



Proteomics sets up single-cell and single-molecule solutions

 Check for updates

New instruments are emerging in single-cell and single-molecule proteomics. This will offer labs a choice of technologies.

By Vivien Marx

Mass spectrometry instruments are being engineered for labs working with tiny samples such as those in single-cell proteomics. Separately, benchtop instruments are emerging to sequence proteins in a massively parallel, single-molecule way¹.

Both approaches aspire to deliver information about the identity, quantity and dynamic range of proteins in samples faster; in higher throughput; with greater depth, sensitivity and quantitative accuracy; and in easier-to-use ways than currently typical. New centers are setting up machine parks focused on single-cell proteomics, such as at the Parallel² Technology Institute initiated by

Northeastern University researcher Nikolai Slavov with support from philanthropies such as Schmidt Futures. Investment in this domain overall is high and job hunters can find ads aplenty.

One protein sequencing company, Quantum-Si, has launched its Platinum instrument. Nautilus Biotechnology, to give scientists a sense of its instrument, is running samples from scientists who entered a First Access Challenge and returning data to them.

For now, there's a parallel universe of sorts between the mass spectrometry-based single-cell proteomics and non-mass-spec single-molecule methods. These areas and approaches "don't yet overlap," says Brigham Young University researcher Ryan Kelly.

The mass spec-based approaches still lack single molecule sensitivity, he says, and single-molecule sequencing hasn't been applied to single cells. "It is an obvious direction towards convergence, but we are not there yet," he says.

Says Matthias Mann, a proteomics researcher at the Max Planck Institute of Biochemistry, the "tech bubble" has meant great funding opportunities for the new protein sequencing companies. Among the challenges they face, he says, is the "dynamic range problem." With plasma, for example, for each protein of interest they would have to sequence millions of albumin proteins. "So I am not shaking in my boots until I see some convincing evidence," he says. Protein

sequencing has yet to deliver breakthroughs, but he doesn't rule out future ones.

Coming onto the scene

In single-cell proteomics with mass spec, changes in experimental design, sample preparation and approaches to protein separation have delivered dramatic advances in measurement sensitivity, says Kelly. The instruments themselves, such as those from Thermo Fisher Scientific and Bruker, have improved too, "so I think single-cell measurements will be accessible to any proteomics lab in the next few years."

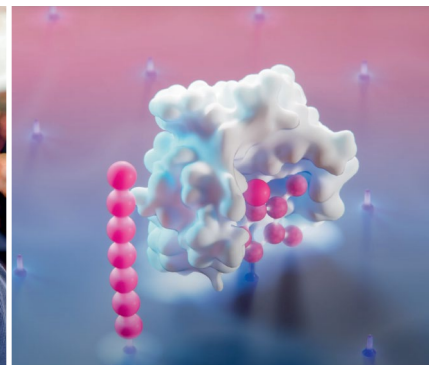
Single-molecule protein sequencing companies include Erisyon, Glyphic Biotechnologies, Nautilus Biotechnology, Oxford Nanopore Technologies (ONT) and Quantum-Si. Genomics has "taken off" in ways proteomics has not, but this disparity can change, says Stanford University researcher Parag Mallick, chief scientist at Nautilus Biotechnology, which he co-founded. Just as any biologist can now perform genome sequencing, he hopes the same will become true with protein analysis.

The future is not one in which every biologist has to be a mass spectrometrist, says Joshua Yang, co-founder and CEO of Massachusetts Institute of Technology (MIT) spinout Glyphic Biotechnologies, which is readying its benchtop protein sequencer. Given that mass spec is complex and takes skilled operators, it's not analysis a biologist can do on her or his own. Existing proteomic analysis systems are not well adapted to analyzing multiple proteins at once or proteins at low concentration, says Yang. "Generally, systems today can only do one of the two," he says. A lab studying neurological disease with mass spec needs to know which proteins to expect in a given sample. With protein sequencing, it can detect all the different types of proteins, including rare proteins. A single-molecule detection system is, he says, "inherently the best sensitivity you can get, right?"

Mass spec is indeed expensive and specialized, says Kelly, which is why protein sequencing companies can generate excitement by promising alternatives. But, he says, "I believe rumors of mass spectrometry's demise have been greatly exaggerated."

Protein sequencing promise

Erisyon applies an approach from the labs of the University of Texas at Austin scientists Edward Marcotte and Eric Anslyn². Jag Swaminathan, a former postdoc in the Marcotte lab, co-founded the company and serves as its chief technology officer. In the



Joshua Yang (left) and Daniel Estandian co-founded the MIT spinout Glyphic Biotechnologies. Small molecules conjugate to the N-terminal of a tethered peptide and tether it to the surface. The peptide's last amino acid is cleaved but remains tethered. A fluorophore-conjugated antibody attaches and generates a signal.

platform, every sample goes through multiple rounds of Edman degradation, which is classic protein sequencing chemistry that dates to the 1950s. As Erisyon's CEO Talli Somekh says, first a sample's proteins are digested into millions of peptides around 15 amino acids long. Fluorophores label select amino acids – for example, the cysteine signal is blue, tryptophans are yellow-orange, carboxylic acids are red. The peptides are immobilized on a glass surface via their C termini. The flow cell is imaged and then exposed to Edman chemistry to cleave the N-terminal amino acid.

Imaging reveals what has changed. For instance, says Somekh, a step-drop in fluorescence in the yellow-orange channel indicates a tryptophan. The next round may show no step-drop, which indicates an unlabeled amino acid. The cycle repeats 15 times and generates a 'fluorosequence' that is run against a database to identify the protein the peptide originated from. To deduce identity, not all amino acids need to be labeled.

In the Glyphic Biotechnologies instrument, denatured proteins are tethered to a glass surface. Next comes ClickP chemistry: a small molecule conjugates the tethered peptide at its N-terminus to the surface, and a change in buffer conditions leads the amino acid at that terminus to be cleaved off, says Yang. It stays tethered on the surface away from the original protein. A fluorophore-conjugated antibody attaches to this amino acid and generates a signal specific to the cleaved amino acid. The cycle repeats.

Protein sequencing gives researchers "greater confidence than any other technology," says Yang. Protein quantification comes from raw read counts of the tethered and sequenced proteins. The instrument also

delivers details about post-translational modifications (PTMs). A lab can use an antibody raised to detect a particular PTM. Multiple PTMs on the same protein, such as those on a cancer-associated protein with co-occurring PTMs, can be detected. "That cannot be easily done with technologies today, because mass spec fragments a protein entirely," says Yang. Glyphic is developing a bioinformatics suite with two-tiered data output. One readout is a list of the identified and quantified proteins. Another tier lets labs dig deeper into the results bioinformatically.

He and his team at Nautilus Biotechnology thought about how best to capture the dynamic range of proteins on a single-molecule level with a high-throughput device, says Mallick. The technology needs to be "incredibly sensitive so you can capture those very rare events," he says.

The platform's flow cell is a nanofabricated protein array that captures intact protein molecules. Each protein lives "in its own independent universe," says Mallick. It's tethered in place on the array, which "changes the nature of the quantification problem," he says, because each peptide is counted. "At the single-molecule level, identification is quantitation," he says. The instrument probes the proteins in multiple rounds of optical and fluidic measurement using multi-affinity protein binders and machine learning algorithms. A result can show that a protein is a tau 2N isoform; another readout might indicate a protein is phosphorylated at amino acids 181 and 214. Between each probing cycle, there's a wash.

Because the protein stays intact, each molecule can be probed repeatedly, says Mallick, which is not possible with other approaches.

One lab might explore the difference between serum troponins in the context of heart disease research; another might use the instrument to find proteoforms, which are variants of proteins born from the same gene, en masse at the single-molecule level. Such work could contribute to the [Human Proteoform Atlas](#), initiated by Northwestern University researcher Neil Kelleher and colleagues.

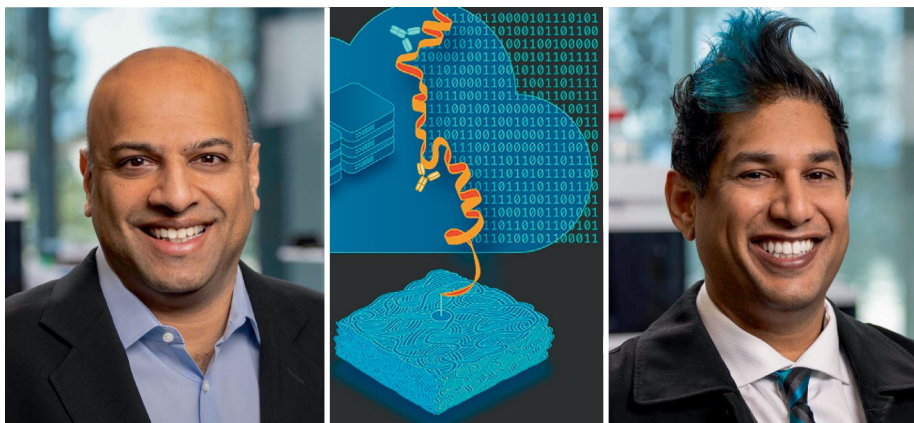
The company's proprietary probes recognize short epitopes, typically three or four amino acids in length. Given the long-standing challenges around antibody specificity, says Mallick, the instrument is designed to ask more general questions of a protein. As the reagents probe the short protein epitopes, a pattern of epitope binding is built. The instrument's machine learning algorithm infers which protein is compatible with this pattern of binding. "Even though each probing itself is not super-specific, the combination of asking all of those questions is shockingly specific," he says.

Quantum-Si's instrument, Platinum, is being shipped. The idea behind the technology³, says Brian Reed, who heads research at the company, is to make protein sequencing accessible to the average research lab with a small-footprint, low-cost instrument. Scientists might use Platinum to assess proteins of interest and focus on specific targets in the lower end of the protein concentration's dynamic range or use it to compare samples in terms of the relative abundance of proteins or proteoforms.

The process starts by digesting proteins into peptides. Next comes a single reaction without any exchange of fluids on the nanofabricated CMOS chip on which the peptides are immobilized in individual nanowells, says Reed. Each nanowell holds a single molecule. Built into the well is optical waveguide circuitry.

A solution flows in with dye-labeled recognizers – proprietary N-terminal amino acid binding proteins – and aminopeptidases. In a sequencing run, the aminopeptidases cleave one amino acid at a time. After each cleavage event, the recognizers bind and unbind each newly exposed N-terminal amino acid. The chip records these events. "This real-time approach eliminates the need for complex and expensive chemistry and fluidics associated with cyclical approaches like Edman degradation or iterative probing with affinity reagents," says Reed.

To distinguish the different recognizers, the instrument uses not wavelength but fluorescence lifetime, a property of dyes that corresponds to the time a molecule spends in the



Stanford University researcher Parag Mallick (right) co-founded the protein sequencing company Nautilus Biotechnology with entrepreneur Sujal Patel. The instrument probes proteins immobilized on an array with multi-affinity protein binders. A pattern of binding to short epitopes emerges, and a machine learning-based algorithm infers which protein the pattern is compatible with.

excited state before it emits a photon. Using fluorescence lifetime makes design and manufacture of the instrument and the CMOS chip easier, he says. The instrument can distinguish among numerous fluorophores, and no optical filter layer on the CMOS chip is needed.

The kinetics of on-off binding events enables PTM detection because each PTM type influences the on-off binding properties in a characteristic way that the software captures. By deploying kinetics, one avoids the need for a new recognizer for every PTM, which, he says, keeps the system simple and maximizes the peptide sequence information obtained.

New and renewed

ONT's sequencers work with DNA and RNA, and the company is readying the platform for protein sequencing with various chemistries. It's straightforward to put a protein through a nanopore as long as, for example, it's charged, but "proteins are difficult," said chief technology officer Clive Brown in a presentation. Some are uncharged; proteins are folded and may be modified with sugars, all of which makes for "tricky complications."

In ONT's approach to nanopore sequencing, when a single strand of DNA moves through a nanopore, the translocation through the pore is controlled by an enzyme, a DNA motor. Once passed through the nanopore, the motor protein lets go of the strand and the pore is ready for the next fragment. One approach ONT researchers have been tinkering with is to use a DNA-peptide conjugate and the motor already in use for DNA and RNA sequencing. They want to get a distinct signal for each

peptide, which is the "core hard problem in protein sequencing," says Brown. But "we've got it."

Although nanopore sequencing is still in its infancy and she has not yet seen de novo protein sequencing with a nanopore platform, once that feat is achieved, "it will unlock the next generation of proteomics," says Jayde Aufrecht, a research fellow in the Department of Energy's Environmental Molecular Sciences Laboratory at Pacific Northwest National Laboratory. "Specifically, this would increase the identification of low abundance proteins." Aufrecht's research sits at the intersection between nanotech and ecology; she studies microbial communities and characterizes host-microbiome dynamics.

She's watching work in labs such as Cees Dekker's at the Delft University of Technology and Shuo Huang's at Nanjing University, who use nanopore platforms for reading peptide sequences.

Aufrecht and her colleagues have a project underway that uses nanopore sequencing to identify functional classes of proteins. Oligonucleotides modified for click chemistry are linked to activity-based probes. The oligos, when read by the sequencer, act as barcoded reporters. "The idea is that a protein will covalently bind to an activity-based probe if that probe targets the protein's specific function," she says. The approach is not direct protein sequencing, but once that is possible, this assay would offer a protein sequence and identify the protein's functional class at the same time. This could, for example, be a way to annotate unknown proteins.

The team devised workarounds given that ONT's platform is optimized for DNA and RNA and tweaked the sample preparation, "but if we can get it working then I think the results will be worth it," she says. It will enable high-throughput analysis of protein function and help to characterize microbial phenotypes in environmental samples.

Researchers are engineering nanopores in different ways, such as by chemically tailoring the nanopore's interior wall⁴. This converts the nanopore into a nanoreactor. At the Kavli Institute of Nanoscience at Delft University of Technology, among the projects Cees Dekker and his team are running are ones to engineer nanopores as sensing devices or use them to fingerprint and sequence proteins at the single-molecule level. "It is hard to beat the sensitivity of nanopores," says Dekker. Nanopores can measure single copies of a protein, which mass spec "sure can't." The nanopore size is reproducible. Nanopores made of proteins, as with ONT's approach, have many advantages, says Dekker. His group pioneered possibilities with hybrid pores in which biological pores are inserted into solid-state ones.

One issue for de novo protein sequencing with nanopores is that although the tightly spaced amino acids pass through one by one, around eight amino acids are probed simultaneously, says Dekker. A nanopore is exquisitely sensitive to the tiniest changes in amino acid sequence. The scientists were able to distinguish peptides with and without PTMs with no need for PTM-specific labels. The PTMs were three amino acids apart from one another, a detection task mass spec would not have been up to, he says.

For a recent preprint⁵, not yet peer reviewed, they detected post-translational modifications at the single-molecule level on immunopeptides found on the cell surface that have phosphorylation patterns associated with cancer. For analysis, the peptide of interest is assembled with a DNA molecule attached to either end and a DNA motor enzyme ratchets the molecule through the nanopore. The Dekker team has also pushed peptides through the nanopore multiple times to obtain multiple readouts of the same molecule⁶, an approach Dekker sees as "very, very important" for many applications given that it boosts readout accuracy to practically 100%. Development of true de novo sequencing of proteins with nanopores or other approaches will take longer to realize, he says.

It's a big job

No instrument can currently analyze the entirety of a single cell's proteome. But, says



A protein is born when genes lead to transcripts that encode amino acid sequences. Amino acids interact with one another and the string folds into three-dimensional shapes. The proteome's diversity is gargantuan.

Jennifer van Eyk, a Cedars Sinai Medical Center researcher, not all of a lab's questions will require knowing the entirety of a proteome. The specific task at hand will determine which technology and instrument to use. What's certain is that analyzing many single cells in parallel means getting a handle on gargantuan protein heterogeneity and numbers.

Each year, Van Eyk and colleagues analyze thousands of different types of samples in

which the focus is usually on between 5 and 1,000 proteins per cell. Even at that level, when analyzing many single cells, "the numbers become huge," she says. Van Eyk has her own lab, is separately associated with a core facility and, among other activities, evaluates technology more generally for research and clinical applications. Cells from the same tissue can contain quite different numbers of proteins. This is, she says, sometimes due to biology but can also be shaped by flow cytometry's shear forces. Two cells from the same tissue can differ by as many as half their proteins. As labs troubleshoot these issues, she is sure that single cells have much proteomic insight to offer.

Her lab studies the heart and its cells, which have been studied for decades. Much is known about the cells' electrophysiology, contraction patterns and mitochondria. "Yet the proteomes are not known," she says. Some findings about the proteomics of heart cells appear to be "breaking dogma," about these cells, she says. For single-cell proteomic analysis, she uses the Bruker timsTOF SCP mass spec system. Mass spec has long been used in clinically focused labs for analyzing metabolomes and proteomes. As she and colleagues move forward in single-cell analysis, she is glad to see improvements but believes "additional breakthroughs" are needed.

For her needs and those of her colleagues working at the intersection of research and clinic, methods need to be robust, offer high throughput and provide unambiguous protein identification and quantification across the sample's dynamic range. New entrants must face many non-technology-related issues, such as clinical approvals or ways to fit technology into a drug discovery and development pipeline. She has not yet tried a protein sequencer but is curious to know if they can handle complex mixtures and to learn whether samples have to be enriched in certain proteins.

Analyzing single cells by mass spec differs in a few ways from experiments with bulk samples. Generally, in mass spec, sample molecules are ionized, for example by electrospray, by exposing the liquid sample to high voltage. The ions reach the mass analyzer chamber in a vacuum. There, ions are separated according to mass-to-charge ratio and detected, and a digital readout is generated. Bulk samples deliver plentiful ions, but ion flux is lower with single-cell samples. Too few ions can mean too low coverage of a proteome.

Right now, single-cell proteomics analysis takes place mainly in labs devoted to

mass spec, says Christoph Krisp, an application scientist at Bruker. But he believes that mass spectrometers can migrate widely into research and clinical labs to help them delve deeply into the single-cell heterogeneity of tissues. Bruker has a range of instruments in the timsTOF series, which stands for trapped ion mobility spectrometry time of flight. One model images tissue at 5 micrometer pixel sizes, close to single-cell resolution. The timsTOF SCP instrument is for low sample amounts, and Bruker sees it as the “workhorse for single-cell proteomics,” says Krisp. The SCP in the name might hint at single-cell proteomics, but Bruker has decided to not spell it out as such because the instrument is for any lab with small samples, such as those who work in immunopeptidomics or phosphoproteomics.

The timsTOF SCP system emerged in collaboration with the Danish company Evosep and the Mann lab. Using a prototype, the Mann lab identified 1,400 proteins from a single cell sample. These days, he says, in the right conditions, they can get up to 4,000 from a single cell.

Says Mann, his lab didn't start this project aiming at single-cell proteomics. It happened as they working on instrument development to analyze single cell types in tissues, he says. “We were kind of taken by surprise by the interest in this.”

At the outset, his team found the Bruker timsTOF to be quite sensitive, he says. Together, they optimized it further and increased ion current four- to five-fold. “But we got an even larger increase from the chromatography,” says Mann.

“But we got an even larger increase from the chromatography,” says Matthias Mann.

The Evosep chromatography system uses low-pressure pumps to premix a gradient. The peptides are pushed over the analytical column by a single high-pressure pump. “This is a very robust system to run thousands of samples in a reproducible manner for clinical cohorts, for example,” he says. The team saw that this system could be tuned to decrease the flow rate from 1 microliter per minute, which is used in clinical analysis, to just

100 nanoliters per minute. The issue is that all other systems use binary pumps, which are tricky at low flow rates, he says.

According to electrospray theory, such a decrease in flow rate should increase sensitivity by a factor of ten. “And it did,” says Mann. Evosep's Evotips sample loading system has turned out, he says, to be “perfect for single cells.” Peptides can be eluted in ‘nanopackages’ of only 20 nanoliters, which obviates the need for specialized ‘nanoreaction chambers’ or the like. In recent work that has not been peer reviewed, he and his team characterized single cell shapes from tissue with this system⁷.

“It was indeed fun to see how all this came together without even having the single cells in mind in the first place,” says Mann. The team likes the system's robust performance and use it day in and day out. For the future, he believes the timsTOF SCP and Evosep system is suited to cancer diagnosis and treatment recommendations applying Deep Visual Proteomics⁸, which combines artificial-intelligence-driven image analysis of cellular shape, single-cell or single-nucleus laser microdissection and mass spectrometry. That, he says, is “our long-term goal.”

“If I wanted to do something different than MS-based proteomics, I would put my money on methods that have a direct deep sequencing and barcoding readout,” says Mann. Such an approach works, for example, “relatively well” in the proximity extension assay from Olink, which is a multiplexed immunoassay platform that uses low sample volumes and involves oligonucleotide–antibody pairs in which the oligos are reporters. An alternative might be a super-resolution imaging approach, says Mann.

Says Utrecht University proteomics researcher Albert Heck, “I love what is going on with the single-molecule sequencing companies,” and he and his lab collaborate with them. What these companies promise to deliver feels akin to what he and the mass spec-based proteomics community promised two decades ago, when there was excitement over biomarker discovery through mass spec-based proteomics with surface enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF MS) and other approaches. “We did deliver, although about 20 years later than initially proposed,” says Heck. “I feel it may be the same with the single-molecule sequencing world,”

he says. “But it is beautiful biophysics and biochemistry.”

It's remarkable, says Kelly, that scientists can use mass spec now to rapidly and robustly quantify more than 3,000 proteins from a single cell. It was a few hundred proteins just a few years ago. “But despite how far we've come, there is still so much more room to run,” says Kelly. Getting there will take improved separations: lower flow rates and narrower peaks will make a big difference. Mass spec data acquisition can be improved to more effectively use what the instrument can deliver. Data analysis, he says, can be “tailored to the unique characteristics of single-cell mass spectra.”

“Despite how far we've come, there is still so much more room to run,” says Ryan Kelly.

Mass spec and non-mass-spec protein sequencing technologies, says Yang, can be complementary. “I don't believe that protein sequencing is going to replace every instance of mass spectrometry that's out there,” he says. Each system has pros and cons. Instead of seeing one as the final choice, “we believe that there are different applications for different technologies,” he says.

Mallick also sees the technologies as complementary. The new protein sequencing technology lacks the maturity of mass spec and its three decades of development, he says. But silver stain gels for detecting proteins by gel electrophoresis didn't kill the Coomassie blue gel and “RNA-seq didn't kill qPCR,” he says. “Depending on the question, there's a tool.”

Vivien Marx ✉
Nature Methods.
✉ e-mail: v.marx@us.nature.com

Published online: 10 March 2023

References

1. Alfaro, J. A. et al. *Nat. Methods* **18**, 604–617 (2021).
2. Swaminathan, J. et al. *Nat. Biotechnol.* **36**, 1076–1082 (2018).
3. Reed, B. R. et al. *Science* **378**, 186–192 (2022).
4. Ying, Y.-L. et al. *Nat. Nanotechnol.* **17**, 1136–1146 (2022).
5. Nova, I. C. et al. Preprint at *bioRxiv* <https://doi.org/10.1101/2022.11.11.516163> (2022).
6. Brinkerhoff, H., Kang, A. S. W., Liu, J. & Dekker, C. *Science* **374**, 1509–1513 (2021).
7. Rosenberger, F. A. et al. Preprint at *bioRxiv* <https://doi.org/10.1101/2022.12.03.518957> (2022).
8. Mund, A. et al. *Nat. Biotechnol.* **40**, 1231–1240 (2022).