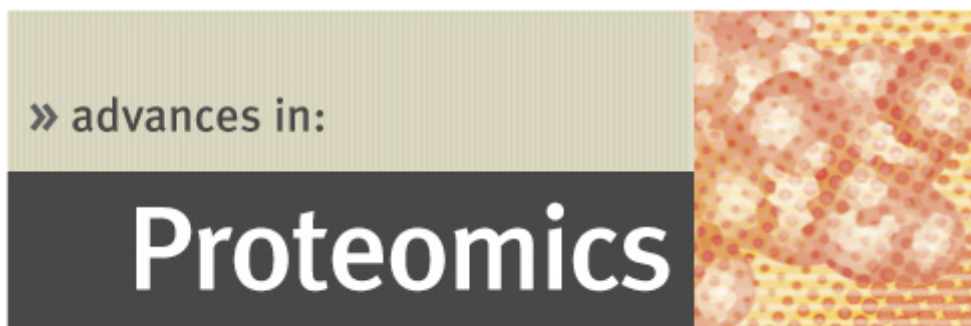


Advances In : Proteomics – Puzzling Out Proteins' Structures



To determine the structures of large numbers of proteins, researchers rely on updated versions of well-tried experimental techniques. But the new approach of computer modeling has started to gather momentum.

by Peter Gwynne and Gary Heebner

CONTENTS

- [Protein expression](#)
- [Protein crystallization](#)
- [X-ray crystallography](#)
- [NMR spectroscopy](#)
- [Software for structure determination](#)
- [Computer modeling](#)

This is the third of four special supplements this year on Advances in Proteomics. The first two appeared in the 25 March and 29 April issues of Science and the final one will appear in the 16 September issue.

Inclusion of companies in this article does not indicate endorsement by either AAAS or Science, nor is it meant to imply that their products or services are superior to those of other companies.

Proteins have complex, three-dimensional structures whose dynamic twists and turns ultimately determine their function. They exist in huge numbers; each human cell typically contains 100,000 or more proteins. Not surprisingly, then, characterizing the

ADVERTISERS

Biacore AB
systems for protein interaction analysis in basic research, drug discovery and development
+46 18 675700
<http://www.biacore.com>

Elchrom Scientific AG
separation and purification products for genomics and proteomics research
+41 41 747 25 50
<http://www.elchrom.com>

Fuji Photo Film Co., Ltd. – Life Science Products Division
QuickGene-810 nucleic acid isolation system, and BAS, LAS

structures of all the proteins in existence demands an eclectic mix of tools and technologies. Researchers who set out to determine proteins' structures and functions rely on techniques that include protein expression, protein crystallization followed by X-ray crystallography, and nuclear magnetic resonance (NMR) spectroscopy. Making sense of the readings that those methods produce demands automated techniques and computer analysis.

Such structural studies have value beyond the collection of fundamental knowledge. The knowledge of accurate molecular structures is a prerequisite for rational drug design and for structure-based functional studies to advance the development of effective therapeutic agents and drugs.

Proteins differ widely in the degree of difficulty that scientists encounter when they try determine their structures. "Assuming that the protein is of a suitable size and well behaved, it's quite straightforward," says Douglas Meinhart, product manager, NMR product lines and manager of the analytical laboratory at **JEOL USA**. "But the interesting ones are not well behaved." Determining the structures of badly behaved proteins "is still difficult to do," adds Carsten Mang, product manager for proteomics at **Hamilton Life Science Robotics**. "It's more an art than a science. But the tools are getting better."

The basic approaches to determining proteins' structures are familiar to bench scientists. "The technologies have been around for many years," says Nigel Darby, vice president of R&D for protein separations at **GE Healthcare**. "They have been applied to small molecules. But proteins are bigger and orders of magnitude more complex than small molecules." To permit them to tackle proteins, vendors have had to upgrade those methods. "The improvements in tools and technologies for protein structure determination have been absolutely essential for the ability of the structural biology community to produce protein structures," says Lance Stewart, vice president of **deCODE biostructures**.

Data Driven Research

The large numbers of samples and conditions involved in these experiments require not only automated techniques but also powerful computers and software programs to make sense of all the information they create. After all, this is a data driven – rather than a hypothesis driven – field of research. "The data flood is incredible," declares Rudy Potenzzone, senior vice president for product management at **Ingenuity Systems**. "Without software," adds Thomas Billert, CEO of **Jena Bioscience**, "you could not solve a protein structure."

Experimental means of determining proteins' structures rely on diffraction techniques. "X-ray crystallography and NMR are the two main methods," Meinhart says. "You'll also run across people doing it with neutrons and even electron beams." X-ray crystallography, which accounts for about 85 percent of known protein structures, "is the only method that can rapidly produce high resolution structures," Stewart adds. NMR, however, has the advantage of investigating proteins in their natural environments, in solution, rather than in the solid state. "It also gives you the ability to do kinetics and look at dynamics," Meinhart continues.

A significant new approach promises to complement those techniques. "The whole field of protein structure determination has been entirely experimental up to now," explains Ying Xu, professor of biochemistry and molecular biology at the **University of Georgia**. "Until very recently the cost per structure determination was over \$100,000. Even a recent reduction to \$50,000 to \$70,000 means that it's still very expensive. So we need another way." That way is

and FLA imaging systems
+81 3 3406 2201
<http://lifescience.fujifilm.com>

Leica Microsystems AG
[Germany]
instruments and systems for
imaging analysis, digital cameras
+49 6441 290
<http://www.leica-microsystems.com>

Leica Microsystems [USA]
847-405-0123

Syngene [Europe]
digital imaging and analysis
products for 2-D gel electrophoresis
+44 (0)1223 727123
<http://www.2dymension.com>

Syngene [USA]
301-662-2863

Takara Bio, Inc.
kits and reagents for molecular
biology research, including
genomics and proteomics
+81 77 543 7247
<http://www.takara-bio.com>

computer modeling. “We’re working on computational approaches to solving proteins’ structures that complement experimental techniques,” Xu states.

Exemplifying the emergence of computer modeling, the **National Institutes of Health** (NIH) recently called for proposals for technology to develop high-accuracy tools for modeling proteins’ structures. The effort will join the NIH-funded Protein Structure Initiative, which has sparked the experimental determination of multiple proteins’ structures over the past half decade.

Purifying Proteins

For X-ray crystallography, the first step involves obtaining enough purified protein to allow analytical methods to be applied. That isn’t always easy. Scientists need a method that will produce a protein with the correct functional properties as well as one that they can crystallize successfully. Sometimes they need to test several expression constructs and expression systems before they find the combination that yields enough protein to move to the next phase: growing a crystal.

The situation has improved markedly in recent years. “The whole area of protein expression and purification, which had been a bottleneck, has undergone tremendous advances in the past 10 to 15 years,” says GE’s Darby. Several companies, including **Invitrogen**, **MP Biomedicals**, and **Sigma-Aldrich**, have developed novel protein expression and purification systems based on inserting small peptide sequences into a specific protein that can be identified with a molecule that recognizes the tag.

Before they can analyze a protein sample, scientists must purify it to remove contaminating substances. Desalting columns that use gel filtration methods and dialysis, from such companies as Sigma-Aldrich and **Stratagene**, provide effective tools to clean up protein samples.

GE Healthcare’s ÄKTExpress automated, fully integrated chromatography system provides parallel, multidimensional, high throughput purification of tagged proteins. By providing up to 50mg of proteins of very high purity, ÄKTExpress maximizes the chances of successful crystallization. Its high throughput permits researchers to purify a large number of target proteins in a short time. The system integrates hardware, software, and chromatography media in a user-friendly automated system. “The whole idea is to make a formidable operation – purifying proteins in parallel via multiple steps – a plug and play operation,” Darby says. “We’ve tried to make the technology as straightforward as possible, so that someone can go into the lab and purify large numbers of proteins in large quantities without being a real expert.”

An Esoteric Bottleneck

As its name indicates, X-ray crystallography requires crystals – preferably extremely pure – of the entity under examination. “It’s still the main bottleneck – the most esoteric part of the whole process,” says Billert of Jena Bioscience. “You need to screen a very large number of conditions before you end up with microcrystals of an unknown protein. This may take just a couple of days, but if you’re unlucky it may take years.”

A good crystal usually has an extremely regular crystalline structure (referred to as low mosaicity) and a diameter of the order of several hundreds of microns. Crystallographers can deal with smaller crystals (down toward 50 microns diameter), but they need higher level sources of X-rays, such as synchrotrons, for structural analysis.

Proteins present crystallographers with a unique problem. Their molecules are large, somewhat unstable, and likely to possess several folds that complicate their development into crystals. In hopes of finding successful conditions for growing crystals with the quality essential for X-ray diffraction, researchers typically set up several multiwell plates with different conditions in each well. They can use any of several methods, including the sitting drop vapor diffusion and hanging drop vapor diffusion methods, to grow the crystals. Weeks or months after setting up the plates, the researcher may find that crystallization has occurred. If the resulting crystals are

not of diffraction grade, the team must then undertake a series of optimization steps to improve their quality.

Companies such as **Emerald Biosystems**, **Hampton Research**, and Jena Bioscience offer supplies for screening and optimizing the crystallization conditions for proteins. “We are trying to become a one-stop shop for crystallography,” says Billert. “We offer screens, kits for crystal optimization and protein surface engineering, crystallization plates, and much more.” Several of the company’s products are also designed for high throughput work. And like other vendors in this field, Jena emphasizes user-friendliness. “The screens we offer are based on database research, to find the conditions that work best,” Billert continues. “We try to offer intelligence to make life easier for the scientist.”

Workstations and Robots

Another group of suppliers presents a different way to simplify protein scientists’ lives. Several companies have developed automated workstations and robots to help with the repetitive work required to set up large numbers of sample replicates in varying conditions. Manufacturers such as **Apogent** (now a part of Fisher Scientific), **Caliper**, Emerald Biosystems, and **PerkinElmer Life and Analytical Sciences** offer a wide range of systems that scientists can adapt to routine laboratory procedures and studies of protein crystallization.

Hamilton Life Science Robotics has designed its MICROLAB STAR automated workstation specifically for protein crystallization. It helps to meet researchers’ needs for high throughput screening of various conditions for protein crystallization, and can provide large, high-quality crystals suitable for X-ray diffraction. “It can handle all types of liquids, polar and nonpolar,” Mang says. “Our robots work like handheld pipettes, with which scientists are familiar in the lab. There are no syringes, no pumps, and no tubing. All problems related to dilution of samples with system liquids are gone. This technology also allows us to monitor each pipetting step, therefore achieving previously unknown process control.”

The workstation features accurate pipetting and positioning of its liquid handling apparatus. The patented CO-REtip attachment technology transfers this precision right into the pipetting tip. That permits scientists to use it with very small (1,536-well) microwell plates. The system’s software checks for sufficient plates and reagents before beginning its routine to ensure that it can perform the desired functions without intervention from users.

Actual Atomic Arrangements

X-ray crystallography uses X-rays to determine the precise arrangements of atoms in crystalline specimens. Crystallographic systems generally include dedicated computers with associated hardware and software for instrument control, data reduction, solution and refinement of molecular structures, and the display and plotting of final results.

The approach can provide reliable answers to several structure related questions, from the nature of protein folds to the atomic details of bonding. The method has no limit on the size of the molecule or complex under study. “And once a reasonable crystal form has been worked out for the protein of interest, then the determination of protein-drug complexes can be worked out almost as fast as the diffraction data can be collected,” deCODE biostructures’ Stewart says.

Thermo Electron (formerly Applied Research Laboratories) offers a line of X-ray instruments for seamless integration into laboratory and total automation solutions. And several companies, including **Cengent Therapeutics** and **Rigaku**, grow crystals as a service for researchers. These companies take a purified protein and deliver a diffraction quality crystal or even a solved structure.

deCODE biostructures provides contract research services related to the use of X-ray crystallography to determine the structures of protein-ligand complexes. “The gene-to-structure collaborative services we offer are of great benefit to startup or young organizations that want to apply structure-based drug designs in their product development plans but would rather not

outlay the capital investments needed to install the capability for determining crystal development structure,” Stewart says.

The company's capabilities include gene synthesis, construct engineering, large scale protein production and purification, high throughput crystallization screening, small molecule pre-formulation screening, X-ray diffraction data collection, model building, and refinement, using laboratory automation and software developed through its own research efforts and available commercially for industry use. “I believe we have the most sophisticated set of leading-edge tools that you can find,” Stewart declares. “We built these tools ourselves, which is unprecedented for the industry.”

Chemical Shifts

NMR, discovered in 1945, uses radio waves to induce transitions between energy levels in atomic nuclei. The frequency of the radio waves is equivalent to the energy difference between the two levels. The field has experienced spectacular and continuing growth since 1960 with the development of pulsed Fourier-transform NMR and multidimensional NMR spectroscopy.

The method's power stems from the fact that the nuclear magnetic resonance frequency of most molecules changes very subtly as their chemical environment changes. Scientists can combine tables of these “chemical shifts” with other NMR information such as peak integration and coupling constants to solve complete three-dimensional structures of molecules in solution without using X-ray crystallography. They can also use NMR to study time dependent phenomena such as intramolecular dynamics in macromolecules, reaction kinetics, and protein folding.

The method is inevitably information intensive. “You have to interpret the NMR spectra and perform a guess of the family structure to which the protein belongs and feed those coordinates into molecular modeling that runs them around and aims to minimize the energy and compare them with the NMR data,” JEOL's Meinhart explains. “That can all be done in computers, but it still takes a matter of weeks or months computer time.”

Technical advances have increased the sensitivity and resolution of NMR. Progress in the theoretical and practical capabilities of the technology has led to an increasingly efficient use of the information content of NMR spectra. Coupled with developments in such molecular biology methods as recombinant protein expression, the method has become a valuable tool in protein analysis. Companies that offer NMR systems include **Bruker AXS** and **Varian**, in addition to JEOL. “We supply our ECA instrument at various field strengths for protein structural work,” Meinhart says. “It can accommodate up to six radio frequency channels; for protein work you typically need four. We also use our Delta NMR processing software – a full featured package for multidimensional data. You can walk in with your sample, put it in the magnet, get a full suite of calibrations, and go into the protein data collection automatically.”

From Databanks to Computer Models

The structures of solved proteins are stored in the Protein Databank, a database that researchers can access for comparison against new, unknown proteins. Structural biologists use software to determine structures and to understand the relationship between structure and function as they analyze and modify proteins and peptides. **Bio-Rad Laboratories**, **Crystal Impact**, and **Emerald Biosystems** are among the vendors of specialized software packages for data management and analysis of protein structures. Other information technology companies, including **Accelrys** and **MDL Information Systems**, offer wide ranges of software applications for use in determining proteins' structures and discovering new drugs.

Ingenuity Systems applies two products, its Pathways Analysis solution and its Pathways Knowledge Base, to work on protein structures. “We come in at the functional studies stage,” Potenzzone says. “Once you've determined what protein is relevant, we capture information about what it's doing, where it's located, what its binding modes are, and what kind of compounds interact with it. In effect, this is the next stage of determination beyond the structure. If you have similarly acting proteins, we might be able to give you some clues as to

where else to look. It's like a dynamic database.”

Databases and information technology have also smoothed the way to the next major advance in determining proteins' structures: computer modeling. The University of Georgia's Xu explains the basis of the approach. “People have found that in nature the number of unique protein folds is quite small – perhaps a few thousand fold families,” he says. “So if you solve one structure in a fold family, you can model the structures of the others in the family, using the solved structure as a structural template. People have found that about 90 percent of the structures solved in the past three years fall into fold families at least one of whose members have had their structures previously solved.”

That observed fact lays the foundation for computer modeling. “After you get enough representatives from fold families, the rest of the structures in the families will be able to be modeled by computers,” Xu continues. “With modeling, it becomes a matter of identifying the fold family a protein belongs to and then aligning the amino acid sequence alignment of a target protein with the solved structure from the same family to give you a placement of the amino acids in the structure template. From that point on you can refine the structure using the more classical energy minimization techniques.”

Only a Start

The process of modeling proteins' structures has only just started, and it still relies heavily on experimental structure determination. “Theoretically, once you know the structure of a member of your folding family you can model other family members' structures,” Xu notes. “But the larger number of accurate structures you can get for a family, the better will be the accuracy of your modeling prediction.” Computational structure predictions still have a long way to go to reach the accuracy level of experimental structure solutions such as X-ray crystallography and NMR. “We are probably still years away from consistently reaching the accuracy level of 1.5 Å obtained by experimental structures,” Xu estimates. Nevertheless, he adds, “The economic argument is that this method has to work because it costs too much per structure to do it experimentally. The scientific argument is that it will work.”



Peter Gwynne (pgwynne767@aol.com) is a freelance science writer based on Cape Cod, Massachusetts, U.S.A. Gary Heebner (gheebner@cell-associates.com) is a marketing consultant with Cell Associates in St. Louis, Missouri, U.S.A.

WEBLINKS

ADVERTISERS

Biacore AB

systems for protein interaction analysis in basic research, drug discovery and development
+46 18 675700
<http://www.biacore.com>

Elchrom Scientific AG

separation and purification products for genomics and proteomics research
+41 41 747 25 50
<http://www.elchrom.com>

FEATURED COMPANIES

Accelrys - a subsidiary of

Pharmacopeia
scientific software
<http://www.accelrys.com>

Apogent Technologies, Inc. (a Fisher Scientific Company)

automated workstations
<http://www.apogent.com>

Bio-Rad Laboratories

software for protein structure analysis

Fuji Photo Film Co., Ltd. – Life Science Products Division
QuickGene-810 nucleic acid isolation system, and BAS, LAS and FLA imaging systems
+81 3 3406 2201
<http://lifescience.fujifilm.com>

Leica Microsystems AG [Germany]
instruments and systems for imaging analysis, digital cameras
+49 6441 290
<http://www.leica-microsystems.com>

Leica Microsystems [USA]
847-405-0123

Syngene [Europe]
digital imaging and analysis products for 2-D gel electrophoresis
+44 (0)1223 727123
<http://www.2dymension.com>

Syngene [USA]
301-662-2863

Takara Bio, Inc.
kits and reagents for molecular biology research, including genomics and proteomics
+81 77 543 7247
<http://www.takara-bio.com>

<http://www.bio-rad.com>

Bruker AXS
<http://www.bruker-axs.de>

Caliper Life Sciences
automated workstations
<http://www.caliperls.com>

Cengent Therapeutics, Inc.
crystallographic services
<http://www.cengent.com>

Crystal Impact
software for protein structure analysis
<http://www.crystalimpact.com>

deCODE biostructures (a subsidiary of deCODE Genetics)
proteomics services - protein structure determination
<http://www.decodebiostructures.com>

Emerald Biosystems, Inc.
protein crystallization kits and reagents
<http://www.emeraldbiostructures.com>

GE Healthcare
protein separation and purification systems
<http://www.gehealthcare.com>

Hamilton Life Science Robotics
laboratory automation systems
<http://www.hamiltonrobotics.com>

Hampton Research Corporation
protein crystallization kits and reagents
<http://www.hamptonresearch.com>

Ingenuity Systems
scientific software
<http://www.ingenuity.com>

Invitrogen Corporation
protein expression systems
760-603-7200
<http://www.invitrogen.com>

Jena Bioscience GmbH
protein crystallization kits and reagents
<http://www.jenabioscience.com>

JEOL-USA, Inc.
NMR spectrometers
<http://www.jeol.com>

MDL Information Systems, Inc.
scientific software
<http://www.mdl.com>

MP Biomedicals
protein expression systems
<http://www.mpbio.com>

National Institutes of Health
government research organization
<http://www.nih.gov>

PerkinElmer Life and Analytical Sciences
automated workstations
<http://las.perkinelmer.com>

Rigaku/MSC
X-ray systems for protein crystallography
<http://www.rigakumsc.com>

Sigma-Aldrich Corporation
protein expression systems
<http://www.sigma-aldrich.com>

Stratagene
protein purification kits and reagents
<http://www.stratagene.com>

Thermo Electron Corporation
X-ray systems for protein crystallography
<http://www.thermo.com>

University of Georgia
university
<http://www.uga.edu>

Varian, Inc.
NMR spectrometers
<http://www.varianinc.com>

Note: Readers can find out more about the companies and organizations listed by accessing their sites on the World Wide Web (WWW). If the listed organization does not have a site on the WWW or if it is under construction, we have substituted its main telephone number. Every effort has been made to ensure the accuracy of this information. Inclusion of companies in this article does not indicate endorsement by either AAAS or *Science*, nor is it meant to imply that their products or services are superior to those of other companies.

**This article was published
as a special advertising section**

in the 27 May 2005 issue of *Science*

**GetInfo – Improved online
reader service!**

Search more easily for *Science* advertisers
and their products. Do all your product research at -
science.labvelocity.com

Visit <http://www.science-benchtop.org>
to find this article as well as past special advertising sections.

▲ PAGE TOP

Copyright © 2005 by The American Association for the Advancement of
Science. All rights reserved.