nature biotechnology

Unfoldase-mediated protein translocation through an α-hemolysin nanopore

Jeff Nivala, Douglas B Marks & Mark Akeson

Using nanopores to sequence biopolymers was proposed more than a decade ago[1]. Recent advances in enzyme-based control of DNA translocation[2] and in DNA nucleotide resolution using modified biological pores[3] have satisfied two technical requirements of a functional nanopore DNA sequencing device. Nanopore sequencing of proteins was also envisioned[1]. Although proteins have been shown to move through nanopores[4–6], a technique to unfold proteins for processive translocation has yet to be demonstrated. Here we describe controlled unfolding and translocation of proteins through the α-hemolysin (α-HL) pore using the AAA+ unfoldase ClpX. Sequence-dependent features of individual engineered proteins were detected during translocation. These results demonstrate that molecular motors can reproducibly drive proteins through a model nanopore—a feature required for protein sequence analysis using this single-molecule technology.

Protein sequencing using nanopores is technically challenging for two reasons: (i) both tertiary and secondary structures must be unfolded to allow the denatured protein to thread through the nanopore sensor with amino acid residues in single-file order; and (ii) processive unidirectional translocation of the denatured polypeptide through the nanopore electric field must be achieved despite a nonuniform charge along the polypeptide chain. To address these challenges, we have devised a general method for enzyme-controlled unfolding and translocation of native proteins through a nanopore sensor using the protein unfoldase ClpX.

ClpX is a component of the ClpXP proteasome-like complex that is responsible for the targeted degradation of numerous protein substrates in *Escherichia coli* and other organisms[7]. Within this complex, ClpX forms a homohexameric ring that uses ATP hydrolysis to unfold and translocate proteins through its central pore and into a proteolytic chamber (ClpP) for degradation. We chose *E. coli* ClpX for this nanopore application because it generates sufficient mechanical force (~20 pN) to denature stable protein folds, and because it translocates along proteins at a rate suitable for primary sequence analysis by nanopore sensors (up to 80 amino acids per second)[8,9].

For each experiment a single α-HL nanopore was inserted into a 30-μm-diameter lipid bilayer between two compartments (termed *cis* and *trans*), each filled with 100 μl of a 200 mM KCl buffer optimized for ClpX function and supplemented with 5 mM ATP (**Fig. 1a**). A patch-clamp amplifier applied a constant 180 mV potential between two Ag/AgCl electrodes (*trans* side +) and recorded ionic current through the nanopore as a function of time. This electrophysiology setup is fundamentally the same as one used in earlier experiments to identify a protein-conducting channel in the endoplasmic reticulum[10]. Substrate proteins were added to the *cis* solution at ~1 μM final concentration, and ~100 nM of a linked ClpX variant (Online Methods) was present in the *trans* solution.

For our initial experiments, we used a modified version of the ubiquitin-like protein Smt3 (ref. 11) as the substrate. Smt3 comprises 98 amino acids arranged into four β-strands and a single α-helix. To facilitate nanopore analysis, we modified the engineered Smt3 protein, designated 'S1', in two ways. First, it was appended with a 65-amino-acid-long glycine/serine tail including 13 interspersed negatively charged aspartate residues (**Supplementary Fig. 1**). This unstructured polyanion was designed to promote capture and retention of S1 in the electric field across the nanopore. Based on its crystal structure[12], the Smt3 folded domain is predicted to sit on top of the α-HL vestibule. Second, the appended polyanion was capped at its C terminus with the ssrA tag, an 11-amino-acid ClpX-targeting motif[13]. This ssrA peptide tag allowed ClpX to specifically bind to the C terminus of the protein when it threaded through the pore into the *trans* compartment (**Fig. 1b** and **c**,**i**). Experiments conducted in bulk phase confirmed that ClpX unfolds and translocates proteins appended with this unique polyanion tag in an ATP-dependent manner (**Supplementary Fig. 2**).

Representative ionic current traces for capture and translocation of protein S1 in the presence of ClpX and ATP are shown in **Figure 2a** and **Supplementary Figure 3a**. From the open channel current of ~34 ± 2 pA (**Fig. 2a**,**i**), S1 capture resulted in a current drop to ~14 pA (**Fig. 2a**,**ii**). This stable current lasted for tens of seconds and was observed in the presence or absence of ClpX and ATP added to the *trans* compartment. This is consistent with the Smt3 structure held stationary atop the pore vestibule by the electrical force acting on the charged polypeptide tail in the pore electric field. In the presence of ClpX and ATP, this initial ionic current state was often followed by a progressive downward current ramp reaching an average of ~11 pA with a median duration of 4.2 s (**Fig. 2a**,**iii** and **Supplementary Fig. 4**). This current ramp was observed with protein S1 a total of 45 times over ~5.5 h of experimentation when ClpX and ATP were present; in contrast, the ramp was never observed after ionic current state ii (**Fig. 2a**,**ii**) when ClpX or ATP was absent from the *trans* solution over ~2.3 and ~1.7 h of experimentation, respectively. In a majority of

Nanopore Group, Department of Biomolecular Engineering, University of California, Santa Cruz, Santa Cruz, California, USA. Correspondence should be addressed to M.A. (makeson@soe.ucsc.edu).

Received 3 October 2012; accepted 11 January 2013; published online 3 February 2013; doi:10.1038/nbt.2503

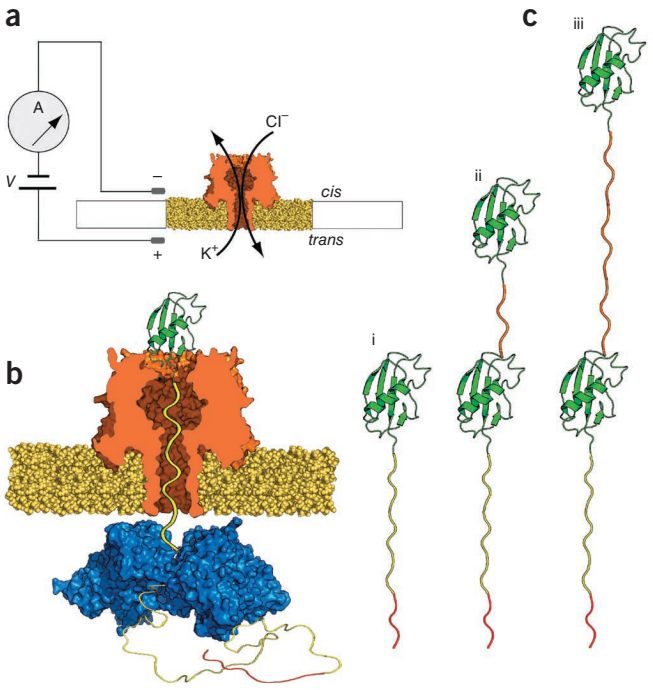


Figure 1 Experimental set-up. (**a**) Nanopore sensor. A single α-HL pore is embedded in a lipid bilayer separating two polytetrafluoroethylene wells each containing 100 μl of 0.2 M KCl solution at 30 °C. Voltage is applied between the wells (*trans* side +180 mV), causing ionic current flow through the channel. Current diminishes in the presence of a captured protein molecule. (**b**) Protein capture in the nanopore. A model protein bearing an Smt3 domain (green) at its N terminus is coupled to a charged flexible linker (yellow) with an ssrA tag (red) at its C terminus. As a result of the applied voltage, the charged, flexible tag is threaded through the pore into the *trans*-side solution until the folded Smt3 domain prevents complete translocation of the captured protein. ClpX present in the *trans* solution binds the C-terminal ssrA sequence. Fueled by ATP hydrolysis, ClpX translocates along the protein tail toward the channel, and subsequently catalyzes unfolding and translocation of the Smt3 domain through the pore. (**c**) Engineered proteins used in this study. S1, a protein bearing a single N-terminal Smt3-domain coupled to a 65-amino-acid-long charged flexible segment capped at its carboxy terminus with the 11 amino acid ClpX-targeting domain (ssrA tag) (**i**); S2-35, similar to S1 but appended at its N terminus by a 35-amino-acid linker and a second Smt3 domain (**ii**); S2-148, identical to S2-35 except for an extended 148-amino-acid linker between the Smt3 domains (**iii**). The linker lengths in this panel are not to scale.

events, the ClpX-dependent ramping state terminated with an abrupt ionic current decrease to ~3.9 pA (**Fig. 2a**,**iv**). The median duration for state iv was ~700 ms before it ended in a rapid increase to open the channel current (**Fig. 2a**,**i′**). As an additional control, we constructed a variant of S1 appended with three additional amino acid residues at the C terminus (protein S1-RQA, **Supplementary Fig. 1**). Because ClpX recognition of the ssrA tag is dependent upon the C terminus α-carboxyl group, the additional residues placed between the tag sequence and the C terminus thereby inhibit ClpX binding[14]. In the presence of ATP and ClpX, we never observed the ramping state with protein S1-RQA over ~1.7 h of experimentation (data not shown).

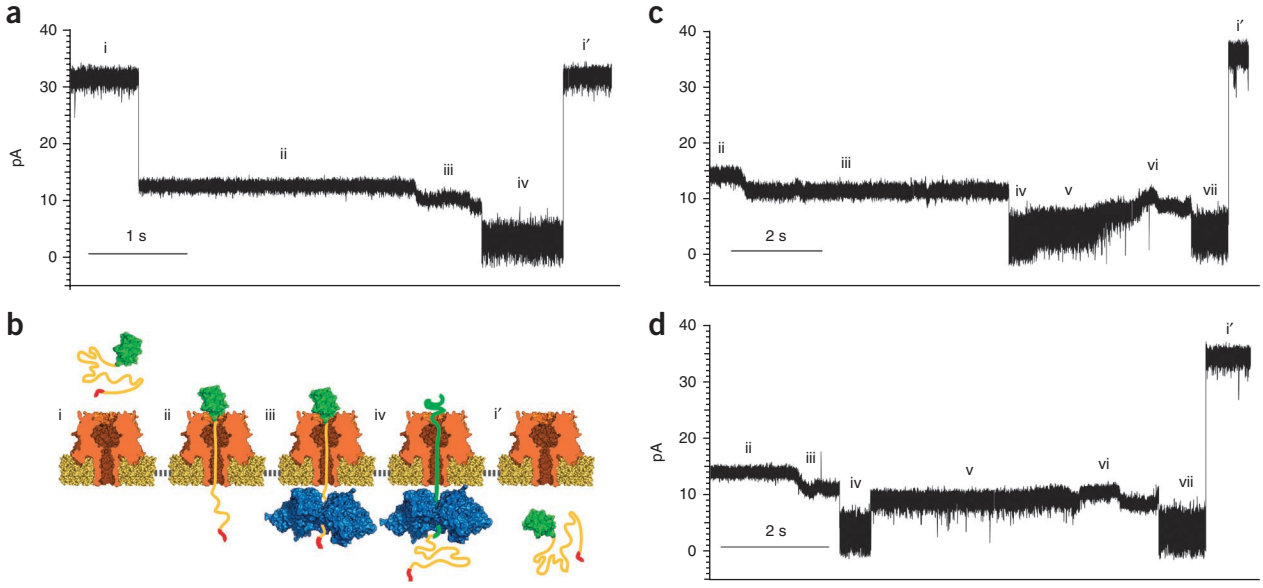


Figure 2 Ionic current traces during ClpX-mediated protein translocation. (**a**) S1 translocation. Open channel current through the α-HL nanopore under standard conditions (~34 ± 2 pA, RMS noise 1.2 ± 0.1 pA) (**i**). Capture of the S1 substrate. Upon protein capture, the ionic current drops to ~14 pA (~0.7 pA RMS noise) (**ii**). ClpX-mediated ramping state. The ionic current decreases to ~10 pA and is characterized by one or more gradual amplitude transitions. This pattern is only observed in the presence of ClpX and ATP (*trans* compartment) (**iii**). Smt3 domain unfolding and translocation through the nanopore (~3.8 pA, 1.7 pA RMS noise) (**iv**). Return to open channel current upon completion of substrate translocation to the *trans* compartment (**i′**). (**b**) Working model of ClpX-mediated translocation of S1. Roman numerals used to label panels correspond to ionic current states in **a**. (**c**) S2-35 translocation. Open channel current (**i**) is not shown. States ii–iv are identical to states ii-iv in **a**. Gradual increase in ionic current to ~10 pA. In our working model this corresponds to a transition from Smt3 domain translocation to linker region translocation (**v**). A second putative ramping state that closes resembles ramping state iii (**vi**). A second putative Smt3 translocation state with ionic current properties that closely resemble state iv (**vii**). Return to open channel current (**i′**). (**d**) S2-148 translocation. Ionic current states i-iv and vi-i′ were nearly identical to those states for S2-35 translocation in **c**. (**v**) In our working model, this ionic current state corresponds to translocation of the 148-amino-acid linker. Its amplitude is ~3 pA higher than the S2-35 linker amplitude (~9 pA), and it has a median duration ~2.5 fold longer than the comparable S2-35 state v. Translocation events that included ramping state iii were observed 62 times for protein S2-35 (7.3 h of experimentation), and 66 times for protein S2-148 (4.3 h of experimentation), when ClpX and ATP were present. In the absence of ClpX, these ramping states were never observed for S2-35 (1.7 h of experimentation) and S2-148 (1.2 h of experimentation).

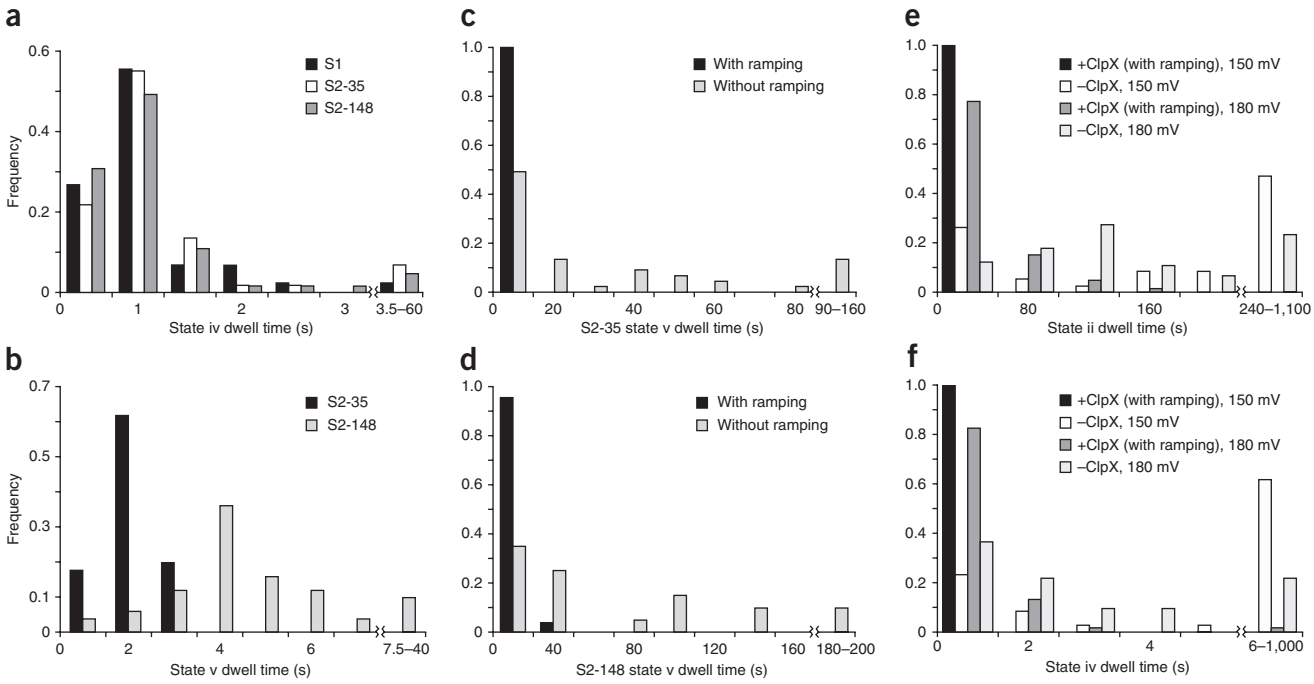


Figure 3 Ionic current state dwell times during translocation of model proteins through the nanopore. (**a**) Comparison of putative Smt3 translocation (state iv) dwell times for three model proteins. Values are from events that included the ClpX-dependent ramping state (**Fig. 2**,**iii**). Black bars: median = 0.71 s, interquartile range (IQR) = 0.41, $n = 45$. Gray bars: median = 0.64 s, IQR = 0.47, $n = 60$. White bars: median = 0.63 s, IQR = 0.40, $n = 65$. (**b**) Comparison of putative linker region (state v) dwell times for S2-35 and S2-148 proteins. Values are from events that included the ClpX-dependent ramping state (**Fig. 2**,**iii**). Black bars: median = 1.52 s, IQR = 0.68, $n = 50$. Gray bars: median = 3.62 s, IQR = 2.03, $n = 50$. (**c**) State v translocation dwell times for S2-35 events. Black bars: median = 1.52 s, IQR = 0.68, $n = 50$. Gray bars: median = 11.45 s, IQR = 39.53, $n = 45$. (**d**) State v translocation dwell times for S2-148 translocation events. Black bars: median = 3.62 s, IQR = 2.03, $n = 50$. Gray bars: median = 37.07 s, IQR = 89.80, $n = 20$. (**e**) State ii dwell times for protein substrates at two voltages. Black bars: median = 4.89 s, IQR = 6.72, $n = 104$. White bars: median = 165.50 s, IQR = 351.75, $n = 34$. Gray bars: median = 17.26 s, IQR = 31.08, $n = 173$. Light gray bars: median = 90.0 s, IQR = 112.0, $n = 72$. (**f**) State iv dwell times for the S1 protein substrate at two voltages. Black bars: median = 0.33 s, IQR = 0.40, $n = 104$. White bars: median = 8.25 s, IQR = 26.82, $n = 34$. Gray bars: median = 0.65 s, IQR = 0.45, $n = 52$. Light gray bars: median = 1.74 s, IQR = 3.12, $n = 41$.

Based on these data we hypothesized that ClpX served as a molecular machine that used chemical energy derived from ATP hydrolysis to pull the S1 protein through the nanopore. In the proposed process, an open channel (**Fig. 2b**,**i**) captures protein S1 with the Smt3 segment perched above the pore vestibule with the slender, charged polypeptide tail segment extended into the pore lumen, and the ssrA tag in the *trans* compartment (**Fig. 2b**,**ii**). In this ionic current state, ClpX is not bound to S1 or, alternatively, is bound but is still distant from the pore; ClpX advances along the S1 strand toward the *trans*-side orifice of the α-HL pore until it makes contact (**Fig. 2b**,**iii**). At this time, ionic current decreases owing to proximity of ClpX to the pore; under the combined force exerted by ClpX and the pore electric field, the Smt3 structure atop the pore is sequentially denatured, thus allowing the polypeptide to advance through the nanopore (**Fig. 2b**,**iv**). In this state, the ionic current has decreased because larger amino acids (or Smt3 secondary structures) have entered the pore lumen. This ionic current state persists until the S1 protein is completely pulled into the *trans* compartment resulting in a return to the open channel current (**Fig. 2b**,**i′**).

This model makes a testable prediction. If the observed states are due to processive movement of polypeptide segments into the pore lumen driven in part by ClpX, then changing the protein primary structure should result in sequential ClpX- and ATP-dependent changes in the ionic current pattern. In particular, addition of a second Smt3 domain should result in a second ramping state (**Fig. 2a**,**iii**) followed by a second Smt3 translocation state centered at ~4 pA (**Fig. 2a**,**iv**). As a test, we fused a flexible glycine/serine-rich 35 amino acid linker to the N terminus of the S1 protein and capped this with a second Smt3 domain (protein S2-35, **Fig. 1c**,**ii** and **Supplementary Fig. 1**). Thus, the single folded-component sequence of S1 (that is, Smt3) is repeated twice in S2-35.

When protein S2-35 was captured in the nanopore with ClpX and ATP present in the *trans* compartment, an ionic current pattern with eight reproducible states was observed (**Fig. 2c** and **Supplementary Fig. 3b**). The first four states (**Fig. 2c**,**i–iv**) were identical to states i–iv caused by S1 translocation (compare **Fig. 2a** and **c**). This similarity included ramping state iii that is diagnostic for ClpX engagement, and the Smt3-dependent state iv. However, beginning at state v, the S2-35 pattern diverged from the S1 pattern (compare **Fig. 2a** and **c**). That is, following Smt3 translocation state iv, a typical S2-35 ionic current trace did not proceed to the open channel current but instead transitioned to a ~6.3-pA state with a median duration of 1.5 s (**Fig. 2c**,**v**). This was followed by a ~8.5-pA state (**Fig. 2c**,**vi**) that closely resembled ramping state iii, and a subsequent ionic current state that closely resembled the putative Smt3 translocation state iv (**Fig. 2c**,**vii**). In other words, consistent with our model, the putative ClpX-bound and Smt3-dependent states that were observed once during S1 events (**Fig. 2a**) were observed twice during S2-35 events (**Fig. 2c**). These analogous states for the two constructs shared nearly identical amplitudes, root mean square (RMS) noise values and durations.

This dependence of ionic current on protein structure is consistent with ClpX-driven protein translocation through the nanopore. As an additional test, we re-examined ionic current state v observed during

S2-35 translocation. This state is consistent with movement of the 35-amino-acid linker through the nanopore based on two observations: (i) its average ionic current is measurably higher than that of surrounding states (**Fig. 2c**) as expected for an amino acid sequence with few bulky side chains; and (ii) in the time domain, state v occurs between Smt3-dependent states iv and vi as expected, given its position along the S2-35 primary sequence (**Fig. 1c**,**ii** and **Supplementary Fig. 1**).

If state v corresponds to translocation of the polypeptide linker under ClpX control, then changes in the length and composition of this linker should result in duration and current amplitude changes. To test this, we designed a third protein in which the S2-35 linker region was appended with an additional 113 amino acids, yielding a final construct consisting of two Smt3 domains separated by an extended 148-amino-acid flexible linker (protein S2-148, **Fig. 1c**,**iii** and **Supplementary Fig. 1**). As predicted, when this protein was captured in the nanopore under standard conditions in the presence of ClpX and ATP, eight reproducible states similar to S2-35 events were observed (**Figs. 2d**,**3a** and **Supplementary Figs. 3c**,**4–6**). Importantly, however, the S2-35 and S2-148 events differed substantially at state v (compare **Fig. 2c** and **d**). That is, the S2-148 state v had a higher mean residual current than did S2-35 (~9 versus ~6 pA, respectively), and a median duration ~2.5 fold longer than that of S2-35 state v (**Fig. 3b**). The increased duration of S2-148 state v relative to S2-35 state v was anticipated and consistent with the model described in **Figure 2**. The increased current level was likely due to differences in linker amino acid composition between the two proteins (S2-35 linker: 51% Gly, 34% Ser, 15% other; S2-148 linker: 34% Gly, 32% Ser, 19% Ala, 15% other). However, confirmation of this hypothesis will require systematic testing of the relationship between amino acid identity and resistance to ionic current through the pore lumen.

Prior studies have shown that proteins can translocate through nanopores under an applied voltage without the assistance of processive enzymes[4–6,15–17]. This was also the case in our experiments, that is, translocation of the three model proteins was observed in the absence of ClpX- or ATP-dependent mechanical work performed on captured strands (**Supplementary Fig. 7**). However, these ClpX- and ATP-independent translocation events lacked the diagnostic ramping states (**Fig. 2**), and they were measurably longer and more variable in duration than were ClpX-mediated translocation events (**Fig. 3c–f**). This is consistent with an unregulated translocation process that depends upon random structural fluctuations of the captured protein molecule and intermittent electrical force acting on amino acid segments with variable charge density in the pore electric field. This model predicts that ClpX-dependent translocation will be relatively unaffected by changes in applied voltage. This proved to be true. State ii and iv dwell times for ClpX- and ATP-dependent S1 translocation events acquired at 150 mV were comparable to those acquired at 180 mV (**Fig. 3e**,**f**). At both voltages, these events were consistently faster and more narrowly distributed than ClpX- or ATP-independent events. Thus, ClpX activity (not voltage) is dominating the unfolding and translocation process.

In summary, we have demonstrated enzymatic control of protein unfolding and translocation through the α-HL nanopore. Segments of each protein could be discerned based on sequence-dependent features as the protein passed through the ~50-Angstrom-long transmembrane pore lumen. This proof-of-concept study is analogous to early work on nanopore DNA sequencing in which T7 DNA polymerase moved model DNA templates through the α-HL pore over short distances[18]. As was true for DNA translocation control in that early work, protein translocation control for bench-top sequencing applications will require technical improvements. This is worthwhile because an optimized nanopore device could provide long reads of individual native protein strands.

METHODS

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

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

The authors thank Oxford Nanopore Technologies (Oxford, UK) for supplying α-HL heptamers, and A. Martin (UC Berkeley) for supplying ClpX-related expression plasmids and for helpful discussions on their use. R. Abu-Shumays, D. Bernick, K. Lieberman and H. Olsen commented on drafts of the manuscript. This work was supported by a UC startup grant to M.A., and by equipment purchased previously using National Human Genome Research Institute grant R01HG006321. The ClpX-ΔN_3 BLR expression strain was obtained from A. Martin (UC Berkeley), as was a his-tagged ClpP expression strain.

AUTHOR CONTRIBUTIONS

J.N. conceived and designed the project, performed the protein engineering and production, conceived and performed experiments and co-wrote the manuscript. D.B.M. performed and conceived experiments, analyzed data and co-wrote the manuscript. M.A. co-wrote the manuscript, designed the nanopore platform and directed the project.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

Published online at http://www.nature.com/doifinder/10.1038/nbt.2503.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Church, G.M., Deamer, D.W., Branton, D., Baldarelli, R. & Kasianowicz, J. Characterization of individual polymer molecules based on monomer-interface interaction. US patent 5,795,782 (1998).
2. Cherf, G.M. *et al.* Automated forward and reverse ratcheting of DNA in a nanopore at 5-Å precision. *Nat. Biotechnol.* **30**, 344–348 (2012).
3. Manrao, E.A. *et al.* Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nat. Biotechnol.* **30**, 349–353 (2012).
4. 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).
5. Talaga, D.S. & Li, J. Single-molecule protein unfolding in solid state nanopores. *J. Am. Chem. Soc.* **131**, 9287–9297 (2009).
6. Merstorf, C. *et al.* Wild type, mutant protein unfolding and phase transition detected by single-nanopore recording. *ACS Chem. Biol.* **7**, 652–658 (2012).
7. Baker, T.A. & Sauer, R.T. ClpXP, an ATP-powered unfolding and protein-degradation machine. *Biochim. Biophys. Acta* **1823**, 15–28 (2012).
8. Aubin-Tam, M.E. *et al.* Single-molecule protein unfolding and translocation by an ATP-fueled proteolytic machine. *Cell* **145**, 257–267 (2011).
9. Maillard, R.A. *et al.* ClpX(P) generates mechanical force to unfold and translocate its protein substrates. *Cell* **145**, 459–469 (2011).
10. Simon, S.M. & Blobel, G. A protein-conducting channel in the endoplasmic reticulum. *Cell* **65**, 371–380 (1991).
11. Johnson, E.S., Schweinhorst, 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.* **16**, 5509–5519 (1997).
12. Sheng, W. & Liao, X. Solution structure of a yeast ubiquitin-like protein Smt3: the role of structurally less defined sequences in protein-protein recognitions. *Protein Sci.* **11**, 1482–1491 (2002).
13. Gottesman, S., Roche, E., Zhou, Y. & Sauer, R.T. The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes Dev.* **12**, 1338–1347 (1998).
14. Kim, Y. *et al.* Dynamics of substrate denaturation and translocation by the ClpXP degradation machine. *Mol. Cell* **5**, 639–648 (2000).
15. Christensen, C. *et al.* Effect of charge, topology and orientation of the electric field on the interaction of peptides with the α-hemolysin pore. *J. Pept. Sci.* **17**, 726–734 (2011).
16. Movileanu, L. Interrogating single proteins through nanopores: challenges and opportunities. *Trends Biotechnol.* **27**, 333–341 (2009).
17. Oukhaled, G. *et al.* Unfolding of proteins and long transient conformations detected by single nanopore recording. *Phys. Rev. Lett.* **98**, 158101 (2007).
18. Olasagasti, F. *et al.* Replication of individual DNA molecules under electronic control using a protein nanopore. *Nat. Nanotechnol.* **5**, 798–806 (2010).

ONLINE METHODS

Materials. A covalently linked trimer of an N-terminal truncated ClpX variant (ClpX-ΔN_3) was used for all ClpX nanopore experiments[19]. This variant maintains near-WT ClpX activity levels on ssrA-tagged substrates while staying associated at lower working concentrations. The ClpX-ΔN_3 BLR expression strain was obtained from A. Martin (UC Berkeley), as was a his-tagged ClpP expression strain. ClpX protein expression was induced at an A_{600} of ~1 by addition of 0.5 mM IPTG, and incubated at 23 °C with shaking for 3–4 h. Cultures were pelleted, resuspended in lysis buffer (50 mM NaH2PO4 pH 8, 300 mM NaCl, 100 mM KCl, 20 mM imidazole, 10% glycerol, 10 mM BME) 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). His-tagged ClpP protein expression was induced at an A_{600} of ~0.6 by addition of 0.5 mM IPTG, and incubated at 30 °C with shaking for 3–4 h; purification was done on a Ni^{2+}-NTA affinity column. To construct substrate protein S1, DNA encoding the 76 amino acid tail (GGSSGGSGGSGSSGDGGSSGGSGGSGSSGDGGSSGGS GGDGSSGDGGSDGDSDGSDGDGDSDGDDAANDENYALAA) was constructed by PCR and cloned into pET-SUMO vector (Invitrogen) at the T/A-cloning site, fusing the tail sequence onto the Smt3 sequence 3′ end. S2-35 was constructed by PCR-based addition of DNA encoding the 35 amino acid linker (GGSGSGGSGSGGSGSQNEYRSGGSGSGGSGSGGSG) to the 5′ end of the S1 Smt-3 sequence. This linker-modified S1 gene was then cloned into pE-SUMO vector (LifeSensors) at the BsaI site, fusing the added linker and S1 sequence to the 3′ end of the pE-SUMO Smt3 sequence. DNA for the S2-148 linker addition (GGSGSAGSGASGSSGSEGSGASGSAGSGSAGSRGSGASGSAGSGSAGSGG AEAAKEAAKEAAKEAAKEAAKAGGSGSAGSAGSASSGSDGSGASGSAGS GSAGSKGSGASGSAGSGSSGS) was constructed by PCR, and cloned into the S2-35 vector within the 35 amino acid linker region by the Gibson assembly method. DNA for the GFP and titin-I27 domains of fluorescent protein S2-GT were extracted by PCR from a GFP-titin-I27V15P-ssrA[20] expression vector obtained from A. Martin, and cloned into the S2-35 vector by Gibson assembly. These engineered proteins were expressed in BL21 (DE3)*. Expression was induced at ~0.6 A_{600} by addition of 0.5 mM IPTG, and incubated at 37 °C with shaking for 4–6 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). The protein substrates and enzymes were flash frozen in small aliquots following purification and stored at −80 °C.

Nanopore experiments. All experiments were performed in PD buffer containing 200 mM KCl, 5 mM $MgCl_2$, 10% glycerol and 25 mM HEPES-KOH pH 7.6. Setup of the nanopore device and insertion of an α-HL nanopore into a lipid bilayer have been described[2]. Briefly, a single α-HL nanopore was inserted into a lipid bilayer that separates two wells that each contained 100 µl of PD buffer. A constant 180 mV potential was applied across the bilayer and ionic current was measured through the nanopore between Ag/AgCl electrodes in series with an integrating patch clamp amplifier (Axopatch 200B, Molecular Devices) in voltage clamp mode. Data were recorded using an analog-to-digital converter (Digidata 1440A, Molecular Devices) at 100 kHz bandwidth in whole-cell configuration then filtered at 2 kHz using an analog low-pass Bessel filter. Experimental conditions were prepared by the daily preparation of PD/ATP 5 mM and PD/ATP 4 mM. ClpX was diluted 1:10 in PD/ATP 5 mM for a final concentration of 30–100 nM $ClpX_6$ in 4.5 mM ATP final. ClpX solution was used to fill the entire system before isolation of a single α-HL nanopore. Upon insertion, the *cis* well was perfused with ~6 mL PD/ATP 4 mM. Experiments were conducted at 30 °C with 1–2 µM substrate added to the *cis* well. In the presence of ClpX, protein substrate capture events were ejected with reverse polarity due to pore clogs or after a predetermined duration. Those capture events that remained in state ii for a duration of ~90 s or longer were considered to be inaccessible to ClpX and were ejected. Voltage-induced translocations were frequently ejected to prevent clogging and to increase the efficiency of data collection. A single nanopore experiment is defined as the time during which ionic current data were acquired from one α-HL nanopore in an intact bilayer before termination by bilayer rupture, loss of channel conductance or completion of a preset number of translocation events. Experiments were conducted over multiple days for each substrate and condition tested. Prior to each experiment, protein samples that were to be used on that day were freshly thawed before the start of the experiment.

ClpX bulk phase assays. The protein S1 degradation time course was performed in PD buffer with 6 µM S1, 150 nM $(ClpX\text{-}\Delta N_3)_2$, 300 nM $ClpP_{14}$ and an ATP-regeneration mix (4 mM ATP, 16 mM creatine phosphate, and 0.16 mg/mL creatine phosphokinase) at 37 °C. Samples were visualized by SDS-PAGE gel (Any kD Mini-PROTEAN TGX, Bio-Rad) and stained with Coomassie blue. Fluorescence assays were performed in PD buffer with 1 µM S2-GT, 150 nM $(ClpX\text{-}\Delta N_3)_2$, 300 nM $ClpP_{14}$ and an ATP-regeneration mix. Reactions were incubated at 37 °C in a Stratagene Mx3005P measuring emission at 516 nm (excitation at 492 nm).

19. Martin, A., Baker, T.A. & Sauer, R.T. Rebuilt AAA+ motors reveal operating principles for ATP-fuelled machines. *Nature* **437**, 1115–1120 (2005).
20. 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* **114**, 511–520 (2003).

Unfoldase-mediated protein translocation through an α-hemolysin nanopore

Jeff Nivala, Douglas B. Marks, and Mark Akeson

Supplementary Figures 1-7

S1:

MGSSHHHHHHGSGLVPRGSASMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKT
TPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGGGGSSGGS
GGSGSSGDGGSSGGSGGSGSSGDGGSSGGSGGDGSSGDGGSDGDSDGSDGDGDSDGDD
AANDENYALAA

S2-35:

MGHHHHHHGSLQDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFA
KRQGKEMDSLRFLYDGIRIQADQAPEDLDMEDNDIIEAHREQIGGGGSGSGGSGSGGSGSQN
EYRSGGSGSGGSGSGGSGMGSSHHHHHHGSGLVPRGSASMSDSEVNQEAKPEVKPEVKPE
THINLKVSDGSSEIFFKIKKTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMED
NDIIEAHREQIGGGGSSGGSGGSGSSGDGGSSGGSGGSGSSGDGGSSGGSGGDGSSGDGG
SDGDSDGSDGDGDSDGDDAANDENYALAA

S2-148:

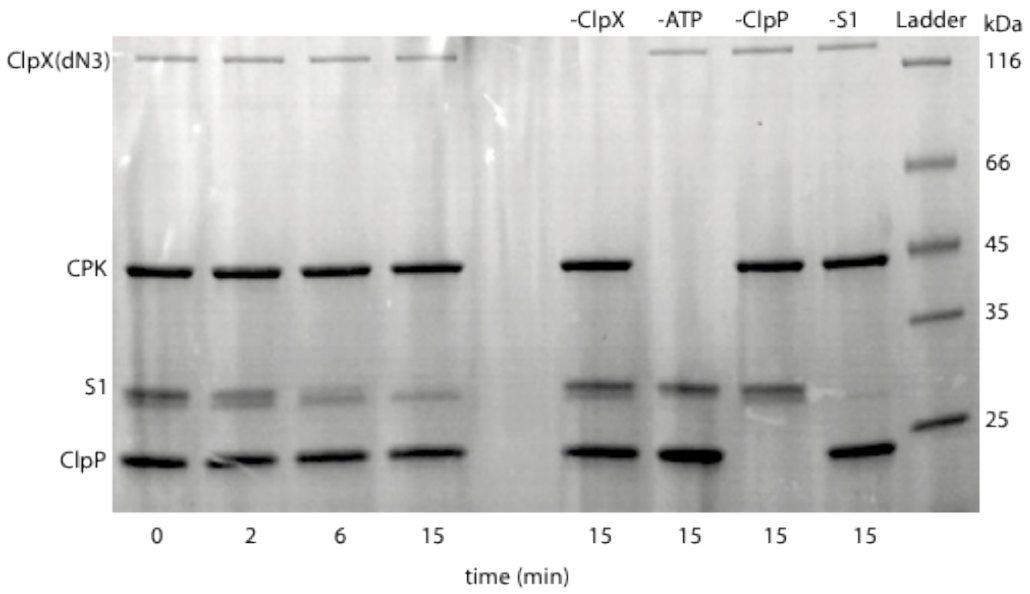
MGHHHHHHGSLQDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFA
KRQGKEMDSLRFLYDGIRIQADQAPEDLDMEDNDIIEAHREQIGGGGSGSGGSGSGGSGSQN
EYRSGGGGSGSAGSGASGSSGSEGSGASGSAGSGSAGSRGSGASGSAGSGSAGSGGAEA
AKEAAKEAAKEAAKEAAKAGGSGSAGSAGSASSGSDGSGASGSAGSGSAGSKGSGASGSA
GSGSSGSSGGSGMGSSHHHHHHGSGLVPRGSASMSDSEVNQEAKPEVKPEVKPETHINLKV
SDGSSEIFFKIKKTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHR
EQIGGGGSSGGSGGSGSSGDGGSSGGSGGSGSSGDGGSSGGSGGDGSSGDGGSDGDSD
GSDGDGDSDGDDAANDENYALAA

S1-RQA:

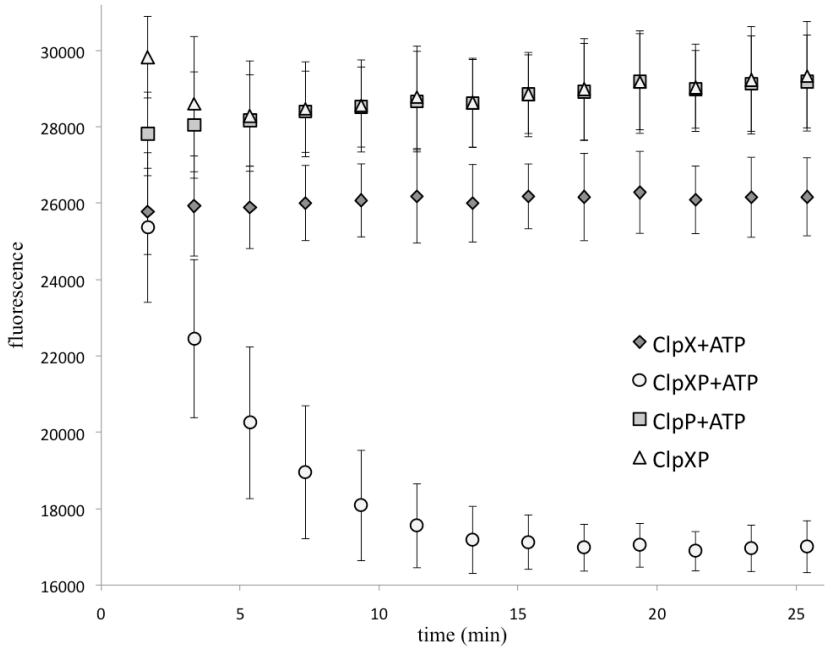
MGSSHHHHHHGSGLVPRGSASMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKT
TPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGGGGSSGGS
GGSGSSGDGGSSGGSGGSGSSGDGGSSGGSGGDGSSGDGGSDGDSDGSDGDGDSDGDD
AANDENYALAARQA

Supplementary Figure 1 Amino acid sequences of substrate proteins used in this study. Green, Smt3 domains; yellow, charged tail; red, ssrA tag; orange, linker region; black, affinity purification tag regions; blue, additional residues added to prevent ClpX binding to the ssrA tag.

a

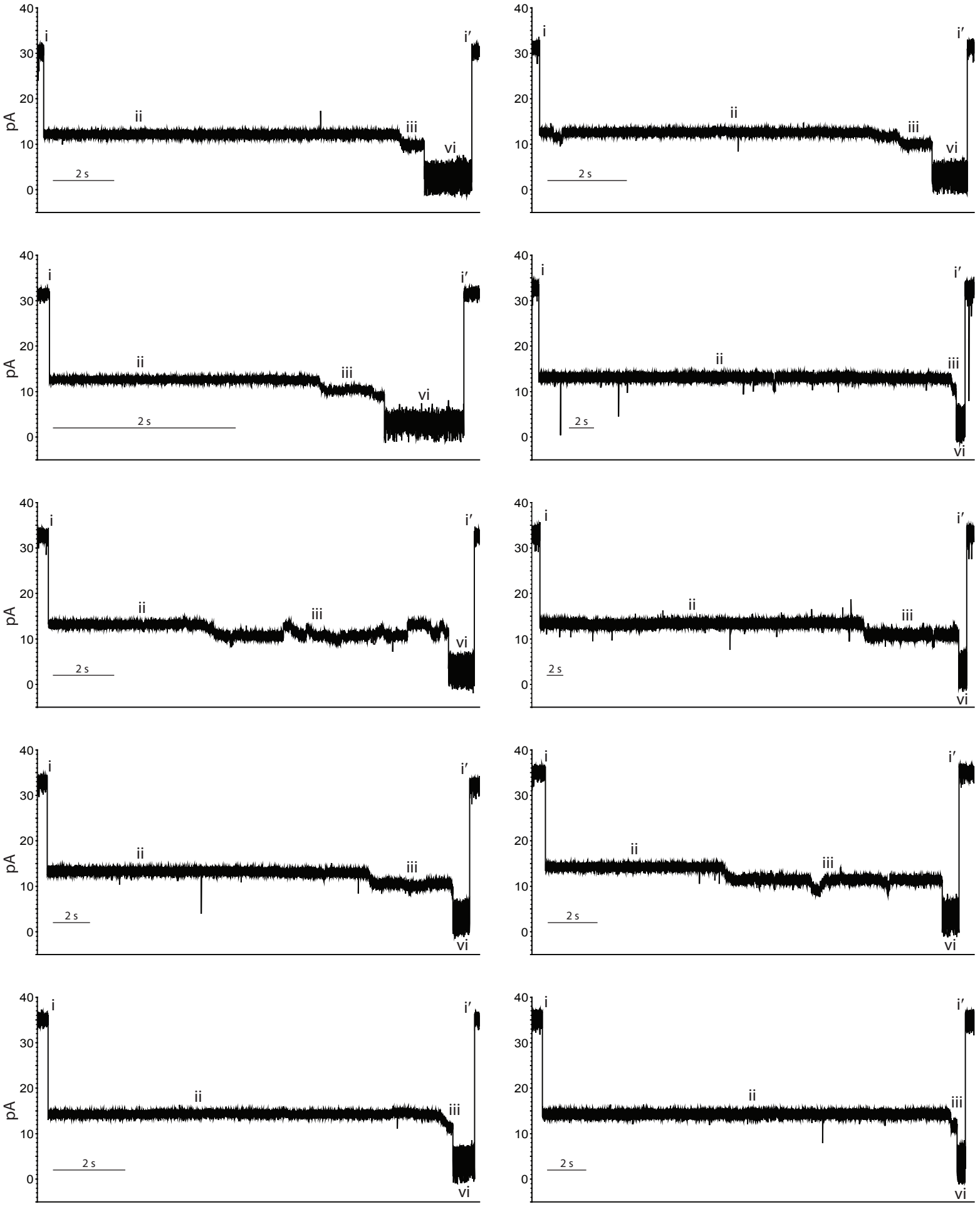


b

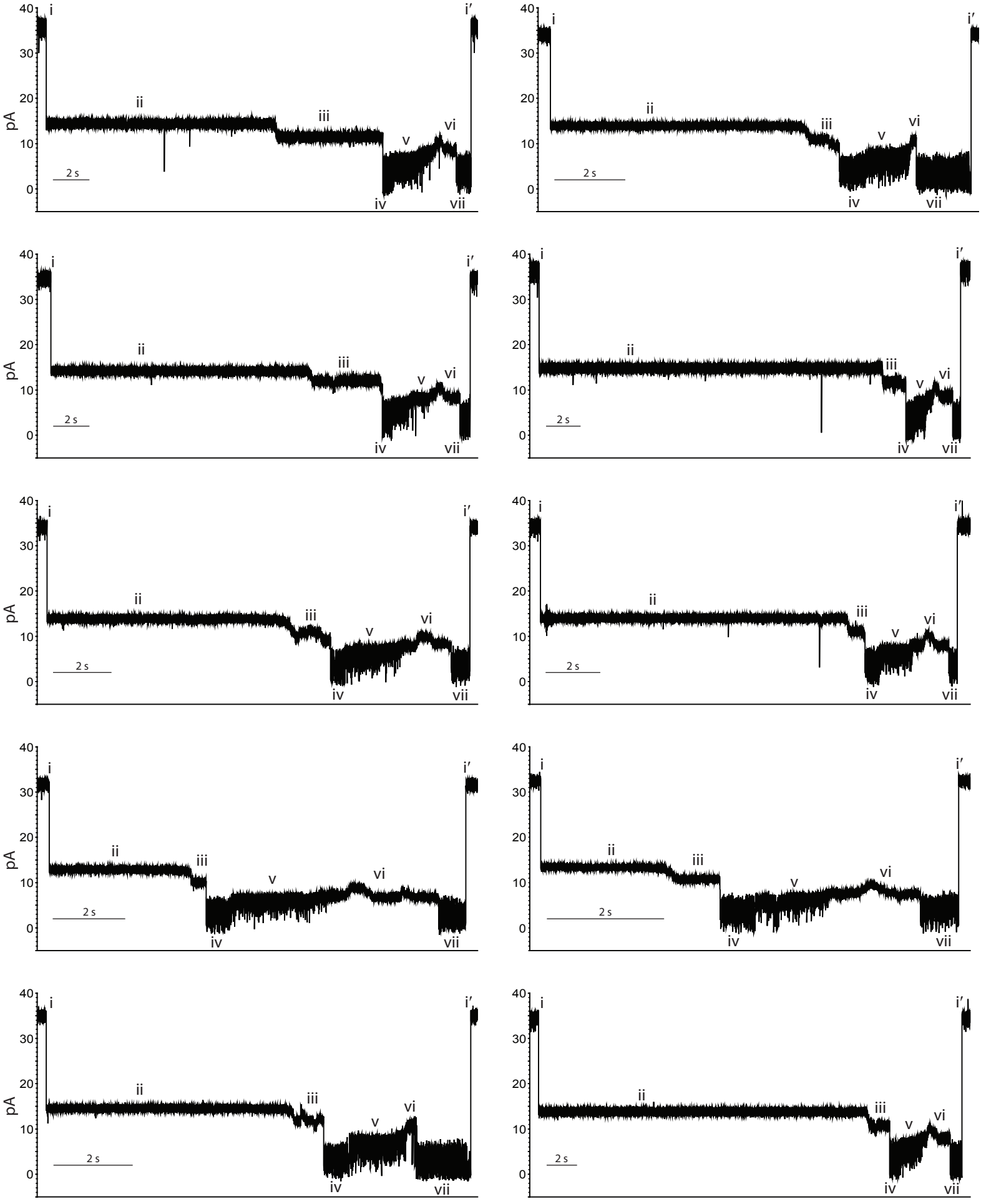


Supplementary Figure 2 Bulk phase assays of ClpX/ATP-dependent unfolding and translocation of substrate proteins bearing the long, charged ssrA-tagged C-terminal tail (see Online Methods for details). (**a**) SDS/PAGE gel showing substrate protein S1 degradation by ClpXP in the presence of ATP. Lanes 1-4 are a time course of S1 digestion in the presence of ClpX, ClpP, and ATP. Reactions minus ClpX (lane 6), minus ClpP (lane 7), or minus ATP (lane 8) showed no comparable degradation. Lane 9 is absent S1. (**b**) ClpX(P)/ATP-dependent quenching of a titin-GFP-tagged variant of the substrate protein S2-35 (C-terminus>charged-flexible tail>Smt3>titin>GFP>Smt3>N-terminus). ClpX+ATP n=4, ClpXP+ATP n=4, ClpP+ATP n=4, ClpXP(no ATP) n=3.

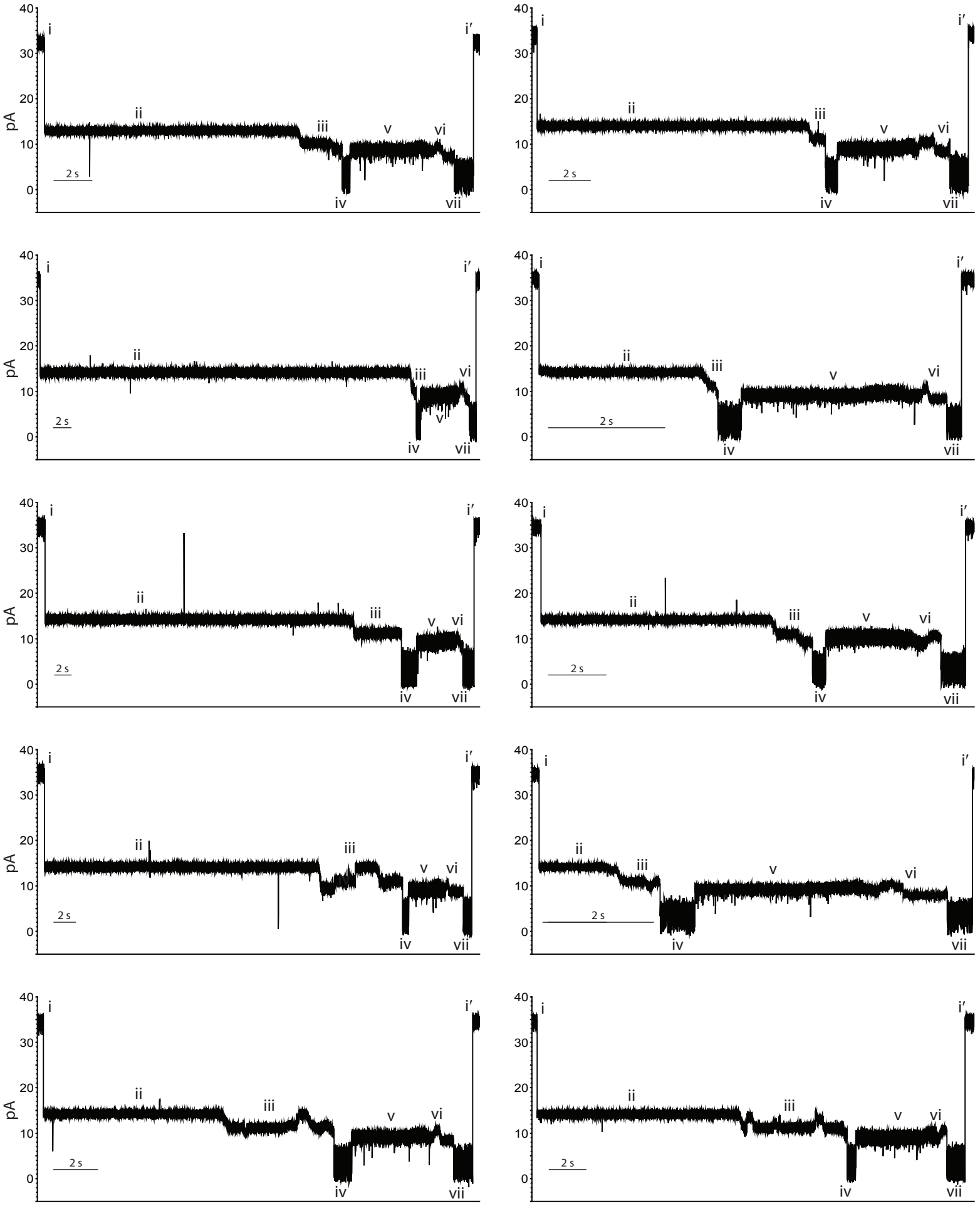
Supplementary Figure 3 Example traces arising from ClpX/ATP-dependent protein substrate translocation. (**a**) ClpX/ATP-dependent S1 translocation events.

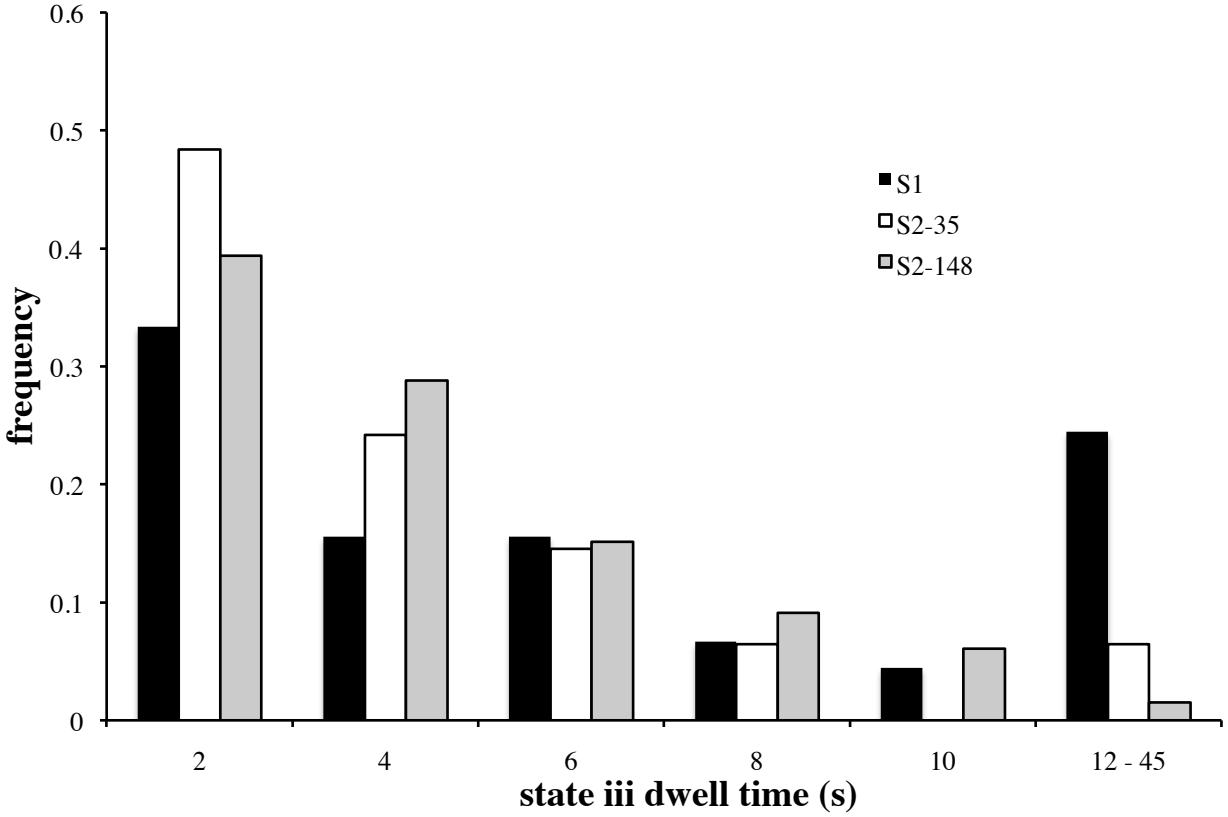


Supplementary Figure 3 Example traces arising from ClpX/ATP-dependent protein substrate translocation. (**b**) ClpX/ATP-dependent S2-35 translocation events.

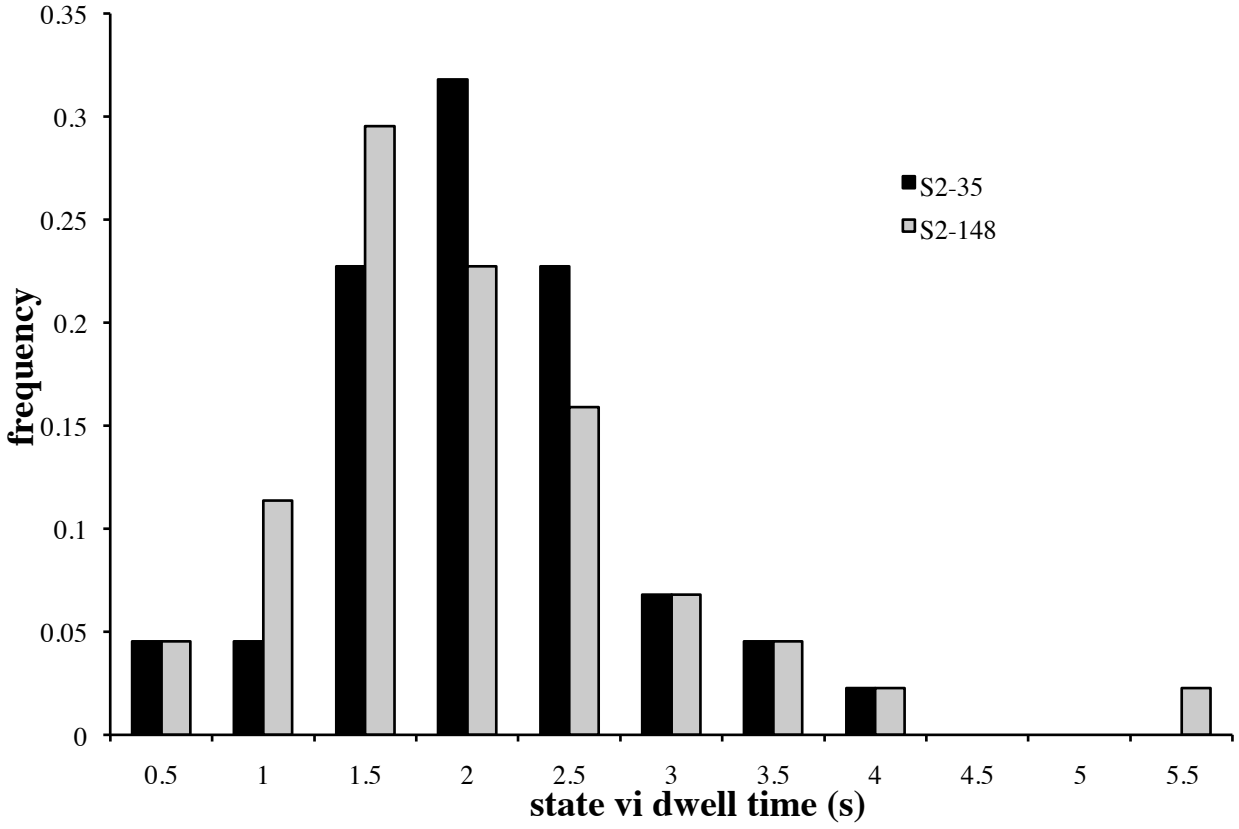


Supplementary Figure 3 Example traces arising from ClpX/ATP-dependent protein substrate translocation. (**c**) ClpX/ATP-dependent S2-148 translocation events.

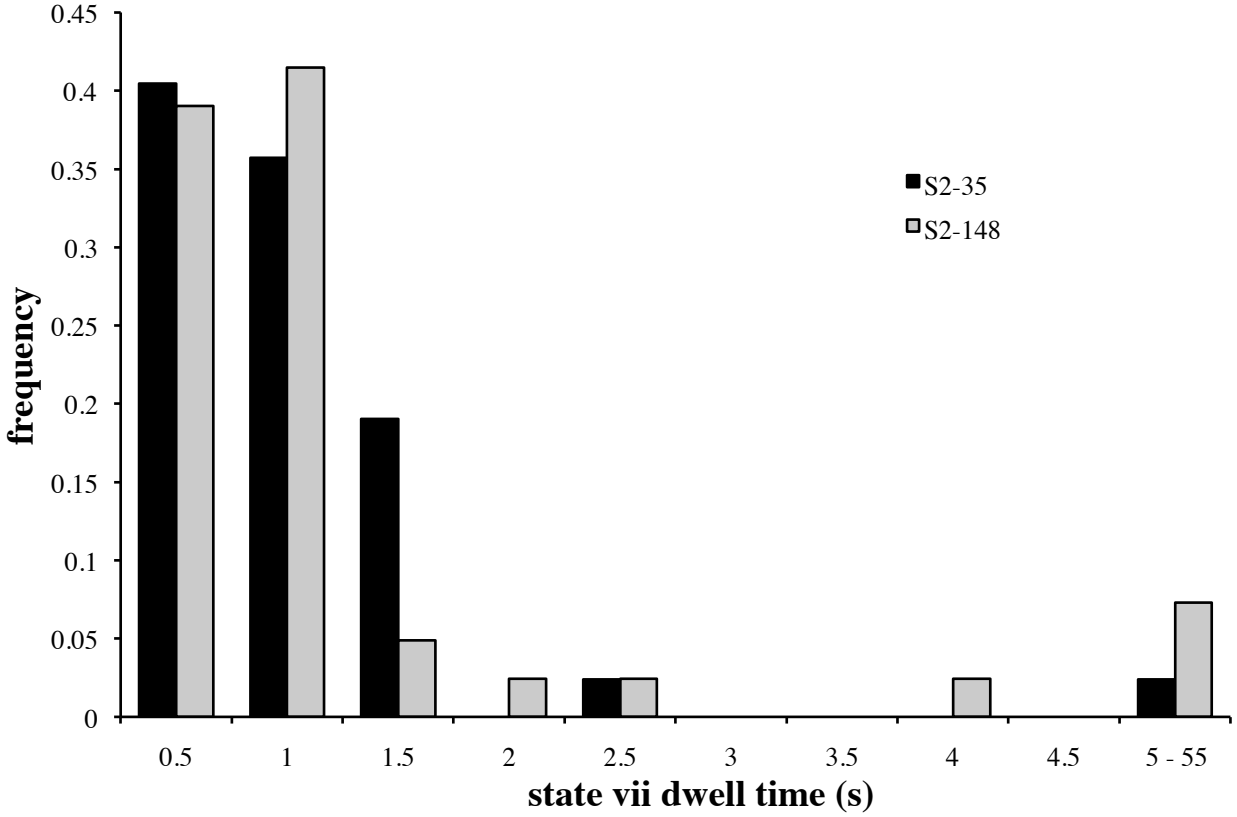




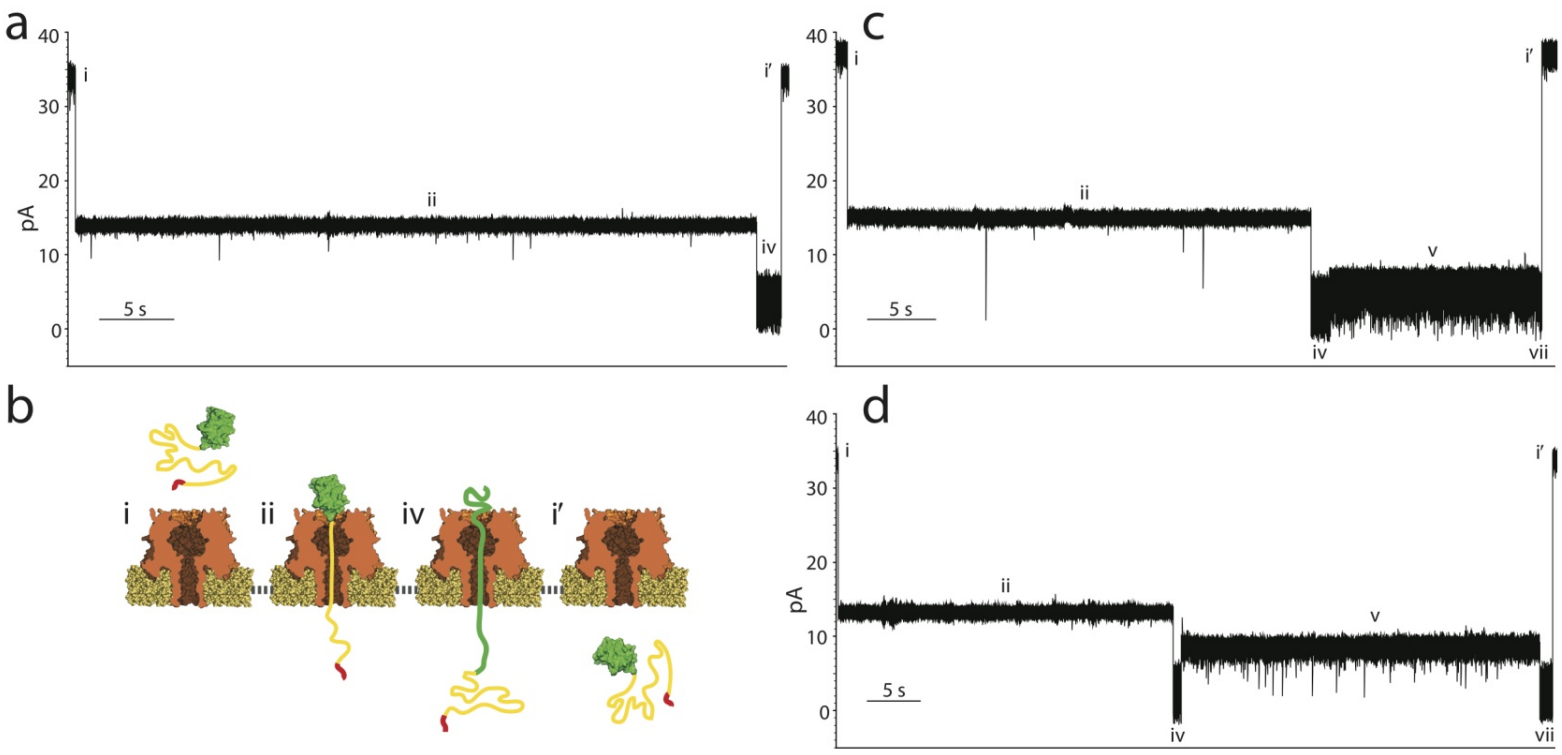
Supplementary Figure 4 Comparison of ionic current state iii dwell times for ClpX/ATP-dependent translocation events. Black bars: median = 4.15 s, IQR = 7.52, n = 45. White bars: median = 2.12 s, IQR = 3.23, n = 62. Gray bars: median = 2.76 s, IQR = 3.98, n = 66.



Supplementary Figure 5 Comparison of ionic current state vi dwell times for ClpX/ATP-dependent translocation events. Black bars: median = 1.80 s, IQR = 0.97, n = 44. Gray bars: median = 1.61 s, IQR = 0.97, n = 44.



Supplementary Figure 6 Comparison of ionic current state vii dwell times for ClpX/ATP-dependent translocation events. Analyzed events included ClpX/ATP-dependent ramping states iii and vi. Black bars: median = 0.63 s, IQR = 0.93, n = 42. Gray bars: median = 0.66 s, IQR = 0.90, n = 41.



Supplementary Figure 7 Ionic current traces showing translocation of the three model proteins absent ClpX/ATP-dependent mechanical work (no ramping states iii/vi). Following state ii, all protein substrates we examined eventually unfolded and translocated due to the 180 mV applied potential. All events exhibited more widely distributed state dwell times compared to ClpX-mediated events (**Fig. 3**). (**a**) S1 translocation. Note the absence of state iii compared to **Figure 2a**. (**b**) Model of ClpX/ATP-independent protein S1 translocation. Cartoons i-i′ correspond to ionic current states i-i′ in **a**. (**c**) S2-35 translocation. Note absence of ramping states iii and vi compared to **Figure 2c**. (**d**) S2-148 translocation. Note absence of ramping states iii and vi compared to **Figure 2d**.

CLPX CATALYZED PROTEIN UNFOLDING AND TRANSLOCATION THROUGH A BIOLOGICAL NANOPORE

by

Douglas B. Marks

A Thesis Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Bachelor of Sciences

in Bioengineering

at

The University of California, Santa Cruz

June 2012

CLPX CATALYZED PROTEIN UNFOLDING AND TRANSLOCATION

THROUGH A BIOLOGICAL NANOPORE

by

Douglas B. Marks

A Thesis Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Bachelor of Sciences

in Bioengineering

at

The University of California, Santa Cruz

June 2012

Faculty Mentor Date

School of Engineering Approval Date

ABSTRACT

CLPX CATALYZED PROTEIN UNFOLDING AND TRANSLOCATION THROUGH A BIOLOGICAL NANOPORE

by

Douglas B. Marks

The University of California, Santa Cruz 2012
Under the Supervision of Mark Akeson

Single-molecule analysis with biological nanopores has the potential for widespread sequencing applications. DNA-bases can be distinguished by the residual current through a nanopore when polynucleotide translocation is coupled to the activity of a processive motor. This achievement provides the proof of principle for sequencing linear biomolecules, particularly DNA, RNA and protein. Here, we present a novel technique for enzyme catalyzed protein translocation through a nanopore. ClpX is an ATP-dependent unfoldase in the AAA+ family from *E. coli*, which we coupled with the biological nanopore α-hemolysin to simultaneously unfold and translocate protein substrates. Proteins of interest are prepared for analysis with an engineered C-terminal sequence consisting of a polyglycine-serine-aspartate linker and ClpX recognition tag, ssrA. The negative charge of this unstructured tail, contributed by the aspartate residues, allows these modified proteins to be threaded through α-hemolysin under voltage. Steric hindrance of the protein folded domain and lack of concentrated internal charge prevent immediate translocation after capture. In the presence of ATP, ClpX will bind the exposed ssrA tag and catalyze

processive unfolding and translocation of the substrate protein through the nanopore. We used residual current traces to differentiate three substrates based on their engineered structural features. Working directly at the single-molecule level, the data we produced is unprecedented for protein structural analysis. This technique is fundamental to developing the world's first protein sequencing device.

Faculty Mentor Date

Recent advances in single-molecule analysis using biological nanopores have been instrumental in the development of a novel DNA sequencing technology[1, 2]. Coupling the motor activity of a processive enzyme to DNA translocation through a nanopore is fundamental to achieving the temporal resolution necessary to correctly sequence nucleotide identity[3]. The successes of nanopore DNA sequencing illustrate the potential for porin structures as versatile biomolecular sensors[4]. An increasing need for techniques that directly analyze RNA and protein motivates methods that adapt nanopores for these uses. While nanopore RNA sequencing only faces the barrier of a compatible processive motor, proteins will need a robust enzyme and an initial capture strategy that requires a charged, intrinsically unstructured terminus. Polypeptide sequence analysis with a nanopore will circumvent the shortcomings of current techniques that require a high copy number and can only resolve small fragments. Protein unfolding and translocation through a nanopore with a molecular motor is the first step towards a revolutionary protein sequencing and analysis platform. We demonstrate this achievement herein by using the processive *Escherichia coli* unfoldase ClpX to translocate single protein molecules through an alpha-hemolysin nanopore. We apply this novel technique to a set of engineered SUMO-based substrates that yield residual current signal characteristics indicative of enzyme-catalyzed translocation (Figure 1C).

ClpX is an *E. coli* molecular motor with a homohexameric ring structure that can bind, unfold, and translocate tagged proteins in an ATP-dependent manner[5, 6]. The conserved, approximately 230 amino acid ATPase module, integral to ClpX function, places this enzyme in the AAA+ superfamily. This module is characterized by catalyzing ATP hydrolysis to overcome energy barriers to unfold and translocate macromolecules[7]. ClpX initiates unfolding of a targeted protein first by binding the short peptide recognition tag ssrA (n-AANDENYALAA-c, 11 aa)

through specific amino acid contacts with the C-terminal LAA[8, 9]. Mechanical force generated in the ATPase module is leveraged to overcome enthalpic and entropic barriers and dismantle the folded domains[10]. Directed, rigid-body structural modulations repeatedly shift pore-loops inside the ClpX ring that advance along the protein substrate by making steric contacts with the peptide backbone and side chains[11]. The GYVG loop, in particular, is highly conserved in AAA+ proteins and functionally implicated in substrate translocation and protein unfolding[12]. In the cell, ClpX binds to the tetradecameric double-ring protease ClpP, forming the ClpXP protein degradation complex[13]. Flexible N-terminal loops protruding from each ClpP monomer facilitate plastic ring-ring interactions that allow ClpX hexamers to bind on both faces of the ClpP double-ring[14, 15]. Cellular proteins marked with the C-terminal ssrA tag will be bound by ClpX on either side of ClpP and translocated bidirectionally into the ClpP lumen[13]. Fourteen serine protease domains, one for each ClpP monomer, will coordinate the hydrolysis of peptide bonds, resulting in amino acid fragments that diffuse laterally through gaps between each ClpP ring[16].

In order to successfully translocate proteins through a biological nanopore using ClpX, we conceived a protein domain suitable for voltage-driven threading. To initiate substrate protein translocation we designed and engineered an intrinsically unstructured C-terminal linker region with substantial negative charge. This tail was composed of alternating glycine and serine residues with intermittently placed aspartates that increase in frequency toward the C-terminus (see SUMO-PolyN sequence below). The amino acid sequence of this region provided a soluble, yet unfolded linker appropriate for threaded capture in alpha-hemolysin. Carboxylate groups of numerous aspartate residues provided the negative charge and allowed the substrate to be

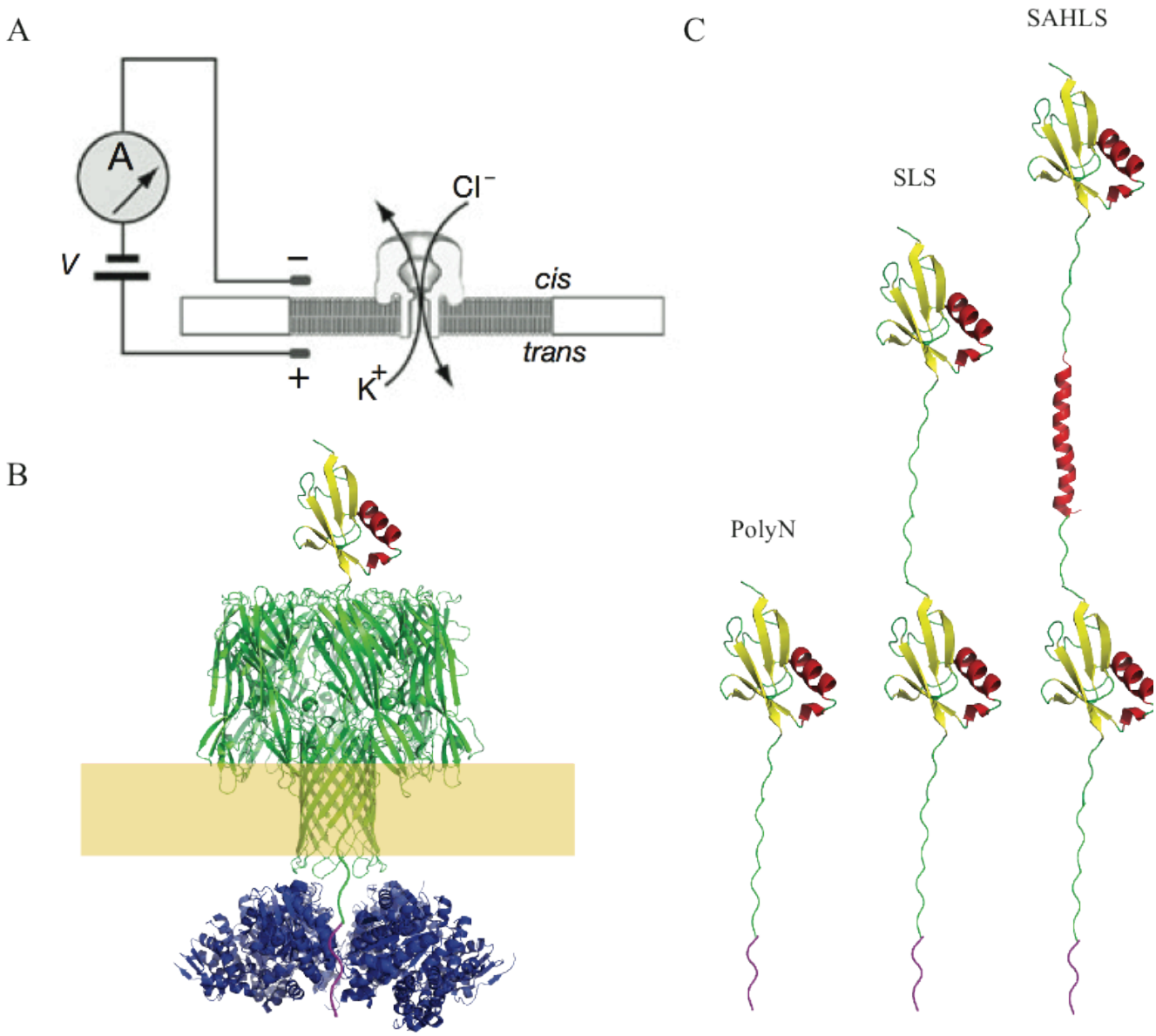


Figure 1 | The biological nanopore apparatus, molecular orientation of ClpX translocation events, and structure of engineered protein substrates. The experimental set-up and basic conditions were used for all recordings. **A)** Nanopore device consists of a single, isolated alpha-hemolysin channel inserted into an artificial planar lipid bilayer separating two wells of PD buffer. An applied voltage across the membrane produces an ionic current that diminishes in the presence of a captured protein molecule. **B)** Substrate protein (PolyN shown) is added to the *cis* well and is captured in alpha-hemolysin (green). ClpX (blue) isolated to the *trans* compartment binds the exposed ssrA tag (purple) and catalyzes unfolding and translocation into the *trans*. **C)** Engineered substrates designed for capture and subsequent translocation by ClpX. Folded domains are Smt3 (SUMO), and the linker regions are shown in green. Each substrate carries the C-terminal ssrA recognition tag (purple) to allow binding and procession by ClpX. [Illustrations are not to scale. Panel A adapted from Cherf et al. 2012[1], panels B and C are PyMol renders of PDB entries 7AHL (α-HL), 3HTE (ClpX), 1EUV (Smt3) and 2I7U (helix).]

captured in a nanopore at low voltages (~100 mV). The charged tail was drawn by the applied potential and translocated through the pore, in microseconds or less for 76 amino acids, until a folded domain N-terminal to the linker was held atop the *cis* side vestibule (Figure 1B). Tertiary structure stability, combined with the size and lack of concentrated negative charge of the folded domain, prevented the threaded protein from immediately translocating due to the applied voltage. The captured protein was oriented with the linker tail threaded through the nanopore and the C-terminal end exposed on the *trans* side, while the folded domain rests in the *cis* compartment and prevents further translocation of the substrate (Figure 1B). The first substrate we engineered for capture by voltage threading was comprised of the SUMO domain (Smt3) and the 76 amino acid C-terminal unstructured linker described previously. The downstream sequence terminated in ssrA (wt sequence, or "natural"), resulting in the full construct Smt3-PolyGSD76ssrA[Natural] (SUMO-PolyN, or PolyN for short). We initially experimented with a modified ssrA tag that replaced non-essential residues with aspartates[9], but found the additional negative charge was the C-terminus was not necessary to efficiently capture the protein. Capture of this protein substrate under voltage provided a *trans* exposed ssrA tag that was recognized, bound and processed by a ClpX hexamer (Figure 1B). Although separated from the substrate folded domain by alpha-hemolysin, therefore unable to leverage specific loops that direct unfolding[12], the force generated by ClpX was sufficient to unfold SUMO with relative ease after binding. ClpX mediated translocation occurred in the same direction as the current (voltage polarity) and produced characteristic amplitude trace signatures caused by increased blockades during the passage of folded domains (Figure 2). These levels and behaviors were distinguishable by eye, in real-time, from the initial capture state based on amplitude and root mean square noise levels, duration and state transitions.

A biological nanopore apparatus[1] was used to conduct experiments (Figure 1A). A single alpha-hemolysin (α-HL) nanopore isolated in an artificial lipid bilayer established a channel between two compartments of solution. These wells are termed *cis* and *trans* to denote α-HL orientation and contain the negative and positive electrodes, respectively. An integrating patch-clamp amplifier (Axopatch 200B, Molecular Devices) applied a constant 180 mV potential between the electrodes during substrate protein capture and translocation. This fundamental setup was used in successive experiments that probed various substrates for translocation signal differences. The set of substrates were increasingly complex variations on the basic SUMO-linker structure (Figure 1C). However, each of these variations depended on an identical aspartate-charged linker to be captured in α-HL by voltage-driven threading. Events were initiated exclusively by this mechanism: substrate protein in the *cis* well is threaded through α-HL to expose a recognition tag to ClpX(P) isolated in the *trans* compartment. As expected, the capture amplitudes were consistent for each substrate under constant conditions. All experiments were performed in a ClpXP optimized protein degradation (PD) buffer[11] containing 200 mM KCl, 5 mM $MgCl_2$, 10% glycerol, and 25 mM HEPES-KOH pH 7.6. Adenosine triphosphate (ATP) disodium salt was added to 5 mM in PD buffer aliquots daily to minimize spontaneous hydrolysis in solution. ClpX catalysis is ATP-dependent[5, 6], and both ClpX and ATP were present in all experiments in which characteristic translocation signals were observed (Figures 2 - 4). Control experiments run in the absence of ATP, ClpX or both did not yield the signal characteristics indicative of a ClpX(P) catalyzed event (Figure 5). However, the capture amplitudes were not affected by these permutations, which lends evidence to substrate structure as the major determinant of residual current.

Experiments were initially run with expressed ClpX in the *trans* compartment, and ClpP was also present in a lower concentration as a co-purifying contaminant from *E. coli*. Therefore, experiments performed with this condition were labeled as ClpX(P) to denote uncertainty in the presence of an accompanying ClpP oligomer. To test the effect of ClpP on the substrate, if any, controls were run with ClpP present in the absence of ClpX. Characteristic translocation signals were not observed in the absence of ClpX. However, control experiments where ClpX was absent and captures were not ejected long past the expected duration to binding (~ 30 sec) eventually shed light on the possibility of a voltage-induced translocation. These events did not contain characteristic signal elements and presented state durations with wide distributions (Figure 5). As expected, voltage-induced translocations of SUMO-based substrates with the PolyGSD-ssrA tail occurred after long captures in PD buffer with no ClpX or ATP added. In later experiments, individually expressed ClpP was supplemented to the ClpX(P) solution in an attempt to increase translocation efficiency[17]. These ClpXP events were distinct from ClpX(P) only in the assumption that ClpP was always present in a catalyzed translocation. However, it had also been reported that ClpX hydrolyzes ATP at a rate 2-3 times that of the ClpXP complex[18]. Anion exchange chromatography with an FPLC apparatus produced pure ClpX that could be tested on the nanopore. No differences in characteristic signal were observed in the absence of ClpP, so all subsequent experiments were run with pure ClpX to leverage the higher ATP hydrolysis rate and corresponding force production.

The principal experiment demonstrated the first instance of enzyme catalyzed protein translocation and determined reliable signal characteristics. The primary engineered protein substrate, Smt3-PolyGSD76ssrA-Natural (PolyN) (Figure 1C), was designed to test voltage driven threading of the aspartate-charged tail and yield a short, simple signal should ClpX(P)

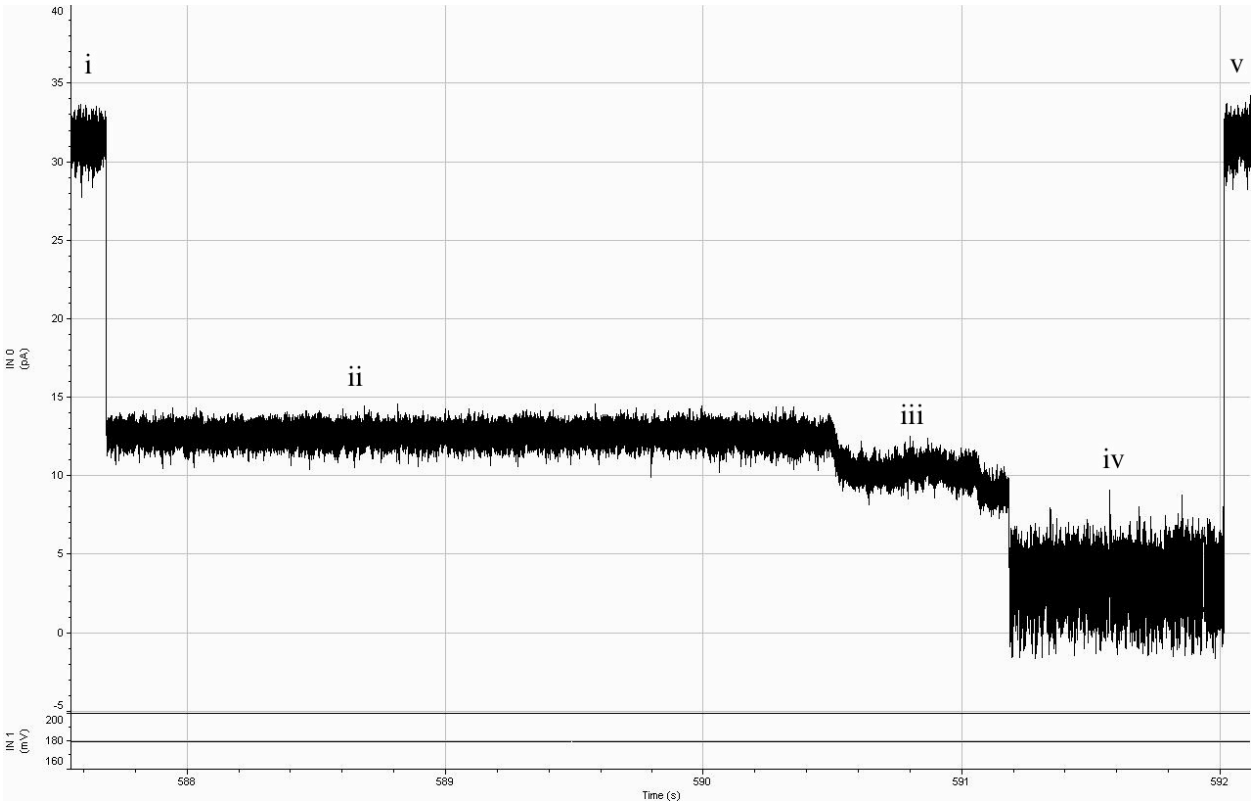


Figure 2 | Characteristic trace of ClpX(P) catalyzed translocation of a PolyN folded substrate through an alpha-hemolysin (α-HL) nanopore. Translocations were observed using ~100 nM ClpX complex in 4.5 mM ATP *trans* compartment conditions. i) Open channel current through α-HL in PD/ATP buffer at 30ºC at a constant 180 mV is ~ 35 pA with 1.3 pA RMS. ii) Capture of the PolyN substrate is indicated by the state change to a constant average of 14 pA with 0.8 pA RMS. iii) ClpX(P) bound events were identified by the presence of a "ramping" state, often reaching amplitudes below 10 pA in one or more transitions. iv) Smt3 folded domain translocation through a nanopore. Characterized by an average of 3.8 pA and 1.8 pA RMS, and a median duration of 775 ms, this state signature was dominated by the presence of disordered Smt3 polypeptide in the α-HL vestibule. v) Return to OCC signals the completion of substrate translocation as ClpX moved the peptide N-terminus from pore and into the *trans* solution.

successfully bind, unfold and translocate. The Smt3 folded domain is comprised of 106 amino acids arranged into four ß-strands and a single alpha helix. This small structure is suitable for translocation to completion in a reasonable time (< 1 s), yet is large and stable enough to rest atop the *cis* exposed rim of α-HL while consistently allowing substantial current (> 1/3 of the open channel current [OCC]) to pass at a reliable amplitude (Figure 2, state ii). The transition from state i to state ii in Figure 2 was indicative of the unstructured, charged tail of Poly-N threading through the pore. This process was concentration dependent and was not rate limiting

to observe translocation at any concentration tested. 1 μM PolyN was added to the *cis* well to begin an experiment and observe captures, but this concentration is negligible because only one molecule at a time will be functionally available to ClpX(P) after threading through α-HL. At this concentration, it was common to observe PolyN capture from OCC in less than 1 second after sustained reversed polarity, while state ii lasted 46 seconds on average for capture events in which translocation (Figure 2, state iv) was observed. During the capture (Figure 2, state ii) of a single PolyN molecule, association of a *trans* side ClpX hexamer was signaled by a "ramping" state that was immediately followed by a translocation state. Characteristic ramping (Figure 2, state iii) was named for the relatively continuous, decreasing amplitude transitions that preceded the translocation state. The constant capture amplitude in state ii "ramps" down to one or more discreet levels over a median duration of 4.3 seconds, and consistently ends with a sharp transition into the translocation signature (Figure 2, state iv). As expected, the translocation state was always followed by return to the OCC. Dissociation of any captured molecule or complex should only be possible by substrate translocation to the *trans* compartment, especially given the constant voltage and direction of ClpX activity. The experimental design, conditions and results provided evidence of the first enzyme-catalyzed protein translocation through a nanopore.

A second substrate molecule was engineered to further test whether or not the observed signal was ClpX catalyzed protein translocation. In particular, we wanted to know how an internal unstructured region of sequence would affect the event trace mid-translocation. Such a feature would contribute a new element to the signal and yield some insight into the behavior of the "ramping" state transition. Sequence for an additional SUMO domain and linker were concatenated to the 5′ end of the Poly-N gene construct by PCR synthesis. The full construct Smt3-PolyGS35-Smt3-PolyGSD76ssrA (SUMO-linker-SUMO-PolyN, or SLS) expanded on

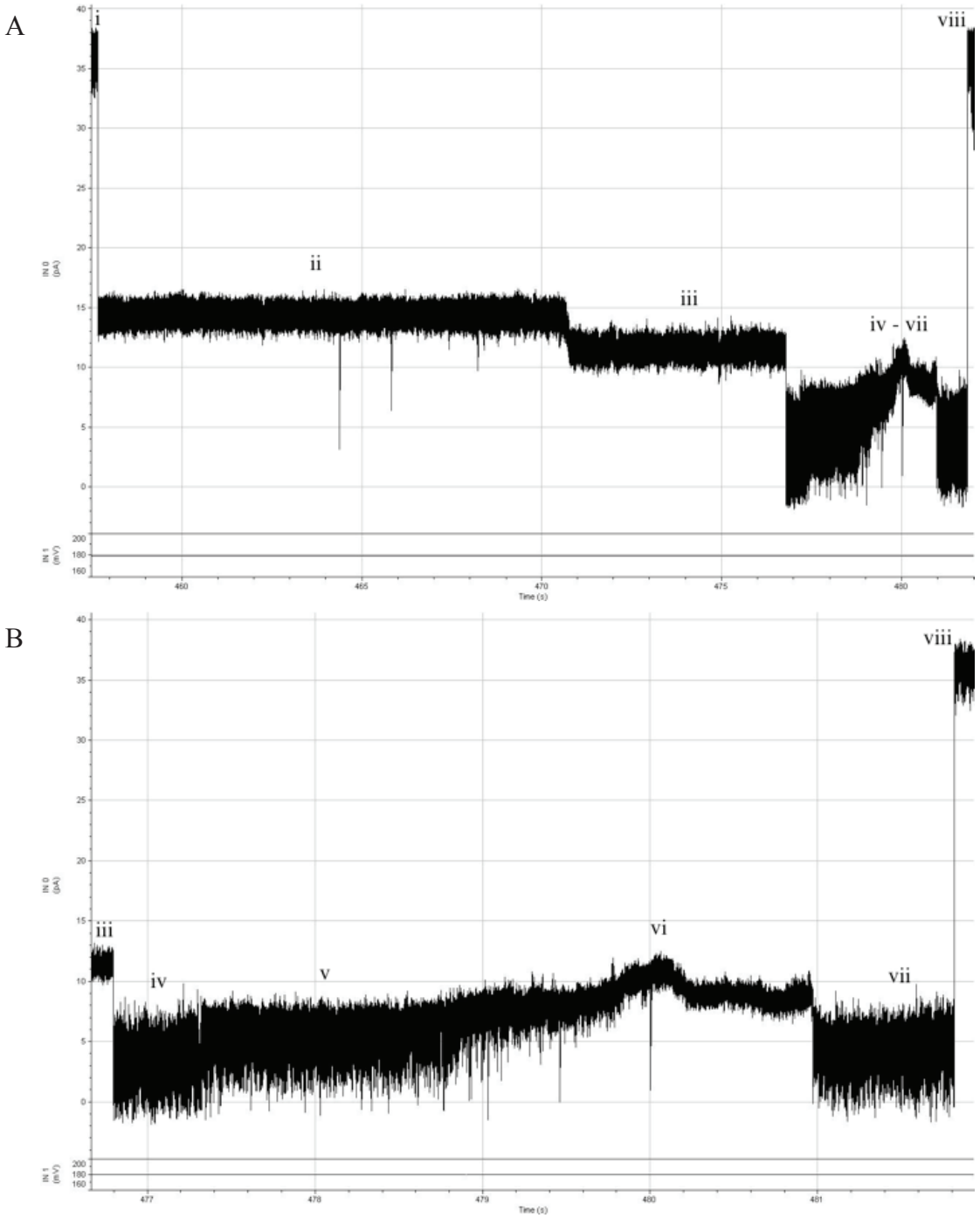


Figure 3 | SLS substrate translocation mediated by ClpX(P). Experimental conditions are identical to those used for the data presented in Figure 2. **A)** Full event trace from capture to release in the *trans* compartment. States i – iv are consistent with Poly-N signal as in Figure 2. **B)** Zoom-in on the SLS translocation signature. iv) The amplitude, RMS noise, and duration are statistically similar to the equivalent state in Figure 2. v) The signal characteristics transition away from the Smt3 folded domain translocation. vi) The linker state displays close resemblance to state iii. vii) A second translocation state with identical characteristics to state iv is observed directly before the event terminates with a return to open channel current in state viii.

PolyN structure with two additional N-terminal regions (Figure 1C). Due to ClpX working from the C-terminus, we expected the resulting translocation signature to be near identical to the PolyN trace up to the end of state iv (Figure 2). We hypothesized that a current signature very similar to state iii, which would correspond to the initial binding of ClpX to an unstructured region, would follow this state. Similarly, we postulated that the progression from late state iii to state v should be replicated at the end of an SLS translocation event. The presence of an Smt3 folded domain preceded by a GS-linker at the N-terminus of SLS made this substrate virtually identical to PolyN, in terms of structural orientation about α-HL, after the first Smt3 domain was processed by ClpX. Figure 3 shows a characteristic trace of an SLS translocation mediated by ClpX(P). As predicted, the event progression through state vi (Figure 3A) followed the same pattern as the Poly-N translocation signature. We believe that state v (Figure 3B) represented a transition from folded domain to unstructured linker translocation. This data may suggest that ClpX mediated translocation resulted in gradual linearization of the linker region as it is pulled through the pore. ClpX exerts force on the substrate to achieve translocation and unfolding, which should apply tension until the barrier to unfolding is broken[10]. The amplitude, RMS noise, and ramping behavior of state vi (Figure 3B) mimicked the properties of initial binding observed in state iii. Accordingly, the signal produced during state vi represented a similar ClpX and substrate orientation relative to α-HL. This claim was supported by state vii (Figure 3B), which shows a second Smt3 translocation signal that is consistent with both state iv and PolyN (Figure 2, state iv).

The third substrate we tested expanded on SLS structure. We modified the linker region of SLS to examine the relationship between this structure and the resulting signal (Figure 3B, states v and vi). The final engineered substrate included 112 additional amino acids between

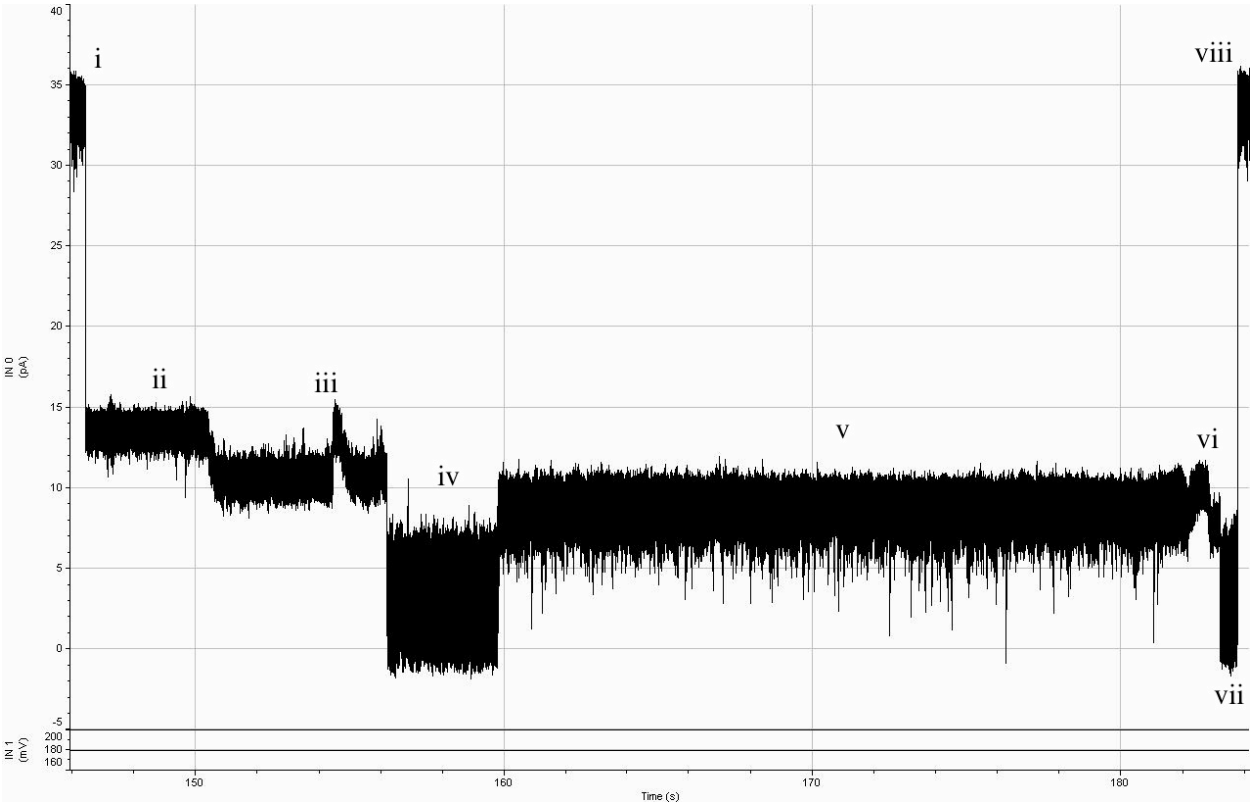


Figure 4 | ClpX catalyzed translocation of SAHLS. Unlike the characteristic SLS translocation, near maximum amplitudes of the translocation signature were reached immediately in the transition state. Event progress was identical to PolyN and SLS through state iv (Figures 2 and 3). State iii in this trace displayed a "failed" ramping event where the current amplitude returned to the capture level before ramping again and initiating Smt3 translocation. v) Transition state amplitude is ~3 pA higher than SLS (9-10 pA), RMS noise is comparable. vi) Peak amplitudes in the linker state are only slightly elevated compared to state v.

Smt3 domains of the original SLS design. This insert included putative alpha-helical sequence (six repeats of N-EAAK-C) surrounded by GS-linker (Figure 1C). We named this substrate SUMO-alpha helix linker-SUMO (SAHLS) to reflect the change to linker structure. This modification was chosen to increase the rigidity and structural orientation of the linker region, which allowed us to directly test the increase in tension hypothesis stated previously. If the signal produced by SLS translocation in the linker region was due to linearization under increased tension, then SAHLS translocation should not increase in amplitude between state v and state vi. We observed a ~3 pA increase in the transition state amplitude (Figure 4, state v) that was near maximal in the SAHLS signal post-ramping. Unlike SLS, this transition state did not increase in

amplitude significantly before reaching the characteristic linker state. Furthermore, the peak amplitudes of the linker states in both SLS (Figure 3B, state vi) and SAHLS (Figure 4, state vi) were consistent (~10 pA). We believe that the introduction of an alpha helix to the substrate linker structure prevented ClpX from applying tension to this region during translocation. The structural rigidity resulted in maximal amplitudes for the duration of the SAHLS linker translocation that was only seen after gradual increase in SLS. We hypothesized that intrinsic orientation of the linker allowed consistently higher residual current compared to a disoriented GS-linker that linearized under tension. This data suggested that an alpha helical structure in the linker region might have prevented excess obstruction in the vestibule of α-HL. The increased rigidity and orientation of the SAHLS linker, compared to SLS, resulted in a decreased blockade during translocation. Additionally, the combined duration of the transition and linker states of SAHLS (Figure 4, states v and vi) was significantly longer on average compared to the equivalent states of SLS (Figure 3B, states v and vi). This increased duration represented the additional time required for ClpX to translocate the 112 amino acids in this region. For a second time, we observed the expected outcome in the translocation signature using directed changes to SUMO-based substrates.

As mentioned previously, voltage-induced translocations were observed for all three substrates tested. These translocation signatures were all similar in that they did not display any ramping behavior before the translocation state began (Figure 5). The lack of this feature lead us to consider the ramping state as the primary criteria indicative of ClpX binding and subsequent activity. These current signatures occurred in the presence of ClpX along with the characteristic events described above. However, in the absence of ClpX or ATP, these event types were exclusively observed. This relationship of experimental conditions and current signatures lent

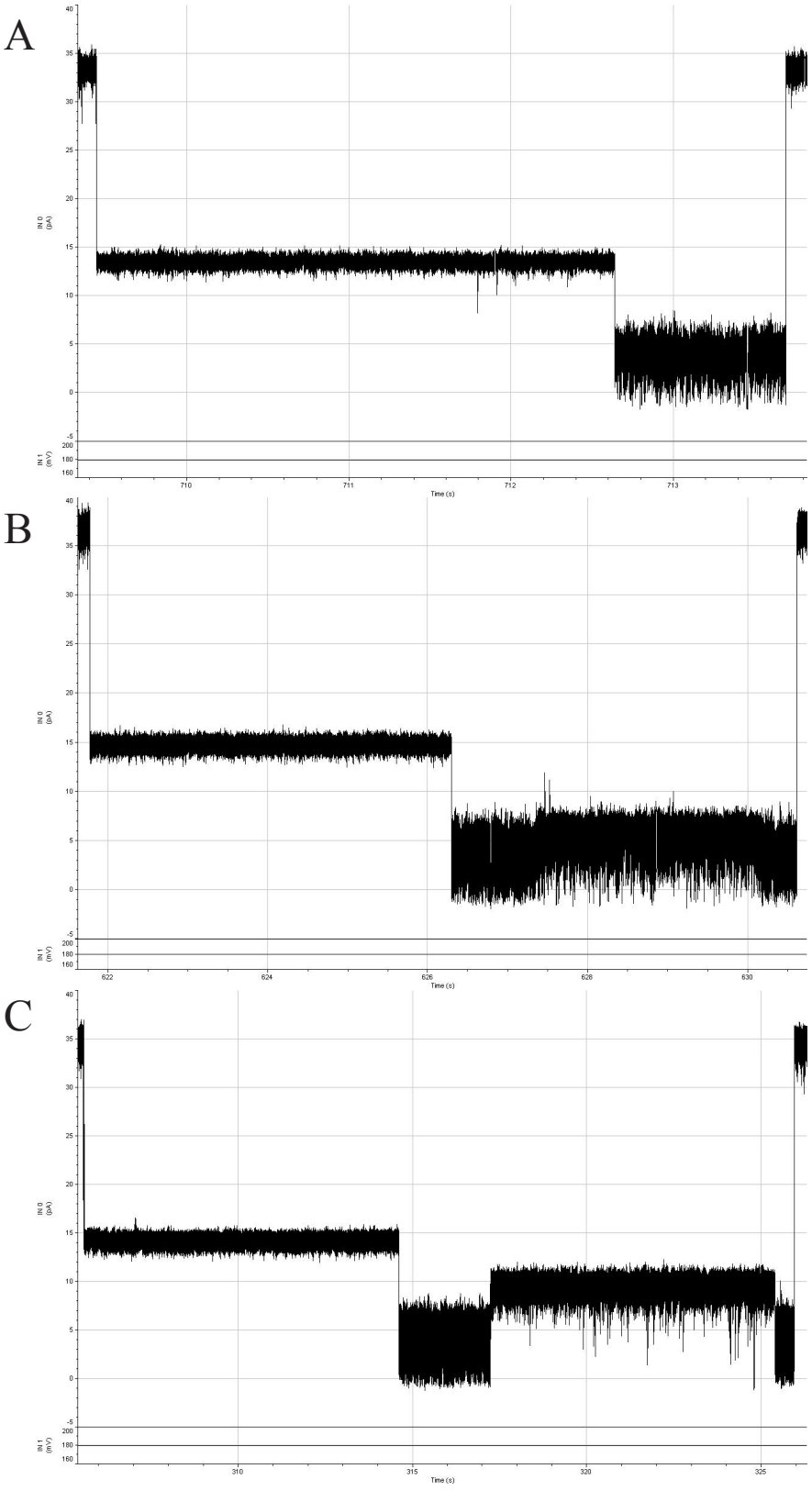


Figure 5 | Voltage-induced translocation of SUMO-based substrates. After a highly variable capture duration (< 5 sec – > 2 min), all substrates tested will eventually unfold and translocate due to the applied voltage. No ramping is observed, signal features are lost from SLS and SAHLS, and all states have widely distributed durations. **A)** PolyN. Compared to Figure 2, state iii is absent and state iv has a longer duration on average. **B)** SLS. Ramping of state iii and linker region of state vi (Figure 3) are not present. **C)** SAHLS Exhibits similar behavior to SLS with the corresponding states omitted (Figure 4). Note that the transition amplitudes for both substrates are consistent with ClpX mediated translocations as in Figures 3 and 4.

evidence to their voltage-induced mechanism. It was consistent that this type of translocation event would occur under any experimental conditions, while the signatures we believe to be indicative of ClpX activity were only observed in the presence of ClpX and ATP. Statistically, states in these events displayed significantly greater variability in duration compared to characteristic ClpX mediated translocation events. This is consistent with an unregulated translocation process that relies heavily on random structural fluctuations and charge-voltage affects rather than the relatively constant ATP hydrolysis rate and mechanical force imparted by a molecular motor. To further test the voltage translocation mechanism, control experiments were run in PD buffer absent ClpX and ATP at a lower, constant 120 mV. With this condition, translocation signatures were observed more infrequently and had longer durations on average compared to standard experiments at 180 mV. The decrease in force from the diminished voltage resulted in a larger barrier to initiating and completing the unfolding and translocation. Additionally, the linker states of SLS and SAHLS were not present during voltage-induced translocation events (Figure 5, B and C). This observation was consistent with an applied tension hypothesis stated previously, and presence of the linker state subsequently became the secondary criteria for identifying ClpX translocations of SLS and SAHLS. Without ClpX applying tension to the substrate at it traverses α-HL, the linker region could not become adequately linearized to produce the characteristic amplitude and RMS noise. It was also consistent that the SLS and SAHLS substrates displayed their respective transition states at the appropriate current amplitude values. Voltage associated forces caused the first Smt3 domain to translocate towards the *trans* well, at which point the linker region began to bias the residual current in the transition state. This state should be consistent with ClpX catalyzed event progression because the orientation of the polypeptide sequence relative to the α-HL vestibule and constriction will dominate the

current regardless of the presence of ClpX. The complications introduced by voltage-induced translocations of SUMO-based substrates will be mitigated in future experiments by analyzing more diverse substrates. Proteins with stable folded domains and less concentrated internal negative charge, such as GFP, will be instrumental in conducting the next phase of ClpX nanopore translocation experiments.

Current methods directed at developing nanopores as protein sensors mainly focus on binary outcomes or characterizing intrinsic properties. For example, the cutting-edge publications in protein sensing research with nanopores revolve around strategies to modify the nanopore to detect binding of a specific protein[4, 19, 20]. Previous studies have developed strategies to observe capture, trapping or unfolding of target proteins in nanopores, again relying on specific modifications and conditions to produce directed outcomes[21- 23]. While these experimental designs can be informative about properties of the particular protein of interest, they lack demonstration of breakthrough proof of principal results with widespread applications. Furthermore, single molecule protein sequencing is completely untouched by the present research paradigm. This may be correlated to the perceived difficulties of achieving a truly functional nanopore DNA sequencing platform, which would make the prospect of a protein analog seem near impossible to any reasonable scientist. However, the successes and advances of motor control and base reading of DNA in a nanopore, this year alone, leave little doubt that a revolutionary, commercialized device is just around the corner[1, 2].

The technique we present here, even in its fledgling stages, is a fundamental breakthrough that opens the door to a new era of versatile, informative protein analysis. The current challenges that limit the progress of this project are similar to those faced in the history of nanopore DNA analysis. These shared difficulties are the key advantage in expanding the

utility of this technology. The past obstacles experienced by researchers studying DNA with nanopores serve at beacons to guide and accelerate the progress of this protein analysis analog. The most impactful breakthroughs to date in the pursuit of nanopore DNA sequencing have improved on two essential components of the system: a processive motor to control translocation and a mutant nanopore to increase sensitivity and resolution. The use of phi29 bacteriophage DNA polymerase to control the rate of DNA translocation and leverage processive DNA replication on a nanopore paved the way for methods that couple motor control to translocation analysis[3]. The development of M2-NNN-MspA as a next-generation porin to accurately distinguish DNA bases using differences in residual ionic current has transformed the dream of nanopore DNA sequencing into a tangible reality[2, 24]. These specific aspects of a nanopore system that analyzes DNA sequence represent attainable targets to functionally developing a protein-sequencing platform.

Porin structures and translocase enzymes are ubiquitous in nature. Because these structures provide essential functions, interaction with extracellular environments and regulation of cellular composition, they can be found in all branches of life and with near endless variations. The task is daunting yet simple; discovery of the ideal translocase enzyme and porin structure will push the potential of a single-molecule protein-sequencing device to the cusp of actualization. Continuation of this work, in tandem with testing and characterization of new protein components, will yield rapid advances that scale and accelerate with the number of talented individuals devoted to further realizing this technology. Simultaneously, the final stages of development and optimization of a functional nanopore DNA sequencing device will lead the way by drawing a road map of the potential difficulties and their solutions. Based on the initial results of experimentation, the precedent set by the successful development of a DNA analysis

analog, and the world of amazing proteins waiting to be adapted to this cause, I am confident that single-molecule protein sequencing with a nanopore is an eventual reality.

Methods.

All synthetic genes were constructed using DNA oligonucleotides (IDT) and PCR synthesis with Titanium Taq PCR kits (ClonTech). The PolyN insert was ligated into a pET SUMO vector (Invitrogen). The SLS and SAHLS genes were constructed from PolyN amplicons and additional oligonucleotides, then ligated in pE SUMO vectors (Life Sensors). Plasmid clones for His_6-tagged ClpX-ΔN_3 (residues 62-243 concatenated with L20 linkers)[25] and ClpP were provided by Andreas Martin at the University of California, Berkeley. All plasmids were transformed into *E. coli* BL21 (DE3) chemocompetent cells and induced with IPTG. Cell lysates were His_6 purified on Ni-Nta Agarose (Qiagen) affinity columns and by buffer exchange in Amicom Ultra-4 Centrifugal Filter units (Millipore). ClpX-ΔN_3 expression was additionally purified using an anion exchange affinity column and a BioLogic DuoFlow FPLC apparatus (Bio-Rad). Buffer conditions for purification and storage of Clp proteins were as described by the Sauer Lab Wiki (PD buffer as described previously). Heptameric α-HL was provided by Oxford Nanopore Technologies. Diphytanoyl-phosphatidylcholine (DPhPC) lipid was purchased from Avanti Polar Lipids.

Setup of the nanopore device and insertion of an α-HL nanopore into a lipid bilayer have been described[26]. Briefly, a single α-HL nanopore was inserted into a lipid bilayer that separates two wells that each contained 100 µl of PD buffer (pH 7.6). A constant 180 mV potential was applied across the bilayer and ionic current was measured through the nanopore between Ag/AgCl electrodes in series with an integrating patch clamp amplifier (Axopatch 200B, Molecular

Devices) in voltage clamp mode. Data were recorded using an analog-to-digital converter (Digidata 1440A, Molecular Devices) at 100 kHz bandwidth in whole-cell configuration then filtered at 2 kHz using an analog low-pass Bessel filter. Experimental conditions were prepared by the daily preparation of PD/ATP 5 mM and PD/ATP 4 mM. ClpX was diluted 1:10 in PD/ATP 5 mM for a final concentration of 30-100 nM ClpX in 4.5 mM ATP final. ClpX solution was used to fill the entire system before isolation of a single α-HL nanopore. Upon insertion, the *cis* well was perfused with ~6 mL PD/ATP 4 mM. Experiments were conducted at 30 °C with 1-2 μM substrate added to the cis well. Protein substrate capture events were ejected with reserve polarity due to pore clogs or after a predetermined duration. Voltage-induced translocations were frequently ejected to prevent clogging and to increase the efficiency of data collection. A single nanopore experiment is defined as the time during which ionic current data were acquired from one α-HL nanopore in an intact bilayer before termination by bilayer rupture, loss of channel conductance or completion of a preset number of translocation events. (Adapted from Cherf et al. 2012)[1].

Acknowledgements.

To Mark Akeson and Jeff Nivala – you guys made this possible! To Sarah Shashaani for sticking by me through thick and thin – I wouldn't have come this far, or enjoyed it all so much along the way without you. Thank you for all the guidance and opportunities...

GAME ON!

Protein sequences.

ClpX (wt, 424 aa, 46,356 Da)
MTDKRKDGSGKLLYCSFCGKSQHEVRKLIAGPSVYICDECVDLCNDIIREEIKEVAPHRE
RSALPTPHEIRNHLDDYVIGQEQAKKVLAVAVYNHYKRLRNGDTSNGVELGKSNILLIG
PTGSGKTLLAETLARLLDVPFTMADATTLTEAGYVGEDVENIIQKLLQKCDYDVQKAQ
RGIVYIDEIDKISRKSDNPSITRDVSGEGVQQALLKLIEGTVAAVPPQGGRKHPQQEFLQV
DTSKILFICGGAFAGLDKVISHRVETGSGIGFGATVKAKSDKASEGELLAQVEPEDLIKFG
L IPEFIGRLPVVATLNELSEEALIQILKEPKNALTKQYQALFNLEGVDLEFRDEALDAIAK
KAMARKTGARGLRSIVEAALLDTMYDLPSMEDVEKVVIDESVIDGQSKPLLIYGKPEAQ
QASGE

ClpP (wt, 207 aa, 23,187 Da)
MSYSGERDNFAPHMALVPMVIEQTSRGERSFDIYSRLLKERVIFLTGQVEDHMANLIVA
QMLFLEAENPEKDIYLYINSPGGVITAGMSIYDTMQFIKPDVSTICMGQAASMGAFLLTA
GAKGKRFCLPNSRVMIHQPLGGYQGQATDIEIHAREILKVKGRMNELMALHTGQSLEQI
ERDTERDRFLSAPEAVEYGLVDSILTHRN

SUMO-PolyN (195 aa, 19,613 Da)
MGSSHHHHHHGSGLVPRGSASMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKI
KKTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGG
GGSSGGSGGSGSSGDGGSSGGSGGSGSSGDGGSSGGSGGDGSSGDGGSDGDSDGSDGD
GDSDGDDAANDENYALAA

SLS-PolyN (338 aa, 34,767 Da)
MGHHHHHHGSLQDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLM
EAFAKRQGKEMDSLRFLYDGIRIQADQAPEDLDMEDNDIIEAHREQIGGGGSGSGGSGS
GGSGSQNEYRSGGSGSGGSGSGGSGMGSSHHHHHHGSGLVPRGSASMSDSEVNQEAKP
EVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQ
ADQTPEDLDMEDNDIIEAHREQIGGGGSSGGSGGSGSSGDGGSSGGSGGSGSSGDGGSS
GGSGGDGSSGDGGSDGDSDGSDGDGDSDGDDAANDENYALAA

SAHLS-PolyN (450 aa, 43,567 Da)
MGHHHHHHGSLQDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLME
AFAKRQGKEMDSLRFLYDGIRIQADQAPEDLDMEDNDIIEAHREQIGGGGSGSGGSGSG
GSGSQNEYRSGGGGSGSAGSGASGSSGSEGSGASGSAGSGSAGSRGSGASGSAGSGSAG
SGGAEAAKEAAKEAAKEAAKEAAKAGGSGSAGSAGSASSGSDGSGASGSAGSGSAGS
KGSGASGSAGSGSSGSSGGSGMGSSHHHHHHGSGLVPRGSASMSDSEVNQEAKPEVKP
EVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQADQT
PEDLDMEDNDIIEAHREQIGGGGSSGGSGGSGSSGDGGSSGGSGGSGSSGDGGSSGGSG
GDGSSGDGGSDGDSDGSDGDGDSDGDDAANDENYALAA

References.

1. Cherf, G.M. *et al.* Automated forward and reverse ratcheting of DNA in a nanopore at 5-Å precision. *Nat. Biotechnol.* **30,** 344–348 (2012).
2. Manrao, E.A. *et al.* Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nat. Biotechnol.* **30,** 349–353 (2012).
3. Lieberman, K.R. *et al.* Processive replication of single DNA molecules in a nanopore catalyzed by phi29 DNA polymerase. *J. Am. Chem. Soc.* **132,** 17961–17972 (2010).
4. Howorka, S. & Siwy, Z.S. Nanopores as protein sensors. *Nat. Biotechnol.* **30,** 506-507 (2012).
5. Lee, C. *et al.* ATP-Dependent Proteases Degrade Their Substrates by Processively Unraveling Them from the Degradation Signal. *Mol. Cell* **7,** 627–637 (2001).
6. Baker, T.A., & Sauer, R.T. ClpXP, an ATP-powered unfolding and protein-degradation machine. *Bioc. et Biop. Acta* **1823**, 15–28 (2012).
7. Neuwald, A.F., Aravind, L., Spouge, J.L. & Koonin, E.V. AAA+: a class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res.* **9,** 27–43 (1999).
8. Gottesman, S., Roche, E., Zhou, Y. & Sauer, R.T. The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes Dev.* **12,** 1338–1347 (1998).
9. Flynn, J.M. *et al.* Overlapping recognition determinants within the ssrA degradation tag allow modulation of proteolysis. *Proc. Natl Acad. Sci. USA* **98,** 10584–10589 (2001).
10. Maillard, R.A *et al*. ClpX(P) Generates Mechanical Force to Unfold and Translocate Its Protein Substrates. *Cell* **145,** 459–469 (2011).
11. Martin, A., Baker, T. A. & Sauer, R. T. Pore loops of the AAA+ ClpX machine grip substrates to drive translocation and unfolding. *Nature Struct. Mol. Biol.* **15,** 1147–1151 (2008).
12. Glynn, S.E., Martin, A., Nager, A.R., Baker, T.A., & Sauer, R.T. Crystal structures of asymmetric ClpX hexamers reveal nucleotide-dependent motions in a AAA+ protein-unfolding machine. *Cell* **139,** 744–756 (2009)
13. Yu, A.Y.H & Houry, W.A. ClpP: A distinctive family of cylindrical energy-dependent serine proteases. *FEBS Letters* **581,** 3749–3757 (2007)
14. Gribun, A. *et al.* The ClpP Double Ring Tetradecameric Protease Exhibits Plastic Ring-Ring Interactions, and the N Termini of Its Subunits Form Flexible Loops That Are Essential for ClpXP and ClpAP Complex Formation. *J. Bio. Chem.* **280,** 16185–16196, (2005).
15. Bewley, M.C., Graziano, V., Griffin, K. & Flanagan, J.M. The asymmetry in the mature amino-terminus of ClpP facilitates a local symmetry match in ClpAP and ClpXP complexes. *J. Struc. Bio*. **153,** 113–128 (2006).
16. Wang, J., Hartling, J.A., and Flanagan, J.M. The structure of ClpP at 2.3 Å resolution suggests a model for ATP-dependent proteolysis. *Cell* **91***,* 447–456 (1997).
17. Aubin-Tam, M-E. *et al.* Single-Molecule Protein Unfolding and Translocation by an ATP-Fueled Proteolytic Machine. *Cell* **145,** 257–267 (2011).
18. 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*. **15,** 139–145 (2008).

19. Rotem, D., Jayasinghe, L., Salichou, M. & Bayley, H. *J. Am. Chem. Soc.* **134,** 2781–2787 (2012).
20. Wei, R., Gatterdam, V., Wieneke, R., Tampe, R. & Rant, U. *Nat. Nanotechnol.* **7,** 257–263 (2012).
21. Han, A. *et al.* Sensing protein molecules using nanofabricated pores. *App. Pys. Lett.* **88,** 093901 (2006).
22. 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)
23. Talaga, D.S. & Li, J. Single-Molecule Protein Unfolding in Solid State Nanopores. *J. Am. Chem. Soc.* **131,** 9287–9297 (2009).
24. Butler, T.Z., Pavlenok, M., Derrington, I.M., Niederweis, M. & Gundlach, J.H. Single-molecule DNA detection with an engineered MspA protein nanopore. *Proc. Natl. Acad. Sci. USA* **105,** 20647–20652 (2008).
25. Martin, A., Baker, T.A. & Sauer, R.T. Rebuilt AAA+ motors reveal operating principles for ATP-fueled machines. *Nature* **437,** 1115–1120 (2005).
26. Branton, D. *et al.* The potential and challenges of nanopore sequencing. *Nat. Biotechnol.* **26,** 1146–1153 (2008).