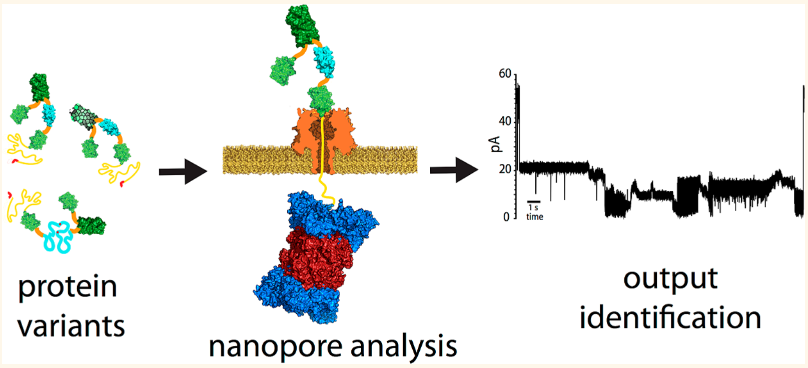
ARTICLE

Discrimination among Protein Variants Using an Unfoldase-Coupled Nanopore

Jeff Nivala, Logan Mulroney, Gabriel Li, Jacob Schreiber, and Mark Akeson*

Nanopore Group, Department of Biomolecular Engineering, University of California, Santa Cruz, California 95064, United States

ABSTRACT



Previously we showed that the protein unfoldase ClpX could facilitate translocation of individual proteins through the α-hemolysin nanopore. This results in ionic current fluctuations that correlate with unfolding and passage of intact protein strands through the pore lumen. It is plausible that this technology could be used to identify protein domains and structural modifications at the single-molecule level that arise from subtle changes in primary amino acid sequence (*e.g.*, point mutations). As a test, we engineered proteins bearing well-characterized domains connected in series along an ∼700 amino acid strand. Point mutations in a titin immunoglobulin domain (titin I27) and point mutations, proteolytic cleavage, and rearrangement of beta-strands in green fluorescent protein (GFP), caused ionic current pattern changes for single strands predicted by bulk phase and force spectroscopy experiments. Among these variants, individual proteins could be classified at 86−99% accuracy using standard machine learning tools. We conclude that a ClpXP-nanopore device can discriminate among distinct protein domains, and that sequence-dependent variations within those domains are detectable.

KEYWORDS: nanopore · protein translocation · protein sequencing · single-molecule · unfoldase · α-hemolysin

Elucidation of protein expression, structure, and function is a cornerstone of experimental biology. Traditional protein analysis tools (antibody recognition, bulk phase biochemistry, chromatography, NMR spectroscopy, X-ray crystallography) have been supplemented recently by state-of-the-art mass spectrometry.[1] Nonetheless, comprehensive proteome analysis has not achieved the throughput and impact of genome sequence analysis that has grown exponentially since the 1970s. This growth was driven initially by Sanger sequencing[2] and then by massively parallel sequencing platforms that now can deliver accurate human genome assemblies for ∼$1000 at centralized core facilities. This difference between proteomics and genomics is in part due to relative complexity.[3] But it is also a technical matter: proteins are terminal products of gene expression; therefore, protein sequencing is not amenable to template-dependent replication and amplification that is the core of commercial DNA and RNA sequencing techniques.

Nanopore analysis of proteins has gained momentum in research laboratories as the historical focus of nanopore research (DNA sequencing) approaches commercial implementation.[4−6] These nanopore experiments fall into three main categories: (i) experiments that detect protein transport and protein unfolding as domains are electrophoretically translocated through biological[7−11] or solid state[12−18] nanopores; (ii) experiments that interrogate protein−pore,[19−23] protein−protein,[24−26] and protein−ligand[27−29] interactions; and more recently (iii) experiments that detect native proteins by covalently attaching aptamers to the pore.[30,31]

* Address correspondence to makeson@soe.ucsc.edu.

Received for review September 4, 2014 and accepted November 17, 2014.

Published online November 17, 2014
10.1021/nn5049987

A relatively new addition to the first category are nanopore experiments that potentiate protein capture and unfolding by attaching a polyanion to the end of the protein strand which is captured in the electric field.[32−35] This technique has enabled the detection of unfolding intermediates[33,34] and post-translational modifications.[35]

Another strategy to drive proteins processively through nanopores is to use an enzyme motor. For example, the AAA+ enzyme ClpX is a bacterial unfoldase that recognizes proteins that have been tagged by the cell for remodeling or degradation.[36] It unfolds substrate proteins by repeated ATP-fueled mechanical pulling attempts which, when coincident with transient stochastic reductions in substrate structural stability, result in denaturation and translocation of the protein through the enzyme's narrow hexameric ring.[37,38] In an initial nanopore study, model proteins bearing a polyanion tail and the ClpX-binding ssrA tag were added to an aqueous volume bathing an alpha-hemolysin (αHL) pore in a synthetic phospholipid bilayer.[32] Individual copies of these engineered proteins were captured in the pore lumen in response to an applied voltage. ClpX hexamers in the *trans* bath (on the opposite side of the bilayer from protein substrate addition) were able to bind to the ssrA tag once it entered the *trans* compartment. ClpX-facilitated translocation of the captured protein ensued in an ATP-dependent manner. During translocation, ionic current patterns correlated with three processes: (i) rapid and unimpeded movement of unstructured segments of the captured protein through the pore lumen; (ii) pauses when the nanopore/ClpX complex contacted stable ubiquitin-like Smt3 domains engineered into the protein substrate; and (iii) sequential unfolding of the Smt3 domains coincident with translocation.

These results suggest that nanopore devices could be used for sequential protein analysis and identification. To examine this further, we designed experiments to answer two questions. First, can the nanopore device discriminate among distinct protein domains in series along individual protein strands as they are driven through the pore sensor? And second, can the nanopore device detect variations within these protein domains, *e.g.*, structural modifications arising from point mutations, truncations, and rearrangements? Such changes are common to pathogenic protein variants,[39−42] and they should be detectable if protein sequence, stability, and unfolding pathways do in fact account for the ionic current patterns we observe.

We addressed these questions using ~700 amino-acid-long engineered proteins bearing well-characterized folded domains. The motor used in these experiments was ClpXP, a proteasome-like complex of *Escherichia coli* and other prokaryotes that is capable of unfolding hundreds of different protein substrates by generating mechanical pulling forces greater than 20 pN.[36−38] We found that specific point mutations, proteolytic cleavage, and sequence rearrangements in these domains resulted in detectable ionic current pattern changes. Naive Bayes-derived decision boundaries applied to our data resulted in single protein identification at 86.4−98.7% accuracy.

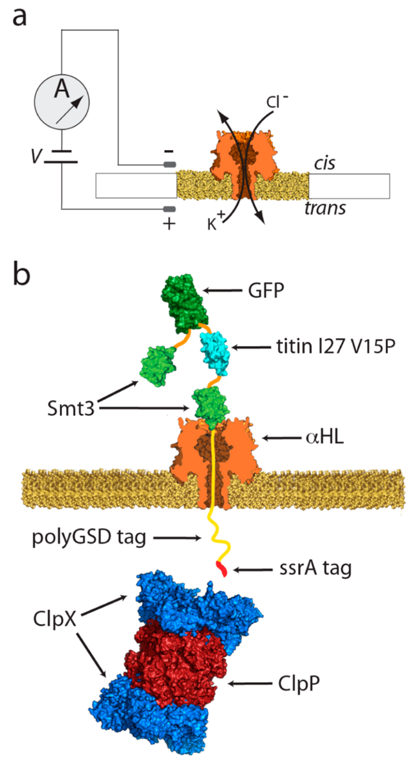


Figure 1. Experimental setup. (a) Nanopore sensor. A single αHL pore is embedded in a lipid bilayer separating two polytetra-fluoroethylene wells each containing 100 μL of 0.3 M KCl solution at 30 °C. Voltage is applied between the wells (*trans* side +180 mV), causing ionic current flow through the channel. (b) Protein S2-GT capture in the nanopore. S2-GT is a model protein bearing four folded domains: Smt3 (light green), GFP (dark green), and titin I27 V15P (cyan), coupled to a negatively charged flexible polyGSD region (yellow) and an ssrA tag (red) at its C-terminus. As a result of the applied voltage, the negatively charged polyGSD tag is threaded through the pore into the *trans*-side solution until the first folded Smt3 domain prevents further translocation of the captured protein. ClpXP present in the *trans* solution binds to the C-terminal ssrA sequence. Fueled by ATP, ClpXP translocates along the protein tail toward the channel, and catalyzes sequential unfolding and translocation of the entire multidomain protein through the pore.

RESULTS

Experimental Setup and Optimization. The standard nanopore setup is illustrated in Figure 1a. Briefly, an individual wild type αHL pore was inserted into a horizontal lipid bilayer that separates two chambers each containing ~100 μL of a 0.3 M KCl solution at pH 7.6. Ionic current through the nanopore was measured using a patch clamp instrument (Axopatch 200B) connected to a Ag/AgCl electrode in each chamber and an applied constant voltage of 180 mV. As individual polypeptide strands translocated through the pore, monovalent ionic current was reduced depending upon the composition of the protein fragment that occupied the pore lumen. The data are reported as a continuous time series of ionic current states.

In the present study, our setup included two modifications that increased the number of protein strands that could be analyzed per experiment and doubled the efficiency of translocation for each captured protein relative to our original ClpX study.[32] These were (1) addition of a conventional ATP-regeneration mixture to the *trans* compartment to maintain a constant ATP concentration over time (see Materials and Methods); and (2) supplementation of the ClpX motor with ClpP to form the ClpXP complex. In the bacterial cell, when an ssrA-tagged protein is unfolded by ClpX, it is threaded into the lumen of an associated compartmentalized peptidase, ClpP, where it is degraded.[43] We found that protein translocations driven by the ClpXP protease were less prone to long off-pathway stalls and slips than were translocations driven by ClpX alone, and doubled the fraction of complete translocation events (18.1 and 36.1% for ClpX and ClpXP, respectively). This observation is consistent with previous studies showing that ClpXP is a more robust unfoldase than is ClpX.[37,38] In addition, trimming of the substrate protein by ClpP in the *trans* compartment reduced the frequency of irreversible protein captures in the αHL pore.

Nanopore Analysis of Reference Protein S2-GT. The reference protein used for our experiments, "S2-GT", was an ~700 amino acid (aa) strand composed of four folded domains connected by short aa linkers. Based on crystal structures, the individual protein domains (ubiquitin-like protein (Smt3), titin fragment (titin I27), and green fluorescent protein (GFP)) were too large to pass through the αHL pore without unfolding of their native tertiary structures.[44−46] Each strand was capped at its carboxy terminus by an aa polyanion (polyGSD), and the ClpX-recognition ssrA motif (Figure 1b, and Supporting Information, Figure S1). At 180 mV applied potential, the polyGSD tail threaded into the pore and the ssrA tag became accessible to ClpXP in the *trans* solution (Figure 1b).

Ionic current traces that arise from ClpXP/ATP-dependent translocation of the S2-GT protein are illustrated in Figure 2a and b (also see Supporting Information, Figure S2). The capture frequency was ~0.24/s at 1 uM S2-GT and 180 mV applied voltage. Each event began at ~53 pA (open channel current Figure 2b,i), followed by a drop to ~22 pA upon S2-GT capture in the pore (Figure 2b,ii). Backflow of the protein out of the pore into the cis chamber was never observed once state ii was observed. This ionic current state persisted until ClpXP bound to the ssrA tag and began pulling on the polyGSD tail. Pulling caused a gradual current decrease (Figure 2b,iii), followed by a sudden drop to a mean current of ~9.0 pA (Figure 2b,iv) characterized by high variance (RMS noise = 2 pA). This pattern was quantitatively consistent with previous work which correlated states iii and iv with preunfolding dwell of the Smt3 atop the pore orifice, followed by ClpXP-mediated unfolding and translocation. Voltage-mediated unfolding and translocation of the C-terminal Smt3 domain of S2-GT was occasionally observed during control experiments in the absence of ClpXP or ATP. However, these events lacked the characteristic ClpX/ATP-dependent ionic current state iii. Additionally, these voltage-mediated events never proceeded past ionic current state iv and would often irreversibly clog the pore. Because a second Smt3 domain was included near the amino terminus of S2-GT, we predicted that this ionic current pattern would be repeated at the end of each complete ClpXP-mediated translocation event. This prediction was supported by our data. Notably, the last state prior to return to open channel current (Figure 2b,xi), shared nearly identical characteristics with state iv (mean current: iv = 9.0 ± 1.8 pA, xi = 8.7 ± 1.9 pA; mean RMS noise: iv = 2.0 ± 0.4 pA, xi = 1.9 ± 0.3 pA; median dwell time: iv = 680 ms Q1, Q3 = 490, 930 ms, xi = 560 ms Q1, Q3 = 310, 790 ms; also see Supporting Information, Figure S3).

Given these Smt3-dependent ionic current "bookends", it was logical that intervening ionic current states v-ix would correlate with processing of the titin I27 and GFP domains. Thus, we developed a model in which these two domains also contributed unique preunfolding and translocation current states (Figure 2c). If correct, the characteristics of each of these states should be domain-dependent, and variations within those domains should cause predictable changes in their ionic current signatures enabling their discrimination from the reference protein.

Nanopore Analysis of the Titin I27 Domain and a Destabilized Mutant. In our model, we assigned ionic current states v and vi to titin I27 because it is the next domain along the S2-GT polypeptide that would contact the nanopore following translocation of the C-terminal Smt3 domain (ionic current state iv). Quantitative evidence supports this assignment. Ionic current state v (putative preunfolding of titin I27) had a median dwell time of 22 s (Q1, Q3 = 10, 36 s). This long dwell time is consistent with single-molecule optical tweezer experiments which showed that titin I27 V15P is resistant to ClpXP-mediated unfolding (having an average preunfolding dwell time of 17 s).[47] Further, state v exhibited ionic current level repeats (arrows in Supporting Information, Figure S2) suggestive of small steps and slips of ClpXP as it attempted to advance along the polypeptide strand against a significant energy barrier. Immediately following a successful unfolding attempt, the ionic current shifted abruptly to state vi. Analysis of state vi dwell time (median = 1.3 s, Q1, Q3 = 1.0, 1.7 s) suggests that ClpXP pulls the unfolded titin I27 domain through αHL at an average

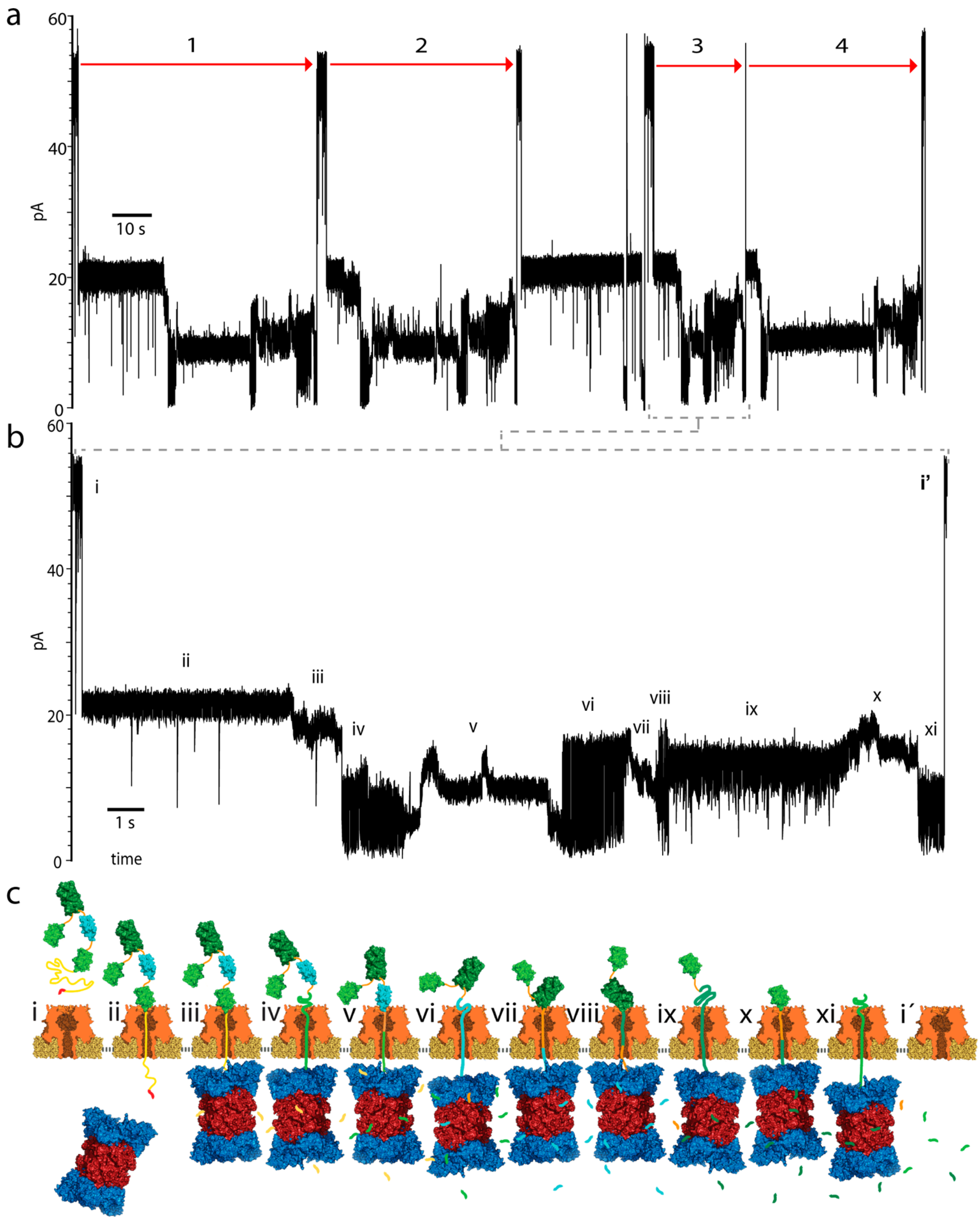


Figure 2. Ionic current traces during ClpXP-mediated protein S2-GT translocation. (a) Four consecutive S2-GT translocation events. The gap between the third and fourth events corresponds to protein captures that were ejected from the nanopore by briefly reversing voltage polarity. (b) Expanded view of ionic current states during S2-GT translocation. Open channel current through the αHL nanopore under standard conditions (mean ∼53 pA, current RMS noise 1.2 pA) (i). Initial capture of the S2-GT substrate (mean current ∼22 pA, current RMS noise 0.7 pA) (ii). ClpXP-mediated C-terminal Smt3 preunfolding (mean current ∼19 pA, current RMS noise 1.7 pA) (iii). C-terminal Smt3 domain unfolding and translocation through the nanopore (mean current ∼9.6 pA, current RMS noise 2.0 pA) (iv). Ionic current transition into the titin I27 V15P preunfolding state. Several discrete current levels are typically observed (mean current ∼9.5 pA, current RMS noise 2.3 pA) (v). Unfolding and translocation of the titin I27 V15P domain through the nanopore (mean current ∼14 pA, current RMS noise 4.6 pA) (vi). The GFP preunfolding state. Several discrete current levels are typically observed (mean current ∼14 pA, current RMS noise 2.0 pA) (vii). Extraction of the C-terminal beta strand 11 of GFP (mean current ∼11 pA, current RMS noise 3.7 pA) (viii). Global unfolding and translocation of GFP (mean current ∼15 pA, current RMS noise 1.4 pA) (ix). N-terminal Smt3 preunfolding state (mean current ∼15 pA, current RMS noise 1.7 pA) (x). N-terminal Smt3 domain unfolding and translocation through the nanopore (mean current ∼9.3 pA, current RMS noise 2.0 pA) (xi). Return to open channel current upon completing translocation of the entire S2-GT protein to the *trans* compartment (i′). (c) Working model of ClpXP-mediated S2-GT translocation. Roman numerals assigned to each panel correspond to ionic current states in b.

rate of 64 aa/s. This rate is similar to previous studies that established a maximum ClpXP translocation rate of 60−70 aa/s.[38,48]

If these assignments are valid, it follows that changes in the stability of the titin I27 domain would result in detectable changes in ionic current state v.

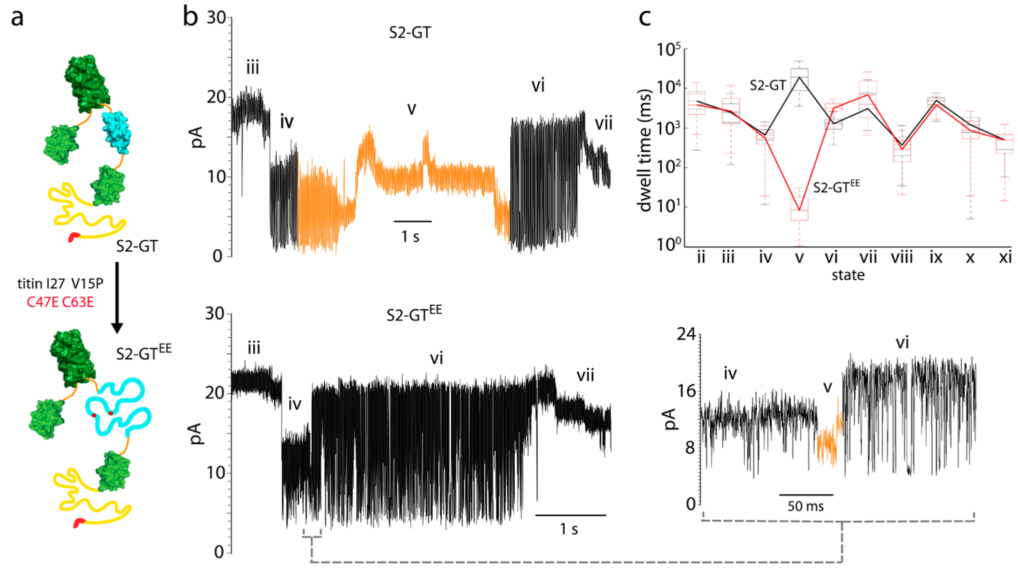


Figure 3. Ionic current state v is dramatically changed by two point mutations in the titin I27 V15P domain of S2-GT. (a) Cartoon depiction of the two proteins that were compared in this experiment. S2-GT is at the top and a modified version bearing two point mutations within the titin I27 V15P domain (S2-GT^{EE}: C47E C63E) is at the bottom. Modifications at C47 and C63 are known to destabilize titin I27 tertiary structure. (b) Ionic current states iii-vii of representative S2-GT (top) and S2-GT^{EE} (bottom) translocation events. State v for each event is colored orange. (c) A parallel coordinates plot comparing median dwell times for ionic current states ii–xi of S2-GT (black, *n* = 91 translocations) and S2-GT^{EE} (red, *n* = 93 translocations). The median state v dwell time of S2-GT^{EE} is ~3.5 orders of magnitude shorter than the comparable state v median dwell time of S2-GT (7.9 ms (Q1, Q3 = 4.4, 16.7 ms) and 22 s (Q1, Q3 = 10, 36 s), respectively). The boxes show the 1st, 2nd, and 3rd quartile median values. The whiskers extend to the lowest and highest value within 1.5 IQR of the 1st and 3rd quartile medians, respectively.

As a test, we constructed an S2-GT variant (S2-GT^{EE}) where two buried cysteines (C47, C63) were mutated to glutamic acid residues (Figure 3a). These side chain alterations are similar to carboxymethylation of C47/C63 and mutation of those cysteines to aspartic acid that are known to destabilize the titin I27 domain.[49,50] As anticipated, S2-GT^{EE} state v dwell times were several orders of magnitude shorter than were S2-GT state v dwell times (Figure 3b, Supporting Information, Figures S4 and S5). Additionally, the S2-GT^{EE} state vi median dwell time was ~2 times longer than that of the parent structure (median = 3.4 s, Q1, Q3 = 3.0, 4.4. Supporting Information, Figure S5). One explanation for this observation is that (partial) refolding of protein domains in the *trans* compartment helps to drive translocation of the domain through the pore. Such an effect would be attenuated in the destabilized titin I27 variant. The other ionic current states remained relatively unchanged between the two constructs (Figure 3c, Supporting Information, Figure S6).

Nanopore Analysis of the GFP Domain and a "Superfolder" Variant. After translocation of titin I27, the next domain along the S2-GT strand is GFP. Accordingly, we predicted that states vii–ix would correlate with preunfolding and translocation of GFP (Figure 2b and 2c). As was the case for ionic current state v (titin preunfolding), ionic current state vii often contained repeated current levels suggestive of small steps and slips of ClpXP as it attempted to unfold GFP. This state (mean current 13.4 pA, current RMS noise 2.0 pA) had a median dwell time of 3.9 s (Q1, Q3 = 2.1, 8.7 s. Also see Supporting Information, Figure S7), similar to the median unfolding time found in previous single-molecule studies of ClpXP-mediated unfolding of GFP (~6 s).[51] State vii ended with an abrupt and irreversible transition to state viii (mean current 11.0 pA, current RMS noise 3.1 pA, and median dwell time 270 ms, Q1, Q3 = 120, 550 ms). This state is consistent with a step along the GFP unfolding pathway that corresponds to extraction of the 11th beta-strand that precedes global GFP unfolding.[48,51] State viii was followed by a distinct shift to state ix characterized by a higher mean ionic current (14.8 pA) with relatively low noise (RMS noise = 1.4 pA). We reasoned that state ix corresponds to translocation of the unfolded GFP domain through αHL at 34–58 aa/s following successful ClpXP-mediated GFP unfolding (Supporting Information, Figure S7).

As an initial test to determine if altering the GFP domain would change ionic current states vii–ix, we engineered an S2-GT construct (S2-G^{SF}T) in which the GFP domain was replaced by a "superfolding" GFP variant (Figure 4a and Supporting Information, Figure S1). Superfolder GFP (GFP^{SF}) contains 11 point mutations which increase its resistance to chemical denaturants and which help maintain GFP fluorescence when beta-strands that form the functional core are permuted.[52] As anticipated, the characteristics of S2-G^{SF}T events were altered relative to S2-GT events (Figure 4b and Supporting Information, Figures S8 and S9). Most notably, S2-G^{SF}T events exhibited a three

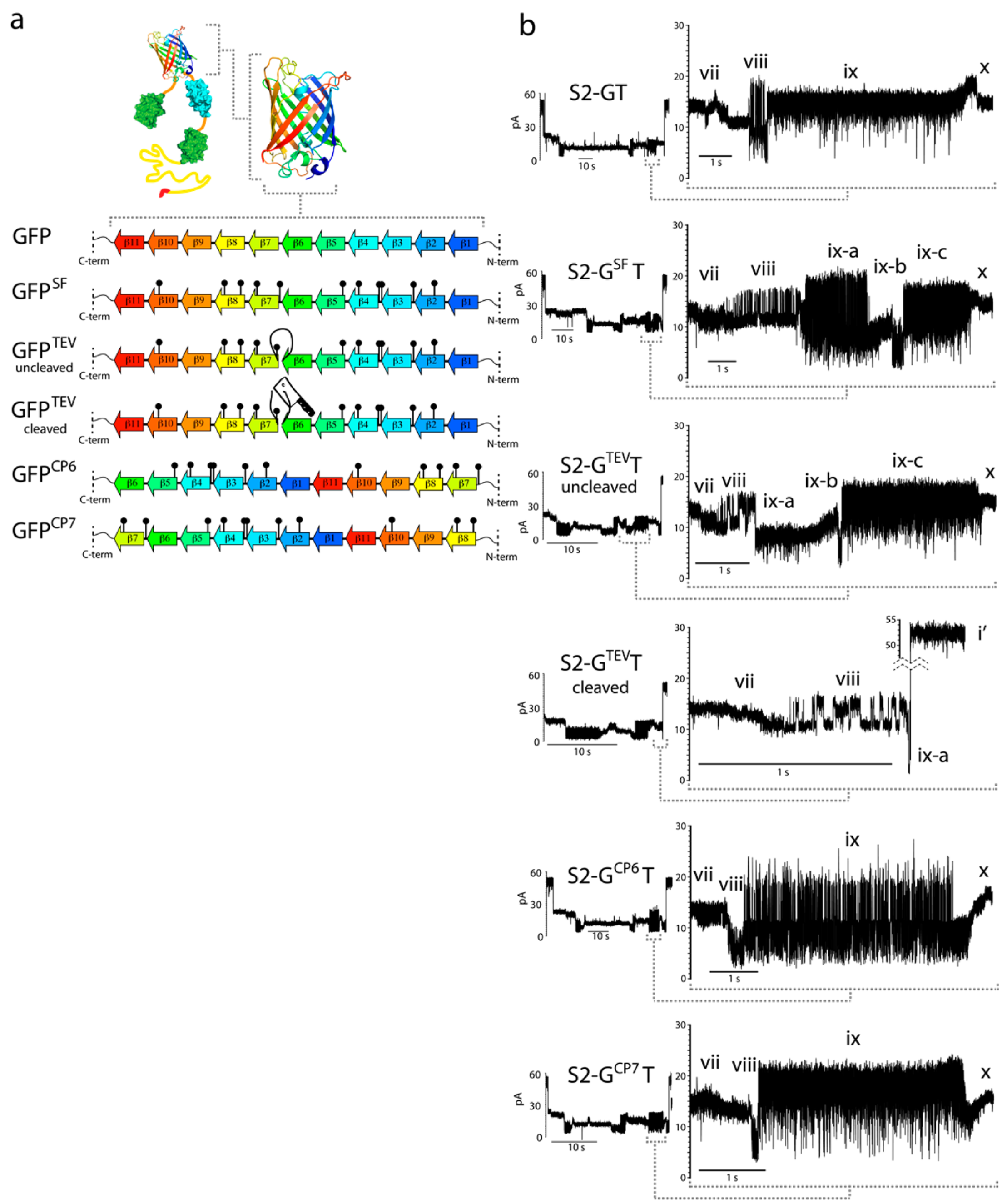


Figure 4. Ionic current signatures for S2-GT GFP variants. (a) Beta-strand connectivity of the GFP domain within proteins S2-GT (GFP), S2-G^{SF}T (GFP^{SF}), cleaved and uncleaved S2-G^{TEV}T, S2-G^{CP6}T (GFP^{CP6}), and S2-G^{CP7}T (GFP^{CP7}). Each colored arrow represents a beta-strand. GFP^{SF} (superfolder GFP) contains 11 point mutations (black markers) compared to GFP. GFP^{TEV} contains a TEV protease cleavage site between the 6th and 7th beta-strands of GFP^{SF}. GFP^{CP6} and GFP^{CP7} are circular permutations of GFP^{SF} between the 6th/7th and 7th/8th beta-strands, respectively. (b) Representative ClpXP-mediated GFP variant translocation events with expanded views of GFP-dependent ionic current states vii–ix. Ionic current states vii and viii are preunfolding of GFP. State ix is translocation of the unfolded GFP domain. Ionic current states ix for S2-G^{SF}T and S2-G^{TEV}T include three unique substates (ix-a, ix-b, and ix-c) that correspond to translocation of unfolded beta-strands 11 → 7 (ix-a), preunfolding of an intermediate (ix-b), and translocation of the unfolded intermediate beta-strands 6 → 1 (ix-c). Cleavage of S2-G^{TEV}T with TEV protease terminates the event following a brief ionic current state ix-a.

level ionic current pattern within state ix (states ix-a, ix-b, and ix-c) that was absent in S2-GT events. We argue that this pattern reflects additional preunfolding (ix-b) and translocation (ix-c) states arising from an additional GFP unfolding intermediate. This argument is based on two facts: (1) Single-molecule optical tweezer experiments have revealed a short-lived GFP unfolding intermediate in which beta-strands 6 → 1 maintain their tertiary structure following initial ClpXP-mediated unfolding of GFP beta-strands 11 → 7.[38] (2) The combined dwell times of states ix-a (median = 1.5 s, Q1, Q3 = 1.2, 2.0) and ix-c (median = 3.5 s, Q1, Q3 = 2.8, 4.2)

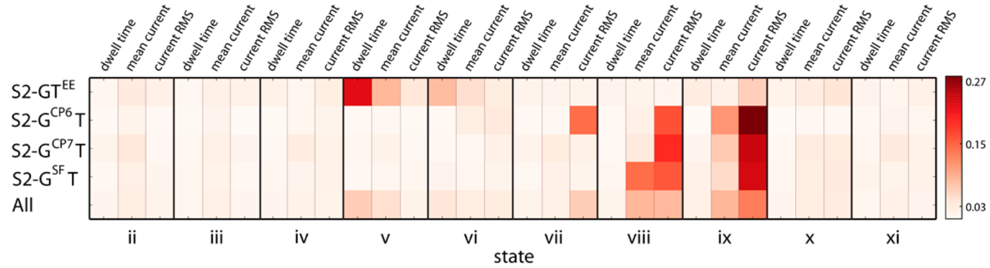


Figure 5. Identification of ionic current states important for discriminating between S2-GT variants. Each column represents a feature (dwell time, mean current, or current RMS noise) for each current state (ii–xi). Each row is one of the S2-GT variants compared against S2-GT using a binary comparison, with the exception of the last row, which is a multiclass comparison of all the S2-GT constructs together. Each row is normalized and sums to 1. The "heat" of each square (scale at right) represents the relative importance of that feature as determined by a forest of extremely randomized trees (see Materials and Methods). State v dwell time was the most important feature for discriminating between proteins S2-GT and S2-GT^{EE}. The current RMS noise of states viii and ix were the most important for discriminating between S2-GT and GFP variant proteins S2-G^{SF}T, S2-G^{CP6}T, and S2-G^{CP7}T.

is 5.0 s, which is similar to the dwell time of state ix (median = 5.5 s, Q1, Q3 = 4.3, 6.3) for translocation of GFP beta-strands 11 → 1 in S2-GT events (Supporting Information, Figure S10). Thus, ionic current states ix-a and ix-c are consistent with sequential translocation of GFP^{SF} beta-strands 11 → 7 and 6 → 1 separated by an intermediate preunfolding state (ix-b) that is not observable in the S2-GT strand.

Protease Cleavage of "Superfolder" GFP. If ionic current states ix-a and ix-b/c are dependent on GFP^{SF} beta-strands 11 → 7 and 6 → 1, respectively, then cleaving the polypeptide chain between the two regions should terminate translocation events at state ix-a. To test this, we inserted a Tobacco Etch Virus (TEV) protease cleavage site between the sixth and seventh beta-strands of GFP^{SF} (protein S2-G^{TEV}T, Figure 4a and Supporting Information, Figure S1). Consistent with our prediction, the uncleaved protein retained states ix-b → xi, while cleavage with TEV protease resulted in events terminating at state ix-a (Figure 4b, Supporting Information, Figures S11 and S12).

Surprisingly, the cleaved GFP^{SF} beta-strands 11 → 7 translocation rate (∼6000 aa/s) was much faster than expected for ClpXP-mediated translocation (∼50 aa/s). A likely explanation is that cleavage and separation of the 275 N-terminal amino acids in the *cis* compartment reduced hydrodynamic drag on the translocating strand.[53] This would allow the relatively weaker electrophoretic force to drive translocation of the cleaved strand at a high rate absent ClpXP activity.

Structural Rearrangements of "Superfolder" GFP. If ionic current state ix is sensitive to GFP sequence as we assert, then rearrangement of beta-strand order should cause the state ix current pattern to change as well. To examine this, we constructed variants of the S2-G^{SF}T protein in which the GFP^{SF} domain was circularly permuted between the sixth and seventh beta-strands (protein S2-G^{CP6}T) and between the seventh and eighth beta-strands (S2-G^{CP7}T) (Figure 4a and Supporting Information, Figure S1).[50] Representative ClpXP-mediated ionic current traces for proteins S2-G^{CP6}T and S2-G^{CP7}T are shown in Figure 4b (see also Supporting Information, Figures S13 and S14). State ix of S2-G^{CP6}T and S2-G^{CP7}T events differed from the three state current pattern observed in S2-G^{SF}T events, displaying a single nearly homogeneous current state similar to S2-GT events in dwell time, but differing in current mean and current variance (Figure 4b and Supporting Information, Figures S15 and S16). These results are consistent with a model where ionic state ix is sensitive to the sequence topology of GFP translocation. State viii characteristics for S2-G^{CP6}T and S2-G^{CP7}T events also differed from S2-GT and S2-G^{SF}T events (Figure 4b). This is not surprising because these structural rearrangements altered the identity of the first beta-strand extracted during unfolding.

Discrimination among Protein Variants Using Naive Bayes Classifiers. One motivation for this research was to develop a nanopore device that uses sequential ionic current measurements to identify individual proteins. As a test of the ClpXP-αHL prototype, we quantified the accuracy of calls among five of the S2-GT variants examined in this study.

Values for three parameters (dwell time, average current amplitude, and standard deviation of the current amplitude) were collected for states ii-to-xi within each complete translocation event (S2-GT: 91 events, S2-GT^{EE}: 93 events, S2-G^{SF}T: 78 events, S2-G^{CP6}T: 73 events, and S2-G^{CP7}T: 82 events). To determine which of these 30 features were useful for protein classification, we performed a random forest analysis (see Materials and Methods). First, pairwise comparisons were performed between the null construct, S2-GT, and each of the four variants independently (Figure 5). As expected, the most important feature for distinguishing between S2-GT and S2-GT^{EE} was dwell time for the titin I27 preunfolding state (v). Also, as expected, features that were important for distinguishing between S2-GT and its GFP variants centered on states vii–ix (preunfolding and translocation of GFP). In these cases, average current amplitude

TABLE 1. Confusion Matrix for Discriminating between S2-GT Variants Using a Multiclass Naive Bayes Classifier[a]

	predicted class				
actual class	S2-GT	S2-GT^{EE}	S2-G^{CP6}T	S2-G^{CP7}T	S2-G^{SF}T
S2-GT	**98.68**	0.05	0.05	1.15	0.05
S2-GT^{EE}	1.13	**96.57**	1.13	1.13	0.05
S2-G^{CP6}T	0.07	0.07	**95.63**	2.80	1.43
S2-G^{CP7}T	0.06	0.06	12.22	**86.38**	1.28
S2-G^{SF}T	2.62	0.06	0.06	3.90	**93.35**

[a] Each cell represents the percent probability of classifying a particular S2-GT variant (left column labels) as any of the five variants (top row labels). The diagonal (bold text) represents the correct classification.

and current standard deviation proved to be important features.

We then reframed the question and asked which features were important for classification of a given translocation event when comparing all five protein constructs against one another simultaneously. Predictably, this analysis yielded eight useful features (row labeled "All" in Figure 5) that were a composite of the features identified by pairwise comparisons.

To estimate the accuracy with which we could call a given translocation event, we used these eight pertinent features and a Naive Bayes classifier[54] to establish a confusion matrix for S2-GT and the four variants compared in Figure 5. Naive Bayes classifiers are suitable for this data set because they do not require tuning of parameters and thus avoid unnecessary complexity for preliminary tests. Upon building a confusion matrix using maximum *a posteriori* estimates, we found that there was an 86.4−98.7% chance of making an accurate call for each protein variant (Table 1).

DISCUSSION

This study was motivated by two questions pertaining to sequential protein analysis using nanopores. First, can we distinguish between different protein domains in series along individual strands as they are driven through the αHL pore by an enzyme motor? Two lines of evidence indicate that we can. (i) Preunfolding states differed among Smt3, GFP, and titin I27. For example, the preunfolding dwell time for titin I27 was substantially longer than preunfolding dwell times for GFP and Smt3 (Figure 3c, and Supporting Information, Figures S3, S5, and S7). This is consistent with titin I27's characteristic resistance to mechanical denaturation.[47,49] Further, GFP displayed a three-state unfolding pathway, distinct from titin I27 and Smt3 two-state pathways (Figure 4b). (ii) Overall, individual domain translocation states had ionic current signatures that were quantitatively distinguishable from one another based on current mean and RMS noise, while dwell times were consistent with domain size (Figure 2b and Supporting Information, Figures S6, S9, S10, S15, S16, and Table S1).

Second, can we detect variants of these domains in single proteins, *e.g.*, structural modifications arising from point mutations, truncations, and rearrangements? A number of results demonstrate that we can. (i) Destabilizing point mutations within titin I27 caused predictable changes to its ionic current pattern (S2-GT^{EE}, Figure 3). (ii) Eleven point mutations within GFP^{SF} (S2-G^{SF}T) modified its unfolding dynamics relative to GFP (S2-GT), and allowed us to identify a second unfolding intermediate within GFP^{SF} (Figure 4b, state ix-b). (iii) Proteolytic cleavage of S2-G^{TEV}T truncated S2-G^{TEV}T translocation events at a predicted position within the GFP-dependent ionic current state (ix, Figure 4b). (iv) Circular permutations of GFP (S2-G^{CP6}T and S2-G^{CP7}T) resulted in ionic current signatures that differed from the reference protein, and that did not display unfolding intermediates found in S2-G^{SF}T (Figure 4b, states xiii and ix-b). (v) A Naive-Bayes classifier demonstrated our ability to discriminate between these variants (Table 1).

Three of the protein variants we tested are representative of variant classes commonly associated with disease states:[39,40] (i) The destabilizing point mutations we analyzed in titin I27 (S2-GT^{EE}) are similar to point mutations within titin Ig domains that cause cardiomyopathy.[55,56] (ii) Mutations that stabilize protein unfolding intermediates (similar to the intermediate observed in superfolder GFP (GFP^{SF})) contribute to diseases such as amyloidosis.[57] (iii) A truncated variant (cleaved S2-G^{TEV}T) demonstrate that the ClpXP-nanopore device can detect truncations derived from early termination or proteolytic processing that are each associated with disease.[42] Additionally, the circular permutants (S2-G^{CP6}T and S2-G^{CP7}T) suggest that we could also detect chimeric oncoproteins[58] and isoforms associated with cancer.[41]

One limitation of the current technology is that the polyGSD-ssrA tag needed for capture and ClpXP binding was engineered into the expressed proteins we analyzed. Practical applications will require a method to conjugate the tag to endogenous proteins.[59,60] A second limitation of our current approach is that we are not able to predict *a priori* the ionic current pattern for a given protein, and must instead rely upon patterns established empirically in each case. However, as the number of analyzed proteins increases, we expect that ionic current patterns will emerge that are characteristic of domain classes. This could facilitate assembly of composite patterns for *de novo* protein identification.

CONCLUSIONS

In summary, an unfoldase-coupled nanopore sensor can discriminate among distinct protein domains and among variants of those domains. Compared to protein mass spectrometry, the ClpXP-nanopore device has the advantages of single-molecule resolution,[61] and analysis of unfragmented protein strands. These could be important because an estimated 2/3rds to 4/5ths of eukaryotic proteins are comprised of multiple domains,[62] with each domain having potentially many unique modified forms.[63]

MATERIALS AND METHODS

ClpX Expression and Purification. A covalently linked trimer of an N-terminal truncated ClpX variant (ClpX-ΔN3) was used for all ClpX nanopore experiments.[64] ClpX protein expression was induced at an A_{600} of ~0.6 by addition of 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), and incubated at 23 °C with shaking for 3 h. Cultures were pelleted, resuspended in lysis buffer (50 mM NaH_2PO_4 pH 8, 300 mM NaCl, 100 mM KCl, 20 mM imidazole, 10% glycerol, 1 mM dithiothreitol (DTT)) and lysed by vortexing with glass beads. After centrifugation and filtration of the lysate, the protein was purified on a Ni^{2+}-NTA affinity column (Thermo) and an Uno-Q anion exchange column (Bio-Rad). ClpX was then flash frozen in small aliquots and stored at −80 °C.

ClpP and S2-GT Constructs Expression and Purification. DNA for the GFP and titin I27 domains of fluorescent protein S2-GT were extracted by PCR from a GFP-titin-I27 V15P-ssrA expression vector obtained from A. Martin (UC Berkeley), and cloned into the S2−35 vector by Gibson assembly. GFP^{SF}, GFP^{CP6} and GFP^{CP7} DNA was obtained from A. Nager and K. Schmitz (MIT) in the form of expression plasmids and subsequently PCR-extracted and cloned into the S2-GT expression plasmid by replacement of the S2-GT GFP domain *via* Gibson assembly. The $S2\text{-}GT^{EE}$ and $S2\text{-}G^{TEV}T$ mutants were constructed by Gibson assembly using mutagenic oligos and PCR. These engineered proteins and a his-tagged ClpP were expressed in BL21 (DE3)*. Expression was induced at ~0.6 A_{600} by addition of 0.5 mM IPTG, and incubated at 30 °C with shaking for 3−4 h. Cultures were pelleted, resuspended in lysis buffer and lysed *via* vortexing with glass beads. After centrifugation and filtration of the lysate, the protein was purified on a Ni^{2+}-NTA affinity column (Thermo) and an SD200 size exclusion column (GE). Cleaved protein $S2\text{-}G^{TEV}T$ was digested with TurboTEV (Nacalai USA) for 24 h at room temperature. The proteins were flash frozen in small aliquots following purification and stored at −80 °C.

Nanopore Experiments. All experiments were performed in protein translocation buffer (PT buffer) containing 300 mM KCl, 10 mM $MgCl_2$, 10% glycerol, 1 mM DTT, 1 mM EDTA, 5 mM ATP, and 10 mM HEPES-KOH pH 7.6. Setup of the nanopore device and insertion of an αHL nanopore into a lipid bilayer have been described.[65] Briefly, a single αHL nanopore is inserted into a lipid bilayer that separates two wells each containing 100 μL of PT buffer. The ionic current through the nanopore was measured between Ag/AgCl electrodes in series with an integrating patch clamp amplifier (Axopatch 200B, Molecular Devices) in voltage clamp mode with a constant 180 mV potential across the bilayer. Data were recorded at 100 kHz bandwidth in whole cell configuration using an analog-to-digital converter (Molecular Devices), then filtered at 2 kHz using an analog lowpass Bessel filter. Experimental conditions were prepared by the daily preparation of PT buffer. $ClpX_6$ and $ClpP_{14}$ were diluted in PT buffer for a final concentration of 100 nM and 300 nM, respectively, in addition to an ATP regeneration mix (8 mM creatine phosphate and 0.08 mg/mL creatine phosphokinase). The ClpXP/ATP-regeneration mix solution was used to fill the entire system before isolation of a single αHL nanopore. Upon insertion, the *cis* well was perfused with ~6 mL PT buffer. Experiments were conducted at 30 °C with ~1 μM substrate added to the *cis* well. In the presence of ClpXP, protein substrate capture events were ejected with reverse polarity due to pore clogs or after a predetermined duration. Capture events that remained in state ii for a duration of ~45 s or longer were considered to be inaccessible to ClpXP and were ejected. Any voltage-induced translocations (events lacking state iii) were ejected to prevent clogging and to increase the efficiency of data collection.

Feature Selection. Dwell times for each state were determined by manual inspection and segmentation of the ionic current traces according to changes in the ionic current mean and RMS noise. Example segment lines appear in the characteristic S2-GT traces shown in Supporting Information, Figure S2.

For every event, the mean, standard deviation, and duration of states ii through xi were taken, resulting in 30 features and 417 events. The identity of each event is also known, since data on each protein variant were collected individually. The Gini importance was calculated for each feature using a forest of extremely randomized trees. Features that performed higher than a null model assuming equal importance for each feature were kept. The open-source scikit-learn[66] (version 0.14.1) command used to build the forest of extremely randomized trees was as follows:

ExtraTreesClassifier (n_estimators = 500, max_features = 6, max_depth = None, min_samples_split = 1, random_state = 42).

Naive Bayes Classification. Using feature selection, eight features were selected automatically as being useful in a five-class classification, with each protein variant being a class. We estimated the error rate using stratified 5-fold cross validation with a Gaussian Naive Bayes classifier. Simply, the classifier is trained on 80% of the data, and predicts labels for the remaining 20%. This process is repeated four more times using a unique 20% each time, predicting labels for each event. This strategy was performed a total of 20 times to ensure the accuracy measurement was not an outlier. A confusion matrix is generated by comparing the predicted labels to the known identity of each event and using Maximum *A Posteriori* (MAP) estimates with a symmetric Dirichlet prior in which one is added to each of the counts before normalization to get probabilities for each cell.

Conflict of Interest: The authors declare the following competing financial interest(s): Oxford Nanopore Technologies has licensed intellectual property related to this work. M.A. is a consultant to Oxford Nanopore Technologies.

Acknowledgment. Oxford Nanopore Technologies (Oxford, UK) supplied alpha-hemolysin nanopores and funded this work (SC20130149). ClpXP-related expression plasmids were obtained from A. Martin (UC Berkeley). A. Nager and K. Schmitz (MIT) provided genes containing circular permutations of GFP. D. B. Marks performed preliminary nanopore experiments.

Supporting Information Available: Figures S1−S13 contain protein sequences, additional example ionic current traces, and supporting data figures. This material is available free of charge *via* the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

1. Henzel, W. J.; Billeci, T. M.; Stults, J. T.; Wong, S. C.; Grimley, C.; Watanabe, C. Identifying Proteins from Two-Dimensional Gels by Molecular Mass Searching of Peptide Fragments in Protein Sequence Databases. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 5011–5015.
2. Sanger, F.; Nicklen, S.; Coulson, A. R. DNA Sequencing with Chain-Terminating Inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **1977**, *74*, 5463–5467.
3. Altelaar, A. F. M.; Munoz, J.; Heck, A. J. R. Next-Generation Proteomics: Towards an Integrative View of Proteome Dynamics. *Nat. Rev. Genet.* **2013**, *14*, 35–48.
4. Movileanu, L. Interrogating Single Proteins through Nanopores: Challenges and Opportunities. *Trends Biotechnol.* **2009**, *27*, 333–341.
5. Oukhaled, A.; Bacri, L.; Pastoriza-Gallego, M.; Betton, J.-M.; Pelta, J. Sensing Proteins through Nanopores: Fundamental to Applications. *ACS Chem. Biol.* **2012**, *7*, 1935–1949.

6. Cressiot, B.; Oukhaled, A.; Bacri, L.; Pelta, J. Focus on Protein Unfolding Through Nanopores. *Bionanoscience* **2014**, *4*, 111–118.
7. Oukhaled, G.; Mathé, J.; Biance, A.-L.; Bacri, L.; Betton, J.-M.; Lairez, D.; Pelta, J.; Auvray, L. Unfolding of Proteins and Long Transient Conformations Detected by Single Nanopore Recording. *Phys. Rev. Lett.* **2007**, *98*, 158101.
8. Pastoriza-Gallego, M.; Rabah, L.; Gibrat, G.; Thiebot, B.; van der Goot, F. G.; Auvray, L.; Betton, J.-M.; Pelta, J. Dynamics of Unfolded Protein Transport through an Aerolysin Pore. *J. Am. Chem. Soc.* **2011**, *133*, 2923–2931.
9. Merstorf, C.; Cressiot, B.; Pastoriza-Gallego, M.; Oukhaled, A.; Betton, J.-M.; Auvray, L.; Pelta, J. Wild Type, Mutant Protein Unfolding and Phase Transition Detected by Single-Nanopore Recording. *ACS Chem. Biol.* **2012**, *7*, 652–658.
10. Payet, L.; Martinho, M.; Pastoriza-Gallego, M.; Betton, J.-M.; Auvray, L.; Pelta, J.; Mathé, J. Thermal Unfolding of Proteins Probed at the Single Molecule Level Using Nanopores. *Anal. Chem.* **2012**, *84*, 4071–4076.
11. Mereuta, L.; Roy, M.; Asandei, A.; Lee, J. K.; Park, Y.; Andricioaei, I.; Luchian, T. Slowing down Single-Molecule Trafficking through a Protein Nanopore Reveals Intermediates for Peptide Translocation. *Sci. Rep.* **2014**, *4*, 3885.
12. Fologea, D.; Ledden, B.; McNabb, D. S.; Li, J. Electrical Characterization of Protein Molecules by a Solid-State Nanopore. *Appl. Phys. Lett.* **2007**, *91*, 539011–539013.
13. Talaga, D. S.; Li, J. Single-Molecule Protein Unfolding in Solid State Nanopores. *J. Am. Chem. Soc.* **2009**, *131*, 9287–9297.
14. Firnkes, M.; Pedone, D.; Knezevic, J.; Döblinger, M.; Rant, U. Electrically Facilitated Translocations of Proteins through Silicon Nitride Nanopores: Conjoint and Competitive Action of Diffusion, Electrophoresis, and Electroosmosis. *Nano Lett.* **2010**, *10*, 2162–2167.
15. Oukhaled, A.; Cressiot, B.; Bacri, L.; Pastoriza-Gallego, M.; Betton, J.-M.; Bourhis, E.; Jede, R.; Gierak, J.; Auvray, L.; Pelta, J. Dynamics of Completely Unfolded and Native Proteins through Solid-State Nanopores as a Function of Electric Driving Force. *ACS Nano* **2011**, *5*, 3628–3638.
16. Cressiot, B.; Oukhaled, A.; Patriarche, G.; Pastoriza-Gallego, M.; Betton, J.-M.; Auvray, L.; Muthukumar, M.; Bacri, L.; Pelta, J. Protein Transport through a Narrow Solid-State Nanopore at High Voltage: Experiments and Theory. *ACS Nano* **2012**, *6*, 6236–6243.
17. Plesa, C.; Kowalczyk, S. W.; Zinsmeester, R.; Grosberg, A. Y.; Rabin, Y.; Dekker, C. Fast Translocation of Proteins through Solid State Nanopores. *Nano Lett.* **2013**, *13*, 658–663.
18. Larkin, J.; Henley, R. Y.; Muthukumar, M.; Rosenstein, J. K.; Wanunu, M. High-Bandwidth Protein Analysis Using Solid-State Nanopores. *Biophys. J.* **2014**, *106*, 696–704.
19. Movileanu, L.; Schmittschmitt, J. P.; Scholtz, J. M.; Bayley, H. Interactions of Peptides with a Protein Pore. *Biophys. J.* **2005**, *89*, 1030–1045.
20. Mohammad, M. M.; Prakash, S.; Matouschek, A.; Movileanu, L. Controlling a Single Protein in a Nanopore through Electrostatic Traps. *J. Am. Chem. Soc.* **2008**, *130*, 4081–4088.
21. Mohammad, M. M.; Movileanu, L. Excursion of a Single Polypeptide into a Protein Pore: Simple Physics, but Complicated Biology. *Eur. Biophys. J.* **2008**, *37*, 913–925.
22. Niedzwiecki, D. J.; Grazul, J.; Movileanu, L. Single-Molecule Observation of Protein Adsorption onto an Inorganic Surface. *J. Am. Chem. Soc.* **2010**, *132*, 10816–10822.
23. Singh, P. R.; Bárcena-Uribarri, I.; Modi, N.; Kleinekathöfer, U.; Benz, R.; Winterhalter, M.; Mahendran, K. R. Pulling Peptides Across Nanochannels: Resolving Peptide Binding and Translocation through the Hetero-Oligomeric Channel from Nocardia Farcinica. *ACS Nano* **2012**, *6*, 10699–10707.
24. Sexton, L. T.; Horne, L. P.; Sherrill, S. A.; Bishop, G. W.; Baker, L. A.; Martin, C. R. Resistive-Pulse Studies of Proteins and Protein/Antibody Complexes Using a Conical Nanotube Sensor. *J. Am. Chem. Soc.* **2007**, *129*, 13144–13152.
25. Han, A.; Creus, M.; Schürmann, G.; Linder, V.; Ward, T. R.; de Rooij, N. F.; Staufer, U. Label-Free Detection of Single Protein Molecules and Protein−Protein Interactions Using Synthetic Nanopores. *Anal. Chem.* **2008**, *80*, 4651–4658.
26. Madampage, C. A.; Andrievskaia, O.; Lee, J. S. Nanopore Detection of Antibody Prion Interactions. *Anal. Biochem.* **2010**, *396*, 36–41.
27. Movileanu, L.; Howorka, S.; Braha, O.; Bayley, H. Detecting Protein Analytes That Modulate Transmembrane Movement of a Polymer Chain Within a Single Protein Pore. *Nat. Biotechnol.* **2000**, *18*, 1091–1095.
28. Lieberman, K. R.; Cherf, G. M.; Doody, M. J.; Olasagasti, F.; Kolodji, Y.; Akeson, M. Processive Replication of Single DNA Molecules in a Nanopore Catalyzed by Phi29 DNA Polymerase. *J. Am. Chem. Soc.* **2010**, *132*, 17961–17972.
29. Harrington, L.; Cheley, S.; Alexander, L. T.; Knapp, S.; Bayley, H. Stochastic Detection of Pim Protein Kinases Reveals Electrostatically Enhanced Association of a Peptide Substrate. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, E4417–26.
30. Rotem, D.; Jayasinghe, L.; Salichou, M.; Bayley, H. Protein Detection by Nanopores Equipped with Aptamers. *J. Am. Chem. Soc.* **2012**, *134*, 2781–2787.
31. Soskine, M.; Biesemans, A.; Moeyaert, B.; Cheley, S.; Bayley, H.; Maglia, G. An Engineered ClyA Nanopore Detects Folded Target Proteins by Selective External Association and Pore Entry. *Nano Lett.* **2012**, *12*, 4895–4900.
32. Nivala, J.; Marks, D. B.; Akeson, M. Unfoldase-Mediated Protein Translocation through an α-Hemolysin Nanopore. *Nat. Biotechnol.* **2013**, *31*, 247–250.
33. Rodriguez-Larrea, D.; Bayley, H. Multistep Protein Unfolding During Nanopore Translocation. *Nat. Nanotechnol.* **2013**, *8*, 288–295.
34. Rodriguez-Larrea, D.; Bayley, H. Protein Co-Translocational Unfolding Depends on the Direction of Pulling. *Nat. Commun.* **2014**, *5*, 4841.
35. Rosen, C. B.; Rodriguez-Larrea, D.; Bayley, H. Single-Molecule Site-Specific Detection of Protein Phosphorylation with a Nanopore. *Nat. Biotechnol.* **2014**, *32*, 179–181.
36. Baker, T. A.; Sauer, R. T. ClpXP, an ATP-Powered Unfolding and Protein-Degradation Machine. *Biochim. Biophys. Acta* **2012**, *1823*, 15–28.
37. Aubin-Tam, M.-E.; Olivares, A. O.; Sauer, R. T.; Baker, T. A.; Lang, M. J. Single-Molecule Protein Unfolding and Translocation by an ATP-Fueled Proteolytic Machine. *Cell* **2011**, *145*, 257–267.
38. Maillard, R. A.; Chistol, G.; Sen, M.; Righini, M.; Tan, J.; Kaiser, C. M.; Hodges, C.; Martin, A.; Bustamante, C. ClpX(P) Generates Mechanical Force to Unfold and Translocate Its Protein Substrates. *Cell* **2011**, *145*, 459–469.
39. Bross, P.; Corydon, T. J.; Andresen, B. S.; Jørgensen, M. M.; Bolund, L.; Gregersen, N. Protein Misfolding and Degradation in Genetic Diseases. *Hum. Mutat.* **1999**, *14*, 186–198.
40. Ferrer-Costa, C.; Orozco, M.; de la Cruz, X. Characterization of Disease-Associated Single Amino Acid Polymorphisms in Terms of Sequence and Structure Properties. *J. Mol. Biol.* **2002**, *315*, 771–786.
41. Wei, J.; Zaika, E.; Zaika, A. P53 Family: Role of Protein Isoforms in Human Cancer. *J. Nucleic Acids* **2012**, *2012*, 687359.
42. García-Sierra, F.; Mondragón-Rodríguez, S.; Basurto-Islas, G. Truncation of Tau Protein and Its Pathological Significance in Alzheimer's Disease. *J. Alzheimer's Dis.* **2008**, *14*, 401–409.
43. Kim, Y. I.; Burton, R. E.; Burton, B. M.; Sauer, R. T.; Baker, T. A. Dynamics of Substrate Denaturation and Translocation by the ClpXP Degradation Machine. *Mol. Cell* **2000**, *5*, 639–648.
44. Johnson, E. S.; Schwienhorst, I.; Dohmen, R. J.; Blobel, G. The Ubiquitin-Like Protein Smt3p Is Activated for Conjugation to Other Proteins by an Aos1p/Uba2p Heterodimer. *EMBO J.* **1997**, *16*, 5509–5519.
45. Improta, S.; Politou, A. S.; Pastore, A. Immunoglobulin-Like Modules from Titin I-Band: Extensible Components of Muscle Elasticity. *Structure* **1996**, *4*, 323–337.
46. Ormö, M.; Cubitt, A. B.; Kallio, K.; Gross, L. A.; Tsien, R. Y.; Remington, S. J. Crystal Structure of the Aequorea Victoria Green Fluorescent Protein. *Science* **1996**, *273*, 1392–1395.

47. Cordova, J. C.; Olivares, A. O.; Shin, Y.; Stinson, B. M.; Calmat, S.; Schmitz, K. R.; Aubin-Tam, M.-E.; Baker, T. A.; Lang, M. J.; Sauer, R. T. Stochastic but Highly Coordinated Protein Unfolding and Translocation by the ClpXP Proteolytic Machine. *Cell* **2014**, *158*, 647–658.
48. Martin, A.; Baker, T. A.; Sauer, R. T. Protein Unfolding by a AAA+ Protease Is Dependent on ATP-Hydrolysis Rates and Substrate Energy Landscapes. *Nat. Struct. Mol. Biol.* **2008**, *15*, 139–145.
49. Kenniston, J. A.; Baker, T. A.; Fernandez, J. M.; Sauer, R. T. Linkage Between ATP Consumption and Mechanical Unfolding During the Protein Processing Reactions of an AAA+ Degradation Machine. *Cell* **2003**, *114*, 511–520.
50. Gur, E.; Sauer, R. T. Recognition of Misfolded Proteins by Lon, a AAA(+) Protease. *Genes Dev.* **2008**, *22*, 2267–2277.
51. Sen, M.; Maillard, R. A.; Nyquist, K.; Rodriguez-Aliaga, P.; Pressé, S.; Martin, A.; Bustamante, C. The ClpXP Protease Unfolds Substrates Using a Constant Rate of Pulling but Different Gears. *Cell* **2013**, *155*, 636–646.
52. Pédelacq, J.-D.; Cabantous, S.; Tran, T.; Terwilliger, T. C.; Waldo, G. S. Engineering and Characterization of a Superfolder Green Fluorescent Protein. *Nat. Biotechnol.* **2006**, *24*, 79–88.
53. Storm, A. J.; Storm, C.; Chen, J.; Zandbergen, H.; Joanny, J.-F.; Dekker, C. Fast DNA Translocation through a Solid-State Nanopore. *Nano Lett.* **2005**, *5*, 1193–1197.
54. Zhang, H. The Optimality of Naive Bayes. In *FLAIRS Conference*; AAAI Press: Miami Beach, FL, 2004; pp 562−567.
55. Herman, D. S.; Lam, L.; Taylor, M. R. G.; Wang, L.; Teekakirikul, P.; Christodoulou, D.; Conner, L.; DePalma, S. R.; McDonough, B.; Sparks, E.; *et al.* Truncations of Titin Causing Dilated Cardiomyopathy. *N. Engl. J. Med.* **2012**, *366*, 619–628.
56. LeWinter, M. M.; Granzier, H. L. Titin Is a Major Human Disease Gene. *Circulation* **2013**, *127*, 938–944.
57. Ma, B.; Nussinov, R. The Stability of Monomeric Intermediates Controls Amyloid Formation: Abeta25−35 and Its N27Q Mutant. *Biophys. J.* **2006**, *90*, 3365–3374.
58. Croce, C. M. Oncogenes and Cancer. *N. Engl. J. Med.* **2008**, *358*, 502–511.
59. Gilmore, J. M.; Scheck, R. A.; Esser-Kahn, A. P.; Joshi, N. S.; Francis, M. B. N-Terminal Protein Modification through a Biomimetic Transamination Reaction. *Angew. Chem., Int. Ed. Engl.* **2006**, *45*, 5307–5311.
60. Scheck, R. A.; Francis, M. B. Regioselective Labeling of Antibodies through N-Terminal Transamination. *ACS Chem. Biol.* **2007**, *2*, 247–251.
61. Lubec, G.; Afjehi-Sadat, L. Limitations and Pitfalls in Protein Identification by Mass Spectrometry. *Chem. Rev.* **2007**, *107*, 3568–3584.
62. Han, J.-H.; Batey, S.; Nickson, A. A.; Teichmann, S. A.; Clarke, J. The Folding and Evolution of Multidomain Proteins. *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 319–330.
63. Prabakaran, S.; Lippens, G.; Steen, H.; Gunawardena, J. Post-Translational Modification: Nature's Escape from Genetic Imprisonment and the Basis for Dynamic Information Encoding. *Wiley Interdiscip. Rev.: Syst. Biol. Med.* **2012**, *4*, 565–583.
64. Martin, A.; Baker, T. A.; Sauer, R. T. Rebuilt AAA + Motors Reveal Operating Principles for ATP-Fuelled Machines. *Nature* **2005**, *437*, 1115–1120.
65. Cherf, G. M.; Lieberman, K. R.; Rashid, H.; Lam, C. E.; Karplus, K.; Akeson, M. Automated Forward and Reverse Ratcheting of DNA in a Nanopore at 5-Å Precision. *Nat. Biotechnol.* **2012**, *30*, 344–348.
66. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; *et al.* Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2012**, *12*, 2825–2830.