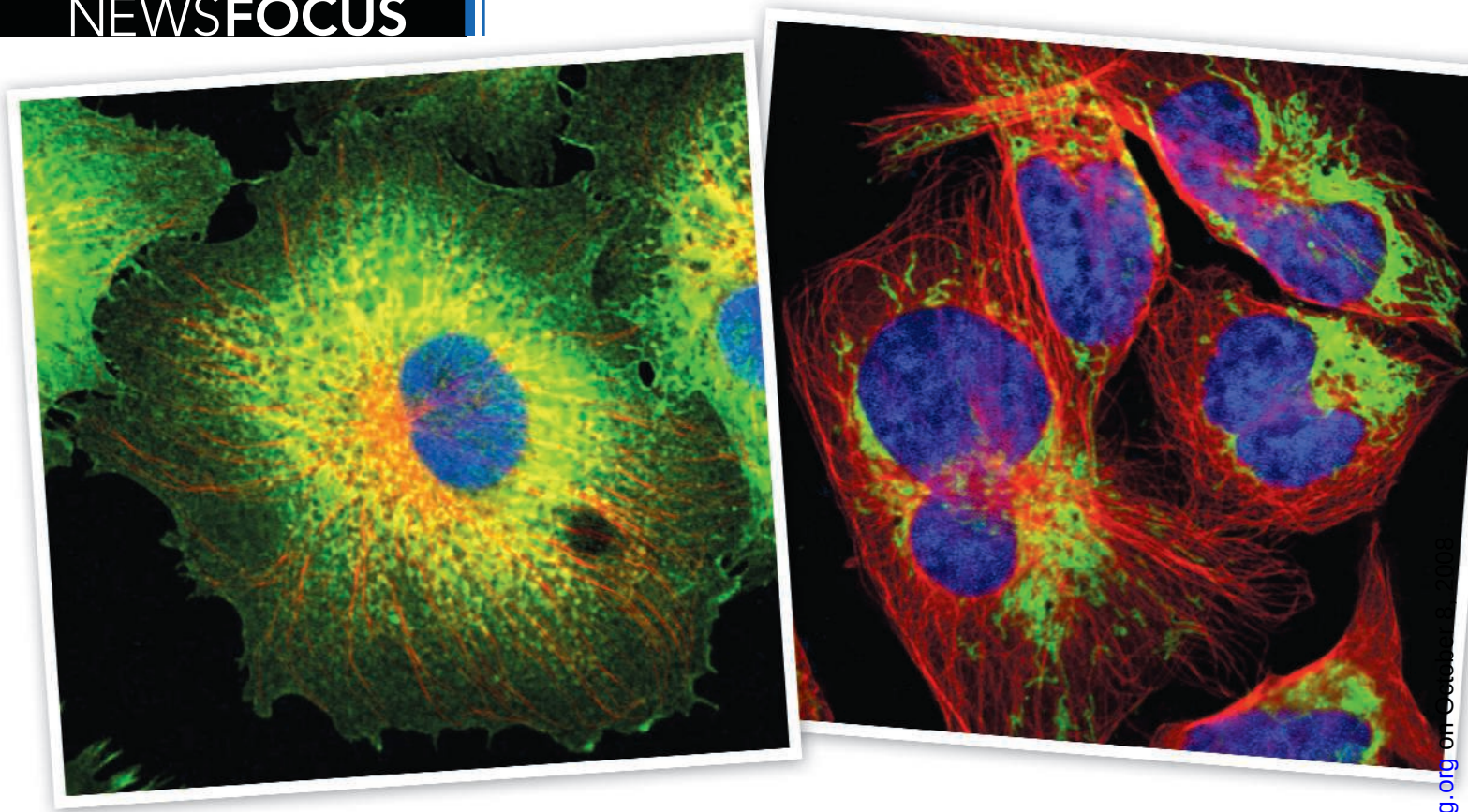


Proteomics Ponders Prime Time

Improved technologies for tracking thousands of proteins at once have spawned talk of a full-scale project to reveal all the proteins in each tissue—but the price tag would be daunting

AMSTERDAM, THE NETHERLANDS—He's too polite to come right out and say it, but Amos Bairoch thinks that much of the data generated by proteomics groups over the past decade is junk. Following the completion of the human genome project, proteomics labs set out to survey all the proteins expressed in different cells and tissues, in essence, putting meat on the bone of the genome. Mass spectrometers and other tools turned out gigabytes of data that purported to identify large numbers of proteins and fed them to Bairoch, who heads Swiss-Prot, a massive database that houses the latest findings on proteins of all stripes. Today, most of those data are ignored, Bairoch says, because the readings were too imprecise to make positive identifications. Throughout the years, many casual observers of the field dismissed proteomics as a waste of time and money. "People thought [the technology] was ready 10 years ago. But they didn't see good results and got disenchanted," Bairoch says.

Today, however, Bairoch's databases and others like them are filling up with terabytes of information that he calls "much better." The upshot: Proteomics is finally coming of age. With the help of better instrumentation and refined techniques, the top proteomics labs can identify and quantify more than 6000 distinct proteins from individual cells and tissues at a time. Now that these labs can cast such a wide net, many proteomics researchers say the time is ripe to undertake a full-scale human proteome project (HPP) to survey the landscape of proteins present in dozens of different human tissues. If successful, such a project would reveal which proteins are actually expressed in different types of cells and tissues, and at what levels, and the network of proteins they communicate with. That knowledge could offer researchers innumerable insights into how organisms convert their genetic blueprint into life and perhaps lead to breakthroughs in biology and medicine. "We are at the point where we can talk about doing this in 8 to 10 years," says Mathias Uhlen, a microbiologist and proteomics expert at the Royal Institute of Technology in Stockholm, Sweden.

It's not just talk. Uhlen and other proteomics leaders gathered here last month to weigh plans for an HPP and to sound out representatives of science funding agencies that would need to pony up the hundreds of millions—if not billions—of dollars needed to pull it off. Most of the responses suggested that tight science budgets make a new mega-sized international science project unlikely anytime soon. Nevertheless, even without a coordinated international HPP, the field is moving so fast that "it's happening already," says Matthias Mann, a proteomics expert at the Max Planck Institute of Biochemistry in Martinsried, Germany.

Spotted history

Many researchers probably assume an international proteome effort started years ago. The availability of the human genome sequence in 2001 told researchers how many proteins are likely to be out there and the exact sequence of amino acids they should look for. The race was on, amid plenty of hype. "Everyone was interested in proteomes," says Mann.

But there were problems, lots of them. For starters, proteins are chemically far more heterogeneous and complex than DNA and RNA. It was relatively easy for researchers to

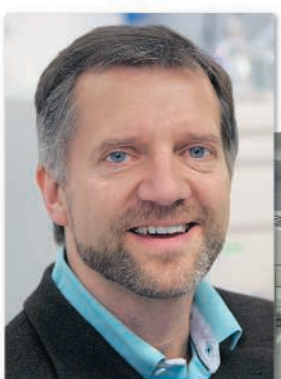
CREDITS: UHLEN AND LUNDBERG/THE HUMAN PROTEIN ATLAS

Published by AAAS

Revealing. Fluorescent antibodies flag the locations of different proteins in cells, offering clues to those whose functions are unknown.

create a single, robust, and standardized sequencing technology to decode the genetic blueprint of humanity. But no single machine could tell researchers everything they wanted to know about proteins. Worse, although each cell contains the same complement of genes, the abundance of different proteins varies widely. One milliliter of blood, for example, contains about 1 picogram of cell-signaling molecules called interleukins and about 10 billion times that amount of a protein called serum albumin. Such plentiful proteins can mask the signals of their rare brethren.

Still, the lure of proteins was undeniable. Whereas genes are life's blueprint, proteins are the bricks and mortar from which it is built. Identify a critical protein in a disease process, and it could serve as a target for a multibillion-dollar drug to fight diabetes or heart disease. Fluctuations in the amounts of some proteins could serve as "biomarkers" to alert doctors to the onset of cancer or Alzheimer's disease. In the early part of this decade, companies flocked to the field, raising and spending hundreds of millions of dollars. But it quickly became clear that the technology was immature. After several years of trudging down blind alleys, most of the companies that were formed to hunt for biomarkers and drug targets were either folded or merged out of existence (see sidebar, p. 1760).



The news wasn't much better in academia. Take an early example from the Human Proteome Organisation (HUPO), which was launched in 2001 to coordinate international proteomics research and bring order to the unruly field. In 2004, HUPO launched its Plasma Proteome Project (PPP) to survey blood proteins and propel the search for candidate biomarkers. HUPO sent identical blood samples to research groups around the globe, each of which conducted its own analysis with its own homegrown version of the technology. "It was a big disaster," says John Yates, a chemist and mass spectrometry (MS) expert at the Scripps Research Institute in San Diego, California. "There was no quality control. Then the data came back, and it was just a mess," he says.

Unfortunately, PPP and other early efforts raised expectations that they would produce a shortcut for finding novel biomarkers for a wide variety of diseases. "The plasma proteome [project] made the search for biomarkers look like a slam dunk," says Jan Schnitzer, who directs the vascular biology and angiogenesis program at the Sidney Kimmel Cancer Center in San Diego. "But it hasn't delivered." That failure and the failure of proteomics as a whole to deliver on its promise, Uhlen adds, "is a history which is still haunting us."

> ***"The biology community at large has to show they really need this. If they can't, why should they fund this?"***
>
> **—AMOS BAIROCH, SWISS-PROT**

HUPO has since promoted uniform standards for everything from how to collect and process blood and tissue samples to the proper methodologies for screening them and analyzing the data. And PPP is now taking a more targeted approach to discovering proteins.

The standards have helped, but they haven't solved all the problems. A study last year compared the ability of 87 different labs to use MS to identify correctly 12 different proteins spiked into an *Escherichia coli* sample. No lab got them all, and only one correctly identified 10 of the 12, says Thomas Nilsson, a proteomics researcher who splits his time between Göteborg University in Sweden and McGill University in Montreal, Canada. In a follow-on study completed this year, only six of 24 labs correctly identified 20 spiked proteins. "That again is quite depressing," Nilsson says. "So what are the chances [for success] of high-throughput proteomics as a distributed effort?" Nilsson asked attendees in Amsterdam.

Perhaps surprisingly, Nilsson says he thinks they are decent. This year's study, he explains, shows that most errors in MS-based analyses arise not because the technology can't spot the proteins researchers are looking for but because software programs often misidentify them.

A big part of the problem, says John Bergeron, a proteomics expert at McGill University, rests with simple statistics. To identify proteins using MS, researchers first chop a sample of proteins into smaller fragments called peptides. Those peptides are fed into a mass spectrometer, which ionizes them and shoots them through a chamber. The time it takes for the ions to "fly" through the chamber reveals the atomic weight of the peptides, which in turn reveals their identities. Computer programs then compare them with a full list of the organism's genes, which code for those peptides and their proteins. If a peptide matches the protein code in only one gene, it is a hit and it is a unique identifier of the protein.

The problem is that not all peptides are successfully ionized in each experiment, so some don't enter the chamber. Even if the same lab runs a sample of proteins through the machine twice, Bergeron says, 33% of the proteins identified will appear to be different between the two runs. To minimize such sampling error, MS labs now typically

Pacesetter. Thanks to better mass spectrometers and software, researchers such as Matthias Mann (*inset*) can now identify thousands of proteins in a single experiment.

CREDITS: MAX PLANCK INSTITUTE FOR BIOCHEMISTRY

Published by AAAS

run samples through their machines as many as 10 times. Today, MS groups also look for more than one unique peptide to confirm the identity of a protein. Those changes, together with other emerging standards, show that "these are problems that can be addressed," Schnitzer says.

A new approach

Such successes are also convincing proteomics leaders that the technology is mature enough to go after a full-scale HPP. Although details remain in flux, the generally agreed-upon plan is to identify one protein for each of the estimated 20,400 human genes. Bairoch reported at the meeting last month that Swiss-Prot has logged what is currently known about each gene, such as the primary proteins a particular gene produces and their function. Proteins for about half of the genes have never been seen, Bairoch stated.

There are far more proteins than genes, because proteins can be spliced together from multiple genes, and once synthesized, they can later be cut down in size or modified with other chemical groups. Trying to find all those variants in all tissues is a task that will likely take decades, Uhlen says. Sticking to one protein for each gene provides a defined endpoint to the project and would create a "backbone" of all human proteins that can be continually fleshed out.

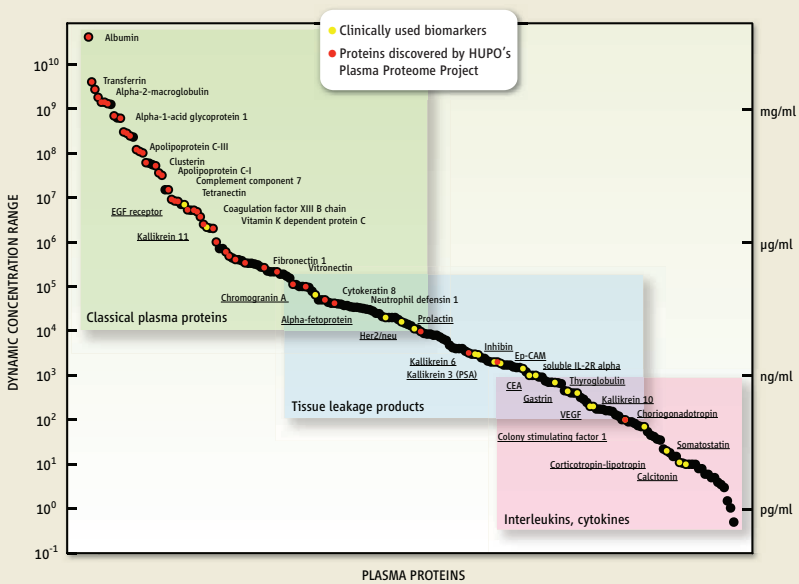
Another possible goal is to create one antibody for every protein in HPP. Because antibodies typically bind to one target and nothing else, researchers can use them to fish out proteins of interest and track their locations in cells and tissues. That would offer clues to the functions of the thousands of proteins for which little is known. Uhlen and colleagues in Sweden launched just such a global antibody project in 2005. And in Amsterdam, Uhlen reported that the catalog now contains more than 6000 antibodies against distinct human proteins, more than one-quarter of the complete set. At the current rate of new antibody production, Uhlen says his team will finish the task in 2014. More money, he says, would undoubtedly speed the effort.

A third project would track which proteins "talk" to one another. To find a protein's partners, researchers create thousands of identical cell lines and insert into each one a chemical tag linked to a different protein. They can use the tag to pull that protein out of the cell at a specific point in its life cycle, along with any other proteins, bits of RNA or DNA, or a metabolite that it is

Will Biomarkers Take Off at Last?

One much-heralded application of proteomics—detecting proteins that are markers for specific diseases—has long been a dream deferred. "It has been extremely difficult to find those proteins that are biomarkers," says Ruedi Aebersold, a proteomics expert at the Swiss Federal Institute of Technology in Zurich, Switzerland, and the Institute for Systems Biology in Seattle, Washington. But after years of disappointments, proteomics researchers say they're cautiously optimistic.

When proteomics caught fire earlier this decade, scientists hoped that mass spectrometry (MS) and other technologies would help them sift through the thousands of proteins in blood and other body fluids to identify a rare protein that indicated the presence of a disease. Researchers could then use these biomarkers to spot diseases in their formative, treatable stages. But the demise of several companies, such as GeneProt and Large Scale Biology, that jumped into the field revealed that nailing down biomarkers is harder than it sounds.



Needles in a haystack. The abundance of different proteins in blood varies by more than 10 orders of magnitude. Most commercially used biomarkers (yellow dots) are present in only minute quantities in blood, below the level at which most proteins are detected (red dots).

One problem is that blood—the most common hunting ground—is difficult to work with. Levels of different proteins in blood vary by 10 orders of magnitude, and the abundant proteins often mask the presence of rare ones. Unfortunately, mass spectrometry, the best tool for casting a wide net to search for proteins, hasn't been sensitive enough to spot the rare ones. "Most clinically used biomarkers are at nanogram [per milliliter] levels or below," Aebersold says. At the meeting, Aebersold reported a new strategy for targeting protein fragments called N-linked glycopeptides, which commonly make up cell-surface receptors and thus are more likely to be shed into the blood. This targeting allowed Aebersold's team to spot proteins down to nanogram-per-milliliter levels and thereby track them to look for possible links to diseases. Aebersold says he's hopeful that similar, more focused, studies will improve prospects for the biomarker hunters.

Better instrumentation won't solve all the problems. Techniques such as MS that survey thousands of different compounds inevitably turn up false positives: proteins that change their abundance in lockstep with a disease just by chance. That means candidate biomarkers must be validated through clinical trials, which can cost tens of millions of dollars—and most of them fail. "To be accepted by [regulatory] agencies, it's almost as costly as developing a new drug," says Denis Hochstrasser, the director of laboratory medicine at Geneva University Hospital in Switzerland. Because diagnostics companies, unlike drugmakers, typically can't charge lofty premiums for their new tests, they have less incentive to develop biomarker tests. Michael Snyder, a proteomics expert and cell biologist at Yale University, says that despite these challenges, he's hopeful that improving proteomics technologies will generate novel biomarkers—"just not on the same time frame as people thought."

–R.F.S.

CREDIT: RALPH SCHIESS/INSTITUTE OF MOLECULAR SYSTEMS BIOLOGY, ETH ZURICH

Published by AAAS

bound to. Bioinformatics experts can then weave together the partners for each protein to construct a complete communication network of the proteins in the cell.

Such protein-interaction networks have been worked out in exquisite detail in yeast and other organisms. But it has been hard to insert the chemical tags reliably into human cell lines. Over the past decade, however, researchers around the globe have shown that different lentiviruses readily insert tagged proteins into a wide variety of human cells. At the meeting, Jack Greenblatt of the University of Toronto in Canada said he has proposed a project to insert one tagged protein for each of the 20,400 genes, the first step to a complete human proteome interaction map. The project is now under review by Genome Canada, the country's national genome sciences funding agency. Greenblatt adds that working with human cell lines isn't perfect, because these lines are typically made up of non-normal cells that have been immortalized. His group is also performing related studies in mice, which can be grown into adult animals, and the interaction networks can be compared with those found in the human cell lines. Other projects could be added to HPP as funding permits. They could include a catalog of all the modified proteins, such as splice variants and phosphorylated proteins, Bergeron says.

Finding the money

How much will it take to complete the wish list? Opinions vary, but somewhere in the neighborhood of $1 billion is a common guess. Michael Snyder, a yeast biologist at Yale University, thinks that's too little. "This is going to require a bigger budget than that," he says.

Whatever the projection, it was enough to make those with the money blanch. Funding agencies around the world are already collectively spending hundreds of millions of dollars on proteomics technologies and centers. They're also already committed to several international big biology projects such as the International HapMap, the International Cancer Genome Consortium, and the Knockout Mouse Consortium, which are putting the squeeze on tight budgets. "From a funding viewpoint from the U.S. context, now is not the right time," says Sudhir Srivastava, who directs proteomics initiatives at the U.S. National Cancer Institute in Rockville, Maryland. "If this was 5 years ago when the NIH [National Institutes of Health] budget was doubling. …" Srivastava trails off.

Still, Uhlen and others say they are hopeful that funding agencies will keep the field moving quickly. "We don't have to have $1 billion from the start," Uhlen says. "With the Human Genome Project, it took 5 years for the funding agencies to put serious money into it. I don't think we should expect funding agencies to jump on board until we have proven the technology."

To do that and make the cost more palatable, HUPO leaders are mulling a pilot project to catalog all the proteins produced by chromosome 21, the smallest human chromosome, which has 195 genes. Although the cost of such a project isn't known, "I think there almost certainly would be interest," says Roderick McInnes, director of the Institute of Genetics at the Canadian Institutes of Health Research in Ottawa.

"We are at the point where we can talk about doing this in 8 to 10 years."

—MATTIAS UHLEN, ROYAL INSTITUTE OF TECHNOLOGY

At the meeting, proteomics expert Young-Ki Paik of Yonsei University in Seoul, South Korea, said the Korean government is considering funding a similar proposal for a Korean-based pilot project on chromosome 13, the second-smallest human chromosome, with 319 genes. Paik says he and his colleagues have proposed a 10-year, $500 million initiative that is currently being considered by the Korean Parliament. A decision is expected in October. If it is funded, Bergeron says it will be a major boost to the field and could help catapult Korea into the forefront of proteomics.

Some researchers are skeptical of going chromosome by chromosome, however. "In gene sequencing, that approach worked," Bairoch says. "You could separate out the work by chromosome. But it doesn't make sense for proteins. There is no [body] fluid or [tissue] sample organized by chromosome." Ruedi Aebersold, an MS expert with a joint appointment at ETH Zurich and the Institute for Systems Biology in Seattle, Washington, agrees. "I'm not a big fan of going chromosome by chromosome," he says. MS machines, he notes, identify whatever proteins show up regardless of the chromosomes they came from.

Whatever path they take to an HPP, proteomics leaders will need to find true believers beyond those already in the flock. "The biology community at large has to show they really need this," Bairoch says. "If they can't, why should they fund this?" Uhlen, Bergeron, and other HUPO leaders agree. And they argue that current demonstrations of the technology are starting to build the case.

At the Amsterdam meeting, for example, Mann reported that recent advances in instrumentation and software have enabled his group to identify the complete yeast proteome in one shot—in just a few days. That feat took months of painstaking effort when it was first accomplished by traditional methods 5 years ago. Mann also described the use of a technique his team first reported last year to monitor changes in the yeast proteome, including levels of individual proteins, between two different states. In one example, Mann's team compared yeast cells with a diploid (double) set of chromosomes to cells with the haploid (single) set undergoing sexual reproduction. The study quantified for the first time the suite of proteins that orchestrate sexual reproduction in yeast. Mann says the technique opens the door to studying proteomewide differences between healthy and diseased cells, developing and mature cells, and stem cells and differentiated cells. "There is no end to what you can compare," Mann says. "Every lab can ask these questions."

In an another study, Uhlen reported using his antibodies to track global protein expression in human cells. He and his colleagues have shown that fewer than 1% of all proteins are expressed in only one tissue. That implies, he says, that tissues are differentiated "by precise regulation of protein levels in space and time, not by turning expression on and off." Aebersold also reported that his lab has devised a scheme for detecting proteins expressed at the level of just a single copy per cell.

"These are unbelievable advances, and they show we can take on the full human proteome project immediately," Bergeron says. Not everyone has turned that corner, but Bergeron and others say that they are confident that time is coming soon. As Pierre LeGrain, director of life sciences at the French Commissariat à l'Energie Atomique in Gif-sur-Yvette, sums it up: "Most of us feel the human proteome project is going to happen, though we don't know how."

–ROBERT F. SERVICE

CREDIT: JENS LASTHEIN