

STRUCTURAL PROTEOMICS: THE RELENTLESS PURSUIT OF PROTEIN SHAPE

In an editorial in the *Journal of Proteome Research*, structural biologist Raymond Stevens and proteomicist John Yates lament the confusion of terms surrounding their work. “Structural Genomics,” they say, is restricted mostly to the United States. Europe and other countries use “Structural Proteomics” to acknowledge a broader functional perspective to their work. Whatever you call it, structural ‘omics efforts have had an indelible impact, and not just on structure databases. Their technological output has altered the way structural biology specifically—and protein chemistry in general—is done.
By Jeffrey M. Perkel

To gauge the technological impact of structural proteomics efforts, take a look at the Protein Structure Initiative (PSI) Structural Genomics Knowledge Base (SG-KB) Technology Portal (cci.lbl.gov/kb-tech). Listed are more than 100 different innovations, from reagents and robotics to informatics and data management solutions—everything needed to take the traditionally laborious process of structural biology, whether using X-ray crystallography or nuclear magnetic resonance (NMR), and make it fast, efficient, and high throughput.

“In terms of technology, there are, for the first time, a whole group of very talented people coming together to optimize structure-determination pipelines as part of their goal,” says SG-KB director Helen Berman, Board of Governors Professor of Chemistry and Chemical Biology at **Rutgers University**.

Take the work of just one so-called PSI phase 2 (PSI-2) specialized center, the **Center for High-Throughput Structural Biology (CHTSB)**. (PSI is a 10-year, NIH-funded effort to determine the structure of all protein folds. Phase 2 comprises four large-scale structure-determination factories and six specialized centers dedicated to more vexing problems like membrane proteins, mammalian proteins, and protein-protein complexes.)

CHTSB includes researchers at six institutes in New York, California, and Canada. Their overall focus, says co-investigator Michael Malkowski, senior research scientist at the **Hauptman-Woodward Medical Research Institute (HWI)** in Buffalo, New York, is “technology development with respect to sample preparation from protein production through the diffraction experiment.”

These teams are attacking that pipeline at multiple points: coexpressing and crystallizing protein-protein complexes; preparing membrane proteins in yeast; developing crystallization additives; cryoprotecting crystals against intense X-ray beams; growing crystals inside capillaries to ease manipulation; and developing tools for crystallization screening, optimization, and image analysis.

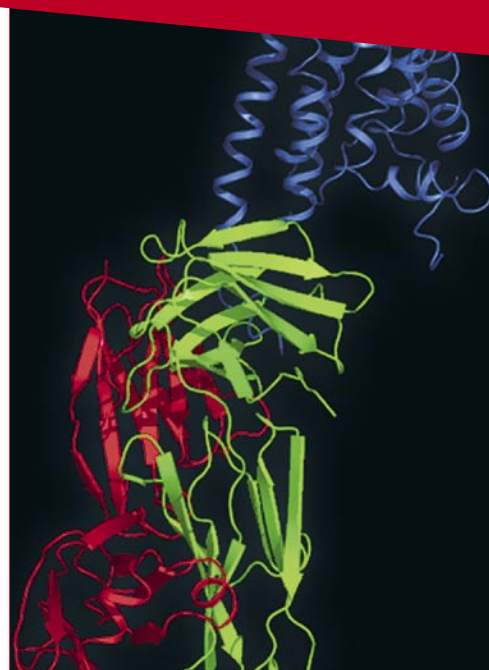
“Each is a piece of a puzzle,” says Malkowski, “and when you put them together you form an automated system that identifies initial crystallization leads, optimizes them, grows production-sized crystals in a capillary, freezes them, mounts them, and collects data.” He adds, “We don’t have that yet, but that’s what we are building toward.”

Putting Technologies through Their Paces

Structural genomics marries the high throughput sensibilities of genomics with the more contemplative process of structural biology. Structural proteomics, depending on whom you ask, is either the same process, or an offshoot that focuses on function. The PSI has largely fixated on maximizing “fold space” coverage without regard to biological significance. International efforts have eschewed that approach, elevating biology over coverage. Whatever the focus, structural biology pipelines initially constricted every step of the process.

“We needed better strategies at all steps—cloning, expression, purification, crystallization, and structure determination,” says Aled Edwards, director of the **Structural Genomics Consortium (SGC)**, an international effort with labs in Canada, the UK, and Sweden. “Every process needed improvement.”

Because each protein is “idiosyncratic,” Edwards says, what works for one may not work for another. Simply relying on anecdotal advice from the biologist down the [continued](#) >



“In terms of technology, there are, for the first time, a whole group of very talented people coming together to optimize structure-determination pipelines as part of their goal.”



High-field NMR systems at the National Magnetic Resonance Facility

Look for these Upcoming Articles

Genomics 2 — September 19

Lab on a Chip/Microfluids — November 7

Cell Signaling — December 5

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.

Proteomics

“It would take a long time to put just one sample in front of the beam. If you can put the samples into a container and have a robot that can do that for you, you can speed it up and have less error.”



hall is unlikely to work. “There are thousands of new technologies,” adds Edwards. “Most will be doorstops in five years. But of all these many, many, many technologies, which are actually useful?”

The solution, he says, is to put these technologies through their paces in a controlled fashion. “That’s what the field is doing,” he says, “looking at the panoply of technologies, seeing which work, so you don’t end up chasing your tail.” Importantly, the community archives that knowledge in searchable public databases like TargetDB and PepecDB, which track not only what works, but also what doesn’t.

Earlier this year members of the SGC (including Edwards) and 13 other structural genomics facilities around the world published their collective wisdom in a paper entitled “Protein production and purification” (*Nat Methods*, 5:147-53, 2008). Andrzej Joachimiak, director of the **Midwest Center for Structural Genomics (MCSG)**, calls the paper “a Maniatis manual for protein purification.” Edwards, who was corresponding author, says the report represents “the final UN-negotiated solution,” a probability-based consensus protocol for all protein jocks, not just structural biologists.

From Bad Media, a Breakthrough

The protocol contains such suggestions as developing multiple distinct constructs for each target; using ligation-independent cloning (commercialized as **Clontech’s** In-Fusion 2.0 PCR Cloning kits); and the unified use of expression tags for easy purification with robots such as **GE Healthcare’s** AKTA-express.

Another tip: try F. William Studier’s autoinduction media (available from **EMD Biosciences** as Novagen Overnight Express.). “It is a technique everyone should be using,” says Edwards.

Recombinant protein expression in bacteria is commonly controlled via the lactose control circuit. Place a *lac* operator in front of a gene of interest, and the gene will remain silent until you add the allolactose analog, IPTG. But IPTG must be added at just the right moment, and different cultures grow at different rates. “How do you get all the cultures growing at the same rate and at the same time to get optimal expression of these proteins in parallel?” asks Studier, of the **Brookhaven National Laboratory**.

In trying to solve that problem, Studier observed that some of his cultures induced on their own, without IPTG. He discovered that one of his media components was contaminated with trace lactose. Making use of this observation, Studier first designed a rich formulation in which cultures could grow to high optical density (OD) but never induce (noninducing medium). He then developed a precise blend of glucose, lactose, and glycerol in which the cells autoinduce upon hitting log phase. And they would do so while cranking out protein at higher levels than usual, because the cells can grow to much greater densities—up to OD 20, compared to OD 3 in Luria broth. “So all you have to do is grow 96 cultures in noninducing medium, then inoculate those into autoinducing medium, grow overnight, and collect induced cultures in the morning,” he says.

Biophysical Considerations

It’s not enough to express a protein; it also must be soluble, and

properly folded. At the **Joint Center for Structural Genomics (JCSG)** researchers employ a battery of methods to ensure desirable biophysical properties are met, says director Ian Wilson. Size-exclusion chromatography indicates whether the protein is soluble or aggregating. SDS-PAGE reveals protein purity. And deuterium-exchange mass spectrometry highlights disordered regions that could cause trouble downstream. “You label for a few seconds with deuterium, quench at low pH, digest with protease, and look for which regions are disordered by how much it is exchanging with the deuterium in the solution,” Wilson explains. “Only the disordered regions incorporate deuterium, because it is so fast.”

Such data can be invaluable. Brian Kobilka of **Stanford University**, who solved the structure of the beta-2 adrenergic receptor, found (via separate methods) that a loop between transmembrane domains 5 and 6 was “floppy,” inhibiting crystallization. His team overcame that via parallel approaches, one involving an antibody against the C-termini of these two helices, and the other, with Ray Stevens of the **Scripps Research Institute**, replacing that loop with a stable T4 lysozyme domain.

Others optimize structural work by tweaking protein length or surface chemistry. Researchers at the **European Molecular Biology Laboratory (EMBL)** outstation in Grenoble, France, site of the European Synchrotron Radiation Facility (ESRF), use a biotinylatable tag in their library construction and screening procedures to find soluble domains, says facility director Stephen Cusack.

Researchers at the MCSG, JCSG, and SGC improve crystal formation using reductive methylation and limited proteolysis. The former alters protein surface chemistry to allow different crystal packing and improve diffraction properties, while the latter removes disordered regions in situ. Both have been tested on large numbers of proteins, says Joachimiak—400 in the case of reductive methylation, yielding 30 structures that were otherwise unattainable.

Accelerating NMR

Structure determination by NMR has also benefited from ‘omics efforts. HIFI-NMR, a “reduced dimensionality” approach developed at the PSI-2 **Center for Eukaryotic Structural Genomics (CESG)**, reduces the time required to acquire spectral assignment data by a factor of 10, says principal investigator John Markley of the **University of Wisconsin, Madison**. Instead of blindly collecting data by standard methods, this approach leverages prior information at each step to get the job done in the most efficient manner.

Using NMR chemical shift data from the **Northeast Structural Genomics Consortium**, Yang Shen and Ad Bax from the **National Institutes of Health**, with Oliver Lange and David Baker at the **University of Washington**, validated a novel computational method (CS-ROSETTA) that predicts protein structure from chemical shift data. Their new approach thereby eliminates the need for both side-chain assignments and one of NMR’s lengthiest steps, nuclear Overhauser effect experiments.

“There are algorithms available to predict chemical shifts from structure,” says Markley. “[Bax] is using [CS-ROSETTA] to go the other way—from the chemical shifts you get a starting approximation of what the 3D conformation of the protein is.” Then, using Baker’s ROSETTA energetics algorithm, the software computes a likely final structure. In March, Bax and Baker applied CS-ROSETTA to 25 protein sequences; they estimated it can cut the structure-generation time by 50 percent.

Robotics

As with any ‘omics enterprise, automation and robotics are integral to structural proteomics. HWI researchers have developed robotics

Featured Participants

Berkeley Center for Structural Biology
bcsb.als.lbl.gov

Brookhaven National Laboratory
www.bnl.gov

CellFree Sciences
www.cfsciences.com

Center for Eukaryotic Structural Genomics
www.uwstructuralgenomics.org

Center for High-Throughput Structural Biology
www.chtsb.org

Clontech
www.clontech.com

Columbia University
www.columbia.edu

EMBL
www.embl.org

EMD Biosciences
www.emdbiosciences.com

Fluidigm
www.fluidigm.com

GE Healthcare
www.gehealthcare.com

Hauptman-Woodward Medical Research Institute
www.hwi.buffalo.edu

Israel Structural Proteomics Center
www.weizmann.ac.il/ISPC

Joint Center for Structural Genomics
www.jcsg.org

Midwest Center for Structural Genomics
www.mcsg.anl.gov

National Institutes of Health
www.nih.gov

New York Consortium on Membrane Protein Structure
www.nycomps.org

New York SGX Research Center for Structural Genomics
www.nysgxrc.org

Northeast Structural Genomics Consortium
www.nesg.org

Oxford Protein Production Facility
www.oppf.ox.ac.uk

Rigaku
www.rigaku.com

Rutgers University
www.rutgers.edu

Scripps Research Institute
www.scripps.edu

Stanford University Research Institute
www.stanford.edu

Structural Genomics Consortium
www.thesgconline.orr

TTP LabTech
www.ttplabtech.com

University of Limerick
www.ul.ie

University of Washington
www.washington.edu

University of Wisconsin, Madison
www.wisc.edu

to automate crystallization trials. “We can do 1,536 conditions in about 20 minutes in a single plate,” Malkowski says. “And then we image these samples every week for a month.”

Such automation means researchers can do more with less, and more accurately. “When I was a postdoc, I would set up crystallization droplets of 20 to 30 μ l. Now we use 200 to 500 nl,” says Joachimiak.

Crystallization robots are available from **Rigaku** and **Fluidigm**, among others. Joachimiak uses **TTP LabTech’s** mosquito. So does Joel Sussman, director of the **Israel Structural Proteomics Center** and coeditor of *Structural Proteomics and Its Impact on the Life Sciences* (World Scientific, 2008). In fact Sussman’s lab is extensively automated, with robots for cloning, protein purification, and crystal visualization. At ESRF, where Sussman’s team runs diffraction experiments, robots even mount their crystals into the beamline.

According to Paul Adams, head of the **Berkeley Center for Structural Biology**, automounting robots are especially useful for screening to see which crystals diffract best. Traditionally, “It would take a long time to put just one sample in front of the beam,” he says. “If you can put the samples into a container and have a robot that can do that for you, you can speed it up and have less error.”

Eukaryotic Troubles

Structural genomics centers have deposited over 6,600 structures in PDB. Yet challenges remain. Of 25,662 targets selected at the JCSG, 20,865 have been cloned, 20,546 expressed, 1,337 crystallized, and 686 solved—a 3.3 percent success rate. **The New York SGX Research Center for Structural Genomics** puts up comparable

numbers: 486 structures from 8,849 selected targets (5.5 percent).

Eukaryotic proteins represent a particular challenge. For a variety of reasons, including posttranslational modifications, size, and domain structure, eukaryotic proteins are tougher nuts to crack than prokaryotic ones. As of June 1, 2008, the 10 PSI-2 centers have solved 1,716 protein structures, but just 197 of those are eukaryotic; the rest are either prokaryotic (1,500) or viral (19).

“Quite often proteins require glycosylation to be made properly,” says Ray Owens, director of the **Oxford Protein Production Facility (OPPF)**, Oxford, UK. “So you cannot make them in *E. coli*. They need to be made in eukaryotic cells, which will authentically glycosylate them.” OPPF researchers use transient transfection of human embryonic kidney (HEK293) cells. Because most glycoproteins are secreted, purification is a snap: just collect the supernatant. The problem, Owens says, is “the chemical heterogeneity of the glycoproteins.” To skirt that issue team members use inhibitors like kifunensine to freeze glycoprotein sugar chains in a more homogeneous form, and endoglycosidases to pare these glycans back.

CESG researchers use a cell-free eukaryotic wheat germ extract from **CellFree Sciences** for their protein expression. “We find that we get roughly twice the number of successful targets produced by wheat germ extract than with *E. coli*,” says Markley. It is also faster, less expensive, and involves easier protein purification, because sufficient protein for a structure determination is isolated from milliliter reactions, rather than in liter quantities of cells.

Membrane Proteins

Membrane proteins also present unique structural challenges, says Scripps’ Stevens. First, because the membrane represents such a small fraction of total cellular volume, they are present at relatively low levels. More important, the membrane itself is integral to their structure. “We have to solubilize them, which puts them in a very unstable state,” he says.

Yet detergent interferes with the protein packing that is essential to crystallization. Stevens has developed or implemented several techniques to work with membrane proteins, including a high throughput “lipidic cubic phase” crystallization process first developed by Martin Caffrey of the **University of Limerick**, Ireland, that was key to solving the beta-2 adrenergic receptor structure.

Researchers at the **New York Consortium on Membrane Protein Structure (NYCOMPS)**, a PSI-2 specialized center, have alternative methods, says director Wayne Hendrickson of **Columbia University**. “We have developed a pipeline process analogous to the kind that has been effective at large-scale centers for soluble proteins to produce membrane proteins and analyze them by crystallography and NMR,” he says.

For instance, the consortium uses ultraviolet absorbance and light scattering to test which detergents work best and which proteins are in the proper oligomeric forms. More recently, they developed a generic antibody-based approach to extensively incorporate selenomethionine residues into protein complexes, key to solving X-ray diffraction “phase problems.”

Using tools such as these, NYCOMPS researchers have produced over three thousand membrane proteins in the past year, resulting in five structures, with another seven or eight “in determination.” “We are very much in early stages,” Hendrickson says.

And despite the thousands of structures pouring out of high throughput labs around the world, so, it is safe to say, is structural proteomics in general.

Jeffrey Perkel is a freelance writer based in Pocatello, Idaho.

DOI: 10.1126/science.opms.p0800027