

Fast forward

Proteomics raises new challenges in protein purification. Technologies well adapted to isolate individual proteins get a makeover to tackle large numbers of samples. Laura Bonetta investigates.

Proteomics aims to identify the cellular functions of all proteins encoded by the genome of an organism. Protein structure determination, proteome-wide functional screens, and the identification of protein interactions are just a few of the proteomics applications requiring thousands of purified proteins as a starting point. Researchers now have many reagents and off-the-shelf kits for the isolation of proteins to 90% purity. The challenge facing researchers is to adapt these methods to purify hundreds of different proteins in parallel using robotic systems. Several years ago a small number of academic institutions and biotechnology companies responded to the challenge by developing pipelines for high-throughput protein purification (see “Proteomics at Harvard”). Some of the reagents and instruments required for these methods are making their way to the marketplace.

HT purification

One widely used method for purifying hundreds of proteins in parallel involves their expression in a heterologous system. Once a target protein is identified, the corresponding

complementary DNA is cloned into an expression vector for producing the protein in *Escherichia coli* or another organism. The purification of the expressed recombinant protein from bacteria typically involves cell lysis, incubation of the lysate with an affinity resin, washes and elution. Additional purification steps, such as ion-exchange and size-exclusion chromatography, are required for protein crystallography (see “Protein Purification for Structural Proteomics”).

One bottleneck in high-throughput protein production is the expression of enough properly folded protein in bacterial cells. Proteins expressed in *E. coli* can accumulate as inactive, misfolded aggregates called inclusion bodies. Proteins in this form can be purified under denaturing conditions, but need to be refolded into their native conformations. A number of groups have developed procedures for expressing proteins in insect cells or cell-free extracts to get around some of these problems (see “Cell or Cell-Free?”).

Each protein is expressed in different amounts and has different properties. Thus, to be able to apply the same purification protocol



PALL CORPORATION

Splitter: Pall's AcroPrep 96 Filter Plate.

across the range of proteins, researchers have engineered proteins with generic tags that will bind to an affinity ligand. Widely used tags include a small peptide of six histidine residues (6xHis), a calmodulin-binding peptide, the streptavidin friendly biotin, the cellulose-binding tag, the maltose-binding protein (NusA), and glutathione *S*-transferase (GST). Vectors designed for expressing tagged proteins can be purchased from several manufacturers includ-

PROTEOMICS AT HARVARD

Founded in the spring of 1999, the Harvard Institute of Proteomics (HIP) at Harvard Medical School aims to provide tools to determine the function of every protein encoded by the human genome and appropriate model and disease organisms. To this end, HIP scientists are building collections of genes from humans and organisms including *Saccharomyces cerevisiae*, *Vibrio cholerae*, *Yersinia pestis*, *Pseudomonas aeruginosa* and *Bacillus anthracis*, as well as some mouse and viral genes. In addition, a number of projects are focused on specific groups of medically relevant genes.

The Breast Cancer 1000 Project has developed a repository of clones for 1,000 genes that contribute to the onset of breast cancer, a subset of which have already been tested in functional assays to identify cDNAs that induce cancer-like phenotypes.



Another project aims to clone the entire range of human kinases. The expression clones generated at HIP, along with the technology to use them, will be made available to all researchers.

Joshua LaBaer (pictured), co-founder and director of the institute, says his group has developed methods for the high-throughput purification of proteins

expressed in *Escherichia coli* and downstream processes to determine their functions or properties. “In one project, we are purifying all the proteins produced by the organism *Francisella tularensis*, the parasite that causes tularaemia,” says LaBaer. The organism’s proteome consists of about 1,600 proteins, all of which have been expressed in *E. coli* and purified. The purified proteins are then used in high-throughput functional screens “to find proteins that produce an immune response”. A similar study focuses on the genome of the bacterium *V. cholerae*, which can cause cholera in humans. Such studies are not only yielding important drug targets but laying the groundwork for studies targeting the much more complex human proteome.

LaBaer says that two bottlenecks affecting the purification of proteins from

Y. pestis, the plague bacterium, are the “the ability to express large or hydrophobic proteins in bacteria, and avoiding the inclusion-body problem.” Partly for this reason the institute is now moving towards the development of protein arrays, with proteins being synthesized on the array rather than spotted on them. In their protocol, plasmid DNA is spotted on the array and genes are then transcribed and translated in a cell-free system. The resulting proteins, hundreds of them per chip, are immobilized *in situ* and can be used to test for protein-protein interactions and other functional assays. Although the technology is not yet ready for primetime, LaBaer says they are having success with it. The development of these types of arrays may alleviate the need to produce and purify proteins from *E. coli*, as scientists will be able to conduct functional studies directly on the arrays. L.B.

ing Invitrogen (Carlsbad, California), Novagen (Madison, Wisconsin), Roche Applied Science (Indianapolis, Indiana) and others.

Some companies are now developing high-throughput methods for purifying proteins that are not tagged. "Typically most scientists work with tagged proteins through the discovery process. However, it is possible to lose some information because of tag interference in the biological assays. Removing the tag means removing some of the question marks," says Lisa Bradbury, director of R&D for proteomics and cell therapy at Pall Corporation of East Hills, New York.

Purification tool bag

Several companies sell reagents for every step of the purification process — from cell lysis, to affinity purification resins, to desalting and concentrating reagents — adapted to high-throughput protocols that use standard 96-well plates. "We were recognized for our protein affinity-purification line. Now we are configuring it into a format that is automation-friendly," says Craig Smith, vice-president of R&D at Pierce Biotechnology of Rockford, Illinois (part of Fisher Scientific). Among the company's offerings are the SwellGel Discs, pellets of dehydrated support that can be distributed into filter plates. When the protein is



High-capacity purification: Teledyne Isco's BioOptix 10.

added, the disc rehydrates to an affinity gel that binds tagged proteins. "We have taken out variability by having a dried pellet," says Smith. More recently, the company's Zeba Micro Desalt Spin Columns were configured to 96 well-plate format for desalting many small samples in parallel.

Pall Corporation also sells reagents for each step of the purification process. The AcroPrep 96 Filter Plate can be used for any type of chromatography, including immobilized metal affinity chromatography, in a multiwell platform. The format facilitates the optimization of various purification parameters, including the choice of metal ion and resin, the sample-to-resin ratio, and the elution conditions. Other reagents adapted to a high-throughput

environment include the Mustang 96-well ion-exchange plates and Nanosep Centrifugal Devices.

QIAGEN, based in Hilden, Germany, offers the Superflow 96 BioRobot, a medium-scale purification kit for 6xHis-tagged proteins. "You load your pellets onto a robot, press a button, and you will have product in two hours," says Frank Schäfer, associate director for R&D protein expression/proteomics. According to the company, purification can be done in denaturing or native conditions and yields 4 mg of protein for each of 96 wells. "It is one system and one process for every sample," says

Schäfer. QIAGEN sells a similar system that runs up to 24 samples in parallel to purify up to 30 mg of protein for large-scale applications.

Novagen's host of reagents for high-throughput protein purification starts with the expression step. The Overnight Express Autoinduction Systems allow fully automated, parallel protein expression from *E. coli* cultures, and achieve high cell densities without monitoring growth or induction by the addition of isopropyl-beta-D-thiogalactopyranoside. "The medium contains a blend of carbon sources optimized for tightly regulated, uninduced growth and automatic induction of protein expression," says Anthony Grabski, R&D group leader for protein purification. The RoboPop Purification Kits contain reagents for the extraction and purifica-

PROTEIN PURIFICATION FOR STRUCTURAL PROTEOMICS

Structural genomics, or structural proteomics, aims to provide three-dimensional information for all proteins. This information can be used to ascribe function to a protein and to reveal or invalidate drug targets. There are structural proteomics projects in both the public and private sectors around the world, including Germany, Japan and the United States.

In 2000, the US National Institute of General Medical Science in Bethesda, Maryland, launched an initiative to determine the structures of about 10,000 proteins representing different structural ('fold') families over the next decade. The studies will result in a public resource linking amino-acid sequence, structural and functional information, eventually allowing scientists to make the three-dimensional atomic structures of most proteins easily obtainable from their corresponding sequences.

One of several participating centres is the Genomics Institute of the Novartis Research Foundation in San Diego, a member of the Joint Center for Structural Genomics (JCSG). Scott Lesley (pictured below, left), who heads the institute's proteomics unit, says his group has processed 50,000 protein samples from more than 20,000 unique expression clones. The pipeline starts with a small screen protocol



that uses samples in a 96-well format to evaluate how the targets behave at different steps in the procedure. "Those that are not behaving well after expression, purification or crystallization are set aside and then run through one of several salvage pathways," says Lesley. Such pathways include using different expression systems (vectors and cells), making point mutations in the cDNA to increase the likelihood it will be expressed,

and trying different tags. "All the bacterial expression and purification is done in parallel and the majority is automated by custom instrumentation," says Lesley.

The targets that perform well are applied to a large-scale purification protocol to yield

10–20 mg of protein. The JCSG has already deposited 239 structures in the protein database — mainly from the bacterium *Thermotoga maritima*, the mouse, and the yeast *Saccharomyces cerevisiae* — representing at least 19 new fold families. "Most of the tools are in place, and we can increase both output and success rate by implementing salvage approaches in high throughput," says Lesley.

Similar efforts are underway in other countries. The Protein Structure Factory in Germany, an initiative of the German Human Genome Project and structural biologists from the Berlin area, targets human proteins for structure determination. "Our current focus is to determine the structures of particular protein-protein or protein-ligand complexes," says Christoph Scheich of the Max Planck Institute for Molecular Genetics in Berlin.

tion of His or GST fusion proteins by magnetic or filtration-based affinity purification protocols, all in a 96-well format. The kits have been validated on both PerkinElmer and Tecan robotic liquid handlers.

Other providers of reagents for high-throughput protein purification include GE Healthcare of Little Chalfont, UK, with its Tricorn High Performance Columns designed for high-resolution purification of proteins, peptides and other biomolecules. Sartorius of Göttingen, Germany, introduced the Vivapure 8-to-96 well cobalt chelate kit for the simultaneous purification of multiple His-fusion proteins by affinity chromatography. And St Louis-based Sigma-Aldrich's HIS-Select iLAP Plates are plates coated with cell-lysis reagents and a HIS-Select nickel chelate matrix allowing for cell lysis, protein capture, and assay of a His-fusion protein in a single well.

As well as purification kits, several companies offer products to help screen the outcome of each step in the purification pipeline. Novagen's RoboPop Solubility Screening Kit contains a filtration plate that retains insoluble inclusion bodies while allowing soluble proteins to be collected for rapid quantitation and analysis. "You can screen expression conditions for soluble protein before you proceed to purification," says Grabski. In January, the company released the iFold Protein Refolding System 1 to screen different refolding conditions in parallel. The system includes inclusion-body purification reagents and 92 different refolding buffers. "Structural proteomics

groups have harvested most of the low-hanging fruits, so now they will have to focus on difficult proteins that are insolubly expressed," says Grabski. Another product for monitoring the purification procedure is the Protein 200-HT2 assay of Agilent Technologies (Palo Alto, California), which allows the identification, sizing and quantification of proteins from 14 kD to 200 kD in size. "It allows researchers to replace SDS-PAGE procedures in the lab," says product manager Carsten Buhmann. Finally, some reagents and instruments have been designed to help researchers hone in, for example, on medically relevant targets, which can then be purified and studied (see "Target Identification").

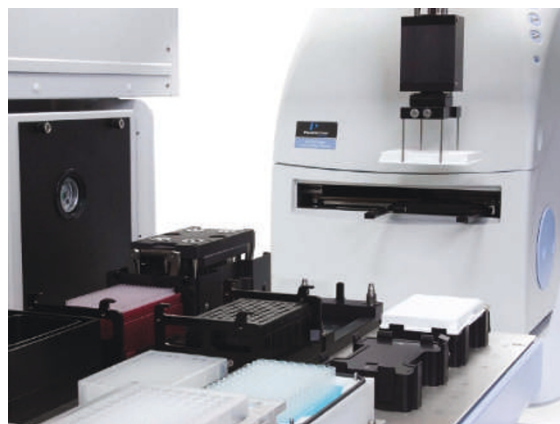
Getting automated

The available reagents and kits can be used to purify several hundred proteins in a 96-well format at considerable speed. However, the tedious and error-prone nature of manually performed high-throughput operations calls for automation of the process. "When a robot is doing the pipetting there are no mistakes, so you can have greater consistency and reproducibility," says David Daniels, applications marketing manager for Beckman Coulter in Fullerton, California. Liquid handlers perform all the steps of protein purification from loading cell cultures to obtaining a protein in solu-

tion, in under four hours.

Two leading liquid handlers are Beckman Coulter's Biomek 3000 and Biomek FX instruments, with plate-deck configurations that can handle 10 and 18 plate positions, respectively. The Biomek 3000 will process two 96-well plates in 2–3 hours; the FX twice as many.

Promega in Madison, Wisconsin, automated high-throughput protein purification using its MagneHis Protein Purification System and the Biomek FX instruments. The MagneHis reagent contains a cell-lysis solution that allows resuspension and lysis of bacterial cell pellets without sonication and centrifugation. Magnetic pre-charged nickel particles are then used to isolate His-tagged proteins from the cell lysate. The MagneHis particles bound to the target proteins are captured on a MagnaBot 96



Mix and match: PerkinElmer's JANUS workstation.

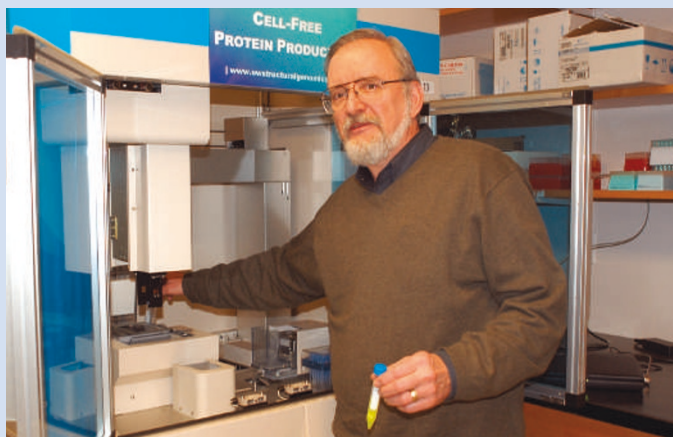
PERKINELMER

CELL OR CELL-FREE?

Only a fraction of proteins can be overproduced in *E. coli* in sufficient yield and without the formation of inclusion-body aggregates or the proteolytic degradation of expressed proteins. Alternative expression systems include cell cultures from eukaryotic organisms, such as insect cells, and cell-free, *in vitro* protein expression.

The latter is the focus of John Markley's group at the Center for Eukaryotic Structural Genomics at the University of Wisconsin, Madison. Collaborating with Ehime University and CellFree Sciences, both in Japan, Markley and colleagues developed a pipeline using wheat germ cell-free protein translation as a way to produce proteins for nuclear magnetic resonance (NMR) structure determination.

Cell-free systems simplify the purification efforts, as "only the protein of interest is expressed



and labelled, thus the background is cleaner", says Markley (pictured). For the NMR studies, his group has been getting twice as many folded proteins using the cell-free system than with expression in *E. coli*. The cell-free method also requires smaller volumes, avoiding lengthy concentration steps, and it lends itself to the labelling of proteins,

which is a requirement for NMR. "But there is a steep learning curve for learning how to do a cell-free system," says Dmitriy Vinarov, who is responsible for high-throughput production at the centre. One downside of cell-free systems is the expense of the reagents, especially when success rates are not high. "About 79% of human proteins will be expressed,

about one half of those will produce protein in sufficient quantities for structural studies, and half of those will remain stably folded for NMR studies," says Markley.

Cell-free and cell-based systems are not mutually exclusive. "Each has unique advantages and capabilities so we will continue to do both," explains Markley. By using a new cloning system from Promega the researchers have been able to transfer target cDNAs from cell-free to cell-based expression systems to achieve greater flexibility. "We still have a lot to learn. It is working quite well for us but we still have improvements to make," says Vinarov. "The technology is not quite primetime."

Cell-free systems are available from companies such as Roche Applied Science, QIAGEN and Invitrogen, as well as from CellFree Sciences.

R. DAVIES

L.B.

Magnetic Separation device — a plate that can be hooked up to the Biomek instruments.

In January, PerkinElmer (Boston, Massachusetts) launched its modular and scalable JANUS Automated Workstation. “Many units can be configured at the time of sale,” says Nance Hall, business unit leader for automation and liquid handling. “With JANUS, you can scale and change the system not only when you buy, but also in the future.” Users can also program the selection of numerous dispense heads and formats. “It is as though you go into a kitchen and you are asked if you need a cup or a teaspoon,” explains Hall. “In the kitchen you will use both.” Likewise, the JANUS allows users to choose different tools automatically depending on the volume range and microplate densities.

Another popular choice among liquid handlers is the Freedom EVO automated liquid handler from Tecan in Durham, North Carolina. Others include the Sciclone ALH Workstation from Caliper Sciences (Hopkinton, Massachusetts), the BioCube System from Proteodyne (Windsor, Connecticut) and the 925 PC Workstation and GX-281 Liquid Handler from Gilson (Middleton, Wisconsin). “You can use over 300 standard racks and many more custom ones,” says Gilson’s spokesperson Greg Robinson.

Going large

For crystallization and other functional studies, a researcher needs more protein than is possible in 96-well plate format. For such purposes, a number of parallel-purification sys-



Just swell: Pierce Biotechnology’s SwellGell discs.

tems that can accommodate larger volumes of cells are starting to come onto the market.

GE Healthcare’s AKTExpress is a fully automated chromatography system for purification of His- and GST-tagged proteins, yielding up to 50 mg of target protein. Affinity chromatography is the first step of all protocols, but as a second step it is possible to choose between desalting or gel filtration, and an ion-exchange step can be added too. Automatic tag removal is possible in all multi-step protocols.

Teledyne Isco (Lincoln, Nebraska) launched the BioOptix 10 last year, a ten-channel parallel-purification system for high-capacity protein purification with subsequent fraction collection. This means it can purify protein

from ten different samples by ion exchange, affinity or size-exclusion chromatography. The instrument includes a high-capacity fraction collector (20 to 60 fractions per sample). Ten independently controlled pumps can be programmed with different gradient conditions for rapid identification of columns and conditions that give optimal results. “Instead of doing it sequentially, a process that can take 4–5 days, you can load different columns and run the instrument over lunch,” says John Urh, product manager for chromatography

Another player in this arena, QIAGEN’s BioRobot Protein LS System, has the capacity for the parallel purification of up to 24 large-scale cultures in less than three hours, and PerkinElmer’s JANUS system can be configured for large-scale purification protocols.

New reagents and instruments have allowed researchers to set up pipelines for high-throughput protein production of a scale and capacity that match the needs of their individual labs. As a result of these advances, hundreds of proteins from different organisms have been purified and their structures and functions determined. But keeping abreast of this rapidly changing field will require ongoing innovation. “Several years ago everyone was talking about genomics, now the focus is on proteomics. More and more we are seeing research moving toward specific proteins,” says PerkinElmer’s Hall. “The challenge is trying to keep up with the technologies.”

Laura Bonetta is a freelance writer based in the Washington DC area.

TARGET IDENTIFICATION

High-throughput protein-purification pipelines are becoming part of many proteomics efforts. But how to choose the targets to put through the pipeline? For those who want to focus on targets relevant to a disease-related process, it is important to catalogue where, when and to what extent a protein is expressed. Currently the main method for determining the protein complement of a given cell or tissue uses two-dimensional (2D) gel electrophoresis or liquid chromatography to resolve the proteins in the sample, and mass spectrometry to then identify the individual proteins. A problem with this approach is that it is difficult to identify rare proteins, because cell extracts are dominated by a few very abundant proteins.

Beckman Coulter has developed a number of innovations for protein fractionation to address

this problem. “We first selectively remove the most abundant proteins from a sample,” says Jerry Feitelson, marketing manager for the company. The ProteomeLab IgY-12 partitioning kits selectively remove the 12 most abundant proteins — albumins and immunoglobulins — from human/primate serum or plasma. These proteins make up 95% of the total protein expressed by cells. So removing them increases the chances of identifying proteins that are not highly expressed and are probably more interesting to researchers. The technology uses antibodies bound to inert beads, which are then packed in liquid-chromatography columns or, for smaller samples, spin columns. “The advantages of our reagents are the greater capacity, increased specificity and the ability to bind across species,” says Feitelson. The enriched material can then be collected and further fractionated

with the Beckman Coulter ProteomeLab PF 2D protein fractionation system. The instrument divides proteins in two dimensions: first, in a 2D gel, samples are separated by isoelectric focusing and size, and then automatically injected into a liquid-chromatography column. The process is fully automated. “One run takes ten hours, so you can do two in one day,” says Feitelson. Detailed protein maps can be constructed for easy comparison between two samples using the Proteome Lab software suite. Interesting proteins can then be further analysed by cutting out the corresponding band, for example, and putting it into a mass spectrometer.

Pall Corporation has also introduced the Enchant Life Sciences kits for albumin depletion and immunoglobulin G purification.

L.B.



H. DEWALD/BECKMAN COULTER