

Research teams around the world are trying to speed up protein structure determination—
and running into many barriers

Tapping DNA for Structures Produces a Trickle

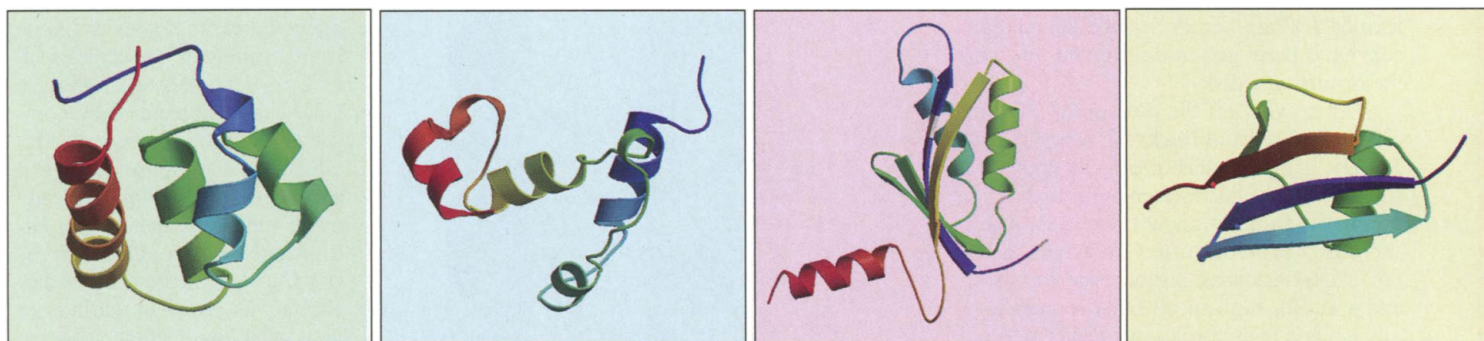
BERLIN, GERMANY—The two dozen labs that signed up for a venture called “structural genomics” several years ago had hoped to be pumping out a stream of results right about now. Their goal, set in 2000, was to follow the lead of the Human Genome Project, ramp up quickly, and have each lab solve hundreds of new protein structures per year. It was a bold idea, but no one knew whether it would be possible to automate the research to this degree. So when research teams met to compare notes here last

the first opportunity for the 30 or so publicly funded structural genomics projects around the world to compare their data in a public forum. So far, these projects have targeted more than 18,000 proteins but solved the structures of only about 200. Says Naomi Chayen, a protein crystallization expert at Imperial College, London: “The number of structures is disappointingly low.”

Although the numbers suggest that the latest field to embrace industrial-scale biology is floundering, the mood at the confer-

protein structures at high speed. The U.S. National Institutes of Health (NIH) followed suit in September 2000, funding seven “structural genomics” pilot projects around the country through NIGMS. Last year, NIGMS added two more centers. Similar pilot projects got started in France, Germany, Canada, the United Kingdom, and South Korea. More recently, Switzerland, China, and Finland have jumped in.

All of this adds up to a major new source of support for structural biology. Japan alone



Proteins on parade. An international collaboration drew on proteins from a variety of organisms to produce these structures.

month,* they were disappointed to learn that everyone was having plumbing troubles. Their pipelines have sprung leaks, and instead of delivering a flood of results, so far they're delivering just a trickle.

Consider the trials of Ian Wilson, a structural biologist at the Scripps Research Institute in La Jolla, California. Wilson heads one of nine U.S. pilot projects that are trying to dramatically speed up the three-dimensional atomic mapping of proteins. Wilson's team began 2 years ago by automating the numerous steps involved in crystallizing proteins and collecting the x-ray data needed to solve their structures. According to data presented at last month's meeting, Wilson's group has encountered difficulties at each stage of the process. Although the group started with more than 1870 protein targets, they've generated only 23 completed protein structures.

Wilson's group is not alone. Other groups presented similar results here in what was

ence was surprisingly upbeat. “We had to start somewhere,” Wilson says. And the high rate of attrition, he says, “is just what we expected.” Structural biologists working with one protein at a time have faced these grim attrition rates for years, Wilson adds. Compared to the traditional output rate, the new programs are doing very well.

Still, most experts acknowledge that the challenges are formidable. “We knew this would be hard, and it is,” says John Norvell, who coordinates the nine structural genomics pilot projects funded by the National Institute of General Medical Sciences (NIGMS) in Bethesda, Maryland. It is still an open question whether this research can be carried out in production-style facilities, but nobody is yet losing faith.

Picking up the pace

The stampede into structural genomics began 5 years ago when the Japanese government announced plans to build the RIKEN Genomic Sciences Center in Yokohama; it was designed among other things to turn out

has promised to spend \$100 million a year for the next 5 years in support of the RIKEN effort and eight new centers. NIGMS is chipping in another \$50 million a year for its centers, and the European Community has added about another \$4 million a year on top of the tens of millions spent by member countries. Over the next 5 years, governments will spend roughly one-quarter the cost of the entire Human Genome Project, all just to see whether large-scale protein mapping is feasible. But then again, the genome project itself was far from a sure bet in its first few years.

The cost is no surprise. Protein mappers have long known that speeding up structure analysis would be difficult. “This is much harder to do than sequencing DNA,” says Thomas Terwilliger, an x-ray crystallographer at the Los Alamos National Laboratory (LANL) in New Mexico. Like genes, proteins consist of a linear chain of building blocks. But each of these protein chains—composed of amino acids—folds into a complex 3D web, and the shape determines the chemical function. Because researchers can't reliably

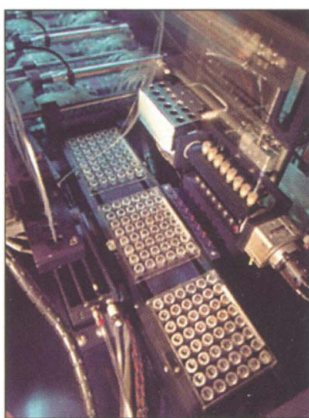
CREDITS: (ALL STRUCTURES) NORTHEAST STRUCTURAL GENOMICS CONSORTIUM

* International Conference on Structural Genomics, 10–13 October.

predict protein structure from the amino acid sequence, they must physically map each one. They do this by bombarding crystals of a protein with powerful x-ray beams or by bouncing radio-frequency signals off a solution of proteins in a nuclear magnetic resonance (NMR) machine, using the reflected energy to measure surfaces.

Powerful incentives are motivating this work. For basic researchers, mapping a protein can be the key to determining exactly how it functions. By scanning large numbers of structures, they hope to gain insights into how families of similar proteins evolved. Drug designers use protein maps to help tailor pharmaceuticals to block or enhance a protein's chemical activity.

But the new strategy is a departure from



Numbers game. Researchers hope that robots designed to handle protein crystallization and other processes will boost output.

a synchrotron to collect the x-ray data. NMR teams, meanwhile, must ensure that their proteins are stable for the extended period of time required for NMR scans.

The attrition is severe at each step. Take the numbers cited by the Northeast Structural Genomics Consortium (NESGC), a collaboration among eight institutions led by NMR expert Gaetano Montelione of Rutgers University in Piscataway, New Jersey. At the Berlin meeting, NESGC members reported that to date they have pursued 5187 DNA targets, cloned

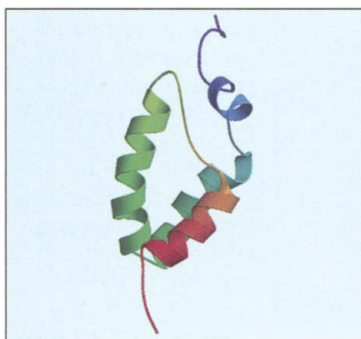
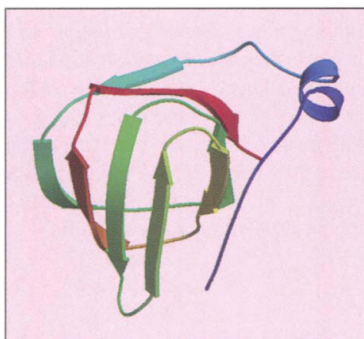
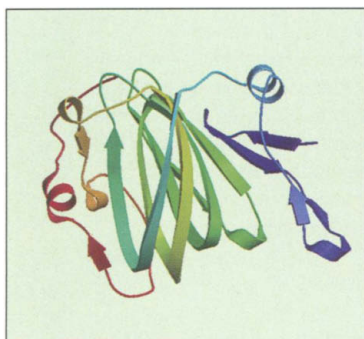
1675 of them, and expressed 1295 as proteins. But only 773 were soluble. Consortium members purified 719 of their proteins, but they crystallized only 94. So far they have determined 50 structures, 22 using

says Montelione. "What we are seeing is that high-throughput is not enough," Chayen says. "What we need is higher output."

Still, Montelione and most others at the meeting say that they expect output will improve rapidly as groups gain experience and bring new technology on line. The current numbers, NIGMS's Norvell says, "are only the initial look" and reflect the fact that most groups have only recently started to produce any structures at all. Just because proteins haven't yielded structures on the first try, "that doesn't mean they're out of the pipeline," says Norvell. Individual groups will likely take more than one crack at solving the difficult ones. William Studier, a biologist at Brookhaven National Laboratory in Upton, New York, agrees: "The phase at which the results are really going to come is still a year or two away."

Plugging holes

New robots and clever biological tricks might solve some of the problems researchers are now facing, according to sci-



tradition. The ability to sequence full genomes "has turned structural biology on its head," says Chris Sander, a computational biologist at the Memorial Sloan-Kettering Cancer Center in New York City. Instead of picking targets because they are biologically interesting, structural genomics researchers can now scan genome databases for stretches of DNA encoding genes of completely unknown function, hunt down their proteins, study the results, and perhaps discover entirely new realms of biology in the process.

Leaky pipes

Getting this new method to work is like mastering a frustrating computer game: Every time you reach the next level, a new challenge awaits you. Research teams must first engineer *Escherichia coli* and other organisms to produce the proteins they want to characterize. Then they have to purify them and see if they are soluble in water, an essential step for creating the crystals used in x-ray studies and the liquid solutions used in NMR research. X-ray teams then try to coax the proteins to form well-ordered crystals, which they take to

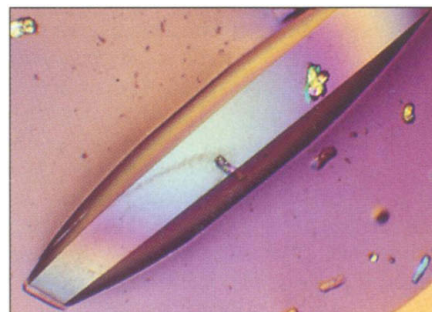
x-ray analysis and 28 using NMR. All the pilot projects report similar statistics. "Right now everything is a bottleneck," says Sander.

One of the biggest problems is a basic one: coaxing *E. coli* and other organisms to express the right proteins and getting them in a soluble form. Researchers are successful about half the time when they try to express bacterial proteins in bacterial vectors. They typically have a lower success rate (20% to 30%) getting bacteria to copy eukaryotic proteins, such as those from yeast or humans, in part because they often require "chaperone" molecules and other factors to encourage proper folding. Some groups are trying to express eukaryotic proteins in yeast and other eukaryotic organisms, but these organisms are widely viewed as finicky and tricky to handle. The result, says Montelione, is that "most of the structures we're seeing right now are bacterial proteins."

Researchers continue to have trouble getting proteins to form the crystals needed for x-ray studies, which produce the lion's share of 3D structures. "We take a big hit in crystallization and optimization of crystals,"

entists at the meeting. Chayen, for example, described a new scheme for growing protein crystals in oil that yields faster and higher quality results, increasing the odds of obtaining precise atomic maps.

At RIKEN, biochemist Shigeyuki Yokoyama is working with another promising technique, known as cell-free protein synthesis, that breaks open cells to isolate and capture the protein-producing ribosomes, eliminating most of the rest of the cellular machinery. Intact cells often cannot handle the large



Crystal medium. Using oil to contain crystals as they grow may remove a production barrier.

Big Biology Without the Big Commotion

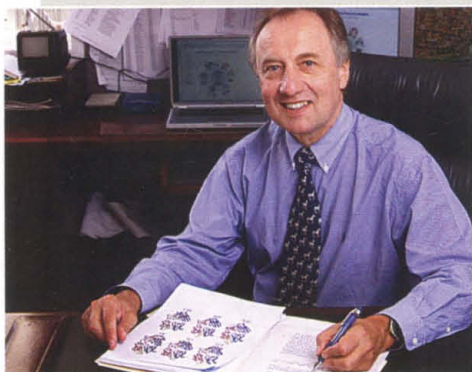
Like its forebear, the Human Genome Project, structural genomics is an exercise in industrial-scale biology. But in at least one respect the two projects seem different: To date, structural genomics has not experienced the noisy head-to-head competition between academic groups and companies that roiled the genome project. If anything, relations are downright cozy—to the extent that two structural genomics companies are even members of U.S. public projects and also plan to deposit some of their private results in a public database.

Companies were quick to gear up their own high-speed efforts to map protein structures. Private start-ups raised approximately \$500 million, an amount that approaches what governments around the world have promised through 2005. But the public and private efforts are largely pursuing separate goals. Whereas academic groups are looking to catalog the diversity of proteins in nature, companies are focusing on targets for new drugs. Says Stephen Burley, chief scientist of Structural GenomiX in San Diego, California: "One is going for breadth, the other is going for depth."

At a recent meeting in Berlin (see main text), Burley and Eric Adam of Syrrx, another San Diego-based structural genomics company, revealed a sampling of the companies' efforts and early results,

showing off their formidable analytical power. Syrrx, for example, turned out 56 structures of unique proteins in the last 8 months and has deliv-

Good bet. Studying new protein families may lead to major drug finds, Ian Wilson says.



quantities of protein that they're asked to produce for structural studies; by stripping away most of the cell, Yokoyama has solved this toxicity problem. His method also allows researchers to fiddle with the protein production mixture—adding chemical cofactors and other enzymes that help proteins fold properly—to boost the likelihood of producing working proteins. "It's very promising," says Montelione. "For the future, all groups will have to work on that technology."

Another key step in preparing material for imaging is to obtain properly folded proteins; Geoffrey Waldo of LANL reported a new way to do this. The technique links each gene for the protein of interest to the gene for a protein-based light emitter, called green fluorescent protein (GFP). The two proteins are linked in such a way that the proper folding of the first causes GFP to fold into a light-emitting shape. As a result, by simply shining light on a panel of growing cells, researchers can tell which proteins are likely expressed and folded correctly. Researchers can also use the approach to screen thousands of mutant versions of proteins to see if tinkering with wild-type proteins creates novel folding

ered a total of 80 structures since the company formed in 1999. Structural GenomiX has already banked more than 100 structures. Taken together, that's about the same number produced by the nearly 30 publicly financed programs worldwide and roughly 10 times the number produced by a typical major pharmaceutical firm in a year. Both Burley and Adam say their companies have yet to hit full speed.

Aside from their focus on putative drug targets, the primary difference in the company efforts is scale. According to Adam, who has recently left Syrrx, the company has already burned through about \$70 million on robotics and other automation technology. The company still has \$50 million in the bank, he says, meaning that its financial backing is roughly half what the U.S. National Institutes of Health plans to spend on its nine structural genomics centers combined over 5 years. It's still early days, but many academics are impressed by the fleet of automation tools that kind of money can buy. "I'm stunned by the technology development," says Wim Hol, a structural biologist at the University of Washington, Seattle.

Adam says that the structures Syrrx has produced so far are either protein kinases or proteases, popular drug targets. Some appear to play key roles in diabetes and breast, colon, prostate, and skin cancers, according to Adam, who adds that the company is working to develop drugs that inhibit them and then team up with major pharmaceutical companies to push the compounds through clinical trials. The goal, he says, is to move compounds into the clinic by 2004. Structural GenomiX has similar goals, Burley says, and it is also pursuing novel kinases that appear to be involved in cancer.

Generating lots of protein structures doesn't guarantee a company a new blockbuster drug, says Ian Wilson, a structural biologist at the Scripps Research Institute in La Jolla, California. However, he adds, the strategy is "a reasonable bet." Companies want to find new proteins because they hope this will give them a head start on designing new classes of drugs, which is often the key to turning out a blockbuster. It's a possibility that entices academics and investors alike

—R.F.S.

patterns that make it easier to conduct structural studies. "This could be a very attractive method to a lot of people here," says Udo Heinemann, who heads the Protein Structure Factory, Germany's primary structural genomics project, located in Berlin.

Engineering teams, meanwhile, are setting up new high-speed robots and computer software designed to produce, purify, and crystallize proteins. They also will be used to scan for the best quality crystals, collect x-ray and NMR data at high speeds, and turn those data into final 3D structures. Whether or not they achieve the production targets, says Montelione, "the new technology development will be very important and valuable for all of structural biology."

Questions of scale

Whether these promises will translate into a flood of new structures remains a big unknown. Structural genomics "hasn't proven itself yet," Sander says. And a funding crunch might be on the horizon.

When NIH launched its protein structure effort in September 2000, it originally set a goal of producing structures for 10,000 unique proteins in 10 years. That was a tall

order considering that only some 2000 such independent proteins had been mapped in the past 4 decades. The agency funded nine centers through 2005 as pilot projects to test out new high-throughput technologies. But NIGMS will soon confront some tough decisions, Norvell says, noting that "by the end of 5 years, we certainly won't be where we need to be." The common view is that 5 years from now, each center will likely be able to produce 100 to 200 protein structures a year, or about 1500 in total. As a result, the expectation all along, says Scripps's Wilson, has been that NIH would select a subset of centers to scale up. Norvell adds that over the next year an NIH advisory committee will begin trying to sort out the best way to proceed after 2005. The answer, he says, likely won't come until early 2004, about the same time the current pilot efforts should be hitting their stride.

Similar projects around the globe also face uncertainty. Science budgets are particularly tight in France and Germany at the moment. For many of these programs, it seems, the question is whether the payoff will come in time to convince funding agencies to stick with structural genomics' hefty price tag.

—ROBERT F. SERVICE

CREDIT: ALAN MCPHEE/BIONMEDICAL GRAPHICS/THE SCRIPPS RESEARCH INSTITUTE