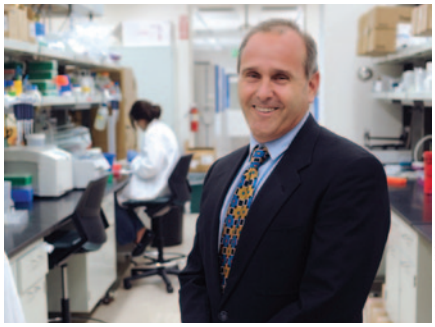
Targeted proteomics

Analysis of a preselected group of proteins delivers more precise, quantitative, sensitive data to more biologists. Vivien Marx reports.

Although the number and identity of protein-coding genes in humans and many other organisms are known to a certain level of approximation, the numbers of proteins produced by each of these genes remains a mystery. Further complicating matters, given the many possible splice forms and post-translational modifications, the potential number of proteins is "staggering," says Arizona State University researcher Josh LaBaer, who is also president-elect of the US Human Proteome Organization. A protein is also dynamic. "It's phosphorylated this minute; it's not phosphorylated the next minute," he says. This is fascinating science, but it makes proteins in a complex, dynamic sample hard to precisely measure.

Understanding disease-related changes, for example, calls for reliable, quantitative ways of assessing protein levels, and mass spectrometers are instruments able to nail that task. But the data from so-called discovery proteomics experiments in which mass spectrometry is used to identify a large number of proteins in a sample are not always useful to biologists. Enter targeted proteomics, in which the analysis focuses on a subset of proteins of interest in a sample—an approach that has been steadily gaining traction over the last few years.



J. LaBaer/ASU

Targeted proteomics points the field toward high-throughput biology, says Josh LaBaer.

"I personally can't wait until we stop hearing about someone describing how big of a list of proteins, peptides or phosphopeptides they detected," says one researcher critical of discovery proteomics who did not wish to be identified. Proteomics has been doing "my list is bigger than your list" for far too long. "It is way more important to measure the one right protein than 10,000 wrong ones."

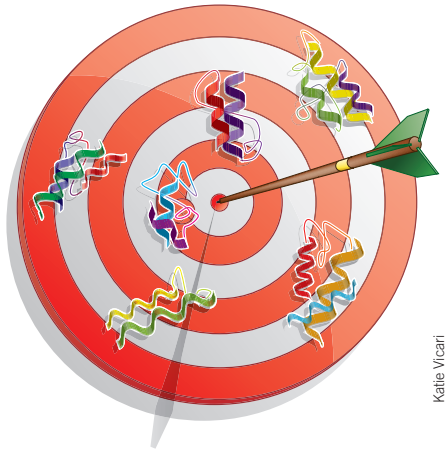
Scientists wanting to follow well-founded hunches about dozens or hundreds of proteins seek a focused, reproducible, quantitative view of a small subset of the whole proteome in their lab vials. High-throughput biology experiments, which include DNA sequencing, genome analysis and gene expression analysis, are generating massive data sets pertaining to particular genes and pathways active in disease or in signaling processes of interest. The shifting of proteomics closer to data mountains in biology is taking it "exactly where it needs to go," says LaBaer. Targeted proteomics helps researchers build on this knowledge and focus their experiments on the subset of proteins important to their line of inquiry.

ETH Zurich

Ruedi Aebersold hopes many laboratories will adopt targeted proteomics.

Targeted proteomics using mass spectrometry promises to deliver data to help address specific biological questions in a way that makes it fundamentally unlike discovery proteomics.

Analysis of any given sample with a discovery-based approach runs into challenges. Two people might do the same experiment, or one person might do the same experiment twice, but "the results are not inherently the same," says Ruedi Aebersold, from the Institute of Molecular Systems Biology at the Swiss Federal Institute of Technology in Zurich. Neither person is necessarily wrong: the contradiction stems from their measurement of different subsets of the whole proteome, he says. "Because the space to sample is so huge, then the mass spectrometer pulls out, every time, a slightly different subset."



Katie Vicari

Targeted proteomics detects proteins of interest with high sensitivity, quantitative accuracy and reproducibility.

"What I like about targeted proteomics is that you answer the question that you are interested in," says Michael MacCoss of the University of Washington. Added significance comes from a "mental shift that our field is beginning to take," he says. Rather than trying to detect all the proteins in a mixture, as in a discovery-based approach, a targeted approach "lets us build quantitative assays to specifically answer hypothesis-driven questions," he says.

The 2012 Method of the Year "tips the hats to the biologists" in their quest to probe, detect, identify and quantify specific aspects of the complex and vast proteome under many conditions, says Brad Gibson from

the Buck Institute for Research on Aging in Novato, California. For instance: knowledge of proteins and their concentrations in a sample can reveal that a signaling pathway is active, giving biologists clues about how the changing components fulfill a physiological function. In clinical applications, to take another example, specific proteins in the blood of a lab animal can indicate how a tumor responds to a drug candidate.

Riding the workhorse

Although the application of mass spectrometry in targeted proteomics studies is relatively recent, it has largely depended on a mature workhorse: the triple quadrupole mass spectrometer. The instrument was originally developed in the late 1970s[1].

A native to analytical chemistry, the triple quadrupole has been widely used in pharmaceutical research, giving researchers ways to standardize how they identify and quantify drug metabolites in blood, says Rob Moritz, an Aebersold collaborator from the Institute for Systems Biology in Seattle. "It's a very highly reproducible standard technique that's used across the industry." The instrument's precision has been a necessity for drug companies, which have worked with mass spectrometer manufacturers to obtain the reproducible, precise measurement results needed to pass muster with the US Food and Drug Administration.

S. Carr /Broad Institute

"The giants who preceded us were the folks in the small-molecule field," says Steven Carr.

"The giants that preceded us were the folks in the small-molecule field," agrees Steven Carr, who directs the proteomics platform at the Broad Institute of MIT and Harvard. To get to less complex mixtures for analysis, the chemists removed proteins from their samples. Researchers in targeted proteomics need the proteins, but they, too, must also reduce sample complexity. They use mass spectrometry to do so.

As Carr explains, adoption of targeted mass spectrometry approaches for peptides dates back to the 1980s, such as in work by University of Tennessee researcher Dominic Desiderio, who sought a specific and selective way to assay peptides found in the brain[2].

Besides the availability of the triple quadrupole, the development of techniques such as electrospray ionization, used to produce gas-phase ions, and other general mass spectrometry advances also paved the way for the adoption of mass spectrometry for targeted peptide and protein analysis.

The triple quad allows researchers to use a processing technique called selected reaction monitoring (SRM), also known as multiple reaction monitoring (MRM), which focuses the mass spectrometer for the detection of a preselected group of analytes.

This SRM approach is similar in concept to immunoassays, in which antibodies latch onto and therefore identify proteins in a specific fashion. Some scientists point out that immunoassays allow greater sensitivity of detection than does mass spectrometry, and there are supporters for each approach (**Box 1**). But antibody-based methods are not easy to scale up. And immunoassays have their limitations because antibodies have varying availability and quality. "I think people would love to have it so that we don't have to rely on antibodies anymore," says MacCoss.

For finding a single protein in 10,000 patients, however, antibody-based tech-

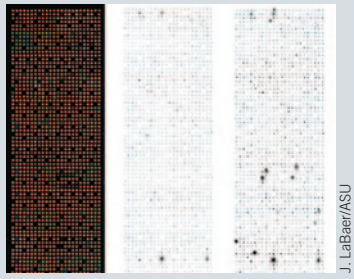
BOX 1 ANTIBODIES TARGET PROTEINS

Historically, researchers seeking quantitative information on proteins and peptides have relied on antibodies or other affinity reagents to provide the results they need. One challenge is that antibodies do not exist for all proteins, and those that do exist vary widely in their quality. A number of current initiatives are addressing this shortcoming, such as the Human Protein Atlas (http://www.proteinatlas.org/), an effort to generate at least two validated antibodies for each protein-coding gene in the human genome; Antibodypedia (http://www.antibodypedia.com/), a database of available antibodies developed in collaboration with Nature Publishing Group; and the US National Institutes of Health Protein Capture Reagents program (http://commonfund.nih.gov/proteincapture/), which focuses on developing renewable affinity reagents.

As a way to study many proteins in a targeted way, LaBaer uses protein arrays, which do not measure the actual protein in a sample but rather use the immune response to find the proteins. They can, for example, help to study antibodies against so-called "self-proteins," he says, which have been found to play a role in autoimmune disease, diabetes and cancer.

LaBaer sees value in hybrid approaches that use combinations of antibodies and mass spectrometry to target specific proteins. Peptides ionize with different efficiencies, he says, which could allow some of them to go missing in the mass spectrometry analysis. And even highly sensitive mass spectrometry–based methods such as SRM still have difficulty detecting low-abundance proteins as compared to antibodies. "If it's a rare thing you're looking for, you could miss it," he says.

One such hybrid approach, explains LaBaer, is stable isotope standard capture with anti-peptide antibodies (SISCAPA)[9], developed by Leigh Anderson, who directs the Plasma Proteome Institute in Washington, DC. This method, in which an antibody is used to pull targeted peptides out of a sample, is one that researchers Amanda Paulovich at the Fred Hutchinson Cancer Research Center and Steven Carr at the Broad Institute harness in their proteomics research.



J. LaBaer/ASU

Protein arrays can use the immune response to detect and find disease-specific signals. After confirmation of consistent protein display (left), identical arrays are probed with plasma from a healthy woman (center) and a breast cancer patient (right).

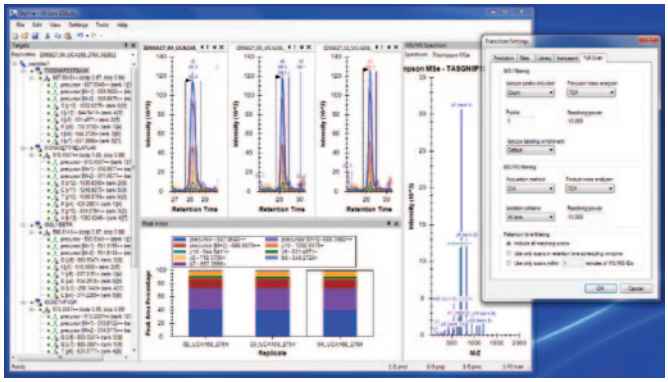
SISCAPA antibodies allow researchers to multiplex experiments and gain analysis depth, too. "In a one-step process, you can pull out things that are at the bottom of the nanogram-per-milliliter range in blood, which is a very challenging biofluid to deal with," says Carr.

Antibodies work with sensitivity and specificity, he says. With mass spectrometry, "we have to do analytical handstands in order to get down to really low detection limits."

niques "will not easily be beat," says Albert Heck, a mass spectrometrist at Utrecht University in The Netherlands. To find 10,000 proteins in a single experiment, discovery-based mass spectrometry is most appropriate, he says. The SRM technology fills a niche between antibody-based detection and discovery-based mass spectrometry and "may cover everything in between, typically about 100 proteins in 1,000 patients or conditions," he says.

Out of the quagmire

In the discovery-based proteomic approach, researchers must plow through databases, matching up the experimental spectra to *in silico*–derived spectra from protein sequence databases to infer which proteins are in a sample, explains Moritz.



M. MacCoss/Univ. Washington

Skyline is open-source software that works across mass spectrometry platforms to help users build targeted methods and analyze the resulting data.

"Quagmire proteomics" is how he refers to the discovery-based approach. It identifies highly abundant proteins in a straightforward manner, "but we could never get below those" to the low-abundance proteins, he says, which are often biologically important.

The complex data analysis steps in discovery-based proteomics have kept the ranks of those proficient in the art and the science quite thin. "It's a small number of labs that generate the bulk of the data," Aebersold says. "It should be the opposite." Recent advances and applications highlight that targeted proteomics is poised to affect the way a growing number of biologists detect and quantify proteins of interest. In contrast to discovery-based proteomics, targeted proteomics simplifies the downstream analysis, doing away with the complex bioinformatics required by discovery proteomics to analyze the mass spectrometry data. "I believe a much larger number of people can do it," he says.

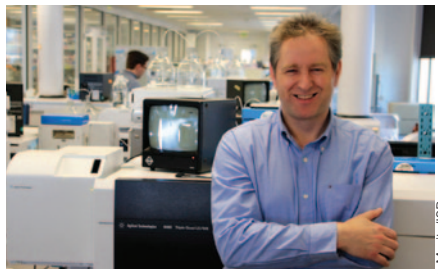
Aebersold has high hopes for targeted proteomics. "My dream would be that it is headed into many laboratories." Those labs might be currently using western blots to confirm hypotheses. "I would hope they would use targeted mass spectrometry," he says. Another application area is in systems biology, where scientists study networks of proteins under many conditions and need consistent data and high sensitivity for complex samples[3].

Whether a targeted or discovery-based workflow is used, the experimental analysis steps are similar: proteins are first preprocessed into smaller components—short amino acid sequences called peptides—that are ionized, separated and sorted according to mass and charge and then shot through a detector that quantifies the ions and delivers peptide spectra to the researcher.

With the SRM approach, scientists first determine which proteins they are interested in for a given sample. Then they program the triple quadrupole with so-called SRM assays, which enable the selection of certain signature peptides known as proteotypic peptides. These peptides not only uniquely represent the target proteins but also have favorable physicochemical properties, allowing the mass spectrometer to more readily detect them[4]. The mass spectrometer thus fragments and analyzes only those peptides that it has been programmed to process.

Being sensitive

Sensitivity is a major advantage of the SRM approach over discovery-based methods. "If we tell the mass spectrometer to only look at the fragments we know are there, then we can dive into the dynamic range and get down to these very low–copy number proteins," of which there may only be ten or fewer per cell, Moritz says. Proteotypic peptide selection and assay development are crucial to ready the technique to "be the most sensitive for the proteins you want to



R. Moritz/ISB

Targeted approaches are a way out of "quagmire proteomics," says Rob Moritz.

study, using the best identifier of the protein that you think of," says Heck.

Reproducibility is another advantage researchers cite for targeted proteomics over discovery-based approaches. SRM assays can perform "exactly the same from lab to lab, machine to machine, so there's no need to go back and do this database searching anymore," says Moritz. Every lab can use the same assays, which helps to make results comparable across labs, he says.

Carr and his colleagues demonstrated the reproducibility of SRM assays in eight laboratories across the United States with different instrument platforms in place, reporting "multisite robustness" and "very good" precision for blood-sample analysis[5]. Gibson, a coauthor on the paper, says he and others had expected it to work, but no study had shown this degree of reproducibility before. Carr says that the study "demonstrated to the community that this is not something that only a few specialized laboratories can do" and that SRM technology delivers "value because of the ability to target and repeatedly measure."

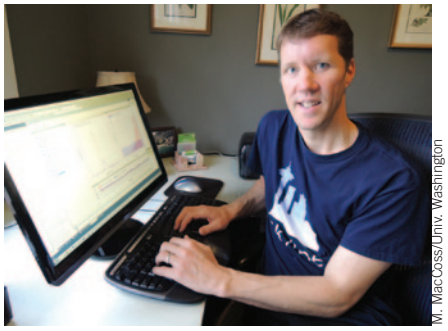
Another advantage of SRM is that, once SRM assays have been generated, it is "much, much faster" than a typical discovery-based experiment, and measurements can "be highly multiplexed so you can measure several hundred and now even several thousand peptides in a single analysis," says Aebersold. New developments over the past few years have enabled rapid proteotypic peptide selection and SRM assay development for new proteins or new splice forms[6].

The quantitative and targeted nature of this approach links proteomics more explicitly to hypothesis-driven research in biology, such as when researchers have hunches about a signaling pathway, says Aebersold. Certain proteins may be involved, and they might increase, decrease or be phosphorylated, all of which become testable hypotheses. Scientists can use SRM to obtain credible and reproducible results, "but you don't

discover a new protein," Aebersold says. If a scientist knows little or nothing about a pathway, then targeted proteomics "is not the method for you."

Meeting and greeting the technology

Over the years, "vendors have done a great job" at improving the sensitivity, dynamic range and specificity of these high-resolution and accurate mass spectrometry instruments for proteomics, says MacCoss. These technology strides, however, have yet to be met by the right analysis tools for targeted proteomics. As a result, he, software developer Brendan MacLean and their colleagues created Skyline[7], open-source software that works across mass spectrometry platforms and helps users build SRM methods. When they use the tool to visually assess their data, they are able to focus on biology rather than the method, MacCoss says.

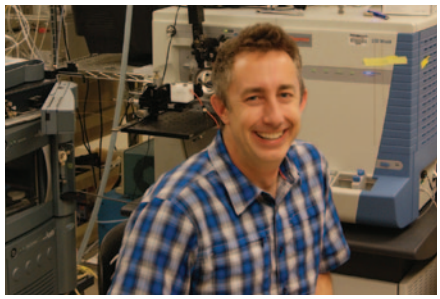


M. MacCoss/Univ. Washington

Brendan MacLean, software developer in the MacCoss lab, managed development of the widely used software tool Skyline.

Skyline, which has been installed 13,000 times, is software that "nearly everyone in the field is using as an indispensable tool for setting up and performing the assays," says Heck. It's becoming a kind of gold standard, says Gibson. As targeted proteomics techniques progress, new analysis bottlenecks emerge that call for new tools and methods.

Other resources moving targeted proteomics ahead include online portals and compendia, such as the SRM Atlas (http://www.srmatlas.org/). As Moritz explains, for the last few years he, Aebersold and other proteomics researchers have built and validated over 170,000 SRM assays for human, mouse and yeast proteins that will soon be available to the broader community via the website.



M. MacCoss/Univ. Washington

"What I like about targeted proteomics is that you answer the question that you are interested in," says Michael MacCoss.

A swath of options

Another new development over the last year has been the emergence of 'SWATH' mass spectrometry, developed by Aebersold in collaboration with mass spectrometer manufacturer AB SCIEX. "We have always known we were frustrated by dynamic range," Gibson says. Traditional discovery-based experiments use a data-dependent approach, in which the instrument's control software selects the top peaks in a mass spectrometry run for further fragmentation and identification.

The SWATH method relies on data-independent acquisition (DIA), a concept originally developed in John Yates's lab at The Scripps Research Institute[8]. Technology developments are helping propel it forward because newer instruments can perform many more scans per second and are more sensitive, Gibson says. With DIA, all peptides within a predetermined mass window are fragmented, which allows a "march up the mass range," giving a "more comprehensive sampling" of many more peptides.

"I think that data-independent acquisition is a great way to interrogate your data," says MacCoss, who believes DIA will become a significant method for proteomics in the near future. The SWATH approach moves targeted proteomics into a higher-throughput space, allowing scientists to analyze more proteins and peptides than the SRM approach. MacCoss, along with other scientists, is pursuing software tool development for DIA, evolving Skyline for these more comprehensive data acquisition strategies.

Separately, Gibson and his colleagues are exploring how to use SRM scheduling to include more peptides in a single experimental run. A mass spectrometer will cycle through the crowd of peptides in the sample every second for an hour, he says. Building on the knowledge that certain peptides ionize at certain times, he and his colleagues target a time window to capture the subset of peptides. Gibson says this approach offers the chance to "quadruple the number of peptide analytes that could be effectively targeted."

Carr is exploring an alternative approach using fractionation methods coupled with mass spectrometry. "The idea is to present a less complex sample to the mass spectrometer," he says. Normally, the most abundant proteins dominate mass spectrometry analysis, hiding the less abundant proteins. Fractionation allows researchers to sample deeper into the proteome to find proteins that might be overlooked or even undetectable initially.

As technology and methods develop and mature, Carr believes sensitivity, specificity, an improved ability for quantification and high levels of confidence about the achieved analytical results will push targeted proteomics ahead. "It's been a long, fun ride, and it's only going to continue to get better."

1. Yost, R.A. & Enke, C.G. *J. Am. Chem. Soc.* **100**, 2274–2275 (1978).
2. Desiderio, D.M. & Kai, M. *Biomed. Mass Spectrom.* **10**, 471–479 (1983).
3. Picotti, P., Bodenmiller, B., Mueller, L.N., Domon, B. & Aebersold, R. *Cell* **138**, 795–806 (2009).
4. Mallick, P. *et al. Nat. Biotechnol.* **25**, 125–131 (2007).
5. Addona, T.A. *et al. Nat. Biotechnol.* **27**, 633–641 (2009).
6. Stergachis, A.B., MacLean, B., Lee, K., Stamatoyannopoulos, J.A. & MacCoss, M.J. *Nat. Methods* **8**, 1041–1043 (2011).
7. MacLean, B., *et al. Bioinformatics* **26**, 966–998 (2010).
8. Venable, J.D., Dong, M.Q., Wohlschlegel, J., Dillin, A. & Yates, J.R. III. *Nat. Methods* **1**, 39–45 (2004).
9. Anderson, N.L. *et al. J. Proteome Res.* **3**, 235–244 (2004).

Vivien Marx is technology editor for *Nature* and *Nature Methods*. (v.marx@us.nature.com).