IOP Publishing Journal of Physics: Condensed Matter
J. Phys.: Condens. Matter **30** (2018) 204002 (14pp) https://doi.org/10.1088/1361-648X/aababe

Protein sequencing via nanopore based devices: a nanofluidics perspective

Mauro Chinappi[1] **and Fabio Cecconi**[2]

[1] Dipartimento di Ingegneria Industriale, Università di Roma Tor Vergata, via del Politecnico 1, 00133 Roma, Italy
[2] Istituto dei Sistemi Complessi (CNR), Via dei Taurini 19, 00185 Roma, Italy

E-mail: mauro.chinappi@uniroma2.it and fabio.cecconi@roma1.infn.it

Received 25 February 2018, revised 18 March 2018
Accepted for publication 29 March 2018
Published 25 April 2018

Abstract

Proteins perform a huge number of central functions in living organisms, thus all the new techniques allowing their precise, fast and accurate characterization at single-molecule level certainly represent a burst in proteomics with important biomedical impact. In this review, we describe the recent progresses in the developing of nanopore based devices for protein sequencing. We start with a critical analysis of the main technical requirements for nanopore protein sequencing, summarizing some ideas and methodologies that have recently appeared in the literature. In the last sections, we focus on the physical modelling of the transport phenomena occurring in nanopore based devices. The multiscale nature of the problem is discussed and, in this respect, some of the main possible computational approaches are illustrated.

Keywords: single-molecule sensing, electrohydrodynamics, biomolecule translocation, nanofluidics, peptides

(Some figures may appear in colour only in the online journal)

1. Introduction

Nanopores displayed a great versatility in the framework of bio-analytical applications, as they can work as single-molecule sensors able to detect, analyze and even manipulate nanoscale constructs. The working principle of nanopore sensing system is quite simple. A nanopore is embedded in a membrane that separates two chambers of a micro/nanofluidic device. The molecules to be analyzed are added to the solution in one of the chambers. When a single molecule interacts with the nanopore, e.g. translocating through it or bumping its entrance, it alters one, or more, properties of the system that can be recorded by appropriate instruments. In the last years, nanopore sensors have emerged as powerful, alternative tools for the detection and the analysis of various chemical compounds and biomolecules, such as RNA, DNA, peptides and proteins at single molecule level. The most relevant success of nanopore sensing is the DNA sequencing. A nanopore based device, commercialized few years ago, represented a turning point in the nucleic acid sequencing for its portability and capability to get long reads (hundreds of kilobases) [1]. The next obvious step in the research, that apparently seemed at hand, was adapting nanopore sensing to another class of important macromolecules: the proteins. This goal has proven to be more challenging than expected. The present review aims at presenting the state of the art of the nanopore protein sequencing field and at discussing the open issues from the side of physical modelling of transport phenomena at the nanoscales.

Nanopores employed for single molecule analysis can have different sizes and shapes. The length of biological channels is comparable to (or slightly larger than) the thickness of the lipid membrane in which they are embedded, $\sim$4 nm, while solid state nanopores in Silicon Nitride (SiN) are usually drilled in 10–20 nm thick substrates, see figures 1(c)–(g). Regardless of the actual pore length, a crucial element is the effective length of the so called sensing region. Indeed, although usually several monomers can engage the pore at the same time, only few of them notably affect the measurement. This is particularly relevant for techniques involving the pore conductance measurement (resistive pulse approach), where, the electrical field is more intense at the narrowest section of

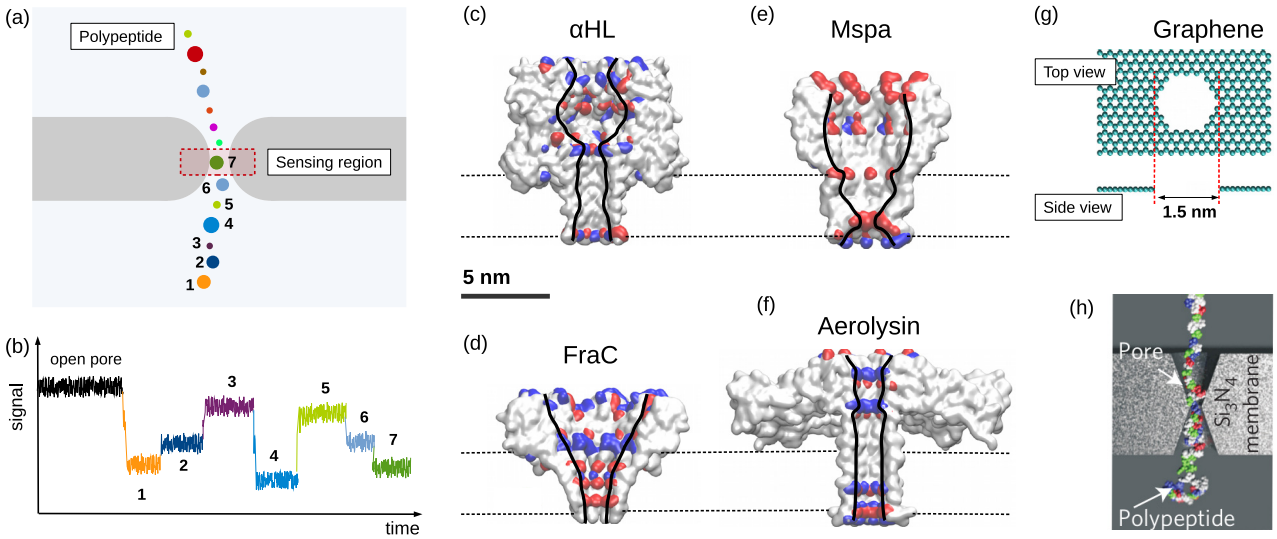

Figure 1. Nanopore sensing. ((a)–(b)) The interaction of a molecule with a pore induces variations in the system properties (i.e. ionic conductance of the pore, transverse current) that can be recorded by appropriate instruments. Generally, the signal is mainly influenced by the interaction of the molecule with a specific region of the pore, termed sensing region. Ideally, for sequencing purposes the size of the sensing region has to be comparable to monomer size. When amino acids enter the pore sequentially, the signal trace will show a series of steps each one of them corresponding to a single monomer. ((c)–(h)) Examples of nanopores. In the resistive pulse approach, the sensing region coincides with the pore constriction, as the electric field is higher in that section. For 2D nanopores the sensing region is the whole pore and, in principle, both transverse current along the membrane and ionic current through the pore can be simultaneously measured. Panel (h) is reprinted by permission from Macmillan Publishers Ltd: Nature Nanotechnology [2], copyright 2016. Crystal structure for Mspa, αHL, FraC and Aerolysin are taken from OPM database [3] based on the PDB entries 1UUN [4], 7AHL [5], 4TSY [6] and 5JZT [7]. VMD [8] software was used to draw panels (c)–(g). Positive and negative residues (at physiological pH) are indicated in blue and red, respectively.

the pore and, consequently, despite the number of monomers that can occupy the pore is quite large (a 10 nm pore has room for ∼30 amino acids) the signal is dominated by the relatively small number of monomers in the constriction. For instance, for the Mspa channel, the estimated sensing region is ∼0.6 nm and, for the DNA, the signal is due to four adjacent nucleotides only [9]. The same argument applies to solid state nanopores drilled in SiN membranes, where the pore has a hourglass shape that focuses the electric field inside its narrowest section [10]. Remarkably, 2D substrates such as graphene [11–14] or Molybdenum disulfide, MoS_2, [15, 16] sheets have sub-nanometer sensing regions able, in principle, to accommodate just a single monomer.

The presence of the molecule in the sensing region can affect different properties of the pore that can be potentially used to discriminate the residues. The two most explored possibilities are the ionic current flowing through the pore (resistive pulse) and the transverse tunneling current along the substrate plane [17–19]. In addition, other approaches have been recently proposed, such as measurement of ionic transverse currents [20] and of mechanical stress along the membrane [21] and combined optical-electrical recording [22].

In this review, we focus on the resistive pulse that, to date, is the most popular technique. Nevertheless, some discussions (e.g. translocation and capture control) can be, in principle, extended to other sensing strategies. The typical set-up in a resistive sensing is reported in figure 2. A single nanopore connects the two chambers of an electrolytic cell. When a voltage ΔV is applied across the two chambers, positive and negative ions migrate towards negative and positive electrodes, respectively, generating an electric current I. If the pore surface is charged, the positive and negative ion distributions differ. In particular, the pore walls attract ions of opposite charge (counterions) and repel ions of the same charge (coions). In this condition, the unbalance between the positive and the negative charge fluxes gives raise to a net water motion: the electroosmotic flow. When a macromolecule interacts with the pore, the above scenario is drastically altered. The main mechanism is the reduction of the effective pore cross section, resulting in a decrease of the current (current blockade). In addition, the charges on the molecule surface perturb the ion distributions, and, accordingly, modify the electrohydrodynamics close to the pore. Figure 2 sketches a resistive pulse set-up for a positively charged pore for which the negative ion flux $\mathbf{J}^-$ is more intense than the positive one $\mathbf{J}^+$. Consequently, the electroosmotic velocity $\mathbf{u}$ is opposite to the electrical field $\mathbf{E}$. Beyond electrohydrodynamical effects, the scenario is more complex as macromolecules can chemically interact with the pore.

Resistive pulse nanopore based approaches have been employed for several protein and peptide analysis such as determination of the folding state [23, 24], analysis of structural changes [25–27] and protein or peptide aggregation [28, 29]. To date, despite the efforts of the wide and heterogeneous community, a nanopore based protein sequencing protocol is still not available.

The two main ideal requirements that a nanopore sequencing device has to fulfill are (i) *signal-to-monomer matching*; each signal has to be unambiguously associated to a specific position in the protein sequence and (ii) *distinguishability*; the

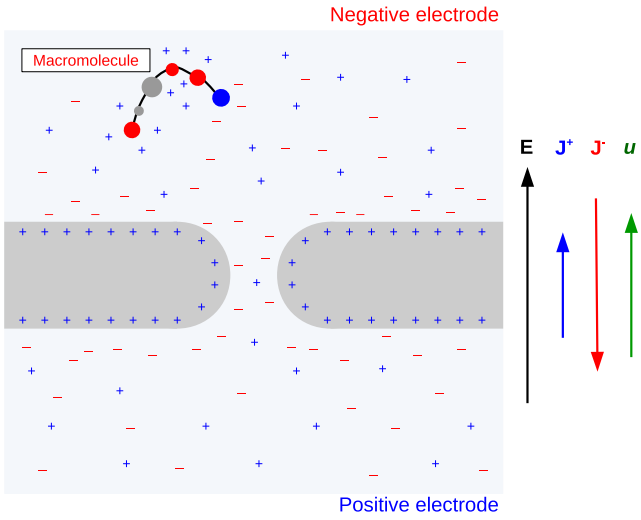

Figure 2. Resistive pulse. The nanofluidic device for resistive pulse is constituted by two chambers of an electrolytic cell which are connected by a single nanopore. Molecules to be analyzed are released in one of the chambers. In general, surface charges are present at the pore walls, hence, the ion distribution close to the nanopore is not homogeneous. An applied external voltage ΔV generates a non-homogeneous electrical field $\mathbf{E}$. Assuming an isolating membrane, the electrical field lines converge into the pore and, the magnitude of $\mathbf{E}$ is larger inside the pore. In such an electroosmotic system, several transport phenomena occurs, such as: (i) positive (negative) ions migration toward the negative (positive) electrode resulting in an average ions flux $\mathbf{J}^+$ ($\mathbf{J}^-$) (ii) each ion drags its surrounding water shell, hence, if the ionic fluxes are unbalanced ($\mathbf{J}^- \neq \mathbf{J}^+$) a net water flux of velocity $\mathbf{u}$ (electroosmotic flow, EOF) sets in. The figure reports the case of a positively charged pore so that EOF direction is opposed to the external electrical field. (iii) macromolecules can move under the effect of $\mathbf{E}$ (electrophoresis or dielectrophoresis) and under electroosmotic flow (Stokes drag). Moreover, macromolecules can chemically interact with the pore surface. When macromolecules are close to the pore, they strongly modify the ionic and the electroosmotic flux, either directly by steric hindrance or indirectly by affecting the ion distribution near the pore.

signal level associated to a single amino acid (AA) has to allow the unambiguous identification of the AA. The next two sections will review the recent results and the proposed approach for both signal-to-monomer matching and distinguishability. Then, we briefly discuss the protein capture and the translocation control, while the last section is dedicated to the possible theoretical and computational approach to the study of transport phenomena relevant for nanopore protein sequencing.

2. Signal-to-monomer matching

The unambiguous correspondence between a specific interval of the signal trace and a single amino acid (AA) in a protein chain is a desirable requirement for protein sequencing. An ideal scenario would consider the protein placed in one of the two sides of the electrolytic chamber which unfolds and sequentially translocates towards the other side. Once the unfolded protein engages the pore, the translocation needs to be strictly progressive, that is, a monomer that left the sensing region, should not come back. The translocation speed has to be as uniform as possible allowing, in principle, to segment the current trace into regular intervals corresponding to the different AAs.

Several issues make this ideal scenario quite challenging to be achieved. First, proteins are not uniformly charged, thus electrophoresis cannot be straightforwardly used to control translocations. Proteins with net (positive or negative) charge can enter the pore and, eventually, translocate under the action of an external field. Nevertheless, the inhomogeneous charge does not guarantee that the translocation speed is uniform and that the same residue does not cross the sensing region back and forth several times. Second, different portions of the same protein chain can interact with the pore with different affinities making the translocation rate not constant even for a constant drive.

An interesting method to overcome these obstacles was proposed by Kennedy *et al* [2]. They employed sodium dodecyl sulfate (SDS), an anionic compound that, in combination with heat and reducing agents (e.g. BME β-mercaptoethanol) is able to denaturate the protein providing a negative charged shell to its backbone. Although the exact structure of the SDS-protein aggregate is not available, a recent computational and experimental work [32] indicates that SDS induces a considerable loss of the protein structure and that SDS-treated proteins translocate through the nanopore in the direction prescribed by the applied electrical field. Kennedy *et al* [2] showed that current trace occurring in a sub-nanometer solid state nanopore, figure 3(a), presents a number of peaks that is close to the number of AAs of the analyzed protein, see figures 3(b) and (c). This evidence suggests that proteins actually translocated across the pore and that the ionic current fluctuations carry information on the protein sequence. More specifically, the authors related the current blockade to the volume of quadromers (four consecutive AAs in the sequence) and developed an algorithm to infer the protein sequence from the current trace. Although the proposed approach is not sensitive enough to discriminate among all the AAs it can be successfully employed for protein identification [33].

Another recent proposed strategy amounts to using protein-pore interaction to control translocation. Stepwise translocation of polypeptides inside graphene nanopores has been analyzed via atomistic simulation by Wilson *et al* [30]. The authors show that hydrophobic residues of an unfolded polypeptide systematically adhere to a graphene sheet. After adhesion, charged peptides can be unidirectionally driven inside the nanopore by a voltage bias ΔV, while a pressure bias is needed for translocating neutral peptides. Figure 3(d) displays a snapshot of the simulated system. The pore substrate is a three-layer graphene sheet and the peptide is constituted by a regular repeat of a hydrophobic AA (Phenylalanin, F, cyan) and a negatively charged AA (Aspartic Acid, D, red). The time evolution of the number of AAs transported through the pore by an electric field is reported in figure 3(e). The results of [30] suggest that a proper calibration of the ratio between hydrophobic and charged residues can be used to tune the translocation speed. Such a calibration can be achieved by altering the pH of the solution that would result

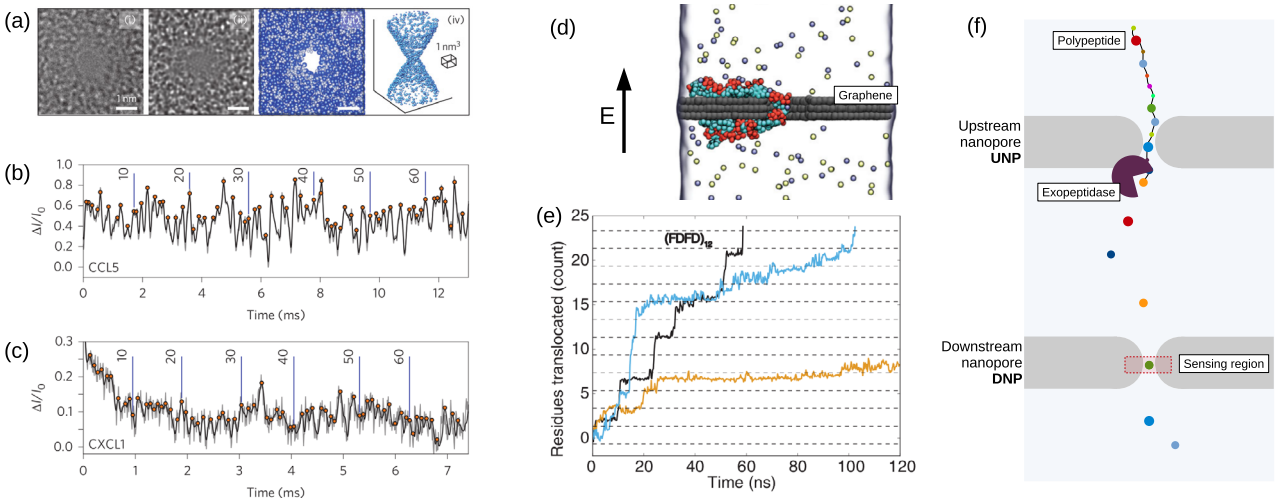

Figure 3. Signal-to-monomer matching and translocation control. (a) Nanopore employed in Kennedy *et al* [2], nominal diameter 0.7 nm: TEM images (a_i), multislice simulations (a_{ii}), 2D projections (a_{iii}) and 3D representations (a_{iv}). (b) and (c) Current blockage traces $\Delta I/I_0 = (I_0 - I)/I_0$, where I_0 indicates the open pore current associated to the translocation of CCL5 (68 AA peptide chain) and CXCL1 (71 AAs). The grey trace represent the raw data, the black line is the smoothed data and the orange circles are the peaks in the trace. The numbers of peaks are $N_{CCL5} = 67$ and $N_{CXCL1} = 69$ in good agreement with the number of AAs in the peptides. The images were reprinted by permission from Macmillan Publishers Ltd: Nature Nanotechnology [2], Copyright 2016. (d) Sketch of the simulation set-up from Wilson *et al* [30]. A polypeptide translocate through a graphene nanopore. Grey: three layer graphene substrate; cyan: Phenylalanin; red: Aspartic Acid; spheres: K^+ and Cl^- ions (only one of ten is reported). Water molecules are not shown. (e) Time evolution of the number of amino acids transported through the pore in presence of an electric field E for three simulation replicas. Adapted from [30] John Wiley & Sons. © 2016 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim. (f) Tandem pore scheme proposed by Sampath [31], the peptide chain enters the upstream pore (UNP) and it is sequentially cut by an exopeptidase enzyme. The cut and isolated AAs then engage the downstream pore (DNP).

in a change of the AA charges without affecting the hydrophobic/hydrophilic nature of the residues. Recently, the same principle was tested via all-atom molecular dynamics simulation for a Molybdenum Disulphide (MoS_2) nanopore [16]. Wilson *et al* [30] simulations also allowed to partially associate the current trace with the identities and the conformations of residues inside the pore. This hints that stepwise translocation of an unfolded peptide can produce a ionic current carrying information on the AA sequence. This positive outcome is currently challenged by two observations reported in [30]: (i) different current levels can be associated with the same sequence of residues inside the pore but in different conformations and (ii) the number of residues inside the pore is not constant. Both these issues can be perhaps solved with a proper tailoring of the nanopore (e.g. a smaller pore might accommodate a quite constant number of residues and reduce the number of possible conformations).

An alternative strategy to preserve the chain order of translocating monomers is to sequentially cut single AAs from one protein terminal and then to import them one by one in a nanopore where the electric signal is measured. This approach, inspired by the exonuclease DNA nanopore sequencing method of [34], was recently theoretical analyzed by Sampath [31]. The proposed device is sketched in figure 3(f). The idea is based on a tandem nanopore system where a first nanopore (upstream nanopore, UNP) unfolds the protein and pushes it toward an exopeptidase that cuts the last AA, that, in turn, is imported in a second pore (downstream nanopore, DNP) where the current signal is recorded. Although recently deviced double nanopore set-up [35] and double-barrel capillary system [36] can potentially alleviate the technical difficulties in fabricating the tandem cell device proposed by Sampath [31], practical obstacles make the above approach very challenging. First, it relies on the existence of exopeptidases able to cut residues regardless their identity. Then, the different diffusivities and charges of AAs do not guarantee that the translocation from UNP to DNP preserves the chain order. The latter issue could be tackled by tuning the pH of the solution to confer charges of the same sign to all the AAs, or by generating a net solvent flow, for instance, using pressure gradients or electroosmosis. Another limiting factor lies in the long-tail nature of the arrival time statistics of the monomers at the entrance of the second nanopore. In this respect, the theoretical analysis by Maulbetsch *et al* [37] on the probability of Brownian particle reshuffling can be useful in selecting suitable parameters.

As a final comment, we would like to mention several techniques to control and to slow down translocation that, although still not directly applied to protein sequencing, can be likely adapted in the near future. Methods to increase the residence time in the pore include the employment of a pressure-voltage biased pore [38], the control of the balance between electrophoretic and electroosmotic flow [39–41], the employment of protein unfoldase [42], plasmonic trapping [43], DNA capture and transport in heated nanopores [44].

3. Distinguishability

Nanopore protein sequencing strictly requires that each AA has to be unambiguously associable to a specific current level. To clarify which are the practical issues implied by this requirement, we consider the following ideal situation. In an

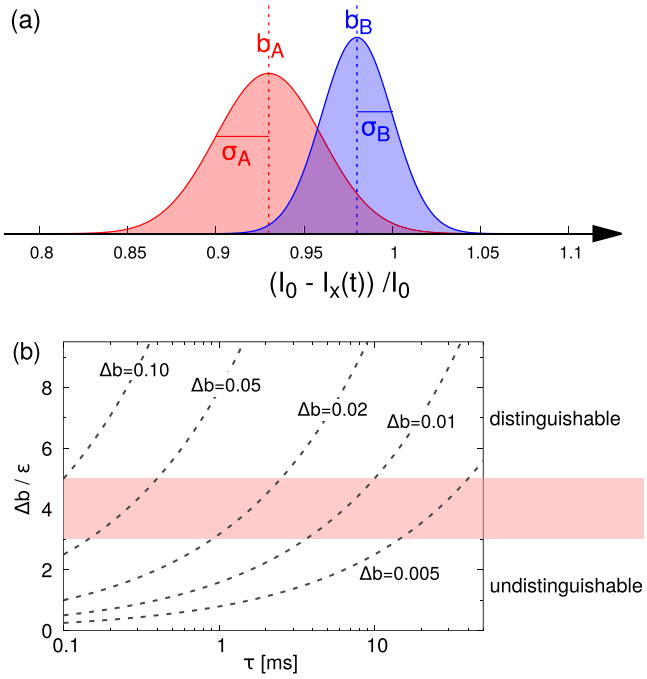

Figure 4. Distinguishability. (a) Current blockade distribution for two monomers. Red (blue) curve corresponds to $b_{A(B)} = 0.93$ (0.98), $\sigma_{A/(B)} = 0.03$ (0.025). (b) Ratio between the current blockade difference for two residues, Δb, and its statistical error, ε, as a function of the residence time τ from the simplified model discussed in section 3, equation (2). Each dashed line corresponds to a different value of Δb. The value of $\Delta b/\epsilon$ provides a first indication on the capability of the sensing system to distinguish between two amino acids: low $\Delta b/\epsilon$, indistinguishable, high $\Delta b/\epsilon$ distinguishable. The red shaded area corresponds to an intermediate range where the amino acid calling would depend on the statistical threshold selected for significance. The curve is drawn at a sampling frequency $f = 10$ KHz and a relative noise $\sigma = 0.02$, that are typical values for α-hemolysin.

experiment, the current trace is a time series $I(t_i)$, sampled with frequency f at times $t_i = i/f$. Let $I_x(t_i)$ indicate the current trace when the residue x engages the sensing region. We define

$$b_x(t_i) = \frac{I_0 - I_x(t_i)}{I_0} \tag{1}$$

the relative current blockade trace due to monomer x, with I_0 the average open pore current. In the following, b_x and σ_x will indicate the average and standard deviation of equation (1), respectively.

In general, the probability distributions of $b_x(t_i)$ for two different residues may overlap, see figure 4(a). In this respect, the distinguishability question can be addressed as follows: 'Given amino acids A and B, what are the values of b_A, b_B and σ_A, σ_B, for which A and B can be distinguished?' This issue cannot be solved without taking into account the residence times τ_A, τ_B that residues A and B spend into the pore sensing region (residence times).

To answer this question, we need to estimate the statistical error of the average current blockade b_x that reads $\epsilon_x = \sigma_x/\sqrt{\tau_x f}$, assuming $\tau_x f$ independent samples. For the sake of simplicity, we take $\sigma_A = \sigma_B = \sigma$ and $\tau_A = \tau_B = \tau$. Two residues are distinguishable if the ratio of the difference $\Delta b = |b_A - b_B|$ between their average current blockades over the statistical error ε lies above a certain threshold. Figure 4(b) reports the ratio

$$\frac{\Delta b}{\epsilon} = \frac{\Delta b}{\sigma}\sqrt{\tau f} \tag{2}$$

as function of the residence time τ for different values of Δb where we assumed $f = 10$ KHz, $\sigma \simeq 0.02$ (i.e. the noise is 2% of the open pore signal), that are typical values for an experimental $\alpha-$hemolysin set-up [45]. The dashed lines correspond to different values of Δb. The criterion for distinguishability amounts to choosing a given threshold for $\Delta b/\epsilon$ according to the following rule of thumb: generally, for $\Delta b/\epsilon < 3$ averages are not considered significantly different; for $\Delta b/\epsilon > 5$ distinction is considered certain, while, in between, any claim might be ambiguous. For example, with reference to figure 4(b), taking a reference time $\tau = 10$ ms (that would allow to sequence a single protein in few seconds) we find that AAs are distinguishable when $\Delta b \geqslant 0.01$, i.e. difference in the current levels of just 1% of the open pore current can be resolved. For a ten time faster sequencing, residence time $\tau = 1$ ms, $\Delta b \geqslant 0.03$. These values pose challenges concerning the actual possibility to sequence proteins with biological pores by resistive pulse approach. Indeed, also in the ideal case in which the sensing region is occupied only by a single residue, the average current blockade b_x of all the 20 AAs should be sufficiently different. This could be particularly challenging for the isomers Leucine and Isoleucine (same formula but different structure) that can give rise to very similar current blockades, also for small residues (e.g. Alanine, Glycine and Serine) that are expected to provide weak signals.

As a final comment, we point out that, in general, the rigorous formulation of the distinguishability problem is more complex than what reported here. An interesting discussion can be found in [20].

3.1. Amino acid fingerprinting

A first step to assess the protein sequencing capability of a specific pore is determining the average current blockade b_x for each AA. A route to get this preliminary information, is to set-up experiments where a stretched homopeptide, i.e. a chain composed by a repeat of the same AA, is trapped inside the pore sensing region for a residence time long enough to lower the error on the estimation of the current blockade b_x. A strategy to achieve this goal was recently proposed by Asandei *et al* [46]. The approach requires a certain polarity of the molecule, a condition that can be achieved, for instance, by attaching a positive and a negative tail to the ends of a neutral chain resulting in a tri-block molecule, see figure 5(a). The capture of the molecule is favored by the artificial dipole which tends to align along the field lines focused into the pore. Indeed, as soon as the molecule approaches the pore, it experiences a gradually increasing importing force due to the larger electrical field. When also the other charged tail enters inside the pore, an opposite couple of forces generates a sort

of 'tug-of-war' resulting in a stable trapping of the central part of the molecule into the pore. Another convenient effect of the artificial polarity is the stretching of the chain along the pore axis. Since the intensity of the electrical field can be easily controlled by changing ΔV, both capture and escape rates can be properly tuned. A crucial point of this technique is that the residence time increases with ΔV. In term of the above discussion on distinguishability, working with larger ΔV has a twofold advantage: on the one hand, it enhances the signal to noise ratio, on the other hand, it increases the number of samples as the residence time becomes longer. Last, but not least, also the capture rate increases with ΔV.

The experimental proof of principle by Asandei *et al* [46] used two polypeptides made of three consecutive blocks of 12 identical monomers each, namely $(E)_{12}-(N)_{12}-(R)_{12}$ and $(E)_{12}-(Q)_{12}-(R)_{12}$, where E,N,R and Q indicate Glutamin Acid, Asparagine, Arginine and Glutamine, translocating in an αHL pore. The central portion, $(N)_{12}$ or $(Q)_{12}$, is neutral while the two ends, $(E)_{12}$ and $(R)_{12}$, carry opposite charges at physiological pH. The average current blockades b_x for the two peptides were found different beyond the statistical error: $b_N = 0.092 \pm 0.002$ and $b_Q = 0.097 \pm 0.001$. More recently, a similar experiment [47] was repeated with other peptides where the central neutral residues were Alanine and Triptophan. Also in this case, significant differences in current blockage b where associated to the different peptides [47]. These results were also supported by atomistic simulations. A detailed analysis of the current trace revealed a richer behavior: for both Ala- and Trp-peptides, the current trace shows a characteristic two-level signal. The interpretation of the two-level fingerprint and its possible application to peptide recognition are currently open issues.

Interestingly, very recently Piquet *et al* [48] showed that Aerolysin nanopore is able to discriminate between two different ten-residue long homopeptides made by Arginine (R) and Lysine (K) and one heteropolymeric peptides, (K)_5-(R)_5). The authors also show that the same set-up allow a very precise accurate size-discrimination between Arginine homopeptides of different length mixed in solution.

4. Capture rate

The control of the capture rate from the bulk to the pore is a further challenge in nanopore protein analysis. Two main effects contribute to the molecule transport: diffusion and external forcing. To assess which is the prevailing one, it is customary to introduce the Péclet number as

$$Pe = \frac{\tau_d}{\tau_f} \tag{3}$$

where

$$\tau_d = \frac{l^2}{D} \quad , \quad \tau_f = \frac{l}{v_f} \tag{4}$$

are the diffusive and forcing time scales with D the protein diffusion coefficient, v_f the velocity induced by the external forcing and l the scale distance between the protein position and the pore entrance.

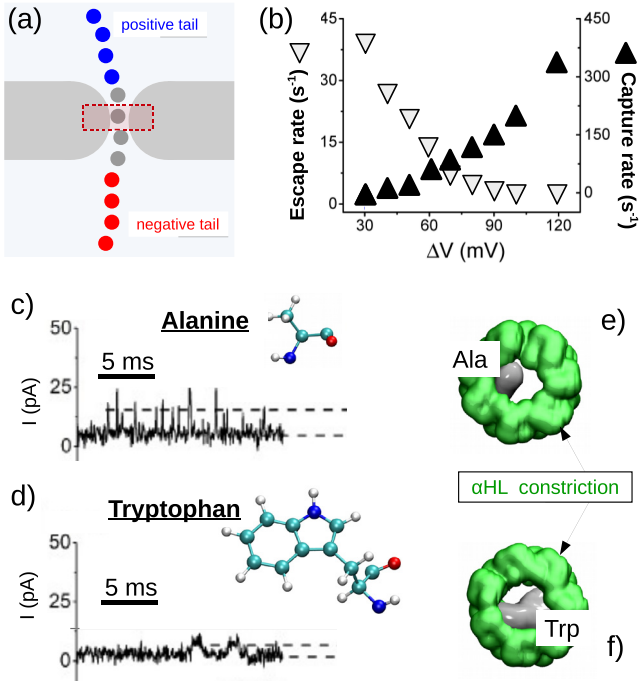

Figure 5. Amino acid fingerprinting. (a) A neutral homopeptide can be trapped inside a nanopore by adding a positive and a negative tail (nanopore tweezer, dielectrophoretic trapping [46, 49]). (b) For such a construct, capture rate (filled upper triangles) increases with ΔV while escape rate (empty lower triangles) decreases. Hence, at high ΔV the residence time increases. The data refer to the 36AA polypeptide $(E)_{12}(N)_{12}(R)_{12}$ and are adapted from Asandei *et al* [46]. ((c), (d)) The same approach was also used for the 30AA peptides $(E)_{12}(A)_6(R)_{12}$ and $(E)_{12}(W)_6(R)_{12}$ where the central region is constituted by Alanine (A) or Tryptophan (W). The current blockade results more intense for W ($b_W \simeq 97\%$) with respect to Alanine ($b_A \simeq 94\%$). Interestingly, the current trace of both A and W presents two different levels that, in principle, could be employed to set-up additional analysis for characterizing the residues. All atom molecular dynamics simulations suggest that the main contribution to current blockade is the clogging of the sensing region. The panels (e)–(f) show a top view of the αHL constriction (green region) where the grey surfaces correspond to the 3 residues (A or W) closer to it. Panels (c)–(f) are adapted with permission from [47]. Copyright 2017 American Chemical Society. VMD software was used for panels (e)–(f) [8].

For electrophoretic forcing, the particle velocity can be estimated as

$$v_f = \mu q E \tag{5}$$

where q is the protein charge, μ its mobility and E the magnitude of the electrical field. It is worth mentioning that, for high conducting electrolytes, q is not the bare charge as screening effects have to be taken into account. This topic goes beyond the aim of the present review. Equation (5) is obtained by balancing the viscous and electrical forces acting on the molecule. If the membrane can be assumed isolating, the streamlines of the electrical field $\mathbf{E}(\mathbf{x})$ focus into the pore and, as a consequence, away from it, the magnitude of E decreases with l^2, see figure 6. As a first approximation, considering the pore entrance as a spherical electrode [50], we have

$$E \simeq \frac{\Delta V}{L}\frac{d^2}{l^2}\,, \tag{6}$$

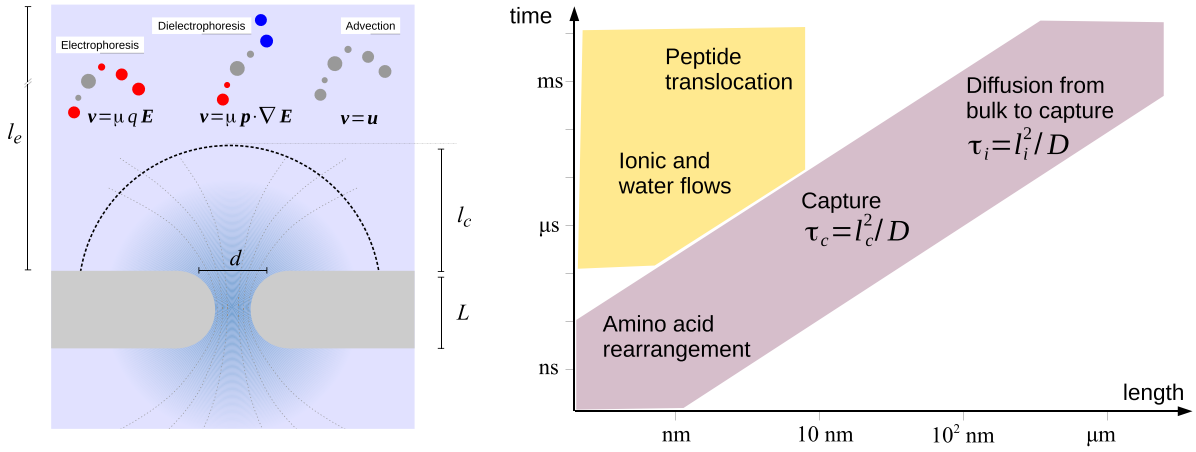

Figure 6. Time and length scales. (a) Far from the pore, external forces are small and the molecule motion is driven by Brownian diffusion. Once a molecule enters the capture region (size l_c), a net force, due to electro-phoresis, dielectrophoresis and/or fluid advection, brings the molecule close to the pore. In this region, the dynamics is the result of a competition between forcing (that attracts the molecule towards the pore) and diffusion. (b) Typical time and length scales for the different transport phenomena involved in protein sequencing. The rearrangements of the single amino acids in the pore sensing region occur at molecular scale (ns and nm). The ionic currents flowing through the pore are usually sampled at decades/hundreds of KHz, so the typical time scale to get a stable signal is $\sim 10\ \mu s \div 1$ ms, while a complete translocation of an unfolded protein requires several milliseconds [2]. Capture scale l_c is decades of nm (see text) and the associated time scale is $1 \div 10\ \mu s$, while the time scale between captures can be estimated as $\tau_b = l_i^2/D$, where l_i is the inter-molecule distance in the bulk, see section 5.

d and L being the diameter and length of the pore[3]. Hence, the Péclet number reads

$$Pe = \frac{d^2 \mu q}{Dl}\frac{\Delta V}{L} = \frac{qd^2}{l k_B T}\frac{\Delta V}{L}\,, \tag{7}$$

where in the last equality we employed the fluctuation-dissipation relation $D = \mu k_B T$ with k_B the Boltzmann constant and T the temperature. Far enough from the pore (large l) $Pe \ll 1$, thus the transport is dominated by diffusion. Near the pore $Pe \gg 1$ and the transport become dominated by the driving forcing. The crossover between these two regimes is found by setting $Pe \simeq 1$ then defining the size l_c of the capture region as

$$l_c \simeq \frac{qd^2}{k_B T}\frac{\Delta V}{L}\,. \tag{8}$$

The corresponding time scale can be estimated as $\tau_d = l_c^2/D$. As an example, for an $\alpha-$hemolysin pore, $d \simeq 2$ nm, $L \simeq 10$ nm, and considering $\Delta V = 100$ mV and a charge of $q = 10e$ (with e the electron charge), we obtain a capture length $l_c \simeq 13$ nm.

A similar argument can be repeated also when the main external drive is not electrophoretic, but, dielectrophoretic or advective transport. For dielectrophoresis, the average velocity induced by a applied (not-uniform) electrical field $\mathbf{E}$ on a molecules of dipole $\mathbf{p}$ can be approximatively expressed as [51]

$$\mathbf{v_f} \simeq \mu \mathbf{p}\cdot\nabla\mathbf{E}\,. \tag{9}$$

The dipole can be either permanent, the molecule is intrinsically polar, or induced by the external field (dielectric polarization). In the following, only permanent dipoles are considered, therefore, the size of the capture region is

$$l_c \simeq d\sqrt{\frac{p}{k_B T}\frac{\Delta V}{L}}\,, \tag{10}$$

where $p = |\mathbf{p}|$ and $|\mathbf{p}\cdot\nabla\mathbf{E}| \simeq pE/l_c$. The value of l_c strongly depends on the dipole moment of the molecule. Just to set the order of magnitude, for a single AA, assuming the two terminals were both charged (one positive and one negative), $p \simeq 0.5 \times 10^{-28}$ Cm and for $\Delta V = 100$ mV, $d = 2$ nm and $L = 10$ nm, the capture region size is $l_c \simeq 5$ nm.

Molecules can be also advected towards the pore by a net water flow generated either by electroosmosis or by an applied pressure gradient. As inertia is negligible, the particle velocity $\mathbf{v_f}$ swiftly settles to the flow velocity $\mathbf{u}$. Indicating the volumetric flow rate as Q, $|\mathbf{u}|$ goes as Q/l^2, thus, imposing $Pe \simeq 1$, the size of the capture region can be estimated as

$$l_c \simeq \frac{Q}{D}\,. \tag{11}$$

For a spherical particle of diameter $d = 5$ nm, the diffusivity amounts to $D = 3\pi\nu\rho d \simeq 10^{-11}$ m^2 s^{-1} with μ and ρ being the kinematic viscosity and density of water. Using as a reference value for electroosmotic flow $Q \simeq 3 \times 10^{-19}$ m^3 s^{-1} from [52][4], we get $l_c \simeq 6$ nm.

[3] Equation (6) is derived assuming that (i) the electrical field intensity inside the pore is $E = \Delta V/L$ a condition that is strictly valid only if access resistance are neglected and the pore section is constant, (ii) the pore entrance is a spherical electrodes of diameter d and (iii) outside the pore, the field lines are radial. In addition, equation (6) is valid only 'far' from the pore, i.e. for $l > d$ [49].

[4] The value is calculated for $\Delta V = 100$ mV using the data reported in [52] for the order of magnitude of the electroosmotic conductance.

In several cases, for protein capture, all the three mentioned effects are relevant and can compete or cooperate each other. An example of competition between electroosmosis and electrophoresis was recently presented in [41] where the electroosmosis was shown to induce peptide capture against electrophoresis.

A small capture region implies low capture rates and, consequently, from a practical point of view, it means that the protein concentration in solution has to be high enough to enhance the number of translocation events. A way to widen the capture length l_c is by increasing the effective protein charge q. In this respect, the method proposed in the already mentioned work by Kennedy *et al* [2], section 2, based on the the employment of SDS molecules, is particularly promising. Similarly, chemical modifications of chain termini by adding charged tails, as shown by Rodriguetz-Larrea and Bayley [53] and Nivala and Akeson [42] are able to confer a net charge to the proteins which allows the capture of even folded structures. In this scenario, the click addition of a single strand DNA tail to a peptide terminal introduced by Biswas *et al* [54], can potentially play a crucial role in the field. In principle, the Biswas *et al* [54] technique can be employed for adding a positive tail (e.g. a Hystidine or Lysine tail) on the other peptide termini creating the strong dipole needed for nanopore tweezer approach [46, 49] presented in section 3.1. The Biswas *et al* [54] technique can also be applied to the 'selectivity tags' method recently proposed by Hoogerheide *et al* [55], where two regions of different but uniform charge density at the ends of the polypeptide enable the discrimintion between polypeptide translocation and retraction.

5. Theoretical modelling

Resistive pulse involves several concomitant mass and charge transport phenomena occurring at different time and length scales. The discussion of section 4 showed that the external driving is relevant only in a small region near the pore (capture region). Hence, far from the pore, the motion is dominated by diffusion. Since diffusion is a too slow process, in practical situation, relatively high values of protein concentration are used to enhance the capture rate. For instance, for μM concentration, the average inter-molecule distance turns to be $l_i = \rho_m^{-3} \simeq 10^2$ nm with ρ_m the number density of the proteins, leading to a time scale $\tau_b = l_i^2/D \simeq 0.01 \div 1$ ms, where we used $D = 10^{-9} \div 10^{-11}\mathrm{m}^2\ \mathrm{s}^{-1}$ [56]. On the contrary, when the molecule is in the capture region, the dynamics is dominated by the direct transport (electrophoresis, dielectrophoresis or advection) that contrasts diffusion. Once the protein engaged the pore, chemical and local electrical interactions become so strong that the dynamics is ruled by atomistic details acting on scales of ns and nm.

Figure 6(b) reports the different phenomena on a time-scale length-scale chart. The violet shaded region includes all phenomena that somehow are accessible to the simulation of the dynamics upon choosing a proper time and space scale resolution (atomistic, coarse grained or continuous). In the most commonly employed classical atomistic force fields [57–59] the shortest time and length scales that need to be solved are fractions of *picoseconds* and *Angstroms*. To date, it is possible to (relatively easily) perform all-atom simulations for systems of characteristic size of few decades of *nanometers* for decades/hundreds of *nanoseconds*. For instance, in the simulations by Wilson *et al* [30], discussed in section 2, the system has a hexagonal cell of inner radius $\simeq$5 nm and height 13 nm and the typical run covers 250 ns. Similarly, during the capture stage, where atomistic details are less relevant, one can simplify the molecule as a rigid-body (or a bead-and-spring chains) where the tiniest relevant length scale is the size of the molecule (or the bead). Also in this case, it is possible to explore the full time evolution and catching the interesting information.

The computational approach is more critical in the yellow region of the chart where the phenomena still involve nanometer scale but occur at time scales which are not accessible to the atomistic simulations. In such conditions, one-to-one matching between experiments and simulations is not possible and other approaches are needed. Remarkably, the yellow area just includes effects that are fundamental for protein sequencing. In the following, we present a discussion on three commonly employed approaches for the description of nanopore systems and their recent applications to protein sequencing.

5.1. Atomistic simulation

In a typical classical all-atom molecular dynamics (MD) simulation of a nanopore system, the membrane containing the nanopore separates two liquid reservoirs filled by water and ions. For computational reasons, periodic boundary conditions are used in x, y and z direction and, the application of a voltage difference ΔV between the two reservoirs is replaced by a constant and homogeneous electrical field of intensity $E = \Delta V/L_z$, with L_z the size of the period box perpendicular to the membrane. Although these computational setting does not strictly correspond to the experimental conditions, a wide literature showed that crucial quantities, such as the open pore conductance [52, 60, 61] reasonably match with experimental data.

Early simulations of ions transport through nanopore dates back to the beginning of the century. Crozier *et al* [62] carried out simulations of ion transport through a model nanopore, while Aksimentiev *et al* [63] and Yeh and Hummer [64] performed the first simulations of nucleic acid transport through solid state nanopores. The ability of MD to provide molecular interpretations of experimental data [32, 65–67] and to propose new ideas in nanopore sensing [21, 30, 68], can be hardly underestimated. Beside the above mentioned control strategy by Wilson *et al* [30], MD was employed to explore the possibility to combine electrical and mechanical signals [21], to unravel the multistep translocation of proteins in nanopores [65, 69, 70] to quantify electroosmotic flow [52, 71–73], and to test the structural stability of oligomeric pores [74].

A crucial factor limiting the employment of MD simulation in nanopore protein sensing is the computational cost.

Although dedicated supercomputers allow the dynamics of proteins to be simulated on *ms* time scales [75, 76], the typical runs span length scales of 5–30 nm and time scales of 10–500 ns. In the framework of nanopore sensing, this means that also the quantitative estimation of current blockade intensities at experimental ΔV remains a hard problem. Indeed, to achieve a sufficiently low error to discriminate different current levels at the voltage used in experiments, one should perform runs of milliseconds which is far from standard available computational resources. The viable option is increasing the external driving in order to get a larger current signal. This enforces to work with ΔV that are, in some cases, one order of magnitude larger that experimental ones.

The inaccessibility to long enough time scales of all-atom MD involves another fundamental issue. The protein-pore interaction is often characterized by the presence of several long-standing metastable states where proteins may get stuck in specific conformations. In principle, the correct sampling on the conformational space can be achieved with free-energy calculations [77] however, often these methods are out of the computational capabilities of currently available computational power. Hence, researchers rely on empirical ad-hoc approaches that, hopefully, lessen the sampling problem. A brute-force solution, is running replicas of the same set-up [65, 69] and clustering the resulting conformations in order to extract the relevant and reproducible structural features. Approach based on the parallel running of several replicas, have been successful in characterizing translocations and ion current blockades of DNA through Mpsa pore [76, 78]. Another possibility consists in using a precalculated electrostatic field acting only on the translocating molecule [79]. This approach, dubbed Grid-Steered MD, was shown to induce the translocation of polymers on computational simulable time scales reducing the strain of the solute conformation that is usually introduced by other translocation protocols [79]. Beside the above mentioned difficulties, different blockade intensities were recently observed for different peptides dwelling in αHL [47], see section 3, and an ongoing research are exploring the capability of αHL to distinguish among different homopeptides [80].

5.2. Coarse-graining and mesoscale approach

To overcome the time and length scale limitation of all-atom MD technique one acceptable compromise is resorting to coarse-graining models (CGM) that introduce a drastic simplification in the finer chemical details of both pores and biopolymers. In CGM, atomic force fields are replaced by the so-called united-atom force fields, where a single unit (particle) replaces a group of atoms by incorporating their chemical, mechanical and physical properties [81]. This approach results reliable as long as, over the time scales of interest, the neglected degrees of freedom can be considered as 'frozen'. In the essence, the speed-up of simulations is pursued by abandoning the atomic description at the price of a more qualitative, yet meaningful, picture.

We often question the actual power of CGM in the biological problems and, in particular, in biopolymer sequencing. It is in fact undeniable that CGMs, due the drastic simplification, neglect important chemical details which may be crucial factors to distinguish the single monomers of the biopolymers. A possible role of CGM in protein sequencing is the exploration of strategies for controlling sequential polymers translocation (signal-to-monomer matching, see section 2) and to study the polymer capture, see section 4. Both problems can be tackled via non-equilibrium runs, where an external forcing, mimicking for instance the applied voltage, acts on the polymer, or by analyzing the free-energy profile associated to the protein transport through the nanopore.

5.3. Non equilibrium CG simulations

A common approach describes the protein as a chain of N interacting beads the dynamics of which is governed by the Langevin equation

$$m_i\ddot{\mathbf{r}}_i = -\gamma_i\dot{\mathbf{r}}_i + \mathbf{F}_i^{\mathbf{int}} + \mathbf{F}_i^{\mathbf{pore}} + \mathbf{F}_i^{\mathbf{ext}} + \boldsymbol{\eta}_i \tag{12}$$

where $\mathbf{r_i}$ indicates the position of the ith bead, m_i its mass, γ_i the friction and $\boldsymbol{\eta}_i$ the noise (satisfying the fluctuation-dissipation relation). The total force is decomposed into the contribution $\mathbf{F}_i^{\mathbf{pore}}$ due to the pore-protein interaction, the external forcing $\mathbf{F}_i^{\mathbf{ext}}$ and the intra-chain interaction $\mathbf{F}_i^{\mathbf{int}}$. For unfolded proteins, $\mathbf{F}_i^{\mathbf{int}}$ can be modelled as a simple bead and spring chain plus excluded volume interactions, while for folded proteins more complex approaches [82] such as the Gō-like models [83, 84] and the Sorenson–Head-Gordon (SHG) model [85] are needed. The electrolytic solution is not explicitly taken into account: it only acts on the system as a thermal bath. CGMs were widely employed in studying the polymer capture and translocation [86–91]. Although quantitative information on the time scales is lost, useful qualitative transport properties are captured (we refer the interested reader to [92] for a thorough discussion on CGM for polymers and colloids). Due to the relatively small computational cost of a single run, unlike all-atom MD, CGMs allow collecting a huge statistics of events, therefore, they are particularly suitable for characterizing the statistical properties of translocations.

5.4. Free-energy profiles

The relatively small computational cost of CGMs makes it convenient their employment for calculating free-energy landscapes of translocation processes. In polymer capture and translocation, usually a single collective variable Q describing the stage of the translocation (e.g. $Q = 0$, untranslocated molecule, $Q = 1$ complete translocation) can be used. Hence, the free-energy profile $G(Q)$ readily provides information on the energy barriers and on possible metastable states that translocating molecules can experience along their pathway along the pore. A crucial choice is the definition of the proper reaction (collective) coordinate Q, which must be as much as representative of the system evolution, in order to monitor the presence of on-pathway intermediates. There are different computational methods able to reconstruct a free-energy profile as a function of Q, a detailed description goes beyond the

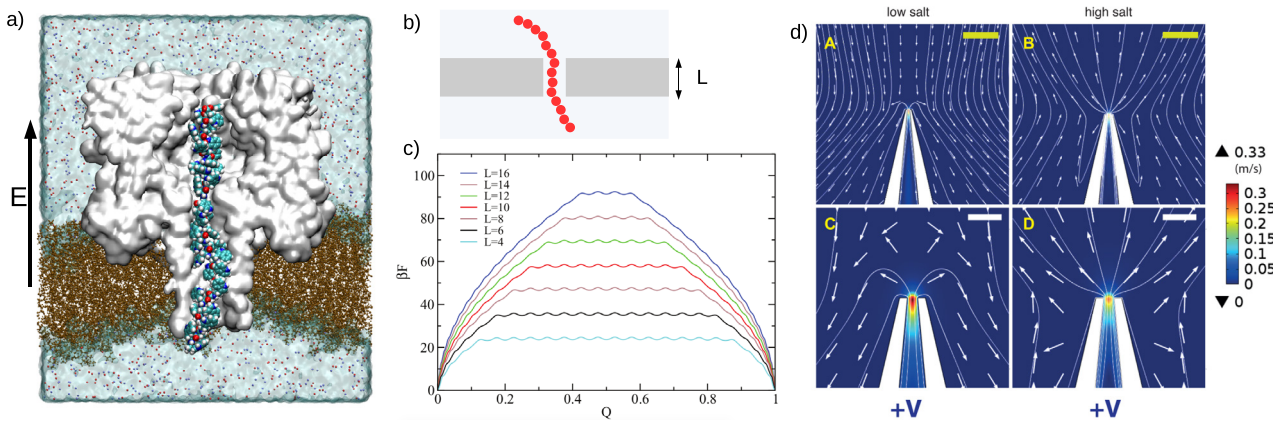

Figure 7. Computational approaches. (a) Sketch of an all-atom MD simulation of a Tryptophan homopeptide in an αHL pore [80]. An external homogeneous and constant electrical field acts orthogonal to the lipid membrane. Water molecules are not explicitly shown. (b) Coarse grained models. The protein is represented as a sequence of beads, each corresponding to an amino acid. (c) Example of free-energy profiles reported in Polson *et al* [90] showing the sensitivity to the sliding of a polymer constituted by $N = 20$ monomers inside a pore. Data refer to different pore length L. Each bump corresponds to the sliding of a single monomer through the pore. The collective variable Q is defined as $Q = 1/N \sum_i^N Q_i$ where $Q_i = 0$ for monomers still outside the pore, $Q_i = 1$ for translocated monomers, while, for monomers inside the pore, $Q_i = z_i/L$ with z_i being the axial coordinate along the pore. Reprinted with permission from [90]. Copyright 2013, AIP Publishing LLC. (d) Finite element simulation of electroosmotic flow for a 15 nm diameter pore for 10 mM (Low salt) and 100 mM (High salt). Yellow and white bars correspond to 200 nm and 50 nm, the arrows indicate flow direction. Panel (d) is adapted with permission from [94]. Copyright 2014 American Chemical Society. VMD [8] software was used for panel (a).

purpose of the present review, the interested reader can refer to the following works [77, 86, 93].

From simple physical considerations, one may expect the following general properties of $G(Q)$: (i) $G(Q)$ increases as the polymer becomes increasingly confined by the nanopore as a result of the reduction in conformational entropy; (ii) after polymer has filled the nanopore, $G(Q)$ should remain approximately constant and local minima indicate the presence of metastable states in which the protein can be trapped; (iii) the height of the free-energy barrier grows with pore length L, as, the longer the pore the larger its confinement power. Point (i) results crucial to the capture phase. Indeed, the higher the entrance free-energy barrier, the lower the probability to be captured. Point (ii), instead, is important to both signal-to-residue matching and distinguishabilty. Indeed, the presence of minima separated by barriers larger than the thermal energy, k_BT, indicates that the protein can stall into the pore. Remarkably, Polson *et al* [90], have shown that a free-energy profile can exhibit oscillations with a period that corresponds to a longitudinal movement of one monomer through the pore, see figure 7(c). The magnitude of the oscillations depends on the geometry of the pore. It is interesting to remark, as Polson and coworkers did, that when the pore longitudinal size exceeds the chain length, the wiggling effect is lost, indicating that shorter pores can be effective also for short peptides. This feature suggests that a unidirectional and step-like translocation is possible.

5.5. Continuum modelling

Continuum models do not take into account neither the motion of individual molecules of the fluid (water and ions) nor the Brownian motion of the macromolecules. This is, potentially, a strong limitation to their direct application to the analysis of the flow inside the nanopore and of the macromolecule capture. However, as we will see, continuum models can be successfully employed in nanopore sensing research down to nanometer scale to get crucial qualitative insight on the shape of electric and fluid dynamics field and to formulate preliminary quantitative models.

5.5.1. Electrohydrodynamics formulation. For simplicity, let us consider an electrolyte where only two ion species are present. At the continuum level, fluid velocity $\mathbf{u}$, pressure p, ion number densities n^+ and n^-, electrical potential Φ and electric field $\mathbf{E} = -\nabla\Phi$ are ruled by the following set of partial differential equations [51, 95]:

$$\nabla \cdot \mathbf{u} = 0 \tag{13}$$

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} = -\frac{1}{\rho}\nabla p - \nu \nabla^2 \mathbf{u} + \frac{\rho_e}{\rho}\mathbf{E} \tag{14}$$

$$\frac{\partial n^\pm}{\partial t} + \mathbf{u} \cdot \nabla n^\pm = -\nabla \cdot \left[D^\pm \nabla n^\pm + n^\pm \mu^\pm e z^\pm \mathbf{E}\right] \tag{15}$$

$$\nabla^2 \Phi = -\frac{\rho_e}{\epsilon} \tag{16}$$

where, ρ and ν are the density and the kinematic viscosity of the solvent (both assumed to be constant and homogeneous), μ^+ and μ^- the ion mobilities, $D^+ = \mu^+ k_B T$ and $D^- = \mu^- k_B T$ their diffusion coefficients, z^+ and z^- their ionic valences, e the electron charge, $\epsilon = \epsilon_0 \epsilon_e$, with ϵ_0 the vacuum dielectric constant and ϵ_e the relative permittivity of the solution and $\rho_e = e(|z^+|n^+ - |z^-|n^-)$ is the local charge. The complete system is often referred as Navier–Stokes Poisson–Nernst–Planck (NS-PNP) [94, 96]. If more than two ionic species are present, it is sufficient to add appropriate transport equations in the form of equation (15) and the corresponding contributions

to the local charge ρ_e. Since all the fluid properties (ν, ϵ_r, ρ) are taken to be constant (i.e. independent of the charge carrier concentrations) and equal to solvent one (water), the model is strictly valid for dilute solutions. In addition, usually, the non linear term $\mathbf{u}\cdot\nabla\mathbf{u}$ in the Navier–Stokes equation is negligible at low Reynolds number and, if no fast external forcing occurs, also the time derivative can be ignored, thus the equations (13)–(16) can be solved for the steady state.

Appropriate boundary conditions have to be provided. Concerning the fluid, in micro and nanofluidic devices a suitable expression for the tangential component of the velocity $\mathbf{u}$ is given by the partial slip Navier boundary condition [97], which for planar walls reads $\mathbf{V}_s = L_s\partial\mathbf{u}_\pi/\partial n$ where $\mathbf{V}_s$ is the fluid velocity with respect to the wall, $\mathbf{u}_\pi$ the tangential component of the velocity field, n the direction normal to the wall and L_s the so called slip length [97]. Atomistic simulations and experiments suggest that the L_s for water on smooth surfaces hardly exceeds 1–2 nm [98–100]. In the systems we are interested in for biopolymer sensing, the diameter of the pore is on the nanometer scale, hence, slip lengths of the order of $L_s \simeq 1$ nm can lead to relevant modifications of the flow pattern. However, in general, the materials usually employed for fabricating the nanopore are hydrophilic [101], thus allowing, as a first approximation, to safely assume the usual no-slip boundary condition.

Concerning the Poisson equation (16), Gauss's law implies the tangential component of $\mathbf{E}$ to be continuous at the liquid-solid interface and a jump in the normal component

$$\epsilon_0\epsilon_e\mathbf{E}_e\cdot\hat{\mathbf{n}} = \epsilon_0\epsilon_w\mathbf{E}_w\cdot\hat{\mathbf{n}} + q_w\,, \tag{17}$$

where the subscript e and w denote the electrolyte and the solid wall and q_w is the surface charge distribution at the interface. In our case, $\epsilon_e \simeq 80$ for liquid water while $\epsilon_w \sim 5$ for the wall ($\epsilon_w \simeq 7.5$ for silicon nitride and $\epsilon_w \simeq 3$ for lipid membranes [102]), hence, as a first approximation we can reduce equation (17) to

$$\frac{\partial\Phi}{\partial n} = -\frac{q_w}{\epsilon_0\epsilon_e}\,. \tag{18}$$

Analytical solutions of the NS-PNP model exist only in few simple cases. In particular, for planar or circular channels with homogeneous charge and an electrical field applied along the channel axis, analytical expressions are found for both ionic and electroosmotic flows [51]. For general systems, numerical solutions are needed [94, 96]. Although continuum models cannot be directly employed to analyze the protein capture and translocation, they can be used to get a reliable estimation of the force acting on proteins in the capture region. Indeed, following [96], for colloid transport in conical nanopore, it is possible to solve a set of steady state simulations with the protein in different position and, for each of them, to calculate the electrophoretic and electroosmotic forces. As an example, figure 7(d) reports the results for an electrohydrodynamic simulation for a electrolytic KCl solution in a glass nanopore [94]. The surface is negatively charged and, as expected, the flow inside the pore is directed outward. However, interestingly, the direction of the flow outside the pore depends on the salt concentration.

It is worth noting that Poisson–Nernst–Planck (PNP) system, i.e. equations (15) and (16) with no advection, possibly modified to include ion size effects or wall repulsion, have been widely applied to nanopores [103–105]. PNP models provide useful insights when electroosmotic effect are negligible, a condition that is often associated with a low surface charge, however, for high surface charge the electrosmotic flow can be so high to overwhelm the ionic currents [106, 107].

As a final comment, we can mention that continuum models are also employed for deriving approximated analytical expressions that are widely used in the interpretation of experimental results. For instance, reduced continuum models provide approximated analytical expression for field inside and outside the pore [49, 50] and for electroosmotic flow [108] and, hence, can be useful in preliminary analysis of protein capture. Another common employment of reduced model is to provide a first estimation of the size of the molecule monomers associated to a current blockade level. However, due to the lacking of several crucial ingredients (the main one being the atomistic nature of the ionic flow in the sensing region), these models hardly goes beyond the intuitive results that current blockade increases with the size of the amino acids clogging the sensing region.

6. Conclusions

Although several recent advances in peptide translocation control and nanopore fabrication have been proposed in the last few years, the goal of nanopore sequencing of proteins is not yet at hand. In this short review, we tried to summarize the principal theoretical approaches and the corresponding issues that researchers have to tackle when dealing with a problem proven to be so challenging already at a theoretical and computational level, where, in principle, all the variables and the various components can be kept under control. Technical innovations and novel device design can hopefully help to find reliable solutions to both the two fundamental requirements we discussed (signal-to-monomer matching and distinguishability) and the effort of a wide and multidisciplinary community is highly needed. On the nanopore fabrication side, possible innovation stems from recently proposed 'zero-depth' pore architecture, obtained by the intersection of two crossed superimposed parallelepiped channel, by Arjmandi-Tash *et al* [109]. These new-generation of pores can potentially be employed for reducing the size of the sensing region. Similarly, combined approaches that merge two different signals (e.g. longitudinal and transversal current) can help in improving the signal distinguishability.

The pathway towards nanopore protein sequencing is rich of possible intermediate achievements that can be beneficial or even ground-breaking for other parallel proteomic employment of nanofluidic devices. Among others, it is worth mentioning the single-molecule protein identification [33, 110], and detection of post-translational modifications [111, 112] and DNA-protein interaction [113]. In such a burgeoning field, a crucial role can be played by a more fundamental understanding of the electrodynamical transport phenomena

under confinement. Indeed, it is undeniable that a reliable sequencing protocol based on current measurements can not disregard a robust knowledge of the complex electrochemical phenomena taking place when a protein crosses a nanopore.

Acknowledgments

We would like to thank Mr Alberto Gubbiotti, Mr Giovanni Di Muccio and Professor Tudor Luchian for useful discussions.

ORCID iDs

Mauro Chinappi https://orcid.org/0000-0002-4509-1247

References

[1] Jain M, Olsen H E, Paten B and Akeson M 2016 The Oxford nanopore minION: delivery of nanopore sequencing to the genomics community *Genome Biol.* **17** 239
[2] Kennedy E, Dong Z, Tennant C and Timp G 2016 Reading the primary structure of a protein with 0.07 nm^3 resolution using a subnanometre-diameter pore *Nat. Nanotechnol.* **11** 968–76
[3] Lomize M A, Lomize A L, Pogozheva I D and Mosberg H I 2006 OPM: orientations of proteins in membranes database *Bioinformatics* **22** 623–5
[4] Faller M, Niederweis M and Schulz G E 2004 The structure of a mycobacterial outer-membrane channel *Science* **303** 1189–92
[5] Song L *et al* 1996 Structure of staphylococcal α-hemolysin, a heptameric transmembrane pore *Science* **274** 1859–65
[6] Tanaka K, Caaveiro J M, Morante K, González-Mañas J M and Tsumoto K 2015 Structural basis for self-assembly of a cytolytic pore lined by protein and lipid *Nat. Commun.* **6** 6337
[7] Iacovache I, De Carlo S, Cirauqui N, Dal Peraro M, Van Der Goot F G and Zuber B 2016 Cryo-EM structure of aerolysin variants reveals a novel protein fold and the pore-formation process *Nat. Commun.* **7** 12062
[8] Humphrey W *et al* 1996 VMD: visual molecular dynamics *J. Mol. Graph.* **14** 33–8
[9] Laszlo A H, Derrington I M and Gundlach J H 2016 MspA nanopore as a single-molecule tool: from sequencing to SPRNT *Methods* **105** 75–89
[10] Van den Hout M, Hall A R, Wu M Y, Zandbergen H W, Dekker C and Dekker N H 2010 Controlling nanopore size, shape and stability *Nanotechnology* **21** 115304
[11] Merchant C A *et al* 2010 DNA translocation through graphene nanopores *Nano Lett.* **10** 2915–21
[12] Heerema S J and Dekker C 2016 Graphene nanodevices for DNA sequencing *Nat. Nanotechnol.* **11** 127
[13] Goyal G, Lee Y B, Darvish A, Ahn C W and Kim M J 2016 Hydrophilic and size-controlled graphene nanopores for protein detection *Nanotechnology* **27** 495301
[14] Barati Farimani A, Heiranian M, Min K and Aluru N R 2017 Antibody subclass detection using graphene nanopores *J. Phys. Chem. Lett.* **8** 1670–6
[15] Feng J, Liu K, Bulushev R D, Khlybov S, Dumcenco D, Kis A and Radenovic A 2015 Identification of single nucleotides in MoS_2 nanopores *Nat. Nanotechnol.* **10** 1070
[16] Chen H, Li L, Zhang T, Qiao Z, Tang J and Zhou J 2018 Protein translocation through a MoS_2 nanopore: a molecular dynamics study *J. Phys. Chem.* C **122** 2070–80
[17] Di Ventra M and Taniguchi M 2016 Decoding DNA, RNA and peptides with quantum tunnelling *Nat. Nanotechnol.* **11** 117–26
[18] Lagerqvist J, Zwolak M and Di Ventra M 2006 Fast DNA sequencing via transverse electronic transport *Nano Lett.* **6** 779–82
[19] Rossini A E, Gala F, Chinappi M and Zollo G 2018 *Nanoscale* **10** 5928–37
[20] Boynton P and Di Ventra M 2016 Sequencing proteins with transverse ionic transport in nanochannels *Sci. Rep.* **6** 25232
[21] Farimani A B, Heiranian M and Aluru N 2015 Electromechanical signatures for DNA sequencing through a mechanosensitive nanopore *J. Phys. Chem. Lett.* **6** 650–7
[22] Gilboa T and Meller A 2015 Optical sensing and analyte manipulation in solid-state nanopores *Analyst* **140** 4733–47
[23] Oukhaled G, Mathe J, Biance A, Bacri L, Betton J, Lairez D, Pelta J and Auvray L 2007 Unfolding of proteins and long transient conformations detected by single nanopore recording *Phys. Rev. Lett.* **98** 158101
[24] Talaga D and Li J 2009 Single-molecule protein unfolding in solid state nanopores *J. Am. Chem. Soc.* **131** 9287–97
[25] Van Meervelt V, Soskine M, Singh S, Schuurman-Wolters G K, Wijma H J, Poolman B and Maglia G 2017 Real-time conformational changes and controlled orientation of native proteins inside a protein nanoreactor *J. Am. Chem. Soc.* **139** 18640–6
[26] Varongchayakul N, Huttner D, Grinstaff M W and Meller A 2018 Sensing native protein solution structures using a solid-state nanopore: unraveling the states of VEGF *Sci. Rep.* **8** 1017
[27] Waduge P, Hu R, Bandarkar P, Yamazaki H, Cressiot B, Zhao Q, Whitford P C and Wanunu M 2017 Nanopore-based measurements of protein size, fluctuations, and conformational changes *ACS Nano* **11** 5706–16
[28] Martyushenko N, Bell N A, Lamboll R D and Keyser U F 2015 Nanopore analysis of amyloid fibrils formed by lysozyme aggregation *Analyst* **140** 4882–6
[29] Giamblanco N, Coglitore D, Janot J-M, Coulon P E, Charlot B and Balme S 2018 Detection of protein aggregate morphology through single antifouling nanopore *Sensors Actuators* B **260** 736–45
[30] Wilson J, Sloman L, He Z and Aksimentiev A 2016 Graphene nanopores for protein sequencing *Adv. Funct. Mater.* **26** 4830–8
[31] Sampath G 2015 Amino acid discrimination in a nanopore and the feasibility of sequencing peptides with a tandem cell and exopeptidase *RSC Adv.* **5** 30694–700
[32] Restrepo-Pérez L, John S, Aksimentiev A, Joo C and Dekker C 2017 SDS-assisted protein transport through solid-state nanopores *Nanoscale* **9** 11685–93
[33] Kolmogorov M, Kennedy E, Dong Z, Timp G and Pevzner P A 2017 Single-molecule protein identification by sub-nanopore sensors *PLoS Comput. Biol.* **13** e1005356
[34] Clarke J, Wu H-C, Jayasinghe L, Patel A, Reid S and Bayley H 2009 Continuous base identification for single-molecule nanopore DNA sequencing *Nat. Nanotechnol.* **4** 265
[35] Pud S, Chao S-H, Belkin M, Verschueren D, Huijben T, van Engelenburg C, Dekker C and Aksimentiev A 2016 Mechanical trapping of DNA in a double-nanopore system *Nano Lett.* **16** 8021–8
[36] Cadinu P *et al* 2017 Single molecule trapping and sensing using dual nanopores separated by a zeptoliter nanobridge *Nano Lett.* **17** 6376–84
[37] Maulbetsch W, Wiener B, Poole W, Bush J and Stein D 2016 Preserving the sequence of a biopolymers monomers as they enter an electrospray mass spectrometer *Phys. Rev. Appl.* **6** 054006

[38] Hoogerheide D P, Lu B and Golovchenko J A 2014 Pressure–voltage trap for DNA near a solid-state nanopore *ACS Nano* **8** 7384–91
[39] Huang G, Willems K, Soskine M, Wloka C and Maglia G 2017 Electro-osmotic capture and ionic discrimination of peptide and protein biomarkers with FraC nanopores *Nat. Commun.* **8** 935
[40] Boukhet M, Piguet F, Ouldali H, Pastoriza-Gallego M, Pelta J and Oukhaled A 2016 Probing driving forces in aerolysin and α-hemolysin biological nanopores: electrophoresis versus electroosmosis *Nanoscale* **8** 18352–9
[41] Asandei A, Schiopu I, Chinappi M, Seo C H, Park Y and Luchian T 2016 Electroosmotic trap against the electrophoretic force near a protein nanopore reveals peptide dynamics during capture and translocation *ACS Appl. Mater. Interfaces* **8** 13166–79
[42] Nivala J, Marks D B and Akeson M 2013 Unfoldase-mediated protein translocation through an α-hemolysin nanopore *Nat. Biotechnol.* **31** 247–50
[43] Belkin M, Chao S-H, Jonsson M P, Dekker C and Aksimentiev A 2015 Plasmonic nanopores for trapping, controlling displacement, and sequencing of DNA *ACS Nano* **9** 10598–611
[44] Belkin M and Aksimentiev A 2016 Molecular dynamics simulation of DNA capture and transport in heated nanopores *ACS Appl. Mater. Interfaces* **8** 12599–608
[45] Mereuta L, Roy M, Asandei A, Lee J K, Park Y, Andricioaei I and Luchian T 2014 Slowing down single-molecule trafficking through a protein nanopore reveals intermediates for peptide translocation *Sci. Rep.* **4** 3885
[46] Asandei A, Chinappi M, Lee J-k, Seo C H, Mereuta L, Park Y and Luchian T 2015 Placement of oppositely charged aminoacids at a polypeptide termini determines the voltage-controlled braking of polymer transport through nanometer-scale pores *Sci. Rep.* **5** 10419
[47] Asandei A, Rossini A E, Chinappi M, Park Y and Luchian T 2017 Protein nanopore-based discrimination between selected neutral amino acids from polypeptides *Langmuir* **33** 14451–9
[48] Piguet F, Ouldali H, Pastoriza-Gallego M, Manivet P, Pelta J and Oukhaled A 2018 Identification of single amino acid differences in uniformly charged homopolymeric peptides with aerolysin nanopore *Nat. Commun.* **9** 966
[49] Chinappi M, Luchian T and Cecconi F 2015 Nanopore tweezers: voltage-controlled trapping and releasing of analytes *Phys. Rev.* E **92** 032714
[50] Wanunu M, Morrison W, Rabin Y, Grosberg A Y and Meller A 2009 Electrostatic focusing of unlabelled DNA into nanoscale pores using a salt gradient *Nat. Nanotechnol.* **5** 160–5
[51] Bruus H 2008 *Theoretical Microfluidics* (Oxford: Oxford University Press)
[52] Bonome E L, Cecconi F and Chinappi M 2017 Electroosmotic flow through an α-hemolysin nanopore *Microfluid. Nanofluid.* **21** 96
[53] Rodriguez-Larrea D and Bayley H 2013 Multistep protein unfolding during nanopore translocation *Nat. Nanotechnol.* **8** 288–95
[54] Biswas S, Song W, Borges C, Lindsay S and Zhang P 2015 Click addition of a DNA thread to the n-termini of peptides for their translocation through solid-state nanopores *ACS Nano* **9** 9652–64
[55] Hoogerheide D P, Gurnev P A, Rostovtseva T K and Bezrukov S M 2018 Real-time nanopore-based recognition of protein translocation success *Biophys. J.* **114** 772–6
[56] Brune D and Kim S 1993 Predicting protein diffusion coefficients *Proc. Natl Acad. Sci.* **90** 3835–9
[57] MacKerell A D, Brooks B, Brooks C L, Nilsson L, Roux B, Won Y and Karplus M 1998 Charmm: the energy function and its parameterization *Encyclopedia Comput. Chem.* (https://doi.org/10.1002/0470845015.cfa007)
[58] Oostenbrink C, Villa A, Mark A E and Van Gunsteren W F 2004 A biomolecular force field based on the free enthalpy of hydration and solvation: the gromos force-field parameter sets 53a5 and 53a6 *J. Comput. Chem.* **25** 1656–76
[59] Wang J, Wolf R M, Caldwell J W, Kollman P A and Case D A 2004 Development and testing of a general amber force field *J. Comput. Chem.* **25** 1157–74
[60] Aksimentiev A and Schulten K 2005 Imaging α-hemolysin with molecular dynamics: ionic conductance, osmotic permeability, and the electrostatic potential map *Biophys. J.* **88** 3745–61
[61] Bhattacharya S, Muzard J, Payet L, Mathé J, Bockelmann U, Aksimentiev A and Viasnoff V 2011 Rectification of the current in α-hemolysin pore depends on the cation type: the alkali series probed by molecular dynamics simulations and experiments *J. Phys. Chem.* C **115** 4255–64
[62] Crozier P S, Henderson D, Rowley R L and Busath D D 2001 Model channel ion currents in NaCl-extended simple point charge water solution with applied-field molecular dynamics *Biophys. J.* **81** 3077–89
[63] Aksimentiev A, Heng J B, Timp G and Schulten K 2004 Microscopic kinetics of DNA translocation through synthetic nanopores *Biophys. J.* **87** 2086–97
[64] Yeh I-C and Hummer G 2004 Nucleic acid transport through carbon nanotube membranes *Proc. Natl Acad. Sci. USA* **101** 12177–82
[65] Di Marino D, Bonome E L, Tramontano A and Chinappi M 2015 All-atom molecular dynamics simulation of protein translocation through an α-hemolysin nanopore *J. Phys. Chem. Lett.* **6** 2963–8
[66] Mathé J, Aksimentiev A, Nelson D R, Schulten K and Meller A 2005 Orientation discrimination of single-stranded DNA inside the α-hemolysin membrane channel *Proc. Natl Acad. Sci. USA* **102** 12377–82
[67] Gopfrich K, Li C-Y, Ricci M, Bhamidimarri S P, Yoo J, Gyenes B, Ohmann A, Winterhalter M, Aksimentiev A and Keyser U F 2016 Large-conductance transmembrane porin made from DNA origami *ACS Nano* **10** 8207–14
[68] Chandramouli B, Di Maio D, Mancini G and Brancato G 2016 Introducing an artificial photo-switch into a biological pore: a model study of an engineered α-hemolysin *Biochim. Biophys. Acta* **1858** 689–97
[69] Bonome E L, Lepore R, Raimondo D, Cecconi F, Tramontano A and Chinappi M 2015 Multistep current signal in protein translocation through graphene nanopores *J. Phys. Chem.* B **119** 5815–23
[70] Luan B, Huynh T, Li J and Zhou R 2015 Nanomechanics of protein unfolding outside a generic nanopore *ACS Nano* **10** 317–23
[71] Luan B and Aksimentiev A 2008 Electro-osmotic screening of the DNA charge in a nanopore *Phys. Rev.* E **78** 021912
[72] Mao M, Ghosal S and Hu G 2013 Hydrodynamic flow in the vicinity of a nanopore induced by an applied voltage *Nanotechnology* **24** 245202
[73] Suk M E and Aluru N 2014 Ion transport in sub-5-nm graphene nanopores *J. Chem. Phys.* **140** 084707
[74] Desikan R, Maiti P K and Ayappa K G 2017 Assessing the structure and stability of transmembrane oligomeric intermediates of an α-helical toxin *Langmuir* **33** 11496–510
[75] Shaw D E *et al* 2008 Anton, a special-purpose machine for molecular dynamics simulation *Commun. ACM* **51** 91–7
[76] Bhattacharya S, Yoo J and Aksimentiev A 2016 Water mediates recognition of DNA sequence via ionic current blockade in a biological nanopore *ACS Nano* **10** 4644–51
[77] Tuckerman M 2010 *Statistical Mechanics: Theory and Molecular Simulation* (Oxford: Oxford University Press)

[78] Bhattacharya S, Derrington I M, Pavlenok M, Niederweis M, Gundlach J H and Aksimentiev A 2012 Molecular dynamics study of MspA arginine mutants predicts slow DNA translocations and ion current blockades indicative of DNA sequence *ACS Nano* **6** 6960–8
[79] Wells D B, Abramkina V and Aksimentiev A 2007 Exploring transmembrane transport through α-hemolysin with grid-steered molecular dynamics *J. Chem. Phys.* **127** 09B619
[80] Rossini A E, Di Muccio G, Di Marino D, Zollo G and Chinappi M 2018 in preparation
[81] Voth G A 2008 *Coarse-Graining of Condensed Phase and Biomolecular Systems* (Boca Raton, FL: CRC Press)
[82] Guardiani C, Livi R and Cecconi F 2010 Coarse grained modeling and approaches to protein folding *Curr. Bioinform.* **5** 217–40
[83] Clementi C *et al* 2000 Topological and energetic factors: what determines the structural details of the transition state ensemble and 'en-route' intermediates for protein folding? An investigation for small globular proteins *J. Mol. Biol.* **298** 937
[84] Gō N and Scheraga H A 1976 On the use of classical statistical mechanics in the treatment of polymer chain conformations *Macromolecules* **9** 535–42
[85] Sorenson J M and Head-Gordon T 2000 Matching simulation and experiment: a new simplified model for simulating protein folding *J. Comput. Biol.* **7** 469–81
[86] Huang L, Kirmizialtin S and Makarov D 2005 Computer simulations of the translocation and unfolding of a protein pulled mechanically through a pore *J. Chem. Phys.* **123** 124903
[87] Menais T, Mossa S and Buhot A 2016 Polymer translocation through nano-pores in vibrating thin membranes *Sci. Rep.* **6** 38558
[88] Sarabadani J, Ghosh B, Chaudhury S and Ala-Nissila T 2018 Dynamics of end-pulled polymer translocation through a nanopore *Europhys. Lett.* **120** 38004
[89] Ghosh B and Chaudhury S 2017 Influence of the location of attractive polymer–pore interactions on translocation dynamics *J. Phys. Chem.* B **122** 360–8
[90] Polson J M, Hassanabad M F and McCaffrey A 2013 Simulation study of the polymer translocation free energy barrier *J. Chem. Phys.* **138** 024906
[91] Muthukumar M and Kong C 2006 Simulation of polymer translocation through protein channels *Proc. Natl Acad. Sci.* **103** 5273–8
[92] Padding J and Louis A 2006 Hydrodynamic interactions and brownian forces in colloidal suspensions: coarse-graining over time and length scales *Phys. Rev.* E **74** 031402
[93] Ammenti A, Cecconi F, Marini Bettolo Marconi U and Vulpiani A 2009 A statistical model for translocation of structured polypeptide chains through nanopores *J. Phys. Chem.* B **113** 10348
[94] Laohakunakorn N, Thacker V V, Muthukumar M and Keyser U F 2014 Electroosmotic flow reversal outside glass nanopores *Nano Lett.* **15** 695–702
[95] Schoch R B, Han J and Renaud P 2008 Transport phenomena in nanofluidics *Rev. Mod. Phys.* **80** 839
[96] Rempfer G, Ehrhardt S, Holm C and de Graaf J 2017 Nanoparticle translocation through conical nanopores: a finite element study of electrokinetic transport *Macromol. Theory Simul.* **26** 1600051
[97] Lauga E, Brenner M and Stone H 2007 Microfluidics: the no-slip boundary condition *Springer Handbook of Experimental Fluid Mechanics* (Berlin: Springer) pp 1219–40
[98] Chinappi M and Casciola C 2010 Intrinsic slip on hydrophobic self-assembled monolayer coatings *Phys. Fluids* **22** 042003
[99] Huang D, Sendner C, Horinek D, Netz R and Bocquet L 2008 Water slippage versus contact angle: a quasiuniversal relationship *Phys. Rev. Lett.* **101** 226101
[100] Attard P 2013 More reliable measurements of the slip length with the atomic force microscope (arXiv:1302.6630)
[101] Barhoumi H, Maaref A and Jaffrezic-Renault N 2010 Experimental study of thermodynamic surface characteristics and pH sensitivity of silicon dioxide and silicon nitride *Langmuir* **26** 7165–73
[102] Gramse G, Dols-Perez A, Edwards M, Fumagalli L and Gomila G 2013 Nanoscale measurement of the dielectric constant of supported lipid bilayers in aqueous solutions with electrostatic force microscopy *Biophys. J.* **104** 1257–62
[103] Modi N, Winterhalter M and Kleinekathoefer U 2012 Computational modeling of ion transport through nanopores *Nanoscale* **4** 6166–80
[104] Simakov N A and Kurnikova M G 2010 Soft wall ion channel in continuum representation with application to modeling ion currents in α-hemolysin *J. Phys. Chem.* B **114** 15180–90
[105] Chaudhry J H, Comer J, Aksimentiev A and Olson L N 2014 A stabilized finite element method for modified Poisson–Nernst–Planck equations to determine ion flow through a nanopore *Commun. Comput. Phys.* **15** 93–125
[106] Chinappi M and Malgaretti P Charge polarization and eddies formation in electro-osmotic fluxes across varying-section channels 2018 in preparation
[107] Malgaretti P, Pagonabarraga I and Rubi J M 2014 Entropic electrokinetics: recirculation, particle separation, and negative mobility *Phys. Rev. Lett.* **113** 128301
[108] Gu L-Q, Cheley S and Bayley H 2003 Electroosmotic enhancement of the binding of a neutral molecule to a transmembrane pore *Proc. Natl Acad. Sci.* **100** 15498–503
[109] Arjmandi-Tash H, Bellunato A, Wen C, Olsthoorn R C, Scheicher R H, Zhang S-L and Schneider G F 2018 Zero-depth interfacial nanopore capillaries *Adv. Mater.* **30** 1703602
[110] Yao Y, Docter M, Van Ginkel J, de Ridder D and Joo C 2015 Single-molecule protein sequencing through fingerprinting: computational assessment *Phys. Biol.* **12** 055003
[111] Rosen C B, Rodriguez-Larrea D and Bayley H 2014 Single-molecule site-specific detection of protein phosphorylation with a nanopore *Nat. Biotechnol.* **32** 179–81
[112] Dong Z, Kennedy E, Hokmabadi M and Timp G 2017 Discriminating residue substitutions in a single protein molecule using a sub-nanopore *ACS Nano* **11** 5440–52
[113] Celaya G, Perales-Calvo J, Muga A, Moro F and Rodriguez-Larrea D 2017 Label-free, multiplexed, single-molecule analysis of protein–DNA complexes with nanopores *ACS Nano* **11** 5815–25