NEWS

Structural Genomics Offers High-Speed Look at Proteins

With the human genome nearly sequenced, the next frontier is understanding the shape and function of the encoded proteins

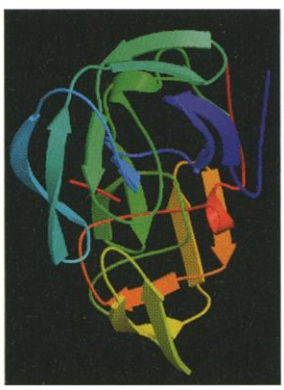
The research was only a tiny victory in the ongoing war against AIDS. But it showed the power of a new technique that's likely to have a major impact on the world of drug discovery.

In the late 1980s, researchers at the pharmaceutical giant Merck and elsewhere managed to snap a series of pictures of the HIV protease, the protein that helps the deadly virus replicate. The pictures, three-dimensional (3D) close-ups, revealed the precise locations of all of the protein's component atoms. The scientists used those maps to design drugs that would interact with the protease and gum up its works. Their research has helped to cut drastically the number of AIDS deaths in developed countries in the last few years.

Such "rationally designed" drugs are a step up from the trial-and-error process that has produced most drugs on the market today. One limiting factor has been the lack of detailed structural maps of the target proteins. But the current number of 1500 unique proteins for which these maps exist—about 1% of the estimated total in humans—is set to rise dramatically as powerful computers, robots, and other high-tech tools are enabling researchers to obtain structures for hundreds of proteins in the time it used to take to get one. Says Aled Edwards, a biochemist at the University of Toronto in Canada: "Structural biology has turned the corner."

The new direction is a souped-up version of the field called structural genomics. It applies high-speed techniques to make a systematic survey of all protein structures, cataloging the common ways in which proteins fold. That information, in turn, could eventually lead to computer programs capable of predicting the shape and function of any protein from the simple linear sequence of A's, G's, C's, and T's in genes. Now, with this newfound efficiency in sight, researchers have begun to think audaciously about surveying the complete spectrum of protein structures in the same way that geneticists have used DNA sequencing machines to decode entire genomes. The approach is expected to extend the genomics revolution from a catalog of genes to a catalog of the 3D shapes of the proteins for which the genes code.

This untold wealth of biochemical information at the atomic level is also likely to provide drug designers with both new targets for drugs and hints for making them successful, as it did with HIV protease inhibitors. But some researchers see a darker side, too. They worry that the nascent field will spawn a high-stakes race between publicly funded academic researchers and private companies, similar to the competition in the genomics community over the last decade.



Crystal clear. Governments and companies are rushing to fund efforts in structural genomics that promise to unlock the secrets of thousands of proteins—like this structure of HIV protease, which allowed researchers to design inhibitors to block replication of the virus.

PUBLICLY FUNDED ACADEMIC PROJECTS

Country	Center	Funding (in millions)
U.S.	NIH Structural Genome Centers	$20
Japan	NMR Park, Yokohama	$49
Germany	Protein Structure Factory, Berlin	$20
Canada	Univ. of Toronto	$23

CORPORATE PLAYERS

Company	Location	Focus
Genomics Institute of Novartis Research Found.	La Jolla, CA	High-throughput x-ray crystallography
Structural Genomix	San Diego, CA	" " "
Syrrx	La Jolla, CA	" " "
Chalon Biotech	Toronto, Canada	" " "
Structure Function Genomics	Princeton, NJ	" " " and high-speed NMR
Inpharmatica	London, U.K.	Structural informatics
Structural Bioinformatics	San Diego, CA	" " "
De Novo Pharmaceuticals	Cambridge, U.K.	" " "

SOURCE: PROTEIN DATA BANK; *SCIENCE*

Taking a fresh look

Historically, 3D protein structures have been created with the help of techniques such as x-ray crystallography. But the process is slow. To obtain a crystal structure, for example, structural biologists must first isolate proteins of interest, purify them, and coax them to form crystals. Then they must subject them to beams of x-ray photons, measure the way those photons ricochet off the crystal, and work out the configuration of atoms that must be present in the crystal to produce the diffraction pattern. For many proteins, this process has taken months or even years. As a result, structural biologists have tended to seek out crystal structures for proteins for which the biochemical function was already known in order to better understand how the molecules accomplished their tasks.

All that is beginning to change. Today, virtually every aspect in the structure-determining process is being automated and scaled up. Researchers are now creating high-throughput systems for cloning genes, expressing proteins, growing crystals, and collecting x-ray data (*Science*, 27 August 1999, p. 1342). More powerful x-ray beams at synchrotrons around the world have improved the quality of diffraction data, and better computers and software have made interpreting results both easier and faster. "All these things have suddenly made high-throughput crystallography possible," says David Eisenberg, a structural biologist at the University of California, Los Angeles. Similar advances are propelling work in nuclear magnetic resonance (NMR) spectroscopy, a related approach to determining structures. The upshot, says Steven Almo, an x-ray crystallographer at the Albert Einstein College of Medicine in New York City, is that "we can do structural genomics on the same scale that people do genome sequencing."

Genomics is also bringing a sea change in structural biology. "The genome project has changed our attitude completely," says Tom Terwilliger, a crystallographer at the Department of Energy's Los Alamos National Laboratory in New Mexico. Gene sequencing, he

Early Successes Hint at Big Payoff, But the Road to New Drugs Is Long

With researchers racing to set up structural genomics efforts in both academia and industry, two unanswered questions loom over the field. How well will this highly experimental approach work? And will it lead to the development of new drugs? "It's early days in the field," says John Moult, who heads a budding structural genomics group at the University of Maryland's Biotechnology Institute in Rockville. "But we are very encouraged."

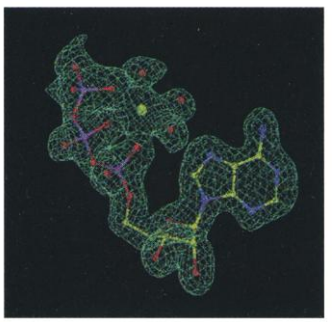
Moult's team is attempting to glean structural information from dozens of proteins in *Haemophilus influenzae*, a microbe that causes bacterial meningitis, among other diseases. The bug's full gene sequence was worked out in 1995 by researchers at The Institute for Genomic Research in Rockville, Maryland. Then, starting with these genes, Moult's group cloned 55 of them into *Escherichia coli*, coaxed bacterial hosts to overexpress them, purified the proteins, crystallized them, and used x-ray crystallography to resolve their structures.

Moult and colleagues have run just a few proteins through this gauntlet, but the results have come quickly. One newly identified protein folds in a manner similar to the anticancer protein endostatin. And its shape suggests that it spends its days binding to either DNA or RNA, project member Osnat Herzberg reported last summer at the American Crystallographic Association meeting in Buffalo, New York. Other protein structures that the team has resolved but not yet reported have been even more revealing of function, she says.

That structural knowledge clearly speeds the process of understanding how the protein functions inside the cell, asserts Herzberg. Without structural information, she says, "if you don't know anything about the function, you have to do tens of thousands of assays to understand the biochemistry. That's not practical."

The Maryland team is not the first to tease out structural information from unfamiliar proteins using this approach. In the 22 December 1998 issue of the *Proceedings of the National Academy of Sciences*, Sung-Hou Kim and his colleagues at the University of California, Berkeley, described a protein from a microbe called *Methanococcus jannaschii*, which lives in the extreme conditions found near volcanic vents. From the structure, they deduced that the protein was either an energy-producing enzyme or a molecular on-off switch. The group performed a couple of quick biochemical assays and found that it is an on-off switch. "The conclusion is that structural information is powerful for inferring functional information of unknown proteins," says Kim. Meanwhile, other structural genomics centers around the globe are beginning to turn up equally encouraging results.

Although the structural genomics approach promises important information on a protein's basic biology, it's not clear whether such information would benefit companies searching for new pharmaceutical and agricultural compounds. True, the three-dimensional structure of the HIV protease made it possible for researchers to design the hugely successful AIDS drugs called protease inhibitors. But that may be an exception, some caution. One concern is that the bulk of the high-speed structure-determining techniques may work only on proteins that make the least interesting targets for new drugs. Most drugs are designed to target proteins that straddle cell membranes and help to control the molecular traffic of cells. But membrane proteins are exceedingly difficult to crystallize even one at a time and thus aren't ready for these new high-speed techniques. As a result, researchers using these techniques will focus initially on nonmembrane proteins to hone their skills and hope to generate enough good science to justify the effort.



Clued in. Berkeley researchers used this structure to determine protein's role as a molecular switch.

Although most drugs are still discovered through painstaking trial and error rather than design, Chris Sander, the chief science information officer at Millennium Pharmaceuticals in Cambridge, Massachusetts, is convinced that such techniques are rapidly becoming obsolete. "The fraction [discovered by design] will increase steadily but surely," he predicts. If he's right, the race to resolve protein structures could get very hot, very soon. **–R.F.S.**

CREDIT: FROM ZAREMBINSKI *ET AL., PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES* 95, 15189 (DECEMBER 1998)

explains, has revealed the linear sequence of amino acids for thousands of proteins for which no function has yet been discovered. To determine their functions, researchers are now looking to structural genomics for help.

The idea is to take closeup photos of each protein and study them for clues to what that protein might be doing. By reversing the starting and ending points, the approach turns structural biology "on its head," says Mark Swindells, scientific director of Inpharmatica. This U.K.-based start-up is looking to provide pharmaceutical companies with computer predictions of important protein structures. It's akin to asking a car mechanic to troubleshoot an unfamiliar engine using engineering diagrams of key components: A component that looks like a fuel valve is not likely to work as a spark plug.

A handful of preliminary research projects are already beginning to show that 3D crystal structures can provide important insights into the role of proteins inside cells (see above). "There's a widespread feeling that if more structures were available, cell biology and medicine would move more rapidly," says Eisenberg.

Indeed, governments and industry are ramping up efforts to search for such structures en masse (see tables). Although most efforts include both experimental work and computer modeling techniques, it's the verifiable 3D maps that provide the most reliable information. This fall the U.S. National Institutes of Health (NIH) will announce up to six winners of a $20 million competition for structural genomics pilot centers. If fruitful, the investment is seen as a likely first step toward a larger program. Last month, the Ontario provincial government announced plans to spend $23 million over 3 years on a structural genomics center at the University of Toronto. And three more countries—Japan, Germany, and the United Kingdom—either have funding in place or are considering plans for similar investments.

For industry, structural genomics promises not only a wealth of new drug targets but also help in eliminating those not likely to be useful. The Genomics Institute of the Novartis Research Foundation in La Jolla, California, is keen to improve its high-throughput approach to x-ray crystallography, says director Peter Schultz. Researchers there are working to automate virtually every aspect of the process, from finding proteins to solving their structures. And a company spin-off, Syrrx, is looking to commercialize the approach.

Syrrx is one of a handful of biotech start-ups in the field. A few, including Structural Genomix in San Diego, intend to resolve the structures of thousands of unique proteins over the next several years, providing information that could be useful to clinical researchers in many fields. If they succeed, their efforts would roughly equal the number of protein structures resolved over the last 4 decades.

Hard choices

Even at this relatively blistering pace, however, 3D structure determination and related structural genomics projects remain too slow and costly to provide an atomic map for all human proteins, at least anytime soon. "Structural genomics is a lot harder than sequencing," says Terwilliger. Whereas the tools used in sequencing can decode virtually any gene, proteins are more finicky. They require more attention to find the right conditions under which they will be expressed, purified, and coaxed into forming a crystal.

That complexity also makes the work expensive—it costs roughly $100,000 to resolve a single protein structure. At that price, researchers may need to be content in the near term with structures for fewer than 10% of the human protein, forcing experts to choose from among their preferred targets. NIH officials say one desired outcome from the pilot centers is technology that would drastically lower the cost of resolving new structures.

So far there is no consensus in academia on which proteins to pursue. One camp, says Terwilliger, aims to resolve the structures of specific unknown proteins to gain clues about their biochemical function. A second emphasizes the need to survey general shapes of proteins by lumping those thought to be similar into families and obtaining the structure of at least one member of each. A third approach seeks structures for as many proteins of a given organism as possible, such as pathogenic bacteria. And a fourth—the approach most similar to conventional structural biology—homes in on proteins thought to be important for understanding diseases and basic biochemistry.

No matter what the strategy, one concern for academics is that few research groups have the expertise to combine all the pieces of a structural genomics effort. "A single investigator really can't do much on his or her own," says Emil Pie, a crystallographer at the University of Toronto. One result is that academic groups that produce proteins are teaming up with others that specialize in crystallography and bioinformatics. One early pilot project that's looking at proteins from a heat-loving archaeon, for example, involves 17 labs at seven institutions, including the University of California, Los Angeles, Los Alamos National Laboratory in New Mexico, and the University of Auckland in New Zealand.

The situation is a bit better in industry because it's often easier for companies to marshal both money and expertise to tackle specific problems. "We have the ability to build [high-throughput] tools that are difficult for the university sector to [build]," says Schultz of Novartis, who also retains an academic appointment at The Scripps Research Institute in La Jolla, California. Another reason, say Shultz and others, is that industrial researchers are more narrowly focused on short-term goals. "We're not looking to define protein fold space," says Tim Harris, president and CEO of Structural Genomix. "We will work on protein families we know are of interest to industry and [on those] we predict they will be interested in." One such example, says Harris, are protein kinases, which regulate information traffic inside cells and have been shown to be involved in a wide variety of diseases such as leukemia.

One big unknown is whether public and private efforts will dovetail or if, as happened with sequencing, the sectors will race each other to bank their data. Harris says he believes the focused industrial work "will complement what the public domain is doing"—namely, going after the big picture of how proteins fold. Others aren't so sure. "We may see a replay of the situation in genome sequencing," says Chris Sander of the Massachusetts Institute of Technology's Center for Genome Research and of Millennium Pharmaceuticals in Cambridge, Massachusetts.

In the genome arena, says Sander, industry has competed with the publicly funded genome project in a scramble for new medicines, diagnostics, and agricultural products. Although Sander believes that this duplication has accelerated the pace of genome research, Terwilliger sees a downside for structural genomics. If industry keeps half of the data secret, he says, "there will be an awful lot of waste and duplication." Asked whether academic groups will have access to his company's structural data, Harris is noncommittal: "We anticipate that at least some of the structures will be available to the public domain."

Next month NIH is sponsoring a meeting at the Wellcome Trust's Sanger Centre for genome sequencing near Cambridge, U.K., to coordinate international financing of structural genomics as well as to address issues surrounding intellectual property and access to data. If all goes as planned, the result will be an agreement akin to one that grew out of a 1994 meeting in Bermuda that helped promote the coordinated international effort to sequence the human genome, says John Norvell, who is heading up the structural genomics program at the National Institute of General Medical Sciences on the NIH campus in Bethesda, Maryland. But even such an agreement is unlikely to give researchers all the help they will need as they travel along the new road of structural genomics.

–ROBERT F. SERVICE

NEWS

Malaria Researchers Wait For Industry to Join Fight

A huge toll of illness and death, a dire need for new treatments, and rapid scientific progress are inspiring researchers to battle malaria. But most drug companies balk at investing in what they see as a Third World disease

By most indicators, the time is right for the pharmaceutical industry to make a major push against malaria. Existing treatments are failing in the face of rising resistance from both the parasite and its mosquito vectors. At the same time, the first results of a major effort to sequence the genome of the parasite have identified some chinks in its armor that could lead to promising new drugs. And there's no shortage of patients: Each year brings as many as a half-billion new cases, including more than 1 million deaths, most among children.

But there's a problem: The drug industry is not interested. Companies need to make money to satisfy shareholders, and as staggering as malaria's toll may be, medicines to combat or prevent malaria do not set an investor's heart beating faster. The obvious target populations live in sub-Saharan Africa and South Asia, where most patients can't afford expensive new drugs. Wealthy tourists from industrialized countries offer a market, but, at $100 million to $250 million annually, it is modest in pharmaceutical terms. "The commercial return is negligible," says Simon Campbell, a retired senior vice president and head of drug discovery at Pfizer Inc., "so it really is not cost efficient for pharmaceutical companies to undertake malaria research on their own." Another major factor keeping companies out of malaria research is insufficient patent protection in the developing world, says Bert Spilker, senior vice president of the Pharmaceutical Research and Manufacturers of America (PhRMA) in Washington, D.C. "It is difficult for companies when countries pirate drugs," he says.

Add to those obstacles the declining potency of current treatments and the paucity of drugs in the pipeline, some of which may not pan out, and the result is that some parts of the world may soon have nothing with which to counter the scourge. "It's a paradoxical situation," says Dyann Wirth, director of the Harvard Malaria Initiative. "It's a very exciting