



**HAL**  
open science

## A Route toward Protein Sequencing using Solid-State Nanopores Assisted by Machine Learning

Andreina Nohemi Urquiola Hernandez, Adrien Nicolai, Christophe Guyeux,  
Patrick Senet

► **To cite this version:**

Andreina Nohemi Urquiola Hernandez, Adrien Nicolai, Christophe Guyeux, Patrick Senet. A Route toward Protein Sequencing using Solid-State Nanopores Assisted by Machine Learning. Journées Calcul Données, Oct 2022, Dijon, France. hal-03893453

**HAL Id: hal-03893453**

**<https://hal.science/hal-03893453>**

Submitted on 11 Dec 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A Route toward Protein Sequencing using Solid-State Nanopores Assisted by Machine Learning

Andreina Nohemi Urquiola Hernandez<sup>1</sup>, Adrien Nicolai<sup>1,\*</sup>, Christophe Guyeux<sup>2</sup> and Patrick Senet<sup>1</sup>



<sup>1</sup> Laboratoire Interdisciplinaire Carnot de Bourgogne, UMR 6303 CNRS-Université de Bourgogne Franche-Comté

<sup>2</sup> Laboratoire FEMTO-ST, UMR 6174 CNRS-Université de Bourgogne Franche-Comté  
adrien.nicolai@u-bourgogne.fr



## Abstract

Solid-State Nanopores made of 2-D materials such as MoS<sub>2</sub> have emerged as one of the most versatile sensors for single-biomolecule detection, which is essential for early disease diagnosis (**biomarker detection**). One of the most promising applications of SSN is DNA and protein sequencing, at a low cost and faster than the current standard methods. The detection principle relies on measuring the relatively small variations of ionic current as charged biomolecules immersed in an electrolyte traverse the nanopore, in response to an external voltage applied across the membrane. The passage of a biomolecule through the pore yields information about its structure and chemical properties, as demonstrated experimentally particularly for DNA molecules. Indeed, protein sequencing using SSN remains highly challenging since the protein ensemble is far more complex than the DNA ensemble [1]. In this work, we performed extensive unbiased all-atom classical Molecular Dynamics simulations to produce data of translocation of biological peptides through single-layer MoS<sub>2</sub> nanopores ( $D = 1.3$  nm). Peptide made of 12 different amino acids from the different families (non polar/hydrophobic, polar/neutral, basic and acidic) were chemically linked to a short polycationic charge carrier. First, ionic current time series were computed from MD and peptide-induced blockade events were extracted and characterized using structural break detection. Second, clustering (unsupervised learning) of ionic current subdrops and duration using Gaussian Mixture Model was applied. Using this technique, we demonstrate that each amino acid presents a large diversity of ionic current characteristics, however, charged amino acids were distinguished from the others. These promising findings may offer a route toward protein sequencing using MoS<sub>2</sub> solid-state nanopores.

## Solid-State Nanopores

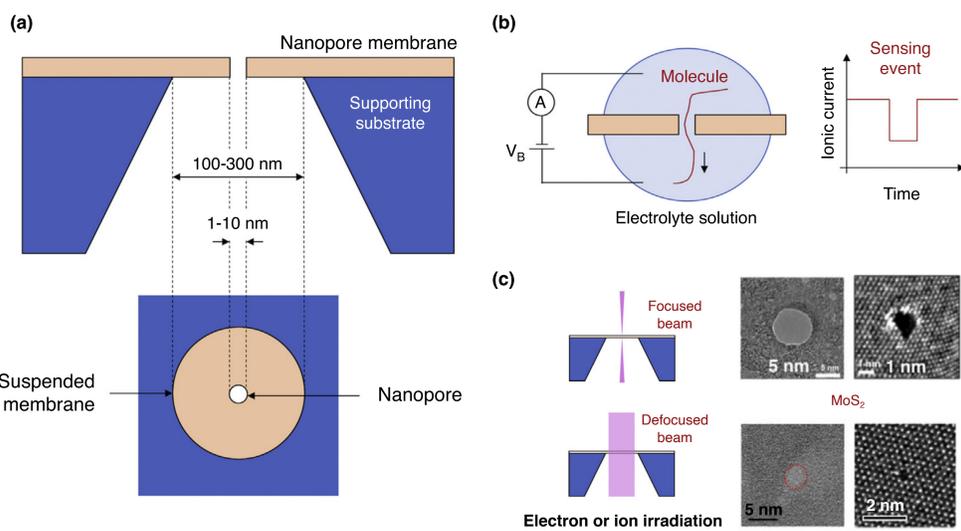
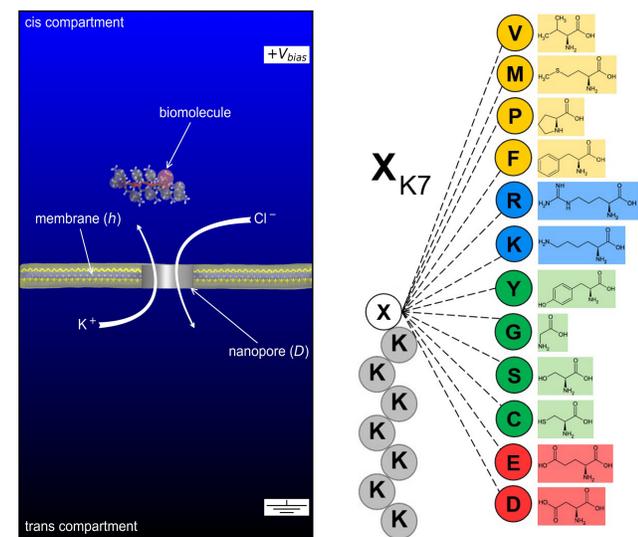
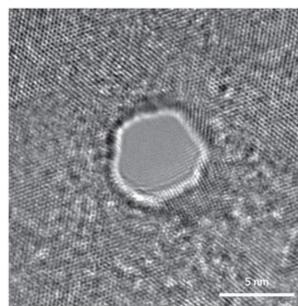


Figure: Schematic of typical nanopore sensor chips and nanoporous membranes [2]. (a) Cross-sectional and top view of a silicon-based or 2D material membrane with a nanopore suspended over a circular aperture in a supporting substrate (e.g. silicon, glass). (b) Voltage-induced translocation of an elongated charged biomolecule through a single nanopore immersed in an electrolyte solution (left) giving rise to ionic current changes, representing a sensing event (right). (c) Controlled fabrication of single 2D nanopores and ordered nanopore arrays using electron or ion irradiation on MoS<sub>2</sub>.

## all-atom classical Molecular Dynamics in explicit solvent

- ▶ GROMACS software package (2018.2)
- ▶ Machines: AMD EPYC 7302 @ 3Ghz (2 processors, 16 cores/processor)
- ▶ Size of the system: 100,000 atoms - Total simulation time: 150  $\mu$ s
- ▶ Scaling: 150 ns / day on 256 cores - Total CPU time: 6 millions of hours

single-layer MoS<sub>2</sub>  
D = 1.3 nm  
h = 0.3 nm



## Results

### Ionic Current Time Series

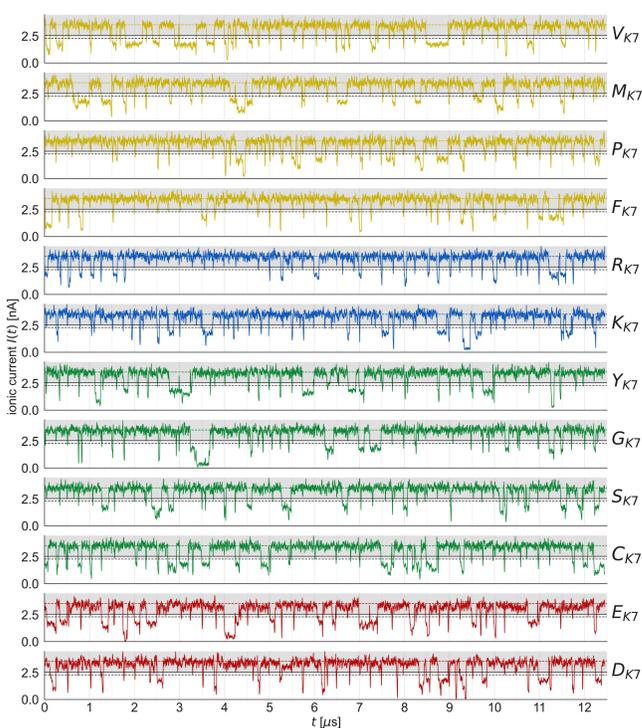


Figure: Ionic current (nA) vs. time (in  $\mu$ s) recorded during MD simulations.

### Big data: statistical analysis of fingerprints

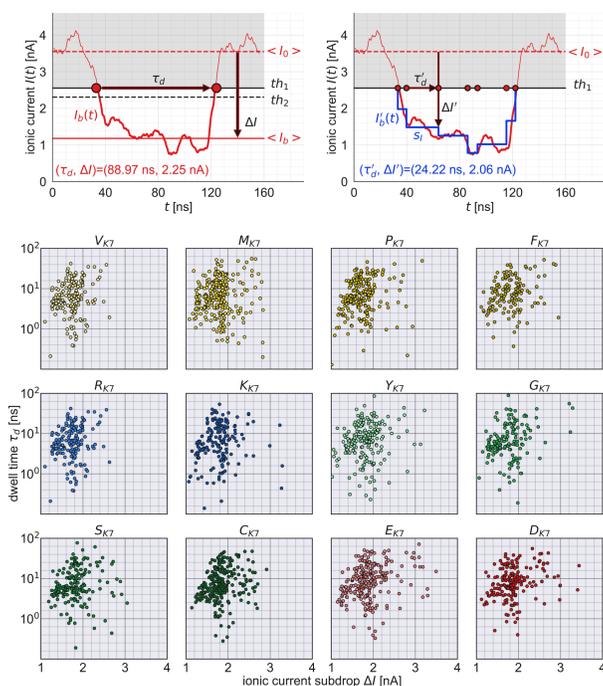


Figure: Top panels: Detection and characterization of peptide-induced blockade events. Bottom panel: Ionic current subdrops (in nA) vs. dwell time (in ns) detected for each amino acid during MD.

### Unsupervised Learning: GMM Clustering

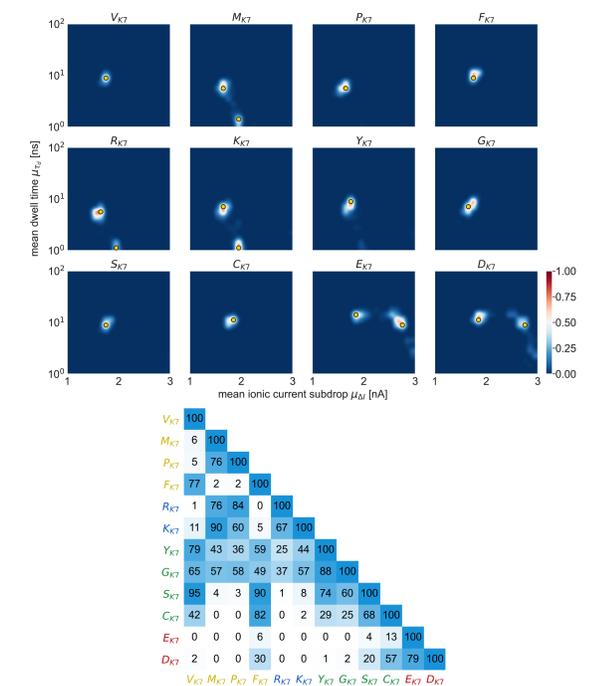


Figure: Top panel: 2-D PDFs of Gaussian means for subdrops and dwell time extracted from Gaussian Mixture Model clustering. Bottom panel: Matrix of similarity between 2-D PDFs. Values are given in percent..

## References

- [1] A. Nicolai and P. Senet. Challenges in Protein Sequencing using MoS<sub>2</sub> Nanopores, **2022**. In Single Molecule Sensing Beyond Fluorescence. Nanostructure Science and Technology. Springer, Cham.
- [2] G. Danda and M. Drndic. Two-dimensional nanopores and nanoporous membranes for ion and molecule transport. Current Opinion in Biotechnology **2019**, 55: 124–133.

## Acknowledgements

The simulations were performed at DSI-CCuB (Université de Bourgogne). The work is part of the project NANO-NEURO-MED (2019–2022) and SEPIA (2021–2024) supported by the EIPHI Graduate School (contract ANR-17-EURE-0002), the Conseil Régional de Bourgogne Franche-Comté and the European Union through the PO FEDER-FSE Bourgogne 2014/2020 and 2021/2027 programs.

