Protein sequencing via nanopore based devices: a theoretical/numerical perspective

Mauro Chinappi [a], Fabio Cecconi [b]

[a] Dipartimento di Ingegneria Industriale, Università di Roma Tor Vergata, 00133, Roma, Italia
[b] CNR-Istituto dei Sistemi Complessi UoS Sapienza, 00185, Roma, Italia.

e-mail: mauro.chinappi@uniroma2.it

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

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 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. A widely used approach is the so called resistive pulse technique where a voltage is applied between the two sides of the membrane and the molecule-pore interaction induces a change in the system electric conductance.

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 three main ideal requirements that a nanopore protein sequencing device has to fulfill are (i) capture; the protein should move from the bulk to the pore region, (ii) signal-to-monomer matching; each signal has to be unambiguously associated to a specific position in the protein sequence and (iii) distinguishability; the signal level associated to a single amino acid (AA) has to allow the unambiguous identification of the AA.

We are currently working on all the three issues [2,3] using computational and theoretical approaches. A main challenge in the modelling of resistive pulse for protein sequencing is the entanglement among hydrodynamic/electric/chemical effects occurring at different time and length scales. 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. On the contrary, when the molecule is close to the pore (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 1(c) reports the main phenomena on a qualitative time-scale length-scale chart. The violet shaded region includes all phenomena that somehow are accessible to the explicit simulation of the dynamics upon choosing a proper time and space scale resolution (atomistic, coarse grained or continuous) while the selection of 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 directly accessible to brute force atomistic simulations. Remarkably, the yellow area just includes effects that are fundamental for protein sequencing. In this contribution, we will discuss our ongoing works on the modelling of protein capture focusing, on particular, on Brownian models for dielectrophoretic and electro-osmotic trapping.

[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] Chinappi, M.; Cecconi, F. (2018) *Protein Sequencing via Nanopore Based Devices: A Nanofluidic Perspective*. J. Phys. Condens. Matter 2018, 30, 30.

[3] Di Muccio G, Rossini AE, Di Marino D, Zollo G, Chinappi M (2019). *Insights into protein sequencing with an α-hemolysin nanopore by atomistic simulations.* Accepted for publication on Scientific Reports.

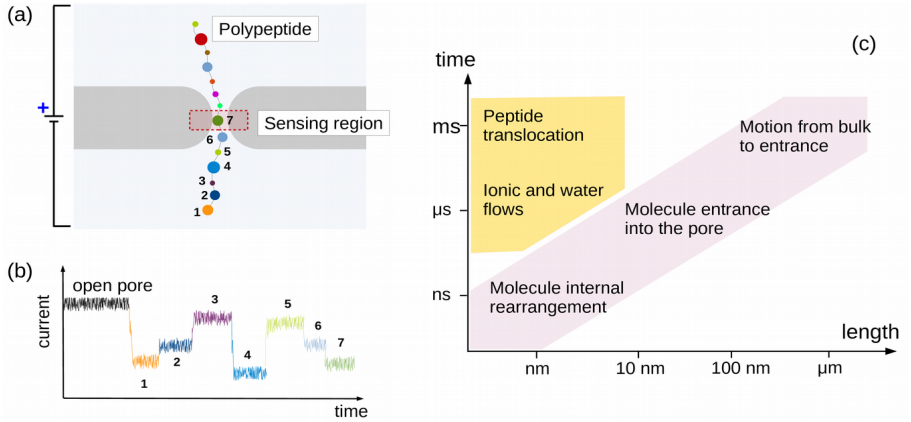

Figure 1. Nanopore protein sequencing. a-b) A voltage is applied between the two reservois and the interaction of a protein with a nanopore induces variations in the ionic current. Generally, the signal is mainly influenced by the interaction of the molecule with a specific region of the pore (the constriction), 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 (b). c) Typical time and length scales for the different transport phenomena involved in nanopore protein sequencing. The rearrangements of the single amino acids in the 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 ∼10μs ÷ 1ms, while a complete translocation of an unfolded protein several requires milliseconds. Capture instead involves larger space and time scale ranging from bulk diffusive scales (motion of the protein from the bulk to the pore region) to decades of nm (capture region close to the pore entrance). Adapted from [2].