

Difficult to express	429
Homing in on easier targets	431
Unflagging focus	432
Box 1: Hitting the sweet spot	430

Making membrane proteins for structures: a trillion tiny tweaks

Monya Baker

Researchers try multiple means to get high-quality membrane proteins for X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy studies.

As much as 30% of human genes code for membrane proteins. But of more than 65,000 structures in the RCSB Protein Data Bank, fewer than a dozen full-size human membrane proteins are represented. The low numbers are not for lack of interest. Membrane proteins help coordinate pretty much everything a cell does; knowing their structures could help reveal how they do so. Researchers at pharmaceutical companies are keenly aware that about half of approved therapeutics target human membrane proteins and are hopeful that structural information can help design better drugs.

The desire for and difficulty of determining membrane protein structures are so great that when the US National Institute of General Medical Sciences (NIGMS) created six specialized centers to work on difficult problems in protein structure, two were devoted to membrane proteins. These six centers are part of a much larger network of centers and projects launched by the NIGMS Protein Structure Initiative (PSI). Grants for the next round of PSI will be announced in July 2010, and membrane proteins remain a big thrust of the program. (In collaboration with Nature Publishing Group, the PSI also produces the Structural Genomics Knowledgebase, a portal to PSI resources and current research.)

Membrane proteins are inherently hard to make and characterize. Even when expressed at high levels, membrane proteins do not purify well. The lipids surrounding proteins in cell membranes interfere with both crystallography and nuclear magnetic resonance (NMR) spectroscopy; when not embedded in the lipid

bilayer, though, membrane proteins usually lose their three-dimensional structure.

A few membrane proteins are expressed in high-enough quantities to be collected from natural sources. Generally, though, getting amounts of protein suitable for structural studies requires researchers to engineer vectors that overexpress proteins and to clone them into a cell expression system. The volumes of cell culture required to obtain just a few milligrams of protein can run into liters. Purification procedures that work for one protein may fail even for closely related proteins.

Multiple specialized techniques for stabilizing and manipulating proteins come into play depending on the protein itself and whether researchers plan to use X-ray crystallography, NMR spectroscopy or other techniques to probe protein structure. Every step of the process needs to be optimized, says Raymond Stevens of The Scripps Research Institute, who has solved several difficult membrane protein structures. "Everybody wants to understand what was the one technological breakthrough," says Stevens. "There wasn't one. There were actually about fifteen."

That is to be expected, says Stephen Burley, a distinguished scholar at the Lilly Biotechnology Center and former chief

scientific officer of SGX, which developed techniques for high-throughput crystallization for human drug targets. The tools

and tricks for studying membrane proteins are different from those used for soluble proteins, he says, but the strategy is the same: working on all aspects of the process. "The key to success is the order in which you try everything."

Difficult to express

Researchers who want to make human membrane proteins generally cannot depend on the standard protein production workhorse, *Escherichia coli*. "*E. coli* is extremely good at making membrane proteins," says Chuck Sanders,

a biochemist at Vanderbilt University. "The problem is that they usually can't fold them." That is because bacterial machinery for folding proteins and facilitating disulfide-bond formation in proteins is quite different from that found in eukaryotes. Still, working with *E. coli* is about an order of magnitude faster and cheaper than working with mammalian protein expression systems. It is also much easier to genetically manipulate and grow in culture than eukaryotes. And unlike all eukaryotes except the yeast *Pichia pastoris*, *E. coli* readily incorporate various isotope-labeled amino acids into proteins, rendering them suitable for structural NMR spectroscopy studies. Consequently, many researchers



Raymond Stevens at The Scripps Research Institute uses crystal structures and other techniques to study how membrane proteins function.



Brian Kobilka

Brain Kobilka at Stanford University studies the structure and activity of G protein-coupled receptors, with an emphasis on the beta-2-adrenergic receptor.

continue to hunt for conditions that could transform the prokaryote into a suitable membrane protein production factory.

Sanders keeps *E. coli* in cold, nutrient-poor conditions; slower-growing bacteria seem to produce better-folded proteins. Adding lipids from higher organisms to the bacteria's growth medium also boosts yield of functional protein. Sanders says his approach has been successful for producing the amyloid beta precursor protein, which is associated with Alzheimer's disease, as well as peripheral myelin protein 22, mutations in which cause Charcot-Marie-Tooth disease and other neuropathies. But he has not had similar luck using *E. coli* to produce two classes of proteins of intense interest to drug companies: G protein-coupled receptors (GPCRs) and ion channels. Still, with so much of the human membrane proteome unexplored, says Sanders, using a fast, economical system to find tractable proteins makes sense. "If the protein expresses, we'll work on it; if it doesn't, we'll move on."

Some researchers are trying to do away with cellular production of proteins altogether. Cell-free protein synthesis systems extract protein production machinery from cell lysates and allow researchers to add protein genes and reagents. They are particularly useful for making proteins that are difficult to express or purify from cells. They also readily incorporate isotope-labeled amino acids useful for NMR spectroscopy studies.

Researchers at the University of Wisconsin-Madison are using a cell-free system that uses wheat germ lysates to produce GPCRs and other membrane proteins, a technology commercialized by CellFree Sciences, which supplies both reagents and protein-production services. Companies such as Promega Corporation also offer a similar technology. Wheat-germ extracts can produce tens or even hundreds of milligrams of membrane proteins for NMR spectroscopy and X-ray crystallography structural analysis, says Masaki Madono, head of sales and marketing for CellFree Sciences. "The challenge is how to make them soluble for subsequent purification." Even so, cell-free systems have an advantage, he says, because solubilizing agents such as detergents and liposomes can be added to the system without toxic effects.

Dan Luo at Cornell University has created a cell-free system using a hydrogel made from highly branched DNA ligated

to genes coding for the desired proteins. This sequesters reactants in a confined space and protects genes from degradation; one study showed the approach was 300 times more efficient than using cell-free systems in solution¹. Luo, who cofounded the company DNANO to commercialize the technology, says the gel pads have been used to make at least five membrane proteins, with some at concentrations as high as 1 mg ml⁻¹. (Except for cell-surface protein CD38, identities of the proteins are confidential.)

Researchers are also working on improved versions of cloning vectors, robotics systems, labeling systems, affinity tags and modifications to protein sequence, not just for membrane proteins but for other challenges such as glycosylated proteins (Box 1). Nonetheless, Aled Edwards, who directs the International Structural Genomics Consortium, says he has yet to see any easy, generalizable new solutions. "There's not much that

BOX 1 HITTING THE SWEET SPOT

A particularly challenging class of proteins to study is glycosylated proteins, proteins that are posttranslationally modified to contain sugars. Cells use these modifications to detect pathogens, sort proteins into the right intracellular locations and perform other functions. Proteins made by nonmammalian cells have different glycosylation patterns; even in mammalian cells, the modifications on proteins are usually heterogeneous, making structural studies very difficult or even impossible. Researchers have tried various approaches to reproduce homogenous mammalian glycosylation patterns; at least two yeast strains have been engineered to produce somewhat human-like glycosylation patterns. Deane Mosher at the University of Wisconsin is making glycosylated thrombospondin in insect cells for X-ray crystallography studies. Ian Wilson at The Scripps Research Institute has used a baculovirus expression system to produce glycosylated proteins suitable for X-ray crystallography studies.

This year, researchers led by Markus Aebi of the Institute of Microbiology, Eidgenössische Technische Hochschule (ETH) Zurich and Lai-Xi Wang at the University of Maryland School of Medicine, described a reliable way to produce one of the most common mammalian posttranslational modifications, called N-linked glycosylation, in which a sugar group is added to asparagine³. They started with the unusual bacteria *Campylobacter jejuni*. Unlike almost all other prokaryotes, *C. jejuni* glycosylates its proteins, albeit in a way that is different from the glycosylation in mammals. By deleting some parts of the *C. jejuni* glycosylation machinery and cloning the rest into *E. coli*, the researchers created a bacterial strain that produced transgenic proteins with an assortment of glycans, including the desired linkage. Crucially, the undesired bacterial glycans could be removed and modified enzymatically to create human proteins with highly homogenous glycosylation. Wang says he is currently identifying and optimizing the best enzymes for making homogenous glycoproteins, but the technique has not yet produced proteins other than antibodies. Like most in the field, Wang's team is focusing on making potentially therapeutic proteins; Aebi also continues to focus on immunoproteins, but, he says, collaborations are already underway to produce glycoproteins for structural analysis by NMR spectroscopy.

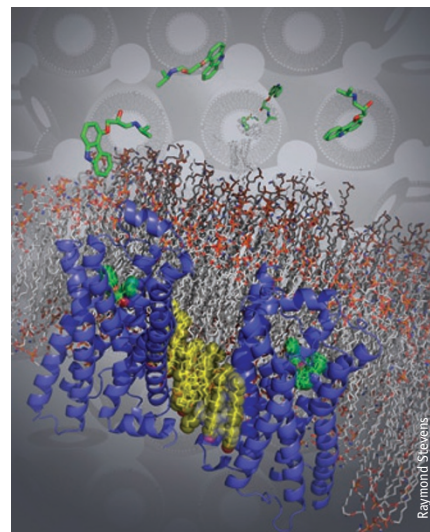
makes me stop my lab and investigate a new methodology,” he says. “They are incremental advances, if anything.” He has been in the field a long time, he says, and his best recommendation for success is planning for a long slog. “We’re plodding along, honing the systems we’ve had in place over the past decade, learning how to better use them. That’s the trend, to take what we know already and apply.”

Homing in on easier targets

As part of the PSI efforts, Robert Stroud at the University of California, San Francisco and colleagues are working to create a high-throughput pipeline for eukaryotic membrane proteins. With the goal of solving structures by crystallography, Stroud’s team picked the yeast strain *Saccharomyces cerevisiae* as an expression system because seven of the 13 integral membrane protein structures solved at the time using recombinant proteins had been produced in this organism. They chose the protein-extraction detergent

n-dodecyl- β -D-maltopyranoside using a similar rationale. They used polyhistidine affinity tags to capture proteins for purification. However, they adjusted conditions to suit membrane proteins by attaching tags to the C termini rather than the N termini of proteins and by cleaving tags using proteases selected to work in detergents. They used size-exclusion chromatography to eliminate proteins that had misfolded or aggregated.

The researchers chose 384 proteins from the approximately 6,600 in the yeast proteome based on a combination of bioinformatics and curation for diverse membrane proteins with at least three transmembrane helices. Of the first 96 proteins thus characterized, they deemed 23 to be of high-enough quality for subsequent studies, a rate that is similar to that recently reported for a set of globular prokaryotic proteins, which are considered the easiest targets. Next, the researchers tackled ten human membrane transporters in the solute carrier superfamily and obtained four



High-resolution structure of the beta-2-adrenergic receptor, a G protein-coupled receptor important for basic research and medicine.

proteins in sufficient quantity and quality for further studies. Although this high-throughput approach promises to produce

TECHNOLOGY FEATURE

more structures of human membrane proteins, it will still exclude many interesting proteins, says Stroud. Its efficiency relies on subjecting all proteins to a highly standardized workflow rather than optimizing conditions for each protein.

Adam Godzik, a professor of bioinformatics and systems biology at the Burnham Institute, has conducted a comprehensive survey of which soluble prokaryotic proteins have produced crystals or failed to do so and used machine learning to predict which proteins are most likely to produce crystals. “If you come to us with a ‘difficult’ protein, we can classify the difficulty, and we can tell you what is most likely to fail,” he says. For soluble proteins, his algorithms can even advise researchers which bits of a protein to truncate to boost chances of getting a crystal structure. He is making plans to undertake a similar project for membrane proteins but anticipates that collecting the data will be much more difficult. Data for successful crystallization of membrane proteins are limited, and patterns may not be predictive for membrane proteins as a class. “Transmembrane proteins are like a negative definition,” says Godzik. “It’s like saying someone is nonwhite. There are many different ways to be nonwhite. The proteins may have nothing to do with each other than that they are sitting in the membrane.”

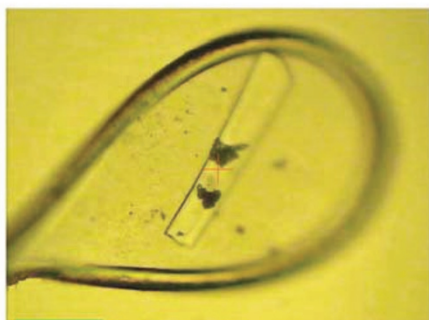
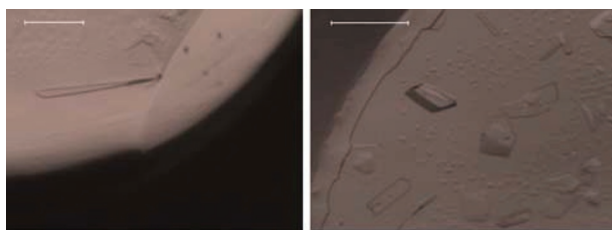
Unflagging focus

Researchers who feel compelled to study a particular protein have little use for tools that predict which are the easiest to produce or purify. Brian Kobilka of Stanford University became fascinated by adrenergic receptors, a family of GPCRs that respond to adrenalin, when he completed his medical residency. “So many of the drugs we used to treat people in the intensive care unit worked on adrenergic receptors. Some drugs could increase blood pressure if it was too low, and other drugs could lower blood pressure if pressure was too high. I was fascinated that there was this one family of proteins that could control so much physiology.” Kobilka worked on the beta-2-adrenergic receptor for 17 years before publishing its first crystal structure².

Like most scientists studying GPCRs, Kobilka uses a cell line originally derived from army moth ovaries to make proteins for structural studies. This system requires expensive medium but produces relatively high levels of the receptor. Still, getting enough protein is a far cry from a crystal or NMR spectroscopy structure. “For a long time people thought that the main problem was expression levels, and if you could express receptors at a high enough level, lo and behold they would crystallize,” says Fiona Marshall, chief scientific officer of Heptares, which has generated



Wheat-germ technology from CellFree Sciences can be incorporated into robots for protein synthesis.



Heptares Therapeutics

Heptares is using stabilizing mutagenesis to obtain structures from crystals of G protein-coupled receptors including the adenosine A2a receptor (top; scale bar, 0.2 mm) and the beta 1-adrenergic receptor (bottom; scale bar, 100 μ m).

crystal structures of GPCRs and is using them for drug discovery. When optimizing expression, says Marshall, researchers need to carefully watch expression and purification conditions to make sure that the proteins are properly folded.

For NMR spectroscopy studies, the problem with using detergents is the formation of protein-free micelles that create artifacts, and some researchers are turning to embedding proteins into lipid nanodiscs originally developed by Steven Sligar at University of Illinois at Urbana-Champaign. But most membrane proteins are still solubilized with detergents for both NMR spectroscopy and X-ray crystallography studies. These molecules include chains of hydrophobic chemical

groups that can help stabilize the proteins. But detergents that work best for keeping proteins stable are also the worst for crystallization, says Marshall. “If you have a long detergent, it’s like the cell membrane, and the protein feels reasonably happy, but then you can’t crystallize it because all the protein is covered by detergent. As you make the detergent shorter, more of the protein is exposed.”

Several researchers are developing better detergents as well as screening methodologies that quickly determine which detergent works best with a particular membrane protein. However, scientists studying GPCRs have developed alternate ways to stabilize the protein. Heptares uses a method developed by cofound-

er Chris Tate of the Medical Research Council (MRC) Laboratory of Molecular Biology in Cambridge, UK to identify stability-boosting mutations in various GPCRs selected as promising drug targets. Among other techniques, Kobilka has used an antibody to two transmembrane helices to stabilize a particularly floppy part of the protein. For another crystal structure, Kobilka replaced that disordered section with a domain from the T4 lysozyme domain and used a small-molecule ligand to stabilize the inactive conformation. Stevens, who cofounded the company Receptors, also uses small-molecule ligands that stabilize various conformations of GPCRs.

Researchers are careful to confirm ideas from structural analysis with functional studies of proteins, but structures can provide insight that is otherwise unobtainable. For instance, comparing the structures of rhodopsin, adenosine A2a receptor and beta-adrenergic receptors revealed that the same ligand can bind related proteins in very different orientations, so modeling from one receptor to another may not always be valid. Despite the value of structural studies, time and money are limiting factors, says Stevens. “This is extremely expensive work. We need to continue working on methods to reduce the cost.”

1. Park, N., Um, S.H., Funbashi, H., Xu, J. & Luo, D. *Nat. Mater.* **8**, 432–437 (2009).
2. Rasmussen, S.G.F. *et al. Nature* **450**, 383–387 (2007).
3. Schwarz, F. *et al. Nat. Chem. Biol.* **6**, 264–266 (2010).

Monya Baker is technology editor for *Nature* and *Nature Methods* (m.baker@us.nature.com).

SUPPLIERS GUIDE: COMPANIES OFFERING SERVICES AND SUPPLIES FOR STRUCTURAL STUDIES OF PROTEINS

Company	Web address	Company	Web address
Accelrys	http://www.accelrys.com/	DiscoverX	http://www.discoverx.com/
Affymetrix	http://www.affymetrix.com/	Douglas Instruments Ltd.	http://www.douglas.co.uk/
Agilent	http://www.agilent.com/	EMD4Biosciences	http://www.emdbiosciences.com/
Applied Biosystems	http://www.appliedbiosystems.com/	Emerald BioSystems Inc.	http://www.emeraldbiosystems.com/
Assay Designs	http://www.assaydesigns.com/	Fluidigm	http://www.fluidigm.com/
BD Biosciences	http://www.bdbiosciences.com/	Formulatrix	http://www.formulatrix.com/
Beckman Coulter	http://www.beckman.com/	GE Healthcare	http://www.gehealthcare.com/
BioFocus	http://www.biofocusdpi.com/	Geneservice	http://www.geneservice.co.uk/
Bio-Rad	http://www.bio-rad.com/	Genetix	http://www.genetix.com/
BioTek	http://www.biotek.com/	GenScript	http://www.genscript.com/
BMG Labtech	http://www.bmglabtech.com/	Hamamatsu	http://www.hamamatsu.com/
Bruker Daltonics	http://www.bdal.com/	Hamilton Company	http://www.hamiltoncompany.com/
Caliper Life Sciences	http://www.caliperls.com/	Heptares Therapeutics	http://www.heptares.com/
Cambridge Bioscience	http://www.bioscience.co.uk/	Hudson Robotics, Inc.	http://www.hudsoncontrol.com/
Cellomics	http://www.cellomics.com/	Imagenex	http://www.imgenex.com/
CEREP	http://www.cerep.fr/	Invitrogen	http://www.invitrogen.com/
ChanTest	http://www.chantest.com/	JEOL	http://www.jeol.com/
Cisbio	http://www.htf.com/	Lectinity	http://www.lectinity.com/
Clontech	http://www.clontech.com/	Matrix Science	http://www.matrixscience.com/
Corning	http://www.corning.com/	Meso Scale Discovery	http://www.meso-scale.com/
		Millipore	http://www.millipore.com/
		Molecular Devices	http://www.moleculardevices.com/
		Molecular Dimensions	http://www.moleculardimensions.com/
		Multispan	http://www.multispaninc.com/
		New England Biolabs	http://www.neb.com/
		Oxford Diffraction	http://www.oxford-diffraction.com/
		Pall Corporation	http://www.pall.com/
		patoBios	http://www.patobios.com/
		PerkinElmer Life Sciences	http://las.perkinelmer.com/
		Pierce Biotechnology	http://www.piercenet.com/
		Promega	http://www.promega.com/
		Protein Sciences Corp.	http://www.proteinsciences.com/
		Protein Technologies	http://peptideinstruments.com/
		Qiagen	http://www1.qiagen.com/
		R&D Systems	http://www.rndsystems.com/
		Receptos, Inc.	http://www.receptos.com/
		Rigaku	http://www.rigaku.com/
		Roche Applied Science	http://www.roche-applied-science.com/
		Sigma-Aldrich	http://www.sigmaaldrich.com/
		Sussex Research	http://www.sussex-research.com/
		Takara Bio Inc.	http://www.takara-bio.com/
		Tecan Group	http://www.tecan.com/
		TGR BioSciences	http://www.tgr-biosciences.com.au/
		Thermo Scientific	http://www.thermo.com/
		Varian	http://www.varianinc.com/
		Waters	http://www.waters.com/