Full-length single protein molecules tracking and counting in thin silicon channels

Shilo Ohayon[1,&], Liran Taib[1,&,], Navneet C. Verma[1,&], Marzia Iarossi[1,&], Ivy Bhattacharya[1], Barak Marom[1], Diana Huttner[1] and Amit Meller[1,2]*

[1]Department of Biomedical Engineering, Technion -IIT, Haifa, Israel, postcode:3200003
[2]Russell Berrie Nanotechnology Institute, Technion -IIT, Haifa, Israel, postcode:3200003

[&]These authors contributed equally
*Corresponding author, ameller@technion.ac.il

Emerging single-molecule protein sensing techniques are ushering in a transformative era in biomedical research. Nevertheless, challenges persist in realizing ultra-fast full-length protein sensing, including loss of molecular integrity due to protein fragmentation, biases introduced by antibodies affinity, identification of proteoforms and low throughputs. Here, we introduce a single-molecule method for parallel protein separation and tracking, yielding multi-dimensional molecular properties used for their identification. Proteins are tagged by chemo-selective dual amino-acid specific labels and are electrophoretically separated by their mass/charge in custom-designed thin silicon channel with subwavelength height. This approach allows us to analyse thousands of individual proteins within a few minutes by tracking their motion during the migration. We demonstrate the power of the method by quantifying a cytokine panel for host-response discrimination between viral and bacterial infections. Moreover, we show that two clinically-relevant splice isoforms of VEGF can be accurately quantified from human serum samples. Being non-destructive and compatible with full-length intact proteins, this method opens up ways for antibody-free single-protein molecule quantification.

1. Introduction

Next-generation and single-molecule DNA sequencing technologies have sharply reduced the cost and time of high-throughput sequencing, consequently transforming our ability to perform whole genome studies and to quantify mRNA expression levels down to single-cells. To date, however, the ability to directly quantify proteins' copy numbers and proteoforms has remained an outstanding challenge.[1–3] This has largely stemmed from the fact that unlike DNAs, proteins cannot be amplified enzymatically and are natively folded into highly complex structures. Furthermore, proteins often appear in several different proteoforms, including sequence variations, splice variants and post-translational modifications (PTMs).[4] This diversity of proteoforms is exceptionally significant for our understanding of the molecular basis for cellular functions and is clinically relevant. Presently, most protein identification methods rely on fingerprinting short proteolytically-produced peptides instead of whole proteins or rely on using pairs of antibodies for immuno-assays protein sensing. Despite significant progress,[5–9] these approaches are fundamentally limited. For example, peptides do not always uniquely fingerprint specific proteins or provide accurate quantifications of proteoforms.[4] Immunosorbent assays can reach single-molecule sensitivity but falls short of high-multiplexing, unbiased quantification of many different target proteins in a single measurment.[10]

To address these challenges, a wave of single-molecule sensing technologies has recently emerged. [1,5,11] Particularly, single-molecule techniques such as fluorosequencing by Edman degradation, Fluorescence Resonance Energy Transfer (FRET), and nanopores have already shown noticeable progress toward achieving single-molecule protein analysis.[12–20] At the same time, most of these

emerging methods are in their infancy and significant challenges have remained unsolved. Specifically, the ability to analyse full-length proteins with single-molecule resolution has remained an outstanding goal.[21–23] Separation of proteins by their mass or mass to charge ratio can greatly facilitate their identification, particularly if this can be combined with *amino-acid specific chemistries* that provide protein-specific information with single-molecule resolution.[24]

Hereby, we introduce a general method for single protein molecule separation, tracking, identification, and quantification based on multi-dimensional molecular feature extraction in subwavelength deep solid-state channels with an ultra-high length/height aspect ratio. The planar geometry of the channels is designed to load and deliver the proteins towards the imaging area by electrophoresis, whilst the channel depth is optimized to image their motion over time within a high-resolution focal area. To identify full-length proteins, we apply dual-colour biochemical conjugation of amino acid residues combined with single protein molecule separation and sensing. We demonstrate the ability to

2

dynamically track thousands of single protein molecules, leading to an extremely fast, quantitative discrimination among protein species.

2. Ultra-high aspect ratio solid-state fluidic device for single protein molecule sensing.
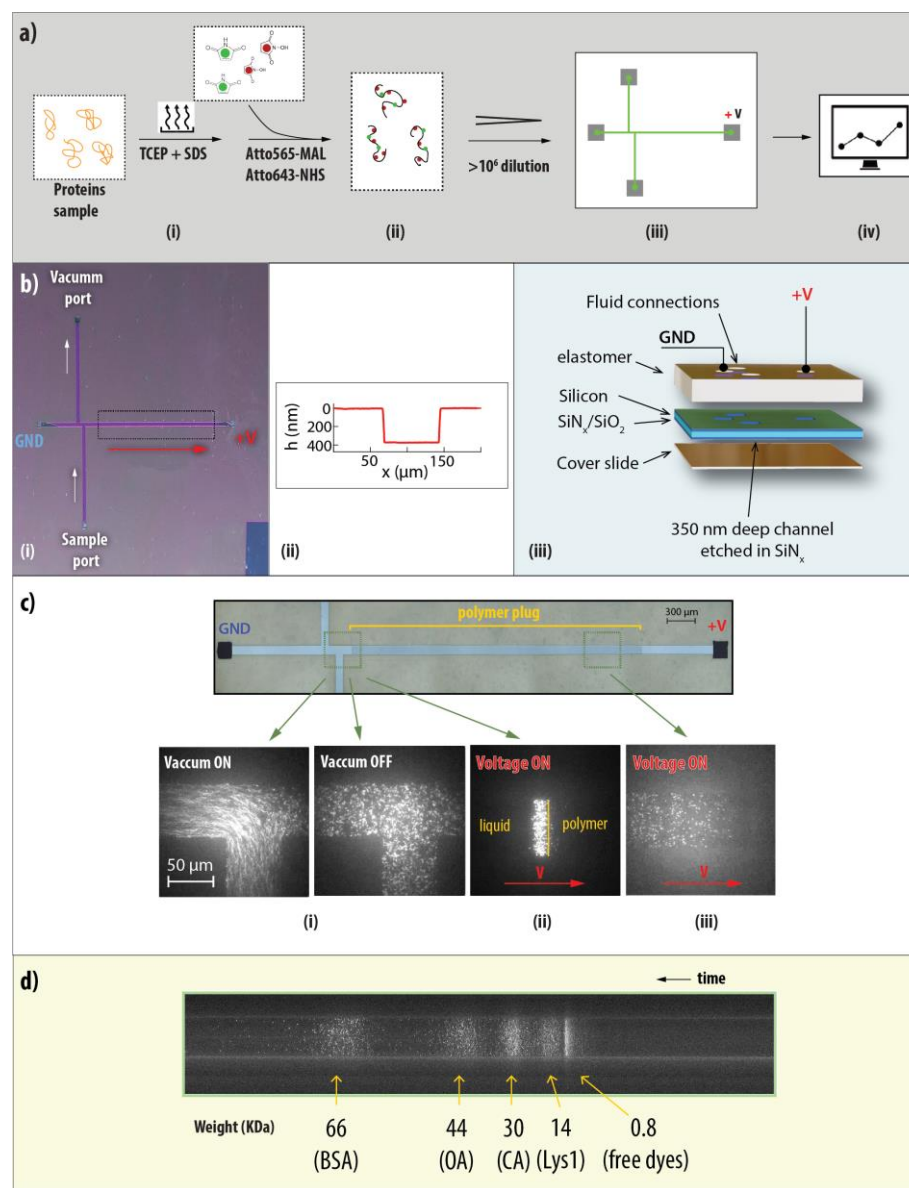
Figure 1 | Full-length, single protein separation and sensing in a nano-channel device. a A schematic illustration of the four steps for single-protein molecule identification in the thin channel device. (i) and (ii) Proteins denaturation and dual colour, specific amino acids labelling. (iii) High-throughput analysis using a custom nanochannel device, and (iv) single molecule tracking and feature extraction. **b** Silicon-based nanochannel for single protein molecule separation imaging and identification. (i) A micrograph showing the nanochannel device, viewed from the cover slide side. (ii) 350 nm deep, 4 mm long channels are etched in the SiN_x layer using RIE, and through holes ports are formed using wet etch prior to bonding of glass slide to from a solid channel. (iii) A schematic side view of the device, showing the glass slide, silicon chip and an elastomer layer bonded at the back of the device to facilitate fluid connections and electrodes placement. **c** Top: a magnified white light view of the main channel showing the polymer plug section in which proteins are separated by mass/charge. Bottom: Fluorescence images of: (i) the labelled proteins during loading into the double "T" junction, (ii) stacking at the liquid/polymer interface, and (iii) during electrophoretic migration in the polymer plug (reproduced from ref [26]). **d** Fluorescence Kymograph of a run consisting of 4 labelled proteins with different molecular weights, visibly separated along the channel.

Th

Our protein identification method consists of four steps, described schematically in **Figure 1a**. First proteins are denatured (i) using a combination of reducing agent (TCEP), anionic surfactant (SDS) and heat. This step is necessary to permit near complete labelling of all Cysteines (C) using Maleimide and Lysine (K) using NHS-ester reactive dyes (ii, Green and Red, respectively). To that end, we modified chemical conjugation procedures to maximize the degree of labelling (DoL: the fraction of K or C labelled per available residues per protein). In most cases we reach 100% DoL for both dyes, and in all cases the DoL > 60%, as confirmed by UV-Vis and bulk gel analyses (**Supporting Information**). Samples are diluted by at least 6 orders of magnitude and loaded to the device (iii), where molecules are separated by molecular weight and imaged using a high sensitivity dual-colour imaging system. Finally, the capture videos are analysed to obtain single-molecule tracking using a custom algorithm (iv), which produces detailed 4D profiles of each and every molecule as function of time: their position, velocity, intensity in the Green and the Red channels.

A key feature of our method is the ultra-high aspect ratio channel (greater than 10,000:1 Length/Height) that enables not only a separation of single protein molecules along the channel length but also *tracking* them *one by one* over time (**Figure 1b**). Surface immobilization of proteins is a powerful method to facilitate single molecule sensing over time using optical methods.[25] However, protein immobilization does not allow simple protein mass separation, and limits the method's throughput. On the other hand, free diffusion of small proteins in solution highly complicates single molecule tracking, as they quickly drift out of focus. To overcome these limitations, we introduce devices featuring sub-wavelength solid channels that physically constrain the proteins within a high-resolution focal area of a few hundreds of nanometers in the *z*-direction. Furthermore, the nano-channel is selectively fitted *in-situ* with a polymeric plug (marked by a dotted rectangle in **Figure 1b** left and **1c** top panels) that drastically slows down their Brownian motion or electro-migration. Hence, this design permits: (i) very fast introduction of thousands of proteins into the device and their delivery towards the separation section, and (ii) electrokinetic mass-based separation of proteins, imaged in real-time with high SNR (signal to noise ratio) permitting single-molecule resolution (**Figure 1c,d**). Importantly, the high SNR achieved at the polymeric plug allows *single-particle tracking* of the electro-kinetically migrating proteins. The nonlinear dynamical migration of the proteins in the gradient polymeric matrix separates them by their mass, producing dynamical velocity tracks for each protein, which adds characteristic information for each protein species.

The devices were fabricated in-house by chemical dry etching selected areas in low-stress silicon-nitride (SiN$_x$) layers deposited on silicon wafers (see **Figure S1**), and hermetically sealing the front side of the chip supporting the channels with glass slides, which is anodically fused to the SiN$_x$ layer. The back side of the chip is bonded to a PDMS layer to create fluidic connections (**Figure 1b iii**). We found that the solid material fabrication is necessary to support the high aspect ratio of the device, without collapsing.[26] The filling zone of the devices consists of a double "T" offset junction, which defines a precise sample volume loading of 4 pL. This is an outstanding feature of our device since ultra-small volume analysis represents an ongoing challenge for current methodologies.

Figure 1c (top) shows a white-light picture of the double "T" offset junction and the polymer plug section used for protein separation. The bottom panels display fluorescence microscopy images during: (i) vacuum-based loading of the hundreds of labelled proteins, (ii) electrophoretic stacking of proteins at the liquid/polymer interface, and (iii) protein migration and separation by mass in a >3 mm long separation channel with a height of ~350 nm (a profilometer scan of the channel is shown in **Figure 1b ii**). The device's internal walls were coated with an anti-stick layer to minimize protein sticking and the effects of an electro-osmotic flow (Methods). The channel is then filled with uncured polyacrylamide and is selectively polymerized *in-situ*, using a UV light illuminated through a microscope objective lens, only at the designated "plug" section by mounted the chip on a precision XY motorized stage. Controlling the polymerization location and the UV dose allows us to form a *gradient gel density* along the channel, leaving the feed-through sections unpolymerized for fast sample loading. Single-molecule imaging is performed through the glass cover slide in the "imaging zone" located at a distance of roughly 2.5 mm along the channel.

4

To image the devices, we constructed a custom dual-colour single molecule sensing setup that produces uniform illumination of the imaging zone using two laser excitations (532 nm and 640 nm, **Figure S2**). The laser exposure times were synchronized with an EM-CCD frame acquisition periods and were alternated to allow independent excitation of "green" (Atto 565) and "red" (Atto 643) fluorophores. The emitted light was collected using a high NA microscope objective (Olympus 60x/1.45), split to two channels, filtered and imaged by an EM-CCD camera. The device was placed on a piezo-driven Z axis stage equipped also with long-travel X-Y motors, to allow its precise placement at the imaging zone as confirmed using a white light image). To electrokinetically drive the proteins through the channel, a steady voltage was applied using a computer-controlled electrometer. The system was monitored in real-time using a custom LabView program, which in addition to fully controlling all parts of the system was also used to directly stream the videos to disk from the EM-CCD. **Figure 1d** is a single-colour kymograph showing the ability of the nanochannel to separate proteins by mass: four different proteins ranging from 14 kDa to 66 kDa, are analysed and clearly resolved. The kymograph was produced by stacking a single line of camera pixels across the channel as a function of time (right to left).

3. **Single molecule dynamic tracking enables protein discrimination.**

5

The low profile of the channels in the z direction ensures that the proteins remain in focus during their migration, as shown in **Supporting Video S1.** We find that this and the high SNR in the single-molecule imaging are essential features for the development of a single-particle tracking algorithm, used to numerate each protein migrating through the camera frame (**Figure 2a**). To produce continuous particle trajectories,
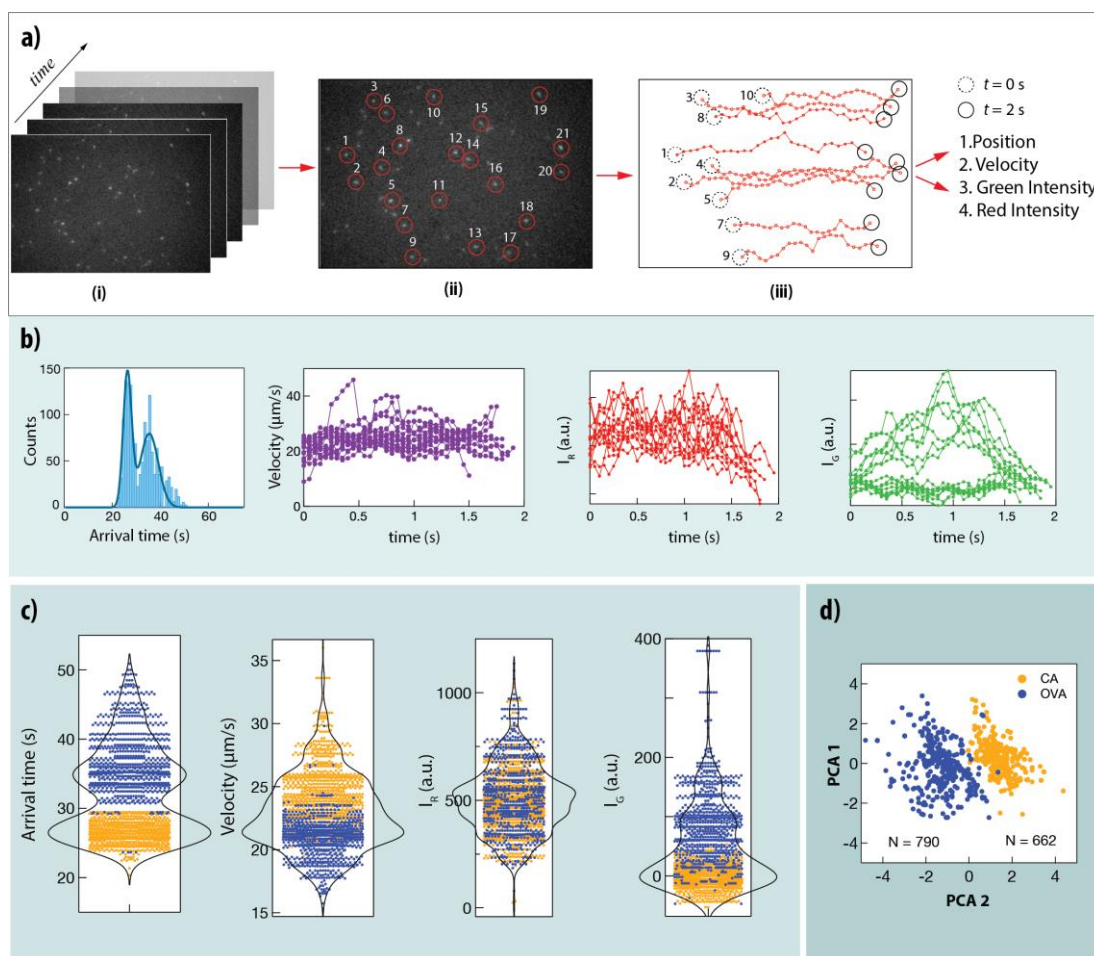
Figure 2 | Single protein molecule tracking and identification in solid-state channels with dual-colour labelling. a Single protein molecule trajectories along the channel are enabled by capturing high-resolution videos (i) of a portion of the channel and performing single particle identification (ii) and tracking of each molecule over time (iii). Alternate lasers (565 nm and 640 nm) excitations yield two independent trajectory plots for the green and red fluorophores (a single channel is shown for simplicity). **b** The protein molecules trajectories are used to calculate their arrival time histogram (light blue), their in-frame velocity profile (purple), and their fluorescence intensity profiles when excited by the red and by the green lasers (red and green). **c** The mean values of the single molecule tracks (N = 1,452) are plotted as "violin" graphs, and are subjected to a Gaussian Mixture Model clustering, which annotated two groups. The faster protein (yellow) is identified as the CA (M_w post labelling = 46.78 kDa) and slower protein (blue) as OVA (M_w post labelling = 66.65 kDa), consistent with the fact that CA shows zero green emission (it has no C residues), whereas OVA shows a high green emission. Both clusters show similar red emission, as expected due to similar number of K. **d** Principal Component Analysis (PCA) of the 4D information allowed simplified representation in a 2D graph, clearly showing two distinct clusters, as annotated by GMM.

6

we first localize the particles in each frame (panel ii) and then apply a routine based on Kalman filter that connects the most likely trajectories throughout the entire frame stack (panel iii and Methods).[27,28] **Supporting Video S1** shows representative dynamical trajectory tracing of a small number of molecules (for clarity), noting that our full data typically contains hundreds of single molecule trajectories. Importantly the dynamic single molecule trajectories contain rich information separated to 4 channels: the molecule's position (in the *x* direction), the molecule's velocity, the molecule's intensity when excited by the Green laser (I_G) and the molecule's intensity when excited by the Red laser (I_R).

To validate the entire process and analysis tools and to demonstrate the method's capacity to extract 4D information, we prepared mixtures of two dually labelled proteins with easily distinguishable mass differences: Carbonic anhydrase - CA (30 kDa), and Ovalbumin - OVA (44.2 kDa). These proteins harbour similar numbers of K residues (18 and 20, respectively) but different numbers of C residues (0 and 6, respectively, **Table 1**). A difference of roughly 20 kDa in M_w post-labelling is easily discernible in bulk SDS-PAGE (**Figure S3**). Prior to loading of samples to the chip in the designed port, the sample is diluted to a typical in-channel concentration of 1 to 100 pM for single-molecule sensing. A voltage is applied between the two platinum electrodes (GND and "+V"), causing proteins to quickly stack at the buffer/gel interface and begin their slow migration in the polymerized section of the channel.

Table 1 | Details of the proteins used in this study

In the top panel proteins properties used for validation of the method and for discrimination and quantification of the VEGF isoforms. In the bottom panel properties of the four cytokines, used for host-response analysis. The molecular weight post labelling is estimated based on full labelling approximation.

Protein	M_w (kDa)	Cysteines	Lysines	M_w post labelling
CA	30	0	18	46.78
OVA	42	6	20	66.65
VEGF 121	14.06	9	7	26.29
VEGF 165	19.1	16	11	39.49

Cytokine	M_w (kDa)	Cysteines	Lysines	M_w post labelling
IP-10	8.65	4	10	20.51
TRAIL	19.6	1	11	30.49
IL-6	20.98	4	14	36.57
CRP	23.2	2	13	36.59

As shown in **Supporting Video S2** two discernible populations of single molecules particles appear. The particle trajectories were used to produce 4 graphs (**Figure 2b**): the histogram of the individual proteins' arrival times from the beginning of the channel gel plug up to the observation zone ("arrival time", light blue), the proteins' in-frame velocities (purple), I_R (red), and I_G (green). For clarity, in **Figure 2b** we show only a dozen of representative single protein tracks out of roughly 1,452 proteins collected in < 60 s experiment. Notably, the proteins' the *in-frame* mean velocity is not constant along the channel and depends on their location along the channel. The arrival time (t_a) is a sum of two main terms, which both vary non-linearly with proteins' M_w: the entry time of proteins into the gel plug, and their in-gel migration time. This nonlinearity is further magnified by applying a gel *density gradient* during the UV polymerization. Consequently, t_a is found to be highly sensitive to the proteins' mass/charge ratio and the DoL.

To analyse the results the mean values of the 4 graphs in **Figure 2b** accumulated for 1,452 single-molecules, were plotted as "violin plots" histograms (**Figure 2c**). Each spot on these plots represents the mean values of a single protein trajectory along the channel. In principle, it is possible to fit these distributions *independently* to a mathematical model representing the two species. However, this would not take full advantage of the fact that the information is obtained from a single experiment. Inspired by previous literature,[29] we treat the whole data using a Gaussian Mixture Model (GMM), which "tags" each particle to its most likely cluster, based on global minimization fit in all four graphs. The global GMM analysis annotated each particle as either CA (yellow) or OVA (blue). Starting from the right panel, we see two clear peaks, where the faster proteins (arrival time of roughly 28 s) are denoted as the CA and the heavier proteins (OVA) arrival time are roughly 35-40 s. Consistently, at

7

the green laser excitation the near zero proteins are the CA (yellow) which have *no* C residues and the rest are OVA (6 C). The red laser excitation panel is randomly mixed, consistent with the fact that the two proteins harbour similar numbers of K residues (18 and 20, respectively). Finally, the GMM correctly annotated the faster proteins on the Velocity plot as the CA (yellow) with mean speeds around 25 μm/s or more, and the slower proteins as the OVA (blue) velocities around 22 μm/s or less. A 2D Principal Component plot of the four-dimensional information is reported in **Figure 2d**, showing clear separation of the data into two distinct clusters of events. We can use this classification to numerically count the two proteins with high degree of confidence. **Figure 2d** shows a very small number of misidentified particles (< 0.4 %). To validate the two-cluster classification of the GMM analysis we performed Calinski Harabasz statistical analysis using our 4D tracking data. Our results (**Figure S8**) confirm that two clusters are statistically most probable attributing high confidence to our analysis method.

4. Single protein molecule quantification of a cytokine panel

Next, we used our method to quantify the relative numbers of a set of cytokines of biomedical relevance. Discrimination between bacterial and viral infections is often challenging due to the similarity of their clinical symptoms, but is crucial to avoid excessive use of antibiotics and prolonging recovery time due to inappropriate treatments.[30] Diagnostic tests, which are routinely carried out in the clinic, include culture, serology, and nucleic acid-based tests (such as RT-PCR), may assist clinicians in determining the source of infection and, subsequently, the most accurate treatment, but are all directed towards pathogen identification.[31] An alternative or complementary diagnostic strategy is to analyse the host's immune response to the infection.[32] This strategy bypasses the need to determine whether an identified pathogen is the direct cause of the infection or an unrelated colonizer as well as potentially provides better diagnostic outcomes in the case of more complicated cases, such as mixed infections with both virus and bacteria.[33] Circulating host biomarkers can be monitored by enzyme-linked immunosorbent-assay (ELISA) at the point-of-care, however accurate and unbiased quantification remains to be a challenge. Interleukin 6 (IL-6) and C-reactive protein (CRP) are such biomarkers, up-regulated in bacterial infections, which are often used to support pathogen-based diagnosis in the clinic. Large-scale proteomics screen performed recently introduced a host-induced viral biomarker, TNF-related apoptosis-inducing ligand (TRAIL).[33] Additionally IP-10 is a smaller cytokine, which showed elevated levels in both bacterial and viral infections, with a more pronounced increase in the latter case. The combined signature of CRP, IP-10, and TRAIL cytokines resulted in an accurate and robust differential diagnosis of acute bacterial or viral infections. This study, validated in subsequent clinical studies,[34–36] prompted us to explore the possibility of *single molecule and antibody-free* discrimination among CRP, IP-10, TRAIL, and IL-6 cytokines while quantifying their relative abundance.

First, we optimized our system by using two or three cytokines, which were easier to identify (**Figure S12 and S13**): IL6 and TRAIL have very similar M_w, as reported in **Table 1**, but are different in their number of Cysteines (4 and 1, respectively), whereas IP10 is a much lighter protein (8.6/20.5 kDa pre/post labelling) and harbours 4 C like IL6. As shown in **Figure S13**, IL6 and IP10 are readily separated by all four metrices independently. In contrast, both IP10 and Trail harbour a similar number of Lysines (10 and 11, respectively), whereas IL6 harbours only a single K. **Figure S12** summarizes the channel measurement using IL6, IP10 and Trail. As expected, the lightest protein IP10 (orange markers) separates readily in both velocity and the arrival time, whereas IL6 and TRAIL (brown and blue markers, respectively) display markedly different green signal. GMM analysis of the entire data set was used to identify and count the each of the three cytokines.

Next, we performed experiments with all four cytokines in a mixture, shown in **Figure 3**. The fourth cytokine (CRP) molecular weight is 23/37 kDa (pre/post labelling) and it harbours 2 C and 13 K, which should allow it to be discriminated by the FRET fluorescence intensity from IL6, as shown in **Figure S13** lower panel. Bulk SDS-PAGE analysis of each protein after dual labelling along a molecular weight ruler is shown in panel a (left-hand panel shows the green excitation and the right-hand panel the red excitation). Using the molecular weight calibration, the gel suggests that the proteins are labelled as expected (see **Table S2**). This was further confirmed by separately analysing each of the proteins in the UV-Vis spectrometer for CRP and TRAIL (**Figure S5**) and for IP10 and IL6 by

8

migration shift estimation post C labelling. The right-most lane in the gel shows all four cytokines in a mixture. Notably, despite the very high quantity of the proteins used in this <u>bulk</u> analysis, a clear identification of the 4 cytokines in bulk is unpractical.

We performed a single-molecule protein analysis of >1,000 proteins in our devices. As before, single particle tracking was performed leading to 4 violin plots (**Figure 3b**). Global GMM classification annotated the proteins as graphically displayed in the PCA plot (**Figure 3c**, left panel). Going back to **Figure 3b**, we tagged the proteins based on the GMM classification: from the in-frame velocity and the arrival time graphs, we can easily identify the lighter protein IP10 marked in yellow, as well as the heaviest protein CRP, marked in red. The green intensity is extremely useful in separating TRAIL with its single C (blue) from the other two proteins harbouring 4 C, but not from CRP containing 2 C (red). IL6 shows both high green and high red intensities (in brown), and TRAIL (in blue) exhibits relatively low green and low-medium red intensities, as expected. The GMM classification directly counts the number of each of the proteins in the data set, as presented in the right-hand panel of **Figure 3c**. The relative numbers of all 4 cytokines shown in this figure is directly proportional to their concentration ratios at the post-labelling stage (see Methods). Given that for the single-molecule analysis we use a minute volume (sub nL) and concentration (~10 pM) our method potentially offers an enhanced sensitivity by several orders of magnitude compared with current methods to accurately differentiate between bacterial and viral infections, without reliance on antibodies during the sensing/quantification stage.

9
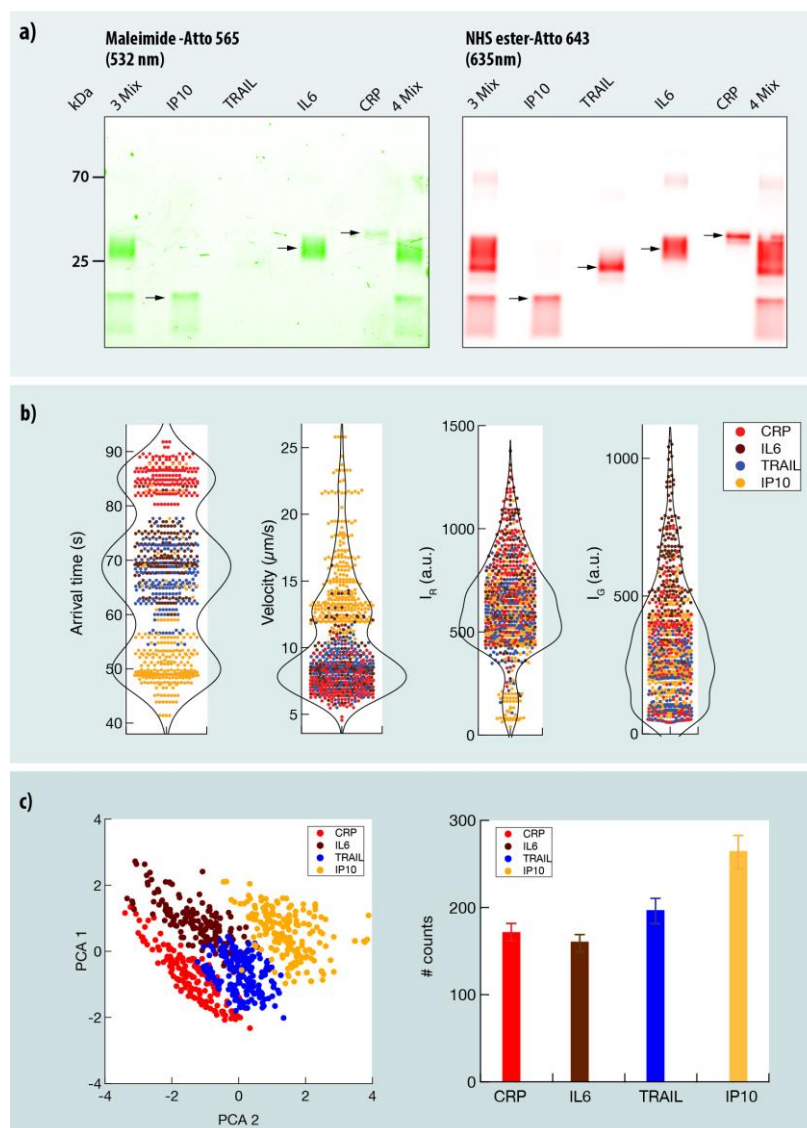
Figure 3 | Antibody-free, single protein molecule quantification of a 4 cytokines panel.
a SDS-PAGE image of dual-colour labelled cytokines panel (IP10, Trail, IL6, CRP). All
proteins were labelled as described in Methods to >70% yield. **b** Violin plots of the 4 cytokine
mix run in channels. Despite the similarity in M_w among some of the proteins, our method
can distinguish among the 4 cytokines. **c** Left: PCA plot with cytokine annotation based on
the GMM analysis. The counts of each of the 4 proteins are shown on the right, permitting
a direct quantification of the sample.

5. **Single-molecule quantification of VEGF isoforms.**

Our method can resolve and quantify small differences in the proteins' M_w as well as their C and K
amino-acid composition. Consequently, it can be applied to resolve and count full-size proteoforms
that are not easily distinguished by MS or immunosorbent methods. To demonstrate this strength, we
analyse two closely related isoforms of the Vascular Endothelial Growth Factor-A (VEGF-A) protein,
which rise from alternative splicing of the VEGFA gene. The ratio between two isoforms, VEGF121
and VEGF165, which is relevant in various cancerous processes[37] was quantified either when spiked
into human serum or endogenously using our method. VEGF has been associated with multiple
physiological and pathological processes ranging from vasculogenesis and angiogenesis to vascular
disease in the retina and cancer.[38] In addition to being subjected to proteolytic regulation, VEGF
expression is diversified using alternative splicing.[39] To date, most quantitative studies on VEGF

10

isoforms relied on mRNA expression,[40] however mRNA expression levels do not always correlate with the actual protein content. Notably, the discrimination among VEGF proteoforms is often

11

complicated by the lack of specific antibody pairs to allow an unbiased quantification of the various
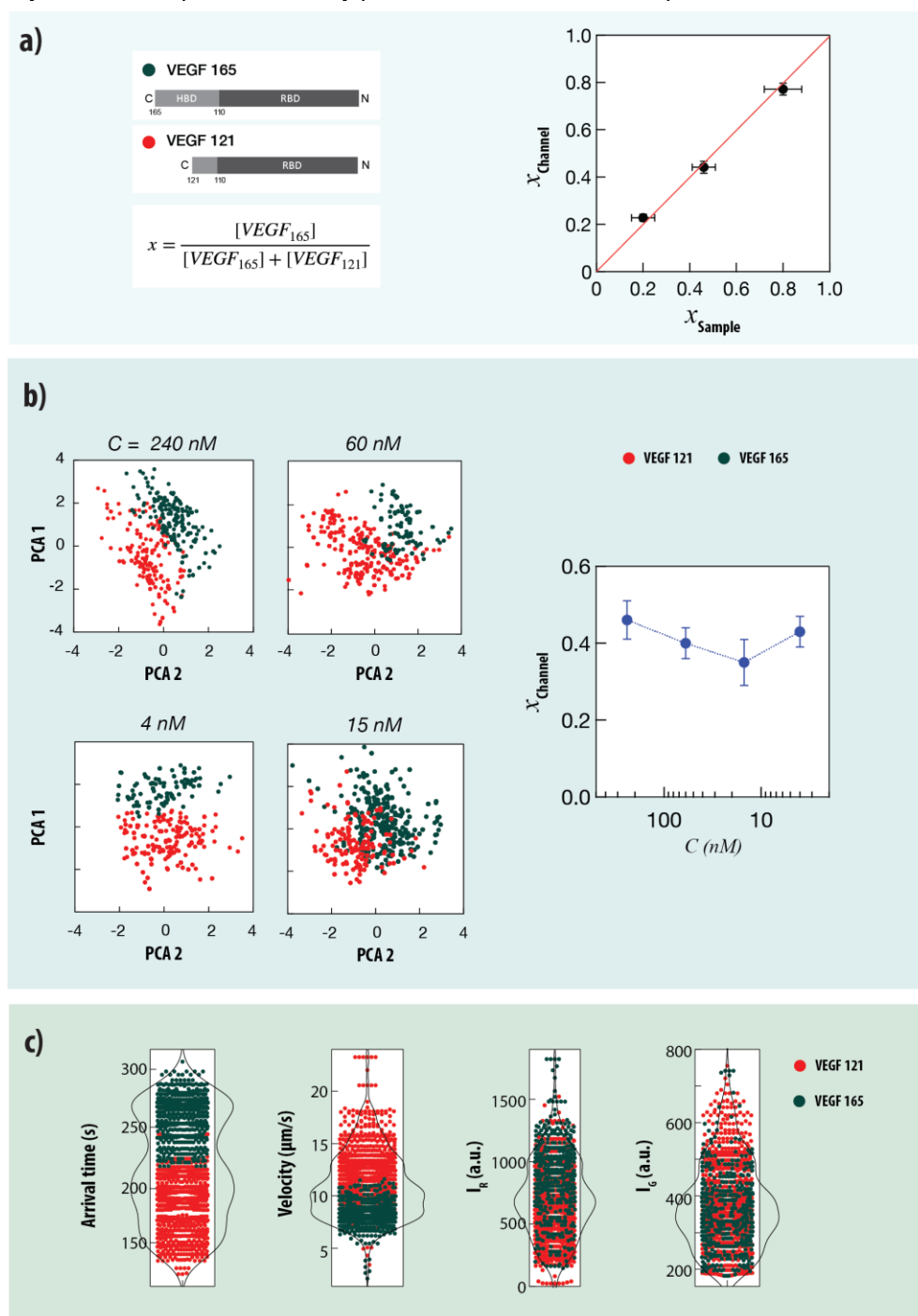
Figure 4 | VEGF isoforms separation and digital quantification in buffer and from human serum. **a** Three mixtures of dually labelled VEGF121 and VEGF165 with different relative concentrations $x = C_{165}/(C_{165} + C_{121})$ analyzed using the channel-based device. The 4D information (**Figure S10**) is used to cluster the data using GMM and calculate the ratio of two isoforms. Right: Comparison of the prepared VEGF isoform samples mixture ratio with the single-molecule counting results. A linear regression fit yield a slope of 1.00 ± 0.06 suggesting that the single-molecule counting analysis of the VEGF isoform is quantitative. **b** Spike-in of human sera samples with mixture of the VEGF isoforms at 4 different initial concentrations (from 240 nM down to 4 nM) having the same sample ratio $x = 0.4$. The resulting PCA plots are shown (the 4D information is provided in **Figure S11**). Right: The resulting single-molecule isoform quantification pulled down from the spiked human sera showing the recovery of the isoform ratio as expected $x = 0.40 ± 0.02$. **c** Single molecule quantification of endogenous VEGF isoforms from human serum, yielding ratio $x = 0.43 ± 0.02$ (n = 1,107). Violin plots results of the 4D tracking analysis are shown.

types,[40,41] particularly in physiological samples, in which VEGF resides in small amounts. Yet there is increasing evidence that different VEGF isoforms play different physiological roles, often via isoform selective co-receptors.[39] The VEGF121 and VEGF165 isoforms were also reported to bind VEGF receptors with different affinities and to play different roles in different types of malignancies.[42,43]

To show that our method can provide biologically and clinically useful insights from a clinical sample, we targeted VEGF121 and VEGF165 in buffer, in spiked serum sample and endogenously in human serum. To test the quantification accuracy of our method, we first individually labelled VEGF165 and VEGF121 using both Maleimide and NHS ester reactive dyes and analysed three different stoichiometric mixtures of the two isoforms with $x = C_{165}/(C_{165} + C_{121})$ using our method (**Supporting Video S3**). Our results, shown in **Figure 4a**, suggest that the two VEGF isoforms are readily separated based the 4D data (**Figure S10**).

We clearly observe a correlation between the fast-moving proteins and lower red intensities. This is expected given that VEGF 121 harbours smaller number of K residues as compared with VEGF165 (**Table 1**). Discrimination in the green channel is not observed, possibly since most of the Atto565 emission is transferred to Atto643 dye via FRET and the two proteins exhibit nearly undistinguishable FRET values. As before, GMM clustering of the results provides a tool to mark and identify the two VEGF isoforms from the single molecule tracking information. We used the isoforms' clustering results to calculate the in-channel measured ratio $x_{Channel}$ as a function of the prepared mixture ratio x_{Sample} as shown in the right panel of **Figure 4a**. The results show a strong quantitative correlation with a slope of unity (1.00 ± 0.06), proving that our method can quantify the VEGF isoform with high accuracy and low bias. Notably, despite unavoidable channel-to-channel variations (primarily due to slightly different degrees of polymer cross-linking), that can alter the proteins' kinetics among independent measurements, we observe a linear relationship between the prepared VEGFs isoforms stoichiometry (**Figure 4a**) and in-channel measurement, as well as stable dependence of the stoichiometry at different total VEGF concentrations (**Figure 4b**).Given that our method can measure the stoichiometric ratio of the VEGF isoforms, we next challenged its ability to do so starting with a clinical sample of spiked human sera. To that end we spiked human sera with VEGF isoform mixtures at four different total concentrations (from 240 nM down to 4 nM) and measured the recovered ratio. The spiked sera samples were first depleted of the most abundant proteins using an affinity column ("High-Select Top14"). Then we recovered the two VEGF isoforms using magnetic beads conjugated in house to a single monoclonal antibody targeting the same N-terminal epitope on both isoforms (Methods). As no additional antibodies are used to differentiate among the isoforms, selectivity biases are minimized. Results for four different total VEGF isoforms concentrations prepared in sera at a ratio $x = 0.4$, are shown in **Figure 4b**. Discrimination among the two recovered isoforms is straightforward using the 4D data analysis (**Figure S11**). In each case about 10^3 single molecule trajectories were collected within ~60 seconds yielding a measured ratio $x = 0.40 \pm 0.02$ in good agreement with the spiked value.

Finally, we used our method to detect endogenous VEGF isoforms from the human sera. The workflow for the preparation of the sample is similar to the one developed for the set of spike and recovery VEGF experiments (but obviously no spike-in). The summary of our results is shown in **Figure 4c**, where about 1,100 single molecule tracks were collected within a few minutes. As before, the separation between the VEGF121 and VEGF165 groups in terms of arrival time, average velocity and red emission were clear leading to straightforward GMM based clustering. From these results we measured the ratio of VEGF121 to VEGF165 to be 1.33:1 ($x = 0.43 \pm 0.02$) in the female human serum sample used. These results show that our method is capable of the identification of proteins in low concentrations from clinical samples relevant for diagnostic challenges.

6. Discussion
The development of methods for full-size protein identification capable of rapid, multiplexed and high-sensitivity detection is on great demand but still facing challenges. Emerging immunosorbent techniques offer single-molecule sensitivity but suffer from quantification biases when used for detection of many proteins simultaneously, and limited proteoform discrimination. Here we combine molecular weight separation of *full-length* individual proteins, with molecular quantification of the number of specific amino-acids to precisely identify protein molecules within a complicated protein

13

mixture. Four main aspects are the pillars of our method: (1) Crafting sub-wavelength heightchannel devices through lithographic methodologies and selectively equipping them with gradient polymer plugs. (2) Near complete dual-colour chemo-selective labelling of proteins targeting K and C residues. (3) The acquisition of dynamic protein videos, staged using a custom-built experimental setup. (4) Custom-made computer programs to cluster proteins based on 4-D information obtained from the single particle tracking of each protein molecule in the channel. A key feature of the device lies in its ability to slow down the protein's movement and introducing an element of nonlinearity to their motion, which is driven by voltage inside the polymer plug. As a result, we were able to measure dynamical trajectories over time of single protein molecules, used to uniquely identify proteins of similar mass and discriminate between them even in complex biological media. This approach offers several advantages compared to other single-molecule protein analysis based on nanopores, as we have shown analysis and discrimination of full-length proteins and the parallelization of the readout by tracking the motion of thousands of proteins within a few minutes.

To illustrate the methods' power, we demonstrated a quantitative discrimination among two *VEGF isoforms* in recombinant mixtures, spiked human sera samples with extremely low sample volume and concentrations, and endogenous VEGF proteins. Unlike the classical "sandwich" ELISA assay, our method evades the need for isoform-specific "reporter" antibody, while still permitting single-molecule sensing with high throughput. Since our method involves only chemical amino-acid specific conjugation, it is less susceptible to quantification biases and lend itself towards greater multiplexing ability. New conjugation chemistries targeting additional amino-acid residues[1] or even proteins PTMs[44] can further enhance the method's scope to address an extremely broad range of biological questions with single-molecule precision. To further demonstrate the method mutiplexibility we sensed and quantified a cytokine panel, relevant for differentiation between viral and bacterial infections.[33] Notably, this method can be integrated to enhance whole proteome screening and PTM mapping prior to literally any other single molecule sensing technique, including nanopore based protein sequencing, sm-FRET based protein recognition, fluorosequencing, as well as other emerging approaches involving N-terminal binders, or even future MS profiling.[12]

7. Methods

7.1 Single particle tracking algorithm
We acquired video-clips of the imaging zone using a custom LabView code that synchronizes the alternate dual laser excitation with the EM-CCD (Andor iXon) exposure. The clips are 512 x 512 pixels in size at a frame rate of 20 fps. Each protein can be seen as a bright particle moving across the channel. To achieve single-particle tracking we first determine the proteins' location and their area size (number of pixels it is spread on) in each frame. The images are corrected for background based on a mask consisting of the mean value of each pixel from the first 100 frames in the experiment, in which no proteins are present. Then we apply a spatial filter (Laplacian of Gaussian) on the image in order to find the border lines of the objects in the image. With this information the images are segmented creating a binary mask representing the particles in the image. This allows counting the number of objects (proteins) in the image, finding their area size, and calculating their intensity and centre of gravity. The centre of gravity corresponds to the protein location in the image and is calculated using:

$$Gravity\ center\ (x) = \sum_{i=1}^{n} x_i * \frac{pixel\ value_i}{energy}$$

The *x* index of the centre of gravity is calculated by the sum of the x index of each pixel (i) from the protein segment multiplied by the photon count of that pixel divided by the total energy of the protein segment (the total counts of the particle). The same equation is used to also calculate the *y* index of the centre of gravity.
The mean pixel value of 5 x 5 pixels around the protein centre of gravity calculates the intensity of the protein. At the end of this step, for each image in the experiment, we achieve a list of all detected proteins with their location, size area, and intensity to be used for the next tracking step.
Next, the video clips are divided into batches of 120 frames each and an overlap of 6 frames between each batch. The location and size of the proteins in a batch are the input for the tracking algorithm. In

14

this multi-object tracking algorithm, we used the Kalman filter to attribute each location to a certain protein. The algorithm collects the initial locations of the proteins and by assuming that an object moves in accordance with a particular motion model (such as constant velocity or constant acceleration), the Kalman filter forecasts the protein's next location. The process noise and measurement noise are taken into account by the filter. Here, the process noise is the difference between the protein's real motion and its motion model. Similarly, the detection error is referred to as the measurement noise using this algorithm, from which the trajectory of each single protein is defined. Further individual protein trajectory is monitored and if the protein is not detected for a few frame numbers (a threshold frame number), it will predict its location using the Kalman filter until it appears again. In case the protein does not appear again, the algorithm will discard this proteins trajectory. The advantage of Kalman filter is that it can predict the location of a protein even when its fluorescence emission is too weak to be detected in each frame. The final output is the trajectory of many proteins in a batch. In a single batch, it is possible to have a very high number of objects with similar properties, hence complicating accurate tracking. To overcome this issue, we filtered unlikely trajectories, in which a protein stepped against the applied electrical field. From the proteins tracking we extract the three single particle trajectories: (1) in-frame velocity, (2) fluorescence intensities when excited by red laser (640 nm), (3) fluorescence intensities when excited by green laser (532 nm). We find that all proteins (except CA) exhibited medium to high FRET levels when excited by the green laser.

7.2 Data clustering and classification
Initial estimation of the classification values was based on K-means using the expected number of clusters. With these initial values we performed Gaussian Mixture Model (GMM) classification which provides: i) The mean value of each parameter per cluster, ii) the covariance matrix of the parameters and clusters, and iii) proportion of each gaussian distribution. Each 4D data point can then be assigned a probability value for a certain cluster based on its cartesian proximity to the gaussian distribution. Principle component analysis (PCA) is used only to display the 4D data in the 2 main coordinates, when each data point is being colored based on the cluster it belongs to.

7.3 Materials
EDTA-treated sera samples (NEGSMPL-P-100) were obtained from healthy subjects from RayBiotech. High-Select™ Top14 Abundant Protein Depletion Resins (catalogue A36370) and Dynabeads MyOne Tosylactivated (catalogue 65501) were purchased from Thermo Fisher Scientific. Mouse IgG kappa (clone MG1-45) and Mouse anti-VEGF antibody (clone L308D10) were purchased from Biolegends. Atto565 maleimide (AD 565-45) and atto643 NHS ester (AD 643-35) were purchased from Atto-tech GmbH. DMSO was purchased from (Sigma). TCEP was purchased from SIGMA. The details of all proteins used in this study are provided in **Table S1**. VivaSpin concentrators with a 30kDa cut-off were purchased from Cytivia. Pre-cast 4-12% Bis-Tris SDS-PAGE were purchased from Thermo Fisher Scientific. 10% SDS was purchased from Biorad.

7.4 Dual-colour labelling of proteins
Proteins were resuspended in Maleimide labelling buffer consisting of 10 mM sodium phosphate, 150 mM NaCl and 1 mM SDS (pH 7.4). Disulfide bonds were reduced using 0.5 mM TCEP at 37°C for 30 minutes and denatured at 95°C for 5 minutes. After cooling down at RT for 15 minutes, atto565 maleimide, dissolved in DMSO or in buffer, was added to the protein samples at a ratio of 2.6x fold per C residue and incubated at 25°C for overnight with shaking at 300 rpm. Subsequently, the samples were dialysed against 10 mM sodium phosphate, 150 mM NaCl and 1 mM SDS (pH 8.4, adjusted with 0.2M sodium bicarbonate (pH 9)). Atto643 NHS ester, dissolved in DMSO or in buffer, was added to the protein samples at a ratio of 20x fold per K residue and incubated at 25 ˚C for 1h inside a heat block with shaking at 300 rpm. The labelled samples were diluted 10x fold with 10 mM sodium phosphate and 1 mM SDS (pH 8.4) buffer to adjust the NaCl concentration to 15 mM. Unconjugated fluorophores were removed using at least six buffer exchange washes with the protein concentrator VivaSpin of a 30 kDa cut-off. The final volume of the labelled proteins was adjusted with NHS ester labelling buffer containing 10 mM sodium phosphate, 15 mM NaCl and 1 mM SDS (pH 8.4). A detailed explanation on how the labelling yield was estimated/calculated is provided in the SI. Qualitative analysis for the labelled proteins was carried out by separating a small fraction of the

15

dually labelled protein samples on a 4-12% Bis-Tris SDS-PAGE and imaged using the Pharos scanner (Biorad) with laser excitation at either 532 or 635 nm.

7.5 Conjugation of BSA or antibodies to tosylactivated Dynabeads

40 µg of either mouse IgG or mouse anti-VEGF antibody were conjugated to 1 mg of tosylactivated beads according to the manufacturer protocol for 16h at 37 °C on a tube rotator. The blocking step was performed in 1X PBS with 0.05% Tween-20 and 0.5% BSA for overnight at 37°C on a tube rotator. Subsequently, the beads were washed three times in 1X PBS with 0.05% Tween-20. Finally, the beads were resuspended at 2.5 mg/ml final concentration using 1X PBS containing 0.05% Tween-20 and 0.02% sodium azide and kept at 4°C for further usage. Similarly, 2 mg of BSA were conjugated to 1 mg of tosylactivated beads according to the manufacturer protocol to obtain BSA-conjugated tosylactivated beads.

7.6 Immunoprecipitation of VEGF165 and VEGF121 isoforms from human serum

Human sera samples from healthy subjects were spiked with different initial concentrations of VEGF121 and VEGF165. These samples were subjected to high abundant protein depletion using the Thermo Scientific High-Select™ Top14 Abundant Protein Depletion Resin according to the manufacturer protocol. The diluted high abundant proteins-depleted plasma was adjusted to contain a final concentration of 1X phosphate buffer saline (PBS) supplemented with 10% glycerol and 0.05% Tween-20 and was subjected to centrifugation at 14,000 g at 4 °C for 10 minutes. The resulting supernatant was incubated with BSA-conjugated Dynabeads (Thermo Fisher) for 1 h at 4 °C on a tube rotator. Total protein of the Top14-depleted plasma sample was quantified using the BCA reagent assay as compared to a BSA standard curve. Typically obtained protein concentrations were ~0.15 mg/ml. After this pre-clearing step, the BSA-conjugated Dynabeads were collected on a magnet, the supernatant was transferred into a new tube, and subjected to centrifugation at 10,000 g at 4 °C for 2 minutes. Subsequently, 10 mg of total protein of the this cleared sera supernatant was used per immunoprecipitation experiment. The samples were incubated with 2-4 µg of either mouse IgG or mouse anti-VEGF antibody-conjugated to tosylactivated Dynabeads (prepared at 0.04 mg antibody per mg beads) at 4°C for overnight on a tube rotator. Anti-VEGF antibody-conjugated beads were collected on a magnet for 2 minutes and the unbound supernatant was transferred into a new tube and kept for analysis. The immunoprecipitated proteins bound to the beads were washed 5 times with 1XPBS containing 10% glycerol and 0.05% Tween-20 and finally eluted using 20 µl of 0.1M Tricine buffer (pH 3) at RT for 10 min. The eluted samples were neutralized to physiological pH using NaOH, mixed at 1:1 ratio with 2x MAL labelling buffer and subjected to dialysis against 1x MAL labelling buffer. For gel analysis, the eluted sample was mixed with final 1X Laemmli sample buffer containing 25 mM Tris-HCl, pH 6.8, 10% (w/v) SDS, 10% (v/v) glycerol and 0.1% (w/v) bromophenol blue, denatured at 95 °C for 5 minutes and subjected to separation on 4-12% Bis-Tris SDS-PAGE using 1x MES buffer at 150V for 1h under non-reducing conditions. The proteins were fixed at RT for 2h on a shaking platform using 40% ethanol and 10% glacial acetic acid. Subsequently, the gel was stained with 1X Flamingo stain in milliQ water for overnight at RT on a shaking platform. Finally, the gel images were acquired using the Pharos scanner (Biorad) with laser excitation at 532 nm. For single protein analysis using the nano-channel device, the eluted samples were subjected to the dual-colour labelling protocol with Atto565 maleimide and Atto643 NHS ester.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Data Availability

The main data supporting the findings of this study are available within the Article and its Supplementary Information. The raw data generated in this study are available from the corresponding author upon reasonable request.

Acknowledgements

16

Author Contributions
S. Ohayon, L. Taib, N. C. Verma, I. Bhattacharya and M. Iarossi performed all measurements and analysed the data, D. Huttner and B. Marom prepared protein samples, A. Meller designed and supervised the research. All authors co-wrote the manuscript.

Competing interests
The authors declare no competing interests.

References

[1] J. A. Alfaro, P. Bohländer, M. Dai, M. Filius, C. J. Howard, X. F. van Kooten, S. Ohayon, A. Pomorski, S. Schmid, A. Aksimentiev, E. V Anslyn, G. Bedran, C. Cao, M. Chinappi, E. Coyaud, C. Dekker, G. Dittmar, N. Drachman, R. Eelkema, D. Goodlett, S. Hentz, U. Kalathiya, N. L. Kelleher, R. T. Kelly, Z. Kelman, S. H. Kim, B. Kuster, D. Rodriguez-Larrea, S. Lindsay, G. Maglia, E. M. Marcotte, J. P. Marino, C. Masselon, M. Mayer, P. Samaras, K. Sarthak, L. Sepiashvili, D. Stein, M. Wanunu, M. Wilhelm, P. Yin, A. Meller, C. Joo, *Nat Methods* **2021**, *18*, 604.

[2] M. Eisenstein, *Nature* **2023**, *613*, 794.

[3] V. Marx, *Nat Methods* **2023**, *20*, 350.

[4] L. M. Smith, N. L. Kelleher, *Science (1979)* **2018**, *359*, 1106.

[5] M. J. MacCoss, J. A. Alfaro, D. A. Faivre, C. C. Wu, M. Wanunu, N. Slavov, *Nat Methods* **2023**, *20*, 339.

[6] H. M. Bennett, W. Stephenson, C. M. Rose, S. Darmanis, *Nat Methods* **2023**, *20*, 363.

[7] A. Singh, *Nat Methods* **2021**, *18*, 856.

[8] L. Cohen, N. Cui, Y. Cai, P. M. Garden, X. Li, D. A. Weitz, D. R. Walt, *ACS Nano* **2020**, *14*, 9491.

[9] D. M. Rissin, C. W. Kan, T. G. Campbell, S. C. Howes, D. R. Fournier, L. Song, T. Piech, P. P. Patel, L. Chang, A. J. Rivnak, E. P. Ferrell, J. D. Randall, G. K. Provuncher, D. R. Walt, D. C. Duffy, *Nat Biotechnol* **2010**, *28*, 595.

[10] J. Van Damme, G. Opdenakker, S. Van Damme, S. Struyf, *Eur Cytokine Netw* **2023**, *34*, 1.

[11] L. Restrepo-Pérez, C. Joo, C. Dekker, *Nat Nanotechnol* **2018**, *13*, 786.

[12] J. Swaminathan, A. A. Boulgakov, E. T. Hernandez, A. M. Bardo, J. L. Bachman, J. Marotta, A. M. Johnson, E. V Anslyn, E. M. Marcotte, *Nat Biotechnol* **2018**, *36*, 1076.

17

[13] H. Brinkerhoff, A. S. W. Kang, J. Liu, A. Aksimentiev, C. Dekker, *Science (1979)* **2021**, *374*, 1509.

[14] E. C. Yusko, B. R. Bruhn, O. M. Eggenberger, J. Houghtaling, R. C. Rollings, N. C. Walsh, S. Nandivada, M. Pindrus, A. R. Hall, D. Sept, others, *Nat Nanotechnol* **2017**, *12*, 360.

[15] H. Ouldali, K. Sarthak, T. Ensslen, F. Piguet, P. Manivet, J. Pelta, J. C. Behrends, A. Aksimentiev, A. Oukhaled, *Nat Biotechnol* **2020**, *38*, 176.

[16] S. Zhang, G. Huang, R. C. A. Versloot, B. M. H. Bruininks, P. C. T. de Souza, S. J. Marrink, G. Maglia, *Nat Chem* **2021**, *13*, 1192.

[17] T. Ensslen, K. Sarthak, A. Aksimentiev, J. C. Behrends, *J Am Chem Soc* **2022**, *144*, 16060.

[18] P. Shrestha, D. Yang, T. E. Tomov, J. I. MacDonald, A. Ward, H. T. Bergal, E. Krieg, S. Cabi, Y. Luo, B. Nathwani, A. Johnson-Buck, W. M. Shih, W. P. Wong, *Nat Nanotechnol* **2021**, *16*, 1362.

[19] B. D. Reed, M. J. Meyer, V. Abramzon, O. Ad, P. Adcock, F. R. Ahmad, G. Alppay, J. A. Ball, J. Beach, D. Belhachemi, A. Bellofiore, M. Bellos, J. F. Beltrán, A. Betts, M. W. Bhuiya, K. Blacklock, R. Boer, D. Boisvert, N. D. Brault, A. Buxbaum, S. Caprio, C. Choi, T. D. Christian, R. Clancy, J. Clark, T. Connolly, K. F. Croce, R. Cullen, M. Davey, J. Davidson, M. M. Elshenawy, M. Ferrigno, D. Frier, S. Gudipati, S. Hamill, Z. He, S. Hosali, H. Huang, L. Huang, A. Kabiri, G. Kriger, B. Lathrop, A. Li, P. Lim, S. Liu, F. Luo, C. Lv, X. Ma, E. McCormack, M. Millham, R. Nani, M. Pandey, J. Parillo, G. Patel, D. H. Pike, K. Preston, A. Pichard-Kostuch, K. Rearick, T. Rearick, M. Ribezzi-Crivellari, G. Schmid, J. Schultz, X. Shi, B. Singh, N. Srivastava, S. F. Stewman, T. R. Thurston, P. Trioli, J. Tullman, X. Wang, Y. C. Wang, E. A. G. Webster, Z. Zhang, J. Zuniga, S. S. Patel, A. D. Griffiths, A. M. van Oijen, M. McKenna, M. D. Dyer, J. M. Rothberg, *Science (1979)* **2022**, *378*, 186.

[20] M. Filius, R. van Wee, C. de Lannoy, I. Westerlaken, Z. Li, S. H. Kim, C. de Agrela Pinto, Y. Wu, G.-J. Boons, M. Pabst, D. de Ridder, C. Joo, *Nat Nanotechnol* **2024**, DOI 10.1038/s41565-023-01598-7.

[21] N. Soni, N. Freundlich, S. Ohayon, D. Huttner, A. Meller, *ACS Nano* **2022**, *16*, 11405.

[22] P. Martin-Baniandres, W.-H. Lan, S. Board, M. Romero-Ruiz, S. Garcia-Manyes, Y. Qing, H. Bayley, *Nat Nanotechnol* **2023**, *18*, 1335.

[23] L. Yu, X. Kang, F. Li, B. Mehrafrooz, A. Makhamreh, A. Fallahi, J. C. Foster, A. Aksimentiev, M. Chen, M. Wanunu, *Nature Biotechnology 2023 41:8* **2023**, *41*, 1130.

[24] S. Ohayon, A. Girsault, M. Nasser, S. Shen-Orr, A. Meller, *PLoS Comput Biol* **2019**, *15*, e1007067.

[25] C. V. de Lannoy, M. Filius, R. van Wee, C. Joo, D. de Ridder, *iScience* **2021**, *24*, 103239.

[26] A. Zrehen, S. Ohayon, D. Huttner, A. Meller, *Sci Rep* **2020**, *10*, 15313.

18

[27] M. Yavuz Yiice, A. Erdogan, A. Jonas, A. Kiraz, *Frontiers in Optics 2011/Laser Science XXVII (2011), paper FTuH5* **2011**, FTuH5.

[28] Q. Wang, W. E. Moerner, *ACS Nano* **2011**, *5*, 5792.

[29] A. H. Squires, E. Atas, A. Meller, *PLoS One* **2015**, *10*, 1.

[30] M. Largman-Chalamish, A. Wasserman, A. Silberman, T. Levinson, O. Ritter, S. Berliner, D. Zeltser, I. Shapira, O. Rogowski, S. Shenhar-Tsarfaty, *PLoS One* **2022**, *17*, e0277401.

[31] A. Van Den Bruel, M. J. Thompson, T. Haj-Hassan, R. Stevens, H. Moll, M. Lakhanpaul, D. Mant, *Bmj* **2011**, *342*, d3082.

[32] O. Ramilo, A. Mejías, *Cell Host Microbe* **2009**, *6*, 199.

[33] K. Oved, A. Cohen, O. Boico, R. Navon, T. Friedman, L. Etshtein, O. Kriger, E. Bamberger, Y. Fonar, R. Yacobov, R. Wolchinsky, G. Denkberg, Y. Dotan, A. Hochberg, Y. Reiter, M. Grupper, I. Srugo, P. Feigin, M. Gorfine, I. Chistyakov, R. Dagan, A. Klein, I. Potasman, E. Eden, *PLoS One* **2015**, *10*, e0120012.

[34] I. Srugo, A. Klein, M. Stein, O. Golan-Shany, N. Kerem, I. Chistyakov, J. Genizi, O. Glazer, L. Yaniv, A. German, D. Miron, Y. Shachor-Meyouhas, E. Bamberger, K. Oved, T. M. Gottlieb, R. Navon, M. Paz, L. Etshtein, O. Boico, G. Kronenfeld, E. Eden, R. Cohen, H. Chappuy, F. Angoulvant, L. Lacroix, A. Gervaix, *Pediatrics* **2017**, *140*, e20163453.

[35] C. Papan, A. Argentiero, O. Adams, M. Porwoll, U. Hakim, E. Farinelli, I. Testa, M. B. Pasticci, D. Mezzetti, K. Perruccio, A. Simon, J. G. Liese, M. Knuf, M. Stein, R. Yacobov, E. Bamberger, S. Schneider, S. Esposito, T. Tenenbaum, *J Med Virol* **2023**, *95*, e28113.

[36] C. Papan, A. Argentiero, M. Porwoll, U. Hakim, E. Farinelli, I. Testa, M. B. Pasticci, D. Mezzetti, K. Perruccio, L. Etshtein, N. Mastboim, E. Moscoviz, T. I. Ber, A. Cohen, E. Simon, O. Boico, L. Shani, T. M. Gottlieb, R. Navon, E. Barash, K. Oved, E. Eden, A. Simon, J. G. Liese, M. Knuf, M. Stein, R. Yacobov, E. Bamberger, S. Schneider, S. Esposito, T. Tenenbaum, *Clinical Microbiology and Infection* **2022**, *28*, 723.

[37] M. S. Chesnokov, P. A. Khesina, D. A. Shavochkina, I. F. Kustova, L. M. Dyakov, O. V. Morozova, N. S. Mugue, N. E. Kudashkin, E. A. Moroz, Y. I. Patyutko, N. L. Lazarevich, *PeerJ* **2018**, *2018*, DOI 10.7717/peerj.4915.

[38] R. S. Apte, D. S. Chen, N. Ferrara, *Cell* **2019**, *176*, 1248.

[39] P. Vempati, A. S. Popel, F. Mac Gabhann, *Cytokine Growth Factor Rev* **2014**, *25*, 1.

[40] B. Eymin, A. Boudria, C. Abou-Faycal, in *Molecular Mechanism of Angiogenesis* (Eds.: J.-J. Feige, G. Pages, F. Soncin), Springer, Verlag, France, **2014**, pp. 421–442.

19

[41]   M. Yamakuchi, M. Okawa, K. Takenouchi, A. Bibek, S. Yamada, K. Inoue, K. Higurashi, A. Tabaru, K. Tanoue, Y. Oyama, S. Higashi, C. Fujisaki, H. Kanda, H. Terasaki, T. Sakamoto, Y. Soga, T. Hashiguchi, *PLoS One* **2023**, *18*, DOI 10.1371/journal.pone.0284131.

[42]   M. Kazemi, A. Carrer, S. Moimas, L. Zandonà, R. Bussani, B. Casagranda, S. Palmisano, P. Prelazzi, M. Giacca, L. Zentilin, N. De Manzini, S. Zacchigna, *Cancer Gene Ther* **2016**, *23*, 125.

[43]   S. B. Mamer, A. Wittenkeller, P. I. Imoukhuede, *Scientific Reports 2020 10:1* **2020**, *10*, 1.

[44]   B. Emenike, J. Donovan, M. Raj, *J Am Chem Soc* **2023**, *145*, 16417.

[45]   T. Caliñski, J. Harabasz, *Communications in Statistics* **1974**, *3*, 1.

20
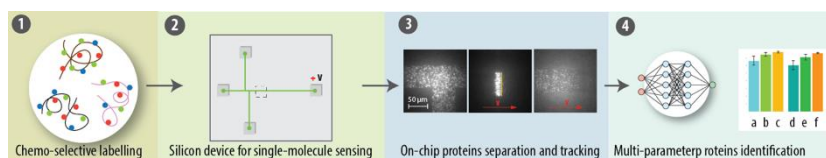
Rapid full-length, single protein molecules separation, tracking and identification in ultra-high aspect ratio silicon devices. Proteins are tagged by chemo-selective dual amino-acid specific labels and are electrophoretically separated by their mass/charge in custom thin silicon channel with subwavelength height. This approach allows the analysis of thousands of individual proteins within a few minutes by tracking their motion during the migration.