Nanopore-Based Proteome Profiling Within Reach, Study Suggests

May 09, 2017 | Adam Bonislawski

Premium

NEW YORK (GenomeWeb) – Researchers at the University of California, San Diego and the University of Notre Dame have developed a computational approach to nanopore-based protein identification that suggests large-scale proteomic profiling via nanopores could be possible.

Described in a study published this week in *PLOS Computational Biology*, the approach identifies proteins by analyzing the distinct electrical signals produced when these molecules pass through a nanopore, and could, in theory, allow researchers to characterize large numbers of proteins in complex mixtures, said Pavel Pevzner, professor of computer science at UCSD and senior author on the study.

The idea is not a new one, as a number of researchers have for years been pursuing nanopore-based protein analysis, which, were it realized, could enable protein measurements with single-molecule level sensitivity. However, while the field has met with some success, analysis of complex protein mixtures akin to what is currently done using techniques like mass spectrometry has remained a distant goal.

For instance, last year a team led by Michael Mayer, chair of biophysics at the University of Fribourg, used a nanopore to distinguish between a glucose-6-phosphate dehydrogenase protein and a G6PDH protein complexed with an IgG antibody. Mayer noted upon publication of the study that, while he hoped nanopores would ultimately prove suitable for large-scale protein characterization, such a development was still a long way off.

In 2014, a team led by Oxford Nanopore co-founder Hagan Bayley (working independently of the company), demonstrated the ability of a nanopore sensor to distinguish between differentially phosphorylated forms of the protein thioredoxin.

Like Mayer, Bayley at the time suggested that while nanopores might be used for highly targeted protein analysis in the relatively near term, characterization of complex protein mixtures was a longer term project.

"We aren't even close to doing that at the moment," he said. "I wouldn't say it's an impossible goal, but it is a bit of a stretch."

Pevzner said, though, that the *PLOS CB* study indicates that nanopore-based analysis of relatively complex protein mixtures could be closer than previously thought.

The key, he said, is using machine learning to analyze the information proteins generate when they translocate through a nanopore. Applying machine learning techniques, the researchers were able to identify distinct signals that Pevzner said could enable large-scale nanopore protein analysis.

His current optimism stands in contrast to his feelings shortly after embarking on the project. Pevzner came to nanopores after spending years developing tools for mass spec-based proteomic analysis — top-down proteomics in particular — and he said in comparison to that field, nanopore-based proteomics appeared intractable.

"In the beginning of this project, the data was so noisy that we almost thought we should give up," he said. "I have been working for almost 10 years now on top-down mass spectrometry, and in comparison with protein identification by top-down mass spectrometry, which by now is almost a mature area, it looked like there was no hope that nanopores could produce a comparable signal."

However, Pevzner said, when he and his colleagues applied machine learning, and random forest analysis, specifically, to their work, "all of a sudden the structure of the signal emerged."

When molecules pass through a nanopore, they alter the electric current across the nanopore. These changes in current will be different depending on the characteristics of the molecule passing through the nanopore, making it possible, in theory, to identify molecules by registering the current changes.

Oxford Nanopore has successfully applied nanopore sequencing to nucleic acid analysis. Protein work, though, is still at a much earlier stage.

"This field is at the very beginning," Pevzner said, noting that this presented certain challenges to his group's efforts, particularly in terms of accessing nanopore data they could use to develop their computational approach.

"Many people, including us, have analyzed DNA nanopore data … but few people have access to [protein] data," he said.

Even after he and his colleagues had obtained data to work with, "it took us a year, roughly, to figure out what was the best technique to use," Pevzner said. "There are various approaches, and some are not as good as others, but ultimately with the random forest model it became clear that the signal is there."

Also key to the method, Pevzner said, is using data from a number of protein translocations, which helps further reduce noise.

"When you have two, three, 10 spectra, the noise cancels itself out," he said.

Pevzner said he envisioned that nanopore-based proteomics experiments would be structured essentially the same as mass spec-based analyses, with researchers building reference databases of protein nanopore "spectra" to which they then match their experimental data.

"You run a sample, and instead of, say, top-down mass spectra, you generate nano-spectra," he said, noting that, though the data produced by the nanopore is not technically a "spectrum," he and his colleagues have termed it such given their mass spec background.

"Then you have a protein database, and you simply compare the nanospectras you have generated [to those in the database], and you find the best match," he said. "Then you compute the statistical significance, and if the significance is good you report it as your identification. The protocol is the same as mass spec, though the way you compute the signal is completely different."

He said that the quality of matches he and his colleagues are able to achieve currently indicate that their approach could be used for protein mixtures containing in the range of 1,000 or fewer different proteins.

"For any sample with 1,000 proteins, you need [matches to have] a p-value of 10^{-5} or lower," he said. "Our p values go to 10^{-5}, which means it is basically ready for simple protein mixtures, like, say, a bacteria proteome."

Pevzner said, in fact, that Notre Dame researcher Greg Timp, his co-author on the *PLOS CB* paper, plans to use the approach to generate a "large-scale protein nanopore dataset" in the next several months.

The field is still in its infancy, Pevzner said, but, he added, mass spec-based proteomic techniques also began modestly.

"If you look at mass spectrometry and the classic SEQUEST [paper from Scripps Research Institute professor] John Yates, it all started with a ridiculously small number of mass spectra, and look at how this has expanded to billions of spectra," he said. "So, this is where [bottom-up mass spec] was 20 years ago."

Filed Under   **Proteomics & Protein Research**

We recommend

UCSD Team Develops De Novo Sequencing, Dereplication Method to Decipher Nonribosomal Peptides
GenomeWeb, 2009

UCSD Team Frees Protein Identification From Database Searching Via Spectral Networks
GenomeWeb, 2007

Researchers Develop Improved Method for Modeling Protein Complexes via Crosslinking MS
GenomeWeb, 2016

ASU's Lindsay Broadening Recognition Tunneling Efforts to Include Protein Analysis
GenomeWeb, 2014

Nanopores May Enable Single-Molecule Protein Analysis, New Study Suggests
GenomeWeb, 2016

Young Investigator Profile: Sangtae Kim
GenomeWeb, 2015

Powered by