The upstart genomics firm is set to embark on an ambitious new effort in protein analysis, but it faces a huge challenge and some stiff competition from companies already in the field

Can Celera Do It Again?

J. Craig Venter and Michael Hunkapiller proved the skeptics wrong once. In 1998, the two teamed up in a bold plan to sequence the human genome by the end of 2001, a full 4 years ahead of the finish date then projected by the publicly funded Human Genome Project. Venter, head of Celera Genomics Corp. in Rockville, Maryland—the firm created to carry out this mission—was the brains behind a novel genome sequencing approach. Hunkapiller, head of the scientific equipment powerhouse PE Biosystems, provided brawn: a suite of the company's experimental high-speed DNA sequencers. Despite strong doubts among genome researchers that Celera could pull it off, the project appears to be just months away from completion.

Now the pair is preparing to push Celera beyond the genome to conquer the next frontier—the identification of proteins involved in human disease. To do so, Celera is launching a major effort in "proteomics," an effort to identify all the proteins expressed in an organism and then track their ebb and flow. The move, says Venter, is the next logical step in understanding the role of all the genes they've decoded. It is also a critical step in developing novel drugs and tailoring medical care to the genetic makeup of individuals. To pay for the new initiative, Celera raised $944 million in a stock offering earlier this month, much of which will be devoted to proteomics.

This time, however, Venter and Hunkapiller are tackling a much more complex problem, and they will be facing competition from companies ranging from small start-ups to big pharma that have already started proteome projects of their own. "I don't think they are more credible than anybody else," says Mattias Mann, chief scientific officer of Protana, a proteomics company in Odense, Denmark.

But Venter, never one to understate his ambitions, boasts: "We're going to dominate in our own way. We're going to have the biggest facility and the biggest database." He concedes that the huge amount of work needed to understand proteins and their interactions inside cells guarantees that many academic labs and other companies will also be players in the field. But, he says, "we're building a Celera-scale proteomics facility" capable of identifying up to 1 million proteins a day.

Plans for the new facility are still coming together, Venter says, but it will likely consist of a fleet of up to 100 machines, including high-speed mass spectrometers for protein analysis, as well as additional protein separation devices. Celera also plans to boost the capacity of its $100 million supercomputer—which currently holds some 50 terabytes of genome data—by a factor of 10 to handle the expected torrent of protein data.

KEY CORPORATE PLAYERS IN PROTEOMICS

Company	Location	Approach
Celera	Rockville, MD	Databases
Incyte Pharmaceuticals	Palo Alto, CA	Databases
GeneBio	Geneva, Switzerland	Databases
Proteome Inc.	Beverly, MA	Databases
PE Biosystems	Framingham, MA	Instrumentation
Ciphergen Biosystems	Palo Alto, CA	Protein arrays
Oxford GlycoSciences	Oxford, U.K.	2D gel/MS*
Protana	Odense, Denmark	2D gel/MS
Genomic Solutions	Ann Arbor, MI	2D gel/MS
Large Scale Proteomics Corp.	Rockville, MD	2D gel/MS

* 2D gel electrophoresis and mass spectrometry.

Venter will have Hunkapiller's help getting his operation up to speed. Last week, PE—the parent company of both Celera and PE Biosystems—announced that it will form a proteomics research center at its PE Biosystems site in Framingham, Massachusetts, to create new high-speed machines. As part of that initiative, PE officials plan to pursue two technologies—one developed by Denis Hochstrasser and colleagues at the University of Geneva in Switzerland, the other by Ruedi Aebersold at the University of Washington, Seattle—that aim to do for protein analysis what the high-speed gene sequencers did for genome work.

Outside observers say that with PE Biosystems' backing, Celera's move into proteomics is likely to be pivotal for the emerging field. "*The* genomics company stuck its flag in the arena of proteomics," says William Rich, president and CEO of Ciphergen, a proteomics company based in Palo Alto, California. "It sends a message to the protein people that the gene people are not going to sit around and wait" for proteomics companies to give them the information they're looking for.

In fact, the march of genomics companies into proteomics is well under way. Incyte Pharmaceuticals, one of Celera's chief genomics rivals, is 2 years into an extensive partnership with Oxford GlycoSciences (OGS), a proteomics company based in Oxford, England. OGS creates protein profiles of different tissues in both healthy and diseased states, and Incyte incorporates this information into a proteome database that it markets to pharmaceutical companies. Incyte signed up its first subscriber to its proteomics database last fall, giving it "a significant lead over other companies in developing proteomics databases," asserts Incyte CEO Roy Whitfield. And like Celera, the company is also flush with cash. Incyte recently raised $620 million on the stock market, much of which is intended to bolster their proteomics work, says Whitfield.

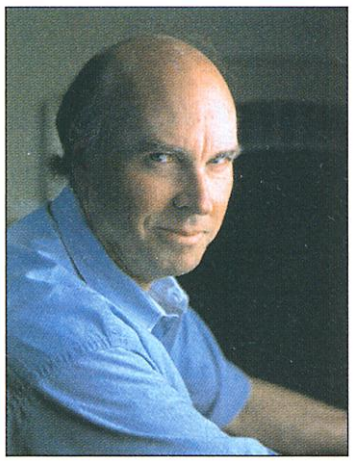
Other companies are pushing into proteomics as well. Virtually every major pharmaceutical company has a proteomics effort under way, says Hanno Langen, who directs proteomics research at Hoffmann–La Roche in Basel, Switzerland. Moreover, small proteomics firms, such as Genomic Solutions of Ann Arbor, Michigan, and Large Scale Proteomics of Rockville, Maryland, are gearing up for initial public stock offerings to raise money for expanded research.

That makes Celera a late entry into the field, and it has some catching up to do. But it is betting nearly $1 billion that it can close the gap.

The next step

This move toward understanding proteins has emerged from the increasing recognition

among genomics and pharmaceutical researchers that identifying DNA, or even messenger RNA (mRNA)—the nucleotide messengers that signal cells to produce a particular protein—is not enough. Neither DNA nor mRNA can identify how much protein is produced inside a cell or what it does once created. Although researchers initially hoped that the presence of a large amount of a particular mRNA meant that copious quantities of the corresponding protein were being produced, "there is significant evidence that there is not necessarily a correlation between mRNA levels and protein levels," says Philip Andrews, a proteomics researcher at the University of Michigan, Ann Arbor. Other factors complicate the picture as well, he adds. For instance, chemical modifications such as phosphorylation play a key role in controlling protein activity; these modifications cannot be detected by screening nucleotides. "The genome tells you what could theoretically happen" inside the cell, explains Raj Parekh, the chief scientific officer at OGS. "Messenger RNA tells you what might happen, and the proteome tells you what is happening."



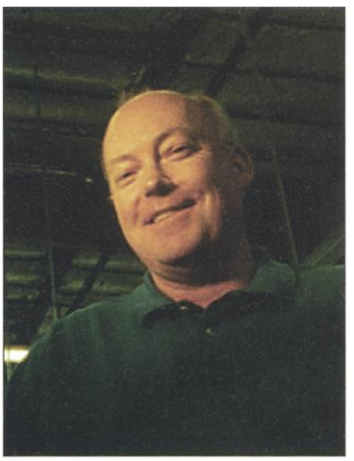
J. Craig Venter: "We're going to dominate in our own way."

Figuring out what's happening at the protein level won't be easy even for Celera. "Proteomics is a much more difficult problem than genomics," says Andrews. Whereas the human genome remains largely unchanged among individuals, he explains, the expression of proteins varies widely. Protein expression changes dramatically from one tissue to another and even within single tissues over time as a person ages. What's more, thousands of chemical modifications occur after proteins are created that alter their enzymatic activity, binding ability, how long they remain active, and so on. Although there may be only some 100,000 human genes, the myriad of modifications may give rise to 10 million to 20 million chemically distinct proteins in a cell, says Andrews.

This complexity, Andrews and others say, makes it almost meaningless to consider a human proteome project—akin to the human genome project—to identify all proteins in every tissue. The best researchers can do is try to focus on changes in key proteins, such as those involved in disease and development. For that reason, skeptics argue that even with Celera's deep pockets it will not be able to sweep aside the competition. "I think it will be very hard for any company to be the dominant proteomics company, much more so than in genomics," says Mann of Protana. Venter agrees—in principle. But he adds that few competitors will be able to match Celera's industrial approach. "We'll be working through every tissue, organ, and cell," he says.

Brute force

In the current proteomics rush, most companies are taking more or less the same brute-force approach to determining which proteins are present in various tissues, a technique called two-dimensional (2D) gel electrophoresis. Researchers start with a protein extract from a tissue of interest and then add it to a sheet of polymer Jell-O. By applying electric fields across the length and width of the sheet, they separate proteins by their electric charge and size. The result is a series of up to several thousand spots, each containing one or more types of proteins.



Michael Hunkapiller promises mass spectrometry machines "orders of magnitude" faster.

Once segregated into separate spots and stained, the proteins are typically cut from the gel one by one, chopped into fragments with an enzyme called trypsin, and dried. Then they are fed into a mass spectrometer that weighs each fragment, forming what amounts to a mass fingerprint of the protein's fragments. From that fingerprint, researchers can work out the likely combination of amino acids comprising it and then compare that to a genomics database to identify the corresponding DNA sequence. With the DNA sequence in hand, they can get a clearer identification of the protein. Researchers can then monitor changes in the expression of that protein to see whether it correlates with a disease state or perhaps a drug response.

All of this takes time. Today's top-of-the-line 2D gel operations, complete with robots to cut apart the gels and computers to analyze the information, can study a couple of thousand proteins a day. As a result, it can still take months to figure out which proteins change their expression in one set of tissues.

It's here that Venter and Hunkapiller hope to make a difference. PE Biosystems produces mass spectrometers, among other things. And according to Hunkapiller, mass spectrometers now in the prototype stage have the potential to go "orders of magnitude" faster than the current variety, making it possible to analyze hundreds of thousands of proteins per day. That scale of improvement in gene-sequencing machines enabled Celera and others "to ask whole new questions," says Hunkapiller. "If we can do the same thing in the protein area, it should have the same effect." By analyzing a large proportion of proteins in cells instead of just a select few, "you will see the things going on that maybe weren't so obvious," predicts Hunkapiller.

But this newfound speed will create bottlenecks upstream and require a faster front-end procedure to separate proteins and feed them into the high-speed mass spectrometer. This is where PE is looking to Geneva's Hochstrasser for help. Last fall, Hochstrasser and his colleagues published work on a new laser-based "molecular scanner" that automates the process of moving separated proteins from the gels to the mass spectrometers. The system applies an electric field perpendicular to the gel sheet; this field draws the proteins through two membranes. The first membrane is studded with trypsin, which digests all the proteins in a gel simultaneously as they move past. The fragments are then trapped by a second membrane, which is fed to the mass spectrometer. Finally, a laser marches along the membrane firing a steady stream of pulses in micrometer-sized steps, each of which blasts protein fragments into the mass spectrometer for fingerprinting. Here, too, the speed should reach tens of thousands of proteins per day, says Hochstrasser.

Working in tandem, the molecular scanner and a high-speed mass spectrometer could be a powerful combination, says Andrews. "The idea is spectacular. But whether it will perform as it needs to remains to be seen," he cautions. OGS's Parekh, for one, is skeptical. A big problem, he says, is how to quantify the amount of protein in each spot on the gel. This is necessary for identifying which proteins

CREDITS: (TOP TO BOTTOM) MARTY KATZ; WILLIAM MCLEOD

change in a stable way with disease—itself a first step to identifying molecular markers of disease and potential drug targets.

The time-consuming approach of gel cutting does allow such quantification. But that's typically not the case when the spots are transferred, or blotted, to another membrane, says Parekh. "Some proteins don't blot. Others are lost in the process. So as soon as you blot, you lose the quantity information," he says.

PE is betting that Aebersold's technology will help. In the October 1999 issue of *Nature Biotechnology*, Aebersold and his University of Washington colleagues reported a new approach that uses stable isotopes to quantify numerous proteins in cell extracts with mass spectrometry. The advance is "outstanding," says Hochstrasser, because it allows mass spectrometers to pin down protein levels from large numbers of proteins at once—a difficult proposition with today's technology. At this stage, it is not clear whether the new technique will work with a molecular scanner and a high-speed mass spectrometer. To find out, Aebersold is already working on joint research projects with PE scientists.

Even if it takes a while to get the mass spectrometers up to top speed, Celera can still make considerable progress, says Venter, as the company will be using other proteomics tools as well. A key approach, he says, will be to create antibodies to all proteins. These antibodies can then be used to fish out of a sample both targeted proteins and those they interact with. That, Venter says, will help Celera build up a database of how proteins interact with each other in complex biochemical pathways—information that is likely to be valuable to drug companies aiming to intervene in those pathways at specific points.

Incyte's Whitfield says he is not fazed by Celera's entry into the field. Even if PE and Celera manage to pull all these pieces together and launch a high-speed proteomics effort, other companies will also be developing their own high-speed approaches, he says: "We all understand that faster, cheaper, better is the way to go." With Celera preparing to enter the field, Whitfield adds, "I'm sure there is going to be great competition."

–ROBERT F. SERVICE

TECHNOLOGY CONTROLS

Space Scientists Decry Stricter Export Rules

Congressional moves to tighten technology regulations could restrict collaborations with non-U.S. scientists

When Stanford University needed a proton detector for a NASA-funded satellite that would probe Einstein's theory of relativity, it ordered one from a Hungarian scientist living in Ireland. Last summer the detector was shipped to a Lockheed Martin facility near Denver for testing. But company officials, citing new rules on the export of sensitive satellite technology, said the scientist was not welcome without express permission from the State Department—a notoriously costly and time-consuming process. Stanford scrambled successfully to find another company that would allow the designer to test the instrument, and work continues on the payload (*Science*, 10 March, p. 1726).

This incident and other episodes—including one in which universities were reluctant to bid on a program for fear of breaking the rules—have left researchers worried about the status of international collaborations. Lockheed's hard line stems from rules drawn up last spring by the State Department, at Congress's insistence, that are designed to prevent sensitive technology from falling into the wrong hands. Researchers fret that a strict interpretation of the rules could have dire consequences, including jail if, for example, they discuss spacecraft designs with foreign graduate students without prior approval. "This is a terrible problem," fumes Stanford physicist and engineer Brad Parkinson. "It flies in the face of reason." But officials at State insist that scientists are exaggerating the threat and that the regulations should not affect research in any dramatic way. "It's an overreaction," says William Lowell, chief of State's office of defense trade controls. "I'm sure we can address 95% of these problems."

The dispute is the latest battle in a long-running war between defenders of national security and scientists who work in sensitive areas. In the early 1980s, for example, scientists fought attempts by the Reagan Administration to regulate the flow of research equipment and data overseas through U.S. arms traffic regulations, which control the import and export of technology, such as satellite technology, with military applications. A 1985 executive order marked a truce by declaring that the problem would be handled by classifying sensitive materials as secret rather than by imposing restrictions on exports.

Under wraps. New rules on sensitive technology affect access to payloads such as Gravity Probe B.

CREDIT: R. UNDERWOOD/LOCKHEED MARTIN MISSILES AND SPACE

But in 1998 Congress told the Administration to take a tougher stance on technology exports, citing the alleged transfer of sensitive satellite technology to China in the course of that country's launching of U.S.-built payloads. As part of the annual authorization of defense programs, it ordered the Administration to shift responsibility from the Commerce Department, which had exercised oversight since the early 1990s, to the State Department. Commerce traditionally seeks ways to promote trade and has a license exemption for technology connected with fundamental research, while the State Department has a reputation for taking a more hawkish position on the transfer of technology.

The new rules put spacecraft technology, with the exception of information in the public domain, on the department's roster of technologies deemed militarily sensitive. That includes not just hardware but also any technical data and the act of providing technical data to foreigners on the design, manufacture, use, and repair of an item. "Under existing rules, [even] speech can be deemed subject to licensing," says one official at the National Research Council (NRC).

A small cadre of policy-minded researchers have begun to protest what they see as a grave danger to scientists with overseas collaborators or who purchase equipment from abroad.