Adaptive nanopores: A bioinspired label-free approach for protein sequencing and identification

Andrea Spitaleri[1,2,§], Denis Garoli[3,4,§], Moritz Schütte[5], Hans Lehrach[5,6], Walter Rocchia[1] (✉), and Francesco De Angelis[3] (✉)

[1] CONCEPT Lab, Istituto Italiano di Tecnologia, Via Morego 30, Genova, I-16163, Italy
[2] Center for Omics Sciences, IRCCS San Raffaele Scientific Institute, Milan, Via Olgettina 58, Milano, I-20132, Italy
[3] Plasmon Nanotechnology Unit, Istituto Italiano di Tecnologia, Via Morego 30, Genova, I-16163, Italy
[4] AB ANALITICA s.r.l., Via Svizzera 16, I-35127 Padova, Italy
[5] Alacris Theranostics GmbH, Max-Planck-Strasse 3, D-12489 Berlin, Germany
[6] Max Planck Institute for Molecular Genetics, Ihnestrasse 63-73, D-14195 Berlin, Germany
[§] Andrea Spitaleri and Denis Garoli contributed equally to this work.

ABSTRACT

Single molecule protein sequencing would tremendously impact in proteomics and human biology and it would promote the development of novel diagnostic and therapeutic approaches. However, its technological realization can only be envisioned, and huge challenges need to be overcome. Major difficulties are inherent to the structure of proteins, which are composed by several different amino-acids. Despite long standing efforts, only few complex techniques, such as Edman degradation, liquid chromatography and mass spectroscopy, make protein sequencing possible. Unfortunately, these techniques present significant limitations in terms of amount of sample required and dynamic range of measurement. It is known that proteins can distinguish closely similar molecules. Moreover, several proteins can work as biological nanopores in order to perform single molecule detection and sequencing. Unfortunately, while DNA sequencing by means of nanopores is demonstrated, very few examples of nanopores able to perform reliable protein-sequencing have been reported so far. Here, we investigate, by means of molecular dynamics simulations, how a re-engineered protein, acting as biological nanopore, can be used to recognize the sequence of a translocating peptide by sensing the "shape" of individual amino-acids. In our simulations we demonstrate that it is possible to discriminate with high fidelity, 9 different amino-acids in a short peptide translocating through the engineered construct. The method, here shown for fluorescence-based sequencing, does not require any labelling of the peptidic analyte. These results can pave the way for a new and highly sensitive method of sequencing.

KEYWORDS

nanopores, single molecule sequencing, protein sequencing, luorescence resonance energy transfer (FRET), amino-acids, fluorescence

1   Introduction

The "sequence–structure–function" paradigm for proteins embodies one of the biggest scientific discoveries about the functioning of Life. The DNA sequence in each gene contains the information necessary to build proteins, via a mechanism where specific triplets of nucleic acids code for the different amino-acids. In addition to this golden rule for translation, proteins' fundamental building blocks can be altered by alternative splicing or post-translational modifications. Mutations in specific positions can greatly impact on protein structure and function. Therefore, the identification of protein sequence can significantly impact our comprehension of Life and our ability to prevent, diagnose, and treat most of the human pathologies. At present, a limited number of methods are available for protein sequencing and the standard technology is still based on mass spectrometry [1, 2]. The major intrinsic drawbacks of these methods are the limit of detection and the dynamic range [3]; in particular, with them it's not possible to achieve single-molecule resolution. The protein concentration in humans spans from picograms/mL up to milligrams/mL. Consequently, the large amount of molecules required to satisfy the detection limits of standard mass spectrometers make the analysis of low concentration proteins extremely challenging. To overcome these limitations and to pursue the final aim of single-cell analysis, during the last few years single-molecule protein sequencing methods started to emerge [4]. Current efforts for protein sequencing and identification are inspired from the outstanding results achieved in DNA sequencing [5]. Unfortunately, the complex nature of proteins makes single molecule sequencing not trivial at all. Difficulties arise from different sides. First, the sequence is constituted of 20 different building blocks, i.e. the amino-acids, against only 4 bases which compose DNA. Second, amino-acids do not have complementary bases as in the case of nucleotide and, consequently, there are no polymerase chain reaction-like amplification methods. These

---

facts make the detection of 20 distinguishable signals from the different amino-acids tremendously challenging, regardless of the read-out method of choice. Some exciting ideas have been recently reported and the reader can find a comprehensive review on approaches to single-molecule protein sequencing in Restrepo-Perez et al. [4]. The methods reported in the literature are based on single molecule techniques as nanopores [6–13], fluorescence [14] and tunneling currents across nanogaps [15, 16]. Every method has its own advantages and drawbacks. Regarding optical methods, which are based on "color discrimination", it is likely impossible to distinguish about 20 different fluorescence spectra. Similarly, electrical readouts, which reflect charge flux through a pore partially occupied by the analyte, would hardly discriminate so many current levels spanning a narrow range of values. Only very recently, two reports demonstrated, by means of numerical simulations, that optimized solid-state [11] and biological nanopore [17] are able to distinguish among high numbers of amino-acids. Other recent works proposed to take advantage of comprehensive protein sequence databases which enable the identification of a protein by labelling and detecting a limited number of amino-acids [14, 18]. By knowing this limited subset of the whole sequence is then possible to select the corresponding protein within the available database. However, these approaches cannot identify specific post-translational modifications or point mutations, which is a fundamental need of proteomics and personalized medicine.

Hence, despite the outstanding results achieved for sequencing DNA and recent brilliant works on single molecule protein detection and sequencing [4, 7–10, 14, 17, 19–21], the translation of these approaches to whole sequencing of individual proteins, especially if available in a small amount, appears still a tremendous challenge. In parallel with the current efforts it is necessary to conceive radically new ways to approach the problem and to reconsider it from a different point of view. As it often happens, "Nature" can inspire us: actually, biomolecular recognition participates every biological process and one can say it is at the basis of life. Biomolecular recognition is mostly performed by proteins, which evolved to perform this fascinating ability. Recognition, and subsequent binding, are regulated by a delicate balance of opposing phenomena, such as the desolvation penalty occurring upon binding and the direct electrostatic interaction, which is stronger in absence of the screening aqueous medium [22]. Molecular interactions can be extremely specific thus enabling to accomplish highly selective recognition tasks.

By taking inspiration from these long-studied phenomena, here we propose and investigate the concept of an "adaptive biological nanopore" as the foundation for a novel class of devices able to discriminate the amino-acidic sequence of a translocating polypeptide. While this protein construct is not a conventional nanopore, we chose this term since the proposed scheme is analogue to that of other nanopores, such as α-hemolysin etc., in which DNA can flow through [23]. However, the sensing mechanism is radically different. We show, by means of molecular dynamics (MD) simulations, that a properly designed protein construct working as a pore can dynamically "re-shape" according to the specific amino-acidic sequence that is flowing through it. Then, by monitoring the spatial rearrangements of the construct one can infer the sequence of the translocating amino-acids. Based on this behavior, we named it "adaptomorphic pore" or simply "adaptive nanopore". By taking advantage of MD simulations, we designed a prototypical adaptive pore and we investigated its interactions with simple poly-peptides. Machine learning analysis confirmed that the deformations occurring to the nanopore are specific

for each single amino-acid with a full precision (recall precision equal to 1). In the second part of the work, we propose a potential experimental method for tracking the spatial re-arrangement of the pore by means fluorescence resonance energy transfer (FRET) spectroscopy. Even though the precision obtained from classification decreases to an average value of 0.59 and 0.42 for more realistic four- and two-channel setups, respectively, the method still appears robust, feasible and with a lot of room for improvements.

2 Methods

The concept of an adaptive pore and its use in the present context can be better conveyed if one thinks that the question: "What is the shape of a molecule?" does not have an unambiguous answer as the concept of shape is intrinsically connected to that of interaction and there are several levels of interaction occurring at the molecular scale.

Probably the most intuitive one is that of steric interaction, one can represent each atom as a hard sphere, with a radius that could be defined as the distance of closest approach for another atom. Based on this simple description and by using a similar representation also for the solvent probe, back in 1971 it was defined the "solvent excluded surface" (SES) of a molecule immersed in a solvent [24]. An example is shown in Fig. 1 (left panel) for a sequence of 9 different amino acids, namely: ARGinine, GLutamiNe, GLUtamic acid, GLYcine, HIStidine, IsoLEucine, SERine, TRiPtophan, and TYRosine.

The SES (left panel, semitransparent boundaries) encloses the space region where a solvent probe, represented as a hard sphere, cannot enter because it would otherwise overlap with some of the van der Waals (vdW) spheres representing a peptide, or a protein, atom. Both vdW and SES definitions rely on the intuitive and simple ideas of contact and impenetrability of two solid bodies. However, in addition to the contact, direct electrostatic forces, such as charge–charge and charge–dipole play a fundamental role in specific molecular interactions [25] and therefore also in the definition of "shape". A graphical representation of the electrostatic potential iso-surface is given in Fig. 1 (right panel). These representations make clear that, according to the considered types of interaction, each amino acid has a different shape both with respect to the others and to itself in the other dimension. Interestingly, current simulative approaches, such as for instance molecular dynamics, can use intermolecular interactions to drive the time evolution of a possibly very complex molecular system and, especially, to sample different conformational states consistently with fundamental thermodynamic control variables such as temperature and pressure. In these simulations, the estimation of
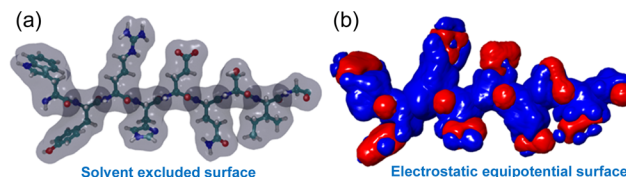
Figure 1 Different interaction fields lead to a multifaceted concept of shape. (a) Ball-and-stick representation of a polypeptide containing 9 different amino acids (ACE-WYRHEQSIG-NME, where ACE and NME are acetyl and N-methyl groups, respectively). The semitransparent surface set the boundaries of the region excluding the presence of the solvent described as a spherical probe of 1.4 Å radius. (b) Electrostatic equipotential surfaces at $+k_B T/e$ and $-k_B T/e$ values, generated by the same amino-acid sequence. Electrostatic force provides a different kind of "shape", which is more selective for different chemical entities (here k_B is the Boltzmann's constant, T the absolute temperature assumed to be at 273 K, and e is the unit charge).

macroscopic observables is done together with the analysis of mechanisms at the atomistic level [26]. Among the most relevant types of interaction for the present purposes, we would like to mention solvent-mediated ones, which play a central role in determining molecular behavior.

To substantiate the proposed concept of an adaptive biological pore, we chose a molecular construct originating from the modification of the 2-Cys Prx I protein from Schistosoma mansoni. SmPrxI is a member of the family of 2-Cys peroxiredoxins (PRXs), enzymes with redox and molecular chaperone activity [27]. Under physiological conditions, ten SmPrxI monomeric subunits assemble into ring-like complexes with thickness, internal and external diameters of 5, 6 and 13 nm, respectively (see Fig. 2). Importantly, SmPrxI shows features that are analogue to those of synthetic biological nanopores that are consolidated approaches for molecular translocations and analyses [23, 28, 29]. Among other aspects, its structure allows for the translocation of a polypeptide through its central channel and it can be easily arranged and bound to a solid-state substrate [30]. The SmPrxI-based construct was designed to fulfill the following main requirements: i) obtaining a set of structurally stable molecular probing "tips" pointing towards the center of the internal hole of the pore that could possibly act as sensors and anchor points for further functionalization; ii) allowing the expression of the biomolecular system via standard molecular biology means, avoiding *ad hoc* chemical modifications. As shown in Fig. 2, the N-terminal of each monomer was prolonged with a 16aa-long sequence inspired to the well-known three-dimensional Leucine zipper structural motif [21]. Details about how the elongation has been designed can be found in Section S2 in the Electronic Supplementary Material (ESM). We included in this added sequence, which has a strong propensity for the helical structure, two cysteine residues. One cysteine residue is located at the new N-terminal so that it can form a disulfide bridge with the corresponding residue of the closest adjacent monomer, stabilizing the leucine zipper structure. The second cysteine residue is located around the middle of the helix, and serves as a possible anchor point for chemical functionalization.
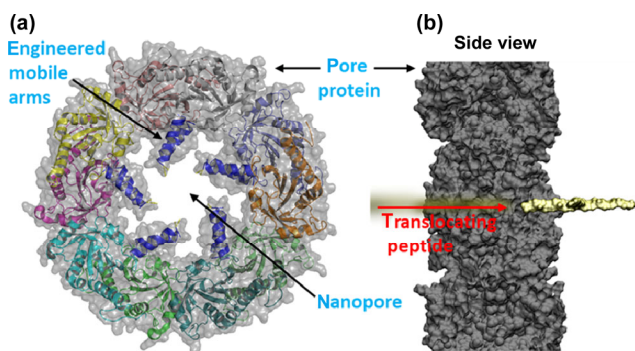
Figure 2 Protein construct embodying the concept of an automorphic pore. The structure is a modification of the SmPrxl member of the peroxiredoxin family. (a) Modified SmPrxl protein. The N-terminal of each monomer of the decamer is prolonged with a helical stretch extracted from the "leucine zipper motif", to stabilize the interaction with the companion stretch on the adjacent monomer. The helices pointing to the center of the pore are meant to be especially sensitive to the passing of an analyte through the pore. Numbers indicate the distances between the tips, in angstroms. (b) Side view of the construct. In yellow, a polypeptidic analyte during its translocation.

3   Results and discussion

To ascertain whether this kind of construct can act as a suitable adaptive nanopore, we investigated its interaction with short

poly-peptides that are translocating along its axis, using MD simulations (details can be found in Section S1 in the ESM). Important to be noted, protein translocation velocity does not have typical values but strongly depends on the translocation strategy and on the protein itself. One of the most crucial difficulties in sequencing via a nanopore is controlling the translocation of the analyte, in our case the polypeptide. Recently, several studies on this topic have been reported. It has been demonstrated that in nanopore-based setups proteins can be unfolded, polypeptides can be stretched, and the translocation can be controlled to some extent by means of external driving forces. Significant examples where a protein, such as ClpX, is used as unfoldase can be found in Refs. [6, 14]. Another strategy makes use of positive and negative tails added at the two peptide ends to pull and stretch the peptide by means of an applied voltage [31]. Sodium dodecyl sulfate (SDS) has also been used to denature the protein and facilitate its electrophoretic translocation [7, 32]. A further denaturation and translocation strategy is reported in Ref. [33].

As a proof of concept, we considered the effect of 9 different endeca-peptides having the following composition, GGGGGXGGGGG, where G stands for the Glycine amino acid and X can be one of the amino acids listed in Table 1. Then, we built 10 simulation boxes, one for each peptide plus a reference with no peptide in it, where the forces between the modified SmPrxl and a peptide located in the middle of its pore were studied for 1 μs. In order to account for the different possible orientations of the side chain of the central amino acids with respect to the pore, on which one has no experimental control, we divided the 1 μs simulation in 8 slots of 125 ns, each of them starting from a different orientation. The orientation is defined as the rotation angle of the Cα–Cβ bond around the axis of the pore. A residence time of 125 ns per residue is consistent with translocation velocities reported in Refs. [6, 7, 31, 33–39].

3.1   Construct deformation analysis

As a measure of the deformation of the construct due to the presence of the endeca-peptide, the equilibrium position of the 5 tips pointing to the center of the pore was chosen. Their locations were represented by the center of the disulfide bridge between the cysteine residues located at the N-terminal of each monomer of the modified SmPrxl protein (see Fig. 2(a)). Equilibrium positions were estimated by averaging along the afore-mentioned eight simulation slots, each of them being 125ns long. While more elaborated, and possibly better performing, approaches could be envisioned, such as averaging only over the more stable configurations observed during each slot, we preferred for this proof of concept a simpler and less user-dependent criterion. The descriptors were analyzed via two machine-learning approaches, namely support-vector machine (SVM) and random forest (RF), to build classifiers that, by

Table 1   List of the simulated translocating peptides

Endeca-peptide
[GLY]$_5$-TRP-[GLY]$_5$
[GLY]$_5$-TYR-[GLY]$_5$
[GLY]$_5$-ARG-[GLY]$_5$
[GLY]$_5$-HIS-[GLY]$_5$
[GLY]$_5$-GLU-[GLY]$_5$
[GLY]$_5$-GLN-[GLY]$_5$
[GLY]$_5$-SER-[GLY]$_5$
[GLY]$_5$-ILE-[GLY]$_5$
[GLY]$_5$-GLY-[GLY]$_5$

observing the deformation of the modified SmPrxl protein, can identify which of the 9 endeca-peptides was located in the pore, and to evaluate the discriminative power of this inference process. For sake of brevity, details regarding the simulations and data analyses are reported in Section S3 in the ESM.
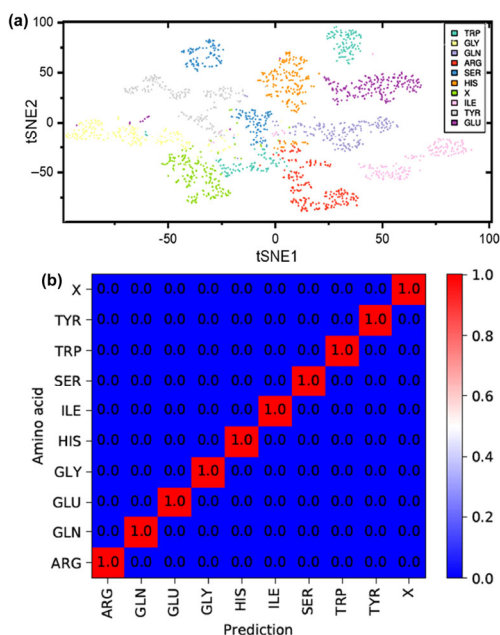
The main results of the data analysis are shown in Fig. 3. They clearly corroborate our hypothesis that a different amino-acidic composition induces a different deformation of the construct. The graphical representation (Fig. 3(a)) clearly suggests a robust discrimination power. Training of a support-vector machine confirmed this by resulting in perfect discrimination ability (precision and recall of 1) for all amino acids on the test set. Importantly, such scores are achieved simulating individual molecules thus making this method promising for reaching single-molecule sensitivity with no need of amplification.

3.2 Deformation transduction

As anticipated, one possible mean to track pore deformation and transduce it into a measurable signal is FRET spectroscopy. To this aim, a proper functionalization of the pore is necessary [30, 40]. Here, we suggest exploiting a second cysteine anchor, added on purpose to each monomer of the construct, and to attach to it a linker ending with a fluorescent dye. FRET signal intensity depends on the sixth power of the spatial separation and relative orientation between two molecular dyes, called donor and acceptor (D and A, respectively).

In fact, the dependence of FRET efficiency E on the D–A distance R follows this functional behavior: $[1 + (R/R_0)^6]^{-1}$, where R_0 is the Forster radius, which also depends on the specific dye pair and their relative orientation. This strong dependence makes FRET able to recognize spatial variations of the D–A positions with a sensitivity of few Angstroms in space and few μs in time [41]. A suitable functionalization of the nanopore with a set of DA pairs enables to dynamically correlate FRET emission variations to the nature of the amino-acids located in the pore center, exploiting the deformation that follows the passage of the analyte.
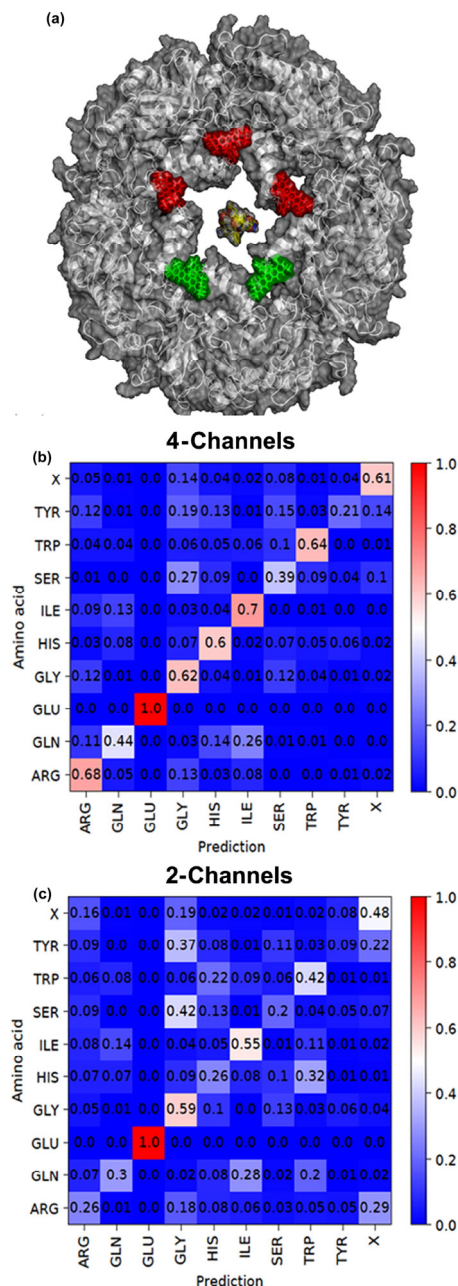
The multimeric structure of the protein construct allows conceiving the functionalization of the construct with 5 FRET

dyes (2 donors and 3 acceptors). While the individual measurements of all the possible interactions between these dyes are practically unfeasible, we investigated two intermediate cases as proofs of concept.

In the first, we consider two donor/acceptor dyes which can mutually interact via homo-FRET and can act as donor for the other 3 acceptor dyes, which are assumed to have distinguishable emission spectra. We will call this arrangement 4-channel system. The second differs from the first only for the fact that the emission spectra of the acceptors are not distinguishable. We call this arrangement 2-channel system and one can envision its realization by a prototype that consists of 2 adjacent ATTO 425 fluorescent dyes, which can act both as donors and acceptors, followed by 3 ATTO 647N acceptor dyes (see Fig. 4).

Figure 3 (a) Discrimination using tSNE dimension reduction performed over the average positions of the 5 tips of the construct along the molecular dynamics simulation. (b) Confusion matrix describing the discriminative capability of the tips positions using support vector machine classification. X describes the case where no peptide is located in the nanopore.

Figure 4 Detecting construct example. (a) Representation of the construct, in grey the molecular surface of the pore construct, in red the three acceptor dyes, in green the two donor/acceptor dyes, in the center the poly-peptidic analyte. (b) and (c) Confusion matrices representing the classification capability of the 4-channel and 2-channel systems, respectively. *Y*-axis shows the inserted amino acid in the simulation and *x*-axis the predicted amino acid on the test set. Numbers on the grid indicate the percentage of each amino acid-prediction combination.

In this way the homo-FRET between the two ATTO 425 would coexist with the regular FRET between each 425-647N pair, providing a 2-channel emitting system. The choice of two rather rigid types of dye is instrumental to maximize their sensitivity to the deformation of the construct due to the passage of the polypeptide through the channel, therefore reducing background noise. The Forster radius of both FRET pairs, 3.6 and 4.3 nm, respectively, is suited for the diameter of the channel, around 6 nm. Details on dyes and linkers characterization and simulation are presented in Section S3 in the ESM.

The same MD simulations used to ascertain the adaptive response of the modified SmPrxI were also analyzed to estimate the discriminant capability of the signals emitted by these multi-channel systems. This task required the computational identification of the equilibrium positions of the dyes, as well as the corresponding low energy conformations of the linkers. This is known to be a significant challenge, due to the long relaxation times that characterize their dynamics, which hamper a sufficient exploration in a MD run [42]. We addressed this issue using specifically accelerated sampling approaches that proved effective in solving similar problems the field of Structure Based Drug Design. Related details, which led to the starting conformation for our simulations are described in Section 3 in the ESM, while the MD analysis is detailed in Section S4 in the ESM. The estimate of the per couple FRET efficiency based on the MD simulation was done via the md2fret analysis tool [43].

The overall performance of the 4-channel was obtained by considering the signals obtained by the homo-FRET couple and these emitted by the 3 acceptor dyes. This led to the discrimination shown in Fig. 4, with an average precision and recall of 0.59. In the 2-channel system, where the ATTO-425 homo-FRET and the combined efficiency of the superposed FRET signals emitted by the three 647N dyes are considered (see Section S6 in the ESM), precision and recall decrease to 0.42.

In both cases, GLU shows perfect discrimination, followed by ARG and ILE. Analysis of these results suggests that, as expected, the greatest sensitivity is provided by size and charge. It is interesting to note that in the simulated construct the sequence of the arms that have been added to the original SmPrxI protein has not yet been optimized for performance, leaving plenty of room for improvement.

4   Conclusions

In conclusion, by exploiting molecular dynamics simulations and machine learning, we showed that a properly designed protein construct can act as a nanopore able to recognize the "shape" of individual amino-acids composing a polypeptide that translocates through the pore. The experimental scheme is analogue to that used with other biological and solid-state nanopores, but the sensing method is radically different and exploits the specificity of the interactions which is at the basis of the recognition and functional binding in biomolecular systems. During the translocation process, the atomistic inter-actions between the polypeptide and the construct cause a reshaping of the nanopore. The latter will therefore assume transient spatial configurations that are specific to the amino acids flowing through. Hence, the information regarding the sequence is transferred, or indeed temporarily copied, from the peptide to the nanopore with an ideal accuracy approaching 100%. This information can be transduced into readable signals by conjugating, for instance, the pore with well-defined dyes acting as DA couples for FRET spectroscopy. Our

analyses of calculated FRET signals showed that the accuracy of identification decreases to an average of 0.6 among the considered amino-acids for the 4-channel system and 0.42 for the 2-channel system. In the present work a proof of concept is presented, where one among several different possible configurations is considered and some assumptions are made. In a practical implementation, many factors, that have only been sketched here, should be carefully evaluated. Among these factors there are the rigidity of the tips, the correct functionalization, and the effect of noise on the accuracy of the measurements. However, this also means that there is a lot of room for improvements with respect to the construct presented here, including the possibility to explore other schemes for transduction. For instance, being the experimental setup similar to that of most used biological nanopores (such as α-hemolysin, MspA, etc. [29]), the FRET recording could be combined with electrical read-outs thus potentially increasing the discrimination power. We want to emphasize that these numbers come from the MD and ML analyses of few/single translocation events. It does suggest that a real system has the capability of discriminating the sequence by carrying out experiments with few/single molecules. In other words, single molecule sensitivity can be reached. In light of the lack of amplification technologies, the latter is an essential feature for enabling single cell technologies and personalized medicine protocols. Importantly, the proposed method is label free.

In summary, in light of system optimization and further developments, the reported data clearly suggest that the approach may lead to protein sequencing platform with per-formance comparable to those of current methods for DNA sequencing still working in label free conditions and with few molecule sensitivity.

References

[1] Steen, H.; Mann, M. The abc's (and xyz's) of peptide sequencing. *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 699–711.

[2] Shimonishi, Y.; Hong, Y. M.; Kitagishi, T.; Matsuo, T.; Matsuda, H.; Katakuse, I. Sequencing of peptide mixtures by edman degradation and field-desorption mass spectrometry. *Eur. J. Biochem.* **1980**, *112*, 251–264.

[3] Domon, B.; Aebersold, R. Options and considerations when selecting a quantitative proteomics strategy. *Nat. Biotechnol.* **2010**, *28*, 710–721.

[4] Restrepo-Pérez, L.; Joo, C.; Dekker, C. Paving the way to single-molecule protein sequencing. *Nat. Nanotechnol.* **2018**, *13*, 786–796.

[5] Ameur, A.; Kloosterman, W. P.; Hestand, M. S. Single-molecule sequencing: Towards clinical applications. *Trends Biotechnol.* **2019**, *37*, 72–85.

[6] Nivala, J.; Marks, D. B.; Akeson, M. Unfoldase-mediated protein translocation through an α-hemolysin nanopore. *Nat. Biotechnol.* **2013**, *31*, 247–250.

[7] Kennedy, E.; Dong, Z. X.; Tennant, C.; Timp, G. Reading the primary structure of a protein with 0.07 nm³ resolution using a subnanometre-diameter pore. *Nat. Nanotechnol.* **2016**, *11*, 968–976.

[8] Kolmogorov, M.; Kennedy, E.; Dong, Z. X.; Timp, G.; Pevzner, P. A. Single-molecule protein identification by sub-nanopore sensors. *PLoS Comput. Biol.* **2017**, *13*, e1005356.

[9] Wilson, J.; Sloman, L.; He, Z. R.; Aksimentiev, A. Graphene nanopores for protein sequencing. *Adv. Funct. Mater.* **2016**, *26*, 4830–4838.

[10] Asandei, A.; Rossini, A. E.; Chinappi, M.; Park, Y.; Luchian, T. Protein nanopore-based discrimination between selected neutral amino acids from polypeptides. *Langmuir* **2017**, *33*, 14451–14459.

[11] Farimani, A. B.; Heiranian, M.; Aluru, N. R. Identification of amino acids with sensitive nanoporous MoS₂: Towards machine learning-based prediction. *npj 2D Mater. Appl.* **2018**, *2*, 14.

[12] Ohayon, S.; Girsault, A.; Nasser, M.; Shen-Orr, S.; Meller, A. Simulation of single-protein nanopore sensing shows feasibility for whole-proteome identification. *PLoS Comput. Biol.* **2019**, *15*, e1007067.

[13] Rosen, C. B.; Rodriguez-Larrea, D.; Bayley, H. Single-molecule site-specific detection of protein phosphorylation with a nanopore. *Nat. Biotechnol.* **2014**, *32*, 179–181.

[14] Van Ginkel, J.; Filius, M.; Szczepaniak, M.; Tulinski, P.; Meyer, A. S.; Joo, C. Single-molecule peptide fingerprinting. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 3338–3343.

[15] Zhao, Y. N.; Ashcroft, B.; Zhang, P. M.; Liu, H.; Sen, S. M.; Song, W. S.; Im, J.; Gyarfas, B.; Manna, S.; Biswas, S. et al. Single-molecule spectroscopy of amino acids and peptides by recognition tunnelling. *Nat. Nanotechnol.* **2014**, *9*, 466–473.

[16] Ohshiro, T.; Tsutsui, M.; Yokota, K.; Furuhashi, M.; Taniguchi, M.; Kawai, T. Detection of post-translational modifications in single peptides using electron tunnelling currents. *Nat. Nanotechnol.* **2014**, *9*, 835–840.

[17] Ouldali, H.; Sarthak, K.; Ensslen, T.; Piguet, F.; Manivet, P.; Pelta, J.; Behrends, J. C.; Aksimentiev, A.; Oukhaled, A. Electrical recognition of the twenty proteinogenic amino acids using an aerolysin nanopore. *Nat. Biotechnol.* **2020**, *38*, 176–181.

[18] Swaminathan, J.; Boulgakov, A. A.; Marcotte, E. M. A theoretical justification for single molecule peptide sequencing. *PLoS Comput. Biol.* **2015**, *11*, e1004080.

[19] Venkatesan, B. M.; Bashir, R. Nanopore sensors for nucleic acid analysis. *Nat. Nanotechnol.* **2011**, *6*, 615–624.

[20] Hu, R.; Tong, X.; Zhao, Q. Four aspects about solid-state nanopores for protein sensing: Fabrication, sensitivity, selectivity, and durability. *Adv. Healthc Mater.*, in press, DOI: 10.1002/adhm.202000933.

[21] Huang, J. A.; Mousavi, M. Z.; Giovannini, G.; Zhao, Y. Q.; Hubarevich, A.; Soler, M. A.; Rocchia, W.; Garoli, D.; De Angelis, F. Multiplexed discrimination of single amino acid residues in polypeptides in a single SERS hot spot. *Angew. Chem., Int. Ed.* **2020**, *59*, 11423–11431.

[22] Sheinerman, F. B.; Norel, R.; Honig, B. Electrostatic aspects of protein–protein interactions. *Curr. Opin. Struct. Biol.* **2000**, *10*, 153–159.

[23] Cao, C.; Long, Y. T. Biological nanopores: Confined spaces for electrochemical single-molecule analysis. *Acc. Chem. Res.* **2018**, *51*, 331–341.

[24] Lee, B.; Richards, F. M. The interpretation of protein structures: Estimation of static accessibility. *J. Mol. Biol.* **1971**, *55*, 379-400, IN3-IN4.

[25] Honig, B.; Nicholls, A. Classical electrostatics in biology and chemistry. *Science* **1995**, *268*, 1144–1149.

[26] Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: Oxford, 2017.

[27] Angelucci, F.; Miele, A. E.; Ardini, M.; Boumis, G.; Saccoccia, F.; Bellelli, A. Typical 2-Cys peroxiredoxins in human parasites: Several physiological roles for a potential chemotherapy target. *Mol. Biochem. Parasitol.* **2016**, *206*, 2–12.

[28] Wang, H. Y.; Li, Y.; Qin, L. X.; Heyman, A.; Shoseyov, O.; Willner, I.; Long, Y. T.; Tian, H. Single-molecule DNA detection using a novel SP1 protein nanopore. *Chem. Commun.* **2013**, *49*, 1741–1743.

[29] Ying, Y. L.; Cao, C.; Long, Y. T. Single molecule analysis by biological nanopore sensors. *Analyst* **2014**, *139*, 3826–3835.

[30] Giovannini, G.; Ardini, M.; Maccaferri, N.; Zambrana-Puyalto, X.; Panella, G.; Angelucci, F.; Ippoliti, R.; Garoli, D.; De Angelis, F. Bio-assisted tailored synthesis of plasmonic silver nanorings and site-selective deposition on graphene arrays. *Adv. Opt. Mater.* **2020**, *8*, 1901583.

[31] Asandei, A.; Chinappi, M.; Lee, J. K.; Seo, C. H.; Mereuta, L.; Park, Y.; Luchian, T. Placement of oppositely charged aminoacids at a polypeptide termini determines the voltage-controlled braking of polymer transport through nanometer-scale pores. *Sci. Rep.* **2015**, *5*, 10419.

[32] Restrepo-Pérez, L.; John, S.; Aksimentiev, A.; Joo, C.; Dekker, C. SDS-assisted protein transport through solid-state nanopores. *Nanoscale* **2017**, *9*, 11685–11693.

[33] Cressiot, B.; Braselmann, E.; Oukhaled, A.; Elcock, A. H.; Pelta, J.; Clark, P. L. Dynamics and energy contributions for transport of unfolded pertactin through a protein nanopore. *ACS Nano* **2015**, *9*, 9050–9061.

[34] Chinappi, M.; Luchian, T.; Cecconi, F. Nanopore tweezers: Voltage-controlled trapping and releasing of analytes. *Phys. Rev. E* **2015**, *92*, 032714.

[35] Pastoriza-Gallego, M.; Breton, M. F.; Discala, F.; Auvray, L.; Betton, J. M.; Pelta, J. Evidence of unfolded protein translocation through a protein nanopore. *ACS Nano* **2014**, *8*, 11350–11360.

[36] Kim, H. J.; Choi, U. J.; Kim, H.; Lee, K.; Park, K. B.; Kim, H. M.; Chi, S. W.; Lee, J. S.; Kim, K. B. Translocation of DNA and protein through a sequentially polymerized polyurea nanopore. *Nanoscale* **2019**, *11*, 444–453.

[37] Mereuta, L.; Roy, M.; Asandei, A.; Lee, J. K.; Park, Y.; Andricioaei, I.; Luchian, T. Slowing down single-molecule trafficking through a protein nanopore reveals intermediates for peptide translocation. *Sci. Rep.* **2014**, *4*, 3885.

[38] Plesa, C.; Kowalczyk, S. W.; Zinsmeester, R.; Grosberg, A. Y.; Rabin, Y.; Dekker, C. Fast translocation of proteins through solid state nanopores. *Nano Lett.* **2013**, *13*, 658–663.

[39] Rodriguez-Larrea, D.; Bayley, H. Multistep protein unfolding during nanopore translocation. *Nat. Nanotechnol.* **2013**, *8*, 288–295.

[40] Ayub, M.; Bayley, H. Engineered transmembrane pores. *Curr. Opin. Chem. Biol.* **2016**, *34*, 117–126.

[41] Roy, R.; Hohng, S.; Ha, T. A practical guide to single-molecule FRET. *Nat. Methods* **2008**, *5*, 507–516.

[42] Dimura, M.; Peulen, T. O.; Hanke, C. A.; Prakash, A.; Gohlke, H.; Seidel, C. A. M. Quantitative FRET studies and integrative modeling unravel the structure and dynamics of biomolecular systems. *Curr. Opin. Struct. Biol.* **2016**, *40*, 163–185.

[43] Hoefling, M.; Grubmüller, H. In silico FRET from simulated dye dynamics. *Comput. Phys. Commun.* **2013**, *184*, 841–852.