Proteomics aims to chart the ebb and flow of tens of thousands of proteins at once to produce snapshots of life inside cells. For now, the technology isn't there. But this young field is growing up fast

High-Speed Biologists Search For Gold in Proteins

PROTEOMICS

This special news focus looks at the promise and roadblocks of biology's latest wellspring.

The scene from the picture windows in GeneProt's third-floor conference room looks downright leisurely: vineyards atop rolling hills with the Swiss Alps beyond. But inside this Geneva-based biotech upstart, it's all about speed.

In a labyrinth of rooms in GeneProt's first- and second-floor laboratories, four different kinds of bench-top robots—24 machines in all—steadily work together in silence. The robots, some weighing as much as 150 kilograms with arms that whir in all directions, carefully isolate a mix of proteins from a tissue sample, separate them into clumps of identical proteins, chop members of each clump into fragments, and place them into an array of wells on tiny metal plates. Technicians feed these plates into a series of 51 mass spectrometers worth over $150,000 each; every second, each of these refrigerator-sized machines spits out a fingerprint of a protein fragment based on its mass. A supercomputer then compares each fingerprint to a database to identify the amino acids it contains. Then, within minutes, it reassembles the jumble of fragments to identify the proteins from which they came. The result: a list of thousands of proteins present in the starting sample. A few years ago, identifying just one of these proteins often took years. Today it takes hours.

GeneProt execs are betting that by comparing such lists from diseased and normal tissues, they will be able to identify which proteins are the most important in various diseases—and therefore make the best targets for new medicines. Although GeneProt started building this futuristic lab only last year, company officials say they've already fingered six proteins that could serve as drugs themselves or as targets for other compounds.

GeneProt's lab is just one of many converging on biology's biggest boom industry: proteomics. The goal of this new "-omic" is no less than to catalog the identity and function of all the proteins in living organisms. In terms of complexity, proteomics makes genomics look like child's play. Instead of an estimated 30,000 to 40,000 genes, protein experts think that humans have somewhere between 200,000 and 2 million proteins. What's more, whereas genes remain essentially unchanged through life, proteins are constantly changing, depending on the tissues they're in, a person's age, and even what someone ate for breakfast.

Researchers are hell-bent on tracking down proteins, says GeneProt co-founder and chief scientist Keith Rose, because proteins, not genes, are where the action is. Whereas genomics offers a look at the blueprints for life, proteomics reveals the nuts and bolts. Defective proteins are responsible for the chemistry that leads to a range of diseases from cancer to Alzheimer's. And blocking or boosting these proteins offers the straightest shot to finding the next blockbuster drug, says Sam Hanash, a proteomics expert at the University of Michigan, Ann Arbor: "The proteome can bring a lot of the fruit that the genome could not."

Eye on the prize. Fast protein analysis offers hope for finding blockbuster drugs.

Both money and hype are flowing fast and furious. Dozens of new companies have sprung up in the past few years to either search for proteins en masse or sell tools to the protein prospectors. Most pharmaceutical giants, such as GlaxoSmithKline and Pfizer, have launched their own proteomics efforts as well; all are racing to find and patent proteins. In a time of tight markets and wary investors, proteomics companies have attracted more than $530 million in venture capital funds in the past 22 months. Stock offerings have raised hundreds of millions more.

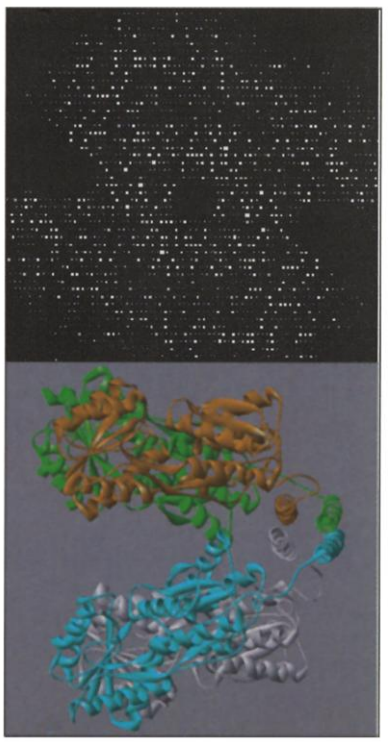
"You're talking about an absolute explosion of interest in proteomics in industry," says Raj Parekh, who directs proteomics research at Oxford GlycoSciences (OGS) in the United Kingdom. "Proteomics, a word almost no one discussed two years ago, has become the new darling of the investment community," life sciences market watcher G. Steven Burrill, CEO of Burrill & Co. in San Francisco, wrote recently.

But despite the deep pockets, "there is still a bit of snake oil in this field," says Phil Andrews, a proteomics expert at the University of Michigan, Ann Arbor. In private, few researchers deny that identifying all the body's proteins might be a lot harder to achieve than the industry's public relations suggests. The technology to pull it off simply doesn't exist yet, and competition is stiff for those proteins that can be nabbed with current technology. And if proteomics does turn up new targets, what's to say they will be any easier to develop into drugs than the targets already out there? But for now, the allure is so compelling that few want to dwell on the gritty underside.

Grab bag

Just what is a proteome? Ask a dozen experts, and you will get a dozen different answers. Most commonly, it means an organism's complete set of proteins in every form they assume. But with proteins winking in and out of existence, what you see depends

ILLUSTRATION: C. SLAYDEN



X-ray Crystallography

This tried-and-true technique maps the atomic structure of proteins. Researchers overexpress a protein, purify it, and coax the individual proteins to line up in crystals. They then use a beam of x-rays to produce a diffraction pattern (*top*), which helps determine the final structure.

CREDITS: (X-RAY DIFFRACTION PATTERN) M. LEWIS/UNIVERSITY OF PENNSYLVANIA; ILLUSTRATIONS: C. SLAYDEN

on when you look and what tools you use. "In genomics, the end point is well defined: the full sequence of an organism's DNA. With proteomics that's completely different. It's an attempt to capture the dynamics of a living system," says Ruedi Aebersold, a proteomics expert and co-founder of the Institute for Systems Biology in Seattle, Washington. That makes it all but impossible to define a single proteome that can be tackled in a large-scale, systematic way akin to the genome project—although there is a nascent attempt at international collaboration to do something along those lines.

Nor is there likely to be a single technology that will dominate the field—as robotic gene sequencers did for genomics—nor a single corporate juggernaut like Celera Genomics of Rockville, Maryland. That's because unlike genes, proteins vary widely in their chemical behaviors, making it difficult to come up with one technique that works equally well on all proteins.

The result is a balkanized landscape in which different groups—mostly companies at this point—are chipping away at different pieces of the puzzle, all with the hope of finding the next blockbuster drug. Some want to know what proteins are expressed in diseased versus normal tissue. Others have set their sights on how proteins interact and what they do. Still others are determining the three-dimensional structure and function of proteins. The field is so vast, the goal so expansive, that there is room for everyone, says Denis Hochstrasser, a proteomics pioneer at the University of Geneva in Switzerland and another co-founder of GeneProt.

The new frontier

No one could even consider studying proteins en masse—instead of one by one—until the mid-1970s and 1980s, when technologies for separating mixtures of proteins and tracking which proteins bind to one another first made their debut. But what truly got the field going, says GeneProt's Rose, was the mass of genomics data churned out by the Human Genome Project and huge advances in computing power in the late 1990s. Suddenly, researchers could identify almost any protein they fished out of a tissue sample. All they needed to do was translate a fragment of the protein's amino acid sequence into DNA's code of A's, T's, G's, and C's; this information could then be used to search a computer database for the gene that made it along with the identity of the complete protein. Backed with the right robotics and supercomputers, researchers can now analyze hundreds of thousands of proteins in a tissue sample in a few months.

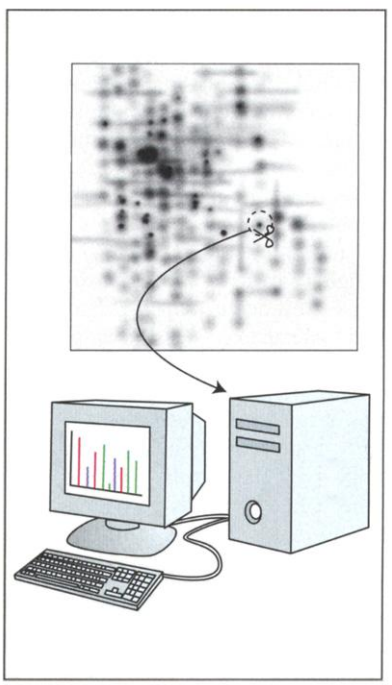
These newfound capabilities promise a major expansion in the number of potential drug targets. "Today the whole industry operates on 500 protein [drug] targets. There are thought to be between 10,000 and 20,000 protein targets. So the whole race [in proteomics]—and it is a race as all the technologies pile on—is how are we going to find those," says OGS chief Michael Kranda. Even so, Kranda readily acknowledges that the real roadblock in the pharmaceutical industry is not a lack of novel targets but the difficulty and expense of turning them into marketable drugs. Matthias Mann, chief scientist of MDS Proteomics in Toronto, Canada, agrees, adding that companies such as GeneProt that are relying on heavy firepower up front might find themselves disappointed in the end. "I don't think the race is going to be won by the number of machines a company is using," Mann says.

All this has created a land-grab mentality, similar to that among genomics companies in the 1990s as they raced to patent genes. "The driving force is to crank through as many proteins as possible to patent them" and claim as much intellectual-property real estate as possible, says Ian Taylor, who heads proteomics efforts at PerkinElmer Life Sciences in Cambridge, U.K. OGS, for example, expects to file some 4000 patents by year's end, each of which will represent a protein whose function is known and linked to disease, says Kranda. At least some of the protein patents seem likely to duplicate existing gene patents, setting the stage for court battles as genomics and proteomics companies try to protect their turf (see p. 2082).

Searching for answers

Researchers are pursuing the race using three fundamental approaches. Two are basically goosed-up versions of long-known technologies for separating mixtures of proteins and watching their interactions. The third is x-ray crystallography, an even older technique for mapping the structure of proteins in atomic detail.

For now, the dominant technology for separating proteins—and the workhorse for companies such as GeneProt, OGS, and



2D Gel Electrophoresis

After a mixture of proteins is placed in a gel, the proteins are separated in one direction by their charges and in the perpendicular direction by their molecular weights. Proteins of interest are then cut from the gels, purified, and broken into fragments. These fragments are sent to a mass spectrometer, which measures their atomic masses. Masses from the fragments are then used to identify the protein.

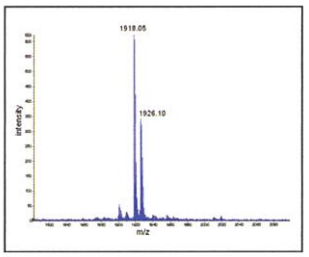
Large Scale Biology Corp. of Vacaville, California—is two-dimensional (2D) gel electrophoresis (see sidebar, p. 2075). The technique uses electric fields to pull proteins through a slab of gelatin to separate them from one another by their molecular weight and charge. Proteins of interest are then cut from the gels, chopped into fragments, and handed off to a mass spectrometer and computer for identification. Companies use 2D gels to try to detect differences in protein expression between tissues—comparing, say, cancerous and healthy liver tissue.

Among the recent deals in this red-hot sector, last November GeneProt and Novartis Pharmaceuticals announced the largest proteomics agreement to date. For an up-front $43 million and the promise of another $41 million over 4 years, GeneProt agreed to run three studies for Novartis to identify drug targets. According to company literature, Large Scale Biology, meanwhile, is using its version of the technology to assemble a database of proteins called the Human Protein Index, "the protein equivalent of the Human Genome Project."

Other companies are using a technology called the yeast–two-hybrid method to map protein-protein interactions (see sidebar, p. 2077). This approach uses known "bait" proteins to bind unknown "prey" and thereby reveal which proteins interact; this information provides insight into the function of the unknown captives. By repeating such experiments en masse, investigators can work out the tangled protein interaction networks in cells. And because these networks reveal the chain of communication cells use to survive and thrive, asserts Sudhir Sahasrabudhe, chief of research at Myriad Genetics of Salt Lake City, Utah, the technique offers the fastest way to home in on potential drug targets.

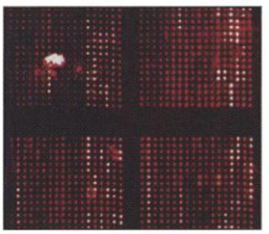
In April, Myriad trumpeted a $185 million collaboration with Hitachi and Oracle as an effort to "map the human proteome" in 3 years. "This project represents a bold leap toward the future of drug development," Myriad president and CEO Peter Meldrum declared at the time.

Proteomics 2.0: The View Ahead



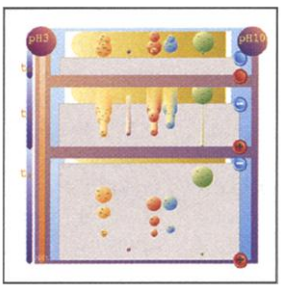
Isotope-Coded Affinity Tags Pioneered by Ruedi Aebersold at the Institute for Systems Biology in Seattle, Washington, this technique enables researchers to chemically tag specific proteins in two separate samples with distinct heavy and light isotopes. By then tracking their relative abundance with a mass spectrometer, researchers get a quantitative measure of how protein expression changes with disease.

Protein Chips On these chips researchers lay down a checkerboard-like grid of molecules designed to capture specific proteins at specific sites. They then use fluorescent probes or other means to detect where proteins bind on the grid. And because they know the identity of the probes at each spot on the grid, this reveals which protein they have captured. Although protein chips have been slow to develop, researchers expect that in time they could become a fast way to scan samples for hundreds or thousands of different proteins (see p. 2080).



Microfluidics Researchers at several universities and companies have created silicon, glass, and plastic chips, engineered to harbor networks of sample holders, channels, and reaction chambers to carry out the complex sequence of steps needed to prepare protein samples for analysis by mass spectrometers and other analytical equipment. Because the chips are fast and can work with very small samples, they have the potential to dramatically improve the speed and sensitivity of proteomic analyses.

Differential In Gel Electrophoresis This technique, commercialized by Amersham Pharmacia Biotech, offers researchers a global view of how protein expression changes between two samples. Proteins from one sample are all tagged with a single fluorescent compound, while those from another sample are tagged with a different colored fluorescent dye. The two samples are then mixed together and the individual proteins are separated on a single two-dimensional gel; this separates proteins in one direction by their charges and in the perpendicular direction by their molecular weights. A quick look at the gel reveals whether separate spots show both colors—meaning that a protein is expressed in both samples—or just a single color that shows which sample harbors the protein. **–R.F.S.**



A third set of research groups, meanwhile, is rushing to automate high-speed x-ray crystallography and related approaches to map the atomic landscapes of proteins. The techniques, collectively known as structural genomics, aren't as widespread as other approaches to proteomics. But a more moderately paced version of the technique is widely used as a key step in designing drugs to interact at specific sites in proteins. Over the last 40 years, those efforts have generated some 2000 unique protein structures in public databases. Tim Harris, who heads Structural GenomiX in San Diego, California, claims his company alone will more than double that number in just 5 years.

Here, too, the dealmaking has been brisk. According to Raymond Stevens of the Scripps Research Institute in La Jolla, California, structural genomics efforts have raised more than $500 million since 1999. About half of that, he says, comes from publicly financed programs in the United States, Japan, and Europe. The rest is private money raised to back start-up companies such as Syrrx and Structural GenomiX in the United States and Astex Technology in the U.K.

Speed bumps

Upon closer inspection, the lofty goals and broad statements that have accompanied announcements of these deals come with important caveats. Myriad, for example, says it won't actually find all the protein-protein interactions in cells. Rather, the company will track down all it can with the yeast–two-hybrid and mass spec approaches. Likewise, Large Scale Biology's human proteome database will survey only tissues the company deems relevant to finding new drugs and diagnostic markers. One reason that the reality is likely to be less impressive than the hype is that all three frontline proteomics technologies suffer from serious limitations.

The 2D gels, in particular,

CREDITS: (TOP TO BOTTOM) INSTITUTE FOR SYSTEMS BIOLOGY; H. ZHU *ET AL.*; GYROS; AMERSHAM BIOSCIENCES

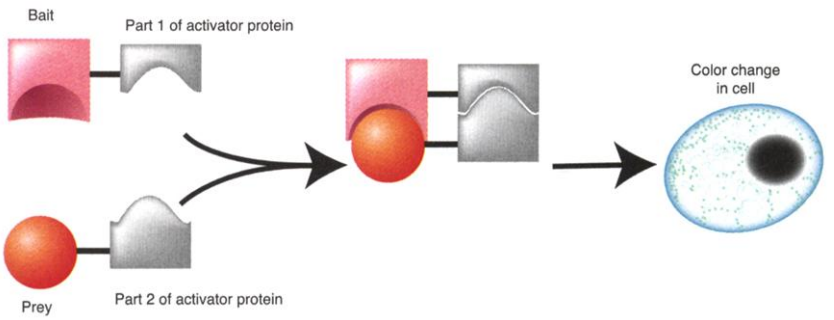
face big drawbacks. The gels typically do a poor job of separating proteins that are either small or very large. They also stumble when it comes to separating out proteins normally embedded in cell membranes, often the best drug targets. And despite improvements, mass spectrometers still have a hard time seeing proteins that are expressed only in minute quantities, many of which could be key protein markers for cancer or other diseases. The 2D gel/mass spec approach "is not based on tomorrow's technology," says MDS's Mann.

The yeast–two-hybrid method is also limited in what it can accomplish. For one, the technique detects proteins that interact when placed inside yeast cells, outside their natural environment. And although many of the interactions might prove interesting from a basic science perspective, few are likely to be related to disease. Moreover, "interactions and abundance [of particular proteins] change over time," says the Institute for Systems Biology's Aebersold. So although mapping the links between proteins is "attractive," Aebersold says, "it is by no means sufficient to explain the biology of a cell."

For structural studies of proteins, the biggest drawback is speed. Even with cutting-edge robotics, individual companies or academic collaborations can hope to resolve only several hundred protein crystal structures a year. And although this might prove useful to drug designers, it's impossible to imagine how researchers might survey an entire proteome with the technology.

Finally, all three techniques falter when it comes to the chemical tags that proteins receive after they emerge from the ribosome where they are built. These "post-translational modifications," such as the addition of phosphate or sulfate groups, can have profound effects on a protein's function, says OGS's Parekh. But deciphering these chemical tags can be painstaking work. "We don't have good tools for looking at these in a high-throughput way," says David Hachey, a proteomics specialist at Vanderbilt University in Nashville, Tennessee.

GeneProt's Rose and others claim that there are technological fixes for some of the problems. For example, another long-used separation technology called liquid chromatography does a better job of scanning for small proteins. Although several companies are adding the technique to their arsenals, Celera, for one, is so convinced of its superiority that it is relying exclusively on it for separating proteins. Still, no one denies that all these technologies will overlook many, if not most, proteins. "But we will be able to detail very, very many," adds Rose. And for now, he and others think that they'll generate plenty of moneymaking discoveries. And new tools are on the way, they say, as academic and corporate researchers continue to push the envelope on characterizing proteins (see sidebar, p. 2076). "The technology is getting 10 times better every year," says Mann.



Yeast–Two-Hybrid

Researchers insert a gene in yeast for a "bait" protein alongside DNA for half of an "activator" protein. The other half of the activator DNA is then inserted alongside DNA for random "prey" proteins. The yeast cells are then grown up and the proteins are allowed to interact. If bait and prey proteins bind, the two halves of the activator protein be close enough to work together to turn on another yeast gene that turns the cell blue, signaling a match.

ILLUSTRATION: C. SLAYDEN

Whither academics

With much of proteomics dominated by biotech and drug companies, many researchers wonder whether the field is already beyond the reach of academics. "All the really big studies are being done in the private sector," says Aebersold. Mann adds that "it's difficult for a university group to mount a sustained effort in proteomics."

But Hanash argues that the breadth of proteomics and the inability of one technology to answer all the questions will ultimately play to the strength of academia. "It's the exact opposite of the genome project," where Celera was able to match the public gene-sequencing effort with a single centralized sequencing lab, says Hanash. Large-scale commercial proteomics efforts "are very limited in what they can accomplish," he adds. "It will be a plurality of efforts that brings the payoff in proteomics."

One academic collaboration is already under way, the Alliance for Cell Signaling (ACS). Launched last year by Alfred G. Gilman at the University of Texas Southwestern Medical Center in Dallas, this consortium is made up of more than 50 experts in 21 institutions. It aims to track all the proteins that carry out communication in two types of cells, antibody-producing B lymphocytes and heart muscle cells. That's not technically a full-scale proteomics effort, because the alliance isn't tracking all the proteins in those cells. However, says Michigan's Andrews, "it is a doable project that makes sense for academics."

Many researchers are already beginning to look to ACS as a model for a publicly funded proteome effort, in part to ensure that not all proteomics data are locked up by companies. At a meeting in Washington, D.C., in October, academic leaders in proteomics huddled with representatives from the National Institutes of Health and other government funding agencies, as well as from proteomics companies, to discuss launching a coordinated initiative. The group decided to proceed with caution, recommending pilot projects in three areas: profiling protein expression in selected tissues, detailing proteins' functions, and creating new bioinformatics tools to handle the deluge of data. The projects, says Michigan's Hanash, will likely follow ACS's lead and focus on specific research topics, such as looking at cytoskeletal proteins or those involved in key organelles such as the energy-producing mitochondria. "We want to build the encyclopedia one chapter at a time," he says.

If there's any lesson from the early work in this field, it's that the proteomics encyclopedia isn't likely to come together quickly. And when it does, its pages will be scattered in databases around the world. It won't hold all the answers for understanding life inside the cell. And it won't remove most obstacles to turning drugs into products. But in the modern world of high-speed biology, being first to discover something means opportunity: for glory, for profits, and for the right to lead the field in new directions. For all its limitations, proteomics has just become biology's latest wellspring of opportunity.

–ROBERT F. SERVICE