were chemically modified at a single position 9 (lysine) either by acetylation (H3A) or trimethylation (H3M). Post-translational modifications (PTMs) such as these are interesting because they introduce new functional groups into the peptide chain that extend protein chemistry beyond the twenty proteinogenic AAs and have been implicated in epigenetics[33,34].

Typically, when the denatured protein was introduced with SDS and BME on the *cis* side of a pore blockades were observed in the open pore current (Fig. 1d) that were attributed to the translocation of single molecules (Fig. 1e and Supplementary Fig. 4). To facilitate comparisons the distribution of the blockades was classified by the fractional change in the pore current relative to the open pore value ($\Delta I/I_0$) and the duration of the blockade (Δt). The aggregate data were represented by normalized heat maps of the probability density functions (PDF) reflecting the number and distribution of blockades (Fig. 1f–k). To account for the differences in the duration and fractional blockades, it was assumed that each translocation explored different aspects of the pore topography: different alignments of a rigid, rod-like protein relative to an irregular pore with a sub-nanometre-cross-section or different trajectories for hydrated ions moving through a blockaded pore of a comparable size. Thus, the blockade distribution was attributed to factors that are related to conformational noise such as a persistent, native-like topology in the denatured protein unravelling in the pore[35], the initial configuration of the molecular termini relative to the pore or different orientations (N-terminus versus C-terminus or yaw/twist about the vertical axis) of the rigid, rod-like molecule relative to the pore topography.

The fractional change in current can be simply related to the ratio of the molecular volume to the pore volume, that is $\Delta I/I_0 = f\Delta V_{mol}/V_{pore}S$ where f measures the molecular shape and orientation and S is a size factor that accounts for distortions in the electric field that occur when the molecule is comparable in size to the pore[36]. This idea was used to set expectations for the measurements of the fractional blockade current (Supplementary Note 3 and Supplementary Table 2). For example, a denatured protein of about 26 residues would span a 10-nm-thick membrane so that $\Delta V_{26AA}/V_{pore} = 0.044 \propto \Delta I/I_0$. However, if the current was crowded in the same topography, then only about four AAs would span a thickness of 1.5 nm, reducing the effective pore volume V_{pore}^{eff}, so that $\Delta V_{4AA}/V_{pore}^{eff} = 0.64\ nm^3/3.72\ nm^3 = 0.17 \propto \Delta I/I_0$. These expectations were borne out in heat maps derived from the ionic blockade distributions associated with CCL5, BSA, H3 and H3A translocations collected from different pores (Fig. 1f–k and Supplementary Fig. 5a,b,d). For example, for CCL5 with a 1 V bias (Fig. 1f–h), the median $\Delta I/I_0 = 0.07$ occurs at a median $\Delta t = 400$ µs for a pore with a $(1.4 \pm 0.1) \times (1.6 \pm 0.1)\ nm^2$ cross-section. Likewise, for BSA with a 1 V bias (Fig. 1i), the median $\Delta I/I_0 = 0.07$ occurs at a median $\Delta t = 400$ µs for the same pore. On the other hand, for CCL5 and the same bias, the median blockade in a $(0.5 \pm 0.1) \times (0.6 \pm 0.1)\ nm^2$ cross-section pore improves substantially to $\Delta I/I_0 = 0.38$, but occurs at nearly the same median $\Delta t = 330$ µs. Moreover, although the distributions broaden, there was very little change in the median duration with diameter. In particular, for a pore with a $(0.3 \pm 0.1) \times (0.4 \pm 0.1)\ nm^2$ cross-section, the blockade distribution extended over a range from 100 µs $< \Delta t <$ 70 ms and $0.25 < \Delta I/I_0 < 1$. The median fraction improved to $\Delta I/I_0 = 0.47$ at a median $\Delta t = 520$ µs. Thus, the data acquired from CCL5 and BSA were found in the range between $\Delta V_{26AA}/V_{pore}$ and $\Delta V_{4AA}/V_{pore}^{eff}$. Likewise, the fractional blockade data acquired from H3N and its conjugates were found in the range between $\Delta V_{20AA}/V_{pore}$ and $\Delta V_{4AA}/V_{pore}^{eff}$. For H3N and a 0.7 V bias in another pore with a $0.5 \times 0.6\ nm^2$ cross-section, the median fractional blockade current measured was smaller, $\Delta I/I_0 = 0.12$, with a median $\Delta t = 2$ ms, consistent with the lower voltage. These data were all interpreted to be consistent with the idea that the current was either crowded near the pore waist or that the molecular volume was larger than that ascribed to a linear chain of AAs (possibly due to SDS bound to the protein).

Taken together, these observations support the assertion that $\Delta I/I_0$ was a measure of the molecular volume occluding the pore. Thus it was reasoned that subnanopore sensitivity to the molecular volume would promote the chemical specificity necessary to discriminate between proteins and even AA residues based on their volume. To test this notion, two pure protein solutions—one containing CCL5 and another BSA—were analysed using a pore with a $1.4 \times 1.6\ nm^2$ cross-section at a 1 V bias[37]. To illustrate that the PDFs are distinct, a contour that represents the PDF from CCL5 (Fig. 1f) was juxtaposed on the PDF heat map corresponding to BSA (Fig. 1i). The point-by-point difference between the PDFs, that is $(PDF_{protein\ 1} - PDF_{protein\ 2})^2$, revealed the dissimilarities. By integrating these differences over the entire blockade current space, a metric of the statistical distance between the two was obtained—namely Δ, which is related to the energy distance[38]. The energy distance between PDFs that represent the BSA and CCL5 measured with the same pore was $\Delta = 1.6 \times 10^{-4}$, which indicated very different distributions. In contrast, two peptides that differ only by a single PTM (H3N and H3A) analysed separately

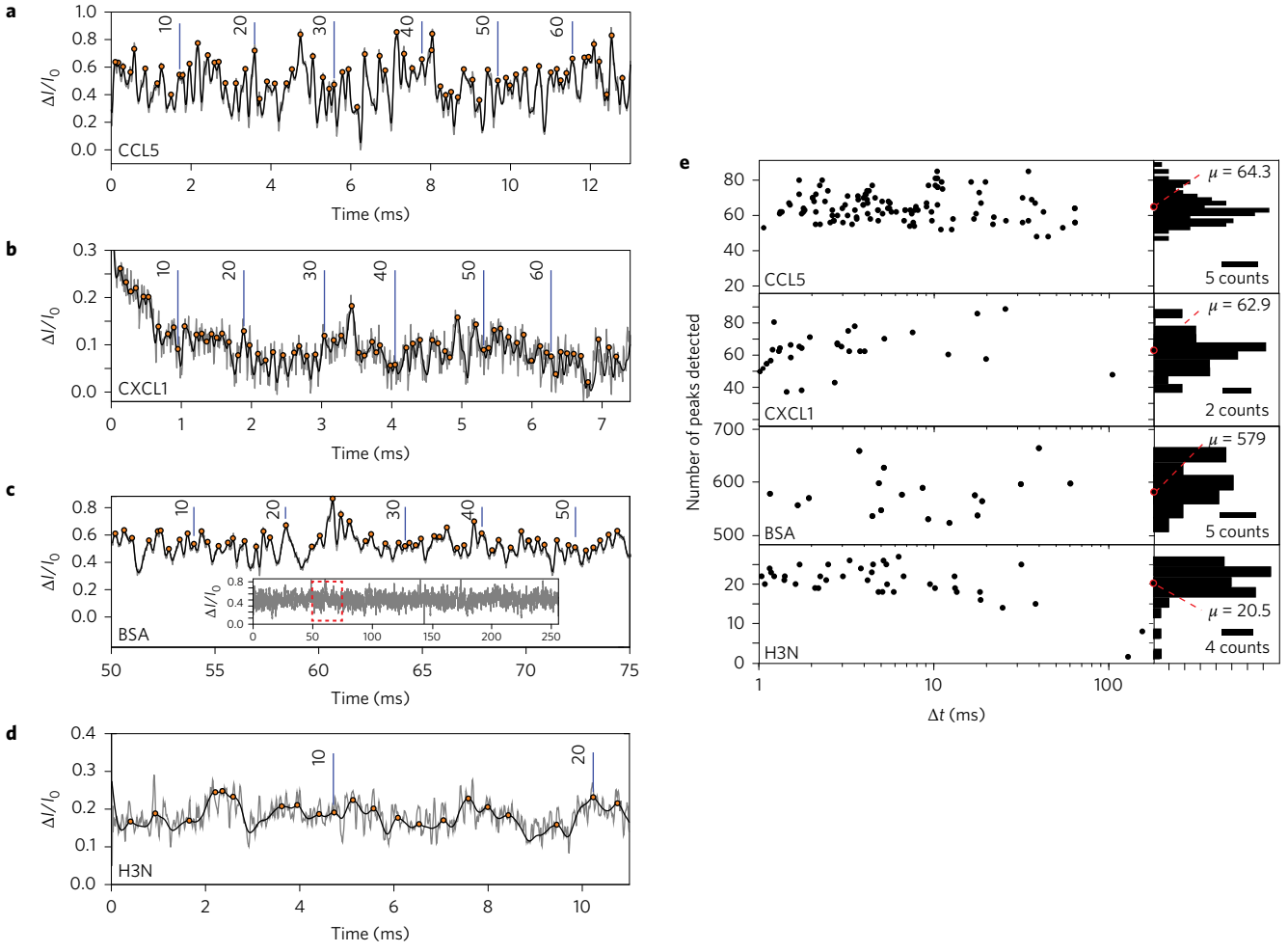

Figure 2 | Detecting amino acids in a single protein using a subnanopore. a, An expanded view is shown of a single blockade (from Fig. 1d) illustrating nearly regular fluctuations in the blockade current associated with a single CCL5 molecule translocating through a subnanopore with a 0.5×0.6 nm^2 cross-section. The fluctuations were attributed to individual residues entering (and exiting) the pore one at a time. The grey trace represents unfiltered, unfitted raw data, whereas the black line is the smoothed data. The orange circles identify the peaks in the trace that were fitted with a Gaussian. Counting this way gives $N_{CCL5} = 67$. **b**–**d**, As for **a**, these plots show expanded views that illustrate nearly regular fluctuations, the number of which is associated with the number of AAs in mature CXCL1 (**b**), BSA (**c**) and H3N (**d**) through subnanopores with 0.5×0.6 nm^2 cross-sections, respectively. For the traces depicted, using a Gaussian fit to identify the fluctuations produces counts of $N_{CXCL1} = 69$, $N_{BSA} = 593$ and $N_{H3N} = 21$, respectively. The inset in **c** shows a current blockade that is associated with single BSA translocation; the portion of the trace highlighted by the red dashed outline is reflected in **c**. **e**, The number of fluctuations tallied using a Fourier analysis from blockades with different duration is shown for the same four proteins as in **a**–**d**. The tallies were independent of the blockade duration, but were dependent on the number of residues in the proteins. Alternatively, when a Gaussian algorithm with the same settings was applied generally to the same data, this approach yielded on average $N_{CCL5} = 60 \pm 29$ peaks for CCL5, $N_{CXCL1} = 62 \pm 26$ for CXCL1 and $N_{H3N} = 22 \pm 12$ peaks for H3N, which are all consistent with the number of AAs in the mature proteins within the margins of error. BSA required changes to the settings, but still yielded $N_{BSA} = 601 \pm 152$ peaks, in agreement with the residues that constitute the mature protein.

using the same pore with a 0.5×0.6 nm^2 cross-section were ostensibly indiscernible (Fig. 1j,k) because $\Delta = 3.9 \times 10^{-5}$. These simple tests indicated that stochastic sensing could discriminate proteins, but it was not specific enough to differentiate between species that differ by only a single PTM.

Reading quadromers from fluctuations in a single blockade

Strikingly, scrutiny of the fluctuations observed within each blockade exposed signatures of the primary structure that allowed for discrimination of the proteins. Each blockade within a subset of the distribution from $\Delta I/I_0 > 0.30$ for $1 < \Delta t < 70$ ms, comprising the majority of blockades, revealed nearly regular fluctuations, the number of which corresponded with the number of residues in each type of protein (Fig. 2a–d and Supplementary Fig. 6). By normalizing the frequency of fluctuations observed by the duration of the blockade, a tally of the number of fluctuations in the blockades that are associated with translocations of CCL5 yielded $N_{CCL5} = 64.3 \pm 3.3$, regardless of the blockade duration (Fig. 2e), which coincided with the 68 AA residues in the protein. Likewise for CXCL1, we tallied a similar number of fluctuations, $N_{CXCL1} = 62.9 \pm 9.3$, close to the 71 residues in the protein. In contrast, $N_{BSA} = 579.0 \pm 64$, was observed under the same conditions when BSA blockaded the pore, which agreed (within the error) with the 583 AAs that constitute the protein. Finally, far fewer fluctuations, $N_{H3N} = 20.5 \pm 1.3$, were observed when H3N was impelled through a subnanopore, corresponding to a peptide with only 21 residues. Although Fourier analysis was the primary means for counting the number of fluctuations in a blockade, another tally (see Methods) that used a Gaussian fit to the peaks (orange circles in Fig. 2a–d) gave similar results.

Because of the correspondence between the number of fluctuations and AA residues in the protein, it was postulated that each fluctuation reflected a read of the AA sequence of a single protein

molecule. As all of the blockades were recorded using the same current amplifier bandwidth, the observation that both the number and the pattern of fluctuations within a blockade persisted over a range of durations and fractional currents for most of the blockades (Supplementary Fig. 7a,b) precluded random noise as the sole explanation. The analysis of the fluctuation patterns indicated two distinct groups, which showed similar peak maxima under temporal inversion (Supplementary Fig. 7c). This was interpreted as evidence of two almost equivalent but opposite translocation directions—either N-terminus or C-terminus first. Therefore, all of the blockades were sorted into two groups and the second was inverted in time, depending on the relative correlation of their observed peaks. A simple binomial *t*-test, assuming a null hypothesis of 50% of the events flipped for random noise, indicated that the number of flipped events observed was $p < 1 \times 10^{-6}$ for all proteins, given the sample sizes used. Therefore, the data consistently showed a preferential direction for the translocation (N-terminus first). Other discrepancies between the patterns observed in the majority of the blockades were attributed to misreads such as a skip or multiple reads of the same AA. Misreads give rise to a different tally of fluctuations or gross irregularities, which may be associated with conformational noise. For example, lags were also observed in the fluctuation pattern (Supplementary Fig. 7d). These accounted for 5–25% of the blockades, depending on the bias and the protein, and were ascribed to time-consuming reconfigurations of the molecular termini just before insertion into the pore. Lags were culled from the distribution. Finally, multilevel events were also observed (especially if the concentration of SDS was <0.0001%), but only rarely, accounting for <5% of blockades typically. This may be associated with the protein unfolding in the pore (Supplementary Fig. 7e), and these events were also culled.

The amplitudes of the fluctuations observed in different blockades from the same protein were highly correlated with each other. This assertion was rigorously tested by comparing the fluctuations in each blockade. As neither the duration nor the fractional current were perfectly uniform, each blockade was first normalized in time and the average fractional current was zeroed for comparison (Fig. 3a–c and Supplementary Fig. 8a). A consensus was then formed from the average of a number of blockades—each associated with the translocation of a single molecule. The mean Pearson product-moment correlation coefficient between the consensus and an individual blockade was $C = 0.42$ for CCL5 (Fig. 3a), 0.55 for CXCL1 (Fig. 3b), 0.67 for H3N (Fig. 3c) and 0.23 for BSA (Supplementary Fig. 8a) Thus the fluctuations persisted even after averaging, unlike the open pore current noise (Supplementary Fig. 9).

Because the fluctuations were correlated and nearly regular, it was inferred that the SDS must impart a nearly uniform negative charge density along the polypeptide, resulting in a consistent electric force that worked to impel the molecule through the pore. It is likely that counter-ions moved along with the SDS–protein aggregate to minimize Coulomb repulsion, but because the pore has a diameter comparable to a hydrated ion and SDS, the motion of the ensemble through it would be impeded by steric hindrances. Thus the coincidence between the number of fluctuations and AAs in the protein and the near-regularity of the patterns were consistent with a tightly choreographed, turnstile motion of AAs through the pore in which a single AA stalled repeatedly in a well-defined conformation and then eventually proceeded through the pore due to the electric force on the molecule (Z. Dong, E. Kennedy and G. Timp, manuscript in preparation, and ref. 39).

If each fluctuation in a blockade reflected one AA entering the pore while another leaves, then it was reasoned that the amplitude of the fluctuation should be attributed to the occluded volume associated with the residues in the pore. Because the pore current was crowded and most of the potential dropped near the waist, it was further argued that each fluctuation should measure the occluded volume due to 3–5 AAs in the waist, with the exception of the first and last fluctuations at the inception and termination of a blockade. These were interpreted as a reduced sum of the AAs.

Consistent with this reasoning, consensuses formed by averaging together the normalized and binned blockades were found to be highly correlated with models for the peptide chains in which each residue was represented by its volume. Taking estimates obtained from crystallography data[40], the primary structure of the protein was translated into a sequence of AA volumes (Supplementary Table 3). To account for the current crowding in the pore waist, a moving average with a window size, k, ranging from $k = 3$ to 5 AAs (depending on the topography) was then performed on the sequence of volumes to create a model for comparison. The models based on AA volumes were found to be correlated with the empirical consensuses, with agreement that improved as the number of blockades increased. To illustrate the argument, error maps were produced by partitioning 400 CCL5 blockades into 17 consensuses (Fig. 3a, bottom), each of which was then compared with the model. The agreement for each read was expressed as a percentage and subsequently identified as a correct call, depending on whether the agreement was greater than 20%. In this way, the 17 consensuses for CCL5 exhibited an average read accuracy of 59.4%, whereas the entire 400-blockade consensus produced a mean read accuracy of 65.2%. In contrast, for CCL5, the consensus correlation to a $k = 3$ model was $C = 0.75$.

Likewise, the correlation between the $k = 5$ model for CXCL1 with the 45-blockade consensus was $C = 0.51$, with a mean read accuracy of 84.7%. The performance with the shortest peptide, H3N, was extraordinary. The correlation of a $k = 3$ model with a single event was 68%, but a 52-blockade consensus showed only two positions out of 21 outside the 20% tolerance, representing a read accuracy of 90%. However, all of these compare favourably with BSA, for which the correlation of the $k = 5$ model with the 41-blockade consensus was 0.35, with an associated mean read accuracy of 68.4% (Supplementary Fig. 8a). The lower correlation for BSA probably reflects more misreads, that is, skips and repeats. Because the true read accuracies found for each protein were more than six standard deviation occurrences with respect to the noise (see Methods), it was fair to assume that each fluctuation represented a low-fidelity read measuring the occluded volume associated with 3–5 AA residues in the pore waist and that the accuracy improved with coverage.

Regardless of the coverage, read fidelity could still be compromised by systemic errors. A further analysis of the read fidelity and cross-correlation between proteins exposed several interesting trends. For example, it was apparent from the error maps (Fig. 3a–c and Supplementary Fig. 8a) that the correlations with the model do not accumulate randomly; for example, discrepancies were consistently found near positions 1, 11, 19, 21, 31, 53 and 58 for CCL5, positions 6, 12, 29 and 30 for CXCL1 and positions 11 and 12 for H3N. By assigning errors to the *a priori* sequence of cumulative AA volumes present at each position, the volume errors for each AA were calculated (Fig. 3d–g and Supplementary Fig. 8b,c) that, in combination, indicated the sources. In particular, negatively charged AAs (D, E) repeatedly showed the highest read errors for these three proteins. On the other hand, the two (positively charged) lysines (K) at position 54–55 both exhibited a mean read accuracy of 92% for CCL5 (Fig. 3a). Finally, AAs with small volumes (A, C, G and S) were frequently misread, which can be rationalized as they represent <10% of the effective pore volume. Thus, the model based on AA volumes was imperfect, which doubtless affected the interpretation of the blockade current.

The correspondence between the volume model and the fluctuations in a blockade was predicated on the idea that the blockade measured quadromer volumes. To test this hypothesis further,

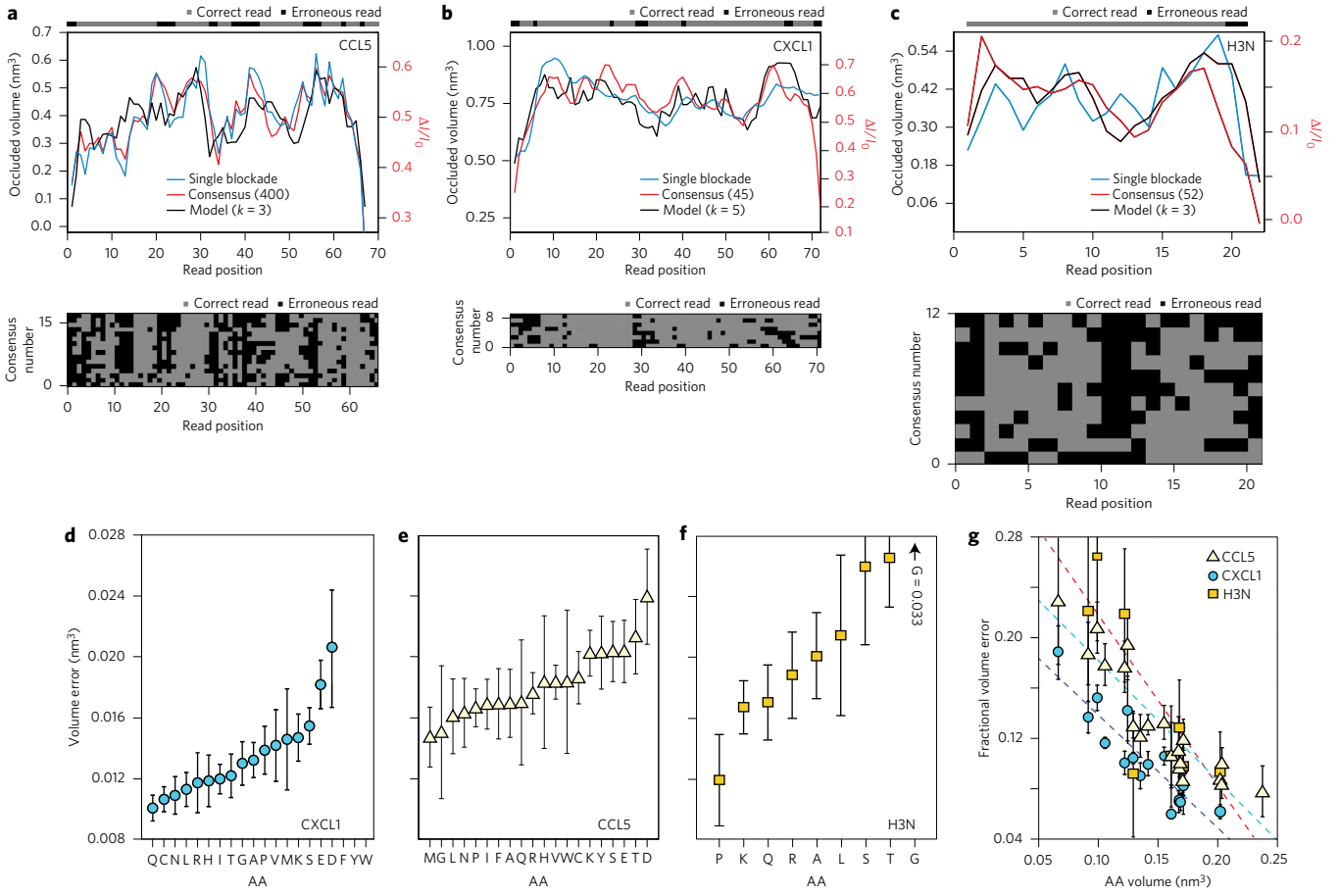

Figure 3 | Protein sequence analysis. a, A 400-blockade consensus is shown for CCL5 through a pore with a 0.5×0.6 nm^2 cross-section (red), juxtaposed with an occluded volume model (assuming $k = 3$, black) and a single highly correlated event ($C = 0.69$, blue). The error map above the plot indicates the read fidelity. Positions where the read departs from the model less than 20% are represented in grey as correct reads. A grey-scale error map illustrating correct reads and misreads is presented below; the AA positions where the read departs from the model less than 20% of the time are represented in grey, whereas black reflects a misread. The percentage of correct reads for CCL5 was 65.2%. **b**, As for **a**, but showing a 45-blockade consensus (red) for CXCL1 through a different subnanopore with a similar 0.5×0.6 nm^2 cross-section, juxtaposed with an occluded volume model (assuming $k = 5$, black) and a single (blue) event ($C = 0.67$). Despite the lower number of blockades, the read fidelity improves for CXCL1, as evident from the error map below. The percentage of correct reads for CXCL1 was 84.7%. **c**, As for **a**, but showing a 52-blockade consensus for H3N through a subnanopore with a 0.5×0.6 nm^2 cross-section juxtaposed with an occluded volume model (assuming $k = 3$, black) and a single (blue) event ($C = 0.68$). The percentage of correct reads for H3N was 90%, as shown in the error map below. **d**–**f**, The volume errors for different residues. By assigning positional errors to the *a priori* sequence associated a read position, the absolute mean read error at each AA residue can be calculated. The errors were then converted to volume differences via the model for the AA residues that constitute CXCL1 (**d**), CCL5 (**e**) and H3N (**f**). **g**, Fractional mean read error as a function of volume of a residue in CXCL1, CCL5 and H3N. Smaller AAs are a dominant source of error. The dotted lines are least-square fits to the data and provide a guide to the eye.

extra (negative control) experiments were performed using the same pore with a $(0.5 \pm 0.1) \times (0.6 \pm 0.1)$ nm^2 cross-section to analyse R30 and E20 (Fig. 4a and Supplementary Fig. 10). Two aspects of the data acquired from these homopolymers indirectly supported the idea that the blockade current measured quadromer volumes. First, the homopolymer blockades lacked fluctuations, whereas the mean number of fluctuations tallied in a blockade were $N_{E20} = 22.5$ for E20 (Fig. 4a, inset) and $N_{R30} = 34$ for R30 (Supplementary Fig. 10, inset), corresponding to the number of AAs in the respective peptides. The tallies were imprecise presumably due to a lack of variation of the molecular volume as the peptide moved through the waist (Fig. 4a, bottom). Second, regardless of the position within the blockade, the measured fractional blockade associated with R30 ($\Delta I/I_0 = 0.18$) was larger than that observed for E20 ($\Delta I/I_0 = 0.14$) and the ratio corresponded closely to the ratio of quadromer volumes. Thus the fractional blockade currents of the two homopolymers were consistent with measurements of the actual AA volumes.

Although some are very similar, because each AA has a unique volume, discriminating volume differences between successive quadromers is the key to sequencing. To test the sensitivity, a third control experiment was performed using a subnanopore with a $(0.4 \pm 0.1) \times (0.5 \pm 0.1)$ nm^2-cross-section to analyse the R-G (Fig. 4b and Supplementary Fig. 10b) Steps observed in the blockade current were attributed to the volume differential between the R-block and the G_3S-block. Moreover, the correlation of the model with single blockades was very high: that is, $C = 0.90$ and 0.96 for the two single blockades shown in the figure, whereas the median of the single blockades was $C = 0.4$, and a consensus of 58 blockades revealed a correlation with the volume model of $C = 0.95$. In addition, the slope associated with the $k = 4$ model corresponded closely with the transition between what was assumed to be the R- and G_3S-blocks, implying quadromer reads. Presumably, for this reason, it was difficult to distinguish G from S because each quadromer had the same volume. Finally, whereas fluctuations could be detected in a consensus formed from the differences

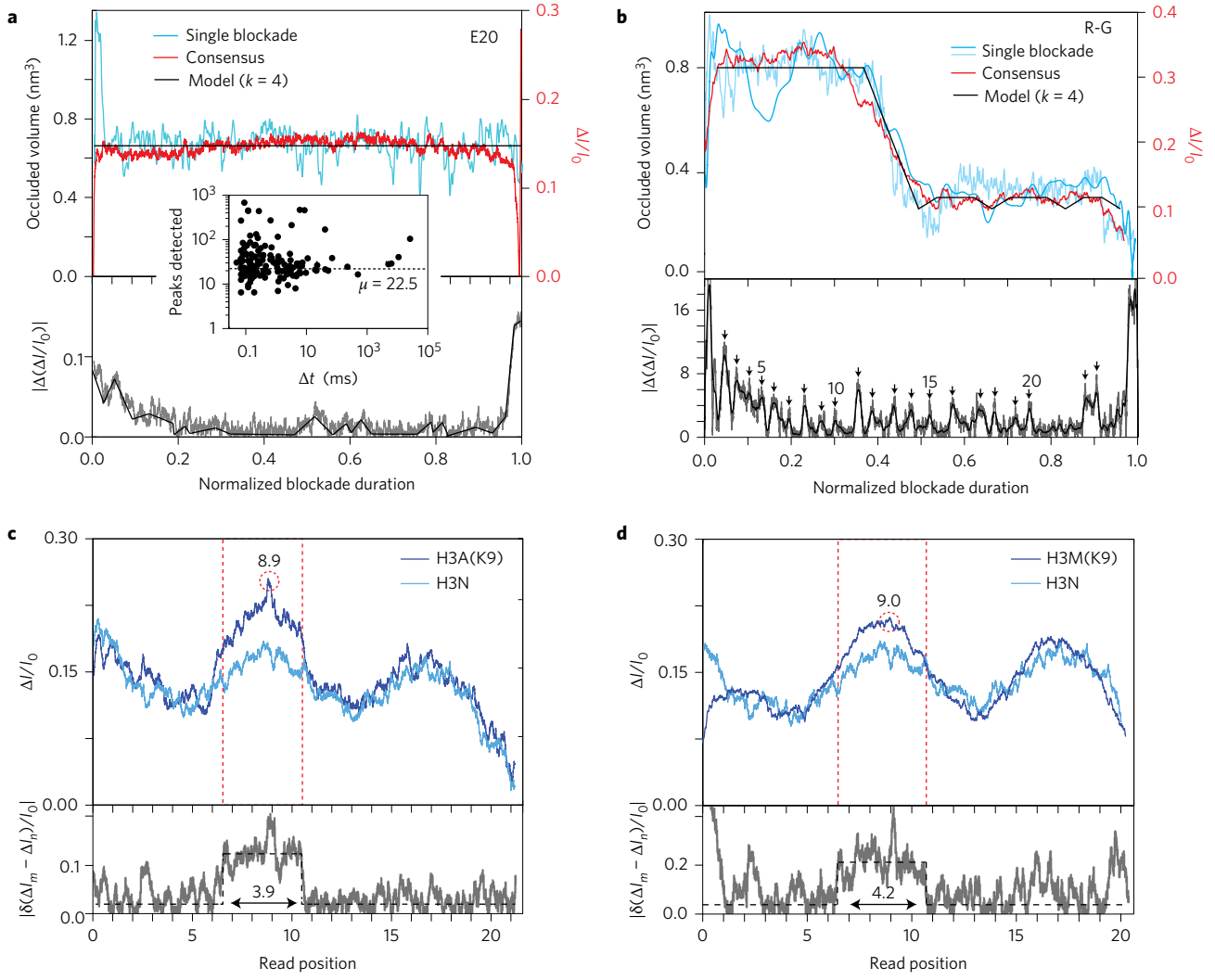

Figure 4 | Testing the sensitivity of a subnanopore to molecular volume. a, A 182-blockade consensus (red) is shown, juxtaposed with an occluded volume model (assuming $k = 4$, black) and a single blockade (blue), all plotted versus normalized blockade duration for the E20 homopolymer. The data were acquired at 0.5 V using a subnanopore with a 0.5×0.6 nm^2 cross-section in an electrolyte containing 250 mM NaCl and 0.01% (weight/volume) (w/v) SDS. The magnitude of the consensus of fluctuations obtained from the difference between individual blockades and the volume model is shown below. Inset: The number of fluctuations tallied from blockades with different durations for E20. Although the mean number ($N_{E20} = 22.5$) was independent of the blockade duration, it was imprecise. **b**, As for **a**, but showing a 58-blockade consensus for the block copolymer R-G through a subnanopore with a 0.4×0.5 nm^2 cross-section (red) juxtaposed with an occluded volume model (assuming $k = 4$, black) and single highly correlated blockades ($C = 0.90$ and 0.96, blue). The fluctuations in the block co-polymer blockade (delineated by the arrows in the figure shown below) persist after averaging. **c**,**d**, The effects of single-site chemical modifications to the histone H3 tail peptide on the blockade fluctuation amplitude are illustrated. Native H3N (light blue, 304 events) and K9-acetylated H3A (dark blue in **c**, 231-blockades, scaled to native H3) and K9-methylated H3M (dark blue in **d**, 958-blockades, scaled to native H3N) consensuses were formed, juxtaposed on the same plots and then compared. It was observed that the fluctuation amplitudes were enhanced between positions 6 and 11 indicating an increased occluded volume there. The differences between the native and modified consensus traces (grey) are presented below, showing a broad top-hat-like increase in fractional blockade (dotted black line) associated with the H3A/H3M site. When fitted, the consensus difference indicated that a single site modification resulted in changes in occluded volume that ranged over 3.9 AA residues (18.5% of the trace) for single-site H3A and 4.2 AA residues for single-site H3M. Although single-site resolution was not indicated, single-site modifications can be clearly observed. Furthermore, near the centre of the consensus difference a prominent fluctuation peak was also evident at position 8.9, which was tentatively attributed to H3A at position K9. Likewise, a weaker fluctuation was evident at position 9.0 to H3M at K9.

between each blockade and the model (Fig. 4b, bottom), they were weak and irregular. Thus, the quadromer volume model accounted for the volume difference between the two blocks that comprise the co-polymer with <0.5 nm^3 resolution, but alignment within a block was frustrated by weak, irregular fluctuations—presumably because of the similarities between the quadromers within a block.

The improved correlation with the model that develops from a consensus suggested that read fidelity could be enhanced by increased coverage. To discover the extent of coverage required, PTMs, H3A and H3M, of a specific residue of H3N were analysed. In this context, the blockade current measurements were particularly interesting because the changes associated with the occluded volumes were expected to be like those associated with glycine. For comparison, three consensuses were formed: one associated with 304-blockades of H3; another with 231-blockades from H3A; and a third with 958-blockades from H3M (Fig. 4c,d). Each was acquired from nominally 0.5-nm-diameter pores.

The juxtaposition of consensuses clearly showed the positional sensitivity of the fractional blockade current. The fractional blockade associated with both PTMs was enhanced between read positions 6 and 11. In addition, a prominent feature was observed near read position 9 in H3A measured relative to H3N, in correspondence with the expected change on position K9 due to acetylation. In contrast, a depression appeared near read position K9 in H3M measured relative to H3N. A fit of the difference in the fractional blockade between the PTM variants and native traces to a simple top-hat form revealed differences over a range of 3.9 positions (Fig. 4c, bottom), which corresponds closely to the FES estimate obtained for the number of AAs in the waist and substantiates the claim that each read reflects about four AAs. Likewise, the difference between the H3M and H3N traces extends over a range of 4.2 read positions (Fig. 4d, bottom). Moreover, the increase in the peak fractional blockade amplitude was $\Delta(\Delta I/I_0) < 0.07$, which corresponds to an estimated volume change of about $\sim 0.07\ \mathrm{nm}^3$ associated with H3A or H3M that was comparable to the volume of glycine. It follows then that a subnanopore has sufficient sensitivity to detect the smallest AA residue. However, to unequivocally discriminate between the volumes of all twenty AAs (to discriminate F-phenylalanine from Y-tyrosine, in particular, for which $\Delta V = 0.0002\ \mathrm{nm}^3$) a much higher resolution would be required.

Conclusion

Fluctuations in the blockade current associated with the translocation of a single denatured protein through a subnanopore were associated with low-fidelity reads of quadromers in the primary structure. As for the nanopore sequencers used for DNA, low-fidelity reads with multiple monomers that affect the blockade current do not pose an insurmountable problem for sequencing proteins (Supplementary Note 4)—there are algorithms that may be adapted to decode the AA sequence[41–43]. However, subnanopore sensitivity remains an open question. When the consensus was large enough, the reads were sensitive enough to detect PTMs on a single residue, but not sensitive enough to discriminate between all of the AAs.

Methods

Methods and any associated references are available in the online version of the paper.

Received 14 September 2015; accepted 6 June 2016; published online 25 July 2016

References

1. Wilhem, M. *et al.* Mass-spectrometry-based draft of the human proteome. *Nature* **509,** 582–587 (2014).
2. Altelaar, A. F. M., Munoz, J. & Heck, A. J. R. Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat. Rev. Genet.* **14,** 35–48 (2013).
3. Chandramouli, K. & Qian, P.-Y. Proteomics: challenges, techniques and possibilities to overcome biological sample complexity. *Human Genom. Proteom.* **2009,** 239204 (2009).
4. Ohshiro, T. *et al.* Detection of post-translational modifications in single peptides using electron tunneling currents. *Nature Nanotech.* **9,** 835–840 (2014).
5. Movileanu, L., Howorka, S., Braha, O. & Bayley, H. Detecting protein analytes that modulate transmembrane movement of a polymer chain within a single protein pore. *Nature Biotechnol.* **18,** 1091–1095 (2000).
6. Mohammad, M. M., Prakash, S., Matouschek, A. & Movileanu, L. Controlling a single protein in a nanopore through electrostatic traps. *J. Am. Chem. Soc.* **130,** 4081–4088 (2008).
7. Talaga, D. S. & Li, J. Single-molecule protein unfolding in solid-state nanopores. *J. Am. Chem. Soc.* **131,** 9287–9297 (2009).
8. Wei, R., Gatterdam, V., Wieneke, R., Tampe, R. & Rant, U. Stochastic sensing of proteins with receptor-modified solid-state nanopores. *Nature Nanotech.* **7,** 257–263 (2012).
9. Freedman, K. J. *et al.* Chemical, thermal, and electric field-induced unfolding of single protein molecules studied using nanopores. *Anal. Chem.* **83,** 5137–5144 (2011).
10. Fologea, D., Ledden, B., McNabb, D. S. & Li, J. Electrical characterization of protein molecules by a solid-state nanopore. *Appl. Phys. Lett.* **91,** 053901 (2007).
11. Cressiot, B. *et al.* Protein transport through a narrow solid-state nanopore at high voltage: experiments and theory. *ACS Nano* **6,** 6236–6243 (2012).
12. Nelson, E. M., Kurz, V., Shim, J., Timp, W. & Timp, G. Using a nanopore for single molecule detection and single cell transfection. *Analyst* **137,** 3020–3027 (2012).
13. Nivala, J., Marks, D. B. & Akeson, M. Unfoldase-mediated protein translocation through an α-hemolysin nanopore. *Nature Biotechnol.* **31,** 247–250 (2013).
14. Rodriguez-Larrea, D. & Bayley, H. Multistep protein unfolding during nanopore translocation. *Nature Nanotech.* **8,** 288–295 (2013).
15. Rosen, C. B., Rodriguez-Larrea, D. & Bayley, H. Single-molecule site-specific detection of protein phosphorylation with a nanopore. *Nature Biotechnol.* **32,** 179–181 (2014).
16. Merstorf, C. *et al.* Wild type, mutant protein unfolding and phase transition detected by single-nanopore recording. *ACS Chem. Biol.* **7,** 652–658 (2012).
17. Nivala, J., Mulroney, L., Li, G., Schreiber, J. & Akeson, M. Discrimination among protein variants using an unfoldase-coupled nanopore. *ACS Nano* **8,** 12365–12375 (2014).
18. Laszlo, A. H. *et al.* Decoding long nanopore sequencing reads of natural DNA. *Nature Biotechnol.* **32,** 829–833 (2014).
19. Jain, M. *et al.* Improved data analysis for the MinION nanopore sequencer. *Nature Methods* **12,** 351–356 (2015).
20. Timp, W. *et al.* Think small: nanopores for sensing and synthesis. *IEEE Access* **2,** 1396–1408 (2014).
21. Laver, T. *et al.* Assessing the performance of the Oxford Nanopore Technologies MinION. *Biomol. Detect. Quant.* **3,** 1–8 (2015).
22. Li, J., Fologea, D., Rollings, R. & Ledden, B. Characterization of protein unfolding with solid-state nanopores. *Protein Pept. Lett.* **21,** 256–265 (2014).
23. Sigalov, G., Comer, J., Timp, G. & Aksimentiev, A. Detection of DNA sequences using an alternating electric field in a nanopore capacitor. *Nano Lett.* **8,** 56–63 (2008).
24. Ho, C. *et al.* Electrolytic transport through a synthetic nanometer-diameter pore. *Proc. Natl Acad. Sci. USA* **102,** 10445–10450 (2005).
25. Ohtaki, H. & Radnai, T. Structure and dynamics of hydrated ions. *Chem. Rev.* **93,** 1157–1204 (1993).
26. Reynolds, J. A. & Tanford, C. Binding of dodecyl sulfate to proteins at high binding ratios. possible implications for the state of proteins in biological membranes. *Proc. Natl Acad. Sci. USA* **66,** 1002–1003 (1970).
27. Ibel, K. *et al.* Protein-decorated micelle structure of sodium dodecyl-sulfate-protein complexes as determined by neutron scattering. *Eur. J. Biochem.* **190,** 311–318 (1990).
28. Samso, M., Daban, J.-R., Hansen, S. & Jones, G. R. Evidence for sodium dodecyl sulfate/protein complexes adopting a necklace structure. *Eur. J. Biochem.* **232,** 818–824 (1995).
29. Mattice, W. L., Riser, J. M. & Clark, D. S. Conformational properties of the complexes formed by proteins and sodium dodecyl sulfate. *Biochemistry* **15,** 4264–4272 (1976).
30. Lundahl, P., Greijer, E., Sandberg, M., Cardell, S. & Eriksson, K. O. A model for ionic and hydrophobic interactions and hydrogen-bonding in sodium dodecyl sulfate-protein complexes. *Biochim. Biophys. Acta* **873,** 20–26 (1986).
31. Gudiksen, K. L., Gitlin, I., Moustakas, D. T. & Whitesides, G. M. Increasing the net charge and decreasing the hydrophobicity of bovine carbonic anhydrase decreases the rate of denaturation with sodium dodecyl sulfate. *Biophys. J.* **91,** 298–310 (2006).
32. Westerhuis, W. H. J., Sturgis, J. N. & Niederman, R. A. Reevaluation of the electrophoretic migration behavior of soluble globular proteins in the native and detergent-denatured states in polyacrylamide gels. *Anal. Biochem.* **284,** 143–152 (2000).
33. Nakayma, J.-I., Rice, J. C., Strahl, B. D., Allis, C. D. & Grewal, S. I. S. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* **292,** 110–113 (2001).
34. Wang, Z. *et al.* Combinatorial patterns of histone acetylations and methylations in the human genome. *Nature Genet.* **40,** 897–903 (2008).
35. Shortle, D. & Ackerman, M. S. Persistence of native-like topology in a denatured protein in 8 M urea. *Science* **293,** 487–489 (2001).
36. Qin, Z. P., Zhe, J. A. & Wang, G. X. Effects of particle's off-axis position, shape, orientation and entry position on resistance changes of micro-Coulter counting devices. *Meas. Sci. Technol.* **22,** 045804 (2011).
37. Niedzwiecki, D. J., Grazul, J. & Movileanu, L. Single-molecule observation of protein adsorption onto an inorganic surface. *J. Am. Chem. Soc.* **132,** 10816–10822 (2010).
38. Székely, G. J. & Rizzo, M. L. A new test for multivariate normality. *J. Multivar. Anal.* **93,** 58–80 (2005).
39. Nelson, E. M., Li, H. & Timp, G. Direct, concurrent measurements of the forces and currents affecting DNA in a nanopore with comparable topography. *ACS Nano* **8,** 5484–5493 (2014).
40. Perkins, S. J. Protein volumes and hydration effects. *Eur. J. Biochem.* **157,** 169–180 (1986).

41. Timp, W., Comer, J. & Aksimentiev, A. DNA base-calling from a nanopore using a Viterbi algorithm. *Biophys. J.* **102,** L37–L39 (2012).
42. Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33,** 1–22 (2010).
43. Kolmogorov, M., Kennedy, E., Dong, Z., Timp, G. & Pevzner, P. Single-molecule protein identification by subnanopore sensors. Preprint at https://arxiv.org/abs/1604.02270v1 (2016).

Acknowledgements

This work was partially supported by a grant from the National Science Foundation (DBI 1256052), the Keough–Hesburgh Professorship and the Walther Cancer Foundation. We gratefully acknowledge conversations with J. Barthel and critical readings of the manuscript by W. Timp and T. Tanaka.

Author contributions

G.T. conceived the experiments. E.K. and Z.D. designed the experiments in consultation with G.T. E.K., Z.D. and C.T. performed the experiments and analysed the data. E.K. developed the algorithms used to analyse the current blockade data, in particular. G.T., E.K. and Z.D. co-wrote the manuscript, but all of the authors discussed the results and commented on it.

Additional information

Supplementary information is available in the online version of the paper. Reprints and permissions information is available online at www.nature.com/reprints. Correspondence and requests for materials should be addressed to G.T.

Competing financial interests

The authors declare no competing financial interests.

Methods

Subnanopore fabrication. Pores with subnanometre cross-sections were sputtered through thin silicon nitride membranes uing a tightly focused, high-energy electron beam carrying a current ranging from 300 to 500 pA (post-alignment) in a scanning transmission electron microscope (STEM, FEI Titan 80–300, Hillsboro, Oregon) with a Super-TWIN pole piece and a convergence angle of 10 mrad. For example, a 398 pA beam was used to sputter a nominally 0.3-nm-diameter pore in 50 s. The silicon nitride was deposited by low-pressure chemical vapour deposition directly on the top surface of a polished silicon handle wafer and a membrane was revealed using an EDP (an aqueous solution of ethylene diamine and pyrocatechol) chemical etch through a window on the polished back surface of the handle. The thickness of the membranes, which ranged from $t = 8$ to 12 nm, was measured *in situ* using electron energy loss spectroscopy before sputtering. The roughness of the membrane, measured with custom-built silicon cantilevers (Bruker, Fremont, CA) with 2 nm radius tips, was estimated to be <0.5 nm root-mean-square.

Multislice image simulations. As the information limit of the Titan STEM was 0.11 nm, to accurately assess the topography, each micrograph was imitated by multislice simulations. TEM images of the pores were simulated using the Dr. Probe software package[44]. The simulation procedure started by creating an atomistic model of the structure. An approximation to an amorphous Si_3N_4 membrane was created by randomly filling a tetragonal $5 \times 5 \times 10$ nm^3 (x–y–z) cell with Si and N atoms. The total number of atoms was determined by the volume (250 nm^3), the density of stoichiometric Si_3N_4 (3.44 g cm^{-3}) and the molecular weight of Si_3N_4 (140.28 g mol^{-1}). Atoms that were closer together than 0.16 nm were removed from the structure. To create a bi-conical pore with an elliptical cross-section at the waist, atoms were selectively extracted from the volume within a border defined by the following mathematic model:

$$\sqrt{(x-2.5+c)^2+(y-2.5)^2}+\sqrt{(x-2.5-c)^2+(y-2.5)^2}=2a$$

where $a = a_0 + \tan(\alpha)|z-5|$, $b = b_0 + \tan(\alpha)|z-5|$ and $c = \sqrt{a^2-b^2}$ and where x, y and z denote the coordinates of each atom (in nm), α is the cone angle, a/b are the major/minor axes of the ellipse and c is the eccentricity.

To prepare the model structures for the calculation of dynamic electron diffraction by means of the multislice algorithm[45], the input cells were partitioned into 40 equidistant slices along z. Phase gratings of the slices were calculated on grids with 512×512 pixels in x and y for 300 keV incident electrons using the elastic and absorptive form factors and Debye–Waller factors to account for the thermal motion of the atoms[46]. The multislice calculations yielded an exit-plane wavefunction that is consistent with the specified model of the structure. Based on the exit-plane wavefunctions, TEM images were constructed using a phase-contrast transfer function that is consistent with the microscope and defocus, assuming instrumental parameters for the aberration coefficient of $C_s = 0.9$ mm and an objective aperture size of 150 µm at an acceleration voltage of 300 kV. The optimal defocus is supposed to effectively cancel the spherical aberration to produce a contrast transfer function in which spatial frequencies higher than the point resolution can be filtered out to facilitate the interpretation of the image. Because the defocus was uncertain, a defocus series from −50 to 50 nm was calculated for comparison with the actual images. The TEM image calculations account for the partial temporal coherence with a focus spread of about 4 nm and for the partial spatial coherence with a 0.4 mrad semi-angle of convergence.

Microfluidics. The silicon chip supporting the membrane with a single pore through it was bonded to a polydimethylsiloxane (PDMS, Sylgard 184, Dow Corning) microfluidic device formed using a mould-casting technique. The PDMS microfluidic device was formed from a thoroughly stirred 10:1 mixture of elastomer (siloxane) with a curing agent (cross-linker) cast in a mould composed of DSM Somos ProtoTherm 12120 plastic (Fineline Prototyping, Raleigh, North Carolina) and then degassed and cross-linked at 75 °C for 2 h. The microfluidic device consisted of two microchannels (each 250×75 µm^2 in cross-section) connected by a via that is 75 µm in diameter. The small via was created using a fine needle to penetrate a thin PDMS layer immediately above the pore. The diameter of the via was measured relative to a micrometre calibration grid (Ted Pella, Inc) in an inverted optical microscope (Zeiss Observer Z1). The small via has the extra benefit of reducing the parasitic capacitance due to the silicon handle wafer supporting the silicon nitride membrane and thereby diminishing the dielectric component of the electrical noise.

A tight seal was formed between the silicon chip containing the silicon nitride membrane with the pore in it and the PDMS *trans* side of the microfluidic channel with a plasma-bonding process (PDS-001, Harrick Plasma, Ithaca, New York). The chip was plasma-bonded to the *trans* side of the PDMS microfluidic using a (blue-white) 25 W oxygen plasma (PDS-001, Harrick Plasma, Ithaca, New York) for 30 s. The *cis* channel was likewise sealed to a clean 75×25 mm^2 glass slide of 1 mm thickness (VWR, Radnor, PA) using the same bonding strategy. To ensure a >100 GΩ seal to the PDMS, the backside of the silicon chip was painted with PDMS, and then the ensemble was heat-treated at a temperature of 75 °C for 30–60 min. Subsequently, two separate Ag/AgCl electrodes (Warner Instruments, Hamden, Connecticut) were embedded in each channel to independently electrically address the *cis* and *trans* sides of the membrane. Likewise, the two microfluidic channels were also connected to external pressure and fluid reservoirs through polyethylene tubing at the input and output ports. The port on the *cis* side was used to convey proteins to the pore. Finally, the sealing protocol was tested against a nominally 10 nm thick silicon nitride membrane without a pore in 200 mM NaCl (pH 7.5) for >4 weeks without failure; the leakage current was <12 pA at 1 V.

Low-noise electrical measurements. To perform current measurements, the subnanopore was first wetted by immersion in a 200–300 mM NaCl electrolyte for 1 d. Subsequently, a transmembrane voltage was applied using the Ag/AgCl electrodes and the corresponding pore current was measured at 22 ± 0.1 °C using an Axopatch 200B amplifier with the output digitized by a DigiData 1440 data acquisition system (DAQ, Molecular Devices, Sunnyvale, California) at a sampling rate of 100–250 kHz. The filter cutoff frequency was not an ideal gauge of the measurement circuit bandwidth because of the quality factor (for example, even with an eight-pole, 50 kHz Bessel filter, features can still be resolved at 80 kHz). Instead, we used the impulse response as an indicator of sensitivity. It was observed that a 20 pA transient response to a 20 µs impulse in the current was clearly observable. The Clampex 10.2 (Molecular Devices, Sunnyvale, California) software was used for data acquisition and analysis. The background noise level of a pristine subnanopore was typically 12 pA-rms in a 250 mM NaCl solution at −0.7 V.

To measure a blockade current, a bias ranging from −0.3 V to −1 V was applied to the reservoir (containing 75 µl of the electrolytic solution and 75 µl of a 2× concentrated solution of the protein and denaturant) relative to the ground in the channel. The nine recombinant, carrier-free proteins used in these experiments were reconstituted according to the protocols offered by the manufacturers (Active Motif, R&D Systems and Epicypher). Typically, the protein was reconstituted at high (100 µg ml^{-1}) concentration in phosphate buffered saline (PBS) without adding BSA to avoid false readings. From this solution, aliquots diluted to 2× the concentration of the denaturant with 200–500 pM protein, 20–100 µM BME, 250 mM NaCl with $2–5 \times 10^{-3}$% SDS were vortexed and heated to 75 °C for 15–60 min. The solution was allowed to cool and added in 1:1 proportion with the (75 µl) electrolyte in the reservoir and allowed to sit for >30 min. Data was recorded in three-minute-long acquisition windows. Generally, no blockades were observed beyond the noise in controls that comprised the electrolyte and the denaturants (SDS and BME), which were heated to 75 °C and then cooled without the protein. Infrequently, short-duration events were observed in the controls, but these were easily culled due to the band-limited duration of the blockade.

Protein adsorption was observed on the silicon nitride membrane. To clear the membrane and reuse the pore, the microfluidic device was flushed with 18 MΩ de-ionized water for 2 d. To establish that it was clear of protein, the open pore current noise was evaluated again before an experiment. It was reused only if the noise returned to that observed in the pristine pore. If a pore became clogged with protein, the data set was cropped using Clampfit (Axon). When a clog occurred, both the channel and reservoir were flushed with 18 MΩ de-ionized water for at least 5 min to clear the pore. If this procedure failed to clear it, a 0.1% SDS solution was flushed through the channel to disperse latent aggregated protein in an attempt to recover the pore. In this way, sequencing data was acquired from one subnanopore for >28 d.

Signal estimation. Data handling involved five steps:

(1) Selection of blockades of sufficient duration from the raw current trace. Blockades were initially extracted from current traces recorded with a 10 kHz eight-pole Bessel filter using OpenNanopore—but not always reliably[47], and so we resorted to custom MATLAB code. These codes allowed for the manual removal of multilevel events and open pore regions that were incorrectly categorized as true events. The settings for OpenNanopore were optimized by manual inspection of the open pore noise and the blockades. ΔI, I_0 and Δt were calculated for each event. Events with sufficient duration to detect single AAs (assuming linear velocity) were selected according to the average number of intraevent peaks observed for a given protein, C, and the acquisition bandwidth cutoff, D, that is, $\tau > 2C/D$. Within this subset, blockades exhibiting a mean amplitude that was both five standard deviations (5σ) above the noise and within 10% of the mean expected per cent blockade were selected.

(2) Fitting of fluctuations within a blockade to peaks. Custom MATLAB code was written to interrogate events for an initial number of fluctuations. Fourier analysis of the blockade allowed for the detection of peak frequencies within a blockade, which were then compared with the peaks observed in the open pore current for an equivalent duration (Supplementary Fig. 6). Within a broad window, covering at minimum ±50% of the number of peaks expected for example, 300–900 peaks for BSA), the maximum peak difference between the open pore and a blockade was determined. This value (in Hertz) was then converted to an estimated number of peaks given the event duration. The average number of peaks observed for a given protein was typically less than 10% different from the known number of AAs for each protein. To validate these results, a custom algorithm was develop in MATLAB to automatically count

peaks. The algorithm worked by fitting the data to an array of Gaussian peaks. First, all of the local maxima were identified and categorized depending on whether or not they conformed to a Gaussian peak profile. Peak positions that were not close to a Gaussian maxima were discarded. Second, if two or more peaks were too close together (assigned within the boundary of the same Gaussian), all but one was discarded. The average number of peaks was then determined from all of the events for each protein. However, it was observed that the number of fluctuations found for this approach scaled with the event duration due to the increased number of noise-related peaks in longer dwelling-time events, which gave rise to the large range of the values obtained (evident in their larger standard deviations). To visualize the fluctuations, events were smoothed using a smoothing spline algorithm that is included in the MATLAB Curve Fitting Toolbox. The stiffness of the spline was adjusted until the number of fluctuations in the smoothed event equaled the number of AAs in the protein.

(3) Rescaling of blockades in time to the same number of data points and current level for averaging. Assuming that the peaks were periodic in time, all of the events were linearly re-sampled either as 10,000 data points or into N bins, where N is the average number of peaks observed per event. For the purpose of averaging, all of the events were scaled to contribute equally to the final consensus traces.

(4) Alignment of blockade translocation directions. Before averaging was performed, it was noted that blockades comprised two distinct groups, which showed similar peak maxima under temporal inversion. This observation was interpreted as evidence of two equivalent translocation directions and so all events were sorted into two groups and the second was inverted in (normalized) time. The event blockades were then renormalized according to the median blockade per cent and averaged.

(5) Renormalization of the consensus traces for comparison to the occluded volume model. The model developed for the occluded volume shows variations (in cubic nanometres) as a function of position. However, the events were recorded in picoamperes. The scaling of picoamperes to cubic nanometres was necessary to compare the model and the consensus events, and can be directly inferred using both the pore geometry and open pore current. However, we chose to linearly scale the ordinate of the events to the volume model using a Nelder–Mead method search.

Contours, maps and error assignments. Contours were created according to the density of the data points in the logarithmic duration fractional blockade space, based on a kernel density function whereby every data point contributes a two-dimensional Gaussian to the cumulative contour, which was then normalized in z such that the entire volume of all of the contributing data integrated to one. The energy distance for two such contours was calculated from the net sum of the squared differences between the two normalized density functions.

Error maps (such as those shown in Fig. 3 and Supplementary Figs 8 and 9) were used to show regions of agreement and disagreement between the model (V) and a consensus (C) as a function of read position. Regions of error were coded in greyscale. If the error was greater than a given threshold, $E = |(C_{1\ldots n} - V_{1\ldots n})/V_{1\ldots n} \geq T$, it was indicated as black and elsewhere it was indicated as grey when considered consistent with the model. Similarly, errors as a function of the read position were found by contributing the vector of the errors (as described above) at each site to all of the possible AAs recorded at that read position. For example, consider a single event where an error of 6% for read position 5 could have arisen from any part of the position trimer {4,5,6}. After these assignments, all of the possible errors on every AA were then summed and normalized to the total observed error on the event and plotted as a function of AA and AA volume for each protein.

The number of fluctuations and their amplitudes informed on the sequence by reading quadromer volumes. However, there are several qualifications required on the read accuracy. First, owing to the current crowding at the pore waist, each read probably reflects the occluded volumes associated with multiple AA residues. Thus, the number of correct reads obtained for each protein (CCL5: 65.2%, BSA: 68.4%, CXCL1: 84.7% and H3: 90%) does not reflect the accuracy with which single residues can be identified. The threshold for a correct read was chosen to be 20%, which means that (on average) ±20% of the optimally ranged and fitted random noise would fit the model because 40% of all of the data will fall within its threshold boundary. So, to what extent is the read accuracy (77% on average) statistically significant? To establish a null dataset for comparison with this value, the regions of open pore current recorded from the pure electrolyte (250 mM NaCl) were sampled with a distribution of blockade durations that reflects the measured distributions. These false events were then optimally flipped, ranged and fitted to the model for CCL5 and their read accuracies were found to have a Gaussian distribution with a mean of $\mu = 38.6\%$ and a standard deviation of $\sigma = 5\%$ for 20 runs. According to these values, the true read accuracies found for each protein were more than 6σ occurrences with respect to the noise.

FES. Simulations of vacated (open) pores that ignored the atomistic details of the structure and electrolyte were used to examine the distribution of the electrostatic potential and current (Supplementary Fig. 2). FESs of the electric field and the electro-osmotic flow were performed using COMSOL (v4.2a, COMSOL Inc., Palo Alto, California), following a Poisson–Boltzmann formalism described elsewhere[48]. Briefly, the applied transmembrane potential φ and the potential ψ due to charges in the pore and on the protein are decoupled from one another and solved independently. The relationship between ψ and the charge carriers, Na^+ and Cl^-, is given by the Poisson equation, $\nabla\psi = -\rho/\varepsilon\varepsilon_0$, where ρ, ε and ε_0 are the volume charge density and the relative and vacuum permittivities, respectively. The charge density is given by $\rho = F\sum_i z_i c_i$, where $F = 96{,}485$ C mol^{-1} is the Faraday constant, z_i is the valence and c_i is the molar concentrations of ionic species i. The distribution of ions close to charged surfaces satisfies the Boltzmann distribution; thus, the charge density is given by $c_i = c_{0,i}\exp(-z_i e\psi/k_B T)$, where c_0 is the molar concentration far from the subnanopore (that is, the bulk concentration), e is the electric charge, $k_B = 1.38 \times 10^{-23}$ J K^{-1} is the Boltzmann constant, and $T = 298$ K is the temperature.

Electro-osmotic flow is expressed by the Navier–Stokes equation, $\eta\nabla^2 u - \nabla p - F\sum_i z_i c_i \nabla V = 0$, where $V = \varphi + \psi$, η is the viscosity, p is the pressure and u is the velocity vector. The transport of ionic species is described by the Nernst–Planck equation given by $D_i\nabla^2 c_i + z_i\mu_i c_i \nabla^2 V = u \cdot \nabla c_i$, where D_i is the diffusion coefficient and μ_i is the ionic mobility of the ith species. In this treatment, u, V and c_i are coupled between equations. The relationship between the surface charges σ_s and the zeta potential ζ is given by the Grahame equation[49]: $\sigma_s(\zeta) = \sqrt{8c_0\varepsilon\varepsilon_0 k_B T}\sinh(e\zeta/2k_B T)$. The boundary conditions for the system are given in Supplementary Table 1.

Code availability. Much of the code used for analysis in this Article leverages routines available in MATLAB or online (such as mygaussfit.m, which was used to find the Gaussian parameters associated with each fluctuation). A refined subset of these codes can be found at https://github.com/fenderglass/Nano-Align, which are described in detail elsewhere (see ref. 43). These refined versions were developed in cooperation with M. Kolmogorov and P. Pevzner.

References

44. Barthel, J. Dr. Probe-High-resolution (S)TEM image simulation software. http://www.er-c.org/barthel/drprobe, version 1.6 (2015).
45. Cowley, J. M. & Moodie, A. F. The scattering of electrons by atoms and crystals. I. a new theoretical approach. *Acta. Cryst.* **10**, 609–619 (1957).
46. Weickenmeier, A. & Kohl, H. Computation of absorptive form factors for high-energy electron diffraction. *Acta Cryst.* **A 47**, 590–597 (1991).
47. Raillon, C., Granjon, P., Graf, M., Steinbocka, L. J. & Radenovic, A. Fast and automatic processing of multi-level events in nanopore translocation experiments. *Nanoscale* **4**, 4916–4924 (2012).
48. Kurz, V., Nelson, E. M., Shim, J. & Timp, G. Direct visualization of single-molecule translocations through synthetic nanopores comparable in size to a molecule. *ACS Nano* **7**, 4057–4069 (2013).
49. Grahame, D. C. The electrical double layer and the theory of electrocapillarity. *Chem. Rev.* **41**, 441–501 (1947).