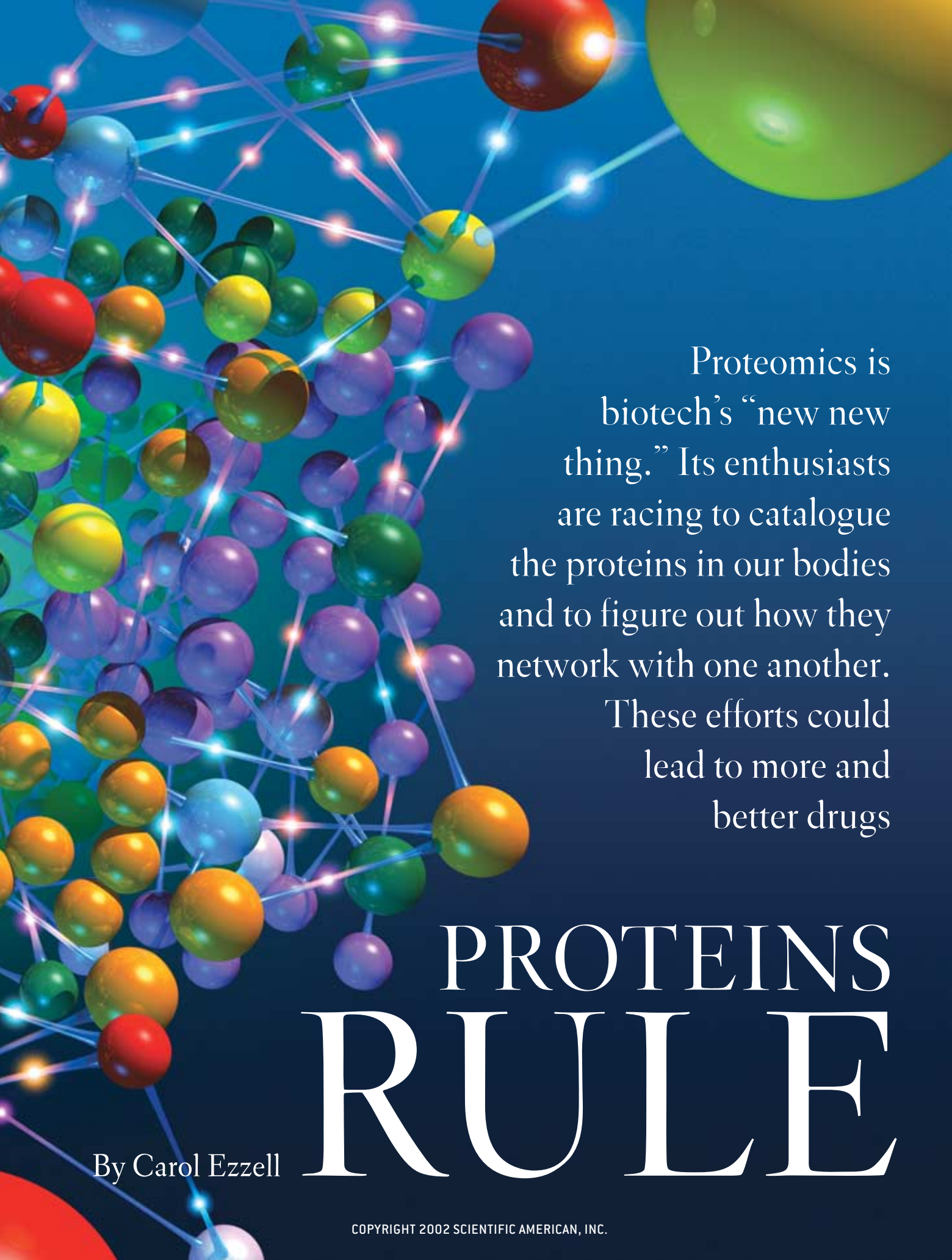
NETWORKS of proteins pervade all cells.



Proteomics is biotech's "new new thing." Its enthusiasts are racing to catalogue the proteins in our bodies and to figure out how they network with one another. These efforts could lead to more and better drugs

# PROTEINS RULE

By Carol Ezzell

# Move over, human genome,

your day in the spotlight is coming to a close. Researchers are now concentrating on the human proteome, the collective body of proteins made by a person's cells and tissues. The genome—the full set of genetic information in the body—contains only the recipes for making proteins; it's the proteins that constitute the bricks and mortar of cells and that do most of the work. And it's proteins that distinguish the various types of cells: although all cells have essentially the same genome, they can differ in which genes are active and thus in which proteins are made; likewise, diseased cells often produce proteins that healthy cells don't, and vice versa.

Accordingly, corporate and academic scientists are looking to catalogue all human proteins and uncover their interactions with one another. The goal is to devise better drugs with fewer side effects.

Reaching that goal won't be a walk in the park, though: proteins are even more difficult to study than genes, and biotech companies are still struggling to come up with the best techniques and instruments for the task. Nevertheless, a race of sorts is on, with at least one company predicting that within three years it will have deciphered the human proteome, an important step in piecing together the myriad interactions among the individual proteins. Meanwhile federal programs are offering money to academic scientists to study the proteomes of cancer cells and serum, the watery component of human blood.

Researchers have already made some important strides: in January two groups reported that they had made maps of how all the proteins interact in baker's yeast, a popular model for studying cell biology. Other scientists announced in February that they had used proteomics techniques to devise an accurate early test for ovarian cancer.

Proteomics is set to become big business. According to investment analysts at Frost & Sullivan, the worldwide market for proteomics instruments, supplies and services will reach roughly $5.6 billion by 2005, up from only $700 million in 1999—and that doesn't include income generated from drugs or diagnostics developed as a result of proteomics approaches. Proteomics could also be vital to the future of the pharmaceutical industry, says Jessica Chutter, managing director and co-director of biotechnology for Morgan Stanley. The industry spent $30 billion on R&D in 2000, she states, but only 30 drugs were approved that year. "Pharmaceutical companies are dependent on proteomics and like technologies to overhaul their entire drug development process, or they will not survive," she claims.

## Overview/*Proteomics*

- Now that the Human Genome Project is completed, scientists are turning to deciphering the networks of proteins within cells and tissues. But proteins are much more complex than genes and more difficult to study.
- Investors have poured hundreds of millions of dollars into companies that are devoted to producing proteomics equipment or to developing drugs or diagnostic tests based on proteomics techniques.
- Determining the three-dimensional structures of proteins allows researchers to find sites where proteins are most vulnerable to drugs.

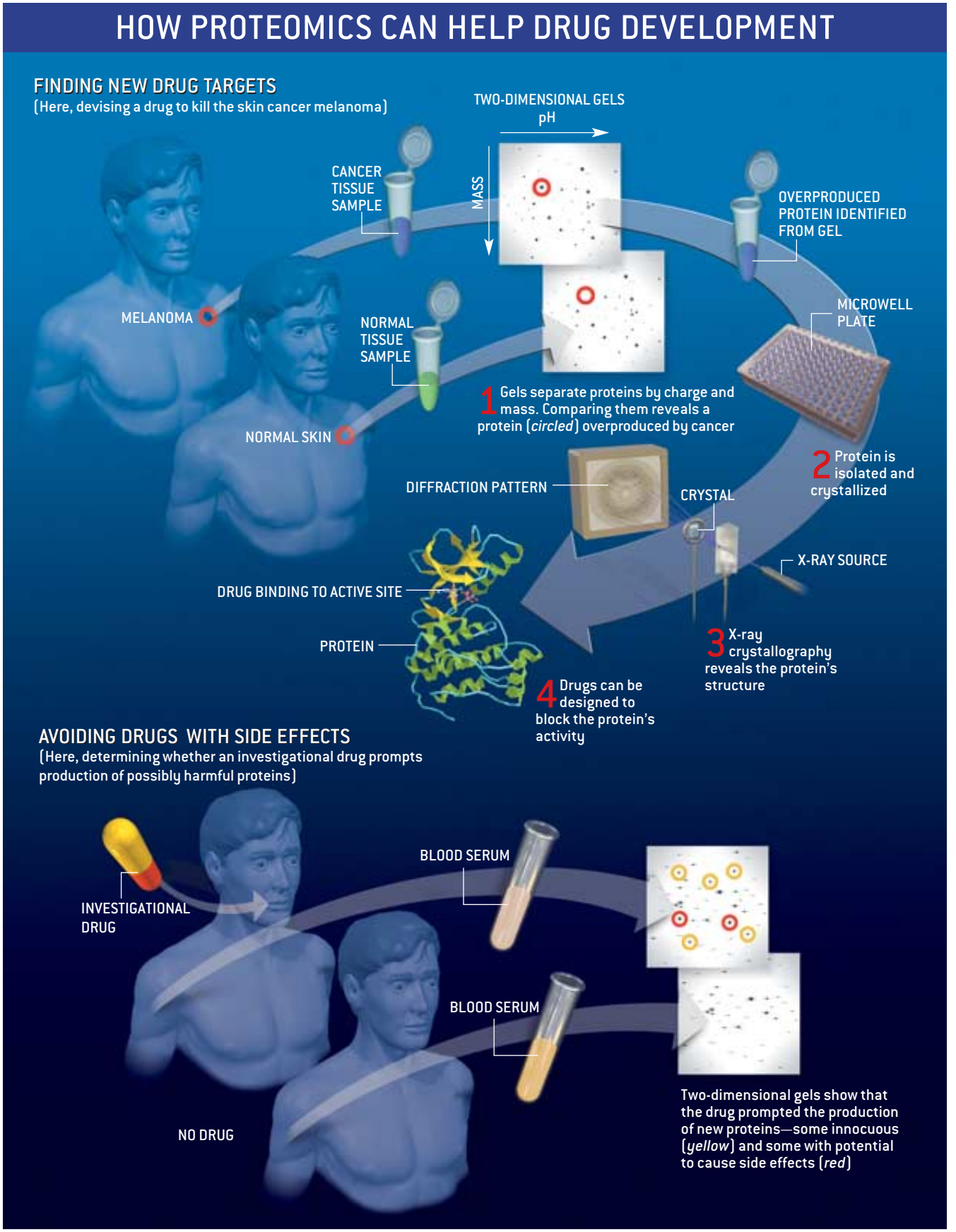
## "Genes Were Easy"

THE TERM "PROTEOME" was coined in 1994 by Marc R. Wilkins, vice president and head of bioinformatics at Proteome Systems in Sydney, Australia, to mean the protein complement encoded by a genome. (The jazzy "-ome" and "-omics" suffixes have proliferated in biology to the point that a Web site now lists dozens of terms carrying the appellations.) The exact definition of proteomics varies depending on whom you ask, but most scientists agree that it can be broken down into three main activities: identifying all the proteins made in a given cell, tissue or organism; determining how those proteins join forces to form networks akin to electrical circuits; and outlining the precise three-dimensional structures of the proteins in an effort to find their Achilles' heels—that is, where drugs might turn their activity off or on.

These tasks sound straightforward, but the title of a 2001 conference on proteomics says it all: "Human Proteome Project: 'Genes Were Easy.'" Some of the hoopla that surrounded the backbreaking, but now essentially completed, Human Genome Project gave the impression

SLIM FILMS [*preceding pages*]

# HOW PROTEOMICS CAN HELP DRUG DEVELOPMENT




SLIM FILMS; PROTEIN/DRUG STRUCTURE COURTESY OF DUNCAN McREE *Syrrx*

that knowing the sequence of the roughly three billion code letters, or DNA base pairs, occurring in the human genome—and specifically knowing the sequences in the protein-coding units (the genes)—would lead to an understanding of the proteins themselves.

Unfortunately, the proteome is much more complicated than the genome. The DNA "alphabet" consists of four chemical bases known by their first letters: adenine (A), cytosine (C), guanine (G) and thymine (T). Proteins, in contrast, are constructed from 20 building blocks, called amino acids. Genes specify which amino acids should be strung together to form a given protein. But even when scientists know the amino acid sequence of a protein, they cannot necessarily deduce what the protein does or which other proteins it engages with. Nor can they always forecast its three-dimensional structure with absolute accuracy. Unlike genes, which are linear, proteins fold into shapes that, in some cases, defy prediction.

Moreover, cells usually modify proteins by adding sugars or fats, or both, to them in ways that can be hard to anticipate as well. To produce a protein encoded by a newly discovered gene, scientists cannot merely string together amino acids in the order dictated by the gene; often they must also ensure that the proper fat and sugar modifications are made. And to determine how a protein behaves, researchers must also take into account that some proteins dissolve in water, whereas others act normally only in an oily environment or have regions that are embedded in fat-filled cell membranes.

That's not the end of the complexity. Although most researchers agree that the genome contains roughly 40,000 genes, a typical cell makes hundreds of thousands of distinct proteins. To understand the proteome, scientists have to learn the characteristics of all of those proteins. Simply making use of the data from the Human Genome Project, which finally put to rest the old dogma that one gene encodes one protein, will not do the trick. Clearly, one gene can somehow give rise to many different proteins.

Despite the complexities, proteomics researchers seem optimistic. "Some 30 to 50 percent of human proteins are unknown and of unknown function," admits Alma L. Burlingame of the University of California at San Francisco. But, he adds, "we now have the capacity to identify the protein components of human beings rather rapidly. It's tractable and will occur over the next couple of years."

## Proteo-Factories

WHEN SCIENTISTS WANT to find out which proteins are present in selected cells or tissues, they usually turn to two techniques: two-dimensional gel electrophoresis and mass spectrometry. With 2-D gels, scientists add a mixture of proteins to the edge of thin gel that separates proteins in one direction according to their size and in a perpendicular direction according to their electrochemical charge [*see illustration on preceding page*]. Because any given protein has a characteristic size and charge, each one shows up as a discrete dot on the gel. Researchers can cut individual dots from the gels to identify the proteins they contain using other techniques. And they can look for proteins made by one tissue but not another by comparing the dot patterns of gels made from the two tissues.

Mass spectrometry employs magnets or electrical fields to resolve distinct proteins according to the masses of their constituent atoms. The results are displayed as peaks on a graph. Neither the 2-D gel technique nor mass spectrometry is ideal, however. Two-dimensional gels are notoriously difficult to run and can't distinguish very large or very small proteins or those that protrude through membranes, and mass spectrometry is expensive (more than $500,000 per machine) and sometimes fails to detect rare proteins.

Nevertheless, several companies are preparing refined versions of these methods for use in industrial-size operations on the order of the ones that made the Human Genome Project possible. The workhorse of that project was the ABI 3700, a DNA sequencer from Applied Biosystems. (Applied Biosystems is now part of Applera, which also includes Celera Genomics, the company that tied with the government consortium in completing the human sequence in 2000.) In January, Applied Biosystems unveiled its mass-spectrometry-based 4700 Proteomics Analyzer and announced an agreement with PerkinElmer and Millipore to provide an automated system for running and analyzing 2-D gels. Company executives hope the automation will allow scientists to do in days what used to take months or years.

But whether these new systems will be the proteomics standards remains to be seen. "There's not going to be one tool that's going to be dominant to the industry," says Darlene Solomon, who oversees life-sciences research and development for Agilent, an Applera competitor. "There's so much to proteomics."

Meanwhile such companies as Myriad Genetics in Salt Lake City, GeneProt in Geneva, Large Scale Biology in Vacaville, Calif., and MDS Proteomics in Toronto have geared up with custom proteomics plants of their own, some of which employ robotics techniques borrowed from the automotive industry. Last year Myriad announced that it had joined forces with Hitachi and Oracle in a $185-million deal to decipher the entire human proteome in three years, a program that officially began in January of this year. Celera, for its part, has raised almost $1 billion for its proteomics efforts. Celera's founder, J. Craig Venter, stepped down as president in January, however, and the company announced it was looking for a replacement who had more expertise in drug development. The

Unfortunately, the PROTEOME is much MORE COMPLICATED than the genome.

ROBOTIC WORKSTATION in a proteomics facility borrows technology from the assembly lines of the automotive industry. This one is configured to automate repetitive tasks such as pipetting and changing the growth medium—steps involved in growing cell cultures, a prerequisite for proteomics studies. Other such modules industrialize protein isolation and identification to ready samples for further study.

move was widely interpreted as an indication that the company would swing away from a business model based on selling access to its genomics—and proteomics—data to other companies and toward one in which it would devise its own drugs.

Critics of such grand projects have pointed out that there is no single human proteome: the pancreas makes a very different set of proteins than the brain does, for instance, and many variables, such as whether someone has just had a glass of wine, can affect the types of proteins the body produces. "Every state—plus or minus disease, plus or minus drug—is a different proteome," explains Michael F. Moran, chief scientific officer of MDS Proteomics.

In other words, listing human proteins takes you just so far. To understand what proteins do in the body and to develop useful drugs, you need to know how the mix of proteins varies from one cell type to another and within a cell as conditions change. You also need to know how proteins collaborate to carry out a cell's various activities.

## Listening In on the Network

MORAN'S COMPANY is focused on this last task—examining how proteins hobnob with one another to form chains of biochemical reactions or make molecular machines such as the spindle that pulls two cells apart during cell division. "Proteins are assembled into networks," he says. "If you had to learn one thing about a protein, it would be what other proteins it interacts with."

In the January 10 issue of *Nature,* scientists from MDS Proteomics and the University of Toronto—and those in an independent group from Cellzome and the European Molecular Biology Laboratory, both in Heidelberg, Germany—reported coming up with a new strategy to find hundreds of such protein interactions in yeast. Their approach involves attaching bits of DNA that encode sticky "tags" to hundreds of selected yeast genes. The researchers can then isolate the proteins made from the modified genes, along with any proteins that have bound to them, by grinding up the yeast and pouring the slurry through a column of microscopic beads that can bind only to the sticky tags. After running the protein complexes through a mass spectrometer and analyzing the results, the scientists found that

MDS PROTEOMICS

## A PROTEOMICS ROLODEX

**CELERA** (an Applera Corporation Business)
www.celera.com
Rockville, Md.
Proteomics efforts focus on comparing normal and diseased tissue to find disease-related proteins that could be targeted with monoclonal antibodies, cellular immunotherapy or small-molecule drugs.

**cellzome**
www.cellzome.com
Heidelberg, Germany
Founded by scientists at the European Molecular Biology Laboratory. Characterizes cellular protein complexes and is developing maps of protein interactions and pathways under varying experimental conditions.

**CuraGen Corporation**
www.curagen.com
New Haven, Conn.
Screens libraries of genes to identify interactions among the proteins they encode and to further understand how genes function in disease. Holds a key patent on the yeast two-hybrid method for outlining protein networks.

**GeneProt™**
U.S. headquarters: North Brunswick, N.J.
www.geneprot.com
Identifies, characterizes and synthesizes select human proteins for use in the discovery and development of new therapeutics. Second industrial-scale proteomics facility to open in 2002 at New Jersey headquarters.

**hybrigenics**
www.hybrigenics.com
Paris
Has developed industrial-scale technology for identifying, selecting and validating drug targets. Goal is to build a pipeline of disease "biomarkers" and small-molecule and antibody drug products.

**LARGE SCALE BIOLOGY**
www.lsbc.com
Vacaville, Calif.
Building a linked family of comprehensive databases of the human proteome, including protein markers for use in diagnosing and monitoring disease. Has both in-house and collaborative programs for drug development.

**mds proteomics**
www.mdsproteomics.com
Toronto
Identifies, selects and validates protein targets for both antibody and small-molecule therapeutics. Focuses on cancer, particularly on receptors and intracellular signaling, the networks of messages within cells.

**MYRIAD**
www.myriad.com
Salt Lake City
Determines the functions of individual proteins and how proteins form the complexes that constitute enzymatic machines and signaling circuits. Uses yeast two-hybrid and mass-spectrometry techniques.

**SGX**
www.stromix.com
San Diego
Uses high-throughput technology to determine protein structures of key targets within protein families. Employs bioinformatics techniques to virtually "dock" candidate drugs into the binding sites of drug targets.

**syrrx**
www.syrrx.com
San Diego
Generates three-dimensional protein structures and uses them to discover new drugs. The Syrrx Protein Structure Factory couples molecular tools with robotics to determine protein structures on an industrial scale.

*SOURCE: Company materials*

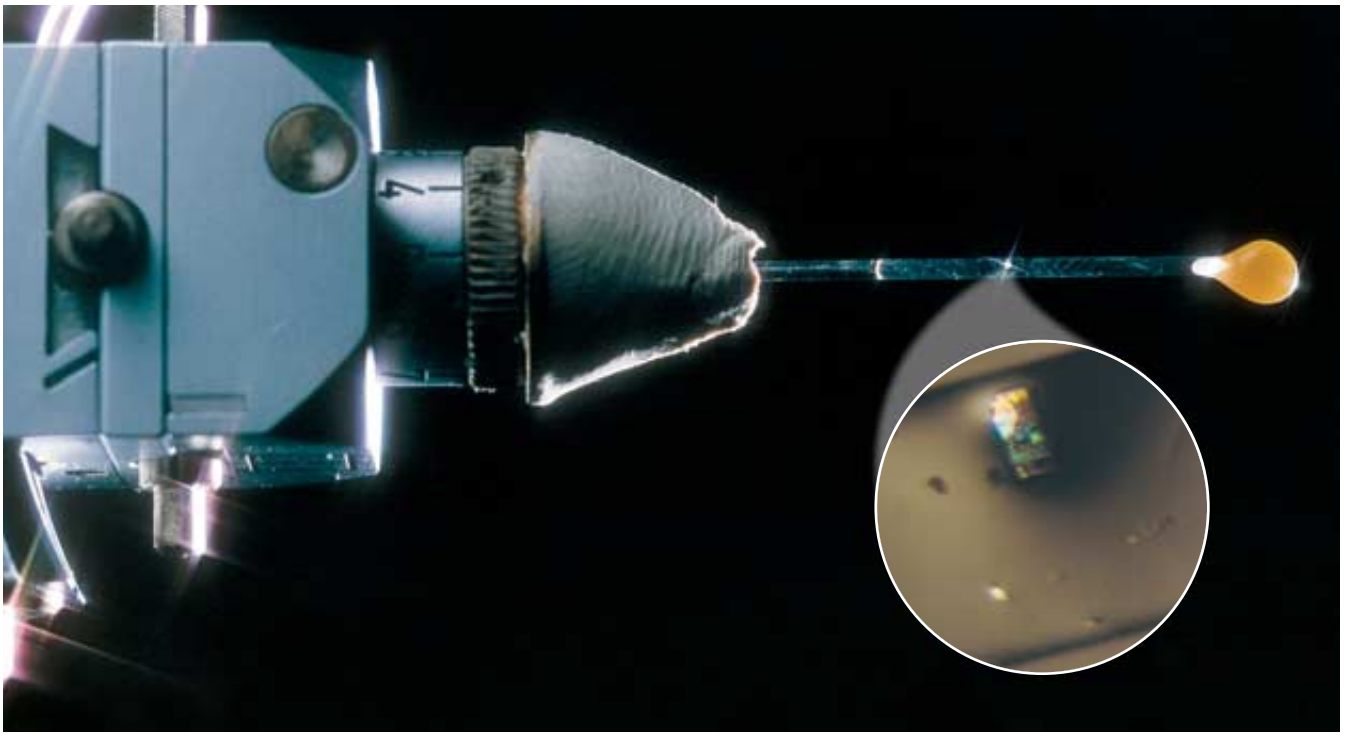
more than 90 percent of the complexes they isolated contained proteins of unknown function. What is more, up to 80 percent of the proteins interacted with at least one other protein, demonstrating the intricacy of the biochemical network within cells.

MDS Proteomics now plans to use the technique on the human proteome. Because the yeast proteome project took only a matter of weeks, company officials predict that they can produce an initial snapshot of the proteome of a human cell within a year. It is not yet clear what type of human cell they will study and under which conditions, however.

The public sector is also gearing up for proteomics. Academic researchers led by Samir M. Hanash of the University of Michigan have established the Human Proteome Organization (HUPO), which aims to link public proteome projects much as the Human Genome Project tied together academic labs deciphering the human genome. One of HUPO's first goals will be to determine the proteins present in blood serum.

The National Cancer Institute (NCI) and the Food and Drug Administration (FDA) have joined in a separate effort to focus on using proteomics to develop more targeted treatments and more reliable diagnostics for cancer. In the program, which was announced in July 2001, researchers will analyze tumor cells from individual patients to come up with a roster of proteins present in cancer cells but not in normal ones. They will also search for protein "markers" that correlate with more aggressive cancers, perhaps leading to better diagnostic tests.

Emanuel Petricoin, co-director of the NCI/FDA program, and his colleagues at the agencies and at Correlogic Systems in Bethesda, Md., recently demonstrated the promise of a proteomics approach to diagnosing cancer. In a paper published February 8 on the Web site of the journal the *Lancet*, the researchers report that they were able to compare the patterns of proteins present in the blood serum of patients with and without ovarian cancer. Through the comparison, they correctly identified all of 50 women who had ovarian cancer. Their test yielded only



X-RAY CRYSTALLOGRAPHY requires growing a pure crystal (*inset*) of the protein under study. Here a crystal of CD4, the protein that serves as a gateway for the AIDS virus to infect immune cells, is held in a tiny tube sealed with a ball of wax. The tube will be bombarded with x-rays to yield a pattern that scientists can interpret to determine the three-dimensional structure of an individual molecule of the protein.

three false positives among the samples of women who did not have the cancer.

Catalogues and maps of protein-protein interactions are only two thirds of proteomics; determining the shapes of proteins is equally important. The classic technique is x-ray crystallography, in which scientists purify proteins, allow them to grow into crystals and then bombard the crystals with x-rays. By analyzing how the x-rays bounce off the individual atoms of a protein, researchers can deduce how the protein is put together and can map its overall three-dimensional shape.

## The Shape of Things to Come

X-RAY CRYSTALLOGRAPHY was once something of a cottage industry and required access to the x-ray beam line of a synchrotron. These often enormous rings, which can be miles in diameter, have historically been used by physicists to accelerate atomic particles. The x-rays are produced as part of that process. But advances in x-ray lasers have led to tabletop devices that can be used in labs.

Two companies—Syrrx and Structural GenomiX (SGX), both in San Diego—have now taken x-ray crystallography industrial. "Today everything is done robotically," explains Nathaniel David, cofounder and director of business development at Syrrx. Like the companies that have automated the protein discovery process, Syrrx has borrowed techniques from the automotive industry. Indeed, it brought in consultants from General Motors to automate its 84,000-square-foot facility, where everything from protein purification to crystallization is done on an assembly line. Besides its own x-ray lasers, the company has a dedicated beam line at the Advanced Light Source at Lawrence Berkeley Laboratory. Structural GenomiX has a similar deal with the Advanced Photon Source at Argonne National Laboratory, where it has built a beam line.

Such structural information could be bankable. Oxford GlycoSciences in England is betting that it can tie up the patent rights to a significant portion of the human genome and proteome using proteomics data. Last December the company filed patent applications for 4,000 human proteins, a move that could shake up how intellectual property is defined in biotechnology. In the past, companies sought to patent DNA sequences and the single protein that they predicted would be encoded by them. But because the same gene can make a range of proteins, claims based on the proteins themselves could be more valuable and offer a way to get around patents on the DNA sequences held by competitors. If so, the courts could be one more arena where genes will have to move over in favor of proteins. SA

*Carol Ezzell is a staff editor and writer.*

### MORE TO EXPLORE

**High-Speed Biologists Search for Gold in Proteins.** Robert F. Service in *Science,* Vol. 294, pages 2074–2077; December 7, 2001.

**Separation Anxiety: Why Proteomics Can't Let Go of 2D Gels.** Aaron J. Sender in *Genome Technology,* No. 16, pages 34–39; December 2001.

For a glossary of the proliferating biotechnology terms ending in "-ome" and "-omics," visit **www.genomicglossaries.com/content/omes.asp**

KAY CHERNUSH