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 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-

PALL CORPORATION

Splitter: Pall's AcroPrep 96 Filter Plate.

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.