

A NANOPORE SYSTEM FOR SINGLE-MOLECULE PROTEIN SEQUENCING

Allocation: Illinois/872.6 Knh

PI: Aleksei Aksimentiev¹

Collaborators: Kumar Sarthak¹, Hadjer Ouldali², Fabien Piguet², Abdelghani Oukhaled², Tobias Ensslen³, Jan C. Behrends³

¹University of Illinois at Urbana–Champaign

²University of Cergy–Pontoise

³University of Freiburg

EXECUTIVE SUMMARY

Single-molecule protein sequencing would provide unprecedented insights into cellular processes and the diseases that arise from protein malfunction. As compared to nucleic acid sequencing, there are no robust methods that can read the protein sequence without any challenges. In collaboration with the Oukhaled Lab at the University of Cergy–Pontoise and the Behrends Lab at the University of Freiburg, the research team is developing a nanopore system that can identify the type of amino acids that pass through the nanopore from a modulation of the nanopore ionic current—a crucial first step toward realizing nanopore protein sequencing. The researchers have already shown that a biological nanopore aerolysin can identify thirteen out of the 20 different amino acids. Using atomistic molecular dynamics (MD) simulations, the team aims to further engineer the aerolysin nanopore to enable identification of all 20 amino acids and to increase the identification fidelity by extending the time that amino acids spend within the nanopore. Successful outcome of this project will pave the way toward the first single-molecule protein sequencing method.

RESEARCH CHALLENGE

Cells express many thousands of proteins to perform a diverse set of complex tasks that are essential to the health of a living organism. A technology providing inexpensive yet sensitive and quantitative identification of the proteins in a cell will help numerous researchers elucidate biological processes, including the molecular origins of many diseases [1]. Existing protein binding and precipitation assays are relatively low-throughput, and their application is typically limited to detection of a few predetermined protein species [2,3]. However, protein sequencing methods aim to identify and quantify proteins and posttranslational modifications in a high-throughput manner without knowing *a priori* which proteins might be in a sample. The sequencing method has to be robust as even a single mutation in the amino acid sequence can lead to devastating diseases.

Sequencing a protein is not as straightforward as in nucleic acids because of the presence of 20 amino acids as compared to just four in DNA as well as the ability of protein to form complex 3D structures. Furthermore, protein function may sensitively depend on small chemical modifications to its residues, which

are difficult to characterize using existing experimental methods. Development of a method to sequence proteins as rapidly and accurately as scientists now sequence nucleic acids will revolutionize proteomics and provide the ultimate diagnostic tool.

METHODS & CODES

To determine the molecular mechanism of the blockade current modulation in aerolysin, the researchers simulated ion transport through aerolysin using the all-atom MD method. The aerolysin pore was placed in a DPhPC (1,2Diphytanoyl-sn-glycero-3-phosphocholine) lipid bilayer and solvated in a 1 M potassium chloride solution. The total number of atoms in the system was roughly 450,000.

The electrostatic potential map of the channel revealed the presence of a constant potential compartment in the middle section of the aerolysin stem, separated from the *cis* and *trans* entrances of the channel by electrostatic potential barriers. Building on this observation, the team simulated peptide transport through the aerolysin pore. In these simulations, the transport of peptides is facilitated by external forces that help the peptides to overcome local translocation barriers [4]. Averaging over all such forces at different positions in the pore yields the free energy landscape of peptide translocation through the pore [5]. All MD simulations were carried out using NAMD [6].

RESULTS & IMPACT

This ongoing project has already determined a rough free-energy landscape underlying peptide transport through an aerolysin nanopore. Repeat MD simulations have identified the location of the two major barriers to peptide transport, one each at the entrance and exit of the aerolysin nanopore. Importantly, the simulations have shown that the nanopore volume confined by the two traps acts as a single molecule trap that fully confines the translocating peptide. By combining the resulting all-atom MD trajectories with their steric exclusion model of ionic current blockades [7], the researchers have shown that the long residence of the peptides within the nanopore trap gives rise to well-defined and highly reproducible ionic current blockade values. Equipped with the knowledge of the microscopic mechanism underlying the ionic current blockades in aerolysin, the team is now in the position to suggest point mutations to aerolysin nanopores that would further increase the fidelity of amino acid identification.

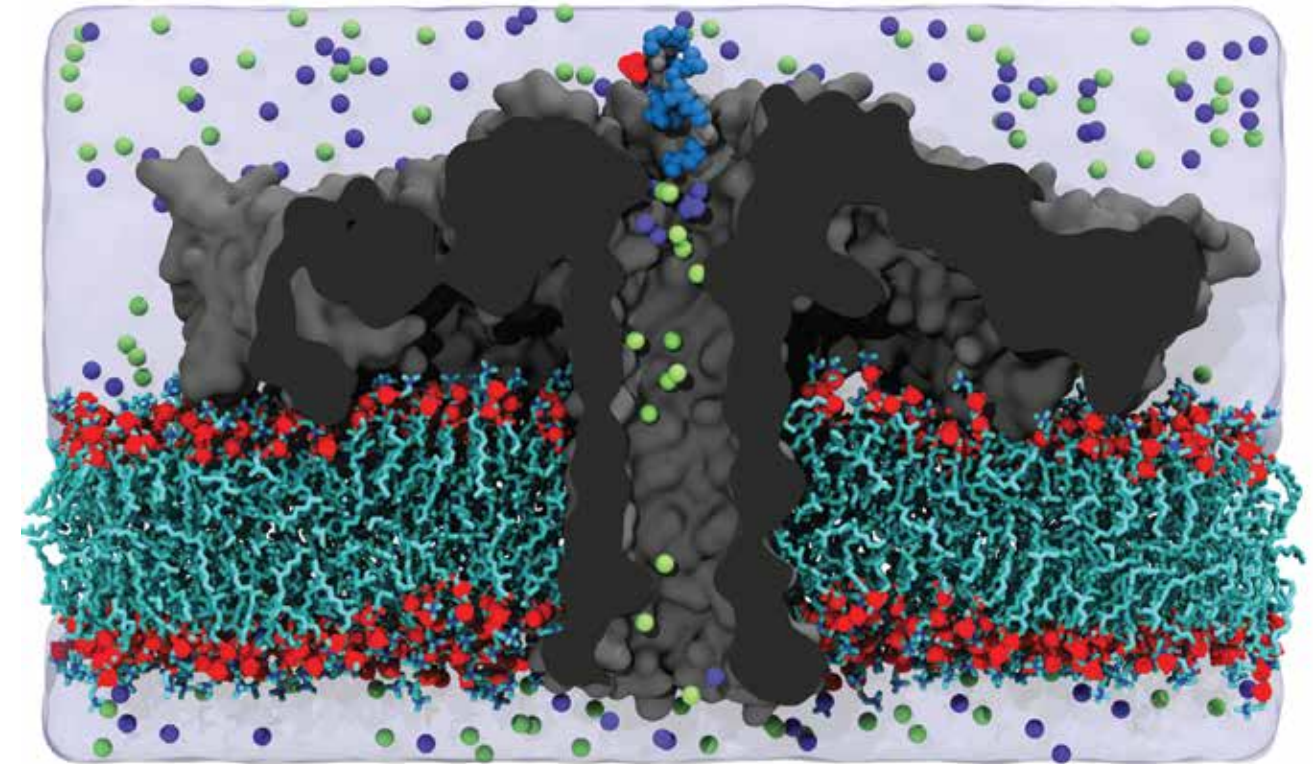


Figure 1: Nanopore sequencing system consisting of an aerolysin channel (cutaway molecular surface) embedded in a lipid membrane (cyan and red) and submerged in electrolyte (blue semitransparent surface, green and purple spheres). A single amino acid analyte (red) attached to a carrier peptide (blue) is captured by the aerolysin nanopore.

WHY BLUE WATERS

To get an accurate estimate of the ionic current and fast convergence of the free-energy landscape, the team needs explicit-solvent all-atom MD simulations that are run in parallel for exchange of structure information between different windows. Because of the long timescales and parallel MD simulations of a 0.5-million-atom system, this project is computationally demanding. The Blue Waters petascale system is one of a few supercomputers in the world with the computational power sufficient to carry out fully atomistic enhanced sampling simulations of peptide translocation through a nanopore. The large number of XK nodes on Blue Waters with graphics processing unit accelerators connected by the fast Gemini interconnect makes it one of the best publicly available systems for performing large-scale parallel MD simulations of protein systems.

PUBLICATIONS & DATA SETS

H. Ouldali *et al.*, “Electrical recognition of the twenty proteinogenic amino acids using an aerolysin nanopore,” *Nat. Biotech.*, vol. 38, pp. 176–181, 2020.