Express yourself

Stephen Chambers, Vertex Pharmaceuticals, 130 Waverly Street, Cambridge, MA 02139-4242, USA; e-mail: Stephen_Chambers@vpharm.com

The 7th annual *Protein Expression* meeting (12–13 January 2004, San Diego, USA), which was organized by the Cambridge Healthtech Institute (http://www.healthtech.com) as part of its Protein Information Week, brought together over 200 researchers from academia and the biopharmaceutical industry to address the major issues facing protein expression.

Bill Studier of Brookhaven National Laboratories (http://www.bnl.gov/world) presented the keynote address on his now ubiquitous T7 promoter system, which is used for protein expression in *Escherichia coli* and formed the basis for the development of the universally used pET system (Novagen; http://www.novagen.com). Almost twenty years after its introduction, the T7 promoter system is still the system of choice for the majority of *E. coli* expression. Studier also discussed how the T7 promoter system has evolved, describing the development of an auto-induction media that was specifically tailored to the T7 promoter and was designed to meet the demands of high-throughput expression.

Much of the conference was devoted to the problem of protein production for crystallography, with sessions discussing protein expression for structural determination and difficult to express proteins.

Protein expression for structural genomics

Protein expression has a direct impact on the successful outcome of crystallography. The overexpression of proteins greatly facilitates purification and subsequent protein production for crystallography. Conversely, poor or insoluble expression will impede the purification of a homogeneous protein sample, which is necessary for crystallography. Although soluble expression alone does not guarantee crystallization, no or low expression, and thus the probable failure to obtain pure protein, is often cited as the main reason that proteins are not crystallized. In the postgenomic era, the protein production bottleneck in the crystallographic determination of protein structure has been aggravated by the abundance of genomic information that is available – there is no shortage of proteins to be expressed or structures to be determined.

Nowhere is this more clearly illustrated than in structural genomics, where whole genomes are being expressed to provide structural information that establishes function and identifies novel folds. Protein expression on this scale requires miniaturization and automation to achieve the levels of throughput required. At the forefront of such efforts is the Protein Structure Initiative (PSI; http://www.nigms.nih.gov/psi) research program that is funded by the National Institute of General Medical Sciences (NIGMS; http://www.nigms.nih.gov). Several researchers from the PSI consortium described high-throughput processes for cloning, expression and purification of protein for structural determination purposes.

Ming Luo of the University of Alabama at Birmingham (http://sgce.cbse.uab.edu) and Rosalind Kim of Lawrence Berkley National Laboratories (http://www.lbl.gov) described their approaches to the *E. coli* expression, purification and crystallization of proteins from the *Caenorhabditis elegans* and *Mycobacterium tuberculosis* genomes, respectively. Although the combination of miniaturization, automation and process design appears to have addressed the throughput problem, all of the PSI centers seem to be falling short in terms of actual output [1].

The greatest hindrance in the transition from gene to crystal is seen as the inability to express soluble protein. Scott Lesley of the Genomics Institute of the Novartis Research Foundation (http://web.gnf.org/index.shtml) gave an impressive account of the high-throughput *E. coli* expression of the *Thermotoga maritima* genome. Practising what was described as a multi-tiered approach, this process introduces a focused effort and methods for salvaging difficult to express proteins, including refolding and alternative expression systems, which directly addresses the problem of low levels of expression. One of the most promising salvage techniques presented involved using deuterium-exchange mass spectrometry to identify regions of protein disorder, which could potentially interfere with crystallization.

Alternative expression systems

Dmitry Vinarov of the University of Wisconsin at Madison (http://uwstructuralgenomics.org) discussed the use of a high-throughput wheat germ cell-free expression system for structural genomics. Wheat germ cell-free protein synthesis, which was developed at Ehime University (Japan; http://www.ehime-u.ac.jp), has the advantage of uncoupled transcription–translation, producing slower and more favorable protein folding compared to other cell-free systems typically using coupled transcription–translation. Much of the

1359-6446/04/$ – see front matter PII: S1359-6446(04)03085-5

interest in this system is because of the ability to express soluble forms of some proteins that are not expressed in a soluble form in *E. coli*. In addition, the use of cell-free expression systems is often promoted because of the ease with which these systems produce isotopically labeled proteins for NMR structure determinations and seleno-methionine labeled proteins for phase determination in X-ray crystallography studies.

Unlike structural genomics, structure-based drug design focuses on the structural determination of therapeutically relevant proteins and thus demands high output with little opportunity for failure. In this different environment, the emphasis is on high output. Stephen Chambers of Vertex Pharmaceuticals (http://www.vrtx.com) used *E. coli* and insect-cell high-throughput expression systems to identify and deliver protein for structure-based drug design. Parallel expression of mammalian targets in *E. coli* and insect cells resulted in all proteins being expressed, with the majority produced as soluble proteins in insect cells.

Philip Laible of Argonne National Laboratories (http://www.bio.anl.gov/laible) provided an excellent account of the fundamental problems encountered with membrane protein expression (not enough membrane) and made a strong case for membrane expression in *Rhodobacter*, which have an abundance of membrane invaginations compared with *E. coli*. Several bacterial membrane proteins were expressed in this organism, with the expression levels of many of the proteins reaching concentrations of >10 mg/ml. These results suggest that *Rhodobacter* is a promising system for the production of membrane proteins.

Refolding

Given the relatively poor performance of *E. coli* expression systems in producing soluble protein, a number of talks described alternative strategies for refolding protein from the inclusion bodies that are often produced. Paul Ramage of the Novartis Institutes for Biomedical Research (http://www.nibr.novartis.com) gave one of the most interesting talks. Following an overview of the state-of-the-art technology available in this area, Ramage went on to describe a novel approach to the critical problem in refolding – monitoring protein 'foldedness'. By coupling field-flow fractionation to light scattering, Ramage was able to develop and screen a matrix of buffers to determine the best conditions for protein refolding. Xinli Lin of ProteomTech (http://www.proteomtech-inc.com) has a formidable track record of refolding insoluble proteins that have been produced in *E. coli*. Lin has adapted his past successes into an automated system for the high-throughput screening of protein-refolding conditions. Given that estimates of the levels of insoluble proteins expressed in *E. coli* ranged from 40% to 60%, each of the refolding strategies presented at the conference understandably generated considerable interest.

Protein fusions

As a perennial favorite of the conference, protein fusions are often cited as a universal remedy for insoluble protein expression. A number of talks presented data on the utility of expressing a protein as a fusion to improve solubility, thereby facilitating purification. Pascal Braun of the Institute of Proteomics, Harvard Medical School (http://www.hip.harvard.edu) described the development of a high-throughput protein purification method that uses a number of the more commonly used fusion-Tags, including six consecutive histidine residues (His), glutathione-S-transferase (GST), maltose-binding protein (MBP) and calmodulin-binding peptide (CBP). An assessment of the Tags (when used in *E. coli*) identified GST- and MBP-Tags to be superior to His- and CBP-Tags, with GST-Tags having a 50% success rate in purifying human proteins expressed in *E. coli*.

Tauseef Butt of LifeSensors (http://www.lifesensors.com) and Philip Bryan of the Center for Advanced Research in Biotechnology (http://www.carb.nist.gov) described successful fusions with small ubiquitin-like modifier (SUMO) and a subtilisin-proregion fusion, respectively.

Something of a clash of cultures (and uses) was evident with one speaker. Harry Meade of GTC Biotherapeutics (http://www.transgenics.com) thought he had gone back in time 20 years. Meade described the successful expression of a number of complex protein therapeutics, including vaccines, hormones, antibodies and soluble receptors, in transgenics. With no problem expressing an abundance of soluble protein in herds of lactating transgenic goats, Meade was somewhat bemused as he heard those involved in structural genomics wrestle with the difficulties of producing soluble protein in *E. coli*. This problem is one that the protein therapeutics arm of the biotechnology industry resolved long ago in favor of using eukaryotic systems, which are more adept at managing the complexities of refolding and producing soluble protein. The argument for using an *E. coli* expression system is that this system combines the robust expression and simplicity of protein production, making it the ideal expression system to deliver protein for crystallography. In addition, *E. coli* lacks the post-translational machinery of eukaryotic systems and, therefore, is more suited to producing a homogeneous protein sample. This is the key to the devotion of the crystallographer to using *E. coli* as an expression system. Furthermore, the

value given to homologous protein drives the use of *E. coli* for structure determination purposes.

Automation

The automation of protein expression enables researchers to cope with the proliferation of protein expression strategies available and a number of presentations described equipment that would perform this function. Grant Cameron of NextGen Sciences (http://www.nextgensciences.com) described the Expressionfactory™, an instrument that automates cloning, expression and purification within a single platform. When combined with a powerful proprietary information management system, different combinations of expression vectors and hosts can be used to express protein, thus enabling the parallel exploration of different growth and purification strategies. Nina Forsberg of Amersham Biosciences (http://www.amershambiosciences.com) presented recent developments in the AKTA-3D purification system, which is capable of automating affinity chromatography, gel-filtration and ion-exchange chromatography. The AKTA-3D system can automatically purify six different His- or GST-tagged proteins to >95% purity, with yields of 50 mg. Throughput can be maximized by using four AKTA modules in parallel to enable the purification of up to 16 proteins in a single run.

Concluding remarks

The manufacture of recombinant protein therapeutics and the production of target proteins for drug screening have dominated previous protein expression conferences but, with the development of structural genomics and proteomics, that emphasis is changing. The challenge now of providing protein content for structural studies and protein interactions has altered the character of protein expression and, as a consequence, the conference. This change is reflected in the content of many of the talks, which now encompass employing automation in the protein expression process, thus dictating more universal and generic strategies for expression, in comparison to the highly customized strategies that were previously used.

Although the PSI has been criticized for not immediately delivering the wealth of new structures once thought possible, it has made an impact on the development of new technologies for protein expression, production and analysis. Just as the Human Genome project aided the development of DNA technology, the PSI is advancing the technologies that underpin structural genomics, and in particular protein expression.

Reference

1 Service, R.F. (2002) Tapping DNA for structures produces a trickle. *Science* 298, 948–950