Bioscience Technology

Advantage Business Media
100 Enterprise Drive
Rockaway, NJ, 07866

E-mail to a colleague

Printer Friendly Format

Articles

New Methods Deliver High-Throughput Proteomics

by Gina Shaw

You could say that proteomics is to genomics what the New York City phone directory is to Cincinnati's. With "only" 22,000 genes, the human genome is a relatively easy tome to wrap your arms around not so the proteome, with approximately 400,000 proteins for which these genes code. If high-throughput methodologies were essential to mapping the human genome, they're even more critical to identifying, characterizing and understanding the functions of the proteome.

That starts at the mass spectrometry level the standard method for protein quantitation. "The development of MALDI matrix-assisted laser desorption/ionization mass spectrometry enabled the development of proteomics as a field, because without these two ionization methods, we really had no chance of detecting very large biomolecules in a mass spectrometer, which is the essence of proteomics," says Akos Vertes, co-director of George Washington University's Institute for Proteomics Technology and Applications. "What you find today in a proteomics lab are these two ionization methods, in combination with separation methods and the mass spectrometer, that allows you to do peptide mapping and de novo sequencing."

But MALDI mass spectrometry has its limitations centered largely around sample preparation. Vertes calls the process that creates the matrix protein sample "blender methods." "You blend some tissue into a juice, a soup, that is completely homogenized and you immediately lose all the spatial information. You can use some extreme experimental procedures to try to make up for that, but by and large, today's high-throughput proteomics does not provide spatial distributions."

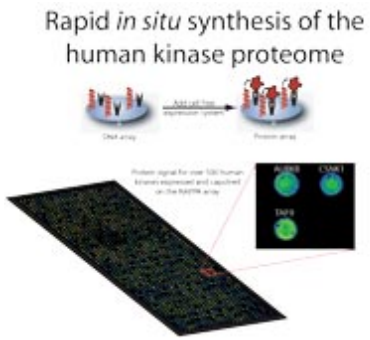
That's changed with the advent of imaging MALDI mass spectrometry, a process first developed by Richard Caprioli, Ph.D., Director of the Mass Spectrometry and Proteomics Research Center of the Vanderbilt School of Medicine. Frozen tissue slices on a MALDI target plate are thawed and fixed, then covered with a tissue crust and scanned, in the vacuum environment of the mass spectrometer, using MALDI equipment with a finely tuned laser focus, to collect spectra point-by-point.

Manufacturers have stampeded to offer the new technology. Applied Biosystems got out in front in 2004 when it released a novel tissue imaging technology for the QSTAR XL Hybrid LC/MS/MS System, developed in collaboration with Caprioli. Within a few months of each other in the spring and summer of 2006, Bruker, Applied Biosystems, Thermo, and Waters each announced the release of imaging mass spectrometry products.

Can imaging mass spec work in vivo?

But to enhance its utility in proteomics, Vertes wants to take imaging mass spectrometry to the next level allowing it to work in vivo. He points to what he sees as three major limitations in current imaging mass spec technology: the requirement for a vacuum, the matrix itself, and the limitations in spatial resolution.

"All the classic mass spectrometry has happened, for half a century or more, in a vacuum. We want to move away from that, because it's not conducive to life, and we want to work with systems in vivo," Vertes says. "The matrix itself is also an obstacle. There are about three to four



***Click* to enlarge.**

dozen matrix materials for MALDI mass spec, and people have to develop skills with them. Also, if we want to work in vivo if you mix an earthworm with a matrix, that's the end of the earthworm."

Vertes reports that his lab has solved the first two problems, with something called Atmospheric Pressure Molecular Imaging by Infrared MALDI Mass Spectrometry. The feasibility of atmospheric pressure MALDI has already been demonstrated; new subfemtomole detection limits and streamlined sample handling have helped to address issues of reduced sensitivity compared to vacuum MALDI. And by using an infrared rather than a UV laser, he notes, scientists can use the simplest of all matrices: water.

"As water is a native component of plant and animal tissues, the ability to perform AP IR-MALDI using water as the matrix enables the in vivo investigation of complex biological systems in their natural state." Their recent paper in *Analytical Chemistry* (Y. Li, B. Shrestha and A. Vertes, "Atmospheric Pressure Molecular Imaging by Infrared MALDI Mass Spectrometry," *Anal. Chem.*,2007, 79, 523-532) demonstrates this the feasibility of this technique, imaging the molecular distributions of the major water-soluble components in various fruits.

Naturally, there are limitations. "When we compare mass spectra taken using UV with an artificial matrix with our mass spectra, we see different things the first shows proteins and peptides, while we preferentially see metabolites," says Vertes. "So in a way, they are complementary, but right now, we are working on a method that allows us to also see the proteins, using only infrared and a natural water matrix."

A more technologically complex challenge one the lab is still working on is the limit in spatial resolution. The classical lens that focuses the laser beam in MALDI mass spec is, in principle, about 160-some nanometers. "In practice, you never reach the diffraction limits. You have maybe 80-micron focal spots, which is appropriate for tissue imaging, but not if you want to go subcellular," says Vertes.

He is now investigating an idea that landed his laboratory a grant from the Keck Foundation to combine the scanning near field principle with imaging mass spectrometry. "Instead of using a conventional lens to focus the light, you can launch the laser light into an optical fiber sharpened to a very fine point at the other end, down to 100 nanometers. If you bring this sharp fiber close enough to the surface of your sample, you can ablate tiny surface areas that can be submicron. But since you remove so much less material, how do you detect that?. We can ablate a submicron surface area, and we can do atmospheric pressure infrared molecular imaging. But how do we bring these two together? That's the next challenge."

Speeding toward structure and function

Identifying, quantifying and mapping proteins is just the beginning. High-throughput approaches to studying the structure and function of proteins are at the heart of today's proteomics efforts.

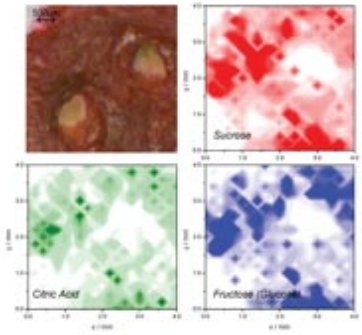


Figure 2. Optical image of a

"We're using robotics, information technology and efficient chemistries to ask questions about proteins and their throughput more efficiently than in the past thousands at a time," says Joshua LaBaer, M.D., Ph.D., director of the Harvard Institute of Proteomics. "It's like the old joke about the guy looking for his keys under the streetlight even though he dropped them in the gutter, because that's where the light is. The problem is that the light in molecular biology is shed on only about two or three thousands genes out of 25,000. If we don't start lighting up the field more broadly, then we're always going to be looking at the same area not necessarily where the keys are."

One way of "lighting up the field" is the use of protein microarrays. LaBaer's lab uses a familiar tool called functional/target protein arrays, which print the proteins themselves on a microarray to identify interactions, activities, binding, and so on. But the old-fashioned method of doing this purifying the proteins and spotting them directly on the array is very challenging, especially in high-throughput. "It's done, but

strawberry skin (top left) compared to AP IR-MALDI images for three major components, sucrose (top right), glucose/fructose (bottom right), and citric acid (bottom left) around embedded seeds. The false color intensities were assigned according to integrated areas of the associated potassiated peaks. Although the distributions of these water-soluble components are similar, the position of the two seeds, with lower concentrations of these compounds, is clearly discernible. Reprinted with permission from Atmospheric Presssure Molecular Imaging by Infrared MALDI Mass Spectrometry, Li, Y., Shrestha, B. and Vertes, A. Analytical Chemistry 79(2) 523-532 (2007). Copyright 2007, American Chemical Society. Click to enlarge.

it's expensive and you run the risk that the proteins you've just purified are going to slowly denature on the array," LaBaer says.

So his lab has developed a chemistry that allows them to print the gene for a protein directly on the array and then synthesize the protein in situ. "We attach the gene to the surface, and at the same time, a capture agent that will recognize the tag on the protein when it's made. We also add an extract that has a transcription and translation mix to the array, which will transcribe and translate the protein directly on the array surface," LaBaer says.

That approach first described in *Science* in 2004 works quite well in his academic lab, LaBaer says, but the big challenge was scaling it up to a high-throughput level. Unlike the DNA microarray world, which is pretty much "plug and go" at this point, protein microarrays of the type he is working on have little supportive infrastructure. "We're so new at this, there are no devices we can use, because no one else is doing it other than our collaborators. Every step we want to take, we need to build an instrument for it."

But after years of work, the goal is in sight. "We can make pretty high-density arrays and get very efficient protein translations we can get proteins made on 95% of what we print and upwards," LaBaer reports. "Now, with our high density and a good yield, we can convert this into functional experiments looking at serum responses to proteins, functional activities of proteins on the array, and protein-protein interactions."

Centering on protein families

Of course, the thousand-pound gorilla of proteomics these days is the Protein Structure Initiative, a ten-center, ten-year effort to determine 10,000 protein structures, launched by the National Institutes of Health's Institute of General Medical Sciences in 2000. Four larger-scale centers are expected to coordinate their efforts to produce the structures of more than 4,000 protein families, while six specialized centers are in the process of developing innovative methods for producing and determining the structures of proteins that traditionally have been difficult to study. So far, more than 1,700 structures from both simple and complex organisms have been deposited in the Protein Data Bank (www.pdb.org), an international structural database.

Two new public resources have just been added to the initiative: the PSI-Materials Repository and the PSI Knowledgebase. Led by LaBaer at the Harvard Institute of Proteomics, the PSI-MS will store and ship PSI-generated clones, which can be used to make specific proteins for studies on their structure and function. Researchers will be able to order clones for a minimal fee to cover processing, handling, and shipping. Launched in November, the PSI-MS is expected to be fully operational by fall 2007. Around the same time, the initiative will launch a "Knowledgebase" a headquarters for structural information generated by its centers. For every protein, scientists will be able to find the best available information about the structure and biological function, and experimental details about each stage of the protein structure process.

One of the Initiative's specialized centers is the New York Consortium on Membrane Protein Structure, headquartered at the New York Structural Biology Center, a cooperative research center founded by ten leading New York research institutions. Determining membrane protein structure is particularly challenging because of the difficulties involved in obtaining sufficient quantities of these proteins in their purest forms an essential step in visualizing their three-dimensional shapes.

"The hard part is to find the conditions that will stabilize a protein when it's extracted from the membrane. That requires using detergent molecules to do that, and you usually need to try a variety of detergent molecules," says Wayne Hendrickson, Ph.D., a professor of biochemistry and molecular biophysics at Columbia University and the principal investigator of the membrane protein structure consortium. "This normally means that you must go through a lot of effort to produce and purify the protein and carry out this analysis."

But the New York consortium has developed a new approach using a fluorescence assay. Attaching a green

fluorescent protein reporter, or a variant, to the c-terminus of a protein so that it will be dominantly cytoplasmic, the center is making proteins from families using a structural genomics approach. "We're taking several sequences that are members of the family from a variety of organisms or paralogs, and then producing them, each tagged with a fluorescent protein. As a result, you can carry out size exclusion chromatography on the unpurified material, which makes the process extremely efficient. You can see your protein and its properties in the midst of everything else in this separation technique, and from that, in quite a high-throughput manner, you can identify those proteins that will be advantageous in bringing forward in scale-up."

The vast size of the proteome renders it virtually impossible at least with today's methodologies to obtain all the protein structures associated with the human genome (or any other genome). But that may not be necessary. "There's a relationship among the products of different genomes, such that if you know the sequences and have organized them into sequence families, at a certain level of sequence similarity there's a pretty high probability that the folded structure will be similar across that family," says Hendrickson. "So we can, somewhat realistically, aspire to an ultimate determination of representatives of all of the families of sequences from all of the organisms, and all of life."