DI

BI

DETECTION OF AMINO ACIDS WITH HIGHLY SENSITIVE MOS_2 NANOPORES: TOWARD MACHINE LEARNING-BASED PREDICTIVE MODELS

Allocation: Illinois/135 Knh
PI: Narayana Aluru[1]

[1]University of Illinois at Urbana-Champaign

EXECUTIVE SUMMARY

Many human diseases such as cancer, diabetes, and digestive disorders can be caused by malfunctioning ribosomes that synthesize defective proteins. In this regard, protein sequencing, or the precise identification of the exact sequence of amino acids that comprise a protein, promises to enable breakthrough advances in the early diagnosis of such diseases. Using the Blue Waters supercomputer, we performed extensive molecular dynamics simulations and obtained 66 microseconds (μs) of atomic trajectories that demonstrate that a single-layer molybdenum disulfide (MoS_2) nanopore can be used to detect and accurately identify individual amino acids in a polypeptide chain. With the aid of machine-learning techniques, we identified key features and clustered both the ionic current and the time each of the 20 standard amino acids spends in the nanopore and identified unique fingerprints of the signals. Using advanced machine-learning classification techniques, we predicted the amino acid type of over 2.8 million hypothetical sensors.

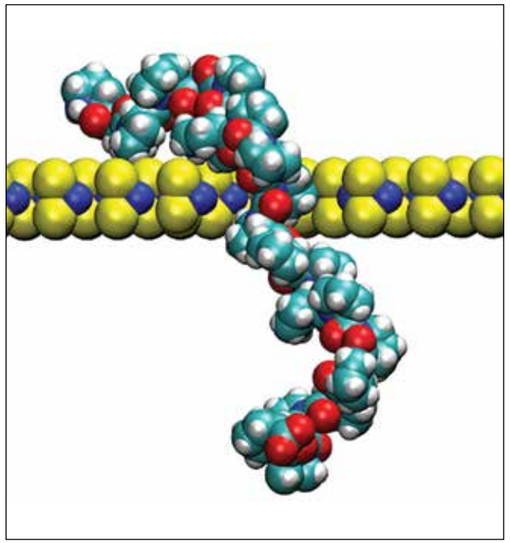

Figure 1: A snapshot of proline polypeptide translocation through the MoS_2 nanopore.

RESEARCH CHALLENGE

DNA sequencing using nanopore technology has significantly evolved over the last few years. Oxford Nanopore Technologies, for example, is currently fabricating a USB device called MinION that can sequence DNA in a matter of just a few hours. Both biological and synthetic nanopores have already been employed for label-free, high-resolution sequencing of DNA. Sequencing of proteins is another active area of research that promises to bring even more advances to personalized human healthcare. The main challenges associated with identification of biological molecules using nanopores are the low signal-to-noise ratio, pore degradation, unique identification of individual molecular units in real time, and the high speed of molecular movement through a nanopore [1,2]. Designing biological and synthetic nanopores with predefined properties for molecular transport is one of the most challenging problems in biotechnology. In this study, we showed that a nanopore in an ultrathin MoS_2 membrane can be used to detect and identify all of the 20 standard amino acids in the translocating proteins.

METHODS & CODES

We performed molecular dynamics simulations using the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS), an open-source classical molecular dynamics code for simulation of matter in liquid, solid, and gas phases. In our simulations we employed three different interatomic potentials: Tersoff, Lennard–Jones, and long-range Coulombic potential. Each simulation box comprised of about 32,000 atoms contained a monolayer membrane of MoS_2, an amino acid chain, water molecules, and ions. The amino acid chain was pulled through a nanopore using an external force. Fig. 1 shows a proline chain translocating through the nanopore in the MoS_2 membrane.

RESULTS & IMPACT

In this study, we showed that a nanopore in a single-layer membrane of MoS_2 can be used to detect and accurately identify individual amino acids in a polypeptide chain. Using extensive molecular dynamics simulations with a total simulation time of 66 μs, we identified key features and clustered the ionic current and the time each amino acid spends in the nanopore. The amino acids were then clustered into different groups based on their physical properties (e.g., size, polarity, and hydrophobicity). Using machine learning, we classified amino acids for any future ionic current and residence time readouts. We found that Logistic Regression, Nearest Neighbor, and Random Forest machine learning classifiers result in predictions of amino acid types with accuracies of 72.45%, 94.55%, and 99.6%, respectively.

Protein sequencing is an active area of research that promises to enable the early detection of cancer and other diseases. In fact, proteomic fingerprinting is believed to be as crucial for determining the health status of a human body as genomic sequencing. The proposed high-precision, single-base resolution and fast biomolecular sequencing using nanopore technology can lead to the fabrication of inexpensive personal healthcare devices that will help provide targeted healthcare. This will enable the quickly emerging medical fields of predictive and personalized medicine and will mark a significant leap forward for clinical genomics and proteomics.

WHY BLUE WATERS

We performed 4,103 molecular dynamics (MD) simulations of systems of up to 50,000 atoms and obtained 66 μs of molecular trajectories. Such expensive computations would not have been possible without a petascale supercomputer such as Blue Waters. LAMMPS, the MD package we use in our simulations, scales almost linearly with the number of cores on Blue Waters.

PUBLICATIONS & DATA SETS

Barati Farimani, A., M. Heiranian, and N.R. Aluru, Identification of Amino Acids with Sensitive Nanoporous MoS2: Towards Machine Learning-Based Prediction. *npj | 2D Materials and Applications*, DOI:10.1038/s41699-018-0060-8.