Single-molecule peptide fingerprinting

Jetty van Ginkel[a,b], Mike Filius[a,b], Malwina Szczepaniak[a,b], Pawel Tulinski[a,b], Anne S. Meyer[a,b,1], and Chirlmin Joo (주철민)[a,b,1]

[a]Kavli Institute of Nanoscience, Delft University of Technology, 2629HZ Delft, The Netherlands; and [b]Department of Bionanoscience, Delft University of Technology, 2629HZ Delft, The Netherlands

Edited by Alan R. Fersht, Gonville and Caius College, Cambridge, United Kingdom, and approved February 21, 2018 (received for review May 1, 2017)

Proteomic analyses provide essential information on molecular pathways of cellular systems and the state of a living organism. Mass spectrometry is currently the first choice for proteomic analysis. However, the requirement for a large amount of sample renders a small-scale proteomics study challenging. Here, we demonstrate a proof of concept of single-molecule FRET-based protein fingerprinting. We harnessed the AAA+ protease ClpXP to scan peptides. By using donor fluorophore-labeled ClpP, we sequentially read out FRET signals from acceptor-labeled amino acids of peptides. The repurposed ClpXP exhibits unidirectional processing with high processivity and has the potential to detect low-abundance proteins. Our technique is a promising approach for sequencing protein substrates using a small amount of sample.

single-molecule FRET | ClpXP | protein analysis | peptides | single-molecule protein sequencing

Proteomic analyses provide essential information on molecular pathways of cellular systems and the state of a living organism (1). Therefore, for understanding of biological processes and their (dys)regulation, including disease, it is critical to monitor the protein composition of cells by sequencing (i.e., determination of the amino acid sequence). Mass spectrometry is currently the first choice for protein sequencing. However, mass spectrometry analysis often fails to recognize minor species embedded among other dominant species since sequence prediction is made through analysis of complex spectral peaks (2). As many cellular proteins exist in low abundance (3), it is difficult to obtain large-scale proteomic information. DNA sequencing presents similar challenges, but they are overcome by amplifying DNA samples until a high signal-to-noise ratio is achieved. This solution cannot be applied to protein analysis since there is no natural machinery that can amplify proteins.

Single-molecule techniques have the potential to provide radically new protein-sequencing tools that can quantify cellular proteins with accuracy as high as for mass spectrometry while requiring sample amounts as small as a single cell. However, despite several recent explorations (4–8), bona fide single-molecule protein sequencing has not yet been achieved due to the complexity that arises from primary protein sequences. Whereas DNA consists of only four building blocks (A, G, C, T), proteins are built from 20 distinct amino acids. Independent of the readout method of choice, full protein sequencing would require the detection of 20 distinguishable signals, which has so far not been demonstrated in single-molecule detection. Recently, our team and another have computationally demonstrated that readout of only a subset of the 20 building blocks is sufficient to identify proteins at the single-molecule level (9, 10). In brief, the number of protein species in an organism is finite and predictable. Through bioinformatics-based comparison with proteomics databases, ordered detection of only two types of amino acids can still allow for protein identification. For example, ordered detection of cysteine and lysine residues, which can be modified using orthogonal chemistries, is sufficient to sequence the human proteome (10). We named this approach "single-molecule protein fingerprinting" to distinguish it from full protein sequencing. Here, we demonstrate a proof of concept of a single-molecule fingerprinting technology that reads out fluorescently labeled amino acids of synthetic peptides and a model cellular protein.

To obtain ordered determination of fluorescently labeled amino acids, we needed a molecular probe that can scan a peptide in a processive manner. We adopted a naturally existing molecular machinery, the AAA+ protease ClpXP from *Escherichia coli*. The ClpXP protein complex is an enzymatic motor that unfolds and degrades protein substrates. ClpX monomers form a homohexameric ring ($ClpX_6$) that can exercise a large mechanical force to unfold proteins using ATP hydrolysis (11, 12). Through iterative rounds of force-generating power strokes, $ClpX_6$ translocates substrates through the center of its ring in a processive manner (13, 14), with extensive promiscuity toward unnatural substrate modifications, including fluorescent labels (15–17). Protein substrates are recognized by $ClpX_6$ when they display specific disordered sequences, such as the 11-aa C-terminal ssrA tag (18). $ClpX_6$ targets substrates for degradation by feeding them into $ClpP_{14}$, a homotetradecameric protease that contains 14 cleavage sites and self-assembles into a barrel-shaped complex that encloses a central chamber (19).

Results

Single-Molecule Fingerprinting Platform. To immobilize ClpXP ($ClpX_6P_{14}$) for single-molecule imaging, we biotinylated $ClpX_6$ and bound $ClpX_6P_{14}$ to a PEG-coated quartz surface through biotin–streptavidin conjugation (Fig. 1*A*). A combination of total internal reflection fluorescence microscopy (TIRF) and alternating laser excitation (ALEX) imaging (20, 21) was used to monitor individual ClpXP complexes bind, translocate, and degrade dye-labeled substrates in real time.

To detect the progression of fluorescently labeled amino acids through the $ClpX_6$ pore with nanometer-scale accuracy, we

Significance

Protein sequencing remains a challenge for small samples. A sensitive sequencing technology will create the opportunity for single-cell proteomics and real-time screening for on-site medical diagnostics. To resolve protein identity, we previously developed a computational algorithm that analyzes the ordered sequence of only two types of amino acids within a protein species. Through modification of a biological nanomachine, here, we developed single-molecule fluorescence technology to linearize protein molecules and to read signals from labeled amino acids in an ordered manner. This proof of concept of single-molecule fingerprinting will open the door to single-molecule protein sequencing and pave the road toward the development of a new, fast, and reliable diagnostic tool.

Author contributions: J.v.G., A.S.M., and C.J. designed research; J.v.G., M.F., M.S., and P.T. performed research; J.v.G. and M.F. analyzed data; and J.v.G., M.S., A.S.M., and C.J. wrote the paper.

Conflict of interest statement: J.v.G., C.J., and A.S.M. hold a patent ("Single molecule protein sequencing"; WO2014014347).

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: anne@annemeyerlab.org or c.joo@tudelft.nl.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1707207115/-/DCSupplemental.

Published online March 12, 2018.

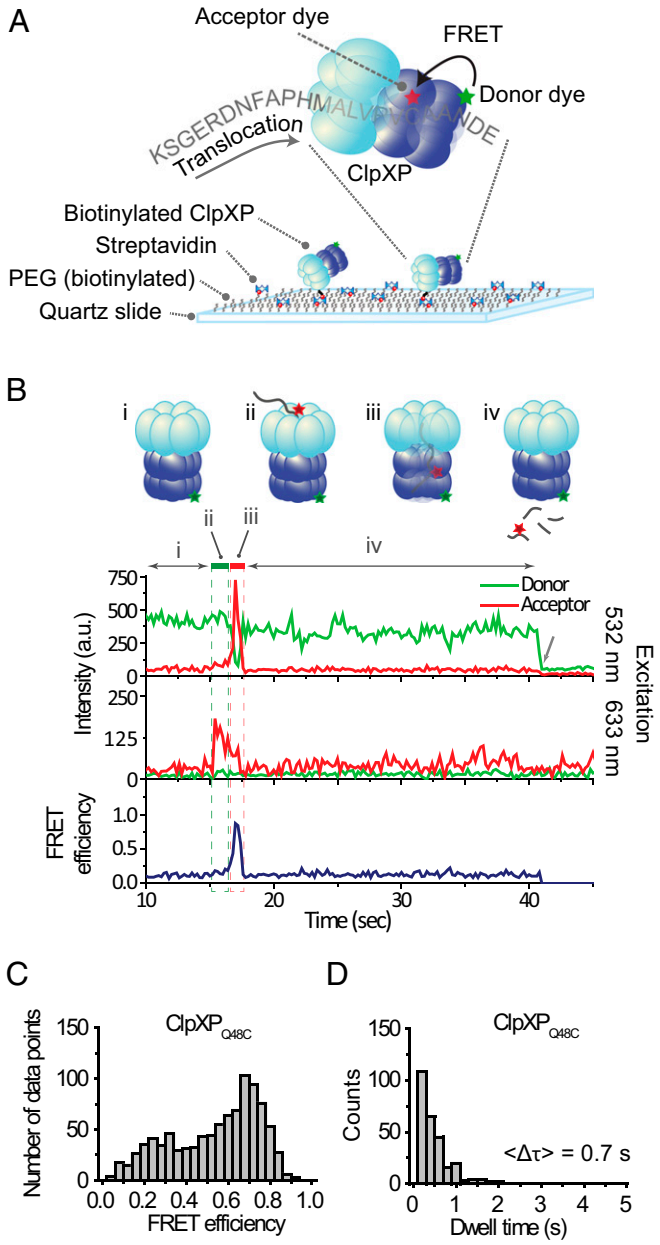

Fig. 1. Single-molecule observation of ClpXP translocation. (*A*) Schematics of the single-molecule fingerprinting platform. Donor-labeled ClpXP is immobilized on a PEG-coated slide via biotin–streptavidin conjugation. $ClpX_6$ recognizes an acceptor-labeled substrate (K-38-C_{Cy5}-ssrA, Table S1) and translocates it into the $ClpP_{14}$ chamber, upon which FRET occurs. (*B*) A typical fluorescence time trace. (*i*) The donor signal is from Cy3-labeled ClpXP (*Top* trace) upon green excitation (532 nm). (*ii*, green box) The sudden appearance of acceptor signal (time ∼16 s) during acceptor-direct excitation with red (633 nm) reports on binding of acceptor-labeled substrate to ClpXP (*Middle* trace). (*iii*, red box) The high FRET (time ∼17 s) reports on the presence of the substrate in $ClpP_{14}$ (*Top* and *Bottom* traces). (*iv*) Loss of fluorescence signal indicates the release of the substrate. The arrow at time ∼40 s indicates the photobleaching of Cy3. (*C*) FRET distribution of stage *iii* (n = 239). (*D*) Dwell time distribution of stage *iii* (n = 239).

employed FRET (Förster resonance energy transfer) (22–24). We used two different types of model substrates for fingerprinting— short synthetic peptides and a small protein (the titin I27 domain). These substrates were labeled with acceptor fluorophores and were also appended with the ssrA tag. We constructed a FRET scanner by adding a fluorophore (donor) to the $ClpP_{14}$ chamber (Fig. 1*A*). We introduced cysteines to the Q48 residue of ClpP ($ClpP_{Q48C}$), A139 ($ClpP_{A139C}$), or F31 ($ClpP_{F31C}$), labeled them with maleimide-functionalized fluorophores (Fig. S1*A*), and evaluated the suitability for FRET-based substrate detection. $ClpP_{Q48C}$ and $ClpP_{A139C}$ showed higher FRET than $ClpP_{F31C}$ (Fig. S1*B*). Among the first two, $ClpP_{Q48C}$ was chosen for our final scanner due to its higher efficiency of fluorophore labeling (*Methods*).

The donor fluorophores on $ClpP_{Q48C}$ are located near the center of the $ClpP_{14}$ chamber, which is ∼12 nm away from the substrate entry portal of $ClpX_6$ (Fig. S1*A*) (25, 26). This distance is longer than the Förster radius of a standard single-molecule FRET pair (∼5 nm). This physical separation enabled us to selectively detect signals from only the fluorophores (acceptors) on a protein substrate that had been translocated through a $ClpX_6$ central channel. We obtained FRET time traces reporting on translocation, as shown in Fig. 1*B*, by presenting a labeled peptide substrate to immobilized ClpXP complexes. The sudden appearance of acceptor signal during direct acceptor excitation indicated binding of acceptor-labeled peptide to ClpXP (Fig. 1*B*, *Middle* trace, stage *ii*). The subsequent appearance of a high FRET state indicated translocation of the substrate by $ClpX_6$ into the $ClpP_{14}$ chamber (Fig. 1*B*, stage *iii*). When a slowly hydrolyzable ATP analog (ATPγS) was used, the probability of high-FRET appearance decreased by one order of magnitude (Fig. S2*D*). Loss of FRET signal occurred upon the release of the dye-labeled peptide fragment (Fig. 1*B*, stage *iv*). When a cleavage inhibitor [diisopropyl fluorophosphate (DFP)] was used (27) (Fig. S2 *A* and *B*), the dwell time of high FRET increased 3.5-fold (Fig. S2*C*).

Our single-molecule fingerprinting concept requires detection of the order of fluorophores on a single substrate. To demonstrate fingerprinting, we functionalized a peptide with one type of fluorophore (Cy3) at the N-terminal site and a second type of fluorophore (Cy5) on an internal cysteine residue. We monitored the order in which the two fluorophores passed through Alexa488-labeled $ClpP_{14}$ (Fig. 2*A*). The positions of the Cy3 and Cy5 fluorophores relative to the ssrA tag on the substrate should dictate the order of Alexa488-Cy3 FRET and Alexa488-Cy5 FRET signals since an ssrA-tagged substrate was translocated through $ClpX_6$, starting from its C terminus. Fig. 2*B* depicts a representative time trace obtained from a substrate ($_{Cy3}NH_3$-17-C_{Cy5}-ssrA). The simultaneous appearance of Cy3 and Cy5 signals upon direct excitation with 532 nm and 637 nm (Fig. 2*B*, *Middle* and *Bottom*, t ∼ 40 s, indicated with arrows) indicates binding of a substrate containing both labels. In the FRET trace (Fig. 2*B*, *Top*), Alexa488-Cy5 FRET (marked * in the time trace) was observed before Alexa488-Cy3 FRET (marked °). This observation confirms that the ClpXP fingerprinter reads an ssrA-tagged substrate from the C-terminal to the N-terminal site.

We applied this fingerprinting scheme to the titin I27 domain. We labeled two Cys residues of titin (Cys64 and Cys80) with acceptor fluorophores. Because we did not have control over which dyes were attached to which Cys residues, we tagged both residues with the same dye, Cy5. Using Cy3 as a donor, we observed two separate FRET peaks within the time trajectories (Fig. S3*A*). The time interval between the two peaks was elongated when ATPγS was mixed with ATP (Fig. S3*B*), indicating that the two peaks represented the sequential probing of Cys80 and Cys64 residues.

We extracted the length of time that Cy3 and Cy5 acceptor fluorophores were engaged with ClpXP ($\Delta\tau_{Cy3}$, $\Delta\tau_{Cy5}$). We observed positive differences in dwell time ($\Delta\tau_{Cy3\text{-}Cy5} = \Delta\tau_{Cy3} - \Delta\tau_{Cy5}$, $\langle\Delta\tau_{Cy3\text{-}Cy5}\rangle = 3.5$ s) for a substrate with N-terminal Cy3 labeling and internal Cy5 labeling (Fig. 2*C*, gray, $_{Cy3}NH_3$-17-C_{Cy5}-ssrA). For a substrate with exchanged dye positions ($_{Cy5}NH_3$-17-C_{Cy3}-ssrA), we observed negative differences (Fig. 2*C*, white, $\langle\Delta\tau_{Cy3\text{-}Cy5}\rangle = -3.9$ s). Thus, dye-labeled amino acids located

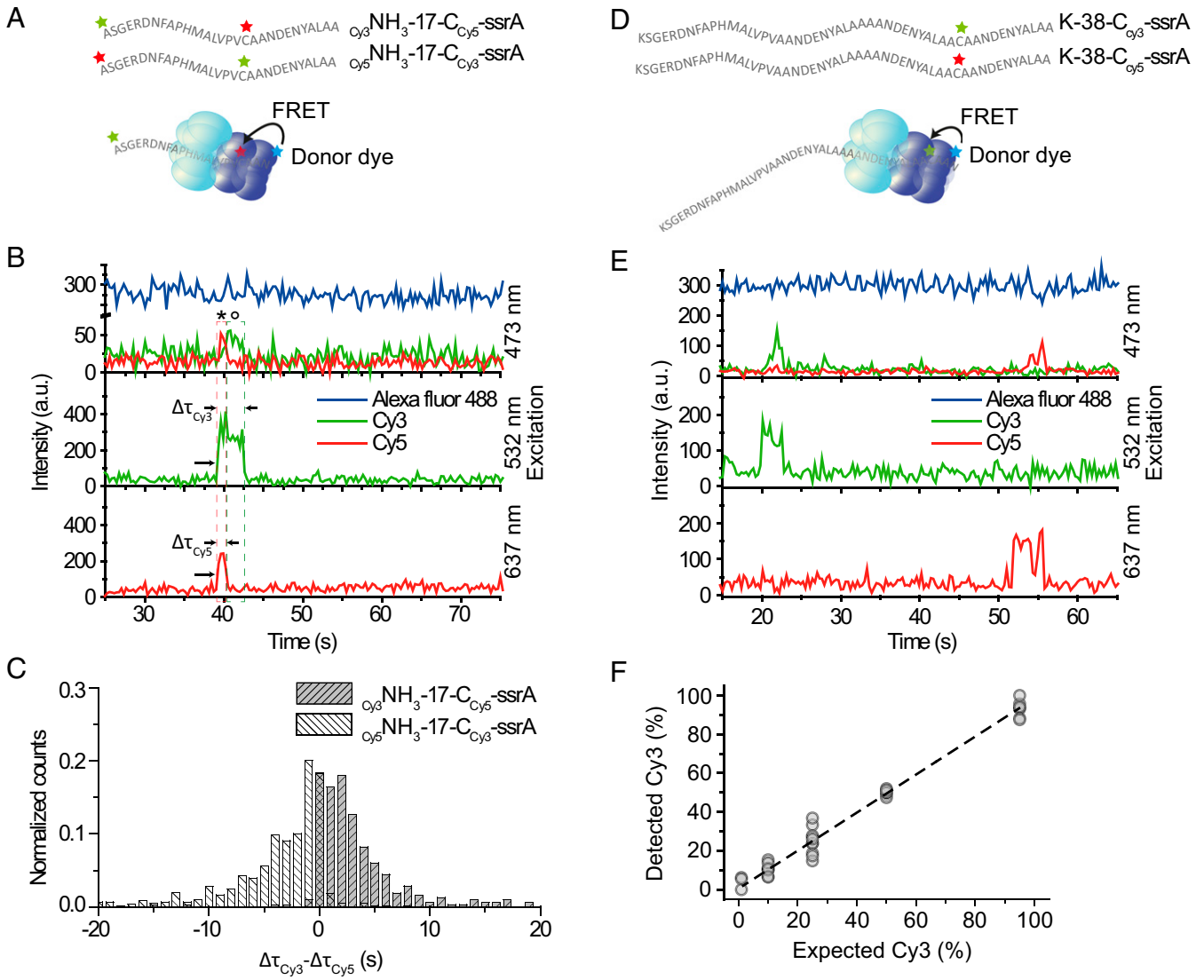

Fig. 2. Single-molecule fingerprinting. (*A*) Substrates with two acceptor dyes were labeled at the N-terminal end and on cysteine residues (Table S1). The green star stands for Cy3, red for Cy5, and blue for Alexa488. (*B*) A typical time trace for $_{Cy3}$-NH$_3$-17-C$_{Cy5}$-ssrA from three-color ALEX (*Top*) showed FRET between Alexa488 and Cy3; and Alexa488 and Cy5 upon excitation with blue laser light (473 nm). Concurrent signals from Cy3 (*Middle*) and Cy5 (*Bottom*) upon direct excitation, respectively, with green (532 nm) and red (637 nm). For clarity, an arbitrary offset of 200 a.u. was applied to the Alexa488 trace, and the sum of the Cy3 and Cy5 signals was plotted (*Middle*). For the original trace, see Fig. S1*E*. The symbol * indicate Alexa488-Cy5 FRET. The symbol ° indicates Alexa488-Cy3 FRET. The arrows indicate the start and end of each FRET. (*C*) Comparison of acceptor dwell times. The dwell time of Cy5 ($\Delta\tau_{Cy5}$) was subtracted from that of Cy3 ($\Delta\tau_{Cy3}$) for each event from $_{Cy3}$-NH$_3$-17-C$_{Cy5}$-ssrA (gray) ($n = 318$). The mean of the distribution is 3.5 ± 4.90 (s). In white is the same analysis for $_{Cy5}$-NH$_3$-17-C$_{Cy3}$-ssrA ($n = 610$). The mean of the distribution is -3.9 ± 5.67 (s). (*D*) Substrates labeled with one acceptor dye. Cysteine residues of the substrates (K-38-C-ssrA) were labeled with either Cy3 or Cy5. The green star stands for Cy3, red for Cy5, and blue for Alexa488. (*E*) A time trace from substrates in *D*. At time ~20 s, a Cy3-labeled substrate binds (*Top*, Alexa488-Cy3 FRET) and 532-nm direct excitation (*Middle*). At time ~50 s, a Cy5-labeled substrate binds (*Top*, Alexa488-Cy5 FRET) and 637-nm direct excitation (*Bottom*). An offset of 200 a.u. applied to Alexa488. (*F*) The percentage of processed Cy3-labeled substrates was plotted against the expected percentage of Cy3-labeled substrates. The line is a linear fit (slope of 0.98 ± 0.02, intercept 0.58 ± 0.79, $R^2 = 0.99$). Data points are from 100-s recordings, repeated 10 times per condition (each $n = 24.4 \pm 1.50$, in total $n = 1{,}194$).

closer to the C-terminal ssrA tag were retained in the ClpXP complex for shorter amounts of time than labeled amino acids located more closely to the N terminus. We can conclude that our fingerprinter can detect dyes in an order matching the amino acid sequence. The ordered disappearance of the Cy3 and Cy5 signals further implies that uncleaved or partially cleaved substrate does not accumulate within the ClpP$_{14}$ chamber, which would otherwise hamper accurate fingerprinting.

Performance of the Single-Molecule Fingerprinter. A single-molecule fingerprinter should perform without any bias to fluorophores and with high dynamic range. To determine the sensitivity of our fingerprinter, we performed a population study in which ClpP$_{14}$ was labeled with donor fluorophore (Alexa488) and substrate peptides were singly labeled with either Cy3 or Cy5 as an acceptor fluorophore (Fig. 2 *D* and *E*). We mixed Cy3- and Cy5-labeled substrates in varying proportions (1:99, 10:90, 25:75, 50:50, 95:5) and quantified the number of translocation events. We observed a linear relationship between the percentage of Cy3-labeled substrates we detected versus the expectation, with an offset of $0.58 \pm 0.79\%$ and a slope of 0.98 ± 0.02 (adjusted $R^2 = 0.99$) (Fig. 2*F*). We conclude that both FRET pairs are detected with equal sensitivity and that our FRET scanner has the potential to detect low-abundance proteins.

Our previous computational analysis indicated that the precision of our fingerprinting method would be enhanced if the distance between labeled cysteine and lysine residues could additionally be determined, as well as their order (10). A uniform speed of the scanner, represented by $ClpX_6$, is crucial to extract distance information. To determine whether the processing time of ClpXP is proportional to the length of protein substrates, we determined the processing times (the dwell time of fluorescence signals emitted by Cy5 labels on substrates, upon direct excitation) for three peptides [29, 40, and 51 amino acids (aa) in length] (Table S1) and monomeric (119-aa) and dimeric (210-aa) versions of titin (all labeled at Cys) (Table S1). Plotting the total time ($\Delta\tau$) (Fig. 3*A*) that a substrate was bound and processed by ClpXP versus the length of the substrates showed a linear increase, with an average processing speed of 23.9 aa per second (Fig. 3*B* and Fig. S4), which agrees with previous results obtained from both bulk (28) and single-molecule assays (11, 12, 29). We obtained a similar processing speed of 14.5 aa per second (translocation of 16 aa from labeled Cys64 to labeled Cys80 for 1.1 s) from the doubly labeled titin (Fig. S3*B*). In Fig. 3*B*, the *y* axis offset of 4.2 s reports on the initial docking phase and the eventual retention within $ClpP_{14}$. Our data indicate that the ClpXP fingerprinter has the potential to determine both the order and spacing distance of labeled residues.

Unidirectional translocation is also of utmost importance for our technology. Backtracking of ClpXP would result in insertion errors in the observed fingerprint and thus reduce the detection precision. To evaluate the occurrence of backtracking, we

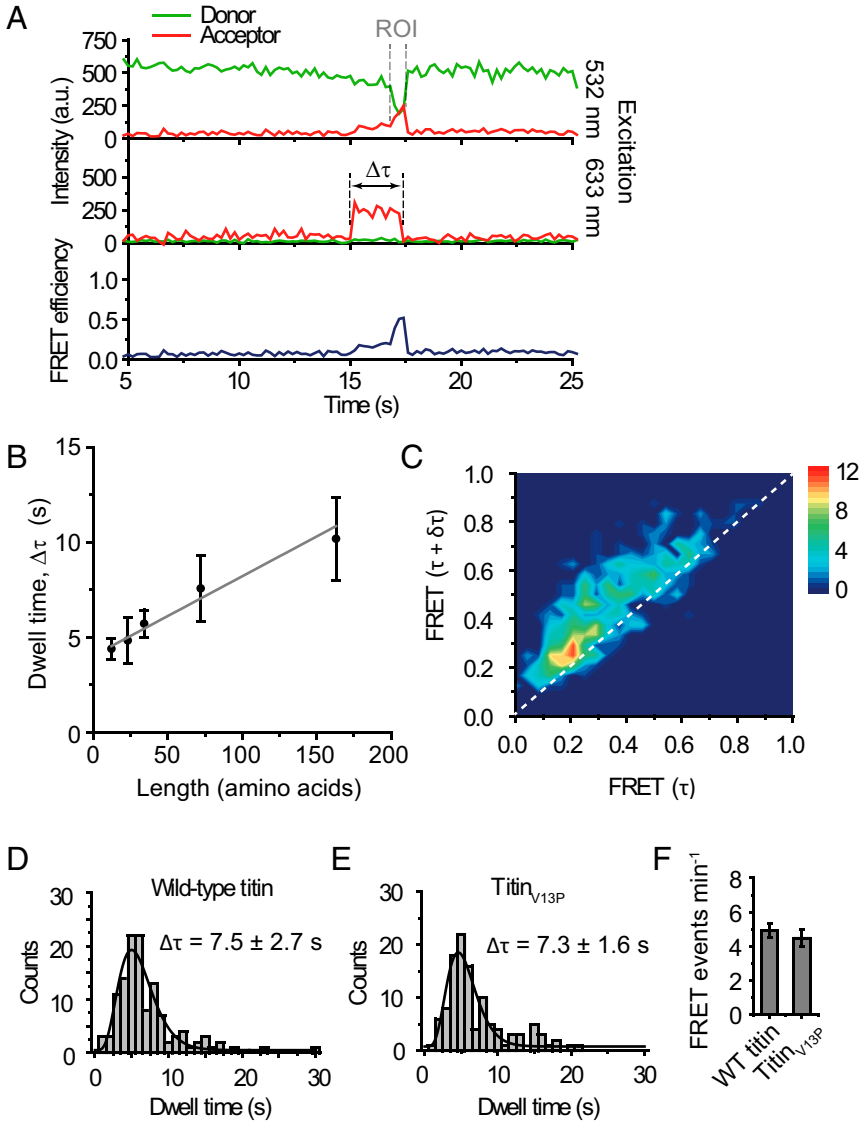

Fig. 3. ClpXP performs unidirectional scanning with a constant speed. (*A*) Representative time trace. ROI (region of Interest) is where the FRET efficiency gradually increases. $\Delta\tau$: the total docking time. (*B*) Total dwell time vs. substrate length. The average time, $\langle\Delta\tau\rangle$, was obtained by fitting data in Fig. S4 with a gamma distribution. Five different substrates were used: K-16-C-ssrA ($n = 227$), K-16-C-11-ssrA ($n = 131$), K-16-C-22-ssrA ($n = 290$), titin monomer ($n = 85$), and titin dimer ($n = 81$). The substrate length is the number of amino acids between the C terminus and a dye the most proximal to the N terminus. Error bars obtained by bootstrapping with 1,000 resamples. A linear fit results in an offset of 4.0 ± 0.20 s and a speed of 23.9 ± 2.86 aa per second. (*C*) Transition density plot. FRET change was analyzed by measuring $FRET_{t=\tau}$ and $FRET_{t=\tau+\delta\tau}$, with $\delta\tau = 0.4$ s, for every point in ROI. The dotted line represents $FRET_{\delta\tau} = FRET_{\tau+\delta\tau}$. K-38-C-ssrA was used ($n = 44$). (*D* and *E*) Total dwell times ($\Delta\tau$) for WT titin ($n = 124$) and $titin_{V13P}$ ($n = 114$). $\Delta\tau = 7.5 \pm 2.7$ s and 7.3 ± 1.6 s were obtained, respectively, by fitting with a gamma distribution. Errors obtained by bootstrapping with 1,000 resamples. WT titin, $n = 123$. $Titin_{V13P}$, $n = 112$. (*F*) The number of traces showing FRET events for WT titin and $titin_{V13P}$. Error bars are SDs from 15 measurements.

determined the change in FRET over time during processing of peptide substrates. We created a 2D heat map by plotting the change of FRET over a given time interval $\delta\tau$. In Fig. 3*C*, FRET ($t = \tau + \delta\tau$) versus FRET ($t = \tau$) is deposited for every time point along a time trace reporting on translocation [Fig. 3*A*, region of interest (ROI)]. We set $\delta\tau = 0.4$ s, a time scale longer than our time resolution (0.2 s) but shorter than the average translocation time (0.7 s, Fig. 1*D*), to visualize the gradual increase of FRET. Any backtracking of $ClpX_6$ along the substrate would result in momentary FRET decrease during translocation, which would appear as FRET ($t = \tau + \delta\tau$) values lower than FRET ($t = \tau$) (population below the diagonal). We observed FRET ($t = \tau + \delta\tau$) higher than FRET ($t = \tau$) (upper diagonal population) for a major fraction (92.5%) of the data points. The remaining fraction is likely due to the backtracking of ClpX, the statistical noise of the fluorescence signals, and the photoblinking of acceptor dyes. This degree of experimental error is predicted not to interfere with the ability to extract length information according to our computational simulation (10).

A single-molecule protein fingerprinter should be able to process any structural element of a protein. Single-molecule force spectroscopy studies of ClpXP have shown that $ClpX_6$ stalls on substrates with rigid secondary structures (30, 31), which would inhibit the extraction of sequence information. We therefore explored the possibility of disrupting such tightly folded structures to enable fingerprinting. Perturbation of cysteine residues in the titin protein has been shown to interfere with the secondary structure of the protein, making it behave as an unstructured polypeptide chain (32, 33). We purified the I27 domain of both WT titin, known to make $ClpX_6$ stall (30), and $titin_{V13P}$, a variant that is still folded but is degraded at a rate close to denatured titin (33). By fluorophore labeling the cysteine residues of WT titin and $titin_{V13P}$, we sought to determine the degree of structural influence of the cysteine-dye conjugation on ClpXP processing. We obtained equivalent total dwell times for processing stable WT titin ($\Delta\tau = 7.5 \pm 2.7$ s, Fig. 3*D*) and $titin_{V13P}$ ($\Delta\tau = 7.3 \pm 1.6$ s, Fig. 3*E*). A similar number of both substrates were processed by ClpXP within our time interval of observation (Fig. 3*F*), indicating that $ClpX_6$ can process labeled WT and V13P substrates with the same efficiency. These results suggest that preparing substrates for sequencing by labeling cysteine residues (and likely lysine residues as well) might sufficiently destabilize their protein structures. This will allow for fingerprinting of any protein regardless of structural stability.

Discussion

We have demonstrated a FRET-based detection platform utilizing an AAA+ protease as a scanner of peptides and proteins. In our approach, we conjugate fluorophores to thiol groups of cysteine residues and amine groups of the N-terminal site (which can be extended to lysine residues) because these chemical groups can be labeled with high efficiency and specificity. Our platform, however, is not limited to these two modifications. With appropriate chemistry, one could target other residues or even posttranslational modifications. Detection of these moieties could be implemented by extending our current three-color FRET scheme to four-color FRET (21).

Several outstanding perspectives remain in order for our method to be directly applied to a protein sequencing technology. First, for proteomics analysis, our sequencing technique has to work for all cellular proteins without sequence bias. ClpX, a core of our platform, only recognizes substrates displaying specific sequence tags including ssrA. The substrate selectivity and specificity of ClpX would need to be broadened equally for every substrate, perhaps through targeted mutations in the substrate-recognition loops of the ClpX channel, or the use of engineered adaptor proteins (e.g., modified SspB) that nonspecifically deliver substrates to ClpX. Second, a challenge of cellular protein analysis is to detect low-abundance proteins within a complex sample, such as a clinical tissue sample. The depth of the sequencing coverage might be increased by removing housekeeping proteins chromatographically (34). Third, to cover the whole proteome in a reasonable amount of time, the throughput should be enhanced. Under the standard conditions used in this work (10 nM substrate, 512 × 512 pixel camera, $ClpX_6$), we obtained ~10 productive reactions per minute per imaging area. By using a complementary metal-oxide semiconductor (CMOS) camera that has a larger number of pixels (e.g., 2,084 × 2,084 pixels) (35), as well as a zero-mode waveguide platform that allows for single-molecule imaging of a higher concentration of substrate (e.g., 1 μM) (36), the throughput would be improved by a factor of ~1,000. We also observed that productive reactions (a trace ending with high FRET) make up only ~10% of the total population (Fig. S2*D*). We suspect that this low yield is due to the lack of adaptor proteins that facilitate binding, such as sspB, in our experiments. By adding adaptor proteins and replacing the hexameric linked ClpXΔN with WT, monomeric ClpX, which contains the N-terminal domain essential for binding adaptor proteins, we expect that the percentage of the productive population will reach near 100%. When our sequencer is improved with these changes, we expect to cover 1× of a single human-cell proteome (~10^8 proteins) in nearly 10 h (10 events per minute × 16 × 100 × 10 ≈ 10^7 events per hour). In practice, sequencing errors will occur, and 1× coverage might not be sufficient to accurately determine a protein fingerprint. However, cellular protein copy numbers typically range from 10^0 to 10^7, resulting in an effective coverage of 10^0–10^7× per protein species. Reliable coverage of the lowest abundance protein species would require increased multiplexing beyond 1×, which could be achieved through further equipment improvements.

Our method has the potential to scan full-length proteins from end to end without the need for fragmentation. Sequencing substrates are processed at a constant speed, allowing for accurate protein identification (10). In this proof-of-concept study, we show our capability to detect low-frequency subpopulations of differentially labeled substrates, as well as our capacity to detect distinct acceptor fluorophores on a single substrate in a sequential manner. The platform we present here has the capability to transform proteomics from a basic research tool into an invaluable asset to clinical diagnostics.

Methods

Constructs used in this study are listed in Table S1. Additional methods may be found in *SI Methods*.

Single-Molecule Sample Preparation. To reduce the nonspecific binding of proteins, acidic piranha-etched quartz slides (G. Finkenbeiner) were passivated with two rounds of polyethylene glycol [mPEG-Succinimidyl Valerate, MW 5000, Laysan, followed by $MS(PEG)_4$, Piercenet] as described previously (37). After assembly of a microfluidic flow chamber, slides were incubated with 5% Tween 20 for 10 min (38), and excess Tween 20 was washed with T50 buffer (10 mM Tris·HCl, pH 8.0, 50 mM NaCl), followed by 1-min incubation with streptavidin (0.1 mg/mL; Sigma). Unbound streptavidin was washed with 100 μL of T50 buffer, followed by 100 μL of PD buffer (25 mM Hepes, pH 8.0, 5 mM $MgCl_2$, 40 mM KCl, 0.148% Nonidet P-40, 10% glycerol). A $ClpX_6$:$ClpP_{14}$ = 1:3 molar ratio was used to ensure ClpXP complex formation with a 1:1 molar ratio (39). Then, 30 nM $ClpX_6$ and 90 nM $ClpP_{14}$ (either WT or mutant) were preincubated for 2 min at room temperature in the presence of 10 mM ATP in PD buffer. After preincubation, the sample was diluted 10 times in PD buffer to reach an expected final ClpXP complex concentration of 3 nM. The diluted sample was applied to the flow chamber and incubated for 1 min. Unbound ClpXP complexes were washed with 100 μL of PD buffer containing 1 mM ATP. Then, 10 to 20 nM acceptor-labeled substrate was introduced to the flow chamber in the presence of an imaging buffer (0.8% dextrose [Sigma], 1 mg/mL glucose oxidase [Sigma], 170 mg/mL catalase [Merck], and 1 mM Trolox [(±)-6-hydroxy-2,5,7,8-tetramethylchromane-2-carboxylic acid, 238813] [Sigma]). Donor-labeled $ClpP_{14}$ added into a chamber without $ClpX_6$ led to very few nonspecifically immobilized ClpP protein

complexes, ruling out any nonspecific adsorption of $ClpP_{14}$ to the surface (Fig. S1*B*). All experiments were performed at room temperature (23 ± 2 °C).

Single-Molecule Fluorescence. Single-molecule fluorescence measurements were performed with a prism-type total internal reflection fluorescence microscope. For two-color measurements, Cy3 molecules were excited using a 532-nm laser (Compass 215M-50; Coherent), and Cy5 molecules were excited using a 633-nm laser (25 LHP 928; CVI Melles Griot). Fluorescence signals from single molecules were collected through a 60× water immersion objective (UplanSApo; Olympus) with an inverted microscope (IX71; Olympus). Scattered light from the 532-nm and 633-nm laser beams was blocked by a triple-notch filter (NF01-488/532/635; Semrock). The Cy3 and Cy5 signals were separated with a dichroic mirror (635 dcxr; Chroma) and imaged using an EM-CCD camera (Andor iXon 897 Classic; Andor Technology).

For three-color measurements, Alexa488 molecules were excited using a 473-nm laser (OBIS 473 nm LX, 75 mW; Coherent), Cy3 molecules were excited using a 532-nm laser (Sapphire 532 nm, 100 mW; Coherent), and Cy5 molecules were excited using a 637-nm laser (OBIS 637 nm LX, 140 mW; Coherent). Fluorescence signals from single molecules were collected through a 60× water immersion objective (UplanSApo; Olympus) with an inverted microscope (IX73; Olympus). The 473-nm laser beam was blocked by a 473-nm long pass filter (BLP01-473R-25; Semrock), the 532 nm laser beam was blocked by a 532-nm notch filter (NF03-532E-25; Semrock), and the 637-nm laser beam was blocked by a 633-nm notch filter (NF03-633E-25; Semrock). The Alexa488, Cy3, and Cy5 signals were separated by dichroic mirrors (540dcxr and 635 dcxr; Chroma) and imaged using an EM-CCD camera (Andor iXon 897 Classic; Andor Technology).

Data Acquisition. Samples were excited alternatingly with different colors and using a custom-made program written in Visual C++ (Microsoft). A series of CCD images with an exposure time of 0.1 s were recorded. The time traces were extracted from the CCD image series using an IDL (ITT Visual Information Solution) algorithm that identifies fluorescence spots with a defined Gaussian profile and with signals above the average of the background signals. Colocalization between Alexa488, Cy3, and Cy5 signals was carried out with a custom-made mapping algorithm written in IDL. The extracted time traces were processed using Matlab (MathWorks) and Origin (Origin Lab).

ACKNOWLEDGMENTS. We thank Marek Noga, Ivo Severins, and Anna Haagsma for technical support; Stanley Chandradoss and Margreet Docter for building the multicolor optical setup; Luuk Loeff and Misha Klein for assisting with data analysis; and Tim Blosser, Laura Restrepo, Margreet Docter, Stephanie Heerema, Noortje de Haan, and Viktorija Globyte for critical reading of the manuscript. $ClpX_6$ expression plasmids were a generous gift from Andreas Martin. ClpP expression plasmids were a generous gift from Tania Baker and Robert Sauer. Titin expression plasmids were a generous gift from Victor Muñoz and Jörg Schönfelder. C.J. and A.S.M. were funded by the Foundation for Fundamental Research on Matter (Grants Projectruimte 12PR3029 and Vrije Programma SMPS).

1. Harper JW, Bennett EJ (2016) Proteome complexity and the forces that drive proteome imbalance. *Nature* 537:328–338.
2. Zubarev RA (2013) The challenge of the proteome dynamic range and its implications for in-depth proteomics. *Proteomics* 13:723–726.
3. Talapatra A, Rouse R, Hardiman G (2002) Protein microarrays: Challenges and promises. *Pharmacogenomics* 3:527–536.
4. Nivala J, Marks DB, Akeson M (2013) Unfoldase-mediated protein translocation through an α-hemolysin nanopore. *Nat Biotechnol* 31:247–250.
5. Rosen CB, Rodriguez-Larrea D, Bayley H (2014) Single-molecule site-specific detection of protein phosphorylation with a nanopore. *Nat Biotechnol* 32:179–181.
6. Zhao Y, et al. (2014) Single-molecule spectroscopy of amino acids and peptides by recognition tunnelling. *Nat Nanotechnol* 9:466–473.
7. Ohshiro T, et al. (2014) Detection of post-translational modifications in single peptides using electron tunnelling currents. *Nat Nanotechnol* 9:835–840.
8. Kennedy E, Dong Z, Tennant C, Timp G (2016) Reading the primary structure of a protein with 0.07 nm^3 resolution using a subnanometre-diameter pore. *Nat Nanotechnol* 11:968–976.
9. Swaminathan J, Boulgakov AA, Marcotte EM (2015) A theoretical justification for single molecule peptide sequencing. *PLoS Comput Biol* 11:e1004080.
10. Yao Y, Docter M, van Ginkel J, de Ridder D, Joo C (2015) Single-molecule protein sequencing through fingerprinting: Computational assessment. *Phys Biol* 12:055003.
11. Aubin-Tam M-E, Olivares AO, Sauer RT, Baker TA, Lang MJ (2011) Single-molecule protein unfolding and translocation by an ATP-fueled proteolytic machine. *Cell* 145:257–267.
12. Maillard RA, et al. (2011) ClpX(P) generates mechanical force to unfold and translocate its protein substrates. *Cell* 145:459–469.
13. Sen M, et al. (2013) The ClpXP protease unfolds substrates using a constant rate of pulling but different gears. *Cell* 155:636–646.
14. Thompson MW, Singh SK, Maurizi MR (1994) Processive degradation of proteins by the ATP-dependent Clp protease from Escherichia coli. Requirement for the multiple array of active sites in ClpP but not ATP hydrolysis. *J Biol Chem* 269:18209–18215.
15. Barkow SR, Levchenko I, Baker TA, Sauer RT (2009) Polypeptide translocation by the AAA+ ClpXP protease machine. *Chem Biol* 16:605–612.
16. Burton RE, Siddiqui SM, Kim YI, Baker TA, Sauer RT (2001) Effects of protein stability and structure on substrate processing by the ClpXP unfolding and degradation machine. *EMBO J* 20:3092–3100.
17. Kolygo K, et al. (2009) Studying chaperone-proteases using a real-time approach based on FRET. *J Struct Biol* 168:267–277.
18. Keiler KC, Waller PR, Sauer RT (1996) Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. *Science* 271:990–993.
19. Baker TA, Sauer RT (2012) ClpXP, an ATP-powered unfolding and protein-degradation machine. *Biochim Biophys Acta* 1823:15–28.
20. Kapanidis AN, et al. (2004) Fluorescence-aided molecule sorting: Analysis of structure and interactions by alternating-laser excitation of single molecules. *Proc Natl Acad Sci USA* 101:8936–8941.
21. Lee J, et al. (2010) Single-molecule four-color FRET. *Angew Chem Int Ed Engl* 49:9922–9925.
22. van Oijen AM (2011) Single-molecule approaches to characterizing kinetics of biomolecular interactions. *Curr Opin Biotechnol* 22:75–80.
23. Lamichhane R, Solem A, Black W, Rueda D (2010) Single-molecule FRET of protein-nucleic acid and protein-protein complexes: Surface passivation and immobilization. *Methods* 52:192–200.
24. Bae W, Choi MG, Hyeon C, Shin YK, Yoon TY (2013) Real-time observation of multiple-protein complex formation with single-molecule FRET. *J Am Chem Soc* 135:10254–10257.
25. Flanagan JM, Wall JS, Capel MS, Schneider DK, Shanklin J (1995) Scanning transmission electron microscopy and small-angle scattering provide evidence that native Escherichia coli ClpP is a tetradecamer with an axial pore. *Biochemistry* 34:10910–10917.
26. Kim DY, Kim KK (2003) Crystal structure of ClpX molecular chaperone from Helicobacter pylori. *J Biol Chem* 278:50664–50670.
27. Maurizi MR, Clark WP, Kim SH, Gottesman S (1990) Clp P represents a unique family of serine proteases. *J Biol Chem* 265:12546–12552.
28. Martin A, Baker TA, Sauer RT (2008) Protein unfolding by a AAA+ protease is dependent on ATP-hydrolysis rates and substrate energy landscapes. *Nat Struct Mol Biol* 15:139–145.
29. Shin Y, et al. (2009) Single-molecule denaturation and degradation of proteins by the AAA+ ClpXP protease. *Proc Natl Acad Sci USA* 106:19340–19345.
30. Cordova JC, et al. (2014) Stochastic but highly coordinated protein unfolding and translocation by the ClpXP proteolytic machine. *Cell* 158:647–658.
31. Nivala J, Mulroney L, Li G, Schreiber J, Akeson M (2014) Discrimination among protein variants using an unfoldase-coupled nanopore. *ACS Nano* 8:12365–12375.
32. Iosefson O, Nager AR, Baker TA, Sauer RT (2015) Coordinated gripping of substrate by subunits of a AAA+ proteolytic machine. *Nat Chem Biol* 11:201–206.
33. Kenniston JA, Baker TA, Fernandez JM, Sauer RT (2003) Linkage between ATP consumption and mechanical unfolding during the protein processing reactions of an AAA+ degradation machine. *Cell* 114:511–520.
34. Han X, Aslanian A, Yates JR, 3rd (2008) Mass spectrometry for proteomics. *Curr Opin Chem Biol* 12:483–490.
35. Juette MF, et al. (2016) Single-molecule imaging of non-equilibrium molecular ensembles on the millisecond timescale. *Nat Methods* 13:341–344.
36. Levene MJ, et al. (2003) Zero-mode waveguides for single-molecule analysis at high concentrations. *Science* 299:682–686.
37. Chandradoss SD, et al. (2014) Surface passivation for single-molecule protein studies. *J Vis Exp* e50549.
38. Pan H, Xia Y, Qin M, Cao Y, Wang W (2015) A simple procedure to improve the surface passivation for single molecule fluorescence studies. *Phys Biol* 12:045006.
39. Singh SK, et al. (2001) Functional domains of the ClpA and ClpX molecular chaperones identified by limited proteolysis and deletion analysis. *J Biol Chem* 276:29420–29429.